ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΤΕΡΖΗ ΣΤΕΛΛΑ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ

Ο Ψηφιακός Μετασχηματισμός και η εφαρμογή του Text Mining στη Ασφαλιστική Βιομηχανία

Επιβλέπων : Πουλούδη Αθανασία – Αναπληρώτρια Καθηγήτρια

Υποβληθείσα ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος (MSc) στη Διοικητική Επιστήμη και Τεχνολογία

Αθήνα, Οκτώβριος 2019

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Η σελίδα αυτή είναι σκόπιμα λευκή.



Βεβαίωση εκπόνησης Διπλωματικής εργασίας

«Δηλώνω υπεύθυνα ότι η συγκεκριμένη μεταπτυχιακή εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών του ΠΜΣ στη Διοικητική Επιστήμη και Τεχνολογία του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του Οικονομικού Πανεπιστημίου Αθηνών έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών στην Ελλάδα ή το εξωτερικό. Η εργασία αυτή έχοντας εκπονηθεί από εμένα, αντιπροσωπεύει τις προσωπικές μου απόψεις επί του θέματος. Οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης διπλωματικής αναφέρονται στο σύνολό τους, δίνοντας πλήρεις αναφορές στους συγγραφείς, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο».

(Υπογραφή)

.....

Τερζή Στέλλα

Φοιτητής MSc στη Διοικητική Επιστήμη και Τεχνολογία

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Περίληψη

Η εξόρυξη δεδομένων, είτε αυτά είναι αριθμητικά είτε σε μορφή κειμένου και η ανάλυση τους στην συνέχεια έχει πλέον μετατραπεί σε επιτακτική ανάγκη για κάθε εταιρεία, καθώς οι εταιρείες έχουν κατανοήσει την σημαντικότητα τους και την βοήθεια που μπορούν να τους προσφέρουν στις διαδικασίες τους. Αν και καθυστερημένα σε σχέση με άλλους κλάδους, η ασφαλιστική βιομηχανία έχει αντιληφθεί και αυτή την συγκεκριμένη ανάγκη, Για να φτάσει όμως μια ασφαλιστική εταιρεία στην εξόρυξη των δεδομένων που χρειάζεται πρέπει να έχει και τα ανάλογα τεχνολογικά εργαλεία, ενώ για την ανάλυση τους πρέπει να έχει ήδη αποφασίσει τι θέλει να δει μέσα από αυτά και πως θα τα συνδυάσει ώστε να της προσφέρουν την καλύτερη εικόνα. Σε αυτά θα βοηθήσει ο ψηφιακός μετασχηματισμός οποίος θα προσφέρει την τεχνολογία, αλλά και την αλλαγή κουλτούρας που χρειάζεται μια ασφαλιστική εταιρεία για αναθεωρήσει τους στόχους της και την στρατηγική της, ώστε να εκμεταλλευτεί ανάλογα τα δεδομένα της. Ένα σημαντικό εργαλείο είναι η εξόρυξη δεδομένων κείμενου (text mining), το οποίο μπορεί να χρησιμοποιηθεί από τον ασφαλιστικό κλάδο τόσο σε εσωτερικές λειτουργίες όσο και για την αύξηση του customer engagement.

Λέξεις Κλειδιά: ψηφιακός μετασχηματισμός, ασφαλιστική βιομηχανία, δεδομένα κειμένου, text mining

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Πίνακας περιεχομένων

1 Εισαγωγή

1.1 Οι ασφαλιστικές εταιρείες και οι διαδικασίες τους.....1

1.2 Αντικείμενο διπλωματικής……2

1.3 Οργάνωση κειμένου……,2

2 Σχετικές Εργασίες

2.1 Ψηφιακός μετασχηματισμός και Ασφαλιστική Βιομηχανία…… 3

2.1.1 Ψηφιακός μετασχηματισμός και Customer Engagement ……4

2.1.2 Ψηφιακός μετασχηματισμός και εσωτερικές λειτουργίες ασφαλιστικών εταιρειών...9

2.1.3 Ψηφιακός μετασχηματισμός και εσωτερικές λειτουργίες ασφαλιστικών εταιρειών.10

2.2 Text Mining και Ασφαλιστική Αγορά……12

3 Θεωρητικό Υπόβαθρο

3.1 Μέθοδοι Text Mining.....16

4 Λειτουργίες της εταιρικής πλατφόρμας και στόχος της ανάλυσης των δεδομένων της

4.1 Πλατφόρμες Customer Service.....22

4.2 Περιγραφή εταιρικής πλατφόρμας.....22

5 Προ-επεξεργασία δεδομένων, εφαρμογή και αξιολόγηση του μοντέλου ταξινόμησης.

6 Τεχνικές λεπτομέρειες.....30

6.1 Λεπτομέρειες υλοποίησης 30

7 Επίλογος

7.1 Σύνοψη και συμπεράσματα 36

7.2 Μελλοντικές επεκτάσεις.....36

8 Βιβλιογραφία.....37

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1

Εισαγωγή

1.1 Οι ασφαλιστικές εταιρείες και οι διαδικασίες τους

Η ασφαλιστική βιομηχανία και οι εταιρείες που δραστηριοποιούνται σε αυτήν έχουν καθυστερήσει την εφαρμογή της διαδικασίας του ψηφιακού μετασχηματισμού. Οι περισσότερες διαδικασίες εξακολουθούν να βασίζονται στο χαρτί, με το προσωπικό να αναγκάζεται να κάνει συμβατικές αλλαγές γραπτώς, το οποίο σημαίνει ότι το εργατικό δυναμικό μιας ασφαλιστικής εταιρείας ξοδεύει το μεγαλύτερο μέρος του χρόνου εργασίας του μη παραγωγικά σε μη σημαντικά καθήκοντα. Συνεπώς σπαταλάτε χρόνος και δεσμεύονται πόροι, οι οποίοι θα μπορούσαν να εκμεταλλευτούν πιο παραγωγικά. Επιπλέον, οι επαναλαμβανόμενες διαδικασίες στις οποίες ο ανθρώπινος παράγοντας κυριαρχεί ενέχουν μεγαλύτερο κίνδυνο σφάλματος. Οι εργαζόμενοι καταπονούνται και είναι λιγότερο ικανοί να συγκεντρωθούν με την πάροδο του χρόνου, εφόσον δεν έχουν τα κατάλληλα εργαλεία να τους βοηθήσουν να ολοκληρώσουν την εργασία τους αποτελεσματικότερα. Προβληματική φαίνεται να είναι και η περιοχή της ανίχνευσης απάτης, καθώς αρκετές περιπτώσεις δεν γίνονται αντιληπτές, με αποτέλεσμα οικονομικό κόστος, αλλά και κόστος για την φήμη της εταιρείας που θα αντιμετωπίσει μια τέτοια κατάσταση. Στα παραπάνω προβλήματα που αντιμετωπίζουν οι ασφαλιστικές εταιρείες έρχεται να προστεθεί και η όλο και πιο έντονη δυσαρέσκεια των πελατών που παρατηρούν σε όλες τις καθημερινές εργασίες τους την τεχνολογική εξέλιξη και την μετάβαση στην ψηφιακή εποχή, αλλά όχι στα ασφαλιστικά τους προϊόντα και στις συναλλαγές τους με τον συγκεκριμένο πάροχο.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1.2 Αντικείμενο διπλωματικής

Η ασφαλιστικές εταιρείες παραμένουν αρκετά προσκολλημένες στις παραδοσιακές διαδικασίες τους. Αρκετές διαδικασίες περιέχουν την χρήση χαρτιού και αποθήκευση του στην συνέχεια, κάτι το οποίο αποδεικνύεται ότι εμπεριέχει κόστος χώρου αν διατηρηθεί μέσα στην ίδια την εταιρεία ή χρηματικό αν επιλεχθεί κάποια άλλη εταιρεία, η οποία θα διαχειριστεί την αποθήκευση και διατήρηση των αρχείων. Επιπλέον το ανθρώπινο στοιχείο είναι πολύ έντονο, το όποιο είναι αρνητικό καθώς ο ανθρώπινος παράγοντας ενέχει μεγαλύτερες πιθανότητες λάθους. Το ανθρώπινο δυναμικό μπορεί να παίξει καλύτερα τον ρόλο του ελεγκτή στις διαδικασίες και αν αυτές ολοκληρώνονται σωστά από τεχνολογικά εργαλεία, παρά να έχει τον ρόλο του διεκπεραιωτή με τους κινδύνους και το άγχος που επιφέρει αυτός ο ρόλος. Επίσης οι εσωτερικές διαδικασίες, όπως το underwriting δεν έχουν τα απαραίτητα δεδομένα διαθέσιμα εφόσον συνήθως έρχονται σε μορφή ερωτηματολογίου συμπληρωμένο από τον ασφαλισμένο, το οποίο ερωτηματολόγιο εκτός ότι δεν μπορεί να καλύψει όλες τις περιπτώσεις και πιθανόν να μην συμπληρωθεί σωστά ή να συμπληρωθεί ψευδώς, επεξεργάζεται μόνο από τον underwriter οπότε υπάρχει ξανά η πιθανότητα λανθασμένης εκτίμησης. Μια άλλη σημαντική εσωτερική διαδικασία για μια ασφαλιστική εταιρεία είναι η ανίχνευση απάτης, η οποία τις περισσότερες φορές έρχεται να αντιμετωπίσει την κατάσταση ενώ έχει γίνει ήδη μερική ζημία και δεν την προλαμβάνει, καθώς και λειτουργεί δειγματοληπτικά. Παράλληλα με όσα αναφέραμε οι απαιτήσεις των πελατών αυξάνονται συνεχώς και η άμεση ανταπόκριση είναι απαραίτητη ώστε οι πελάτες να διατηρηθούν. Ο τρόπος αντιμετώπισης των παραπάνω φαίνεται να είναι ο ψηφιακός μετασχηματισμός των ασφαλιστικών εταιρειών. Ο ψηφιακός μετασχηματισμός για μια εταιρεία σημαίνει υιοθέτηση τεχνολογίας για μεγάλο μέρος των διαδικασιών της και αλλαγή κουλτούρας. Στην ουσία ο ψηφιακός μετασχηματισμός μια εταιρείας δεν σταματά ποτέ λόγω τον εξελίξεων, αλλά ενσωματώνεται στην εταιρεία. Στην διαδικασία της μετάβασης της περιλαμβάνονται λήψη δεδομένων σε ηλεκτρονική μορφή από πλατφόρμες ή εφαρμογές, εκμετάλλευση του machine learning, των predictive analytics,της τεχνητής νοημοσύνης και των τεχνικών πρόβλεψης συμπεριφοράς των πελατών. Τα δεδομένα που συλλέγονται από τον νέο τρόπο δραστηριοποίησης των εταιρειών και η εκμετάλλευση τους βοηθά στο customer engagement και customer service εφόσον τα χαρακτηριστικά και οι ανάγκες των πελατών είναι πλέον πιο ακριβή και με την κατάλληλη επεξεργασία ευδιάκριτα στα χέρια των εταιρειών.

Μεγάλος όγκος αυτών των δεδομένων έρχεται στην μορφή κειμένου, για αυτό και επιλέξαμε από το πεδίο της ανάλυσης δεδομένων να ασχοληθούμε με την εφαρμογή της εξόρυξης δεδομένων από κείμενο (text mining) και τι θα προσφέρει στις διαδικασίες της ανίχνευσης απάτης, της διαχείρισης των αποζημιώσεων και στο κομμάτι του customer engagement και customer service. Μετά από ανάλυση του θεωρητικού υπόβαθρου του text mining και προκειμένου να δούμε στην πράξη την λειτουργία του, ζητήθηκαν αιτήματα πελατών από πλατφόρμα ασφαλιστικής εταιρείας που χρησιμοποιείται για την διαδικασία των αποζημιώσεων και την εξυπηρέτηση του ασφαλισμένου. Τα αιτήματα αυτά είναι σε

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

μορφή κειμένου και αφού παραμετροποιήθηκαν και προ-επεξεργάστηκαν, εφαρμόστηκε μοντέλο ταξινόμησης, το οποίο αυτό που κατάφερε είναι να ταξινομεί τα αιτήματα σε κλάσεις και βάση αυτών των κλάσεων να ορίζεται το e-mail το οποίο θα λαμβάνει ο ασφαλισμένος μέχρι την οριστική επίλυση του αιτήματός του, καθώς και υπήρξαν παράπονα για τους χρόνους ανταπόκρισης. Το μοντέλο αξιολογήθηκε βάση μετρικών αξιολόγησης επιδεικνύοντας αρκετά μεγάλη ακρίβεια.

1.3 Οργάνωση κειμένου

Στο κεφάλαιο 2 αναφερθήκαμε στον ψηφιακό μετασχηματισμό των εταιρειών και συγκεκριμένα στον ψηφιακό μετασχηματισμό των εταιρειών που δραστηριοποιούνται στον ασφαλιστικό τομέα. Επίσης αναφέραμε εργαλεία, διαδικασίες και καινοτόμες ιδέες, τα οποία είναι μέρος του μετασχηματισμού και θα έχουν αντίκτυπο στο customer engagement και στις διαδικασίες του underwriting και της ανίχνευσης απάτης. Μέρος του δεύτερου κεφαλαίου περιέχει αναφορά στον ρόλο της ηγεσίας στον ψηφιακό μετασχηματισμό μιας εταιρείας. Στην ενότητα 2 του κεφαλαίου 2, ασχοληθήκαμε με την εφαρμογή του text mining στην ανίχνευση απάτης, την διαχείριση αποζημιώσεων και στο customer engagement και customer service. Στο κεφάλαιο 3 αναλύσαμε το θεωρητικό υπόβαθρο πίσω από το text mining δεδομένων κειμένου. Το κεφάλαιο 4 περιέχει αναφορά στις πλατφόρμες customer service και περιγραφή την πλατφόρμας από την οποία προέρχονται τα δεδομένα τα οποία επεξεργαστήκαμε στο κεφάλαιο 5 και αναπτύξαμε το μοντέλο μας πάνω σε αυτά. Στο κεφάλαιο 6 αναγράφεται ολοκληρωμένος ο κώδικας που συντάχθηκε και το κεφάλαιο 7 περιλαμβάνει την συνόψιση της διπλωματικής εργασίας και τις μελλοντικές επεκτάσεις της. Τέλος, στο κεφάλαιο 8 βρίσκουμε την βιβλιογραφία που χρησιμοποιήθηκε για την διπλωματική εργασία.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

2

Σχετικές Εργασίες

2.1 Ψηφιακός μετασχηματισμός και Ασφαλιστική Βιομηχανία

Με τον όρο ψηφιακό μετασχηματισμό αναφερόμαστε στην υιοθέτηση ψηφιακών διαδικασιών και εργαλείων σε έναν οργανισμό, ο οποίος σχετίζεται κυρίως με τους πελάτες του και το πως η αναδιαμόρφωση της στρατηγικής του θα ικανοποιήσει τις ανάγκες τους. Οι εταιρείες που αποφασίζουν να προχωρήσουν σε αυτόν στην ουσία επανεξετάζουν ολόκληρη την δομή τους και την σχέση με τους πελάτες τους. Ο ψηφιακός μετασχηματισμός αλλάζει τις υπάρχουσες διαδικασίες ,δημιουργεί νέες και έχει την ικανότητα να αναδιαμορφώσει το επιχειρηματικό σκηνικό αφού δημιουργεί νέα τμήματα αλλά και νέες εταιρείες. Ας μιλήσουμε λίγο πιο συγκεκριμένα για τα οφέλη που μια επιχείρηση αποκομίζει από τον ψηφιακό μετασχηματισμό. Αρχικά τα μέλη του οργανισμού έρχονται πιο κοντά και μαθαίνουν το νόημα της συνεργασίας εφόσον αναλαμβάνουν μεταξύ τους να συνεργαστούν όσο το δυνατόν καλύτερα ώστε να ολοκληρώσουν την μετάβαση. Ολόκληρη η νοοτροπία της επιχείρησης αλλάζει, η αυστηρή ιεραρχία παύει να ισχύει και οποιαδήποτε βαθμίδα έχει λόγο και εκφέρει τις ιδέες της με αποτέλεσμα να προωθείται η καινοτομία και οι πειραματισμοί. Όσον αφορά την εικόνα της προς τους πελάτες της, ο ψηφιακός μετασχηματισμός δίνει την δυνατότητα στις επιχειρήσεις να διευκολύνουν τις συναλλαγές μαζί τους και να τους προσφέρουν την καλύτερη δυνατή εξυπηρέτηση, εφόσον θα λειτουργούν πλέον με τα απαραίτητα εργαλεία. Το σημαντικότερο, όμως είναι να κατανοήσουν οι επιχειρήσεις, οι οποίες θα μπουν σε αυτήν την διαδικασία ότι δεν είναι απλά ένα ακόμα έργο με αρχή, μέση και τέλος ,αλλά μια συνεχής προσπάθεια βελτίωσης και εναρμόνισης με τις εξελίξεις της ψηφιακής εποχής.(Michelle Cyca,2018)

Στο κεφάλαιο αυτό θα ασχοληθούμε με τον ψηφιακό μετασχηματισμό στην ασφαλιστική βιομηχανία συγκεκριμένα ,η οποία βρίσκεται στο κατώφλι της ψηφιακής επανάστασης και μιας νέας εποχής. Ο ασφαλιστικός κλάδος έχει πλέον κατανοήσει ότι μια απλή έκπτωση ασφαλίστρων ή η αντιγραφή του ανταγωνιστή δεν είναι λύσεις που θα καταστήσουν βιώσιμη μια εταιρεία που δραστηριοποιείται στον συγκεκριμένο κλάδο και της ασκούνται δυνάμεις τις οποίες δεν μπορεί να αγνοήσει και προέρχονται κυρίως από τους καταναλωτές της. Πλέον το πελατολόγιο μιας ασφαλιστικής εταιρείας έχει αναδιαμορφωθεί και κατά σημαντικό ποσοστό αποτελείται από νεότερης ηλικίας άτομα, τους λεγόμενους Generation Y,των οποίων η αγοραστική δύναμη είναι αυξανόμενη. Οι πελάτες αυτοί έχουν

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

διαφορετικές προσδοκίες από την ασφαλιστική τους εταιρεία ,οι οποίες είναι η ευρεία χρήση τεχνολογίας, η καινοτομία και η κατανόηση των αναγκών τους. Δεδομένου ότι μακροπρόθεσμα δεν θα αποτελούν πλέον ποσοστό του πελατολογίου αλλά το πελατολόγιο αυτό καθ' αυτό τους κατατάσσει αυτόματα στην ομάδα καταναλωτών που πρέπει να ληφθούν υπόψιν για την δημιουργία νέων ασφαλιστικών προϊόντων. Η νέα σύνθεση του πελατολογίου φαίνεται να επηρέασε και τα ποσοστά τα οποία αναφέρθηκαν στο World Insurance Report για το 2019.(i-scoop Team,2019).

Σύμφωνα με το World insurance report μόνο ένα μικρό ποσοστό καταναλωτών ασφαλιστικών προϊόντων ,λιγότερο από 15% ,πιστεύει ότι το συμβόλαιο και οι όροι του είναι διαμορφωμένο ώστε να καλύψει τους νέους κινδύνους που κάθε εποχή ενέχει (π.χ περιβαλλοντική αλλαγή,cyber attacks). Αυτό σημαίνει ότι το υπόλοιπο ποσοστό είναι πελάτες οι οποίοι δεν έχουν εμπιστοσύνη στην ασφαλιστική τους εταιρεία και αναμένουν να τους διαψεύσει και να αναλογιστεί από που προκύπτει αυτή η κατάσταση και να την αντιστρέψει.Το συγκεκριμένο report επίσης κατέληξε ότι λιγότερο από 50% των ασφαλιστικών εταιρειών έχει κατανοήσει ότι οι ανάγκες και οι κίνδυνοι διαφέρουν πλέον από δεκαετία σε δεκαετία ,ακόμα και από χρόνο σε χρόνο, ενώ το 40% των ασφαλιστικών εταιρειών που δραστηριοποιούνται στον τομέα Ζωής και Υγείας έχει αντιληφθεί ότι είναι επιτακτική η ανάγκη δημιουργίας νέων προϊόντων. Επίσης η έρευνα που έγινε για το World insurance report έδειξε ότι ενώ το 55% των πελατών είναι έτοιμο να δεχθεί την καινοτομία από την ασφαλιστική βιομηχανία και σε ποσοστά 37% και 55% αντίστοιχα είναι διαθέσιμοι να παραχωρήσουν επιπλέον στοιχεία στην ασφαλιστική τους εταιρεία και είναι διατεθειμένοι για αύξηση ασφαλίστρων εφόσον το πρόγραμμα τους θα έχει δημιουργηθεί για τις υπάρχουσες ανάγκες τους και προβλέπει και την κάλυψη νέων αναγκών, ενώ μόνο το 26% των ασφαλιστικών εταιρειών έχει ξεκινήσει να ερευνά νέα επιχειρηματικά μοντέλα και καινοτόμες εναλλακτικές ,τα οποία κάνουν πράξη την πελαντοκεντρικότητα . Τα παραπάνω ποσοστά δείχνουν ξεκάθαρα πόσο έχουν αλλάξει οι προσδοκίες των πελατών και ειδικά των νέων πελατών οι οποίοι είναι σαφώς νεότεροι ,μυημένοι στην τεχνολογία και αναθρεμμένοι σε έναν ταχύτατα αναπτυσσόμενο περιβάλλον. (Team of Capgemini and EFMA,2019)

Αντιλαμβανόμαστε λοιπόν ότι ήδη οι ασφαλισμένοι είναι έτοιμοι για την αλλαγή που είναι απαραίτητη να συμβεί στο επιχειρησιακό μοντέλο των ασφαλιστικών εταιρειών. Η ύπαρξη όμως ενός επιχειρησιακού μοντέλου που θέλει να θέσει τον πελάτη στο κέντρο του χωρίς όμως να μετρά την διατηρισημότητα του στην εταιρεία και την ικανοποίηση του δεν μπορεί να εδραιωθεί και να επιφέρει αποτέλεσμα. Η ασφαλιστική βιομηχανία φαίνεται να έχει ανάγκη την αύξηση του customer engagement.(Hugh Terry et al.,2019)

2.1.1 Ψηφιακός μετασχηματισμός και Customer Engagement

Με τον όρο customer engagement αναφερόμαστε στην αλληλεπίδραση του πελάτη με την εταιρεία και το πόσο ισχυρή είναι. Συγκεκριμένα οι ορισμοί που έχουν δοθεί από τους ειδικούς Paul Greenberg και Eric Peterson είναι αντίστοιχα οι εξής:

"Customer engagement είναι η συνεχιζόμενη αλληλεπίδραση μεταξύ εταιρείας και πελάτη, που προσφέρεται από την εταιρεία και επιλέγεται από τον πελάτη."(Paul Greenberg for Hubspot,2014),

"Customer engagement είναι μια εκτίμηση του βαθμού και του βάθους της αλληλεπίδρασης τ)ων πελατών σε ένα σύνολο προκαθορισμένων στόχων. "(Eric Peterson ,2006)

Η αλληλεπίδραση αυτή είναι η δίοδος για την συλλογή δεδομένων των ασφαλισμένων, τα οποία η ασφαλιστική βιομηχανία όπως και αρκετές βιομηχανίες μέχρι σήμερα δεν έχουν αντιληφθεί την σημαντικότητα τους και την συμβολή που θα μπορούσαν να έχουν στην εξέλιξη τους. Τα δεδομένα όμως προκειμένου να συλλεχθούν ,επεξεργαστούν και να οδηγήσουν σε συμπεράσματα είναι απαραίτητο να υπάρξουν προσθήκες στην τεχνολογία, κατανόηση της πελατοκεντρικής στρατηγικής και ριζική αλλαγή μια νοοτροπίας ότι οι ασφαλιστικές εταιρίες απλά τρέφονται από την ύπαρξη κινδύνου και πείθουν πελάτες να αποκτήσουν τα συμβόλαια τους. Κατανοούμε λοιπόν ότι ο ψηφιακός μετασχηματισμός και το customer engagement είναι αλληλένδετα, ο μετασχηματισμός δρομολογείται βάση των αναγκών των πελάτων και τα νέα εργαλεία και το νέο στρατηγικό πλάνο που θα επιφέρει ο μετασχηματισμός είναι ενέργειες οι οποίες θα δείξουν το ενδιαφέρον προς τον πελάτη και θα διατηρήσουν την εμπιστοσύνη του.

Η προσθήκη νέων τεχνολογικών εργαλείων στις διαδικασίες των ασφαλιστικών εταιρειών υποβοηθείται από InsurTech εταιρείες, οι οποίες προσφέρουν τις υπηρεσίες τους, καθώς αρκετές ασφαλιστικές εταιρείες δεν έχουν την δυνατότητα να υποστηρίξουν μόνες τους την μετάβαση σε ένα νέο data-driven επιχειρηματικό μοντέλο. Μερικά παραδείγματα InsurTech εταιρειών σε όλο τον κόσμο είναι τα .(Hugh Terry et al.,2019) :

- Atidot — προγνωστική πλατφόρμα ανάλυσης που χρησιμοποιεί τεχνητή νοημοσύνη, μηχανική μάθηση και μεγάλα δεδομένα για να παρέχει στον ασφαλιστή μεγαλύτερη γνώση των δεδομένων του.
- Cover Genius — προϊόν XCover επιτρέπει στους μεταφορείς να μεταφέρουν προϊόντα πέρα από τα σύνορα, αναλαμβάνοντας την προσαρμογή του προϊόντος σε 20 γλώσσες και πάνω από 70 χώρες.
- eBaoTech — συνεργάζεται με περισσότερους από 100 φορείς, πράκτορες και φορείς οικοσυστημάτων σε περισσότερες από 30 χώρες που αναπτύσσουν συναφή ασφαλιστικά μοντέλα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Flamingo — χρησιμοποιεί τεχνητή νοημοσύνη για να βελτιώσει τις εμπειρίες των πελατών για να βελτιώσει την αποτελεσματικότητα και την ανάπτυξη της κίνησης.
- Pega — αμερικανική εταιρεία που αναπτύσσει λογισμικό για τη διαχείριση πελατειακών σχέσεων, την αυτοματοποίηση ψηφιακών διαδικασιών και τη διαχείριση επιχειρηματικών διαδικασιών.
- Zhong An — είναι η online ασφαλιστική εταιρεία της Κίνας και ιδρύθηκε το 2013 . Από την αρχή, έχει αποκτήσει 460 εκατομμύρια χρήστες και έχει γράψει περισσότερες από 5,8 δισεκατομμύρια πολιτικές.

Η εξωτερική ανάθεση σε τρίτους είναι ένας λογικός τρόπος για να ξεπεραστούν οι περιορισμοί ενός οργανισμού σε απαρχαιωμένες διαδικασίες του, οι οποίες παραμένουν και θεωρούνται λειτουργικές και να παρακάμψουν τα εσωτερικά κενά στην ικανότητα του οργανισμού που να διαχειριστεί την αλλαγή. Η συνεργασία αυτή δημιουργεί ευκαιρίες και ανοίγει το πνεύμα της εταιρίας σε πολιτιστικές αλλαγές ,νέους τρόπους σκέψης και καινοτόμες ιδέες. Συνεπώς ,δεν αναφερόμαστε απλά στην υιοθέτηση νέων τρόπων εργασίας μέσω νέας τεχνολογίας, αλλά μια εντελώς διαφορετική εφαρμογή της διαδικασίας ανάπτυξης προϊόντων η οποία θα έχει τα δεδομένα που χρειάζεται ώστε να προβλέψει με συνέχειες δοκιμές και ακρίβεια τα αποτελέσματα αυτών των προϊόντων. Με αυτήν την προσέγγιση, όπου η εικόνα του αποτελέσματος λόγω των διαθέσιμών δεδομένων είναι πιο σαφής, η εταιρία και οι άνθρωποί της δεν συνδέονται υπερβολικά με ένα έργο που δεν λειτουργεί και για το οποίο η επιτυχία φαίνεται απίθανη ή υπερβολικά δαπανηρή .(Hugh Terry et al.,2019)

Η συνεργασία με Insurtech εταιρείες, σημαίνει ότι νέα τεχνολογία και εργαλεία θα είναι διαθέσιμα σε διαδικασίες όπως αυτές των πωλήσεων ,των αποζημιώσεων και των πληρωμών, οι οποίες όταν είναι εύρυθμες προσφέρουν μια πολύ εικόνα της εταιρείας και αυξάνουν την δέσμευση των πελατών. Κατανοούμε, λοιπόν, ότι τα εργαλεία τα οποία υιοθετούνται στα πλαίσια του ψηφιακού μετασχηματισμού σκοπό στην αύξηση του customer engagement αλλά παράλληλα στην απόκτηση όγκου δεδομένων και καλύτερη εξυπηρέτηση του πελάτη.

Πωλήσεις

Ο συνδυασμός τεχνολογίας και δεδομένων που μπορούν να αναλυθούν επιτρέπει την εξατομικευμένη κάλυψη με παραμέτρους όπως η διαμόρφωση προϊόντων βάση της ζήτηση για συγκεκριμένη κάλυψη του ασφαλισμένου την συγκεκριμένη χρονική στιγμή. Ο ασφαλισμένος δηλαδή δεν θα χρεώνεται για μία κάλυψη που δεν χρειάζεται και αυτή που έχει πραγματικά ανάγκη δεν είναι στην διάθεση του. Η

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

προσαρμογή των προϊόντων βάση ζήτησης ωφελεί και τον ασφαλισμένο και την εταιρεία που τον διατηρεί ικανοποιημένο. Επίσης ο συνδυασμός όγκου δεδομένων και τεχνολογίας δίνει την δυνατότητα ταξινόμησης και κατηγοριοποίησης των πελατών όχι μόνο βάση των δημογραφικών χαρακτηριστικών τους αλλά και βάση συμπεριφορικών χαρακτηριστικών, όπως η συνεργασία με τον συγκεκριμένο ασφαλιστή ή οι αγορές προϊόντων που έχει πραγματοποιήσει μέσα στα χρόνια συνεργασίας. Με αυτή την πληροφορία στα χέρια της η ασφαλιστική εταιρεία θα μπορεί να κάνει πράξη το cross-selling και να παρέχει τον κατάλληλο ασφαλιστή με την κατάλληλη εμπειρία αλλά και χαρακτηριστικά στο πελάτη. Συνεπώς, καταλήγουμε τελικά ο πελάτης να είναι ο πυρήνας της εταιρείας και η δέσμευση (engagement) του ο στόχος, ενώ οι πωλήσεις που είναι στοχευμένες με έναν διαμεσολαβητή που ταιριάζει στο προφίλ του πελάτη έχουν μεγαλύτερες πιθανότητες επιτυχίας και διατηρισημότητας στο μέλλον. (Hugh Terry et al.,2019)

Μερικές Insurtech εταιρείες που παρέχουν υπηρεσίες και εργαλεία και βοηθούν σε αυτό το βήμα είναι οι εξής (Hugh Terry et al.,2019):

- Polywizz,- Βρετανική νεοσύστατη εταιρεία η οποία αναλύει ολόκληρο το χαρτοφυλάκιο ενός πελάτη μέσω deep learning και στη συνέχεια επιβεβαιώνει τα αποτελέσματα με τον διαμεσολαβητή ώστε να καθορίσει αν ο πελάτης είναι ασφαλισμένος σύμφωνα με τις ανάγκες του , έτσι ο διαμεσολαβητής έχει ένα καλύτερο εργαλείο για να προχωρήσει.
- Atidot- προγνωστική πλατφόρμα ανάλυσης που χρησιμοποιεί τεχνητή νοημοσύνη, μηχανική μάθηση και μεγάλα δεδομένα για να παρέχει στον ασφαλιστή μεγαλύτερη γνώση των δεδομένων του.
- Flamingo - χρησιμοποιεί τεχνητή νοημοσύνη για να βελτιώσει τις εμπειρίες των πελατών
- Pega- αμερικανική εταιρεία η οποία αναπτύσσει λογισμικό για τη διαχείριση πελατειακών σχέσεων, την αυτοματοποίηση ψηφιακών διαδικασιών και τη διαχείριση επιχειρηματικών διαδικασιών.
- Element- insurtech εταιρεία που βοηθά τις ασφαλιστικές εταιρείες να δημιουργήσουν ψηφιακές υπηρεσίες.
- Instanda- cloud-based πλατφόρμα για την δημιουργία ασφαλιστικών προϊόντων
- Gefen Technologies- λογισμικό ως πλατφόρμα υπηρεσιών που επιτρέπει στους ασφαλιστικούς συμβούλους να πλησιάσουν και να επικοινωνήσουν με τους πελάτες.

Αποζημιώσεις

Ένα σημαντικό κομμάτι για την δέσμευση των πελατών μιας ασφαλιστικής εταιρείας είναι η άμεση και σωστή διευθέτηση των αποζημιώσεων της. Η ταχύτητα με την οποία γίνεται ο διακανονισμός της αποζημίωσης και η εξόφληση της αυξάνει την καλή εμπειρία του πελάτη και όχι μόνο τον διατηρεί αλλά προσελκύει και νέους, συνεπώς η σταδιακή κατάργηση της χειρωνακτικής διαδικασίας και η δημιουργία πλατφορμών είναι μονόδρομος. Οι δύο διαδικασίες καλό θα ήταν να συνυπάρξουν στο χρονικό διάστημα της ενσωμάτωσης.

Η διαδικασία της άμεσης αποζημίωσης δημιουργεί ένα αίσθημα ασφάλειας για αυτό και την αναζητούν οι περισσότεροι πελάτες ,οπότε αντιλαμβανόμαστε ότι όσο αυξάνεται αυτό το αίσθημα τόσο αυξάνεται και το αίσθημα της δέσμευσης .Μια ιδέα για την διατήρηση αυτού του συναισθήματος είναι η ασφαλιστική εταιρεία να μην δρα πλέον όταν ο κίνδυνος έχει επέλθει αλλά να προλαμβάνει αυτόν τον κίνδυνο.

Αυτό θα τα καταφέρει με τα δεδομένα τα οποία θα συλλεχθούν από τους ασφαλισμένους με την έναρξη του συμβολαίου τους και την εφαρμογή της τηλεϊατρικής ώστε να έχει όσο το δυνατόν περισσότερες πληροφορίες αλλά και στην συνέχεια με την διαδικασία των αποζημιώσεων από όπου μπορεί να ελέγξει τις εξετάσεις ,τις νοσηλείες και κτλ που πραγματοποιούνται. Επίσης συλλογή δεδομένων μπορεί να πραγματοποιηθεί με χρήση εφαρμογών όπου ο κάθε ασφαλισμένος θα μπορεί με δική του συναίνεση να καταγράφει τα αποτελέσματα εξετάσεων του ,καθημερινές μετρήσεις τιμών κτλ και με χρήση φορητών συσκευών, οι οποίες συλλέγουν δεδομένα όπως ο καρδιακός ρυθμό ανάπαυσης, τις ώρες ύπνου, συνήθειες όπως το κάπνισμα ή το αλκοόλ και να τις παρέχει οικειοθελώς στην ασφαλιστική του εταιρεία.



Εικόνα 1,Φορητές Συσκευές

Όλα αυτά τα δεδομένα που θα έχουν συλλεχθεί μπορούν να βοηθήσουν την ασφαλιστική εταιρεία να κατέχει τον ρόλο του συμβούλου προς τον ασφαλισμένο για έναν πιο υγιεινό τρόπο ζωής, προλαμβάνοντας μέχρι και επικίνδυνες παθήσεις ,για παράδειγμα με αυξημένες τιμές σε χοληστερίνη ή πίεση να προτείνει έναν γιατρό του δικτύου της για επίσκεψη ή να υπενθυμίζει από μια ηλικία και μετά σε άνδρες και γυναίκες τις εξετάσεις που θα έπρεπε να πραγματοποιήσουν. Με αυτόν το τρόπο έχει οφέλη και η εταιρεία και ο ασφαλισμένος. Ο πελάτης πιθανόν να παραμείνει υγιής και να αποτρέψει κάποιο σοβαρότερο νόσημα λόγω έγκαιρης παρέμβασης ,ενώ η εταιρεία δημιουργεί μια πληρέστερη βάση δεδομένων, έρχεται πιο κοντά με τον ασφαλισμένο και εφόσον αποτρέπονται σοβαρότερα περιστατικά εξοικονομούνται αποζημιώσεις. (Sam Draper,2017)

Παρακάτω μπορούμε να δούμε ορισμένα παραδείγματα εταιρειών και προϊόντων .(Hugh Terry et al.,2019):

- Babylon/Prudential - Ψηφιακές υπηρεσίες υγείας βασισμένες σε τεχνητή νοημοσύνη.
- Good Doctor - Προϊόν τηλεϊατρικής της Ping An.
- Medopad - εταιρεία η οποία ασχολείται με ανάπτυξη αλγορίθμων για τη βελτίωση της διάγνωσης και της εκτίμησης του κινδύνου.
- Muang Thai - αναπτύξε το προϊόν BaoWan για την παρακολούθηση του διαβήτη και την ανταμοιβή της διαχείρισης της κατάστασης.
- Wellthy Therapeutics - ψηφιακή θεραπευτική βασισμένη σε τεχνητή νοημοσύνη.

Συναλλαγές/Πληρωμές

Αναλύσαμε παραπάνω πως θα αυξηθεί η δέσμευση των πελατών μέσω της διαδικασίας των αποζημιώσεων και της πρόληψης σε θέματα υγείας ,δεν είναι όμως μόνο αυτά τα οποία ενδιαφέρουν έναν ασφαλισμένο .Αρκετές φορές η πρόληψη μπαίνει σε δεύτερη μοίρα κυρίως από μικρότερες ηλικίες και αυτό που προέχει είναι η ευελιξία και για την ακρίβεια η ευελιξία στο θέμα την πληρωμής όταν αναφερόμαστε σε ασφαλιστικά συμβόλαια ,δεδομένου ότι οι περισσότερες συναλλαγές γίνονται πλέον μέσω κινητού τηλεφώνου ή μέσω κάποιας εφαρμογής .Η διαδικασία των χρηματικών συναλλαγών κάθε ασφαλισμένου με την εταιρεία θα μπορούσε να προστεθεί σε μια πλατφόρμα, η οποία θα χρησιμοποιείται για την διαδικασία των αποζημιώσεων με αποτέλεσμα να δημιουργηθεί ένα οικοσύστημα όπου όλες οι ανάγκες του πελάτη θα ικανοποιούνται μέσω μιας πλατφόρμας ή εφαρμογής ,δηλαδή οι πληρωμές του, οι αποζημιώσεις του, η αγορά νέων προϊόντων του, η αναζήτηση νέων

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

προϊόντων του θα διενεργούνται όλα μέσω αυτής της πλατφόρμας. Όλα αυτά τα δεδομένα που θα συλλέγονται καθημερινά από τις κινήσεις των πελατών θα μπορούν να επεξεργαστούν στην συνέχεια από την εταιρεία ώστε να αποκτήσει τις πληροφορίες που χρειάζεται για το πελατολόγιο της. Η άριστη γνώση του πελατολογίου μέσα από τις καθημερινές συναλλαγές του με την εταιρεία είναι απαραίτητη ώστε να γίνουν οι κατάλληλες ενέργειες με τις οποίες θα διατηρηθεί.(Research2Guidance,2018)

Εταιρείες που έχουν δημιουργήσει οικοσυστήματα από πλατφόρμες ή εφαρμογές είναι οι εξής .(Hugh Terry et al.,2019):

- Ant Financial - θυγατρική της Alibaba, είναι η εταιρεία fintech της Κίνας με την υψηλότερη αξία, η οποία εκμεταλλεύεται την Alipay, τη μεγαλύτερη online πλατφόρμα πληρωμών για κινητά. Επίσης μέσω της Alipay έχει αναπτύξει τεχνολογία πληρωμών αναγνώρισης προσώπου.
- Tencent - κινεζικός όμιλος ετερογενών δραστηριοτήτων και η μεγαλύτερη εταιρεία τυχερών παιχνιδιών και κοινωνικών μέσων στον κόσμο
- Zhong An - που ιδρύθηκε το 2013, είναι η ασφαλιστική εταιρεία της Κίνας, η οποία έχει μόνο online παρουσία 460 εκατομμύρια χρήστες και περισσότερα από 5,8 δισεκατομμύρια συμβόλαια.

Η απόφαση για τον ψηφιακό μετασχηματισμό μιας εταιρείας επηρεάζεται ιδιαίτερα από εξωτερικούς παράγοντες, όπως ο ανταγωνισμός, η τεχνολογική εξέλιξη και οι εξελίξεις στον τομέα της επιχειρηματικής ευφυίας, αλλά κυρίως από τις ανάγκες που εκφράζουν οι πελάτες της. Το αποτέλεσμα της μετάβασης αυξάνει το customer engagement μακροπρόθεσμα, εξάλλου αυτός είναι και ο στόχος της. Η μετάβαση αυτή είτε υποστηρίζεται από την ίδια την εταιρεία αν έχει το κατάλληλο εργατικό δυναμικό με την απαραίτητη τεχνογνωσία, είτε με την συνεργασία με άλλες εταιρείες, παραδείγματα των οποίων αναφέρθηκαν παραπάνω, οι οποίες αναλαμβάνουν να φέρουν σε πέρας την αλλαγή. Με τον έναν ή με τον άλλον τρόπο, υπάρχει η λύση η ασφαλιστική βιομηχανία να αποδεχθεί ότι πρέπει να ακολουθήσει το ρεύμα των εξελίξεων και να αλλάξει την κουλτούρα της.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

2.1.2 Ψηφιακός μετασχηματισμός και εσωτερικές λειτουργίες ασφαλιστικών εταιρειών

Ο ψηφιακός μετασχηματισμός των ασφαλιστικών εταιρειών και η συλλογή δεδομένων δεν θα βοηθήσει μόνο στον τομέα του customer engagement. Η τεχνολογία που θα είναι πλέον διαθέσιμη θα βοηθήσει και σε εσωτερικές λειτουργίες της εταιρείας όπως είναι οι λειτουργείες του underwriting και του ελέγχου απάτης, οι οποίες είναι και οι δύο πολύ σημαντικές για την εύρυθμη λειτουργία της.

Underwriting

Η διαδικασία του underwriting με την υιοθέτηση τεχνολογίας και την δημιουργία εταιρικών πλατφορμών και εφαρμογών μπορεί σε μεγάλο μέρος της να αυτοματοποιηθεί. Η λήψη στοιχείων για την δημιουργία συμβολαίων σε ηλεκτρονική μορφή θα διευκολύνει τον έλεγχο τους, όπως έλλειψη στοιχειών ή την επεξεργασία ερωτηματολογίου σε μορφή πολλαπλής επιλογής ή σε μορφή κειμένου και ανάλογα με τα αποτελέσματα του πρώτου ελέγχου η διαδικασία θα συνεχίζεται σύμφωνα με τις δικλίδες που έχουν οριστεί μέχρι και την έκδοση του ασφαλιστηρίου. Η αυτοματοποίηση προλαμβάνει το ανθρώπινο λάθος και έχει ως αποτέλεσμα σωστότερη λήψη αποφάσεων ,ταχύτερες διαδικασίες επεξεργασίας των στοιχείων και μείωση των χρονικών στόχων εξυπηρέτησης με ταυτόχρονη μείωση του κόστους.

Ένα παράδειγμα λογισμικού το οποίο έχει δημιουργηθεί για να βοηθήσει ασφαλιστικές εταιρείες να αυτοματοποιήσουν τη διαδικασία του underwriting μέσω εφαρμογής machine learning διατίθεται από την εταιρεία Daisy Intelligence.Το λογισμικό αυτό φορτώνεται με δεδομένα πελατών (ηλικία, πίεση αίματος) και στην συνέχεια παρέχει μεμονωμένες προτάσεις τιμολόγησης με βάση τους παράγοντες κινδύνου του καθενός.

Η πρόσβαση σε μεγαλύτερο όγκο δεδομένων ευνοεί και την εφαρμογή predictive analytics,τεχνικών ελέγχου αξιοπιστίας των πελατών (credibility techniques),τεχνικών για την πρόβλεψη συμπεριφοράς των πελατών και ανάλυση συναισθήματος (sentiment analysis),με σκοπό την δράση των εταιρειών άμεσα σε ότι αφορά τις εξελίξεις της ασφαλιστικής και όχι μόνο βιομηχανίας και την αναδιαμόρφωσή τους όταν χρειάζεται .

Όσον αφορά τις φορητές συσκευές όπως αναφέρθηκε και παραπάνω δεν θα ωφελήσουν μόνο τον ασφαλισμένο αλλά και την ασφαλιστική εταιρεία, αφού η συλλογή δεδομένων θα είναι πιο εύκολη και φυσικά ακριβέστερη αφού θα λαμβάνεται ατόφια. (Mike Adler, Gary Plotkin, Praatek Saxena,2017)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Έλεγχος απάτης

Ο ψηφιακός μετασχηματισμός των ασφαλιστικών εταιρειών ενέχει τον κίνδυνο δημιουργίας νέων τρόπων για να διενεργηθεί απάτη. Ένα σημαντικό πρόβλημα το οποίο αντιμετωπίζουν οι εταιρείες της ασφαλιστικής βιομηχανίας οι απάτες που είτε μέσω νέων συμβολαίων είτε από ήδη υπάρχοντα συμβόλαια μέσω της διαδικασίας αποζημιώσεων, με σκοπό την υπεξαίρεση χρηματικών ποσών και ξεπλύματος χρημάτων από παράνομες δραστηριότητες. Η χρήση τεχνολογίας και ο μετασχηματισμός των διαδικασιών σε ψηφιακές δημιουργεί νέους τρόπους εξαπάτησης όπως δημιουργία ψεύτικων προφίλ με διάφορα στοιχεία διαφορετικών πελατών αλλά πραγματικό ΑΜΚΑ, «σπάσιμο» λογαριασμών με αλλαγή IBAN και καταθέσεις ψευδών δικαιολογητικών προς αποζημίωση ή εκμετάλλευση λογαριασμών με επενδυτικά προγράμματα και συσσωρευμένα κεφάλαια.(Tim Sandle,2018)

Εκτός από την σημαντική οικονομική ζημία που προκαλούν τέτοια περαστικά, πληγείτε και η φήμη της εταιρείας προς τους καταναλωτές της ,το οποίο μακροπρόθεσμα θα επιφέρει οικονομικές επιπτώσεις.

Η λύση για την αποφυγή των περιστατικών απάτης φαίνεται να είναι υπαρκτή μέσω της βιομετρίας συμπεριφοράς (behavioral biometrics).Η συμπεριφορική βιομετρία ασχολείται με την δημιουργία μοτίβων ανθρώπινης συμπεριφοράς. Μερικά στοιχεία που συνθέτουν ένα μοτίβο είναι τα χαρακτηριστικά χρήσης του πληκτρολογίου και του ποντικιού, η ανάλυση βημάτων και φωνής και η ανάλυση υπογραφής. Τα μοτίβα συμπεριφοράς των νόμιμων χρηστών με αυτών που επιχειρούν να διαπράξουν την απάτη διαφέρουν με αποτέλεσμα η αναγνώριση τους να είναι ευκολότερη. Οι παράνομοι χρηστών παρουσιάζουν στοιχεία όπως εξοικείωση με δημιουργία λογαριασμών και ανεπτυγμένες δεξιότητες στη χρήση ηλεκτρονικού υπολογιστή.(Team of Biocatch,2017).

2.1.3 Ο ρόλος την ηγεσίας στον Ψηφιακό Μετασχηματισμό της ασφαλιστικής βιομηχανίας

Ο ψηφιακός μετασχηματισμός όπως αναφέραμε είναι μια μεταβατική κατάσταση όπου το επιχειρησιακό μοντέλο και η κουλτούρα μιας εταιρείας αλλάζει ριζικά λόγω υιοθέτησης της τεχνολογίας στις διαδικασίες τις εταιρείας .Η εφαρμογή της τεχνολογίας και η εκπαίδευση σε αυτήν χρειάζεται την τεχνογνωσία η οποία παρέχεται από ειδικούς ,αλλά η αποδοχή της καινοτομίας ως κουλτούρα χρειάζεται την στήριξη της ηγεσίας ενός οργανισμού για να γίνει αναπόσπαστο κομμάτι του. Ο ψηφιακός μετασχηματισμός μια εταιρείας μπορεί να αποτύχει αν η αλλαγή δεν είναι ευπρόσδεκτη από την ηγεσία λόγω αμφιβολιών για το ρίσκο που αναλαμβάνει ,λόγω μη κατανόησης της πολυπλοκότητας της μετάβασης στην νέα εποχή και έλλειψης ικανοτήτων να πείσει την ομάδα του για αυτό.(Douglas Karr,2018).

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Επίσης σημαντικό ρόλο παίζει και ο τύπος της ηγεσίας που ασκείται και αν αυτός ταιριάζει με την στρατηγική που έχει ορισθεί για το μετασχηματισμό.(Allio, 2015) .

Οι τύποι ηγεσίας είναι οι εξής:

- **Transormational Leadership**

Αυτός ο τύπος ηγεσίας παρακινεί τους εργαζόμενους να παίρνουν πρωτοβουλίες, να προτείνουν ιδέες και αφήνουν χώρο στην καινοτομία χωρίς να φοβούνται ότι θα "ταράξουν" τις δεδομένες διαδικασίες. Στην ουσία δημιουργούν εργαζόμενους αυτοελεγχόμενους με εμπιστοσύνη στην κρίση τους και τις ενέργειες τους πάνω στο αντικείμενο της εργασίας τους .Αυτό αποτελεί διαφορετική εμπειρία και για τους εργαζόμενους ,οι οποίοι αφήνουν πίσω από ένα καθεστώς μικροδιαχείρησης(micromanagement) και αποκτούν περισσότερο χώρο δράσης και έκφρασης.(Sarah K. White,2018)

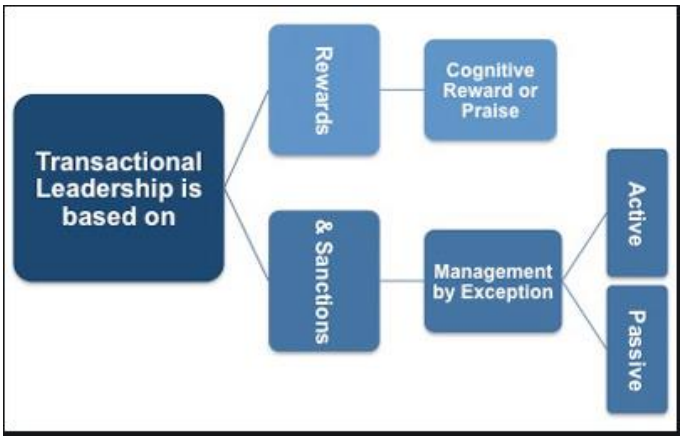


Εικόνα 2,Transformational Leadership

- **Transactional Leadership**

Στον τύπο αυτόν ηγεσίας οι εργαζόμενοι δεν έχουν το περιθώριο πρωτοβουλίας ,αντίθετα δρουν βάση αυστηρών διαδικασιών οι οποίες δύσκολα διαφοροποιούνται και ακολουθούν το επιχειρηματικό σχέδιο βάση κανόνων και εντολών από την ηγεσία χωρίς παρέκκλισή.

Η ηγεσία καθορίζει τους κανόνες και τις διαδικασίες και φροντίζει να παραμένουν αμετάβλητες και οι ενέργειες των ομάδων να στηρίζονται σε αυτές . Προκειμένου να εδραιωθεί η νοοτροπία αυτή και να δοθεί έμφαση στην σημαντικότητα της εφαρμογής υπάρχει η ανταμοιβή και η τιμωρία ως κίνητρο.(Kendra Cherry,2019).



Εικόνα 3,Transactional Leadership

- **Innovative (participative) leadership**

Ο συγκεκριμένος τύπος ηγεσίας χαρακτηρίζεται από την συμμετοχή όλων των υπαλλήλων μιας εταιρείας στις αποφάσεις της. Οι υπάλληλοι είναι ενήμεροι για τις δραστηριότητες της εταιρείας και κάθε μελλοντικής της κίνηση εγκρίνεται από την πλειοψηφία (George N. Root III,2019).Τα χαρακτηριστικά αυτής της διοίκησης απέχουν πολύ από την παραδοσιακή ιδέα σχετικά με την ηγεσία, η οποία θέτει ως άξονα την ιεραρχία ,και το πως αυτή ασκείται σε ένα οργανισμό και όσο ριζοσπαστική και αν μοιάζει οι εταιρείες που την υιοθέτησαν κέρδισαν την εκτίμηση των καταναλωτών ,καθιερώνοντας την ως μια δυναμική ενναλακτική ηγεσίας. Η λήψη των αποφάσεων ίσως να καθυστερεί λόγω της διαδικασίας αλλά υπάρχουν πλεονεκτήματα που καθιστούν την καθυστέρηση

αποδεκτή (Anastasia for Cleverism,2019). Οι εργαζόμενοι της εταιρείας αποδέχονται αρχικά ευκολότερα οποιαδήποτε αλλαγή λόγω ότι έχουν υπάρξει μέρος αυτής εγκρίνοντας την, συνεπώς νιώθουν ελεύθεροι να προχωρήσουν σε νέες προτάσεις και ιδέες ενώ παράλληλα ενδιαφέρονται περισσότερο για το για το αντίκτυπο τους. Επίσης η αίσθηση της σημαντικότητας που προσφέρει στον εργαζόμενο αυτός ο τρόπος ηγεσίας αυξάνει το επίπεδο διατηρισημότητας υπαλλήλων στην εταιρεία και οι έριδες μειώνονται, ενώ αυξάνεται η συνεργασία εφόσον πλέον υπάρχει κοινός στόχος (George N. Root III,2019).Συμπεραίνουμε, λοιπόν ,ότι το κίνητρο προς τον εργαζόμενο αποτελεί τον πυρήνα για το συγκεκριμένο μοντέλο ηγεσίας.



Εικόνα 4,Participative Leadership

Δεν υπάρχει όμως ένα μόνο ένας τύπος ηγεσίας ο οποίος ταιριάζει σε κάθε οργανισμό και θα έχει θετικό αντίκτυπο στην διαδικασία ψηφιακού μετασχηματισμού της .Η μείξη τους και η αλληλοσυμπλήρωση τους θα ήταν ιδανική ,ενώ θα πρέπει να ευθυγραμμίζεται με την οργανωτική νοοτροπία και θα πρέπει να παρέχεται στους εργαζόμενους η απαραίτητη κατάρτιση, γνώση και καθοδήγηση ώστε να απορροφήσουν την αλλαγή πιο αποτελεσματικά. Ο τύπος ηγεσίας θα πρέπει να ενισχυθεί και με άλλα χαρακτηριστικά όπως είναι η επιχειρηματική γραμμή της εταιρείας όσον αφορά την μεταβασή της στην νέα εποχή της όσο και χαρακτηριστικά του ηγέτη.(Dr. Mouhamadou Sow & Dr. Solomon Aborbie,2018). Συνεπώς για να είναι πετυχημένος ο ψηφιακός μετασχηματισμός, ο οργανισμός θα πρέπει να έχει διαθέσιμους πόρους (υλικούς ή άυλους) και να έχει ορίσει τον τρόπο λήψης αποφάσεων (τύπο ηγεσίας) ενώ θα πρέπει να έχει δημιουργήσει ένα μελλοντικό πλάνο μετά τον μετασχηματισμό σχετικά με τους στόχους, την βιωσιμότητα του και την ανταγωνιστικότητα του στον ψηφιακό κόσμο (Dahlstrom et al., 2017). Επίσης η εταιρεία που θα αποφασίσει να προχωρήσει σε ψηφιακό

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

μετασχηματισμό θα πρέπει να είναι έτοιμη να προχωρήσει και σε ριζικές αλλαγές στην οργανωτική δομή, στις διαδικασίες και την κουλτούρα της. Για να συμβεί όμως αυτό, ένας δημιουργικός ηγέτης με καινοτόμες ιδέες και την προσωπικότητα να τις στηρίξει είναι πολύ σημαντικός παράγοντας για να ολοκληρωθεί και να επιφέρει μακροπρόθεσμα αποτελέσματα. (Demirkan et al., 2016). Ο ίδιος ηγέτης πρέπει να έχει την ικανότητα να πείσει την ομάδα του για τα αποτελέσματα του μετασχηματισμού, προσφέροντας κίνητρο και επεξήγηση πως αυτά θα ωφελήσουν την προσωπική εργασία και εξέλιξη τους. Παράλληλα πρέπει να έχει την συναισθηματική ευφυΐα να αναγνωρίσει μέχρι που φτάνουν οι δυνατότητες του και να δείξει εμπιστοσύνη σε ικανότερους συναδέλφους με περισσότερες γνώσεις(Dahlstrom et al., 2017).

Καταλήγουμε ότι οι εταιρείες που υποβάλλονται σε ψηφιακό μετασχηματισμό πρέπει να διαθέτουν τα κατάλληλα εργαλεία για την υλοποίηση του. Οι ηγέτες των οργανισμών πρέπει να κατανοήσουν αυτήν την διαδικασία, να αποδεχτούν την αλλαγή και να κατανοήσουν ότι ο τύπος ηγεσίας τους θα πρέπει να ευθυγραμμίζονται με τη στρατηγική της εταιρείας, ώστε να είναι πιθανότερο να είναι πιο αποτελεσματικά. Επίσης είναι σημαντικό οι ηγέτες που μπορούν να προσαρμοστούν στην αλλαγή αποτελεσματικά, να υποστηρίξουν μια διαφορετική νοοτροπία και να έχουν τα χαρακτηριστικά να αντέξουν τις προκλήσεις κατά τη διάρκεια της φάσης υλοποίησης Ταυτόχρονα, πρέπει να αναπτύξουν ένα αποδεκτό επίπεδο εμπιστοσύνης στους υπαλλήλους που έχουν συγκεκριμένες γνώσεις σχετικά με αυτού του είδους τις αλλαγές, ώστε να μην υπάρχει εμπόδιο στην δημιουργικότητα και την καινοτομία και να υποστηρίζονται ώστε να κάνουν τη διαφορά κατά τη διάρκεια και μετά τον ψηφιακό μετασχηματισμό και μακροπρόθεσμα για τον οργανισμό.

2.2 Text Mining και Ασφαλιστική Αγορά

Ο ψηφιακός μετασχηματισμός των ασφαλιστικών εταιρειών, θα έχει ως αποτέλεσμα ένα τεράστιο όγκο δεδομένων από πολλά διαφορετικά κανάλια ,τα οποία διαχωρίζονται σε δομημένα και μη δομημένα δεδομένα. Τα μη δομημένα δεδομένα είναι αυτά τα οποία έρχονται σε μορφή κειμένου και μπορούν να αναλυθούν και να εξαχθούν πληροφορίες από αυτά μέσω τεχνικών text mining,για το οποίο θα μιλήσουμε εκτενέστερα στα επόμενα κεφάλαια και αφορά την εξόρυξη πληροφορίας από δεδομένα κειμένου. Τα αριθμητικά δεδομένα δεν είναι πλέον τα μοναδικά που μπορούν να δώσουν απαντήσεις, αντίθετα οι τεχνικές text mining μπορούν να εξορύξουν πληροφορίες από ένα κείμενο, όπως το συναίσθημα που περιέχει ή το ύφος γραφής του.

Η ασφαλιστική βιομηχανία είναι ένας τομέας ο οποίος μπορείς να επωφεληθεί σημαντικά με την εφαρμογή και την ενσωμάτωση του text mining στις διαδικασίες της. Σε καθημερινή βάση, οι

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ασφαλιστικές εταιρείες συγκεντρώνουν τεράστιους όγκους κειμένων από πολλά κανάλια, όπως το ηλεκτρονικό ταχυδρομείο, οι ιστότοποί τους, τα δίκτυα κοινωνικής δικτύωσης ή πλατφόρμες μέσω των οποίων εξυπηρετούνται οι πελάτες τους. Τα κείμενα αυτά περιέχουν από παράπονα και αιτήματα μέχρι πολιτικές, εκθέσεις εμπειρογνωμόνων και ιατρών, αξιώσεις και άλλα στοιχεία που σχετίζονται με τον πελάτη. (Noel Bambrick,2016).

Δεδομένου ότι η στρατηγική των ασφαλιστικών εταιρειών είναι κατά κύριο λόγω πελατοκεντρική ,η εξόρυξη κειμένου μπορεί να γίνει ένα δυνατό εργαλείο στο χέρια της για την κατανόηση των αναγκών του πελάτη της ,την άποψη τους ,την δυσαρέσκεια ή ικανοποίηση τους ,ακόμα και την πιθανότητα να λήξει την συνεργασία του μαζί τους,

Customer Service και Customer Engagement

Η χρήση των text analytics είναι μια ξεκάθαρη επένδυση στην ανάπτυξη του Customer Engagement και Customer Service ,στην καλύτερη δόμηση διαδικασιών και ένα βήμα προς την κουλτούρα της καινοτομίας που λείπει από την συγκεκριμένη βιομηχανία. Η δυνατότητα που προσφέρουν στην επεξεργασία της φυσικής γλώσσας και των μη δομημένων δεδομένων από αυτήν κάνουν εφικτή την εφαρμογή chat bots σε μια ασφαλιστική πλατφόρμα και τα δεδομένα που συγκεντρώνονται από αυτά μπορούν να αναλυθούν ώστε να γίνουν προσωποποιημένες προτάσεις στον ασφαλισμένο, να κανονιστεί ραντεβού με γιατρό ή να γίνουν υπενθυμίσεις σχετικά με εκκρεμότητες. Με αυτόν τον τρόπο προστίθεται σημαντική αξία στην εμπειρία του πελάτη με έξυπνη, περισσότερο στοχευμένη και 24ωρη υποστήριξη .Το ίδιο μπορεί να πραγματοποιηθεί με οποιαδήποτε άλλη γραπτή επαφή (π.χ.μέσω e-mail) ή ακόμα και τηλεφωνική αν η συνομιλία μετατραπεί σε μορφή κειμένου. Επίσης, η άποψη και το συναίσθημα του ασφαλισμένου απέναντι στα προιόντα ή την εταιρεία ανιχνεύεται μέσω email, κοινωνικών δικτύων, feedback πεδίων, live chat, γενικού περιεχομένου στο Internet. Η ανάλυση τους, δίνει την δυνατότητα να εντοπιστούν νέες τάσεις στον τομέα, να προσδιοριστεί η αντίληψη του κοινού απέναντι στα προϊόντα της εταιρείας και την φήμη της, κατί το οποίο είναι πολύ σημαντικό ώστε να διατηρηθεί η καλή φήμη και να βελτιωθεί αυτή που πάσχει πριν δημιουργηθεί μεγαλύτερη ζημιά. Τα σχόλια των πελατών που περιέχουν σημαντικά υψηλά επίπεδα αρνητικών συναισθημάτων μπορούν να μεταφερθούν στις κατάλληλες ομάδες για να τους βοηθήσουν να εστιάσουν τις προσπάθειές τους στα προβλήματα που φαίνεται να υπάρχουν μέσα από αυτά. Τέλος, ο χρόνος ανταπόκρισης σε ερωτήματα ή παράπονα μειώνεται καθώς πραγματοποιείται για την αυτόματη δρομολόγηση και κατηγοριοποίηση των ερωτημάτων των πελατών.(Noel Bambrick,2016).

Το Customer Engagement είναι ένα μόνο κομμάτι στο οποίο η χρήση του text mining και των text analytics θα έφερνε θετικά αποτελέσματα. Μερικοί ακόμα τομείς είναι οι εξής:

Ανίχνευση απάτης

Το φαινόμενο της απάτης στην ασφαλιστική αγορά είναι αρκετά συχνό όσον αφορά την διαδικασία των αποζημιώσεων.

Μέχρι τώρα η αντιμετώπιση της είχε την εξής διαδικασία:

- Ανίχνευση μιας ύποπτης αποζημίωσης.
- Ανάλυση της αποζημίωσης και των στοιχείων από τα ανάλογο τμήματα της εταιρείας και τους εμπλεκόμενους στην διαδικασία αποζημίωσης.
- Επισήμανση της αποζημίωσης ως ψευδής και στην συνέχεια άρνηση της καταβολής της και ίσως δικαστικές ενέργειες.

Η διαδικασία αυτή λειτουργεί δειγματοληπτικά χωρίς να λαμβάνονται υπόψιν όλα τα δεδομένα από όλες τις πηγές τις εταιρείας, ειδικά τα μη δομημένα δεδομένα σε μορφή κειμένου, με αποτέλεσμα πολλές απάτες να λαμβάνουν μέρος απαρατήρητες.

Οι δηλώσεις αποζημιώσεων και τα κοινωνικά δίκτυα είναι πηγές που μπορεί να εκμεταλλευτεί η ασφαλιστική βιομηχανία για την αποφυγή απάτης μέσω των αποζημιώσεων. Στο κείμενο των δηλώσεων αποζημιώσεων υπάρχουν πληροφορίες οι οποίες δεν καταχωρούνται σε ένα εταιρικό σύστημα και μπορούν να κάνουν την διαφορά με λεπτομέρεια που έχει καταγραφεί αλλά δεν έχει επεξεργαστεί. Μιλώντας πιο τεχνικά αυτή η λεπτομέρεια είναι που θα παίξει το ρόλο της μεταβλητής που θα αλλάξει ολόκληρο το μοντέλο της ανίχνευση της απάτης.(James Ruotolo,2013).Επίσης μπορούν να ανιχνευθούν κοινά πρότυπα κειμένων σε πολλαπλές δηλώσεις ατυχημάτων των οποίων η ανάλυση πιθανόν να δράσει συμπληρωματικά στην απόδειξη της απάτης. .(Noel Bambrick,2016).

Διαχείριση Αποζημιώσεων

Η ανάλυση του κειμένου που περιέχεται σε μια δήλωση αποζημίωσης είναι άλλη περιοχή για τη χρήση του text mining . Ανεξάρτητα από το εισερχόμενο κανάλι, οι αποζημιώσεις ανάλογα με το περιεχόμενο τους μπορούν να ταξινομηθούν αυτόματα ανάλογα με τα προϊόντα, τις υπηρεσίες και κυρίως και τη βαρύτητά τους ,προκειμένου να τεθούν σε προτεραιότητα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Επίσης η ανάλυση θα ήταν χρήσιμη ως συμπληρωματική πρόβλεψη ,στην συχνότητα των αποζημιώσεων, για παράδειγμά νοσηλείας οι οποίες κοστίζουν περισσότερο σε μια ασφαλιστική εταιρεία ,και κατ' επέκταση στα δυνητικά έξοδα της.

Για να διασφαλιστεί ότι η πληροφορία που φτάνει στην ασφαλιστική εταιρεία δεν είναι ελλιπής έχουν αρχίσει να εμφανίζονται B2B προμηθευτές ,οι οποίοι μπορούν να εξοπλίσουν μια ασφαλιστική εταιρεία με λογισμικό το οποίο μπορεί να μεταφέρει μια ομιλία ενός πελάτη κατά τη διάρκεια τηλεφωνικής κλήσης να συμπληρώσει αυτόματα τη δήλωση αποζημίωσης του ασφαλισμένου, ενώ αν έχει παραλειφθεί κάποια δήλωση με στοιχεία που δεν έχουν συμπληρωθεί θα μπορούσε να επικοινωνήσει αυτόματα με τον ασφαλισμένο μέσω e-mail.Με αυτόν τον τρόπο οι πληροφορίες είναι ακριβέστερες και ταχύτερες το οποίο σημαίνει πιο άμεση εξυπηρέτηση και διευκόλυνση του ανθρώπινου χρήστη μέσα στην διαδικασία.(Raghav Bharadwaj,2019)

Τα οφέλη του text mining στην ασφαλιστική βιομηχανία αφορούν τους ασφαλισμένους πελάτες ,αλλά και τις ασφαλιστικές εταιρείες. Από την πλευρά της εταιρείας αυξάνεται η αποδοτικότητα των υπαλλήλων και μειώνεται ο φόρτος εργασίας εφόσον υπάρχουν αποδοτικότερα εργαλεία που υποστηρίζουν την εργασία τους , ενώ η έγκαιρη αντιμετώπιση της απάτης είναι ευκολότερη και το κόστος που προέρχεται από αυτήν μειώνεται σε μεγάλο βαθμό. Η εισροή δεδομένων που αφορά την αγορά βοηθά την εταιρεία να την κατανοήσει και να οργανώσει ανάλογα την επιχειρηματική στρατηγική της. Επίσης ,εφόσον η εργασία των υπαλλήλων είναι αποδοτικότερη μειώνεται ο χρόνος ανταπόκρισης προς τον πελάτη, με αποτέλεσμα την ικανοποίηση του η οποία αυξάνει τους δείκτες διατηρισημότητας του και ταυτόχρονα δημιουργεί την απαραίτητη φήμη για εισροή νέου πελατολογίου (Jose C. Gonzalez,2014).

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3

Θεωρητικό Υπόβαθρο

Το Text Mining είναι ένας διεπιστημονικός τομέας, ο οποίος συνδυάζει τις τεχνικές της εξόρυξης από δεδομένα (Data Mining), την μηχανική μάθηση (Machine Learning), την επεξεργασία της φυσικής γλώσσας (Natural Language Processing), την ανάκτηση πληροφορίας (Information Retrieval), την εξαγωγή πληροφορίας (Information Extraction) και την διαχείριση γνώσης (Knowledge Management),με σκοπό να κατανοήσει την πληροφορία που προέρχεται από ένα σύνολο κειμένων. Οι πηγές των κειμένων προέρχονται από μηνύματα ηλεκτρονικού ταχυδρομείου ή δημοσιεύσεων σε κοινωνικά μέσα ,από ,ημι-μορφοποιημένο κείμενο (html, xml κ.τ.λ) και μορφές εγγράφων όπως OpenDocument,,Microsoft Word ,Microsoft Powerpoint .Στα κείμενα αυτά ,τα οποία συλλέγονται από τις διάφορες πηγές που αναφέραμε χρησιμοποιούνται αυτοματοποιημένοι μέθοδοι με στόχο την κατανόηση τους και την εξαγωγή συμπερασμάτων από αυτά.(Jeffrey L. Solka,2008)

3.1 Μέθοδοι και Τεχνικές Text Mining

Το πρώτο βήμα για την εφαρμογή text mining είναι να καθοριστεί ο τρόπος με τον οποίο θα γίνει η ανάκτηση της πληροφορίας. Υπάρχουν τρείς κύριες μέθοδοι για αυτή την διαδικασία και είναι οι εξής:

Based Method (TBM)

Phrase Based Method (PBM)

Concept Based Method (CBM)

- **Term Based Method (TBM):** Με έννοια term (όρος) εννοούμε την λέξη αυτή η οποία είναι σημαντική μέσα στο κείμενο και είναι απαραίτητη ως πληροφορία για την ανάλυση του κειμένου .Ο κάθε όρος συσχετίζεται με μια τιμή γνωστή ως βάρος ,η οποία καθορίζει την σημαντικότητα του όρου στο κείμενο. Ο συντελεστής βαρύτητας του όρου W υπολογίζεται από την εξής μαθηματική πράξη Wi=TF(ti,d)* IDF(ti) ,όπου TF(ti,d) (Term Frequency) είναι ο αριθμός των φορών που εμφανίζεται ένας όρος (term) σε ένα έγγραφο (document) και όπου

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

IDF είναι ο αριθμός των εγγράφων (document) που ο όρος(term) εμφανίζεται τουλάχιστον μια φορά. Η αντίστροφη συχνότητα εγγράφου του όρου (IDF) είναι χαμηλή εάν εμφανίζεται σε πολλά έγγραφα και υψηλότερη εάν ο όρος εμφανίζεται μόνο σε ένα έγγραφο.(S.Brindha et al,2016).Το γεγονός όμως ότι πολλές λέξεις έχουν την ίδια σημασία ή μια λέξη έχει πολλές έννοιές δημιουργεί μειονέκτημα στην μέθοδο αυτή.(Sonali Vijay Gaikwad et al,2014)

- **Phrase Based Method (PBM)** : Η μέθοδος αυτή ανακτά πληροφορία από φράσεις μέσα στο κείμενο, οι οποίες είναι σαφέστερες και λιγότερο αμφίσημες από τις,αλλά η αρνητική πλευρά είναι ότι έχουν μικρότερη συχνότητα εμφάνισης και πιθανόν είναι να περιέχουν πολλές περιττές πληροφορίες μέχρι να φτάσει ο αναλυτής σε αυτήν που τον ενδιαφέρει πραγματικά. (Sonali Vijay Gaikwad et al,2014).Η μέτρηση της συχνότητας σε αυτήν μέθοδο γίνεται με την παρακάτω μαθηματική συνάρτηση :

$$Ctf = \frac{\sum_{n=1}^{sn} ctfn}{sn}$$

  - ➔ ctf είναι ο αριθμός που εμφανίζεται η έννοια c (concept) στην δομή της πρότασης s (sentence).
  - ➔ sn είναι ο συνολικός αριθμός των προτάσεων σ που περιέχουν την έννοια c στο έγγραφο d.

Συνεπώς το Ctf είναι ο μέσος όρος των εμφανίσεων των εννοιών c στον συνολικό αριθμό προτάσεων του εγγράφου d.Η έννοια c η οποία έχει μεγάλο Ctf value είναι και περισσότερο χρήσιμη στην ανάλυση του κειμένου. (S.Brindha et al,2016)

- **Concept Based Method (CBM):**Στην concept based μέθοδο κάθε έννοια του εγγράφου συνδυάζεται με άλλες έννοιες και συγκρίνεται και με έννοιες προηγούμενων επεξεργασμένων εγγράφων .Για να γίνει περισσότερο κατανοητό ,στην term based μέθοδο αναλύεται η σημαντικότητα του όρου στο έγγραφο ενώ στην concept based μέθοδο ,η σημαντικότητα του όρου καθορίζεται και από την σημαντικότητα του στην πρόταση που περιέχεται αλλά και σε άλλα έγγραφα που μπορεί να εμφανίζεται. Με αυτόν τον τρόπο η ανάλυση είναι πληρέστερη και πιο ολοκληρωμένη αφού όχι μόνο βοηθά στην εύρεση σημαντικών εννοιών σε ένα έγγραφο αλλά και όμοιων σε άλλα έγγραφα του κειμένου.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Αυτός είναι και στόχος της concept based ανάλυσης, η ομοιότητα που υπάρχει μεταξύ των εγγράφων και η εξόρυξη της . Υπάρχουν τρεις σημαντικοί παράγοντες : η έννοια ή οι έννοιες (concepts) οι οποίες καθορίζουν την σημαντικότητα της πρότασης, η συχνότητα εμφάνισης μια έννοιας στην πρόταση ,η οποία δείχνει ποσό σημαντική είναι για την αντίληψη του νοήματος της πρότασης αλλά και του κυρίως θέματος του εγγράφου και ο αριθμός των εγγράφων που περιέχουν τις έννοιες που αναλύονται και χρησιμοποιείται για τη διάκριση μεταξύ των υπόλοιπων εγγράφων στον υπολογισμό της ομοιότητας.(S.Brindha et al,2016).

Αφού η πληροφορία έχει συλλεχθεί με μια ή παραπάνω από τις μεθόδους που αναφέραμε, στην συνέχει πραγματοποιείται μια προ-επεξεργασία με μια σειρά τεχνικών, οι οποίες βοηθούν στην βελτιστοποίηση και την αύξηση της αποτελεσματικότητας της πληροφορίας μέσω της μείωσης του λεξιλογίου και συνεπώς του μεγέθους των εγγράφων. Η προ-επεξεργασία αποτελείται από τα παρακάτω στάδια (Mehdi Allahyari et al,2017):

Αφαίρεση της δομής των κειμένων: Στο στάδιο αυτό, τα κείμενα οποιασδήποτε μορφής (HTML, PDF, κτλ), αναπαρίστανται ως απλό κείμενο (.txt) και στη συνέχεια επεξεργάζονται με σκοπό να απλοποιηθεί η δομή τους. Ουσιαστικά, παύει να υφίσταται η παραγραφοποίηση, η ύπαρξη θεματικών ενοτήτων και κεφαλαίων.

Λημματοποίηση (Tokenization): Ένα κείμενο αποτελείται από λήμματα (tokens), τα οποία μπορεί να είναι λέξεις, αριθμοί, σύμβολα ή σημεία στίξης. Επειδή οι αριθμοί, τα σύμβολα και τα σημεία στίξης δεν προσφέρουν καμία πληροφορία και δεν έχουν καμία σχέση με το περιεχόμενο του κειμένου μπορούν να αναγνωριστούν σε αυτό το σημείο και να αφαιρεθούν και παραμένουν μόνο οι λέξεις του κειμένου.

Αφαίρεση των Stopwords: Ως stopword θεωρείται ένας όρος ο οποίος από τη μια έχει μεγάλη συχνότητα εμφάνισης μέσα στο κείμενο και από την άλλη δεν σχετίζεται με το περιεχόμενο του κειμένου. Ένας τέτοιος όρος μπορεί να είναι μία πρόθεση, ένα άρθρο, ή ακόμη και ένας σύνδεσμος δύο προτάσεων. Οι όροι αυτοί λειτουργούν ως «θόρυβος» με κίνδυνο τη μείωση της απόδοσης του συστήματος, συνεπώς πρέπει να αφαιρούνται.

Λεξικογραφική Ανάλυση (Pos Tagging): Στην διαδικασία της λεξικογραφικής ανάλυσης αναγνωρίζεται το μέρος του λόγου που αποτελεί κάθε λέξη του κειμένου (ουσιαστικό, ρήμα, επίθετο, κτλ) και επιλέγονται κυρίως τα ουσιαστικά διότι φέρουν τη πιο σημαντική πληροφορία των κειμένων.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Stemming: Η διαδικασία του stemming έχει ως σκοπό τον αυτόματο χειρισμό των καταλήξεων των λέξεων και την μείωση αυτών στις ρίζες τους και γίνεται συνήθως με αφαίρεση των οποιοδήποτε συνημμένων επιθεμάτων και των προθεμάτων .Για τη διαδικασία του stemming έχουν αναπτυχθεί αρκετοί αλγόριθμοι με γνωστότερο τον αλγόριθμο του Porter .

Μετά την εφαρμογή των παραπάνω βημάτων, πρέπει τα κείμενα να αναπαρασταθούν σε μία μορφή που να είναι επεξεργάσιμη και η πιο γνωστή μέθοδος είναι η διανυσματική αναπαράσταση. Στην αναπαράσταση αυτή δημιουργείται ένας διανυσματικός χώρος, όπου κάθε κείμενο μπορεί να αναπαρασταθεί ως ένα διάνυσμα. Ο διανυσματικός αυτός χώρος αποτελείται από τόσες διαστάσεις όσοι είναι και οι μοναδικοί όροι των κειμένων. Έτσι λοιπόν, στην διανυσματική αναπαράσταση κάθε κείμενο ως αναπαρίσταται ως διάνυσμα από ένα σύνολο όρων. Για το λόγο αυτό γίνεται και η προ-επεξεργασία που περιεγράφηκε παραπάνω, προκειμένου να εντοπισθούν οι μοναδικοί όροι που θα αποτυπώσουν το νόημα του κειμένου, οι οποίοι στη συνέχεια θα αποτελέσουν τις διαστάσεις του διανυσματικού χώρου.

Υπάρχουν δύο βασικοί τρόποι που χρησιμοποιούνται για τη διανυσματική αναπαράσταση κειμένων:

Boolean Model : Σε ένα Boolean διανυσματικό μοντέλο η κάθε διάσταση του διανυσματικού χώρου μπορεί να πάρει τη τιμή 1 ή τη τιμή 0. Όπως προαναφέραμε κάθε όρος ενός κειμένου αντιστοιχεί σε μία διάσταση. Επομένως, η τιμή 1 σε μία διάσταση σημαίνει ότι ο όρος εκείνος εμφανίζεται στο κείμενο, ενώ η τιμή 0 σημαίνει ότι ο όρος αυτός δεν υπάρχει στο κείμενο.

Term-Weight Model: Σε ένα Term – Weight διανυσματικό μοντέλο σημαντικό ρόλο παίζει η συχνότητα εμφάνισης των όρων στα κείμενα. Αυτό σημαίνει ότι σε κάθε διάσταση του διανυσματικού χώρου αντιστοιχεί μία τιμή, η οποία υποδηλώνει τη συχνότητα εμφάνισης του εκάστοτε όρου στο κείμενο. Μία αρκετά παρόμοια μέθοδος αυτού του τύπου είναι το Term Frequency Inverse Document Frequency Weighting (TF – IDF). (Mehdi Allahyari et al,2017)

Μετά την διαδικασία της προ-επεξεργασίας των δεδομένων ακολουθούν οι παρακάτω τεχνικές text mining:

- Εξαγωγή Πληροφοριών (Information Extraction)
- Ταξινόμηση (Classification)
- Ομαδοποίηση (Clustering)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Συνόψιση (Summarization)
- Απεικόνιση Πληροφορίας (Information Visualization)

Εξαγωγή Πληροφοριών (Information Extraction):

Η διαδικασία της εξαγωγής πληροφοριών βοηθά στην μετατροπή του κειμένου από ένα σύνολο αδόμητων δεδομένων σε μια δομημένη μορφή. Το λογισμικό το οποίο έχει επιλεχθεί να χρησιμοποιηθεί αναγνωρίζει σημαντικές φράσεις ή λέξεις κλειδιά μέσα στο κείμενο και το πως σχετίζονται μεταξύ τους. Αν για παράδειγμα σε ένα μεγάλης έκτασης κείμενο περιέχονται πληροφορίες σχετικά με πρόσωπα ,τοποθεσίες και ημερομηνίες ,η διαδικασία της εξαγωγής πληροφορίων θα συνδέσει τα πρόσωπα αυτά με τις ανάλογες τοποθεσίες και ημερομηνίες και θα εξαιρέσει την πληροφορία που δεν έχει χρησιμότητα για τον ενδιαφερόμενο. (Sonali Vijay Gaikwad et al,2014)

Ταξινόμηση (Classification):

Η διαδικασία της ταξινόμησης αναγνωρίζει τα βασικά στοιχεία ενός κειμένου αντιστοιχίζοντας τα σε μια η περισσότερες κατηγορίες βάση της συχνότητας εμφάνισης τους στο κείμενο που αναλύεται. Τα δεδομένα, δηλαδή το κείμενο το οποίο θα αναλυθεί χωρίζεται σε ένα train και ένα test αρχείο τα οποία περιέχουν έγγραφα (documents) του κειμένου. Ο σκοπός είναι να βρεθεί το κατάλληλο μοντέλο ταξινόμησης το οποίο θα εκπαιδευθεί στο train αρχείο ώστε να μπορεί να αντιστοιχήσει τα δεδομένα του test αρχείου στην σωστή κατηγορία. Το ποσοστό των σωστά ταξινομημένων εγγράφων προς το σύνολο των εγγράφων αξιολογεί το μοντέλο και ονομάζεται accuracy.(Mehdi Allahyari et al,2017)

Μοντέλα ταξινόμησης είναι τα Naive Bayesian classifier,Nearest Neighbour classifier, Decision tree classifiers and Support Vendor Machines,τα οποία θα αναλύσουμε παρακάτω.

Naive Bayesian classifier: Το μοντέλο Naïve Bayesian είναι βασισμένο στην θεωρία πιθανοτήτων (probabilistic model) και είναι το πιο διαδεδομένο. Βασίζεται στην παραδοχή ότι η κατανομή ενός όρου είναι ανεξάρτητη από την κατανομή ενός άλλου όρου και χρησιμοποιεί το train αρχείο εγγράφων ώστε να εκτιμήσει τις παραμέτρους και στην συνέχεια αντιστοιχεί τα έγγραφα του test αρχείου στις πιθανότερα "ταιριαστές" κατηγορίες. (Mehdi Allahyari et al,2017)

Nearest Neighbour classifier: Αυτό το μοντέλο ταξινομεί τα έγγραφα του κειμένου βάση μετρικών απόστασης (distance-based measures),π.χ. η Ευκλείδεια απόσταση, καθώς θεωρεί ότι η ομοιότητα μεταξύ των εγγράφων αυξάνεται όταν η απόσταση τους είναι πιο κοντινή στην κατηγορία που ανήκουν. (Mehdi Allahyari et al,2017)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Decision tree classifiers: Τα δέντρα απόφασης ταξινομούν τα δεδομένα ιεραρχικά βάση συνθήκης σε ένα χαρακτηριστικό, π.χ. ύψος, φύλο, ηλικία το οποίο παίζει το ρόλο κόμβου και η εφαρμογή της συνθήκης δημιουργεί υποσύνολα. Στην περίπτωση του κειμένου, ένας όρος παίζει τον ρόλο του κόμβου με συνθήκη την ύπαρξη ή όχι του όρου στο έγγραφο. (Mehdi Allahyari et al,2017)

Support Vendor Machines (SVM):Τα SVM αλγόριθμοι ταξινόμησης είναι μια μορφή γραμμικών μοντέλων ταξινόμησης, οι οποίοι ταξινομούν τα έγγραφα του κειμένου που αναλύεται βάση της γραμμικής συσχέτισης των εγγράφων (Mehdi Allahyari et al,2017). Επιπλέον,όμως, οι SVM αλγόριθμοι έχουν την δυνατότητα να εκτελούν αποτελεσματικά μη γραμμική ταξινόμηση χρησιμοποιώντας αυτό που ονομάζεται kernel trick.Με την έννοια kernel trick αναφερόμαστε στα kernel functions, τα οποία μεταφέρουν τα δεδομένα σε high-dimensional spaces και έχει βρεί εφαρμογή στα δεδομένα που αντλούνται από κείμενο (Sunil Ray, 2017).Οι high dimensional spaces προκύπτουν ως ένας τρόπος μοντελοποίησης των dataset με πολλά χαρακτηριστικά. Ένα τέτοιο σύνολο δεδομένων μπορεί να αναπαρίσταται σε ένα χώρο (space) όπου κάθε εγγραφή να αντιπροσωπεύεται ως σημείο στον χώρο με τη θέση του να ορίζεται ανάλογα με τις τιμές χαρακτηριστικών του (Skillicorn, David B.,2012)

Ομαδοποίηση (Clustering):

Η τεχνική της ομαδοποίησης χρησιμοποιείται ώστε να ομαδοποιηθούν φράσεις ή λέξεις του κειμένου ,οι οποίες έχουν το ίδιο περιεχόμενο. Δημιουργούνται τα λεγόμενα clusters τα οποία περιέχουν έγγραφα παρόμοιου περιεχομένου. Η διαφορά με την ταξινόμηση είναι ότι ένα έγγραφο μπορεί να εμφανίζεται σε παραπάνω από ένα cluster,οπότε διασφαλίζεται ότι αν ένα έγγραφο είναι χρήσιμο θα συμπεριληφθεί στην ανάλυση. (Sonali Vijay Gaikwad et al, 2014).Ένας συχνά χρησιμοποιούμενος αλγόριθμος για την δημιουργία clusters είναι ο K-means.Ένα παράδειγμα για το πως δημιουργεί clusters ο K-means μπορούμε να δούμε παρακάτω όπως παρουσιάστηκε στο άρθρο «A Brief Survey of TextMining : Classification, Clustering and Extraction Techniques, MehdiAllahyari et al,2017»:

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ALGORITHM 1: k-means clustering algorithm

Input : Document set $\mathcal{D}$, similarity measure $\mathcal{S}$, number k of cluster

Output : Set of k clusters

initialization

Select randomly k data points as starting centroids.

while *not converged* **do**

  Assign documents to the centroids based on the closest similarity.

  Calculate the the cluster centroids for all the clusters.

end

return *k clusters*

Συνόψιση (Summarization) :

Η συνόψιση ενός κειμένου είναι χρήσιμη στην προσπάθεια κατανόησης του περιεχομένου ενός κειμένου, με σκοπό να καταλάβει ο αναλυτής αν το κείμενο αυτό καλύπτει ή όχι τις απαιτήσεις του και να προχωρήσει σε βαθύτερη ανάγνωση του. Η τεχνική της συνόψισης δημιουργεί σε λίγο χρόνο την περίληψη που θα έκανε ο χρήστης ,καθιστώντας της σε ένα σημαντικό εργαλείο για μεγάλη έκτασης κείμενα καθώς βοηθά στην μείωση της έκτασης και των λεπτομερειών, διατηρώντας όμως τα βασικά του σημεία και το συνολικό νόημα. (Sonali Vijay Gaikwad et al, 2014)

Απεικόνιση Πληροφορίας (Information Visualization):

Στόχος της απεικόνισης πληροφοριών είναι να οργανώσει μεγάλες πηγές κειμένου σε μία οπτική ιεραρχία. Αυτό δίνει τη δυνατότητα της περιήγησης, η οποία είναι πιο σημαντική σε σχέση από μία απλή αναζήτηση και καθιστά πιο εύκολη τη διαδικασία περιορισμού μεγάλου όγκου κειμένων.

Η τεχνική αυτή, χρησιμοποιείται προκειμένου να δημιουργηθεί μία γραφική απεικόνηση των εγγράφων. Χρησιμοποιεί flags στο κείμενο για την επισήμανση χρήσιμων εγγράφων ή ομάδας εγγράφων ή για να υποδείξει ότι σε κάποιο σημείο το κείμενο έχει πιο συμπαγή μορφή .Η προσέγγιση αυτή είναι πιο φιλική και ελκυστική στον χρήστη που προσπαθεί να αντλήσει πληροφορίες και αποτελεί χρήσιμο εργαλείο όταν απαιτείται ο περιορισμός ενός μεγάλου όγκου εγγράφων και η αναζήτηση σχετικών θεματικών ενοτήτων.(Sonali Vijay Gaikwad et al,2014)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

4

Λειτουργίες της εταιρικής πλατφόρμας και στόχος της ανάλυσης των δεδομένων της

4.1 Πλατφόρμες Customer Service

Οι πλατφόρμες για την εξυπηρέτηση πελατών χρησιμοποιούνται όλο και περισσότερο από τις ασφαλιστικές εταιρείες στα πλαίσια του ψηφιακού μετασχηματισμού. Οι πλατφόρμες αυτές συγκεντρώνουν μεγάλο όγκο μη δομημένων δεδομένων στα οποία μέσω text mining μπορούν να πραγματοποιηθούν ποικίλες ενέργειες:

- ✓ Κατηγοριοποίηση αιτημάτων πελατών και δημιουργία μοντέλου για πρόβλεψη των επόμενων.
- ✓ Φιλτράρισμα των κατηγοριών και θεμάτων που έχουν μεγαλύτερη συχνότητα.
- ✓ Ταξινόμηση των αναφορών
- ✓ Παρακολούθηση του συναισθήματος των πελατών σε πραγματικό χρόνο,η οποία μαζί με την κατηγοριοποίηση θεμάτων, δημιουργούν τις συνθήκες για παρακολούθηση αν οι γενικές μετρήσεις ικανοποίησης επηρεάζονται σημαντικά από αλλαγές του περιβάλλοντος (οικονομικές ,πολιτικές ,τεχνολογικές κ.τ.λ) .
- ✓ Εφαρμογή τεχνητής νοημοσύνης ώστε σε πραγματικό χρόνο να υπάρχει δυνατότητα ανάπτυξης συνομιλίας για άμεση δράση απέναντι στον αρνητικό πελάτη.
- ✓ Δημιουργία ενεργειών παρακολούθησης για τα θέματα που δεν μπορούν να επιλυθούν άμεσα. Η ανάλυση του θέματος μπορεί να συσχετισθεί με διαφορετικά αυτοματοποιημένα e-mails τα οποία θα αποστέλλονται άμεσα προς ικανοποίηση του πελάτη.
- ✓ Συγκεντρώνοντας τα δεδομένα όπως τα δημογραφικά στοιχεία και το θέμα για το οποίο επικοινώνησε ο κάθε πελάτης και αναλύοντας δημιουργείται μια συνολική εικόνα για το πελατολόγιο και τα προβλήματα τα οποία το με αποτέλεσμα οι μακροπρόθεσμες επιχειρηματικές αποφάσεις και το χτίσιμο μιας επιχειρηματικής στρατηγικής να είναι ευκολότερα και ασφαλέστερα ,με μικρότερο το στοιχείο της αβεβαιότητας.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

4.2 Περιγραφή της εταιρικής πλατφόρμας

Τα δεδομένα τα οποία θα αναλύσουμε στο επόμενο κεφάλαιο με τεχνικές text mining προέρχονται από customer service πλατφόρμα ασφαλιστικής εταιρείας η οποία χρησιμοποιείται από ασφαλισμένους και administrator χρήστες της εταιρείας .Τα στοιχεία των ασφαλισμένων (ονοματεπώνυμο ,e-mail κ.τ.λ) ενημερώνονται από το σύστημα καταχώρησης της ασφαλιστικής. Η πρόσβαση στην πλατφόρμα για τους ασφαλισμένους ξεκινά πραγματοποιώντας την εγγραφή τους και στην συνέχεια έχουν εικόνα για τα συμβόλαια που διαθέτουν ,τις παροχές τους και πρόσβαση σε έντυπο υλικό όπως ο πίνακας παροχών του συμβολαίου, φορολογικές αποδείξεις σχετικές με τις αποζημιώσεις τους κ.τ.λ. .Επίσης μέσω αυτής την πλατφόρμας ανεβαίνουν σε μορφή pdf ή jpeg τα δικαιολογητικά τα οποία είναι απαραίτητα για να προχωρήσει μια αποζημίωση συμπληρώνοντας τον αριθμό συμβολαίου και το είδος της πάθησης και φτάνουν στον διαχειριστή ώστε να συνεχίσει manual την διαδικασία για να την εγκρίνει ή να την απορρίψει βάση των όρων του κάθε συμβολαίου. Η συγκεκριμένη διαδικασία μέσω της πλατφόρμας διαρκεί το μέγιστο πέντε ημέρες ,ενώ οι αποζημιώσεις που παραλαμβάνονται σε hard copy χρειάζονται είκοσι ημέρες.

Στην ίδια πλατφόρμα υπάρχει και πεδίο στο οποίο ο ασφαλισμένος μπορεί να στείλει αίτημα στην ασφαλιστική. Μερικά ενδεικτικά αιτήματα είναι αλλαγή/διόρθωση προσωπικών στοιχείων ,e-mail,ερωτήσεις σχετικά με εξετάσεις ή αποζημιώσεις, αξία εξαγοράς συνταξιοδοτικού συμβολαίου του.

Ο administrator χρήστης έχει την δυνατότητα να απαντά σε αυτά τα αιτήματα είτε ότι ολοκληρώθηκαν είτε για να ζητήσει επιπλέον πληροφορίες αν χρειάζεται .Επίσης μπορεί να ανεβάσει και να κατεβάσει έντυπο υλικό ,να ξεμπλοκάρει χρήστες της πλατφόρμας οι οποίοι έχουν κλειδωθεί λόγω εισαγωγής λανθασμένου κωδικού και να τρέξει reports που αφορούν την τα ποσοστά χρήσης της πλατφόρμας.

Σχετικά με τα αιτήματα των ασφαλισμένων λόγω ότι αφορούν κατά ένα μεγάλο μέρος αλλαγές προσωπικών στοιχείων, υπήρξε η πρόταση να δοθεί η δυνατότητα στον ασφαλισμένο να κάνει μόνος του τις αλλαγές αλλά απορρίφθηκε λόγω ρίσκου της ταυτοποίησης .Δεδομένου ότι τα αιτήματα τα οποία λαμβάνονται στην πλατφόρμα απαντώνται από συγκεκριμένο υπάλληλο παράλληλα με άλλες εργασίες ,το αποτέλεσμα είναι να υπάρχουν αρκετά παράπονα σχετικά με τους χρόνους ανταπόκρισης.

Προκειμένου να υπάρχει ένδειξη ότι το αίτημα του κάθε ασφαλισμένου έχει παραλειφθεί και επεξεργάζεται, τα αιτήματα των πελάτων παραμετροποιήθηκαν ώστε να επεξεργαστούν με τεχνικές text mining και να κατηγοριοποιηθούν ώστε να αποστέλλεται το ανάλογο μήνυμα στον ασφαλισμένο.

Το πρόβλημα που αντιμετωπίσαμε με τα δεδομένα είναι ότι είναι σε ελληνική γλώσσα ,ενώ η R που χρησιμοποιήθηκε ως γλώσσα για την επεξεργασία των δεδομένων και την δημιουργία του μοντέλου περιέχει βιβλιοθήκες σε αγγλικά ,οπότε προχωρήσαμε σε μια αρκετά χρονοβόρα μετάφραση αυτών των αιτημάτων.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

5

Προ-επεξεργασία δεδομένων, εφαρμογή και αξιολόγηση του μοντέλου ταξινόμησης

Τα αιτήματα της πλατφόρμας της πλατφόρμας, όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο ήταν σε ελληνική γλώσσα, οπότε και μεταφράστηκαν ώστε να επεξεργαστούν.

	A
1	Request
2	Group retirement
3	Change mail address
4	Certificate of taxation for retirement and life insurance
5	REGULAR UPDATE OF PRIVATE 7119678
6	change of personal email
7	Change of management
8	Allagi Kartas Xreosis
9	Allagi Kartas Xreosis 2
10	request to change home address
11	change of address
12	Visit-Diagnosis 21/5/2018
13	Magnetic 1/6/2018 Athens Bioclinic
14	PREMIUM
15	Dermatologist Visit
16	test by IT
17	CHANGE OF HOUSEHOLD ADDRESS AND CORRESPONDENCE
18	CHANGE OF HOUSEHOLD ADDRESS AND CORRESPONDENCE
19	CHANGE OF CORRESPONDENCE-COMMUNICATION INFORMATION
20	Pension
21	SUBMITTED REimbursement RELEASE
22	CHANGE OF DATA

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Για την επεξεργασία των δεδομένων χρησιμοποιήθηκε η γλώσσα προγραμματισμού R και το περιβάλλον RStudio και τα βήματα μου ακολούθησαν ήταν τα εξής:

Εγκατάσταση πακέτων της R :

Εγκαταστάθηκαν τα παρακάτω πακέτα:

```
install.packages("ggplot2")
install.packages("SnowballC")
install.packages("tm")
install.packages("caTools")
install.packages("tidyverse")
install.packages("e1071")
```

```
install.packages("tidytext")
install.packages("wordcloud")
install.packages("topicmodels")
install.packages("qdap")
install.packages("syuzhet")
install.packages("kernlab")
```

```
requests<- read.csv("requests.csv", stringsAsFactors = FALSE)
```

Διαχωρισμός του dataset σε train και test dataset:

Ορίσθηκε "σπάσιμο" σε 70%-30%,δηλαδή από το αρχικό αρχείο επιλέγονται τυχαία το 70% των αιτημάτων και τοποθετούνται στο train dataset και το 30% στο test dataset.Ο διαχωρισμός αυτός γίνεται ώστε το μοντέλο το οποίο θα χρησιμοποιήσουμε παρακάτω για την πρόβλεψη του είδους των αιτημάτων μας να εκπαιδευθεί στο train dataset και στην συνέχεια να εφαρμοστεί στο test dataset και να ελεγχθεί η απόδοση και ακρίβεια του. Το μοντέλο δεν έχει επεξεργαστεί ποτέ τα δεδομένα του test, οπότε η απόδοση που προκύπτει είναι ένας καλός οδηγός για το τι θα προβλέψει το μοντέλο όταν θα εφαρμοστεί σε νέα δεδομένα. Το ποσοστό που επιλέχθηκε είναι αρκετά καλό, δεδομένου ότι όσο περισσότερα δεδομένα έχουμε στο train set τόσο καλύτερο είναι το μοντέλο, ενώ παράλληλα υπάρχει ένας ικανός όγκος δεδομένων στο test set, ώστε η εκτίμηση να είναι ακριβέστερη.

```
split <- sample.split(requests$Request, SplitRatio = 0.7)
trainRequests <- subset(requests, split == TRUE)
testRequests <- subset(requests, split == FALSE)
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Παραμετροποίηση των δεδομένων του train set:

Στο αρχείο xls του train set προστέθηκε μια ακόμα στήλη με το όνομα Target, στην οποία υπάρχουν οι αριθμοί 0,1,2. Ο αριθμός 1 αντιστοιχεί στις εκφράσεις που περιέχουν την λέξη change και αντιστοιχούν σε αλλαγή προσωπικών πληροφοριών. Ο αριθμός 2 αντιστοιχεί στις εκφράσεις οι οποίες περιέχουν την λέξη reimbursement (αποζημίωση) και αντιστοιχούν σε αιτήματα σχετικά με αποζημιώσεις. Ο αριθμός 0 αντιστοιχεί στα υπόλοιπα.

```
trainRequests<-read.csv("trainRequests.csv")
str(trainRequests)
```

	Request	Target
1	Group retirement	0
2	Change mail address	1
3	Certificate of taxation for retirement and life insurance	0
4	REGULAR UPDATE OF PRIVATE 7119678	0
5	change of personal email	1
6	Change of management	1
7	Allagi Kartas Xreosis	0
8	Allagi Kartas Xreosis 2	0
9	request to change home address	1
10	Visit-Diagnosis 21/5/2018	0

Προ-επεξεργασία Δεδομένων:

Ξεκινάμε με την δημιουργία του corpus.Το corpus αποτελεί μια συλλογή εγγράφων, κάθε αίτημα αντιστοιχίζεται με ένα έγγραφο.

```
trainRequests$Target <- as.factor(trainRequests$Target)

str(trainRequests)

trainCorpus<- Corpus(VectorSource(trainRequests$Request))
trainCorpus
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Στην συνέχεια τα έγγραφα μετατρέπονται σε μικρούς χαρακτήρες, αφαιρούνται τα σημεία στίξης και οι ειδικοί χαρακτήρες και αφαιρούμε τις stopwords. Stopwords θεωρούνται οι λέξεις όπως “i”, “me”,”my”,”you”. Επίσης αφαιρέσαμε τυχόν επιπλέον κενά μεταξύ των λέξεων που μπορεί να υπάρχουν λόγω την πληκτρολόγησης του χρήστη.

```
#Converting to Lowercase
trainCorpus<-tm_map(trainCorpus,tolower)
inspect(trainCorpus[[5]])

#Removing Punctuation
trainCorpus<-tm_map(trainCorpus,removePunctuation)

#Removing Stopwords
trainCorpus<-tm_map(trainCorpus,removePunctuation)

# Remove Extra Whitespaces
trainCorpus <- tm_map(trainCorpus,stripWhitespace)
trainCorpus<- tm_map(trainCorpus, removeWords, c(stopwords("english")))
```

Επιπλέον πραγματοποιήθηκε το stemming των λέξεων των εγγράφων. Όπως αναφέραμε και στο κεφάλαιο 3, η διαδικασία του stemming έχει ως σκοπό την μείωση των λέξεων στις ρίζες τους και γίνεται με αφαίρεση των επιθεμάτων και των προθεμάτων.

```
#Stemming
trainCorpus <- tm_map(trainCorpus, stemDocument)
```

Μετά από αυτήν την επεξεργασία μπορούμε να εξαγάγουμε τις συχνότητες λέξεων που θα χρησιμοποιηθούν για την πρόβλεψη.

Το πακέτο tm παρέχει μια λειτουργία που ονομάζεται DocumentTermMatrix που δημιουργεί έναν πίνακα όπου οι σειρές αντιστοιχούν σε έγγραφα, στην περίπτωση μας τα αιτήματα των πελατών και οι στήλες αντιστοιχούν σε λέξεις των αιτημάτων.

```
#Bag of Words
frequencies <- DocumentTermMatrix(testCorpus)
frequencies
```

Το αποτέλεσμα της εντολής frequencies είναι το παρακάτω :

```
<<DocumentTermMatrix (documents: 7626, terms: 798)>>
Non-/sparse entries: 21433/6064115
Sparsity           : 100%
```

Μπορούμε να δούμε ότι υπάρχουν 798 λέξεις στο matrix και 7626 έγγραφα ή αιτήματα μετά την προ-επεξεργασία. Με την εντολή inspect βλέπουμε παρακάτω ένα μέρος της μορφή του matrix πίνακα:

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
      Terms
Docs   goumopoulou marinas director jobs 7148495 10494 38059p
  3001           0       0        0    0       0     0       0
  3002           0       0        0    0       0     0       0
  3003           0       0        0    0       0     0       0
  3004           0       0        0    0       0     0       0
  3005           0       0        0    0       0     0       0
  3006           0       0        0    0       0     0       0
  3007           0       0        0    0       0     0       0
  3008           0       0        0    0       0     0       0
  3009           0       0        0    0       0     0       0
  3010           0       0        0    0       0     0       0
```

Παρατηρούμε ότι υπάρχουν πολλά μηδενικά και αυτά τα δεδομένα ονομάζονται αραιά (sparse data).Μπορούμε να δούμε τους όρους οι οποίο εμφανίζονται το λιγότερο 20 φορές μέσα στα αιτήματα των πελάτών και :

```
> findFreqTerms(frequencies, lowfreq = 20)
 [1] "address"      "change"       "pension"      "data"         "email"
 [6] "payment"      "premium"      "contract"     "information"  "request"
[11] "account"      "mobile"       "phone"        "communication" "personal"
[16] "policy"       "tax"          "home"         "medical"      "visit"
[21] "details"      "number"       "insurance"    "name"         "update"
[26] "evidence"     "beneficiary"  "reimbursement" "purchase"    "detail"
[31] "info"         "mail"         "contracts"    "certificate"  "copy"
[36] "credit"       "payments"     "card"         "premiums"     "break"
[41] "beneficiaries" "cancellation" "proof"       "correction"   "group"
```

Συνεπώς από τις από τις 798 λέξεις στον matrix πίνακα, μόνο 45 λέξεις εμφανίζονται τουλάχιστον είκοσι φορές στα αιτήματας.Αυτό σημαίνει ότι έχουμε πιθανώς πολλούς όρους οι οποίοι δεν θα είναι ιδιαίτερα χρήσιμοι για το μοντέλο πρόβλεψής.

Ο αριθμός των όρων είναι σημαντικός για δύο λόγους.Ο ένας είναι υπολογιστικός, δηλαδή περισσότεροι όροι σημαίνουν περισσότερες ανεξάρτητες μεταβλητές, οι οποίες συνήθως σημαίνει ότι χρειάζεται περισσότερος χρόνος για την κατασκευή του μοντέλου. Επιπλέον, το ποσοστό των ανεξάρτητων μεταβλητών στο σύνολο των δεδομένω θα επηρεάσει την αποδοτικότητα του μοντέλου. Για αυτό τον λόγο αφαιρούμε κάποιους όρους που δεν εμφανίζονται πολύ συχνά:

```
testSparse <- removeSparseTerms(frequencies, 0.995)
```

Στην εντολή removeSparseTerms βλέπουμε ότι υπάρχει και ένα ποσοστό, το ποσοστό αυτό ονομάζεται “sparsity threshold” και λειτουργεί ως εξής:

- Το ποσοστό 0,98 σημαίνει ότι θα διατηρήσουμε τους όρους που εμφανίζονται στο 2% ή και περισσότερο των αιτημάτων.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Το ποσοστό 0,99 σημαίνει ότι θα διατηρήσουμε τους όρους που εμφανίζονται στο 1% ή και περισσότερο των αιτημάτων.
- Το ποσοστό 0,995 σημαίνει ότι θα διατηρήσουμε τους όρους που εμφανίζονται στο 0,5% ή και περισσότερο των αιτημάτων.

Χρησιμοποιώντας το 0,5% ως κατώτατο όριο, διατηρούμε τους όρους που εμφανίζονται σε 30 ή παραπάνω αιτήματα πελατών.

Ο πίνακας που βλέπουμε παρακάτω μετά την εφαρμογή της εντολής removeSparseTerms με sparsity threshold 0,995

```
> testSparse
<<DocumentTermMatrix (documents: 7626, terms: 22)>>
Non-/sparse entries: 19018/148754
Sparsity           : 89%
Maximal term length: 13
Weighting          : term frequency (tf)
```

Παρατηρούμε ότι από τους 798 όρους παρέμειναν οι 22 με τους οποίους θα δημιουργήσουμε και το data frame και θα προσθέσουμε την εξαρτημένη μεταβλητή Target την οποία θέλουμε να προβλέψουμε.

```
#Adding the Dependent Variable
trainSparse.df <- as.data.frame(as.matrix(trainSparse))
trainSparse.df$Request<-trainRequests$Request
trainSparse.df$Target<-trainRequests$Target
```

Παρόμοια διαδικασία ακολουθείται και για το test set εκτός από την προσθήκη της εξαρτημένης μεταβλητής.

Εφαρμογή μοντέλου ταξινόμησης :

Εφαρμόσαμε το μοντέλο ταξινόμηση με την εξαρτημένη μεταβλητή Target:

```
df.model<-ksvm(Target~.,data=df.trainSparse,kernel="rbfdot")
summary(df.model)
```

Το μοντέλο το οποίο χρησιμοποιήσαμε είναι το ksvm model.Ο αλγόριθμος ksvm ανήκει στους SVM αλγόριθμους ταξινόμησης, τους οποίους αναφέραμε στο κεφάλαιο 3, και εφαρμόζεται το kernel method rbf ώστε να πραγματοποιηθεί η μη γραμμική ταξινόμηση.

Το μοντέλο εφαρμόζεται και στο test dataset ώστε να αξιολογηθεί η ακρίβεια του και καταλήγουμε στο

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Submission dataset,το οποίο περιέχει τα δεδομένα του test set και την κλάση στην οποία ταξινομήθηκαν από το μοντέλο.

```
df.pred<-predict(df.model, df.testSparse)

con.matrix<-confusionMatrix(df.pred, df.testSparse$Target)
print(con.matrix)

dim(df.testSparse)

df.testSparse$Target <- df.pred
submission <- df.testSparse[,c("Request","Target")]
```

Η ακρίβεια του μοντέλου είναι σε μεγάλο ποσοστό 99,5% και από το confusion matrix πίνακα βλέπουμε ότι 7406 αιτήματα ταξινομήθηκαν σωστά στην κλάση "0", 14924 ταξινομήθηκαν σωστά στην κλάση "1" και 366 ταξινομήθηκαν σωστά στην κλάση "3". Συνεπώς το μοντέλο μπορεί να προβλέψει με ικανοποιητική ακρίβεια την κλάση των αιτημάτων των πελατών.

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1     2
         0  7406    44     3
         1    64 14924     0
         2     0     1   366

Overall Statistics

               Accuracy : 0.9951
                 95% CI : (0.9941, 0.996)
    No Information Rate : 0.6563
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.9894

 Mcnemar's Test P-Value : 0.05255

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Sensitivity            0.9914   0.9970  0.99187
Specificity            0.9969   0.9918  0.99996
Pos Pred Value         0.9937   0.9957  0.99728
Neg Pred Value         0.9958   0.9942  0.99987
Prevalence             0.3275   0.6563  0.01618
Detection Rate         0.3247   0.6543  0.01605
Detection Prevalence   0.3268   0.6571  0.01609
Balanced Accuracy      0.9942   0.9944  0.99591
```

Εφαρμόστηκε και το μοντέλο NaiveBayes στα δεδομένα μας ώστε να συγκριθεί με το ksv μοντέλο και τα αποτελέσματα είναι τα παρακάτω:

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1     2
         0  3631   332    13
         1    64 14627     0
         2  3775    10   356

Overall Statistics

               Accuracy : 0.8161
                 95% CI : (0.811, 0.8211)
    No Information Rate : 0.6563
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6445

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Sensitivity            0.4861   0.9772  0.96477
Specificity            0.9775   0.9918  0.83132
Pos Pred Value         0.9132   0.9956  0.08597
Neg Pred Value         0.7961   0.9579  0.99930
Prevalence             0.3275   0.6563  0.01618
Detection Rate         0.1592   0.6413  0.01561
Detection Prevalence   0.1743   0.6441  0.18156
Balanced Accuracy      0.7318   0.9845  0.89805
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Παρατηρούμε ότι εκτός από την χαμηλότερη τιμή του accuracy, στο confusion matrix φαίνεται να ταξινομούνται λιγότερα σωστά στην κλάση “0” και στην κλάση “2”.

Η παραμετροποίηση η οποία αναφέρθηκε παραπάνω μπορεί να διευρυνθεί ώστε οι κλάσεις της ταξινόμησης να αυξηθούν μελλοντικά και στην συνέχεια οποιοδήποτε αίτημα ασφαλισμένου θα καταγράφεται στην πλατφόρμα θα ταξινομείται απευθείας και θα συνδέεται με το ανάλογο mail που θα έχει ορισθεί για αυτό το αίτημα μέχρι και την επίλυση. Η αμεσότητα αναμένεται ότι θα μειώσει τα παράπονα των πελατών ως προς την καθυστέρηση της επικοινωνίας και θα αναδείξει την πλατφόρμα η οποία δεν έχει κερδίσει ακόμα την εμπιστοσύνη όλων των ασφαλισμένων. Η ταξινόμηση που προτείνεται και το αποτέλεσμα που η εταιρεία μπορεί να επιφέρει με αυτήν δείχνει πως το text mining μπορεί όντως να προσφέρει στο κομμάτι του customer engagement βελτιώνοντας ένα customer service εργαλείο, το οποίο εφαρμόστηκε στα πλαίσια του ψηφιακού μετασχηματισμού της εταιρείας.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

6

Τεχνικές λεπτομέρειες

6.1 Λεπτομέρειες υλοποίησης

Στο κεφάλαιο αυτό παρατίθεται η σύνταξη του κώδικά στο περιβάλλον του R studio

```
install.packages("ggplot2")
install.packages("SnowballC")
install.packages("tm")
install.packages("caTools")
install.packages("tidyverse")
install.packages("e1071")
```

```
install.packages("tidytext")
install.packages("wordcloud")
install.packages("topicmodels")
install.packages("qdap")
install.packages("syuzhet")
install.packages("kernlab")
library("e1071")
```

```
library(ggplot2) # install.packages("ggplot2") if not installed
library(caTools)
library(tm) # install.packages("tm") if not installed
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
library(SnowballC) # install.packages("SnowballC") if not installed
library(tidyverse)
library(caret)
# text mining libraries
library(tidytext)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(topicmodels)
library(data.table)
library(stringi)
library(qdap)
library(dplyr)
library(rJava)
library(syuzhet)
library(kernlab)



requests<- read.csv("requests.csv", stringsAsFactors = FALSE)
str(requests)
set.seed(123)
split <- sample.split(requests$Request, SplitRatio = 0.7)
trainRequests <- subset(requests, split == TRUE)
testRequests <- subset(requests, split == FALSE)
str(trainRequests)
str(testRequests)
#write.csv(trainRequests, file = "trainRequests.csv",row.names=FALSE)
#write.csv(testRequests, file = "testRequests.csv",row.names=FALSE)

trainRequests<-read.csv("trainRequests.csv")
str(trainRequests)
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
trainRequests$Target <- as.factor(trainRequests$Target)

str(trainRequests)

trainCorpus<- Corpus(VectorSource(trainRequests$Request))
trainCorpus


inspect(trainCorpus[[1]])
#Converting to Lowercase
trainCorpus<-tm_map(trainCorpus,tolower)
inspect(trainCorpus[[5]])

#Removing Punctuation
trainCorpus<-tm_map(trainCorpus,removePunctuation)

#Removing Stopwords
trainCorpus<-tm_map(trainCorpus,removePunctuation)

# Remove Extra Whitespaces
trainCorpus <- tm_map(trainCorpus,stripWhitespace)
trainCorpus<- tm_map(trainCorpus, removeWords, c(stopwords("english")))

#Stemming
trainCorpus <- tm_map(trainCorpus, stemDocument)

#Bag of Words
frequencies <- DocumentTermMatrix(trainCorpus)
frequencies
inspect(frequencies[1000:1005, 621:627])
#We can look at what the most popular terms are with the function findFreqTerms
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
findFreqTerms(frequencies, lowfreq = 20)

trainSparse <- removeSparseTerms(frequencies, 0.995)
trainSparse


#the same process for test dataset

testRequests<-read.csv("testRequests.csv")
str(testRequests)


testCorpus<- Corpus(VectorSource(testRequests$Request))
testCorpus


inspect(testCorpus[[1]])
#Converting to Lowercase
testCorpus<-tm_map(testCorpus,tolower)
inspect(testCorpus[[5]])

#Removing Punctuation
testCorpus<-tm_map(testCorpus,removePunctuation)

#Removing Stopwords
testCorpus<-tm_map(testCorpus,removePunctuation)

# Remove Extra Whitespaces
testCorpus <- tm_map(testCorpus,stripWhitespace)
testCorpus<- tm_map(testCorpus, removeWords, c(stopwords("english")))

#Stemming
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
testCorpus <- tm_map(testnCorpus, stemDocument)

#Bag of Words
frequencies <- DocumentTermMatrix(testCorpus)
frequencies
inspect(frequencies[3001:3010, 621:627])
#We can look at what the most popular terms are with the function findFreqTerms
findFreqTerms(frequencies, lowfreq = 20)

testSparse <- removeSparseTerms(frequencies, 0.995)
testSparse


#Adding the Dependent Variable
trainSparse.df <- as.data.frame(as.matrix(trainSparse))
trainSparse.df$Request<-trainRequests$Request
trainSparse.df$Target<-trainRequests$Target
str(trainSparse.df)

testSparse.df <- as.data.frame(as.matrix(testSparse))
testSparse.df$Request<-testRequests$Request


df.trainSparse<-trainSparse.df
str(df.trainSparse)
head(df.trainSparse)
df.testSparse<-trainSparse.df

df.model<-ksvm(Target~.,data=df.trainSparse,kernel="rbfdot")
summary(df.model)
df.pred<-predict(df.model, df.testSparse)
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
con.matrix<-confusionMatrix(df.pred, df.testSparse$Target)
print(con.matrix)

dim(df.testSparse)

df.testSparse$Target <- df.pred
submission <- df.testSparse[,c("Request","Target")]
colnames(submission)[colnames(submission)=="Target"] <- "Prediction"
write.csv(submission, "submission.csv",row.names = FALSE)

df.model1<-naiveBayes(Target~.,data=df.trainSparse)
summary(df.model1)
df.pred<-predict(df.model1, df.testSparse)
con.matrix<-confusionMatrix(df.pred, df.testSparse$Target)
print(con.matrix)
```


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

7

Επίλογος

7.1 Σύνοψη και συμπεράσματα

Ο ψηφιακός σχηματισμός είναι το μέσο για την πρόσβαση στα δεδομένα που είναι διαθέσιμα με σκοπό η επεξεργασία τους να αυξήσει το customer engagement, να μειώσει το ρίσκο και να διευκολύνει τις καθημερινές εργασίες, αλλά οι υπάρχουσες διαδικασίες και τα διαθέσιμα εργαλεία των εταιρειών δεν δίνουν την δυνατότητα για την συλλογή και την επεξεργασία τους. Η εγκατάσταση τους επίσης αποδεικνύεται δυσκολότερη στην πράξη, κυρίως λόγω έλλειψης χρηματικών πόρων. Στην διπλωματική εργασία επικεντρωθήκαμε στον ψηφιακό μετασχηματισμό στην ασφαλιστική βιομηχανία και αναπτύξαμε ιδέες εργαλείων που ο μετασχηματισμός θέτει το υπόβαθρο ώστε να εγκατασταθούν σε μια εταιρεία είτε αναπτύσσοντας το κατάλληλο τμήμα εντός αυτής είτε σε συνεργασία με Insurtech εταιρείες. Επίσης αναπτύξαμε πως τα εργαλεία αυτά θα μπορούν να συλλέξουν πληροφορίες και δεδομένα και πως μέσω αυτών θα διευκολύνουν οι λειτουργίες της επιχείρησης, το ανθρώπινο δυναμικό της και θα ικανοποιηθεί το πελατολόγιο της. Με τον όρο εργαλεία αναφερθήκαμε τόσο σε πλατφόρμες, εφαρμογές, φορητές συσκευές, αλλά και σε τεχνικές όπως αυτή της εξόρυξης δεδομένων από κείμενο (text mining). Την εφαρμογή του text mining την παρατηρήσαμε στην πράξη σε δεδομένα κειμένου από αιτήματα πελατών σε πλατφόρμα customer service ασφαλιστικής εταιρείας και πως αυτή η εφαρμογή θα βοηθήσει στην εικόνα της εταιρείας απέναντι στο πελατολόγιο της.

7.2 Μελλοντικές επεκτάσεις

Η πλατφόρμα της ασφαλιστικής εταιρείας από την οποία συλλέχθηκαν τα δεδομένα σχεδιάζει την προσθήκη πεδίου εντός αυτής για σχολιασμό σχετικά με την πλατφόρμα και τις λειτουργίες της, η οποία θα συνοδεύεται από κλίμακα βαθμολόγησης. Συνεπώς ο κάθε ασφαλισμένος ο οποίος την χρησιμοποιεί θα μπορεί να εκφέρει την άποψη του για αυτή. Αυτό θα γίνει με σκοπό η πλατφόρμα να γίνει καλύτερη και πιο φιλική προς τον χρήστη, αλλά και για την εταιρεία να μάθει την άποψη των ασφαλισμένων για αυτή. Τα δεδομένα αυτά, όπως καταλαβαίνουμε θα είναι σε μορφή κειμένου και σε αυτά μπορούν να χρησιμοποιηθούν οι τεχνικές text mining με σκοπό την ανάλυση συναισθήματος. Η ύπαρξη

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

βαθμολογιών κάνουν τον διαχωρισμό τους ακόμα ευκολότερο αλλά ο όγκος επεξεργασίας των εγγράφων που θα προκύπτουν από το κείμενο θα είναι αρκετά μεγαλύτερος. Συνεπώς θα ήταν αρκετά ενδιαφέρον να δούμε την απόδοση του μοντέλου στα νέα δεδομένα ή την εφαρμογή νέου μοντέλου.

Ενδιαφέρον επίσης θα ήταν να δούμε τον ρόλο της blockchain τεχνολογίας στον ψηφιακό μετασχηματισμό των ασφαλιστικών εταιρειών. Τα smart contracts τα οποία χρησιμοποιούν την τεχνολογία blockchain αυτό-εκτελούνται σύμφωνα με προγραμματισμένους κανόνες και προσφέρουν εύκολη πρόσβαση σε δεδομένα υγείας, οικονομικών, δημογραφικών δεδομένων κ.λ.π για τον ασφαλιστικό κλάδο, με αποτέλεσμα αποδοτικότερες διαδικασίες risk management και underwriting. Άλλα πεδία εφαρμογής είναι αυτά των πληρωμών και της ανίχνευσης απάτης, αλλά έμμεσα επηρεάζει και την εμπειρία του πελάτη σε σχέση με την εταιρεία. Τα πλεονεκτήματα είναι η αύξηση της ταχύτητας των πληρωμών, η μείωση των hedging fees που οι εταιρείες καταβάλλουν για να προστατευθούν από τις συναλλαγματικές διακυμάνσεις στις διεθνείς συναλλαγές, η αποφυγή του ξεπλύματος χρήματος και τα οι πελάτες νιώθουν μεγαλύτερη ασφάλεια εφόσον τα προσωπικά του δεδομένα αποθηκεύονται σε ένα ασφαλές, αποκεντρωμένο σύστημα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

8

Βιβλιογραφία

[1] Hugh Terry, Will Pritchett, Yoshi Makita, Paul Brenchley, Ilanit Adesman, David Milligan , Bo Huang, Gary Plotkin Insuretech: 10 Trends for 2019, 2019.

[2] Research2Guidance, How mobile micro-insurance will impact the health insurance industry — Interview with Jose Martin, Global Health Manager at BIMA, July 2018, available at https://medium.com/@r2guidance.

[3] i-scoop Team, Insurance: digital transformation, digitization and optimization in a disrupted landscape,2019, available at https://www.i-scoop.eu/insurance-industry.

[4] i-scoop Team, Changing customer expectations regarding insurance providers: the Gen Y effect, 2019, available at https://www.i-scoop.eu/insurance-industry.

[5] Alex McEachern, What is Customer Engagement, and Why is it Important?, July 2019, available at https://blog.smile.io .

[6] Mike Adler, Gary Plotkin, Praatek Saxena, Enabling the future of Underwriting,2017, available at https://www.the-digital-insurer.com.

[7] Sam Draper, How Wearable Devices Can Benefit the Life Insurance Industry, July 2018, available at https://www.wearable-technologies.com.

[8] Tim Sandle, Digital Transformation of Insurance to lower fraud risk, January 2018, available at https://www.blumbergcapital.com.

[9] BioCatch Team, Digital Transformation in the Insurance Sector: How Behavioral Biometrics is Paving the Way to Less Fraud, Less Friction and More Functionality, September 2017.

[10] Noel Bambrick What industries are next to be disrupted by NLP and Text Analysis?, September 2016, available at http://blog.aylien.com.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

[11] Michael Elliott ,Judith Vaughan, Why Analyzing Unstructured Data Benefits Every Claims Manager, December 2016, available at https://www.claimsjournal.com.

[12] Ruchi Verma, Sathyan Ramakrishna Mani, Using Analytics for Insurance Fraud Detection, 2019 available at https://www.the-digital-insurer.com.

[13] James Ruotolo, Tapping the Benefits of Text Mining for Fraud, 2013, available at http://insurancetech.com.

[14] Jose C. Gonzalez , The Role of Text Mining in the Insurance Industry, October 2014, available at https://www.meaningcloud.com/blog.

[15] Douglas Karr, Digital Transformation is a Leadership Issue, Not a Technology Issue, available at https://martech.zone.

[16] Raghav Bharadwaj, Document Search and Data Mining in Insurance – Claims Processing, Fraud Detection, and Data Management ,February 2019, available at https://emerj.com/ai-sector-overviews/.

[17] Mouhamadou Sow, Solomon Aborbie, Impact of Leadership on Digital Transformation, August 2018 available at https://www.researchgate.net/.

[18] Kendra Cherry, How a Transactional Leadership Style Works, October 2019 available at https://www.verywellmind.com.

[19] Michelle Cyca, What digital transformation means for your company, November 2018, available at https://slackhq.com.

[20] Oliver Müller, Stefan Debortoli, Iris Junglas, Jan vom Brocke, Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data, December 2016.

[21] Diego Zappa, Mattia Borrelli, Gian Paolo Clemente, Nino Savelli, Text Mining In Insurance: From Unstructured Data To Meaning, 2014.

[22] Jeffrey L. Solk, Text Data Mining: Theory and Methods,2008.

[23] Sonali Vijay Gaikwad, Pramod Patil, Text Mining Methods and Techniques, January 2014.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

[24] S.Brindha, Dr. K.Prabha, Dr. S.Sukumaran, The Comparison of term based methods using text mining, September 2016.

[25] Mehdi Allahyari, Saied Safaei, Elizabeth D.Trippe, Juan B.Gutierrez, A Brief Survey of TextMining: Classification, Clustering and ExtractionTechniques, July 2017.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ