

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία

*Σύστημα υποστήριξης απόφασης για την ανίχνευση
απάτης σε πιστωτικές κάρτες*

Αγάθος Λεωνίδας

Π2015096

p15agat@ionio.gr

Επιβλέπων: Έξαρχος Θέμης

12 Νοεμβρίου 2021

Επιβλέπων

Έξαρχος Θέμης, Επίκουρος καθηγητής
Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Έξαρχος Θεμιστοκλής, Επίκουρος καθηγητής
Ιόνιο Πανεπιστήμιο

Κερμανίδου Κάτια-Λήδα, Αναπληρώτρια Καθηγήτρια
Ιόνιο Πανεπιστήμιο

Κουρουθανάσης Παναγιώτης, Επίκουρος καθηγητής
Ιόνιο Πανεπιστήμιο

Περίληψη

Αυτή η πτυχιακή εργασία, έχει ως στόχο να παρουσιάσει και να αναλύσει από διαφορετικές όψεις, δύο έννοιες οι οποίες έχουν γίνει ευρέως διαδεδομένες. Η μία έννοια είναι η μηχανική μάθηση ή machine learning όπως είναι παγκοσμίως γνωστό, και η άλλη είναι οι απάτες πιστωτικών καρτών. Οι δύο αυτές έννοιες, αποτελούν πλέον βασικές έννοιες στην καθημερινότητα μας ακόμα και αν δεν το κατανοούμε. Αρχικά, συναντάμε καθημερινά τη μηχανική μάθηση η οποία χρησιμοποιείται σε πολλούς τομείς όπως στις υπολογιστικές Επιστήμες, Ιατρική, Οικονομία με μεγάλη επιτυχία. Από την άλλη, η χρήση πιστωτικών και χρεωστικών καρτών αποτελεί άλλο ένα σημαντικό κομμάτι της καθημερινότητας μας, καθώς μέσω αυτών, πραγματοποιούνται οι περισσότερες καθημερινές συναλλαγές. Γι' αυτό το λόγο, οι απάτες σε τέτοιου είδους κάρτες έχουν κάνει την εμφάνιση τους εδώ και πολλά χρόνια, αποτελώντας τον μεγάλο "πονοκέφαλο" τόσο των μεγάλων τραπεζικών οργανισμών, όσο και των επιχειρήσεων αλλά και των καταναλωτών, όπου συνήθως είναι και τα θύματα τέτοιου είδους απατών. Με την εξέλιξη της τεχνολογίας λοιπόν τα τελευταία χρόνια, γίνεται μια μεγάλη προσπάθεια εντοπισμού και αποτροπής τέτοιου είδους απατών καθώς κάθε χρόνο χάνονται τεράστια χρηματικά ποσά λόγω αυτών. Η μηχανική μάθηση, λόγω των δυνατοτήτων που παρέχει, αποτελεί ένα από τα εργαλεία, το οποίο χρησιμοποιείται πλέον ευρέως προκειμένου να αντιμετωπιστούν τέτοιου είδους απάτες. Προς το παρόν, η μηχανική μάθηση έχει αποφέρει πολύ αξιόλογα αποτελέσματα και έχει προσφέρει πολλά στην αντιμετώπιση τέτοιων καταστάσεων, αφού ήδη χρησιμοποιείται εδώ και κάποια χρόνια. Παράλληλα, έχουν γίνει αρκετές έρευνες και έχουν συγγραφεί επίσης πολλές αναφορές για τη χρήση μηχανικής μάθησης προκειμένου να αντιμετωπιστεί το συγκεκριμένο πρόβλημα, κάποιες από τις οποίες θα αναφερθούν και θα αναλυθούν παρακάτω. Η πλειοψηφία των ερευνών, έχουν διαφορετικό τρόπο προσέγγισης μεταξύ τους, χρησιμοποιώντας διαφορετικές τεχνικές, όλες τους όμως καταλήγουν στο συμπέρασμα πως η μηχανική μάθηση έχει πολλές δυνατότητες οι οποίες μένει να αναδειχθούν και να χρησιμοποιηθούν στο μέλλον προκειμένου να αντιμετωπιστεί το πρόβλημα των απατών όσο πιο αποτελεσματικά γίνεται. Η συρραφή της παρακάτω πτυχιακής λοιπόν, έχει ως στόχο την ανάδειξη της σημαντικότητας του προβλήματος, την παρουσίαση διαφόρων στοιχείων γύρω από το πρόβλημα καθώς και την παρουσίαση των διαφορετικών τεχνικών που ακολούθησαν οι ερευνητές που ασχολήθηκαν με αυτό το πρόβλημα. Τέλος, θα παρουσιασθεί μία πρακτική προσπάθεια προκειμένου να δούμε την αποτελεσματικότητα διαφορετικών τεχνικών στην αντιμετώπιση του συγκεκριμένου ζητήματος, χρησιμοποιώντας ρεαλιστικά δεδομένα και στοιχεία συναλλαγών διαφόρων χρηστών.

Πρόλογος και Ευχαριστίες

Η συρραφή αυτής της πτυχιακής εργασίας επιτεύχθηκε με αρκετό κόπο ενώ παράλληλα αφιερώθηκαν πολλές ώρες ελεύθερου χρόνου. Προκειμένου να επιτευχθεί ένα όμορφο αποτέλεσμα, χρειάζεται προσπάθεια και βοήθεια από πολλές πλευρές την οποία και είχα σε μεγάλο βαθμό. Γι' αυτό το λόγο, θα ήθελα να ευχαριστήσω τους δικούς μου ανθρώπους που ήταν κοντά μου, με στήριξαν σε αυτή μου την προσπάθεια και στάθηκαν δίπλα μου στις δυσκολίες που αντιμετώπισα. Παράλληλα, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής κ. Κουρουθανάση και κα. Κερμανίδου όπου με τις γνώσεις και τα εφόδια που μας προσέφεραν όλα αυτά τα χρόνια, βοήθησαν και εκείνοι με τον τρόπο τους να πετύχω αυτό το αποτέλεσμα. Τέλος, θα ήθελα να κάνω μια ξεχωριστή αναφορά και να πω ένα μεγάλο ευχαριστώ, στον επιβλέπον καθηγητή αυτής της προσπάθειας, τον κ. Έξαρχο ο οποίος με βοήθησε πραγματικά τόσο στην σταδιοδρομία μου στο Ιόνιο Πανεπιστήμιο μέσω των μαθημάτων και των διαλέξεων του, όσο και στη συρραφή αυτής της πτυχιακής. Θέλω λοιπόν να τον ευχαριστήσω πραγματικά για την βοήθεια του, την υπομονή του και την κατανόηση του, καθώς χωρίς την βοήθεια του, το αποτέλεσμα σίγουρα δεν θα ήταν το ίδιο.

Κέρκυρα, 12 Νοεμβρίου 2021

Αγάθος Λεωνίδας

Κατάλογος περιεχομένων

Εισαγωγή	6
Κεφάλαιο 1 “Εισαγωγή στις απάτες πιστωτικών καρτών”	7
1.1 Εισαγωγή	7
1.2 Ανίχνευση απάτης σε πιστωτικές κάρτες	7
1.3 Κατηγορίες δραστών	8
1.4 Κατηγορίες απατών στις πληρωμές	9
1.5 Επιπτώσεις απατών στην οικονομία	10
1.6 Σημαντικότητα πρόβλεψης των απατών και μέτρα αντιμετώπισης	13
Κεφάλαιο 2 “Μηχανική μάθηση”	15
2.1 Εισαγωγή στη μηχανική μάθηση (Machine learning)	15
2.2 Χρησιμότητα και λειτουργία της μηχανικής μάθησης	15
2.3 Ιστορία της μηχανικής μάθησης	16
2.4 Εφαρμογές μηχανικής μάθησης στη καθημερινότητα	17
2.5 Μηχανική μάθηση ή απλοί κανόνες τεχνητής νοημοσύνης;	19
2.6 Μηχανική μάθηση και πρόβλεψη απατών σε πιστωτικές κάρτες	21
2.7 Μηχανική μάθηση και ηθική	22
Κεφάλαιο 3 “Βιβλιογραφική ανασκόπηση”	25
3.1 Εισαγωγή και ανάλυση εργασιών	25
Κεφάλαιο 4 “Πειραματική Ανάλυση δεδομένων και γραφήματα”	41
4.1 Περιγραφή και υλοποίηση του πειράματος	41
Κεφάλαιο 5 “Συμπεράσματα ανάλυσης δεδομένων”	57
5.1 Αριθμητικά συμπεράσματα πειράματος	57
5.2 Επίλογος	58
Βιβλιογραφία	60

Εισαγωγή

Στη σύγχρονη κοινωνία στην οποία ζούμε, πολλές από τις συνήθειές μας έχουν αλλάξει, αφού συνεχώς εισάγονται νέοι τρόποι προκειμένου να πραγματοποιούμε διάφορες καθημερινές εργασίες χρησιμοποιώντας την εξέλιξη της τεχνολογίας. Κάτι τέτοιο συμβαίνει και στον κόσμο της οικονομίας, αφού με την εξέλιξη της τεχνολογίας γίνεται προσπάθεια προκειμένου να βρίσκονται συνεχώς νέοι τρόποι για ασφαλέστερες συναλλαγές. Τέτοια παραδείγματα στις μέρες μας αποτελούν η χρήση εφαρμογών στα κινητά τηλέφωνα, οι ανέπαφες συναλλαγές μέσω POS, οι διαδικτυακές συναλλαγές και όποια άλλη συναλλαγή σχετίζεται με την εξέλιξη των τεχνολογικών μέσων τα οποία προσφέρουν μεγαλύτερη ασφάλεια κατά την πραγματοποίηση μιας συναλλαγής. Οι πιστωτικές και χρεωστικές κάρτες, έχουν μπει ριζικά στην καθημερινότητά μας προσφέροντας στον πελάτη κάθε τράπεζας μεγαλύτερη ευκολία και ταχύτητα στις οικονομικές συναλλαγές του. Αυτές οι αλλαγές, έχουν επιφέρει σημαντικά πλεονεκτήματα αλλά και μειονεκτήματα στο ευρύ κοινό, αφού η χρήση τέτοιων τεχνολογιών μπορεί να αποδειχθεί μεγαλύτερος μπελάς απ' ό τι φαίνεται αρχικά. Το πιο σύνηθες φαινόμενο το οποίο θα είναι και το κύριο αντικείμενο μελέτης σε αυτήν την πτυχιακή, είναι το φαινόμενο της απάτης-υποκλοπής σε πιστωτικές-χρεωστικές κάρτες κατά το οποίο υποκλέπτονται σημαντικά στοιχεία της κάρτας του ιδιοκτήτη, με αποτέλεσμα συνήθως την απώλεια χρηματικών ποσών από το λογαριασμό του. Κάτι τέτοιο, αποτελεί μεγάλο πρόβλημα τόσο για τον πελάτη όσο και για την εκάστοτε τράπεζα μιας και οι δύο βγαίνουν ζημιωμένοι από μια τέτοια κατάσταση, ενώ παράλληλα, τρίτες εταιρείες που μπορεί να εμπλέκονται σε μια τέτοια κατάσταση έχουν επίσης σοβαρό πρόβλημα. Γενικότερα, υπάρχουν διάφοροι τρόποι με τους οποίους μπορεί κάποιος να υποκλέψει τέτοιου είδους στοιχεία, ο πιο σύνηθες όμως φαίνεται να είναι οι υποκλοπές μέσω ηλεκτρονικών αγορών που μπορεί να πραγματοποιήσει ο κάτοχος της κάρτας. Τέτοιες περιπτώσεις θα αναλυθούν εκτενέστερα παρακάτω, ενώ παράλληλα θα μελετηθεί και θα αναλυθούν τρόποι αντιμετώπισης για το πρόβλημα το οποίο έχει δημιουργηθεί λόγω της εξέλιξης της τεχνολογίας και παράλληλα των τρόπων υποκλοπής.

Κεφάλαιο 1 “Εισαγωγή στις απάτες πιστωτικών καρτών”

1.1 Εισαγωγή

Το αντικείμενο το οποίο πραγματεύεται αυτή η πτυχιακή εργασία, όπως προαναφέρθηκε, είναι οι απάτες σε πιστωτικές κάρτες αλλά και πως αυτό το φαινόμενο μπορεί να καταπολεμηθεί. Αρχικά, με την έννοια απάτες σε πιστωτικές κάρτες εννοούμε οποιαδήποτε υποκλοπή στοιχείων ή υποκλοπή της ίδιας της κάρτας και χρησιμοποίησή της, χωρίς την άδεια του κατόχου της. Συνήθως, οι υποκλοπές έχουν ως απώτερο σκοπό την εκμετάλλευση των στοιχείων της κάρτας και την χρησιμοποίησή τους προκειμένου ο δράστης να επωφεληθεί οικονομικά από αυτό, συγκεντρώνοντας άλλες φορές μικρότερα και άλλες φορές μεγαλύτερα, χρηματικά ποσά. Τέτοιου είδους απάτες μπορούν να πραγματοποιηθούν με διάφορους τρόπους, τόσο φυσικούς (κλοπή κάρτας) όσο και ηλεκτρονικούς, όπου στις μέρες μας είναι το πιο σύνηθες, το οποίο θα αποτελέσει και το σημείο στο οποίο θα επικεντρωθεί η παρακάτω ανάλυση. Γενικότερα, τέτοιου είδους προσπάθειες ολοένα και αυξάνονται αφού περισσότεροι επιτήδειοι προσπαθούν με διάφορους τρόπους να αποσπάσουν διάφορες πληροφορίες, κάνοντας έτσι τα πράγματα ακόμα πιο δύσκολα όσον αφορά τα μέτρα προστασίας και πρόληψης ενάντια σε τέτοιες περιπτώσεις.

1.2 Ανίχνευση απάτης σε πιστωτικές κάρτες

Ένα από τα μεγαλύτερα ζητήματα στον τομέα αντιμετώπισης του προβλήματος, είναι η διαδικασία που πραγματοποιείται για την αρχική ανίχνευση τέτοιου είδους απατών. Η προσπάθεια ανίχνευσης μιας κάρτας η οποία έχει παραβιαστεί δεν είναι απλή υπόθεση, γι’ αυτό το λόγο υπάρχουν διάφορες τεχνικές, σύμφωνα με τις οποίες αναλύοντας συγκεκριμένα δεδομένα, μπορούμε πλέον να δούμε αν οι πληροφορίες της κάρτα έχουν υποκλαπεί και χρησιμοποιηθεί προς συμφέρον των δραστών. Οι τεχνικές αντιμετώπισης διαφέρουν μεταξύ τους, παρόλα αυτά οι περισσότερες πραγματοποιούνται βάσει προηγούμενων δεδομένων τα οποία έχουν καταγραφεί από διάφορες υποκλοπές σε πιστωτικές κάρτες. Η τεχνική ανίχνευσης-αντιμετώπισης απατών στην οποία επικεντρώνεται η συγκεκριμένη εργασία είναι αυτή που βασίζεται στην μηχανική μάθηση (Machine Learning). Εκμεταλλευόμενοι λοιπόν τα εφόδια που μπορεί να μας δώσει αυτή η τεχνική και αναλύοντας συγκεκριμένα δεδομένα, μπορούμε να έχουμε πολύ ενδιαφέροντα ευρήματα τα οποία μπορεί κανείς να χρησιμοποιήσει προκειμένου να αντιμετωπίσει το πρόβλημα των απατών. Τέτοιες διαδικασίες ανίχνευσης, έχουν υιοθετηθεί από αρκετές τράπεζες οι οποίες μόνο χαμένες μπορούν να βγουν, εάν ένας πελάτης τους

ζημιωθεί μέσω κάποιας υποκλοπής. Γι' αυτό το λόγο, οι τραπεζικοί οργανισμοί επιδιώκουν να παρέχουν όσο το δυνατόν μεγαλύτερη ασφάλεια απέναντι στους πελάτες τους προκειμένου να μην χαθεί το κλίμα εμπιστοσύνης το οποίο υπάρχει, ενώ παράλληλα, οι τράπεζες μπορούν να βγουν αρκετά ζημιωμένες από μια τέτοια κατάσταση.

1.3 Κατηγορίες δραστών

Οι απάτες πιστωτικών καρτών είναι ένα παγκόσμιο πρόβλημα το οποίο προσπαθεί να αντιμετωπιστεί όσο το δυνατόν γίνεται καλύτερα από τους ίδιους τους χρήστες καρτών αλλά και από την πλευρά των χρηματοοικονομικών οργανισμών, οι οποίοι ζημιώνεται εξίσου όταν γίνει κάτι τέτοιο. Γι' αυτό το λόγο κάθε οργανισμός προσπαθεί να διαχειρίζεται όσο δυνατόν καλύτερα τέτοια γεγονότα. Μια βασική ερώτηση που προκύπτει, αναφέρεται στο ποιος θα μπορούσε να είναι πίσω από μια απάτη αλλά και πως απέκτησε πρόσβαση σε τόσο ευαίσθητες πληροφορίες. Ο υπεύθυνος για μια τέτοια πράξη μπορεί να είναι ο καθένας ακόμα και άτομα τα οποία είναι υπεράνω πάσης υποψίας. Έτσι, δεν αποκλείει το ενδεχόμενο, το πρόσωπο που κρύβεται πίσω από την απάτη, να είναι ακόμα και κάποιος εργαζόμενος μιας τράπεζας ο οποίος θα μπορούσε εύκολα να αποκτήσει πρόσβαση σε κάποιο λογαριασμό και να εκμεταλλευτεί την ευκαιρία. Γι' αυτό το λόγο, τόσο οι τράπεζες όσο οι οργανισμοί οφείλουν να ελέγχουν κάθε περιστατικό και να παρέμβουν εγκαίρως. Υπάρχουν βέβαια και περιπτώσεις στις οποίες δεν εμπλέκεται άμεσα μια τράπεζα, αφού η απάτη πραγματοποιείται μέσω διαδικτύου σε online παραγγελίες. Ένα τέτοιο παράδειγμα αποτελούν οι απόπειρες phishing(προσπάθεια “ψαρέματος σημαντικών στοιχείων όπως κωδικοί κλπ), κατά τις οποίες ο απατεώνας προσπαθεί να υποκλέψει κάποια σημαντικά στοιχεία μιας κάρτας και να καταφέρει έτσι να αποσπάσει χρηματικά ποσά. Συνήθως, οι δράστες έχουν αρκετά καλή γνώση πάνω στο αντικείμενο των ηλεκτρονικών υπολογιστών, αφού συνήθως τέτοιου είδους απάτες πραγματοποιούνται μέσω του διαδικτύου. Αρκετοί από αυτούς μπορεί να είναι black hat hackers (κακόβουλοι) οι οποίοι έχουν την τεχνογνωσία να πραγματοποιήσουν επιθέσεις ώστε να καταφέρουν να υποκλέψουν σημαντικές πληροφορίες. Από την άλλη πλευρά, υπάρχουν και τα φυσικά πρόσωπα- κλέφτες, οι οποίοι απλά κλέβουν κάποια πιστωτική κάρτα και την χρησιμοποιούν προς όφελος τους. Τα τελευταία χρόνια έχει παρατηρηθεί άλλο ένα φαινόμενο, το οποίο πραγματοποιείται με την εξέλιξη που υπάρχει πλέον στις ανέπαφες συναλλαγές. Κατά το φαινόμενο αυτό, ο απατεώνας μπορεί να χρησιμοποιήσει ένα τερματικό μηχανήμα ή ακόμα και το κινητό του τηλέφωνο(όπου αυτή η λειτουργία υποστηρίζεται) και φέρνοντάς το κοντά στην τσέπη του θύματος όπου υπάρχει το πορτοφόλι με την ανέπαφη κάρτα, να πραγματοποιήσει μια αγορά χωρίς φυσικά να την έγκριση του θύματος.

1.4 Κατηγορίες απατών στις πληρωμές

Με μία γενικότερη ματιά θα μπορούσαμε να πούμε πως υπάρχουν 6 βασικές κατηγορίες απατών μέσω κάρτας.

- Lost or stolen cards
- account takeover
- counterfeit cards
- never received
- fraudulent application
- collusive merchants
- email order/telephone order fraud

Lost or stolen cards: Είναι μια από τις συνηθέστερες περιπτώσεις όπου κάποιος μπορεί να εκμεταλλευτεί την χαμένη κάρτα ενός θύματος, προκειμένου να του αποσπάσει κάποιο χρηματικό ποσό.

Account takeover : Σε αυτή τη περίπτωση, ο απατεώνας θα προσπαθήσει να αποκτήσει πλήρη πρόσβαση στους διαφορετικούς λογαριασμούς του χρήστη- θύματος (όπως τραπεζικοί λογαριασμού, κάρτες SIM, email κλπ). Στη συνέχεια, αφού βρει τις κατάλληλες πληροφορίες που χρειάζεται, προσπαθεί να αποκτήσει πρόσβαση στους λογαριασμούς τραπεζής του χρήστη με σκοπό την απόσπαση χρηματικών ποσών. Το θύμα,εάν είναι τυχερό, μπορεί να καταλάβει ότι υπάρχει κάποιο πρόβλημα με το λογαριασμό του αφού ίσως παρατηρήσει κάποια περίεργη δραστηριότητα στους ηλεκτρονικούς λογαριασμούς(όπως ασυνήθιστη δραστηριότητα email). Σε αντίθετη περίπτωση,θα παρατηρήσει τα χαμένα χρήματα στο λογαριασμό του όταν θα είναι ήδη αργά. Συνήθως, τα θύματα πέφτουν στη παγίδα μέσω ψευδών διαφημίσεων, μέσω της τεχνικής phishing αλλά και μέσω διαφόρων κακόβουλων λογισμικών.

- Counterfeit cards: Στη συγκεκριμένη περίπτωση, ο δράστης δημιουργεί έναν «κλώνο» της κάρτας του θύματος προκειμένου να τη χρησιμοποιήσει αντί για την αυθεντική με αποτέλεσμα να «ξεγελάσει» αρκετές φορές το σύστημα στο οποίο εισάγει την κάρτα(ATM, αυτόματοι πωλητές).
- Never received : Σε αυτή τη περίπτωση, αναφέρεται η υποκλοπή μιας κάρτας από έναν απατεώνα μέσω ενός λογαριασμού του θύματος και στη συνέχεια αποστέλλεται μια καινούργια προς αντικατάσταση της προηγούμενης, η οποία δεν φτάνει ποτέ στον πραγματικό της ιδιοκτήτη.
- *Application fraud :* Η συγκεκριμένη κατηγορία αναφέρεται στις περιπτώσεις στις οποίες ο απατεώνας κλέβει προσωπικά έγγραφα του θύματος ή δημιουργεί πλαστά, προκειμένου να εμφανιστεί με το όνομα του θύματος και να μπορέσει να εκμεταλλευτεί

μια τέτοια κατάσταση. Ο δράστης, μπορεί να προσπαθήσει να δημιουργήσει έναν λογαριασμό σε τράπεζα με αυτά τα στοιχεία και να εκμεταλλευτεί το ψευδές όνομα του προκειμένου να αποσπάσει χρηματικά ποσά από τη τράπεζα σε βάρος του θύματος.

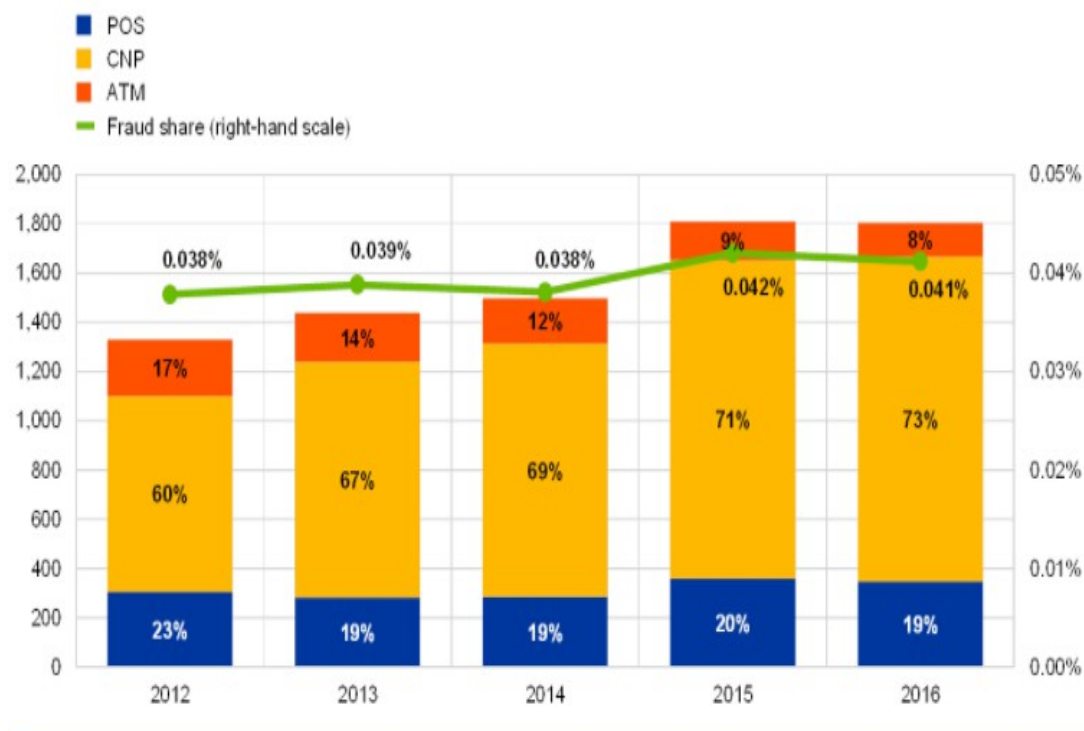
- Collusive merchants: Σε αυτή τη περίπτωση, έχουμε την συνεργασία ενός απατεώνα με τον υπάλληλο μια τράπεζας προκειμένου να υποκλέψουν σημαντικές πληροφορίες για το θύμα και να αποκτήσουν πρόσβαση στον τραπεζικό λογαριασμό του.

Email order/telephone order fraud: Στη τελευταία περίπτωση, έχουμε τις τηλεφωνικές παραγγελίες και τις παραγγελίες μέσω email, όπου και είναι ο πιο σύνηθες τρόπος απατών αφού οι περισσότερες παραγγελίες πραγματοποιούνται μέσω των δύο αυτών τρόπων. Όταν υπάρχει φυσική παρουσία της κάρτας, οι δράστες έχουν πολύ ευκολότερο έργο αφού μπορούν πολύ εύκολα να χρησιμοποιήσουν μιας κλεμμένη κάρτα προκειμένου να αγοράσουν κάτι μέσω email ή τηλεφώνου.

1.5 Επιπτώσεις απατών στην οικονομία

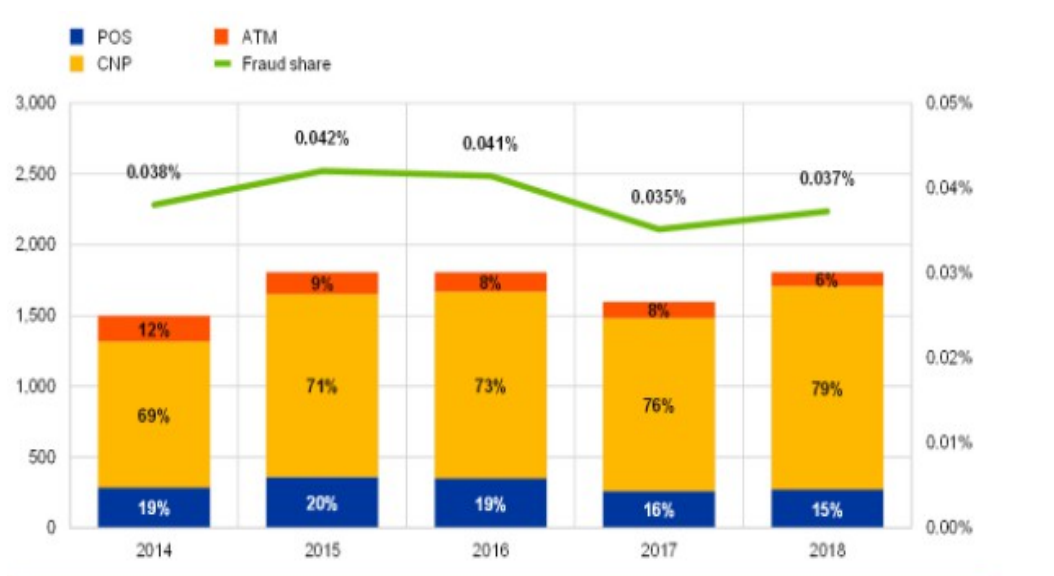
Οι απάτες πιστωτικών καρτών αποτελούν όπως γίνεται αντιληπτό, ένα μεγάλο πρόβλημα στην παγκόσμια οικονομία το οποίο χρήζει αντιμετώπισης. Τα τελευταία χρόνια, οι αριθμοί υποδεικνύουν το αρκετά σοβαρό πρόβλημα που προκύπτει από τέτοιου είδους απάτες. Σύμφωνα λοιπόν με την πέμπτη αναφορά της Ευρωπαϊκής Τράπεζας, η οποία επικεντρώνεται στις ευρωπαϊκές αγορές, το 2016 οι απάτες μεταφράστηκαν σε ζημιά 1.8 δισεκατομμύρια ευρώ η οποία είναι μειωμένη κατά 0,4% σε σχέση με το 2015. Πιο συγκεκριμένα, για το 2016 το 73% των απατών προήλθε από πληρωμές με μη φυσική παρουσία κάρτας (μέσω διαδικτύου, ηλεκτρονικού ταχυδρομείου ή τηλεφώνου, το 19% των συναλλαγών προήλθε από συναλλαγές μέσω τερματικών (ταμειακή μηχανή, POS) ενώ το 8% προήλθε από συναλλαγές που πραγματοποιήθηκαν μέσω ATM. Παρακάτω παρατίθεται αναλυτικό διάγραμμα με τα ποσοστά και τις κατηγορίες των απατών.

left-hand scale: total value (EUR millions); right-hand scale: value of fraud as share of value of transaction (%)



Εικόνα 1 (European Central Bank, 5th Report on Card fraud (2018))[1]

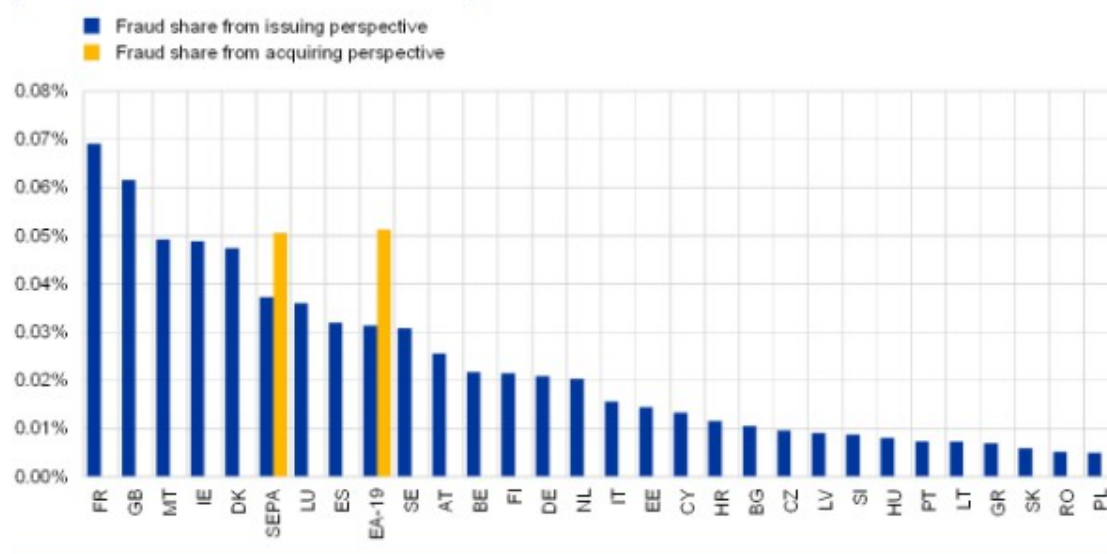
Σύμφωνα με την έκτη αναφορά της Ευρωπαϊκής Τράπεζας, οι συναλλαγές που πραγματοποιήθηκαν μέσω καρτών οι οποίες αποκτήθηκαν εντός ζώνης SEPA(Single Euro Payments Area) και χρησιμοποιήθηκαν παγκοσμίως, ανέρχεται στα 1,80 δισεκατομμύρια ευρώ για το 2018. Παράλληλα, όσον αφορά τις πιστωτικές κάρτες που εκδόθηκαν στη ζώνη του ευρώ και χρησιμοποιήθηκαν παγκοσμίως, η ζημιά ανήλθε στα 0,94 δισεκατομμύρια ευρώ για το 2018. Πιο αναλυτικά, το 79% των απατών προήλθε από συναλλαγές χωρίς την παρουσία της κάρτας (μέσω διαδικτύου, ηλεκτρονικού ταχυδρομείου, τηλεφωνική απάτη), το 15% προήλθε από συναλλαγές σε τερματικά μηχανήματα (ταμειακές, POS), ενώ το 6% προήλθε από συναλλαγές σε ATM. Οι απάτες μέσω μη φυσικής παρουσίας της κάρτας φαίνεται να αυξήθηκαν σε σύγκριση με το 2017 όπου είχαμε ποσοστό 76%, ενώ οι απάτες μέσω τερματικών POS μειώθηκαν (16% το 2017), όπως και οι απάτες μέσω ATM.(8% το 2017)



Εικόνα 2 (European Central Bank, 6th Report on Card fraud (2020))[2]

Σύμφωνα με την ίδια αναφορά και πάλι, μπορούμε να διακρίνουμε σε ποιές χώρες επιτεύχθηκε το μεγαλύτερο ποσοστό απατών, τόσο στη ζώνη του Ευρώ, όσο και στη ζώνη SEPA. Στο παρακάτω γράφημα, η μπλε μπάρα υποδεικνύει τις συναλλαγές με κάρτες οι οποίες εκδόθηκαν στη συγκεκριμένη χώρα και χρησιμοποιήθηκαν οπουδήποτε, ενώ η κίτρινη μπάρα, υποδεικνύει τις συναλλαγές με κάρτες οι οποίες εκδόθηκαν οπουδήποτε και πραγματοποιήθηκαν στη συγκεκριμένη τοποθεσία ή χώρα. Όπως διαπιστώνουμε, η Γαλλία έχει το μεγαλύτερο ποσοστό εκδόσεων καρτών μέσω των οποίων πραγματοποιήθηκαν απάτες, με την Αγγλία και την Μάλτα να ακολουθούν.

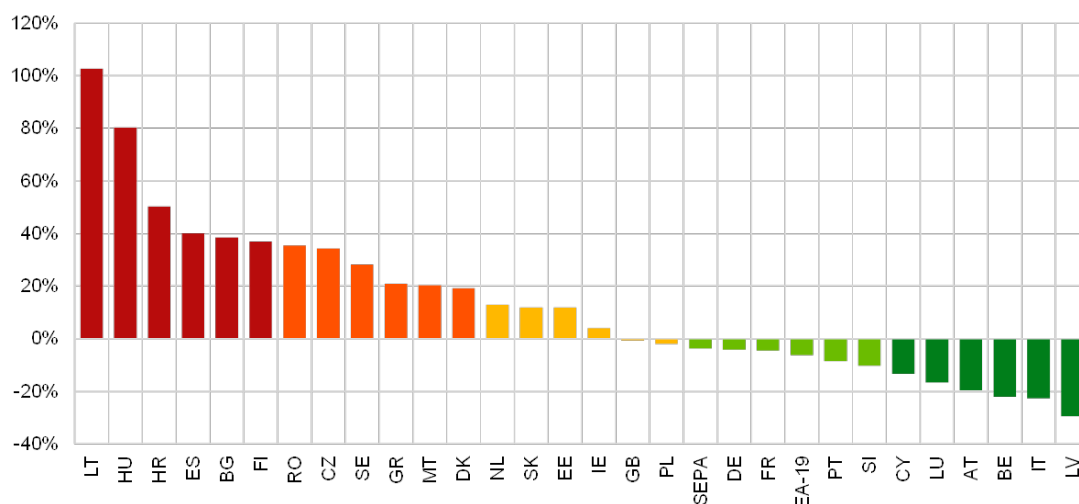
(value of fraud as share of value of transactions)



Εικόνα 3 (European Central Bank, 6th Report on Card fraud (2020))[3]

Ακολουθεί ένα ακόμη γράφημα το οποίο παρουσιάζει το ποσοστό αύξησης και μείωσης της

συνολικής αξίας των απατών που έχουν πραγματοποιηθεί τα τελευταία χρόνια(2014-2018) με την Λιθουανία να είναι πρώτη στη λίστα, κατέχοντας το μεγαλύτερο ποσοστό αύξησης το οποίο αγγίζει το 100%, ενώ ακολουθούν η Ουγγαρία και η Κροατία, με ποσοστό αύξησης 80% και 50% αντίστοιχα.



Εικόνα 4 (European Central Bank, 6th Report on Card fraud (2020))[4]

Οι εκτιμήσεις για το μέλλον, σύμφωνα με τα υπάρχοντα δεδομένα δεν είναι τόσο ενθαρρυντικές, αφού σύμφωνα με εκτιμήσεις της Nilson Report κατά το έτος 2019, στα επόμενα χρόνια και πιο συγκεκριμένα το 2023, οι πληρωμές με κάρτες αναμένεται να φτάσουν στα 57.080 τρισεκατομμύρια, με τις απάτες καρτών να φτάνουν στα 35.67 δισεκατομμύρια. Ο αριθμός χαμένων χρημάτων εκτιμάται ότι θα αυξηθεί περαιτέρω έως το 2027, όπου σύμφωνα και πάλι με την ίδια έρευνα, οι πληρωμές με κάρτες θα φτάσουν τα 71.593 τρισεκατομμύρια, ενώ οι απάτες καρτών θα στοιχίσουν περίπου 40.63 δισεκατομμύρια. [5]

1.6 Σημαντικότητα πρόβλεψης των απατών και μέτρα αντιμετώπισης

Όπως είδαμε και προηγουμένως, το να μπορεί κάποιος να εντοπίζει μια ενδεχόμενη απάτη είναι πολύ σημαντικό θέμα ειδικότερα στις μέρες μας όπου πλέον τέτοια γεγονότα αποτελούν καθημερινότητα. Γι' αυτό το λόγο, ας δούμε για ποιους τομείς και πόσο σημαντική είναι η πρόβλεψη των απατών στις πιστωτικές κάρτες

- Αρχικά, στον απλό πολίτη που έχει στη κατοχή του κάρτες, αφού προφανώς διατρέχει

το μεγαλύτερο κίνδυνο ως επικείμενο θύμα κάποιας απάτης.

- Τις τράπεζες, αφού όπως είδαμε τέτοιου είδους ζημιές μπορούν να επηρεάσουν την οικονομικής τους κατάσταση, να αποφέρουν μεγάλες οικονομικές ζημιές αλλά και να επηρεάσουν σημαντικά την φήμη τους.
- Το κράτος, το οποίο επίσης χάνει χρήματα όταν πραγματοποιούνται τέτοια φαινόμενα, ενώ παράλληλα πλήττεται το κύρος και η αξιοπιστία του με αποτέλεσμα να μειώνονται οι πιθανοί ξένοι επενδυτές στη χώρα.
- Οι επιχειρήσεις, οι οποίες πολλές φορές τυχαίνει να πέσουν θύμα μιας απάτης, πράγμα που μπορεί να γίνει εύκολα λόγω κάποιου λάθους από κάποιον εργαζόμενο. Επομένως, μια απάτη με σκοπό την υποκλοπή χρηματικών ποσών, μπορεί να αποβεί αρκετά επιζήμια για μια επιχείρηση, γι' αυτό και η αντιμετώπιση τέτοιων καταστάσεων αποτελεί προτεραιότητα για πολλές επιχειρήσεις.

Προκειμένου να αποφευχθεί ένα τέτοιο ατυχές συμβάν, υπάρχουν διάφοροι τρόποι με τους οποίους κάποιος θα μπορούσε να προστατευθεί από τέτοιες επιθέσεις, με κάποιες γενικές συμβουλές να παρατίθενται παρακάτω.

- Διατήρηση της κάρτας σε ασφαλές μέρος προκειμένου να μην είναι εύκολη μια ενδεχόμενη επίθεση.
- Αποφυγή διαμοιρασμού στοιχείων της κάρτας.
- Μεγάλη προσοχή στα scam emails (παραπλανητικά email)
- Συχνοί έλεγχοι στις τραπεζικές κινήσεις
- Δυνατοί κωδικοί, (όπου απαιτείται κωδικός πρόσβασης) και συχνή αλλαγή τους.

Κεφάλαιο 2 “Μηχανική μάθηση”

2.1 Εισαγωγή στη μηχανική μάθηση (Machine learning)

Τα τελευταία χρόνια, η μηχανική μάθηση ή αλλιώς machine learning (όπως είναι ευρύτερα γνωστό), είναι ένας τομέας της πληροφορικής ο οποίος έχει εισβάλει για τα καλά στη ζωή μας και σίγουρα θα μας απασχολήσει αρκετά στο μέλλον. Η μηχανική μάθηση έχει εισέλθει στη ζωή μας κυρίως μέσω “έξυπνων συσκευών”, συστήματα δηλαδή τεχνητής νοημοσύνης τα οποία συνήθως είναι απλές καθημερινές συσκευές τις οποίες έχουμε στο σπίτι μας, όπως το κινητό τηλέφωνο, μια “έξυπνη” λάμπα ακόμα και το air condition του σπιτιού μας. Η λειτουργία πολλών από αυτές τις συσκευές είναι άρρηκτα συνδεδεμένη με το machine learning. Ουσιαστικά, το machine learning είναι ο τρόπος με τον οποίο ένα σύστημα επεξεργάζεται συγκεκριμένες πληροφορίες προκειμένου να βγάλει ένα συμπέρασμα ή μια πρόβλεψη σε μια ερώτηση βάσει συγκεκριμένων αλγορίθμων. Παράλληλα, έχει τη δυνατότητα να προβλέψει καταστάσεις βάσει παλαιότερων πληροφοριών που έχουν δοθεί και ακολουθώντας πάλι συγκεκριμένα βήματα των αλγορίθμων που επιλέγονται για το συγκεκριμένο πρόβλημα. Η μηχανική μάθηση χωρίζεται επίσης σε 3 βασικά είδη , την Επιβλεπόμενη μάθηση (Supervised Learning), μη επιβλεπόμενη μάθηση(Unsupervised Learning) και Ενισχυτική μάθηση (Reinforcement Learning).

2.2 Χρησιμότητα και λειτουργία της μηχανικής μάθησης

Η μηχανική εκμάθηση χρησιμοποιείται κάποιες φορές σε συνδυασμό με την τεχνητή νοημοσύνη, ενώ κάποιες φορές προτιμάται αυτόνομη, αφού πολλά συστήματα χρησιμοποιούν τέτοιου είδους προσέγγιση προκειμένου να εξυπηρετούν τους πελάτες τους. Ένα πολύ εύκολο και καθημερινό παράδειγμα είναι η αναζήτηση που πραγματοποιούμε σε μια μηχανή αναζήτησης. Είναι σχεδόν σίγουρο, πως τα αποτελέσματα τα οποία θα λάβουμε είναι εξατομικευμένα και σχετικά με κάποιες προηγούμενες αναζητήσεις μας. Αυτό λοιπόν συμβαίνει μέσω του machine learning, μιας και αναλύοντας προηγούμενες αναζητήσεις, ένα σύστημα μπορεί να κάνει μια πρόβλεψη για το επιθυμητό αποτέλεσμα. Ένα ακόμα πιο ξεκάθαρο παράδειγμα είναι το auto-correct (αυτόματη συμπλήρωση) των αναζητήσεων μας σε κάποια μηχανή αναζήτησης, όπου εκεί το σύστημα «μαθαίνει» από τις πληροφορίες που του παρέχουν οι χρήστες, ώστε αυτό να γίνει «έξυπνότερο» και να μπορεί να είναι πιο ακριβές στην αυτόματη συμπλήρωση προτάσεων προς αναζήτηση. Προκειμένου λοιπόν να λειτουργήσει η μέθοδος της μηχανικής μάθησης, ένα σύστημα θα πρέπει να λάβει πολλά ερεθίσματα αλλά και την

πληροφορία για το τι είναι σωστό και τι λάθος προκειμένου να «μαθαίνει» από τα δύο αυτά αποτελέσματα. Αν για παράδειγμα είχαμε ένα ρομπότ στο οποίο θα θέλαμε να μάθουμε να ξεχωρίζει δυο διαφορετικά αντικείμενα (ένα βάζο και ένα μήλο), θα έπρεπε να του δώσουμε πολλές εικόνες από τα δύο διαφορετικά αντικείμενα με τη πληροφορία του αν είναι όντως αληθές αυτό που το ρομπότ θα μάντεψε(το οποίο αρχικά μαντεύει τυχαία). Έτσι, σταδιακά και όταν αυτό θα έχει αποκτήσει την κατάλληλη εμπειρία και αφού θα έχει ξεχωρίσει το πως μοιάζει το ένα αντικείμενο και πώς το άλλο, τότε θα είναι σε θέση να μπορεί να προβλέψει σωστά. Κάτι τέτοιο όμως, όπως αναφέραμε χρειάζεται μεγάλο αριθμό δειγμάτων προκειμένου το αποτέλεσμα να είναι ακριβέστερο και να αποφευχθεί η τυχαία επιλογή του ρομπότ. Με αυτό το τελείως απλοϊκό παράδειγμα, μπορούμε να καταλάβουμε την ουσία των πραγμάτων, που δεν είναι άλλη από τη χρησιμότητα που μπορεί να μας προσφέρει η μέθοδος του machine learning αλλά παράλληλα και τα προβλήματα τα οποία μπορεί να δημιουργήσει, μιας και πολλές φορές τα δεδομένα τα οποία συλλέγει κανείς, μπορεί να υπόκεινται σε προσωπικές πληροφορίες που κάποιος δεν θα ήθελε να γνωστοποιηθούν ακόμα και για πειραματικούς σκοπούς. Κάπου εδώ λοιπόν καταλαβαίνουμε ότι προκύπτουν περαιτέρω ζητήματα τα οποία θα ανφέρουμε αργότερα και αφορούν την ιδιωτικότητα, όπου πρέπει επίσης να ληφθούν υπόψη.

2.3 Ιστορία της μηχανικής μάθησης

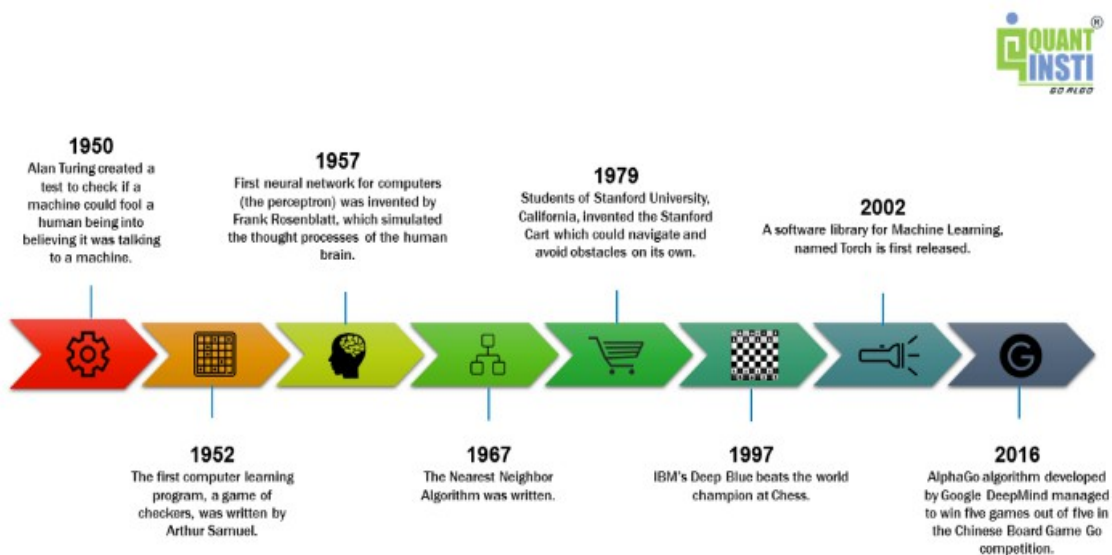
Η πρώτη αναφορά στα νευρωνικά δίκτυα προέρχεται από το 1943, όταν ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts, έγραψαν μια αναφορά σχετικά με τους νευρώνες. Αποφάσισαν λοιπόν, βάση αυτού, να δημιουργήσουν ένα μοντέλο, χρησιμοποιώντας ένα ηλεκτρικό κύκλωμα ξεκινώντας έτσι τα νευρωνικά δίκτυα. Τη χρονιά 1950 ο Alan Turing πραγματοποίησε το Turing test κατά το οποίο, ένας υπολογιστής θα έπρεπε να καταφέρει να πείσει έναν άνθρωπο ότι πίσω από τον υπολογιστή, βρίσκεται κάποιος άνθρωπος. Το 1952 ο Arthur Samuel δημιουργεί ένα παιχνίδι το οποίο έχει τη δυνατότητα να μαθαίνει πληροφορίες καθώς αυτό εξελίσσεται. Το 1958 έχουμε το πρώτο τεχνητό νευρωνικό δίκτυο από τον Frank Rosenblatt ενώ το 1959 οι Bernard Widrow και Macrian Hoff δημιουργούν δύο τέτοια μοντέλα προκειμένου να αναγνωρίζουν απλά, πράγματα ή αντικείμενα για αρχή. Έως το 1970 δεν υπήρξε κάποια μεγάλη πρόοδος. Το 1982 ξεκίνησε και πάλι το ενδιαφέρον για τα νευρωνικά δίκτυα, αφού ο John Hopfield πρότεινε τη δημιουργία ενός δικτύου παρεμφερούς των νευρωνικών δικτύων. Το 1997, ένας ηλεκτρονικός υπολογιστής της IBM κατάφερε να κερδίσει τον παγκόσμιο πρωταθλητή σκακιού. Από εκεί και πέρα, η μηχανική μάθηση γίνεται ολοένα και χρησιμότερη, φτάνοντας στην σημερινή εποχή όπου χρησιμοποιείται ευρέως προκειμένου να μας διευκολύνει. Μερικά μεγάλα έργα στα οποία χρησιμοποιήθηκε το machine learning είναι :

- GoogleBrain(2012)- Το project αυτό είχε σαν στόχο την αναγνώριση μοτίβων σε

εικόνες και βίντεο χρησιμοποιώντας πηγές από το Google.

- DeepFace(2014)- Το project αυτό δημιουργήθηκε από το Facebook, προσπαθώντας να βρει μοτίβα ανάμεσα σε πρόσωπα και να τα αναγνωρίσει.
- DeepMind(2014)-Η εταιρεία εξαγοράστηκε από τη Google και δημιουργήθηκε για να “μάθει” να παίζει βιντεοπαιχνίδια όπως θα έπαιζε ένας άνθρωπος.
- OpenAI(2015)-Η εταιρεία δημιουργήθηκε από τον Elon Musk και κάποιους ακόμα, με σκοπό τη δημιουργία και ανάπτυξη τεχνητής νοημοσύνης από την οποία θα επωφεληθεί η ανθρωπότητα.
- Amazon Machine Learning Platform(2015)- Αποτελεί ένα μέρος των υπηρεσιών του Amazon και από πληροφορίες της ίδιας της εταιρείας, η πλατφόρμα αυτή βοηθάει στη λειτουργία πολλών εσωτερικών συστημάτων, όπως απλές λειτουργίες και προτάσεις αναζήτησης της Alexa(σύστημα φωνητικών εντολών της Amazon), αλλά και πειραματικά τα συστήματα Prime Air και Amazon Go.

Παρακάτω βλέπουμε ένα τυπικό χρονοδιάγραμμα της ιστορίας της μηχανικής μάθησης.



Εικόνα 5 (History of Machine Learning, Medium(2019))[6]

2.4 Εφαρμογές μηχανικής μάθησης στη καθημερινότητα

Η μηχανική μάθηση πλέον αποτελεί καθημερινότητα στη ζωή μας ακόμα και αν δεν το συνειδητοποιούμε αφού πολλές δραστηριότητες της καθημερινότητας μας εμπεριέχουν στοιχεία

τα οποία σχετίζονται με τη μηχανική μάθηση. Ας δούμε λοιπόν κάποιες από τις εφαρμογές της μηχανικής μάθησης οι οποίες βρίσκουν εφαρμογή στην καθημερινότητα.

- Εικονικοί προσωπικοί βοηθοί

Όπως αναφέρθηκε και παραπάνω, η Siri, η Alexa κλπ αποτελούν τους δημοφιλέστερους εικονικούς βοηθούς για Smartphone και tablet, οι οποίοι δέχονται φωνητικές εντολές και εκείνοι απαντάνε ή εμφανίζουν το κατάλληλο αποτέλεσμα. Σε μια τέτοια περίπτωση, τα αποτελέσματα εμφανίζονται, βάση πληροφοριών που έχουν ήδη συλλεγεί σχετικά με το χρήστη και επεξεργαστεί ενώ υπάρχει η δυνατότητα απάντησης σε διάφορες ερωτήσεις. Για όλα τα παραπάνω, υπεύθυνη είναι η μηχανική μάθηση αφού τα αποτελέσματα που παρέχονται, βασίζονται στην «εκπαίδευση» και στον όγκο πληροφοριών που έχουμε δώσει στο σύστημα μας.

- Προβλέψεις κατά την μετακίνηση

Μια ακόμα εφαρμογή της μηχανικής μάθησης συναντάται στις μετακινήσεις, καθώς το GPS ενός αυτοκινήτου θα προσπαθήσει να δώσει στον οδηγό την βέλτιστη λύση αποφεύγοντας την κίνηση. Στη πραγματικότητα, δε μπορεί το GPS να γνωρίζει αν υπάρχει κίνηση ή όχι, παρόλα αυτά μέσω της μηχανικής μάθησης μπορεί να υπάρξει μια σχετική πρόβλεψη βάση προηγούμενων πληροφοριών που έχει συλλέξει στο παρελθόν για συγκεκριμένα σημεία του οδικού δικτύου.

- Υπηρεσίες κοινωνικών δικτύων

Τα κοινωνικά δίκτυα είναι από τις καθημερινές δραστηριότητες που έχουν επηρεαστεί περισσότερο από την μηχανική μάθηση. Αυτό συμβαίνει γιατί μέσω της μηχανικής μάθησης, οι διάφορες πλατφόρμες μπορούν να προσαρμόζουν το περιεχόμενο τους ανάλογα με τις προτιμήσεις του χρήστη. Ένα τέτοιο παράδειγμα αποτελεί το περιεχόμενο του Facebook, το οποίο με τη βοήθεια της μηχανικής μάθησης, εμφανίζει περιεχόμενο σύμφωνα με πρόσφατες αναζητήσεις, προφίλ ανθρώπων που επισκεπτόμαστε συχνότερα, αλλά ανάλογες διαφημίσεις, σύμφωνα με τις προτιμήσεις του χρήστη. Τα τελευταία χρόνια μάλιστα, το Facebook, προσφέρει την υπηρεσία αναγνώρισης προσώπου, κατά την οποία το σύστημα εξετάζει μια φωτογραφία και ενημερώνει τον χρήστη ότι μια φωτογραφία που πιθανόν να βρίσκεται σε αυτήν, έχει αναρτηθεί στο Facebook, αποτελώντας άλλο ένα παράδειγμα για τη χρησιμότητα της μηχανικής μάθησης.

- Μηνύματα ηλεκτρονικού ταχυδρομείου

Η μηχανική μάθηση μπορεί να μας βοηθήσει αρκετά σε έναν ακόμη τομέα, ο οποίος είναι ο εντοπισμός ψευδών, ενοχλητικών ή και επικίνδυνων μηνυμάτων ηλεκτρονικού ταχυδρομείου. Τα email που λαμβάνουμε φιλτράρονται κατάλληλα σύμφωνα με αλγορίθμους μηχανικής μάθησης και αν το αποτέλεσμα δείξει ότι είναι κάτι από τα παραπάνω, τότε το email αντιμετωπίζεται κατάλληλα. Καθημερινά, τα μηνύματα που λαμβάνονται με κάποιο κακόβουλο

λογισμικό ξεπερνούν τα 325.000, επομένως καταλαβαίνουμε πόσο σημαντική και χρήσιμη είναι η χρήση μηχανικής μάθησης για την αντιμετώπιση του προβλήματος.

- Προτεινόμενα προϊόντα σε αναζητήσεις

Όπως ανέφερα και παραπάνω, πολλές φορές μια απλή αναζήτηση ενός προϊόντος, μπορεί στη συνέχεια να μας φέρει πολλά παρόμοια αποτελέσματα, κάτι το οποίο επιτυγχάνεται μέσω της μηχανικής μάθησης. Για παράδειγμα, μετά από κάποια παραγγελία, το σύστημα αυτόματα θα εμφανίσει κάποια προϊόντα- προτάσεις, παρέχοντας στο χρήστη παρόμοια προϊόντα με αυτά που αγόρασε.

- Ανίχνευση διαδικτυακών απατών

Ο συγκεκριμένος τομέας, αποτελεί ένα από τα μεγαλύτερα επιτεύγματα της μηχανικής μάθησης, αφού η ανίχνευση τέτοιων δραστηριοτήτων, μπορεί να επιφέρει αποφυγή απώλειας μεγάλων χρηματικών ποσών, αλλά και να βοηθήσει στην εξασφάλιση ασφαλέστερων συναλλαγών. Η μηχανική μάθηση, βοηθάει κυρίως συγκρίνοντας στοιχεία συναλλαγών με αντίστοιχα στοιχεία απατών-συναλλαγών και με συμπεριφορικά μοτίβα των χρηστών, προκειμένου να βρεθεί κάποια ύποπτη συναλλαγή και να αντιμετωπιστεί κατάλληλα. Το machine learning, κρίνεται απαραίτητο εργαλείο για την ανίχνευση διαδικτυακών απατών γι' αυτό το λόγο και έχει υιοθετηθεί από πολλές μεγάλες εταιρείες οι οποίες ασχολούνται με χρηματικές συναλλαγές.

2.5 Μηχανική μάθηση ή απλοί κανόνες τεχνητής νοημοσύνης;

Όπως έχει αναφερθεί και παραπάνω, η μηχανική μάθηση είναι ένας από τους αποτελεσματικότερους τρόπους να ανιχνεύσουμε απάτες σε πιστωτικές κάρτες. Γιατί όμως η μηχανική μάθηση βοηθάει τόσο πολύ ενώ υπερτερεί σε σχέση με άλλες μεθόδους ανίχνευσης; Αρχικά, κάποιες παλαιότερες μέθοδοι οι οποίες χρησιμοποιούσαν απλούς κανόνες τεχνητής νοημοσύνης, είχαν κατά ένα μέρος πειστικά αποτελέσματα όμως τα προβλήματα ήταν αρκετά. Αυτό που συνέβαινε συνήθως, ήταν η εμφάνιση πολλών περιπτώσεων ως απάτες, ενώ στη πραγματικότητα πίσω από τις συναλλαγές ήταν οι αληθινοί χρήστες. Οι απλοί κανόνες ήταν επίσης αρκετά απόλυτοι, με μικρές δυνατότητες ευελιξίας αφού οι περισσότερες απαντήσεις περιορίζονται σε “ναι/όχι”. Παράλληλα, πολλές φορές μια συναλλαγή από χώρα με μεγαλύτερο δείκτη απατών, αυτόματα θεωρούταν απάτη αν το ποσό ήταν μεγάλο. Όπως καταλαβαίνουμε λοιπόν, οι απλοί κανόνες δεν ήταν πολύ αποτελεσματικοί, άρα θα έπρεπε να υπάρξει μια διαφορετική προσέγγιση στο ζήτημα προκειμένου τα αποτελέσματα να είναι πιο ακριβή και όσο περισσότερο γίνεται, κοντά στην πραγματικότητα. Ας δούμε όμως τις βασικές διαφορές μεταξύ των σύγχρονων απλών κανόνων ΑΙ με τη μηχανική μάθηση. [7]

- Αρχικά, η μηχανική μάθηση είναι πιθανοτική σε αντίθεση με τους απλούς κανόνες ΑΙ οι οποίοι είναι ντετερμινιστικοί. Αυτό πρακτικά σημαίνει ότι τα συστήματα μηχανικής

μάθησης εξελίσσονται και προσαρμόζονται σύμφωνα με τα δεδομένα που τους παρέχονται.

- Μία άλλη διαφορά των δύο μεθόδων είναι η εξέλιξη που μπορεί να έχει η μέθοδος της μηχανικής μάθησης σε αντίθεση με τους απλούς κανόνες οι οποίοι δεν έχουν τόσο μεγάλες πιθανότητες και περιθώρια εξέλιξης, συγκριτικά με τη μηχανική μάθηση.
- Κατά τη μέθοδο της μηχανικής μάθησης, απαιτούνται περισσότερα δεδομένα προκειμένου να έχουμε σωστά αποτελέσματα, σε αντίθεση με την περίπτωση των απλών κανόνων όπου μπορούμε να έχουμε αποτελέσματα με βασικά γνωρίσματα μιας περίπτωσης.
- Οι απλοί κανόνες είναι αμετάβλητα αντικείμενα, δεν χρειάζεται δηλαδή να κάνουμε κάποια αλλαγή στη μορφή τους ενώ στη μηχανική μάθηση μπορεί να χρειαστεί να κάνουμε κάποιες τροποποιήσεις και να το μετατρέψουμε σε κάποια άλλη γλώσσα προγραμματισμού.
- Κατά τη μέθοδο της μηχανικής μάθησης το σύστημα “μαθαίνει” από εμάς προκειμένου το αποτέλεσμα να είναι όσο το δυνατόν ακριβέστερο, πράγμα που δεν συμβαίνει με τους απλούς κανόνες AI. Παρακάτω παραθέτω έναν πίνακα με τα βασικά χαρακτηριστικά του κάθε συστήματος.

Πίνακας 1 (Σύγκριση απλών κανόνων AI με μηχανική μάθηση)[8]

	Απλοί κανόνες	Μηχανική μάθηση
Στόχος	Αναγνώριση συγκεκριμένων χαρακτηριστικών	Αναγνώριση συγκεκριμένων χαρακτηριστικών
Λειτουργία	Βασισμένο σε καθορισμένους κανόνες	Στοιχεία που εισάγονται και χρησιμοποιούνται για εξάσκηση(train) και δοκιμή
Μοντέλο	Εμπειρικά καθορισμένοι κανόνες	Αυτοματοποιημένη δημιουργία
Labeled Data	Όχι	Ναι (για εξάσκηση)
Χειροκίνητη διαδικασία	Απαραίτητη για να εισαχθούν δεδομένα	Ελάχιστη
Κόστος	Ρύθμιση από τον άνθρωπο	Λειτουργία σύζευξης
Ευελιξία	Πολύ χαμηλή	Πολύ υψηλή

Όπως μπορούμε να καταλάβουμε από τον πίνακα αλλά και από τα προηγούμενα γενικότερα συμπεράσματα, η μηχανική μάθηση είναι το μέλλον, χωρίς όμως αυτό να αφήνει τους κανόνες στην άκρη, αφού σε κάποιες περιπτώσεις μπορούν να αποδειχθούν πολύ χρήσιμοι. Σε κάποιες περιπτώσεις είναι αρκετά χρήσιμοι, ωστόσο η μηχανική μάθηση έχει αρκετά πλεονεκτήματα σε σύγκριση με τους απλούς κανόνες AI.

2.6 Μηχανική μάθηση και πρόβλεψη απατών σε πιστωτικές κάρτες

Η μηχανική μάθηση όπως είδαμε και παραπάνω, αποτελεί ένα από τα πιο αποτελεσματικά εργαλεία για την επίλυση διαφόρων προβλημάτων αλλά παράλληλα αποτελεί ένα από τα σύγχρονα μέσα για την επίτευξη διαφόρων λειτουργιών στην καθημερινότητα μας. Όπως είδαμε προηγουμένως, οι απάτες σε πιστωτικές κάρτες αποτελούν μια μάλιστα στη σύγχρονη κοινωνία την οποία οι ειδικοί ψάχνουν τρόπο να περιορίσουν ή και να εξαλείψουν. Γι' αυτό το λόγο λοιπόν, στράφηκαν προς τη μηχανική μάθηση αφού όπως θα δούμε, αποτελεί μια πολύ χρήσιμη λύση σε αυτό το μεγάλο πρόβλημα. Αρχικά, η χρήση μηχανικής μάθησης για την πρόβλεψη απατών, κάνει πολύ ευκολότερη τη ζωή του χρήστη, μιας και σε αντίθετη περίπτωση (μη χρήση μηχανικής μάθησης) ο χρήστης θα έπρεπε να διαβάσει όλα τα δεδομένα να σημειώσει και να κατηγοριοποιήσει όλα τα στοιχεία που έχει στη διάθεση του. Αυτό το οποίο πραγματοποιεί η μηχανική μάθηση και οι αλγόριθμοι της, είναι να βρίσκουν τα κοινά σημεία και τα μοτίβα μεταξύ πολλών δεδομένων προκειμένου να μπορούν στη συνέχεια να τα αναλύσουν. Παράλληλα, μετατρέπουν τις πολλές πληροφορίες σε μια απλή και κατανοητή μορφή για το χρήστη, η οποία είναι έτοιμη για επεξεργασία. Προκειμένου ο χρήστης να έχει τα σωστά αποτελέσματα, θα πρέπει να «προπονήσει» ένα μέρος των δεδομένων προκειμένου να μπορεί να έχει τα ανάλογα αποτελέσματα.

Γιατί όμως θα πρέπει να επιλέξουμε τη μηχανική μάθηση σε σύγκριση με άλλες τεχνικές πρόβλεψης απατών; Παρακάτω παραθέτω κάποια από τα πλεονεκτήματα της μηχανικής μάθησης.

- *Ταχύτητα.* Με τους έως τώρα απλούς κανόνες, πολλές διεργασίες πραγματοποιούνταν πολύ αργά ενώ ο χρήστης έπρεπε επίσης να πραγματοποιήσει χειροκίνητα ενέργειες προκειμένου να έχει κατάλληλα αποτελέσματα. Με την μηχανική μάθηση αυτό το πρόβλημα λύνεται, μιας και οι αλγόριθμοι μαζεύουν τις απαραίτητες πληροφορίες, ενώ με το κατάλληλο training set, οι αλγόριθμοι της μηχανικής μάθησης μπορούν να αποδειχθούν αρκετά αποτελεσματικοί, ακριβείς και ταχύτατοι. Όλα αυτά εξοικονομούν πολύτιμο χρόνο στον χρήστη μιας και αυτός δεν χρειάζεται να κάνει κάτι, ενώ οι διαδικασίες πραγματοποιούνται σε πολύ γρηγορότερο χρόνο συγκριτικά με τους απλούς κανόνες.
- *Κλίμακα.* Η μηχανική μάθηση μας παρέχει ευελιξία σε σχέση με τους απλούς κανόνες αφού γι' αυτούς, ο μεγάλος όγκος δεδομένων αποτελούσε αρκετά μεγάλο πρόβλημα. Με τη μηχανική μάθηση, τα παραπάνω δεδομένα είναι πολύτιμα αφού όσο τα δεδομένα για το training set αυξάνονται, τόσο πιο ακριβές είναι το τελικό μας αποτέλεσμα. Μέσω της μηχανικής μάθησης, έχουμε τη δυνατότητα να επεξεργαζόμαστε τεράστιο όγκο πληροφοριών, κάτι που δε θα μπορούσαμε να βρούμε στους απλούς κανόνες AI.

- Αποδοτικότητα. Είναι προφανές ότι ένα σύστημα αλγορίθμων μπορεί να αποδώσει πολύ καλύτερα και σε μικρότερο χρονικό διάστημα, ειδικότερα αν ο χρήστης έχει να επεξεργαστεί μεγάλο όγκο δεδομένων και στη συνέχεια να ταξινομήσει ανάλογα. Παράλληλα, μπορεί να εντοπίσει διάφορα μοτίβα τα οποία ο άνθρωπος δεν θα μπορούσε να βρει τα οποία παίζουν πολύ μεγάλο ρόλο στις απάτες πιστωτικών καρτών.
- Μεγαλύτερη ακρίβεια. Η μηχανική μάθηση μπορεί να επιφέρει πολύ καλύτερα αποτελέσματα τα οποία κάποιες φορές είναι καλύτερα και από αυτά των ανθρώπων. Αυτό συμβαίνει γιατί με την μηχανική μάθηση, ελέγχονται πολύ περισσότερα κριτήρια και μοτίβα ταυτόχρονα, σίγουρα πολύ περισσότερα απ' όσα θα μπορούσε να ελέγξει ένας άνθρωπος, ενώ παράλληλα τα μοντέλα αυτά έχουν την ιδιότητα να μαθαίνουν από τις πληροφορίες που τους παρέχονται.

Η μηχανική μάθηση σύμφωνα με αρκετές αναλύσεις και μελέτες είναι η κατάλληλη μέθοδος για την αντιμετώπιση και την πρόβλεψη απατών αφού εκτός των άλλων, το πιο σημαντικό χαρακτηριστικό της, είναι η ακρίβεια στα αποτελέσματα τα οποία παράγει, ξεπερνώντας πολλές φορές και τις επιδόσεις ανθρώπων. Αυτό συμβαίνει όπως προανέφερα λόγω της δυνατότητας “εξάσκησης” μέσα από μεγάλο όγκο δεδομένων. Όσο και αν μας φαίνεται παράξενο, υπάρχουν πολλές δραστηριότητες τις οποίες τα μοντέλα μηχανικής μάθησης μπορούν να φέρουν εις πέρας πολύ πιο αποτελεσματικά ακόμα και από κορυφαίους αναλυτές. Ουσιαστικά, ο τομέας που υπερτερεί η μηχανική μάθηση, είναι η διαφορά στην ταχύτητα επεξεργασίας των πληροφοριών σε σχέση με τον όγκο τους ενώ παράλληλα, φαίνεται πως τα τελευταία χρόνια έχει υπάρξει τεράστια πρόοδος στην αναγνώριση εικόνων από μοντέλα μηχανικής μάθησης. Ιδιαίτερα σε αυτόν τον τομέα, από το 2012 μέχρι σήμερα, το ποσοστό λάθους στην αναγνώριση εικόνων, έχει πέσει στο 3,58% ενώ μια ανάλυση από ειδικούς θα μπορούσε να έχει 5% ποσοστό λάθους (Bayes error).

Η δυνατότητα των μοντέλων να αναγνωρίζουν συγκεκριμένα μοτίβα τα οποία μπορεί να μην παρατηρούσε ένας άνθρωπος, είναι μια ακόμα σημαντική δυνατότητα των μοντέλων μηχανικής μάθησης. Παράλληλα, σε συνδυασμό με την δυνατότητα των μοντέλων να διαχειρίζονται μεγάλο όγκο δεδομένων και να “μαθαίνουν” από το παρελθόν και από τα υπάρχοντα δεδομένα, η μηχανική μάθηση φαίνεται να είναι ένα πολύ καλό εργαλείο στην καταπολέμηση του προβλήματος.

2.7 Μηχανική μάθηση και ηθική

Έως τώρα έχουμε δει και αναλύσει αρκετά την μηχανική μάθηση και τα πλεονεκτήματα της, παρατηρώντας πόσο πολύ μπορεί να βοηθήσει ιδιαίτερα στην αντιμετώπιση των απατών πιστωτικών καρτών. Όμως ένα ακόμη ζήτημα το οποίο κυριαρχεί και απασχολεί μεγάλη μερίδα

ανθρώπων, είναι η ηθική γύρω από τη μηχανική μάθηση και γενικότερα την τεχνητή νοημοσύνη. Πολλοί είναι αυτοί οι οποίοι θεωρούν την τεχνητή νοημοσύνη ως καταστροφή για τον άνθρωπο, αφού όπως λένε θα αντικαταστήσει τον άνθρωπο σε πολλούς τομείς της καθημερινότητας, έχοντας καταστροφικές επιπτώσεις. Από την άλλη πλευρά, πολλοί υποστηρίζουν ότι η μηχανική μάθηση και η τεχνητή νοημοσύνη μπορεί να βοηθήσει σημαντικά σε πολλούς τομείς της καθημερινότητας μας, αφού η τεχνολογία συνεχώς εξελίσσεται και προχωρά, με αποτέλεσμα την εξέλιξη της τεχνητής νοημοσύνης σε τομείς όπως η απλή καθημερινότητα, οι επιχειρήσεις αλλά ακόμα και η ιατρική. Έχουμε λοιπόν δύο αντικρουόμενες απόψεις, με την καχυποψία να κυριεύει την μία πλευρά και την απόλυτη εμπιστοσύνη να κυριαρχεί από την άλλη. Η πραγματικότητα στο συγκεκριμένο ζήτημα ίσως κρύβεται κάπου στη μέση. Τα μοντέλα μηχανικής μάθησης χρησιμοποιούν κυρίως πραγματικά δεδομένα ως training data προκειμένου να καταφέρουν να έχουν όσο το δυνατόν μικρότερο ποσοστό λάθους. Χρησιμοποιώντας το παράδειγμα της μηχανικής μάθησης για την εύρεση απατών στις πιστωτικές κάρτες, υποθέτουμε πως στα δεδομένα αυτά, πολύ πιθανόν να βρίσκονται οι συναλλαγές μας μέσω της πιστωτικής μας κάρτας αλλά και διάφορες πληροφορίες σχετικά με αυτές. Η μερίδα των ανθρώπων που τάσσεται κατά της μηχανικής μάθησης και της τεχνητής νοημοσύνης, ουσιαστικά απαιτεί να μην χρησιμοποιούνται τα προσωπικά του δεδομένα για κανένα λόγο, ακόμα και αν πρόκειται να χρησιμοποιηθούν έμμεσα για την δική τους ασφάλεια. Από την άλλη, υπάρχουν πολλοί ακόμα που δεν διαφωνούν με τη χρήση των δεδομένων τους για τέτοιους σκοπούς. Για να υπάρξει λοιπόν μια ισορροπία, θα ήταν καλό να υπάρχουν κάποιες συγκεκριμένες και προκαθορισμένες συνιστώσες στον τρόπο και τον λόγο τον οποίο χρησιμοποιούνται τα προσωπικά μας δεδομένα. Ένα ακόμα βασικό και πολύ σημαντικό ζήτημα που αποτελεί μεγάλο πρόβλημα της μηχανικής μάθησης, είναι το ζήτημα της προκατάληψης. Ένα πολύ απλό και καθημερινό παράδειγμα αποτελεί η μέχρι πρότινος χρήση μόνο γυναικείων φωνών σαν βοηθούς στα κινητά τηλέφωνα, το οποίο άλλαξε πολύ πρόσφατα μετά τη δυσαρέσκεια μεγάλης μερίδας αγοραστών, αφού θεωρούν ως λανθασμένο στερεότυπο τη χρήση της γυναικείας φωνής ως βοηθό. Υπάρχουν πολλά ακόμη τέτοια παραδείγματα αφού τα στερεότυπα και οι προκαταλήψεις δεν παύουν να ισχύουν από τη στιγμή που το μοντέλο “προπονείται” από πραγματικά δεδομένα που εισάγει ο άνθρωπος. Υπάρχει όμως ένα ακόμη σενάριο, το οποίο είναι ακόμη χειρότερο από το πρόβλημα των προκαταλήψεων, το οποίο δεν είναι άλλο από το ζήτημα της σκοπιμότητας. Για άλλη μια φορά, τα παραδείγματα είναι πολλά, όμως θα χρησιμοποιήσουμε ένα παράδειγμα στον τομέα της ιατρικής, ο οποίος είναι καίριας σημασίας. Εάν υποθέσουμε πως πραγματοποιούμε μια εξέταση και το αποτέλεσμα εξάγεται μέσω μοντέλου μηχανικής μάθησης, δεν θα μπορούμε να είμαστε σίγουροι πως το αποτέλεσμα που έβγαλε το μοντέλο, είναι η πραγματική λύση και όχι μία απλή λύση την οποία θα μπορούσε να εκμεταλλευτεί κάποιος οικονομικά με διάφορους τρόπους. Στη μηχανική μάθηση λοιπόν, τα πάντα ξεκινούν από τα δεδομένα με τα οποία θα εκπαιδευτεί κάθε μοντέλο. Εάν καταφέρουμε

να έχουμε ένα αξιόπιστο training set δίχως προκαταλήψεις και σκοπιμότητες, τότε η μηχανική μάθηση μπορεί να μας βοηθήσει πραγματικά. Πλέον, μέσω της μηχανικής μάθησης, επηρεάζονται σε ένα βαθμό οι συνήθειες μας αφού καθημερινά συναντάμε προτάσεις που αφορούν την ταινία που θα παρακολουθήσουμε, την μουσική που θα ακούσουμε αλλά και το κατάλληλο μέρος για να απολαύσουμε τον καφέ μας. Καταλαβαίνουμε λοιπόν πόσο πολύ εμπλέκεται η μηχανική μάθηση στη ζωή μας, όταν όμως υπάρχει η συνεχή παρουσία της σε ότι κάνουμε, θα πρέπει και εμείς να είμαστε πιο προσεκτικοί. Ουσιαστικά, θα πρέπει τα μοντέλα να μας βοηθούν να πάρουμε μια απόφαση και όχι να την παίρνουν αυτά για εμάς καθώς και να μην ασκούν επιρροή στη βούληση μας. Η μηχανική μάθηση αποτελεί μεγάλη καινοτομία και μπορεί πραγματικά να βοηθήσει σε πολλούς τομείς της καθημερινότητας, θα πρέπει όμως και εμείς να κρίνουμε και να πράττουμε σωστά προκειμένου να συνεχίσει να αποτελεί εργαλείο που θα μας βοηθήσει και όχι να αποφασίζει αντί για εμάς. [9]

Κεφάλαιο 3 “Βιβλιογραφική ανασκόπηση”

3.1 Εισαγωγή και ανάλυση εργασιών

Όπως έχουμε δει και παραπάνω, η μηχανική μάθηση αποτελεί ένα πολύ καλό και χρήσιμο εργαλείο για την ανίχνευση απατών σε πιστωτικές κάρτες. Γι' αυτό το λόγο, πολλές είναι οι έρευνες και εργασίες οι οποίες έχουν αφιερωθεί σε αυτόν τον τομέα. Προκειμένου όμως να μπορέσω να εντυφίσω παραπάνω στο συγκεκριμένο θέμα, θα αναλύσω κάποιες εργασίες οι οποίες θεωρώ πως μπορούν να προσφέρουν παραπάνω γνώση επί του θέματος.

Στην πρώτη εργασία [10] που θα δούμε, οι συγγραφείς ασχολούνται με τις τεχνικές τις οποίες χρησιμοποιούν οι επίδοξοι απατεώνες ενώ παράλληλα αναλύει τις τεχνικές με τις οποίες εντοπίζονται οι απάτες μέσω των πιστωτικών καρτών και πως η μηχανική μάθηση βοηθά στην εντοπισμό τους. Αρχικά λοιπόν, καθορίζονται οι τύποι απατών που υπάρχουν προκειμένου να μπορούμε να τους κατηγοριοποιήσουμε. Έχουμε λοιπόν τις εξής :

- Απάτες ασφάλειας, όπου οι εταιρείες εξαπατούν τους πελάτες τους με διάφορους τρόπους.
- Απάτες που αφορούν μετοχές και αξία όπως επενδύσεις υψηλών αποδόσεων, το σχήμα Πυραμίδα, συναλλαγή σε ξένο νόμισμα, επένδυση σε δήθεν ανερχόμενες μετοχές κλπ.
- Ξέπλυμα χρήματος, όπου οι απατεώνες προσπαθούν να δικαιολογήσουν τα λαθραία χρήματα που έχουν στη διάθεση τους.
- Υψηλή απόδοση επένδυσης, κάτι το οποίο συνήθως γίνεται από εταιρείες προκειμένου να παρουσιάζονται ως κερδοφόρες, να απορροφούν το ενδιαφέρον επενδυτών, να έχουν φοροαπαλλαγές, ενώ τα στοιχεία που παρουσιάζει να είναι πλαστά.

Στη συνέχεια ο συγγραφέας μας περιγράφει τις τεχνικές με τις οποίες μπορεί να πραγματοποιηθεί η ανίχνευση απάτης στις πιστωτικές κάρτες. Τέτοιες τεχνικές είναι οι παρακάτω :

- Περιγραφικές ή μη εποπτευόμενες τεχνικές, όπου αναλύονται τα περιγραφικά μοντέλα ως προς τη δομή τους, τις σχέσεις και το τρόπο που συνδέονται μεταξύ τους.
- Τεχνικές πρόβλεψης, όπου σκοπός του είναι το χτίσιμο ενός αναλυτικού μοντέλου ώστε να μπορεί να προβλέπει τα σημεία τα οποία θέλουμε. Σε αυτές τις τεχνικές πρόβλεψης, οι συνηθέστερες είναι η Support Vector Machine, η Naive Bayes και η K-Nearest Neighbor, αφού χρησιμοποιούνται ευρέως.
- Τεχνικές Τεχνητής και Υπολογιστικής Νοημοσύνης, όπου αυτό το κομμάτι

ασχολείται με την σύνδεση των δύο τομέων, αναλύοντας το πως και με ποιο τρόπο μπορεί να βοηθήσει η υπολογιστική δύναμη σε απλά προβλήματα της καθημερινότητας. Οι πιο γνωστές και χρήσιμες τεχνικές είναι οι Genetic Algorithm, Scatter Search, Artificial Immune System κλπ.

Ο συγγραφέας του κειμένου μας παρουσιάζει κάποιες ακόμα έννοιες οι οποίες θεωρούνται υβριδικά μοντέλα και χρησιμοποιούνται σε δυσκολότερες καταστάσεις. Τέτοια μοντέλα είναι το Dempster Shafer Theory, Computational fraud detection model και Locally Weighted Learning. Στη συνέχεια, βλέπουμε μια ανάλυση η οποία σχετίζεται με τον τύπο απάτης ενώ επισημαίνεται ο κατάλληλος τρόπος για να εντοπιστεί.

- Για τον εντοπισμό απάτης ασφάλειας, στο συγκεκριμένο παράδειγμα χρησιμοποιήθηκε ένα μοντέλο λογιστικής παλινδρόμησης και φάνηκε να δίνει κάποια καλά αποτελέσματα.
- Για τον εντοπισμό απάτης πιστωτικής κάρτας, χρησιμοποιήθηκαν πολλοί διαφορετικοί αλγόριθμοι, μιας και υπάρχει μεγάλη γκάμα αλγορίθμων ενώ κάθε ένας αποδίδει καλύτερα για συγκεκριμένες περιπτώσεις.
- Για τον εντοπισμό απατών οικονομικής φύσεως, και πάλι χρησιμοποιήθηκαν αρκετοί διαφορετικοί αλγόριθμοι, όπως για παράδειγμα τον Naive Bayes, τον CART και τον FFS

Τέλος, ο συγγραφέας μας παρουσιάζουν την έρευνα την οποία διεξήγαγαν προκειμένου να δουν ποιος είναι ο αποτελεσματικότερος αλγόριθμος και σε ποια περίπτωση, τα αποτελέσματα των οποίων κατέγραψαν. Για τις απάτες οικονομικής φύσεως, ο καταλληλότερος θα είναι ο Pnn(Probabilistic neural network) με ποσοστό ακρίβειας στο 98.09%, ενώ για την ανίχνευση απατών, κατάλληλος είναι ο Bayes αλλά και ο SVM με πολύ υψηλό ποσοστό ακρίβειας (99,02%, 98,8%). Σύμφωνα με τα συμπεράσματα των συγγραφέων, το υβριδικό μοντέλο όπου χρησιμοποιούνται παλιοί και νεότεροι αλγόριθμοι είναι το κατάλληλο, παρόλα αυτά τα μοντέλα έχουν πολλά περιθώρια εξέλιξης, αφού οι απαραίτητες διαδικασίες για την ανίχνευση απατών πρέπει να γίνονται σε πραγματικό χρόνο.

Στην επόμενη εργασία [11] οι συγγραφείς της έρευνας αρχικά, μας περιγράφουν το μεγάλο πρόβλημα που αντιμετωπίζεται με τις απάτες των πιστωτικών καρτών, αφού αποτελούν τον πιο συνηθισμένο τύπο απάτης με 133.015 περιστατικά να καταγράφονται το 2017. Τα δεδομένα που προσφέρονται όμως από τις πιστωτικές κάρτες είναι αρκετά ανισόρροπα, πράγμα που κάνει την προσπάθεια εντοπισμού και ανίχνευση τους πολύ δυσκολότερη. Τα τελευταία χρόνια γίνεται μια προσπάθεια από τις τράπεζες για την χρησιμοποίηση της EMV κάρτας προκειμένου να κρατούνται περισσότερα στοιχεία και να είναι ευκολότερη η διαδικασία εντοπισμού του απατεώνα. Οι συγγραφείς λοιπόν, προκειμένου να διακρίνουν τον καλύτερο τρόπο εντοπισμού των απατών χρησιμοποιούν Supervised αλλά και Semi-supervised τεχνικές. Τα διάφορα μοντέλα όπως τα Δέντρα αποφάσεων, ο αλγόριθμος Naive Bayes, ο Logistic

Regression χρησιμοποιούνται κυρίως για την ανίχνευση απατών σε πραγματικό χρόνο. Παρόλα αυτά, πολλές φορές λόγω της ανωμαλίας των δεδομένων, μπορεί τα μοντέλα να αποδειχθούν λανθασμένα, γι' αυτό και χρησιμοποιείται ένα μοντέλο κωδικοποίησης το οποίο μπορεί να δημιουργήσει συναλλαγές προκειμένου να βρεθούν οι ανωμαλίες σε σύγκριση με τα δεδομένα που έχουμε. Τα δεδομένα, έχουν ακόμα ένα χαρακτηριστικό που δυσκολεύει την σωστή πρόβλεψη των μοντέλων το οποίο είναι η διαφορετικότητα και η μη ομοιότητα κάθε συναλλαγής με την προηγούμενη (Concept drift). Αυτό έχει ως αποτέλεσμα την υψηλή ανισορροπία στα δεδομένα, κάτι που δεν βοηθάει στην σωστή πρόβλεψη των μοντέλων. Στη συνέχεια, μας δίνεται η περιγραφή ενός data set, κατά το οποίο ένα χρήστης πραγματοποίησε συναλλαγές σε διάστημα δύο ημερών. Τις δύο αυτές ημέρες πραγματοποιήθηκαν 284,807 συναλλαγές ενώ οι 492 (0.172%) ήταν απάτες. Το σύνολο δεδομένων είναι αρκετά ανισόρροπο, ενώ τα περισσότερα χαρακτηριστικά έχουν αντικατασταθεί με χαρακτηριστικές μεταβλητές V1,V2,V3...V28, προκειμένου να μην τίθεται θέμα προσωπικών δεδομένων και απορρήτου. Στη συνέχεια, οι συγγραφείς αναλύουν τη μεθοδολογία που ακολουθείται προκειμένου να αντιμετωπίσουμε το Concept drift.

- Αρχικά, χωρίζουμε τους κατόχους καρτών σε ομάδες με βάση τον αριθμό συναλλαγών τους
- Χρησιμοποιώντας τη μέθοδο Sliding-Window, συγκεντρώνουμε τις συναλλαγές σε ομάδες, βάση συμπεριφοράς των κατόχων καρτών(π.χ. ανάληψη μεγάλων ποσών, ανάληψη μικρότερων ποσών κλπ)
- Εφαρμόζοντας τη μέθοδο Sliding-Window, βρίσκουμε την κατάλληλη ομάδα για τον κάθε κάτοχο, ενώ στη συνέχεια θα εκπαιδεύσουμε τους κατάλληλους ταξινομητές (classifiers) για κάθε ομάδα ξεχωριστά, σύμφωνα και πάλι με τη συμπεριφορά των κατόχων. Οι ταξινομητές είναι πολύ πιθανό να μη δουλέψουν σωστά λόγω της μεγάλης ανισορροπίας.
- Στη συνέχεια, χρησιμοποιούμε την τεχνική SMOTE
- Προκειμένου να αντιμετωπίσουμε την ανισορροπία, μπορούμε να συσχετίσουμε τον ταξινομητή στο αρχικό σύνολο δεδομένων ή να χρησιμοποιήσουμε ταξινομητές μιας κατηγορίας. Έτσι, έχουμε το αποτέλεσμα για την συμπεριφορά κάθε κατόχου.
- Τέλος, το αποτέλεσμα που θα πάρουμε θα είναι και η λύση στο πρόβλημα του Concept drift.

Οι συγγραφείς, μας παρουσιάζουν στο τέλος ένα παράδειγμα στο οποίο χρησιμοποιεί τις τεχνικές Matthews Correlation Coefficient (MCC), SMOTE και One-Class SVM προκειμένου να ελεγχθεί η ισορροπία των δύο ταξινομητών, παίρνοντας τελικά θετικά αποτελέσματα, αφού με αυτές τις τεχνικές είναι εφικτή η ισορροπία των συνόλων δεδομένων, ενώ φαίνεται οι αλγόριθμοι Logistic regression (97,18%), decision tree (97,08%) και random

forest (99,98%) είναι οι αλγόριθμοι που μας έδωσαν τα καλύτερα αποτελέσματα ως προς την ακρίβεια του συστήματος (χρησιμοποιώντας την τεχνική SMOTE).

Στην επόμενη εργασία [12] απάτες πιστωτικών καρτών όπως έχουμε δει αποτελεί ένα μεγάλο πρόβλημα αφού τα χρηματικά ποσά που χάνονται λόγω αυτών, είναι τεράστια. Χαρακτηριστικά, στη Βόρεια Αμερική, υπολογίζονται περίπου στα 3.5 δισεκατομμύρια το 2012, όπου αποτελεί το 0,9% των συνολικών εσόδων της, ενώ όταν πρόκειται για συναλλαγές από παραγγελίες εκτός της Βόρειας Αμερικής, το ποσοστό είναι σχεδόν διπλάσιο αφού αγγίζει το 1,6%. Οι απάτες μέσω πιστωτικών καρτών πολλές φορές πραγματοποιούνται κατά την πληρωμή μιας παραγγελίας. Γι' αυτό το λόγο, οι έμποροι έχουν επίσης ένα μερίδιο ευθύνης προκειμένου να μην πραγματοποιηθεί κάποια απάτη, αφού θα πρέπει να δημιουργήσουν ένα ασφαλές περιβάλλον πληρωμών. Όλα αυτά τα χρόνια, γίνονται προσπάθειες μέσω διαφόρων τεχνολογιών προκειμένου να αντιμετωπιστεί το συγκεκριμένο πρόβλημα. Παρόλα αυτά, ο εντοπισμός των απατών δεν είναι καθόλου εύκολη διαδικασία, μιας και οι απατεώνες συνεχώς προσαρμόζονται και προσπαθούν να βρουν καινούργιους τρόπους απόσπασης χρημάτων, αλλάζοντας συνεχώς τακτικές. Προκειμένου λοιπόν να υπάρξει μια λύση σε αυτό το πρόβλημα, χρησιμοποιούμε την μηχανική μάθηση, επικεντρώνοντας την προσοχή μας στους Supervised Αλγόριθμους, όπως τον Navi Bayes, Logistic regression κλπ. Η συγκεκριμένη έρευνα, αναλύει στη συνέχεια ένα σύστημα εντοπισμού απάτης το οποίο εφαρμόστηκε σε μια etail εταιρεία (διαδικτυακών παραγγελιών). Το συγκεκριμένο σύστημα ερευνά το συνδυασμό μη αυτόματης και αυτόματης ταξινόμησης, βασισμένο σε αληθινά δεδομένα που προέρχονται από την εταιρεία ενώ στο τέλος θα συγκριθεί με τις έως τώρα πρακτικές που χρησιμοποιούνται από την εταιρεία. Προκειμένου να συμβεί αυτό και να υπάρξει το επιθυμητό αποτέλεσμα, ακολουθήθηκαν συγκεκριμένα βήματα. Αρχικά, με τη βοήθεια των υπεύθυνων και μέσα από συνεντεύξεις με τους εμπλεκόμενους, βοήθησαν στην κατανόηση των διαδικασιών και της λειτουργίας της εταιρείας. Μετά από το πρώτο βήμα, ακολούθησε η καταγραφή των δεδομένων των παραγγελιών, ενώ στη συνέχεια πραγματοποιήθηκε η απαραίτητη επεξεργασία των δεδομένων προτού προχωρήσουν στο επόμενο βήμα, το οποίο είναι η εισαγωγή δεδομένων στους αλγορίθμους Logistic regression, Random forest και Support Vector. Στη συνέχεια, περιγράφεται η εταιρεία και πως αυτή λειτουργεί μέσω μιας μελέτης περίπτωσης, κατά την οποία, η εταιρεία είναι μια από τις κορυφαίες στην διαδικτυακή λιανική πώληση προϊόντων μόδας πολυτελείας. Γι' αυτό το λόγο, χρειάζεται έναν αποτελεσματικό τρόπο να αντιμετωπίσει τις απάτες αφού έχει έναν μεγάλο αριθμό ζήτησης προϊόντων. Μέχρι τώρα, τα αιτήματα ζήτησης επαληθεύονταν από την υπεύθυνη ομάδα και όλα λειτουργούσαν χειροκίνητα, κάτι το οποίο η εταιρεία θα ήθελε να αλλάξει την διαδικασία με την αυτοματοποίηση της διαδικασίας (σε ποσοστό 80%), διατηρώντας παράλληλα την ασφάλεια των συναλλαγών (ποσοστό απάτης κάτω του 4.5%). Το σύστημα το οποίο προτάθηκε, λειτουργεί με ένα δείκτη ο οποίος βάσει των δεδομένων που παρέχονται θα αποφασίζει αν μια συναλλαγή φαίνεται ύποπτη προκειμένου να

αναθεωρηθεί. Ο συγγραφέας στη συνέχεια μας δίνει κάποια στοιχεία για τα supervised και unsupervised method, αφού στη συγκεκριμένη έρευνα χρησιμοποιήθηκαν και οι δύο μέθοδοι. Προχωρώντας στην υλοποίηση του συστήματος, ο συγγραφέας περιγράφει τα δεδομένα που χρησιμοποίησε αλλά και το λόγο που το έκανε, ενώ περιγράφει και τα προβλήματα τα οποία αντιμετώπισε. Το δείγμα που χρησιμοποίησε για testing, προήλθε από το 20% των πρόσφατων πωλήσεων και στη συνέχεια περιγράφονται αναλυτικά οι μεταβλητές που χρησιμοποιήθηκαν με τις κατάλληλες προσθήκες και προσαρμογές. Προκειμένου να βρεθεί όμως ποιος αλγόριθμος λειτουργεί καλύτερα, τα δεδομένα χωρίστηκαν προκειμένου να γίνει το cross validation έτσι ώστε να μπορούμε να μετρήσουμε την απόδοση των αλγορίθμων. Χρησιμοποιώντας ένα 10-fold validation στους 3 διαφορετικούς αλγορίθμους έχουμε :

- Random forest, λόγω των περιορισμένων δυνατοτήτων των “κλαδιών”, ο αλγόριθμος δεν μπορεί να προσφέρει πολλά πράγματα αφού η απόδοση του είναι στάσιμη.
- Support Vector Machines, αποτελεί προηγμένη μέθοδο ταξινόμησης, όμως και πάλι δεν υπήρξαν τα απαραίτητα αποτελέσματα αφού για $C=10$ (βέλτιστα αποτελέσματα), η τιμή του αποτελέσματος είναι αρκετά χαμηλή.
- Logistic Regression, λόγω της περιορισμένης μνήμης, δεν ήταν δυνατή η περαιτέρω εκπαίδευση του αλγορίθμου σε δεύτερη και τρίτη αλλαγή παραγγελίας. Η λύση που βρέθηκε ήταν η χρησιμοποίηση του αλγορίθμου Broyden-Fletcher-Goldfarb-Shanno προκειμένου να βελτιωθεί η μνήμη.

Τα αποτελέσματα του cross-validation μας δείχνουν ότι ο αλγόριθμος random forest απέδωσε καλύτερα σε σχέση με τους άλλους, αφού οι άλλοι δύο αλγόριθμοι απέδωσαν περίπου το ίδιο. Στη διαδικασία της δοκιμής, ο αλγόριθμος παρουσίασε μειωμένη απόδοση κάτι που ήταν αναμενόμενο από τους ερευνητές λόγω των προβλημάτων που αναφέρθηκαν παραπάνω. Στη συνέχεια δοκιμάζονται διάφορες προσεγγίσεις όπου κατά την πρώτη προσέγγιση, έχουμε την αυτόματη αποδοχή ή απόρριψη παραγγελιών σύμφωνα με το πόσο ύποπτη είναι, ενώ κατά την άλλη προσέγγιση όλες οι παραγγελίες οι οποίες έχουν ένα συγκεκριμένο score μεγαλύτερου του ορίου που έχει τεθεί, θα γίνονται αποδεκτές ενώ όσες δεν περνούν το όριο θα εξετάζονται χειροκίνητα. Τέλος, προκειμένου να ελεγχθεί το συγκεκριμένο σύστημα στην πράξη, οι υπεύθυνοι το έθεσαν σε λειτουργία εξετάζοντας 50 παραγγελίες που ζητήθηκαν, όπου και αναλύοντας τα στοιχεία κατέληξαν σε ένα συγκεκριμένο αποτέλεσμα (score) και με βάση αυτό κατέταξαν την παραγγελία σαν ύποπτη ή μη ύποπτη με τα αποτελέσματα να φαίνονται αρκετά ενθαρρυντικά για χρήση ενός τέτοιου συστήματος στο μέλλον.

Στην επόμενη εργασία [13], ο συγγραφέας ουσιαστικά με την ανάλυση πολλών διαφορετικών ερευνών, προσπαθεί να δώσει κατευθυντήριες γραμμές προκειμένου να βοηθήσει με τον τρόπο του την επιστημονική κοινότητα να βρει τρόπους ώστε να αντιμετωπιστεί το

πρόβλημα απατών σε πιστωτικές κάρτες το οποίο έχει κοστίσει πολλά δισεκατομμύρια. Επισημαίνει μάλιστα πως πολλές φορές οι απάτες αυτές σχετίζονται με άλλα εγκλήματα όπως εμπόριο ναρκωτικών, τρομοκρατία και εμπόριο όπλων, ώστε να καταλάβουμε πόσο σοβαρή είναι η κατάσταση. Με την εξέλιξη της τεχνολογίας, εξελίσσονται και οι τρόποι με τους οποίους πραγματοποιούνται οι απάτες πιστωτικών καρτών, αφού συνεχώς οι απατεώνες βρίσκουν καινούργια μέσα αλλά και ευκαιρίες. Το έργο αυτής της εργασίας, είναι η ανάδειξη συγκεκριμένων τεχνικών προκειμένου να αντιμετωπιστεί το πρόβλημα απατών μέσω πιστωτικών καρτών, αλλά και να δούμε ποια είναι η καταλληλότερη μέθοδος. Τα στοιχεία που χρησιμοποιούνται στην έρευνα προέρχονται από το 2017 και είναι πραγματικά στοιχεία συναλλαγών. Αρχικά, έχουμε μια περιγραφή και μια ιστορική αναδρομή στο γενικότερο ζήτημα των πληρωμών, μιας και μέσω αυτών πραγματοποιούνται οι απάτες. Το συγκεκριμένο ζήτημα των απατών σε πιστωτικές κάρτες, έχει πάρει μεγαλύτερες διαστάσεις τα τελευταία χρόνια μιας και έχουμε μεγάλη αύξηση των διαδικτυακών παραγγελιών αλλά και των ανέπαφων συναλλαγών γενικότερα. Παράλληλα, τα τελευταία χρόνια έχει προκύψει μια πολύ λανθασμένη άποψη για τους υπεύθυνους των απατών, μιας και οι πράξεις τους δεν θεωρούνται τόσο εγκληματικές από το ευρύ κοινό, αφού οι περισσότεροι πιστεύουν ότι κάτι τέτοιο επηρεάζει μόνο τις μεγάλες τράπεζες και τις κυβερνήσεις, πράγμα όμως που δεν ισχύει σε καμία περίπτωση. Μια τέτοιου είδους απάτη θα έπρεπε να θεωρείται το ίδιο εγκληματική με μια κανονική απάτη ή κλοπή. Στη συνέχεια, ο συγγραφέας μας αναλύει τον τρόπο και τη σειρά με την οποία πραγματοποιείται μια πληρωμή και το πως λειτουργούν τα FMS (Fraud Management System) όπου έχουμε την φραγή των συναλλαγών οι οποίες θεωρούνται ύποπτες ως απάτη. Πολλές από τις μεθόδους οι οποίες ακολουθούνται, αποτελούν μυστήριο για τους ερευνητές, αφού προτιμάται η μη δημοσιοποίηση της μεθοδολογίας την οποία ακολουθούν. Παράλληλα, τα έξοδα για την συνεχή παρακολούθηση ύποπτων περιπτώσεων είναι αρκετά μεγάλα, ενώ το πρόβλημα έλλειψης μεγάλου όγκου δεδομένων για έρευνα, είναι επίσης κάτι που δεν βοηθάει τους ερευνητές. Στη συνέχεια, ο συγγραφέας αναλύει τον τρόπο με τον οποίο πραγματοποιείται ο εντοπισμός των απατών (όσο επιτρέπεται να λόγω της φύσης των δεδομένων) μέσω των συστημάτων ενώ αναφέρει τη μεθοδολογία και τον τρόπο υπολογισμού των μετρήσεων. Παράλληλα, παρατίθενται παραδείγματα από υπολογισμούς που πραγματοποιήθηκαν υπό αληθινές συνθήκες χρησιμοποιώντας πραγματικά δεδομένα ενώ αφιερώνει ένα μεγάλο κομμάτι της έρευνας σε διάφορα στοιχεία τα οποία προκύπτουν από αυτές τις μετρήσεις. Ένα επίσης σημαντικό κομμάτι το οποίο αναλύεται σε αυτή την έρευνα είναι η προσπάθεια για εντοπισμό απατών σε πραγματικό χρόνο, κάτι το οποίο θα βοηθούσε πολύ στην καταπολέμηση του προβλήματος, μιας και πολλές φορές τα αποτελέσματα μιας έρευνας μπορεί να καθυστερήσουν, με ότι αυτό συνεπάγεται. Γι' αυτό το λόγο ο συγγραφέας, αναλύει συγκεκριμένες έρευνες τις οποίες έχει αντλήσει από έγκυρες επιστημονικού περιεχομένου ιστοσελίδες, μέσα από τις οποίες εξάγει συμπεράσματα προκειμένου να βοηθήσει στην ανάπτυξη μεθόδων

καταπολέμησης των απατών. Έτσι λοιπόν, στη συνέχεια αναλύονται εργασίες και έρευνες που έχουν γίνει ξεχωριστά για : Expert system/Decision Trees, Rule Based, Decision Tree/Random Forest, Case-Based Reasoning. Μετά την ανάλυση των προαναφερθέντων, ο συγγραφέας προχωράει σε μια ανάλυση για τα νευρωνικά δίκτυα(Supervised και παρακάτω Unsupervised), δίνοντας επίσης τις διαφορετικές πτυχές τους και πάλι μέσω διαφόρων εργασιών και ερευνών. Προκειμένου να καλύψει όλη την ευρύτερη περιοχή, ο συγγραφέας επιλέγει να αναλύσει επίσης τα παρακάτω: Bayesian Network, Εξελικτική υπολογιστική, το Κρυφό μοντέλο Markov(HMM) Support Vector Machine, Eclectic Methodology. Οι παραπάνω εργασίες και έρευνες δεν αναλύθηκαν για οικονομία χώρου και χρόνου. Τέλος, ο συγγραφέας αναφέρει στα συμπεράσματα του, πως χρειάζεται περαιτέρω έρευνα πάνω στο συγκεκριμένο θέμα προκειμένου να περιοριστεί ή να εξαλειφθεί το φαινόμενο απατών, καθώς θα πρέπει η θεωρία να συμβαδίζει απόλυτα με την πράξη προκειμένου να έχουμε το επιθυμητό αποτέλεσμα.

Σύμφωνα με την επόμενη έρευνα[14], οι εκτιμώμενες αρνήσεις εξυπηρέτησης λόγω λανθασμένης αναγνώρισης ως απάτη, μεταφράζονται περίπου στα 118 εκατομμύρια δολάρια στις Ηνωμένες Πολιτείες, για το έτος 2014 ενώ περίπου 9 εκατομμύρια δολάρια χάθηκαν από πραγματικές απάτες. Γι' αυτό το λόγο, ένα σύστημα αναγνώρισης πραγματικών απατών κρίνεται απαραίτητο και αναγκαίο, προκειμένου να μην χάνονται χρήματα λόγω λανθασμένης αναγνώρισης. Έως τώρα, ένα υβριδικό μοντέλο το οποίο συνδυάζει διάφορες τεχνικές, φαίνεται ως το καταλληλότερο. Παρόλα αυτά, το deep learning έχει έρθει στο προσκήνιο τα τελευταία χρόνια, έχοντας αρκετά καλές αποδόσεις όσον αφορά την αναγνώριση φωνής και γενικά σε θέματα τεχνητής νοημοσύνης, ενώ πολλές εταιρείες το χρησιμοποιούν ήδη για τον εντοπισμό απατών σε πιστωτικές κάρτες. Σε αυτή την έρευνα λοιπόν, μελετάται εάν το deep learning είναι κατάλληλο για την ανάπτυξη μοντέλων ανίχνευσης απατών. Οι συγγραφείς λοιπόν αρχικά παρουσιάζουν το πρόβλημα το οποίο αντιμετωπίζουν, που δεν είναι άλλο από την άρνηση συγκεκριμένων μοντέλων να εξυπηρετήσουν τους πελάτες, λόγω της εσφαλμένης πρόβλεψης για μία συναλλαγή την οποία μπορεί να θεωρήσουν ύποπτη ως απάτη στέλνοντας εσφαλμένα “καμπανάκια” για διερεύνηση της συναλλαγής. Κάθε εταιρεία λοιπόν θα πρέπει να μελετάει τις περιπτώσεις κατά τις οποίες όντως υπάρχει κάποια ύποπτη συναλλαγή και όχι να χαραμίζει πόρους για να μελετάει συναλλαγές οι οποίες εσφαλμένα θεωρούνται από το σύστημα ως απάτες. Γι' αυτό το λόγο, οι συγγραφείς θα αναλύσουν παρακάτω αν το deep learning αποδίδει καλύτερα από το υβριδικό μοντέλο το οποίο επικρατεί έως τώρα. Ενδεικτικά, αποκαλούν το υβριδικό σύστημα ως “πρωταθλητή” ενώ το deep learning μοντέλο ως “δικεδικητή”. Όπως είδαμε παραπάνω, το υβριδικό μοντέλο αποτελεί ένα συνδυασμό διάφορων machine learning μοντέλων όπως τα decision trees, logistic regression κλπ, ενώ από την άλλη, τα deep learning μοντέλα είναι νεοσύστατα και αποτελούνται από μια σειρά βημάτων για την ανεύρεση του καλύτερου μοντέλου η οποία είναι: i)επιλογή υποψήφιων μοντέλων, ii) ρύθμιση

υπερπαραμέτρων, iii) εκπαίδευση μοντέλων, iv) προσθήκη αφαίρεση μοντέλων. Προκειμένου λοιπόν να συγκριθούν οι δύο τεχνικές, χρησιμοποιείται ένα data set από μεγάλη εταιρεία έκδοσης πιστωτικών καρτών της Νότιας Κορέας, στο οποίο περιλαμβάνονται 5709 απάτες ανάμεσα σε 11.025.756 συναλλαγές. Στη συνέχεια, οι συγγραφείς προχώρησαν στις κατάλληλες δοκιμές και συγκρίσεις με τις τεχνικές K-S statistics, AUROC, alert rate precision και recall, ενώ παρέθεσαν πίνακες με όλα τα στοιχεία που παρέχονται από τις μετρήσεις τους. Αφού τελειώσουν με την παράθεση όλων των ευρημάτων τους, ουσιαστικά ανακοινώνουν τον “νικητή” ο οποίος είναι το deep learning μοντέλο, αφού απέδωσε καλύτερα στις δοκιμές, ενώ παράλληλα οι “συναγερμοί” τους οποίους απέδωσε, ήταν αρκετά πιο ακριβείς από το υβριδικό μοντέλο. Κλείνοντας, τονίζουν για άλλη μια φορά τις διαφορές τις οποίες βρήκαν ανάμεσα στα δύο μοντέλα, ενώ αναφέρουν πως η τεχνική του deep learning έχει ακόμα πολλά να αποδώσει στο μέλλον, ενώ έχει ακόμα περιθώρια εξέλιξης.

Η εργασία [15] που θα δούμε στη συνέχεια, επικεντρώνεται στην παρουσίαση και τον καθορισμό των δυσκολιών που ανιχνεύονται στις ήδη υπάρχουσες τεχνικές οι οποίες θεωρούνται κατάλληλες για την ανίχνευση απατών σε πιστωτικές κάρτες. Ο άλλος στόχος της συγκεκριμένης έρευνας, είναι η προβολή καινούργιων τεχνικών και υποσχόμενων κατευθύνσεων προκειμένου να πραγματοποιείται αποτελεσματικά η ανίχνευση απατών σε πιστωτικές κάρτες. Αρχικά λοιπόν, οι συγγραφείς εξηγούν τους τύπους απατεώνων οι οποίοι πραγματοποιούν τέτοιου είδους πράξεις κατατάσσοντας τους σε κατηγορίες. Στη συνέχεια, παρουσιάζονται διάφοροι τύποι απατών οι οποίες πραγματοποιήθηκαν, μαζί με κάποιες λεπτομέρειες για αυτές τις άπατες. Οι συγγραφείς της εργασίας, παρουσιάζουν στη συνέχεια κάποιες αναλυτικές πληροφορίες για τα δεδομένα και τα data set τα οποία χρησιμοποιούνται για τα συστήματα ανίχνευσης απατών, δίνοντας αναλυτικές πληροφορίες για τα ποσοστά εμφάνισης κάθε τύπου απάτης. Οι αναλύσεις των τεχνικών που ακολουθήθηκαν από άλλες εργασίες, πραγματοποιούνται προκειμένου να δοθούν παραπάνω πληροφορίες αλλά και να καθοδηγηθεί αποτελεσματικότερα η έρευνα την οποία πραγματοποιούν οι συγγραφείς. Οι συγγραφείς θα αναλύσουν τις διάφορες τεχνικές και αλγορίθμους οι οποίοι ακολουθούνται, δοκιμάζοντας διαφορετικές προσεγγίσεις προκειμένου να βρουν την κατάλληλη. Τέτοιους αλγόριθμους αποτελούν οι Naive Bayes, Decision Trees, C4.5, CBR κλπ. Παράλληλα, μαζί με τους αλγορίθμους αναφέρονται και διάφορες έρευνες οι οποίες υποστήριξαν τον κάθε αλγόριθμο. Οι συγγραφείς στη συνέχεια αποφασίζουν να εξετάσουν και την πλευρά του Unsupervised learning, αναφέροντας και πάλι διάφορες προσεγγίσεις αλλά και τις εργασίες οι οποίες τις στηρίζουν. Σε μια γενική κριτική των αναφερθέντων τεχνικών, επισημαίνεται πως η θεωρία με την πράξη απέχουν αρκετά, αφού πολλά πράγματα εξαρτώνται από πρακτικά ζητήματα ενώ τονίζεται η σημασία της απλότητας για την εξαγωγή ακριβέστερων αποτελεσμάτων. Η ανίχνευση τρομοκρατίας και τρομοκρατικών ενεργειών, αναφέρεται επίσης

από τους συγγραφείς, όπου και η αντιμετώπιση της είναι απαραίτητη μέσω δεδομένων που προκύπτουν και με την βοήθεια διαφορετικών αλγορίθμων. Παράλληλα, αναφέρεται η ανίχνευση οικονομικού εγκλήματος όπως το ξέπλυμα χρήματος και οι παράνομες συναλλαγές, οι οποίες χρήζουν αντιμετώπισης, η οποία μπορεί επίσης να πραγματοποιηθεί μέσα από τη χρήση αλγορίθμων όπως αλγόριθμοι κοντινότερου γείτονα, δέντρα αποφάσεων κλπ. Τέλος, οι παραβιάσεις και η ανίχνευση ανεπιθύμητων μηνυμάτων, αναφέρονται επίσης στην έρευνα, όπου υπάρχουν ήδη συστήματα τα οποία αποτρέπουν τέτοιες ενέργειες, όμως με τις κατάλληλες τεχνικές ανίχνευσης, αυτά τα συστήματα μπορούν να γίνουν πολύ ισχυρότερα.)

Στη επόμενη εργασία [16], οι συγγραφείς θα χρησιμοποιήσουν ένα data set το οποίο προέρχεται από πραγματικά καταγεγραμμένα δεδομένα, προκειμένου να εξάγουν όσο περισσότερο ρεαλιστικότερα αποτελέσματα μπορούν, χρησιμοποιώντας διαφορετικούς αλγόριθμους supervised machine learning. Παράλληλα, σημαντικό ζήτημα αποτελεί και η επιλογή των καταλληλότερων μεταβλητών προκειμένου να έχουμε τα καλύτερα δυνατά αποτελέσματα. Αρχικά λοιπόν, οι συγγραφείς αναφέρονται στις απάτες πιστωτικών καρτών, εξηγώντας πόσο εύκολα μπορεί κανείς να εξαπατηθεί, παραθέτοντας παράλληλα κάποιες πληροφορίες σχετικά με τις απάτες. Στη συνέχεια, οι συγγραφείς αναφέρουν πως θα χρησιμοποιήσουν 10 μοντέλα μηχανικής μάθησης προκειμένου να συγκρίνουν δείκτες όπως Accuracy, Precision, TPR κλπ. Κατά τη σύγκριση των αλγορίθμων, κάποιοι απ' αυτούς είχαν κακή απόδοση λόγω της ανισορροπίας των δεδομένων, ενώ το καλύτερο αποτέλεσμα φαίνεται να είχαν τα νευρωνικά δίκτυα (unsupervised method) πετυχαίνοντας 95% ανίχνευση απατών χωρίς λανθασμένο συναγερμό. Φαίνεται λοιπόν οι unsupervised methods να μπορούν να αναγνωρίζουν καλύτερα μοτίβα ανίχνευσης απατών. Οι συγγραφείς παραθέτουν στη συνέχεια τα αποτελέσματα ενώ παραθέτουν τη χρησιμότητα που έχει κάθε αλγόριθμος όσον αφορά την ανίχνευση απατών. Τέλος, στα συμπεράσματα αναφέρεται πως ο αλγόριθμος Logistic Regression είναι ο πιο πολλά υποσχόμενος, μιας και απέδωσε καλύτερα στις δοκιμές ανίχνευσης απατών, ενώ δεύτερος έρχεται ο random forest και τρίτος ο XGB classifier

Η επόμενη εργασία [17] ασχολείται και πάλι με το θέμα ανίχνευσης απατών σε πιστωτικές κάρτες ενώ συγκεκριμένα χρησιμοποιεί ένα data set Ευρωπαίων κατόχων πιστωτικών καρτών προκειμένου να εξάγει αποτελέσματα, ενώ αναζητά διαφορετικές τεχνικές που θα βοηθούσαν στην ανίχνευση απατών. Αρχικά λοιπόν, οι συγγραφείς μας περιγράφουν με ποιον τρόπο μπορεί πολύ εύκολα κάποιος να πέσει θύμα υποκλοπής, αφού πλέον δεν χρειάζεται κάποιος να υποκλέψει φυσικά την πιστωτική κάρτα, αφού μπορεί πολύ ευκολότερα να αποκτήσει πρόσβαση σε αυτήν μέσω του διαδικτύου. Τα δεδομένα λοιπόν που χρησιμοποιήθηκαν αρχικά για να πραγματοποιηθεί η έρευνα, όπως προανέφερα προέρχονται από Ευρωπαίους κατόχους καρτών, ενώ έχουν υποστεί την κατάλληλη προ επεξεργασία. Σύμφωνα με τους συγγραφείς, διάφορες τεχνικές ακολουθήθηκαν από διαφορετικούς

ερευνητές, τους οποίους αναφέρουν στην έρευνα τους. Για παράδειγμα, οι K. R. Seeja και M. Zareapoor χρησιμοποίησαν ένα μοντέλο το οποίο ανιχνεύει απάτες μέσα από ένα καθόλου ισορροπημένο data set. Από την άλλη, Y. Sahing και E. Duman χρησιμοποίησαν έναν συνδυασμό αλγορίθμων προκειμένου να πετύχουν το επιθυμητό αποτέλεσμα, ενώ οι Raghavendra Patidar, Lokesh Sharma χρησιμοποίησαν τα νευρωνικά δίκτυα. Οι συγγραφείς στη συνέχεια παραθέτουν τις διάφορες τεχνικές με τις οποίες μπορούν να επεξεργαστούν και να αναλύσουν τα δεδομένα. Οι τεχνικές που αναφέρονται είναι οι Unsupervised Learning, Supervised Learning, Local Outlier Factor, Support Vector Machine. Τέλος, ο Isolation Forest όπως φαίνεται απέδωσε πολύ καλύτερα σε σχέση με τους υπόλοιπους αλγόριθμους, οπότε και οι συγγραφείς αναφέρουν πως θα τον προτιμούσαν έναντι των υπόλοιπων αλγορίθμων οι οποίοι δεν τα πήγαν όσο καλά περίμεναν.

Η επόμενη εργασία [18] ασχολείται με την εφαρμογή supervised και unsupervised αλγορίθμων σε data set με μεγάλη ανισορροπία, προκειμένου να δουν και να συμπεράνουν ποιος αλγόριθμος θα αποδώσει καλύτερα υπό αυτές τις συνθήκες. Όπως γνωρίζουμε λοιπόν, οι πληρωμές μέσω πιστωτικών καρτών είναι μια κυρίαρχη μέθοδος πληρωμής στις μέρες μας αφού οι περισσότερες συναλλαγές πραγματοποιούνται μέσω πιστωτικών καρτών. Όμως, όπως έχουμε δει πολλές φορές, οι κάτοχοι καρτών πολλές φορές πέφτουν θύματα απάτης με αποτέλεσμα κάποιος να αποκτά πρόσβαση στα στοιχεία της κάρτας του. Γι' αυτό το λόγο, υπάρχουν τα FDS (Fraud detection systems) τα οποία είναι αυτοματοποιημένα μηχανήματα τα οποία προσπαθούν να αποτρέψουν τέτοιες απάτες, ενώ ταυτόχρονα προσπαθούν να ειδοποιήσουν με εσωτερικό “συναγερμό” για μια τέτοια συναλλαγή. Σε αυτές τις περιπτώσεις, σημαντικό εργαλείο αποτελεί η μηχανική μάθηση με supervised και unsupervised μεθόδους, η οποία βοηθάει στον εντοπισμό τέτοιων απατών. Σύμφωνα με τους συγγραφείς λοιπόν, ένα ιδανικό σύστημα FDS θα πρέπει να επικεντρώνεται στην κατανόηση συμπεριφορών από τους πελάτες προκειμένου στο μέλλον να μπορεί να αποτρέψει κάποια απάτη, αναγνωρίζοντας μοτίβα συμπεριφορών. Στη συνέχεια, οι συγγραφείς παρουσιάζουν κάποιους αλγορίθμους supervised (random forest, neural networks, deep learning κλπ) και unsupervised (K-means, isolation forest, self-organizing maps κλπ) μεθόδων εξηγώντας που υπερτερεί και πως λειτουργεί ο καθένας. Σχετικά με το dataset το οποίο χρησιμοποιούν, αναφέρουν ότι τα δεδομένα είναι πραγματικά και έχουν συλλεχθεί από κατόχους καρτών ενώ δίνουν κάποιες λεπτομέρειες γι' αυτά. Αφού πραγματοποιήσουν τα απαραίτητα πειράματα, αναφέρουν τελικά πως ο κάθε αλγόριθμος μπορεί να απαντήσει σε διαφορετικά ερωτήματα ενώ σαν γενικότερη εικόνα, οι NN (neural networks), ANN (artificial neural networks), XGBT (extended gradient boosted tree), LR (logistic regression), IF (isolation forest), LOF (local outlier factor) είχαν τα καλύτερα αποτελέσματα, ενώ οι IF και LOF είναι αυτοί που ξεχώρισαν ακόμη περισσότερο με την απόδοσή τους.

Η επόμενη εργασία [19] που θα εξετάσουμε, παρουσιάζει ένα μοντέλο ανίχνευσης απατών σε πιστωτικές κάρτες, το οποίο καθορίζεται από περίπου 150 χαρακτηριστικά ανά επισκέπτη, πράγμα που θεωρητικά το κάνει αρκετά ακριβές στις προβλέψεις του. Οι συγγραφείς λοιπόν αρχικά βασίζονται τη συγγραφή του συγκεκριμένου paper στην κατάλληλη βιβλιογραφία, όπου αναφέρουν εργασίες από τις οποίες πήραν την κατάλληλη καθοδήγηση αναφέροντας παράλληλα γενικές πληροφορίες για τις απάτες πιστωτικών καρτών. Η πρώτη έρευνα που αναφέρουν, προτείνει ένα μοντέλο το οποίο βασίζεται στον εντοπισμό outlier προκειμένου να αποκαλυφθεί μια απάτη, ενώ σύμφωνα με τη δεύτερη έρευνα, χρησιμοποιείται η θεωρία Bayes Minimum Risk προκειμένου να υπάρξει ο κατάλληλος εντοπισμός απατών. Σε δύο ακόμα διαφορετικές προσεγγίσεις, χρησιμοποιούνται απλές μετρήσεις προκειμένου να βρεθούν οι περιπτώσεις απατών, ενώ η άλλη προσέγγιση επικεντρώνεται στην ανίχνευση σε πραγματικό χρόνο. Όσον αφορά τα δεδομένα που χρησιμοποιούνται, για λόγους ιδιωτικότητας δεν παρουσιάζονται ευαίσθητα στοιχεία, ενώ λόγω της μεγάλης ανισορροπίας που υπήρχε, τα δεδομένα δέχτηκαν την κατάλληλη προ επεξεργασία. Τα δεδομένα χωρίστηκαν στη μέση προκειμένου να χρησιμοποιηθούν ως training και testing, ενώ κατά το στάδιο προεπεξεργασίας στα δεδομένα συμπληρώθηκαν οι μηδενικές μεταβλητές, κανονικοποιήθηκαν και τελικά χρησιμοποιήθηκαν οι logistic regression, decision tree και extreme gradient boosting. Στους πίνακες που παρουσιάζουν στο paper, φαίνεται καθαρά η υπεροχή του logistic regression στις μετρήσεις αφού έχουν σημειωθεί οι υψηλότερες βαθμολογίες σε σχέση με τους υπόλοιπους, πράγμα που κάνει τους συγγραφείς να πιστεύουν πως ο αλγόριθμος αποτελεί σημαντικό εργαλείο στην αντιμετώπιση των απατών.

Η επόμενη έρευνα [20], ξεκινά αναλύοντας το πρόβλημα των απατών των πιστωτικών καρτών και το πόσο σημαντικό ζήτημα αποτελεί στις μέρες μας, καθώς οι περισσότερες συναλλαγές πραγματοποιούνται μέσω του διαδικτύου. Στη συνέχεια, τονίζεται η ανάγκη για νέες ή έστω διαφοροποιημένες τεχνικές εύρεσης των απατών, μιας και η συμπεριφορά των απατεώνων μπορεί να αλλάζει συνεχώς προκειμένου να μη γίνονται αντιληπτοί, ενώ από την άλλη πλευρά τα δεδομένα τα οποία συνήθως χρησιμοποιούνται για τέτοιες μετρήσεις έχουν μεγάλη ανισορροπία και δεν είναι ιδανικά για πειραματισμούς. Στη συγκεκριμένη εργασία λοιπόν, οι συγγραφείς θα προσπαθήσουν να δώσουν μια δική τους προσέγγιση η οποία ίσως με τη σειρά της βοηθήσει στην αντιμετώπιση των απατών σε πιστωτικές κάρτες. Γι' αυτό το λόγο, στη συγκεκριμένη έρευνα χρησιμοποιήθηκαν οι naive bayes, k-nearest neighbor και logistic regression σε αρκετά μη κατάλληλα δεδομένα προκειμένου να δούμε ποιος θα αποδώσει καλύτερα. Οι μετρήσεις στο τέλος θα αξιολογηθούν βάσει δεικτών όπως accuracy, sensitivity, specificity, MCC και balanced classification rate. Όσον αφορά την ίδια την απάτη στις πιστωτικές κάρτες, προκειμένου να βρεθεί το καλύτερο αποτέλεσμα, πρέπει πρώτα να μελετηθεί ο τρόπος με τον οποίο αυτές πραγματοποιούνται. Οι απατεώνες λοιπόν, αλλάζουν

συνέχεια τρόπους και μοτίβα κινήσεων προκειμένου να μη γίνονται αντιληπτοί, ενώ οι απάτες γενικότερα χωρίζονται σε εσωτερικές και εξωτερικές απάτες. Για τους παραπάνω λόγους λοιπόν, πρέπει πρώτα να αναλυθούν οι κινήσεις του κατόχου προκειμένου να μπορεί στη συνέχεια να συγκριθεί με την συμπεριφορά του υποκλοπέα. Αφού γίνουν όλες οι παραπάνω διαδικασίες, οι συγγραφείς θα χρησιμοποιήσουν supervised και unsupervised τεχνικές προκειμένου να πραγματοποιήσουν τους πειραματισμούς τους. Το dataset το οποίο χρησιμοποιείται, προέρχεται από στοιχεία Ευρωπαίων κατόχων πιστωτικών καρτών από το Σεπτέμβριο του 2013 και περιέχει 284,807 συναλλαγές με τις απάτες να αποτελούν το 0,172%. Αφού λοιπόν οι συγγραφείς πειραματίστηκαν με όλες τις μεθόδους οι οποίες αναφέρθηκαν, παρουσιάζουν τα αποτελέσματα τους, αναφέροντας πως τα πολύ υψηλά ανισόρροπα δεδομένα χρησιμοποιήθηκαν σε ένα υβριδικό μοντέλο όπου προέκυψαν δύο set και στη συνέχεια οι τρεις classifiers (naive bayes, K-nearest, logistic regression) εξετάστηκαν με τα δύο αυτά set, με τον kNN να αποδίδει καλύτερα από τους άλλους.

Στις μέρες μας, το καταναλωτικό κοινό έχει στραφεί αρκετά στις ηλεκτρονικές αγορές, μιας και είναι πολύ ευκολότερο για κάποιον να τις πραγματοποιήσει. Λόγω αυτής της αύξησης όμως, έχουν αυξηθεί και οι περιπτώσεις απατών σε πιστωτικές κάρτες, αφού πολλοί είναι εκείνοι που προσπαθούν να αποκτήσουν πρόσβαση σε προσωπικά στοιχεία και συνεπώς σε πιστωτικές κάρτες και τραπεζικούς λογαριασμούς του θύματος. Στη συγκεκριμένη έρευνα [21] λοιπόν θα δούμε πως μπορεί να καταπολεμηθεί αυτό το μεγάλο πρόβλημα που έχει δημιουργηθεί με την αύξηση των ηλεκτρονικών συναλλαγών, με τη βοήθεια της μηχανικής μάθησης η οποία μας προσφέρει πολλές δυνατότητες. Η πραγματοποίηση μιας συναλλαγής πραγματοποιείται με δύο τρόπους, είτε με φυσική παρουσία, είτε ηλεκτρονικά όπου και είναι πιο εύκολο για κάποιον επίδοξο απατεώνα να υποκλέψει προσωπικά στοιχεία του θύματος. Επομένως, οι συγγραφείς αναφέρουν πως οι δύο τρόποι για να αντιμετωπιστεί κάτι τέτοιο είναι είτε η πρόβλεψη της απάτης, είτε ο εντοπισμός της την κατάλληλη στιγμή. Γενικότερα, οι απατεώνες προσπαθούν να υποκλέψουν προσωπικά στοιχεία μέσω κλεμμένων καρτών, μέσω Phishing (“ψάρεμα” στοιχείων), Skimming (τρόπος υποκλοπής στοιχείων μέσω ATM), Card not Present (ο απατεώνας προσποιείται ότι είναι ο κάτοχος της κάρτας). Στην έρευνα την οποία εξετάζουμε, θα χρησιμοποιηθούν 6 τεχνικές εξόρυξης οι οποίες θα είναι Classification, Clustering, Prediction, Outlier Detection, Regression, Visualization. Παράλληλα, οι συγγραφείς αναλύουν κάποιες ήδη υπάρχουσες τεχνικές, οι οποίες είναι:

- Artificial Immune System(τεχνική εξόρυξη δεδομένων)
- Bayesian Belief Network(στατιστική ταξινόμηση βασισμένη σε θεώρημα Bayes)
- Logistic Regression (δυαδική ταξινόμηση βασισμένη σε στατιστική μέθοδο)
- Neural Network(υπολογιστική προσέγγιση του ανθρώπινου εγκεφάλου)

- Support Vector Machine (μετατροπή προβλήματος σε υψηλών διαστάσεων χώρο χαρακτηριστικών)
- Genetic Algorithm (εμπνέεται από την ιδέα της εξέλιξης πληθυσμού)
- Decision Trees (δυναμικό δέντρο για την ταξινόμηση δεδομένων)
- Self-organising Map (παραλλαγή του Neural Network)
- Hybrid Methods (συνδυασμός μεθόδων)

Αφού παρουσιάσουν τα αποτελέσματα των αναλύσεων τους σε σχετικό πίνακα, οι συγγραφείς καταλήγουν στο συμπέρασμα πως οι supervised techniques και hybrid approach της μηχανικής μάθησης είναι οι κατάλληλες τεχνικές για την ανίχνευση απατών. Επισημαίνουν παράλληλα, πως η μηχανική μάθηση είναι η κατάλληλη προσέγγιση για την ανίχνευση απατών, καθώς παρέχει υψηλή ακρίβεια και ποσοστό ανίχνευσης.

Κατά την τελευταία δεκαετία έχει παρατηρηθεί μεγάλη αύξηση στις διαδικτυακές αγορές, λόγω της ευκολίας με την οποία πλέον μπορεί κάποιος να πραγματοποιήσει μια αγορά, όπως προαναφέρθηκε. Αυτό όμως έχει ως συνέπεια την αύξηση των περιπτώσεων κατά τις οποίες οι χρήστες εξαπατώνται και πέφτουν θύματα επίδοξων απατεώνων με απώτερο σκοπό την υποκλοπή προσωπικών στοιχείων αλλά και χρημάτων. Προκειμένου να αντιμετωπιστεί αυτό το φαινόμενο, οι συγγραφείς της συγκεκριμένης έρευνας αναφέρουν πως θα πρέπει να αναγνωρισθούν και να αποτυπωθούν τα μοτίβα συμπεριφοράς των κατόχων καρτών, όσον αφορά τις συναλλαγές, προκειμένου να μπορέσουν μέσω συγκεκριμένων μεθόδων να αναγνωρίσουν πότε μια συναλλαγή αποτελεί απάτη. Στη συγκεκριμένη έρευνα [22], συνδυάζονται η σύγκριση των μοτίβων συμπεριφορών με συγκεκριμένες τεχνικές προκειμένου να βρεθεί αν μια συναλλαγή αποτελεί απάτη. Αρχικά, οι συγγραφείς αναφέρουν την σημαντικότητα της ανίχνευσης μιας απάτης καθώς εκτός από τους κατόχους καρτών, μεγάλη ζημιά υπάρχει και για τις μεγάλες επιχειρήσεις οι οποίες πλήττονται από τέτοιες καταστάσεις. Στη συγκεκριμένη εργασία λοιπόν, θα αναλυθούν τεχνικές ανίχνευσης απατών σύμφωνα με αλγόριθμους της μηχανικής μάθησης όπως Logistic Regression, K-nearest Neighbour, Naive Bayes, Decision Trees, Neural Network- Chebyshev Functional Link Artificial Neural Networks, Multi-Layer Perceptron βασισμένα σε TPR, TNR, FPR, FNR, accuracy, sensitivity, specificity. Στη συνέχεια, αναλύεται ο τρόπος προσέγγισης και οι αλγόριθμοι που χρησιμοποιήθηκαν σε διάφορες έρευνες τις οποίες οι συγγραφείς χρησιμοποίησαν ως βιβλιογραφία, περιγράφοντας συνοπτικά τον τρόπο λειτουργίας του καθενός. Οι αλγόριθμοι και οι προσεγγίσεις που θα χρησιμοποιηθούν από τους συγγραφείς, περιγράφονται στη συνέχεια της έρευνας, όπου περιγράφεται αναλυτικά και η χρησιμότητα τους στην ανίχνευση απατών. Συγκεκριμένα, περιγράφεται η χρήση και λειτουργία :

- Meta-Classification strategy (χρησιμοποίηση decision trees, Naive-basyesian και K-nearest neighbour)
- Hybrid Sampling (αφού γίνει data sampling)
- Credid Card Fraud Detection using Naive Bayes (δυναμική ταξινόμηση μέσω Naive Bayes)
- Credit Card Fraud Detection using Logistic Regression (χρησιμοποιείται για την εκτίμηση της πιθανότητας δυαδικής απόκρισης)
- Credit Card Fraud Detection using K-nearest neighbour (χρησιμοποιεί μετρήσεις για την ταξινόμηση του)
- Decision Trees (ταξινόμηση βάσει αποφάσεων)

Ξεχωριστό κεφάλαιο αποτελούν τα νευρωνικά δίκτυα, κατά το οποίο περιγράφεται η λειτουργία τους και η μεθοδολογία την οποία χρησιμοποιούν οι συγγραφείς. Το data set το οποίο χρησιμοποιείται περιλαμβάνει δεδομένα Ευρωπαίων κατόχων καρτών μέσα στα οποία υπάρχουν 492 συναλλαγές-απάτες από τις συνολικά 284.807 συναλλαγές. Προχωρώντας στην ανάλυση των αλγορίθμων και στα αποτελέσματά τους, συμπεραίνουμε πως ο K-nearest neighbour έχει τα μεγαλύτερα ποσοστά όσον αφορά τους δείκτες accuracy, sensitivity και specificity σε σχέση με τους υπόλοιπους, με τον Naive Bayes να έρχεται δεύτερος στις μετρήσεις.

Λόγω της όλο και αυξανόμενης ζήτησης στα ηλεκτρονικά καταστήματα και γενικότερα στις ηλεκτρονικές συναλλαγές, έχουν αυξηθεί και οι απάτες μέσω πιστωτικών καρτών. Γι' αυτό το λόγο, τον τελευταίο καιρό η μηχανική μάθηση έχει αρχίσει να διαδίδεται ακόμα περισσότερο, ενώ πολλοί αναζητούν λύση στο πρόβλημα των απατών πιστωτικών καρτών, μέσω αυτής. Στη συγκεκριμένη εργασία [23], θα εξεταστούν τρεις διαφορετικές προσεγγίσεις με τους αλγόριθμους Random Forest, Support Vector Machine και Logistic Regression προσπαθώντας να εντοπίσουν απάτες σε dataset το οποίο προέρχεται εξ ολοκλήρου από συναλλαγές κατόχων καρτών, δίνοντας έναν παραπάνω ρεαλισμό στα αποτελέσματα τα οποία θα εξάγουν. Αρχικά, οι συγγραφείς αναφέρουν κάποιες λεπτομέρειες οι οποίες σχετίζονται με την στατιστική απεικόνιση του προβλήματος, μιας και τα χρήματα που χάθηκαν από απάτες κατά το έτος 2016, άγγιξαν τα 1.8 δισεκατομμύρια, το οποίο είναι κατά 92% μεγαλύτερο από το 2012. Γι' αυτό το λόγο, έχουν αναπτυχθεί διάφορες τεχνικές οι οποίες προσπαθούν να εντοπίσουν μια απάτη. Το πρώτο στάδιο αυτή της τεχνικής είναι η αυτοματοποιημένη αναγνώριση και επικύρωση κάρτας, στο οποίο αν δεν εντοπιστεί κάτι ύποπτο, η κάρτα περνάει τον έλεγχο και προχωράει στο δεύτερο στάδιο. Το δεύτερο στάδιο (Predictive model) προσπαθεί να αναγνωρίσει μια απάτη από συγκεκριμένα χαρακτηριστικά και μοτίβα της συναλλαγής, σε σχέση με παλαιότερες που

είχαν πραγματοποιηθεί από τον ίδιο κάτοχο. Ουσιαστικά, το δεύτερο επίπεδο θα είναι αυτό το οποίο μετράει περισσότερο και πολλοί κάτοχοι παρέχουν πληροφορίες (feedback) προκειμένου να βοηθήσουν στην καταπολέμηση αυτού του φαινομένου, αφού αυτό είναι το στάδιο στο οποίο συνήθως εντοπίζονται οι απάτες. Ο εντοπισμός απατών λοιπόν μπορεί να πραγματοποιηθεί μέσω εξόρυξης δεδομένων και της μεθόδου ταξινόμησης (Classification), το οποίο θα είναι και το θέμα συζήτησης στη συνέχεια. Όσον αφορά τα δεδομένα που έχουν οι ερευνητές στη διάθεση τους, πολλές φορές δεν είναι ακριβή λόγω της ευαισθησίας του συγκεκριμένου θέματος ενώ χαρακτηρίζονται πολλές φορές από μεγάλο ποσοστό ανισορροπίας (Imbalanced data). Προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα, κατά το στάδιο της προεπεξεργασίας, πραγματοποιούνται οι κατάλληλες μετατροπές και η χρησιμοποίηση διαφόρων τεχνικών προκειμένου τα δεδομένα να επέλθουν σε μια καλύτερη για επεξεργασία, μορφή. Τέτοια παραδείγματα αποτελούν η λειτουργία SMOTE, αφαίρεση outliers κλπ. Από την άλλη πλευρά, προκειμένου να μπορεί ένα σύστημα να αναγνωρίζει τις συνήθειες αλλά και συγκεκριμένα μοτίβα των χρηστών, δεν θα πρέπει αυτό να είναι στατικό αλλά να αλλάζει και να εξελίσσεται συνεχώς μέσω της μηχανικής μάθησης. Στη συνέχεια, οι συγγραφείς θα αναλύσουν τα δεδομένα τα οποία έχουν στη διάθεση τους, με βάση τους δείκτες precision, recall και φυσικά TP, FP, TN, FN. Τα δεδομένα όπως ανέφερα και πριν, αποτελούν πραγματικά στοιχεία Ευρωπαίων κατόχων (284.807 συναλλαγές συνολικά) με τις 492 περιπτώσεις να αποτελούν απάτες. Οι αλγόριθμοι που χρησιμοποιούνται είναι οι :

- Random forest (διαφοροποίηση του δέντρου αποφάσεων)
- Support Vector Machine (μετατροπή προβλήματος σε γραμμικό και ευκολότερο)
- Logistic Regression (πιθανοτικό μοντέλο)

Αφού πραγματοποιήθηκαν λοιπόν τα κατάλληλα πειράματα(static και incremental), οι συγγραφείς συμπέραναν πως σε όλες τις περιπτώσεις ο SVM είχε τη χειρότερη απόδοση(και στις δύο περιπτώσεις, ενώ οι άλλοι δύο (RF,LR) είχαν μικρές διαφορές στις συγκρίσεις μεταξύ τους με τον LR να αποδίδει πολύ ικανοποιητικά σε συνθήκη incremental. Ως μελλοντική επέκταση, οι συγγραφείς αναφέρουν πως θα ήθελαν να ασχοληθούν με μεγαλύτερα dataset προκειμένου να έχουν μια ακόμα καθαρότερη εικόνα.

Η συνεχώς αυξανόμενη τάση διαδικτυακών αγορών έχει ως άμεσο αποτέλεσμα την αύξηση των περιστατικών απατών σε πιστωτικές κάρτες, κάτι που είναι ολέθριο και για τις μεγάλες επιχειρήσεις εκτός από τους κατόχους των καρτών. Οι απάτες αυτές μπορούν να εντοπιστούν, μέσω σύγκρισης συμπεριφορών των κατόχων καρτών με προηγούμενες συναλλαγές η οποία πραγματοποιείται με τη βοήθεια της μηχανικής μάθησης. Στη συγκεκριμένη έρευνα [24], οι συγγραφείς εξετάζουν διάφορες μεθόδους προκειμένου να μπορούν να αναγνωρίσουν τέτοιου είδους απάτες. Το πρόβλημα των απατών λοιπόν, ξεκινά

όταν ο κάτοχος χάσει την κάρτα του ή στην περίπτωση που κάποιος αποκτήσει πρόσβαση σε αυτήν, χωρίς την συναίνεση του κατόχου. Ο απατεώνας, θα χρησιμοποιήσει στη συνέχεια την κάρτα προκειμένου να αποσπάσει κάποιο χρηματικό ποσό από τον λογαριασμό του κατόχου. Ανά καιρούς, έχουν αναπτυχθεί διάφορα μοντέλα τα οποία βοηθούν στην εξεύρεση των συναλλαγών-απατών προκειμένου να περιοριστεί όσο γίνεται το φαινόμενο. Τέτοια μοντέλα είναι Different Fraud Location Models, Dempster-Shafer theory and Bayesian learning, Hiddedn Markov Model Fraud detection, Detection of Credit Card Fraud using Fuzzy, Bayesian and Neural Networks, Fraud detecion using Association rules, CARDWATCH, Fraud detection using SOM&PSO, Genetic Algorithm and Scatter Search, Fraud detection using Decission Tree and SVM, Fraud Detection usinr Rule based an expert framework, Fraud detection using Neural Network, Game Theory approach, Fraud detection using Peer Group analysis. Στην απέναντι όχθη, οι απατεώνες, προκειμένου να καταφέρουν να υποκλέψουν προσωπικά δεδομένα και να αποσπάσουν χρηματικά ποσά από τους λογαριασμούς των θυμάτων, χρησιμοποιούν τεχνικές μέσω διαδικτύου όπως το site cloning (κλώνος μιας πραγματικής σελίδας απ' όπου υποκλέπονται προσωπικά δεδομένα), False merchant areas (υποκλοπή προσωπικών στοιχείων) ενώ στην έρευνα αναφέρονται επίσης οι τεχνικές : Lost/stolen cards, Account takeover, Cardholder-not-present, Fake and counterfeit cards, Errasing the alluring strip, Making a fraud card, Skimming, Phising. Οι τεχνικές που ακολουθούνται σύμφωνα με τους συγγραφείς, προκειμένου να προβλεφθούν οι απάτες χωρίζονται σε

- Descriptive or Unsupervised Techniques (Graphical outlier detection procedure, OLAP, Statistical outlier detection, Break point analysis, Association rule analysis, Clustering, Self-Organizing Maps)
- Predictive Techniques (Linear Regression, Decision Trees, Neural Networks, Support Vector Machines, Naive Bayes, Rule Based Classification)
- Artificial & Computational Intelligence Techniques (Genetic Algorithm, Particle swarm improvmenet, Hidden Markov Model, Fuzzy rationale, Game Theory, Dempster Shafer Theory, Entropy)

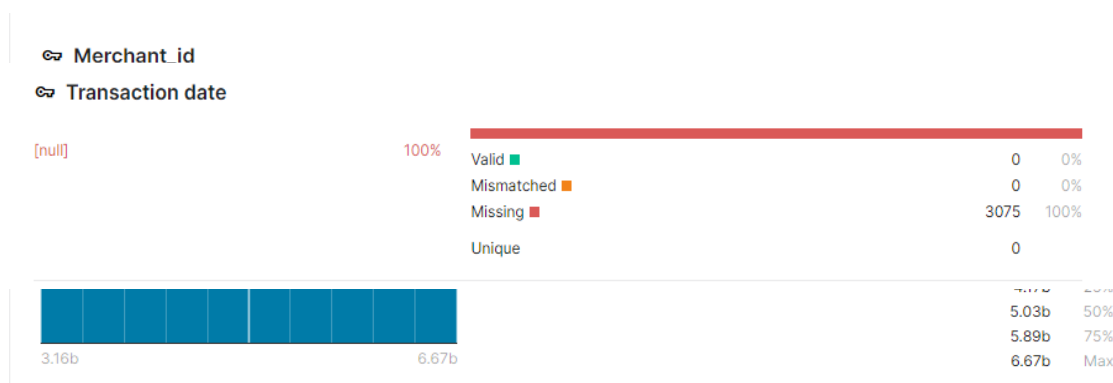
Τέλος, οι συγγραφείς δίνουν κάποιες χρήσιμες συμβουλές προκειμένου να αποτρέψει κανείς μια πιθανή περίπτωση απάτης, με φράσεις όπως “Carry your card independently from your wallet, Remember your PIN and decimate a paper warning, When enetering you PIN utilize your free hand and your body to protect the number”.

Κεφάλαιο 4 “Πειραματική Ανάλυση δεδομένων και γραφήματα”

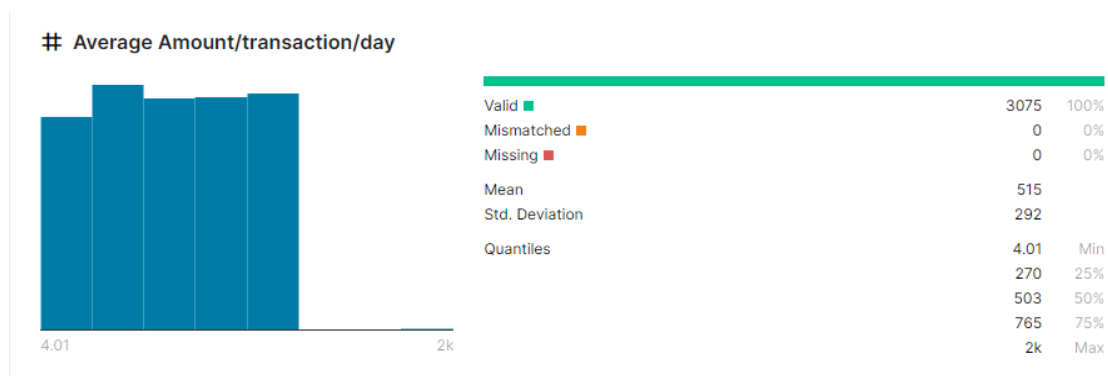
4.1 Περιγραφή και υλοποίηση του πειράματος

Αφού προηγήθηκε η βιβλιογραφική και θεωρητική ανάλυση του θέματος το οποίο πραγματεύεται η εργασία, θα ακολουθήσει μια βαθύτερη ανάλυση η οποία θα περιέχει πρακτική ανάλυση και πειραματισμό με τα data set τα οποία θα χρησιμοποιηθούν. Προκειμένου να μπορέσει να χρησιμοποιηθεί το συγκεκριμένο σύνολο δεδομένων, επισκεφθήκαμε τον ιστότοπο του Kaggle, μια ιστοσελίδα η οποία είναι ευρέως γνωστή για την μεγάλη ποικιλία σε εργασίες και data set, που διαθέτει. Στη συγκεκριμένη ιστοσελίδα βρήκαμε λοιπόν το σύνολο δεδομένων το οποίο μπορεί να βοηθήσει στην πραγματοποίηση διαφόρων πειραμάτων αφού όπως θα δούμε στη συνέχεια, περιέχει έναν μεγάλο αριθμό δεδομένων. Το πρόγραμμα επεξεργασίας δεδομένων που θα χρησιμοποιηθεί, είναι το Weka. Το Weka είναι ένα ανοιχτού κώδικα λογισμικό, το οποίο δημιουργήθηκε από το Πανεπιστήμιο του Waikato στη Νέα Ζηλανδία, και αποτελεί ένα από τα πιο διαδεδομένα λογισμικά ανάλυσης δεδομένων, αφού περιέχει εργαλεία οπτικοποίησης και αλγορίθμους για την πρόγνωση μοντέλων και την ανάλυση δεδομένων. Ο στόχος των πειραματισμών που θα πραγματοποιηθούν είναι η διερεύνηση της αποτελεσματικότητας που μπορούν να έχουν οι διάφοροι αλγόριθμοι στον εντοπισμό απατών μέσω πιστωτικών καρτών και πως μπορούν γενικά η μηχανική μάθηση να βοηθήσει στο συγκεκριμένο ζήτημα.

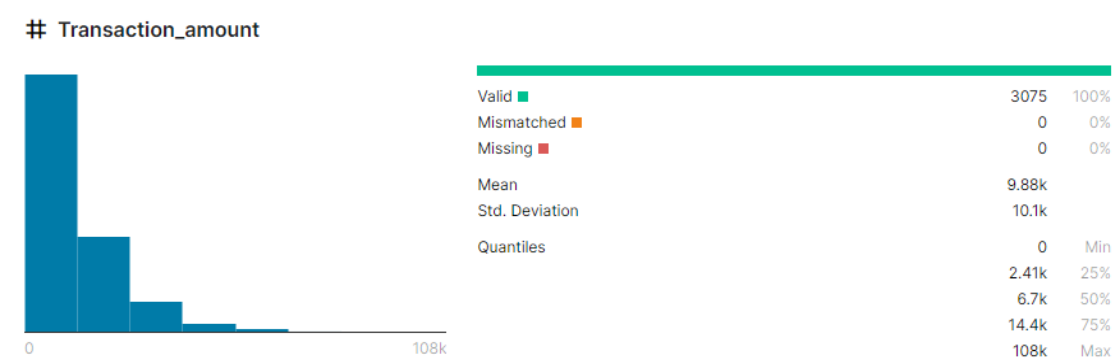
Το data set το οποίο θα χρησιμοποιηθεί, αποτελείται από δεδομένα τα οποία περιγράφουν διαφορετικούς παράγοντες μιας συναλλαγής. Παρακάτω, παραθέτονται τα attributes μαζί με τις λεπτομέρειες που παρέχονται για το καθένα.



Εικόνα 6 (Αναγνωριστικό εμπόρου(πελάτη)-Ημερομηνία συναλλαγής)[25]



Εικόνα 7 (Μέσες δαπάνες-συναλλαγές ανά μέρα)[25]



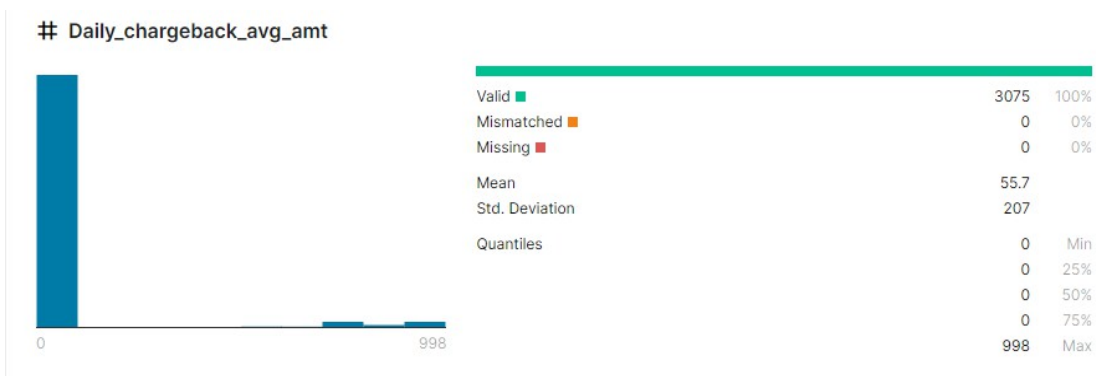
Εικόνα 8 (Ποσό συναλλαγής)[25]



Εικόνα 9 (Συναλλαγή που πραγματοποιήθηκε σε διαφορετικό νόμισμα)[25]



Εικόνα 10 (Αριθμός άρνησης συναλλαγής ανά ημέρα) [25]



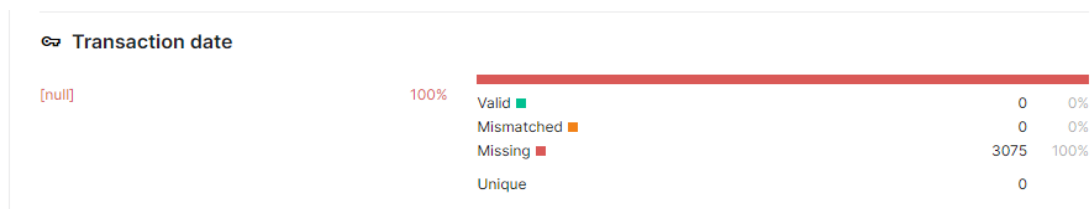
Εικόνα 11 (Μέση καθημερινή αντιστροφή χρέωσης(αμφισβήτηση χρέωσης)) [25]

✓ isHighRiskCountry

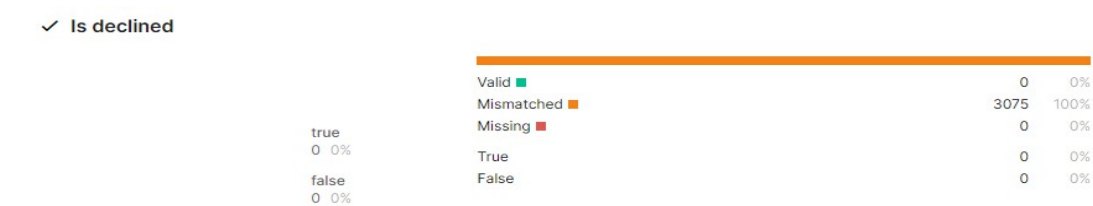
		Valid	0 0%
		Mismatched	3075 100%
		Missing	0 0%
true	0 0%	True	0 0%
false	0 0%	False	0 0%

Εικόνα 12 (Αποτελεί χώρα υψηλού κινδύνου) [25]

Εικόνα 13 (Ημερομηνία συναλλαγής) [25]

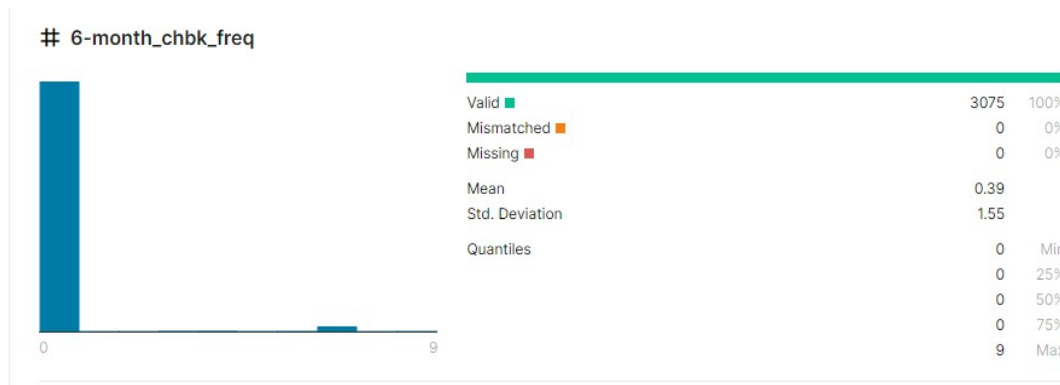


Εικόνα 14 (Μέσο όρο 6μηνης αντιστροφή χρέωσης(αμφισβήτηση χρέωσης))[25]



Εικόνα 15 (Η συναλλαγή απορρίφθηκε))[25]

Εικόνα 16 (Συχνότητα αντίστροφης χρέωσης ανά εξάμηνο)[25]

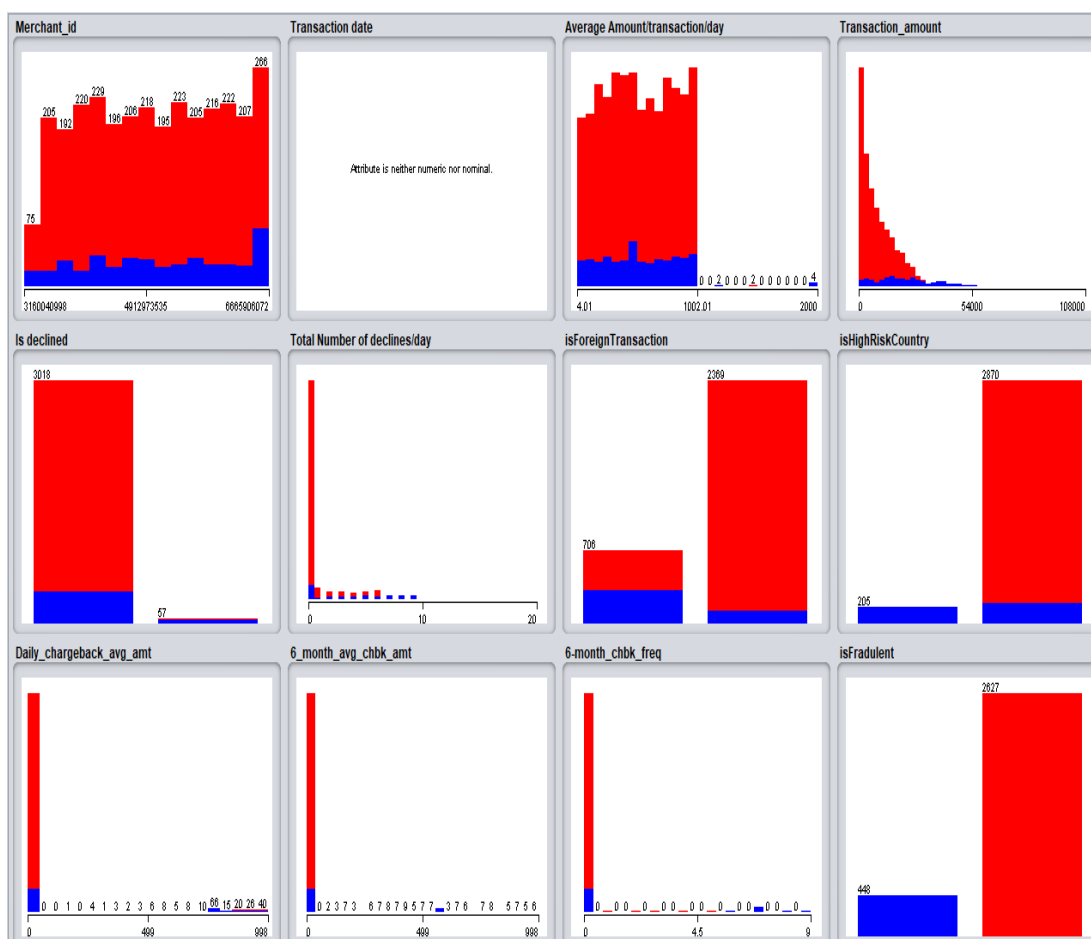


✓ isFradulent

		Valid	0 0%
		Mismatched	3075 100%
		Missing	0 0%
true	0 0%	True	0 0%
false	0 0%	False	0 0%

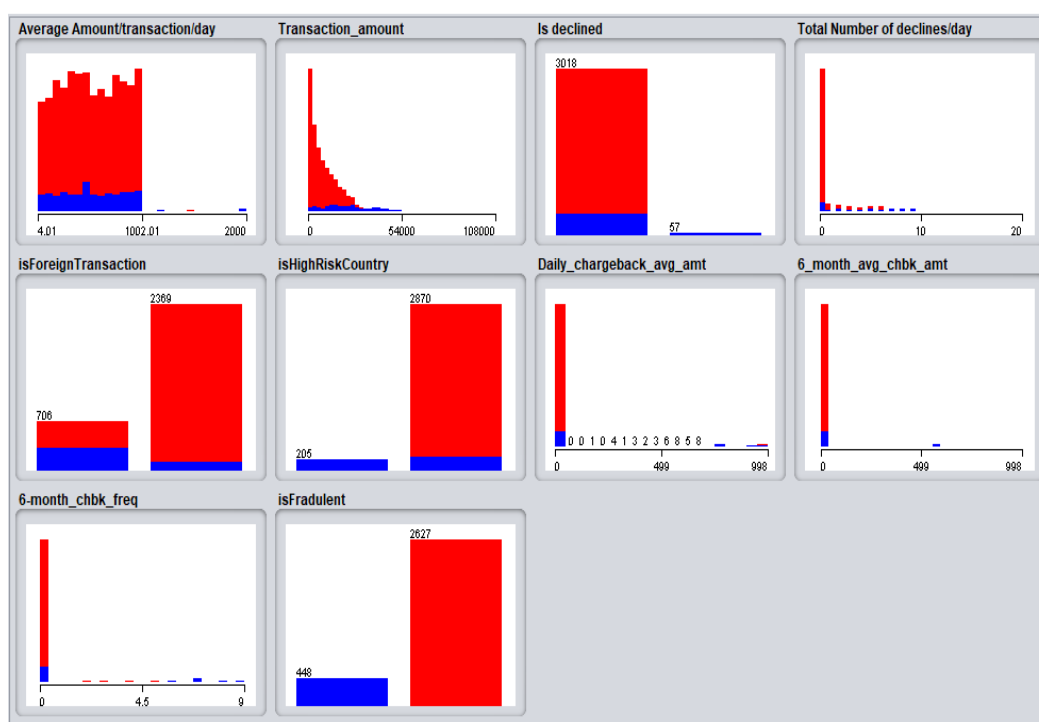
Εικόνα 17 (Η συναλλαγή αποτελεί απάτη)[25]

Προκειμένου να πειραματιστούμε με τα δεδομένα, αρχικά εισήχθηκαν στο περιβάλλον του Weka. Μια ψηφιοποίηση των attributes του συνόλου δεδομένων (χωρίς επεξεργασία) παρατίθεται παρακάτω.



Εικόνα 18 (Ψηφιοποίηση Attributes)

Όπως φαίνεται παραπάνω, τα attributes 'Transaction date' αλλά και το 'Merchant id' δεν προσφέρουν κάποια χρήσιμη πληροφορία γι' αυτό και αφαιρέθηκαν εντελώς από τα πειράματα, αφού το πρώτο δεν περιέχει καθόλου δεδομένα, ενώ το δεύτερο δεν περιέχει κάποια χρήσιμη πληροφορία. Επομένως, τα συγκεκριμένα attributes αφαιρέθηκαν από τα πειράματα που πραγματοποιήθηκαν, αφού προκάλεσαν προβλήματα στην ταξινόμηση των attributes. Στη συνέχεια, πριν περάσουμε στην προεπεξεργασία των δεδομένων, μπορούμε να παρατηρήσουμε στα παραπάνω γραφήματα πως τα δεδομένα που έχουμε στη διάθεση μας είναι αρκετά μη ισορροπημένα. Τα δεδομένα επίσης δεν έχουν missing values(τιμές που λείπουν από το σύνολο δεδομένων), επομένως δεν χρειάζεται να κάνουμε κάποια κίνηση γι' αυτό. Προκειμένου να εξετασθεί η περίπτωση που τα δεδομένα χρειάζονται “καθαρισμό”, δοκιμάσαμε τον εντοπισμό και την αφαίρεση των τιμών outliers και extreme values, τις τιμές δηλαδή που έχουν ενταχθεί λανθασμένα στο σύνολο δεδομένων. Κάτι τέτοιο όμως προκάλεσε ακόμη ένα πρόβλημα, αφού όταν αφαιρέθηκαν αυτές οι τιμές, από το σύνολο δεδομένων εξαλείφθηκε μεγάλο μέρος των δεδομένων που είχαν μικρή τιμή, κάτι το οποίο θα μας επιφέρει λανθασμένα αποτελέσματα. Επομένως αποφασίσαμε να μην αφαιρέσουμε τις τιμές outliers και extreme values, αφού χρειαζόμαστε τις τιμές με χαμηλή συχνότητα εμφάνισης.



Εικόνα 19 (Τελική επιλογή Attributes)

Τελειώνοντας με το στάδιο της προ επεξεργασίας δεδομένων, προχωρήσαμε στο στάδιο της ταξινόμησης προκειμένου να δούμε τα αποτελέσματα των διαφορετικών αλγορίθμων που θα δοκιμασθούν, για να δούμε την αποτελεσματικότητα των συγκεκριμένων αλγορίθμων. Παρακάτω λοιπόν παραθέτονται τα αποτελέσματα της ταξινόμησης (classify) με τους διαφορετικούς αλγορίθμους ταξινόμησης σε συνδυασμό με την επιλογή χαρακτηριστικών (attribute selection). Το attribute σύμφωνα με το οποίο πραγματοποιήθηκε η ταξινόμηση, είναι το isFraudulent το οποίο έχει τις τιμές “yes” και “no”, ανάλογα με το αν μια συναλλαγή είναι απάτη ή όχι. Ταξινομώντας λοιπόν το σύνολο δεδομένων με βάση το χαρακτηριστικό IsFraudulent (γνώρισμα κλάσης), θα προβλεφθεί αν μια συναλλαγή αποτελεί απάτη, με τη βοήθεια και των υπόλοιπων χαρακτηριστικών του data set, αλλά και διάφορες μεταβλητές γύρω από αυτά όπως την ακρίβεια, την ανάκληση κλπ. Θα παραθέσουμε τα αποτελέσματα χρησιμοποιώντας πρώτα όλα τα χαρακτηριστικά για την ταξινόμηση ενώ στη συνέχεια, θα επιλεγεί ο κατάλληλος αριθμός χαρακτηριστικών που κάθε φορά θα είναι ο βέλτιστος προκειμένου να έχουμε τα βέλτιστα αποτελέσματα. Προκειμένου να πραγματοποιηθεί η ταξινόμηση μέσω επιλογής χαρακτηριστικών, θα χρησιμοποιηθεί η μέθοδος CfsSubsetEval για αξιολόγηση και η μέθοδος Best First ως μέθοδος αναζήτησης προκειμένου σε κάθε περίπτωση να βρεθεί ο βέλτιστος αριθμός χαρακτηριστικών αλλά και ο συνδυασμός των καταλληλότερων χαρακτηριστικών για το βέλτιστο αποτέλεσμα. Αυτός ο συνδυασμός προέκυψε μετά από διάφορους πειραματισμούς με τις υπόλοιπες μεθόδους αξιολόγησης-αναζήτησης, ενώ χρησιμοποιήθηκε η μέθοδος του Cross-validation σε 10 folds. Ο συγκεκριμένος συνδυασμός

48



Classifier Model
Naive Bayes Classifier

Attribute	Class	
	Y (0.15)	N (0.85)
=====		
Average Amount/transaction/day		
mean	531.6178	512.1911
std. dev.	320.1068	286.675
weight sum	448	2627
precision	0.662	0.662
Transaction_amount		
mean	22855.794	7662.9991
std. dev.	15201.4276	6868.2676
weight sum	448	2627
precision	36.3392	36.3392
Is declined		
N	401.0	2619.0
Y	49.0	10.0
[total]	450.0	2629.0
Total Number of declines/day		
mean	3.6964	0.434
std. dev.	3.5603	1.3417
weight sum	448	2627
precision	2	2
isForeignTransaction		
Y	322.0	386.0
N	128.0	2243.0
[total]	450.0	2629.0
isHighRiskCountry		
Y	204.0	3.0
N	246.0	2626.0
[total]	450.0	2629.0
Daily_chargeback_avg_amt		
mean	248.9937	22.8113
std. dev.	376.7838	135.4322
weight sum	448	2627
precision	7.5606	7.5606
6_month_avg_chbk_amt		
mean	181.84	15.8301
std. dev.	290.0183	100.5238
weight sum	448	2627
precision	6.8828	6.8828
6-month_chbk_freq		
mean	2.0558	0.1081
std. dev.	3.2035	0.7092
weight sum	448	2627
precision	1	1

Εικόνα 22 (Naive Bayes με χρήση όλων των χαρακτηριστικών)

Naive Bayes Classifier

Attribute	Class	
	Y (0.15)	N (0.85)
=====		
Transaction_amount		
mean	22855.794	7662.9991
std. dev.	15201.4276	6868.2676
weight sum	448	2627
precision	36.3392	36.3392
Total Number of declines/day		
mean	3.6964	0.434
std. dev.	3.5603	1.3417
weight sum	448	2627
precision	2	2
isForeignTransaction		
Y	322.0	386.0
N	128.0	2243.0
[total]	450.0	2629.0
isHighRiskCountry		
Y	204.0	3.0
N	246.0	2626.0
[total]	450.0	2629.0
6-month_chbk_freq		
mean	2.0558	0.1081
std. dev.	3.2035	0.7092
weight sum	448	2627
precision	1	1

Εικόνα 23 (Naive Bayes με χρήση 5 χαρακτηριστικών(βέλτιστος αριθμός χαρακτηριστικών))

- Ενδεικτικό παράδειγμα Random Forest με χρήση όλων των χαρακτηριστικών. Τα αποτελέσματα δεν παραθέτονται σε πλήρη μορφή για οικονομία χώρου και χρόνου, γι' αυτό και παρατίθεται ενδεικτικά το συγκεκριμένο κομμάτι των αποτελεσμάτων του αλγορίθμου.

isForeignTransaction = Y

```
| Transaction_amount < 25500.07
| | Transaction_amount < 9473.1
| | | isHighRiskCountry = Y : Y (53/0)
| | | isHighRiskCountry = N
| | | | 6-month_chbk_freq < 6.5
| | | | | Total Number of declines/day < 7
| | | | | | Average Amount/transaction/day < 279.9
| | | | | | | Total Number of declines/day < 2.5
| | | | | | | | Average Amount/transaction/day < 275.27
| | | | | | | | | Average Amount/transaction/day < 112.5
| | | | | | | | | | Total Number of declines/day < 1
| | | | | | | | | | | Average Amount/transaction/day < 55.2
| | | | | | | | | | | | Average Amount/transaction/day < 44.72 : N (9/0)
| | | | | | | | | | | | | Average Amount/transaction/day >= 44.72 : Y (1/0)
| | | | | | | | | | | | | | Average Amount/transaction/day >= 55.2 : N (25/0)
| | | | | | | | | | | | | | | Total Number of declines/day >= 1
| | | | | | | | | | | | | | | | Average Amount/transaction/day < 68.39 : N (1/0)
| | | | | | | | | | | | | | | | | Average Amount/transaction/day >= 68.39 : Y (2/0)
| | | | | | | | | | | | | | | | | | Average Amount/transaction/day >= 112.5 : N (33/0)
| | | | | | | | | | | | | | | | | | | Average Amount/transaction/day >= 275.27 : Y (3/0)
| | | | | | | | | | | | | | | | | | | | Total Number of declines/day >= 2.5 : N (11/0)
| | | | | | | | | | | | | | | | | | | | | Average Amount/transaction/day >= 279.9 : N (172/0)
| | | | | | | | | | | | | | | | | | | | | | Total Number of declines/day >= 7 : Y (3/0)
```

- Ενδεικτικό παράδειγμα Random Forest με χρήση 5 χαρακτηριστικών (βέλτιστος αριθμός)

isHighRiskCountry = Y

| Transaction_amount < 35970.53 : Y (194/0)

| Transaction_amount >= 35970.53

| | Transaction_amount < 36159.03 : N (3/0)

| | Transaction_amount >= 36159.03 : Y (36/0)

isHighRiskCountry = N

| 6-month_chbk_freq < 5.5

| | Transaction_amount < 29915.05

| | | isForeignTransaction = Y

| | | | Total Number of declines/day < 6.5

| | | | | Total Number of declines/day < 1.5

| | | | | | 6-month_chbk_freq < 2.5

| | | | | | | Transaction_amount < 8525.65

| | | | | | | | Transaction_amount < 1752.47

| | | | | | | | | 6-month_chbk_freq < 0.5

| | | | | | | | | | Total Number of declines/day < 0.5

| | | | | | | | | | | Transaction_amount < 1708.13 : N (77/0)

| | | | | | | | | | | Transaction_amount >= 1708.13 : Y (1/0)

| | | | | | | | | | | Total Number of declines/day >= 0.5 : N (1/0)

| | | | | | | | | | | 6-month_chbk_freq >= 0.5 : N (1/0)

| | | | | | | | | | | Transaction_amount >= 1752.47 : N (124/0)

| | | | | | | | | | | Transaction_amount >= 8525.65

| | | | | | | | | | | Transaction_amount < 8637.87 : Y (3/0)

| | | | | | | | | | | Transaction_amount >= 8637.87

| | | | | | | | | | | 6-month_chbk_freq < 1

| | | | | | | | | | | Total Number of declines/day < 0

- Ενδεικτικό παράδειγμα JRip με χρήση όλων των χαρακτηριστικών

JRIP rules:

=====

(isHighRiskCountry = Y) => isFradulent=Y (205.0/2.0)

(Total Number of declines/day >= 7) => isFradulent=Y (84.0/2.0)

(6-month_chbk_freq >= 6) => isFradulent=Y (95.0/8.0)

(isForeignTransaction = Y) and (Transaction_amount >= 26704.41684) => isFradulent=Y (28.0/3.0)

(isForeignTransaction = Y) and (Transaction_amount >= 13074.61064) and (Average Amount/transaction/day <= 452.309247) => isFradulent=Y (15.0/0.0)

(Transaction_amount >= 21143.86074) and (Average Amount/transaction/day <= 699.512957) => isFradulent=Y (9.0/0.0)

(isForeignTransaction = Y) and (Average Amount/transaction/day <= 276.825748) and (Transaction_amount >= 5680.128626) => isFradulent=Y (10.0/3.0)

=> isFradulent=N (2629.0/20.0)

Number of Rules : 8

- Ενδεικτικό παράδειγμα JRip με χρήση 5 χαρακτηριστικών(βέλτιστος αριθμός)

JRIP rules:

=====

(isHighRiskCountry = Y) => isFraudulent=Y (205.0/2.0)

(Total Number of declines/day >= 7) => isFraudulent=Y (84.0/2.0)

(6-month_chbk_freq >= 6) => isFraudulent=Y (95.0/8.0)

(isForeignTransaction = Y) and (Transaction_amount >= 26704.41684) => isFraudulent=Y
(28.0/3.0)

(isForeignTransaction = Y) and (Transaction_amount >= 14392.05359) and (Total Number of declines/day >= 4) => isFraudulent=Y (15.0/5.0)

(6-month_chbk_freq >= 1) and (Transaction_amount >= 22436.79651) => isFraudulent=Y
(10.0/3.0)

=> isFraudulent=N (2638.0/34.0)

Number of Rules : 7

Παρακάτω, παραθέτονται τα αποτελέσματα που προέκυψαν από την ταξινόμηση του data set με τους διαφορετικούς αλγόριθμους που παραθέτονται προκειμένου να γίνει η ανάλογη σύγκριση και αξιολόγηση τους.

Πίνακας 2 (Αποτελέσματα χρήσης όλων των χαρακτηριστικών για ταξινόμηση)

	J48	Naive Bayes	Random Forest	K-nearest neighbors	JRIP
Correctly identified instances (%)	98.17%	94.92%	98.47%	97.98%	97.91%
Precision	98,2%	95%	98,5%	98%	97,9%
Recall	98,2%	94,9%	98,5%	98%	97,9%
TP	98,2%	94,9%	98,5%	98%	97,9%
FP	5,9%	14,2%	5,1%	6,3%	7%
F-measure	98,2%	95%	98,5%	98%	97.9%

Πίνακας 3 (Αποτελέσματα χρήσης βέλτιστου αριθμού χαρακτηριστικών για ταξινόμηση)

	J48	Naive Bayes	Random Forest	K-nearest neighbors	JRIP
Correctly identified instances (%)	97.49%	95.93%	97.39%	97.07%	97.33%
Precision	97,5%	96,2%	97,4%	97.1%	97,3%
Recall	97,5%	95,9%	97,4%	97.1%	97,3%
TP	97,5%	95,9%	97,4%	97.1%	97,3%
FP	9,5%	7,2%	7,5%	9,2%	9,3%
F-measure	97,5%	96%	97,4%	97.1%	97,3%

Τα αποτελέσματα που παρατέθηκαν στους πίνακες προέρχονται από τη στήλη Weighted Avg των αποτελεμάτων των αλγορίθμων.

Κεφάλαιο 5 “Συμπεράσματα ανάλυσης δεδομένων”

5.1 Αριθμητικά συμπεράσματα πειράματος

Τα πειράματα στα οποία χρησιμοποιήθηκαν όλα τα χαρακτηριστικά για ταξινόμηση, έδειξαν πως ο Random Forest είχε το καλύτερο αποτέλεσμα όσον αφορά τις σωστά ταξινομημένες περιπτώσεις, μετρώντας 98.47 % με TP Rate, Precision, Recall και F-measure στο 98,5% και FP Rate 5,1%. Δεύτερος καλύτερος αλγόριθμος, έρχεται ο J48 με 98.17% όσον αφορά τις σωστά ταξινομημένες περιπτώσεις, ενώ μετρά 98,2% σε TP Rate, Precision, Recall, F-measure και ποσοστό 5,9% σε FP Rate. Λίγο χαμηλότερα, βρίσκουμε τα αποτελέσματα του K-nearest neighbors με ποσοστό σωστά ταξινομημένων περιπτώσεων στο 97.98%, ενώ όσον αφορά τα υπόλοιπα χαρακτηριστικά, το TP Rate, Precision, Recall, F-measure βρίσκονται στο 98% και το FP Rate στο 6,3. Χαμηλότερα στην κατάταξη έχουμε τον JRIP με ποσοστό σωστά ταξινομημένων μεταβλητών στο 97.91 % ενώ στα υπόλοιπα χαρακτηριστικά, TP Rate, Recall, Precision και F-measure στο 97,9 και FP Rate 7%. Τελευταίος στην λίστα, έρχεται ο Naive Bayes με ποσοστό σωστά ταξινομημένων μεταβλητών στο 94.92%, Recall, TP Rate στο 94,9, Precision και F-measure στο 95% και FP Rate στο 14,2%.

Όσον αφορά τα πειράματα στα οποία χρησιμοποιήθηκαν 5 χαρακτηριστικά η κατάταξη των αλγορίθμων ως προς τα βέλτιστα αποτελέσματα, ο J48 είχε το καλύτερο αποτέλεσμα όσον αφορά τις σωστά ταξινομημένες περιπτώσεις με ποσοστό 97.49%, ενώ TP Rate, Precision, Recall, F-measure έρχονται σε ποσοστό 97,5% και FP Rate σε 9,5%. Δεύτερος στη σειρά έρχεται ο Random Forest με ποσοστό σωστά ταξινομημένων περιπτώσεων στο 97,39%, ενώ TP Rate, Precision, Recall, F-measure έρχονται σε ποσοστό 97,4% και με FP Rate σε ποσοστό 7,5%. Τρίτος στη σειρά έρχεται ο JRIP με ποσοστό σωστά ταξινομημένων περιπτώσεων στο 97,33%, TP Rate, Precision, Recall, F-measure έρχονται σε ποσοστό 97,3% και FP Rate σε ποσοστό 9,3%. Στη συνέχεια συναντάμε τον K-nearest neighbors με ποσοστό σωστά ταξινομημένων μεταβλητών στο 97,07%, TP Rate, Recall, Precision, F-measure στο 97.10% και FP Rate 9,2%. Τέλος, έχουμε τον Naive Bayes με ποσοστό σωστά ταξινομημένων μεταβλητών στο 95,93%, ενώ Recall, TP Rate στο 95,9%, F-measure με ποσοστό 96%, Precision με 96,2% και FP Rate 7,2%.

Μελετώντας τα αποτελέσματα που μας δόθηκαν από τους αλγόριθμους, μπορούμε να δούμε πως όλοι πέτυχαν ποσοστό άνω του 90% όσον αφορά τις σωστά ταξινομημένες μεταβλητές. Όμως στη συγκεκριμένη περίπτωση δεν μπορούμε να βασιστούμε μόνο στο ποσοστό σωστά ταξινομημένων μεταβλητών για να βγάλουμε συμπεράσματα, καθώς στο αρχικό data set, οι περιπτώσεις που έχουμε απάτη(448 περιπτώσεις) είναι αρκετά λιγότερες από αυτές που δεν έχουμε κάποια απάτη(2627 περιπτώσεις), στο σύνολο των 3075 συναλλαγών. Γι'

αυτό θα πρέπει να λάβουμε υπόψη και άλλες μεταβλητές προκειμένου να εξάγουμε συμπεράσματα. Για παράδειγμα, σύμφωνα με τα ποσοστά το οποίο εξάγονται από τα confusion matrix μπορούμε να αντιληφθούμε πολλά περισσότερα πράγματα για τα δεδομένα. Η στήλη που παρουσιάζει αρκετό ενδιαφέρον είναι η FN(false negative) αφού αυτή παρουσιάζει τις περιπτώσεις στις οποίες μια απάτη ταξινομήθηκε ως μη απάτη, αλλά στην πραγματικότητα ήταν απάτη. Όπως μπορούμε να δούμε την καλύτερη απόδοση(μιλώντας πάντα για τη χρήση όλων των χαρακτηριστικών) είχε ο Random Forest ο οποίος είχε 26 λανθασμένα αρνητικές περιπτώσεις ενώ ο Naive Bayes είχε την χειρότερη απόδοση με 72 λανθασμένα αρνητικές περιπτώσεις. Βλέπουμε λοιπόν πως ο Random Forest, ο οποίος προέβλεψε 422 απάτες ,άλλες 26 φορές προέβλεψε λάθος, πράγμα που σημαίνει πως μέσα από τις συνολικές συναλλαγές, 26 φορές πραγματοποιήθηκε απάτη χωρίς να ανιχνευθεί από το σύστημα. Παράλληλα, βλέπουμε το ποσοστό λανθασμένων προβλέψεων να είναι χειρότερο με την επιλογή χαρακτηριστικών, καθώς έχουμε 38 λανθασμένες προβλέψεις. Το σύστημα μας λοιπόν μπορούμε να πούμε πως απέδωσε αρκετά καλά σύμφωνα με τα ποσοστά που είδαμε. Οι αριθμοί πραγματικών απατών που δεν ανιχνεύθηκαν δεν είναι ιδιαίτερα υψηλοί σαν απλά ποσοστά, δεν παύουν όμως να αποτελούν πρόβλημα, καθώς 26 ή ακόμα χειρότερα, 38 απάτες μπορεί να προκαλέσουν μεγάλο πρόβλημα τόσο στους κατόχους καρτών, τόσο και στους διάφορους οργανισμούς (τράπεζες, επιχειρήσεις κλπ). Όπως μπορούμε να συμπεράνουμε,τα ποσοστά που πέτυχαν οι αλγόριθμοι είναι καλά σύμφωνα με τις δυνατότητες που παρέχει και το data set (μεγάλη ανισορροπία δεδομένων), υπάρχουν όμως και πολλά περιθώρια βελτίωσης προκειμένου τέτοιες απάτες να μειωθούν έως και να εξαλειφθούν αφού ακόμα και οι αριθμοί οι οποίοι είναι μικροί, μπορούν να προκαλέσουν ζημιά.

5.2 Επίλογος

Κλείνοντας αυτή την εργασία και λαμβάνοντας υπόψιν όλα τα παραπάνω, μπορούμε να συμπεράνουμε πως η μηχανική μάθηση είναι ένα πολύ σημαντικό εργαλείο στην αντιμετώπιση των απατών στις πιστωτικές κάρτες. Το ακόμα σημαντικότερο γεγονός είναι οι δυνατότητες εξέλιξης που έχει ο συγκεκριμένος τομέας κάτι το οποίο φάνηκε στις βιβλιογραφικές αναλύσεις, αφού οι περισσότεροι συγγραφείς επισήμαναν τις μεγάλες και αρκετές δυνατότητες που υπάρχουν προκειμένου να αντιμετωπιστεί με μεγαλύτερη ακρίβεια και το πρόβλημα των απατών. Όπως είδαμε, μπορούμε να χρησιμοποιήσουμε διαφορετικές προσεγγίσεις και διαφορετικούς αλγόριθμους για κάθε περίπτωση, ενώ σημαντικό ρόλο παίζει και το σύνολο των δεδομένων που μελετάμε ώστε να επιλέξουμε την κατάλληλη προσέγγιση. Στις περισσότερες περιπτώσεις, τα υβριδικά μοντέλα φάνηκαν πιο χρήσιμα αφού οι ερευνητές είχαν τη δυνατότητα να συνδυάσουν διαφορετικούς αλγορίθμους μαζί. Στο γενικότερο πρόβλημα των απατών στις πιστωτικές κάρτες, είδαμε πως πρέπει να γίνουν κατάλληλες ενέργειες από όλες τις

πλευρές(κράτος,τράπεζες,ιδιοκτήτες καταστημάτων, πελάτες) προκειμένου να περιοριστεί το συγκεκριμένο πρόβλημα, αφού ήδη αποτελεί μεγάλο πρόβλημα εδώ και χρόνια, ενώ όπως φαίνεται θα απασχολήσει την παγκόσμια οικονομία. Αυτό συμβαίνει λόγω των μεγάλων ποσών που έχουν υποκλαπεί μέσω των απατών, χρίζοντας την υποκλοπή καρτών ένα από τα σημαντικότερα προβλήματα στην παγκόσμια οικονομία. Όπως είδαμε, πολλές χώρες αντιμετωπίζουν σοβαρά προβλήματα αφού τα χαμένα ποσά λόγω απατών είναι πολύ μεγάλα ενώ παράλληλα ακόμα και οι μεγάλες τράπεζες παγκόσμιας εμβέλειας έχουν πληγεί από τέτοιου είδους υποθέσεις. Κλείνοντας, θα ήθελα να αναφερθώ για άλλη μια φορά στο πόσο προσεκτικοί πρέπει να είναι οι απλοί καταναλωτές με τις προσωπικές τους πληροφορίες, καθώς είναι πολύ εύκολο για κάποιον απατεώνα να υποκλέψει και να αποσπάσει μεγάλα χρηματικά ποσά χρησιμοποιώντας τέτοιες πληροφορίες. Γι' αυτό το λόγο, θα πρέπει ο καθένας να αναλογιστεί σε ποιον επιτρέπει να έχει πρόσβαση σε ευαίσθητες πληροφορίες, ενώ θα είναι συνετό οι καταναλωτές να ενημερώνονται σωστά από έγκυρες πηγές προκειμένου να “προλάβουν” τέτοιου είδους καταστάσεις και να μην χρειαστεί να έρθουν σε δυσάρεστη θέση αντικρίζοντας κάποια απάτη εις βάρους τους. Η σωστή ενημέρωση και η πρόληψη λοιπόν, είναι τα καταλληλότερα εργαλεία προκειμένου να αντιμετωπίσει ένας απλός καταναλωτής μια τέτοιου είδους απειλή προστατεύοντας την ιδιωτικότητα αλλά και τον τραπεζικό του λογαριασμό, πράγμα πολύ σημαντικό στις μέρες μας.

Βιβλιογραφία

- [1] <https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport201809.en.html>
- [2] <https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport202008~521edb602b.en.html>
- [3] <https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport202008~521edb602b.en.html>
- [4] <https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport202008~521edb602b.en.html>
- [5] <https://nilsonreport.com/mention/407/1link/>
- [6] <https://medium.com/analytics-vidhya/fundamental-omachine-learning-ada28afa1bd3>
- [7] <https://becominghuman.ai/the-key-differences-between-rule-based-ai-and-machine-learning-8792e545e6>
- [8] https://www.researchgate.net/figure/Comparison-of-properties-between-rule-based-and-machine-learning-ML-based-approaches_tbl2_291419651
- [9] <https://hub.packtpub.com/machine-learning-ethics-what-you-need-to-know-and-what-you-can-do/>
- [10] Sadgali, N. Sael, F. Benabbay, “Performance of machine learning techniques in the detection of financial frauds”, 'Procedia Computer Science' Vol.148 (2019) p. 45–54
- [11] Vaishnavi N. Dornadulaa, Geetha S. (2019), “Credit Card Fraud Detection using Machine Learning Algorithms”, 'Procedia Computer Science' Vol. 165 (2019) p.631–641
- [12] N. Carneiro, G. Figueira, M.Costa (2017), “A data mining based system for credit-card fraud detection in e-tail”, 'Decision Support Systems' Vol.95(2017) p. 91-101
- [13] Nick F. Rymman-Tubb, P. Krause, W. Garn (2018), “How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark”, 'Engineering Applications of Artificial Intelligence' Vol. 76, p.130-157
- [14] E. Kim, J. Lee, H. Shin, H. Yang, S. Cho, S. Nam, Y. Song, J. Yoon, J.Kim (2019) “Champon-challenger analysis for credit card fraud detecion:Hybrid ensemble and deep learning”, 'Expert Systems with Applications' (2019) Vol.128 p.214-224
- [15] C. Phua, V. Lee, K. Smith, R. Gayler (2010) “A Comprehensive Survey of Data Mining-based Fraud Detection Research”
- [16] S. Dhankhad, E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," 2018 IEEE International Conference on Information Reuse and Integration

(IRI), 2018, pp. 122-125, doi: 10.1109/IRI.2018.00025

[17]P.Kumar, F. Iqbal (2019)"Credit Card Fraud Identification Using Machine Learning Approaches", Published in: 2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT)2019, pp. 1-4, doi: 10.1109/ICIICT1.2019.8741490

[18]S. Mittal, S. Tyagi (2019)"Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection", Published in: [2019 9th International Conference on Cloud Computing, Data Science & Engineering \(Confluence\)](#)2019, pp. 320-324, doi: 10.1109/CONFLUENCE.2019.8776925.

[19] T. Choudhury, G. Dangi, T. Pal Singh, A.Chauhan, A. Aggarwal(2018)"An Efficient Way to Detect Credit Card Fraud Using Machine Learning Methodologies"Published in: [2018 Second International Conference on Green Computing and Internet of Things \(ICGCIoT\)](#)2018, pp. 591-597, doi: 10.1109/ICGCIoT.2018.8753077.

[20]John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare (2017), "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis"Published in: [2017 International Conference on Computing Networking and Informatics \(ICCNI\)](#), 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.

[21]R. R. Popat, J.Chaudhary (2018)"A Survey on Credit Card Fraud Detection using Machine Learning", Published in:[2018 2nd International Conference on Trends in andInformatics\(ICOEI\)](#),2018, pp. 1120-1125, doi:10.1109/ICOEI.2018.8553963

[22]D. Dighe, S. Patil, S. Kokate (2018)"Detection of Credit Card Fraud Transactions using Machine Learning Algorithms and Neural Networks: A Comparative Study"Published in :[2018 Fourth International Conference on Computing Communication Control and Automation \(ICCUBEA\)](#),2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697799.

[23]M. Puh, L. Brkic(2019), "Detecting Credit Card Fraud Using Selected Machine Learning Algorithms"Published in:[2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics \(MIPRO\)](#), 2019, pp. 1250-1255, doi: 10.23919/MIPRO.2019.8757212.

[24]S. Bakshi (2018)“Credit Card Fraud Detection: A classification analysis”Published in:[2018 2nd International Conference on I-SMAC \(IoT in Social, Mobile, Analytics and Cloud\)\(I-SMAC\)I-SMAC \(IoT in Social, Mobile, Analytics and Cloud\) \(I-SMAC\)](#), 2018,pp. 152-156, doi: 10.1109/I-SMAC.2018.8653770.

[25]<https://www.kaggle.com/shubhamjoshi2130of/abstract-data-set-for-credit-card-fraud-detection>