ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF BUSINESS

DEPARTMENT OF MANAGEMENT SCIENCE AND TECHNOLOGY

ELTRUN: THE E-BUSINESS RESEARCH CENTER

Ph.D. Thesis

Indoor localization systems for retail stores: An Artificial Intelligence Location Analytics Approach

by

Vasilis Stavrou

A thesis submitted for the degree of Doctor of Philosophy (Ph. D.)

Thesis supervisor:

Assoc. Prof. Katerina Pramatari

Athens,

November 2019

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ELTRUN: ΕΡΓΑΣΤΗΡΙΟ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΜΠΟΡΙΟΥ ΚΑΙ ΕΠΙΧΕΙΡΕΙΝ

Διδακτορική Διατριβή

Συστήματα Εντοπισμού Θέσης σε Καταστήματα Λιανεμπορίου: Μια Προσέγγιση Τεχνητής Νοημοσύνης για Χωρική Αναλυτική

του

Βασίλειου Σταύρου

Η παρούσα διδακτορική διατριβή υποβλήθηκε για την απονομή του τίτλου του Διδάκτορα

Επιβλέπουσα:

Αναπλ. Καθ. Κατερίνα Πραματάρη

Αθήνα
Νοέμβριος 2019

This page was intentionally left blank

Declaration

This research has been co-financed by various industry projects.

Publications

Journal Papers

1. Stavrou, V., Bardaki, C., Papakyriakopoulos, D., Pramatari, K. (2019), An Ensemble Filter for Indoor Positioning in a Retail Store Using Bluetooth Low Energy Beacons, Sensors 19(20). doi: 10.3390/s19204550
2. Triantafyllou S., Koutsokera L., Stavrou V., Griva A., (2018), Enrich customer experience and support decision making using IoT technologies in a grocery retail store, *Astrolavos, Scientific Journal of New Technologies (Hellenic Mathematical Society),* 28, 60-71.

Papers in Non-Refereed Conferences and Workshops

1. Batziakoudi K., Griva A., Stavrou V., Pramatari K., (2019), The value of collaborative analytics in contemporary retail. Evidence from a real-life case study, *Proceedings of 15th Student Conference of Management Science and Technology*, 14 May 2019, Athens, Greece.
2. Triantafyllou, S., Koutsokera, L., Stavrou V., Griva, A., (2017), Enhance shopping experience and support decision making leveraging BLE beacons in a grocery retail store, *Proceedings of 14th Student Conference of Management Science and Technology*, 27 April 2017, Athens, Greece.
3. Kalaidopoulou, K., Koutsokera L., Stavrou V., Griva, A., (2016), Investigating shopping visits patterns across different store types: The case of a grocery retail chain, *Proceedings of 13th Student Conference of Management Science and Technology*, 12 May 2016, Athens, Greece.

Working Papers and Work under Review

1. Stavrou V., Papakiriakopoulos D., Pramatari, K., Doukidis G. "Indoor location analytics for marketing intelligence: Combining motion, purchase and shopper data to extract in-store behavioral insights", Journal of Retailing and Consumer Services (under review)

2. Stavrou, V., Bardaki, C., Griva A., Pramatari, K. "Deploying a Retail Location-based Coupon Recommendation Application: Guidelines and Lessons Learnt". Computers in Industry (working paper)
3. Papakyriakopoulos D., Griva, A., Stavrou V. "Graph mining to extract shopping missions: A case from DIY industry". (working paper)

Invited Articles in Business Magazines

1. Gavalas, L., Griva, A., Pramatari, K., Stavrou, V. (2018), What can be really achieved by the cooperation of a Research Center with an Industry Partner?, Self Service, 481 (in Greek)

Research related Awards

- 2017: Self Service Excellence Awards, Best Practices in the FMCG - for the project AB ShopMate – Beacon enabled store, with cooperation to AB Vassilopoulos, Ahold Delhaize Group.

ACKNOWLEDGEMENTS

Ένα μεγάλο ταξίδι φτάνει αισίως στο τέλος του και κοιτώντας πίσω, το κομμάτι των ευχαριστιών είναι ίσως αυτό που με δυσκόλεψε περισσότερο να γράψω, κυρίως γιατί προσπάθησα να μην ξεχάσω κάποιον από τους συνταξιδιώτες αυτής της διαδρομής. Το ταξίδι του διδακτορικού είναι μια συνεχής μεταβολή είτε προς τα πάνω, που αρχίζεις να πιστεύεις στην έρευνα σου, κι άλλοτε προς τα κάτω, που αμφιβάλεις για τα πάντα. Εκεί είναι που ο ερευνητής χρειάζεται την μεγαλύτερη υποστήριξη και καθοδήγηση. Ευχαριστώ λοιπόν όλους αυτούς για την υπομονή, την υποστήριξη και την καθοδήγηση που έδειξαν με τον τρόπο τους κατά τη διάρκεια αυτής της διαδρομής. Άλλωστε όλες αυτές οι εμπειρίες είναι που σε εξελίσσουν και σε διαμορφώνουν πρώτα ως άνθρωπο και μετά ως επαγγελματία κι ερευνητή.

Θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα μου, Αναπληρώτρια Καθηγήτρια Κατερίνα Πραματάρη για την εμπιστοσύνη και τη στήριξη κατά τη διάρκεια αυτής της διαδρομής. Η θετική ενέργεια και η αισιοδοξία που διακατέχουν την Κατερίνα είναι δύο βασικά εφόδια απαραίτητα για να πετυχαίνεις τους στόχους σου. Ιδιαίτερα σημαντικό σε όλα αυτά είναι να μην ξεχνάς πως είναι να είσαι άνθρωπος. Κατερίνα σε ευχαριστώ που πάνω από όλα μου έδειξες τι σημαίνει να είσαι πρώτα Άνθρωπος και μετά όλα τα υπόλοιπα.

Έχω την τιμή και τη χαρά να έχω στην τριμελή επιτροπή μου τον Καθηγητή Γεώργιο Δουκίδη, τον οποίο θα ήθελα να ευχαριστήσω θερμά για την καθοδήγηση και τις προτάσεις του κατά τη διάρκεια της εκπόνησης αυτής της έρευνας. Ο κύριος Δουκίδης είναι ένας διορατικός άνθρωπος που έχει την ικανότητα να αφουγκράζεται τόσο την έρευνα όσο και την αγορά, ώστε να παρέχει κατευθύνσεις προς καινοτόμα και ενδιαφέροντα θέματα. Κύριε

Δουκίδη σας ευχαριστώ θερμά, καθώς μέσα από την αλληλεπίδραση και τις συζητήσεις μαζί σας, πέραν της διαμόρφωσης της διδακτορικής έρευνας, διεύρυνα τόσο τους ορίζοντες μου όσο και τον τρόπο σκέψης μου πάνω σε διάφορα θέματα που θεωρώ πως με εξέλιξαν σαν άνθρωπο.

Κλείνοντας τις ευχαριστίες προς την τριμελή μου επιτροπή, θα ήθελα να ευχαριστήσω θερμά τον Λέκτορα Δρ. Δημήτρη Παπακυριακόπουλο για την καθοδήγηση και τις συμβουλές του κατά τη διάρκεια εκπόνησης του διδακτορικού. Ανατρέχοντας τη διαδρομή του διδακτορικού θυμάμαι τις συζητήσεις μας, το mentoring που μου έκανες πάνω σε θέματα στα οποία χρειαζόμουν ενίσχυση και την άμεση συνεργασία σε ερευνητικά. Ο Δημήτρης αποτελεί έναν μέντορα τόσο όσο άνθρωπος όσο και ως ακαδημαϊκός. Δημήτρη σε ευχαριστώ θερμά για όλα.

Δεν θα μπορούσα να παραλείψω τις ειδικές ευχαριστίες προς την Επίκουρη Καθηγήτρια Κλεοπάτρα Μπαρδάκη για τη συνδρομή της σε αυτό το ταξίδι. Η Κλεοπάτρα είναι ο αφανής ήρωας αυτής της διδακτορικής διατριβής, καθώς ήτανε πάντα εκεί και με καίριες παρεμβάσεις συνέβαλε στην ολοκλήρωση αυτής της έρευνας. Κλεοπάτρα σε ευχαριστώ θερμά για την καθοδήγηση σου και την καλή θέληση που έδειχνες πάντα για να μας βοηθήσεις όλους μας.

Συνεχίζοντας, θα ήθελα να ευχαριστήσω όλα τα μέλη του Eltrun που με τον τρόπο τους συνείσφεραν σε αυτή τη διδακτορική διατριβή. Ειδική αναφορά θα ήθελα να κάνω στα πιο στενά μέλη της ερευνητικής ομάδας που δουλέψαμε όλο αυτόν τον καιρό, την Αναστασία και τον Λέων. Κατερίνα και Μαριάννα, οφείλω να ομολογήσω πως θα μου λείψουν οι βραδινές συζητήσεις μας για το διδακτορικό.

Θερμές ευχαριστίες στην οικογένεια μου για την ανιδιοτελή της συνδρομή καθώς είναι πάντα εκεί για να στηρίξει και να ακούσει. Οι συμβουλές ήτανε πάντα χρήσιμες και μου έδιναν πάντα κουράγιο στο να συνεχίζω και να πετυχαίνω τους στόχους μου.

Κλείνοντας, θερμές ευχαριστίες στους φίλους μου (κοντινοί και μακρινοί γεωγραφικά) που ήτανε πάντα εκεί όταν χρειαζόταν για να ακούσουν προβληματισμούς και να δώσουν συμβουλές. Και να μην ξεχνάμε ότι όταν η ζωή σε τραβάει πίσω είναι για να σε πετάξει ακόμα πιο μπροστά.

Με εκτίμηση,

Βασίλης Σταύρου

ΕΠΙΤΕΛΙΚΗ ΣΥΝΟΨΗ

Η συμπεριφορά του καταναλωτή μέσα στο κατάστημα αποτελεί ερευνητική περιοχή ενδιαφέροντος για περισσότερα από 60 χρόνια και έχει ως στόχο να εντοπίσει διάφορα μοτίβα που μπορούν να προσφέρουν αξία (Sorensen et al., 2017). Τα τελευταία χρόνια, η συγκεκριμένη ερευνητική περιοχή έχει λάβει ιδιαίτερη προσοχή όχι μόνο από τους διευθυντές των καταστημάτων αλλά και από ερευνητές και επαγγελματίες του χώρου (Ferracuti et al., 2019, Sorensen et al., 2017). Η κατανόηση της συμπεριφοράς του καταναλωτή μπορεί να οδηγήσει σε σημαντικές πληροφορίες για τους λιανέμπορους και να ενισχύσει την εμπειρία των καταναλωτών μέσα στο κατάστημα. Παρόλο που η συμπεριφορά των αγοραστών είναι πιο εύκολο να εξεταστεί στα ψηφιακά περιβάλλοντα, τα τελευταία χρόνια τα φυσικά καταστήματα έχουν αρχίσει να λαμβάνουν ιδιαίτερη προσοχή λόγω της εξέλιξης της τεχνολογίας και των σύγχρονων εργαλείων που παρέχονται για την εξέταση και ανάλυση της συμπεριφοράς του καταναλωτή στο κατάστημα, κάνοντας χρήση Χωρικής Αναλυτικής (Location Analytics).

Τα Μεγάλα Δεδομένα (Big Data) και η Επιχειρηματική Αναλυτική (Business Analytics) αποτέλεσαν αρωγό στη δημιουργία νέων δυνατοτήτων και γνώσης στο λιανεμπόριο, χάρη στην αυξημένη ποιότητα των δεδομένων και τις δυνατότητες των εφαρμογών που αναπτύσσονται. Αυτές οι δυνατότητες πηγάζουν από τη χρήση και το συνδυασμό νέων πηγών δεδομένων, την εφαρμογή στατιστικών εργαλείων, την αξιοποίηση γνώσης πεδίου και τον εμπλουτισμό με γνώσεις από τη θεωρία (Bradlow et al., 2017). Οι ακαδημαϊκοί και ειδικοί από την αγορά αναγνωρίζουν το γεγονός πως η προσθήκη μιας επιπλέον διάστασης στα επιχειρηματικά δεδομένα μπορεί να ενισχύσει σημαντικά την ποιότητα της εξαγόμενης γνώσης (Pick et al., 2017). Η διάσταση αυτή αναφέρεται στη θέση/τοποθεσία. Η τοποθεσία παρέχει μια επιπλέον προοπτική στην Επιχειρηματική Αναλυτική (Business Analytics) και μπορεί να διευκολύνει τους υπεύθυνους για τη λήψη αποφάσεων τόσο στο να ανακαλύψουν νέα γνώση όσο και στο να αποκτήσουν ένα μέσο αποτελεσματικής επικοινωνίας όσον αφορά την οπτικοποίηση της πληροφορίας.

Η Χωρική Αναλυτική (Location Analytics) αναφέρεται στην εξειδικευμένη ανάλυση χωρικών δεδομένων προκειμένου να κατανοηθούν μοτίβα και σχέσεις σε φαινόμενα που σχετίζονται με τη θέση (Pick et al., 2017; Senior et al., 2007). Οι λιανεμπορικές επιχειρήσεις μπορούν να επωφεληθούν από τη γνώση σχετικά με τη θέση των πελατών μέσα στο κατάστημα, όπως για παράδειγμα οι περιοχές που τείνουν περισσότερο να περιηγούνται, προκειμένου να οργανώσουν καλύτερα τις δραστηριότητες λειτουργίας τους (Mou et al., 2018) και να βελτιώσουν την εμπειρία των πελατών μέσα στο κατάστημα (Yaeli et al., 2014).

Η έλευση του Διαδικτύου των Πραγμάτων (Internet of Things - IoT) (Ali et al., 2015) διευκόλυνε τον εντοπισμό της θέσης του πελάτη στο κατάστημα. Με τον τρόπο αυτό, η κατανόηση της συμπεριφοράς του καταναλωτή στα καταστήματα λιανεμπορίου καθίσταται όλο και πιο ουσιώδης και εφικτή από ποτέ. Για το σκοπό αυτό, το λιανεμπόριο αξιοποιεί την εξέλιξη της τεχνολογίας και του Διαδικτύου των Πραγματών (IoT), μαζί με την Επιχειρηματική Ευφυία (Business Intelligence), προκειμένου να δημιουργήσει αξία για τους οργανισμούς εστιάζοντας σε σημαντικές πληροφορίες σχετικά με τη θέση και κίνηση των πελατών μέσα στο κατάστημα (Fink et al., 2017).

Τυπικές πηγές που είναι ικανές να διευκολύνουν την ανίχνευση τοποθεσίας στο κατάστημα του αγοραστή αποτελούν τα κινητά τηλέφωνα μέσω των τεχνολογιών Wi-Fi, Bluetooth, RFID και οι κάμερες (Rai et al., 2011). Τα χωροχρονικά δεδομένα που προκύπτουν από αυτές τις τεχνολογίες συνήθως μεταφράζονται σε διαστάσεις όπως ο χρόνος, η τοποθεσία και το αντικείμενο παρακολούθησης. Ο χρόνος υποδεικνύει τη χρονική σήμανση των δεδομένων, η τοποθεσία αναφέρεται στον χώρο όπου ανιχνεύεται το αντικείμενο και ο όρος "αντικείμενο παρακολούθησης" αναφέρεται στον χρήστη ή το προϊόν του οποίου η κίνηση καταγράφεται. Προκειμένου να επιτευχθεί η καταγραφή της κίνησης του αντικειμένου, συνήθως απαιτείται ένας μηχανισμός καταγραφής θέσης. Για παράδειγμα, η κίνηση των χρηστών μπορεί να καταγραφεί είτε χρησιμοποιώντας μια κινητή συσκευή είτε μια εφαρμογή για κινητά τηλέφωνα που επικοινωνεί με την ασύρματη υποδομή. Αυτά τα δεδομένα μπορούν να επεξεργαστούν περαιτέρω, να μετασχηματιστούν και να χρησιμοποιηθούν για τη δημιουργία αξίας χρησιμοποιώντας τεχνικές ανάλυσης δεδομένων θέσης.

Οι νέες αυτές τεχνολογίες έχουν αλλάξει δραματικά το χώρο του λιανεμπορίου. Οι λιανέμποροι αρχικά αντιμετώπισαν την απειλή των διαδικτυακών ανταγωνιστών οι οποίοι χωρίς ιδιαίτερο κόστος μπόρεσαν να εκμεταλλευτούν τις πληροφορίες των πελατών τους σχετικά με τις αγορές που έχουν πραγματοποιήσει καθώς και τις προτιμήσεις τους, προκειμένου να τους στοχεύσουν καλύτερα με κατάλληλες προσφορές. Στη συνέχεια, έχοντας οικοδομήσει τις δικές τους δυνατότητες στο διαδίκτυο, οι λιανέμποροι πρέπει να αναπτύξουν τις δεξιότητες τους στο πανκαναλικό μάρκετινγκ (omni-channel marketing) ώστε να προσφέρουν τις προσεγγίσεις του διαδικτυακού καταστήματος μέσα στο φυσικό κατάστημα. Πολλές νέες τεχνολογίες έχουν εμφανιστεί τα τελευταία χρόνια, όπως οι Αναλυτική Βίντεο (Video Analytics), η Αναλυτική μέσω Ασύρματου δικτύου (Wi-Fi Analytics) καθώς και τα Bluetooth Low Energy Beacons προκειμένου να βοηθήσουν τους λιανέμπορους να βελτιώσουν τόσο την εμπειρία χρήστη των φυσικών τους καταστημάτων όσο και την κερδοφορία τους. Ωστόσο, λόγω του ότι οι τεχνολογίες αυτές δεν έχουν εφαρμοστεί πλήρως και λόγω των ερωτημάτων που εγείρονται σχετικά με τη χρήση τους, οι λιανέμποροι τις δοκιμάζουν πιλοτικά σε πρώτο στάδιο αντί για πλήρη υλοποίηση κι εφαρμογή τους στα καταστήματα τους, ώστε να διερευνήσουν τις δυνατότητές τους.

Όπως η Αναλυτική Διαδικτύου (Web Analytics) είναι ένα βασικό εργαλείο στον παγκόσμιο ιστό, η Χωρική Αναλυτική (Location Analytics) θα αποτελέσει απαραίτητο στοιχείο για το σχεδιασμό, τη διαχείριση και τη μέτρηση των εμπειριών μέσα στα φυσικά καταστήματα (Harvard Business Review, 2015). Η Χωρική Αναλυτική (Location Analytics) μπορεί να προσφέρει μεγάλο αντίκτυπο στην αποτελεσματική σχεδίαση διάταξης καταστημάτων, τις αποτελεσματικές ενέργειες μάρκετινγκ, τις λειτουργίες και τον προγραμματισμό της στρατηγικής με βάση τα δεδομένα τοποθεσίας που συλλέγονται από τα φυσικά καταστήματα λιανεμπορίου μέσω έξυπνων τεχνολογιών.

Με βάση την εξέλιξη της τεχνολογίας και τις δυνατότητες που αναδεικνύονται μέσω αυτής, οι λιανέμποροι πρέπει να σκεφτούν νέες ιδέες ώστε να διατηρήσουν τους πελάτες τους, καθώς ένας πελάτης με υψηλότερες προσδοκίες είναι πιθανότερο να μεταπηδήσει σε κάποιον ανταγωνιστή που μπορεί να του προσφέρει καλύτερη εμπειρία. Αρκετοί εξειδικευμένοι λιανέμποροι χρησιμοποιούν Αναλυτική Βίντεο (Video Analytics) προκειμένου να μελετήσουν τις διαδρομές που ακολουθούν και τις ενέργειες που προβαίνουν οι πελάτες τους ώστε να σχεδιάσουν ανάλογα το κατάστημά τους. Παράλληλα, οι μεγάλοι λιανέμποροι διαθέτουν εφαρμογές για την πλοήγηση των πελατών τους μέσα στο κατάστημα, βοηθώντας τους με αυτό τον τρόπο να εντοπίσουν αυτό που αναζητούν.

Παρά τη δημοτικότητα της Χωρικής Αναλυτικής (Location Analytics) και των υπηρεσιών βάσει τοποθεσίας (Location-based Services), και οι δύο περιπτώσεις αντιμετωπίζουν ένα κοινό πρόβλημα. Αυτό το πρόβλημα είναι η ακρίβεια του μηχανισμού εντοπισμού που χρησιμοποιούν. Προκειμένου να λειτουργούν αποτελεσματικά, αυτοί οι μηχανισμοί απαιτούν τον εντοπισμό θέσης του χρήστη με μεγάλη ακρίβεια. Το ζήτημα της αποτελεσματικής ανίχνευσης της θέσης του χρήστη συναντάται τόσο στην βιβλιογραφία όσο και ως ζήτημα που αντιμετωπίζουν οι επιχειρήσεις όταν προσπαθούν να αναπτύξουν τέτοιους μηχανισμούς σε πραγματικές περιπτώσεις, κι όχι σε πειραματικά περιβάλλοντα μόνο.

Από την ακαδημαϊκή οπτική, το ζήτημα του εντοπισμού θέσης, ειδικά σε εσωτερικούς χώρους, αποτελεί μια αναδυόμενη πρόκληση. Παρόλο που δεν είναι τόσο δύσκολο και πολύπλοκο να εντοπιστεί η ακριβής θέση ενός κινούμενου αντικειμένου έξω από ένα κτίριο (π.χ. χρησιμοποιώντας το GPS και τεχνικές geofencing (Kriz et al, 2016)), αυτό αρχίζει να γίνεται ένα αρκετά δύσκολο εγχείρημα όταν υλοποιείται σε εσωτερικούς χώρους. Τα καταστήματα λιανεμπορίου και τα εμπορικά κέντρα γενικά αποτελούν ένα γόνιμο έδαφος για την ανάπτυξη συστημάτων εντοπισμού θέσης και τα αντίστοιχα ερευνητικά έργα ασχολούνται κυρίως με δύο σημαντικά ζητήματα: (α) την υποδομή εντοπισμού θέσης που αποτελείται από μια τεχνική εντοπισμού θέσης και μια ασύρματη τεχνολογία (π.χ., Jin et al., 2013; He et al., 2016)) και (β) τις παρεχόμενες υπηρεσίες που προσφέρονται από αυτή την υποδομή.

Το ζήτημα του εντοπισμού θέσης σε εσωτερικά περιβάλλοντα αντιμετωπίζεται από τη βιβλιογραφία ως πολύ απαιτητικό και η επίτευξη υψηλής ακρίβειας εντοπισμού είναι κοινός στόχος, όπως αναγνωρίζεται από την ακαδημαϊκή κοινότητα (He et al., 2016; Shin et al., 2015). Ειδικότερα, Οι Lymberopoulos et al. (2015) υποστηρίζουν ότι το πρόβλημα της του εντοπισμού της θέσης σε εσωτερικά περιβάλλοντα εξακολουθεί να παραμένει άλυτο και αναδεικνύουν τη σημασία της χρήσης ρεαλιστικών προσεγγίσεων που αντισταθμίζουν την επιθυμητή ακρίβεια εντοπισμού με παράλληλα χαμηλό κόστος υλοποίησης.

Από την επιχειρηματική οπτική, η έρευνα έχει δείξει ότι παρά την πρόοδο της τεχνολογίας, απαιτείται αυξημένο κόστος, επιπρόσθετο υλικό και πολύπλοκες εγκαταστάσεις για την αποτελεσματική επίτευξη εντοπισμού θέσης σε εσωτερικά περιβάλλοντα. Τα πληροφοριακά συστήματα διαχείρισης πληροφορίας (Management Information Systems) μαζί με τα δεδομένα της επιχείρησης περιέχουν πληροφορίες θέσης. Ωστόσο, εξακολουθούν να μην εκμεταλλεύονται αυτές τις πληροφορίες για να εξάγουν γνώση. Επιπλέον, πραγματοποιώντας ένα ημι-δομημένο focus group, μαζί με επιπλέον συνεντεύξεις με ενδιαφερόμενους και χρήστες συστημάτων εντοπισμού θέσης σε εσωτερικά περιβάλλοντα, αναδείχθηκε το γεγονός ότι ο αποτελεσματικός εντοπισμός θέσης αποτελεί μια πρόκληση που πρέπει να αντιμετωπιστεί.

Συνεπώς, και οι δύο προσεγγίσεις (ακαδημαϊκή και επιχειρηματική) καταλήγουν στο γεγονός ότι απαιτείται αποτελεσματικός εντοπισμός θέσης σε εσωτερικά περιβάλλοντα προκειμένου να εξαχθεί αξιόπιστη γνώση Χωρικής Αναλυτικής (Location Analytics) και να σχεδιαστούν αποδοτικές υπηρεσίες βάσει θέσης του καταναλωτή. Στο πλαίσιο αυτής της διατριβής προτείνεται ένα σύστημα και μια προσέγγιση μηχανικής μάθησης για τον εντοπισμό θέσης σε εσωτερικά περιβάλλοντα λιανεμπορίου.

Με βάση τις παραπάνω παρατηρήσεις εγείρονται διάφορα ερωτήματα όπως τα ακόλουθα:

- Ποιες είναι οι κατάλληλες τεχνολογίες για εντοπισμό θέσης σε εσωτερικά περιβάλλοντα;
- Πόσο αποτελεσματική είναι η τεχνολογία Bluetooth Low Energy Beacons για εντοπισμό θέσης σε εσωτερικά περιβάλλοντα;
- Πόσο αποτελεσματικό είναι το Wi-Fi για εντοπισμό θέσης σε εσωτερικά περιβάλλοντα;
- Πώς να σχεδιαστούν και να αναπτυχθούν συστήματα εντοπισμού θέσης σε εσωτερικά περιβάλλοντα;

Για το σκοπό αυτό, στόχος της παρούσας διατριβής είναι να απαντήσει στα ακόλουθα ερευνητικά ερωτήματα:

- RQ1. Πώς μπορούμε να επιτύχουμε αποτελεσματικό εντοπισμό θέσης σε εσωτερικά περιβάλλοντα κάνοντας χρήση χωροχρονικών δεδομένων;
- RQ2. Πώς να σχεδιαστούν και να αναπτυχθούν συστήματα εντοπισμού θέσης σε εσωτερικά περιβάλλοντα;

Για να αντιμετωπιστούν αυτά τα ερευνητικά ερωτήματα, (α) υιοθετούμε ως μεθοδολογική προσέγγιση το μοντέλο Design Science (Hevner et al., 2004), (β) αναπτύσσουμε ένα σύστημα εντοπισμού θέσης για εσωτερικά περιβάλλοντα και (γ) υιοθετούμε μια προσέγγιση μηχανικής μάθησης που εκτελεί εντοπισμό θέσης ως αποτέλεσμα αυτής της μελέτης. Συνοπτικά, σχεδιάζουμε και αναπτύσσουμε ένα σύστημα που παράγει και επεξεργάζεται δεδομένα σήματος και αναπτύσσουμε μια προσέγγιση μηχανικής μάθησης για τον εντοπισμό θέσης που μπορεί να εφαρμοστεί σε χωροχρονικά δεδομένα από συσκευές Διαδικτύου των Πραγμάτων (IoT). Αξιολογούμε δύο διαφορετικές ασύρματες τεχνολογίες ((α) Wi-Fi και (β) Bluetooth Low Energy Beacons) και εφαρμόζουμε τεχνικές τεχνητής νοημοσύνης για την αντιμετώπιση των ερευνητικών ερωτημάτων. Επιπλέον, για την αντιμετώπιση των ερευνητικών ζητημάτων προτείνουμε ένα artifact συστήματος το οποίο είναι υπεύθυνο για τη δημιουργία, την καταγραφή και την επεξεργασία των δεδομένων για τον εντοπισμό θέσης σε εσωτερικά περιβάλλοντα. Το αποτέλεσμα της προτεινόμενης προσέγγισης είναι η θέση του χρήστη της ασύρματης υποδομής μέσα στο κατάστημα.

Εφαρμόζουμε αυτήν την προσέγγιση σε δύο διαφορετικές περιπτώσεις. Η πρώτη περίπτωση αφορά την τεχνολογία Bluetooth Low Energy Beacons, ενώ η δεύτερη αφορά τεχνολογία Wi-Fi. Στη συνέχεια αξιολογούμε τα ευρήματα από κάθε περίπτωση χρησιμοποιώντας τεχνική αποτίμηση της επίδοσης της και επίσης εξετάζουμε την επιχειρηματική ερμηνεία των αποτελεσμάτων. Για το σκοπό αυτό, αξιοποιούμε την αξιολόγηση βάσει δεδομένων και την αξιολόγηση βάσει χρηστών, προκειμένου να αξιολογήσουμε τα αποτελέσματα της προσέγγισης εντοπισμού θέσης.

Τέλος, παραθέτουμε μια σειρά πρακτικές εφαρμογές με βάση τη διάσταση της θέσης, όπως αυτή προκύπτει από τα χωροχρονικά δεδομένα. Οι πρακτικές εφαρμογές διαχωρίζονται σε δύο μεγάλες κατηγορίες. Λειτουργίες Λιανεμπορίου (Retail Operations) και Πληροφορίες για αξιοποίηση

Μάρκετινγκ (Marketing Insights). Όσον αφορά το πρώτο, παρουσιάζουμε μια σειρά δεικτών επίδοσης (KPIs) με βάση τη θέση του καταναλωτή και τις πτυχές της Χωρικής Αναλυτικής (Location Analytics) που θα μπορούσαν να ενδιαφέρουν τους λιανέμπορους και τους επαγγελματίες. Επιπλέον, εξάγουμε μοτίβα συμπεριφοράς με βάση τη θέση του πελάτη και παρέχουμε τη γνώση που εξαγάγαμε και αφορά σχεδιαστές συστημάτων εντοπισμού θέσης για εσωτερικά περιβάλλοντα στο σύγχρονο λιανεμπόριο, καθώς και μια λεπτομερή περιγραφή των παραγόντων που πρέπει να λαμβάνουν υπόψη κατά το σχεδιασμό τέτοιων συστημάτων. Όσον αφορά τις πληροφορίες γι' αξιοποίηση μάρκετινγκ (Marketing Insights), πραγματοποιούμε κατάτμηση πελατών βάσει των τοποθεσιών που επισκέπτονται μέσα στο κατάστημα και παρέχουμε μια σειρά από δείκτες επίδοσης (KPIs) για περεταίρω αξιοποίηση από τους ειδικούς του χώρου.

Οι συνεισφορές αυτής της εργασίας συνοψίζονται στα ακόλουθα. Όσον αφορά την θεωρητική συνεισφορά, αυτή η διατριβή προτείνει ένα σύστημα και μια προσέγγιση μάθησης μηχανής που εκτελεί εντοπισμό θέσης σε εσωτερικά περιβάλλοντα. Συνεχίζει με τη διερεύνηση τεχνικών εντοπισμού θέσης σε περιβάλλοντα λιανεμπορίου, καθώς μέχρι στιγμής υπάρχει περιορισμένη βιβλιογραφία πάνω σε αυτή την περιοχή και εντοπίζει παράγοντες που επηρεάζουν προσεγγίσεις και συστήματα εντοπισμού θέσης. Η προτεινόμενη προσέγγιση επιτυγχάνει σφάλμα εντοπισμού θέσης μέχρι 2 μέτρα για την τεχνολογία BLE Beacons και 3 μέτρα για το Wi-Fi. Επιπλέον, η μελέτη των ensemble classifiers αναδεικνύει την αποδοτικότητά τους σε σύγκριση με τη χρήση μεμονωμένων classifiers, καθώς αντιμετωπίζουν καλύτερα τις προκλήσεις του περιβάλλοντος του λιανεμπορίου. Τέλος, αυτή η διατριβή προτείνει ένα καινοτόμο δείκτη απόδοσης (KPI) που εφαρμόζεται για την τεχνολογία BLE Beacons και αναφέρεται στην ταχύτητα του καταναλωτή.

ABSTRACT

Shopper in-store behavior has been an area of interest for more than 60 years, trying to uncover patterns of shopper behavior (Sorensen et al., 2017). Over the last few years it has received great attention not only from store managers but also from researchers and practitioners (Ferracuti et al., 2019; Sorensen et al., 2017). The understanding of shopper behavior can lead to essential insights for retailers and enrich the in-store shopper experience. Although, shopper behavior is easier to be examined in digital environments, the physical stores draw more attention over the years due to the evolution of technology and the modern tools provided for examining in-store shopper behavior, powered by location analytics.

Big Data and Business analytics have enabled new capabilities and insights in retailing thanks to the increase in data quality and application possibilities coming from a mix of new data sources, a smart application of statistical tools and domain knowledge combined with theoretical insights (Bradlow et al., 2017). Scholars and industry partners acknowledge the realization that by adding an additional dimension to business data can significantly enhance the quality of the extracted insights (Pick et al., 2017). The dimension refers to location. Location provides a new context in the perspective of Business Analytics and can facilitate decision makers to discover new insights and, also, a means of an effective communication in term of visualization.

Location analytics (LA) refers to the specialized spatial data analysis in order to understand spatial patterns and relationships in location referenced phenomena (Pick et al., 2017; Senior et al., 2007). Retail business can benefit from having insights regarding the in-store locations that their customers tend

to browse during their shopping trip in order to better organize their operation activities (Mou et al., 2018) and, also, improve customer experience (Yaeli et al., 2014).

The advent of IoT (Ali et al., 2015) has facilitated the detection of customer's in-store location. Thus, understanding customer behavior in indoor retail environments gets more and more essential and feasible than ever. To this end, the retail industry exploits the evolution of technology and IoT, along with Business Intelligence in order to create value for the organization, by focusing on the essential information of customer's in-store location (Fink et al., 2017).

Typical sources that are able to facilitate the detection of the in-store location of the shopper are mobile phones via Wi-Fi, Bluetooth, and RFID technology, and cameras (Rai et al., 2011). Spatiotemporal data occur from these technologies and typically they are translated to dimensions such as time, location, and tracking object. Time indicates the timestamp of the data; location refers to the place that the object is detected and with the term tracking object is implied the user or the tracked product. Tracking objects usually require a tracking mechanism; for example, users can be tracked either by utilizing a mobile device or a mobile application that communicates with the wireless infrastructure. These data can be further processes, transformed and be used to create value using location analytics techniques.

These new technologies have dramatically changed the retail landscape. Retailers initially faced the threat of online competitors without retail store costs who were able to leverage detailed customer shopping and preference information to better target promotions. Now, after building their own online capabilities, retailers need to master omni-channel marketing to sync the analytical, highly targeted approaches of online with the hands-on, experiential environment of in-store. Several new technologies such as video analytics, Wi-Fi analytics and beacons have emerged to help retailers optimize their bricks and mortar store experience and profitability, but there are still many questions and for the most part, retailers are piloting these solutions rather than committing to full rollouts.

Just as web analytics is an essential tool on the Web, location analytics will become a must-have for designing, managing, and measuring offline experiences (Harvard Business Review, 2015). Location analytics can have a great impact on efficient layout design, effective marketing actions, operations and strategy planning based on location data gathered from physical retail stores.

Based on this technology advancement and the emerging capabilities retailers have to come up with new ideas to hold on to their customers because a customer with higher expectation is more likely to shift to other competitors for a better experience. Specialty retailers use video analytics to study their customer trails and actions helping them design their store accordingly; while big retailers have even designed an app to navigate their customers inside the store helping them locate what they are looking for.

Despite the popularity of location analytics and location-based services, they both face a similar problem; i.e. the accuracy of the localization mechanism they use. In order to perform effectively, such mechanisms require detecting accurately the position of the user. The issue of effective detection of the user position is encountered both in academic literature and also a business issue when trying to deploy such mechanisms in real world cases.

From the academic perspective, the issue of user-positioning, especially in indoor environments, is an emerging challenge. Although it is not so difficult and complex to locate the exact position of a moving object outside a building (e.g. using GPS signals and geofencing techniques (Kriz et al, 2016)), this becomes a quite challenging task in indoor environments. Retail stores and shopping malls in general provide fertile ground for the development of indoor positioning systems and relevant research works address two major issues: (a) the localization infrastructure shaped by a positioning technique and a wireless technology (e.g. (Jin et al., 2013; He et al., 2016)) and (b) the provided services on the top of the infrastructure.

The indoor positioning issue appears in the relevant literature as a very challenging one and the achievement of high localization precision (accuracy) is a common objective shared by various scholars (He et al., 2016; Shin et al., 2015). Lymberopoulos et al. (2015) argue that the indoor location problem still remains unsolved and stress the importance of the employment of a realistic approach that would counterbalance the desired localization accuracy with low cost.

From the business perspective, research has indicated that despite the advancement of technology, increased costs, additional hardware and complex deployments are required in order to perform efficient indoor tracking. Management Information Systems contain location information along with the enterprise data, but still lack to exploit this information to extract knowledge. In addition, a semi-structured focus group, along with extra interviews with indoor positioning system stakeholders and experts indicated effective indoor positioning is a challenge that should be dealt.

Both approaches (academic and business) come down to one thing; effective indoor positioning is required in order to extract reliable location analytics and be able to design effective location-based services. In the context of this thesis a system artifact and machine learning indoor positioning approach are proposed that aim at extracting effectively the position of a user in an in-store retail environment.

Based on these observations various questions may arise:

- Which are the suitable technologies indoor positioning?
- How efficient are BLE beacons for indoor positioning?
- How efficient is Wi-Fi for indoor positioning?
- How to design and deploy indoor positioning systems?

To this end, in this dissertation we aim to answer the following research questions:

- RQ1. How can we perform efficient indoor positioning from spatiotemporal data?
- RQ2. How to develop indoor positioning systems?

To address these research questions, (a) we adopt as methodological backbone the design science paradigm (Hevner et al., 2004), (b) deploy an indoor positioning system and (c) consider a machine learning approach that performs indoor positioning as outcome of this study. In a nutshell, we design and deploy a system that generates, and process wireless signal data and we develop a machine learning indoor positioning approach that can be applied on spatiotemporal data from IoT access points. We assess two different wireless technologies (i.e. Wi-Fi and Bluetooth Low Energy Beacons) and apply Artificial Intelligence techniques to address the questions. In addition, to address the research questions we propose a system artifact which is responsible for generating, capturing and processing the data for indoor positioning. The outcome of the proposed approach is the indoor position of the user of the wireless infrastructure.

We apply this approach to two different cases. The first case involves Bluetooth Low Energy Beacons technology, while the second one involves Wi-Fi technology. We then evaluate the findings from each case using technical evaluation and examine the business interpretation of the outcomes. To this end, we utilize data-driven and user-based evaluation in order to assess the results of the indoor positioning approach.

Finally, we preset a series of practical implications based on the location dimension of spatiotemporal data. The practical implications are broken down to two major categories; i.e. Retail Operations and Marketing Insights. Regarding Retail Operations, we present a series of location based KPIs and present aspects of location analytics that could be of interest for retail managers and practitioners. In addition, we extract location-based behavioral patterns and provide the lessons learnt for indoor location-based systems designers in contemporary retail, along with a detailed description of the factors that designers should take into consideration when designing indoor location-based systems. Regarding Marketing insights, we perform location-based shopper segmentation and provide a series of marketing analytics KPIs for further exploitation.

The theoretical contributions of this thesis are summarized to the following. This thesis proposes a system artifact and a machine learning approach that performs indoor positioning. It moves forward by exploring localization and Artificial Intelligence techniques in retail environments, as there are few works that examine this area and then identifies factors that affect indoor positioning approaches and systems. The proposed approach achieves localization error of 2 meters for the BLE Beacons technology and 3 meters for Wi-Fi technology. In addition, the study of the ensemble classifiers indicates their effectiveness against single ones, as they deal with the challenges of retail environments. Finally, this thesis provided an innovative KPI for BLE Beacon infrastructures that refers to shopper speed.

TABLE OF CONTENTS

DECLARATION II

PUBLICATIONS II

ACKNOWLEDGEMENTS IV

ΕΠΙΤΕΛΙΚΗ ΣΥΝΟΨΗ VII

ABSTRACT XVII

TABLE OF CONTENTS XXIV

LIST OF FIGURES XXVI

LIST OF TABLES XXVIII

LIST OF ABBREVIATIONS XXIX

1 INTRODUCTION 30

1.1 RESEARCH MOTIVATION 30
1.2 RESEARCH OBJECTIVE & QUESTIONS 33
1.3 RESEARCH METHODOLOGY OVERVIEW 36
1.4 THESIS OUTLINE 40

2 RESEARCH BACKGROUND 45

2.1 LITERATURE REVIEW 46
2.1.1 Indoor Positioning 46
2.1.2 Artificial Intelligence for indoor positioning 70
2.1.3 Location Analytics 72
2.2 RESEARCH GAPS 81

3 RESEARCH METHODOLOGY 85

3.1 SYSTEM DESIGN AND IMPLEMENTATION 85
3.2 DESIGN SCIENCE RESEARCH APPROACH 86
3.3 MULTIPLE CASE STUDIES DESIGN 91

4 AN INDOOR POSITIONING SYSTEM AND A MACHINE LEARNING APPROACH FOR INDOOR POSITIONING IN IOT ENVIRONMENTS 95

4.1 SYSTEM DESIGN AND IMPLEMENTATION 96
4.2 BUSINESS, DATA AND SETTING UNDERSTANDING AND PREPARATION LAYER 98
4.3 DATA MODELING AND EVALUATION LAYER 100
4.3.1 Modeling 101
4.3.2 Evaluation 103
4.4 RESULTS TRANSLATION 105

5 CASE A: APPLICATION OF INDOOR POSITIONING USING BLE BEACONS TECHNOLOGY 106

5.1 INTRODUCTION TO THE CASE STUDY 106
5.2 SYSTEM DESIGN AND IMPLEMENTATION 107
5.3 BUSINESS, DATA AND SETTING UNDERSTANDING AND PREPARATION LAYER 109

5.4 MODELING **114**
5.5 CLASSIFIERS' EVALUATION **116**
5.5.1 Classification Performance Comparison *117*
5.5.2 Proposed Filter *119*
5.6 DATA-DRIVEN EVALUATION **124**
5.6.1 Field Study: Positioning Data vs. Purchase Data *125*
5.6.2 Process and Resulting Behavioral Patterns *125*
5.6.3 Setting and Data Collection *130*
5.6.4 Findings *131*

6 CASE B: APPLICATION OF INDOOR POSITIONING USING WI-FI TECHNOLOGY 138

6.1 INTRODUCTION TO THE CASE STUDY **138**
6.2 SYSTEM DESIGN AND IMPLEMENTATION **139**
6.3 BUSINESS, DATA AND SETTING UNDERSTANDING AND PREPARATION LAYER **141**
6.4 MODELING **147**
6.5 EVALUATION **148**
6.5.1 Classification Performance Comparison *150*
6.5.2 Proposed Filter *154*
6.6 USER-BASED EVALUATION **158**

7 MANAGERIAL IMPLICATIONS 161

7.1.1 Retail Operations *161*
7.1.2 Marketing Insights *186*

8 CONCLUSIONS AND FURTHER RESEARCH 199

8.1 RESEARCH OUTCOMES **201**
8.2 THEORETICAL CONTRIBUTIONS **202**
8.3 FINDINGS SUMMARY **207**
8.3.1 Case A: BLE Beacons Summary *207*
8.3.2 Case B: Wi-Fi Summary *209*
8.3.3 Findings/Implications Overview *212*
8.4 LIMITATIONS AND FUTURE RESEARCH **213**

REFERENCES 215

APPENDIX A: INDICATIVE IOT DATA SET STRUCTURE 237

APPENDIX B: FOCUS GROUP 238

LIST OF FIGURES

Figure 1. Research approach....37
Figure 2. Research outline....44
Figure 3. Related work components....46
Figure 4. Techniques comparison....49
Figure 5. Technologies comparison....55
Figure 6. Indoor location-based services factors....59
Figure 7. Fixed and variable length surfaces....61
Figure 8. Forced path layout....62
Figure 9. Grid layout....63
Figure 10. Racetrack layout....64
Figure 11. Straight layout....64
Figure 12. Diagonal layout....65
Figure 13. Angular layout....66
Figure 14. Free flow layout....67
Figure 15. Mixed layout....68
Figure 16. Geometric layout....68
Figure 17. Boutique layout....69
Figure 18. Architecture modules....85
Figure 19. Research approach....88
Figure 20. Indoor positioning system....96
Figure 21. Machine Learning approach for indoor positioning....96
Figure 22. System architecture layers....97
Figure 23: Data and processing challenges....100
Figure 24. BLE Indoor positioning system overview....107
Figure 25. Location-based service overview....109
Figure 26. Detailed map of the ground floor of the Grocery Store....111
Figure 27. Detailed map of the first floor of the Grocery Store....112
Figure 28. BLE Beacon dataset format....113
Figure 29: Ensemble classifier....120
Figure 30. Classifiers ROC space graph....121
Figure 31. ROC curve....123
Figure 32. Positioning error for filter....123
Figure 33: Segment1 areas of interest....127
Figure 34: Segment2 areas of interest....128
Figure 35: Segment3 areas of interest....128
Figure 36: Segment4 areas of interest....129
Figure 37: Segment5 areas of interest....129
Figure 38: Segment6 areas of interest....130
Figure 39: Segment7 areas of interest....130
Figure 40. Wi-Fi indoor positioning system overview....139
Figure 41. Store 1 – Ground floor fingerprinting....142
Figure 42. Store 1 – First floor fingerprinting....142
Figure 43. Store 1 – Second floor fingerprinting....142
Figure 44. Store 1 – Ground floor area division....143
Figure 45. Store 1 – First floor area division....143
Figure 46. Store 1 – Second floor area division....143

Figure 47. Store 2– Fingerprinting areas. 144
Figure 48. Store 2 –Area division. 144
Figure 49. Store 3 – Fingerprinting areas. 145
Figure 50. Store 3 –Area division. 145
Figure 51. Wi-Fi raw data 146
Figure 52. Wi-Fi dataset transformed format 147
Figure 53. Classifiers ROC space graph 156
Figure 54. ROC curve..156
Figure 55. Positioning error for filter 157
Figure 56. Time spent 158
Figure 57. Area visits...159
Figure 58. Shopper flows 160
Figure 59: Heatmap representation 166
Figure 60. Indoor positioning system overview 168
Figure 61. Coupon recommendation overview 170
Figure 62. Fixed and variable length surfaces 181
Figure 63. Indoor location-based service generic architecture 183
Figure 64. Challenges and architecture's modules..184
Figure 65. Factors affective each phase of the architecture 185
Figure 66: Indicative area-based shopper segment 187
Figure 67: Speed-based shopper segmentation 189
Figure 68. Business Analytic framework 192
Figure 69: Cluster 3 drill-down buying analysis..196

LIST OF TABLES

Table 1. Overview of studies that utilize Wi-Fi technology 51
Table 2. Overview of studies that utilize BLE Beacon technology 53
Table 3. Overview of studies that utilize RFID technology 54
Table 4. Indoor positioning issues overview 80
Table 5. Case studies overview 93
Table 6. Modelling format 102
Table 7. Dataset structure 114
Table 8. Number of observations per area ranked by area size 115
Table 9. Classifiers metrics comparison 118
Table 10. Ensemble classifier metrics comparison 121
Table 11: Segmentation features 126
Table 12: Segment characteristics 127
Table 13: Segment overall characteristics 135
Table 14: Segment purchase descriptive statistics 136
Table 15: Cluster shoppers' impulse buying characteristics 136
Table 16. Dataset structure 148
Table 17. Store 1 – Ground Floor APs comparison 151
Table 18. Store 1 – First Floor APs comparison 151
Table 19. Store 1 – Second Floor APs comparison 151
Table 20. Store 2 APs comparison 151
Table 21. Store 2 classifiers' performance 152
Table 22. Store 1 – Ground floor classifiers' performance 152
Table 23. Store 1 – First floor classifiers' performance 153
Table 24. Store 1 – Second floor classifiers' performance 153
Table 25. Ensemble classifier metrics comparison 155
Table 26. Symbols and definitions 163
Table 27: Business Intelligence dimensions 193
Table 28. Findings overview 213

LIST OF ABBREVIATIONS

TERMS	**ABBREVIATIONS**
Angle of Arrival	AoA
Artificial Intelligence	AI
Bluetooth Low Energy Beacons	BLE Beacons
Business Analytics	BA
Design Science Research	DSR
Direction of Arrival	DoA
Fast Moving Consumers Goods	FMCG
Global Positioning System	GPS
Information Systems	IS
Internet of Things	IoT
Key Performance Indicator	KPI
Location Analytics	LA
Management Information System	MIS
Not Only Standardized Query Language	noSQL
Point of Sales	POS
Radio Frequency Identification	RFID
Received Signal Strength	RSS
Standardized Query Language	SQL

1 INTRODUCTION

This opening chapter introduces the reader to the purpose and goals of this doctoral dissertation. It begins by laying the motivation for undertaking this research and by positioning its topic within its research context. Subsequently, it sets pinpoints pertinent research gaps and questions. Then it shortly presents the research methodology and concludes by providing the dissertation outline.

1.1 Research Motivation

Shopper in-store behavior has been an area of interest for more than 60 years, trying to uncover patterns of shopper behavior (Sorensen et al., 2017). Over the last few years it has received great attention not only from store managers but also from researchers and practitioners (Ferracuti et al., 2019; Sorensen et al., 2017). The understanding of shopper behavior can lead to essential insights for retailers and enrich the in-store shopper experience. Although, shopper behavior is easier to be examined in digital environments, the physical stores draw more attention over the years due to the evolution of technology and the modern tools provided for examining in-store shopper behavior, powered by location analytics.

Big Data and Business analytics have enabled new capabilities and insights in retailing thanks to the increase in data quality and application possibilities coming from a mix of new data sources, a smart application of statistical tools and domain knowledge combined with theoretical insights (Bradlow et al., 2017). Scholars and industry partners acknowledge the realization that by adding an additional dimension to business data can significantly enhance the quality of the extracted insights (Pick et al., 2017). The dimension refers to location. Location provides a new context in the perspective of Business Analytics and can facilitate decision makers to discover new insights and, also, a means of an effective communication in term of visualization.

Location analytics (LA) refers to the specialized spatial data analysis in order to understand spatial patterns and relationships in location referenced phenomena (Pick et al., 2017; Senior et al., 2007). Retail business can benefit from having insights regarding the in-store locations that their customers tend to browse during their shopping trip in order to better organize their operation activities (Mou et al., 2018) and, also, improve customer experience (Yaeli et al., 2014).

The advent of IoT (Ali et al., 2015) has facilitated the detection of customer's in-store location. Thus, understanding customer behavior in indoor retail environments gets more and more essential and feasible than ever. To this end, the retail industry exploits the evolution of technology and IoT, along with Business Intelligence in order to create value for the organization, by focusing on the essential information of customer's in-store location (Fink et al., 2017).

Typical sources that are able to facilitate the detection of the in-store location of the shopper are mobile phones via Wi-Fi, Bluetooth, and RFID technology, and cameras (Rai et al., 2011). Spatiotemporal data occur from these technologies and typically they are translated to dimensions such as time, location, and tracking object. Time indicates the timestamp of the data, location refers to the place that the object is detected and with the term tracking object is implied the user or the tracked product. Tracking objects usually require a tracking mechanism; for example, users can be tracked either by utilizing a mobile device or a mobile application that communicates with the wireless infrastructure. These data can be further processes, transformed and be used to create value using location analytics techniques.

Despite the popularity of location analytics and location-based services, they both face a similar problem; i.e. the accuracy of the localization mechanism they use. In order to perform effectively, such mechanisms require detecting accurately the position of the user. The issue of effective detection of the user position is encountered both in academic literature and also a business issue when trying to deploy such mechanisms in real world cases.

From the academic perspective, the issue of user-positioning, especially in indoor environments, is an emerging challenge. Although it is not so difficult and complex to locate the exact position of a moving object outside a building (e.g. using GPS signals and geofencing techniques (Kriz et al, 2016)), this becomes a quite challenging task in indoor environments. Retail stores and shopping malls in general provide fertile ground for the development of indoor positioning systems and relevant research works address two major issues: (a) the localization infrastructure shaped by a positioning technique and a wireless technology (e.g. (Jin et al., 2013; He et al., 2016)) and (b) the provided services on the top of the infrastructure.

The indoor positioning issue appears in the relevant literature as a very challenging one and the achievement of high localization precision (accuracy) is a common objective shared by various scholars (He et al., 2016; Shin et al., 2015). Lymberopoulos et al. (2015) argue that the indoor location problem still remains unsolved and stress the importance of the employment of a realistic approach that would counterbalance the desired localization accuracy with low cost.

From the business perspective, research has indicated that despite the advancement of technology, increased costs, additional hardware and complex deployments are required in order to perform efficient indoor tracking. Management information systems contain location information along with the enterprise data, but still lack to exploit this information to extract knowledge. In addition, a semi-structured focus group, along with extra interviews with indoor positioning system stakeholders and experts indicated effective indoor positioning is a challenge that should be dealt.

Both approaches (academic and business) come down to one thing; effective indoor positioning is required in order to extract reliable location analytics and be able to design effective location-based services. In the context of this thesis we propose a machine learning indoor positioning approach that aims at extracting effectively the position of a user in an in-store retail environment.

1.2 Research Objective & Questions

Considering the above, the starting point and the ultimate objective of this research is to advance the understanding of indoor positioning. By examining this issue various questions may arise:

- Which are the suitable technologies indoor positioning?
- Which is the suitable granularity level of indoor localization?
- How efficient are BLE beacons for indoor positioning?
- How efficient is Wi-Fi for indoor positioning?

- Which is the appropriate number of access points for efficient indoor positioning (Wi-Fi and BLE beacons)?
- How to design and develop systems that perform indoor positioning?
- How to segment locations for indoor positioning?
- How can we extract knowledge by examining the path followed during the shopping trip?
- How can we extract KPIs based on the location attribute?

Such questions may not only refer to IS researchers but also to practitioners and indoor positioning system designers who examine ways to improve the positioning accuracy and thus offer effective location-based services to shoppers.

By going through the related literature, our first objective is to examine and understand existing indoor positioning approaches and systems. Then we seek for the factors that affect the input, the design and the results of such systems and we identify several research gaps. Afterwards, our second goal was to identify how these research gaps are translated into industry open issues and business problems. Via performing desk research, we identified that business people still struggle to adopt and use indoor positioning systems due to positioning accuracy issues and factors that add complexity to the deployment, as many systems require additional hardware and costs in order to provide the best possible accuracy. Our third goal was to identify how indoor positioning is translated to practical implication for decision makers.

Summarizing the research background, the main research questions of this thesis are formulated as following:

- RQ1: How can we perform efficient indoor positioning from spatiotemporal data?
- RQ2. How to develop indoor positioning systems?

To answer this research questions, we develop an indoor positioning artifact and a machine learning indoor positioning approach that can be applied on spatiotemporal data from IoT access points. We assess two different wireless technologies (i.e. Wi-Fi and Bluetooth Low Energy Beacons) and apply Artificial Intelligence techniques to address the question. In addition, to address the research questions we propose a system artifact which is responsible for generating, capturing and processing the data for indoor positioning. The outcome of the proposed approach is the indoor position of the user of the wireless infrastructure.

We deployed this approach to two heterogeneous retail environments cases and validated the results through technical performance. Then we go beyond by examining solely the technical performance and utilize data-driven and user-based evaluation in order to assess the performance of the indoor positioning output. This way we demonstrate the approach generalizability.

The first case concerns spatiotemporal data from a two-floor store from a major Greek FMCG retailer, utilizing BLE Beacon technology as wireless infrastructure. The second case concerns spatiotemporal data from a major Greek electronics retailer.

Following emerged the need to evaluate not only the approach but also the results it generates i.e. indoor position of the user. To do so, we evaluated via machine learning metrics the performance of the output and we go beyond and assess the performance using other approaches. We assess the first case via a field study, by utilizing purchase data and examine whether or not the basket indicates correlation to the shopping trip (via the shopper locations), Then, we assess the second case by conducting user evaluation with practitioners who use such systems in their daily routine.

This thesis moves forward and includes also the business and user acceptance challenges, highlighting the role of the application context and aspires to provide realistic and more holistic, not just technical, guidelines on prospective researchers and designers of location-based services for retail stores and other contexts, as well as encourage them to embrace such design initiatives.

The proposed approach can be the key element of an indoor location-based service; thus, another goal our interdisciplinary research is to highlight those factors that do not affect the indoor positioning approach. Furthermore, this thesis aims to share the development and application challenges encountered; the factors identified affecting the system's quality and performance when deploying an indoor location-based service. Then, it shows how to handle all the arising issues of indoor positioning. Finally, we present a series of practical implications that derive from positioning data that can the support decision making process.

1.3 Research Methodology Overview

Given our research objective and the aforementioned research questions, we adopt as methodological backbone the design science paradigm (Hevner, March, Park, & Ram, 2004) and we consider a machine learning approach that performs indoor positioning as outcome of this study. In design science, the researcher creates and evaluates IT (Information Technology) artifacts and/or theories intended to solve identified organizational problems. The knowledge base is composed of foundations and methodologies used to develop the artifact.

The proposed artifact is evaluated, assessed and refined via simulation, experiment, case study, field study etc. Owing the lack of prior systematic research on the visit segmentation topic this research is based on multiple case studies design. In more detail, the proposed indoor positioning approach is been evaluated by applying it into real data derived from two case studies. Below we present the basic components of design science research and how are addressed in the current dissertation (Figure 1).

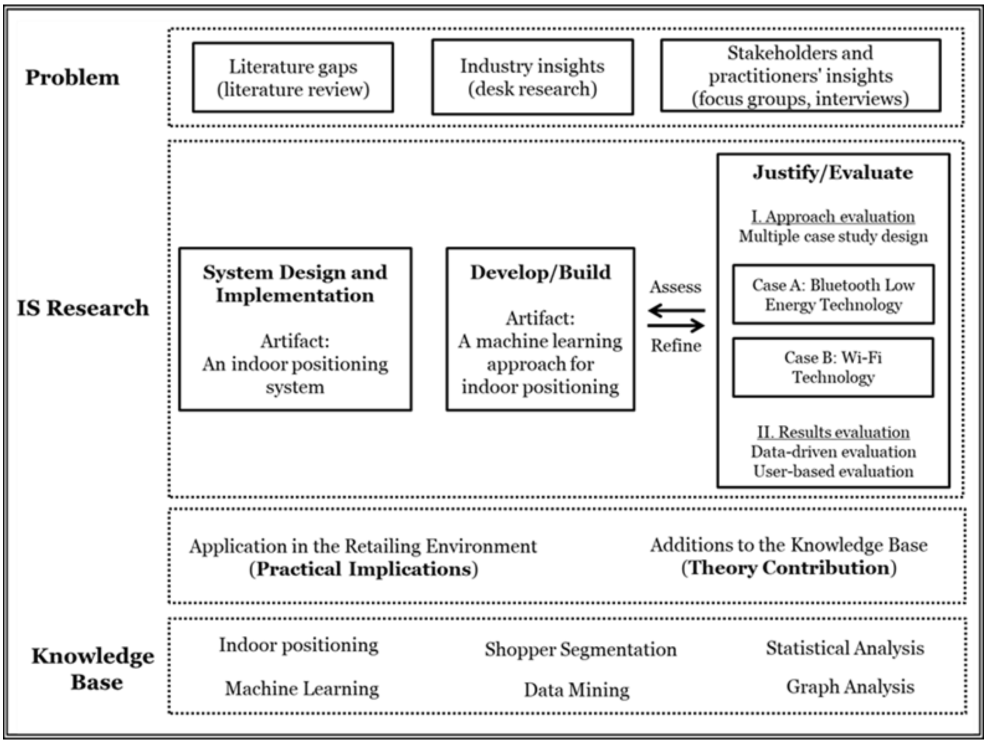


Figure 1. Research approach

(A) Problem definition

This dissertation aims to solve a business problem/need in the retailing environment, which is to perform efficient indoor localization and extract insights in order to identify the reason behind the shopper trajectory during the shopping trip. To better define this problem, it follows the steps below:

- **Literature gaps**: To set the research setting firstly we conducted a review of the pertinent literature. This way we specified the research questions which are related with the indoor localization concept and we pinpoint the research gaps and the purpose of this research.
- **Industry insights**: Having laid the foundations upon which this doctoral research will be grounded, then we investigated various open issues and business problems industry people face, when they try to conceptualize the term of indoor positioning. We found out that the "indoor positioning" is translated into the need of proper level of detection (e.g. X-Y coordinates, store aisle, area, store segment) to identify shoppers' in-store positions and exploit them in business terms.
- **Stakeholders and professionals' insights**: Afterwards to better understand and conceptually define indoor positioning in retail domain, we conducted a semi-structured focus group discussion with 22 participants from both retailers and suppliers in order to understand the way they apprehend indoor positioning in the retail content. We further conducted interviews with retail system users that utilize indoor positioning and location analytics software and approaches in order to evaluate the performance of the proposed indoor positioning approach of this thesis.

(B) Develop/Build

Then we develop and evaluate a technology-based solution that is relevant to the above research problem. In this research the **developed artifact is an approach**, providing a certain manner to handle the appropriate machine learning **to extract the indoor position.**

(C) Justify/Evaluate

Then, we put the approach in practice to evaluate it and realize if it can solve the original problem. This evaluation includes two steps: (i) Approach evaluation and (ii) results evaluation. Regarding the first one owing the lack of prior systematic research on the indoor positioning area, to address this objective the research is based **on multiple case studies design**. In more detail, **the proposed approach is been evaluated** by **applying it into** real data derived from **two case studies.** In both cases we conducted a data-driven evaluation and a user-based evaluation in order to gain insights regarding the business performance of the approach. In the data-driven evaluation we conduct a field study to identify whether or not the shopping journey data can reveal knowledge regarding the purpose of the visit via the purchase data. To do so, we compare the paths followed from 100 shoppers and compare the results to their purchase data. Regarding the user-based evaluation we examine real world data along with indoor positioning system users and assess the quality of the extracted information.

(D) Knowledge base

To build the proposed approach, we used both theoretical foundation and methodologies. Theory knowledge inputs in the developed approach. In more detail in our approach we utilize machine learning and data mining techniques such as clustering and classification, data mining algorithms such as k-means were used to develop the approach. Statistical analysis was used to evaluate the performance of the machine learning algorithms and detect the ones that perform significantly better than the others. Finally, we utilize graph analysis in order to examine shopper paths and shopper segmentation, as we try to detect shopper segments based on the way they navigate into the store.

1.4 Thesis Outline

There are seven (7) chapters that constitute this dissertation as follows:

- Chapter 1 (Introduction)

  This chapter introduces the readers to the main concepts of this research, i.e. indoor positioning. It communicates the research's motivation, as well as the key research questions. Last, but not least, this chapter shortly describes the research approach according to which the main research objective and question are answered.

- Chapter 2 (Research background)

  It is critical to be cognizant of the rationale for the relevance of the work. Therefore, in this chapter an extensive literature review is presented to pinpoint the research gaps. We go through the topics of indoor positioning, machine learning approaches for indoor positioning and then examine the literature in the area of location analytics. The detected research gaps are translated into industry open issues and business problems. Then, stakeholders and user-based insights via interviews are used as tools for highlighting the significance of this study and for validating the main research objective. Finally, and conclude with the main business and research gaps.

- Chapter 3 (Research methodology)

  The aim of this chapter is to present the research methodology employed to address the research objectives and answer the research questions. Given our research objective and the aforementioned research questions, we adopt as methodological backbone the Design Science Research (DSR) paradigm (Hevner et al., 2004) and we consider a machine learning approach that performs indoor positioning as outcome of this study. For collecting data for the various steps of DSR, three different cases studies are selected and presented (multiple case study design). Thus, firstly we present and explain DSR, and then we explain Multiple Case Study rationale. Afterwards, we describe in detail how we adopt these two approaches into the design of this dissertation.

- Chapter 4 (A machine learning approach for indoor positioning)

  In this chapter we describe in detail the proposed data-driven approach that may be used to perform indoor positioning. In high level it includes the following phases/layers: (a) data understanding and preparation, (b) modelling and evaluation (c) results translation. The major input of our approach is data related to spatiotemporal data (e.g. access points and timestamps along with signal strength). The output is indoor positioning areas that are further utilized to extract patterns of similar behaviour regarding in-store browsing.

- Chapter 5 (Case A: Application of indoor positioning using BLE Beacons technology)

  Here, we put our proposed location analytics approach in practice demonstrating how it achieves the original goal, i.e. detect shoppers' in-store positions in a BLE Beacon environment. This first case concerns spatiotemporal data from BLE Beacons technology of a major Greek FMCG retailer. We demonstrate how to machine learning approach is applied on BLE Beacons and finally we evaluate the results via two ways; i.e. the technical evaluation of the algorithms and then via a filed study that examines the output of the indoor positioning approach in correlation with the purchased data by the shoppers.

- Chapter 6 (Case B: Application of indoor positioning using Wi-Fi technology)

  Respectively, in the second case, we utilized spatiotemporal data from Wi-Fi technology of a major Greek electronics retailer. We demonstrate how to machine learning approach is applied on Wi-Fi data. Unlikely to BLE Beacons case we do not have the flexibility to alter the wireless infrastructure in order to improve the signal quality. Finally, we evaluate the results via two ways; i.e. the technical evaluation of the algorithms and then via a user-based evaluation. Indoor positioning system users provide us with input regarding the outcomes of the positioning approach, enhancing this way the quality of the results.

- Chapter 7 (Managerial Implications)

  Chapter 7 overviews the managerial implications of this research. The practical implications are broken down to Retail Operations (e.g. location-based KPIs, lessons learnt for indoor location-based systems designers in contemporary retail) and Marketing Insights (e.g. location-based behavioural patterns, location-based shopper segmentation).

- Chapter 8 (Conclusions and Further Research)

  Chapter 8 describes the main outcomes of this research. Then, it presents and discusses the research's contribution to theoretical knowledge along with its practical value. In addition, this chapter summarizes the findings and implications of each case study. Finally, the research limitations are pointed out and avenues for further research are recommended.

To support reading comprehension, the following figure (Figure 2) presents thesis' structure according to the research methodology, as described in the previous sub-section.

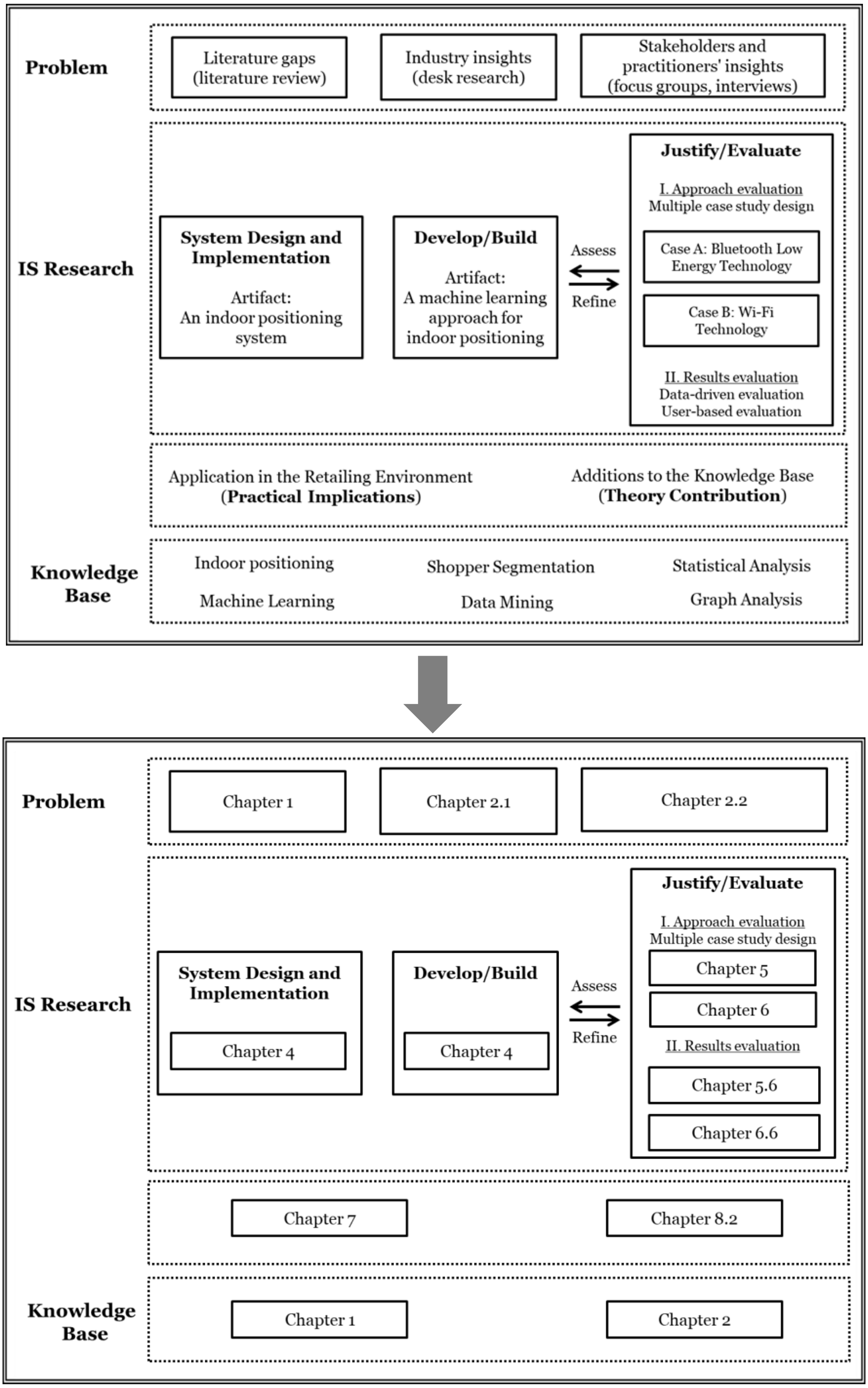


Figure 2. Research outline

2 RESEARCH BACKGROUND

It is critical to be cognizant of the rationale for the relevance of the work. Therefore, an extensive literature review, an industrial exploratory research and customer insights via focus groups and questionnaires have been used as tools for highlighting the significance of this study.

The basic axis of the literature review refers to indoor positioning, to the intelligence techniques that are applied to perform it and then factors that affect indoor positioning. The first level of discussion is the indoor positioning techniques and approaches upon which intelligence is applied to perform efficient positioning. Then, to top level of discussion is location analytics which generate value and insights from spatiotemporal data. Thus, we further review studies in the field of location analytics, as they introduce the concept and the value of the "location" dimension. In location analytics scope we examine, also, shopper trajectory studies, as the sequence of position data forms shopper paths. The location analytics area literature review continues with studies that refer to shopper segmentation. Shopper segmentation is an area that has receives great attention by scholars; however, the dimension of location may provide additional segmentation approaches that rely on location. As the shopper position is meaningful for researchers and decision-makers we go through the retail KPIs literature, in order to examine and extract useful indicators. Another topic of interest is the granularity level in terms of the precision of the tracking object's position. We conclude this section by reviewing the research gaps that this research aims to contribute.

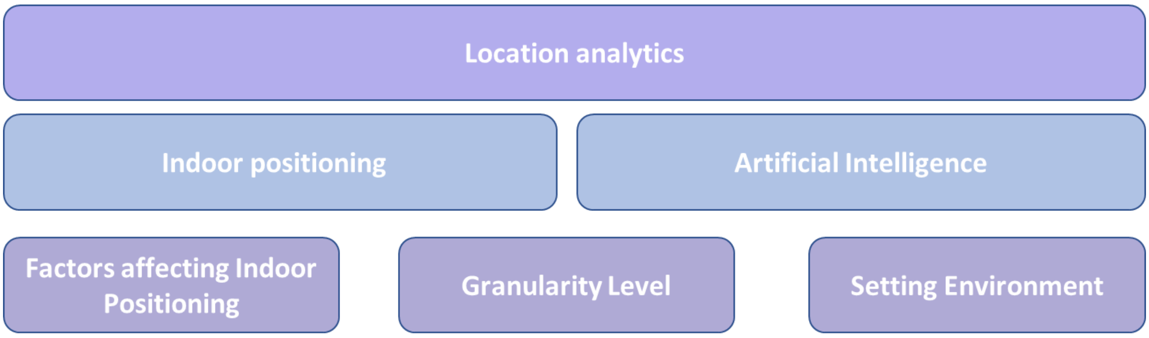


Figure 3. Related work components

2.1 Literature Review

2.1.1 Indoor Positioning

The popularization of location-based technologies, such as Wi-Fi, Bluetooth and radio-frequency identification (RFID), has facilitated users to track and observe several objects moving into indoor environments (Civilis et al., 2005; Jin et al., 2013). At the same time, massive amounts of data are generated while these objects move around in various environments. However, several challenges may arise; a common one regards how we can utilize these data to identify the position of an object within indoor environments. Over the past years, researchers have proposed various techniques and approaches to tackle this issue and effectively determine or track the location of an object in an indoor environment. These studies can be classified into lab experiments (e.g. (Wang et al., 2015), (Palipana et al., 2017)) and real-world cases (e.g. (Yim, 2008), (Campos et al., 2014)), where indoor localization techniques are proposed for controlled and operational environments, respectively. Although lab experiments achieve significant localization results, the proposed techniques do not perform as efficiently in real world scenarios; thus, indoor positioning is a problem that remains unsolved. Following are presented studies referring to indoor position techniques and then describe the current wireless localization technologies that serve such purposes.

2.1.1.1 Indoor positioning techniques

Various techniques have been developed to determine the geometrical placement or position of an object in an indoor environment. The location is estimated by taking into consideration the distance as well as the angle between the transmitters and the tracked objects in the environment (Curran et. al, 2011). In addition, signal strength is used as a means of creating signatures for each location, in order to identify an object's location (Lassabe et al., 2005; Youssef & Youssef, 2007). The most common techniques used for indoor localization are triangulation and trilateration. Albeit, techniques such as fingerprinting, proximity and other machine learning-based approaches are developed to mitigate measurement errors. Each technique has its advantages and disadvantages; thus, the selection of the most appropriate one highly depends on the application context, while in some cases, using more than one technique and algorithms simultaneously may lead to better performance. The most common techniques of indoor localization (Turgut et al., 2016) are described below.

(Tri)Lateration: The lateration (Turgut et al., 2016) method utilizes the distance between the localized transmitters and the moving object to estimate the object's location. The minimum number of transmitters required for accurate location determination is three. When three transmitters are used to estimate the location/position of an object, the technique is called trilateration. When more than three transmitters are used, the process is called multilateration.

(Tri)Angulation: The angulation (Turgut et al., 2016) method utilizes the angles between the localized transmitters and the moving un-localized object to estimate the object's location. When at least two angles of a moving object from two localized transmitters are measured to estimate the object position, the process is called triangulation. In this case, trigonometric laws of sines and cosines are used to estimate the position of the object.

Fingerprinting: Fingerprinting is the most popular method of localization with high applicability in complex environments due to its high accuracy and low complexity compared to other methods (Subedi and Pyun, 2017; He and Chan, 2016). It utilizes the received signal strength (RSS) as an approximate metric to determine the indoor location. To apply the fingerprinting technique, first we collect the signal strengths by the localization technology (fingerprints) for each possible location of the environment in a database (offline phase) and, thus, we formulate the reference fingerprinting map (RFM) (Zuo et al., 2018). Then, in the online phase, an object's signal collected from a location in real time is com-pared with the fingerprints in the RFM, to solve for the location of the object (Zuo et al., 201816; Wang et al., 2015). In this phase, fingerprinting is based on classification algorithms and methods, such as neural networks (Galvan-Tejada et al., 2014; Mazan and Kovarova, 2015), decision trees (Yim, 2008), k-nearest neighbors (Liu et al., 2007), support vector machines (Liu et al., 2007) and random forests (Breiman, 2001), which predict the object's current location based on the fingerprint database (Li et al., 2019; Carbonell et al., 1983). Since fingerprinting relies on signal strength, problems that may occur are related to signal variations deriving from communication issues, such as fading, interference or even from environmental factors. Fingerprinting is applied on a fixed environment and each applied process is effective only for the deployed environment, meaning that deploying an existing setting on another environment will not work.

Proximity: This technique determines whether an un-localized object is close to a localized transmitter. By referring to closeness to known transmitters in indoor environments, one can reduce the cost and effort for localization (Kim et al., 2015).

Dead reckoning: This technique is based on calculating the current position of a moving object by using a previously known position, and forecast the next position based on estimated patterns (such as speed and the route followed) (Pratama and Hidayat, 2012; Sharp and Yu, 2014).

	Trilateration	Triangulation	Fingerprinting	Proximity	Dead Reckoning	Hybrid Algorithms
Measurements Techniques	RSS, ToA, TDoA	AoA, DoA	RSS	RSS	Velocity acceleration	Hybrid
Accuracy	Medium	Medium	High	Medium	Medium	High
Time Cost	Low	Low	High	Low	Low	Medium
Distance	Low	Low	Medium	Medium	Low	High
Algorithm type	Deterministic	Deterministic	Probabilistic	Deterministic	Deterministic	Probabilistic Deterministic
Specification	Time based	Direction based	Range based	Range based	Time based	Hybrid

Figure 4. Techniques comparison

One may notice the vast amount of techniques used for indoor positioning. All approaches refer to actions on the surface of the indoor environment. As movement is not a trivial problem, position determination has been tried to be resolved in the literature as referred before. Some approaches emphasize on pattern discovery, while others on the techniques in order to be able to detect object's position. Thus, we observe the existence of location intensive techniques (e.g. trilateration) and data intensive ones (e.g. fingerprinting) that can be used depending on the case needs.

2.1.1.2 Indoor positioning wireless infrastructure technologies

The involved technologies used to track an object while it is moving around in indoor environments focus mostly on (A) Wi-Fi, a popular wireless networking technology (Youssef et al., 2003), (B) Bluetooth, a wireless technology standard for communication over short distances (Haartsen, 2000), and (C) RFID, a technology based on radio-frequency identification via electromagnetic fields that can identify and monitor tags that are attached to objects of interest (Shepard, 2005; Srivastava, 2004). At this point, we should mention that GPS technology is not appropriate, as signal fails to reach indoor environments (Turgut et al., 2016; Enge and Misra, 1999; Kourogi, 2006).

Wi-Fi: One of the main problems of Wi-Fi is signal attenuations (Morales et al., 2014), which is a common cause of faulty measurements in indoor positioning. Apart from signal processing (Dardari et al., 2015), numerous studies have examined indoor positioning approaches that either survey indoor position systems or examine attributes that can improve accuracy (Liu et al., 2007). Regarding accuracy improvement, Torres-Sospedra et al. (2015) focus on the optimal distance function and conclude to the selection of the best configuration for an indoor positioning algorithm. Machine learning techniques have also been used for indoor positioning, such as decision trees (Yim, 2008), unsupervised labelling on sequential data (Perez et al., 2016), unsupervised clustering for multi-floor environments (Campos et al., 2014) and Online Sequential Extreme Learning (Zou et al., 2014; Zou et al., 2015). Bayesian models are also utilized and eliminate the problem of training data;

thus, they are more effective in terms of the cost of gathering data (Madigan et al., 2005). Other studies examine indoor positioning techniques regarding fingerprinting (Lin and Lin, 2005) and classifiers to achieve effective and more accurate indoor location (Villarubia et al., 2013; Galván-Tejada et al., 2014). Finally, requirements and challenges for IoT are also a field of interest for indoor positioning (Turgut et al., 2016).

Authors	Techniques
Lin and Lin (2005)	Fingerprinting
Liu et al. (2007)	Triangulation / Fingerprinting / Proximity
Yim J. (2008)	Decision trees
Campos et al. (2014)	Fingerprinting
Galván-Tejada et al. (2014)	Random Forest / KNN/ Neural Networks
Morales et al. (2014)	Faulty measurements
Zou et al. (2014;2015)	Online Sequential Extreme Learning
Torres-Sospedra et al. (2015)	Fingerprinting
Perez et al. (2016)	Unsupervised learning
Turgut et al. (2016)	Trilateration/ Fingerprinting / Proximity/ PDR/Hybrid
Duque Domingo et al. (2017)	Fingerprinting, Synchronized Euclidean distance
Cao et al. (2019)	Fingerprinting
Wang et al. (2019)	Fingerprinting/ Clustering/ signal weighted Euclidean distance
Zhang et al. (2019)	Convolutional Neural Network/ Gaussian Process Regression

Table 1. Overview of studies that utilize Wi-Fi technology

Bluetooth: Bluetooth is a recent technology in the field of indoor positioning and has been applied both for in-door localization, as well as for location proximity presenting challenges that should be taken into consideration (Cabero et al., 2014).

More specifically, Bluetooth-based indoor positioning has been conducted by combining BLE beacons with a pedestrian dead reckoning (PDR) technique to provide meter-level positioning and estimate current position by using a previously determined position (Li et al., 2015). Moreover, they have developed a range-based localization system based on stigmergy that relies on Received Signal Strength (RSS) of BLE beacon packets (Subhan et al., 2011; Palumbo et al., 2015). On the other hand, Diaz et al. (Diaz et al., 2010) introduced an indoor Bluetooth-based localization system (titled Bluepass) that achieves localization at room-level based on a signal coverage density method. Alike, Bobek et al. (2015) dealt with the issue of indoor positioning by determining location at room-level.

Due to the nature of Bluetooth and the signals emitted, the most widely used and efficient technique used for Bluetooth-based indoor positioning is fingerprinting (e.g. Mazan and Kovarova, 2015; Subhan et al., 2011; Kriz et al., 2016; Subedi et al., 2017; Mohsin et al., 2019). For example, Subedi et al. (2017) combines fingerprinting with weighted centroid localization aiming to reduce the total number of collected fingerprints and, thus, improve the time required for the positioning process. One of the most recent studies combines fingerprinting with geometric techniques to pinpoint the movements of patients in a BLE beacon-enabled hospital room (Mohsin et al., 2019). Moreover, when utilizing Bluetooth signals, a commonly used classification technique for indoor positioning is neural networks. Mazan and Kovarova (2015) collect fingerprints (signal strengths by the beacon devices) and employ an artificial neural network to determine the user's location. Respectively, employing multiple neural networks has achieved better accuracy by handling the variability of the strength of the signals transmitted by Bluetooth devices (Altini et al., 2010).

Authors	Techniques
Altini et al. (2010)	Neural Networks
Diaz et al. (2010)	Signal Coverage Density Method
Subhan et al. (2011)	Fingerprinting
Cabero et al. (2014)	Proximity
Kim et al. (2015)	Time windows and frequency
Li et al. (2015)	Dead Reckoning
Palumbo et al. (2015)	Stigmergy
Mazan and Kovarova (2015)	Neural Networks
Bobek et al. (2015)	Rule learning
Faragher and Harle (2015)	Fingerprinting
Paek et al. (2016)	Geometric Adjustment
Kriz et al. (2016)	Fingerprinting
Castillo-Cara et al. (2017)	Supervised learning
Yohan & Winata (2018)	Proximity
Ferreira et al. (2018)	Geometric Adjustment
Zuo et al., (2018)	Graph optimization
Ke et al. (2018)	Multilateration
Subedi & Pyun (2019)	Fingerprinting
AL-Madani et al. (2019)	Fuzzy logic
Mohsin et al. (2019)	Fingerprinting

Table 2. Overview of studies that utilize BLE Beacon technology

RFID: RFID (Radio Frequency Identification) technology has also been utilized for indoor positioning. Wang et al. (2015) propose an indoor localization scheme that relies on curve fitting and location search. Their approach consists of two phases; an offline phase where the environment is divided into smaller areas and fingerprints are generated for each one and an online phase where the area is determined and then location search algorithms are deployed to determine the area. Calderoni et al. (2015) utilize a hierarchical structure of classifiers. Huang et al. (2015) introduce an approach that reduces the RSSI drift, localization, error, computational complexity and deployment cost of indoor systems by using Kalman-filter drift removal and Heron-bilateration location estimation. Zou et al. (2013) introduce two algorithms to provide higher localization accuracy; i.e. weighted path loss and extreme learning machine. Montaser and Moselhi (2014) propose an approach for indoor location identification and material tracking for construction projects. Lastly, respectively to material tracking, RFID technology has been used for object tracking in the Supply Chain (Bose and Pal, 2005; Keller et al., 2014).

Authors	Techniques
Zou et al. (2013)	Extreme Learning
Keller et al. (2014)	Classification
Montaser & Moselhi (2014)	Triangulation/ Proximity
Calderoni et al. (2015)	Random Forest classifiers
Huang et al. (2015)	Kalman-filter drift removal/ Heron-bilateration
Wang et al. (2015)	Curve fitting
Xu et al. (2017)	Bayesian probability and K-Nearest Neighbor
Liu, F. and Zhong, D. (2018)	Glowworm Swarm Optimization with semi-supervised online sequential extreme learning
Shen et al. (2019)	Angle of arrival method and spinning antenna

Table 3. Overview of studies that utilize RFID technology

Hybrid methods: Apart from solely using one technology, approaches that combine Wi-Fi and Bluetooth technology to detect a user's location have also been found in the literature (Vu et al., 2011; Galvan-Tejada et al., 2013). Vu et al. (2011) present a framework for constructing predictive models of users' movements by combining Wi-Fi and Bluetooth traces in order to detect where the moving object is. Mirowski et al. (2013) utilize Wi-Fi, Bluetooth, Long Term Evolution (a 4G wireless broadband technology), and magnetic signals in order to achieve instant recovering of the moving object position.

	BEACONS	GPS	WI-FI	NFC	RFID
Recommended for	In/near-store and micro-location use-cases	Macro-location and out of store use-cases	In-store use-cases	Close proximity, secure interaction	In-store use-cases
Potential uses	In-aisle notifications and offers, in-store navigation, hands-free payment	Near-store notifications and offers, pre-arrival customer 'check-in'	In-aisle notifications and offers, in-store navigation, hands-free payment	Payments, product tagging	product tagging
Ease of set up and maintenance	Medium	Medium-high	Medium	Medium	Medium
Range	Medium	Long	Medium-low	Close	Medium-low
Accuracy	Medium	Medium-low	Medium	High	Medium
Ease of use for consumer	Medium	Medium	Medium-high	Medium-high	Medium-high
Energy efficiency on consumer device	Medium-high	Medium-low	Medium-high	High	Medium-high

Figure 5. Technologies comparison

2.1.1.3 Factors Affecting Indoor Positioning

Having examined studies across domains, we specialize to the challenges of implementing an indoor positioning service for location-based coupon recommendation, based on Bluetooth Low Energy Beacons technology, using a case study in the retail sector and compare the factors that affected our work in correlation with the factors examined in the literature. We perform an across domain and technology study in order to examine the challenges and issues encountered during the development of indoor location-based services and detect patterns among them. Then we discuss the roadmap followed to overcome the encountered challenges and we present the lessons learnt when developing such systems.

Over the years, the evolution of technology in indoor environments (i.e. Wi-Fi, RFID, and BLE Beacons) has facilitated the development of indoor positioning mechanisms in order to deploy appropriate services. Several works in literature refer to indoor location-based services and applications. However, the need for high accuracy in each case is often limited by several constraints that should be overcome in order these mechanisms to perform efficiently.

Indoor positioning services have been developed for cases like museums (Kuflik et al., 2011), hospitals (Calderoni et al., 2015; Yang et al., 2015), tourism (Curran and Smith, 2006), elderly and disabled people (Marco et al., 2008), smart buildings (Lin et al., 2016), couponing and infomediation (Zou and Huang, 2015) and for generic purposes based on different technologies such as Bluetooth (Diaz et al., 2010), Wi-Fi (Au et al., 2013) and RFID (Huang et al., 2015). Each case is characterized by different requirements in terms of accuracy and cost and is deployed based on the specific needs for each of the cases. By reviewing the indoor positioning literature and studies regarding the development of indoor positioning systems we detect and examine several factors that seem to affect performance varying from the simple metric of positioning accuracy to the acceptance of the positioning system from the stakeholders.

The design and development of a location-based service should take into consideration both technical and business-related requirements and challenges rising from the domain of the service (Dhar and Varshney, 2011) and address them respectively in order the service to perform efficiently. Each domain encounters different kind of challenges in order to perform indoor localization effectively (Lymberopoulos et al., 2015). For example, museums are environments that the exhibits are dense and serve plenty of information; thus, requiring an accurate visitor position (Kuflik et al., 2011) so as the location-based service to perform efficiently. To this end, positioning accuracy is an important factor for this domain. Moreover, the indoor positioning mechanism in the museum domain requires also availability and stability to provide the required functionality. Hospital environments (Calderoni et al., 2015; Yang et al., 2015), have similar requirements differentiating to the fact that the system detects the room that the patient is located. Thus, another factor that affects a positioning mechanism is the granularity level of the position detection (i.e. x-y coordinates, area or room level).

Literature (Nuaimi and Kamel, 2016; Mainetti et al., 2014) acknowledges the following problems and challenges when implementing indoor positioning systems and algorithms which are (a) the accuracy of the system meaning that the more accurate to the closest calculated position the more efficient is the system, (b) the range of coverage meaning that the wider the range the better is the performance of the system and (c) the security meaning the mechanisms and techniques used to ensure user privacy. In addition (Xia et al., 2017), introduce the following challenges that should fulfill an indoor positioning system that refer to the following: (d) complexity meaning that low complexity methods have better adaptability to the dynamic change of user's position, (e) robustness meaning that the system achieve high accuracy and precision even when errors occur, (f) scalability, meaning that the system performs efficiently when the environment expands, and (g) cost meaning that low-cost indoor positioning systems are more likely to be adopted.

Apart from the technical challenges of the indoor positioning system, the radiofrequency technology used as infrastructure inserts several challenges that should be overcome. The effects on the signal strength (RSS fading effect) coming from either the (h) human bodies interfering in the environment or from (i) obstacles (Multipath Effect) that affect the signal may cause signal loses and affect the system's accuracy. The (j) number of access points used in the infrastructure plays also an important factor of the quality of system's performance. Too few access points result in inaccurate positioning data, leading to poor performance. However, blindly increasing the access point number can result further costs in terms of infrastructure and processing. Finally, the (k) mobile devices play an important role, as the signal strength in a specific point can be different when measured by different mobile devices. This affects the positioning system, as signal strength is mostly used to detect the position of the user.

An additional challenge that is involved when developing and deploying indoor positioning systems is the user acceptance (Yoon et al., 2017). Users of such systems and services tend to express concerns regarding privacy [Yun et al., 2013], while others tend to seek motivation to use such services (Beeck and Toporowski, 2017; Kwon et al., 2007; Zhu et al., 2017). User acceptance is performed also in terms of logging in an application, by connecting to the store

Wi-Fi network, or by entering a store where they are priory informed that they may be tracked for the provided service.

Finally, the business challenges that should be dealt refer to the requirements set by the desired functionality of the system. Such challenges refer to the required accuracy of the system. Apart from the technical challenge of system's accuracy, the business needs may require different level of accuracy or may require the detection of different areas. For example, in a grocery retail store, the store manager may require different areas division during different time periods which should adapt dynamically. Ice creams section is an area of interest mostly in the summer and the store manager may require knowing which parts of the ice creams section gathers most attention by the customers.

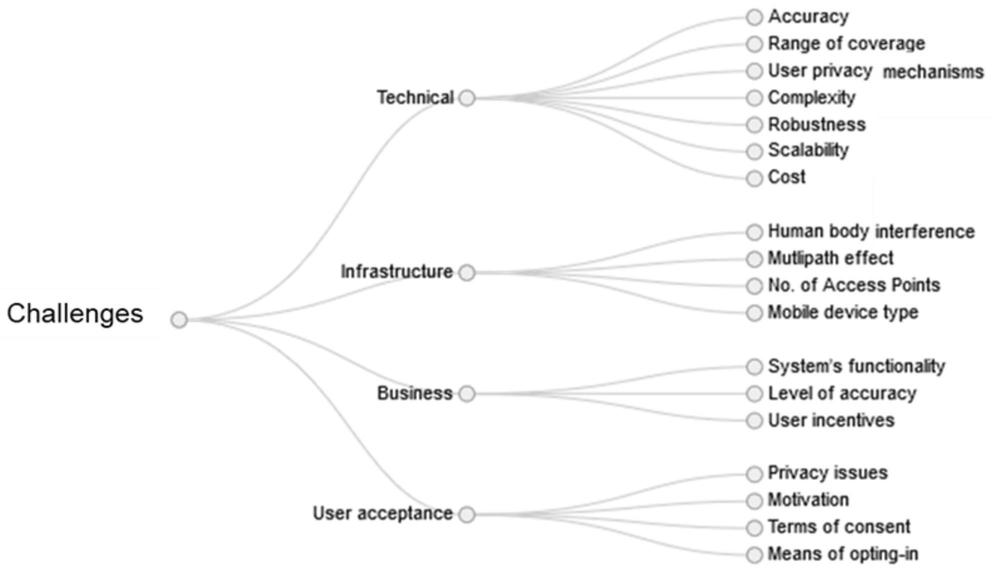


Figure 6. Indoor location-based services factors

On the contrary, during winter this information is not needed in such detail. In addition, the unit of analysis may vary regarding the domain of interest. For example, in a museum the unit of analysis is the visitor, while in a grocery retail store the unit of analysis may be either the trolley-basket or the customer. Last but not least, the incentives (Yoon et al., 2017; Beeck & Toporowski, 2017) that will be offered to the users of the indoor positioning service are also a business challenge as they have to be related to users' interests to motivate them adopt the service.

2.1.1.4 Granularity level

The most important factor when dealing with indoor positioning is the precision of the object's position. Literature examines indoor positioning either in terms of precision in meters, proximity or in terms of area. Thus, technology or other limitations (e.g. the environment) may be a critical issue that affects the applicable level of precision. BLE beacons is a wireless technology where indoor positioning is affected by signal attenuations and fading and should determine the positioning level based on other factors such as the environment, the setting and even the product material and are also related to the business requirements that may occur.

Studies in the literature (Wang et al., 2015) approach the issue of indoor positioning by segmenting the initial surface into equal smaller areas, thus forming a grid which can be then used to detect the exact segment where the object is located and applied iteratively in order to further split an area into smaller ones for higher precision. Using the grid over a surface seems an appropriate solution to detect the location of the object (i.e. fixed-length surface see Figure 7). However, various factors such as the physical layout of the surface or technological requirements or business needs introduce restrictions that may affect the size and dimensions of the area and as a result the detection of the position of the object (i.e. variable-length surface).

The physical layout may lead to the adoption of areas with different sizes due to obstacles or walls that do not allow the sole use of areas with equal dimensions. In addition, technological restrictions may require a minimum threshold of area dimensions, as signal attenuations may be unable to effectively detect the position of the object. Finally, business needs may indicate that some areas are characterized as being of more interest than others. For example, chocolates are more popular during winter, while ice-creams are more popular during summer. As a result, low interest areas can be grouped into bigger-sized areas, while high interest areas can be split into smaller-sized areas.

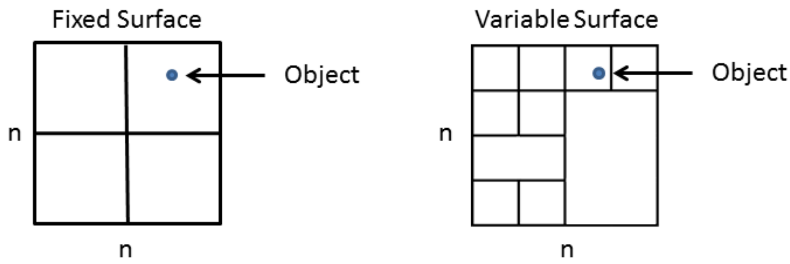


Figure 7. Fixed and variable length surfaces

In the scope of this study we utilize a variable-length surface conceptual model, i.e. areas with different size and dimensions due to limitations that occur from the layout of the store, the deployment of access points and the business needs defined by stakeholders. It's a more realistic and effective conceptual model that can be easily adopted for cases where multiple factors affect an indoor positioning mechanism.

2.1.1.5 Store Layouts

Layout is often the first thing most of us would notice on entering an operation because it governs its appearance. It also determines the way in which transformed resources – the materials, information and customers – flow through the operation (Slack et al., 2010). The most practical layouts are derived from only four basic layout types. These are (a) fixed-position layout, (b) functional layout, (c) cell layout and (d) product layout. Thus, we review literature (Vrechopoulos et. al, 2004; Aghazadeh, 2005; Griffith, 2005; Page et. al, 2019) and conclude to the following ten (10) store layouts.

2.1.1.5.1 Forced path Layout

This layout directs the customer on a predetermined route through the retail store. Research shows that the retailer achieves a uniform and efficient customer flow that promotes higher sales. The advantage of this layout is that every aisle in the store is maximized. With customers exposed to all of the merchandise offered, this design might entice the customer to make an unplanned purchase. However, using this store layout contains to risk of irritating shoppers that have a specific task and desired location in mind. It could overwhelm shoppers by hurrying them through an experience of customers all moving in one direction together, quickly.

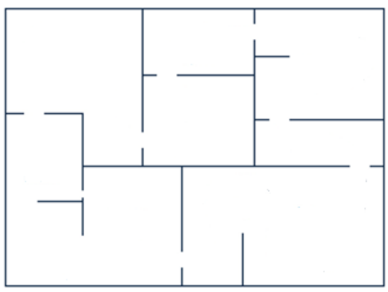


Figure 8. Forced path layout

2.1.1.5.2 Grid Layout

The grid layout is probably one of the most common layouts encountered by most people. These are generally favored by grocery stores, convenience stores, and hardware stores which feature multiple aisles. One of the key benefits of using this type of layout is that it is so ubiquitous your customers will instantly familiar with navigating it. They can also move quickly and efficiently through the space by reading signs which are usually located in the aisles themselves. The grid layout also allows you to display quite a lot of products and allows them to find specific items. Unfortunately, because the layout is so common it is also the most uninspiring. It forgoes creating an experience for customers and they may be confused by the groupings of items. The aisles themselves also create a bottleneck which might make it difficult for customers to squeeze by each other.

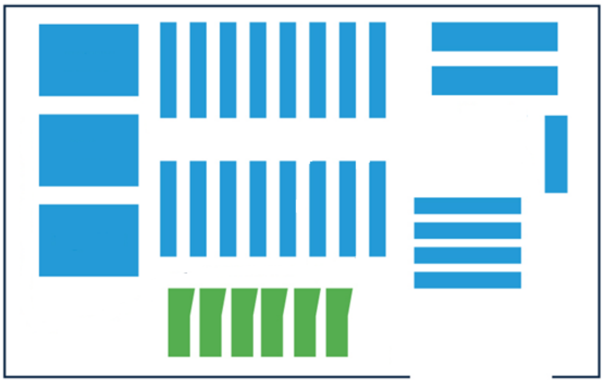


Figure 9. Grid layout

2.1.1.5.3Racetrack Layout

Racetrack is also known as the “loop layout”. It shares a lot in common with the forced-path layout. Featuring only a single path that loops around the retail space, it encourages customers to circle around the store and visit different areas or departments within the retail space. In this configuration, there are a few departments in the center with a path around it and more departments along the walls. Like the forced path, the racetrack creates a more rigid layout for customers to follow which can be frustrating if they’re looking for a specific item. This makes the racetrack unsuitable for stores which require high traffic and quick turnover or customers who need time to consider purchases.

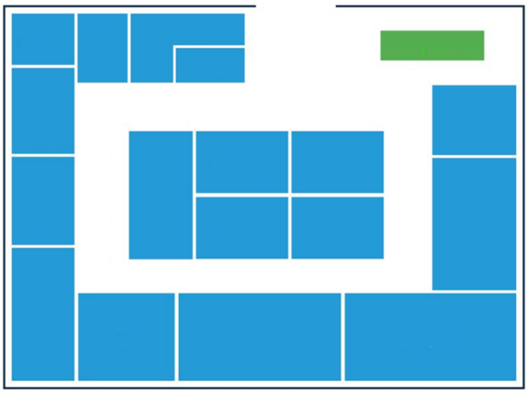


Figure 10. Racetrack layout

2.1.1.5.4Straight Layout

The straight store layout is efficient, simple to plan, and capable of creating individual spaces for the customer. Plus, a basic straight design helps pull customers towards featured merchandise in the back of the store. Merchandise displays and signage is used to keep customers moving and interested. Convenience stores, and small markets use the straight design efficiently. However, the drawback is the simplicity: Depending on how a customer enters the store and moves past the transition zone, it may be more difficult to highlight merchandise or draw them to a specific location.

Figure 11. Straight layout

2.1.1.5.5 Diagonal Layout

The diagonal store layout uses aisles placed at angles to increase customer sightlines and expose new merchandise as customers navigate through the space. A variation of the grid layout, the design helps guide customers to the checkout area. Small stores can benefit from this space management option, and it is excellent for self-service retailers because it invites more movement and better customer circulation.

When the checkout is located in the center and possibly raised up, the diagonal layout offers better security and loss prevention due to the extra sightline effect. The downside of this layout is that it doesn't enable the customer to shortcut toward specific merchandise, and the risk of narrow aisles is higher.

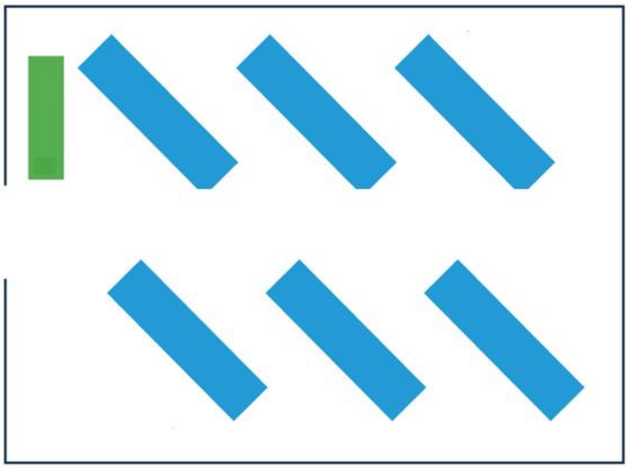


Figure 12. Diagonal layout

2.1.1.5.6 Angular Layout

The name of this design is deceptive, as the "angular" store layout relies on curved walls and corners, rounded merchandise displays, and other curved fixtures to manage the customer flow. Luxury stores use this layout effectively because, according to Herb Sorenson's research from Inside the Mind of the Shopper: The Science of Retailing, customers notice free-standing product displays 100 percent of the time (end cap displays - those at the end of aisles - also get noticed 100 percent of the time).

There is a perception of higher quality merchandise that the angular layout leverages to target the appropriate customer behavior in that environment. And although this design sacrifices efficient space use, because of the rounded displays and limited shelf space, if a retailer has sufficient inventory storage away from the sales floor, this layout is useful in creating a unique perception.

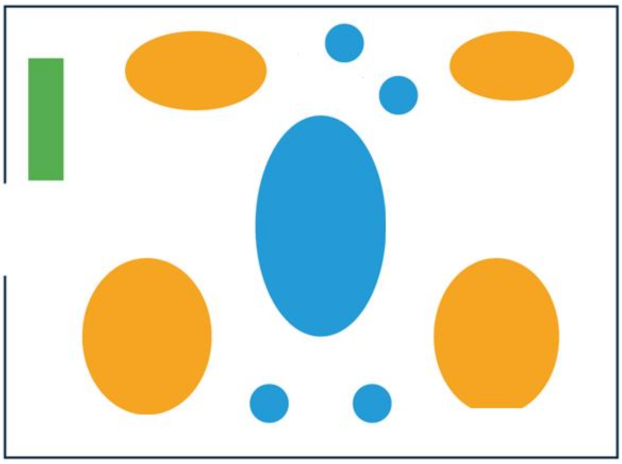


Figure 13. Angular layout

2.1.1.5.7 Free flow Layout

A free flow layout rejects typical design patterns and styles commonly used to influence customer behavior. In a free flow layout, the intent is not to lead the customer using predictable design patterns, displays, or signage. There are no specific design rules followed for this retail store design, and customers have more liberty to interact with merchandise and navigate on their own. For this reason, the free flow layout is sophisticated in its simplicity. Customers feel less rushed in this creative environment. Retail stores look less sterile in the free flow design, and merchandise may seem more intriguing. The only limitation for retailers using this layout is the overall space available, but that doesn't mean that the research on customer navigation behavior and tendencies shouldn't be accounted for as well. The main disadvantage to this experimental design layout is the risk of confusing customers past the point of their preferred behavior and disrupting customer flow.

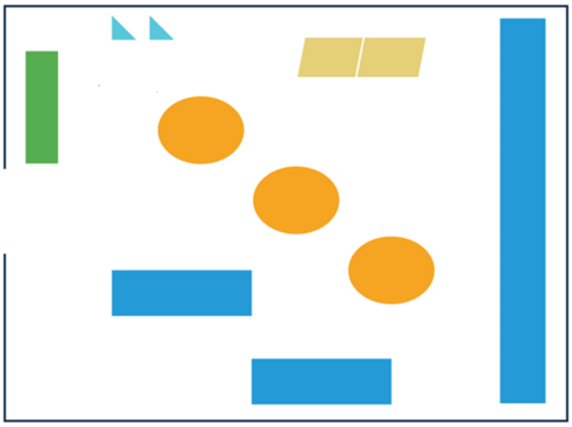


Figure 14. Free flow layout

2.1.1.5.8Mixed Layout

The mixed store layout uses design elements from multiple layouts to create a flexible option for retailers. Department stores use a compelling mix of straight, diagonal, and angular concepts, among other design elements, to create a dynamic flow through a range of departments featuring a variety of merchandise. Large grocery store chains also successfully combine mixed store layout elements. The advantages of combining different store layouts seem apparent, but the space and resource requirements to maintain this design can pose difficulties to retailers.

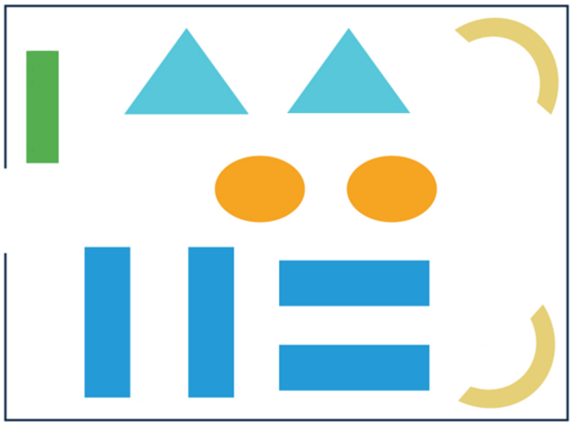


Figure 15. Mixed layout

2.1.1.5.9Geometric Layout

Geometric layout is popular with retailers targeting millennials and Generation Z demographics. A geometric layout offers artistic expression and function when combined with the appropriate displays and fixtures. The unique architecture of some retail stores, including wall angles, support columns, and different ceiling styles mix well with the uniqueness of a geometric layout. Merchandise displays and fixtures of various geometric shapes and sizes. Clothing and apparel stores use a variety of environmental merchandising strategies with the geometric layout to enhance the customer experience.

Figure 16. Geometric layout

2.1.1.5.10 Boutique Layout

The boutique layout (also called shop-in-the-shop or alcove layout) is the most widely used type of free flow layout. Merchandise is separated by category, and customers are encouraged to interact more intimately with like items in semi-separate areas created by walls, merchandise displays, and fixtures. Typically used by boutique clothing retailers, wine merchants, and gourmet markets, this layout stimulates customer curiosity in different brands or themes of merchandise within the overall category. The downsides of the boutique layout include the following factors: (a) Reducing the total display space for merchandise with inefficient space management, (b) encouraging too much exploration of separate areas within the store, and (c) confusing customers past the point of purchasing behavior.

Ultimately, the exploration can distract from customer interaction with the merchandise.

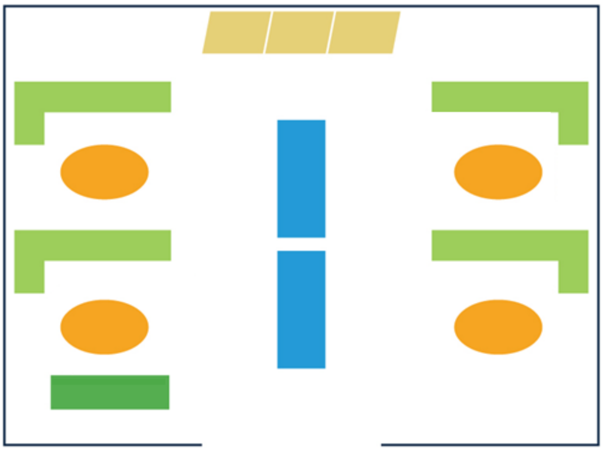


Figure 17. Boutique layout

2.1.2 Artificial Intelligence for indoor positioning

Indoor positioning systems have drawn research attention due to the increasing demands on location-based services in indoor environments (Zou and Jiang, 2016). As techniques such triangulation and trilateration do not perform effectively, fingerprinting is widely used in order to mitigate pervasive indoor multipath effects. The performance of fingerprinting is affected significantly by the device heterogeneity as smartphones, tablets and laptops have different signal receivers. In addition, indoor environmental changes such as the material of products, the number of people browsing the area and the placement of furniture may affect fingerprinting performance, as well. To this end, various approaches that aim to process and remove noise from the received signal strength are utilized to overcome the issue of device heterogeneity.

Towards dealing with fingerprinting challenges, artificial intelligence and machine learning are used to mitigate received signal strength issues regarding device heterogeneity. Machine learning is a field of artificial intelligence and aims at learning by utilizing a vast amount of data, in order to discover patterns and predict an outcome for data received for the first time (Carbonell et al., 1983). Thus, signal fingerprints can be used to extract patterns and reduce faulty measurements.

Machine learning algorithms and techniques such as Neural networks (Galván-Tejada et al., 2014; Mazan and Kovarova, 2015), decision trees (Yim, 2008), k-nearest neighbors (Liu et al., 2007) and random forests (Breiman, 2001 have been used in the literature for indoor positioning. Another well-known technique is Online Sequential Extreme Learning Machine (OS-ELM) which involves feed-forward neural networks for classification (Lan et al., 2009). The most common algorithms and methods for the location comparison and estimation are probabilistic methods, k-NN (k-nearest neighbours), neural networks and Support Vector Machines (Liu et al., 2007). In addition, Decision Trees, Random Forests and Neural Networks are techniques that have been widely used in literature for indoor positioning using Wi-Fi and RFID. Regarding Bluetooth technology, Neural Networks are one of the most popular techniques across the three technologies, as they have been used with Bluetooth technology. Decision Trees, Random Forests and Neural Networks are techniques that have been widely used in literature for indoor positioning using Wi-Fi and RFID. Regarding Bluetooth technology, Neural Networks are one of the most popular techniques across the three technologies, as they have been used with Bluetooth technology.

Deep learning is a modern technique utilized for indoor fingerprinting. Deep Learning often refers to as hierarchical learning is a dynamic algorithm-based technique used for learning several levels of representation as a means of modelling complex relationships among data (Deng and Yu, 2014). Deep learning offers an effective object recognition and location identification in comparison to the classical machine learning pipeline of feature extraction and classification using a different approach like support vector machine (SVM). Deep learning requires the consistent utilization of object position information which is obtained from the relative posture of the person or the object image regardless of the object distance from the camera for effective object detection and location identification (Kumar et al., 2016). In addition, the application of Deep Learning using the CSI of NICs for location identification provides an efficient approach to determining the target location. This efficient location identification approach has resulted in the development of several innovative systems like DeepFi (Wang et al., 2015a), BiLoc (Wang et al., 2017), and PhaseFi (Wang et al., 2015b) among others.

In their work Wang et al. (2015) they built a deep learning mechanism consisting of an offline and online localization phase. In the offline training phase, deep learning is utilized to train all the weights as fingerprints. Moreover, a greedy learning algorithm is used to train all the weights layer-by-layer to reduce complexity. In the online localization phase, we use a probabilistic method based on the radial basis function to obtain the estimated location and effectively reduce the location error compared to BiLoc and PhaseFi.

Machine learning aims to enhance the accuracy of the indoor positioning systems and reduce the required computational cost and time. Goal of machine learning studies is to find the perfect match between user locations from a pre-defined set of captured points. Comparing between the user fingerprint and stored fingerprint is the principle method for positioning, which find the best matching pattern in the signal radio map.

2.1.3 Location Analytics

Location analytics (LA) refers to the contemporary concept of using specialized spatial analysis techniques to understand spatial arrangements, patterns, groupings and relationships in geographically referenced phenomena (Pick et al., 2017). When dealing with retail domain and indoor environments in order to understand shopper behavior, literature highlights the importance of understanding and extracting value from customers' locations for both the enterprises that aim to provide a more personal and compelling shopping experience and the customer that is the user of customer services (Yaeli et al., 2014).

The advent of IoT (Ali et al., 2015) has enabled a series of new capabilities for indoor location-based services and as a series new data sources to process in order to extract value. Sensor data from wearable devices and in-store transmitters and sensors can be further elaborated to extract value and provide new insights (Radhakrishnan et al., 2016). Knowledge that was priory available only for online stores, now is available for physical stores as well. The establishment of analogies between online and physical store facilitates the mapping of these concepts as the physical store is treated as the online store, the link transition is the zone transition in the store, the time spent in a webpage is the time spent in the physical zone and the online checkout is like the store cashiers (Yaeli et al., 2014). Thus, e-commerce analytics can be transferred and applied on physical stores as well, unlocking insights that was priory available only for online stores. Yeali et al. (2014) highlight the importance of the venue information in physical stores so as retailers can optimize their stores' layouts and improve store operations. In their work they use mobile indoor location data for a thorough understanding of customer behavior and present a real-world use case in order to provide smarted physical commerce.

An additional aspect when dealing with location analytics in the classification level of the location information. As presented by (Pick et al., 2017), O'Sullivan and Unwin (2014) provide four levels of location analytics that start from an elementary level of raw data (as they mention "dots on a map"), continue with data analysis that examines relationships among the location data, and conclude to statistical analysis that uses statistical model and spatial modeling that aims to answer questions regarding location data such as customer flows. Similarly to indoor location analytics the classifications levels can be used as depicting the user's position in the store as a dot (level 1), calculating the distance to a zone of interest for the user in the store (level 2), detection of areas of high interest for the users (level 3), and prediction of the users that will visit a particular zone in the next hour (level 4).

2.1.3.1 Shopper Trajectories

Detecting user's trajectory in outdoor environments is much easier as GPS technology facilitates the process and provides accurate results (Yuan et al., 2014). However, indoor environments require the use of other technologies rather than GPS, as it does not perform effectively in indoor environments. For this purpose, technologies such as Wi-Fi, RFID and Bluetooth can be used in order to determine the position of a user in an indoor environment.

Recent approaches involve tracking systems where the cart is track during the shopping trip. In their work (Ferracuti et al., 2019) the authors develop a novel path analysis method and detect the most and least preferred shopper paths in order to explore the real dynamics of the store and its departments. Other approaches (Larson et al., 2005) examine spatial constraints in order to get insights regarding the areas that shoppers visit and detect inaccessible areas due to obstacles or aisles and the preferred areas that shoppers visit.

The knowledge of shopper's in-store position can lead to extraction of the followed path during the shopping session. As a process is like extracting the path followed on an online store by examining the pages that the user visited (Kuo et al., 2005). However, unlike the online environment where the “location” information is more accurate than using a wireless infrastructure in a physical store, the physical stores are characterized by noise in the location data. As a result, the sequence of the location data may be inaccurate. To this end, various approaches have been introduced in the literature in order to tackle this issue. Such approaches include a population-based path relinking algorithm (Ma et al., 2017) that aims at proving effective solutions with a high computational efficiency. In addition, Hidden Markov Models are utilized to analyze sequences and accurately determine the trajectory followed by the user (Popa et al., 2013). Also, the traveling salesman problem (TSP) has been used in order to examine patterns in shopping paths (Hui et al., 2009).

Shopping paths can be further examined as groups and examine their fractal dimensions in order to examine the shopper movement (path) on the purchased behavior [Kaneko and Yada, 2016]. As the authors highlight their research indicated a positive correlation between path length and purchase behavior. In addition, moving patterns and purchase transactions have been also examined by (Tsai et al., 2017) who also highlight the opportunities raised by having the knowledge of where the shoppers moved in the store and what finally bought.

2.1.3.2 Shopper Segmentation

Location analytics can be used in order to detect customer segments of similar behavior. Clustering of customer shopping path in grocery stores leads to the obtainment of movement patterns (Sano et al., 2016). These patterns provide insights regarding demographic features, purchase results and KPIs regarding the time shoppers spend in the stores’ areas. In addition, comparing the movement data to the purchases performed can provide important insights regarding where shoppers go and what they finally buy.

Customer segmentation may be performed by using spatial data (Fan and Zhang, 2009). Thus, relationships between locations and surrounding objects can be performed. Especially in indoor environments, the information regarding locations (i.e. indoor areas) and surrounding objects (i.e. moving customers that are being tracked by a tracking mechanism such as a mobile application) may lead to useful customer segmentation that takes into consideration the areas where customers tend to spend their time during their shopping trip.

Trajectory clustering can reveal useful insights regarding customer behavior. As pointed in the literature several clustering techniques have been used in that often rely on spatiotemporal parameters (Yao et al., 2018). In their study, the authors transform each trajectory into a feature sequence to describe object movements and further employ a sequence-to-sequence auto-encoder to learn fixed-length deep representations. Thus, efficient trajectory clustering may lead to useful insights regarding shopper behavior.

2.1.3.3 Retail KPIs

The knowledge of the position of the shopper can be utilized to extract KPIs and metrics that will facilitate the monitoring and the efficiency of various operations of retail environment. KPIs deriving from the in-store position of the customer may be the number of sales floor zones visited, the basket size, the time spent in a zone area (Kaneko and Yada, 2016). The advent of modern technologies may transform the physical store to an online one (Yaeli et al., 2014). As a result, KPIs priory used only on online stores can now be applied to the physical ones. The time spent in different store areas, the number of different routes followed while browsing or what routes the customer follows can also be considered KPIs for the physical stores. Metrics such as the store's surface coverage becomes available thanks to the tracking ability that technology offers. This information provides insights regarding the percentage of the store that shoppers browse during their shopping trip, enabling this way capabilities for efficient strategies (Silberer et al., 2007; Sorensen, 2017). In addition, the Gross Rating Points (GRP) metric has been developed as a standard metric for advertising exposure and highlight the correlation between frequency and time (Fader and Lodish, 1990). This metrics has been further enhanced and normalized by taking into account the area in square meters for each store area (Ferracuti et al., 2019) and is valued as (People * Average time spent)/m2. This metrics highlights the performance of each area in terms of visit, time spent and the size of the surface. The higher the performance, more traffic is generated for these areas.

Similar studies from web analytics (Gupta et al., 2013; Singal et al., 2014; Waisberg and Kaushik, 2009) can be examined in order to adopt potential KPIs that due to the advent of technology can be applied to physical stores. Thus, metrics such as visits, average time on site and customers per page can be transformed to metrics such as visits to a physical area, average time on physical area, and distinct customers per physical area. These KPIs can be further examined in the way they interact with supply chain performance management (Cai et al., 2009) and define or combine appropriate metrics for the whole supply chain.

2.1.3.4 Indoor positioning efficiency

Despite the popularity of location analytics and location-based services, they both face a similar problem; i.e. the accuracy of the localization mechanism they use. In order to perform effectively, such mechanisms require detecting accurately the position of the user. The issue of effective detection of the user position is encountered both in academic literature and also a business issue when trying to deploy such mechanisms in real world cases.

From the academic perspective, the issue of user-positioning, especially in indoor environments, is an emerging challenge. Although it is not so difficult and complex to locate the exact position of a moving object outside a building (e.g. using GPS signals and geofencing techniques (Kriz et al, 2016)), this becomes a quite challenging task in indoor environments. Retail stores and shopping malls in general provide fertile ground for the development of indoor positioning systems and relevant research works address two major issues: (a) the localization infrastructure shaped by a positioning technique and a wireless technology (e.g. (Jin et al., 2013; He et al., 2016)) and (b) the provided services on the top of the infrastructure.

Bellow we examine the industry insights upon the area of indoor positioning and present the insights we got via interviews with experts and stakeholders of indoor positioning systems.

The indoor positioning issue appears in the relevant literature as a very challenging one and the achievement of high localization precision (accuracy) is a common objective shared by various scholars (He et al., 2016; Shin et al., 2015). Lymberopoulos et al. (2015) argue that the indoor location problem still remains unsolved and stress the importance of the employment of a realistic approach that would counterbalance the desired localization accuracy with low cost.

From the business perspective, research has indicated that despite the advancement of technology, increased costs, additional hardware and complex deployments are required in order to perform efficient indoor tracking. The fact that academics are asked to present approaches that improve indoor positioning accuracy (Lymberopoulos et al., 2015) highlights the fact that even practitioners acknowledge that we cannot have effective accuracy, despite those who claim to have extremely high levels of accuracy. Management information systems contain location information along with the enterprise data, but still lack to exploit this information to extract knowledge. In addition, a semi-structured focus group, along with extra interviews with indoor positioning system stakeholders and experts indicated effective indoor positioning is a challenge that should be dealt.

Both approaches (academic and business) come down to one thing; effective indoor positioning is required in order to extract reliable location analytics and be able to design effective location-based services. In the context of this thesis we propose a machine learning indoor positioning approach that aims at extracting effectively the position of a user in an in-store retail environment.

In order to acquire insights and detect open issues regarding indoor positioning, we conducted interviews, in order to verify the issue that is highlighted by both academia and business.

We conducted interviews with retail experts that utilize and deploy indoor positioning and location analytics software and approaches to extract insights regarding the way they perceive the shopper location and pinpoint open issues regarding indoor positioning. We discussed with professionals and practitioners from the fields of Business Intelligence, Business Analytics and Merchandising who plan on using the location attribute in their routines to support decision-making. Most of them already use platforms and services that provide insights regarding the in-store behavior of their shoppers. However, the highlight the fact that the information they see via the interfaces is significantly problematic in terms of that they cannot use this information, due to accuracy issues. Most of these solutions utilize triangulations which is not an effective technique for wireless infrastructures. In the scope of this thesis we have verified that triangulation is not an effective technique for indoor positioning.

Industry Issues
Positioning accuracy
Triangulation efficiency
Additional hardware to increase accuracy
Unit tracking (trolley-basket or shopper)
Store layout effect
Areas division

Table 4. Indoor positioning issues overview

Other insights rely on the fact that solutions that have significantly better performance in terms of positioning accuracy require additional hardware, complex technology and increased costs. This means that the retailer should install hardware on trolley-baskets and deploy a series of antennas in the store to be able to track the exact position of the shopper. This has an effect on the fact that shoppers tend to leave aside the trolley-basket and continue shopping without it, leading this way to inaccurate results besides the increased accuracy of this approach.

Another issue that was highlighted by the discussion with the experts indicated that the store layout affects the areas division for indoor positioning. Different kind of layouts set different kind of restrictions. The layout per se, may restrict specific areas division. Thus, the examination of store layouts is a factor that should be taken into consideration when performing indoor positioning. Store layouts, also, have an effect on the selection of the appropriate machine learning technique for indoor positioning. For example, instance-based classifiers seem more suitable for free flow layouts, while tree classifiers are more suitable for grid layouts.

2.2 Research Gaps

On the one hand Wi-Fi technology has dominated the field of indoor positioning until now utilizing a vast variety of techniques. On the other hand, Bluetooth technology has become more mature and reliable for indoor positioning and the available studies are increasing. Still, most works are practically lab experiments (e.g. Kim et al., 2015; Wang et al., 2015; Paek et al., 2016; Ke et al., 2018) that haven't validated their indoor positioning approach in a real, fully operational environment. Further, researchers have experimented with indoor localization of subjects moving only in one-floor contexts (e.g. Cabero et al., 2014). Moreover, most studies employ a single positioning technique for performing indoor positioning (e.g. Li et al., 2015; Subhan et al., 2011; AL-Madani et al., 2019; Yohan et al., 2018). They do not compare and evaluate different techniques or even combine them aiming to higher efficiency in indoor positioning. On the contrary, they have experimented with hybrid of different technologies (e.g. Galvan-Tejada et al.

(2013) propose an algorithm to obtain the location of a receiver combining Bluetooth and Wi-fi technologies).

Aspiring to fill these acknowledged gaps in the literature, we have developed and deployed an indoor positioning approach that is applied on two different case studies of shoppers moving in retail stores with different layout characteristics; i.e. BLE beacon and Wi-Fi. We experimented with different placement scenarios of the BLE beacons and we identified the ceiling of the store as the most appropriate one. Next, we applied and evaluated established localization techniques with the purpose to achieve the most efficient position determination of moving customers in the retail store. We applied tri- and multi-lateration, but the results were disappointing and, then, we assessed fingerprinting along with classification methods to determine the customer location at store area-level. We did not consider the separate floors of the store split into equal smaller areas making a grid (i.e. fixed-length surface) to detect the exact store area of the customer. Instead, we identified the constraints of the store's physical layout and the retailer's needs and, thus, we adopted a more realistic, variable-length surface model with store areas of different sizes and dimensions. Similarly, for the Wi-Fi technology we adopt a variable-length surface areas division and utilize fingerprinting (instead of tri-lateration) for efficient indoor positioning.

Since indoor localization in retail environments is still unexplored, this study started by assessing the performance of the most common techniques for indoor positioning (i.e., trilateration and fingerprinting) and established classifiers for Bluetooth Low Energy (BLE) technology. It was found that the random forest is the best classifier. However, this study moved on to proposing and assessing an ensemble filter. The absolute mean and root mean squared errors of the ensemble filter are significantly lower (40.7% and 18% lower, respectively). More specifically, for approximately 70% of our cases (captured events of consumers), the ensemble method results in a localization error of less than 1 m and in 80% of the cases, the localization error is approximately 2 m. On the contrary, for the random forest, in 80% of the cases, the localization error is approximately 2.5 m.

In retail environments, such a deviation is significant, because even 0.5 m away from the actual shopper's position may lead to position him in a different shopping isle and in front of a different store shelve, thus a different product category. Namely, the more accurate localization of consumers, the more accurate and rich insights on the customers' shopping behavior. Consequently, the retailers and the marketing managers will be able to offer more effective customer location-based services

Apart from the research gaps, this study selects Bluetooth and Wi-Fi technologies because it determines accurately user's position compared to RFID technology. RFID is able to track only trolley baskets rather than the shopper per se. Shoppers usually leave their basket trolley and start shopping without carrying it around, leading to being unable to know where the shopper actually is. Thus, RFID cannot provide reliable information and is more important to be able to track shopper's actual location. Therefore, Bluetooth and Wi-Fi technologies are more eligible to enable efficient user tracking (i.e. shoppers using the mobile app) instead of trolley baskets.

Beyond indoor positioning gaps, literature acknowledges the importance and the value of understanding the way that shoppers navigate in the store. The added value from spatiotemporal data is useful in order to extract knowledge for the retailers. Customer segmentation attracts great attention in literature as it is an important tool for the organizations to handle their customers. Similar approaches are used on trying to segment the paths that shoppers follow and thus extract common behavioral patterns among shoppers. As we acknowledge that there is room in handling shopper segmentation with various approaches that can provide value to the retailer, in this paper we introduce a novel approach of segmenting shopper based on the areas navigating during their shopping visit.

In terms of examining in-store behavior and shopping paths, our approach differentiates on the fact that existing solutions utilize tracking mechanisms on trolley baskets. Instead we utilize a mobile application that the user keeps with him/her during the shopping trip. Thus, we avoid the phenomenon of leaving the trolley basket on another aisle and continue shopping on a different aisle. In addition, we aim at identifying the reason that the shoppers enter the store solely by the areas of interest. So far, this process is performed by identifying the need from basket sales data (Griva et al., 2018; Sarantopoulos et al., 2016), however the path to purchase and the areas of interest during the shopping trip may reveal the purpose that the shopper had in mind during the visit.

3 RESEARCH METHODOLOGY

The aim of this chapter is to present the research methodology employed to address the research objectives and answer the research questions. Given our research objective and the research questions, we adopt as methodological backbone the design science paradigm (Hevner et al., 2004) and we consider an indoor positioning system and a machine learning indoor positioning approach as outcome of this study. For collecting data for the various steps of Design Science Research, two different cases studies are selected and presented (multiple case study design). Below both Design Science Research approach and Multiple Case Study design are presented.

3.1 System Design and Implementation

The first step of the research methodology refers to the design and implementation of an indoor positioning system that is responsible for the generation, collection and processing of wireless signal data. The architecture of the system should comprise 3 modules.

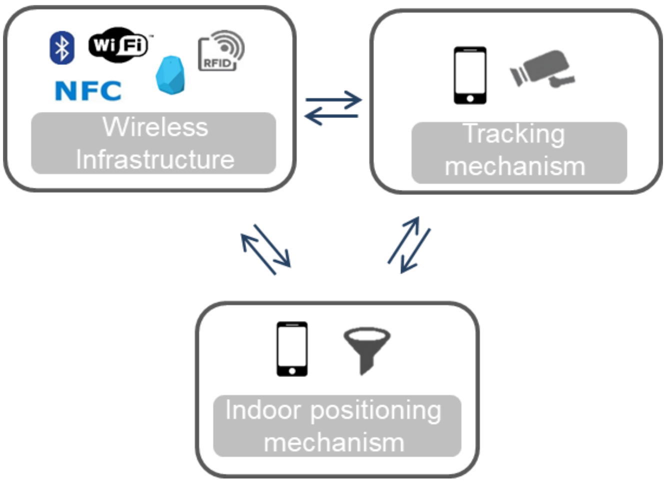


Figure 18. Architecture modules

The first module refers to the wireless infrastructure. The appropriate technology should be selected, and the deployed sensors generate the respective signal data. Following, the tracking mechanism module is responsible for tracking the shopper/user in the retail environment. Such tracking mechanisms may be the mobile phone along with a mobile application that communicates with the wireless infrastructure or cameras that capture the shopper movement. Finally, a module responsible for detecting the in-store shopper position from the wireless signals is required. The indoor positioning module utilizes techniques that perform indoor localization and such techniques involves deterministic or probabilistic approaches. For example, triangulation can be deployed to determine the shopper position using the signals emitted or fingerprinting techniques that based on the captured signals to determine the shopper location.

3.2 Design Science Research Approach

Design Science Research (DSR) is one of the two research paradigms that (Hevner et al., 2004) have recognized. The other research paradigm, called as behavioral-science paradigm, has its roots in natural science research methods and focuses on identifying and explaining the underlying regularities of phenomena or on interpreting human experiences and discourse (Romme, 2003). It seeks to develop and justify theories (i.e., principles and laws) that explain or predict organizational and human phenomena surrounding the analysis, design, implementation, management, and use of information systems. On the other hand, the design-science paradigm has its roots in engineering and the sciences of the artificial (Simon, 1996) guidelines, design principles and technical capabilities through which the analysis, design, implementation and use of information systems can be effectively and efficiently accomplished (Denning, 1997). Such artifacts are not exempt from natural laws or behavioral theories. On the contrary, their creation relies on existing kernel theories that are applied, tested, modified, and extended through the experience, creativity, intuition, and problem-solving capabilities of the researcher (Markus, Majchrzak, & Gasser, 2002). Such artifacts vary from constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems) (Hevner & Chatterjee, 2016; Hevner et al., 2004). (Goes, 2014) highlights the absence of design science research in Top Journals and subscribes to the notion that the IS field needs more design science research. The design science research paradigm increasingly diffuses into the IS community and has gained increasing recognition over the last years (Baskerville, 2008).

In design science, the researcher creates and evaluates IT (Information Technology) artifacts and/or theories intended to solve identified organizational problems. The knowledge base is composed of foundations and methodologies used to develop the artifact. Below we present the basic components of design science research and how are addressed in the current dissertation (Figure 19).

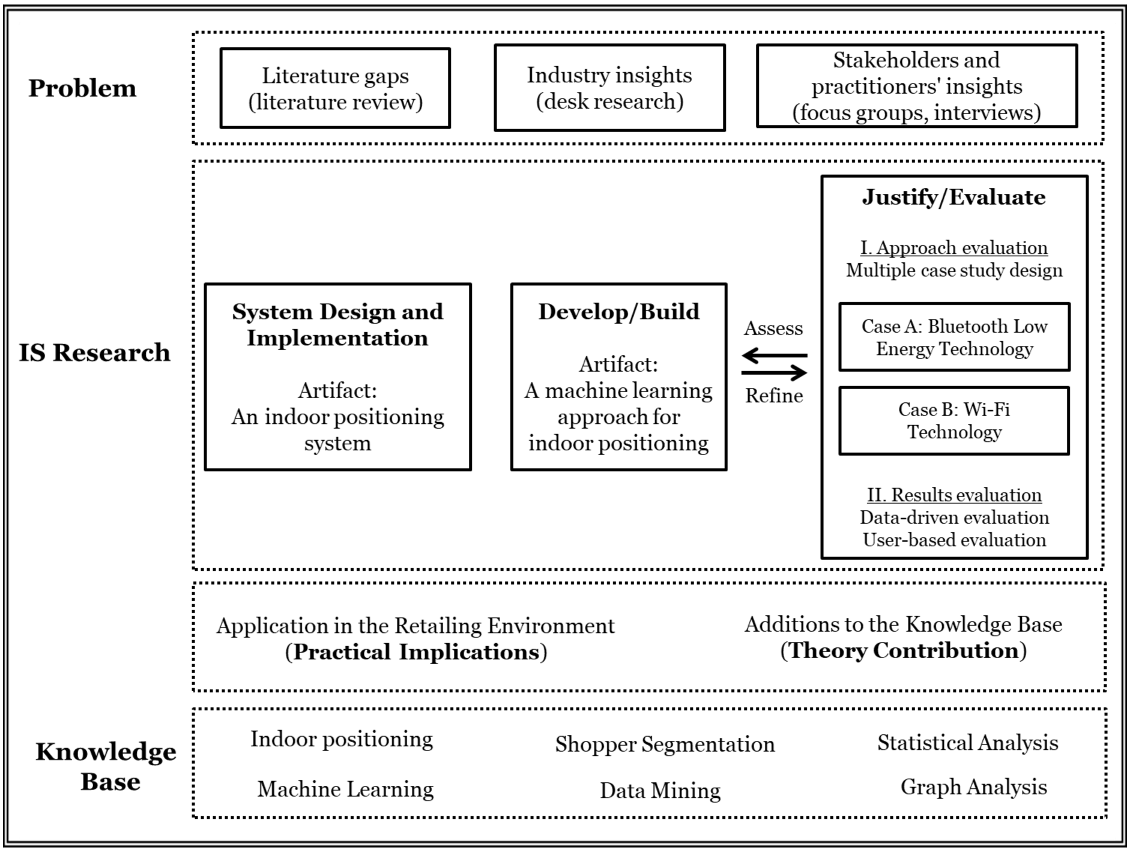


Figure 19. Research approach

(A) Problem definition

This dissertation aims to solve a business problem/need in the retailing environment, which is to perform efficient indoor localization and extract insights in order to extract insights upon the shopper trajectory during the shopping trip. To better define this problem, it follows the steps below:

- **Literature gaps**: To set the research setting firstly we conducted a review of the pertinent literature. This way we specified the research questions which are related with the indoor localization concept and we pinpoint the research gaps and the purpose of this research.
- **Industry insights**: Having laid the foundations upon which this doctoral research will be grounded, then we investigated various open issues and business problems industry people face, when they try to

conceptualize the term of indoor positioning. We found out that the “indoor positioning” is translated into the need of proper level of detection (e.g. X-Y coordinates, store aisle, area, store segment) to identify shoppers’ in-store positions and exploit them in business terms.

- **Stakeholders and professionals’ insights**: Afterwards to better understand and conceptually define indoor positioning in retail domain, we conducted a semi-structured focus group discussion with 22 participants from both retailers and suppliers in order to understand the way they apprehend indoor positioning in the retail content. We further conducted interviews with retail system users that utilize indoor positioning and location analytics software and approaches in order to evaluate the performance of the proposed indoor positioning approach of this thesis.

(B) System Design and Implementation

Then we design and implement an indoor positioning system that will generate and collect wireless signal data upon which we will perform indoor positioning. In this research the developed artifact is system that will facilitate the following artifact of the indoor positioning approach.

(C) Develop/Build

Then we develop and evaluate a technology-based solution that is relevant to the above research problem. In this research the **developed artifact is an approach**, providing a certain manner to handle the appropriate machine learning **to extract the indoor position.**

(D) Justify/Evaluate

Then, we put the approach in practice to evaluate it and realize if it can solve the original problem. This evaluation includes two steps: (i) Approach evaluation and (ii) results evaluation. Regarding the first one owing the lack of prior systematic research on the indoor positioning area, to address this objective the research is based **on multiple case studies design**. In more detail, **the proposed approach is been evaluated** by **applying it into** real data derived from **two case studies.** In both cases we conducted a data-driven evaluation and a user-based evaluation in order to gain insights regarding the business performance of the approach. In the data-driven evaluation we conduct a field study to identify whether or not the shopping journey data can reveal knowledge regarding the purpose of the visit via the purchase data. To do so, we compare the paths followed from 100 shoppers and compare the results to their purchase data. Regarding the user-based evaluation we examine real world data along with indoor positioning system users and assess the quality of the extracted information.

(E) Knowledge base

To build the proposed approach, we used both theoretical foundation and methodologies. Theory knowledge inputs in the developed approach. In more detail in our approach we utilize machine learning and data mining techniques such as clustering and classification, data mining algorithms such as k-means were used to develop the approach. Statistical analysis was used to evaluate the performance of the machine learning algorithms and detect the ones that perform significantly better than the others. Finally, we utilize graph analysis in order to examine shopper paths and shopper segmentation, as we try to detect shopper segments based on the way they navigate into the store.

3.3 Multiple Case Studies Design

Theory building from multiple case studies gained respect as it is suitable for unexplored research areas where it is critical to bring the researcher in close proximity, both conceptually and physically, to the underlying phenomenon, allowing for deeper engagement with the social settings (Fendt & Sachs, 2008). As (Eisenhardt & Graebner, 2007) highlight, “a major reason for the popularity and relevance of theory building from case studies is that it is one of the best (if not the best) of the bridges from rich qualitative evidence to mainstream deductive research.”. Papers that build theory from cases are often regarded as the “most interesting” research (Bartunek, Rynes, & Ireland, 2006).

Selecting cases is an important but difficult aspect of case research. Literature provides some insight into this process (Stake, 1995; Yin, 1994) recommending that the cases should be easy and willing subjects, maximizing what can be learned within limited time. Based on the assertion of (Stake, 1995) “a good instrumental case does not have to defend its typicality”. A good practice in multiple case study design is the cases to follow replication logic. In this regard, although each individual case study represents a “whole” study, in which information is gathered from various sources and conclusions drawn on those facts, the outcomes from one case are compared with the conclusions from the other cases. This indicates that we talk about literal replication expecting that each case shows the same results. (Yin, 1994) proposes the usage of around 2-3 cases for literal replication. The first case can be considered as the pilot case that will help us in deciding the final data collection protocols to be used and the design as a whole. Finally, all the cases can be considered as embedded case studies, as they try to draw conclusions by analyzing sub-units of the study object and not the phenomenon as a whole. The proposed approach is evaluated utilizing two case studies that are applied on two different wireless technologies.

Case A: The first case concerns spatiotemporal data from BLE Beacon technology from a major Greek fast-moving consumer goods (FMCG) retailer. This case was implemented in one retailer's store. During this case, a customized mobile application has been developed in order to provide coupon recommendations to shoppers and also track the in-store browsing of the shopper during the shopping visit. The collected spatiotemporal data that have been collected refer to 100 users. We also obtained their purchase data during the sessions they used the mobile application to accompany their visit. Apart from the data, we obtained the store layout where a study was performed in order to place efficient access point (i.e. BLE Beacon transmitters) to enable to communication of the mobile application with the back-end infrastructure. The store layout of this case comprises two floors characterized by a grid and free flow layout (ground floor) and a pure grid layout (first floor). We utilized the store layouts to deploy appropriately the available BLE Beacon transmitters and identify the areas of interest where the indoor positioning would be performed.

Case B: The second case concerns spatiotemporal data from Wi-Fi technology from a major Greek electronics retailer. This case was implemented in two different stores; while the second store has also been renovated which leads us to treat it as a separate store. Therefore, this case refers to two stores with 5 different store layouts. The first store comprises three floors with grid and diagonal layouts, while the second floor was based on a grid layout and after the renovation it is characterized by a mixed layout composed by grid, free flow, diagonal and angular layouts. During this case we collected data for a time period of two weeks for each store. That led us to collect spatiotemporal data from nearly 900 users that have been connected to the wireless network of the electronics retailer.

	Case A	**Case B**
Domain	FMCG	Electronics
Wireless Infrastructure	BLE Beacons	Wi-Fi
Tracking mechanism	Customized mobile application	Mobile device
Total Users	100	877
Stores	1	2
Different floor layouts	2	5
Layout types	- Grid - Free flow	- Grid - Free flow - Diagonal -Angular

Table 5. Case studies overview

Apart from the collected data, we obtained the layouts of each store in order to study and determine the areas of interest for the analysis. Case study B does not require any special customised mobile application in order to communicate with the infrastructure. On the contrary, this case utilizes solely the connection the wireless network that is offered by the retailer to the store customers. In addition, for this case we did not have available the purchase data of the tracked users, as we process only anonymised data from connections to the wireless routers of each one of the stores.

To evaluate the results of our approach we utilized a data-driven (via a field study) for the BLE Beacons case and a user-based evaluation for the Wi-Fi case.

The data-driven evaluation for the BLE Beacons case involves the analysis of spatiotemporal data from the grocery store to identify the position of the shopping sessions of 100 shoppers. Having the in-store position for each user, we perform location analytics techniques and extract shopper paths and insights regarding their behavioural patterns. We also aim to identify the purpose of the visit based on solely spatiotemporal data and the trajectory followed during the shopping trip. Via the field study, we evaluate the resulting spatiotemporal patterns from the shopping visits and asses their validity by comparing the patterns to the ones occurring by solely the purchase data; thus, we detect an opportunity gap where the areas visited, and the purchases are not overlapping. To achieve this, we exploited two different means i.e. a **mobile app** and the **sales data**. While users shopped and navigated in this store, they used a custom mobile application which disseminated various coupons and was also communicating with the BLE Beacon wireless infrastructure that collected spatiotemporal data during the shopping visit. Via this case study we proved that shopper trajectories indicate the purpose of the shopping visit and also are able to detect opportunity gaps and missing sales, as shoppers tend to spend time to areas of interest and finally do not purchase products from these aisles.

The user-based evaluation for the Wi-Fi case involves the obtain of spatiotemporal data from Wi-Fi routers in order to turn them into indoor positions. Following, we turned these data into heatmap information and KPI representation. The users of indoor positioning systems and software from the electronics retailer evaluated the displayed results and the outcome was found to be effective and could provide enormous value in their store operations.

4 AN INDOOR POSITIONING SYSTEM AND A MACHINE LEARNING APPROACH FOR INDOOR POSITIONING IN IOT ENVIRONMENTS

Our goal is to perform effective indoor positioning in ambient retail environments that are characterize by signal attenuations due to obstacles, store layouts, product packaging and passing by shoppers. To this end, we propose a system artifact and a machine learning approach that aims to detect the position/area that a shopper is located in the store. This approach processes spatiotemporal data originated from various sources and technologies, such as Wi-Fi, Bluetooth and RFID and extracts the location of the browsing users in the retail environment. To this end, an indoor positioning system is required in order to generate data and support the indoor positioning approach.

The approach is summarized to the following phases/layers: (a) data and setting understanding and preparation, (b) modeling and evaluation (c) results translation. The input of the approach is mostly spatiotemporal events and also the store setting that needs to be examined. In addition, various factors may affect the input of the indoor positioning approach, such as obstacles (e.g. walls, promo stands) or product packages and passing customers that affect the transmitted signals.

The output of the approach is the detected indoor position of the moving user into the store. Such knowledge (i.e. the in-store shopper position) can be used to support decision making. Also, various marketing actions or location-based services can be designed upon the knowledge of the in-store shopper position and provide benefits for both the retailer and the customers. The originality of our approach is embodied to the last phase/layer where we translate the indoor positioning results into knowledge to communicate them to the experts.

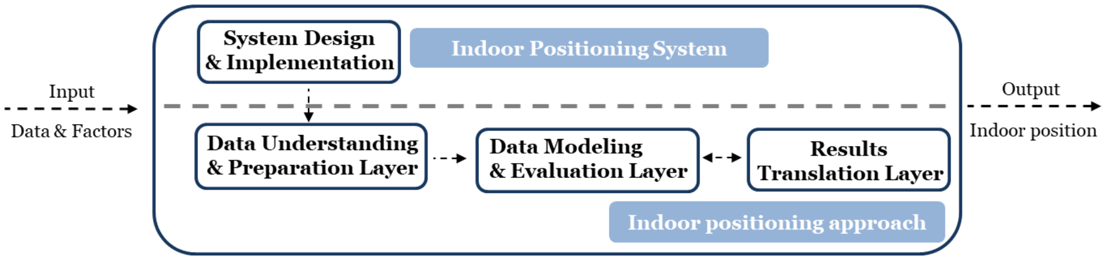


Figure 20. Indoor positioning system

Following, we analyse each layer of the approach. The originality of our approach is embodied in the "Modelling" phase (marked with red in Figure 21) where we employ a filter selection for indoor positioning using an ensemble filter with the best performing algorithms. This phase includes: (a) areas division, (b) fingerprinting, (c) data modeling, (d) classification evaluation and (e) filter selection. Below, we summarize the steps of our approach.

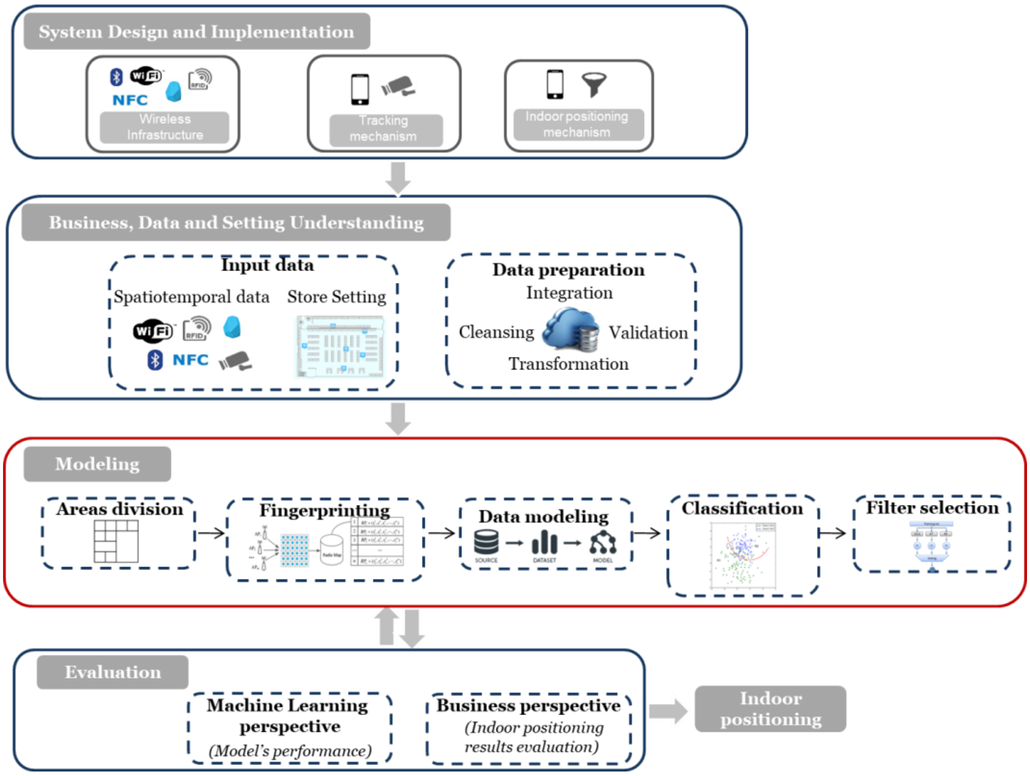


Figure 21. Machine Learning approach for indoor positioning

4.1 System Design and Implementation

The first phase when performing indoor positioning is the design and implementation of the appropriate mechanism that will generate, collect and process signal data from IoT transmitters. The first step requires the selection of the appropriate wireless infrastructure technology (e.g. Wi-Fi, Bluetooth, RFID). Wi-Fi and RFID are the most mature technologies, while Bluetooth and specifically Bluetooth Low Energy (BLE) Beacons are emerging technologies that are expected to be widely adopting for indoor positioning purposes. In addition, BLE Beacons enable additional capabilities regarding location-based services and personalization.

Following to the wireless infrastructure, a tracking mechanism is required in order to associate the infrastructure and the user of the service. When wireless signals are used, a mobile application is a suitable mechanism as it can communicate with the wireless signal (e.g. Wi-Fi and Bluetooth) and via a mobile application implement a location-based service.

Finally, as indoor positioning is a complex issue, the appropriate positioning mechanism is required. The most widespread indoor positioning techniques are triangulation and fingerprinting, while several approaches are used to improve indoor positioning accuracy as described in section 2.

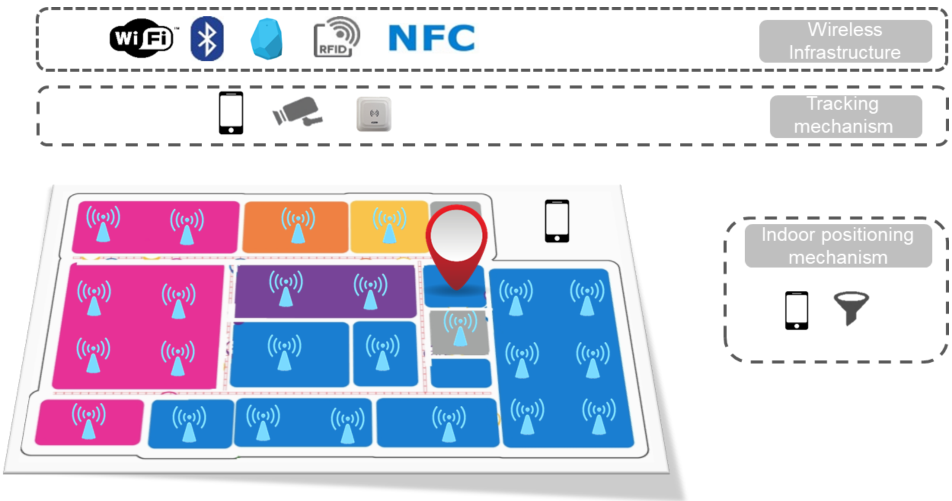


Figure 22. System architecture layers

4.2 Business, Data and Setting Understanding and Preparation Layer

The business goal is to detect efficiently the in-store position of the shopper and determine the appropriate areas that indoor positioning will be performed on. We perform indoor positioning by examining spatiotemporal data from wireless infrastructures such as Wi-Fi, Bluetooth and RFID that are generated by the movement of the user in the store. Apart from the spatiotemporal data, the input data include the setting of the environment; i.e. the map of the store the approach will be applied.

Spatiotemporal data are by nature noisy. This means that just collecting them does not lead to efficient results. To this end, it is required that the data are stored into a data repository (Jagadish et al., 2014). Spatiotemporal data require synchronization and integration to establish a consistent dataset, as data are collected from various sources and channels. Following, data cleansing is necessary for detecting and correcting or removing errors and inconsistencies of the data to improve data quality (Rahm, 2000); and execute data transformations, such as transforming the obtained data from unstructured forms to structured ones.

Data quality is a crucial factor for accurate insights and meaningful value for the stakeholders. When dealing with spatiotemporal event data, the encountered challenges are crucial and should be tackled in order to extract accurate insights. Data quality challenges involve signal fading or attenuations that affect the signal strength leading to problematic data records. Obstacles and human bodies moving in the store can affect the quality of the signal strength and, also, product packaging as the signal emitted is affected by the product materials. Another data challenge refers to the effective floor detection. Most retail stores consist of more than one floors, thus the signals emitted may interfere to the other floors, apart from the one that the access points are placed. In addition, connectivity issues may lead to incomplete collection or storage of the captured data. This phenomenon may occur, as users tend to disconnect from the network due to interruptions or time out session expirations or they may close the mobile application in the case of BLE beacons infrastructure. In such cases, the user session may be terminated unexpectedly or be incomplete.

Following the data quality challenges, the data processing challenges are crucial in order to acquire efficient results. Data cleansing is an important phase of data preparation for processing and the problematic records should be removed or fixed. The quality of the cleansed data will affect significantly the outcome of the data extraction process. After the application of the data cleansing, the appropriate models and algorithms should be applied on the data. The selection of models and algorithms is also a challenge, as different purposes require different algorithms. Supervised (classification) or unsupervised (clustering) algorithms are applied depending the requirements of the analysis and the problem that is being solved. Finally, an important challenge is the interpretation of the findings. Even though the algorithms perform effectively, the finding may not lead to meaningful insights; thus, another algorithm should be examined in order to lead to meaningful outcomes. Nevertheless, a few times, the value creation is a significant challenge as the findings may not lead to a meaningful value regarding the analysis of the data.

More specifically, in the indoor localization process of our case studies we encountered outlier values that we eliminated in the cleansing process and, also, applied additional logic in order to perform efficient floor detection. In addition, we selected supervised learning to create models for location prediction and clustering in order to detect common behavioural patterns from users' movement. The interpretation of the findings played an important role to select the appropriate number of clusters for the pattern detection, so as the findings are interpretable and meaningful.

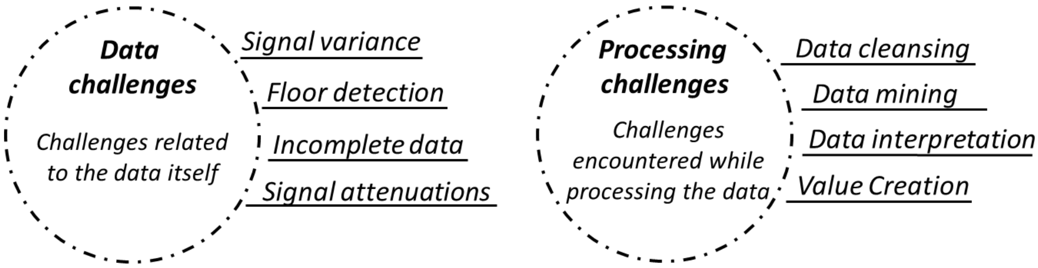


Figure 23: Data and processing challenges

The data processing level is responsible for transforming the location data in sequences of locations and thus extracts the trajectories that the users followed during their shopping sessions in the retail stores. As the trajectories may be incomplete due to potential connectivity issues, we address the issue as the network completion problem [Kim and Leskovec, 2011], and utilize techniques to complete the problematic parts of the user trajectories.

Finally, we end up with data validation after each of the above steps to consolidate the data integrity of the available datasets based in ad hoc criteria selected by the researchers.

4.3 Data Modeling and Evaluation Layer

This phase includes the following tasks: (a) areas division, (b) fingerprinting, (c) data modelling, (d) classification evaluation and (e) filter selection. Below, we summarize the steps of our approach.

4.3.1 Modeling

4.3.1.1 Areas division

The critical part of this process is to identify the areas of interest, upon which indoor positioning will be performed. Thus, apart from solely technical parameters, it is crucial to define these areas along with the stakeholders and decision makers. It is an important part of the process, as area segmentation is an open issue and various approaches are proposed in the literature. In our approach we start by segmenting the areas into the smaller possible segments and then, if the accuracy is not satisfying, start merging segments to form bigger and more efficient areas.

4.3.1.2 Fingerprinting

During fingerprinting a collection is created that contains the generated signal strengths for each possible location of the environment, and then each incoming signal is compared to the initial collection (training set) to detect the location it is referred to (Wang et al., 2015). This technique is comprised of two phases; i.e. training and localization. During the training phase, location information is gathered in order to correlate locations with data regarding signal strength or distance from a transmitter. This face is considered as the offline phase and localization stage as the online phase. During the localization phase the new dataset is collected and compared to those gathered during the offline phase to determine the location.

As fingerprinting is performed using signal strength, the problems that may occur are generated due to signal variations. These variations derive from communication issues such as fading, interference or even from environmental factors. Thus, in some cases, re-training may be required. Furthermore, fingerprinting is widely used despite its pre-deployment complexity, which required huge effort in terms of time. Hence, researchers have tried to reduce that time-consuming phase, which involves the collection of signal strength samples (Hossain et al., 2015). Moreover, this technique is mostly preferred when dealing with Wi-Fi technologies (Yim, 2008). The most common algorithms and methods used in order to estimate a location, based on fingerprinting include probabilistic methods, k-Nearest Neighbours (k-NN), Neural Networks, and Support Vector Machines (Liu et al., 2007). Thus, at this task is important to extract the fingerprints for the desired areas.

4.3.1.3 Data Modeling

The collected data by the fingerprinting task require further processing in order to able to extract value for the tasks that follow next. Fingerprinting stores data in forms of signals and captures information regarding the transmitting access points. In our case we model the collected data in the form described in Table 6. This format enables to have for each area the top 6 closest access points (AP) along with the distance and the signal strength.

	AP id data			Distance (meter)			Signal Strength (db)		
Area	**AP_id_1**	**.**	**AP_id_6**	**Distance_1**	**.**	**Distance_6**	**RSSI_1**	**.**	**RSSI_6**
<area>	<id1>	.	<id6>	<dist 1>	.	<dist 6>	<dB1>	.	<dB6>

Table 6. Modelling format

4.3.1.4 Classification

As our purpose is to determine shopper location, this task aims at experimenting and evaluating the localization techniques used. The most common algorithms and methods for the location comparison and estimation are probabilistic methods, k-NN (k-nearest neighbours), neural networks and Support Vector Machines (Liu et al., 2007). To this end, the classification algorithms we have decided to evaluate are the following: (i) Naïve Bayes (Lewis, 1998), (ii) Support Vector Machines (Hearst et al., 1998), (iii) Logistic Regression (Anderson, 1982), (iv) Decision Trees (J48 algorithm) (Quinlan, 1986), (v) Neural Networks (Multilayer Perceptron model) (Kosko, 1992), (vi) KStar (or K*) (Cleary and Trigg, 1995), (vii) Random Forests (Breiman, 2001).

Decision Trees, Random Forests and Neural Networks are techniques that have been widely used in literature for indoor positioning using Wi-Fi and RFID. Regarding Bluetooth technology, Neural Networks are one of the most popular techniques across the three technologies, as they have been used with Bluetooth technology. Thus, this task selects and evaluates the selected algorithms.

4.3.1.5 Filter Selection

We propose a hybrid solution, as we have observed different performances when dealing with different areas of the store; for some areas, a few classifiers performed better than the others and vice versa. Therefore, instead of using solely one classifier, we decided to form a hybrid approach that involves a voting process that takes into consideration multiple classifiers.

The hybrid method is actually an ensemble method (Dietterich, 2000; Kotsiantis et al., 2006) which involves learning algorithms that form a set of classifiers. We chose the three significantly better algorithms based on the F-measure metric to form the ensemble classifier, which classifies new data by taking a weighted vote of the selected algorithms' predictions.

4.3.2Evaluation

Here, we suggest that the resulted indoor positioning results should be assessed in both business and technical terms. On the one hand, a group of industry experts should assess the validity of the results based on their accumulated experience. If they defy them, we should re-execute the analysis after changing the input dataset.

For communicating the results of our approach to the business experts, we depicted the position results on a map. Following, we captured browsing in the store and then examined the results on the map. If the process does not provide efficient results we should start over, define new areas, select new classification algorithms and perform fingerprinting from scratch.

In terms of technical evaluation, we need to test the classification model's performance. We use the metrics of accuracy, precision, recall and f-measure (or F1 score) (Manning et al., 2008). Accuracy measures the number of correct classifications performed by the classifier. Precision indicates the exactness of the classifier, meaning that higher and lower precision leads to less and more false positive classifications respectively. Recall measures the classifier's completeness. Higher and lower recall means less and more false negative classifications (the data are not assigned as related to an area, although they should be) respectively. Precision and recall are increased at the expense of each other. Thus, they are combined to produce the weighted harmonic mean of both metrics, which is the F-measure.

The evaluation of all algorithms is performed via 10-fold cross-validation (Kohavi, 1995), where the original data are randomly divided into ten equal subsets. Of these ten subsets, one is retained as the validation test of the model and the remaining 9 are used as the training data.

4.4 Results Translation

In order to communicate the results of the indoor positioning approach we visualize the outcome on a store map. Thus, experts are able to assess the quality of the depicted information. This step is important as business decision makers should be able to understand and draw decision upon this information.

To enhance the quality of the provided information, we further suggest the calculation of additional descriptive statistics and KPIs. Such metrics may be the average time spent in an area, the unique customers each area receives and the total visits towards this area. These metrics form a first level of analysis regarding the way that shoppers behave during their browsing into the store. Several actions can be designed or evaluated based on the average time that shoppers spent in certain store areas or the popularity of an area. In addition, building upon the location information, shopper paths can be formed as a sequence of in-store locations. The paths shoppers follow can provide insights and useful information upon the shopper in-store navigation. Such information may reveal opportunities regarding the most popular shopper paths or issues that require further investigation.

Decision makers and stakeholders may investigate the information extracted from spatiotemporal data and evaluate the extracted results. The business evaluation of the positioning approach may be an indicator regarding its performance. The positioning approach may achieve effective performance in terms of accuracy, precision, recall and f-measure, however it is important to extract meaningful results. that help understand shopper behavior and support decision making.

5 CASE A: APPLICATION OF INDOOR POSITIONING USING BLE BEACONS TECHNOLOGY

5.1 Introduction to the case study

The selected operating environment of the indoor positioning system is a grocery retail store. A grocery retail store is a challenging environment due to the presence of technical factors affecting the localization accuracy such as the positions of the transmitters, obstacles (e.g. walls or product signs), and product materials that interact with the signals emitted from them or even the consumers that are interposed between their own receivers and the transmitters . The selected store comprises two floors (i.e. the ground floor and the first floor) with different layout characteristics, where 81 BLE Estimote beacon transmitters were deployed in total. Moreover, a customized smart phone application was developed to track the position of the consumers.

In this context we developed and deployed an indoor positioning system that consists of a BLE-beacon infrastructure and a customized filter that adapts to the specific contextual store requirements, along with a back-end cloud infrastructure that processes and stores the data from the captured mobile device positions. Based on thorough experimentation, it turned out that the proposed system achieves high localization precision, has low deployment complexity, is scalable because each store acquire its own localization model utilizing fingerprinting during the training stage and is responsive because the positioning algorithm can be performed within the time limits suggested by the business requirements.

5.2 System Design and Implementation

For Case A we designed and implemented an indoor positioning system that provides the location of the user. The system relies on Bluetooth technology and in specific on transmitter devices called beacons that emit radio signals to nearby devices. These signals contain information regarding the identifier of the beacon and other data such as its RSSI strength. Beacons are a sufficient technology that can be used for this purpose.

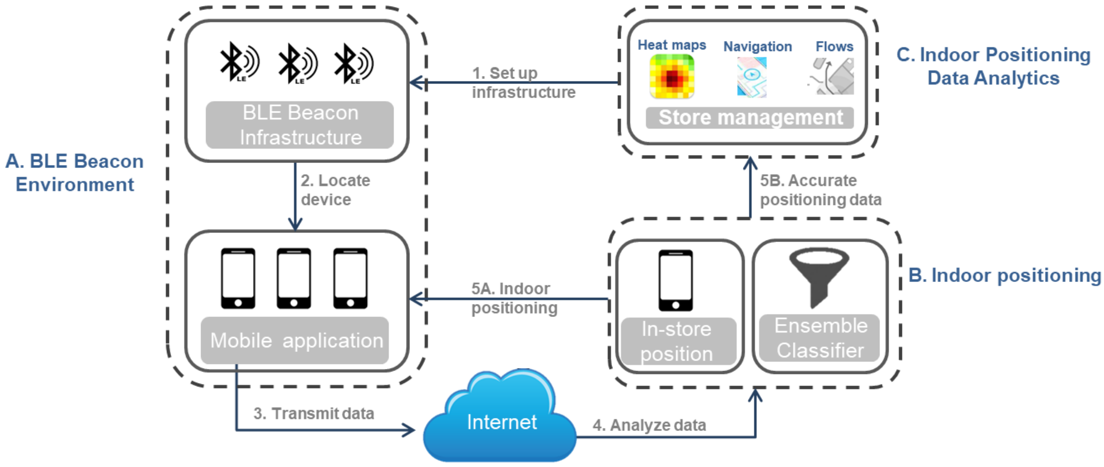


Figure 24. BLE Indoor positioning system overview.

In order to perform indoor positioning using BLE beacons the positioning system's overview is presented in Figure 24. The system consists of the following three major modules: (A) the BLE beacon environment, (B) the indoor positioning mechanism and (C) the indoor positioning data analytics that stores and extracts knowledge from collected data and provides the management of additional location-based services.

More precisely (A) the BLE beacon environment (setting and map) consists of the following modules:

(I) The BLE Beacon Infrastructure layer refers to the beacons that are deployed in the store and the data that are transmitted and captured by the mobile applications of the consumers. It's practically the setting upon which the system is deployed. The BLE beacons can be deployed in various ways in order to achieve the required purpose (e.g. indoor positioning, proximity marketing etc.). Depending on the BLE beacon manufacturer and the available sensors on the beacon device, the environment can provide various form data, such as the id of the beacon device, the strength of the received signal (RSS), the distance between the beacon and the mobile device of the user and information such as light strength, humidity or temperature.

(II) The Mobile Application layer which utilizes an application that captures beacon data during the in-store movement of the user and communicates them via the internet to the backend infrastructure that stores the received data from the BLE beacons and the id of the mobile application. In addition, the mobile application receives the instore position module of the backend infrastructure each time that a request is sent with the captured information from the BLE beacons.

(B) The indoor positioning and mobile app module is responsible for two tasks; i.e. (i) the in-store position of the consumer that communicates with the mobile application to provide the current position and (ii) the positioning filer which processes the collected data from each consumer to provide accurate positioning data for consumer traffic heatmaps and flows visualization. The data captured by the mobile application are stored in the backend infrastructure for further processing, as well as for the generation of the appropriate meta-data. The data from the backend infrastructure are used by other applications or services such as consumer traffic heatmap reports, user navigation and flows visualization (i.e. the administrative and data analytics module).

The (C) indoor positioning data analytics module is responsible for extracting insights and knowledge from the location data that are captured by the mobile devices of the users and stored in the system's infrastructure. This module addresses the appropriate location-based services to the consumers. Essentially, as the consumer navigates in the store environment, the mobile device communicates to the back-end infrastructure the signals it captures from the BLE beacons emitted in the store. These data are processed by the positioning filter in order to determine the consumer position and transmit back the position to the device along with the appropriate service that fits best the position of the consumer (see Figure 25).

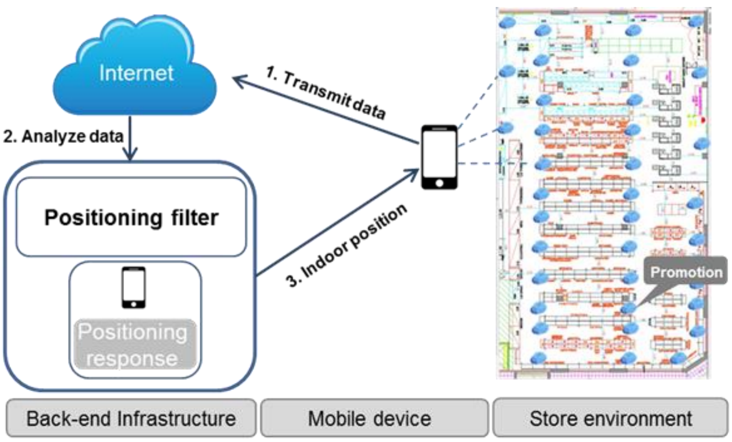


Figure 25. Location-based service overview.

5.3 Business, Data and Setting Understanding and Preparation Layer

The proposed approach was deployed in an area of two floors of 1200 m^2 each. The ground floor was divided in 23 areas and the first floor in 30 areas based on the criteria of technical efficiency and business interest. The selection of the areas has been performed on the basis that the next step of the process is to produce consumer traffic area heatmaps regarding the movement of the consumers within the grocery store. Thus, the stakeholders of the grocery store determine the area selection based on their business needs. Beacon placement was performed in order to create virtual perimeters for each area so as the location of the consumer to be determined more efficiently. The literature proposes several approaches to determine an appropriate transmitter placement within an indoor environment (Chawathe, 2008; Domingo-Perez et al., 2016). To decide the most efficient way in order to deploy the beacon transmitters we had available, we examined a series of combinations in order to detect the most appropriate one.

We began by placing the transmitters on top of the store aisles. However, a lot of problems emerged that prevented the capture of the beacon signal, due to obstacles and products that clerks placed over the transmitters during the replenishment process, which covered them. To overcome this issue, we decided to place the transmitters on the ceiling. The outcome of this placement was more encouraging, as it is less possible for beacons to be covered with products, thus preventing their signal from being efficiently emitted. However, in this case we had to encounter some incidents where beacon signals from the ground floor had been detected, while the consumer was in the first floor, since the consumer's device was located closer to the ground-floor receiver. Figure 26 presents the final deployment of the beacons in the ground floor and Figure 27 presents that of the first floor. Each transmitter is denoted with a red circle. The circles filled with red colour depict the elevation beacons, which are used to facilitate the floor transition. This deployment facilitates the moving object to be visible by all surrounding transmitters and thus the indoor position technique relies on more accurate input data in order to determine the object location. Moreover, at this point we should mention that in order to capture the location of each consumer, we utilized a mobile app that has been installed in his/her Android smart phone.

The BLE Beacon data that are transmitted by the beacon infrastructure and captured by the mobile app during the user's movement and are pushed to the mobile application layer which is responsible for determining the user's position. The data captured by the mobile application (every 0.5 second) are stored in the enterprise cloud for further processing and the generation of appropriate meta-data. The data from the backend infrastructure are used by other applications or services such as heatmap reports, user navigation and flows visualization.

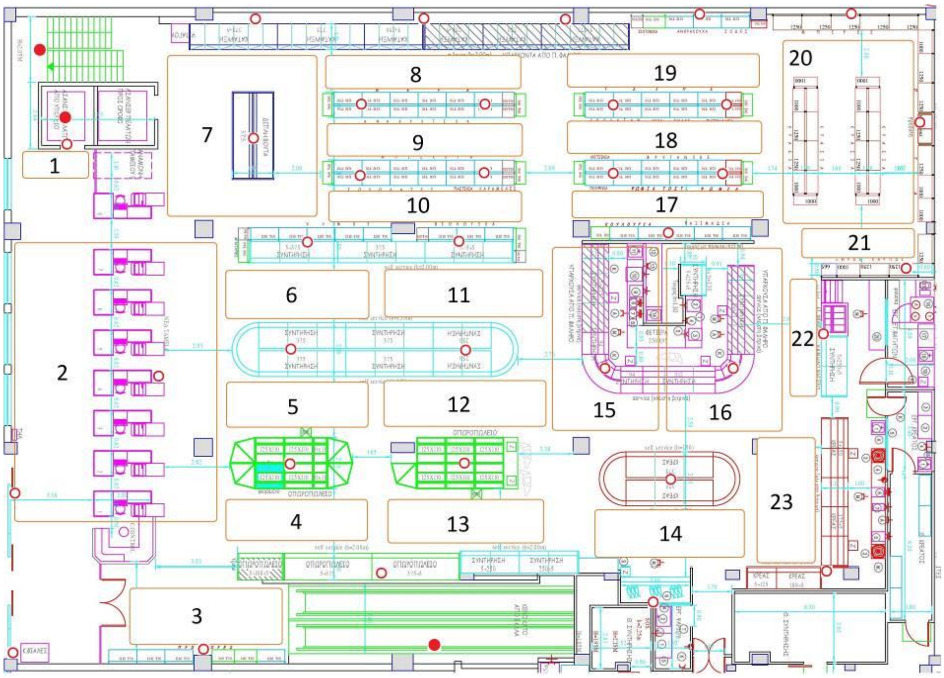


Figure 26. Detailed map of the ground floor of the Grocery Store.

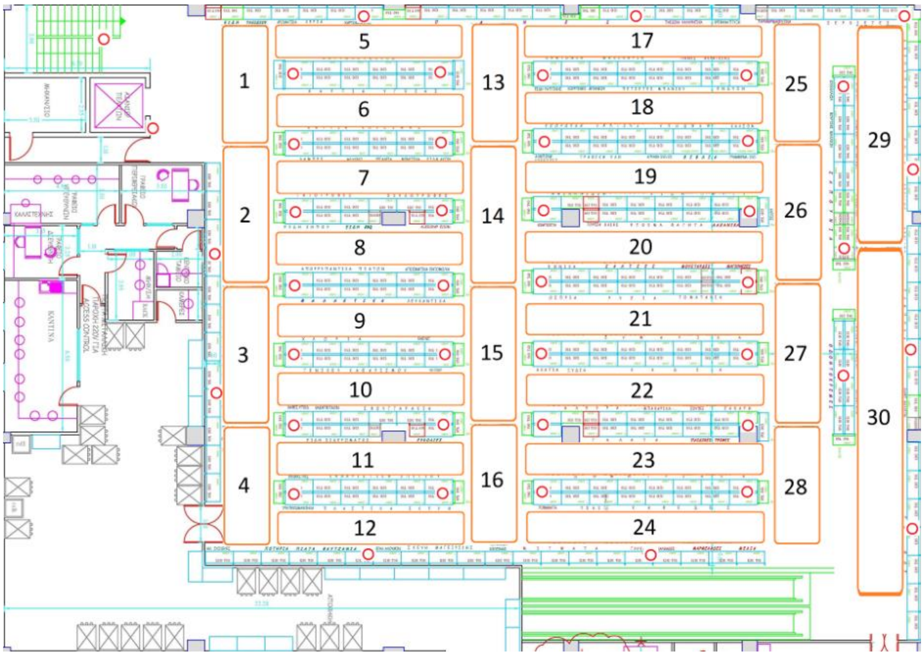


Figure 27. Detailed map of the first floor of the Grocery Store.

In the context of retailing, each shopping visit is recorded as a user session, which contains all the data gathered from the user device. The backend infrastructure is programmed to capture transmissions from nearby beacons every 500 milliseconds, in order to determine user's position. These snapshots are called events and contained in each session. Each event contains for each beacon that is captured by the application, its ID, the distance between the mobile device and the beacon and the strength of the signal that the device receives. The application is able to capture and store more than 6 beacons; however, based on our experiments we have decided to keep up to 6 beacons. Finally, each user session is stored in a NoSQL database in the backend infrastructure.

For each user session (i.e. a shopping visit), all the captured events are stored in it. As mentioned earlier, we keep the six closest beacons to user's device to be able to determine user location. We have experimented with a various number of beacons to estimate location. Although, six beacons seemed the more effective combination and it is aligned with the best practices described in Estimote's forums[1].

Furthermore, we store additional information for each event, which includes:

- the identifier of the beacon that is a unique number assigned to the transmitter (Beacon ID)
- the signal strength that the application captures from the specific beacon (RSSI) in order to determine how close or far is the mobile device from the transmitter, and
- the distance between the transmitter and the moving device.

An indicative snapshot of the dataset format is depicted in Figure 28. The collected dataset for the calibration of the indoor positioning system comprises of 7000 observations and will be discussed later in the paper.


```
{
  "events" : {
    "-KWZ0BHgQ5YzsmZY2Pxh" : {
      "beacons" : [ {
        "distance" : 3.17984020180828,
        "major" : 1,
        "minor" : 137,
        "rssi" : -94,
      }, {
        "distance" : 4.287486797775197,
        "major" : 1,
        "minor" : 136,
        "rssi" : -98,
    },
    "-KWZ0BgFz7XvPHNKlik_" : {
      "beacons" : [ {
        "distance" : 3.772642378975824,
        "major" : 1,
        "minor" : 137,
        "rssi" : -96,
      }
    }
  },
}
```


Figure 28. BLE Beacon dataset format

[1] https://forums.estimote.com

5.4 Modeling

We decided to examine the fingerprinting approach (Wang et al., 2015) and change the granularity level from x-y coordinates to area level. More precisely, we determined to utilize the signal strength and the nearby beacon IDs to determine the location of a user. As mentioned in the "Indoor Positioning Techniques" section, fingerprinting consists of an offline and an online phase.

For the offline phase, we gathered a dataset that establishes the connection between store locations and the corresponding signal strength, as well as the distance from a transmitter. Two researchers have been moving in the store for two hours generating the required movement content using two mobile devices. The areas covered were those presented in Figure 26 and Figure 27. The first device was a Samsung Galaxy J5 and the second was a Samsung A3 and both devices were using Android version 6.0.1. We gathered 7.000 observations that represent the simultaneous movement of two mobile phones in our experiment. The collected data were transformed into a single table where every row is a single observation and relates the store's area with Beacon IDs, Distance from each transmitter and the Signal Strength as depicted in the next figure.

	Beacon id data			**Distance (meter)**			**Signal Strength (db)**		
Area	**Beacon_id_1**	**…**	**Beacon_id_6**	**Distance_1**	**…**	**Distance_6**	**RSSI_1**	**…**	**RSSI_6**
12	139	…	138	2.24	…	4.92	-90	..	-100
24	164	…	203	1.06	…	4.87	-81	…	-101
….	…..	…	…	…	…	…	…	…	…
5	152	…	149	1.67	…	4.85	-86	…	-99

Table 7. Dataset structure

During the offline phase, we exploit machine learning techniques to train our model in order to predict the area where the user is located. Based on our approach, the dependent (class) variable is the Area and the independent variables (features) are the measurements produced by the Beacon IDs, the Distance and the Signal Strength. We have further experimented with the algorithms by examining the most efficient number of beacons to take into consideration.

Area size (m^2)	**Number of areas**	**Store Areas**	**Floor**	**Observations**
57.5	1	20	Ground	154
50	1	7	Ground	150
44	1	2	Ground	165
33	2	15,16	Ground	143
28	1	30	First	162
19.5	1	29	First	148
19	1	23	Ground	80
16	16	5-12,17-24	First	146
15.5	6	8-10,17-19	Ground	115
12.5	8	3-6,11-14	Ground	123
9.5	12	1-4,13-16,25-28	First	75
5	2	21,22	Ground	71

Table 8. Number of observations per area ranked by area size

The data supplied to the different machine learning algorithms is the data set structure exhibited in Table 8. Area is the dependent variable and has 53 different class labels that correspond to specific regions of the store (as Figure 26 and Figure 27 depicts). The selected data-driven approach requires the design of an efficient sampling process with regards to increasing the internal validity of the experimentation. It was decided to collect sufficient observations for all areas in the store based on its surface in m2, thus larger areas would be represented though a larger number of observations as presented in Table 8. The area sizes vary from relatively large (57 m^2) to small (9m^2), the first floor is characterized by a gird layout and at the same time the ground floor was a mix of grid and open layouts. In the forthcoming section, we present the classification algorithms we have used and compare their classification capabilities.

5.5 Classifiers' Evaluation

As mentioned in research background, the most common algorithms and methods for the location comparison and estimation are probabilistic methods, k-NN (k-nearest neighbours), neural networks and Support Vector Machines (Liu et al., 2007). To this end, the classification algorithms that have been used to address the indoor positioning problem are the following: (i) Naïve Bayes (NB) (Lewis, 1998), (ii) Support Vector Machines (SVM) (Hearst et al., 1998), (iii) Logistic Regression (LR) (Anderson, 1982), (iv) Decision Trees (C4.5) (Quinlan, 2014), (v) Multilayer Perceptron Neural Networks model (MLP) (Kosko, 1992), (vi) KStar (K*) (Cleary and Trigg, 1995) and (vii) Random Forests (RF) (Breiman, 2001).

Statistical models (e.g. Naïve Bayes), Decision Trees (e.g. C4.5), Analytical Models (e.g. Support Vector Machines), Instance based models (e.g. K*) and Neural Networks (e.g. MLP) are representative algorithms from the machine learning field and have been employed in the relevant literature as an enabler to achieve higher localization performance results. Therefore, we decided to examine these along with other approaches. Naïve Bayes was selected as it is one of the simplest classification algorithms with strong independence assumptions between the features. Additionally, we selected Logistic Regression to examine the classification problem as a generalized linear model and KStar, an entropy-based algorithm, to investigate their performance in such classification cases. The software used for the experimentation with the algorithms is Weka (Witten et al., 2016). The focus of prior relevant research works is the improvement beacons accuracy in terms of meters while utilizing small amounts of equally sized surfaces. Our study utilizes a variable-length surface to perform indoor position in area-level detection, thus a direct comparison to existing studies is not applicable due to different settings and environments.

5.5.1 Classification Performance Comparison

In this section, we compare the efficiency of each classifier based on the metrics of accuracy, precision, recall and f-measure (or F1 score) (Manning et al., 2008). Accuracy measures the number of correct classifications performed by the classifier. Precision indicates the exactness of the classifier, meaning that higher and lower precision leads to less and more false positive classifications respectively. Recall measures the classifier's completeness. Higher and lower recall means less and more false negative classifications (the data are not assigned as related to an area, although they should be) respectively. Precision and recall are increased at the expense of each other. Thus, they are combined to produce the weighted harmonic mean of both metrics, which is the F-measure.

The evaluation of all algorithms is performed via 10-fold cross-validation (Kohavi, 1995), where the original data are randomly divided into ten equal subsets. Of these ten subsets, one is retained as the validation test of the model and the remaining 9 are used as the training data.

Table 9 summarizes the quality of classification using a set of related metrics such as Accuracy, F-measure, Kappa statistic, Mean Absolute Error and Root Mean Squared error. We also provide the variance in parenthesis for each metric. The results suggest that Random Forest outperforms all the classification algorithms in most of the aspects.

	Classification Algorithms						
Metrics	**NB**	**SVM**	**LR**	**C4.5**	**MLP**	**K***	**RF**
Accuracy	74.08% (1.73)	85.11% (1.48)	85.92% (1.70)	86.84% (1.36)	90.87% (2.01)	93.51% (1.33)	**95.95%** (1.06)
F-Measure (Weighted Avg.)	0.744 (0.10)	0.850 (0.08)	0.859 (0.09)	0.868 (0.06)	0.908 (0.13)	0.935 (0.05)	**0.959** (0.03)
Kappa statistic	0.7341 (0.02)	0.8477 (0.01)	0.8564 (0.02)	0.8651 (0.01)	0.9062 (0.01)	0.9336 (0.01)	**0.9586** (0.01)
Mean Absolute Error	0.0101 (0.03)	0.0371 (0.02)	0.0056 (0.02)	0.0056 (0.01)	0.0043 (0.02)	**0.0026** (0.01)	0.0086 (0.01)
Root Mean Squared Error	0.08891 (0.03)	0.1351 (0.02)	0.0592 (0.02)	0.0662 (0.01)	0.0543 (0.02)	**0.0439** (0.01)	0.0478 (0.01)

Table 9. Classifiers metrics comparison

We use 10-fold cross validation and compare the classifiers using t-paired test to detect which are significantly better than the others (a=.05 level) (Salzberg, 1997). We employ F-Measure as a comparison metric because it is more reliable and provides a good trade-off between Precision and Recall. Again, Random forest statistically outperforms the other classifiers, followed by K* follows, C4.5, Multilayer Perceptron, Logistic Regression, Naïve Bayes and Support Vector Machines in descending order. Random Forest is by design an ensemble learning method meaning that it combines a large number of weak classifiers with respect to the achievement of better classification results. This led us to follow the design principles of Random Forest and suggest an ensemble classification approach to improve the localization accuracy for indoor positioning system.

5.5.2 Proposed Filter

Instead of relying solely on one classifier, we decided to form a hybrid approach (see Figure 29) that involves a voting process that takes multiple classifiers into consideration. The hybrid method is actually an ensemble method (Dietterich, 2000; Kotsiantis et al.2006), which involves learning algorithms that form a set of classifiers. We chose the three significantly better algorithms based on the F-measure metric to form the ensemble classifier, which classifies new data by taking a weighted vote of the selected algorithms' predictions.

More specifically, the proposed ensemble filter is a meta-classifier that performs weighted majority voting among the three selected classifiers (i.e., C4.5, K* and random forest). The ensemble filter predicts the class label y. To do so, a weight w_j is associated with each classifier C_j. The filter formula is the following:

$$y = argmax_i \sum_{j=1}^{m} w_j X_A(C_j\,(x) = i),$$

where X_A is the classification function $[C_j(x) = i \ \in A]$ and A is the set of unique class labels (i.e., the store areas). The outcome of the formula is the class with the arguments with the greater weight (i.e., argmax). In this case, the weights are assigned automatically via the Weka software.

In order to study the accuracy of the proposed ensemble mechanism (external validity experiment) we created a test data set in which we simulated the shopping trip of 5 different consumers, and we collected 2500 observations in total. We compared the actual observations of the test data set with the estimation of the proposed ensemble with regards to evaluate the indoor positioning capabilities. The proposed ensemble method we use (Table 10) seems to achieve a slightly lower accuracy than the highest classifier (i.e. Random Forest), however the absolute mean and root mean squared errors are lower. The accuracy of the ensemble method is almost equal to the highest classifier (i.e. Random Forest), but also the errors of the absolute mean and the root mean squared are lower.

The ensemble and the Random Forest classifiers achieve slightly similar results. None of them is significantly better than the other, thus they behave in a similar way in terms of accuracy. Regarding the mean absolute and root mean squared errors in the case of the ensemble method; these are significantly lower compared to the best classifier (i.e. Random Forest). Lower errors mean that the result is closer to the actual one, leading to more efficient location determination.

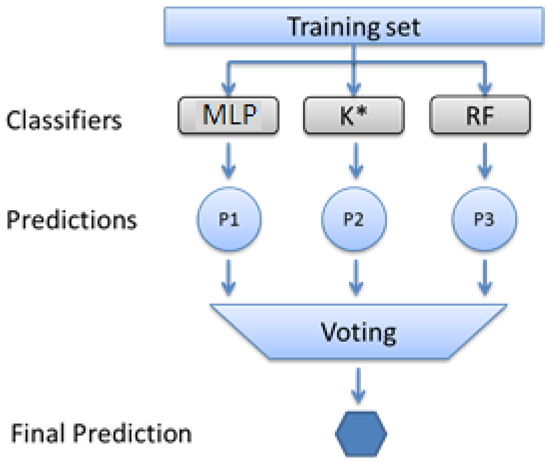


Figure 29: Ensemble classifier

To visualize the performance of all classifiers examined, we utilize receiver operating characteristic (ROC) graphs; a technique used in machine learning and data mining (Fawcett, 2006). Figure 30 presents the trade-off between true positive and false positive rate for each classifier.

Metrics				
Classifier	**MLP**	**K***	**RF**	**Ensemble**
Accuracy	90.87% (2.01)	93.51% (1.33)	95.95% (1.06)	95.78% (1.00)
F-Measure (Weighted Avg.)	0.908 (0.13)	0.935 (0.05)	0.959 (0.03)	0.957 (0.02)
Kappa statistic	0.9062 (0.01)	0.9336 (0.01)	0.9586 (0.01)	0.9569 (0.01)
Mean absolute error	0.0043 (0.02)	0.0026 (0.01)	0.0086 (0.01)	0.0051 (0.01)
Root mean squared error	0.0543 (0.02)	0.0439 (0.01)	0.0478 (0.01)	0.0392 (0.01)

Table 10. Ensemble classifier metrics comparison

The important points in this graph are: (a) point (0,0), which means that the classifier commits no false positive errors but also gains no true positives, (b) point (1,1) where the classifier commits positive classifications unconditionally and (c) point (0,1), which represents a perfect classification (Fawcett, 2006). A classifier is considered to perform better as it moves to the northwest part of the graph and makes positive classifications only with strong evidence, as the false positive rate gets close to the Y axis (Fawcett, 2006).

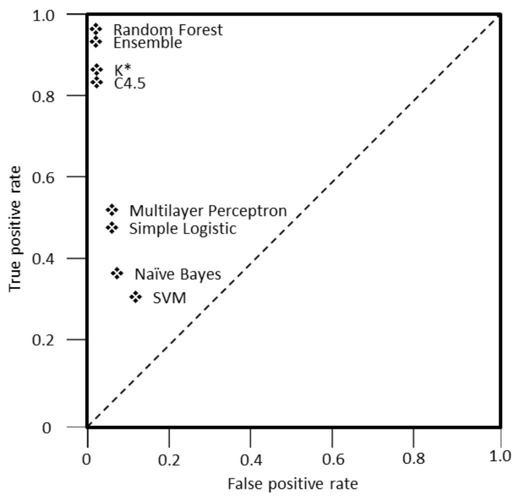


Figure 30. Classifiers ROC space graph

The Random Forest classifier and the ensemble approach are considered to make positive predictions based on strong evidence, as they are close to the upper left point of the graph, followed by K* and C4.5. The rest classifiers (i.e. Multilayer Perceptron, Simple Logistic, Naïve Bayes and Support Vector Machines) tend to have a lower performance than the first ones as their distance from point (0,1) increases.

For further comparison between Random Forest and the proposed ensemble classifiers we transform the multiclass classification of the user's area into a binary classification to check whether the classifiers correctly predict the user's position. Figure 31 presents the ROC curve of Random Forest and ensemble classifiers and demonstrates their training speed. The ensemble classifier is characterized by a better learning ability as the training instances increase up to 0.65 True Positive rate. From 0.65 to 0.80 both algorithms tend to behave the same and above 0.8 Random Forest is slightly better. When both algorithms reach 0.95 of True Positive rate, they have a similar behaviour, with Random Forest being slightly superior by 0.01. The ensemble approach improves detection accuracy; however, this cannot be generalized as it depends to contextual factors such as the store layout, product placement and the devices used for fingerprinting.

Finally, we present the overall cumulative probability and error correlation in Figure 32. Based on our experimentation we conclude that approximately a localization error of 70% of the cases is lower than 1 meter and in 80% of the cases the localization error is around 2 meters. The results confirm prior similar studies that try to improve the detection accuracy when using BLE beacons. On the contrary, for the random forest classifier, in 80% of the cases the localization error is approximately 2.5 m. In retail environments in particular, such deviation is significant, because even 0.5 m away from the actual shopper's position may lead to position him in a different shopping isle and in front of a different store shelve, thus a different product category. Considering the need of retailers to use such indoor positioning systems for knowing the actual shopping trips of customers and offering them personalized services (e.g., promotions designed based on their route in the store), the lower positioning error of our ensemble filter compared to the random forest is significant even though they do not have significant differences in terms of accuracy and F-measure.

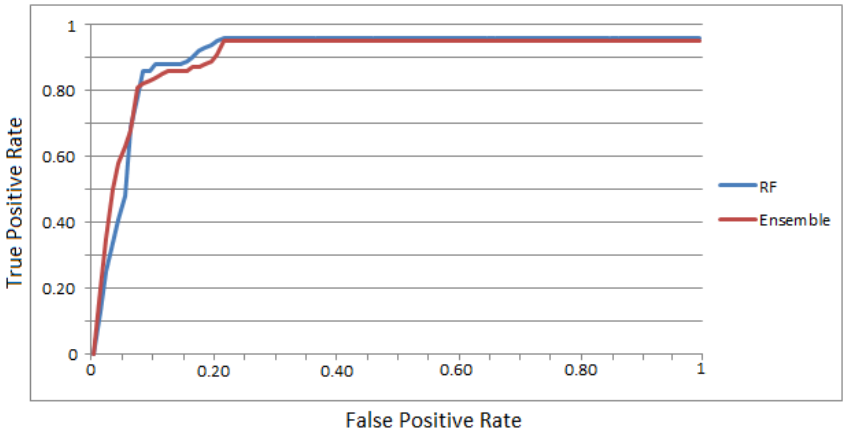


Figure 31. ROC curve

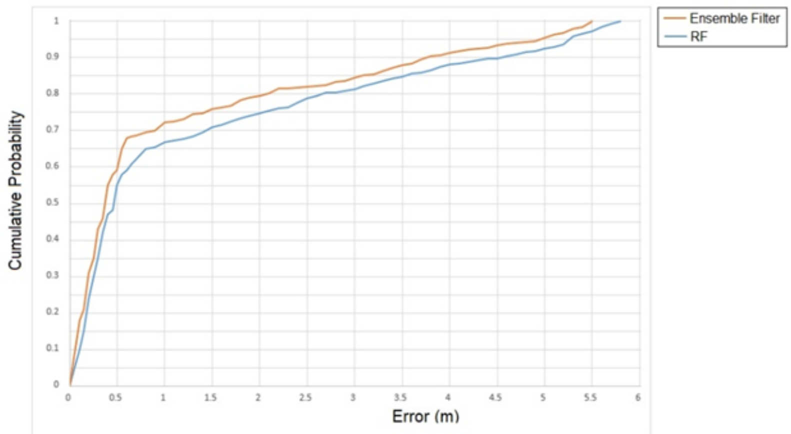


Figure 32. Positioning error for filter

The implementation of the indoor positioning system within a retail grocery is affected by factors such as accuracy, scalability and cost (Liu et al., 2007). These performance metrics raise the question of how easily it can be developed and deployed in the store. The positioning system consists of a fix BLE beacon infrastructure and the shop-specific filter. The infrastructure access points do not require be modifying and the areas can be dynamically redefined based on the daily operations of the store (e.g. promo areas, new product introduction). As a result, the part that needs to be modified is the ensemble filter that is trained based on the dynamic areas of the store. Regarding the cost of war-driving (He and Chan, 2016) for the ensemble filter it can be easily integrated in the daily operations of the grocery store and performed during the daily product check by the store manager.

5.6 Data-driven Evaluation

Apart from the technical evaluation of the classifiers, we go beyond that and evaluate the outcome by utilizing additional data that will help interpret the findings. To this end we exploit location analytics techniques. Location analytics enable tremendous value for retailers, as they can exploit indoor location data in order acquire competitive advantage and make actionable data-driven business decisions. The uncovering of relationships between location data and other data available to the retailer can facilitate a vast number of business processes and operations. One may notice that location analytics can enable significant advantages if used properly; thus, adding the “where” in business data can add value to the organization.

Location analytics applied on spatiotemporal shopper data are able to extract patterns of shopper trajectories. The identification the movement patterns of browsing shoppers in retailing leads to the discovery of frequent paths and preferred areas in the store. Such patterns can be utilized for effective placement of promotions (Grewal, 2011), as it is more likely that shoppers will pass through the promotion area. At the first part of this section we perform a field study that aims to uncover correlations between positioning and purchase data. Then, at the second part of this section we present our major findings.

5.6.1 Field Study: Positioning Data vs. Purchase Data

To evaluate the results of the visit purpose identification we designed a field study for a major grocery Greek FMCG retailer. Firstly, we analysed spatiotemporal data for 100 users that navigated and purchased products in the store. For each shopper session we processed the spatiotemporal results and based on the positioning approach we extracted the areas visited during the shopping trip. Upon the shopper trajectories we perform area-based segmentation and identify the purpose of the visit by solely examining the areas that the shopper spent time during the visit. In addition, we examine another aspect, namely opportunity gap that indicates whether a shopper spent time in an aisle but not made any purchases from there. This parameter indicates an opportunity that under specific actions may lead to product purchases.

5.6.2Process and Resulting Behavioral Patterns

Goal of this part is to exploit the collected data from the positioning system and be able to extract insights regarding shopper behaviour. Thus, the approach used for performing area-based shopper segmentation aims at discriminating shopper segments of similar behaviour patterns based on the areas they visit in the store. To this end, we form a vector containing all the areas visited by the user and, also, the characterization of the residence time of the user in this area in terms of low, medium or high residence. In cases that the user visits the same area more than once, we aggregate the individual visits and sum the overall residence time for the categorization variable.

In this study we utilize as input in shopping trip clustering task the average time spent by the shopper in each area. Using discretization and normalization, it is possible to transform the shopper temporal events to an ordinal variable (e.g. Low, Medium, High) based on the total amount of time spent in the store. The decision to transform the time of a shopping trip into an ordinal variable is based on the idea that a shopper has variable pace and movement patterns during a shopping trip. It is expected that a shopper will move at a relative slower pace when browsing an area of interest while he/she will move relatively faster through areas of no interest. Therefore, using relative time and speed measures allows describing a shopping trip more accurately.

In Table 11 we present the way accumulated data was organized. In detail, each row of the table represents a shopping trip and includes information such as the trip ID, the date of the visit, the start and end time of the shopping trip, and the time spent in different shopping areas. If a shopper hasn't passed from a specific area, then the corresponding cell is marked as Not Available (N/A).

Shopping trip ID	**Date**	**Start time**	**End time**	**Area 1**	**Area 2**	**…**	**Area 52**
SP1	01-05-2019	10:50	11:14	low	high	…	N/A
SP2	01-05-2019	18:23	18:57	high	N/A	…	low
SP3	14-05-2019	09:07	09:32	low	medium	…	low
…	…	…	…	…	…	…	..
SP100	17-05-2019	13:05	13:28	low	medium	…	N/A

Table 11: Segmentation features

During our field experiment each shopper performed only one shopping trip. We performed clustering based on the features presented in Table 11 which indicate the interaction with each store area for each user. The outcome of the clustering process extracted seven discrete shopper segments based on their in-store behaviour. In order to be able to assess the shopper segments we have combined their purchase data as descriptive statistics and be able to extract further insights regarding shopper segments. Table 12 presents the characteristics of each segment in term of shoppers and sales volume.

Segment	Shoppers	Sales volume
Segment 1	25.93%	29.9%
Segment 2	16.05%	25.9%
Segment 3	19.75%	19.5%
Segment 4	9.88%	5.0%
Segment 5	9.88%	2.5%
Segment 6	12.35%	11.9%
Segment 7	6.17%	5.3%

Table 12: Segment characteristics

Segment 1 comprises 25.93% of the shoppers and nearly 30% of the total purchases made by all the shoppers. Figure 33 highlights the areas of interest for Segment 1. These shoppers devoted time in areas that contain products refereeing to detergents, diapers, baby care, bread and pet foods.

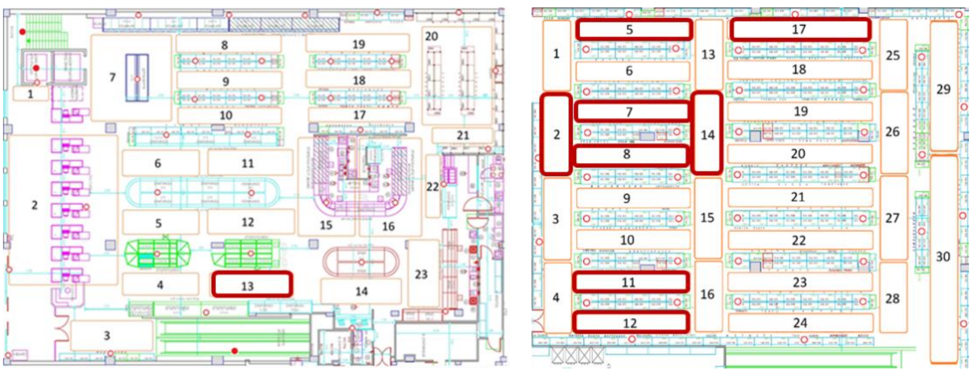


Figure 33: Segment1 areas of interest

Segment 2 comprises 16.05% of the shoppers and nearly 26% of the total purchases made by all the shoppers and Figure 34 highlights its areas of interest. Shoppers of Segment 2 devoted time in areas containing fresh fruits and vegetables, tea, coffee, cold cuts and personal care.

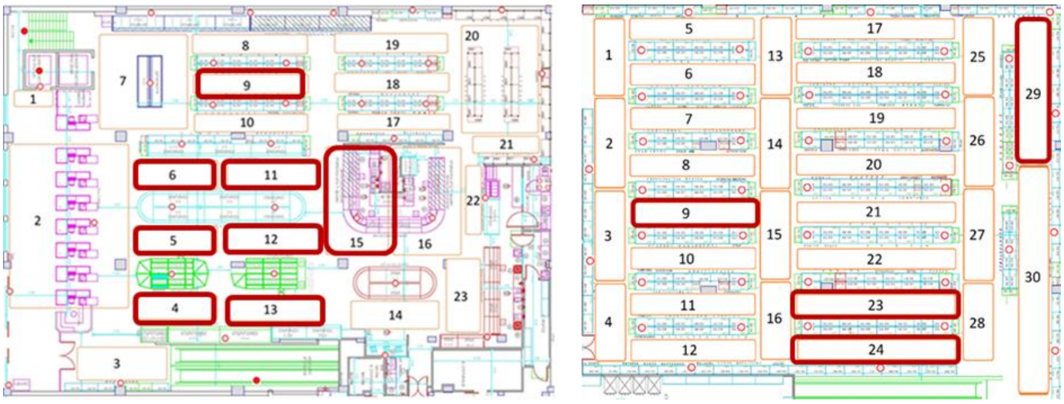


Figure 34: Segment2 areas of interest

Segment 3 comprises 19.75% of the shoppers and 19.5% of the total purchases made by all the shoppers with Figure 35 depicting the areas of interest for the shopper of this segment. These shoppers spent time in areas containing fresh fruits and vegetables, butchery, cold cuts, personal care, oral care, beverages and biscuits.

Figure 35: Segment3 areas of interest

Segment 4 comprises 9.88% of the shoppers and 5% of the total purchases made by all the shoppers and Figure 36 depicting the areas of interest for the shopper of this segment. The interesting part about this segment is that the spend significant time during their shopping trip only in one area; i.e. beverages and beers.

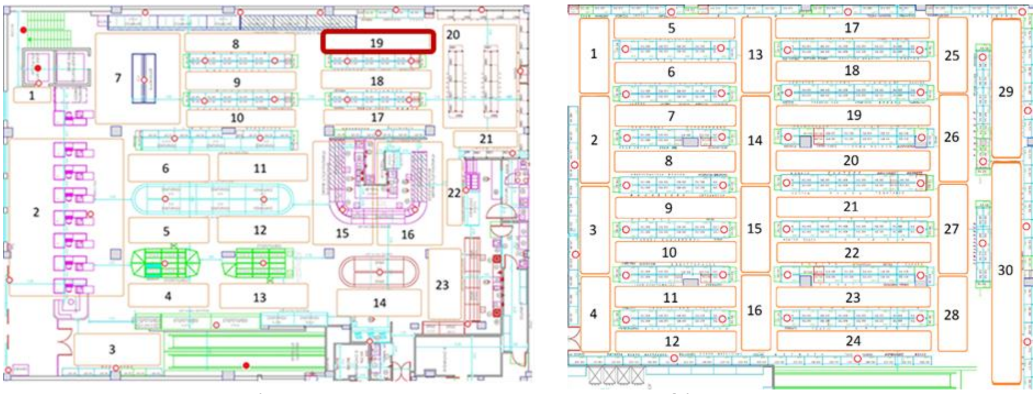


Figure 36: Segment4 areas of interest

Segment 5 comprises 9.88% of the shoppers and 2.5% of the total purchases made by all the shoppers. Figure 37 indicates that these shoppers spent time only on the ground floor. The areas of interest for these shoppers are chocolates, chips, biscuits, fresh fruits and vegetables, yeast, ice-cream, cold cuts, fresh milk, butchery and yogurt.

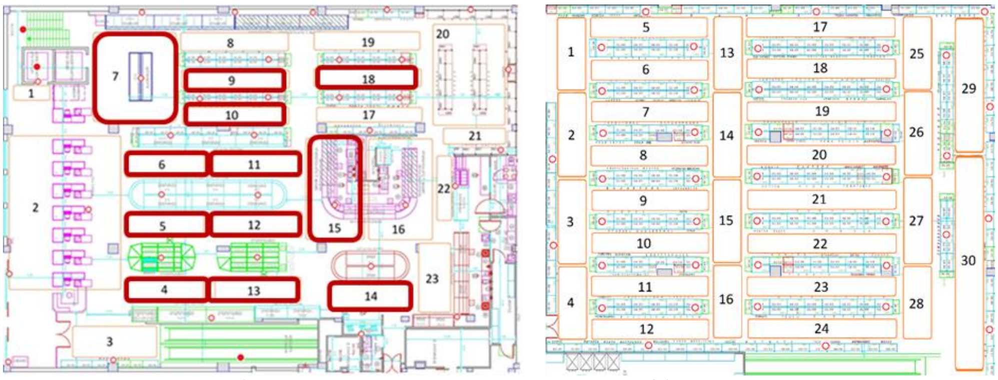


Figure 37: Segment5 areas of interest

Segment 6 comprises 12.35% of the shoppers and nearly 12% of the total purchases made by all the shoppers. Figure 38 indicates that these shoppers spent time only on the first floor, unlikely to the previous shopper segment that spent time solely on the ground floor. The areas of interest for this segment involve curtains, towels, underwear, sauces, cookware, salt and pasta.

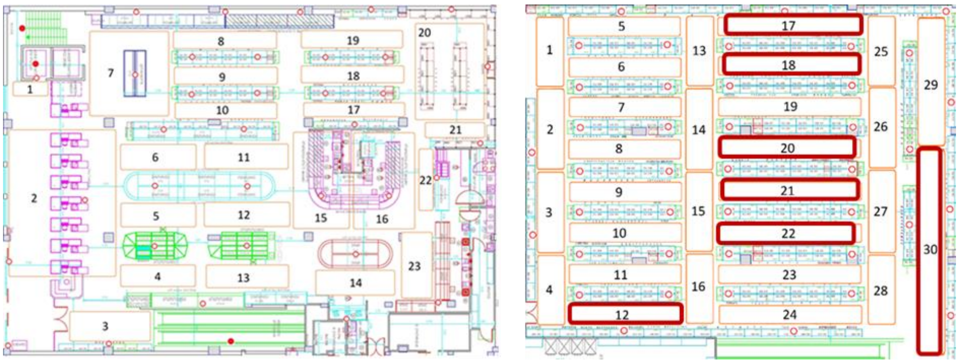


Figure 38: Segment6 areas of interest

Finally, Segment 7 comprises 6.17% of the shoppers and 5.3% of the total purchases made by all the shoppers with Figure 39 depicting the areas of interest for this shopper segment. These areas contain products regarding stationary, toilet paper, salt, sugar, diapers pet food and cold cuts.

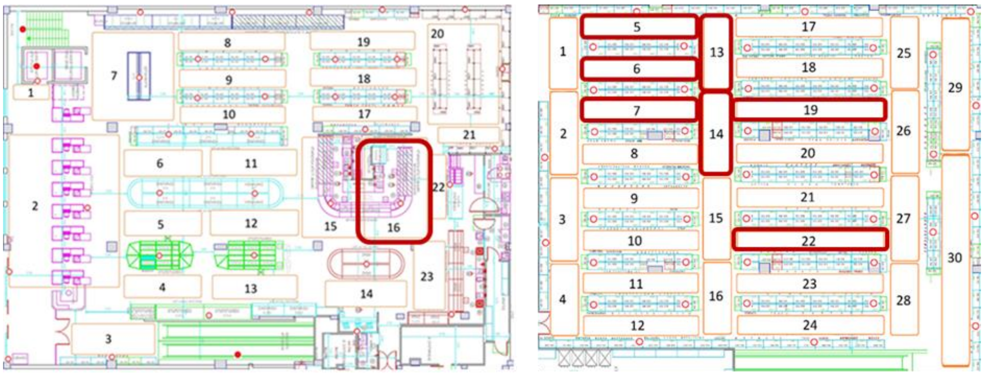


Figure 39: Segment7 areas of interest

5.6.3Setting and Data Collection

The field study was conducted for 100 shoppers that visited the retail store and via a customized mobile application navigated the store while shopping. While the shoppers browsed the store the wireless BLE Beacon infrastructure along with the mobile application gathered data regarding shopper movement. During the experiment we acquired the basket data regarding the products that participants bought.

Thus, from the wireless infrastructure and the mobile application we were able to:

- Track the shopping trip of the user
- Extract the shopper location via the indoor positioning approach
- Calculate shopping trip descriptive statistics such as total duration and KPIs regarding each area such as time spent and visits
- Identify the areas visited during the shopping trip

From the POS data we were able to:

- Identify the purpose of the visit by examining the products purchased
- Identify the time spent in each area where the purchases have been made

5.6.4Findings

The findings of the field study fall into two pillars. On the one hand we demonstrate that the path followed and the areas of interest (i.e. the areas that the shopper spent time regarding the shopping trip) depict the purpose of the shopping visit. In addition, we pinpoint the opportunity gap KPI that indicates shoppers more prone to actions, as they spend much time in store areas but not making any purchases. However, such shoppers seem to be more impulse buyers than others.

5.6.4.1 Association between Positioning Data and Purchased Data

Regarding the correlation between the positioning data and the purpose derived by the purchased products we identify the following:

Segment 1: By comparing their purchase data, these shoppers purchased products regarding food, detergents, confectionary, home and personal care. As a result, this segment uncovers a correlation between the areas of interest where the shopper spent time with the purchases made.

Segment 2: The purchase data of this segment indicates that these shoppers bought products from categories regarding personal care, breakfast and light meal. This segment, also, indicates that their areas of interest correlate to their purchase data.

Segment 3: Their purchase data for this segment indicate that shoppers bought detergents, breakfast and light meal, personal care, snack, beverages, fruits and vegetables. Similar to the previous segments, this one, also, indicates that the areas of interest correlate to the purchased products by the shoppers.

Segment 4: The purchase data by these shoppers indicate that they bought products regarding beverages and beer and products regarding oral care, breakfast and snacks. The fact that they did not spent more time in the areas of the remaining purchases indicates that they had already planned these purchases and the purchase that required more thought was the one involving beverages and beers.

Segment 5: The purchases of these shoppers include beverages, snack, breakfast, confectionary and personal care. Unlike to the previous segments, this one indicates that shoppers visited significantly more aisles than the ones containing the products they bought. This shopper segment indicates shoppers that may have browsed the ground floor of this store for pleasure and avoided to visit the first floor of the store.

Segment 6: These shoppers bought products regarding snacks, personal and home care, and breakfast. These purchases differ significantly from the areas of interest, thus it’s a segment that could have bought products from these areas that shoppers spent time.

Segment 7: The purchase data of these shoppers indicate that they bought products regarding personal care, breakfast, cold cuts and detergents. Segment 7 indicates that shoppers stayed significantly in store areas where they also purchased products, but also spent time in area where they bought nothing.

The findings indicate that in most of the segments, the path followed matches the shopping purpose as extracted from the POS data. In a few cases the shoppers followed a path that did not lead to purchases from all the areas they stayed. Such cases indicate the opportunity to further examine the factors that shoppers spent time in an area, but finally won't purchase anything from there.

5.6.4.2 Impulse Buying Opportunity

Apart from the time spent in store areas we examine the "coverage" KPI and examine the percent of the store surface that each shopper segment browses during the shopping trip. We further break down "coverage" KPI into metrics regarding the time shoppers spend in store's areas. Table 13 depicts each segment along with (i) the percent of the store that browsed spending low time (column: Low time surface), (ii) the percent of the store that spend medium time during the shopping trip (column: Medium time surface), (iii) the percent of the store that spend significant time during the shopping trip (column: High time surface), (iv) the percent of the unvisited store surface (column: Unvisited surface), (v) the percent of the store that purchases have been made (column: Purchases surface), and (vi) the missing sales ratio which is the ratio between "High time surface" and "Purchases surface". Columns "Low time surface", "Medium time surface", "High time surface" and "Unvisited surface" sum to 100% which is the overall store's surface.

Segment 1 and Segment 3 are two segments that their shoppers tend to spend significant time in most of the stores surface. In addition, these segments bought products from more store areas comparing the rest shopper segments (i.e. 31.41% and 20.94% respectively). Segment 4, Segment 5 and Segment 7 are characterized by the least surface store coverage, as these shoppers visited solely the areas containing the products they bought. Segment 6 presents a more balanced behaviour among these two groups.

We further define the Impulse Buying Margin (IBM) metric. This metric indicates how many of the visited areas, where the shopper spent time in, led to purchases. IBM derives from the division of the percentage of the areas that shopper spent time by the percentage of the store surface that shopper's purchases come from. The outcome of the IBM metric reveals to what extend there was impulse/planned buying behaviour. This metric is essential for managers as the information it reveals aids them in recognizing the level of impulse/ planned buying behaviours. Subsequently, managers may exploit the opportunity to turn impulse/planned behaviours into sales using the appropriate strategy.

$$IBM_i = \frac{PSC_i}{SSC_i}$$

Typically, IBM values concentrate around 1 and express the average shopper behaviour where shoppers tend to purchase products from their areas of interest. Extreme values near 0 or much greater than 1 indicate either mostly impulse purchases or strictly planned accordingly. Investigating intermediate values, we observe that values of IBM < 1 (empirically <.77) signal the decision maker that the shopper i devoted a significant amount of time to most of the areas they visited without necessarily making a purchase. On the contrary, values of IBM>1 (empirically >1.30) indicate that shoppers spent less time in each area and made purchases. Therefore, it is evident that IBM<.77 reflects impulse behaviour, whereas IBM>1.30 reflects planned behaviour.

For example, in Segment 1 shoppers spends high time in 41.13% of the store's surface and their purchases come from the 31.41% of the store's surface. This highlights an opportunity of trying to turn this gap into actual sales. Similarly, Segment 4 and Segment 7 present the same opportunity. Segment 2 may have a ratio of 4.6, however it is achieved due to the significant low "High time surface" and thus it cannot be taken into consideration. The similar phenomenon occurs for Segment 5, despite the ratio value of 1.35. As a result, Segment 1, Segment 3 and Segment 6 would be a potential pool of shoppers to be targeted in order to design strategies for turning this opportunity into actual sales.

Segment	Low time surface	Medium time surface	High time surface	Unvisited surface	Purchases surface	IBM
Segment 1	8.27%	36.51%	41.13%	14.09%	31.41%	0.76
Segment 2	32.28%	5.10%	3.28%	59.33%	15.05%	4.6
Segment 3	16.03%	23.60%	21.35%	39.02%	20.94%	0.98
Segment 4	11.75%	7.48%	9.52%	71.24%	6.98%	0.72
Segment 5	15.80%	6.09%	5.01%	73.10%	6.76%	1.35
Segment 6	11.30%	13.98%	12.23%	62.48%	12.65%	1.03
Segment 7	4.56%	7.24%	7.47%	80.74%	6.22%	0.83

Table 13: Segment overall characteristics

In order to further understand the shopper profile behind each segment, we combine the demographic data of the pilot and provide a high-level description for each one (see Table 14). Segment 1 is characterized by females who have come alone to the store and our analysis indicate that the time spent in the store highlights a shopper that has time available to spend in the store and has generated the highest spending among segments. On the contrary, Segment 2 consists of mostly men who have come alone to the store mostly for planned purchases. Segment 3 is characterized by couples (female driven) that cover less store surface than Segment 1 but seems to spend adequate time in store's areas.

Segment	shoppers	Average Value (€)	Average variety	Total items	Total value	Sales volume
Segment 1	25.93%	80.19	16	34.79%	34.91%	29.90%
Segment 2	16.05%	50.81	13	16.71%	13.69%	25.90%
Segment 3	19.75%	65.70	14	23.01%	21.79%	19.50%
Segment 4	9.88%	33.49	8	4.66%	5.55%	5.00%
Segment 5	9.88%	40.91	9	5.21%	6.78%	2.50%
Segment 6	12.35%	59.68	14	13.15%	12.37%	11.90%
Segment 7	6.17%	47.15	12	2.47%	4.89%	5.30%

Table 14: Segment purchase descriptive statistics

Segment 4 is characterized by female shoppers that have come alone to the store for planned purchases but occasionally make purchases from the areas that spent time. Segment 5 comprises males who have come alone to the store for extremely planned purchases as they cover a specific surface of the store. Segment 6 comprises of couples that tend to spend more time in areas of interests, despite the small store surface coverage, and occasionally make impulse purchases from these areas. Finally, Segment 7 is mostly male shoppers who have come alone, browse a very small part of the store and make impulse purchases only if the content of the area matches their interests.

Segment	Gender	Visit companion	Average IBM	Buying behaviour
Segment 1	Female	Alone	0.76	Impulse
Segment 2	Men	Alone	4.6	Planned
Segment 3	Female	Couple	0.98	Partially Planned
Segment 4	Female	Alone	0.72	Impulse
Segment 5	Male	Alone	1.35	Planned
Segment 6	Male	Couple	1.03	Partially Planned
Segment 7	Male	Alone	0.83	Partially Planned

Table 15: Cluster shoppers' impulse buying characteristics

Summarizing this section, in order to describe the characteristics of each shopper segment we went through the areas of interest where shoppers spent more time regarding the average time spent in the store during their shopping trip. As mentioned, in order to validate our findings, we go beyond the area-based segmentation and additionally examine the shopper baskets of these customers. This approach enables the examination of where the shoppers go versus what they buy and detect selling gaps. In most of the cases we verified that indeed the area-based segmentation matched the products that shoppers bought, indicating that this kind of segmentation could be used to extract the shopping purpose of the shopper visit. However, certain shopper segments did not match the segments formed by shopper purchases. The products that shoppers bought indicated potential missing sales for the retailer. Such cases may refer to price comparison where the shopper decided to purchase the products from another retail chain but spent time in the specific area that the product is placed. Decision makers can utilize this knowledge in order to exploit the detected selling gaps. Shoppers that spend time in area with detergents but never buy them would be an opportunity for promotional actions in order to enable these shoppers start purchasing the specific products from this retail store.

6 CASE B: APPLICATION OF INDOOR POSITIONING USING WI-FI TECHNOLOGY

6.1 Introduction to the case study

The selected operating environment of the indoor positioning system is an electronics retail chain. The selected retail stores for this case are two (2); a three-floor store (Store 1) and a single floor store (Store 2) with different layouts. Store 2 has been renovated and designed with a different layout; thus, we treat the new layout as a different store (Store 3). All stores are characterized by different layout characteristics; Store 1 is characterized by grid and diagonal layout, while Store 2 is grid-based layout and Store 3 is a mixed layout (i.e. free flow, diagonal and angular layouts) store. Store 1 has installed 14 Wi-Fi access points and Stores 2 and 3 have 10 Wi-Fi access points each. On the contrary of case A, no customized smart phone application is required. For this case we utilize the connection to the Wi-Fi network each compatible mobile phone makes to access the internet. So instead of processing BLE Beacons signals, we process the Wi-Fi signals along with the connection to the wireless network.

In this context we applied and deployed the indoor positioning approach into a system that consists of a Wi-Fi infrastructure and a customized filter that adapts to the specific contextual store requirements, along with a back-end cloud infrastructure that processes and stores the data from the captured mobile device positions. Based on thorough experimentation, it turned out that the proposed system achieves high localization precision, has low deployment complexity, is scalable because each store acquire its own localization model utilizing fingerprinting during the training stage and is responsive because the positioning algorithm can be performed within the time limits suggested by the business requirements.

6.2 System Design and Implementation

Case B was deployed on an existing wireless infrastructure that could be not modified. To this end, we designed and implemented the indoor positioning system without modifying the Wi-Fi routers. The implemented architecture, unlikely to BLE Beacons, does not require a mobile application to track the user. The Wi-Fi infrastructure captures the interaction between the wireless network and the mobile device and upon these data indoor positioning can be performed.

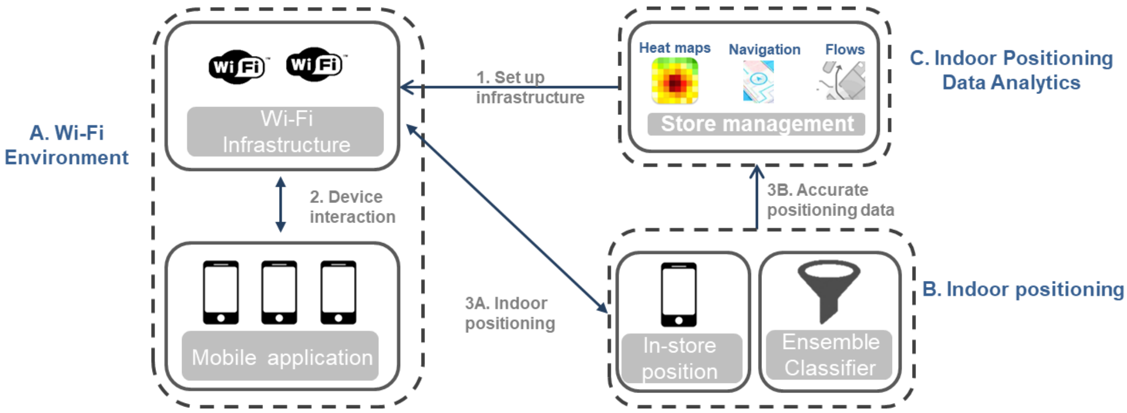


Figure 40. Wi-Fi indoor positioning system overview.

The system in this case consists of the following three major modules: (A) the Wi-Fi environment, (B) the indoor positioning mechanism and (C) the indoor positioning data analytics that stores and extracts knowledge from collected data and provides the management of additional location-based services.

More precisely (A) the Wi-Fi environment (setting and map) consists of the following modules:

(I) The Wi-Fi Infrastructure layer refers to the Wi-Fi routers that are deployed in the store and the data that are transmitted and captured by the wireless card of the mobile device of the consumers. In this case the routers were already placed in fixed positioning and could not be modified. The available data from these routers include the router that the mobile device was connected, the timestamp of the connection, the type of device that was connected and information regarding the connection (e.g. association/disassociation or connection timeout from the access point).

(II) The Mobile Application layer which utilizes the wireless card of the mobile device that communicates with infrastructure when the device is connected to the wireless network. Via the interaction of the device with the wireless network, the infrastructure captures that aforementioned data and captures them for further processing.

(B) The indoor positioning module is responsible for two tasks; i.e. (i) the in-store position of the consumer that communicates with the mobile device to provide the current position and (ii) the positioning filer which processes the collected data from each consumer to provide accurate positioning data for consumer traffic heatmaps and flows visualization. The data captured by the mobile device are stored in the backend infrastructure for further processing, as well as for the generation of the appropriate meta-data. The data from the backend infrastructure are used by other applications or services such as consumer traffic heatmap reports, user navigation and flows visualization (i.e. the administrative and data analytics module).

The (C) indoor positioning data analytics module is responsible for extracting insights and knowledge from the location data that are captured by the mobile devices of the users and stored in the system's infrastructure. This module addresses the appropriate location-based services to the consumers.

6.3 Business, Data and Setting Understanding and Preparation Layer

The proposed approach was deployed in 3 stores, where Store 1 is 2400 m2 and Stores 2 and 3 2400 m^2. Store 1 was divided in 14 areas, while Stores 2 and 3 were divided in 10 areas each, based on the criteria of technical efficiency and business interest. The selection of the areas has been performed on the basis that the next step of the process is to produce consumer traffic area heatmaps regarding the movement of the consumers within the electronics stores. Thus, the stakeholders of the electronics stores determine the area selection based on their business needs. Unlikely Case A where the Beacon placement was performed as the first step, in Case B we could not change the location of the Wi-Fi access points. Thus, we had to define the areas having this constraint in mind. Thus, to decide the most efficient way to split the areas of interest, we examined a series of combinations in order to detect the most appropriate one. We started by gathering signal fingerprints for each store in order to form the final areas. Starting from small blocks we tried to detect the appropriate granularity level until the final area division. The following figures demonstrate the initial areas that fingerprinting was performed and then the final areas that each floor was divided into.

Store 1 comprises 3 floors:

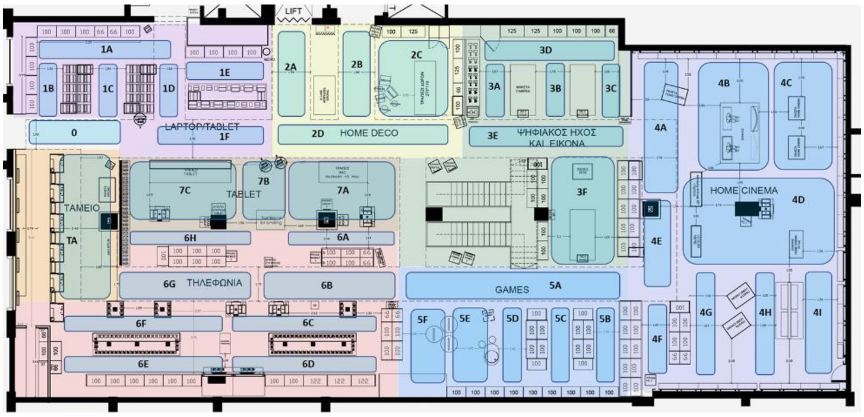


Figure 41. Store 1 – Ground floor fingerprinting.

Figure 42. Store 1 – First floor fingerprinting.

Figure 43. Store 1 – Second floor fingerprinting.

While the areas division is the following:

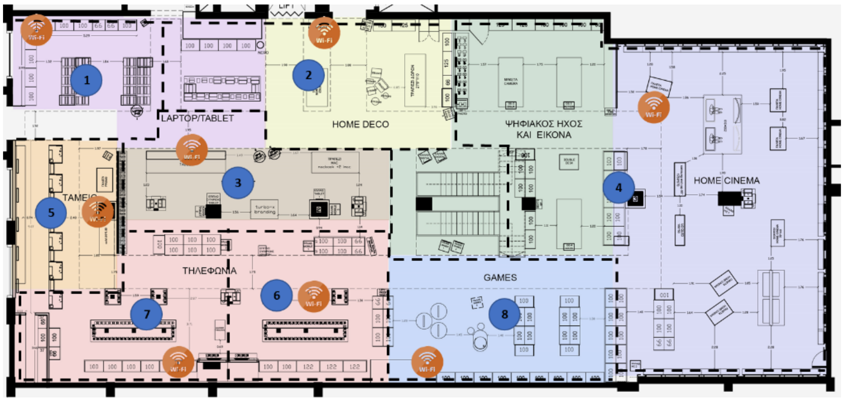


Figure 44. Store 1 – Ground floor area division.

Figure 45. Store 1 – First floor area division.

Figure 46. Store 1 – Second floor area division.

Store 2 is a single-floor store:

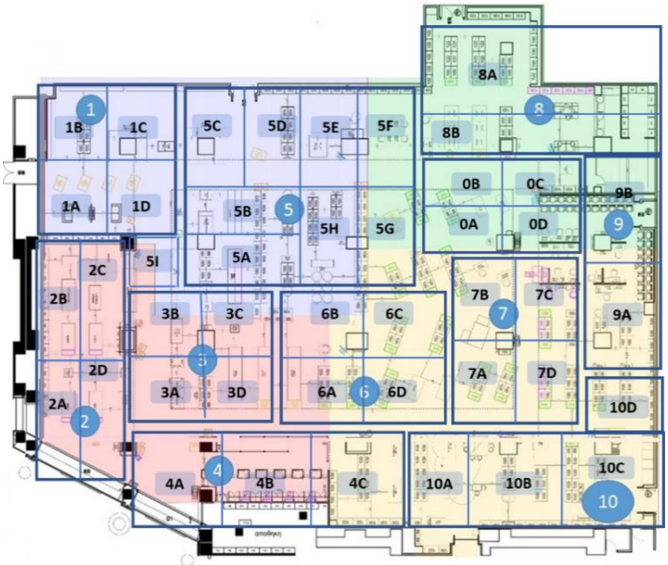


Figure 47. Store 2– Fingerprinting areas.

Figure 48. Store 2 –Area division.

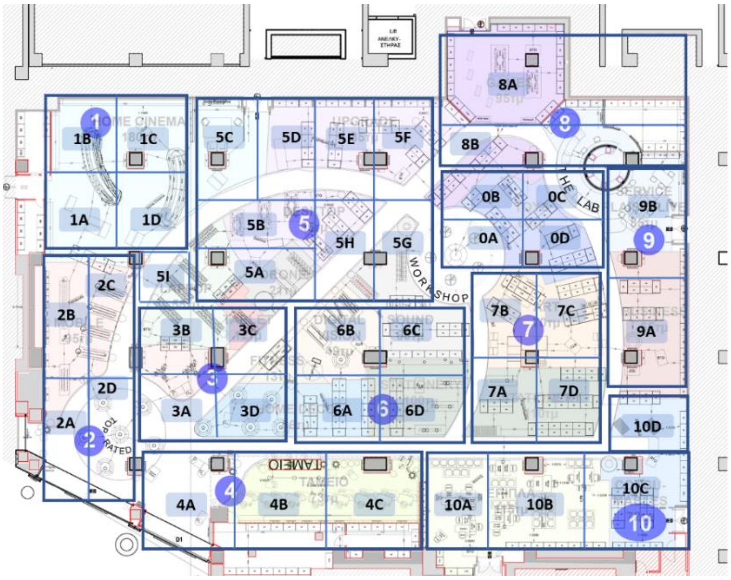


Figure 49. Store 3 – Fingerprinting areas.

Figure 50. Store 3 –Area division.

Each mobile device that is connected to the wireless network communicates with the access point routers. These routers push the association messages from the mobile devices to the back-end software and these signals are used for determining the in-store position. The stored data are captured every one second and are used for the generation of appropriate meta-data.

Every time the mobile device associates with an access point it immediately generated a record contacting information regarding the time, the network name, the type of the event and details regarding the association. These data are depicted in Figure 51 and require further transformation in order to be processed. The raw data are stored in a NoSQL database in the backend infrastructure.

Time (EET) ▾	Access point	SSID	Client	Event type	Details
Feb 21 20:48:40	Metamorfosi - F0 - AP-01	STAFF-WIFI	DESKTOP-A0J9HN1	802.11 disassociation	client has left AP
Feb 21 20:48:39	Metamorfosi - F0 - AP-01	STAFF-WIFI	DESKTOP-A0J9HN1	802.11 disassociation	client has left AP
Feb 21 20:48:39	Metamorfosi - F0 - AP-01	Plaisio_FreeWifi	android-61733c7e0a1c3f33	802.11 disassociation	unknown reason
Feb 21 20:48:37	Metamorfosi - F0 - AP-04	DEMO-WiFi	HUAWEI_P20_lite-fa43c487b	802.11 disassociation	unknown reason
Feb 21 20:48:37	Metamorfosi - F0 - AP-04	DEMO-WiFi	HUAWEI_P20_lite-fa43c487b	WPA deauthentication	radio: 1, vap: 2, client_mac: 04:79:70:EB:58:8C more »
Feb 21 20:48:35	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	WPA authentication	
Feb 21 20:48:35	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	802.11 association	channel: 44, rssi: 39
Feb 21 20:48:35	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	802.11 disassociation	client has left AP
Feb 21 20:48:35	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	WPA deauthentication	radio: 1, vap: 2, client_mac: 6C:C7:EC:B8:90:82 more »
Feb 21 20:48:25	Metamorfosi - F0 - AP-06	DEMO-WiFi	Galaxy-S10	WPA authentication	
Feb 21 20:48:25	Metamorfosi - F0 - AP-06	DEMO-WiFi	Galaxy-S10	802.11 association	channel: 44, rssi: 47
Feb 21 20:48:25	Metamorfosi - F0 - AP-06	DEMO-WiFi	Galaxy-S10	802.11 disassociation	client has left AP
Feb 21 20:48:25	Metamorfosi - F0 - AP-06	DEMO-WiFi	Galaxy-S10	WPA deauthentication	radio: 1, vap: 2, client_mac: 6C:C7:EC:B8:90:82 more »
Feb 21 20:48:18	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	WPA authentication	
Feb 21 20:48:18	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	802.11 association	channel: 44, rssi: 49
Feb 21 20:48:18	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	802.11 disassociation	client has left AP
Feb 21 20:48:18	Metamorfosi - F0 - AP-03	DEMO-WiFi	Galaxy-S10	WPA deauthentication	radio: 1, vap: 2, client_mac: 6C:C7:EC:B8:90:82 more »
Feb 21 20:48:13	Metamorfosi - F0 - AP-05	DEMO-WiFi	MiMIX3-Mi?????	WPA authentication	
Feb 21 20:48:13	Metamorfosi - F0 - AP-05	DEMO-WiFi	MiMIX3-Mi?????	802.11 association	channel: 44, rssi: 38
Feb 21 20:48:13	Metamorfosi - F0 - AP-06		MiMIX3-Mi?????	802.11 association rejected	load: 58, best_ap: 172.16.8.112, best_ap_load: 18 more »
Feb 21 20:48:10	Metamorfosi - F0 - AP-08	DEMO-WiFi	MiMIX3-Mi?????	802.11 disassociation	client was deauthenticated
Feb 21 20:48:10	Metamorfosi - F0 - AP-08	DEMO-WiFi	MiMIX3-Mi?????	WPA deauthentication	radio: 1, vap: 2, client_mac: A4:50:46:5A:0A:21 more »

Figure 51. Wi-Fi raw data

In order to be able to process the data we transform and store the data in a structured format (see Figure 28). Thus, we keep the following information:

- The timestamp of the association.
- The network name. This is useful, when multiple networks are available, and we have to pick the appropriate ones for indoor positioning.
- the signal strength that the application captures from the access point (RSSI) in order to determine how close or far is the mobile device from the transmitter, and

- The event type, in order to discriminate the association and the disassociation from the access point.

6.4 Modeling

We utilize the fingerprinting approach (Wang et al., 2015) and use area level as positioning granularity. Fingerprinting consists of an offline and an online phase. For the offline phase, we gathered a dataset that establishes the connection between store locations and the corresponding signal strength, as well as the distance from a transmitter. Two researchers have been moving in the store for two hours generating the required movement content using three mobile devices. The areas covered were those presented in fingerprinting figures (Figure 41, Figure 42, Figure 43, Figure 47, Figure 49). We gathered 6.800 observations that represent the simultaneous movement of two mobile phones in our experiment. The collected data were transformed into a single table where every row is a single observation and relates the store's area with Beacon IDs, Distance from each transmitter and the Signal Strength as depicted in the next figure.

Timestamp	Area	SSID	Client	Event type	RSSI	RSS
2019-02-21 10:59:32.000	F0AP03	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 5	-91
2019-02-21 11:00:10.000	F0AP06	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 16	-86
2019-02-21 11:00:11.000	F1AP11	Plaisio_FreeWifi	iPhone	802.11 association	channel: 140 rssi: 18	-85
2019-02-21 11:00:59.000	F0AP04	Plaisio_FreeWifi	iPhone	802.11 association	channel: 11 rssi: 10	-89
2019-02-21 11:04:15.000	F0AP04	Plaisio_FreeWifi	iPhone	802.11 association	channel: 112 rssi: 15	-87
2019-02-21 11:04:48.000	F0AP06	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 40	-75
2019-02-21 11:19:17.000	F0AP03	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 46	-72
2019-02-21 11:19:53.000	F0AP05	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 34	-78
2019-02-21 11:23:45.000	F0AP03	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 44	-73
2019-02-21 11:26:46.000	F0AP06	Plaisio_FreeWifi	iPhone	802.11 association	channel: 44 rssi: 30	-80
2019-02-21 11:37:20.000	F0AP02	Plaisio_FreeWifi	iPhone	802.11 association	channel: 40 rssi: 32	-79

Figure 52. Wi-Fi dataset transformed format

Following, we transform these data into the structure presented in Table 16.

	Beacon id data			Distance (meter)			Signal Strength (db)		
Area	**WiFi_id _1**	**…**	**WiFi _id_7**	**Distance_ 1**	**…**	**Distance _7**	**RSSI_1**	**…**	**RSSI_7**
F0AP0	AP0	…	AP3	3.21	…	5.96	-85	..	-89
F0AP3	AP3	…	AP0	1.02	…	6.17	-73	…	-98
….	…..	…	…	…	…	…	…	…	…
F0AP8	AP8	…	1P7	2.03	…	3.25	-92	…	-99

Table 16. Dataset structure

During the offline phase, we exploit machine learning techniques to train our model in order to predict the area where the user is located. Based on our approach, the dependent (class) variable is the Area and the independent variables (features) are the measurements produced by the Wi-Fi routers IDs, the Distance and the Signal Strength. We have further experimented with the algorithms by examining the most efficient number of beacons to take into consideration.

The data supplied to the different machine learning algorithms is the data set structure exhibited in Table 16.Area is the dependent variable and has different class labels that correspond to specific regions of the store. We experiment with 3 stores; thus the area variable is 14 for store 1 and 10 for stores 2 and 3. The selected data-driven approach requires the design of an efficient sampling process with regards to increasing the internal validity of the experimentation. In the forthcoming section, we present the classification algorithms we have used and compare their classification capabilities.

6.5 Evaluation

As mentioned in research background, the most common algorithms and methods for the location comparison and estimation are probabilistic methods, k-NN (k-nearest neighbours), neural networks and Support Vector Machines

(Liu et al., 2007). To this end, the classification algorithms that have been used to address the indoor positioning problem are the following: (i) Naïve Bayes (NB) (Lewis, 1998), (ii) Support Vector Machines (SVM) (Hearst et al., 1998), (iii) Logistic Regression (LR) (Anderson, 1982), (iv) Decision Trees (C4.5) (Quinlan, 2014), (v) Multilayer Perceptron Neural Networks model (MLP) (Kosko, 1992), (vi) KStar (K*) (Cleary and Trigg, 1995) and (vii) Random Forests (RF) (Breiman, 2001).

Statistical models (e.g. Naïve Bayes), Decision Trees (e.g. C4.5), Analytical Models (e.g. Support Vector Machines), Instance based models (e.g. K*) and Neural Networks (e.g. MLP) are representative algorithms from the machine learning field and have been employed in the relevant literature as an enabler to achieve higher localization performance results. Therefore, we decided to examine these along with other approaches. Naïve Bayes was selected as it is one of the simplest classification algorithms with strong independence assumptions between the features. Additionally, we selected Logistic Regression to examine the classification problem as a generalized linear model and KStar, an entropy-based algorithm, to investigate their performance in such classification cases. The software used for the experimentation with the aforementioned algorithms is Weka (Witten et al., 2016). The focus of prior relevant research works is the improvement beacons accuracy in terms of meters while utilizing small amounts of equally sized surfaces. Our study utilizes a variable-length surface to perform indoor position in area-level detection, thus a direct comparison to existing studies is not applicable due to different settings and environments.

6.5.1 Classification Performance Comparison

In this section, we compare the efficiency of each classifier based on the metrics of accuracy, precision, recall and f-measure (or F1 score) (Manning et al., 2008). Accuracy measures the number of correct classifications performed by the classifier. Precision indicates the exactness of the classifier, meaning that higher and lower precision leads to less and more false positive classifications respectively. Recall measures the classifier's completeness. Higher and lower recall means less and more false negative classifications (the data are not assigned as related to an area, although they should be) respectively. Precision and recall are increased at the expense of each other. Thus, they are combined to produce the weighted harmonic mean of both metrics, which is the F-measure. The evaluation of all algorithms is performed via 10-fold cross-validation (Kohavi, 1995), where the original data are randomly divided into ten equal subsets. Of these ten subsets, one is retained as the validation test of the model and the remaining 9 are used as the training data.

We start by examining the efficient number of access points, in order to perform efficient indoor positioning. For each of store settings we provide the metrics of the examined classifiers. We have also experimented with number of areas. Our experiments indicated that the most efficient number of areas in the case of Wi-Fi access points is the number of available access points. Having the prior knowledge of BLE Beacons, and a few indicative sampling evaluations we conclude to the following classification algorithms for experimentation (i.e. Random Forest, C4.5, K* and Naïve Bayes). We skip presenting store 3 as the layout and the access point's positions are the same, thus they lead to similar results.

		Classification			
Aps number	Areas	RF	C.4.5	K*	NB
7	8	70.8	64.9	64.5	47.5
6	8	70.3	63.1	66.7	47.1
5	8	69.9	64.1	68.5	46.5
4	8	68.4	62.8	68.1	45.3
3	8	69.2	62.4	69.4	44.9

Table 17. Store 1 – Ground Floor APs comparison

		Classification			
Aps number	Areas	RF	C.4.5	K*	NB
7	6	44.6	43.5	55.7	54.9
6	6	44.4	42.9	57.3	54.4
5	6	43.9	43.2	58.1	53.8
4	6	42.8	42.1	57.8	52.3
3	6	43.3	42.2	59.2	51.1

Table 18. Store 1 – First Floor APs comparison

		Classification			
Aps number	Areas	RF	C.4.5	K*	NB
7	3	55.1	52.4	63.3	66.1
6	3	54.8	51.8	64.4	65.2
5	3	54.1	52.1	65.7	64.1
4	3	53.7	51.3	64.1	63.5
3	3	53.3	51.2	66.2	62.9

Table 19. Store 1 – Second Floor APs comparison

		Classification			
Aps number	Areas	RF	C.4.5	K*	NB
7	10	83.3	78.0	82.7	83.8
6	10	83.1	78.2	83.6	83.7
5	10	82.9	77.6	83.2	83.4
4	10	80.8	76.5	80.8	80.9
3	10	80.9	76.5	80.9	80.9

Table 20. Store 2 APs comparison

Thus, bellow we present the detailed classifiers metrics for each store and floor.

We summarize the quality of classification using a set of related metrics such as

Accuracy, F-measure, Kappa statistic, Mean Absolute Error and Root Mean Squared error. We also provide the variance in parenthesis for each metric.

The results suggest that Random Forest outperforms all the classification algorithms in most of the aspects.

	Classification Algorithms					
Metrics	**NB**	**LR**	**C4.5**	**MLP**	**K***	**RF**
Accuracy	67.54% (1.92)	67.36% (1.91)	78.09% (1.53)	71.89% (2.25)	82.79% (1.51)	**83.34%** (1.21)
F-Measure (Weighted Avg.)	0.671 (0.15)	0.665 (0.14)	0.775 (0.11)	0.712 (0.18)	0.823 (0.11)	**0.828** (0.08)
Kappa statistic	0.6213 (0.06)	0.6174 (0.06)	0.7437 (0.05)	0.6706 (0.05)	0.799 (0.05)	**0.8051** (0.05)
Mean Absolute Error	0.0767 (0.07)	0.0871 (0.06)	0.0542 (0.05)	0.0655 (0.06)	**0.035** (0.05)	0.0492 (0.05)
Root Mean Squared Error	0.214 (0.06)	0.2073 (0.05)	0.1892 (0.04)	0.21 (0.05)	**0.1731** (0.04)	0.155 (0.04)

Table 21. Store 2 classifiers' performance

	Classification Algorithms					
Metrics	**NB**	**LR**	**C4.5**	**MLP**	**K***	**RF**
Accuracy	47.13% (2.10)	51.71% (2.05)	63.11% (1.63)	63.26% (2.35)	66.78% (1.81)	**70.30%** (1.51)
F-Measure (Weighted Avg.)	0.436 (0.20)	0.489 (0.19)	0.628 (0.16)	0.626 (0.23)	0.668 (0.16)	**0.699** (0.13)
Kappa statistic	0.3045 (0.11)	0.3692 (0.11)	0.5345 (0.10)	0.5328 (0.10)	0.5834 (0.10)	**0.6238** (0.10)
Mean Absolute Error	0.1723 (0.10)	0.1745 (0.09)	0.1205 (0.08)	0.1131 (0.09)	**0.0987** (0.08)	0.1315 (0.08)
Root Mean Squared Error	0.3126 (0.06)	0.297 (0.07)	0.2917 (0.06)	0.2964 (0.07)	**0.2855** (0.06)	0.2482 (0.06)

Table 22. Store 1 – Ground floor classifiers' performance

	Classification Algorithms					
Metrics	**NB**	**LR**	**C4.5**	**MLP**	**K***	**RF**
Accuracy	44.42% (2.12)	42.94% (2.07)	57.39% (1.65)	54.47% (2.37)	57.09% (1.83)	**67.39%** (1.53)
F-Measure (Weighted Avg.)	0.431 (0.22)	0.411 (0.21)	0.572 (0.18)	0.54 (0.25)	0.57 (0.18)	**0.67** (0.15)
Kappa statistic	0.315 (0.13)	0.282 (0.13)	0.4738 (0.12)	0.4373 (0.12)	0.4722 (0.12)	**0.596** (0.12)
Mean Absolute Error	0.2211 (0.12)	0.2276 (0.11)	0.1629 (0.10)	0.1649 (0.11)	**0.1455** (0.10)	0.1647 (0.10)
Root Mean Squared Error	0.3423 (0.08)	0.3385 (0.09)	0.3304 (0.08)	0.3408 (0.09)	**0.3584** (0.08)	0.2789 (0.08)

Table 23. Store 1 – First floor classifiers' performance

	Classification Algorithms					
Metrics	**NB**	**LR**	**C4.5**	**MLP**	**K***	**RF**
Accuracy	54.87% (2.12)	51.87% (2.07)	64.41% (1.65)	65.29% (2.37)	67.24% (1.83)	**73.01%** (1.53)
F-Measure (Weighted Avg.)	0.523 (0.22)	0.476 (0.21)	0.642 (0.18)	0.652 (0.25)	0.671 (0.18)	**0.728** (0.15)
Kappa statistic	0.2359 (0.13)	0.1709 (0.13)	0.4312 (0.12)	0.4482 (0.12)	0.4788 (0.12)	**0.5686** (0.12)
Mean Absolute Error	0.3653 (0.12)	0.3878 (0.11)	0.2748 (0.10)	0.2488 (0.11)	**0.2231** (0.10)	0.2568 (0.10)
Root Mean Squared Error	0.4415 (0.08)	0.4428 (0.09)	0.431 (0.08)	0.4303 (0.09)	**0.4412** (0.08)	0.3552 (0.08)

Table 24. Store 1 – Second floor classifiers' performance

Similarly to Case A, we use 10-fold cross validation and compare the classifiers using t-paired test to detect which are significantly better than the others (a=.05 level) (Salzberg, 1997). We employ F-Measure as a comparison metric because it is more reliable and provides a good trade-off between Precision and Recall. Similarly to Case A, we follow the design principles of Random Forest and suggest an ensemble classification approach to improve the localization accuracy for indoor positioning system.

6.5.2Proposed Filter

Instead of relying solely on one classifier, we decided to form a hybrid approach similar to Case A that involves a voting process that takes multiple classifiers into consideration. We chose the three significantly better algorithms based on the F-measure metric to form the ensemble classifier, which classifies new data by taking a weighted vote of the selected algorithms' predictions.

The proposed ensemble method we use (Table 10) seems to achieve a slightly lower accuracy than the highest classifier (i.e. Random Forest), however the absolute mean and root mean squared errors are lower. The accuracy of the ensemble method is almost equal to the highest classifier (i.e. Random Forest), but also the errors of the absolute mean and the root mean squared are lower.

The ensemble and the Random Forest classifiers achieve slightly similar results. None of them is significantly better than the other, thus they behave in a similar way in terms of accuracy. Regarding the mean absolute and root mean squared errors in the case of the ensemble method; these are significantly lower compared to the best classifier (i.e. Random Forest). Lower errors mean that the result is closer to the actual one, leading to more efficient location determination.

To visualize the performance of all classifiers examined, we utilize receiver operating characteristic (ROC) graphs; a technique used in machine learning and data mining (Fawcett, 2006). Figure 30 presents the trade-off between true positive and false positive rate for each classifier. The important points in this graph are: (a) point (0,0), which means that the classifier commits no false positive errors but also gains no true positives, (b) point (1,1) where the classifier commits positive classifications unconditionally and (c) point (0,1), which represents a perfect classification (Fawcett, 2006). A classifier is considered to perform better as it moves to the northwest part of the graph and also makes positive classifications only with strong evidence, as the false positive rate gets close to the Y axis (Fawcett, 2006).

Metrics				
Classifier	**MLP**	**K***	**RF**	**Ensemble**
Accuracy	65.29% (2.37)	67.24% (1.83)	73.01% (1.53)	74.18% (1.34)
F-Measure (Weighted Avg.)	0.652 (0.25)	0.671 (0.18)	0.728 (0.15)	0.738 (0.02)
Kappa statistic	0.4482 (0.12)	0.4788 (0.12)	0.5686 (0.12)	0.6532 (0.01)
Mean absolute error	0.2488 (0.11)	0.2231 (0.10)	0.2568 (0.10)	0.2154 (0.04)
Root mean squared error	0.4303 (0.09)	0.4412 (0.08)	0.3552 (0.08)	0.2872 (0.05)

Table 25. Ensemble classifier metrics comparison

The Random Forest classifier and the ensemble approach are considered to make positive predictions based on strong evidence, as they are close to the upper left point of the graph, followed by K* and Multilayer Perceptron. The rest classifiers (i.e. C4.5, Simple Logistic, Naïve Bayes and Support Vector Machines) tend to have a lower performance than the first ones as their distance from point (0,1) increases.

For further comparison between Random Forest and the proposed ensemble classifiers we transform the multiclass classification of the user's area into a binary classification to check whether the classifiers correctly predict the user's position. Figure 31 presents the ROC curve of Random Forest and ensemble classifiers and demonstrates their training speed.

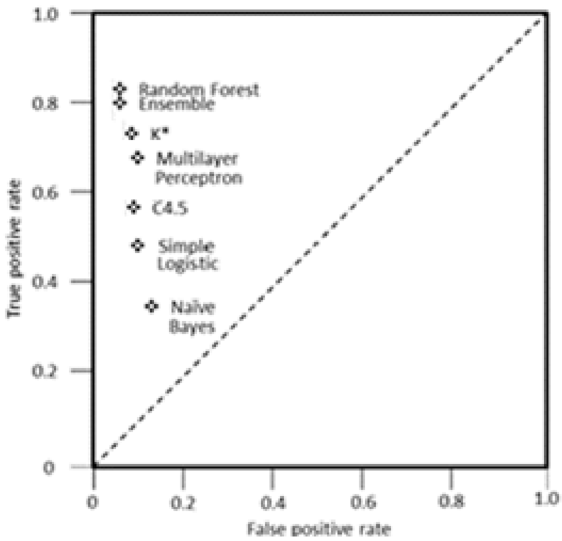


Figure 53. Classifiers ROC space graph

The ensemble classifier is characterized by a better learning ability as the training instances increase up to 0.65 True Positive rate. From 0.65 to 0.80 both algorithms tend to behave the same and above 0.8 Random Forest is slightly better. When both algorithms reach 0.95 of True Positive rate they have a similar behaviour, with Random Forest being slightly superior by 0.01. The ensemble approach improves detection accuracy; however, this cannot be generalized as it depends to contextual factors such as the store layout, product placement and the devices used for fingerprinting.

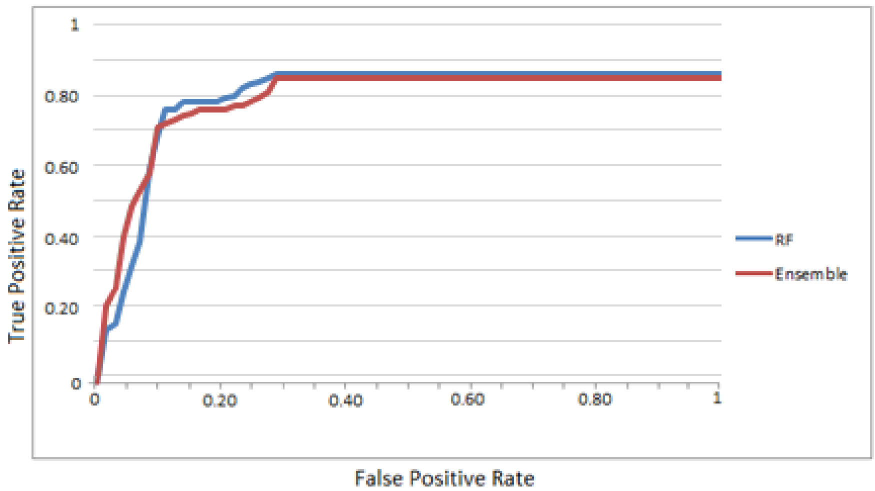


Figure 54. ROC curve

Finally, we present the overall cumulative probability and error correlation in Figure 32. Based on our experimentation we conclude that approximately a localization error of 70% of the cases is lower than 3 meter and in 80% of the cases the localization error is around 4 meters. The results confirm prior similar studies that try to improve the detection accuracy when using Wi-Fi access points.

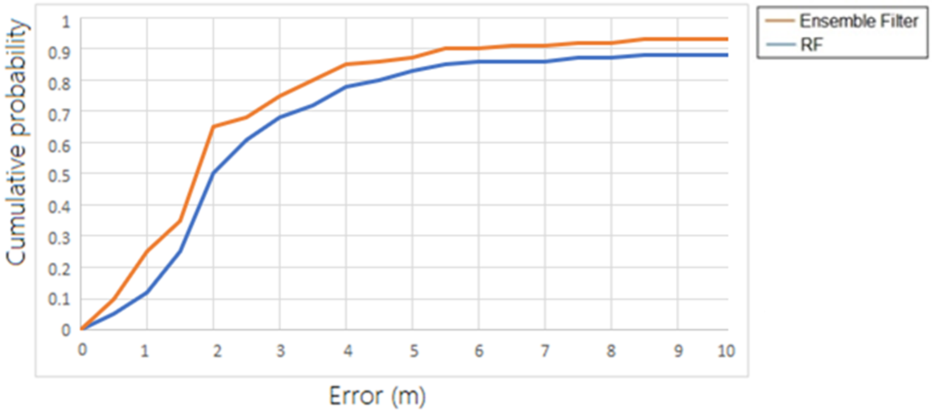


Figure 55. Positioning error for filter

The implementation of the indoor positioning system within a retail electronics store is affected by factors such as accuracy, scalability and cost (Liu et al., 2007). These performance metrics raise the question of how easily it can be developed and deployed in the store. The positioning system consists of a fix Wi-Fi infrastructure and the shop-specific filter. The infrastructure access points do not require be modifying. As a result, the part that needs to be modified is the ensemble filter that is trained based on the dynamic areas of the store. Regarding the cost of war-driving (He and Chan, 2016) for the ensemble filter it can be easily integrated in the daily operations of the electronics retail store and performed during the daily product check by the store manager.

6.6 User-based Evaluation

Apart from the technical evaluation of the classifiers, we go beyond that and evaluate the outcome by utilizing additional user input that will help interpret the findings. To this end, we use the feedback we received by indoor positioning system users to assess the quality of our findings.

In order to assess the performance of the indoor positioning approach we transform the positioning data into KPIs such as, time spent, user visits and area transitions, based on the spatiotemporal data we gathered from the real-time tracking system.

The indoor position of the shopper is a means of extracting the time (Figure 56) that a shopper spends in an area. Also, the sequence of area positions is able to extract the visits an area receives (Figure 57). Finally, we form the shopper flows from the area transitions (Figure 58), based on the data generated from the indoor positioning approach.

Figure 56. Time spent

The extracted time spent for each area was verified by the retailer's team. The qualitative results they pose from the store insights indicate similar time spans for the store areas.

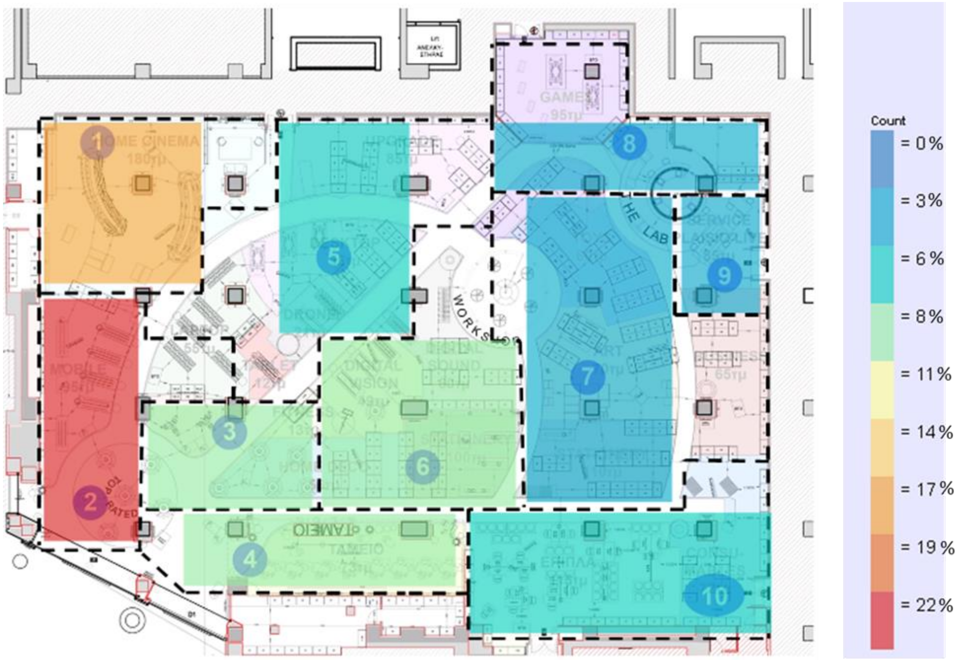


Figure 57. Area visits

In addition, the area popularity (i.e. the area visits) highlights the areas that they know that generate traffic and receive most of the shoppers. Finally, the shopper flows highlight the way that shoppers browse the store. So far, such observation was only available by observation and in a limited way. Using spatiotemporal data, the retailer is able to track larger sets of users in order to extract insightful results.

To this end, the retailer's team that operates and examines such information evaluated the results of our indoor positioning approach. Their feedback indicates high quality results and verifies the robustness of the outcome and the findings of the positioning approach.

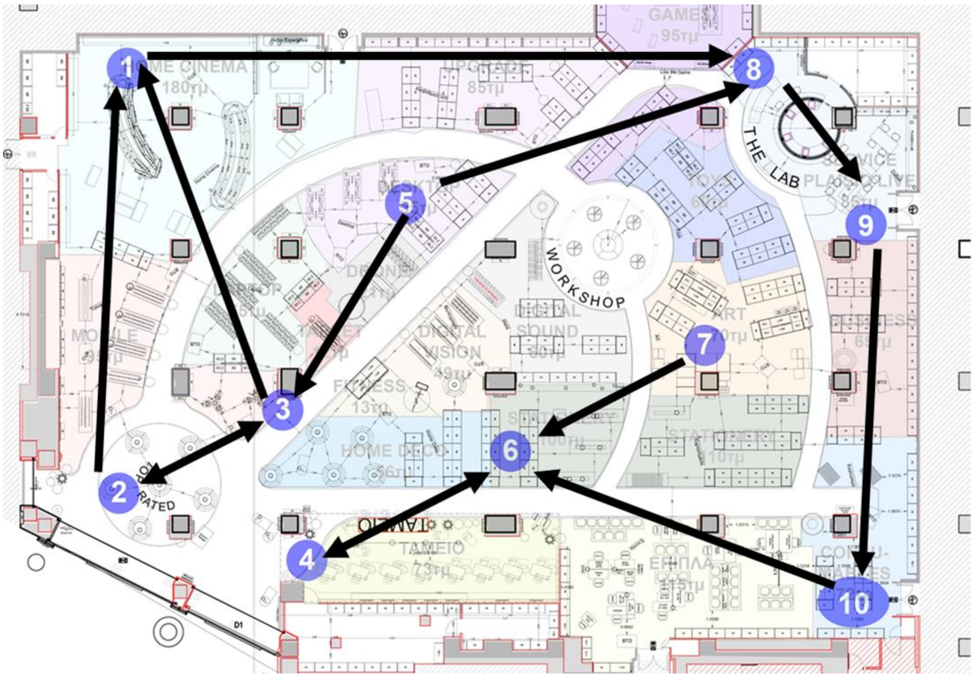


Figure 58. Shopper flows

7 MANAGERIAL IMPLICATIONS

Shopper position can generate enormous insights for decision makers. In the following sub-sections, we present a series of practical implications that derive from shopper indoor position and are broken down to two major categories; i.e. Retail Operations and Marketing Insights. We begin by discussing Retail Operation and discuss insights regarding the value of position in the design of location-based services, as it was extracted by a semi-structured focus group (see Appendix B). We continue with the discussion of location-based KPIs in retailing and the lessons learnt when deploying indoor positioning services. The we refer to Marketing Insights and discuss location-based shopper segmentation which can be of great value for retailers as it highlights the shopper behaviour. In addition, speed-based shopper segmentation can provide insights regarding the way shoppers navigate into the store, and how prone they are to promotional content.

7.1.1 Retail Operations

7.1.1.1 Location-based potential applications

The exploratory study upon the concept of indoor position involved a semi-structured focus group. Our discussion involved 22 experts from both retailers and suppliers in order to understand the way they apprehend indoor positioning and its value to the retail content. The structure of the focus group is further presented in Appendix B. The main concept of the focus group was "The value of location information in designing location-based services in retailing". One of the main goals was to understand what the participants consider as "position". A few ones consider the location in terms of X-Y coordinates; others in terms of aisle and the remaining in terms of broader areas. The discussion continued by exploring ways of exploiting the location information in terms of services. A few highlights of the services that the participants proposed include (a) out of stock prevention, as areas with higher visits may lead to increased sales, (b) extraction of behavioral patterns based on the way shoppers move, and (c) services that will improve customer in-store experience.

An interesting part when designing location-based services is the criteria that will affect the developed service. For example, participants from the side of the retailer consider the cost, the user experience, the increase on sales and increased traffic in certain areas as critical factors that will affect the design of the service.

Marketers consider the user experience, the increase on sales and the brand awareness as the critical factors that will affect the design of the process. They consider that a loyal shopper is loyal to both the retail chain and the brand, leading to sales increase. Finally, from the shopper perspective, the participants considered important the factors of time, value, shopping experience and trust, as the ones that play the major role when designing a service for the shoppers.

7.1.1.2 Location-based KPIs in Retailing

KPIs should be defined appropriate in order to support business goals (Gupta et al., 2013). To this end, they should not be complex, meaning that the decision maker is able to understand it, regardless the department he/she belongs to. In addition, the KPIs should be relevant and applicable to the involved processes and available in short time so as the decisions are made in the appropriate time period. Finally, the KPIs should be instantly useful, meaning it should be useful and used in a short time period.

Trying to extract KPIs aligned to the objective of supporting efficiently the decision-making process; our goal was to extract value from the moving behaviour of the shopper in the store. In order to extract value from the processed data, we calculate a series of KPIs for each location in the store. Table 26 briefly introduces the main symbols that formulate the base for the definition of marketing intelligence metrics.

Symbol	**Definition**
i	Shopping trip [1..N]
j	Area [1..M]
A_{ij}	a_{ij}=1, when the shopper of shopping trip i visits area j a_{ij}=0, when the shopper of shopping trip i does not visit area j
P_{ij}	p_{ij}=1, when the shopper of shopping trip i purchase from area j p_{ij}=0, when the shopper of shopping trip i does not purchase from area j
T_{ij}	The time (t_{ij}) spent in an area j by a shopper during shopping trip i
V_j	The sum of shoppers visiting specified area j

Table 26. Symbols and definitions

(i) **Average time spent** (ATS): This metric depicts the average time shoppers spent in area j during their shopping trips i. We consider as Trip Total time the moving time within store areas, along with the purchasing and waiting time. Even though retailers implement marketing strategies based on store traffic, time spent in each area is an important metric because as literature (Ferracuti et al., 2019) suggests, the higher the time a shopper spends in the store, the higher the possibility the shopper will make purchases. We quantify Average Time Spent as the ratio of the total time spent in the area divided by the total number of shopping trips in this area:

$$ATS_j = \sum_{i=1}^{N} \frac{t_{ij}}{a_{ij}}, \forall\, j \in [1..M]$$

(ii) **Store Surface Coverage** (SSC): The need to identify in-store behavioural patterns led researchers to examine the store area covered on a shopping trip. Recent studies have verified that very few shoppers visit the whole store during a shopping trip (Silberer et al., 2007; Sorensen, 2017), while the visited areas of the store are inversely proportional to the size of the store (Hui et al., 2009). This metric helps to identify the most and least essential areas for a shopper during their shopping trip (Ferracuti et al., 2019). Specifically, SSC metric derives from the division of the areas that a shopper visited during a shopping trip i, by the overall number of store areas. The outcome reflects the percentage of the store surface covered by the shopper.

$$SSC_i = \sum_{j=1}^{M} \frac{a_{ij}}{M}, \forall\, i \in [1..N]$$

(iii) **Purchases surface coverage** (PSC): Similar to the Store Surface Coverage (SSC), this metric helps to identify the percent of the store that a shopper's purchases come from given that the shopper visited the area. PSC metric derives from the division of the areas that shopper purchased from during the shopping trip i, by the overall number of store areas. The outcome reflects the percentage of the store surface where shopper's purchases came from.

$$PSC_i = \sum_{j=1}^{M} \frac{p_{ij}}{M}, \forall\, i \in [1..N]$$

This study suggests that when SSC and PSC are combined emerge patterns that reveal the way shoppers browse the store and the percentage of the store surface where purchases come from.

(iv) **Impulse buying margin** (IBM): This metric indicates how many of the visited areas, where the shopper spent time in, led to purchases. IBM derives from the division of PSC by SSC and the outcome reveals to what extend there was impulse/planned buying behavior. This metric is essential for managers as the information it reveals aids them in recognizing the level of impulse/ planned buying behaviors. Subsequently, managers may exploit the opportunity to turn impulse/ planned behaviors into sales using the appropriate strategy.

$$IBM_i = \frac{PSC_i}{SSC_i}$$

Typically, IBM values concentrate around 1 and express the average shopper behavior where shoppers tend to purchase products from their areas of interest. Extreme values near 0 or much greater than 1 indicate either mostly impulse purchases or strictly planned accordingly. Investigating intermediate values, we observe that values of IBM < 1 (empirically <.77) signal the decision maker that the shopper i devoted a significant amount of time to most of the areas they visited without necessarily making a purchase. On the contrary, values of IBM>1 (empirically >1.30) indicate that shoppers spent less time in each area and made purchases. Therefore, it is evident that IBM<.77 reflects impulse behavior, whereas IBM>1.30 reflects planned behavior.

(v) **Shopper speed**: An additional innovative KPI that refers to how fast shoppers browse the store. We approach shopper speed as a 3-value scale; i.e. low, medium and high value speeds. An issue when calculating shopper speed is that each shopper has a different speed pattern, meaning that a speed of 1m/s may be a fast speed for certain shoppers, but for others may be a slow to medium pace. Thus, we decided to adjust these scales depending on each shopper per se. By examining each shopper individually, we decide which speed ranges refer to low, medium and high speed for him. This KPI is calculated for the BLE beacons case and have not been introduced in literature for this technology before.

These KPIs can be further combined with sales metrics, such as sales per area in order to extract more complex insights like the missing sales. This examines the areas with high number of visits or traffic combined with low sales. We visualize these KPIs into heatmaps (see Figure 59), that depict each metric in a color-scale with high values being close to red color, while low values are close to green color.

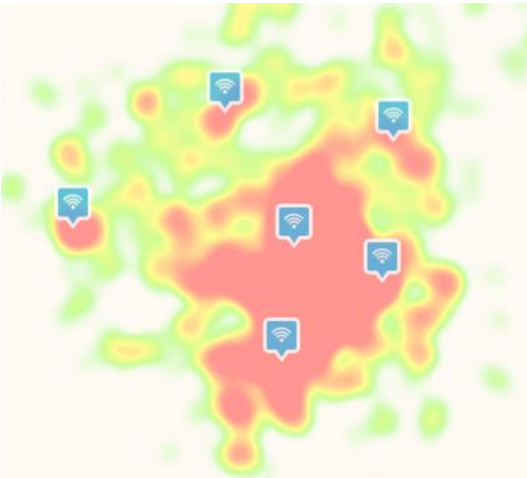


Figure 59: Heatmap representation

7.1.1.3 Lessons learnt and guidelines for indoor positioning systems

This section presents a location-based coupon recommendation service deployed in a retail store, which utilizes Bluetooth Low Energy Beacons technology to track the consumers in the store. The goal of this section is to share the development and application challenges we encountered, the factors we identified affecting the system's quality and performance; and show how we handled all the arising issues. Most studies in the literature either examine solely technical factors that affect indoor positioning (Lymberopoulos et al., 2015; Kuflik et al., 2011) or refer to the techniques increasing the user acceptance when using indoor location-based services (Yoon et al., 2017). Our study moves forward and includes also the business and user acceptance challenges, highlighting the role of the application context. This paper aspires to provide realistic and more holistic, not just technical, guidelines on prospective researchers and designers of location-based services for retail stores and other contexts, as well as encourage them to embrace such design initiatives.

7.1.1.3.1 Management System: Deploying an Indoor Positioning Service for a Coupon Recommendation Application

The presented case study deals with the deploying an indoor positioning service for a coupon recommendation application in a grocery retail store. The service is built upon an indoor positioning system that provides the location of the user. Thus, it is used to provide a coupon recommendation application based on the position of the user. The system relies on Bluetooth technology and in specific on transmitter devices called beacons that emit radio signals to nearby devices. These signals contain information regarding the identifier of the beacon and other data such as its RSSI strength. Beacons are a sufficient technology that can be used for this purpose.

In order to perform indoor positioning using BLE beacons the position system overview is presented in Figure 60. The system consists of the following three major modules: (A) the BLE beacon environment, (B) the indoor positioning mechanism and (C) the data knowledge and coupon management that stores and extracts knowledge from collected data and provides the management service of the coupons provided to the service user.

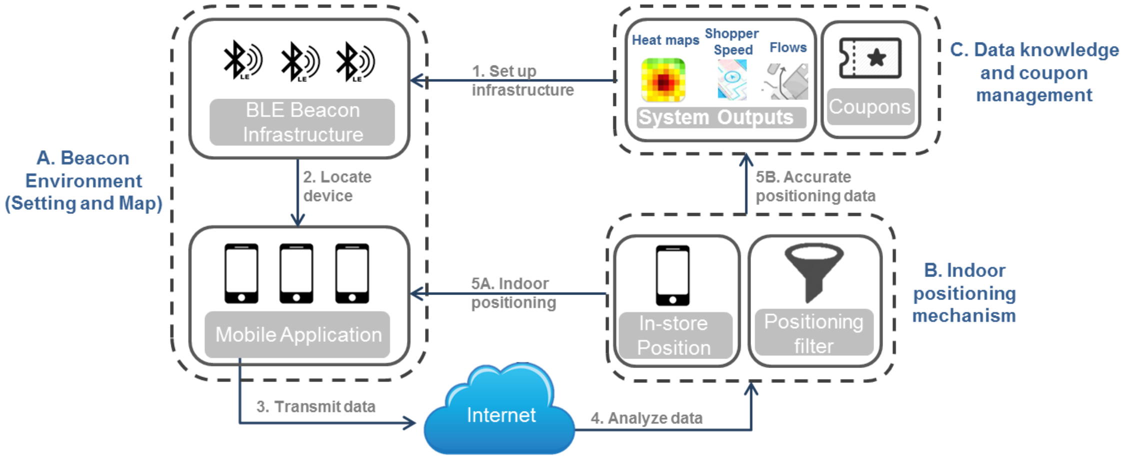


Figure 60. Indoor positioning system overview

More precisely (A) the BLE beacon environment (setting and map) consists of the following modules:

(I) The BLE Beacon Infrastructure layer refers to the beacons that are deployed in the store and the data that are transmitted and captured by the mobile applications of the consumers. It's practically the setting upon which the system is deployed. The BLE beacons can be deployed in various ways in order to achieve the required purpose (e.g. indoor positioning, proximity marketing etc.). Depending on the BLE beacon manufacturer and the available sensors on the beacon device, the environment can provide various form data, such as the id of the beacon device, the strength of the received signal (RSS), the distance between the beacon and the mobile device of the user and information such as light strength, humidity or temperature.

(II) The Mobile Application layer which utilizes an application that captures beacon data during the in-store movement of the user and communicates them via the internet to the backend infrastructure that stores the received data from the BLE beacons and the id of the mobile application. In addition, the mobile application receives the instore position module of the backend infrastructure each time that a request is sent with the captured information from the BLE beacons. When the user is in an appropriate position the respective coupon is offered for redemption via the mobile application.

(B) The indoor positioning and mobile app module is responsible for two tasks; i.e. (i) the in-store position of the consumer that communicates with the mobile application to provide the current position and (ii) the positioning filer which processes the collected data from each consumer to provide accurate positioning data for consumer traffic heatmaps and flows visualization. The data captured by the mobile application are stored in the backend infrastructure for further processing, as well as for the generation of the appropriate meta-data. The data from the backend infrastructure are used by other applications or ser-vices such as consumer traffic heatmap reports, user navigation and flows visualization (i.e. the administrative and data analytics module).

The (C) data knowledge and coupon management module has to do with knowledge management the monitoring and the results of the processes which involves consumer traffic heatmaps, navigation and flows visualization. This module is responsible for extracting insights and knowledge from the data that are captured by the mobile devices of the users and stored in the system's infrastructure. Also, this module is responsible for addressing the appropriate coupon recommendations to the user. As the user navigates in the store environment the mobile device communicates to the back-end infrastructure the signals it captures from the BLE beacons emitted in the store. These data are processed by the positioning filter in order to determine the user position and transmit back the position to the device along with the available coupons to be claimed (see Figure 61).

Goal of a design system is to provide additional Key Performance Indicators (KPIs) to its stakeholders in order to gain insights regarding consumer behaviour. Thus, knowledge such as the time spent in a store area, or the number of shopper visits in a specific store area is information that can be exploited to provide insights.

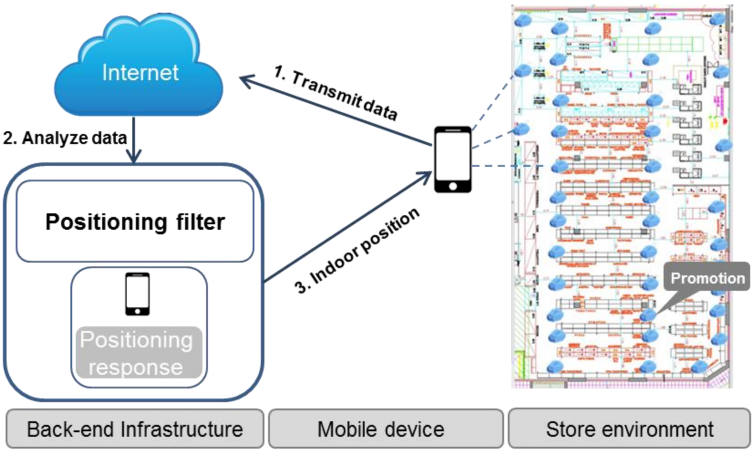


Figure 61. Coupon recommendation overview

In addition, shopper speed is a KPI that can be calculated using the BLE Beacon technology. The number of BLE Beacon transmitters can provide area level localization and as a result calculate the shopper speed by dividing the time spent in the area by its length. The transmitting data can highlight the timespans that the shopper does not move, and thus exclude these data from the speed calculation.

7.1.1.3.2 Challenges and Lessons Learnt

Indoor positioning opens a new era in the location-based services. However, several issues may arise when developing positioning solutions, varying for accuracy issues of the system or the technique that determines that location to the type of the mobile device used. Apart from that, the deployment of an indoor positioning approach to the grocery store required a lot of effort in terms of transmitter placement, positioning algorithms and determining the types of potential applications to operate over BLE beacon transmitters. The lessons we learnt and the knowledge we elicited by the design issues we encountered in each step of the process, while developing the infrastructure presented in the case study section are presented below.

The BLE beacon infrastructure deals with beacon placement, configuration and user's position. During the development of this architecture's module we dealt with several issues some of which are still open in the literature. In the remaining cases we reviewed existing literature and advanced our approach for more effective outcomes. Below we describe the challenges met at each stage of the deployment process and append the lesson learnt while overcoming these challenges.

7.1.1.3.2.1 Process Setup

We began to deploy our positioning mechanism by initially deciding the requirements that the coupon recommendation application would provide. As a first step, we had to develop an indoor positioning system upon which the application would be deployed and offered to the users. The deployment of the indoor positioning system had to do with the following steps:

(a) Beacon placement: The process started by choosing the appropriate number and position of the transmitters for our infrastructure in the store.

(b) Beacon configuration: Following we had to assign IDs and calibrate the signal strength for each one of the BLE beacons. This calibration affects the coverage that each beacon will emit its signal.

(c) Define areas of interest: Having the wireless infrastructure set, we had to decide with the store manager the areas of interest in terms of area accuracy. Thus, we performed indoor positioning in aisle level, meaning that the system detected the presence of the user when browsing in store aisles.

(d) Develop positioning filter: The positioning system operates by having a database of the signals that are emitted in each area and based on them the system can decide where each user is browsing in the store. This process refers to the positioning filter which is a mechanism training by a data set which contains signal strength capturing for each area and based on this knowledge responds the location every time it reads new unknown signal from the mobile device of the user.

(e) Use application for fingerprinting: The capturing of the signals emitted during the development of the positioning filter requires an application to be developed. This application captures signals emitted from BLE beacons and assigns them to the pointed area that is captured.

(f) Develop mobile application: From a user perspective, a mobile application is required in order to interact with the BLE beacon infrastructure, detect user's location (based on the positioning filter), and provide the coupon recommendation service when the user is tracked in specific areas within the store.

(g) Positioning filter: For each store that this service is offered, a separate shop specific filter is developed and store for use.

(h)Refresh training data: In addition, as products may be moved to other aisles or new promo stands may be positioned in the store, a refresh of the positioning filter is required in order to be updated based on the changes that may occur in the store.

(i) Having set the indoor positioning system, we further deployed a backend module with a database of the available coupons for the user. Thus, when detecting the user at specific position, the application would push to the user the appropriate coupon from the backend infrastructure.

7.1.1.3.2.2 Technical Challenges

During the development and deployment of the indoor positioning module of the location-based service we encountered several technical challenges, most of them like the ones pointed out in the related work section.

The procedure followed to deploy our architecture is summarized to the following steps:

(i) *Transmitters placement*: We experimented with various setting formats (grid/open space layout) and placements (on top of the aisles/on the ceiling).

Beacon placement within a store is a challenge that needs to be taken into consideration. The challenge of effective transmitter placement in indoor environments is still open, as there is no single solution that is able to fit every case. Therefore, we decided to place the transmitters based on the needs of our case. We started by placing the transmitters on the store's aisles. As a result, during this initial deployment on the aisles we observed that obstacles either interfere or totally blocked the signal emitted by the transmitters. This occurred due to the replenishment process, where products would be placed over the transmitters and thus blocking the signal emission. Therefore, what we have learnt is that placing beacons over the crowd decreases the chances of signal's interference because of the shoppers between the user device and the beacon.

Thus, installing transmitters on the ceiling seems better than having the transmitters on the aisles' shelves as the line of sight between the transmitter and the shopper is less likely to be disrupted. In addition, sticking beacons on the ceiling reduces the chances of being stolen or pulled off by accident. Regarding the replacement of beacon's battery, it's a process required every 2 or 3 years, so it's a matter of a couple of hours to replace the batteries.

(ii*) Beacon configuration:* We experimented with different calibration settings to achieve the desirable trade-off for battery life and signal strength.

Beacon technology is not yet mature enough. As a result, predefined configurations do not apply to most of the cases that they are deployed. In general, beacons can be configured for indoor location by setting the appropriate value settings such as location packet's broadcasting power and location packet's advertising interval time. Though, it affects significantly the battery life, as the device is operating in a more intense way to broadcast packets in shorter time intervals. Such parameters can improve signal's strength quality and thus, achieve accurate results regarding position determination. However, it is always important to find an adequate trade-off between battery life and signal's strength.

(iii) *Define areas of interest:* We experimented with different granularity levels (i.e. X-Y coordinates/area level) in order to achieve the desirable accuracy. X-Y coordinates did not perform well, so we abandoned using triangulation and moved to an area level approach which behaved significant efficiently.

(iv) *Develop positioning filter*: In order to achieve the desired accuracy, we experimented with various positioning techniques (i.e. trilateration, fingerprinting, machine learning approaches and ensemble classifiers) evaluating their performance based the setting of transmitters we had deployed.

Literature examines trilateration with technologies such as Wi-Fi and RFID. When dealing with Bluetooth technology this technique seems problematic. Thus, we demonstrated that with the current state of this technology trilateration cannot be used for effective user positioning. Trilateration performs well when the error rate of the estimated distance between the fixed point and the tracking object is low. Thus, Beacons cannot be used to determine the exact location of the shopper in terms of x-y coordinates, as signal strength is affected by various factors and cannot be precise enough and it usually greatly deviates from the actual value regarding the distance of the user from the transmitter. Therefore, the distances from the beacon transmitter are mostly unreliable. Hence, what we have noticed is that feeding problematic and inaccurate values into a trilateration algorithm provides as output faulty results leading to the conclusion that trilateration should not be used as is but requires further techniques or technologies in order to get better signal quality. Beacons cannot be used to determine the exact location of the shopper, in terms of X-Y coordinates, as signal strength is affected by various factors and cannot be precise enough regarding the distance of the shopper from the transmitter. In contrast fingerprinting technique has been found to be more effective for indoor localization.

(v) *Develop mobile application*: The development of a mobile application for the location-based service requires to satisfy a few requirements. In our case we developed an android application. As a result, we had to develop the application for a minimum android version to achieve greater compatibility with users' devices. Another issue that we dealt with the mobile application was the network connectivity. In a few cases the mobile device had interruptions to the internet connection. As a result, the application was able to keep a data buffer and communicate the non-transmitted results later when the connectivity was restored.

We observed that the type of the mobile device plays an important role in the localization process. Our mobile application has been developed using the Estimote SDK for Android. This SDK is a library that allows the interaction between the BLE beacons and the mobile device. In order to perform effectively the Android version of the device should be 4.3 or above, as it is required by the developers of the Estimote SDK. Additionally, a device with Bluetooth Low Energy is required to run the mobile application. During our experiments, we encountered an issue with low cost devices. Although the Android version of the device met the requirements, the device would only capture signals from only one beacon, ignoring the remaining. Thus, no accurate estimation of the location is feasible based solely on one BLE beacon. We experimented with other devices too and reached to the conclusion that a device with an average cost would work with no problem in capturing the beacon Bluetooth signals and thus determining shopper's location. Finally, in Android versions greater than 7.0, the Estimote SDK does not allow the monitoring period to be lower than 6 seconds, so as it does not affect battery consumption. Hence, it is possible that a slight delay occurs when trying to locate beacons for the first time of the execution of the application.

In terms of factors that affected our location-based service, we developed a filter that achieves significant accuracy in terms of area level and covers the entire store where the service was deployed. In terms of complexity, the system was easy to deploy and be adopted and was robust as it has been developed to achieve high accuracy and absorve errors that can affect the accuracy. The system is scalable, as it can further expand to a larger space and can be deployed to an entirely different store without significant changes. Also, the cost of the system relies only on the development of the mobile application which is once-off cost and the cost of the infrastructure that depends on the number of transmitters/Aps deployed in the store. In our case we deployed 80 BLE beacon transmitters that are cheaper than other technologies (e.g. Wi-Fi).

7.1.1.3.2.3 Infrastructure Challenges

Apart from the technical challenges we encountered several ones by the wireless infrastructure used for our service (i.e. BLE beacons). As mentioned in the related work section, the wireless infrastructure poses a significant effect on the signal strength that comes from various factors, such as human bodies, obstacles and walls. The setting of our transmitters affected significantly the quality of the generated data, as the human bodies browsing the store affected the signals emitted. In the case we had the transmitters on top of the aisles the interference between the human bodies and the transmitters was much greater and we encountered severe loss of information as the user mobile device could not receive the signal from the closest transmitter. As a result, the indicated location of the user was different from the actual one.

Except the obstacles that affected the quality of the information, we came across another problem; that is the different type of devices used by the users. Mobile devices have different versions of Bluetooth antennas and consequently the value of signal strength received by each device is estimated with different value. To overcome this issue, we used different brands and versions of mobile devices in order to increase the captured dataset that feeds and trains the indoor positioning filter of our system.

The infrastructure challenges are mostly dealt with by developing an efficient positioning filter. Thus, apart from configuring the hardware of a beacon device, the position determination algorithms that receive as input data from the beacon transmitters can play an important role in the accuracy of the result. An important thing we learnt is that by applying different algorithms on position data may further improve the quality of the output of a position mechanism. In the presented case study, we utilize a hybrid approach of algorithms to increase the accuracy of the location determination and absorve infrastructure issues.

7.1.1.3.2.4 User Acceptance Challenges

The acceptance and adoption of a system by its users is an important factor for the success of its functionality. The system deployed in the grocery retail store had to be initially accepted by the employees of the store and from the customers that were the receivers of the service. The employees expressed several issues and objections regarding the radiation emitted by the BLE beacons. As soon as they got informed that BLE beacons operate at the same frequency of Wi-Fi (i.e. 2.4GHz) their objections became minor.

Customer acceptance is another factor that we encountered. During the use of the mobile application, users were informed that their browsing in the store would be recorded by the data captured from the BLE beacons. The user was prompted to consent to the capturing of the transmitter data during the shopping session. As soon as the user accepted the mobile application and the coupon recommendation application was available for use.

User incentives play an important role for the acceptance of location-based services, as user seeks a form of reward or value when consenting to provide personal data for such purposes. The personalized coupons were a productive and useful incentive as the results indicated high redemption. Finally, user friendly interfaces seem to have a positive impact in terms of user acceptance (Basiri et al., 2017), as users are more willing to use the mobile application.

Regarding the coupon redemption, we observed that the location-based service facilitated the redemption process. Users could select any of the available coupons that matched their preferences and the coupon became available each time that the user was near the assigned area. This process facilitated the user to pick coupons via the mobile application when she/he was near the areas of interest during the shopping trip. The effect of this process led to increased basket size by 2.14 products solely purchased because of the provided coupons. Discussing with the users of the service, they highlighted that the location-based coupon recommendation helped them pick only relevant coupons to their shopping trip without having to spend time looking for further information. In addition, they enjoyed the fact that they received coupons based on their preferences.

7.1.1.3.2.5 Business Challenges

The business challenges we encountered had to do with:

Accuracy: The retailer was mostly interested in acquiring knowledge at aisle level, thus the area accuracy we used was an adequate implementation.

Areas division: The most important factor when dealing with indoor positioning is the precision of the object's position. Literature examines indoor positioning either in terms of precision in meters, proximity or in terms of area. Thus, technology or other limitations (e.g. the environment) may be a critical issue that affects the applicable level of precision. BLE beacons is a wireless technology where indoor positioning is affected by signal attenuations and fading and should determine the positioning level based on other factors such as the environment, the setting and even the product material and are also related to the business requirements that may occur.

Studies in the literature (Wang et al., 2015) approach the issue of indoor positioning by segmenting the initial surface into equal smaller areas, thus forming a grid which can be then used to detect the exact segment where the object is located and applied iteratively in order to further split an area into smaller ones for higher precision. Using the grid over a surface seems an appropriate solution to detect the location of the object (i.e. fixed-length surface. see Figure 4). However, various factors such as the physical layout of the surface or technological requirements or business needs introduce restrictions that may affect the size and dimensions of the area and as a result the detection of the position of the object (i.e. variable-length surface).

The physical layout may lead to the adoption of areas with different sizes due to obstacles or walls that do not allow the sole use of areas with equal dimensions. In addition, technological restrictions may require a minimum threshold of area dimensions, as signal attenuations may be unable to effectively detect the position of the object. Finally, business needs may indicate that some areas are characterized as being of more interest than others. For example, chocolates are more popular during winter, while ice-creams are more popular during summer. As a result, low interest areas can be grouped into bigger-sized areas, while high interest areas can be split into smaller-sized areas.

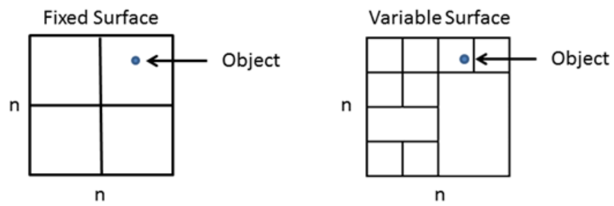


Figure 62. Fixed and variable length surfaces

In our case we utilized a variable-length surface conceptual model, i.e. areas with different size and dimensions due to limitations that occur from the layout of the store, the deployment of BLE beacon transmitters and the business needs defined by stakeholders. It's a more realistic and effective conceptual model that can be easily adopted for cases where multiple factors affect an indoor positioning mechanism.

Unit of analysis: Location-based services can either offer the service to a mobile application or to fixed base stations. In our case (grocery retail store) we decided to track users than trolley baskets. Shoppers usually leave their basket trolley and start shopping without carrying it around, leading to being unable to know where the shopper actually is. As a result, we were able to provide more efficient service to the final user, as the content was served based on the actual in-store position.

Step	implication	Challenges
Beacon placement	Grid layout	Technical
Beacon configuration	Set default settings or trade-off for battery life and signal strength	Technical
Define areas of interest	Create dynamic in-store areas for positioning	Business/Technical
Develop positioning filter	Select appropriate algorithm	Technical
Use application for fingerprinting	Fingerprinting in operational level that can generate training sets for indoor positioning	Technical
Develop mobile application	Provide motives and incentives in order the shopper uses the application	Technical/ User acceptance
Positioning filter	Each store has its training data	Technical/Business
Refresh training data	Data should be up-to-date and accurate regarding potential layout or product changes	Technical/Business

Table 1. Challenged during the development steps

Having developed the positioning system, an additional challenge is how to develop the location-based service for the user. To this end, we organized sessions with potential users in order to gather requirement and design the appropriate service functionality. User input is a critical factor during this process as this functionality will determine whether a user will use the service or not. In the following table we summarize the challenges we encountered in the developments steps and provide the implication for each one.

7.1.1.3.2.6 Discussion

By examining the studies in the literature, we summarize a generic architecture overview that depicts the basic modules that an indoor location-based service consists of (see Figure 5); i.e. (A) Localization infrastructure, (B) Data layer and (C) Administrative and data analytics. The localization infrastructure (A) module consists of the technologies and mechanisms utilized for the localization of the user within the environment of interest. Depending on the technology that is used as wireless infrastructure (e.g. Wi-Fi, RFID) the appropriate localization mechanism is used. For example, in the case of wireless technologies, a mobile user device is required in order to be able to communicate with the infrastructure and push the appropriate information to the user. When the localization infrastructure is camera, then no localization device is required, as the face recognition is used to detect user's location.

The data layer (B) refers to the information that is generated via the interaction of the user with the localization infrastructure. Such information refers to spatiotemporal data that capture the position of the user within the environment. These raw data are further process to extract meta-data to explore further information that can be formed by them. The administrative and data analytics module (C) is responsible for managing information regarding the functionality of the indoor location-based service. For example, in the case of a coupon recommendation service, this module is responsible for setting the coupon to be provided. In addition, this module usually provides insights using data analytics mechanisms and depicts KPIs in order to monitor and evaluate the performance of the service.

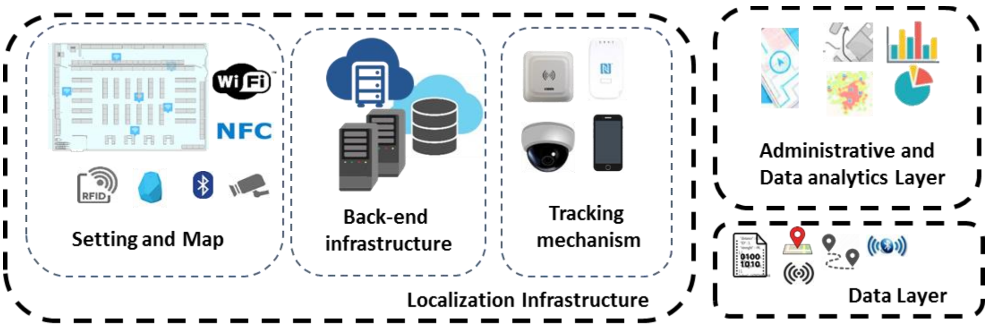


Figure 63. Indoor location-based service generic architecture

Finally, we map each type of challenge to the respective module of the indoor location-based service architecture. The technical challenges refer mostly to the localization infrastructure module as they are technical challenges and issues that should be dealt with at the indoor positioning system level.

The infrastructure challenges deal with mostly the data module and partially with the localization infrastructure module. The quality of raw data is a combination of the quality of the signals emitted by the radiofrequency infrastructure and the way that the access points of the infrastructure are deployed in the setting of the store. The business challenges affect each module of the architecture as they are a combination of the functional and non-functional requirements for the overall performance of the indoor location-based service. Finally, an additional factor of a location-based service is the user. The user interacts with the service and is the final receiver of the offered value. A location-based service without a final user is not able to achieve the goals of its functionality. Thus, the user should be convinced to utilize the provided service. As a result, the user acceptance challenges refer to the actor (user) of the service. Figure 64 presented the mapping between the generic architecture of location-based services and the type of challenges that address each of the modules.

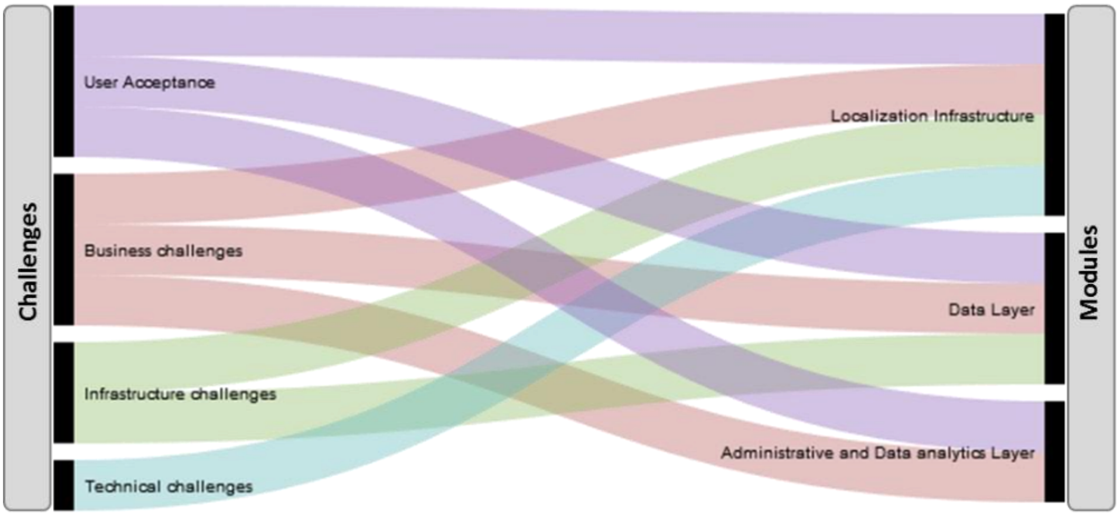


Figure 64. Challenges and architecture's modules

Figure 65 depicts the factors that affect indoor positioning and location-based services grouped by the module of the generic architecture (see Figure 63).

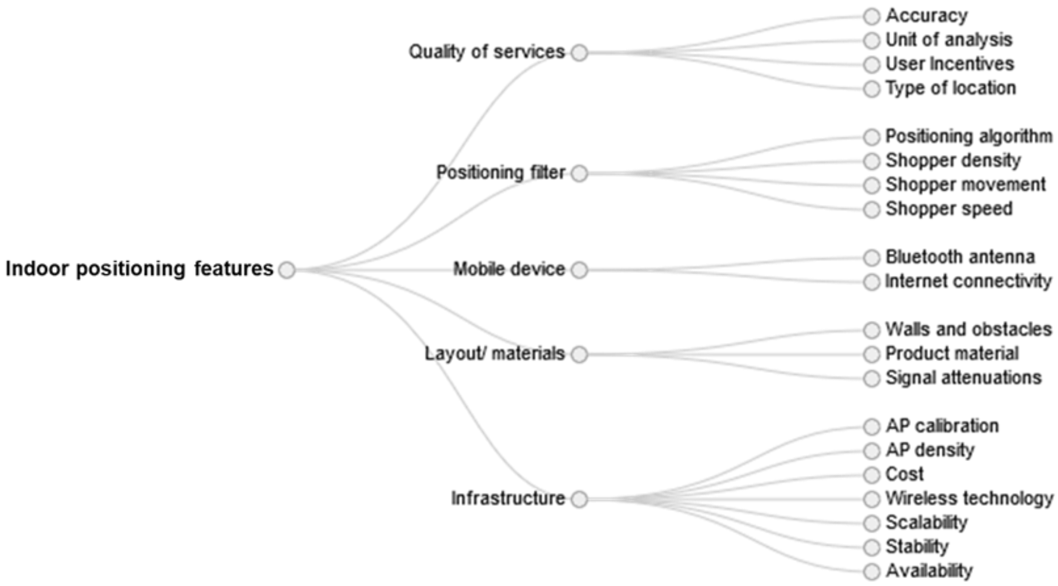


Figure 65. Factors affective each phase of the architecture

As depicted in Figure 65 , the positioning filter is affected by factors such as the selection of the appropriate positioning algorithm. Shopper density is an important factor for the positioning process, as shoppers affect the quality of the emitted signals as they get moving obstacles that interfere between the mobile device and the transmitters. Shopper movement and shopper speed affect the positioning filter as capturing the appropriate signals during the user browsing in the store determines the appropriate location of the user.

Quality of services is affected by system's accuracy, as if the system does not perform based on the actual user location, the value that it provides will not really be interested to the user. Tracking the user, instead of the trolley-basket will provide the actual location of the user as usually users tend to leave the trolley-basket and continue shopping without it. In addition, the appropriate selection of user incentives will affect the quality of the offered service.

The mobile device also affects indoor positioning, as the service should cover a wide range of Bluetooth antennas and have internet connectivity. The device is the service touchpoint as it is the means of interaction with the user. Store layout and product materials have an effect on indoor positioning. Both obstacles (i.e. walls, promo stands, passing by shoppers) and product material affect the signal quality and the caused signal attenuations affect the accuracy of shopper's position.

Finally, the factors that affect the infrastructure have to do with the access point density deployment and calibration. Increased access point density is able to provide greater granularity level in terms of more and smaller areas; however, the overall cost of the service increases. The wireless technology upon which the infrastructure will be based is also a factor that affects indoor positioning. Wi-Fi and Bluetooth are wireless technologies that facilitate indoor tracking with less cost than RFID technology that requires additional equipment (e.g. RFID tags and antennas) in order to perform accurately. Last but not least, the infrastructure should be scalable, stable and highly available in order to provide the location-based service.

7.1.2 Marketing Insights

7.1.2.1 Area-based Shopper Segmentation

Trying to examine potential features of shopper segmentation, we go beyond speed-based segmentation and examine the feature of the residence time in each area. For each shopper individually we examine the average time he/she spends in each in-store area. Thus, we create a benchmarking and detect the areas that stayed longer than usually during the shopping trip. Indicatively, we tried to avoid areas that indicated simple passing through them and tried to detect the time that indicates that the shopper interacted with the products of the specific area. This leads to the extraction of areas that the shopper spent more time than simple passing through the aisle.

Goal of this approach is to perform area-based shopper segment segmentation and discriminate shopper segment of similar behavior patterns based on the areas they visit in the store. We performed clustering based on the areas and the flag that indicates that the shopper interacted with the area and extracted 10 discrete shopper segments based on the behavior of the shoppers. Indicatively, Figure 66 depicts a shopper segment from grocery retail store and the map of the store's ground floor. The highlighted areas indicate the areas that shoppers of this segment tend to spend most of their time during their shopping session.

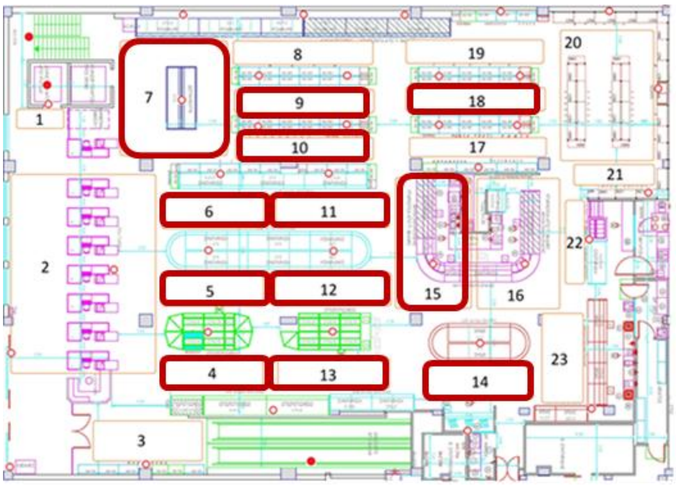


Figure 66: Indicative area-based shopper segment

In this example, shoppers stay longer in areas containing chocolate and chips, biscuits and beverages, ice-creams, juices and butters. The areas that this shopper segment visits indicate a visit for sweet, snacks and beverages. Applying a similar drilldown analysis for each of the shopper segments we aim at identifying the shopping purpose of their visit to the store based on the areas they browse. Thus, via the analysis of the shopper segments, we detect shoppers browsing areas that indicate visit for confectionery products, for meal preparation, detergents, personal care and house cleaning.

Area-based segmentation introduces new patterns regarding the in-store locations shoppers visit and try to discover the reason of the shopping trip solely by spatiotemporal data (i.e. the in-store areas and the path followed). Thus, the combination of the areas that a shopper visits may unravel information regarding the area he/she visits. By combining this information with shopper's purchases, we can detect sales gaps as we extract insights regarding the areas that the shopper actually spends time in but does not buy any products. These missing sales may facilitate decision makers to design their strategies in order to address this phenomenon. This combination of multiple sources of data highlights the opportunities that rise when adding dimensions to existing data sources, turning location analytics into a tool that helps decision makers to turn into value all the available sources of the organization. Finally, the use of two different wireless infrastructures (i.e. Wi-Fi and BLE beacons) indicates that despite the tracking mobile application in the case of BLE beacons, both cases have similar requirements and lead to similar KPIs and findings. The only requirement that must be met for both cases studies is the ability to identify the mapping of locations in the store and an id of the user that navigates in these areas. Then, the results can be further enriched with loyalty data (if available) and extract the described insights and KPIs.

7.1.2.2 Speed-based Shopper Segmentation

Apart from traditional features of customer segmentation, another useful feature that can be used is speed. Speed-based segmentation can reveal segments of shoppers that browse the store with different paces. In our grocery retail store case study, we performed shopper segmentation based on speed feature and extracted four major segments of behavior, which we combined with loyalty card data and the shopping baskets of their trips.

Our first segment indicated users that tend to navigate in the store with low speeds and usually with higher session duration. The loyalty data indicated that these shoppers are mostly women who have visited alone the store and they have come mostly for large (stock) baskets and products regarding meal preparation. This segment tends to spend much time in the store, thus making it more likely to respond to in-store promotional activities.

Figure 67: Speed-based shopper segmentation

The following segment refers to shoppers with low to medium speed pace. These shoppers are mostly couples who have visited together the store and they have come for products regarding their breakfast, personal care and house cleaning. These shoppers have been found to spend less time than women visiting the store alone, albeit they are also more likely to respond to in-store promotional activities.

The third segment refers to users that navigate with medium to high speed in the store. Using loyalty data, these shoppers are mostly men who have visited the store alone. They achieve significantly lower time duration during their shopping visit and come to buy mostly products related to their breakfast and light meal. As they spend less time in the store and navigate with higher pace that the first two segments, these shoppers are less likely to respond to in-store promotional activities. As a result, promotional actions before their visit to the store would have more efficient results. Finally, the fourth shopper segment is the one detected with shoppers navigating the store at high speed pace. These shoppers are mostly families or men who have visited the store alone. These shoppers tend to buy more confectionary products, snacks and beverages and, also, large (stock) baskets. This segment is also a segment that would not be likely to respond to in-store promotional activities; thus, a prior communication is more likely to be effective.

Speed-based segmentation reveals an additional dimension of shopper in-store behavior. Apart from the purchasing behavior, the way that shoppers browse the store may facilitate decision makers to design strategies of targeting shoppers from this perspective and examine the way they interact and respond to these strategies. In addition, stakeholders can experiment on various approaches of affecting shopper speed or improve shopping experience in order to detect potential changes in shopper behaviour.

7.1.2.3 Location analytics for marketing intelligence

One of the major areas of focus in regard to an evolving retail industry is understanding shopper in-store behaviour. In-store behaviour has drawn the attention of scholars over decades (Kollat and Willet, 1967; Donovan et al., 1994; Muruganantham and Bhakat, 2013; Sorensen et al., 2017), as shopper behaviour is constantly a topic of interest for retailers that seek shoppers' patterns and insights of their movement in physical stores. This knowledge not only facilitates retailers to better understand shopper visits, but also enables them to organize their operations more efficiently which consequently delivers higher in-store experience to the shopper.

Retail environments generate massive amounts of data from different channels (physical, web, mobile) (Bradlow et al., 2017) that include data varying from transactions and loyalty schemes to legacy systems. Based on them, researchers and retailers aim to understand and interpret shopper behavior via the formed shopping basket (Tang et al., 2008; Cil, 2012; Sarantopoulos et al., 2016; Griva et al., 2018). However, this approach has its limitations since it only examines shoppers' basket preferences and it lacks considering certain other aspects of shopper behaviour such shopper in store browsing or how the basket is formulated.

Another determinant characteristic of shopper behaviour is the shopper path (Sorensen et al., 2017). The study of shopper path may reveal insights based on the way shoppers navigate the store. Gathering and analysing data related to shopper path enables retailers and marketeers to study shopper in-store trajectories and extract patterns regarding a shopping trip.

In this study we utilize location and business analytics techniques to study how motion data, purchase data and shopper data can be combined to provide insights on shopper behaviour and build on marketing intelligence. To this end, firstly we cluster individual shopping trips and secondly, we associate the results with corresponding purchases. Thereafter, in order to study shopper behaviour in depth, we go beyond focusing solely on shopper attributes or purchases, and associate motion data with purchases and shopper information. Finally, we discuss how the association of these three dimensions (i.e. Basket, Location and Shopper) may lead to valuable information for the decision-makers. An interesting extension of this work includes examining the impulse buying behaviour that emerged through the collected data (Cobb and Hoyer, 1986).

7.1.2.4 Business Analytics for shopping trip investigation

Shopping trips could be studied within a single dimension, which is the most common approach, or combine several dimensions and draw a rich environment. To collect, combine and extract intelligence from such environments requires efficient Business Analytics approaches to turn data to insights. Upon this observation, we rely on three major dimensions (see Figure 68): (a) the Basket (B dimension) that answers the question "what is purchased", (b) the Shopper (S dimension) that addresses the issue "who purchased", and (c) the in-store Location (L dimension) that contemplates how the basked was formed in terms of the followed shopper path.

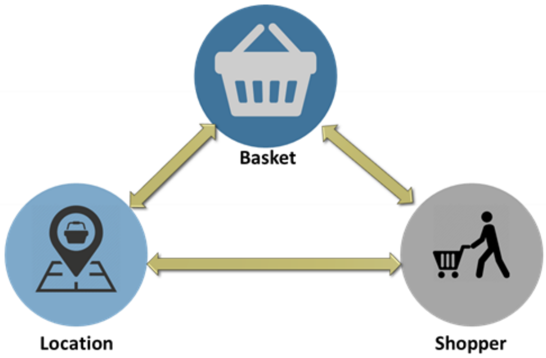


Figure 68. Business Analytic framework

Each dimension includes attributes that can be utilized during the clustering task, to draw a rich experimentation environment. The B dimension includes a vast variety of attributes ranging from quantity purchased to certain product characteristics and it is considered as the most common dimension since it is enabled by the Point-Of-Sales transactional system which is widely available in all retail settings. The S dimension consists of data such as demographics and loyalty schemes. The unique identification of shoppers is contingent on their willingness to utilize a loyalty program; therefore, the S dimension considers both identifiable and anonymous shoppers. The latter group is of great importance because it describes the vast majority of retail transactions while it includes people that visited the store and didn't proceed to any purchase. In the same manner, the L dimension includes attributes related to the online layout or the areas of the physical store. The proposed Business Analytics approach focuses on physical store because we consider it as a challenging research environment that contributes to the omni-channel marketing strategies.

Profoundly, the utilization of more than one dimension augments the search space allowing to discover interesting patterns which enhance the valuable knowledge on the subject of investigation. Table 27: Business Intelligence dimensions summarizes relevant, recent studies according to the employed dimensions and the research approach for each of them.

Dimension(s)	**Research approach**	**Relevant works**
B	Market basket analysis	Tang et al., 2008; Cil, 2012; Griva et al., 2018
L	Trajectory analysis	Larson et al., 2005; Fan and Zhang, 2009; Yao et al., 2018; Ferracuti et al., 2019
S	Customer segmentation	Larson et al., 2005; Hong and Kim, 2012; Miguéis et al., 2012
B-S	Market basket analysis and Loyalty schemes	Griva et al., 2018
B-L	Association of moving locations and shopper baskets	Sano et al., 2016; Tsai et al., 2017
L-S	Path to purchase analysis	Ferracuti et al., 2019

Table 27: Business Intelligence dimensions

In regard to dimension B, several studies suggested market basket analysis (Tang et al., 2008; Cil, 2012; Griva et al., 2018). Trajectory analysis is a common technique to study Shopping trips and conceptually is related to the L dimension of the proposed framework (Yao et al., 2018; Ferracuti et al., 2019).

Following, regarding dimension S, shopper segmentation approaches have been used to analyse and detect patterns among shoppers (Larson et al., 2005; Hong and Kim, 2012; Miguéis et al., 2012). B-S dimensions provide a more detailed view regarding the combination of shoppers and products. On one hand, it is not always feasible to extract insights for both the shopper and the basket, because the actual level of engagement in loyalty schemes is low (Bruneau et al., 2018), therefore the B-S combination is not yet effective for brick-and-mortar settings. On the other hand, utilizing B-L dimensions allows reviewing how a shopping trip forms a basket and overcomes Shopper's data availability issues. The issues are bypassed since all baskets are considered as anonymous even when identifiable. L-S dimensions enable to track the shopper regardless if they have purchased anything or not. Therefore, it is an ideal approach when shopper awareness is a significant variable.

As a result, the Location itself is a very interesting self-explanatory dimension and it can enhance in-store shopping behaviour knowledge since it is not restricted by prior Shopper knowledge. In this paper we suggest a research approach utilizing B, L, S (B-L-S) dimensions and we conduct a field experiment to demonstrate the valuable insights offered by the co-existence of all three dimensions simultaneously.

7.1.2.5 Implications of multiple dimensions

Shopping trips can be thoroughly studied under the prism of B-L-S dimensions. The knowledge extraction using a single dimension has been examined by several researchers (e.g. (Fan and Zhang, 2009; Yao et al., 2018; Ferracuti et al., 2019), while the combination of two dimensions is an emerging research area (Sano et al., 2016; Tsai et al., 2017; Griva et al., 2018; Ferracuti et al., 2019).

However, considering more than 2 dimensions, as presented in this study, has been proven to provide more meaningful insights for the decision makers. Despite the fact that adding dimensions in the analysis, increments the complexity of interpreting the results, it can provide a dynamic potential when properly understood and analyzed. To this end, the proposed dimensions provide both exploratory and explanatory power and pave the ground to knowledge acquisition regarding in-store shopping behaviour. The role of the key decision makers is critical during the whole process. During this research, key persons, when asked to provide their feedback at the early stages of the BATS development process, commonly responded cautiously and with scepticism. Later on, at the implementation phase, when they became familiar with the process, they developed a more favorable and positive attitude. When they started reviewing the results and everything started to make sense it became apparent that they appreciated the potential value of the marketing analytics insights offered by BATS and unlocked the opportunities of this new approach.

Prior research indicates that heterogeneity (Kim and Park, 1997) is a key attribute in terms of shopping trip regularities (or segments). In regard to that, research efforts employee the discovery of homogeneous group of shoppers that share similar behaviours. Kim and Park (1997) identify two shopping trip regularities in order to group shoppers with similar characteristics. Our study goes beyond these two shopping trip regularities and suggests the identification of 7 clusters, which derived from our field experiment data. The quality of the clusters is essential for BATS because all marketing analytics reports utilize them as unit of analysis and report the findings accordingly. In this study, we grouped similar shopping trips and not the shopper, which is the most common and straightforward approach so far. Yet, grouping based on shopping trips/paths, which address specific shopper needs each time, seems to offer important value since it takes into account that shoppers might behave differently between two distinct store visits. The proposed IBM metric sufficiently captures the changing patterns of the in-store shopper behaviour by examining the amount of time spent in the area where the purchases were made.

A general concern pointed in the literature (Kim and Park, 1997) is how to adjust the proposed metrics in the business environment. In our study, we adjust the proposed metrics to detect impulse and planned behaviour. Additionally, we highlight the importance of the role of the key decision maker, as the knowledge on the domain is critical in order to design appropriate marketing actions. For instance, Cluster 3 (see Figure 69) contains a mixture of shopping paths where shoppers tend to be "Partially Planned" or "Impulse". Thus, this is a sample of a cluster where a marketing decision maker can intervene to turn shoppers from "Partially Planned" into "Impulse" buyers. The decision to move towards impulse derives from the pre-existing concentration in the impulse category.

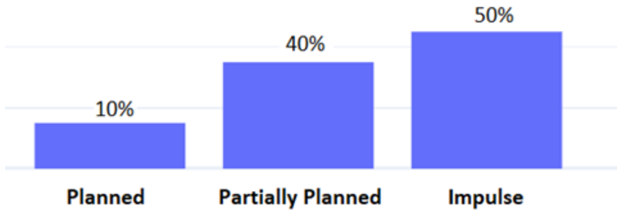


Figure 69: Cluster 3 drill-down buying analysis

Another point is that a shopper may engage in different shopping trips which in turn may have different trip characteristics. As a result, the same shoppers may belong to different clusters, ranging from “Planned” to “Impulse”. Thus, shopping trips can become the moderators that can reveal the shoppers’ needs. It is anticipated that each shopper will perform different kinds of shopping trips according to specific needs/preferences at a given time. In other words, a shopper may manifest planned buying behaviour when visiting the store for some needs and impulse buying behaviour for other needs, while in some occasions “Partially Planned” would more accurately define the shopping trip. In any case, the early identification of a shopping trip enables the decision maker to determine the appropriate action plan that will enhance the shopper’s shopping trip experience.

Finally, it should be mentioned that systems utilizing B, L, S dimensions require the accumulation of personal data which may raise privacy concerns. During our field experiment we requested users’ consent in order to be able to monitor their in-store navigation (via a mobile application). Also, we were granted access to process purchased data and loyalty scheme information by the retailer. A full-scale implementation of B, L, S approach would require a one-time shopper consent in order to be fully aligned with the local legislation regarding personal data processing.

Concluding, it is evident that the practical implications of the combination of B-L-S dimensions may provide useful insights on decision-making. Decision makers are able to track the traffic in specific product categories and areas and design responsive marketing actions. At the same time the decision maker has a tool to benchmark the retail store’s performance and monitor metrics such as the association of the shopper traffic in each area with the respective sales. This way, marketing goals can be measured and assessed using appropriate metrics.

As a result, the proposed BATS can enhance the marketing orientation of a business, as it introduces dimensions and metrics that can be elaborated to improve the performance of marketing actions.

8 CONCLUSIONS AND FURTHER RESEARCH

Location analytics generate enormous insights for decision makers and marketers. They can gain competitive insights and acquire knowledge regarding offline shopper behaviour. Location technology is becoming a key factor for businesses to improve customer experience and increase operational efficiency. Many retail stores expect to install wireless infrastructures in their stores showing their interest in location analytics to support their data-driven decisions. Retailers can employ location analytics to perform actions such as:

(a) Stores' areas performance assessment: Being connected with the customers you get to know where, when and how they go through your store. Retailers can assess the productivity of a sales area eradicating or changing the location of the inefficient areas. For example, if you found out through your analytics that most of the visitors coming to your stores are students of the age group 18-25 you might not want to keep kids related stuff in front. Even the things kept on the top shelf or bottom shelf have different attractiveness index. If planned well our conversion rates and time spent might increase.

(b) In-store personalized shopping experience: Intensifying the level of customization will be the key to profit. As soon as the customers enter the store, they have two questions in their mind and that is, the product availability, if available where they can locate it. These answers can lead to a better assistance and proper offers if they are using their mobile to locate the product. Your repeat customer might help you customize it properly building a brand awareness and credibility among others

(c) Navigation control: If shoppers know where to go as soon as they enter the store, they will visit it more often. In addition, if customer's location inside the store is known, the retailer can make location-based offers in order to achieve an increased basket size.

(d) Customer browsing patterns: With the use of location analytics, an optimal pattern can be achieved by observing the flow of the customer. As in-store patterns change regularly, retailers should be up to date regarding their shoppers' behavior.

(e) Performance monitoring: The advantage of this kind of location tracking is massive. Retailers can easily build data on calculating length of their visit, their duration, their engagement, overall number of customers and their movement in the store.

(f) Location-based services design: The technology improvements can further add value to consumer preferences by offering new services. Notifications can be sent to customers when browsing specific store areas.

(g) Detect selling opportunities: Combine location with sales data retailers can acquire knowledge regarding the areas that customers visit, but not purchasing any products. As a result, specific actions and strategies can be made to exploit such opportunities in order to attract more customers and increase basket size.

Thus, in this final chapter, we present and discuss the research's' contributions to theoretical knowledge along with its practical value. Finally, we present lessons learnt and guidelines when designing and developing indoor positioning systems and conclude with the limitations and the future research of this thesis.

8.1 Research Outcomes

This research deals with the issue of effective indoor positioning. By going through the related literature, our first objective is to examine and understand existing indoor positioning approaches and systems. We concluded to the examination of the following research questions.

- RQ1. How can we perform efficient indoor positioning from spatiotemporal data?
- RQ2. How to develop indoor positioning systems?

To address these questions, we adopt as methodological backbone the design science paradigm (Hevner et al., 2004) consider a machine learning approach that performs indoor positioning as outcome of this study. In a nutshell, we develop a machine learning indoor positioning approach that can be applied on spatiotemporal data from IoT access points. We assess two different wireless technologies (i.e. Wi-Fi and Bluetooth Low Energy Beacons) and apply Artificial Intelligence techniques to address the question. In addition, to address the research questions we propose a system artifact which is responsible for generating, capturing and processing the data for indoor positioning. The outcome of the proposed approach is the indoor position of the user of the wireless infrastructure.

We apply this approach to two different cases. The first case involves BLE Beacon technology, while the second one involves Wi-Fi technology. We then evaluate the finding from each case using technical evaluation and examine the business interpretation of the outcomes. To this end, we utilize data-driven and user-based evaluation in order to assess the results of the indoor positioning approach.

This thesis moves forward and includes also the business and user acceptance challenges, highlighting the role of the application context and aspires to provide realistic and more holistic, not just technical, guidelines on prospective researchers and designers of location-based services for retail stores and other contexts, as well as encourage them to embrace such design initiatives.

A few interesting findings extracted from the valuation of our two cases involve the correlation between Positioning Data and Purchased Data that indicates that the shopping journey reveals the shopping purpose behind the visit, and an interesting KPI that indicates the impulse buying opportunity gap. Shoppers tend to spend time in certain areas, but this does not turn into actual sales. This information can be exploited by decision makers and marketers to design actions to turn this opportunity into sales.

8.2 Theoretical Contributions

One of the main strengths of this thesis is its interdisciplinary nature, as it interweaves three different disciplines: Machine Learning (ML), Location Analytics (LA), and Information Systems (IS). Therefore, the contribution of this thesis from a theoretical perspective is found across these three disciplines.

The following paragraphs summarize and discuss the theoretical contribution of this research:

Develops a machine learning approach that performs indoor positioning.

Literature indicates that the indoor positioning issue is still open and is characterized as a very challenging one. The achievement of high localization precision (accuracy) is a common objective shared by various scholars (He et al., 2016; Shin et al., 2015). Lymberopoulos et al. (2015) argue that the indoor location problem still remains unsolved and stress the importance of the employment of a realistic approach that would counterbalance the desired localization accuracy with low cost.

This study develops and deploys an indoor positioning approach that is applied on two different case studies of shoppers moving in retail stores with different layout characteristics; i.e. BLE beacon and Wi-Fi. The developed approach does not consider the surface of the store split into equal smaller areas making a grid (i.e. fixed-length surface) to detect the exact store area of the customer. Instead, it identifies the constraints of the store's physical layout and the retailer's needs and, thus, adopts a more realistic, variable-length surface model with store areas of different sizes and dimensions.

More specifically, the approach performs fingerprinting and assesses the performance of seven established classifiers for efficient shoppers' localization in a real retail store. Then, it proposes an ensemble filter that considers the three best classifiers and executes a voting process taking a weighted vote of their predictions of the customers' locations in the store. The filter is characterized by significant lower absolute mean and root mean squared errors than the ones of the best single classifier leading to more efficient indoor positioning of consumers. Finally, for the BLE Beacons case, this study achieves localization error around 1m in 70% of the recorded cases reflecting consumers moving in the store, and around 2m in 80% of the cases, while the error for Wi-Fi technology is 3.5 meters.

Explores localization techniques in retail environments.

Since indoor localization in retail environments is still unexplored, this study started by assessing the performance of the most common techniques for indoor positioning (i.e., trilateration and fingerprinting) and established classifiers. It was found that the random forest is the best classifier. However, this study moved on to proposing and assessing an ensemble filter. The absolute mean and root mean squared errors of the ensemble filter are significantly lower (40.7% and 18% lower, respectively). More specifically, for approximately 70% of our cases (captured events of consumers), the ensemble method results in a localization error of less than 1 m and in 80% of the cases, the localization error is approximately 2 m. On the contrary, for the random forest, in 80% of the cases, the localization error is approximately 2.5 m.

In retail environments, such a deviation is significant, because even 0.5 m away from the actual shopper's position may lead to position him in a different shopping isle and in front of a different store shelve, thus a different product category. Namely, the more accurate localization of consumers, the more accurate and rich insights on the customers' shopping behavior. Consequently, the retailers and the marketing managers will be able to offer more effective customer location-based services (e.g., personalized offers, coupons etc. reflecting their recent shopping trips). Overall, the authors propose that it is worth moving forward from the random forest and proposing an ensemble method because better consumer localization results in better decision making in the retail industry.

Proposes a system artifact for indoor positioning in retail environments

Indoor positioning approaches process spatiotemporal data originated from various sources and technologies, such as Wi-Fi, Bluetooth and RFID and extracts the location of the browsing users in the retail environment. To this end, an indoor positioning system is required in order to generate data and support the indoor positioning approach. This thesis proposes a generic architecture for designing and deploying an indoor positioning system based on the available wireless technologies. As spatiotemporal data are by nature noisy, just collecting them to decide the shopper position does not lead to efficient results. Thus, appropriate sensors deployment is required, further for data cleansing is applied and processing mechanisms to transform the data in formats that can be utilized for position determination. This thesis proposes three major modules, i.e. the wireless infrastructure, the tracking mechanism and the indoor positioning mechanism. Finally, on top of the major modules can be deployed an additional module that refers to Data Analytics and extracts insights upon the collected data.

Identifies factors affecting indoor positioning approaches and systems.

Indoor positioning services have been developed for cases like museums (Kuflik et al., 2011), hospitals (Calderoni et al., 2015; Yang et al., 2015), tourism (Curran and Smith, 2006), elderly and disabled people (Marco et al., 2008), smart buildings (Lin et al., 2016), couponing and infomediation (Zou and Huang,

2015) and for generic purposes based on different technologies such as Bluetooth (Diaz et al., 2010), Wi-Fi (Au et al., 2013) and RFID (Huang et al., 2015). Each case is characterized by different requirements in terms of accuracy and cost and is deployed based on the specific needs for each of the cases. By reviewing the indoor positioning literature and studies regarding the development of indoor positioning systems we detect and examine several factors that seem to affect performance varying from the simple metric of positioning accuracy to the acceptance of the positioning system from the stakeholders.

The design and development of a location-based service should take into consideration both technical and business-related requirements and challenges rising from the domain of the service (Dhar and Varshney, 2011) and address them respectively in order the service to perform efficiently. Each domain encounters different kind of challenges in order to perform indoor localization effectively (Lymberopoulos et al., 2015). For example, museums are environments that the exhibits are dense and serve plenty of information; thus, requiring an accurate visitor position (Kuflik et al., 2011) so as the location-based service to perform efficiently. To this end, positioning accuracy is an important factor for this domain. Moreover, the indoor positioning mechanism in the museum domain requires also availability and stability to provide the required functionality. Hospital environments (Calderoni et al., 2015; Yang et al., 2015), have similar requirements differentiating to the fact that the system detects the room that the patient is located. Thus, another factor that affects a positioning mechanism is the granularity level of the position detection (i.e. x-y coordinates, area or room level).

Having examined studies across domains, we specialize to the challenges of implementing an indoor positioning service for location-based coupon recommendation, based on Bluetooth Low Energy Beacons technology, using a case study in the retail sector and compare the factors that affected our work in correlation with the factors examined in the literature. We perform an across domain and technology study in order to examine the challenges and issues encountered during the development of indoor location-based services and detect patterns among them.

8.3 Findings Summary

Summarizing the case studies this section briefly presents each one, along with its limitations and key findings. Finally, a brief overview for both cases is presented.

8.3.1 Case A: BLE Beacons Summary

Case A is characterized by the following steps/implications:

- Assesses indoor positioning techniques for BLE Beacons technology
- Proposes a system for indoor positioning
- Proposes a filter for indoor positioning and achieves localization error of 2 m
- Examines the factors affecting indoor positioning
- Discusses the number and density of BLE Beacon transmitters and their effect on indoor position performance
- Provides guidelines for indoor positioning system designers
- Provides location-based applications for retailers and suppliers

- Applies Artificial Intelligence techniques to connect shopper in-store locations to form the shopper path followed during a shopping trip
- Extracts a series of Location-based KPIs that can assess in-store shopper behavior
- Extracts shopper speed metric when using BLE Beacons technology
- Utilizes shopper movement data and sales data to detect selling opportunities
- Extracts insights upon shopper in-store behavior.
- Utilizes Business Analytics for shopping trip investigation
- Proposes marketing analytics to examine impulse buying behavior

The key findings of Case A are the following:

- BLE Beacons are an efficient technology for indoor positioning when using an adequate number of transmitters.
- Ensemble filters perform effectively that single classifiers and are more flexible to the environmental challenges
- Requires fingerprinting to cover a wide range of mobile devices
- Requires frequent fingerprinting process to adapt environmental changes such as store layout and product placement
- More flexible to business needs, as the transmitters placement allows various area divisions
- BLE Beacons is an emerging technology, thus the guidelines and lessons learnt to system designers could be of high interest

- Location analytics can empower the capabilities in Retail Operations and Marketing Analytics
- Retailers and store managers acquire a mechanism that can facilitate their decision-making process.
- Regarding shopper in-store behavior, shoppers tend to browse more than 60% of the store during a shopping trip.
- Shoppers may hesitate to use mobile applications during their shopping trip to communicate with the wireless infrastructure that rises user acceptance issues.
- Incentives is an effective mechanism to encourage shoppers install and use a mobile application.

Case A uses a flexible infrastructure in terms of being able to change BLE Beacon transmitters placement. In addition, it provides two different floor layouts. Also, this deployment allows to use advanced area granularity (aisle level). Finally, this case had sales data available and enabled one-to-one shopper identification.

On the other hand, case A is characterized by a small participants number (i.e. 100 shoppers). The retail environment affects significantly the emitted wireless signals causing interferences and affect indoor positioning performance. Finally, the mobile device compatibility is a issue that should be dealt with, by using the most widely used Bluetooth antennas.

8.3.2 Case B: Wi-Fi Summary

Case B is characterized by the following steps/implications:

- Assessed indoor positioning techniques for Wi-Fi technology
- Proposed a system for indoor positioning
- Proposed a filter for indoor positioning and achieves localization error of 3 m.
- Examined the factors affecting indoor positioning
- Discusses the effect of fixed Wi-Fi access points on indoor position performance
- Provides guidelines for indoor positioning system designers
- Provides location-based applications for retailers and suppliers
- Applies Artificial Intelligence techniques to connect shopper in-store locations to form the shopper path followed during a shopping trip
- Extracts a series of Location-based KPIs that can assess in-store shopper behaviour
- Examines store performance in terms of store comparison and renovation assessment
- Extracts insights upon shopper in-store behavior

The key findings of Case B are the following:

- Wi-Fi is the most widely used technology for indoor positioning
- Ensemble filters perform effectively that single classifiers and are more flexible to the environmental challenges
- Requires fingerprinting to cover a wide range of mobile devices

- Requires frequent fingerprinting process to adapt environmental changes such as store layout and product placement
- Due to the fixed-location infrastructure issues regarding area division were encountered
- Location analytics can empower the capabilities in Retail Operations and Marketing Analytics
- Retailers and store managers acquire a mechanism that can facilitate their decision-making process
- Regarding shopper in-store behavior, shoppers tend to browse 25-30% of the store during a shopping trip
- Incentives are required to enable users connect to the wireless infrastructure
- No action from the perspective of the shopper is required, but connecting to the wireless network, which is more user friendly that mobile applications

Case B uses a robust number of shoppers (i.e. 877) and is deployed in 5 different store layouts. In addition, these stores had an adequate number of Wi-Fi access points to perform indoor positioning.

On the other hand, case B is characterized by a fixed-location infrastructure as we were unable to modify/move the Wi-Fi access points. This restriction and the limited number of access points led us to form lesser areas than Case A, which hindered the ability to calculate shopper speed effectively. In addition, in this case we had no sales data available and thus we were unable to link sales data to shoppers.

8.3.3 Findings/Implications Overview

Summarizing both cases, we present the following table. The table contains the major findings/implications along with the case they correspond to (Case A, Case B) and a brief description of the finding/implication. The fields mentioned in the table refer to the fields that the implication applies and involve Artificial Intelligence, Machine Learning, Business Analytics, Location Analytics, Retail Operations, Information Systems and Marketing Analytics.

Findings/Implications overview				
Finding/ Implication	Case		Description	Field(s)
Indoor Positioning Filter	Wi-Fi	☑	Filter for effective indoor positioning using ensemble classifiers	Artificial Intelligence Machine Learning
	BLE	☑		
Localization techniques in retail environments	Wi-Fi	☑	Improved localization accuracy in retail environments	Artificial Intelligence Machine Learning Indoor Positioning
	BLE	☑	Examined BLE Beacon technology which is an emerging technology and provide performance insights	
Factors affecting indoor positioning	Wi-Fi	☑	Overviewed the factors that affect indoor positioning in retail environments	Indoor Positioning Information Systems
	BLE	☑		
Location-based applications	Wi-Fi	☑	Presented a series of location-based applications extracted by focus group and interviews with retailers and suppliers	Retail Operations Artificial Intelligence
	BLE	☑		
Location-based KPIs	Wi-Fi	☑	Presented a series of Location-based KPIs and Shopper speed implemented for BLE Beacons	Location Analytics Retail Operations
	BLE	☑		
Lessons learnt for indoor positioning systems	Wi-Fi	☑	The lessons learnt from both cases when deploying indoor positioning systems	Indoor Positioning Retail Operations Information Systems
	BLE	☑		

Area-based shopper segmentation	Wi-Fi	☑	A shopper segmentation approach based on the areas that shoppers spend time	Location Analytics Marketing Analytics
	BLE	☑		
Speed-based shopper segmentation	Wi-Fi	-	A shopper segmentation approach based on the speed that shoppers browse the store (implemented for BLE Beacons)	Business Analytics Marketing Analytics
	BLE	☑		
Business analytics for shopping trip investigation	Wi-Fi	☑	Examine the dimensions of a shopping trip (Basket, Location, Shopper) to discuss the implications of multiple dimensions	Business Analytics Location Analytics Marketing Analytics
	BLE	☑		
Impulse buying behavior	Wi-Fi	-	Introduce KPIs for investigating impulse buying behavior (implemented for BLE Beacons)	Marketing Analytics
	BLE	☑		
Assess channel performance	Wi-Fi	☑	Assess the performance of the channel (physical/digital) a product is sold to determine the most effective one	Marketing Analytics
	BLE	☑		
Store performance	Wi-Fi	☑	Utilize KPIs to assess the performance of each store. Evaluate marketing actions and layouts performance.	Location Analytics
	BLE	☑		

Table 28. Findings overview

8.4 Limitations and Future Research

This research deals with two different wireless technologies. Case A involved BLE Beacon technology while case B Wi-Fi technology. One limitation we had during this research was the unavailability of utilizing hybrid approaches utilizing both technologies in each case, instead of using solely one each time.

Additional limitations refer to the wireless infrastructure in the case of Wi-Fi, as we were not able to move the wireless routers in order to increase the quality of the emitted signals. The deployment of the access points had been made in a way to cover the network connectivity along the store surface. Thus, we had to work on this setting deployment.

The contribution of this thesis focuses on the development of the indoor positioning approach and examines a series of practical implications of location analytics. As future research it would be of high interest to further examine the impact of location analytics in retail operations. This requires further review of the literature and the design of further experiments. In addition, as future research we could further examine and support the finding that the shopping path highlights the purpose of the shopping visit. Apart from data-driven observation, the use of focus group could uncover further insights upon this finding. Finally, impulse buying is an interesting area, as we are able to track the time that a shopper spends in front of a product. Thus, via sales data we can verify whether or not the shopper bought this product.

Impulse buying is also related to the time that shopper spends into the store. By being able to track the shopper speed within the first minutes of the shopping visit we can assume whether the visit can involve planned or impulse buying. Closing, future research may apply to marketing analytics and the way shopping experience and in-store promotions can change the way that shopper interacts within the retail context.

References

Adler, P., & Adler, P. A. (1994). Observational Techniques” dalam Norman K. Denzin dan Yvonna S. Lincoln (ed) Handbook of Qualitative Research.

Aghazadeh, S. M. (2005). Layout strategies for retail operations: A case study. Management Research News, 28(10), 31-46.

Ali, Z. H., Ali, H. A., & Badawy, M. M. (2015). Internet of Things (IoT): Definitions, Challenges and Recent Research Directions. International Journal of Computer Applications (0975--8887) Volume.

Altini, M., Brunelli, D., Farella, E., & Benini, L. (2010). Bluetooth indoor localization with multiple neural networks. In Wireless Pervasive Computing (ISWPC), 2010 5th IEEE International Symposium on (pp. 295-300). IEEE.

Anderson, J., 1982. Logistic regression, In Handbook of Statistics, pp. 169-191, North-Holland, USA.

Au, A. W. S., C. Feng, S. Valaee, S. Reyes, S. Sorour, S. N. Markowitz, ... M. Eizenman. (2013). “Indoor tracking and navigation using received signal strength and compressive sensing on a mobile device.” IEEE Transactions on Mobile Computing, 12(10), 2050–2062.

Bartunek, J. M., Rynes, S. L., & Ireland, R. D. (2006). What Makes Management Research Interesting, and Why Does It Matter? The Academy of Management Journal, 49(1), 9–15.

Basiri, A., E. S. Lohan, T. Moore, A. Winstanley, P. Peltola, C. Hill, ... P.

Figueiredo e Silva. (2017). "Indoor location-based services challenges, requirements and usability of current solutions." Computer Science Review, 24, 1–12.

Baskerville, R. (2008). What design science is not. European Journal of Information Systems, 17(5), 441–443.

Beatty, S. E., & Ferrell, M. E. (1998). Impulse buying: Modeling its precursors. Journal of retailing, 74(2), 169-191.

Beeck, I. and W. Toporowski. (2017). "When location and content matter: effects of mobile messages on intention to redeem." International Journal of Retail and Distribution Management, 45(7–8), 826–843.

Bobek, S., Grodzki, O., & Nalepa, G. J. (2015). Indoor microlocation with BLE beacons and incremental rule learning. In Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference. 91-96.

Bose, I., & Pal, R. (2005). Auto-ID: Managing Anything, Anywhere, Anytime in the Supply Chain. Communications of the ACM, 48(8), 100-106.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. Journal of Retailing, 93(1), 79-95.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Cabero, J. M., Molina, V., Urteaga, I., Liberal, F., & Martín, J. L. (2014). Acquisition of human traces with Bluetooth technology: Challenges and proposals. Ad Hoc Networks, 12, 2-16.

Cai, J., Liu, X., Xiao, Z., & Liu, J. (2009). Improving supply chain performance management: A sys-tematic approach to analyzing iterative KPI accomplishment. Decision support systems, 46(2), 512-521.

Calderoni, L., Ferrara, M., Franco, A., & Maio, D. (2015). Indoor localization in a hospital environment using Random Forest classifiers. Expert Systems with Applications, 42(1), 125-134.

Campos, R. S., L. Lovisolo and M. L. R. De Campos. (2014). "Wi-Fi multi-floor indoor positioning considering architectural aspects and controlled computational complexity." Expert Systems with Applications, 41(14), 6211–6223.

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An overview of machine learning. In Machine learning. Springer, 3-23.

Chawathe, S. S. (2008). Beacon placement for indoor localization using bluetooth. In 2008 11th International IEEE Conference on Intelligent Transportation Systems (pp. 980-985). IEEE.

Čivilis, A., C. S. Jensen and S. Pakalnis. (2005). "Techniques for efficient road-network-based tracking of moving objects." IEEE Transactions on Knowledge and Data Engineering, 17(5), 698–712.

Cleary, J. and Trigg, L.: "K*: An Instance-based Learner Using an Entropic Distance Measure", 12th International Conference on Machine Learning, 108-114, 1995.

Cobb, C. J., & Hoyer, W. D. (1986). Planned versus impulse purchase behavior. Journal of retailing.

Curran, K. and K. Smith. (2006). “A Location-Based Mobile Tourist Guide.” Tourism and Hospitality Research, 6(2), 180–187.

Dardari, D., Closas, P., & Djurić, P. M. (2015). Indoor tracking: Theory, methods, and technologies. IEEE Transactions on Vehicular Technology, 64(4), 1263-1278.

Dedes, G. and A. G. Dempster. (2005). “Indoor GPS positioning.” VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference, 2005., 1, 412–415.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4), 197-387.

Denning, P. J. (1997). A new social contract for research. Communications of the ACM, 40(2), 132–134.

Deva, B., Raschke, P., Garzon, S. R., & Küpper, A. (2017). STEAM: A Platform for Scalable Spa-tiotemporal Analytics. Procedia Computer Science, 109, 731-736.

Dhar, S. and U. Varshney. (2011). “Challenges and business models for mobile location-based services and advertising.” Communications of the ACM, 54(5), 121.

Diaz, J. J., Maues, R. D. A., Soares, R. B., Nakamura, E. F., & Figueiredo, C. M. (2010). Bluepass: An indoor bluetooth-based localization system for mobile applications. In Computers and Communications (ISCC), 2010 IEEE Symposium on (pp. 778-783). IEEE.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Springer Berlin Heidelberg.

Domingo-Perez, F., Lazaro-Galilea, J. L., Wieser, A., Martin-Gorostiza, E., Salido-Monzu, D., & de la Llana, A. (2016). Sensor placement determination for range-difference positioning using evolutionary multi-objective optimization. Expert Systems with Applications, 47, 95-105.

Dong, Q., & Dargie, W. (2012). Evaluation of the reliability of RSSI for indoor localization. In Wireless Communications in Unusual and Confined Areas (ICWCUCA), 2012 International Conference on (pp. 1-6). IEEE.

East, R., & Uncles, M. D. (2008). In praise of retrospective surveys. Journal of Marketing Management, 24(9-10), 929-944.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. Academy of Management Journal, 50(1), 25–32.

Enge, P., & Misra, P. (1999). Special issue on global positioning system. Proceedings of the IEEE, 87(1), 3-15.

Fader, P. S., & Lodish, L. M. (1990). A cross-category analysis of category structure and promotional activity for grocery products. Journal of Marketing, 54(4), 52-65.

Fan, B., & Zhang, P. (2009). Spatially enabled customer segmentation using a data classification method with uncertain predicates. Decision Support Systems, 47(4), 343-353.

Faragher, R., & Harle, R. (2015). Location fingerprinting with bluetooth low energy beacons. IEEE journal on Selected Areas in Communications, 33(11), 2418-2428.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Fendt, J., & Sachs, W. (2008). Grounded theory method in management research: Users’ perspectives. Organizational Research Methods, 11(3), 430–455.

Ferracuti, N., Norscini, C., Frontoni, E., Gabellini, P., Paolanti, M., & Placidi, V. (2019). A business application of RTLS technology in Intelligent Retail Environment: Defining the shopper's preferred path and its segmentation. Journal of Retailing and Consumer Services, 47, 184-194.

Fink, L., Yogev, N., & Even, A. (2017). Business intelligence and organizational learning: An empir-ical investigation of value creation processes. Information & Management, 54(1), 38-56.

Galván-Tejada, C. E., Carrasco-Jiménez, J. C., & Brena, R. F. (2013). Bluetooth-WiFi based combined positioning algorithm, implementation and experimental evaluation. Procedia Technology, 7, 37-45.

Galván-Tejada, C. E., García-Vázquez, J. P., García-Ceja, E., Carrasco-Jiménez, J. C., & Brena, R. F. (2014). Evaluation of four classifiers as cost function for indoor location systems. Procedia Computer Science, 32, 453-460.

Galvan-Tejada, I., Sandoval, E. I., & Brena, R. (2012). Wifi bluetooth based combined positioning algorithm. Procedia Engineering, 35, 101-108.

Goes, P. B. (2014). Big Data and IS Research. MIS Quarterly, 38(3), iii–viii.

Grant, R. M. (2016). Contemporary strategy analysis: Text and cases edition. John Wiley & Sons.

Grewal, D., Ailawadi, K. L., Gauri, D., Hall, K., Kopalle, P., & Robertson, J. R. (2011). Innovations in retail pricing and promotions. Journal of Retailing, 87, S43-S52.

Griffith, D. A. (2005). An examination of the influences of store layout in online retailing. Journal of Business Research, 58(10), 1391-1396.

Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. Expert Systems with Applications, 100, 1-16.

Gupta, R., Mehta, K., Bhavsar, K., & Joshi, H. (2013). Mobile web analytics. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 2(3), pp-288.

Haartsen, J. C. (2000). The Bluetooth radio system. IEEE personal communications, 7(1), 28-36.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their Applications, 13(4), 18-28.

Hevner, A. R., & Chatterjee, S. (2016). Design research in information systems: Theory and practice. Reference Syllabi, Association for Information Systems (AIS), 337–342.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quartely, 28(1), 75–105.

Huang, C. H., L. H. Lee, C. C. Ho, L. L. Wu and Z. H. Lai. (2015). “Real-time RFID indoor positioning system based on kalman-filter drift removal and heron-bilateration location estimation.” IEEE Transactions on Instrumentation and Measurement, 64(3), 728–739.

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Research note—the traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. Marketing Science, 28(3), 566-572.

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Research note—the traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. Marketing Science, 28(3), 566-572.

Hui, S. K., Inman, J. J., Huang, Y., & Suher, J. (2013). The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. Journal of Marketing, 77(2), 1-16.

Hwang, I., & Jang, Y. J. (2017). Process mining to discover shoppers’ pathways at a fashion retail store using a WiFi-base indoor positioning system. IEEE Transactions on Automation Science and Engineering, 14(4), 1786-1792.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86–94.

Jin, B., W. Zhuo, J. Hu, H. Chen and Y. Yang. (2013). “Specifying and detecting spatio-temporal events in the internet of things.” Decision Support

Systems, 55(1), 256–269.

Jovicic, A., J. Li and T. Richardson. (2013). “Visible light communication: Opportunities, challenges and the path to market.” IEEE Communications Magazine, 51(12), 26–32.

Kaneko, Y., & Yada, K. (2016). Fractal dimension of shopping path: Influence on purchase behavior in a supermarket. Procedia Computer Science, 96, 1764-1771.

Keller, T., Thiesse, F., Fleisch, E. (2014). Classification Models for RFID-Based Real-Time Detection of Process Events in the Supply Chain: An Empirical Study. ACM Transactions on Management Information Systems 5(4).

Kim, D. Y., Kim, S. H., Choi, D., & Jin, S. H. (2015). Accurate Indoor Proximity Zone Detection Based on Time Window and Frequency with Bluetooth Low Energy. Procedia Computer Science, 56, 88-95.

Kim, E., S. T. Examiners, J. Collin, M. Binsabbar, J. Yang, Z. Wang, ... D. Namiot. (2014). “An iPhone Application for Providing iBeacon-based Services to Students.” International Journal of Open Information Technologies, 3(3), 57–72.

Kim, M., & Leskovec, J. (2011). The network completion problem: Inferring missing nodes and edges in networks. In Proceedings of the 2011 SIAM International Conference on Data Mining (pp. 47-58). Society for Industrial and Applied Mathematics.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

Kosko, B. (1992). Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence/book and disk. Prentice Hall, Upper Saddle River.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159-190.

Kourogi, M., Sakata, N., Okuma, T., & Kurata, T. (2006). Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/self-contained sensor system. Advances in Artificial Reality and Tele-Existence, 1310-1321.

Kriz, P., Maly, F., & Kozel, T. (2016). Improving indoor localization using bluetooth low energy beacons. Mobile Information Systems, 2016.

Kuflik, T., J. Lanir, E. Dim, A. Wecker, M. Carmel, M. Corra, … O. Stock. (2011). "Indoor Positioning: Challenges and Solutions for Indoor Cultural Heritage Site." ACM Intelligent User Interfaces '11 [IUI] Palo Alto, California, USA, 375–378.

Kumar, A. K. T. R., Schäufele, B., Becker, D., Sawade, O., & Radusch, I. (2016, June). Indoor localization of vehicles using deep learning. In 2016 IEEE 17th international symposium on a world of wireless, mobile and multimedia networks (WoWMoM) (pp. 1-6). IEEE.

Kuo, R. J., Liao, J. L., & Tu, C. (2005). Integration of ART2 neural network and genetic K-means algorithm for analyzing web browsing paths in electronic commerce. Decision Support Systems, 40(2), 355-374.

Kwon, O., K. Choi and M. Kim. (2007). "User acceptance of context-aware services: Self-efficacy, user innovativeness and perceived sensitivity on contextual pressure." Behaviour and Information Technology, 26(6), 483–498.

Larson, J. S., Bradlow, E. T., & Fader, P. S. (2005). An exploratory look at supermarket shopping paths. International Journal of research in Marketing, 22(4), 395-414.

Lassabe, F., Canalda, P., Chatonnay, P., Spies, F., & Baala, O. (2005). A Friis-based calibrated model for WiFi terminals positioning. In Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks. IEEE, 382-387.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15). Springer Berlin Heidelberg.

Li, X., Wang, J., & Liu, C. (2015). A bluetooth/PDR integration algorithm for an indoor positioning system. Sensors, 15(10), 24862-24885.

Lin, K., M. Chen, J. Deng, M. M. Hassan and G. Fortino. (2016). "Enhanced Fingerprinting and Trajectory Prediction for IoT Localization in Smart Buildings." IEEE Transactions on Automation Science and Engineering, 13(3), 1294–1307.

Lin, T. N., & Lin, P. C. (2005). Performance comparison of indoor positioning techniques based on location fingerprinting in wireless networks. In Wireless Networks, Communications and Mobile Computing, 2005 International Conference on (Vol. 2, pp. 1569-1574). IEEE.

Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(6), 1067-1080.

Lymberopoulos, D., Liu, J., Yang, X., Choudhury, R. R., Handziski, V., and Sen, S. (2015). "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned." In Proceedings of the 14th international conference on information processing in sensor networks (pp. 178-189). ACM.

Ma, F., Wang, Y., & Hao, J. K. (2017). Path relinking for the vertex separator problem. Expert Sys-tems with Applications, 82, 332-343.

Madigan, D., Einahrawy, E., Martin, R. P., Ju, W. H., Krishnan, P., & Krishnakumar, A. S. (2005). Bayesian indoor positioning systems. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE (Vol. 2, pp. 1217-1227). IEEE.

Mainetti, L., L. Patrono and I. Sergi. (2014). "A survey on indoor positioning systems." 2014 22nd International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2014, 111–120.

Manley, E. D., Al Nahas, H., & Deogun, J. S. (2006). Localization and Tracking in Sensor Systems. In SUTC (2), 237-242.

Manning, C., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval, Cambridge University Press.

Marco, A., R. Casas, J. Falco, H. Gracia, J. I. Artigas and A. Roy. (2008).

"Location-based services for elderly and disabled people." Computer Communications, 31(6), 1055–1066.

Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A Design Theory for Systems That Support Emergent Knowledge Processes. MIS Quartely, 26(3), 179–212.

Mazan, F., & Kovarova, A. (2015). A Study of Devising Neural Network Based Indoor Localization Using Beacons: First Results. Computing and Information Systems Journal. University of the West of Scotland, 19(1), 15-20.

Mirowski, P., Ho, T. K., Yi, S., & MacDonald, M. (2013). SignalSLAM: Simultaneous localization and mapping with mixed WiFi, Bluetooth, LTE and magnetic signals. In Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on (pp. 1-10). IEEE.

Moiseeva, A., & Timmermans, H. (2010). Imputing relevant information from multi-day GPS tracers for retail planning and management using data fusion and context-sensitive learning. Journal of Retailing and Consumer Services, 17(3), 189-199.

Montaser, A., & Moselhi, O. (2014). RFID indoor location identification for construction projects. Automation in Construction, 39, 167-179.

Morales, J. A., Akopian, D., & Agaian, S. (2014). Faulty measurements impact on wireless local area network positioning performance. IET Radar, Sonar & Navigation, 9(5), 501-508.

Mou, S., Robb, D. J., & DeHoratius, N. (2018). Retail store operations:

Literature review and re-search directions. European Journal of Operational Research, 265(2), 399-422.

Mowrey, C. H., Parikh, P. J., & Gue, K. R. (2018). A model to optimize rack layout in a retail store. European Journal of Operational Research, 271(3), 1100-1112.

Nuaimi, K. Al and H. Kamel. (2016). "A survey of indoor positioning systems and algorithms A Survey of Indoor Positioning Systems and Algorithms." 2011 International Conference on Innovations in Information Technology, (May 2011), 185–190.

O'Sullivan, D., & Unwin, D. (2014). Geographic information analysis. Hoboken, NJ: John Wiley & Sons

Page, B., Trinh, G., & Bogomolova, S. (2019). Comparing two supermarket layouts: The effect of a middle aisle on basket size, spend, trip duration and endcap use. Journal of Retailing and Consumer Services, 47, 49-56.

Palipana, S., B. Pietropaoli and D. Pesch. (2017). "Recent advances in RF-based passive device-free localisation for indoor applications." Ad Hoc Networks, 64, 80–98.

Palumbo, F., Barsocchi, P., Chessa, S., & Augusto, J. C. (2015). A stigmergic approach to indoor localization using bluetooth low energy beacons. In Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on, 1-6.

Perez, I., Pinchin, J., Brown, M., Blum, J., & Sharples, S. (2016). Unsupervised labelling of sequential data for location identification in indoor environments. Expert Systems with Applications.

Pick, J. B., Turetken, O., Deokar, A., & Sarkar, A. (2017). Location Analytics and Decision Support: Reflections on Recent Avancementa, a Research Framework and the Path Ahead. Decision Support Systems, 99.

Pointrlabs. (2016). Beacons: Everything you need to know. URL: http://www.pointrlabs.com/blog/ beacons-everything-you-need-to-know. Accessed 15.09.16.

Popa, M. C., Rothkrantz, L. J., Shan, C., Gritti, T., & Wiggers, P. (2013). Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information. Pattern Recognition Letters, 34(7), 809-819.

Pratama, A. R., & Hidayat, R. (2012). Smartphone-based pedestrian dead reckoning as an indoor positioning system. In System Engineering and Technology (ICSET), 2012 International Conference on (pp. 1-6). IEEE.

Priyantha, N. B., Chakraborty, A., & Balakrishnan, H. (2000). The cricket location-support system. In Proceedings of the 6th annual international conference on Mobile computing and networking. ACM, 32-43.

Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.

Radhakrishnan, M., Sen, S., Subbaraju, V., Misra, A., & Balan, R. (2016). IoT+ Small Data: Trans-forming In-Store Shopping Analytics and Services.

Rahm, E. (2000). Data Cleaning : Problems and Current Approaches. Informatica, 1–11.

Rai, H. G., Jonna, K., & Krishna, P. R. (2011). Video analytics solution for tracking customer locations in retail shopping malls. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 773-776). ACM.

Romme, A. G. L. (2003). Making a Difference: Organization as Design. Organization Science, 14(5), 558–573.

Safavian, R., and Landgrebe, D. (1990). A survey of decision tree classifier methodology.

Salzberg, S. L. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. Data mining and knowledge discovery, 1(3), 317-328.

Sano, N., Tsutsui, R., Yada, K., & Suzuki, T. (2016). Clustering of customer shopping paths in Japa-nese grocery stores. Procedia Computer Science, 96, 1314-1322.

Sarantopoulos, P., Theotokis, A., Pramatari, K., & Doukidis, G. (2016). Shopping missions: An analytical method for the identification of shopper need states. Journal of Business Research, 69(3), 1043-1052.

Scamell-Katz, S. (2012). The art of shopping. LID Publishing Limited.

Senior, A. W., Brown, L., Hampapur, A., Shu, C. F., Zhai, Y., Feris, R. S., ... & Carlson, C. (2007, September). Video analytics for retail. In Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on (pp. 423-428). IEEE.

Seol, S., E. K. Lee and W. Kim. (2017). "Indoor mobile object tracking using RFID." Future Generation Computer Systems, 76, 443–451.

Sharp, I., & Yu, K. (2014). Sensor-based dead-reckoning for indoor positioning. Physical Communication, 13, 4-16.

Shepard, S. (2005). RFID: radio frequency identification. McGraw Hill Professional.

Silberer, G., Büttner, O. B., & Gorbach, A. (2007). Exploring shopping paths and attention behavior at the point of sale. Marketing theory and practice in an inter-functional world, 8.

Simon, H. A. (1996). The sciences of the artificial (3rd ed.). MIT Press.

Singal, H., Kohli, S., & Sharma, A. K. (2014). Web analytics: State-of-art & literature assessment. In Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th Interna-tional Conference- (pp. 24-29). IEEE.

Slack, N., Chambers, S., & Johnston, R. (2010). Operations management. Pearson education.

Sorensen, H., Bogomolova, S., Anderson, K., Trinh, G., Sharp, A., Kennedy, R., ... & Wright, M. (2017). Fundamental patterns of in-store shopper behavior. Journal of Retailing and Consumer Services, 37, 182-194.

Srivastava, B. (2004). Radio frequency ID technology: the next revolution in SCM. Business Horizons, 47(6), pp. 60-68.

Stake, R. E. (1995). The art of case research (Vol. 80). Thousand Oaks, CA: Sage

Publications.

Subhan, F., Hasbullah, H., Rozyyev, A., & Bakhsh, S. T. (2011). Indoor positioning in bluetooth networks using fingerprinting and lateration approach. In Information Science and Applications (ICISA), 2011 International Conference on (pp. 1-9). IEEE.

Sun, G., Chen, J. and Guo, W. (2005). “Signal Processing Techniques in Network-Aided Positioning.” IEEE Signal Processing Magazine, (July), 12–23.

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, Ó., & Huerta, J. (2015). Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Systems with Applications, 42(23), 9263-9278.

Tsai, C. Y., Li, M. H., & Kuo, R. J. (2017). A shopping behavior prediction system: Considering moving patterns and product characteristics. Computers & Industrial Engineering, 106, 192-204.

Turban, E., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2015). “Electronic commerce: A managerial and social networks perspective.” Springer.

Turgut, Z., Aydin, G. Z. G., & Sertbas, A. (2016). Indoor Localization Techniques for Smart Building Environment. Procedia Computer Science, 83, 1176-1181.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing. Journal of retailing, 91(2), 174-181.

Villarubia, G., Rubio, F., De Paz, J. F., Bajo, J., & Zato, C. (2013). Applying classifiers in indoor location system. In Trends in Practical Applications of Agents and Multiagent Systems (pp. 53-58). Springer International Publishing.

Vrechopoulos, A. P., O'keefe, R. M., Doukidis, G. I., & Siomkos, G. J. (2004). Virtual store layout: an experimental comparison in the context of grocery retail. Journal of Retailing, 80(1), 13-22.

Vu, L., Do, Q., & Nahrstedt, K. (2011). Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace. In Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on (pp. 54-62). IEEE.

Waisberg, D., & Kaushik, A. (2009). Web Analytics 2.0: empowering customer centricity. The origi-nal Search Engine Marketing Journal, 2(1), 5-11.

Wang, X., Gao, L., & Mao, S. (2015b). PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach. In 2015 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

Wang, X., Gao, L., & Mao, S. (2017). Biloc: Bi-modal deep learning for indoor localization with commodity 5ghz wifi. IEEE Access, 5, 4209-4220.

Wang, X., Gao, L., Mao, S., & Pandey, S. (2015a). DeepFi: Deep learning for indoor fingerprinting using channel state information. In 2015 IEEE wireless communications and networking conference (WCNC) (pp. 1666-1671). IEEE.

Wang, B., Zhou, S., Liu, W., & Mo, Y. (2015). Indoor localization based on curve fitting and location search using received signal strength. IEEE Transactions on Industrial Electronics, 62(1), 572-582.

Wilson, M. W. (2012). “Location-based services, conspicuous mobility, and the location-aware future.” Geoforum, 43(6), 1266–1275.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Xia, S., Y. Liu, G. Yuan, M. Zhu and Z. Wang. (2017). “Indoor Fingerprint Positioning Based on Wi-Fi: An Overview.” ISPRS International Journal of Geo-Information, 6(5), 135.

Yaeli, A., Bak, P., Feigenblat, G., Nadler, S., Roitman, H., Saadoun, G., ... & Sandbank, T. (2014). Understanding customer behavior using indoor location analysis and visualization. IBM Journal of Research and Development, 58(5/6), 3-1.

Yang, J., Wang, Z., & Zhang, X. (2015). “An ibeacon-based indoor positioning systems for hospitals.” International Journal of Smart Home, 9(7), 161-168.

Yao, D., Zhang, C., Zhu, Z., Hu, Q., Wang, Z., Huang, J., & Bi, J. (2018). Learning deep representa-tion for trajectory clustering. Expert Systems, 35(2), e12252.

Yim, J. (2008). “Introducing a decision tree-based indoor positioning technique.” Expert Systems with Applications, 34(2), 1296–1302.

Yin, R. K. (1994). Case study research: Design and Methods (second edition).

Beverly Hills, CA: Sage Publishing.

Yoon, S., J. (Sunny) Kim and D. J. Connolly. (2018). "Understanding Motivations and Acceptance of Location-Based Services." International Journal of Hospitality and Tourism Administration, 19(2), 187–209.

Young, B., & Hetherington, M. (1996). The literature on advertising and children's food choice. Nutrition & Food Science, 96(5), 15-18.

Youssef, A. M., & Youssef, M. (2007). A Taxonomy of Localization Schemes for Wireless Sensor Networks. In ICWN, 444-450.

Youssef, M. A., Agrawala, A., & Shankar, A. U. (2003). WLAN location determination via clustering and probability distributions. In Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on, 143-150.

Yuan, H., Qian, Y., Yang, R., & Ren, M. (2014). Human mobility discovering and movement inten-tion detection with GPS trajectories. Decision Support Systems, 63, 39-51.

Yun, H., D. Han and C. C. Lee. (2013). "Understanding the Use of Location-Based Service Applications: Do Privacy Concerns Matter?" Journal of Electronic Commerce Research, 14(3), 215–231.

Zeinalipour-Yazti, D. and C. Laoudias. (2017). "The anatomy of the anyplace indoor navigation service." SIGSPATIAL Special, 9(2), 3–10.

Zhu, D. H., Sun, H., & Chang, Y. P. (2017). "How the content of location-based advertisings influences consumers' store patronage intention." Journal of

Consumer Marketing, 34(7), 603–611.

Zou, H., Jiang, H., Lu, X., & Xie, L. (2014). An online sequential extreme learning machine approach to WiFi based indoor positioning. In Internet of Things (WF-IoT), 2014 IEEE World Forum on (pp. 111-116). IEEE.

Zou, H., Lu, X., Jiang, H., & Xie, L. (2015). A fast and precise indoor localization algorithm based on an online sequential extreme learning machine. Sensors, 15(1), 1804-1824.

Zou, H., Huang, B., Lu, X., Jiang, H., & Xie, L. (2015). A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine. IEEE Transactions on Wireless Communications, 15(2), 1252-1266.

Zou, H., Wang, H., Xie, L., & Jia, Q. S. (2013). An RFID indoor positioning system by using weighted path loss and extreme learning machine. In Cyber-Physical Systems, Networks, and Applications (CPSNA), 2013 IEEE 1st International Conference on (pp. 66-71). IEEE.

Zou, X. and K. W. Huang. (2015). "Leveraging location-based services for couponing and infomediation." Decision Support Systems, 78, 93–103.

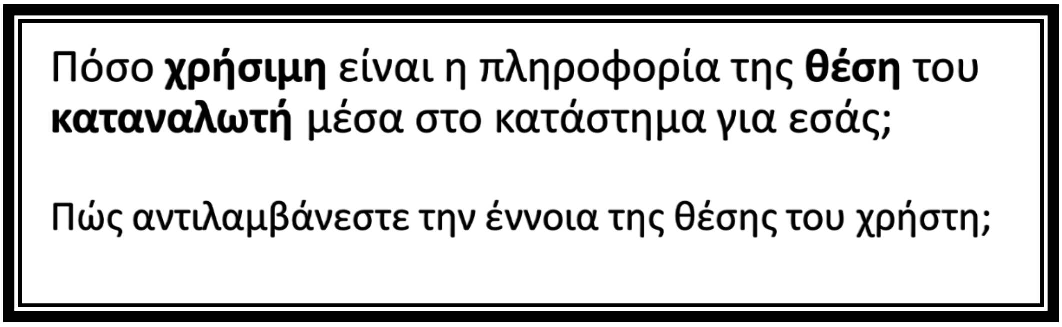
Appendix A: Indicative IoT Data set Structure


```
"session" : {
  "-event1" : {
   "beacons" : [ {
     "distance" : 2.97,
     "ID" : 1,
     "strenght" : -98,
   }, {
     "distance" : 4.28,
     "ID" : 2,
     "strenght" : -98,
   }, {
     "distance" : 6.31,
     "ID" : 3,
     "strenght" : -98,
   }, {
     "distance" : 7.26,
     "ID" : 4,
     "strenght" : -100,
   } ],
   "when" : 1479226996544
  },
  "-event2" : {
   "beacons" : [ {
     "distance" : 2.44,
     "ID" : 5,
     "strenght" : -91,
   }, {
     "distance" : 2.97,
     "ID" : 6,
     "strenght" : -98,
   }, {
     "distance" : 3.17,
     "ID" : 7,
     "strenght" : -94,
   }, {
     "distance" : 3.46,
     "ID" : 8,
     "strenght" : -95,
   } ],
   "when" : 1479226998042
  },
```

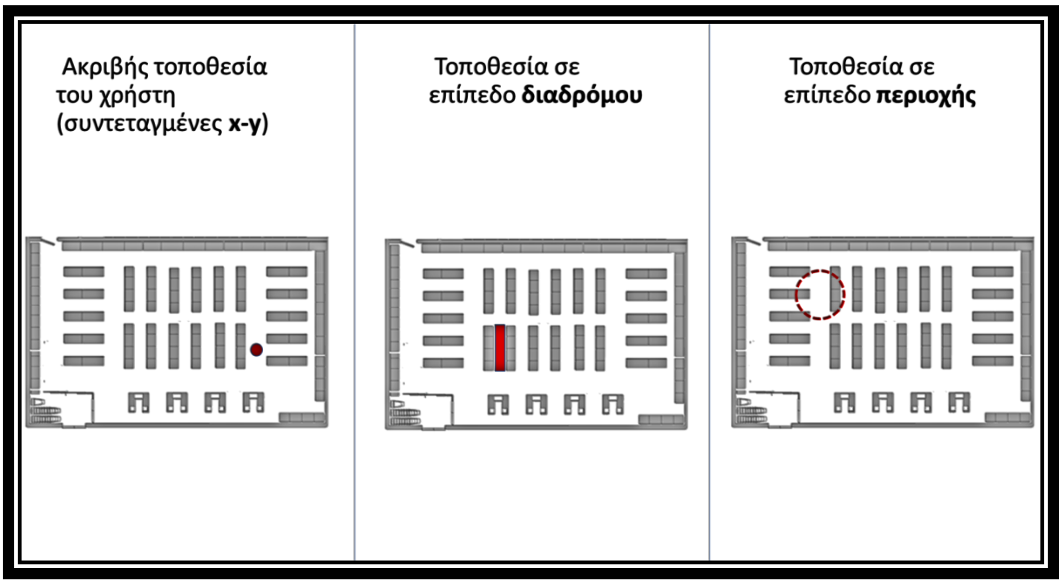
Appendix B: Focus Group

Slide 1

Slide 2

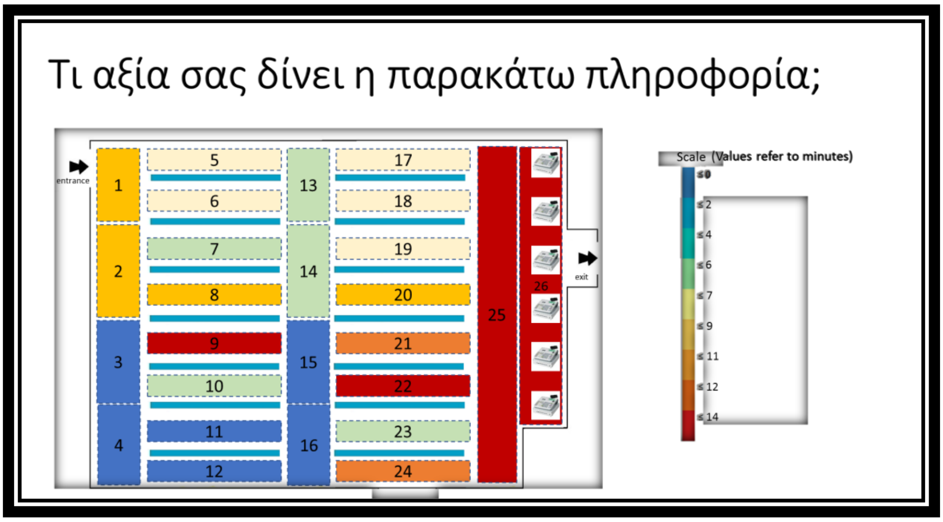


Slide 3

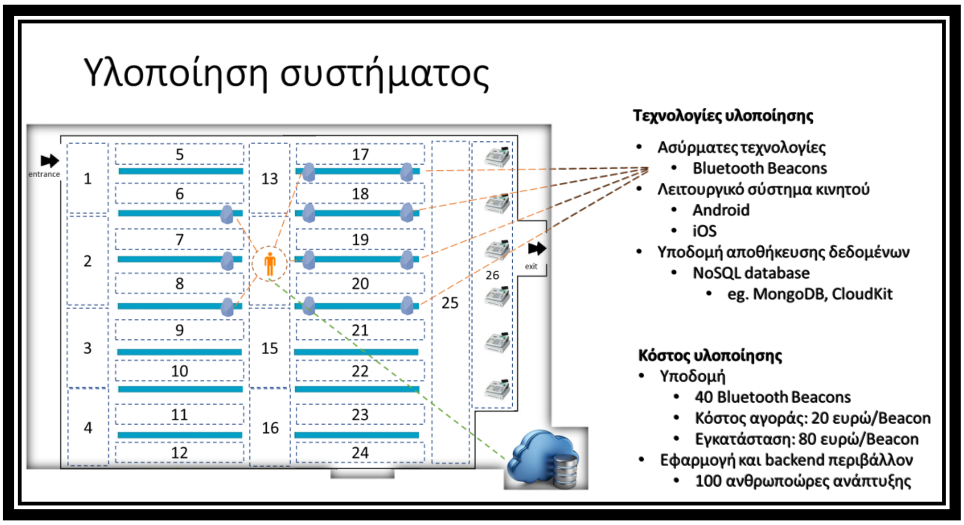


Slide 4

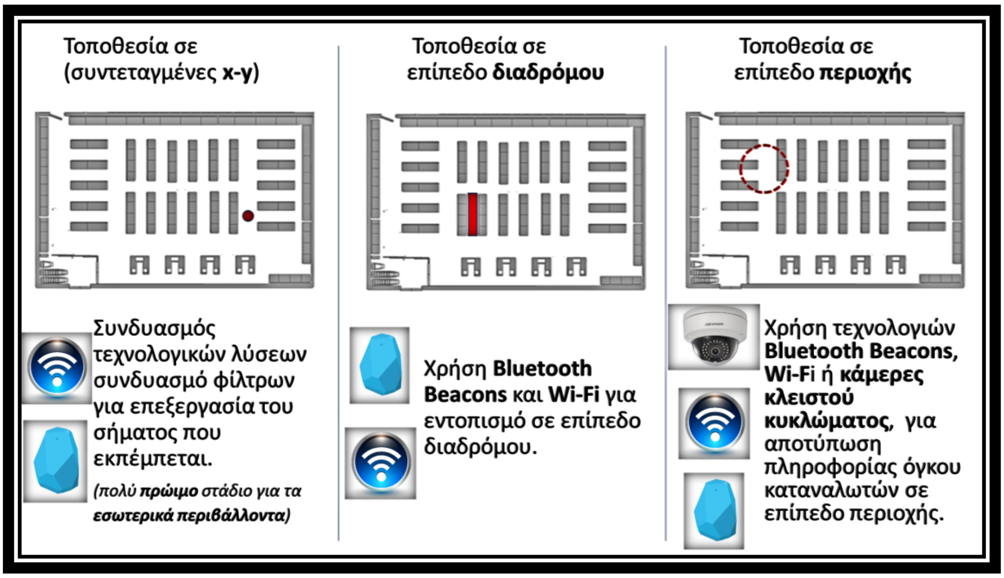
Slide 5



Slide 6



Slide 7



Slide 8

Ευχαριστούμε!