



# ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Μηχανικών Παραγωγής Και  
Διοίκησης

## **“ΜΕΘΟΔΟΛΟΓΙΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΓΙΑ ΤΟΝ ΕΝΤΟΠΙΣΜΟ ΤΗΣ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ ΑΠΑΤΗΣ: ΜΙΑ ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ”**

Διπλωματική εργασία που υπεβλήθη για τη μερική ικανοποίηση  
των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος

**Υπό**

Παναγιώτη Διγενάκη

**Επίβλεψη**

Μιχάλης Δούμπος

Χανιά 2022

Στους γονείς μου....

# ΕΥΧΑΡΙΣΤΙΕΣ

Η εκπόνηση της παρούσας Διπλωματικής εργασίας πραγματοποιήθηκε στα πλαίσια της ολοκλήρωσης της παρακολούθησης του μεταπτυχιακού προγράμματος Διοίκησης Επιχειρήσεων της σχολής «Μηχανικών Παραγωγής και Διοίκησης».

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Μιχάλη Δούμπο, για την καθοδήγηση του και τις πολύτιμες συμβουλές που μου παρείχε καθ' όλη την διάρκεια της εργασίας.

Οφείλω επίσης ένα μεγάλο ευχαριστώ στους γονείς μου, που βρίσκονται πάντα δίπλα μου, με στηρίζουν συνεχώς σε κάθε επιλογή μου και με ενθαρρύνουν να ακολουθώ τα όνειρα μου .

# ΠΕΡΙΛΗΨΗ

Η απάτη συνδέεται τόσο με το οργανωμένο έγκλημα όσο και με την τρομοκρατία και προκαλεί σημαντική οικονομική ζημιά. Από τις πιο συνηθισμένες μορφές απάτης είναι η χρηματοοικονομική απάτη, η οποία μπορεί να διακριθεί σε κατηγορίες, όπως στην παραποίηση χρηματοοικονομικών παραστάσεων, στην φοροδιαφυγή και φορολογική απάτη, στην απάτη πιστωτικών καρτών, στο ξέπλυμα χρήματος, κτλ. Όσοι διαπράττουν απάτη παίζουν ένα δυναμικό παιχνίδι γάτας με το ποντίκι με όσους προσπαθούν να τους εμποδίσουν. Η αποτροπή ενός συγκεκριμένου είδους απάτης δεν σημαίνει ότι οι απατεώνες εγκαταλείπουν, απλώς αποστασιοποιούνται έως ότου δημιουργήσουν νέα τακτική. Συγκεκριμένα βρίσκονται συνεχώς σε επιφυλακή για νέες οδούς για απάτη, για νέες αδυναμίες στο σύστημα. Δεδομένου λοιπόν, ότι τα οικονομικά συστήματα αναπτύσσονται διαρκώς, υπάρχουν πάντα νέες ευκαιρίες προς εκμετάλλευση.

Ο εντοπισμός της χρηματοοικονομικής απάτης γίνεται όλο ένα και πιο επιτακτικό ζήτημα, κατά καιρούς έχουν αναπτυχθεί διάφοροι μέθοδοι για τον εντοπισμό της. Στην παρούσα εργασία θα αναλυθούν οι μέθοδοι τεχνητής νοημοσύνης και ορισμένες στατιστικές τεχνικές.

Λέξεις κλειδιά: Χρηματοοικονομική Απάτη, Εντοπισμός Χρηματοοικονομικής Απάτης, Τεχνητή νοημοσύνη, Στατιστικές Τεχνικές.

# ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1 .....	- 1 -
ΕΙΣΑΓΩΓΗ.....	- 1 -
1.1 Γενικά.....	- 1 -
1.2 Διάρθρωση εργασίας .....	- 2 -
ΚΕΦΑΛΑΙΟ 2 .....	- 3 -
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΑΠΑΤΗ .....	- 3 -
2.1 Εισαγωγή.....	- 3 -
2.2 Ορισμός της απάτης.....	- 5 -
2.3 Ανίχνευση και Πρόληψη Απάτης.....	- 7 -
2.3.1 Ορισμοί .....	- 7 -
2.3.2 Τεχνικές Ανίχνευσης .....	- 8 -
2.3.3 Εποπτευόμενοι μέθοδοι .....	- 9 -
2.3.4 Μη Εποπτευόμενοι μέθοδοι.....	- 10 -
2.4 Περιγραφή ειδών χρηματοοικονομικής απάτης .....	- 11 -
2.4.1 Απάτες σε πιστωτικές κάρτες .....	- 11 -
2.4.2 Παραποιημένες χρηματοοικονομικές καταστάσεις .....	- 13 -
2.4.3 Φοροδιαφυγή και Φορολογική απάτη .....	- 16 -
2.4.4 Ξέπλυμα Χρήματος .....	- 17 -
2.1 Άλλα είδη απάτης .....	- 19 -
ΚΕΦΑΛΑΙΟ 3 .....	- 21 -
ΜΕΘΟΔΟΛΟΓΙΚΟ ΠΛΑΙΣΙΟ .....	- 21 -
3.1 Εισαγωγή.....	- 21 -
3.2 Εξόρυξη Δεδομένων .....	- 21 -
3.2.1 Ορισμός.....	- 22 -
3.2.2 Ιστορική Εξέλιξη.....	- 24 -
3.3 Στατιστικές Τεχνικές .....	- 25 -
3.3.1 Γραμμική Παλινδρόμηση (Linear Regression) .....	- 25 -

3.3.2	Διακριτική Ανάλυση (Discriminant Analysis) .....	- 26 -
3.3.3	Ανάλυση Logit και Probit.....	- 27 -
3.4	Πολυκριτήριες Μεθοδολογίες.....	- 28 -
3.4.1	Μέθοδος UTADIS.....	- 29 -
3.4.2	Μέθοδος MHDIS .....	- 30 -
3.5	Μεθοδολογίες Τεχνητής Νοημοσύνης .....	- 32 -
3.5.1	Δέντρα Αποφάσεων (Decision Trees).....	- 32 -
3.5.2	Νευρωνικά Δίκτυα (Neural Networks) .....	- 34 -
3.5.3	Bayesian Δίκτυα Πειοιθήσεων (Bayesian Belief Networks) . -	36 -
3.5.4	Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines-SVMs).....	- 37 -
3.5.5	Αλγόριθμος των Κ-πλησιέστερων Γειτόνων (K - Nearest Neighbour Method) .....	- 38 -
3.5.6	Γενετικοί Αλγόριθμοι (Genetic Algorithms - GA).....	- 39 -
ΚΕΦΑΛΑΙΟ 4.....		- 41 -
ΚΑΤΑΓΡΑΦΗ ΕΦΑΡΜΟΓΩΝ ΑΠΑΤΗΣ.....		- 41 -
4.1	Εισαγωγή.....	- 41 -
4.2	Γενικά.....	- 41 -
4.3	Μελέτη περίπτωσης- MetaFraud .....	- 45 -
4.4	Μελέτη περίπτωσης-Εξόρυξη Δεδομένων σε πιστωτικές κάρτες. -	48 -
ΚΕΦΑΛΑΙΟ 5.....		- 51 -
ΣΥΜΠΕΡΑΣΜΑΤΑ .....		- 51 -
ΒΙΒΛΙΟΓΡΑΦΙΑ.....		- 52 -

# ΚΕΦΑΛΑΙΟ 1

## ΕΙΣΑΓΩΓΗ

### 1.1 Γενικά

Η απάτη χωρίς αμφιβολία δεν αποτελεί ένα πρόσφατο και μοναδικό φαινόμενο στην σύγχρονη κοινωνία, ούτε καν μοναδικό στην ανθρωπότητα. Σε δραστηριότητες απάτης εμπλέκονται επίσης και ορισμένα είδη ζώων όπως πίθηκοι, χαμαιλέοντες, έντομα ραβδιών και άλλα, αν και ίσως θα έπρεπε να ταξινομηθεί η συμπεριφορά τους ως χειραγωγική αντί για ως δραστηριότητες απάτης. Δεδομένου ότι οι εγκληματικές ή άδικες ενέργειες είναι κατηγορίες των ανθρώπων ή έννοιες που δεν μπορούν να ισχύουν ευθέως για τα ζώα. Πράγματι το εάν οι δραστηριότητες είναι παράνομες ή εγκληματικές εξαρτάται από τους ισχύοντες κανόνες ή τη νομοθεσία η οποία ορίζει ρητά και επίσημα αυτές τις ενέργειες που απαιτούνται για να είναι δυνατή η ταξινόμηση της συμπεριφοράς ως απάτη.

Η απάτη, εξαιτίας της άνθησης του διαδικτύου και της εφεύρεσης άλλων σύγχρονων τεχνολογιών, έχει σημειώσει δραματική αύξηση και συνδέεται με συστήματα που σχετίζονται με όλες τις πτυχές του επιχειρηματικού κόσμου. Μερικά από αυτά τα κοινά συστήματα που συναντώνται, περιλαμβάνουν απάτη με πιστωτικές κάρτες, απάτη ηλεκτρονικών υπολογιστών, απάτη συναλλαγών ηλεκτρονικού εμπορίου, τηλεπικοινωνιακή απάτη και ασφαλιστική απάτη.

Με τον όρο «απάτη» (Fraud) ορίζεται η άδικη ή η εγκληματική εξαπάτηση που αποσκοπεί σε οικονομικό ή προσωπικό όφελος, μέσω της χρήσης ψευδών, λανθασμένων ή ημιτελών εγγράφων, πληροφοριών και αναπαραστάσεων.

Αρχικά έχουν αναπτυχθεί δυο κύριοι μηχανισμοί για την αποφυγή απάτης και ζημιών που οφείλονται σε δραστηριότητες απάτης. Οι μηχανισμοί αυτοί αφορούν τα συστήματα πρόληψης απάτης και ανίχνευσης απάτης. Η πρόληψη της απάτης είναι ο προληπτικός μηχανισμός με σκοπό να σταματήσει την εμφάνιση της απάτης. Τα συστήματα ανίχνευσης απάτης μπαίνουν στο παιχνίδι όταν οι απατεώνες καταφέρνουν να ξεπεράσουν τα συστήματα πρόληψης και ξεκινούν μια συναλλαγή απάτης. Ένας αποτελεσματικός τρόπος αποτροπής και εντοπισμού της απάτης, έχει αποδειχθεί ότι είναι οι τεχνολογίες που βασίζονται στη στατιστική και μηχανική μάθηση, αλλά οι απατεώνες είναι προσαρμοστικοί και συνήθως καταφέρνουν να βρουν τρόπους να τις παρακάμψουν. Οι υπάρχουσες τεχνικές ανίχνευσης απάτης για τις περισσότερες περιπτώσεις απάτης συνήθως μοιράζονται πολύ παρόμοιες αρχές εξόρυξης δεδομένων, οι οποίες είναι δυνατόν να διαφέρουν από πολλές απόψεις με εξειδικευμένες γνώσεις τομέα.

Στην συγκεκριμένη εργασία, θα πραγματοποιηθεί μια ανασκόπηση στην χρηματοοικονομική απάτη καθώς και θα αναφερθούν ορισμένα άλλα είδη απάτης. Επιπλέον θα γίνει αναφορά σε μεθοδολογίες τεχνητής νοημοσύνης για τον εντοπισμό της χρηματοοικονομικής απάτης.

## 1.2 Διάρθρωση εργασίας

Στο επόμενο κεφάλαιο (κεφάλαιο 2) που ακολουθεί, γίνεται μια σύντομη αναφορά στην έννοια της απάτης και σε ορισμένους όρους που σχετίζονται με αυτή. Περιγράφονται επίσης ορισμένες τεχνικές ανίχνευσης απάτης καθώς και οι εποπτευόμενη και μη εποπτευόμενη μέθοδοι. Τέλος παρουσιάζονται οι δημοφιλέστερες μορφές απάτης που έχουν εμφανιστεί.

Στο τρίτο κεφάλαιο πραγματοποιείται μια αναφορά στον ορισμό της εξόρυξης δεδομένων και παρουσιάζεται μια ιστορική αναδρομή της πορείας της. Στο τελευταίο μέρος του κεφαλαίου αναφέρονται οι πιο δημοφιλείς τεχνικές ταξινόμησης για τον εντοπισμό χρηματοοικονομικής απάτης.

Στο τέταρτο κεφάλαιο πραγματοποιείται μια σύνοψη-καταγραφή διάφορων εφαρμογών σε διάφορους τύπους απάτης (πχ. Απάτη πιστωτικών καρτών, παραποίηση οικονομικών καταστάσεων, φοροδιαφυγή, κλπ.). Στην συνέχεια γίνεται μια πιο εκτενής αναφορά σε κάποιες χαρακτηριστικές έρευνες που σχετίζονται με τις πιστωτικές κάρτες και τις οικονομικές καταστάσεις.

Στο πέμπτο κεφάλαιο πραγματοποιείται η ολοκλήρωση της παρούσας εργασίας, παρουσιάζοντας τα συμπεράσματα και τις παρατηρήσεις για μελλοντικά θέματα μελέτης.



## ΚΕΦΑΛΑΙΟ 2

# ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΑΠΑΤΗ

### 2.1 Εισαγωγή

Η χρηματοοικονομική απάτη μπορεί να επιφέρει σοβαρές συνέπειες στην μακροπρόθεσμη βιωσιμότητα ενός οργανισμού καθώς και αρνητικές επιπτώσεις τόσο στους εργαζομένους και στους επενδυτές του, όσο και στην οικονομία συνολικά. Πολλές από τις μεγαλύτερες χρεοκοπίες στην ιστορία κυρίως στην Η.Π.Α αλλά και στην Ευρώπη αφορούσαν εταιρείες που συμμετείχαν σε μεγάλες απάτες.

Ορισμένα από τα μεγαλύτερα σκάνδαλα ήταν της WorldCom, Enron, Lehman Brothers, Bernie Madoff, AIG (American International Group) και Waste Management που κόστισαν πολλά δισεκατομμύρια δολάρια σε επενδυτές και πιστωτές καθώς και την απώλεια χιλιάδων θέσεων εργασίας. Η Waste Management μια δημόσια εμπορική εταιρεία διαχείρισης απορριμμάτων των Ηνωμένων Πολιτειών Αμερικής, είχε αναφέρει πάνω από 1,7 δισεκατομμύρια δολάρια σε πλαστά κέρδη. Η Επιτροπή Κεφαλαιαγοράς καθόρισε ως ένοχο τον ιδιοκτήτη και πρώην Διευθύνων Σύμβουλο της εταιρείας μαζί με πολλά άλλα κορυφαία στελέχη.

Η Enron ήταν μια εταιρεία στον κλάδο της ενέργειας που έδρευε στο Χιούστον του Τέξας. Η εταιρεία έκρυψε δισεκατομμύρια δολάρια επισφαλών χρεών μέσω λογιστικών κενών (πρόκειται για μέθοδο που επιτρέπει σε άτομα ή εταιρείες να μεταφέρουν εισόδημα ή περιουσιακά στοιχεία από φορολογητέες καταστάσεις σε καταστάσεις με χαμηλότερους φόρους ή χωρίς φόρους), ενώ ταυτόχρονα διόγκωνε τα κέρδη της. Η επιτροπή κεφαλαιαγοράς (SEC) διαπίστωσε ότι ο Διευθύνων Σύμβουλος και ο πρώην Διευθύνων Σύμβουλος είχαν διατηρήσει χρέος δισεκατομμυρίων δολαρίων εκτός ισολογισμού της εταιρείας

Η WorldCom ήταν μια Αμερικάνικη εταιρεία στον κλάδο των τηλεπικοινωνιών, η οποία είχε διογκώσει τα περιουσιακά της στοιχεία περίπου κατά 11 δισεκατομμύρια δολάρια, καθιστώντας την ένα από τα μεγαλύτερα σκάνδαλα της ιστορίας. Το σκάνδαλο ανακαλύφθηκε από το τμήμα εσωτερικού ελέγχου της εταιρείας που ανακάλυψε περίπου 3,8 δισεκατομμύρια δολάρια σε λογαριασμούς απάτης. Ο Διευθύνων Σύμβουλος της εταιρείας καταδικάστηκε σε πολλά έτη φυλάκισης για απάτη, συνωμοσία και κατάθεση πλαστών εγγράφων.

Η AIG, μια πολυεθνική ασφαλιστική εταιρεία της Αμερικής, παρατηρήθηκε ότι είχε δεσμεύσει δάνεια ως έσοδα στα βιβλία της και ανάγκαζε τους πελάτες να χρησιμοποιούν ασφαλιστές με τους οποίους η εταιρεία είχε προϋπάρχουσες συμφωνίες αποπληρωμής. Αποκαλύφθηκε από την SEC ότι ο Διευθύνων Σύμβουλος της εταιρείας διέπραξε λογιστική απάτη ύψους σχεδόν 4 δισεκατομμυρίων δολαρίων. Ο Διευθύνων Σύμβουλος της εταιρείας, επίσης κρίθηκε ένοχος για χειραγώγηση των τιμών των μετοχών.

Η Lehman Brothers ήταν μια εταιρεία χρηματοοικονομικών υπηρεσιών με έδρα την Νέα Υόρκη. Πρόκειται για μια από τις μεγαλύτερες επενδυτικές τράπεζες των Ηνωμένων Πολιτειών. Αποκαλύφθηκε ότι η εταιρεία κατά τη διάρκεια της οικονομικής κρίσης του 2008, είχε κρύψει περισσότερα από 50 δισεκατομμύρια δολάρια σε δάνεια. Τα δάνεια αυτά είχαν συγκαλυφθεί ως πωλήσεις μέσω λογιστικών κενών.

Ο Bernie Madoff ήταν ένας πρώην μεσίτης χρηματιστηρίου, ο οποίος διεύθυνε το μεγαλύτερο σχέδιο Ponzi (το σχέδιο Ponzi αναφέρεται σε οτιδήποτε σχετίζεται με χρηματική απάτη που στηρίζεται σε πυραμίδα επενδυτών) στην ιστορία. Ο Madoff διεύθυνε την Bernard L. Madoff Investment Securities LLC. Ανακαλύφθηκε ότι μετά την οικονομική κρίση του 2008, ο Madoff είχε ξεγελάσει τους επενδυτές για περισσότερα από 64,8 δισεκατομμύρια δολάρια

Αξίζει να αναφερθεί ένα σκάνδαλο μιας ελληνικής εταιρείας, της Folli Follie. Η Folli Follie ιδρύθηκε το 1982 στην Ελλάδα. Η εταιρεία ξεκίνησε με την κατασκευή κοσμημάτων και λίγο αργότερα λάνσαρε μια σειρά ρολογιών. Ωστόσο το 2018 μετά από την καταγγελία που έγινε στην Νέα Υόρκη από το fund Quintessential Capital Management, η οποία σχετιζόταν με ψευδή ισολογισμό και την παρουσίαση λιγότερων πωλήσεων από των πραγματικών, ήρθε στο φως η μεγάλη οικονομική απάτη της εταιρείας.

Στο παρόν κεφάλαιο, θα επιχειρηθεί η παρουσίαση του ορισμού της απάτης και θα γίνει μια αναφορά στις έννοιες της ανίχνευσης και πρόληψης. Επιπλέον θα γίνει παρουσίαση των περισσότερο συνηθισμένων ειδών χρηματοοικονομικής απάτης. Πέρα από τα συγκεκριμένα είδη θα αναφερθούν και ορισμένες άλλες μορφές απάτης.

## 2.2 Ορισμός της απάτης

Στη βιβλιογραφία μπορεί κανείς να συναντήσει διάφορους ορισμούς για το τι είναι απάτη. Για παράδειγμα το λεξικό της Οξφόρδης ορίζει την απάτη ως «την εσφαλμένη ή εγκληματική εξαπάτηση που έχει σκοπό να οδηγήσει σε οικονομικό ή προσωπικό όφελος». Οι Van Vlasselaer et al. (2015) παρουσίασαν ένα πιο εμπειριστατωμένο και λεπτομερή χαρακτήρα του πολύπλευρου φαινομένου της απάτης: «Η απάτη είναι ένα ασυνήθιστο, καλά μελετημένο, ανεπαίσθητα συγκαλυμμένο, εξελισσόμενο στον χρόνο και συχνά προσεκτικά οργανωμένο έγκλημα που εμφανίζεται με πολλούς τύπους μορφών». Αυτός ο ορισμός τονίζει πέντε χαρακτηριστικά που σχετίζονται με ιδιαίτερες προκλήσεις, οι οποίες συνδέονται με την ανάπτυξη ενός συστήματος ανίχνευσης απάτης.

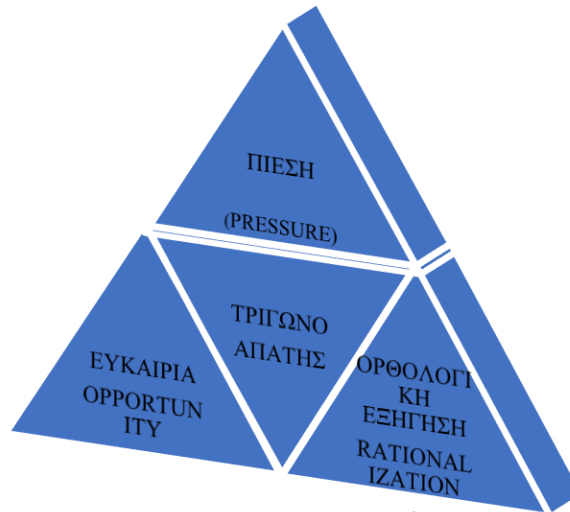
- Το πρώτο χαρακτηριστικό αφορά το γεγονός ότι η απάτη είναι ασυνήθιστη. Το γεγονός αυτό καθιστά δύσκολο τον εντοπισμό της, με αποτέλεσμα να εντοπίζεται μόνο ένας περιορισμένος αριθμός περιπτώσεων απάτης.
- Το δεύτερο χαρακτηριστικό αφορά το γεγονός ότι η απάτη αποκρύπτεται. Συχνά οι απατεώνες προσπαθούν να μην συμπεριφέρονται διαφορετικά από τους υπόλοιπους, ώστε να μην γίνονται αντιληπτοί και να παραμένουν καλυμμένοι.
- Η απάτη είναι εξελισσόμενη στο χρόνο. Οι απατεώνες μαθαίνουν και προσαρμόζονται στις μεθόδους εντοπισμού απάτης, προκειμένου να καταφέρουν να τις ξεπεράσουν. Συνεπώς, απαιτείται συνεχής ανάπτυξη συστημάτων ανίχνευσης απάτης.
- Τέταρτο χαρακτηριστικό αποτελεί ότι η απάτη είναι ένα προσεκτικά οργανωμένο έγκλημα. Συχνά η απάτη δεν είναι ένα μεμονωμένο γεγονός και έτσι για τον εντοπισμό της θα πρέπει να λαμβάνεται υπόψη το πλαίσιο, για παράδειγμα το κοινωνικό δίκτυο των απατεώνων.
- Το τελευταίο χαρακτηριστικό αφορά το γεγονός ότι απάτη μπορεί να συναντηθεί με διαφορετικούς τύπους. Συγκεκριμένα αναφέρεται στο ευρύ σύνολο τεχνικών και προσεγγίσεων που χρησιμοποιούν οι απατεώνες.

Η λέξη «άπατη» υποδηλώνει μια σκόπιμη πράξη που αποσκοπεί στην εξαπάτηση ή παραπλάνηση ενός άλλου μέρους. Μπορεί να ταξινομηθεί σε δυο τύπους: στην απάτη υπαλλήλων και στην απάτη ανώτατης διοίκησης. Γενικά η απάτη ανώτατης διοίκησης περιλαμβάνει τη παραποίηση λογιστικών αρχείων και συναλλαγών ή την εσφαλμένη εφαρμογή των λογιστικών αρχών. Η απάτη από τα ανώτατα στελέχη έχει καταστροφικές επιπτώσεις στους μετόχους και τους υπαλλήλους μιας εταιρείας και μπορεί να καταστρέψει τη φήμη και την αξιοπιστία της (Pai et al., 2011).

Με το να θεωρείται συχνά η απάτη ως οργανωμένο έγκλημα, προϋποθέτει ότι οι απατεώνες συχνά δεν λειτουργούν ανεξάρτητα, έχουν συμμάχους και μπορεί να προκαλέσουν αντιγραφές. Επιπλέον, διάφοροι τύποι απάτης όπως το ξέπλυμα του βρώμικου χρήματος και η απάτη «καρουζέλ» περιλαμβάνουν πολύπλοκες δομές που έχουν δημιουργηθεί με σκοπό τη διάπραξη απάτης με οργανωμένο τρόπο.

Ο Donald R. Cressey το 1953 στο βιβλίο του διατύπωσε μια υπόθεση, μέσω της οποίας προήλθε το τρίγωνο της απάτης (Σχ.2.1). Το τρίγωνο απάτης περιέχει μια πιο περίπλοκη εξήγηση για τα υποκείμενα κίνητρα ή τους οδηγούς για την διάπραξη απάτης και επιπλέον παρέχει μια χρήσιμη εικόνα για το φαινόμενο απάτη. Το μοντέλο αποτελείται από τρία σκέλη που μαζί δημιουργούν την συμπεριφορά απάτης (Baesens et al., 2015):

- Η πίεση (Pressure) είναι το πρώτο σκέλος και αφορά το κύριο κίνητρο για τη διάπραξη απάτης. Ένα άτομο θα διαπράξει απάτη επειδή αντιμετωπίζει μια πίεση ή μια οικονομική ή κοινωνική δυσκολία, που δεν μπορεί να επιλυθεί ή να εκτονωθεί με νόμιμο τρόπο.
- Η ευκαιρία (Opportunity) είναι το δεύτερο σκέλος του μοντέλου και αφορά την προϋπόθεση ότι ένα άτομο έχει την δυνατότητα να διαπράξει απάτη. Συγκεκριμένα, οι δραστηριότητες απάτης μπορούν να διαπραχθούν, μόνο όταν δοθεί η ευκαιρία σε ένα άτομο να επιλύσει δυσκολίες ή προβλήματα με μη νόμιμο ή εγκεκριμένο τρόπο.
- Η ορθολογική εξήγηση (Rationalization) είναι το τρίτο και τελευταίο σκέλος του μοντέλου, αποτελεί τον ψυχολογικό μηχανισμό που εξηγεί γιατί οι απατεώνες δεν απέχουν από την διάπραξη απάτης, αλλά θεωρούν την συμπεριφορά τους αποδεκτή.



**Σχήμα 2.1:** Τρίγωνο απάτης  
(Baesens et al., 2015)

## 2.3 Ανίχνευση και Πρόληψη Απάτης

### 2.3.1 Ορισμοί

Στο σημείο αυτό θα αναφερθούμε σε δυο στοιχεία που αποτελούν βασικά μέρη σχεδόν κάθε αποτελεσματικής στρατηγικής για την καταπολέμηση της απάτης. Συγκεκριμένα θα αναφερθούμε στην ανίχνευση και την πρόληψη της απάτης. Τα δυο αυτά εργαλεία έχουν ξεκάθαρη διαφορά, το πρώτο αφορά μια εκ των υστέρων προσέγγιση ενώ το δεύτερο αφορά μια εκ των προτέρων προσέγγιση.

Η πρόληψη απάτης αναφέρεται σε μέτρα που μπορούν να ληφθούν για την αποφυγή ή τον περιορισμό της. Με τη συνεχή εξέλιξη της τεχνολογίας, έχει δοθεί η δυνατότητα για την ανάπτυξη, μέσω, αλλά και συστημάτων για την πρόληψη απάτης. Η προσέγγιση αυτή έχει ως μειονέκτημα ότι μπορεί να αντιμετωπίσει, και κατά συνέπεια να εμποδίσει, μόνον γνωστές απάτες. Οι απατεώνες όμως, χρησιμοποιώντας προς όφελος τους την εξέλιξη της τεχνολογίας εντοπίζουν συχνά νέους τρόπους απάτης που παρακάμπτουν τους μηχανισμούς της πρόληψης.

Η ανίχνευση απάτης αναφέρεται στην ικανότητα αναγνώρισης ή ανακάλυψης δραστηριοτήτων απάτης. Η διαδικασία της ανίχνευσης ξεκινά τη στιγμή που η πρόληψη απάτης αποτύχει, συνεπώς όταν έχει διαπραχθεί ήδη η απάτη. Ωστόσο δεν πρέπει να θεωρείται ως βοηθητικό εργαλείο αντιμετώπισης της απάτης, διότι δεν είναι δυνατόν να προβλέψουμε πάντα πότε τα συστήματα προφύλαξης θα παρακάμπτονται από τους απατεώνες. Αντίθετα αποτελεί ένα σημαντικό εργαλείο, καθώς καλύπτει το κενό που δημιουργείται από την αποτυχία προφύλαξης έναντι απάτης και κατά συνέπεια της εμφάνισης νέων περιπτώσεων (Baesens et al., 2015).

Για παράδειγμα, σε μια περίπτωση που συναντάμε μεθόδους πρόληψης είναι οι μηχανισμοί Pin και Chips που δημιουργούνται για τα συστήματα πιστωτικών καρτών. Οι μηχανισμοί αυτοί όμως δεν εμποδίζουν τους πιο συνηθισμένους τύπους απάτης, όπως απάτες πιστωτικών καρτών μέσω εικονικών τερματικών POS (Point Of Sale) ή ταχυδρομικών παραγγελιών που ονομάζονται διαδικτυακές απάτες με πιστωτικές κάρτες. Ως αποτέλεσμα, η ανίχνευση της απάτης γίνεται το βασικό εργαλείο και πιθανώς ο καλύτερος τρόπος για να σταματήσουμε τέτοιους τύπους απάτης (Sahin et al., 2013).

Η ανίχνευση απάτης αν και δεν μας προστατεύει από αυτήν είναι σημαντική για την ενίσχυση και τη συνεχή εξέλιξη της πρόληψης. Ως εκ τούτου, τα δυο εργαλεία μπορούν και μάλιστα θα πρέπει να χρησιμοποιηθούν με συμπληρωματικό τρόπο για την επιδίωξη του κοινού στόχου, την μείωση της απάτης.

### 2.3.2 Τεχνικές Ανίχνευσης

Ο εντοπισμός απάτης είναι ένα δύσκολο πρόβλημα, καθώς οι απατεώνες καταβάλλουν κάθε δυνατή προσπάθεια να εμφανίσουν τη συμπεριφορά τους ως σύννομη. Μια ακόμη δυσκολία είναι ότι ο αριθμός των νόμιμων αρχείων είναι πολύ μεγαλύτερος από τον αριθμό των υποθέσεων απάτης. Τέτοια μη ισορροπημένα σύνολα απαιτούν πρόσθετες προφυλάξεις από τον αναλυτή δεδομένων (Carneiro et al., 2016).

Οι απατεώνες πράγματι αναπτύσσουν προηγμένες στρατηγικές για να καλύψουν έξυπνα τα ίχνη τους και να αποφύγουν την αποκάλυψή τους. Επιπλέον προσπαθούν να ενσωματωθούν όσο το δυνατόν περισσότερο στο περιβάλλον μέσα στο οποίο δρουν. Μια τέτοια είδους προσέγγιση θυμίζει τεχνικές «καμουφλάζ». Επομένως δεν πρόκειται για μια ευκαιριακή δραστηριότητα, αλλά για μια προσεκτικά και καλά σχεδιασμένη πράξη. Αυτό έχει ως αποτέλεσμα την ανάγκη για νέες τεχνικές που είναι ικανές να ανιχνεύουν και να αντιμετωπίζουν μοτίβα που αρχικά φαίνονται να μην συμμορφώνονται με τη συνήθη συμπεριφορά.

Οι μηχανισμοί ανίχνευσης που βασίζονται σε τεχνικές μάθησης χωρίς επίβλεψη ή σε περιγραφικές αναλύσεις, στοχεύουν συνήθως στην εύρεση συμπεριφοράς που αποκλίνει από την κανονική συμπεριφορά ή με άλλα λόγια στην ανίχνευση ανωμαλιών. Οι τεχνικές ανίχνευσης ακραίων στοιχείων έχουν μεγάλη αξία και επιτρέπουν τον εντοπισμό ενός σημαντικού ποσοστού απατών. Συγκεκριμένα, μπορεί να επιτρέπουν τον εντοπισμό απάτης που χρησιμοποιεί νέους, άγνωστους μηχανισμούς που καταλήγουν σε ένα νέο μοτίβο απάτης. Τα νέα αυτά μοτίβα δεν ανακαλύπτονται από έμπειρα συστήματα και ως εκ τούτου, η περιγραφική ανάλυση θα μπορούσε να είναι ένα πρώτο συμπληρωματικό εργαλείο που πρέπει να υιοθετηθεί προκειμένου να βελτιωθούν τα συστήματα ανίχνευσης απάτης που βασίζονται σε ειδικούς κανόνες. Ωστόσο, οι περιγραφικές τεχνικές δείχνουν ότι είναι επιρρεπείς στην εξαπάτηση από τις στρατηγικές απάτες που μοιάζουν με καμουφλάζ (Baesens et al., 2015).

Ένας άλλος τύπος συμπληρωματικού εργαλείου αφορά την ανάλυση κοινωνικών δικτύων, η οποία αποτελεί ένα σχετικά πρόσφατο εργαλείο για την καταπολέμηση της απάτης. Η ανάλυση κοινωνικών δικτύων επεκτείνει περαιτέρω τις ικανότητες του συστήματος ανίχνευσης απάτης, μαθαίνοντας και ανιχνεύοντας χαρακτηριστικά συμπεριφοράς απάτης σε ένα δίκτυο συνδεδεμένων οντοτήτων. Επιπρόσθετα επιτρέπει την ανάλυση των σχέσεων μεταξύ των οντοτήτων και κατά συνέπεια μπορεί να συμβάλει στην αποκάλυψη συγκεκριμένων προτύπων που υποδηλώνουν απάτη.

Είναι σημαντικό να τονιστεί ότι οι διαφορετικοί τύποι τεχνικών μπορεί να αλληλοσυμπληρώνονται, καθώς εστιάζουν σε διαφορετικές πτυχές απάτης και δεν πρέπει να θεωρούνται αποκλειστικές εναλλακτικές λύσεις. Για να είναι λοιπόν ένα σύστημα ανίχνευσης και πρόληψης αποτελεσματικό, θα πρέπει να συνδυάζει τα διαφορετικά εργαλεία καθώς έχουν διαφορετικές δυνατότητες και περιορισμούς και συνεπώς η μια τεχνική ενισχύει την άλλη.

### 2.3.3 Εποπτευόμενοι μέθοδοι

Τα μοντέλα που βασίζονται στις εποπτευόμενες μεθόδους, εκπαιδεύονται να διακρίνουν μεταξύ συμπεριφορών απάτης και μη απάτης, έτσι ώστε νέες παρατηρήσεις να μπορούν να αντιστοιχηθούν σε κατηγορίες προκειμένου να βελτιστοποιηθεί κάποιο μέτρο της απόδοσης της ταξινόμησης. Αυτό φυσικά απαιτεί να είναι κανείς σίγουρος για τις πραγματικές κατηγορίες των αρχικών δεδομένων που χρησιμοποιούνται για την κατασκευή των μοντέλων. Η αβεβαιότητα εισάγεται όταν οι νόμιμες συναλλαγές αναφέρονται εσφαλμένα ως απάτη ή όταν οι παρατηρήσεις απάτης δεν προσδιορίζονται ως τέτοιες. Οι εποπτευόμενες μέθοδοι απαιτούν να έχουμε παραδείγματα και των δυο κλάσεων και μπορούν να χρησιμοποιηθούν μόνο για τον εντοπισμό απατών ενός τύπου που έχουν συμβεί στο παρελθόν. Η χρήση τέτοιων μεθόδων συχνά παρουσιάζει δυσκολίες όταν τα μεγέθη των δυο κατηγοριών είναι πολύ διαφορετικά μεταξύ τους. Στα προβλήματα ανίχνευσης απάτης, οι νόμιμες συναλλαγές υπερτερούν κατά πολύ των συναλλαγών απάτης και αυτή η ανισορροπία μπορεί να οδηγήσει σε μοντέλα με περιορισμένη διακριτική ικανότητα (Bolon and Hand, 2001).

Στη βιβλιογραφία, οι περισσότερες μελέτες για τον εντοπισμό χρηματοοικονομικής απάτης εφαρμόζουν εποπτευόμενους μεθόδους μάθησης όπως τα νευρωνικά δίκτυα, τεχνικές δέντρων αποφάσεων και μηχανές διανυσμάτων υποστήριξης. Οι Maes et al. (2002) εφάρμοσαν ένα Bayesian εκπαιδευόμενο νευρωνικό δίκτυο, για τον εντοπισμό απάτης σε πιστωτικές κάρτες. Την τεχνολογία των νευρωνικών δικτύων την εφάρμοσαν επίσης και οι Zaslavsky and Strizhak (2006) στην ερευνά τους για την απάτη σε συναλλαγές με κάρτα. Στο ίδιο πεδίο (απάτη σε πιστωτικές κάρτες), εφαρμόστηκε και μια μέθοδος δέντρων αποφάσεων από τους Sahin et al. (2013), η οποία ελαχιστοποιούσε το κόστος της εσφαλμένης ταξινόμησης. Στον χώρο της παραποίησης χρηματοοικονομικών καταστάσεων οι Pai et al. (2010) χρησιμοποίησαν ένα μοντέλο βασισμένο σε μηχανές διανυσμάτων υποστήριξης. Στον ίδιο χώρο οι Κίρκος et al. (2007) διερεύνησαν την αποτελεσματικότητα των δέντρων αποφάσεων με τον αλγόριθμο ID3, των νευρωνικών δικτύων και των δικτύων Bayesian belief. Οι Lin et al. (2008) πρότειναν ένα μοντέλο νευρωνικού δικτύου ακτινικής βάσης για την ανίχνευση ύποπτων συναλλαγών στο χρηματοοικονομικό τομέα.

### 2.3.4 Μη Εποπτευόμενοι μέθοδοι

Οι μη εποπτευόμενες μέθοδοι αναζητούν τους λογαριασμούς και τους πελάτες των οποίων η συμπεριφορά είναι ασυνήθιστη. Αυτοί οι μέθοδοι, απαιτούν να γίνει μια μοντελοποίηση βασικής κατανομής που αντιπροσωπεύει την κανονική συμπεριφορά και στη συνέχεια να ανιχνευθούν παρατηρήσεις που δείχνουν τη μεγαλύτερη απόκλιση από τον κανόνα. Οι ακραίες τιμές είναι μια βασική μορφή μη τυπικής παρατήρησης που μπορεί να χρησιμοποιηθεί για τον εντοπισμό απάτης. Εξαιτίας των ακραίων τιμών, οι ερευνητές δεν μπορούν με απόλυτη σιγουριά, μέσω της στατιστικής ανάλυσης να εντοπίσουν αν έχει διαπραχθεί απάτη. Ως εκ τούτου, η ανάλυση θα πρέπει να θεωρηθεί ως εργαλείο που προειδοποιεί για το γεγονός ότι μια παρατήρηση είναι πιο πιθανό να αποτελεί απάτη από τις υπόλοιπες, έτσι ώστε στη συνέχεια να διερευνηθεί με περισσότερες λεπτομέρειες. Στόχος της στατιστικής ανάλυσης είναι η διαμόρφωση μιας βαθμολογίας για την πιθανότητα απάτης. Όσο υψηλότερη είναι η βαθμολογία, τόσο πιο ασυνήθιστη είναι η παρατήρηση ή τόσο περισσότερο μοιάζει με προηγούμενες περιπτώσεις απάτης. Υπάρχουν επίσης πολλοί διαφορετικοί τρόποι υπολογισμού των εκτιμήσεων αυτών, εξαιτίας του γεγονότος ότι υπάρχουν πολλοί διαφορετικοί τρόποι με τους οποίους μπορεί να διαπραχθεί η απάτη και πολλά διαφορετικά σενάρια στα οποία μπορεί να συμβεί (Bolton and Hand, 2001).



Αντίθετα με τις εποπτευόμενες μεθόδους οι μη εποπτευόμενες μέθοδοι δεν είναι τόσο διαδεδομένες στον εντοπισμό της χρηματοοικονομικής απάτης. Οι Juszczak et al (2008) στην έρευνα τους για την ανίχνευση απάτης σε πιστωτικές κάρτες χρησιμοποίησαν μια μη εποπτευόμενη μέθοδο την οποία συνέκριναν με μια εποπτευόμενη μέθοδο. Μια ακόμα έρευνα που πραγματοποιήθηκε από τους Quah and Sriganesh (2008) εφάρμοσε τη μεθοδολογία των self-organization map (SOM). Το SOM είναι ένα νευρωνικό δίκτυο που χρησιμοποιεί μη εποπτευόμενη μάθηση για να διαμορφώσει τους νευρώνες του σύμφωνα με την τοπολογική δομή των δεδομένων εισόδου. Ωστόσο έχουν εφαρμοστεί περισσότερο για τον εντοπισμό εισβολής σε υπολογιστή (hacking). Εάν ένας χάκερ αποκτήσει παράνομη πρόσβαση σε ένα λογαριασμό, τότε η εισβολή του ανιχνεύεται από την παρουσία μιας ακολουθίας εντολών που δεν βρίσκονται στο προφίλ των εντολών που πληκτρολογεί ο χρήστης. Οι Qu, Vetter et al. (1998) χρησιμοποίησαν πιθανότητες γεγονότων για να καθορίσουν το προφίλ, ενώ οι Forrest et al (1996) καθώς και οι Kosoresow και Hofmeyr (1997), χρησιμοποίησαν ομοιότητα αλληλουχιών που μπορούν να ερμηνευτούν σε ένα πιθανό πλαίσιο.

## 2.4 Περιγραφή ειδών χρηματοοικονομικής απάτης

Στην ενότητα αυτή θα περιγράψουμε περιπτώσεις χρηματοοικονομικής απάτης. Οι συγκεκριμένες περιπτώσεις επιλέχθηκαν βάσει της συχνότητας εμφάνισής τους, προσπαθώντας να καλυφθεί όσο το δυνατόν μεγαλύτερο μέρος του φάσματος των ειδών απάτης.

### 2.4.1 Απάτες σε πιστωτικές κάρτες

Οι τράπεζες λιανικής έχουν αναπτύξει με επιτυχία πλαστικές κάρτες για να παρέχουν ένα ευρύ φάσμα προϊόντων και τραπεζικών ευκαιριών στους καταναλωτές. Ο θεσμός αυτός συνοδεύτηκε από το πρόβλημα της απάτης με πλαστικές κάρτες. Στο Ηνωμένο Βασίλειο για παράδειγμα, το 2004 υπολογίστηκε ότι οι ζημιές ανήλθαν σε £505 εκατομμύρια. Το 2006 παρατηρήθηκε μια μείωση κατά 5% από τα προηγούμενα έτη, στα £209 εκατομμύρια. Η μείωση αυτή εικάστηκε ότι οφείλεται στην εισαγωγή ελέγχου ταυτότητας PIN από τα συστήματα chip και PIN, που ξεκίνησε στις 14 Φεβρουαρίου 2006. Την ευθύνη για το οικονομικό βάρος της απάτης την απορροφούν οι πιστωτές, οι έμποροι και οι νόμιμοι πελάτες. Οι πιστωτές και οι έμποροι δαπανούν σημαντικούς πόρους για να εξασφαλίσουν τα συστήματα και τις διαδικασίες τους σε μια προσπάθεια να περιορίσουν την ευθύνη τους για τέτοιου είδους κόστη.( Juszczak et al, 2008).

Η απάτη με πιστωτικές κάρτες είναι ένα σοβαρό και αυξανόμενο πρόβλημα. Ειδικότερα στην σημερινή ολοένα και πιο ηλεκτρονική κοινωνία και εξαιτίας της ταχείας προόδου του ηλεκτρονικού εμπορίου, όπου η χρήση πιστωτικών καρτών για αγορές έχει γίνει περισσότερη εύκολη και απαραίτητη. Οι συναλλαγές με πιστωτικές κάρτες έχουν γίνει το de facto πρότυπο για το διαδίκτυο και το ηλεκτρονικό εμπόριο μέσω του διαδικτύου. Ο αυξανόμενος λοιπόν, αριθμός συναλλαγών με πιστωτικές κάρτες έχει δώσει περισσότερες ευκαιρίες στους απατεώνες να υποκλέψουν αριθμούς πιστωτικών καρτών και στη συνέχεια να διαπράξουν απάτη. Ως εκ τούτου όταν οι τράπεζες χάνουν χρήματα λόγω απάτης πιστωτικών καρτών, οι κάτοχοι των καρτών πληρώνουν για όλη αυτή την απώλεια μέσω υψηλότερων επιτοκίων, υψηλότερων προμηθειών και μειωμένων παροχών (Chan et al., 1999).

Οι Bolton και Hand (2001) κατηγοριοποίησαν τις απάτες με πιστωτικές κάρτες σε δυο ομάδες:

- Σε απάτη αιτήσεων, η οποία παρατηρείται όταν οι απατεώνες αποκτούν νέες κάρτες από εταιρείες έκδοσης χρησιμοποιώντας ψευδείς πληροφορίες άλλων ατόμων.
- Σε απάτη συμπεριφοράς, η οποία μπορεί να παρατηρηθεί σε τέσσερα διαφορετικά είδη : κλοπή αλληλογραφίας, κλοπή/χαμένη κάρτα, πλαστή κάρτα και απάτη όπου ο κάτοχος της κάρτας δεν είναι παρών.

Η κλοπή αλληλογραφίας συμβαίνει όταν οι απατεώνες υποκλέπτουν πιστωτικές κάρτες από το ταχυδρομείο προτού φτάσουν στους κατόχους καρτών ή κλέβουν προσωπικά στοιχεία από τις κινήσεις τραπεζών και πιστωτικών καρτών. Όταν οι απατεώνες αποκτούν πιστωτικές κάρτες μέσω κλοπής ή αποκτούν πρόσβαση σε χαμένες κάρτες, παρατηρείται απάτη συμπεριφοράς και συγκεκριμένα το δεύτερο είδος «κλοπή/χαμένη κάρτα». Τα δυο τελευταία είδη απάτης, πλαστές κάρτες και απάτη «ο κάτοχος δεν είναι παρών», παρατηρούνται εξαιτίας της αυξημένης χρήσης των ηλεκτρονικών συναλλαγών. Και στους δυο αυτούς τύπους απάτης, τα στοιχεία της πιστωτικής κάρτας αποκτώνται εν αγνοία των κατόχων της κάρτας και στη συνέχεια είτε γίνονται πλαστές κάρτες είτε οι πληροφορίες χρησιμοποιούνται για την πραγματοποίηση συναλλαγών όπου ο κάτοχος της κάρτας δεν είναι παρών, π.χ. μέσω ταχυδρομείου, τηλεφώνου ή διαδικτύου.

Οι πληροφορίες των κατόχων της κάρτας λαμβάνονται με διάφορους τρόπους, όπως από υπάλληλους που κλέβουν πληροφορίες μέσω μη εξουσιοδοτημένων «swipers», απάτες «phishing» ή μέσω εισβολής σε δίκτυα υπολογιστών της εταιρείας. Το ηλεκτρονικό ψάρεμα «phishing» είναι ένας τύπος επίθεσης κοινωνικής μηχανικής που χρησιμοποιείται συχνά για την κλοπή δεδομένων των χρηστών, συμπεριλαμβανομένων των διαπιστευτηρίων σύνδεσης και των αριθμών πιστωτικών καρτών. Συμβαίνει όταν ένας εισβολέας εξαπατά ένα θύμα για να ανοίξει ένα email ή γενικά να αντλήσει πληροφορίες (Bhattacharyya et al., 2011).

Οι απατεώνες μπορούν επίσης να καταφύγουν στη δημιουργία αριθμών πιστωτικών καρτών χρησιμοποιώντας BIN (Αριθμοί Αναγνώρισης Τραπεζών) των τραπεζών. Επιπλέον στη μέθοδο «phishing», οι απατεώνες λειτουργούν μέσω ενός ιστότοπου με αυθεντική εμφάνιση, όπου διαφημίζει και πουλά αγαθά σε εξαιρετικά μειωμένες τιμές. Ο αγοραστής που δεν γνωρίζει ή είναι ανυποψίαστος υποβάλλει τα στοιχεία της κάρτας του και αγοράζει αγαθά. Στη συνέχεια, οι απατεώνες υποβάλλουν μια παραγγελία σε έναν γνήσιο έμπορο χρησιμοποιώντας τα στοιχεία της κλεμμένης κάρτας. Επιπλέον η κλεμμένη κάρτα χρησιμοποιείται για την αγορά αγαθών ή τη μεταφορά κεφαλαίων σε λογαριασμούς που δεν μπορούν να ανιχνευτούν, παρά μόνο μετά από αρκετές μέρες που ο έμπορος και ο κάτοχος της κάρτας αντιλαμβάνονται την απάτη. Αυτός ο τύπος απάτης προκαλεί αρχική σύγχυση που παρέχει καμουφλάζ στον απατεώνα για να πραγματοποιήσει τις δραστηριότητες του (Quah, 2008).

Ο εντοπισμός απάτης πιστωτικών καρτών θεωρείται γενικά ως πρόβλημα ταξινόμησης εξόρυξης δεδομένων, που στόχο έχει τη σωστή ταξινόμηση συναλλαγών με πιστωτική κάρτα ως απάτη ή μη απάτη. Παρόλο που η ανίχνευση απάτης έχει μακρά ιστορία, δεν έχει εμφανιστεί εκτενής έρευνα στον τομέα αυτό. Ο λόγος είναι ότι οι τράπεζες δεν είναι έτοιμες να αποκαλύψουν ευαίσθητα δεδομένα συναλλαγών πελατών τους για λόγους απορρήτου και ως εκ τούτου οι ερευνητές δεν έχουν στη διάθεσή τους πραγματικά δεδομένα ώστε να μπορούν να πραγματοποιήσουν πειράματα. Επιπλέον, συνήθιζαν να αλλάζουν τα ονόματα των πεδίων, έτσι ώστε οι ερευνητές να μην έχουν ιδέα για τα πραγματικά πεδία. Εξαιτίας αυτής της σπανιότητας πραγματικών δεδομένων, δεν έχουν αναπτυχθεί και περιγραφεί πολλά μοντέλα ανίχνευσης απάτης στην ακαδημαϊκή βιβλιογραφία και ακόμα λιγότερα είναι γνωστό ότι έχουν εφαρμοστεί σε πραγματικά συστήματα ανίχνευσης. Ωστόσο μπορούν να παρατηρηθούν ορισμένες επιτυχημένες εφαρμογές διάφορων τεχνικών εξόρυξης δεδομένων όπως ανίχνευση ακραίων τιμών, νευρωνικό δίκτυο, ταξινομητής Bayes, μηχανή υποστήριξης διανυσμάτων, ασαφή συστήματα, γενετικός αλγόριθμος, K-πλησιέστερος γείτονας και το κρυφό μοντέλο Markov (Seeja and Zareapoor, 2014).

## 2.4.2 Παραποιημένες χρηματοοικονομικές καταστάσεις

Οι οικονομικές καταστάσεις είναι τα βασικά έγγραφα μιας εταιρείας που αντικατοπτρίζουν την οικονομική της κατάσταση. Μια καλά προσεκτική ανάγνωση των χρηματοοικονομικών καταστάσεων μπορεί να δείξει εάν η εταιρεία λειτουργεί ομαλά ή βρίσκεται σε κρίση. Εάν λοιπόν η εταιρεία βρίσκεται σε κρίση, οι οικονομικές καταστάσεις μπορούν να δείξουν εάν το πιο κρίσιμο πράγμα που αντιμετωπίζει η εταιρεία είναι τα μετρητά ή τα κέρδη ή κάτι άλλο.

Όλες οι εισηγμένες εταιρείες υποχρεούνται να δημοσιεύουν τις οικονομικές τους καταστάσεις κάθε χρόνο και κάθε τρίμηνο. Οι μέτοχοι μπορούν να σχηματίσουν μια καλή ιδέα για το οικονομικό μέλλον των εταιρειών μέσω των συγκεκριμένων καταστάσεων και μπορούν να αποφασίσουν αν αξίζει να επενδύσουν. Η τράπεζα επίσης χρειάζεται τις οικονομικές καταστάσεις των εταιρειών προκειμένου να αποφασίσει εάν θα χορηγήσει δάνεια σε αυτές. Με λίγα λόγια, οι χρηματοοικονομικές

καταστάσεις αποτελούν τον καθρέφτη της οικονομικής κατάστασης των εταιρειών. Γενικά, περιλαμβάνουν ισολογισμούς, καταστάσεις λογαριασμού αποτελεσμάτων, καταστάσεις ταμειακών ροών, καταστάσεις κερδών εις νέο και ορισμένες άλλες καταστάσεις. Μια περιγραφή των στοιχείων που παρατίθενται στις διάφορες οικονομικές καταστάσεις δίνεται παρακάτω (Ravisankar et al. , 2010):

- **Ισολογισμός:**

Πρόκειται για μια κατάσταση της λογιστικής αξία ενός οργανισμού σε μια συγκεκριμένη ημερομηνία, συνήθως στο τέλος του οικονομικού έτους. Ένας ισολογισμός αποτελείται από τρία μέρη: περιουσιακά στοιχεία, υποχρεώσεις και ίδια κεφάλαια. Η διαφορά μεταξύ των περιουσιακών στοιχείων και των υποχρεώσεων είναι γνωστή ως «καθαρά περιουσιακά στοιχεία» ή «καθαρή αξία» της εταιρείας.

- **Κατάσταση Λογαριασμού Αποτελεσμάτων Χρήσης:**

Η κατάσταση αποτελεσμάτων χρήσης υποδεικνύει πως τα ακαθάριστα έσοδα (χρήματα που λαμβάνονται από την πώληση προϊόντων και υπηρεσιών πριν αφαιρεθούν τα έξοδα) μετατρέπονται σε καθαρά έσοδα (το αποτέλεσμα μετά από όλα τα έσοδα και έξοδα που έχουν λογιστεί). Ο σκοπός της, είναι να δείξει στους διαχειριστές και στους επενδυτές εάν η εταιρεία κέρδισε ή έχασε χρήματα κατά την υπό εξέταση περίοδο.

- **Κατάσταση Ταμειακών Ροών:**

Η συγκεκριμένη κατάσταση παρουσιάζει τα εισερχόμενα και εξερχόμενα κεφάλαια κατά την διάρκεια μιας συγκεκριμένης περιόδου. Η κατάσταση δείχνει πως οι αλλαγές στους λογαριασμούς ισολογισμού και Εισοδήματος επηρεάζουν τα μετρητά και τα ταμειακά ισοδύναμα. Ως αναλυτικό εργαλείο, η Κατάσταση ταμειακών ροών είναι χρήσιμη για τον προσδιορισμό της βραχυπρόθεσμης βιωσιμότητας μιας εταιρείας, ιδιαίτερα της ικανότητας της να πληρώνει λογαριασμούς.

- **Κατάσταση Κερδών εις Νέο:**

Η Κατάσταση κερδών εις νέο, γνωστή και ως « Κατάσταση ίδιων κεφαλαίων» και «Κατάσταση καθαρών περιουσιακών στοιχείων» για μη κερδοσκοπικούς οργανισμούς, εξηγεί τις αλλαγές στα κέρδη εις νέο της εταιρείας κατά την περίοδο αναφοράς. Αναλύει τις αλλαγές που επηρεάζουν τον λογαριασμό, όπως κέρδη ή ζημιές από λειτουργίες, μερίσματα που καταβάλλονται και οποιαδήποτε άλλα στοιχεία χρεώνονται ή πιστώνονται στα κέρδη εις νέο.

Το 1997, το Συμβούλιο Ελεγκτικών Προτύπων εξέδωσε τη Δήλωση για τα Ελεγκτικά Πρότυπα (SAS) : Εξέταση απάτης σε έλεγχο οικονομικών καταστάσεων. Αυτό το Πρότυπο απαιτεί από τους ελεγκτές να αξιολογούν τον κίνδυνο απάτης κατά

τη διάρκεια κάθε ελέγχου και ενθαρρύνει τους ελεγκτές να λαμβάνουν υπόψη τόσο το σύστημα εσωτερικού ελέγχου όσο και τη στάση της διοίκησης απέναντι στους ελεγχόμενους, όταν κάνουν αυτή την αξιολόγηση. Οι παράγοντες κινδύνου ή οι «κόκκινες σημαίες» που σχετίζονται με παραπονημένες οικονομικές αναφορές μπορούν ομαδοποιηθούν σε τρεις κατηγορίες (Kirkos et al. 2007):

- Τα χαρακτηριστικά και η επιρροή της διοίκησης στο περιβάλλον ελέγχου.
- Οι συνθήκες του κλάδου
- Λειτουργικά χαρακτηριστικά και χρηματοπιστωτική σταθερότητα.

Η απάτη στις οικονομικές καταστάσεις μπορεί να οριστεί ως ουσιώδης παράληψη ή ψευδής δήλωση που προκύπτει από σκόπιμη αδυναμία αναφοράς οικονομικών πληροφοριών σύμφωνα με τις γενικά αποδεκτές λογιστικές αρχές. Όπως αναφέρθηκε από το Κέντρο Ελέγχου Ποιότητας, τα άτομα εμπλέκονται σε απάτες οικονομικών καταστάσεων για διάφορους λόγους, συμπεριλαμβανομένου του προσωπικού κέρδους, της ανάγκης να ανταποκριθούν στις βραχυπρόθεσμες οικονομικές προσδοκίες και της επιθυμίας απόκρυψης κακών ειδήσεων. Οι παραπονημένες οικονομικές καταστάσεις χειραγωγούνται ώστε να είναι πειστικά όμοιες με τις μη παραπονημένες και οι συνήθεις τύποι περιλαμβάνουν ακατάλληλη αναγνώριση εσόδων, υποεκτίμηση/υπερεκτίμηση εσόδων, έξοδα, υποχρεώσεις περιουσιακών στοιχείων και ψευδείς δηλώσεις (ή παραλείψεις) στις υποσημειώσεις των οικονομικών καταστάσεων ή στην συζήτηση και ανάλυση της διοίκησης για μια επισκόπηση (Hajek and Henriques, 2017).

Η απάτη στις οικονομικές καταστάσεις αποτελεί σημαντικό κίνδυνο για τους εξωτερικούς ενδιαφερόμενους, όπως επενδυτές και εμπορικούς δανειστές. Συχνά η ανακάλυψη της , έρχεται ως έκπληξη που έχει άμεσες και σοβαρές επιπτώσεις στην αξία της εταιρείας. Λόγω της εγκληματικής φύσης της απάτης στις οικονομικές καταστάσεις, ο ισχυρισμός απάτης για μια εταιρεία που δεν διαπράττει απάτη μπορεί να έχει σοβαρές συνέπειες. Αν προκύψει αυτός ο ισχυρισμός πιθανότατα θα οδηγήσει σε χαμένη επενδυτική ευκαιρία καθώς και θα μειώσει την πίστη του καταναλωτή στην εταιρεία (Hoogs et al., 2007).

Το κόστος της απάτης στις οικονομικές καταστάσεις είναι τεράστιο, για παράδειγμα στις ΗΠΑ υπολογίζεται \$572 δισεκατομμύρια ετησίως. Εκτός από το άμεσο κόστος, η απάτη στις χρηματοοικονομικές αναφορές επηρεάζει αρνητικά τους επενδυτές και τους εργαζόμενους και υπονομεύει την αξιοπιστία των εταιρικών οικονομικών καταστάσεων, με αποτέλεσμα υψηλότερο κόστος συναλλαγής και λιγότερο αποτελεσματικές αγορές. Οι ελεγκτές τόσο μέσω της αυτορρύθμισης όσο και τις νομοθεσίας, είναι υπεύθυνοι για την παροχή εύλογης διασφάλισης ότι οι οικονομικές καταστάσεις είναι απαλλαγμένες από ουσιώδη ανακρίβεια που προκαλείται από απάτη (Perlos , 2011).

Ο εντοπισμός παραπονημένων χρηματοοικονομικών καταστάσεων μπορεί να θεωρηθεί και αυτό ως πρόβλημα ταξινόμησης. Καθώς έχουν γίνει διαθέσιμες νέες τεχνολογίες τεχνητής νοημοσύνης και εξόρυξης δεδομένων, οι ελεγκτές έχουν

υιοθετήσει ορισμένα από αυτά τα εργαλεία και ορισμένες από τις τεχνικές για να βοηθήσουν στον εντοπισμό απάτης, κυρίως κατά την εξέταση των αριθμητικών δεδομένων των οικονομικών καταστάσεων. Για την βελτίωση της ανίχνευσης απάτης, η πρόσφατη λογιστική έρευνα έχει επικεντρωθεί στην δοκιμή της χρησιμότητας διάφορων αλγορίθμων στατιστικής και μηχανικής μάθησης, όπως λογιστική παλινδρόμηση και τα τεχνητά νευρωνικά δίκτυα στον εντοπισμό απάτης οικονομικών καταστάσεων.

Εκτός από τις προαναφερθέντες μεθόδους έχουν χρησιμοποιηθεί ακόμα, τα δέντρα αποφάσεων και οι μηχανές διανυσμάτων υποστήριξης (SVM). Η λογιστική παλινδρόμηση έχει χρησιμοποιηθεί ως παραδοσιακός ταξινομητής στατιστικής αναφοράς. Μια ποικιλία νευρωνικών δικτύων έχει εφαρμοστεί σχετικά με την ανίχνευση απάτης στις χρηματοοικονομικές απάτες, συμπεριλαμβανομένου του πολυστρωματικού perceptron (MLP), των πιθανοτικών νευρωνικών δικτύων, group method data handling, radial basis function neural networks και growing hierarchical self-organising maps. Τα δέντρα αποφάσεων περιλαμβάνουν τόσο μεμονωμένα δέντρα απόφασης (όπως C4.5 και C5.0) όσο και τα σύνολα δέντρων αποφάσεων, όπως stacking ή το bagging. Ενώ το SVM, παρέχει μια αποτελεσματική γενίκευση στα δεδομένα δοκιμών, υπάρχει δυνατότητα να επιφέρει κακή απόδοση σε θορυβώδη δεδομένα (Perlos, 2011, Hajek and Henriques, 2017).

### 2.4.3 Φοροδιαφυγή και Φορολογική απάτη

Η φορολογική απάτη είναι ένα παγκόσμιο φαινόμενο που επηρεάζει το κοινωνικό σύνολο. Αυτό το φαινόμενο μπορεί να αναφερθεί ως μια σκόπιμη πράξη ψεύδους σε μια φορολογική δήλωση με σκοπό την απόκτηση παράνομου οικονομικού οφέλους και τη μείωση της φορολογικής υποχρέωσης. Μελέτες έχουν εκτιμήσει ότι οι κυβερνήσεις σε όλο τον κόσμο χάνουν περίπου 500 δισεκατομμύρια ετησίως. Οι απώλειες δημοσιονομικών εσόδων είναι ιδιαίτερα μεγαλύτερες σε χώρες χαμηλού έως μεσαίου εισοδήματος, όπως για παράδειγμα τη Λατινική Αμερική, Νότια Ασία και περιοχές της Καραϊβικής, οι οποίες είναι επίσης χώρες που εξαρτώνται περισσότερο από τα φορολογικά έσοδα για τον δημοσιονομικό τους σχεδιασμό (Pérez et al., 2018).

Η φοροδιαφυγή και η φορολογική απάτη αποτελούν συνεχή ανησυχία για τις φορολογικές διοικήσεις, ειδικά όταν αφορούν αναπτυσσόμενες χώρες. Αν και είναι αλήθεια ότι οι φόροι δεν είναι η μόνη πηγή κρατικής χρηματοδότησης, ωστόσο αποτελούν έναν ισχυρό παράγοντα για τη δέσμευση και την αποτελεσματικότητα, με την οποία το κράτος μπορεί να ασκήσει τα καθήκοντά του και να περιορίσει την πρόσβαση σε άλλες πηγές εισοδήματος. Συγκεκριμένα ο φόρος προστιθέμενης αξίας (ΦΠΑ), που εφαρμόζεται σε περισσότερες από 130 χώρες σε διαφορετικά στάδια οικονομικής ανάπτυξης, έχει γίνει βασικό συστατικό των φορολογικών εσόδων, αυξάνοντας περίπου το 25% των παγκόσμιων φορολογικών εσόδων.

Οι εταιρείες πολλές φορές προσπαθούν να αυξήσουν την έκπτωση του φόρου και να μειώσουν την πληρωμή του ΦΠΑ. Για να το πετύχουν αυτό, προσομοιώνουν αγορές

που δεν υπήρξαν ποτέ, συγκεκριμένα λαμβάνοντας πλαστά τιμολόγια. Η πλαστότητα του παραστατικού μπορεί να είναι ουσιώδης, εάν τα φυσικά στοιχεία που συνθέτουν το τιμολόγιο έχουν νοθευθεί ή ιδεολογικά, όταν η ουσία του παραστατικού δεν έχει αλλοιωθεί, αλλά οι πράξεις που καταγράφονται σε αυτό είναι νοθευμένες ή ανύπαρκτες. Το τελευταίο είναι πιο περίπλοκο και δύσκολο να εντοπιστεί επειδή περιλαμβάνει εικονικές συναλλαγές στις οποίες απαιτείται έλεγχος για την εξέταση των βιβλίων πωλήσεων και διασταύρωση των πληροφοριών με τους προμηθευτές. Επιπλέον αυτές οι περιπτώσεις είναι πιο ακριβείς για την Υπηρεσία Εσωτερικών Εσόδων, καθώς απαιτούν μεγαλύτερο χρόνο αφιερωμένο στη συλλογή και τον έλεγχο αποδεικτικών στοιχείων, κάτι που είναι πιο δύσκολο να βρεθεί. Πέρα από την φυσική νοθεία του παραστατικού, γνωστές περιπτώσεις υλικής πλαστογράφησης είναι η χρήση κρεμαστών τιμολογίων, όπου ένα τιμολόγιο παραποιείται για να προσωποποιήσει έναν φορολογούμενο καλής συμπεριφοράς και η χρήση διπλού συνόλου φορολογικών τιμολογίων, με δυο ίδια αριθμημένα τιμολόγια, όπου το ένα από τα δυο είναι φανταστικό και με υψηλότερο ποσό (Gonzalez and Velasquez, 2013).

Ορισμένες κυβερνήσεις έχουν εφαρμόσει την μηχανική Εκμάθηση για την πρόβλεψη φορολογικής απάτης. Για παράδειγμα η Υπηρεσία Εσωτερικών Εσόδων (Internal Revenue Service, IRS) των Ηνωμένων Πολιτειών, εφάρμοσε το σύστημα Return Review Program (RRP). Ο κύριος στόχος του είναι ο εντοπισμός απάτης, εντοπίζοντας δόλιες επιστροφές με χαμηλότερο ποσοστό ψεύδους ανίχνευσης (McKenney, 2017). Η φορολογική Διοίκηση της Χιλής, η οποία εφάρμοσε Δένδρα αποφάσεων, Νευρωνικά Δίκτυα και Μπεϋζιανά Δίκτυα για τον εντοπισμό φορολογούμενων που χρησιμοποιούν πλαστά τιμολόγια (Gonzalez and Velasquez, 2012). Ένα ακόμα παράδειγμα πρόβλεψης φορολογικής απάτης με βάση την Μηχανική Μάθηση προέρχεται από το Ισπανικό Ινστιτούτο Δημοσιονομικών Μελετών. Η μελέτη τους εφάρμοσε Νευρωνικά δίκτυα σε δεδομένα από την Ισπανική Υπηρεσία Εσόδων, με στόχο τους φορολογούμενους που φοροδιαφεύγουν (Lopez et al. , 2019).

#### 2.4.4 Ξέπλυμα Χρήματος

Θα μπορούσε να σημειωθεί ότι ως ένα βαθμό το ξέπλυμα χρήματος συνδέεται με την φορολογική απάτη, αφού τα χρήματα αποκτώνται με παράνομο τρόπο, αποφεύγοντας τον φορολογικό νόμο. Ωστόσο, το συγκεκριμένο φαινόμενο απάτης αξίζει να μελετηθεί ξεχωριστά.

Το ξέπλυμα χρήματος είναι ένα οικονομικό έγκλημα. Με τον όρο «οικονομικό έγκλημα» εννοούμε δραστηριότητες (παράνομες) που παραβιάζουν τα θεμέλια της οικονομίας, της ελεύθερης αγοράς. Περιλαμβάνει χρήση μέσων και θεσμών της χρηματοπιστωτικής αγοράς για την απόκτηση οικονομικών κερδών σε βάρος άλλων παραγόντων της αγοράς. Οι εγκληματίες καταβάλλουν προσπάθειες να εισάγουν τα παράνομα αποκτηθέντα χρήματά τους σε νόμιμες οικονομικές συναλλαγές. Θα μπορούσε να αναφερθεί ότι το ξέπλυμα χρήματος καθιστά δυνατή την μεταφορά χρηματοοικονομικών περιουσιακών στοιχείων από την λεγόμενη γκρίζα αγορά και το εγκληματικό παρασκήνιο στην νόμιμη, επίσημη οικονομία. Με αυτόν τον τρόπο, οι

εγκληματίες μπορούν να χρησιμοποιούν ελεύθερα τέτοια περιουσιακά στοιχεία χωρίς να χρειάζεται να ανησυχούν ότι η εφορία θα αμφισβητήσει την νόμιμη προέλευσή τους και ότι η επιβολή του νόμου ή η διοίκηση των δικαστικών υπηρεσιών θα τα κατασχέσει και θα τους επιβάλλει ποινή.

Υπάρχει επίσης ένας δευτερεύων σκοπός του ξεπλύματος μαύρου χρήματος, είναι η απόκρυψη της ταυτότητας των προσώπων που συντονίζουν τις δραστηριότητες από υπηρεσίες επιβολής του νόμου και απονομής της δικαιοσύνης προκειμένου να διαφύγουν τη δίωξη. Πρέπει να τονιστεί, ότι δεν αποτελεί ένα ενιαίο φαινόμενο αντιθέτως έχει πολλαπλές μορφές και παραλλαγές. Συγκεκριμένα, εξαρτάται από οικονομικούς, πολιτικούς, γεωγραφικούς και κοινωνικούς παράγοντες. Ένα χαρακτηριστικό που μπορεί να θεωρηθεί κοινό όλων των μεθόδων που χρησιμοποιούν οι εγκληματίες είναι η παρουσία πολυάριθμων συναλλαγών μεταξύ πολυάριθμων οντοτήτων, φυσικών ή νομικών προσώπων και των τραπεζικών λογαριασμών τους σε σχετικά σύντομο χρονικό διάστημα. Μια από τις βασικές μεθόδους είναι η ανάμειξη του νόμιμου εισοδήματος με το παράνομο εισόδημα. Σκοπός είναι να καταστεί δύσκολος ο διαχωρισμός των δυο ειδών εισοδήματος που αποκτούν μια ή περισσότερες οντότητες που συνεργάζονται με εγκληματικές ομάδες (Drezewski et al., 2012).

Η νομιμοποίηση εσόδων είναι μια δυναμική δραστηριότητα που προσπαθεί να παρακάμψει δράσεις κατά του ξεπλύματος μαύρου χρήματος. Το ξέπλυμα χρήματος είναι η επεξεργασία εγκληματικών, μαύρων χρημάτων για να συγκαλύψει την παράνομη προέλευση τους και να τα κάνει να φαίνονται νόμιμα και «καθαρά». Επιπλέον στοιχεία έχουν δείξει ότι τα μαύρα χρήματα έχουν χρηματοδοτήσει τρομοκρατικές επιθέσεις παγκοσμίως. Συνεπώς, η έρευνα για την καταπολέμηση της νομιμοποίησης εσόδων από παράνομες δραστηριότητες είναι κρίσιμης σημασίας για την εθνική χρηματοπιστωτική σταθερότητα και τη διεθνή ασφάλεια. Τα πρότυπα συμπεριφοράς και τα δομικά χαρακτηριστικά του δικτύου νομιμοποίησης εσόδων από παράνομες δραστηριότητες (Money Laundering, ML) είναι απαραίτητα για την καταπολέμηση της ML, ωστόσο η παραδοσιακή έρευνα επικεντρώνεται σε νομοθετικές εκτιμήσεις και απαιτήσεις συμμόρφωσης. Η ML σπάνια, έως ποτέ, εκδηλώνεται από ένα μεμονωμένο άτομο, επιχείρηση, λογαριασμό ή συναλλαγή, αλλά μάλλον από ένα μοτίβο συμπεριφοράς που εμφανίζεται με την πάροδο του χρόνου και περιλαμβάνει ένα σύνολο σχετικών οντοτήτων πραγματικού κόσμου (Real-World-Entities, RWE). Το παραδοσιακό καθεστώς καταπολέμησης ML, υποθέτει ότι μια συνηθισμένη συναλλαγή είναι νόμιμη και μια ασυνήθιστη συναλλαγή είναι ύποπτη και ίσως να είναι παράνομη. Επιπλέον επειδή το ξέπλυμα χρήματος είναι αδίκημα κινήτρου παρά έγκλημα δραστηριότητας, όταν δυο διαφορετικά άτομα εμπλέκονται στην ίδια συναλλαγή δεν σημαίνει ότι και τα δυο είναι ένοχα, μπορεί το ένα να είναι και το άλλο να μην είναι.

Το «σύνηθες» και το «ασυνήθιστο» σχετίζονται με τη συμπεριφορά ή το επιχειρηματικό μοντέλο κάποιου, ενώ το «ύποπτο» είναι μια προσωπική κρίση για τη νομιμότητα μιας συναλλαγής. Επιπλέον το «σύνηθες» σημαίνει πιθανότητα αλλά όχι απαραίτητα νόμιμο ενώ το «ύποπτο» υπονοεί ότι υπάρχει μεγαλύτερη πιθανότητα να είναι παράνομο. Όσον αφορά τους όρους, οι ΗΠΑ, οι Κίνα, οι Φιλιππίνες, ο Καναδάς



και οι Κάτω Χώρες χρησιμοποιούν την «αναφορά ύποπτης δραστηριότητας», την «αναφορά ύποπτης συναλλαγής», την «αναφορά καλυμμένης συναλλαγής» και την «αναφορά ασυνήθιστης συναλλαγής», αντίστοιχα για τα συστήματα αναφοράς καταπολέμησης ML τους, υποδηλώνοντας ότι διαφορετικές χώρες μπορεί να έχουν διαφορετική ερμηνεία των λέξεων «ασυνήθιστο» και «ύποπτο» (Gao et al., 2007).

Τα χρήματα που χρειάζονται «ξέπλυμα» αφήνουν ίχνη στα αρχεία συναλλαγών του χρηματοπιστωτικού ιδρύματος, εφόσον χρησιμοποιούνται τραπεζικές ή χρηματοοικονομικές υπηρεσίες. Διαφορετικά, άτομα που κάνουν τέτοιου είδους δραστηριότητες υπάρχει μεγάλη πιθανότητα να χρησιμοποιούν μια εταιρεία κέλφους, πλασματικές συναλλαγές και ψευδή λογιστική για να νομιμοποιήσουν τα παράνομα χρήματα. Ανεξάρτητα από την μέθοδο που εφαρμόζεται από τα άτομα, οι συναλλαγές τους έχουν καταγραφές, παρόλο που μπορεί να μοιάζουν με νόμιμες. Οι καταγραφές αυτές περιλαμβάνουν, επιχειρηματικές οικονομικές καταστάσεις, αρχεία συναλλαγών ή εμπορικά τιμολόγια. Συνεπώς ο τρόπος εντοπισμού αυτών των εγγράφων καθώς και η αντικειμενική ερμηνεία των επιπτώσεων αυτών των δεδομένων, είναι τα βασικά ζητήματα για την ανίχνευση ML (Yang and Wei, 2010).

## 2.1 Άλλα είδη απάτης

Η ανάπτυξη της τεχνολογίας έχει επιφέρει μια ευρεία εξάπλωση και χρήση του Διαδικτύου, το οποίο περιγράφεται ως αιτία της αύξησης της λογοκλοπής. Λογοκλοπή είναι όταν ένα άτομο χρησιμοποιεί λέξεις ή ιδέες κάποιου άλλου ως δικές του (Anderson and Steneck, 2011). Το φαινόμενο αυτό παρατηρείται σε φοιτητές, όπου αντιγράφουν άρθρα ή εργασίες και τα παριστάνουν ως δικά τους στα πλαίσια ενός μαθήματος ή μιας πτυχιακής. Χωρίς αμφιβολία το διαδίκτυο παρέχει μια πλούσια βιβλιοθήκη κειμένων, άρθρων και εργασιών, συνεπώς τους παρέχει μια πρωτοφανής ευκαιρία για λογοκλοπή (Howard, 2006).

Η απάτη αυτή έχει προκαλέσει μεγάλη ανησυχία σε αρκετά πανεπιστήμια που προσπαθούν να εντοπίσουν τρόπους για να την καταπολεμήσουν. Ορισμένα έχουν αναπτύξει δικά τους συστήματα και βάσεις για τον έλεγχο των εργασιών, ενώ άλλα πανεπιστήμια αγοράζουν άδεια από υπηρεσίες, για την πρόσβαση σε συστήματα ανίχνευσης απάτης. Το Turnitin για παράδειγμα είναι μια υπηρεσία ανίχνευσης λογοκλοπής που βασίζεται στο Διαδίκτυο, από το 1997 το λογισμικό της έχει χρησιμοποιηθεί από διάφορα πανεπιστήμια, συγκεκριμένα το χρησιμοποίησαν για να ελέγξουν τα υποβληθέντα έγγραφα σε σχέση με τη βάση δεδομένων και το περιεχόμενο άλλων ιστότοπων (Batane, 2010) .

Ένα άλλο συχνό φαινόμενο απάτης παρατηρείται στον κλάδο της ασφαλιστικής βιομηχανίας, με σκοπό την είσπραξη αποζημιώσεων. Στη βιβλιογραφία παρατηρείται ότι έχουν εφαρμοστεί διάφορες τεχνικές ανίχνευσης για τον εντοπισμό απάτης στον ασφαλιστικό τομέα. Οι Viaene et al. (2005) χρησιμοποίησαν τον ταξινομητή νευρωνικών δικτύων για την ανίχνευση απάτης αξιώσεων ασφάλειας αυτοκινήτου, για προστασία προσωπικών τραυματισμών. Ένας πολύ γνωστός και με ευρείες εφαρμογές

ταξινομητής, ο αφελής ή απλός ταξινομητής Bayes χρησιμοποιήθηκε από τους Viaene et al. (2004), για την διάγνωση απάτης ασφαλιστικών αξιώσεων. Οι Pathak et al. (2005) ανέπτυξαν ένα έμπειρο μοντέλο βασισμένο στην ασαφή λογική, που μπορεί να εντοπίσει και να αξιολογήσει εάν εμπλέκονται στοιχεία απάτης στον διακανονισμό ασφαλιστικών αποζημιώσεων. Στον τομέα ασφάλισης αυτοκινήτων για την ανίχνευση απάτης, οι Nian et al. (2016) εφάρμοσαν φασματική κατάταξη χωρίς επίβλεψη.

# ΚΕΦΑΛΑΙΟ 3

## ΜΕΘΟΔΟΛΟΓΙΚΟ ΠΛΑΙΣΙΟ

### 3.1 Εισαγωγή

Σε μια εποχή έκρηξης πληροφοριών είναι ιδιαίτερα σημαντικό η ανάπτυξη συστημάτων και η εφαρμογή μεθόδων, που θα διαχειρίζονται τον τεράστιο όγκο δεδομένων. Η ανακάλυψη χρήσιμης γνώσης από τη βάση δεδομένων και η μετατροπή των πληροφοριών σε χρήσιμα αποτελέσματα, είναι μια σημαντική πρόκληση που αντιμετωπίζουν οι εταιρείες. Η εξόρυξη δεδομένων (data mining) αποτελεί ένα τέτοιο εργαλείο, όπου είναι απαραίτητο στο σύστημα υποστήριξης αποφάσεων και διαδραματίζει βασικό ρόλο στον εντοπισμό απάτης (Yeh and Lien, 2009).

Η εφαρμογή τεχνικών εξόρυξης δεδομένων (ΕΔ) για τη χρηματοοικονομική ταξινόμηση είναι ένας γόνιμος ερευνητικός τομέας. Οι τεχνικές εξόρυξης (ΕΔ) δεδομένων, που ισχυρίζονται ότι διαθέτουν προηγμένες δυνατότητες ταξινόμησης και πρόβλεψης, θα μπορούσαν να διευκολύνουν τη διαδικασία εντοπισμού απάτης (Kirkos et al., 2007).

Οι πιο συχνά χρησιμοποιούμενες μέθοδοι ταξινόμησης ήταν η λογιστική παλινδρόμηση (logistic regression), τα νευρωνικά δίκτυα (neural networks) και η διακριτική ανάλυση (discriminant analysis), ενώ σε μεταγενέστερες μελέτες εφαρμόστηκαν τα δέντρα αποφάσεων (decision trees), τα δίκτυα Bayes (Bayesian networks) και οι μηχανές διανυσμάτων υποστήριξης (support vector machines, SVM, Abbasi et al., 2012).

Αξίζει να σημειωθεί ότι, στη βιβλιογραφία έχουν παρατηρηθεί και εφαρμογές μεθόδων ομαδοποίησης κατά τη διαδικασία ανίχνευσης απάτης. Η ομαδοποίηση αν και φαίνεται παρόμοια διαδικασία με την ταξινόμηση, παρουσιάζει μια βασική διαφορά. Πρόκειται για μια τεχνική μάθησης χωρίς επίβλεψη, όπου εφαρμόζεται για να ομαδοποίηση παρόμοιες περιπτώσεις με βάση τα χαρακτηριστικά, ενώ η ταξινόμηση είναι μια εποπτευόμενη μάθηση που εφαρμόζεται για τη διάκριση των περιπτώσεων σε προκαθορισμένες ετικέτες με βάση των χαρακτηριστικών.

Στόχος του παρόντος κεφαλαίου, είναι η παρουσίαση και η κατανόηση της έννοιας εξόρυξης δεδομένων όσο το δυνατόν καλύτερα. Στις παρακάτω παραγράφους θα παρουσιαστούν επίσης μερικές από τις βασικότερες μεθοδολογίες για τον εντοπισμό απάτης.

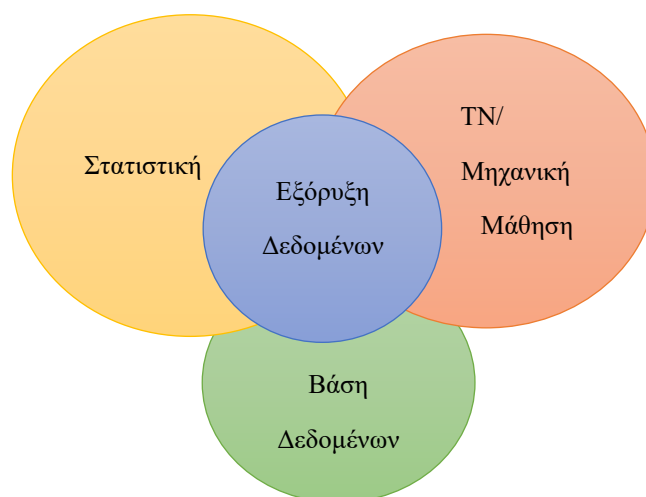
### 3.2 Εξόρυξη Δεδομένων

### 3.2.1 Ορισμός

Γενικά, η εξόρυξη δεδομένων είναι ένας τρόπος ανάπτυξη νοημοσύνης (συγκεκριμένα, ενεργών πληροφοριών ή γνώσεων) από δεδομένα που συλλέγονται, οργανώνονται και αποθηκεύονται.

Η εξόρυξη δεδομένων (Data Mining ,DM) είναι μια επαναληπτική διαδικασία στην οποία η πρόοδος καθορίζεται από την ανακάλυψη μέσω αυτόματων ή μη αυτόματων μεθόδων. Το DM είναι πιο χρήσιμο σε ένα σενάριο διερευνητικής ανάλυσης στο οποίο δεν υπάρχουν προκαθορισμένες αντιλήψεις σχετικά με το τι θα αποτελέσει ένα ενδιαφέρον «αποτέλεσμα». Πολλές μονάδες επιβολής του νόμου και ειδικές ερευνητικές μονάδες, των οποίων αποστολή είναι ο εντοπισμός δραστηριοτήτων απάτης, έχουν εφαρμόσει μεθόδους εξόρυξης δεδομένων με επιτυχία (Kirkos et al., 2007).

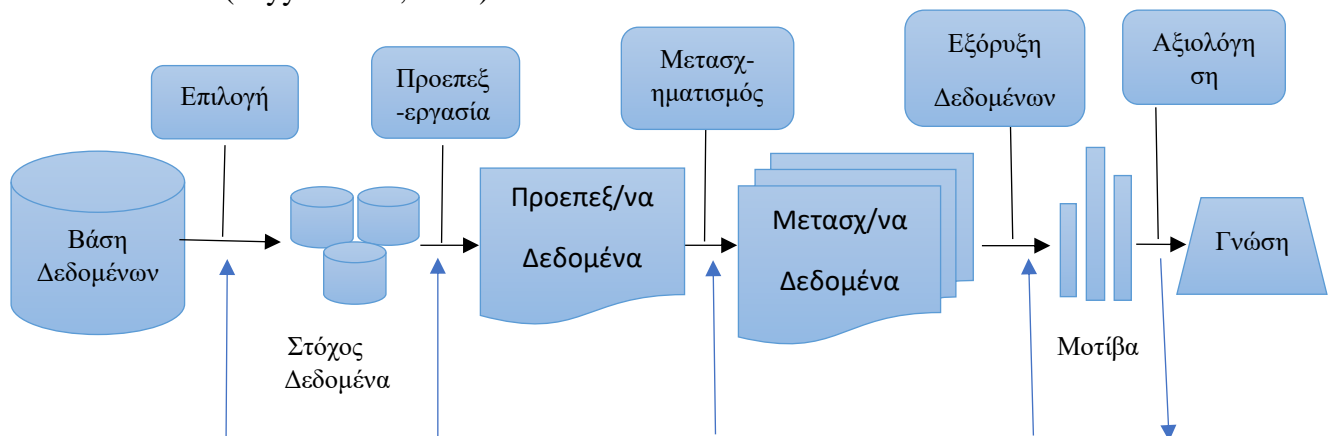
Η εξόρυξη δεδομένων επίσης, είναι γνωστή ως η απόκτηση γνώσεων και ο εντοπισμός ενδιαφερόντων μοτίβων από δεδομένα που είναι αποθηκευμένα σε μεγάλες βάσεις δεδομένων κατά τρόπο ώστε τα μοτίβα και οι γνώσεις να είναι στατιστικά αξιόπιστα, άγνωστα στο παρελθόν και εφαρμόσιμα (Elkan, 2001). Ορίζεται επίσης ως «μια διαδικασία που χρησιμοποιεί τεχνικές στατιστικής, μαθηματικής, τεχνητής νοημοσύνης και μηχανικής μάθησης για την εξαγωγή και τον εντοπισμό χρήσιμων πληροφοριών και έπειτα την απόκτηση γνώσης από μια μεγάλη βάση δεδομένων» (Σχ.3.1, Turban et al., 2007, Frawley et al., 1992).



**Σχήμα 3.1:** Εξόρυξη δεδομένων ως συνδυασμός της στατιστικής, της μηχανικής μάθησης και των βάσεων δεδομένων

Η έννοια της εξόρυξης δεδομένων, συνεπώς από τους ορισμούς που παρατηρούνται στη βιβλιογραφία, φαίνεται να ταυτίζεται με την έννοια της εξόρυξης γνώσης από βάσεις δεδομένων (Knowledge Discovery In Databases-KDD), η οποία δίνει έμφαση στη διαδικασία ανάλυσης δεδομένων παρά στη χρήση συγκεκριμένων μεθόδων ανάλυσης (Fayyad et al., 1996). Η KDD είναι μια διαδικασία που συντίθεται από πέντε βασικά στάδια (Σχ.3.2). Ανάμεσα στα στάδια αυτά παράγονται προϊόντα, τα οποία εφαρμόζονται για την πραγματοποίηση των επόμενων σταδίων. Σημαντικό πριν την έναρξη της διαδικασίας είναι η κατανόηση της αρχικής γνώσης και ο προσδιορισμός στόχων.

1. Στο πρώτο στάδιο γίνεται η επιλογή συνόλου δεδομένων ή ενός υποσυνόλου δεδομένων στο οποίο θα γίνει η εξόρυξη.
2. Στο δεύτερο στάδιο υλοποιείται καθαρισμός και επεξεργασία των δεδομένων που έχει επιλεγθεί στο πρώτο στάδιο.
3. Στο τρίτο στάδιο εφαρμόζονται τεχνικές για τη μείωση διαστάσεων των δεδομένων ή μετασχηματισμοί για τη μείωση του ενεργού αριθμού των υπό εξέταση μεταβλητών για τα δεδομένα. Μετά την ολοκλήρωση του σταδίου αυτού, γίνεται η συσχέτιση των στόχων που έχουν προσδιοριστεί στην αρχή μέσω μεθόδων εξόρυξης όπως, ταξινόμηση ή ομαδοποίηση. Έπειτα πριν την εξόρυξη γνώσης, δηλαδή το επόμενο στάδιο, γίνεται η επιλογή αλγορίθμου και της μεθόδου που θα χρησιμοποιηθούν για την αναζήτηση πρότυπων δεδομένων καθώς και ο καθορισμός των παραμέτρων που θα χρησιμοποιηθούν.
4. Στο τέταρτο στάδιο γίνεται η εξόρυξη δεδομένων. Πριν το επόμενο στάδιο πραγματοποιείται ερμηνεία των πληροφοριών που έχουν προκύψει από τα προηγούμενα στάδια καθώς και η απεικόνιση των αποτελεσμάτων.
5. Στο πέμπτο και τελευταίο στάδιο περιλαμβάνεται η εξαγωγή και η αξιολόγηση γνώσης. Αυτό περιλαμβάνει επίσης τον έλεγχο και την επίλυση πιθανών διενέξεων με γνώσεις που είχαν εξαχθεί παλαιότερα (Fayyad et al., 1996).



**Σχήμα 3.2:** Διαδικασία Εξόρυξης γνώσης από δεδομένα

(Fayyad et al., 1996)

### 3.2.2 Ιστορική Εξέλιξη

Η αναζήτηση πληροφοριών από δεδομένα, εμφανίζεται από την αρχαιότητα, η οποία υλοποιούταν με χειροκίνητο τρόπο. Ωστόσο με τον ταχέως αυξανόμενο όγκο δεδομένων, υπήρξε ανάγκη για αυτοματοποιημένες και πιο αποτελεσματικές προσεγγίσεις εξόρυξης. Το θεώρημα Bayes, η ανάλυση παλινδρόμησης και η εκτίμηση μέγιστης πιθανότητας, αποτελούν μερικές από τις πρώτες τεχνικές που εφαρμόστηκαν για τον εντοπισμό μοτίβων από τα δεδομένα. Με την εξάπλωση, την εμφάνιση και τη συνεχώς αναπτυσσόμενη δύναμη της τεχνολογίας των υπολογιστών, η συλλογή και η αποθήκευση δεδομένων διερευνήθηκαν αξιολογούμενα. Συγκεκριμένα, εξαιτίας της συνεχόμενης αύξησης των όγκων δεδομένων, που μπορούν να διατηρηθούν σε αρχεία υπολογιστών και σε βάσεις δεδομένων, οι χρήστες αυτών των δεδομένων περιμέναν περισσότερες και πιο εξελιγμένες πληροφορίες. Συνεπώς η εξόρυξη δεδομένων εμφανίστηκε για να λύσει αυτές τις ανάγκες (Dunham,2003).

Η πρόοδος στον τομέα της στατιστικής, της τεχνητής νοημοσύνης και της μηχανικής μάθησης, στις αρχές της δεκαετίας του 1950 και των επόμενων δεκαετιών, οδήγησε σε μια μεγάλη ποικιλία εργαλείων ανάκτησης πληροφοριών, όπως νευρωνικά δίκτυα, δέντρα αποφάσεων, γραμμικά μοντέλα ταξινόμησης, πλησιέστεροι γείτονες και πολλών άλλων (Bartschat et al., 2019, Dunham, 2003). Ο όρος «εξόρυξη δεδομένων», επινοήθηκε ουσιαστικά τη δεκαετία του 1980 και κέρδισε αυξανόμενη προσοχή από εκείνη την εποχή και μετά (Bartschat et al., 2019).

Στην δεκαετία του 1990 εμφανίστηκε και χρησιμοποιήθηκε για πρώτη φορά ο όρος «ανακάλυψης γνώσης σε βάσεις δεδομένων» (KDD) , όπου στην πορεία η KDD έλαβε μια συνεχή εξέλιξη από την διασταύρωση των ερευνητικών πεδίων (όπως: μηχανική μάθηση, αναγνώριση προτύπων, βάσεις δεδομένων, κ.α., Fayyad et al., 1996).

Εξαιτίας του μεγάλου όγκου δεδομένων, εκτός από την ανάπτυξη νέων τεχνικών διαχείρισης πληροφοριών από βάσεις δεδομένων, δημιουργήθηκε η ανάγκη για την ανάπτυξη αποθήκες δεδομένων. Παράλληλα βρέθηκαν στο προσκήνιο τα συστήματα υποστήριξης αποφάσεων (DSS), η Διαδικτυακή αναλυτική επεξεργασία (OLAP), οι αλγόριθμοι κανόνων συσχέτισης και ο μετασχηματισμός δεδομένων (Dunham, 2003, Han and Kamber, 2006).

Σήμερα η Εξόρυξη Δεδομένων χρησιμοποιείται σε πολλούς τομείς εφαρμογών όπως η υγεία, οι τηλεπικοινωνίες, η ανάλυση δεδομένων και η χρηματοοικονομική ανάλυση για εντοπισμό απάτης. Επιπλέον στο πλαίσιο των πινάκων δεδομένων, διατίθεται ένα καλά εδραιωμένο εύρος τεχνικών εξόρυξης δεδομένων. Η εξόρυξη δεδομένων πλέον, είναι εντός των δυνατοτήτων πολλών εμπορικών επιχειρήσεων και ερευνητών, χρησιμοποιώντας λογισμικά σε τυπικές επιτραπέζιες μηχανές (όπως το SPSS Clementine, Hicham et al., 2020, Coenen, 2011).

### 3.3 Στατιστικές Τεχνικές

Τα στατιστικά εργαλεία για τον εντοπισμό απάτης είναι πολλά και ποικίλα, καθώς τα δεδομένα από διαφορετικές εφαρμογές μπορεί να ποικίλουν τόσο σε μέγεθος όσο και σε τύπο, ωστόσο έχουν κοινά θέματα. Τέτοια εργαλεία βασίζονται ουσιαστικά στη σύγκριση των παρατηρούμενων δεδομένων με αναμενόμενες τιμές, οι οποίες μπορούν να εξαχθούν με διάφορους τρόπους, ανάλογα με το πλαίσιο. Μπορεί να είναι απλές αριθμητικές περιλήψεις κάποιας πτυχής της συμπεριφοράς, απλές γραφικές περιλήψεις στις οποίες μια ανωμαλία είναι εύκολα εμφανής αλλά και πιο περίπλοκα (πολυμεταβλητά) προφίλ συμπεριφοράς. Οι στατιστικές μέθοδοι ανίχνευσης απάτης μπορεί να είναι υπό επίβλεψη ή χωρίς επίβλεψη (Bolton and Hand, 2002).

Ο στόχος της στατιστικής ανάλυσης, μπορεί να ειπωθεί ότι είναι η απόδοση μιας εκτίμησης (όπου μια υψηλότερη απόδοση μπορεί να θεωρηθεί περισσότερο ύποπτη από μια χαμηλότερη). Όσο υψηλότερη είναι η βαθμολογία, τόσο πιο ασυνήθιστη είναι η παρατήρηση ή τόσο περισσότερο μοιάζει με προηγούμενες τιμές απάτης. Μια πολύ γνωστή μέθοδος είναι ο υπολογισμός βαθμολογιών  $z$ , όπου παρατηρήσεις με απόλυτη τιμή βαθμολογίας  $z$  μεγαλύτερη από 3 μπορούν να θεωρηθούν ως ακραίες τιμές (Baesens et al., 2015, Bolton and Hand, 2002).

Στην ενότητα αυτή, θα πραγματοποιηθεί μια παρουσίαση των πιο συχνών χρησιμοποιούμενων στατιστικών τεχνικών. Συγκεκριμένα θα εστιάσει στις διαδικασίες των μεθόδων της γραμμικής παλινδρόμησης (Linear Regression), της διακριτικής ανάλυσης (Discreet Analysis) και της ανάλυσης logit and probit.

#### 3.3.1 Γραμμική Παλινδρόμηση (Linear Regression)

Η γραμμική παλινδρόμηση είναι αναμφίβολα η πιο συχνά χρησιμοποιούμενη τεχνική για την μοντελοποίηση μιας μεταβλητής συνεχούς στόχου. Αποκαλύπτει τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών (Baesens et al., 2015, Han and Kamber, 2006).

Σε περίπτωση μιας εξαρτημένης μεταβλητής  $y$  και μιας ανεξάρτητης μεταβλητής  $x$ , πρόκειται για το μοντέλο απλής γραμμικής παλινδρόμησης. Στην συγκεκριμένη περίπτωση γίνεται μοντελοποίηση του  $y$  ως γραμμική συνάρτηση του  $x$ , όπου η διακύμανση του  $y$  θεωρείται σταθερή και οι συντελεστές παλινδρόμησης  $a_0$ ,  $a_1$  καθορίζουν την τομή  $y$  (με τον άξονα  $y'$ ) και την κλίση της γραμμής, αντίστοιχα. Οι συντελεστές παλινδρόμησης μπορούν να θεωρηθούν ως βάρη, έτσι ώστε να μπορούμε να γράψουμε την εξίσωση (Han and Kamber, 2006):

$$y = a_0 + a_1x \quad (3.1)$$

Οι συντελεστές αυτοί μπορούν να εκτιμηθούν με την μέθοδο των ελαχίστων τετραγώνων, η οποία εκτιμά την καλύτερη ευθεία γραμμή ως αυτή που ελαχιστοποιεί το σφάλμα μεταξύ των πραγματικών δεδομένων και της εκτιμώμενης γραμμής:

$$a_1 = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^m (x_i - \bar{x})^2} \quad (3.2)$$

$$a_0 = \bar{y} - a_1 \bar{x} \quad (3.3)$$

όπου:

m: σύνολο δεδομένων της μορφής  $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$

$\bar{x}$ : είναι η μέση τιμή των  $x_1, x_2, \dots, x_m$

$\bar{y}$ : είναι η μέση τιμή των  $y_1, y_2, \dots, y_m$

Μια επέκταση της απλής γραμμικής παλινδρόμησης είναι η πολλαπλή γραμμική παλινδρόμηση. Η πολλαπλή γραμμική παλινδρόμηση περιλαμβάνει περισσότερες από μια ανεξάρτητες μεταβλητές. Συνεπώς επιτρέπεται η μοντελοποίηση της εξαρτημένης μεταβλητής  $y$  ως γραμμική συνάρτηση  $n$  μεταβλητών ή χαρακτηριστικών  $x_1, x_2, \dots, x_n$  (Han and Kamber, 2006):

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n + \varepsilon_i, i=1, 2, \dots, n \quad (3.4)$$

όπου  $\varepsilon_i$  είναι σφάλματα που θεωρούνται ανεξάρτητες τυχαίες μεταβλητές ενώ τα  $x_1, x_2, \dots, x_n$  όπως και στο απλό γραμμικό μοντέλο δεν θεωρούνται τυχαίες.

### 3.3.2 Διακριτική Ανάλυση (Discriminant Analysis)

Η διακριτική ανάλυση είναι μια πολυπαραγοντική μέθοδος στατιστικής ταξινόμησης, η οποία χρησιμοποιείται εδώ και δεκαετίες από ερευνητές και επαγγελματίες για την ανάπτυξη μοντέλων ταξινόμησης (Zorounidis and Doumpos, 2002).

Ο στόχος της είναι να ανακαλύψει κανόνες που περιγράφουν το διαχωρισμό μεταξύ ομάδων παρατηρήσεων. Επιπλέον, επιτρέπει την ταξινόμηση νέων παρατηρήσεων σε μια από τις γνωστές ομάδες. Οι κανόνες διάκρισης, στην κλασική προσέγγιση, βασίζονται συχνά στον εμπειρικό μέσο όρο και τον πίνακα συνδιακύμανσης των δεδομένων ή τμημάτων των δεδομένων (Hubert and Driessen, 2004).

Στην γενική περίπτωση όπου η ταξινόμηση περιλαμβάνει  $q$  ομάδες, αναπτύσσονται  $q-1$  γραμμικές συναρτήσεις της μορφής (Zorounidis and Doumpos, 2002):

$$z = a + b_1 g_1 + b_2 g_2 + \dots + b_n g_n \quad (3.5)$$



Όπου:

$g_1, g_2, \dots, g_n$ : είναι τα χαρακτηριστικά που περιγράφουν τις εναλλακτικές  $X_1, X_2, \dots, X_n$

$a$ : είναι ο σταθερός όρος

$b_1, b_2, \dots, b_n$ : είναι οι συντελεστές των χαρακτηριστικών στη συνάρτηση διάκρισης.

Για την εκτίμηση του σταθερού όρου  $a$  και των συντελεστών  $b_1, b_2, \dots, b_n$  λαμβάνουμε υπόψη ότι :

- a. Τα δεδομένα ακολουθούν την πολυμεταβλητή κανονική κατανομή (multivariate normal distribution)
- b. Οι πίνακες διακύμανσης – συνδιακύμανσης για κάθε ομάδα είναι ίσοι

Επειδή όμως οι εκτιμήσεις, επηρεάζονται σε μεγάλο βαθμό από ακραίες τιμές, οι κανόνες διάκρισης γίνονται ακατάλληλες σε μολυσμένα (αλλοιωμένα, κακόβουλα ή τυχαία) σύνολα δεδομένων. Με την εισαγωγή λοιπόν, ισχυρών εκτιμήσεων θέσης και διασποράς σε γενικευμένους κανόνες μέγιστης πιθανότητας σε κανονικές κατανομές, απαιτείται η λήψη ισχυρών κανόνων διάκρισης. Αυτό επιτρέπει τη διάκριση μεταξύ πολλών πληθυσμών, με ίση ή άνιση δομή συνδιακύμανσης και με ίσες ή άνισες πιθανότητες μέλους (Hubert and Driessen, 2004).

### 3.3.3 Ανάλυση Logit και Probit

Η ανάλυση logit και probit προέρχεται από το πεδίο της οικονομετρίας. Για την κατανόηση των εννοιών που αφορούν την ερμηνεία των μοντέλων logit και probit, σημαντικό ρόλο έπαιξε η θεωρία της διακριτής επιλογής που έκανε την εμφάνισή της, την δεκαετία του 1970. Βασίζονται στην ανάπτυξη μιας μη γραμμικής συνάρτησης που μετρά την πιθανότητα ένταξης στην ομάδα, για τις υπό εξέταση εναλλακτικές λύσεις. Η διαφορά μεταξύ των δυο προσεγγίσεων αφορά τη μορφή της συνάρτησης που χρησιμοποιείται. Συγκεκριμένα, η ανάλυση logit χρησιμοποιεί τη λογιστική συνάρτηση, ενώ η ανάλυση probit χρησιμοποιεί την αθροιστική συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής. Σύμφωνα με αυτές τις συναρτήσεις και θεωρώντας ένα πρόβλημα ταξινόμησης δυο ομάδων, η πιθανότητα μιας εναλλακτικής  $x_i$  να ανήκει στην ομάδα  $c_1$  ή  $c_2$  ορίζεται ως εξής (Zorounidis and Doumpos, 2002):

$$1. P_i = F(a + g'_i b) = \frac{1}{1 + e^{-a - g'_i b}} \quad (\text{Logit, 3.6})$$

$$2. P_i = f(a + g'_i b) = \int_{-\infty}^{a + g'_i b} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{z^2}{2}} dz \quad (\text{Probit, 3.7})$$

Όπου:

$P_i$  : η πιθανότητα ένταξης της εναλλακτικής  $i$  στην ομάδα  $c_1$  ή  $c_2$

$a$ : είναι ο σταθερός όρος

$b$ : είναι το διάνυσμα το οποίο περιλαμβάνει τους συντελεστές των χαρακτηριστικών

$g_i$ : αποτελούν τα χαρακτηριστικά που περιγράφουν την εναλλακτική

Το διάνυσμα  $b$  και ο σταθερός όρος  $a$  εκτιμάται, χρησιμοποιώντας την ιδέα της μέγιστης πιθανότητας. Η βελτιστοποίηση μέγιστης πιθανότητας επιλέγει παραμέτρους με τέτοιο τρόπο ώστε να μεγιστοποιεί την πιθανότητα λήψης του δείγματος. Συγκεκριμένα η εκτίμηση των παραμέτρων γίνεται μέσω της μεγιστοποίησης της ακόλουθης συνάρτησης (Baesens et al., 2015, Zorounidis and Doumpros, 2002):

$$\ln L = \sum_{\forall x_i \in c_2} \ln (P_i) + \sum_{\forall x_i \in c_1} \ln (1 - P_i) \quad (3.8)$$

Η μεγιστοποίηση της συνάρτησης πιθανότητας αποτελεί ένα μη γραμμικό πρόβλημα βελτιστοποίησης που συχνά η επίλυση του είναι δύσκολη. Οι Altman et al.(1981) αναφέρουν ότι εάν είναι δυνατή η ανάπτυξη ενός γραμμικού συνδυασμού των χαρακτηριστικών που να διακρίνει τις ομάδες μεταξύ τους, τότε η διαδικασία βελτιστοποίησης δεν συγκλίνει σε μια βέλτιστη λύση.

Για τον ορισμό των ομάδων  $c_1$  ή  $c_2$ , γίνεται εκχώρηση μια δυαδικής μεταβλητής η οποία λαμβάνει τιμές 0-1 έτσι ώστε οι παραπάνω εξισώσεις (3.6, 3.7) να παρέχουν την πιθανότητα ότι μια εναλλακτική εντάσσεται σε μια από της δυο ομάδες. Εάν για παράδειγμα,  $c_1 \rightarrow 0$  και  $c_2 \rightarrow 1$  τότε οι εξισώσεις δίνουν την πιθανότητα ότι η εναλλακτική ανήκει στην ομάδα  $c_2$ , στην αντίθετη περίπτωση ( $c_1 \rightarrow 1$  και  $c_2 \rightarrow 0$ ) δίνεται η πιθανότητα ότι η εναλλακτική ανήκει στην ομάδα  $c_1$ . Σε περίπτωση προβλημάτων ταξινόμησης περισσότερων ομάδων, η ανάλυση logit και probit μπορεί να χρησιμοποιηθεί με την μορφή πολυωνυμικών ή διατεταγμένων μοντέλων, με την διαφορά ότι τα πρώτα μοντέλα θεωρούν έναν ονομαστικό ορισμό για τις ομάδες ενώ τα δεύτερα έναν ορισμό τάξεως (Zorounidis and Doumpros, 2002).

### 3.4 Πολυκριτήριες Μεθοδολογίες

Η πολυκριτήρια ανάλυση (Multicriteria Analysis-MA) είναι συνδεδεμένη με την έννοια της λήψης αποφάσεων πολλαπλών κριτηρίων (Multiple-criteria decision making-MCDM) ή της ανάλυσης αποφάσεων πολλαπλών κριτηρίων (Multiple-criteria decision analysis-MCDA). Πρόκειται για ένα σύνολο μεθόδων που δίνει την δυνατότητα συγκέντρωσης πολλών κριτηρίων αξιολόγησης για την επιλογή, την κατάταξη, την ταξινόμηση ή την περιγραφή ενός συνόλου εναλλακτικών (Zorounidis, 1999).

Στις αρχές ερευνητικά, οι μέθοδοι MCDA εστίασαν την προσοχή τους στην ανάπτυξη μεθοδολογιών επιλογής και κατάταξης, για την υποστήριξη λήψης αποφάσεων. Ενώ, για τη μελέτη προβλημάτων ταξινόμησης, πραγματοποιήθηκε σημαντική έρευνα λίγο αργότερα (Zorounidis and Doumpos, 2002).

Στην ενότητα αυτή, θα γίνει αναφορά σε δυο προσεγγίσεις MCDA, στην μέθοδο UTADIS (Utilites Additives Discriminates) και στη μέθοδο MHDIS (Multi-group Hierarchical Discrimination). Το κύριο πλεονέκτημα τους, είναι ότι δεν κάνουν υποθέσεις όπως οι παραδοσιακές στατιστικές και οικονομετρικές τεχνικές (Pasiouras et. al., 2007).

### 3.4.1 Μέθοδος UTADIS

Η μέθοδος UTADIS προέρχεται από την οικογένεια των μεθόδων UTA (Utilites Additives). Αποτελεί μια μέθοδο τακτικής παλινδρόμησης, που χρησιμοποιεί ως βάση την προσέγγιση διαχωρισμού προτιμήσεων MCDA. Ο στόχος της μεθόδου UTADIS, είναι η ταξινόμηση των εναλλακτικών λύσεων σε προκαθορισμένες κατηγορίες με το ελάχιστο σφάλμα ταξινόμησης. Ένας αναλυτής με την χρήση της μεθόδου, πέρα της ταξινόμησης των εναλλακτικών, μπορεί να καθορίσει ποια εναλλακτική είναι η καλύτερη και ποια η χειρότερη (Zorounidis and Doumpos, 2001).

Συγκεκριμένα στοχεύει στην ανάπτυξη μιας προσθετικής συνάρτησης χρησιμότητας και στην εκτίμηση κατωφλίων για την ταξινόμηση των εναλλακτικών. Θεωρώντας ένα πεπερασμένο σύνολο εναλλακτικών  $A=\{a_1, a_2, \dots, a_n\}$ , οι οποίες για να ταξινομηθούν σε ένα προκαθορισμένο σύνολο ομάδων  $C_1, C_2, \dots, C_n$  με την  $C_1$  να αποτελεί την καλύτερη ομάδα και την  $C_n$  την χειρότερη, αξιολογούνται βάση ενός συνόλου κριτηρίων  $g_1, g_2, \dots, g_n$ . Ως εκ τούτου η γενική μορφή μιας προσθετικής συνάρτησης χρησιμότητας είναι της μορφής (Zorounidis and Doumpos, 2002, Pasiouras et. al., 2007):

$$U(\alpha) = \sum_{i=1}^n p_i u_i(g_i) \quad (3.9)$$

Όπου :

$U(\alpha)$ : είναι η συνάρτηση ολικής χρησιμότητας (global utility)

$g$ : αποτελεί το διάνυσμα κριτηρίων ( $g_1, g_2, \dots, g_n$ )

$p_i$ : αποτελούν τα βάρη των κριτηρίων  $g_i$ , συγκεκριμένα δείχνουν την σημασία κάθε κριτηρίου και ορίζονται με τέτοιο τρόπο ώστε να μην είναι αρνητικοί και να αθροίζουν στην μονάδα ( $p_1+p_2+\dots+p_n=1$ )

$u_i(g_i)$ : είναι η συνάρτηση οριακής χρησιμότητας (marginal utility) του κριτηρίου  $g_i$

Η συνάρτηση ολικής χρησιμότητας μετρά την συνολική επίδοση κάθε εναλλακτικής σε όλα τα κριτήρια, αποδίδοντας μεγαλύτερη βαθμολογία στις εναλλακτικές που ανήκουν στην πρώτη ομάδα  $C_1$  και σταδιακά χαμηλότερη βαθμολογία στις εναλλακτικές που ανήκουν στις κατώτερες ομάδες. Ενώ οι συναρτήσεις μερικής χρησιμότητας δίνουν ένα μηχανισμό για την αποσύνθεση του συνολικού αποτελέσματος, της συνάρτησης ολικής χρησιμότητας, σχετικά με την ατομική αξιολόγηση σε επίπεδο κριτηρίου. Η κλίμακα που κυμαίνονται οι συναρτήσεις ολικής και οριακής χρησιμότητας είναι μεταξύ 0 και 1 (Pasiouras et. al., 2007).

Η ταξινόμηση στην απλή περίπτωση δύο ομάδων πραγματοποιείται συγκρίνοντας την συνάρτηση ολικής χρησιμότητας κάθε εναλλακτικής με ένα σημείο διαχωρισμού, το οποίο ορίζεται στην κλίμακα χρησιμότητας μεταξύ 0 και 1. Οι εναλλακτικές με ολική χρησιμότητα υψηλότερη από το σημείο διαχωρισμού ανήκουν στην ομάδα  $C_1$  ενώ οι εναλλακτικές με βαθμολογία χαμηλότερη από το σημείο διαχωρισμού ανήκουν στην ομάδα  $C_2$  (Zorounidis and Doumpos, 2002).

Σε περίπτωση προβλήματος ταξινόμησης των εναλλακτικών λύσεων σε περισσότερες από δυο ομάδες, έστω  $n$ -ομάδες ( $C_1, C_2, \dots, C_n$ ), επιλέγονται  $n-1$  σημεία διαχωρισμού ( $u_1, u_2, \dots, u_{n-1}$ ). Τα σημεία διαχωρισμού αυτά, αναφέρονται και ως κατώφλια χρησιμότητας τα οποία, όπως και στην περίπτωση ταξινόμησης δυο ομάδων, ορίζονται σε κλίμακα χρησιμότητας μεταξύ 0 και 1. Με βάση τα παραπάνω, η ταξινόμηση των εναλλακτικών πραγματοποιείται σύμφωνα με τους παρακάτω κανόνες ταξινόμησης (Zorounidis and Doumpos, 2002, Pasiouras et. al., 2007):

$$\begin{aligned} & \text{An } U(\alpha) \geq u_1 \text{ τότε } \alpha \in C_1 \\ & \text{An } u_2 \leq U(\alpha) < u_1 \text{ τότε } \alpha \in C_2 \\ & \dots\dots \\ & \text{An } U(\alpha) < u_{n-1} \text{ τότε } \alpha \in C_n \end{aligned} \quad (3.10)$$

### 3.4.2 Μέθοδος MHDIS

Η πολυκριτηριακή ιεραρχική διάκριση (Multicriteria Hierarchical Discrimination-MHDIS) είναι μια μέθοδος μη παραμετρικής διάκρισης, η οποία στηρίζεται στις έννοιες ανάλυσης αποφάσεων πολλαπλών κριτηρίων (MCDA) και στις τεχνικές μαθηματικού προγραμματισμού. Η μέθοδος στοχεύει στην ανάπτυξη μοντέλων διάκρισης, με σκοπό την ταξινόμηση ενός συνόλου εναλλακτικών λύσεων σε προκαθορισμένες ομάδες με τακτικό τρόπο. Για την ανάπτυξη των μοντέλων η MHDIS βασίζεται στην φιλοσοφία της παλινδρόμησης (Spathis et al., 2004).

Σε αντίθεση με την UTADIS μέθοδο, η MHDIS διακρίνει τις ομάδες σταδιακά, ξεκινώντας από τη διάκριση της πρώτης ομάδας από όλες τις υπόλοιπες, και στη συνέχεια προχωρά στη διάκριση μεταξύ των εναλλακτικών λύσεων που ανήκουν στις υπόλοιπες ομάδες (Pasiouras et. al., 2007).

Έστω ένα σύνολο αναφοράς  $A$ , το οποίο αποτελείται από ένα πεπερασμένο πλήθος εναλλακτικών  $(a_1, a_2, \dots, a_n)$ , όπου η κάθε εναλλακτική αξιολογείται σε ένα σύνολο κριτηρίων  $g_1, g_2, \dots, g_n$ , για την ταξινόμησή τους σε ένα σύνολο  $n$ -ομάδων  $(C_1, C_2, \dots, C_n)$ .

Η διαδικασία ιεραρχικής διάκρισης λοιπόν, που εφαρμόζεται στη μέθοδο αποτελείται από  $n-1$  στάδια. Κάθε στάδιο αντιμετωπίζεται ως πρόβλημα ταξινόμησης δυο ομάδων με σκοπό τη διάκριση των εναλλακτικών που ανήκουν στην πρώτη ομάδα  $C_1$  από τις εναλλακτικές των υπόλοιπων ομάδων. Δεδομένου ότι οι ομάδες ορίζονται με τακτικό τρόπο, αυτό ισοδυναμεί στην διάκριση της ομάδας  $C_1$  από το σύνολο των ομάδων  $(C_2, \dots, C_n)$ . Συνεπώς σε κάθε στάδιο της διαδικασίας υπάρχουν δυο επιλογές για την ταξινόμηση μιας εναλλακτικής (Zorounidis and Doumpos, 2002):

1. Η επιλογή ότι η εναλλακτική ανήκει στην πρώτη ομάδα  $C_1$
2. Η επιλογή ότι η εναλλακτική ανήκει στο σύνολο  $(C_2, \dots, C_n)$ .

Όσες εναλλακτικές εντοπίστηκαν να ανήκουν στην ομάδα  $C_1$  (σωστά είτε λανθασμένα) εξαιρούνται από την περαιτέρω ανάλυση. Αυτή η διαδικασία επαναλαμβάνεται μέχρι και το τελευταίο στάδιο ( $n-1$ ), όπου έχουν ληφθεί υπόψη όλες οι ομάδες.

Αξίζει να σημειωθεί ότι στην MHDIS σε σχέση με την μέθοδο της UTADIS, για την ταξινόμηση των εναλλακτικών αναπτύσσονται δυο προσθετικές συναρτήσεις χρησιμότητας :

$$U_k(a) = \sum_{i=1}^n p_{ki} u_{ki}(g_i) \quad (3.11)$$

$$U_{\sim k}(a) = \sum_{i=1}^n p_{\sim ki} u_{\sim ki}(g_i) \quad (3.12)$$

$$K=1,2,\dots,n-1$$

Όπου:

$U_k(a)$  και  $U_{\sim k}(a)$ : είναι οι συναρτήσεις ολικής χρησιμότητας (global utilities)

$g_i$ : είναι το διάνυσμα των κριτηρίων ( $g_1, g_2, \dots, g_n$ )

$p_{ki}$  και  $p_{\sim ki}$ : είναι τα βάρη των κριτηρίων, τα οποία αθροίζονται στην μονάδα

$u_{ki}$  και  $u_{\sim ki}$ : είναι οι συναρτήσεις οριακής χρησιμότητας (marginal utility) για κάθε κριτήριο  $g_i$

Συγκεκριμένα η πρώτη συνάρτηση  $U_k(a)$  περιγράφει τις εναλλακτικές της ομάδας  $C_1$ , ενώ η δεύτερη συνάρτηση  $U_{\sim k}(a)$  περιγράφει τις υπόλοιπες εναλλακτικές που ταξινομούνται στις κατώτερες ομάδες  $(C_2, \dots, C_n)$ . Η εκτίμηση των βαρών των κριτηρίων στις συναρτήσεις χρησιμότητας καθώς και των συναρτήσεων οριακής χρησιμότητας πραγματοποιείται μέσω μαθηματικών τεχνικών προγραμματισμού. Σε κάθε στάδιο συγκεκριμένα, της διαδικασίας ιεραρχικής διάκρισης, επιλύονται δυο

γραμμικά προγράμματα και ένα μικτού ακεραίου για να εκτιμηθούν βέλτιστα οι δυο πρόσθετες συναρτήσεις χρησιμότητας και να ελαχιστοποιηθεί το σφάλμα ταξινόμησης (Pasiouras et al. , 2007).

Με βάση τις παραπάνω συναρτήσεις (3.11), (3.12), ο κανόνας για την ταξινόμηση οποιασδήποτε εναλλακτικής έχει την ακόλουθη μορφή:

$$\begin{aligned}
 & \text{Αν } U_1(a) \geq U_{-1}(a) \text{ τότε } a \in C_1 \\
 & \text{Αλλιώς αν } U_2(a) \geq U_{-2}(a) \text{ τότε } a \in C_2 \\
 & \dots\dots\dots \\
 & \text{Αλλιώς αν } U_{n-1}(a) \geq U_{-n-1}(a) \text{ τότε } a \in C_{n-1} \\
 & \text{Αλλιώς } a \in C_n \qquad \qquad \qquad (3.13)
 \end{aligned}$$

### 3.5 Μεθοδολογίες Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη (Artificial Intelligence-AI), χάρη στην εφαρμογή της στον έλεγχο χρηματοοικονομικής απάτης, έχει επηρεάσει σημαντικά τους στόχους του ελέγχου και τους τρόπους επίτευξης των στόχων. Αν και στο παρελθόν υπήρχαν περιορισμοί εξαιτίας των τότε διαθέσιμων τεχνολογιών, η αυξανόμενη πρόοδος και εφαρμογή των τεχνολογιών AI προκάλεσε ανεξίτηλες αλλαγές στο κοινωνικό περιβάλλον τροποποιώντας τα συμβατικά πρότυπα ζωής και εργασίας. Για την καλύτερη προσαρμογή της τρέχουσας κοινωνίας όπου οι πληροφορίες και η ευφυΐα εξελίσσονται γρήγορα, όλοι οι κλάδοι όπως και ο τομέας του εντοπισμού απάτης (ελέγχου) αναδιαρθρώνουν ή βελτιώνουν τις στρατηγικές και τις διαδικασίες τους.

Οι τεχνολογίες τεχνητής νοημοσύνης δημιουργούν περισσότερη πρόσβαση, σε ένα μεγαλύτερο εύρος δεδομένων. Τα υποκείμενα δεδομένα, μετά την ανάλυση, θα δημιουργήσουν έναν τεράστιο αριθμό χρήσιμων πληροφοριών πολλαπλών ειδών που μπορούν να επαληθεύσουν άμεσα τις πληροφορίες. Στη διαδικασία που είναι σύμφωνη με τη χρονική και χωρική αλληλουχία που είναι εγγενής στα θέματα ελέγχου (π.χ. τάση, δομή, σχέσεις και ούτω καθεξής), οι ελεγκτές και οι οργανισμοί μπορούν να ορίσουν κανόνες κρίσης και περιοριστικούς όρους για να δημιουργήσουν μαθηματικούς ή λογικούς τύπους, οι οποίοι μπορούν να εφαρμοστούν ως τεχνικά μέτρα για την επαλήθευση της πραγματικής κατάστασης (Gao and Han, 2021).

Στην ενότητα αυτή, θα γίνει μια αναφορά σε διάφορες μεθόδους τεχνητής νοημοσύνης. Η επιλογή των μεθόδων αυτών έγινε με βάση τον βαθμό διάδοσης που αυτές εμφανίζουν και με προσπάθεια να καλυφθεί ένα μεγάλο μέρος.

#### 3.5.1 Δέντρα Αποφάσεων (Decision Trees)

Τα δέντρα αποφάσεων (Decision Trees) είναι αλγόριθμοι αναδρομικής κατανομής (Recursive-Partitioning Algorithms, RPAs) που δημιουργούν μια δομή, η οποία μοιάζει

με δέντρο που αναπαριστά μοτίβα σε ένα υποκείμενο σύνολο δεδομένων (Baesens et al., 2015).

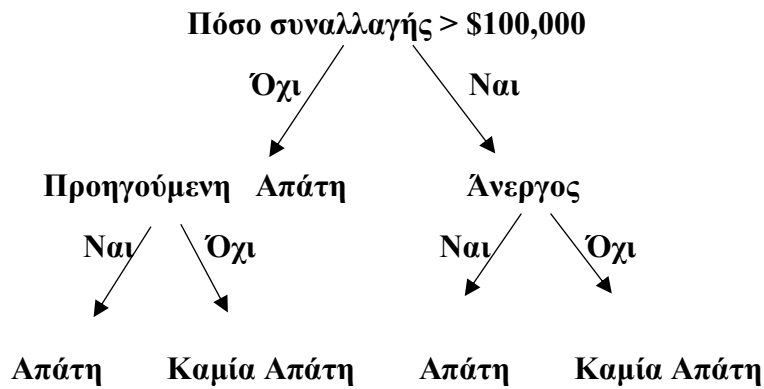
Κάθε κόμβος στη δομή δέντρου, αντιπροσωπεύει μια δοκιμή σε ένα χαρακτηριστικό και κάθε κλάδος αντιπροσωπεύει ένα αποτέλεσμα της δοκιμής. . Με αυτόν τον τρόπο προσπαθεί να χωρίσει τις παρατηρήσεις σε παρόμοιες αποκλειστικές υποομάδες. Συνεπώς το δείγμα διαιρείται διαδοχικά σε υποσύνολα, έως ότου να μην μπορούν οι υποομάδες να διασπαστούν ή η περαιτέρω διάσπαση τους μην επιφέρει κάποια στατιστικά σημαντική διαφορά. Συγκεκριμένα ο πάνω κόμβος είναι ο ριζικός κόμβος που καθορίζει μια συνθήκη δοκιμής, της οποίας το αποτέλεσμα αντιστοιχεί σε έναν κλάδο που οδηγεί σε ένα εσωτερικό κόμβο. Οι τερματικοί κόμβοι του δέντρου εκχωρούν τις ταξινομήσεις και αναφέρονται επίσης ως κόμβοι απόφασης (Kirkos et al., 2007, Baesens et al., 2015).

Η μέθοδος δέντρου αποφάσεων, η οποία προέρχεται από την έννοια συστημάτων μάθησης, ανέπτυξε τους αλγόριθμους ID3 (Iterative Dichotomiser 3) και C4.5 που μπορούν να ασχοληθούν με συνεχή δεδομένα. Στη βιβλιογραφία έχουν αναφερθεί και άλλοι αλγόριθμοι, όπως ο αλγόριθμος CART (Classification & Regression Tree) και ο αλγόριθμος CHAID (Chi-Squared Automatic Interaction Detection). Οι αλγόριθμοι αυτοί, έχουν διαφορά ως προς τον τρόπο που απαντούν στις βασικές αποφάσεις για την κατασκευή του δέντρου (Baesens et al., 2015, Shen et al., 2007 ):

- Απόφαση Διαχωρισμού: Ποια μεταβλητή θα διαχωριστεί σε ποια τιμή;
- Απόφαση Διακοπής : Πότε πρέπει να σταματήσει η προσθήκη κόμβων στο δέντρο
- Απόφαση Ανάθεσης : Ποια κατηγορία να αντιστοιχιστεί σε έναν τερματικό κόμβο.

Θα μπορούσε να αναφερθεί ότι τα δέντρα αποφάσεων αποτελούν εργαλεία πρόβλεψης υποστήριξης αποφάσεων, τα οποία δημιουργούν μια χαρτογράφηση από τις παρατηρήσεις έως τις πιθανές συνέπειες. Τα δέντρα αποφάσεων επίσης, είναι πολύ ισχυρές τεχνικές, οι οποίες μπορούν να χρησιμοποιηθούν για πιο σύνθετα όρια αποφάσεων. Αξίζει να σημειωθεί ότι αποτελούν μια μη παραμετρική μέθοδο, με την έννοια ότι δεν απαιτούνται υποθέσεις κανονικότητας ή ανεξαρτησίας για τη δημιουργία τους. Ωστόσο παρουσιάζουν ένα σημαντικό μειονέκτημα, συγκεκριμένα εξαρτώνται σε μεγάλο βαθμό από το δείγμα που χρησιμοποιήθηκε για να κατασκευαστούν. Κατά συνέπεια μια μικρή απόκλιση στο υποκείμενο δείγμα μπορεί να αποφέρει ένα εντελώς διαφορετικό δέντρο απόφασης (Baesens et al., 2015, Sharma and Panigrahi, 2012).

Για παράδειγμα, παρακάτω φαίνεται ένα δέντρο απόφασης σε μία ρύθμιση ανίχνευσης απάτης (Σχήμα 3.3). Στο συγκεκριμένο παράδειγμα η απόφαση διαχωρισμού είναι το ποσό της συναλλαγής (μεγαλύτερο από \$100000), αν υπάρχει προηγούμενη απάτη ( Ναι ή Όχι) και η μεταβλητή Άνεργος, η οποία διαχωρίζεται με την ένδειξη Ναι ή Όχι. Η απόφαση ανάθεσης, όπου η κατηγορία αντιστοιχεί σε ένα τερματικό κόμβο, είναι ένδειξη απάτη ή καμία απάτη (Baesens et al., 2015).



**Σχήμα 3.3:** Παράδειγμα Δέντρο Απόφασης.

(Baesens et al., 2015)

### 3.5.2 Νευρωνικά Δίκτυα (Neural Networks)

Τα νευρωνικά δίκτυα είναι μη γραμμικά εργαλεία μοντελοποίησης στατιστικών δεδομένων που εμπνέονται από τη λειτουργικότητα του ανθρώπινου εγκεφάλου χρησιμοποιώντας ένα σύνολο διασυνδεδεμένων κόμβων. Τα νευρωνικά δίκτυα εφαρμόζονται ευρέως τόσο στην ταξινόμηση όσο και στην ομαδοποίηση (Sharma and Panigrahi, 2012).

Συχνά αναφέρονται ως τεχνητά νευρωνικά δίκτυα και έχουν αναπτυχθεί από ερευνητές τεχνητής νοημοσύνης ως μια καινοτόμος μεθοδολογία μοντελοποίησης περίπλοκων προβλημάτων. Μπορούμε να φανταστούμε τον ανθρώπινο εγκέφαλο, οποίος αποτελείται από ένα τεράστιο αριθμό οργανωμένων νευρώνων σε ένα εξαιρετικά πολύπλοκο δίκτυο. Κάθε νευρώνας είναι μια μεμονωμένη μονάδα επεξεργασίας, όπου λαμβάνει ένα σήμα εισόδου (ερέθισμα από αισθητήρες σώματος ή σήμα εξόδου από άλλους νευρώνες), το οποίο μετά από μια φάση επεξεργασίας παράγει ένα σήμα εξόδου που μεταφέρεται σε άλλους νευρώνες για περαιτέρω επεξεργασία. Το αποτέλεσμα της συνολικής διαδικασίας είναι η ενέργεια ή η απόφαση που λαμβάνεται σύμφωνα με το αρχικό ερέθισμα (Zorounidis and Doumpos, 2002).

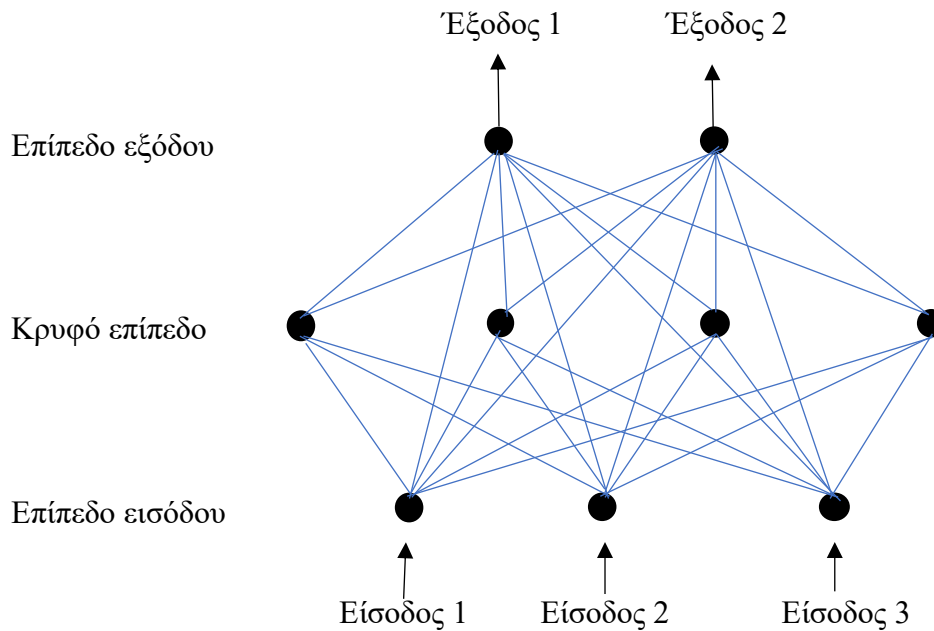
Λόγω ότι αποτελούν μια μη γραμμική σχέση χαρτογράφησης από το χώρο εισόδου στο χώρο εξόδου, τα νευρωνικά δίκτυα μπορούν να μάθουν από τις δοθείσες περιπτώσεις και να συνοψίσουν τις εσωτερικές αρχές των δεδομένων, ακόμη και χωρίς να γνωρίζουν τις πιθανές αρχές δεδομένων που θα ακολουθήσουν (Shen et al., 2007).

Γενικά, κάθε νευρωνικό δίκτυο αποτελεί ένα δίκτυο παράλληλων μονάδων επεξεργασίας (νευρώνων) οργανωμένων σε στρώματα. Μια τυπική δομή ενός νευρωνικού δικτύου (Σχήμα 3.4) περιλαμβάνει τα ακόλουθα δομικά στοιχεία (Zorounidis and Doumpos, 2002):



- Ένα επίπεδο εισόδου, το οποίο αποτελείται από ένα σύνολο κόμβων (μονάδες επεξεργασίας-νευρώνες), έναν για κάθε είσοδο στο δίκτυο.
- Ένα επίπεδο εξόδου που αποτελείται από έναν ή περισσότερους κόμβους ανάλογα με τη μορφή της επιθυμητής εξόδου του δικτύου. Στα προβλήματα ταξινόμησης, ο αριθμός των κόμβων επιπέδου εξόδου καθορίζεται ανάλογα με τον αριθμό ομάδων. Έστω ένα πρόβλημα ταξινόμησης δύο ομάδων ( $C_1, C_2$ ) το επίπεδο εξόδου μπορεί να περιλαμβάνει μόνο έναν κόμβο που παίρνει δύο τιμές, την τιμή 1 για την ομάδα  $C_1$  και την τιμή 2 για την ομάδα  $C_2$  (αυτές είναι αυθαίρετες επιλεγμένες τιμές και είναι δυνατό οποιοδήποτε άλλο ζεύγος). Στη γενική περίπτωση όπου υπάρχουν  $q$  ομάδες, ο αριθμός των κόμβων στο επίπεδο εξόδου συνήθως ορίζεται ως ο μικρότερος ακέραιος που είναι μεγαλύτερος από  $\log_2 q$ . Εναλλακτικά, είναι επίσης δυνατόν να οριστεί ο αριθμός των κόμβων εξόδου ίσος με τον αριθμό των ομάδων.
- Μια σειρά από ενδιάμεσα επίπεδα που αναφέρονται ως κρυφά στρώματα. Οι κόμβοι κάθε κρυφού στρώματος συνδέονται πλήρως με τους κόμβους του επόμενου στρώματος. Επιπλέον, είναι επίσης δυνατό να εξεταστούν πιο περίπλοκες δομές όπου όλα τα στρώματα είναι πλήρως συνδεδεμένα μεταξύ τους. Τέτοιες γενικές δομές δικτύου είναι γνωστές ως πλήρως συνδεδεμένα νευρωνικά δίκτυα. Για τον καθορισμό του αριθμού των κρυφών στρωμάτων δεν υπάρχει κάποιος γενικός κανόνας, αυτό εκτελείται συνήθως μέσω διαδικασιών δοκιμής και σφάλματος.

Τα νευρωνικά δίκτυα μπορούν είναι προσαρμοστικά και μπορούν να δημιουργήσουν στιβαρά μοντέλα. Ένα ακόμα πλεονέκτημα τους είναι ότι, η διαδικασία ταξινόμησης μπορεί να τροποποιηθεί εάν τεθούν νέα βάρη προπόνησης. Από την άλλη παρουσιάζουν αρκετά μειονεκτήματα, όπως η δυσκολία επιβεβαίωσης της δομής, η αποτελεσματικότητα της εκπαίδευσης, η υπερβολική εκπαίδευση κ.λπ. (Sharma and Panigrahi, 2012, Shen et al., 2007).



**Σχήμα 3.4:** Γενική δομή ενός νευρωνικού δικτύου  
(Zorounidis and Doumplos, 2002)

### 3.5.3 Bayesian Δίκτυα Πειοιθήσεων (Bayesian Belief Networks)

Το δίκτυο πειοιθήσεων Bayes (BBN) αντιπροσωπεύει ένα σύνολο τυχαίων μεταβλητών και τις υπό όρους εξαρτήσεις τους χρησιμοποιώντας ένα κατευθυνόμενο άκυκλο γράφημα (Directed Acyclic Graph-DAG). Στο γράφημα αυτό οι κόμβοι αντιπροσωπεύουν τυχαίες μεταβλητές και τα τόξα κωδικοποιούν την εξάρτηση υπό όρους μεταξύ των μεταβλητών (Sharma and Panigrahi, 2012).

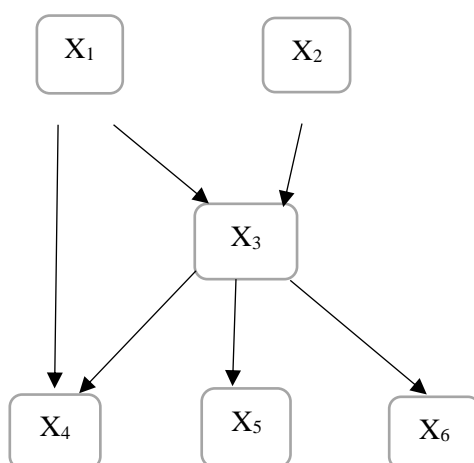
Τα Bayesian δίκτυα πειοιθήσεων χρησιμοποιούν την ταξινόμηση κατά Bayes η οποία βασίζεται στο στατιστικό θεώρημα του Bayes. Το θεώρημα του Bayes παρέχει έναν υπολογισμό για την μεταγενέστερη πιθανότητα. Σύμφωνα με το θεώρημα αν μια υπόθεση  $H$  όπως για παράδειγμα το αντικείμενο  $X$  ανήκει στην κατηγορία  $C$ , τότε η πιθανότητα που έχει η υπόθεση  $H$  είναι:

$$P(H / x) = \frac{(P(x|H) \cdot P(H))}{P(x)} \quad (3.14)$$

Σε περίπτωση που υπάρχουν εναλλακτικές κατηγορίες και το αντικείμενο  $X$  ανήκει σε μια από αυτές τις κατηγορίες  $i$ , τότε ο ταξινομητής Bayes υπολογίζει τις πιθανότητες  $P(C_i/X)$  για όλες τις πιθανές κατηγορίες  $C_i$  και εισάγει το αντικείμενο στην κλάση με την μέγιστη πιθανότητα (Kirkos et al., 2007).

Ένα δίκτυο πεποιθήσεων, πέρα από ένα κατευθυνόμενο άκυκλο γράφημα, ορίζεται και από ένα σύνολο πινάκων πιθανοτήτων υπό όρους. Εάν από ένα κόμβο  $X$  τραβήξουμε ένα τόξο σε έναν κόμβο  $Y$ , τότε ο  $X$  είναι γονικός ή άμεσος προκάτοχος του  $Y$  και ο  $Y$  είναι απόγονος του  $X$ . Μέσο του πίνακα πιθανοτήτων υπό όρους, για κάθε κόμβο, καθορίζεται η πιθανότητα υπό όρους του κόμβου για κάθε πιθανό συνδυασμό των τιμών των προκατόχων του (γονέων του, Han and Kamber, 2006, Kirkos et al., 2007).

Στο Σχήμα 3.5 παρουσιάζεται ένα Bayesian δίκτυο πεποιθήσεων που δείχνει την αιτιώδη σχέση μεταξύ των μεταβλητών. Τα βέλη (τόξα) που λείπουν επιδεικνύουν ότι δυο μεταβλητές είναι ανεξάρτητες ή μια από την άλλη (West and Bhattacharya 2016).



**Σχήμα 3.5:** Παράδειγμα γραφικής αναπαράστασης ενός Μπεϋζιανού δικτύου πεποιθήσεων. (West and Bhattacharya 2016)

### 3.5.4 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines-SVMs)

Οι μηχανές διανυσμάτων υποστήριξης (SVM) χρησιμοποιούν τεχνικές ταξινόμησης, οι οποίες βασίζονται στη στατιστική θεωρία μάθησης. Τα SVM παράγουν ένα δυαδικό ταξινομητή, ο οποίος αποκαλείται «βέλτιστο διαχωριστικό υπερεπίπεδο», μέσω μιας εξαιρετικά μη γραμμικής χαρτογράφησης των διανυσμάτων εισόδου σε ένα χώρο χαρακτηριστικών υψηλών διαστάσεων.

Τα SVM κατασκευάζουν ένα γραμμικό μοντέλο για να εκτιμήσουν μια συνάρτηση απόφασης χρησιμοποιώντας μη γραμμικά όρια κλάσεων με βάση διανύσματα υποστήριξης. Εάν τα δεδομένα είναι γραμμικά διαχωρισμένα, τα SVM εκπαιδεύουν γραμμικές μηχανές για ένα βέλτιστο υπερεπίπεδο που διαχωρίζει τα δεδομένα χωρίς σφάλματα και στη μέγιστη απόσταση μεταξύ του υπερεπίπεδου και των πλησιέστερων σημείων εκπαίδευσης. Τα σημεία εκπαίδευσης που είναι πιο κοντά στο βέλτιστο διαχωριστικό υπερεπίπεδο ονομάζονται διανύσματα υποστήριξης (Pai et al., 2011).

Μετατρέποντας ένα γραμμικό ζήτημα σε ένα χώρο χαρακτηριστικών υψηλότερων διαστάσεων, έχουν τη δυνατότητα να επιλύσουν περίπλοκα, μη γραμμικά προβλήματα όπως η ανίχνευση χρηματοοικονομικής απάτης με γραμμική ταξινόμηση χωρίς να αυξήσουν την υπολογιστική πολυπλοκότητα. Η συνάρτηση που χρησιμοποιείται για τον μετασχηματισμό του συνόλου δεδομένων ονομάζεται συνάρτηση πυρήνα, η οποία μπορεί να θεωρηθεί ως αντιστοίχιση σημείων μεταξύ του χώρου εισόδου και ενός χώρου υψηλότερων διαστάσεων. Η συνάρτηση πυρήνα ορίζεται από (West and Bhattacharya 2016):

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (3.15)$$

Όπου το  $\Phi: X \rightarrow H$ , δείχνει την αντιστοίχιση των σημείων εισόδου  $X$  στο χώρο υψηλότερης διάστασης  $H$ . Μετά την εφαρμογή της συνάρτησης πυρήνα στο σύνολο δεδομένων χρησιμοποιείται ένα υπερεπίπεδο για το διαχωρισμό των κλάσεων, το οποίο έχει τη μορφή:

$$\langle w, x \rangle + b = 0 \quad (3.16)$$

Το υπερεπίπεδο είναι κατασκευασμένο με τέτοιο τρόπο ώστε να μεγιστοποιείται ο διαχωρισμός μεταξύ των δύο κατηγοριών, γεγονός που βοηθά στη μείωση των πιθανών σφαλμάτων που προκαλούνται από την υπερβολική εκπαίδευση. Η ταξινόμηση για μια μηχανή φορέα υποστήριξης μπορεί επομένως να οριστεί ως εξής:

$$\sum_i a_i y_i k(x_i, x) + b = 0 \quad (3.17)$$

Υπάρχουν διάφορες συναρτήσεις πυρήνα που χρησιμοποιούνται όπως η συνάρτηση του Gauss και η πολωνυμική συνάρτηση, ωστόσο η επιλογή της συνάρτησης που θα χρησιμοποιηθεί εξαρτάται από το σύνολο δεδομένων και τις απαιτήσεις της ταξινόμησης (West and Bhattacharya 2016).

### 3.5.5 Αλγόριθμος των K-πλησιέστερων Γειτόνων (K - Nearest Neighbour Method)

Οι K-πλησιέστεροι γείτονες (K-Nearest Neighbour, KNN) είναι μία μη παραμετρική μέθοδος ταξινόμησης. Η δομή του μοντέλου εξαρτάται από το σύνολο των δεδομένων, και χωρίς να υπάρχει καμία υπόθεση για κατανομή δεδομένων. Η ιδιότητα αυτή είναι χρήσιμη διότι τα σύνολα δεδομένων πραγματικού κόσμου δεν τηρούν πάντα τις θεωρητικές παραδοχές (Wang et al., 2020).

Οι ταξινομητές πλησιέστερου γείτονα βασίζονται στην εκμάθηση κατ' αναλογία, δηλαδή, συγκρίνοντας μια δεδομένη πλειάδα δοκιμής με πλειάδες εκπαίδευσης που είναι παρόμοιες με αυτήν. Οι πλειάδες εκπαίδευσης περιγράφονται με  $n$  χαρακτηριστικά. Κάθε πλειάδα αντιπροσωπεύει ένα σημείο σε ένα  $n$ -διάστατο χώρο. Με αυτόν τον τρόπο, όλες οι πλειάδες προπόνησης αποθηκεύονται σε ένα  $n$ -διάστατο

χώρο σχεδίων. Όταν δίνεται μια άγνωστη πλειάδα, ένας ταξινομητής k-πλησιέστερου γείτονα αναζητά το χώρο του σχεδίου για τις k πλειάδες εκπαίδευσης που είναι πιο κοντά στην άγνωστη πλειάδα. Αυτές οι k πλειάδες εκπαίδευσης είναι οι k «πλησιέστεροι γείτονες» της άγνωστης πλειάδας (Han and Kamber, 2006).

Η ταξινόμηση επιτυγχάνεται μέσω του καθορισμού μιας μετρικής απόστασης. Η πιο διαδεδομένη μετρική απόσταση είναι η Ευκλείδεια απόσταση. Έστω δύο σημεία ή πλειάδες,  $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$  και  $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ , η Ευκλείδεια απόσταση αυτών δίνεται από τον τύπο:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (3.18)$$

Με άλλα λόγια για κάθε αριθμητικό χαρακτηριστικό, παίρνουμε την διαφορά μεταξύ των αντίστοιχων τιμών αυτού του χαρακτηριστικού στην πλειάδα  $X_1$  και στην πλειάδα  $X_2$ , την τετραγωνίζουμε και τη συσσωρεύουμε. Η τετραγωνική ρίζα λαμβάνεται από το συνολικό πλήθος των συσσωρευμένων αποστάσεων (Han and Kamber, 2006).

Τα πλεονεκτήματα της μεθόδου είναι ότι η φάση εκπαίδευσης ενός K-πλησιέστερου γείτονα, είναι πολύ πιο γρήγορη από άλλους ταξινομητές και είναι αρκετά απλή, καθώς δεν περιέχει υποθέσεις. Η μέθοδος επίσης, μπορεί να είναι χρήσιμη σε περίπτωση διεξαγωγής προβλήματος πολλαπλών κλάσεων. Η αδυναμία αυτής της μεθόδου είναι ότι η φάση δοκιμής της ταξινόμησης K-πλησιέστερου γείτονα είναι πιο δαπανηρή και πιο αργή, καθώς απαιτεί μεγάλη μνήμη για την αποθήκευση ολόκληρων συνόλων δεδομένων εκπαίδευσης. Επιπλέον, ο KNN είναι ευαίσθητος σε μεγέθη και ακραίες τιμές επειδή επιλέγει άμεσα τους γείτονες με βάση την απόσταση (Wang et al., 2020).

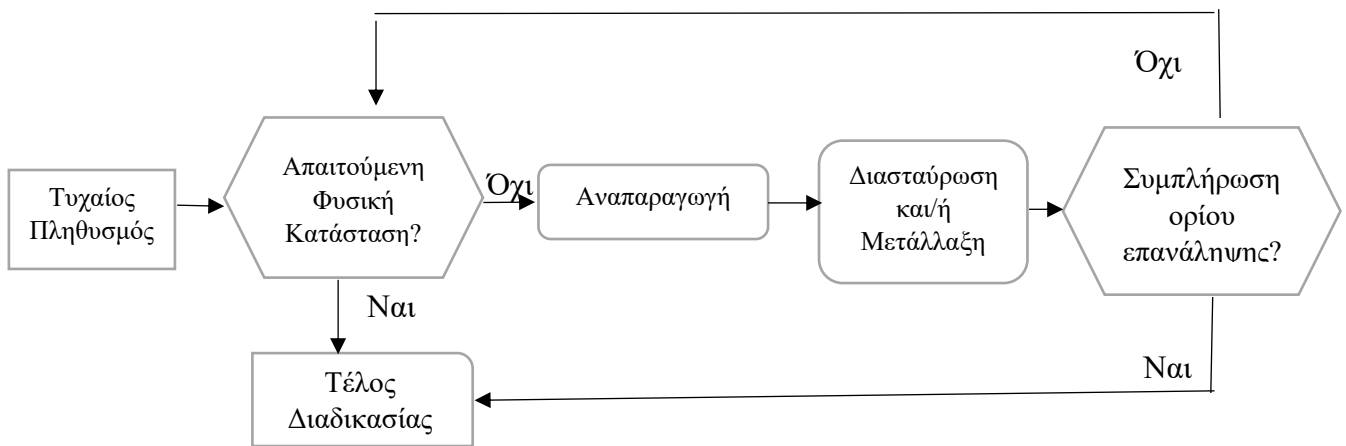
### 3.5.6 Γενετικοί Αλγόριθμοι (Genetic Algorithms - GA)

Οι γενετικοί αλγόριθμοι είναι μια τεχνική βελτιστοποίησης που βασίζεται σε μοντέλα φυσικής επιλογής και εξέλιξης. Το σημείο εκκίνησης όταν ασχολούμαστε με γενετικούς αλγόριθμους είναι ο αρχικός πληθυσμός. Ο πληθυσμός αποτελείται από έναν αριθμό χρωμοσωμάτων (άτομα) που το καθένα αντιπροσωπεύει μια λύση στο πρόβλημα. Κάθε χρωμόσωμα περιλαμβάνει έναν αριθμό γονιδίων που τυπικά κωδικοποιούνται ως δυαδικοί αριθμοί. Το βέλτιστο μέγεθος του πληθυσμού εξαρτάται από την πολυπλοκότητα του προβλήματος που πρέπει να επιλυθεί. Γενικά, ένας πολύ μικρός πληθυσμός μπορεί να οδηγήσει σε κακές λύσεις, ενώ ένας πολύ μεγάλος πληθυσμός θα σπαταλήσει περιττούς υπολογιστικούς πόρους (Höglund, 2017).

Λειτουργούν δημιουργώντας μια τυχαία αρχική γενιά και στη συνέχεια αναπαράγουν κάθε επόμενη γενιά μέσω διάφορων τεχνικών. Η αναπαραγωγή πραγματοποιείται με τη λήψη ζευγαριών χρωμοσωμάτων (γονέων) από την τρέχουσα γενιά και την εφαρμογή διασταύρωσης σε δύο σημεία. Ωστόσο μερικές φορές, μπορεί να συμβούν ορισμένες μεταλλάξεις στα στοιχεία των απογόνων, που με τη σειρά τους

αυξάνουν την ποικιλομορφία στον πληθυσμό. Η ικανότητα των απογόνων μπορεί να μετρηθεί χρησιμοποιώντας μια συνάρτηση φυσικής κατάστασης, το αποτέλεσμα της οποίας καθορίζει ποιους γονείς και απόγονοι επιλέγονται για να εκπροσωπήσουν την επόμενη γενιά. Η διαδικασία (Σχήμα 3.6) αυτή επαναλαμβάνεται μέχρι να περάσει ένας προκαθορισμένος αριθμός γενεών και να επιλεγεί η καλύτερη λύση που έχει βρεθεί μέχρι τότε (West and Bhattacharya, 2016, Duman and Ozcelik, 2011).

Οι γενετικοί αλγόριθμοι επιπλέον, είναι παρόμοιοι με τα νευρωνικά δίκτυα καθώς δεν απαιτούν προηγούμενη γνώση του τομέα του προβλήματος και μπορούν να ανιχνεύσουν τις υποκείμενες σχέσεις μεταξύ των δειγμάτων. Ωστόσο η απόδοση τους εξαρτάται σε μεγάλο βαθμό από τις παραμέτρους που βρίσκονται υπό τον έλεγχο του ερευνητή, οι οποίες απαιτούν προσαρμογές για την αντιμετώπιση του προβλήματος (Ravisankar et al, 2011, Kozeny, 2015).



**Σχήμα 3.6:** Μια τυποποιημένη διαδικασία γενετικού αλγόριθμου.

(West and Bhattacharya, 2016)

## ΚΕΦΑΛΑΙΟ 4

# ΚΑΤΑΓΡΑΦΗ ΕΦΑΡΜΟΓΩΝ ΑΠΑΤΗΣ

### 4.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται μια βιβλιογραφική ανασκόπηση εφαρμογών πολυκριτήριων, στατιστικών και τεχνικών τεχνητής νοημοσύνης για τον εντοπισμό χρηματοοικονομικής απάτης. Αρχικά καταγράφονται ορισμένες γενικές μελέτες επί του θέματος. Έπειτα ακολουθεί μια πιο εκτενής αναφορά σε ερευνητική μελέτη για τον εντοπισμό απάτης χρηματοοικονομικών καταστάσεων. Τέλος καταγράφεται μια ερευνητική μελέτη που αφορά τον εντοπισμό απάτης σε πιστωτικές κάρτες.

### 4.2 Γενικά

Σε αυτό το σημείο παρατίθενται κάποιες ερευνητικές μελέτες, που συναντάμε στη διεθνή βιβλιογραφία και πραγματεύονται θέματα γύρω από την χρηματοοικονομική απάτη.

Για την καταπολέμηση της χρηματοοικονομικής απάτης, οι ερευνητές και ελεγκτικοί οργανισμοί δημοσιεύουν τις αξιολογήσεις τους σε τακτική βάση. Στην προσπάθεια αυτή, χρησιμοποιούν διάφορες τεχνικές εντοπισμού απάτης, αναπτύσσοντας νέα μοντέλα ταξινόμησης ή εφαρμόζοντας τα ήδη υπάρχον.

Ο Chen (2006), πραγματοποίησε έρευνα σχετική με την ανίχνευση απάτης σε πιστωτικές κάρτες (Credit Card Fraud), μέσω ενός νέου και πιο αποτελεσματικού εργαλείου που ονομάζεται δυαδικό σύστημα διανυσμάτων υποστήριξης (BSVS). Το μοντέλο αυτό, βασίζεται στην έννοια των διανυσμάτων υποστήριξης (SV, τα οποία επιλέγονται χρησιμοποιώντας τον γενετικό αλγόριθμο (GA) για τη βελτιστοποίηση της ακρίβειας πρόβλεψης. Χρησιμοποιήθηκε για την αύξηση του πραγματικού αρνητικού (True Negative-TN) ποσοστού, για τον εντοπισμό απάτης πιστωτικών καρτών. Για την αρχική όμως λήψη, υψηλού πραγματικού αρνητικού ποσοστού εφαρμόστηκε η αυτό-οργανωτική χαρτογράφηση (SOM), μέσω της οποίας έγινε εκτίμηση του μοντέλου διανομής των δεδομένων εισόδου.

Μια προσέγγιση γενετικού αλγορίθμου για τον εντοπισμό απάτης στις οικονομικές καταστάσεις, παρουσιάστηκε από τους Hoogs et al. (2007). Η συγκεκριμένη μελέτη χρησιμοποίησε ένα δείγμα που περιλάμβανε μια κατηγορία 51 εταιρειών, που είχαν κατηγορηθεί από την Επιτροπή Κεφαλαιαγοράς (SEC) για εσφαλμένη αναγνώριση εσόδων και μια κατηγορία ομότιμων 339 εταιρειών, οι οποίες ταίριαζαν σε κλάδο και μέγεθος (έσοδα). Συγκεκριμένα εφάρμοσαν το μοντέλο για τον εντοπισμό μοτίβων σε δημόσια διαθέσιμα οικονομικά δεδομένα που είναι χαρακτηριστικά της δόλιας χρηματοοικονομικής αναφοράς. Κατά την διεξαγωγή της μελέτης χρησιμοποιήθηκαν

«Ανακοινώσεις Επιβολής Λογιστικής και Ελέγχου» (AAERs, Accounting and Auditing Enforcement Releases) που δημοσιεύθηκαν από την SEC για τον προσδιορισμό του δείγματος φερόμενων δόλιων εταιρειών.

Οι Gao et al. (2007), πραγματοποίησαν έρευνα που αφορά τη νομιμοποίηση εσόδων από παράνομες δραστηριότητες. Ο σκοπός αυτής της εργασίας ήταν να προτείνει ένα πλαίσιο για την καταπολέμηση της νομιμοποίησης εσόδων από παράνομες δραστηριότητες (AML) που βασίζεται στην εξόρυξη δεδομένων (DM). Αρχικά, τα δεδομένα υποψίας προετοιμάστηκαν χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων. Επιπλέον, οι μέθοδοι DM συγκρίθηκαν με τις παραδοσιακές τεχνικές έρευνας. Στη συνέχεια, τα σπάνια μοτίβα συναλλαγών κατηγοριοποιήθηκαν περαιτέρω ως ασυνήθιστα/ανώμαλα/ανώμαλα και ύποπτα μοτίβα των οποίων η αναγνώριση περιελάμβανε επίσης τον εντοπισμό απάτης/ακραίων στοιχείων.

Το 2008 οι Bai et al., εισήγαγαν μια στατιστική τεχνική, συγκεκριμένα το δέντρο ταξινόμησης και παλινδρόμησης (CART), για τον εντοπισμό και την πρόβλεψη των επιπτώσεων απάτης χρηματοοικονομικών καταστάσεων (Financial Fraud Statement-FFS). Ερεύνησαν τα τεχνάσματα χειραγώγησης των οικονομικών καταστάσεων, τους δείκτες FFS και τις τεχνικές ανίχνευσης FFS σε διεθνή προοπτική. Επιπλέον εξέτασαν δέκα εισηγμένες εταιρείες με γνωστή ιστορία στην απάτη χρηματοοικονομικών καταστάσεων στην Κίνα και πρότειναν βασικούς δείκτες για να δουλέψουν με το CART.

Οι Hashimzade et al. (2010) σε έρευνα σχετική με την απάτη στην φορολογία, αναλύουν τη φορολογική απάτη σε αξιώσεις για εκπτώσεις Φόρου Προστιθέμενης Αξίας (ΦΠΑ) σε εξαγόμενα προϊόντα. Οι παραγωγοί μπορούν να ζητήσουν έκπτωση ΦΠΑ στην αξία των εξαγωγών, επομένως η υπερεκτίμηση της αξίας των εξαγωγών δημιουργεί μια αδικαιολόγητη έκπτωση. Η έρευνα έδειξε ότι η ύπαρξη απάτης δεν επηρεάζει την πραγματική απόφαση παραγωγής της επιχείρησης ούτε τη φορολογική πολιτική της κυβέρνησης.

Την ίδια χρονιά οι Yang και Wei προτείνουν μια προσέγγιση ανίχνευσης νομιμοποίησης εσόδων από παράνομες δραστηριότητες που περιλαμβάνει τρία ξεχωριστά μέτρα ανίχνευσης που εφαρμόζονται ταυτόχρονα, παρέχοντας έναν ενοποιημένο δείκτη για την ελαχιστοποίηση της παραβίασης. Ο δείκτης ενσωματώνει τρία μέτρα ανίχνευσης: (1) το μέτρο απόκλισης του όγκου και της συχνότητας συναλλαγών, (2) το μέτρο ασυνήθιστων πληρωμών προς ή εισπράξεων από άτυπο εμπορικό εταίρο· και (3) τον νόμο του Benford (σύμφωνα με τον οποίο αντί σε ένα σύνολο ψηφίων από το 1 μέχρι το 9, τα ψηφία να έχουν την ίδια συχνότητα εμφάνισης, το πρώτο ψηφίο παρουσιάζει μεγαλύτερη πιθανότητα και καθώς προχωράμε προς το τελευταίο ψηφίο η πιθανότητα μικραίνει), με βάση τον αριθμό των φορών που εμφανίζεται ένα συγκεκριμένο ψηφίο σε μια συγκεκριμένη θέση, σε αριθμούς για την ανίχνευση οικονομικής απάτης. Τέλος σχεδίασαν ένα αριθμητικό τεστ, το οποίο θα έπρεπε να ικανοποιεί κάθε λογική προσέγγιση ανίχνευσης.



Στη μελέτη του Krivko (2010) παρουσιάστηκε το πλαίσιο για ένα υβριδικό μοντέλο, για συστήματα ανίχνευσης απάτης πιστωτικών καρτών. Συγκεκριμένα, η προτεινόμενη προσέγγιση αποτελούσε συνδυασμό στοιχείων εποπτευόμενων και μη εποπτευόμενων μεθοδολογιών με στόχο την αντιστάθμιση των επιμέρους ελλείψεων των μεθόδων. Επιπλέον χρησιμοποιήθηκαν τέσσερα μέτρα απόδοσης, το πρώτο από αυτά ήταν η αναλογία των νόμιμων λογαριασμών που έχουν λανθασμένα χαρακτηριστεί ως δόλιοι προς τους δόλιους λογαριασμούς που έχουν εντοπιστεί σωστά. Το μέτρο αυτό δείχνει πόσοι λογαριασμοί σε ειδοποίηση πρέπει να διερευνηθούν για να βρεθεί ένας δόλιος, χρησιμοποιείται κυρίως στην πράξη. Είναι η αναλογία των δόλιων συναλλαγών που έχουν διαφύγει τον εντοπισμό προς όλες τις συναλλαγές απάτης που έχουν πραγματοποιηθεί σε λογαριασμό. Τέλος τα άλλα δυο μέτρα που χρησιμοποιήθηκαν είναι το ποσοστό απάτης που έχει εντοπιστεί σωστά σε όλους τους λογαριασμούς που έχουν παραβιαστεί και η εξοικονόμηση πόρων που έχει εκχωρηθεί στην εντοπισμένη απάτη.

Ένα χρόνο αργότερα (2011) οι Dechow et al., στόχευσαν στην ανάπτυξη μιας ολοκληρωμένης βάσης δεδομένων οικονομικών ανακρίβειών. Στόχος τους ήταν να περιγράψουν αυτήν τη βάση δεδομένων και να την καταστήσουν ευρέως διαθέσιμη σε άλλους ερευνητές για την προώθηση της έρευνας σχετικά με τις ανακρίβειες κερδών. Επιπλέον ανέλυσαν τα οικονομικά χαρακτηριστικά των εταιρειών που παρουσίαζαν ανακρίβεια και αναπτύξαν ένα μοντέλο για την πρόβλεψη ανακρίβειών. Το αποτέλεσμα αυτής της ανάλυσης ήταν μια κλιμακούμενη πιθανότητα (βαθμολογία F), η οποία μπορεί να χρησιμοποιηθεί ως κόκκινη σημαία της πιθανότητας διαχείρισης κερδών ή ανακρίβειας. Συγκέντρωσαν τη βάση δεδομένων τους μέσω μιας λεπτομερούς εξέτασης εταιρειών που είχαν υποβληθεί σε ενέργειες επιβολής από την Επιτροπή Κεφαλαιαγοράς των ΗΠΑ (SEC) για φερόμενη ανακρίβεια των οικονομικών τους καταστάσεων.

Την ίδια χρονιά (2011), ο Perols σε σχετική ερευνητική μελέτη σχετικά με την ανίχνευση απάτης εταιρειών σε χρηματοοικονομικές καταστάσεις, σύγκρινε την απόδοση έξι δημοφιλών μοντέλων στατιστικής και μηχανικής μάθησης. Ο κύριος πειραματικός παράγοντας ενδιαφέροντος ήταν, ο αλγόριθμος ταξινόμησης. Οι αλγόριθμοι ταξινόμησης ελήφθησαν από το Weka, ένα εργαλείο εξόρυξης δεδομένων ανοιχτού κώδικα. Η χρήση εργαλείου ανοιχτού κώδικα διευκόλυνε την αναπαραγωγή και την επέκταση αυτής της μελέτης. Το Weka εφαρμόζει ένα σχετικά πλήρες σύνολο αλγορίθμων ταξινόμησης, συμπεριλαμβανομένων πολλών από τους πιο δημοφιλείς. Σύμφωνα επομένως με προηγούμενη έρευνα για απάτη οικονομικών καταστάσεων και προηγούμενη έρευνα εξόρυξης δεδομένων σε τομείς με μη ισορροπημένα σύνολα δεδομένων, στη συγκεκριμένη μελέτη επιλέχθηκαν έξι αλγόριθμοι από το Weka: ο αλγόριθμος ταξινόμησης δέντρων αποφάσεων J48, ο αλγόριθμος self-organising maps (SOM), ο αλγόριθμος νευρωνικού δικτύου MultilayerPerceptron, ο αλγόριθμος Logistic, ο αλγόριθμος στοίβαξης (stacking) και ο αλγόριθμος bagging .

Το 2012 οι Drezewski et al., παρουσίασαν ένα σύστημα που υποστηρίζει τον αστυνομικό αναλυτή στη διαδικασία ανίχνευσης της νομιμοποίησης εσόδων από παράνομες δραστηριότητες. Τα κύρια μέρη του συστήματος είναι οι αλγόριθμοι

εισαγωγής δεδομένων και ανάλυσης, όπως ο αλγόριθμος εξόρυξης συναλλαγών και οι αλγόριθμοι εξόρυξης συχνών μοτίβων. Τα αποτελέσματα που προκύπτουν με τη χρήση αυτών των αλγορίθμων μπορούν να οπτικοποιηθούν, ώστε να μπορούν να εξερευνηθούν εύκολα από τον αστυνομικό αναλυτή. Οι συναλλαγές που εντοπίζονται μπορούν να αντιμετωπιστούν ως ύποπτες πράξεις. Τα συχνά μοτίβα που εντοπίζονται χρησιμοποιούνται κυρίως για τον προσδιορισμό των ρόλων των ύποπτων οντοτήτων.

Λίγο αργότερα το 2013, οι González και Velásquez εμφανίζουν στοιχεία που δείχνουν ότι είναι δυνατόν ο χαρακτηρισμός και ο εντοπισμός των πιθανών χρηστών ψευδών τιμολογίων σε ένα δεδομένο έτος, ανάλογα με τις πληροφορίες στην πληρωμή του φόρου τους, την ιστορική απόδοση και τα χαρακτηριστικά τους, χρησιμοποιώντας διαφορετικούς τύπους τεχνικών εξόρυξης δεδομένων. Αρχικά χρησιμοποίησαν αλγόριθμους ομαδοποίησης όπως το SOM και το νευρικό αέριο (neural gas) για τον εντοπισμό ομάδων παρόμοιας συμπεριφοράς στον κόσμο των φορολογουμένων. Στη συνέχεια χρησιμοποιήθηκαν τα δέντρα αποφάσεων, τα νευρωνικά δίκτυα και τα δίκτυα Bayes για τον εντοπισμό των μεταβλητών που σχετίζονται με τη διεξαγωγή απάτης ή/και την απουσία απάτης, τον εντοπισμό προτύπων συσχετιζόμενης συμπεριφοράς και τον καθορισμό σε ποιο βαθμό μπορούν να εντοπιστούν περιπτώσεις απάτης ή/και μη απάτης.

Το 2015, οι Mahmoudi και Duman διερεύνησαν μια γραμμική διάκριση που ονομάζεται διακριτική συνάρτηση Fisher για πρώτη φορά στο πρόβλημα ανίχνευσης απάτης πιστωτικών καρτών. Ο γραμμικός ταξινομητής ή Fisher Linear Discriminant χρησιμοποιεί τη μέθοδο μείωσης διαστάσεων για να βρει το καλύτερο (D-1) υπερεπίπεδο διαστάσεων που μπορεί να διαιρέσει ένα D-διάστατο χώρο, σε δυο ή περισσότερους υποχώρους. Πρόκειται για μια κλασική και δημοφιλής εποπτευόμενη μέθοδος εκμάθησης που χρησιμοποιείται συνήθως στην αναγνώριση προσώπου, αναγνώριση ομιλίας/μουσικής και εξαγωγή χαρακτηριστικών με ορισμένες τροποποιήσεις.

Οι Carneiro et al. (2017) πραγματοποίησαν έρευνα που αφορά τις απάτες πιστωτικών καρτών σε ηλεκτρονικό λιανικό εμπόριο (e-tail). Συγκεκριμένα ανέπτυξαν και εφάρμοσαν ένα σύστημα ανίχνευσης απάτης σε μεγάλο έμπορο ηλεκτρονικού ταχυδρομείου, όπου για λόγους εμπιστευτικότητας δεν αποκάλυψαν το όνομα της εταιρείας. Το σύστημα αυτό διερεύνησε τον συνδυασμό χειροκίνητης και αυτόματης ταξινόμησης, με βάση πραγματικά δεδομένα από τη μελέτη περίπτωσης της εταιρείας. Το τελικό σύστημα αξιολογήθηκε και συγκρίθηκε με την τότε υπάρχουσα πρακτική. Η δουλειά τους συνέσφερε στην βιβλιογραφία με διάφορους τρόπους, παρείχε μελέτη περίπτωσης από το e-tail (ηλεκτρονικό λιανικό εμπόριο), το οποίο είχε σημαντικές διαφορές σε σύγκριση με τον καλά μελετημένο τραπεζικό τομέα, διερεύνησε το συνδυασμό ενός αυτόματου ταξινομητή με τη μη αυτόματη αναθεώρηση, παρείχε πρακτική προοπτική για την πλήρη διαδικασία ανάπτυξης μιας τέτοιας λύσης, παρουσίασε τη σύγκριση διαφορετικών μεθόδων μάθησης υπό επίβλεψη και περιέγραψε την πρακτική ανάπτυξη σε μια εταιρεία e-tail.

Την ίδια χρονιά (2017), οι Hajek και Henriques εξέτασαν το ενδεχόμενο ανάπτυξης ενός βελτιωμένου συστήματος ανίχνευσης χρηματοοικονομικής απάτης συνδυάζοντας συγκεκριμένα χαρακτηριστικά που προέρχονται από οικονομικές πληροφορίες και διαχειριστικά σχόλια στις εταιρικές ετήσιες εκθέσεις. Για την ανάπτυξη αυτού του συστήματος, χρησιμοποιήθηκε τόσο η έξυπνη επιλογή χαρακτηριστικών όσο και η ταξινόμηση, χρησιμοποιώντας ένα ευρύ φάσμα μεθόδων μηχανικής εκμάθησης. Ωστόσο, τα περισσότερα από τα συστήματα ανίχνευσης απάτης που είχαν μελετήσει, αναπτύχθηκαν μεγιστοποιώντας μόνο την ακρίβεια της πρόβλεψης, αγνοώντας την ερμηνευτικότητά τους. Αυτή η πτυχή είχε ιδιαίτερη σημασία, δεδομένου ότι η ανάπτυξη διάφανων μοντέλων καταστάθηκε ζωτικής σημασίας ιδιαίτερα μετά την οικονομική κρίση. Συνεπώς, παρείχαν δύο ερμηνεύσιμα μοντέλα, που βασίζονται στη μέθοδο Naïve Bayes. Ένα BBN (Bayesian Belief Networks) με τιμές "πράσινη σημαία" και "κόκκινη σημαία", το οποίο μπορούσε να χρησιμοποιηθεί για τον υπολογισμό της πιθανότητας απάτης στις οικονομικές καταστάσεις, ενώ ένας πίνακας αποφάσεων/υβριδικό μοντέλο Naïve Bayes πρόσφερε ένα σύνολο κανόνων απόφασης για τον εντοπισμό δόλιων/μη δόλιων επιχειρήσεων.

### 4.3 Μελέτη περίπτωσης- MetaFraud

Οι Abbasi et al. (2012) σε ερευνητική έρευνα, υπό το πρίσμα της ανάγκης για πιο ισχυρές μεθόδους αναγνώρισης, χρησιμοποίησαν μια προσέγγιση της επιστήμης του σχεδιασμού για να αναπτύξουν το MetaFraud. Το MetaFraud είναι ένα πλαίσιο μετα-μάθησης για βελτιωμένη ανίχνευση χρηματοοικονομικής απάτης.

Η μετα-μάθηση είναι μια εξειδικευμένη μορφή μηχανικής μάθησης που χρησιμοποιεί την τεχνογνωσία που αποκτήθηκε μέσω μηχανικής μάθησης ή διεργασιών εξόρυξης δεδομένων για να αυξήσει την ποιότητα των αποτελεσμάτων. Η μηχανική μάθηση παρέχει πλήθος αλγορίθμων για την ολοκλήρωση μιας εργασίας χωρίς να προσφέρει καθοδήγηση σχετικά με το ποιοι αλγόριθμοι να χρησιμοποιηθούν σε ένα δεδομένο πλαίσιο, αντίθετα η μετα-μάθηση παρέχει έναν τρόπο μάθησης, για την ίδια τη διαδικασία μάθησης ώστε να αποκτηθεί γνώση σχετικά με τα υποκείμενα χαρακτηριστικά και συνεπώς οι αλγόριθμοι μπορούν να εφαρμοστούν πιο αποτελεσματικά.

Για την αξιολόγηση του MetaFraud, διεξάχθηκαν μια σειρά πειραμάτων που περιλάμβαναν χιλιάδες νόμιμες και μη νόμιμες εταιρείες. Το πλαίσιο κατέδειξε ότι οι τεχνικές BI (Business Intelligence) που βασίζονται στη μετα-μάθηση μπορούν να ενσωματωθούν μέσω ενός τεχνουργήματος της επιστήμης σχεδιασμού και να πετύχουν σημαντικά υψηλότερη απόδοση στην ανίχνευση χρηματοοικονομικής απάτης από προηγούμενες έρευνες.

Η επιστήμη του σχεδιασμού είναι ένα ισχυρό παράδειγμα που παρέχει συγκεκριμένες προδιαγραφές για την ανάπτυξη τεχνουργημάτων πληροφορικής, συμπεριλαμβανομένων δομών, μοντέλων, μεθόδων και στιγμιότυπων. Στο παράδειγμα της επιστήμης του σχεδιασμού, "οι μέθοδοι ορίζουν τις διαδικασίες, παρέχουν

καθοδήγηση σχετικά με τον τρόπο επίλυσης προβλημάτων, δηλαδή, τον τρόπο αναζήτησης του χώρου λύσης".

Σημαντικές πτυχές της αρχιτεκτονικής και διαδικασίας της MetaFraud περιλαμβάνουν, ισχυρή κατασκευή χαρακτηριστικών, η οποία περιλαμβάνει οργανωτικές και βιομηχανικές συναφείς πληροφορίες και τη χρήση τριμηνιαίων και ετήσιων δεδομένων, και μια μέθοδο συγχώνευσης στοιβαγμένης γενίκευσης, ημι-εποπτευόμενης μάθησης, και προσαρμοστικής/ενεργητικής μάθησης. Επιπλέον παρέχουν ένα μέτρο επιπέδου εμπιστοσύνης, που εντόπισε ένα μεγάλο υποσύνολο περιπτώσεων απάτης σε υψηλό ποσοστό νόμιμων και ανάκλησης απάτης, χρησιμοποιώντας δημόσιες πληροφορίες για διάφορα ενδιαφερόμενα μέρη.

Ένα πλούσιο σύνολο χαρακτηριστικών, πολυάριθμων μεθόδων ταξινόμησης σε επίπεδο βάσης και στοιβάς και ένας προσαρμοστικός αλγόριθμος εκμάθησης, χρησιμοποιήθηκε από το πλαίσιο MetaFraud. Ξεκινώντας με ένα σύνολο ετήσιων και τριμηνιαίων αναλογιών εκκίνησης, προέκυψαν χαρακτηριστικά σε επίπεδο βιομηχανίας και οργάνωσης που βασίζονται στο πλαίσιο για τη δημιουργία των ετήσιων και τριμηνιαίων συνόλων χαρακτηριστικών. Μέσω αυτών των συνόλων χαρακτηριστικών, στόχευαν στη βελτίωση της δηλωτικής μεροληψίας.

Το πλαίσιο μετα-μάθησης στον εντοπισμό χρηματοοικονομικής απάτης βοήθησε τους ερευνητές, εκτός από τα ερευνητικά κενά να αντιμετωπίσουν και τις σχετικές υποθέσεις:

- Ο συνδυασμός ετήσιων οικονομικών μέτρων με χαρακτηριστικά οργανωτικού πλαισίου θα έχει καλύτερη απόδοση από τη χρήση μόνο των ετήσιων οικονομικών μέτρων.
- Ο συνδυασμός ετήσιων χρηματοοικονομικών μέτρων με χαρακτηριστικά σε επίπεδο κλάδου θα έχει καλύτερη απόδοση από τη χρήση μόνο των ετήσιων οικονομικών μέτρων.
- Ο συνδυασμός ετήσιων οικονομικών μέτρων με χαρακτηριστικά σε επίπεδο κλάδου και οργανωτικού πλαισίου θα έχει καλύτερη απόδοση από τη χρήση μόνο των ετήσιων οικονομικών μέτρων.
- Ο συνδυασμός τριμηνιαίων οικονομικών μέτρων με χαρακτηριστικά οργανωτικού πλαισίου θα έχει καλύτερη απόδοση από τη χρήση μόνο των τριμηνιαίων οικονομικών μέτρων.
- Ο συνδυασμός τριμηνιαίων χρηματοοικονομικών μέτρων με χαρακτηριστικά σε επίπεδο κλάδου θα έχει καλύτερη απόδοση από τη χρήση μόνο των τριμηνιαίων χρηματοοικονομικών μέτρων.

- Ο συνδυασμός τριμηνιαίων χρηματοοικονομικών μέτρων με χαρακτηριστικά σε επίπεδο κλάδου και οργανωτικού πλαισίου θα έχει καλύτερη απόδοση από τη χρήση μόνο των τριμηνιαίων χρηματοοικονομικών μέτρων.
- Ο συνδυασμός ετήσιων και τριμηνιαίων λειτουργιών που βασίζονται σε δηλώσεις θα ξεπεράσει τη χρήση μόνο ετήσιων λειτουργιών όσον αφορά την απόδοση εντοπισμού απάτης.
- Ο συνδυασμός ετήσιων και τριμηνιαίων λειτουργιών που βασίζονται σε δηλώσεις θα ξεπεράσει τη χρήση μόνο τριμηνιαίων χαρακτηριστικών όσον αφορά την απόδοση εντοπισμού απάτης.
- Όταν χρησιμοποιούνται ετήσιες ή τριμηνιαίες λειτουργίες που βασίζονται στο περιβάλλον, οι ταξινομητές στοίβας θα έχουν καλύτερη απόδοση από μεμονωμένους ταξινομητές όσον αφορά την απόδοση εντοπισμού απάτης.
- Η χρήση ενός προσαρμοστικού μηχανισμού μάθησης, ικανού να εκπαιδευτεί εκ νέου καθώς γίνονται διαθέσιμες νέες πληροφορίες, θα ξεπεράσει το στατικό αντίστοιχό του όσον αφορά την απόδοση ανίχνευσης απάτης.
- Ένα πλαίσιο μετα-μάθησης που περιλαμβάνει κατάλληλες διατάξεις για τη βελτίωση της δηλωτικής και διαδικαστικής μεροληψίας σε συντονισμό θα ξεπεράσει τις υπάρχουσες μεθόδους όσον αφορά την απόδοση εντοπισμού απάτης.
- Ένα πλαίσιο μετα-μάθησης που περιλαμβάνει γενίκευση και προσαρμοστική μάθηση, θα παρέχει βελτιωμένη διαδικαστική μεροληψία σε σχέση με τις υπάρχουσες ημι-εποπτευόμενες μεθόδους μάθησης που βασίζονται σε σύνολο, με αποτέλεσμα βελτιωμένη απόδοση ανίχνευσης χρηματοοικονομικής απάτης.

Η μελέτη αυτή μέσω των πειραμάτων έδειξε ότι η ενσωμάτωση πληροφοριών σε επίπεδο κλάδου και οργανωτικού πλαισίου, είτε σε ετήσιο είτε σε τριμηνιαίο επίπεδο, μπορεί να βελτιώσει την απόδοση σε σχέση με τη χρήση απλών αναλογιών βάσει ετήσιων δηλώσεων χωρίς πληροφορίες περιβάλλοντος. Επιπλέον έδειξε ότι ο συνδυασμός ετήσιων και τριμηνιαίων πληροφοριών απέδωσε καλύτερα αποτελέσματα καθώς παρέχουν συμπληρωματικές πληροφορίες. Υποστήριξε επίσης την ιδέα, ότι η ικανότητα των ταξινομητών στοίβας να εκμεταλλεύονται διαφορετικούς υποκείμενους ταξινομητές, τους δίνει τη δυνατότητα να βελτιώσουν την απόδοση ταξινόμησης.

Μέσω των πειραμάτων επιπλέον, αποκαλύφθηκε ότι ο προσαρμοστικός αλγόριθμος ημι-εποπτευόμενης μάθησης που προτάθηκε, βελτίωσε περαιτέρω την απόδοση αξιοποιώντας τις υποκείμενες ταξινομήσεις με τα υψηλότερα επίπεδα εμπιστοσύνης,

καθώς και έδειξε ότι συλλογικά το προτεινόμενο πλαίσιο μετα-μάθησης ήταν σε θέση να ξεπεράσει τις σύγχρονες μεθόδους σύγκρισης.

Το πλαίσιο MetaFraud συνέβαλε στον τομέα του εντοπισμού απάτης μέσω των βαθμολογιών εμπιστοσύνης. Συγκεκριμένα, οι βαθμολογίες εμπιστοσύνης που δημιουργήθηκαν από το πλαίσιο αυτό, μπορούν να παρέχουν μια χρήσιμη βοήθεια λήψης αποφάσεων για διάφορες ομάδες ενδιαφερομένων. Διατύπωσαν επίσης ότι το πλαίσιο MetaFraud θα μπορούσε να επηρεάσει το σχεδιασμό και την ανάπτυξη συστημάτων ανίχνευσης χρηματοοικονομικής απάτης που ενσωματώνουν προγνωστικές και αναλυτικές τεχνολογίες επιχειρηματικής ευφυΐας, επιτρέποντας έτσι στους αναλυτές να βγάλουν τα δικά τους συμπεράσματα.

#### 4.4 Μελέτη περίπτωσης-Εξόρυξη Δεδομένων σε πιστωτικές κάρτες.

Οι Bhattacharyya et al. (2011) σε συγκριτική μελέτη, σχετική με τον εντοπισμό απάτης πιστωτικών καρτών μέσω εξόρυξης δεδομένων, αξιολόγησαν δύο προηγμένες προσεγγίσεις εξόρυξης δεδομένων, (1) μηχανές διανυσμάτων υποστήριξης (SVM), και (2) τυχαία δάση, μαζί με τη γνωστή λογιστική παλινδρόμηση, ως μέρος μιας προσπάθειας καλύτερου εντοπισμού.

Η μελέτη αυτή χρησιμοποίησε το σύνολο δεδομένων μιας προγενέστερης έρευνας, που πραγματοποιήθηκε από τον Paasch (2007), το οποίο ελήφθη από μια διεθνή λειτουργία πιστωτικών καρτών. Στην συγκεκριμένη έρευνα χρησιμοποιήθηκαν τεχνητά νευρωνικά δίκτυα (ANN) συντονισμένα από γενετικούς αλγόριθμους (GAs) για τον εντοπισμό απάτης. Αυτό το σύνολο δεδομένων είχε συναλλαγές από πιστωτικές κάρτες 13 μηνών, από τον Ιανουάριο του 2006 έως τον Ιανουάριο του 2007, συγκεκριμένα περίπου 50 εκατομμυρίων συναλλαγών πιστωτικών καρτών σε περίπου ένα εκατομμύριο πιστωτικών καρτών από μία μόνο χώρα.

Για τους σκοπούς της μελέτης ο Bhattacharyya et al. (2011), ονόμασαν το παραπάνω σύνολο όλων των συναλλαγών, σύνολο δεδομένων  $U$ . Έπειτα επέλεξαν ως σύνολο δεδομένων  $A$ , ένα πολύ μικρότερο υποσύνολο του μεγάλου συνόλου δεδομένων, το οποίο περιείχε 2420 γνωστές δόλιες συναλλαγές με 506 πιστωτικές κάρτες. Παρατήρησαν ότι στο σύνολο  $A$ , σχεδόν όλες οι συναλλαγές απάτης αφορούσαν λιανική αγορά σε σύγκριση με το σύνολο  $U$  που περίπου οι μισές συναλλαγές απάτης αφορούσαν λιανική αγορά και οι υπόλοιπες αφορούσαν μη κατευθυνόμενες πληρωμές και επιταγές. Συνεπώς διαχώρισαν το σύνολο δεδομένων  $U$  για να συμπεριλάβουν μόνο τους τύπους συναλλαγών που βρέθηκαν στο σύνολο δεδομένων απάτης  $A$ , το μειωμένο αυτό σύνολο το ονόμασαν  $V$ .

Για την σύγκριση της πρόβλεψης απάτης πιστωτικών καρτών με την χρήση διαφορετικών τεχνικών, ήταν αναγκαία η εύρεση συνόλων πέρα από γνωστών συναλλαγών απάτης και μη ανιχνευόμενων συναλλαγών ή παρατηρούμενων νόμιμων συναλλαγών. Το σύνολο δεδομένων  $A$  είχε περιπτώσεις γνωστών συναλλαγών απάτης,

και από το σύνολο V αποφάσισαν να δημιουργήσουν ένα τυχαίο δείγμα υποτιθέμενων νόμιμων συναλλαγών, εξαιρουμένων όλων των συναλλαγών από το σύνολο δεδομένων A. Επομένως χρησιμοποιώντας τα δύο σύνολα A και V, δημιούργησαν αρχικά το σύνολο δεδομένων B, το οποίο είχε όλες τις συναλλαγές των προηγούμενων συνόλων. Το σύνολο δεδομένων B αποτελούνταν από 37.280 συναλλαγές, εκ των οποίων 2420 ήταν οι συναλλαγές του συνόλου δεδομένων A, που αποτελούσαν τις γνωστές συναλλαγές απάτης.

Στη συνέχεια, δημιούργησαν το σύνολο δεδομένων C, ένα τυχαίο δείγμα παρατηρούμενων νόμιμων συναλλαγών από το σύνολο δεδομένων V μείον τις συναλλαγές του συνόλου δεδομένων B. Το σύνολο δεδομένων C, στο οποίο παρατηρήθηκαν νόμιμες συναλλαγές, αποτελούνταν από 9645 πιστωτικές κάρτες, όπου επιλέχθηκαν με τυχαίο τρόπο. Συνολικά το σύνολο δεδομένων C αποτελούνταν με 340.589 συναλλαγές με πιστωτική κάρτα.

Τα κύρια χαρακτηριστικά των συναλλαγών με πιστωτική κάρτα που απάρτιζαν τα παραπάνω σύνολα, αποτελούνταν από τα χαρακτηριστικά «ημερομηνίας καταχώρησης», «αριθμός λογαριασμού» (ο 16ψηφιος αριθμός πιστωτικής κάρτας), «τύπος συναλλαγής» (προκαταβολή μετρητών, αγορά λιανικής κτλ.), «νόμισμα», «κωδικός εμπόρου» (ο κωδικός κατηγορίας για τον έμπορο για μια δεδομένη συναλλαγή), «συναλλαγή σε ξένο νόμισμα», «ημερομηνία συναλλαγής», «ονόματος εμπόρου», «ονόματος πόλης», «ονόματος χώρας» (τα τρία αυτά χαρακτηριστικά περιέγραφαν τους έμπορους των αντίστοιχων συναλλαγών), «ο κωδικός αναφοράς» (είναι ένας μοναδικός κωδικός για κάθε συναλλαγή). Επιπλέον μια δυαδική μεταβλητή, η σημαία ηλεκτρονικού εμπορίου εξέταζε αν η συναλλαγή ήταν ηλεκτρονικού εμπορίου.

Ο στόχος αυτής της μελέτης, όπως προαναφέρθηκε ήταν η εξέταση της απόδοσης δυο προηγμένων τεχνικών εξόρυξης δεδομένων, των τυχαίων δασών και των μηχανών υποστήριξης δεδομένων, μαζί με τη γνωστή λογιστική παλινδρόμηση. Ήθελαν επίσης να συγκρίνουν την επίδραση της έκτασης της δειγματοληψίας δεδομένων στην απόδοση αυτών των τεχνικών. Για το SVM, χρησιμοποίησαν τη συνάρτηση ακτινικής βάσης Gauss ως συνάρτηση πυρήνα, πρόκειται για έναν πυρήνα γενικής χρήσης με καλά αποτελέσματα απόδοσης. Για τα τυχαία δάση, όρισαν τον αριθμό των χαρακτηριστικών που λαμβάνονται υπόψη σε έναν κόμβο,  $b=\sqrt{B}$ , όπου B είναι τα συνολικά χαρακτηριστικά στα δεδομένα και τον αριθμό δέντρων  $T=200$ . Στη μελέτη, για τα τυχαία δάση, πραγματοποιήθηκε και χρήση της μεθόδου με αριθμό δέντρων  $T=500$ , όπου παρατηρήθηκε ελαφρώς καλύτερη απόδοση αλλά με μεγαλύτερο υπολογιστικό κόστος.

Χρησιμοποίησαν τυχαία υπό-δειγματοληψία για να αποκτήσουν σύνολα δεδομένων εκπαίδευσης με ποικίλες αναλογίες περιπτώσεων απάτης. Εξέτασαν την απόδοση των διαφορετικών αλγορίθμων σε τέσσερα σύνολα δεδομένων εκπαίδευσης με 15%, 10%, 5% και 2% συναλλαγών απάτης. Συγκεκριμένα διέσπασαν το σύνολο δεδομένων A σε δύο υποσύνολα 1237 (51%) και 1183 (49%) συναλλαγών. Το πρώτο σύνολο 1237

συναλλαγών απάτης, το χρησιμοποίησαν για τη συμπλήρωση των τεσσάρων συνόλων δεδομένων μοντελοποίησης με συναλλαγές απάτης και ομοίως το δεύτερο σύνολο 1183 συναλλαγών για τη συμπλήρωση του δοκιμαστικού συνόλου δεδομένων. Στην συνέχεια πραγματοποίησαν δειγματοληψία νόμιμων συναλλαγών από το σύνολο δεδομένων C για να δημιουργήσουν διαφορετικά ποσοστά απάτης στα σύνολα δεδομένων μοντελοποίησης και δοκιμής. Με άλλα λόγια, διατήρησαν τον ίδιο αριθμό συναλλαγών απάτης στα τέσσερα σύνολα δεδομένων μοντελοποίησης, αλλά άλλαξαν τον αριθμό των νόμιμων συναλλαγών από το σύνολο δεδομένων C για να δημιουργήσουν διαφορετικά ποσοστά απάτης.

Οι ερευνητές χρησιμοποίησαν διάφορα μέτρα ταξινόμησης, που είναι ευρέως διαδεδομένα, την ακρίβεια, την ευαισθησία, την ειδικότητα, το μέτρο F, το G-μέσο και την σταθμισμένη ακρίβεια. Η συνολική ακρίβεια δεν είναι επαρκής ως δείκτης απόδοσης όταν υπάρχει σημαντική ανισορροπία κατηγορίας στα δεδομένα, καθώς μια προεπιλεγμένη πρόβλεψη όλων των περιπτώσεων στην πλειοψηφική κατηγορία θα δείξει υψηλή τιμή απόδοσης. Η ευαισθησία και η ειδικότητα μετρούν την ακρίβεια στις θετικές (απάτη) και αρνητικές (μη απάτη) περιπτώσεις. Το μέτρο F δίνει την αρμονική μέση ακρίβεια και ανάκληση και το G-μέσο δίνει τη γεωμετρική μέση ακρίβεια της απάτης και μη απάτης. Η σταθμισμένη ακρίβεια μαζί με το μέτρο F και το G-μέσο παρείχαν στην έρευνα, συνοπτικούς δείκτες απόδοσης αντισταθμίσεων μεταξύ αληθινών θετικών και αληθινών αρνητικών. Λαμβάνουν επίσης υπόψη το μέτρο απόδοσης AUC, το οποίο συχνά θεωρείται ως καλύτερο μέτρο της συνολικής απόδοσης. Το μέτρο AUC μετρά την περιοχή κάτω από την καμπύλη ROC και είναι ανεξάρτητο από συγκεκριμένες τιμές αποκοπής ταξινόμησης.

Η μελέτη διαπίστωσε ότι η μέθοδος τυχαία δάση δείχνει συνολικά καλύτερη απόδοση. Η μέθοδος λογιστικής παλινδρόμησης τα καταφέρνει καλύτερα από τις μηχανές υποστήριξης δεδομένων σε όλα τα μέτρα εκτός από την ειδικότητα, την ακρίβεια και το F. Παρατήρησαν επίσης, ότι η λογιστική παλινδρόμηση παρουσίασε ιδιαίτερα χαμηλή ακρίβεια όταν το ποσοστό απάτης στα δεδομένα εκπαίδευσης βρίσκονταν στο χαμηλότερο επίπεδο. Για το σύνολο δεδομένων εκπαίδευσης με το χαμηλότερο ποσοστό απάτης, οι μηχανές διανυσμάτων υποστήριξης ξεπερνούν την λογιστική παλινδρόμηση και έχουν συγκρίσιμη απόδοση με τα τυχαία δάση(Random Forest-RF). Αυτό το μοτίβο, με το SVM να ταιριάζει με την απόδοση του RF όταν εκπαιδεύεται με το χαμηλότερο ποσοστό περιπτώσεων απάτης στα δεδομένα, παρατηρήθηκε επίσης και για τα άλλα μέτρα.

Η απόδοση της λογιστικής παλινδρόμησης για το μέτρο AUC ήταν αξιοσημείωτη. Ενώ τα τυχαία δάση και οι μηχανές υποστήριξης διανυσμάτων παρουσίασαν μειωμένη AUC με χαμηλότερα ποσοστά απάτης στα δεδομένα εκπαίδευσης. Η λογιστική παλινδρόμηση (Logistic Regression-LR) παρατήρησαν ότι διατηρεί σταθερά καλή απόδοση. Μια σταθερά υψηλή τιμή της AUC για το LR στα σύνολα δεδομένων εκπαίδευσης υποδηλώνει ότι τα μοντέλα LR διατηρούν παρόμοια κατάταξη υποθέσεων, ανεξάρτητα από το επίπεδο υπό-δειγματοληψίας περιπτώσεων μη απάτης στα δεδομένα εκπαίδευσης.



# ΚΕΦΑΛΑΙΟ 5

## ΣΥΜΠΕΡΑΣΜΑΤΑ

Η απάτη όπως αναφέρθηκε και σε προηγούμενη ενότητα μπορεί να οδηγήσει σε μεγάλες οικονομικές επιπτώσεις, ακόμη και στην πτώχευση. Κατά διαστήματα έχουν αποκαλυφθεί μεγάλα οικονομικά σκάνδαλα, που έχουν πλήξει εταιρείες, ολόκληρους κλάδους ακόμα και οικονομίες. Η περίπτωση της Enron αποτελεί χαρακτηριστικό παράδειγμα ενός τέτοιου σκανδάλου. Η Enron συγκεκριμένα ήταν ένας κολοσσός στον κλάδο ενέργειας που εμφάνιζε ογκώδη έσοδα δισεκατομμυρίων δολαρίων, τα οποία ήταν αποτέλεσμα λογιστικής απάτης.

Μια άλλη περίπτωση αποτελεί η κατάρρευση της Lehman Brothers, η οποία πριν υποβάλλει αίτηση πτώχευσης ήταν η τέταρτη μεγαλύτερη επενδυτική τράπεζα στις Ηνωμένες Πολιτείες. Η πτώχευσή της τράπεζας αποτέλεσε την μεγαλύτερη στην ιστορία των ΗΠΑ και το γεγονός που πυροδότησε την παγκόσμια χρηματοπιστωτική κρίση.

Η εμφάνιση γεγονότων όπως της Enron και της Lehman Brothers καθώς και άλλων σκανδάλων όπως της Waste Management, της WorldCom, της AIG (American International Group) και της Bernie Madoff Scandal, οδήγησαν στην επιτακτική ανάγκη για δημιουργία συστημάτων ανίχνευσης και πρόβλεψης της απάτης.

Οι ερευνητές και οι οργανισμοί που αγωνίζονται για την ανίχνευση και πρόβλεψη της χρηματοοικονομικής απάτης, προσπαθούν συνεχώς να αναπτύσσουν νέα ή να εξελίσσουν περαιτέρω τα ήδη υπάρχων συστήματα εντοπισμού και πρόβλεψης απάτης. Διότι οι απατεώνες χάρη στην ραγδαία εξέλιξη της τεχνολογίας έχουν την δυνατότητα να ελίσσονται με μεγαλύτερη ευκολία έναντι των συστημάτων ανίχνευσης και πρόβλεψης και ορισμένες φορές να τα ξεπερνούν. Οι επιστήμες της στατιστικής, της επιχειρησιακής έρευνας, της πληροφορικής και της μηχανικής μάθησης συμβάλλουν στην προσπάθεια δημιουργίας συστημάτων εντοπισμού απάτης μέσω μεθόδων και μοντέλων ταξινόμησης.

Στην παρούσα εργασία έγινε παρουσίαση της έννοιας της απάτης, μεθόδων και μοντέλων ανίχνευσης καθώς και μια ανασκόπηση ερευνών στον τομέα της χρηματοοικονομικής απάτης. Θα μπορούσε να ειπωθεί ότι έχουν πραγματοποιηθεί αξιόλογες μελέτες στα πλαίσια της χρηματοοικονομικής απάτης, καταλήγοντας σε μοντέλα που είναι σε θέση να προβλέπουν ή να ανιχνεύουν δραστηριότητα απάτης στον χρηματοοικονομικό τομέα (πιστωτικές κάρτες, χρηματοοικονομικές καταστάσεις, ξέπλυμα χρήματος, κτλ.).

# ΒΙΒΛΙΟΓΡΑΦΙΑ

Abbasi A., Albrecht C., Vance A., and Hansen J. (2018), “MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud”, *MIS Quarterly*, Vol. 36, No. 4, pp. 1293-1327.

Altman, E.I., Avery, R., Eisenbeis, R. and Stinkey, J. (1981), “Application of Classification Techniques in Business, Banking and Finance”, *Contemporary Studies in Economic and Financial Analysis*, Vol. 3, JAI Press, Greenwich.

Baesens B., Verbeke W., and Vlasselaer V. (2015), “Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques”, *A Guide to Data Science for Fraud Detection*, Wiley J., and SAS Business Series.

Bai B. (2008), “False financial statements: Characteristics of China’s Listed Companies and Cart detecting approach”, *International Journal of Information Technology and Decision Making*, Vol. 7, No. 2 (2008) 339–359.

Bartschat A., Reischl M., and Mikut R. (2017), “Data mining tools”, *WIREs Data Mining Knowl Discov*.

Batane, T. (2010), “Turning to Turnitin to Fight Plagiarism among University Students”. *Educational Technology & Society*, 13 (2),1–12.

Bhattacharyya S., Jha S., Tharakunnel K., and Westland C. (2011) “Data mining for credit card fraud: A comparative study”, *Decision Support Systems* 50, 602–613.

Bolton R.J and Hand D.J. (2001), “Unsupervised Profiling Methods for Fraud Detection”. *Credit Scoring and Credit Control VII*.

Bolton R.J. and Hand D.J. (2002) “Statistical fraud detection: A review”. *Statistical Science* 17 (3), 235-255.

Carneiro N., Figueira G., and Costa M. (2017), “A data mining-based system for credit-card fraud detection in e-tail”, *Decision Support Systems* 95, 91–101.

Chan P.K., Fan W., Prodromidis A.L, and Stolfo J. (1999), “Distributed Data Mining in Credit Card Fraud Detection”, *Florida Institute of Technology and Columbia University*, 67-74.

Cressey, D. R. (1953), “Other People’s Money, A Study of the Social Psychology of Embezzlement”, New York: Free Press.

Chen R.C. (2006), “A new binary support vector system for increasing detection rate of credit card fraud”, *International Journal of Pattern Recognition and Artificial Intelligence* Vol. 20, No. 2, 227–239.

Coenen F. (2011), “Data mining: past, present and future”, *The Knowledge Engineering Review*, Vol. 26:1, 25–29.

Dechow P.M., Ge W., Larson C.R., and Sloan R.G. (2011), “Predicting Material Accounting Misstatements”, *Contemporary Accounting Research*, 28: 17–82.

Drezewski R., Sepielak J., and Filipkowi W. (2012), “System supporting money laundering detection”, *Digital Investigation* 9, 8–21.

Duman E., and Ozcelik M. H. (2011), “Detecting credit card fraud by genetic algorithm and scatter search”, *Expert Systems with Applications* 38, 13057–13063.

Dunham M. *Data mining introduction and advanced topics*, New Jersey, Prentice Hall, 2003.

Elkan C. (2001), “Magical thinking in data mining: Lessons from CoIL Challenge”, department of computer science and engineering 0114 University of California, San Diego La Jolla, California 92093-0114.

Fayyad U., Shapiro G.P., and Smyth P. (1996), “Knowledge Discovery and Data Mining: Towards a Unifying Framework”. *KDD-96 Proceedings*, pp 82-88.

Forrest, S., Hofmeyr S., Somayaji A. and Longstaff T. (1996), “A sense of self for unix processes”. *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, Los Alamitos, CA.

Frawley W., Shapiro G.P., and Matheus C. J. (1992), “Knowledge Discovery in Databases: An Overview”, *AI Magazine* Volume 13 Number 3.

Gao Y., and Han L. (2021), “Implications of Artificial Intelligence on the Objectives of Auditing Financial Statements and Ways to Achieve Them”, Elsevier B.V.

Gao Z., and Ye M. (2007), “A framework for data mining-based anti-money laundering research”, *Journal of Money Laundering Control*, Vol. 10 Iss 2 pp. 170 – 179.

González P.C, and Velásquez J.D. (2013), “Characterization and detection of taxpayers with false invoices using data mining techniques”, *Expert Systems with Applications* 40, 1427–1436.

Hajek P., and Henriques R. (2017), “Mining corporate annual reports for intelligent detection of financial statement fraud –A comparative study of machine learning methods”, *Knowledge-Based Systems* 128, 139–152.

Han, J. and Kamber, M. (2006), “Data Mining: Concepts and Techniques. 2nd Edition, Morgan Kaufmann Publishers”, San Francisco.

Hashimzade N., Huang Z., and Myles G.D. (2010), “Tax fraud by firms and optimal auditing”, *International Review of Law and Economics* 30, 10–17.

Hicham A., Jeghal A., Sabri A., and Tairi H. (2020), “A Survey on Educational Data Mining”, *IEEE Xplore*. Restrictions apply.

Höglund H. (2017), “Tax payment default prediction using genetic algorithm-based variable selection”, *Expert Systems with Applications* 88, 368–375.

Hoogs B., Kiehl T., Lacombe C., and Senturk D. (2007), “A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud”, *Intelligent System in Accounting, Finance and Management*.15,41-56.

Howard R.M. (2005), “Understanding “Internet plagiarism”, *Computers and Composition* 24, 3–15.

Hubert M., and Driessen V. (2004), “Fast and robust discriminant analysis”, *Computational Statistics & Data Analysis* 45, 301 – 320.

Juszczak P., Adams N.M., Hand D.J., Withrow C., and Weston D.J. (2008), “Off-the-peg and bespoke classifiers for fraud detection”, *Computational Statistics and Data Analysis* 52, 4521–4532.

Kirkos E., Spathis C., and Manolopoulos Y. (2007), “Data Mining techniques for the detection of fraudulent financial statements”, *Expert Systems with Applications* 32 995–1003

Kosoresow, A. P. and Hofmeyr S. A. (1997), “Intrusion Detection via System Call Traces”. *IEEE Software* 14(5), 24-42.

Kozeny V. (2015), “Genetic algorithms for credit scoring: Alternative fitness function performance comparison”, *Expert Systems with Applications* 42, 2998–3004.

Krivko M., (2010), “A hybrid model for plastic card fraud detection systems”, *Expert Systems with Applications* 37, 6070–6076.

Lin T., Ji N., and Zhang L. (2008), “A RBF Neural Network model for Anti-Money Laundering”, *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition*, Hong Kong, 30-31.

Lopez C.P., Rodriguez M.D., and Santo S.L. (2019), “Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers”, *Future Internet*, 11, 86.

Maes S., Tuyls K., Vanschoenwinkel B., and Manderick B. (2002), “Credit Card Fraud Detection Using Bayesian and Neural Networks”, *Vrije Universiteit Brussel - Department of Computer Science Computational Modeling Lab (COMO) Pleinlaan 2, B-1050 Brussel, Belgium*.

Mahmoudi N., and Duman E. (2015), “Detecting credit card fraud by Modified Fisher Discriminant Analysis”, *Expert Systems with Applications* 42, 2510–2516.

Nian K., Zhang H., Tayal A., Coleman T., and Li Y. (2016), “Auto insurance fraud detection using unsupervised spectral ranking for anomaly”, *The Journal of Finance and Data Science* 2, 58-75.

Pai P.F., Hsu M.F., and Wang M.C. (2011), “A support vector machine-based model for detecting top management fraud”, *Knowledge-Based Systems* 24, 314–321.

Pasiouras F., Gaganis C., and Doumpos M. (2007), “A multicriteria discrimination approach for the credit rating of Asian banks”, *Annals of Finance*, 351–367.

Pasiouras F., Gaganis C., and Zopounidis C. (2007), “Multicriteria decision support methodologies for auditing decisions: The case of qualified audit reports in the UK”, *European Journal of Operational Research* 180, 1317–1330.

Pathak J., Vidyarthi N., and Summers S. (2005), “A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims”. *Managerial Auditing Journal* Vol. 20 No. 6, pp. 632-644.

Perez B., Roux D., Moreno A., Villamil M. and Figuera C. (2018), “Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach”, *Applied Data Science Track Paper KDD*, 19-23.

Perols J. (2011), “Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms”, *Auditing: A Journal of Practice & Theory*, Vol. 30, No. 2, pp. 19–50

Quah J.T.S., and Sriganesh M. (2008) “Real-time credit card fraud detection using computational intelligence”, *Expert Systems with Applications* 35, 1721–1732.

Qu, D., Vetter B. M., Wang F., Narayan R., Wu S. F., Hou Y. F., Gong F. and Sargor C. (1998), “Statistical Anomaly Detection for Link-State Routing Protocols. Proceedings”, *Sixth International Conference on Network Protocols*, 62-70.

Ravisankar P., Ravi V., Rao G.R., and Bose I. (2011), “Detection of financial statement fraud and feature selection using data mining techniques”, *Decision Support Systems* 50, 491–500.

Sahin Y., Bulkan S., and Duman E. (2013), “A cost-sensitive decision tree approach for fraud detection”, *Expert Systems with Applications* 40, 5916–5923.

Seeja K. R., and Zareapoor M. (2014), “FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining”, *Hindawi Publishing Corporation, The Scientific World Journal*, Article ID 252797, 10 pages.

Sharma A, and Panigrahi P.K. (2012), “A Review of Financial Accounting Fraud Detection based on Data Mining Techniques”, *International Journal of Computer Applications* (0975 – 8887), Volume 39– No.1.

Shen A., Tong R., and Deng Y. (2007), “Application of Classification Models on Credit Card Fraud Detection”, *IEEE*. 1-4244-0885.

Spathis Ch., Doumpos M., and Zopounidis C. (2004), “Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques”, *European Accounting Review*, 11:3, 509-535,

Turban E., Delen D., and Sharda R. (2007), “Business Intelligence and Analytics: Systems for Decision Support”, *Pearson Education*.

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck M., and Baesens, B. (2015), “Gotcha! Network-based Fraud Detection for Social Security Fraud”, *Management Science*, Submitted.

Viaene S., Derring R.A., and Dedene G. (2004), “A Case Study of Applying Boosting Naïve Bayes to Claim Fraud Diagnosis”, *IEEE Transactions on Knowledge and Data Engineering*, Vol 16, NO. 5.

Viaene S., Dedene g., and Derrig R.A. (2005), “Auto claim fraud detection using Bayesian learning neural networks”, *Expert Systems with Applications* 29, 653–666.

Wang Y., Zhang Y., Lu Y., and Yu X. (2020), “A Comparative Assessment of Credit Risk Model Based on Machine Learning”, International Conference on Identification, Information and Knowledge in the Internet of Things, *Procedia Computer Science* 174, 141–149.

West J., and Bhattacharya M. (2016), “Intelligent financial fraud detection: A comprehensive review”, *Computers and Security* 57, 47-66.

Yang S., and Wei L. (2010), “Detecting money laundering using filtering techniques: a multiple-criteria index”, *Journal of Economic Policy Reform*, 13:2, 159-178.

Yeh I.C., and Lien C.H. (2009), “The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients”, *Expert Systems with Applications* 36, 2473–2480.

Zaslavsky V., and Strizhak A. (2006), “Credit Card Detection Using Self-Organizing Maps”, *An International Journal*, Vol 18, 48-63.

Zopounidis C. (1999), “Multicriteria decision aid in financial management”, *European Journal of Operational Research* 119, 404-415.

Zopounidis C., and Doumpos M. (2002), “Multicriteria Decision Aid Classification Methods”, *Kluwer Academic Publishers*.

Zopounidis C., and Doumpos M. (2001), “A preference disaggregation decision support system for financial classification problems”, *European Journal of Operational Research* 130, 402-413.