TECHNICAL UNIVERSITY OF CRETE

SCHOOL OF PRODUCTION ENGINEERING AND MANAGEMENT

IDENTIFICATION OF FRAUDULENT FINANCIAL STATEMENTS USING DATA MINING TECHNIQUES

Thesis submitted in partial accomplishment of the requirements for obtaining a Ph.D.

Maria D. Tragouda

Chania, 2024

Maria Tragouda's thesis is approved by the committee:

Examination Committee Members

Michalis Doumpos, Professor, School of Production Engineering and Management of the Technical University of Crete

Constantin Zopounidis, Professor, School of Production Engineering and Management of the Technical University of Crete

Charalambos Spathis, Professor, Department of Economics Aristotle University of Thessaloniki

Atsalakis George, Assistant Professor, School of Production Engineering and Management of the Technical University of Crete

Tsafarakis Stelios, Assistant Professor, School of Production Engineering and Management of the Technical University of Crete

Gaganis Chrysovalantis, Professor, Department of Economics of the University of Crete

Lemonakis Christos, Assistant Professor, School of Business Administration on SMEs Management at the Hellenic Mediterranean University

To my parents Vasso and Dimitris,

my sister Stella

and

my husband Dimitris

Special Thanks

At this point I would like to sincerely thank everyone who helped in implementation of this thesis.

First, I would like to thank my PhD supervisor, Professor Michalis Doumpos, without whose help it would not have been possible to prepare and accomplish this thesis. I sincerely thank him for his constant support, his valuable advice, his constant interest as well as his patience in reading and correcting it of this thesis.

Particularly important was the contribution of Professor Constantin Zopounidis, who urged me to join the team of the TUC as a PhD candidate and he was who initially believed in my potential development. Throughout my studies he generously provided me with his knowledge and guided me with it his experience.

Moreover, I would like to thank Professor Charalampos Spathis, for his advisory role on the supervisory committee.

Furthermore, I would like to thank the rest members of the examination committee who, with their guidance, helped to correct this thesis.

Finally, from the bottom of my heart I thank my family, my parents Vasso and Dimitris, my sister Stella and my husband Dimitris for the endless love, the non-stop psychological support, their understanding, and patience for all these years. Additionally, I would like to thank my friends Anestia and Giannis for their constant assistance during my studies. I hope that at some point I will be able to repay them a part of what they have done for me.

Brief Curriculum Vitae

Maria Tragouda is a graduate of Department of Mathematics of the Faculty of Science at the Aristotle University of Thessaloniki. She then received a master's degree from the University of Macedonia in Accounting and Finance majoring in Finance. She is currently a PhD candidate at the School of Production Engineering and Management of the Technical University of Crete. Her research interests focus on developing methods to detect financial occupational fraud using data mining techniques. She has published several papers in sites, a chapter of book and a journal with main subject the financial fraud detection. She is now active in the private sector, specializing in strategic procurement development and negotiation of agreements. Her goal is to use the methodology developed in this thesis to evaluate potential and existing suppliers in order to minimize the risks of the collaborations.

Abstract

Although the financial audit controls in companies have advanced over the years, the number of corporate fraud instances is growing, thus raising the need for investigating the factors that can be used as early-warning signals and developing effective systems for identifying financial fraud. In this thesis, financial statements from 133 Greek companies listed in the Athens Stock Exchange over the period 2014 to 2019 are investigated, based on the fraud diamond theory. Financial data and corporate governance variables are used as inputs to data mining techniques to develop models that can identify patterns of irregularities in a company's financial reports. To this end popular machine learning classification algorithms are employed in a novel multi-label classification setting that not only identifies fraudulent cases, but also considers the nature of the auditors' comments. The results indicate that the proposed multi-label approach provides enhanced results compared to binary classification algorithms, avoiding inconsistent outputs with respect to the existence of different forms of manipulation of financial statements.

Keywords: Falsified financial statements, Corporate financial fraud, Fraud diamond, Data mining, Multi-label classification

Περίληψη

Ο εντοπισμός παραποιημένων οικονομικών καταστάσεων είναι μια επίπονη και σημαντική διαδικασία για τη λειτουργία των επιχειρήσεων, των κυβερνήσεων και της παγκόσμιας οικονομίας. Τα τελευταία χρόνια, όλο και περισσότεροι ερευνητές εργάζονται πάνω σε αυτό αναπτύσσοντας καινοτόμα μοντέλα και διαδικασίες. Αρχικά, οι παραποιημένες οικονομικές καταστάσεις είναι οικονομικές καταστάσεις που έχουν εσκεμμένα παραποιηθεί για να εξαπατήσουν τους επενδυτές, τους πιστωτές ή άλλους ενδιαφερόμενους. Αυτό μπορεί να περιλαμβάνει σκόπιμες παραλείψεις, ανακρίβειες εσόδων ή εξόδων ή σκόπιμη εσφαλμένη ταξινόμηση πληροφοριών. Η παραποίηση των οικονομικών καταστάσεων μπορεί να χρησιμοποιηθεί για τη διάπραξη απάτης και μπορεί να οδηγήσει σε νομικές ενέργειες κατά των υπευθύνων.

Υπάρχουν διάφοροι παράγοντες που έχουν αυξήσει την ανάγκη για επιστημονική έρευνα στη διαδικασία εντοπισμού της απάτης, όπως: (1) το πολύ υψηλό κόστος που επιφέρει για τις επιχειρήσεις, τους μετόχους, αλλά και τη συνολική οικονομία μιας χώρας, (2) ο αντίκτυπος στην κοινωνία, (3) ο αθέμιτος ανταγωνισμός μεταξύ των εταιρειών, (4) η πλασματική ανάπτυξη και η στρέβλωση της συνολικής οικονομικής εικόνας που οδηγεί τους λήπτες αποφάσεων σε λάθος αποφάσεις. Επιπλέον, η αποτυχία εντοπισμού των χρηματοοικονομικών απατών οδήγησε σε χρεοκοπίες και σε επιδείνωση της εμπιστοσύνης του κοινού ακόμα και στις επαγγελματικές ελεγκτικές εταιρείες (Whiting et al., 2012).

Αυτό το είδος απάτης έχει αποδειχθεί δύσκολο να εντοπιστεί, κυρίως αφού οι απατεώνες κατέχουν συχνά κρίσιμες διοικητικές θέσεις. Πρώτον, το διοικητικό προσωπικό που εμπλέκεται σε χρηματοοικονομικό παράπτωμα, συχνά καταβάλλει μια συντονισμένη προσπάθεια για να κρύψει την απάτη των οικονομικών καταστάσεων και είναι συχνά οι κύριοι ένοχοι. Δεύτερον, επειδή τα μέλη της διοίκησης κατέχουν θέσεις μεγάλης εμπιστοσύνης, έχουν μεγαλύτερη δύναμη να παρακάμπτουν κρίσιμες εσωτερικές διασφαλίσεις, γεγονός που καθιστά πιο εύκολη τη διενέργεια και δυσκολότερη τον εντοπισμό της απάτης στις οικονομικές καταστάσεις. Τρίτον, οι περισσότερες διαδικασίες εξωτερικού ελέγχου δεν έχουν σχεδιαστεί για να ανιχνεύσουν τη συνεργασία και την πλαστογραφία που χρησιμοποιούν συχνά οι δράστες απάτης στις οικονομικές καταστάσεις για να διαπράξουν και να αποκρύψουν την απάτη.

Επιπλέον, επειδή τα δεδομένα των οικονομικών καταστάσεων είναι πολύ συμπυκνωμένα και συγκεντρωτικά, οι απάτες είναι πιο δύσκολο να εντοπιστούν χρησιμοποιώντας συμβατικές τεχνικές αναλυτικής και στατιστικής μοντελοποίησης. Ωστόσο, ορισμένες πιο πρόσφατες τεχνικές δείχνουν ότι έχουν τις δυνατότητες για να εντοπίσουν πιο πολύπλοκες ανωμαλίες που κρύβονται μέσα στις οικονομικές καταστάσεις. Η εξόρυξη δεδομένων, η οποία προσπαθεί να εξάγει χρήσιμες πληροφορίες από συλλογές μεγάλων δεδομένων, έχει χρησιμοποιηθεί εκτενώς ως εργαλείο ενεργητικής λήψης αποφάσεων. Εξαιτίας των δυνατοτήτων που δίνονται από τη χρήση τεχνικών εξόρυξης δεδομένων, οι ερευνητές ενδιαφέρονται όλο και περισσότερο για τη διαμόρφωση μεταβλητών που θα τους βοηθήσουν να βρουν πρακτικές τάσεις ικανές να εντοπίσουν τις παραποιήσεις στις οικονομικές καταστάσεις.

Αν και οι αρχές της εταιρικής διακυβέρνησης (Corporate Governance-CG) είναι πλέον καθιερωμένες και ευρέως αποδεκτές, ο αριθμός των περιπτώσεων απάτης από εργαζομένους αυξάνεται, εντείνοντας την ανάγκη διεξοδικής διερεύνησης παραγόντων που μπορούν να χρησιμοποιηθούν ως ενδείξεις έγκαιρης προειδοποίησης. Η παραποίηση των οικονομικών καταστάσεων μίας επιχείρησης είναι ένα από τα πιο σημαντικά είδη εταιρικής απάτης. Η αδυναμία των ελεγκτών να ανιχνεύουν οικονομικά σκάνδαλα καθιστά απαραίτητη τη χρήση συγκεκριμένων μεθόδων και εργαλείων ελέγχου (J. W. Lin et al., 2003). Ένας μεγάλος αριθμός ακαδημαϊκών ερευνητών (Dechow et al., 2011; Price et al., 2011) έχει προτείνει διάφορες μεθόδους για τον εντοπισμό χρηματοοικονομικής απάτης, που συχνά βασίζονται σε πληροφορίες που προέρχονται από τις οικονομικές καταστάσεις (Ngai et al., 2011).

Σήμερα, ακόμα κι αν οι έλεγχοι είναι εντατικοί, οι απατεώνες μπορούν να τους παρακάμψουν και οι ελεγκτές δεν μπορούν να εντοπίσουν έγκαιρα τις απάτες. Επομένως, είναι επιτακτική ανάγκη να προσεγγίσουμε τον εντοπισμό απατών μελετώντας τα κίνητρα, την ευκαιρία, την πίεση και την ικανότητα ενός απατεώνα να διεξάγει δόλιες ενέργειες. Αυτά τα τέσσερα χαρακτηριστικά αποτελούν τη θεωρία του διαμαντιού της απάτης (fraud diamond theory) και αυτή η διατριβή υιοθετεί αυτή τη θεωρία για την εφαρμογή μιας προσέγγισης μηχανικής μάθησης. Από την γνώση που υπάρχει έως σήμερα, είναι η πρώτη φορά στην Ελλάδα που εφαρμόζεται μια προσέγγιση της θεωρίας του διαμαντιού για τον εντοπισμό παραποιημένων οικονομικών καταστάσεων χρησιμοποιώντας συνδυασμό οικονομικών δεικτών και εταιρικών μεταβλητών.

Στη συγκεκριμένη διατριβή, χρησιμοποιούνται για την επεξεργασία και την εξαγωγή συμπερασμάτων οι ακόλουθες δυαδικές τεχνικές: τα γενικευμένα προσθετικά μοντέλα (Generalized Additive Models-GAM), ο αλγόριθμος του πλησιέστερου γείτονα (k-Nearest Neighbor - kNN), η λογιστική παλινδρόμηση (Logistic Regression-LR) και ο ταξινομητής δένδρων αποφάσεων (Random Forest-RF). Αυτές οι τεχνικές ασχολούνται με την ταξινόμηση μιας ετικέτας σε παραποιημένη ή μη περίπτωση (Falsified or non-falsified financial statements, FFS/nFFS). Επιπλέον, εξετάζεται μια προσέγγιση ταξινόμησης πολλαπλής ετικέτας με τη χρήση διχοτομικών ταξινομητών καθώς και η τεχνική πολλαπλών ετικετών γνωστή ως Multi Label Technique (ML-kNN), η οποία αποτελεί επέκταση του πλησιέστερου γείτονα.

Η ανάπτυξη μοντέλων δυαδικής ταξινόμησης για τη διάκριση μεταξύ των περιπτώσεων FFS και nFFS ήταν το κύριο επίκεντρο προηγούμενων μελετών για την αναγνώριση της ανίχνευσης παραποιημένων οικονομικών καταστάσεων. Ωστόσο, η ενισχυμένη προσέγγιση που χρησιμοποιήθηκε σε αυτή τη μελέτη υπερβαίνει το βασικό δυαδικό σχήμα για να προσδιορίσει περαιτέρω τη φύση των σχολίων των ελεγκτών. Ως αποτέλεσμα, μια παρατήρηση μπορεί να ανήκει σε περισσότερες από μία κατηγορίες και όχι μόνο σε μία κατηγορία σε ένα σχήμα ταξινόμησης πολλαπλών ετικετών και έτσι προκύπτει η κύρια καινοτομία αυτής της διατριβής.

Σε αυτό το πλαίσιο, η προσέγγιση που παρουσιάζεται σε αυτή τη μελέτη βασίζεται σε εποπτευόμενες τεχνικές εξόρυξης δεδομένων, όπως η λογιστική παλινδρόμηση, τα γενικευμένα προσθετικά μοντέλα, καθώς και οι αλγόριθμοι του πλησιέστερου γείτονα και των δένδρων αποφάσεων. Αυτές οι προσεγγίσεις εφαρμόζονται σε μια ρύθμιση ταξινόμησης για να προσδιοριστεί εάν οι εταιρείες έχουν παραποιήσει οικονομικές καταστάσεις (FFS) ή όχι (nFFS). Τέτοια σχήματα δυαδικής ταξινόμησης έχουν εξεταστεί ευρέως στη σχετική βιβλιογραφία. Ωστόσο, μια διχοτομική ταξινόμηση παραβλέπει σημαντικές πληροφορίες, οι οποίες μπορεί να έχουν πολύτιμες επιπτώσεις στην πράξη.

Μια τέτοια προσέγγιση πολλαπλών ετικετών μπορεί να είναι χρήσιμη για την ανάλυση των παραποιημένων οικονομικών καταστάσεων, καθώς παρέχει στους ελεγκτές και τις ρυθμιστικές αρχές ένα ολοκληρωμένο πλαίσιο μοντελοποίησης που όχι μόνο επιτρέπει τον εντοπισμό χρηματοοικονομικής απάτης, αλλά μπορεί επίσης να παρέχει πληροφορίες για τους τύπους απάτης που σχετίζονται με κάθε συγκεκριμένη περίπτωση.

Πιο συγκεκριμένα, η παραπάνω προσέγγιση βελτιώνει τους μηχανισμούς ελέγχου καθώς τους κάνει πιο συγκεκριμένους και βελτιώνει τον χρόνο των ελεγκτών, χωρίς να χρειάζεται να καταφεύγουν σε διαφορετικά μοντέλα για κάθε τύπο απάτης, τα οποία θα μπορούσαν να είναι δύσκολο να εφαρμοστούν, ιδιαίτερα όταν χωριστά μοντέλα παρέχουν αντικρουόμενες ενδείξεις.

Η μεθοδολογία που υιοθετήθηκε στο παραπάνω πλαίσιο πολλαπλών ετικετών, βασίζεται στο δυαδικό σχήμα συνάφειας (Binary relevance scheme, BR, Zhang *et al.*, 2018), το οποίο περιλαμβάνει την κατασκευή δυαδικών ταξινομητών για κάθε ζεύγος κλάσεων, μέσω μιας προσέγγισης ένα εναντίον όλων. Ένα κοινό πρόβλημα που προκύπτει με τους αλγόριθμους πολλαπλών κλάσεων είναι ότι συχνά αγνοούν τις σχέσεις μεταξύ των κλάσεων, οδηγώντας έτσι σε ασυνεπή και ανούσια αποτελέσματα. Για να αντιμετωπιστεί αυτό το πρόβλημα, εισάγονται δύο απλές στρατηγικές διόρθωσης -που ονομάζονται Διόρθωση ασυνέπειας βάσει περίπτωσης (Case-based Inconsistency Correction-CBIC) και σχήμα meta-learner με βάση τον πλησιέστερο γείτονα (kNN-based Meta-Learner, NNML) που διασφαλίζουν τη συνέπεια των αποτελεσμάτων ταξινόμησης με τη φύση των κατηγοριών FFS/nFFS.

Ο κύριος στόχος αυτής της διατριβής, επομένως, είναι να προτείνει μια μεθοδολογία που χρησιμοποιεί μεθοδολογίες μηχανικής μάθησης και έναν συνδυασμό οικονομικών δεικτών και εταιρικών μεταβλητών που βασίζονται στη θεωρία του διαμαντιού απάτης και να χρησιμοποιήσει αυτό τον συνδυασμό σε τεχνικές εξόρυξης δεδομένων για τον εντοπισμό των παραποιημένων οικονομικών καταστάσεων αλλά και τον προσδιορισμό του τύπου της απάτης. Τα δεδομένα εκπαίδευσης κατηγοριοποιούνται σε FFS ή nFFS λαμβάνοντας υπόψη συγκεκριμένα σχόλια των ορκωτών λογιστών που δημοσιεύονται στις οικονομικές καταστάσεις της κάθε εισηγμένης εταιρίας.

Σε αυτή τη διδακτορική διατριβή, ερευνώνται οι χρηματοοικονομικές καταστάσεις 133 ελληνικών μη χρηματοπιστωτικών εταιρειών εταιριών που είναι εισηγμένες στο Χρηματιστήριο Αθηνών τη χρονική περίοδο 2014-2019. Η ανάλυση βασίζεται σε κοινά χρησιμοποιούμενες χρηματοοικονομικές μεταβλητές καθώς και σε μη χρηματοοικονομικά χαρακτηριστικά που επιλέχθηκαν με βάση τη θεωρία του διαμαντιού απάτης, λαμβάνοντας υπόψη τη διακριτική τους δύναμη. Οι περιπτώσεις παραποιημένων χρηματοοικονομικών καταστάσεων στο δείγμα προσδιορίζονται μέσω της εξέτασης έξι τύπων σχολίων ελεγκτών, τα οποία ομαδοποιούνται σε κλάσεις μέσω μιας διαδικασίας ιεραρχικής ομαδοποίησης για να οριστούν σημαντικές κατηγορίες που αποτελούνται από διαφορετικούς τύπους FFS. Οι κατηγορίες που προκύπτουν αντιστοιχούν στις ετικέτες του προτεινόμενου πλαισίου ταξινόμησης πολλαπλών ετικετών.

Η απόδοση των εφαρμοζόμενων μεθόδων εκτιμάται με διάφορες μετρήσεις; Macro-Precision (PR), Macro-Recall (RC), Macro-F1 (F1), Hamming loss (HL) και ακρίβεια (accuracy-AC). Μια ανάλυση διασταυρούμενης επικύρωσης (cross validation - CV) 10 επαναλήψεων χρησιμοποιείται για να ληφθούν αμερόληπτες εκτιμήσεις της απόδοσης των αλγορίθμων. Όλες οι μετρήσεις απόδοσης για κάθε μέθοδο συγκρίνονται μεταξύ τους για να προσδιοριστεί η μέθοδος με τα καλύτερα αποτελέσματα. Επιπλέον, δύο μεθοδολογίες χρησιμοποιούνται στην ανάλυση για τη διόρθωση αντικρουόμενων αποτελεσμάτων που μπορεί να προκύψουν λόγω μη ταξινόμησης των εταιρειών λαμβάνοντας υπόψη τις απόψεις των ελεγκτών. Παράλληλα, συγκρίνονται και αξιολογούνται τα αποτελέσματα που προέκυψαν με τη χρήση των τεχνικών LR, kNN, GAM και RF, καθώς και με τον αλγόριθμο πολλαπλών ετικετών ML-kNN. Φυσικά, ελέγχεται και η αποτελεσματικότητα των δύο διαδικασιών διόρθωσης (CBIC, NNML).

Λαμβάνοντας υπόψη το περιεχόμενο της διατριβής, εξάγεται το συμπέρασμα ότι είναι σημαντικό να προωθηθούν οι διαδικασίες ανίχνευσης απάτης πέρα από τα παραδοσιακά εργαλεία ελέγχου και να ενισχυθούν αυτές οι διαδικασίες με την καλύτερη κατανόηση των κινήτρων και των πρακτικών των απατεώνων. Η προσέγγιση των πολλαπλών ετικετών ενίσχυσε την ακρίβεια της ταξινόμησης και οδήγησε σε καλύτερη κατανόηση της σημασίας των σχολίων των ελεγκτών. Παράλληλα, οι δύο διορθωτικές διαδικασίες (CBIC, NNML) ενίσχυσαν τις μετρήσεις απόδοσης των μοντέλων. Αν και η βιβλιογραφική ανασκόπηση και το γεγονός ότι η πλειονότητα των απατών προέρχονται από τους διοικητικούς μίας εταιρίας απαιτεί την επιτάχυνση της χρήσης νέων «ανθρωποκεντρικών» θεωριών στην ανίχνευση των παραποιημένων οικονομικών καταστάσεων, όπως η Θεωρία του Διαμαντιού της Απάτης, τα αποτελέσματα αυτής της προτεινόμενης μεθοδολογίας δεν έδωσαν σαφή εικόνα για τη συμβολή των μη χρηματοοικονομικών μεταβλητών που προέρχονται από τη θεωρία του διαμαντιού. Πιθανώς αυτό συνέβη από την επιλογή των συγκεκριμένων μεταβλητών.

Συμπερασματικά, η προτεινόμενη προσέγγιση επιδεικνύει ανώτερη προγνωστική απόδοση και επιτρέπει τον εντοπισμό πρώιμων προειδοποιητικών σημάτων (κόκκινες σημαίες red flags) για τον εντοπισμό των χρηματοοικονομικών απατών. Η μεθοδολογία αυτής της διατριβής μπορεί να προσφέρει σε επενδυτές, εμπειρογνώμονες, ρυθμιστικούς φορείς, ελεγκτές και άλλους μετόχους σημαντικές πληροφορίες, βελτιστοποιώντας την ορθότητα της λήψης των αποφάσεων και βελτιώνοντας τους χρόνους στον έλεγχο των δεδομένων. Τέλος, μπορεί να βοηθήσει τους ακαδημαϊκούς/ερευνητές να δημιουργήσουν σημαντικές επιστημονικές προτάσεις για τη βελτίωση της διαδικασίας ανίχνευσης παραποιημένων οικονομικών καταστάσεων.

Λέξεις-κλειδιά: Παραποιημένες οικονομικές καταστάσεις, Εταιρική οικονομική απάτη, Θεωρία του διαμαντιού, Εξόρυξη δεδομένων, Ταξινόμηση πολλαπλών ετικετών

Table of Contents

Introduction

The detection of falsified financial statements is a painful and important process for the operation of firms, governments, and the global economy. In recent years, more and more researchers have been working on this by developing innovative models and processes. To begin with, falsified financial statements (FFS) are financial statements that have been deliberately misstated to deceive investors, creditors, or other stakeholders. This can include intentional omissions, misstatements of revenue or expenses, or intentional misclassification of information. FFS can be used to commit fraud and can lead to legal action against the responsible parties.

There are a number of factors that have increased the need for scientific research into the process of detecting fraud, such as: (1) the very high costs it brings for firms, shareholders, but also the overall economy of a country, (2) the impact on society, (3) unhealthy competition between companies, (4) fictitious growth and distortion of the overall economic picture leading decision makers to wrong decisions. Moreover, failure to uncover financial fraud has led to bankruptcies, and a deterioration in public faith in professional audit firms (Whiting et al., 2012).

This kind of fraud has proven challenging to detect since fraudsters often hold critical management positions (ACFE, 2022). Firstly, management personnel who engage in financial misconduct, frequently make a concerted effort to hide financial statement fraud and are frequently its primary culprits. Secondly, because members of management hold positions of great trust, they have more power to bypass crucial internal safeguards, which makes it simpler to commit and more difficult to catch financial statement fraud. Thirdly, most external audit processes are not built to find the collaboration and forgery that financial statement fraud perpetrators often use to commit and conceal the fraud.

Furthermore, because financial statement data is heavily condensed and aggregated, frauds are more difficult to hide and harder to spot using conventional analytical and statistical modeling techniques. However, some more recent strategies show potential. Data mining, which tries to extract useful information from big data collections, has been extensively used as a tool for active decision-making. Due to this, academics are becoming more interested in contextualizing data to help them find fascinating and practical trends.

Although the principles of corporate governance (CG) are now well-established and widely accepted, the number of cases of occupational fraud is growing, raising the need to investigate thoroughly factors that can be used as early-warning signals. Financial fraud is among the most important types of corporate fraud. The auditors' weakness to detect financial scandals renders the usage of specific audit methods and tools necessary (J. W. Lin et al., 2003). An ample number of academic researchers (Dechow et al., 2011; Price et al., 2011) has proposed various methods to detect financial fraud, often based on information derived from financial statements (Ngai et al., 2011).

Nowadays, even if the audits are intensive, fraudsters can override them, and auditors cannot detect the frauds on time. Thus, it is imperative to approach the detection of frauds studying the motivation, the opportunity, the pressure, and the ability of a fraudster to conduct fraudulent actions. These four characteristics constitute the fraud diamond theory, and this thesis adopted this theory for the application of a machine learning approach. To the best of my knowledge, it is the first time in Greece that a fraud diamond approach was applied in detecting falsified financial statements using combination of financial ratios and corporate variables.

The main objective of this thesis is to propose a methodology using machine learning methodologies and a combination of financial ratios and corporate variables that rely on fraud diamond theory as inputs to data mining techniques for detecting FFS. The determination of whether a financial statement was falsified or not, was based on specific comments found in the auditors' reports included in each company's financial statements. Moreover, generalized additive models (GAM), the k-nearest neighbor algorithm (kNN), logistic regression (LR), and the random forest algorithm (RF) were utilized in the analysis. These methods deal with single-label classification. Additionally, a multi-label technique called the ML-kNN algorithm was considered, which extends kNN to multi-label situations.

The development of binary classification models to distinguish between FFS and nFFS (non-falsified financial statements) cases was the primary focus of earlier studies on the identification of FFS detection; however, the enhanced approach used in this study goes beyond the basic binary scheme to further identify the nature of the auditors' comments. As a result, an observation may belong to more than one class rather than just one category in a multi-label classification scheme and thus the main novelty of this thesis arises.

Within this context, the approach presented in this study relies on supervised data mining techniques such as logistic regression, generalized additive models, as well as the nearest neighbor and the random forest algorithms. These approaches were applied in a classification setting to identify whether firms have falsified financial statements (FFS) or not (nFFS). Such binary classification schemes have been widely considered in the relevant literature. However, a dichotomic classification overlooks valuable information, which can have valuable implications in practice. For instance, Kim et al. (2016) compared logistic regression, support vector machines, and Bayesian networks in a multi-class approach that distinguishes among three mutually exclusive classes of financial misstatements considering the presence of fraud intention.

Such a multi-label approach can be useful for FFS analysis, as it provides auditors and regulators with a comprehensive modeling framework that not only enables the identification of financial fraud, but it can also provide insights into the types of fraud that are relevant for each particular case, without having to resort to different models for each type of fraud, which could be difficult to implement, particularly when separate models provide conflicting indications.

The methodology adopted in the above multi-label context, is based on the binary relevance scheme (BR, Zhang et al., 2018), which involves the construction of binary classifiers for each pair of classes, through a one-against-all approach. A widespread problem that arises with multi-class algorithms is that they often ignore the relationships between the classes, thus leading to inconsistent and meaningless results. To address this problem, two simple correction strategies -called Case-based inconsistency correction (CBIC) and kNN-based meta learner (kNNML) schemes- were introduced that ensure the consistency of the classification results with the nature of the FFS/nFFS categories.

The methodology was applied to a dataset of 133 Greek non-financial companies from the Greek stock exchange during the period 2014 to 2019. The analysis was based on commonly used financial variables as well as non-financial attributes selected based on the fraud diamond theory, while having taken into consideration their discriminating power. Cases of FFS in the sample were identified through the examination of six types of auditors' comments, which were clustered into groups through a hierarchical clustering process to define meaningful classes consisting of diverse types of FFS. The resulting classes corresponded to the labels of the proposed multi-label classification framework.

The performance of the applied methods was estimated with various metrics; macro-precision (PR), macro-recall (RC), macro F1 (F1), Hamming loss (HL) and accuracy (AC). A cross-validation (CV) analysis with 10 folds (i.e., 10-fold CV) was used to get unbiased estimations of the algorithms' performance. All the performance metrics for each method were compared with each other to identify the method with the best results. Moreover, two methodologies were used in the analysis to correct conflicting outcomes that may derive due to miss classification of firms considering the auditors' opinions. In comparison to the base learners created using LR, kNN, GAM, and RF, as well as the multi-label ML-kNN algorithm, the effectiveness of the two correction schemes had evaluated.

Considering the results of the methodology, the suggested approach demonstrated superior predictive performance and enabled the detection of early warning signs (red flags) at the corporate governance level as well as the financial level to facilitate the audit procedure. The methodology of this thesis can give investors, experts, regulators, auditors, and other stockholders crucial information and may help academics to create significant scientific proposals to enhance the procedure of detecting FFS cases. Accounting professionals and their management level may want to think about investing in these techniques to stop costly frauds for their firms and meet the urgent needs of regulatory authorities and legal requirements such those.

The structure of this thesis is as follows:

The first chapter begins with the definition of financial fraud and the importance of FFS's detection. Through the presentation of some of the most popular financial scandals worldwide and in Greece, and simultaneously, the explanation of how fraudsters can conduct an occupational fraud and who may be the perpetrators, the reader can understand the whole fraudulent action and behavior. Given several examples of the cost that frauds cause in financial and social terms, the section concludes that the detection of

FFS through corporate governance's approaches such as fraud triangle and diamond theory, will provide an effective methodology.

In the second chapter, a detailed literature review is presented concerning the general discussions on occupational fraud. Moreover, there is an extent presentation of applied data mining techniques. Additionally, there is a summary with the financial ratios, corporate and linguistic variables that have been used by other academics and a discussion on findings of the literature review.

In the third chapter there is a thorough presentation of the supervised data mining techniques which were selected for this thesis. Logistic Regression (LR), k-Nearest Neighbor (KNN), Generalized Additives Methods (GAM), Random Forests (RF) methods, a multi-label classification strategy (ML-kNN) and two corrective schemes (CBIC and NNML) were all used in this study to successfully categorize the firms into the FFS or nFFS category. Additionally, all the performance metrics which were applied to this research were presented in this chapter. Moreover, the performance results for each method were compared to one another including two correction schemes, which were employed to resolve inconsistencies that may result from incorrectly classifying firms considering the auditors' findings. The performance of these two schemes was compared to the basic learners built with LR, kNN, GAM, and RF, as well as the multi-label ML-kNN technique and all the results were thoroughly examined.

Chapter four summarizes the results discussed in the previous chapters, presents the limitations of this thesis and suggests proposals for future research. These proposals will evolve the methodology presented in this thesis giving an even more comprehensive tool to those interested in using it such as auditors, lenders, stakeholders, and governance.

Chapter 1. The definition and the importance of financial fraud detection

1.1 Introduction

Among the factors that have contributed to widespread falsification of accounting statements and the so-frequent use of misleading accounting practices to meet unattainable market expectations are the fierce competition in the world of finance, a drop in business performance, and the great pressure on business executives to achieve ever-higher goals (Jennings, 2004). The worldwide business community has been harshly rattled by the exposure of several financial scandals, but in the past several years, cases of false financial statements have gotten disturbingly serious (Humpherys et al., 2011; Kamarudin et al., 2012; Kirkos et al., 2007; Yeh et al., 2010).

Financial fraud is defined as a damaging act that occurs when someone deprives another person of their money or otherwise negatively affects their financial situation by false representations, illegal schemes, or other means. This can be accomplished in several ways, including investment fraud and theft of an identity. Closely to this definition but not the same, as the first is wider than the second, stands the definition for corporate scandals, which are defined as business scandals (usually with political implications), which arise from the publication of misdemeanors of executives of large companies, often with the tolerance or complicity of the competent auditing authorities (Kirkos et al., 2007). In both cases of criminal behavior, these misdemeanors may include falsification of records / documents / financial statements, concealment or deliberate omission of transactions, improper valuations, embezzlements, etc.

Many of the world's most prominent companies and organizations have been hit by large-scale fraudulent fraud of the companies involved (Zerban, 2018). Examples of financial statement fraud that have had a tremendous impact in different international environments are Enron, Parmalat, WorldCom, Freddie Mac, Tyco, Xerox, Lehman Brothers, Satyam, proving that fraud is a phenomenon mostly emerging from strong economies (Carnegie & Napier, 2010). International financial scandals have had multiple adverse effects on the global economy, with rising unemployment rates for the lower and middle class (Abdullahi & Mansor, 2015b) while leading companies, employees, creditors and investors themselves into huge financial losses (Rezaee, 2005). Although these scandals have occurred in recent decades, wide-ranging operational turmoil, conflicting interests and the collapse of governance standards have been the catalyst for the shake-up of stakeholder confidence in the functioning of the financial market and the weakening of its credibility (Abdullahi & Mansor, 2015a; Karpoff, 2021; Rezaee, 2005).

The latest scandals of the international financial scene of Volkswagen (2015) and Toshiba, come to demonstrate the shortcomings that still exist in the implementation of corporate governance rules and to question the role of accounting and auditing as tools that protect investors in obtaining their decisions. Of course, the role played by the auditing companies themselves in fraud is considered crucial, as their mistakes and omissions result to an ineffective assessment of the accuracy and correctness of financial statements. The scandals of previous decades have made it clear that auditors have failed to live up to their role, succumbing to personal ambitions and financial gains.

The immediate consequence was that the auditing profession was severely criticized both for its work and for its responsibilities, while a crisis of trust was created in the public towards auditors in general. Following this climate of doubt and controversy, the need for immediate intervention became urgent, with states being forced to strengthen their legislation (e.g. the Sarbanes-Oxley Act), increasing safeguards against such phenomena and introducing new requirements for companies, their administrations and statutory auditors (Jones, 2011).

Focusing furthermore on the thematic of the present study, "financial fraud" states the activities of criminal groups which misuse financial or payment systems for the purpose of obtaining financial gain. It is a fact that as people become more civilized, violent crimes give way to more 'sophisticated' crimes, called 'Economic Crimes', such as tax evasion, bribery, corruption, embezzlement, money laundering, domestic information on the stock market, breaches of competition law, production of defective or counterfeit products, irregular supply contracts, piracy of music or computer software, industrial espionage, environmental pollution etc., which have received alarming dimensions worldwide. In addition to the purely financial costs for companies and organizations, financial crime also has significant negative effects on employee morale, corporate reputation and business relationships. (Krambia-Kapardis & Tsolakis, 2011) in their book "Financial crimes in business" approach the concept of financial crime mainly in the light of detecting and preventing it, offering the necessary valves (political and judicial) to prevent professional fraud, as they define corruption, misappropriation of assets and falsification of financial statements.

In Greece, in particular, the phenomena of bribery and corruption are considered extremely widespread, especially in the public sector which still stands inaccessible to efficient and thorough evaluation and survey, which is reflected in a recent survey conducted between November 2016 and January 2017 in Europe, the Middle East, India and Africa, on behalf of Ernst & Young and according to which Greece is ranked - based on the views of the surveyed executives - in 3rd place (with the highest rate of corruption after Ukraine and Cyprus), with upward trends compared to previous years.

In the same survey, 81% of business executives consider corruption widespread in Greece, while one in five executives say they have been pressured not to report offenses, to which they have succumbed for fear of either personal safety or development, within the company or in another company in the future, or because of solidarity with colleagues and loyalty to the company[1]. In each country, the definition of fraud differs, but it basically involves a violation of the law and / or a violation of the regulatory framework. In general, "fraud" is the deliberate deception of an individual, deceiving him with lies or by means not revealing the full truth or state of things, in order to obtain a personal, socio-economic-political benefit unjustly (Riahi-Belkaoui, 2003).

Similarly, fraud in the legal sense of the term can be defined as the offense of misleading a person by fraudulent means, such as concealing or falsifying true facts or presenting untrue facts as true, aimed at the direct or indirect benefit of the perpetrator (Kazantzis, 2006). According to Belkaoui and Karnik (1989), "fraud" consists of falsifying documents, recording false transactions in the accounting records or permanently deleting them from them, and concealing information of major importance. As one of the growing problems in modern corporate reality, it is the act performed on an entity by one person or more persons (management, staff or third parties), to obtain illegal benefit.

[1](www.ey.com)

The International Auditing Standard ISA 240 defines "fraud" as any deliberate act by an individual, the management of a company or a third-party involving fraud in order to obtain unlawful or unjust benefits. Fraud as a concept includes features and elements of many disciplines, such as accounting, administration, auditing, but also elements of psychology and criminology. The most important feature for the duration of a fraud is its concealment, noticeable difference from other criminal acts. Some of the most typical cases of financial fraud are bribery, corruption, financial blackmail, conflict of interest, mismanagement of assets and falsification of financial statements (Loumiotis, 2010).

1.2 Significant scandals of the last 20 years

During the last 20 years, a multitude of worldwide scandals uncovered shocking the world financially and socially. Hereby some of the most popular scams globally and in Greece are presented:

WorldCom

One of the greatest stock market champions of the 1990s, WorldCom is now known for having one of the costliest corporate accounting scandals in American history, with an estimated $11 billion in losses as of March. In a cafe in Hattiesburg, Mississippi, in 1983, partners headed by former basketball coach Bernard Ebbers drew up their concept for a worldwide business on a napkin. In 1984, their business, started acting as an interstate reseller. In 1985, Bernard Ebbers was appointed CEO, and in August of 1989, the business went public. At the time, its $40 billion merger with MCI in 1998 was the biggest ever. Investors and Wall Street analysts favored the company. In June 1999, the stock rose to a high of $64.51. CEO Bernard Ebbers was one of the wealthiest men in the United States at the time, according to Forbes.

WorldCom tried to buy out Sprint in October 1999 for $129 billion in stock and debt in a stock buyout. The US Department of Justice rejected the agreement. The company started to break apart at the same time as debt and expenditures piled up, the stock market declined, and long-distance rates and income decreased. In the beginning of 2002, the company made several shocking admissions, which prompted a Chapter 11 petition in July of the same year. WorldCom overstated capital spending by $3.8 billion, which increased cash flow and earnings over the five prior quarters. The actual net loss for 2001 and the first quarter of 2002 was concealed by this.

The SEC filed a civil lawsuit against WorldCom accusing it of committing an enormous accounting fraud amounting to more than $3.8 billion. WorldCom allegedly falsely exaggerated its profits before interest, taxes, depreciation, and amortization (EBITDA) by about $3.055 billion in 2001 and $797 million in the first quarter of 2002 (SEC Annual Report 2002), according to the Commission's complaint.

Enron

The Enron Scandal, which came out in 2003, was a financial scandal involving the Arthur Andersen accounting firm and the energy firm Enron. Following a series of revelations about controversial accounting practices in the 1990s, Enron was on the verge of bankruptcy in November 2001. An attempt to rescue the firm from its destiny from a similar but smaller energy company, Dynegy, failed. Enron filed for bankruptcy on December 2, 2001. As the scandal unfolded, Enron shares plummeted from $90 per share to less than ¢50 per share. And while the company was among the blue chips, this unexpected development had a big impact on the market. Enron's dive came after it was revealed that much of the profits and revenue were the result of agreements with special purpose entities, limited liability companies it controlled.

Due to the aforementioned, a large portion of the company's liabilities were not disclosed in its financial statements. Large debts were kept off balance sheets by CEO Jeff Skilling and former CEO Ken Lay, with the primary outcomes being the company's bankruptcy, the breakup of Arthur Andersen, the loss of $74 billion for shareholders, and the investors experiencing huge losses. The Enron scandal was even cited as the biggest audit failure. In 2002 WorldCom's internal auditing department with the assistance of KPMG team – who had inherited the WorldCom account by Arthur Andersen- revealed that line costs had been underreported by capitalizing rather than expensing and inflated revenues with fake accounting entries and they overstated their assets by over $11 billion. This fraud led to the collapse of the company, 30,000 individuals lost their jobs, while investors lost $180 billion.

Lehman Brothers

On September 15, 2008, Lehman Brothers activates Article 11, i.e., it files for bankruptcy. The bankruptcy filing is still the biggest in American history. Lehman owned assets of more than $600 billion. As the mortgage crisis reached its peak, the company was extremely vulnerable to any decline in property prices as it had mortgaged so many properties that it had ended up as a real estate agency. The bankruptcy caused the Dow Jones Industrial Average to plummet by 4.5 percent per day, which was the biggest decline since the September 11 attacks.

In the financial crisis involving Lehman Brothers, the management sold toxic assets to Cayman Island banks with the intention of later buying them back, giving the impression that Lehman had $50 billion more in cash and $50 billion fewer toxic assets than it actually did. The Lehman's external auditors Ernst and Young were aware of the use of Repo 105, and they fail to disclose its use. In 2009, Satyam, an Indian provider of IT solutions and administration accounting, overstated their revenue by USD 1.5 billion. Ramalinga Raju, the company's founder and chairman, purposely altered the cash balances, margins, and revenues to the tune of 50 billion rupees. The Indian subsidiary of PwC was fined $6 million by the US Securities and Exchange Commission (SEC) for failing to carry out its obligations regarding the auditing of the financial records of Satyam Computer Services in line with the requirements for auditing and standard of conduct.

Folli Follie

As far as domestic cases of corporate financial fraud are concerned, Greece had recently been stigmatized by the Folli Follie scandal by Koutsolioutsos family. In 2018, this scam was revealed[2] after research from the American fund Quintessential Capital Management (QCM). The defendants' criminal activity dates back to at least 2006 and is characterized by an "orgy" of forgery of bank documents, fictitious transactions and alleged subsidiaries of Folli-Follie in Asia, which in fact had no commercial activity.

In particular, as it was established by the audits, Folli – Follie's company in China, for a number of years presented fictitious bank balances in legitimate bank accounts, as well as fictitious bank balances in non-existent fictitious bank accounts and produced

[2] https://bit.ly/3hPZEv0

bank statements to document said balances. Subsequently, Dimitris and Tzortzis Koutsolioutsos, according to the case file, sent to Greece the falsified financial statements of their alleged companies in China, and requested that these be consolidated with the balance sheets of the companies of their other group, which was active in our country, the Europe and North America. Specifically, it was found that the revenue from the Asian subsidiaries was $116,847 million out of the $1,112 billion announced by the company. Moreover, Folli Follie had 341 phantom sales points in order to record higher revenues. It had been concluded that he recorded sales were fake at about 200% up. Thus, according to the case file, they managed to draw up "advantageous business agreements and had easy access to loans through credit institutions, thereby shifting the risk (credit, investment, etc.) knowingly to third parties.

The judges also refered to the actions of the Koutsolioutsos family (father and the son) which aimed to prevent any audit that would reveal their actions. The members of the Judicial Council characteristically report that the two defendants as managers formed an appropriate hierarchical internal structure in such way to cover their fraudulent behavior such as people of their trust and financially dependent on them were chosen as members of the Board of Directors. The purpose of this criminal organization was to provide a misleading image regarding the financial figures of the group inside and outside Greece, to present the company Folli Follie as a healthy, solvent, profitable business, with large annual turnovers, stable annual profits, cash available, prospects for growth and further profitability and to make it an attractive investment, since with the false information he drew up advantageous business agreements and had easy access to loans through credit institutions, thereby shifting the risk (credit, investment, etc.), in knowledge of the members of the criminal organization to third parties.

This significant financial fraud of Folli Follie, caused a huge loss which, although it cannot be precisely determined, is estimated to amount to 413,078,346.17 euros.

Furthermore, it is interesting to be noticed that QCM had raised a big question mark here. Baker Tilly, an internationally renowned auditor, had served as the audit firm for a while, but it had just been replaced by Ecovis, a new company, indicating the importance of changing the company of auditors during the scam years.

Finally, the prosecutor had accused of (apart from of the Koutsolioutsos family and other shareholders for fraud and manipulation):

1. The Capital Market Commission regarding its broader stance in the management of the case, both for the previous period and for the years preceding the 2017 balance sheet, as well as for the general care and determination in the exercise of its duties, i.e. its supervisory role
2. To Legal Entity of Public Law which checks the balance sheets, which as it had appeared from the whole process were altered for several years and over a period of at least 11 years
3. To banks for the loans, they granted for a number of years to the company Folli Follie, without checking the accuracy of the company's balance sheets and financial data.

The above accusations proved that none of the audit controls or systems had been implemented in order to reveal this significant fraud case indicating that the current audit system needs to be improved.

Bank of Crete

In 1979, Giorgos Koskotas worked at the Bank of Crete, handling accounting and computing sections. He misappropriated money, stealing $150,000 from customers' checks and the bank's reserves, resulting in his owning 95% of the bank's shares. Koskotas became President of the Bank's board and executive director in 1985, gaining complete control of the bank and its management. In 1987, Koskotas replaced forged documents in several branches, covering up his 32 million US dollars misappropriation. After being freed from US authorities, Koskotas assembled his bank employees to forge documents, falsely certifying the legality of moneys credited to Koskotas's fictitious personal account. In 1982, Koskotas set up 'Line' as a mass media company, diverting approximately 12 million drachmas into the company. Koskotas took control of his media company, bought newspapers Kathimerini, Evdomi, and Vradinifor, and published positive views on Papandreou's socialist government. He was exposed for fraud and exploited inadequate accounting systems.

Accounting frauds in the 1980s were attributed to the shortcomings of accounting information systems. During that period, Greek banks relied on based on autonomous branches setting up their own ledgers and general ledgers. However, these systems struggled to handle emergent economic events, leading to problems in reconciling inter-branch accounts. This weakness exposed the system's weakness, as initial cash was recorded but no corresponding liability was set up. All major Greek banks faced similar problems, except for Bank of Crete, which had larger percentage balances after Koskotas's involvement increased.

Creta Farm

After the revelation of the Folli Follie fraud, yet another scandal came to prominence in 2018[3]. In the case of Creta Farms there was a dispute between the two big shareholder brothers (C. Domazakis and Emm. Domazakis), which appears to have started in 2014 and centered on the 11 million euros that one brother (Constantinos) gave to the company and this amount had never been recorded in the financial statements. Meanwhile, Delloite announced that Creta Farms Foods S.A, one of the largest pork meat producers and meat-processing company in Greece, committed the following illegal activities: 12 companies affiliated with the company but were not disclosed in the financial statements and illegal outflows of more than 2.9 million euros had been tracked, moving from the company to the two main shareholders during 2014-2018.

Moreover, the big issue that aroused was the huge impairments that sank profitability and net worth, which had as result the company to show in 2018 a negative net position of 4.7 million euros from a positive 8.7 million euros and the parent positive net position of 6.6 million euros from 42.5 million euros. Characteristic example of how investments and real estate were measured in previous years is the fact that, as stated in the financial statements, the value of the buildings and plots of the subsidiary FARMA THESSALIA was valued at 8,3 million euros, which were sold for 3,838 million euros, amount which covered an equivalent loan of Piraeus. Of course, it is not mentioned to whom the specific properties were transferred, resulting in a group-wide loss of approximately 4.5 million euros. Moreover, it is important to be mentioned that having taken into account the cash flows, the group recorded an amount of 3.46 million euros for

[3] https://bit.ly/3zqN7nZ

cash advances and loans to third while it was known that the group was facing difficulties was not able to give such amounts to third parties.

The interesting part in this case is that in the last 15 years, nine financial directors, five accounting managers and several external and internal auditors have left the company. Especially in the last three years until the scandal "erupted", the changes in the financial directors were something unprecedented and probably have not happened again, at least from the point of view of a listed company.

1.3 Ways to convict fraud

Financial crime includes a wide range of offenses, which are directly related to money. Its legislation is included, in addition to the penal code (Law 4619/2019), in various criminal laws. These laws deal with a variety of forms of financial offences, such as competition, privacy, corruption, bribery, corporations, government abusers, smuggling, money laundering, the environment, intellectual property, personal data, tax offenses, debts to the State and stock exchange transactions. Also, financial crime is defined as a wholesome of an illegal activity, which is carried out through businesses and results in the threat or damage to the proper functioning of the economy.

The Greek legal order is aware of crimes that have an economic content or economic significance, as this is particularly the case with crimes against property and property rights (Pitsela, 2011). These acts, however, do not constitute financial crimes, simply because they present an economic dimension. The term "financial crime" is generally accepted to mean something different. The phrase "financial crime" has no accepted definition in the field of criminal law. However, the meaning of this is usually defined by reference to the particular legal good that is infringed upon its performance. The legal property that is to be protected, refers to the economy, and in particular the national economy as a whole or in some of its sub-sectors or institutions, or the economic class. Thus, financial crime is defined as the criminal act standardized by law, which is committed by exploiting the possibilities of the financial system and aims at increasing the property of the perpetrator or another for whom he acts and, as a rule, damages the property of the State, banks, businesses or consumers (Manoledakis, 2005).

In other words, economic crime is defined as the totality of that illegal activity, which is carried out through businesses and has the effect of insulting (threatening or damaging) the proper functioning of the economy or its important branches and institutions. If organized crime and purely fiscal crimes get excluded from the field of financial crime, one will find that the common denominator of economic crimes is the fact that everything is against the legal good, which is the economic order.

A more precise terminology would refer to "financial crime", the insult to the economic order, which is externalized by the acquisition of economic advantages arising either from the development of illegal economic activity, or from the abusive exploitation of economic power. On the other hand, in contrast to financial crime, the financial offense as unlawful conduct, from the point of view of administrative law, constitutes an act of non-compliance of the administered with the requirements of legitimate economic activity. This non-compliance is abstractly assessed by the legislator as dangerous to the economic class. Thus, when this behavior does not exceed the limits, which are set by the legislator and determined qualitatively, so that with the corrective intervention of the state (eg by imposing an administrative sanction), it is possible to force the administrator in compliance with the purposes of the administration, then the conduct does not go beyond the limits of administrative law.

Otherwise, if the administrative offense exceeds the limits of administrative law, it constitutes a criminal conduct, which is assessed as such on the basis of the principles of criminal law. Consequently, the financial size, either as the impugned legal property or as the content of the purpose of the offense, is not in itself sufficient to give that act the character of an economic crime, in the technical sense of the term. The noticeable difference between a financial crime and a crime with a purely financial content or with a purely economic significance, lies precisely in the appearance of the expression of the mechanisms of the economic system, with appropriate handling by the perpetrator. The consequences of the crime are felt, to a greater or lesser extent, as deviations and insults of the economic system itself, because economic crime may ultimately be a pathological consequence of the economic system itself.

Examples of criminal acts that are included conceptually to the above-mentioned term, are the following (Reiner, 2017):

1. Incorrect and excessive invoicing for products and services. This category uses deceptive pricing practices to carry out unauthorized value transfers between importers and exporters of goods and services.
2. Multiple invoicing of goods and services. The same product or service is billed more than once in this type of fraud, and the payments are frequently made using a variety of different banking institutions. This scenario discusses the Ring fraud, a type of financial scam.
3. Maintaining record of income before it is received
4. Generating fictitious income
5. Boosting profits with non-recurring transactions
6. Postponing existing costs until a later time.
7. Failing to record or disclose liabilities
8. Delaying the use of present revenues
9. Transferring future costs to a previous time frame
10. Land grabbing/Real estate scams
11. Public employee bribery and corruption
12. Unlawful international trade
13. Company's scams

The latter two tactics project current-year profits into the future in an effort to look to have a steady income over the years, whereas the first five tricks aim to boost current-year earnings. Financial accounting got increasingly sophisticated along with mergers and acquisitions and advanced financial instruments, such as holding derivatives or engaging in other off-balance activities, making the identification of falsified financial statements (FFS) more challenging than ever. In an overall perspective, financial fraud is expressed through numerous criminal acts, each one affecting a different aspect of economic activities. The present thesis focuses, as it will be analyzed thoroughly in the following chapters, on the financial statement frauds occurring in companies listed in the Athenian stock exchange market.

According to ACFE (2022), there are three main categories of conducting occupational fraud:

I. Corruption

a. Conflict of interest in procurement and sales sectors

b. Invoice bribery and bids rigging
c. Illegal tips
d. Financial Blackmailing

II. Misappropriation of Assets

a. Cash (Theft of cash on hand or receipts or fraudulent payments such as forged checks for personal needs, phony purchases, etc.)
b. Stock and all other assets (misuse and theft, including phony sales and transportation, etc.)

III. Financial statement fraud

a. Overstatements of net worth and net income (such as timing discrepancies, phony revenues, hidden liabilities, incorrect asset estimates, and fraudulent declarations)
b. Understatements of net worth and net income (such as timing issues, understated revenues, inflated liabilities, incorrect asset estimates, and misleading disclosures)

In the majority of cases (86%) an employee steals or uses company resources in an improper manner. However, the median loss from these tactics is often less than \$100,000 each instance. Financial statement fraud, which occurs when an individual intentionally causes a significant error or omission in an organization's financial statements, is the least common (9% of schemes) but most expensive category (\$593,000). Corruption, which includes offenses like bribery, conflicts of interest, and extortion, is categorized according to frequency and losses in the middle. A median loss of \$150,000 is caused by these schemes, which happen in 50% of cases (ACFE, 2022).

In the same survey, the top 5 methods that fraudsters use in order to conceal the scams are presented as follows:

1. Creation of fraudulent physical documents (39%)
2. Change of physical documents (32%)
3. Creation of fraudulent electronic documents or files (28%)
4. Change of electronic documents or files (25%)
5. Destroy or withholding of physical documents (23%)

Moreover, it is an interesting fact that a 12% of fraudster does not proceed in any action to conceal its fraudulent action.

Although financial crime is the subject of worldwide scrutiny, prosecution and investigation, its peculiarities in relation to common criminal crime make it difficult to deal with. Financial crime has many manifestations, it moves intangibly through financial, stock exchange, banking, and other transactions, it appears in a legal form, it can involve civil servants, prosecutors, judiciary and even politicians, resulting in deficits in political will and action to combat it. Dealing with it requires specialization and continuous training of the control and prosecution bodies, as new forms and methods are developed, and in many cases the damage from the commission of the crime seems to be diffused or unclear (Law 4619/2019).

In general, it seems that in all its different manifestations, financial crime has common elements and characteristics, which are the following:

A. Financial crimes, because they are committed mainly in the private sphere, are relatively invisible, as the perpetrators are legally and reasonably present at the crime scene.
B. Financial crimes, because they are usually committed during lawful business and professional activities, constitute a substantial and total abuse of the trust of the parties involved, such as employees, customers, associates, suppliers, banks, the state.
C. Financial criminals can be people who have knowledge, skills, access to internal information networks, knowledge and use of information and media. This makes the crimes more complicated and even more difficult to detect (Jackson et al., 2010)

The prosecution of financial crimes is ambiguous, problematic and the evidence of the crime is often questionable. In most cases of financial crimes, the deceit, the intention, the computational thinking, the pre-planning of the act, the synergy, the formation of a criminal gang is questioned. Also, the rarity of the disclosure of financial crimes and often the impossibility of presenting solid evidence, testimonies, evidence of a crime, leads to the acquittal of the accused due to lack of sufficient evidence, as well as the absence of a previous convicted similar case. In many cases, law enforcement authorities analyze and study documents and evidence, and are unable to have evidence because there are no eyewitnesses. The offense may be ongoing and there may be a lack of temporal or geographical identification, such as fraudulent transactions, concealment of taxable material, bank fraud, and the misuse of confidential information.

Finally, it is characteristic that most of the time, both the prosecuting and the judicial authorities are unable to locate where the money that has been illegally obtained in the execution of the illegal act has been spent or "sent". Therefore, they try, either through the freezing of the bank accounts of those involved and the confiscation of the assets of those convicted of financial crimes, or through a background financial check of the suspects, to identify the misappropriated money came from. The latter is particularly difficult, as the mentioned check can mainly work hypothetically and indicative of where the money may have been spent. Illegal enrichment can be seen in part and under certain conditions with this control. In most cases, however, the proceeds of crime are consumed immediately, diffused into consumption in the form of services and expenses, and in some cases even "fled" illegally abroad, so that their traces are lost forever.

1.4 Profile of fraudster

It is an undeniable fact that if researchers concentrate on bettering their grasp of the fraudster's profile, the volume of scams that are discovered will rise dramatically. According to academics, there is a meaningful relationship between the authority of the fraudster and the extent of the scam. Only 23% of the frauds in the survey conducted by ACFE (2022) were committed by owner/executives, yet the median loss in such cases ($337,000) was substantially higher than losses brought on by managers ($125,000). The corresponding amount by employees is $50,000. Hereby are presented some elements of the ACFE’s survey:

1. The main departments in an enterprise that are keener to proceed in occupational fraud are operations (15%), accounting (12%), executives or top management (11%) and sales (11%).

2. 85% of fraudsters appears red flags for fraudulent behavior.

3. 48% of top management fraudsters delete the evidence.

4. 61% of managers fraudsters create fraudulent evidence.

5. More perpetrators are individuals with higher levels of authority in their companies (62%), such as managers, executives, or owners. This percentage had been increased the current years (in 2012 was 56%).

6. Most scams derive from collaboration of the fraudsters. In 2012 a 58% of revealed scams was conducted by one perpetrator, although in 2022 the same percentage (58%) was for 2 and more fraudsters.

7. The perpetrators are mainly men (73%). Although the losses from men scammers were $200,000 and $91,000 from women in 2012, in 2022 the gap has narrowed and the amounts from men and women perpetrators were $125,000 and $100,000 correspondingly.

8. Long tenure fraudsters steals triple bigger amount than the new ones. For instance, an employee who works in the company 10 years and more, can cause an estimated scam of $127,000 while an employee with a maximum of 5-year-tenure can cost the company approximately $36,000.

9. The 54% of scams had been committed by individuals between 31-45 years old. Contrasted with, median losses tended to increase with age. Only 3% of perpetrators were over 60, yet their typical loss was $800,000, significantly higher than that of any other age group.

10. 65% of people who committed occupational fraud had a university degree or higher. This group's median losses were higher than those of people with lower levels of education.

Additionally, according to KPMG (2013), a 39% of perpetrators are highly respected, a 35% develops a friendly behavior among the professional environment and a 33% are extroverted.

Except for this evidence, researchers focus on deeper characteristics of fraudsters' behavior, and motives which exhibits fraudulent red flags. ACFE (2022) reveals the most common red flags which leads to occupational fraud:

1. Living beyond means (39%)

2. Financial distress (25%)

3. No usually relationship with vendor/client (20%)

4. No behavioral red flags (15%)

5. No desire to share tasks (13%)

6. Intimidation (12%)

7. Family issues such as divorce (11%)

8. “Dealer” attitude (10%)

Less than 10% of the remaining causes are as follows: intense peer or family pressure to succeed, workplace stress, vacation dismissal, social estrangement, prior legal issues, lack of authority, other career barriers, etc. Long tenure individuals are developing more common the categories such as living beyond means (43%), very close relationship with specific vendor/clients (25%) and no willing to share tasks (19%).

Moreover, KPMG (2013) claims that desiring to conceal bad information (22%) corporate competition (23%), market competition (29%) and aggressive sales environment (31%) introducing some important determinants to lead an individual to commit occupational fraud. The same study focuses on specific emotional motivations that explains the reason why scammers conduct fraudulent actions; sense of anger (17%), sense of being with a lower salary than expected (14%), sense of being underestimated (13%) and sense of fear (7%). Furthermore, KPMG’ s researchers have studied the characteristics of personality that may lead to scams; personal profit (47%), greed (42%), sense of superiority (36%), tendency of being dishonest and in collusion (31%) and capability “because I can” (18%). The last characteristic can be more intense because the applied controls are weak giving the advantage to fraudsters to expand their fraudulent action within the company. According to (ACFE, 2022), an absence of internal controls (29%), internal controls being overridden (20%), and a lack of management review (16%) are some of the reasons fraudsters continue to engage in their activities.

Additionally, ACFE (2022) approaches the HR-related issues which are most common between perpetrators; fear of job loss (16%), poor performance (15%), denied increase of salary or promotion (12%), cut of benefits (7%), decrease in salary (6%), actual job loss (6%), unwilling decrease of hours (4%) and demotion (4%).

All the aforementioned of fraudsters' characteristics, behavior, personality, and motivation, provide to researchers an advantage in thoroughly understanding why a perpetrator will proceed in fraudulent action, enhancing auditors' performance in detecting red flags for occupational fraud.

1.5 The cost of financial fraud

Financial fraud is widely known as an excessively costly crime with a multiple negative impact globally. Whoever the scammer is within the company, the followers are accepting the consequences of the fraud once it is revealed: owners/executives, top management, managers, employees, shareholders, external auditors, government, organizations, vendors, clients and of course, the society. The financial costs combining with the negative reputation of failing detecting a fraud, the unemployment causing by companies' closure and the lack of trust in the whole control mechanism lead to a collapse of institutions. Karpoff (2021) develops a third-party enforcement -named The Trust Triangle (TT)- which encourages individuals of not committing financial crimes. The parties are personals ethics, integrity, and culture (TT1), laws, institutions, regulations, and regulators (TT2), market forces and reputational capital (TT3). In the same survey, it is mentioned that a collapse in institutional trust promotes an uptrend in scams while all the three aspects of trust triangle are undermining.

Focusing on financial costs, the global study by the ACFE (2022) analyzed data from 133 countries and estimated that about 5% of annual corporate revenues are lost due to fraud. The largest part of this cost is attributed to financial statements fraud, which, although not as frequent as other types of financial fraud, it has a much higher cost. The same study estimated that the damage from investigative fraud cases exceed $7 billion. In the case of fraud, the average loss was found to be $130,000, whereas the average loss due to corruption an average was $250,000. According to a previous study, Karpoff, Lee, and Martin (2008), when a company's wrongdoing is revealed, it loses 38% of its market value. This finding suggests that the market is responding to a more truthful picture of a firm's financial status. This represents the "true" corporation value adjustment if the company's books had not been falsified.

Moreover, ACFE (2022) proves that the more the perpetrators, the higher the cost. For instance, a median loss of three and more fraudsters is $219,000 and the corresponding amount for one perpetrator is $57,000 with an average period of scam 12 months. According to the same data, an owner or executive commits fraud at a rate that is around three times as fast as that of employees and supervisors. These results highlight how those in the top levels have a lot more ability than those in lower positions to harm the organization. Energy, manufacturing, and government/public administration were the three industries with the highest incidences of corruption, according to (ACFE, 2022).

In their study, Button *et al.* (2015) focused on wide variety of consequences to other fields. Thus, they discovered that some of the major extra expenditures included those for investigations, staff postponement, internal disciplinary actions, external sanctions, continuing staff replacements, incidental charges, and unforeseen expenses. The findings, particularly for initial scams under £25,000, revealed considerable expenditures that are significantly more than the value of the crime.

Research and the experience of professional investigators have shown that the longer a fraud lasts until it is exposed, the more it grows and affects its victims. Specifically, the ACFE study notes that fraud cases that last more than five years, cost 20 times more than those revealed within 6 months. On the other hand, the perpetrators tend to start with small-scale scams and rapidly increase their acts in less than three years. Therefore, it is imperative that companies use fraud prevention and timely repression mechanisms to minimize the damage caused.

Financial statement fraud has affected significant corporations over the past few decades, which has hurt capital markets and reduced shareholder value (Hajek & Henriques, 2017). In particular, Beasley (1996) notes that (1) the average fraudulent firm's stock price decreased by 16.7% in response to the initial press disclosures of an alleged fraud, (2) 28% of fraudulent firms were liquidated or declared bankrupt within two years, (3) 47% were delisted from a national stock exchange, and (4) 62% were impacted by material asset sales. In addition, four of the ten biggest bankruptcies in American history were linked to significant financial frauds (Abbasi et al., 2012).

Financial fraud may, in fact, serve as a good predictor of severe financial issues that lead to bankruptcy (Beneish, 1999). As a result, investors and other participants in the capital market have been extremely concerned about financial statement fraud.

1.6 Corporate governance approaches

It constitutes an admittedly fact that occupational fraud costs significant amounts to a plethora of sectors and, it is understandable that the existing audit controls cannot detect the frauds in a short time exceeding the financial costs. As it has been mentioned in the above paragraphs, it is important auditors to change their approach in detecting FFS by implementing theories concentrated to understanding the motivation of the fraudsters and not only to examine financial ratios.

1.6.1 Fraud triangle theory

"Trusted persons become trust violators when they conceive of themselves as having a financial problem which is non-shareable, are aware this problem can be secretly resolved by violation of the position of financial trust, and are able to apply to their own conduct in that situation verbalizations which enable them to adjust their conceptions of themselves as trusted persons with their conceptions of themselves as users of the entrusted funds or property (Cressey, 1973).*"*

In 1953, Donald Cressey published his survey-among many others- on organized crime, called *Other People's Money: A Study in the Social Psychology of Embezzlement.* Due to his academic studies on sociology and criminology and having an honest interest on why fraudsters get tempted to commit fraud, he conducted extended research that included the interviewing of 200 fraudsters convicted for embezzlement. It was after identifying 3 basic elements that were common to all fraudsters, in relation to their motivation for committing fraud, that Cressey developed his theory on Fraud Triangle.

According to Cressey's theory (1953), there are three conditions that accompany fraud: opportunity, pressure / motivation and rationalization. These three conditions create the concept of the "Triangle of Fraud". It refers to a set of conditions that must exist for a criminal activity or act of fraud to take place. Thus, the three components of the fraud triangle are defined as follows:

1. The *opportunity*, which must be sufficient to commit fraud. Therefore, situations or the environment must allow criminal activity to take place. Opportunity includes generally loose of restrictive measures circumstances that allow the person to move easily towards the thought of committing fraud and eventually daring it.
2. The *rationalization* of the individual, which is the most important element, as this includes the mentality of an individual that leads him or her to commit crime or act unlawfully. By giving oneself excuses for the righteousness or the unharmful of the act, the potential fraudster attempts to justify the act of fraud.
3. The *pressure* that the person receives or the motivations that influence him to commit fraud. No person commits a crime or fraudulent activity without any motive. Motivation is most frequently created by feelings of revenge or distress, or when money is scarce or in great need or when the potential fraudster is being blackmailed.

More specifically, those three factors that create opportunities for fraud, can be examined through an alternative/opposite scope (Brink et al., 2013):

An environment of opportunity is cultivated due to Dorminey *et al.* (2012):

1. Lack of internal control. Absence of segregation of duties, lack of proper organization in a company, absence of physical controls and ineffective supervision are characteristics of an inadequate internal control system that nurtures criminal activities.
2. Ineffective board. When the board of directors is inadequate and weak or lacks independence, senior management is being given an opportunity to conduct fraud as they believe that there is no strong control mechanism to monitor its conduct and practices.
3. Practices of impunity. When there are no penalties for offenders in a company, employees are encouraged to commit fraud as they believe that revealing it will not lead to dismissal or any severe reprimand.

4. Lack of ethical guidance and leadership. There is a need for the competent people, who are usually the management and the senior executives of the departments of a company, to take the responsibility to properly inform and guide those involved in all departments of an organization, most importantly in the accounting circuit. In addition to providing ethical guidance, management must also act as a model of ethical behavior.

Furthermore, rationalization represents a code of moral values or behavior that allows individuals to commit a fraud voluntarily and intentionally, justifying the act. People do not commit a fraud if they cannot justify it according to their own beliefs and morals. For example, executives, to justify an act of fraud, invoke opinions such as that everything should be sacrificed for the good of the business or that all financial units apply similar practices. On the other hand, employees can justify an act by claiming that they feel that they are paid little, that everyone does the same, that one has to do anything in order to move up the hierarchy (Gabbioneta *et al.*, 2013).

The motives and reasons that lead a company to the deterioration of its financial statements come from both the internal (employees, management) and the external environment (industry, state, competition). The growth rate of companies in recent decades and the competition of companies for higher profits, have led both executives and employees themselves to commit fraud either for the benefit or to the detriment of the company. The expectations and motivations that are created are the reasons for cultivating to the individual the desire to commit fraud. The most important motivations / pressures that push individuals, executives, and non-executives, can be located in Mayhew and Murphy (2014):

1. The intense competition of companies
2. The pressure to meet the expectations of investors and the capital markets
3. Financial incentives from the company itself. For example, executives may be better paid or get a raise when the organization’s financial results are satisfactory or present a raising tendency. These goals can easily lead to adopting practices that lead to distorted financial statements or equivalent acts of fraud.
4. The composition of the board of directors.

5. The managing director holding the position of chairman of the board. It is possible that in such a case there is no proper separation of responsibilities between the two roles.
6. The growing growth rate and size of a business.
7. The specialization and high training of the external auditor.

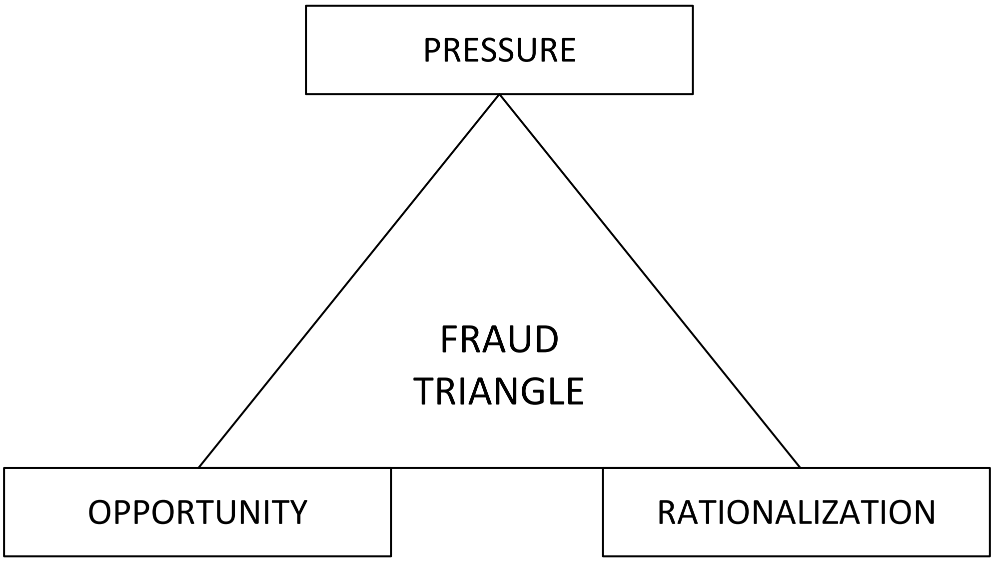

Figure 1: Fraud Triangle

1.6.2 Fraud diamond theory

Completing and evolving the fraud triangle theory, Wolfe and Hermanson (2004) published in "The CPA Journal" a new improved theory that introduced a new parameter to the aforementioned three conditions, that of the capability of the fraudster to conduct fraud. The Fraud Diamond theory, as it was named, is more oriented to individuals that play an important role in committing fraud. Maintaining the three basic elements of Cressey's theory, the advanced fraud diamond theory, adds an additional element-that of the ability of a person to proceed in fraudulent actions- that needs to be detected before deciding on the existence of fraud. In a broader perspective, opportunity paves the way for fraud, while motivation and rationalization allow the individual to go forward with it.

However, it is considered as a prerequisite that the individual must have the ability to recognize the opportunity and seize it, not just once, but repeatedly.

Ability

The perpetrator has the ability; personal characteristics and skills. These may include position / tasks, knowledge and intelligence, self-confidence / selfishness, skills, pressure, ability to lie effectively and immunity to stress. Ability is described as the state in which a person possesses the required traits, abilities, or skills to conduct fraud. These characteristics play a crucial role in whether there is a real possibility of fraud. A prospective scammer must be able to spot the specific opportunity for fraud and have the skills necessary to take advantage of it. Ability is supported by status, cognitive ability, ego, uncontrollable urge, dishonesty, and anxiety (Wolfe and Hermanson, 2004). According to Bressler and Bressler (2007), as reported by Mackevičius and Giriūnas (2013), having motives, opportunities, and justification of one's behavior, does not necessarily mean that a person is able to commit fraud, due to lack of the ability to do so or hide it. This element, according to Albrecht *et al.* (1995) is crucial when large-scale or persistent fraud is being evaluated. Additionally, considering Albrecht *et al.* (1995) paper only someone with a very high level of capacity will be able to comprehend the current internal control techniques and processes, spot its flaws, and use them in planning when enacting fraud.

Position / Operation

A person's position or job within an organization may give them the possibility to create or seize a fraud opportunity that is not available to others, according to Wolfe and Hermanson (2004). Additionally, as people frequently complete a job, like signing banking agreements or opening new vendor accounts, their capacity for fraud grows as they get more familiarity with the protocols and safeguards.

Intelligence / Creativity and Ego

The potential fraudster is someone who can identify and exploit gaps in internal control mechanisms and take advantage of their status, job function, or other privileges. Smart, experienced, creative people with a consistent understanding of the controls and vulnerabilities of the systems they handle are usually the ones committing most of the biggest scams today. ACFE (2022) revealed that 65% of scammers had at least a college degree or higher and as it has been already noted, executives and managers commit the largest losses in occupational fraud.

Additionally, fraudsters are characterized by a strong ego and having faith that their actions won't be noticed and believing that they can get away with it if they are ever found. This self-confidence, reaching to the point of arrogance, can affect a person's evaluation of the pros and cons regarding their involvement in fraud. The estimated cost of deception is higher the more self-assured a person is. One of the most prevalent personality qualities of fraudsters, referring to a related article titled "The human face of fraud," is ego. An individual who is "driven to succeed at all costs, self-centered, confident in themselves, and arrogant" is considered selfish (Duffield and Grabosky, 2001). Additionally, the most frequently seen personality traits include a tendency toward grandeur, a need for adulation, and a lack of compassion for others. People with this disorder believe they are superior or unique and are likely to have inflated views about their accomplishments and abilities.

Coercion, Deceit and Stress

A successful scammer can force others to commit or conceal a fraud (Rudewicz, 2011). A person with a very persuasive personality may be able to persuade others to commit fraud or simply conceal the fact. The fact that "bullies" are a typical personality type among fraudsters is also emphasized. These individuals "make unusual and substantial demands of those who work for him, develop fear rather than respect, and refrain from being bound by the same standards and processes as others" are all examples of bullies.

According to Wolfe and Hermanson (2004) and Rudewicz (2011), a successful scammer must also lie effectively and consistently. To avoid detection, the scammer must present to auditors, investors, and others a realistic, yet untrue, version of facts and convince them of an unrealistic situation. Thus, fraudsters must also can keep track with their lies so that the overall story remains consistent. In the Phai-Mor scam, the auditors claimed that Phar-Mor had formed a group of scammers consisting of executives and former auditors whose mission is to ensure that they are constantly working to hide evidence of fraud. Among other things, the auditors claimed that the fraud team not only lied but also forged documents and "cleaned up" what the auditors found out in order to conceal their acts (Cottrell & Glover, 1993).

Another common characteristic of fraudsters is their ability to handle stressful situations (Manurung & Hadian, 2013). Fraud also requires fraud management over a long period of time, a situation that can clearly be stressful. As the risk of detection has not been eliminated, along with the personal consequences an arrest may induce, there is a constant need to hide the fraud daily. The individual must be able to control his/her anxiety, as committing and concealing an unlawful act can be an extremely stressful process (Rudewicz, 2011).

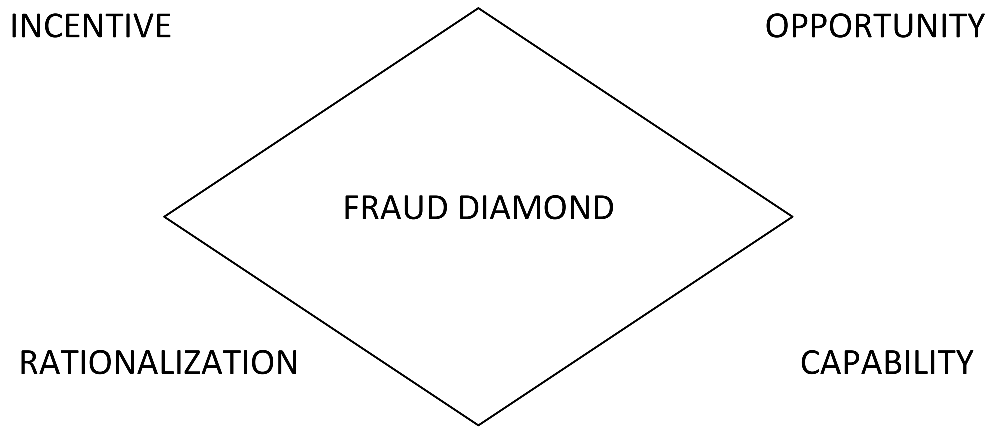

Figure 2: Fraud Diamond

Chapter 2. Literature Review

2.1 Financial Fraud Detection

The negative consequences due to financial statement fraud continue to rise. A vast amount of losses is annually estimated and except for the economic cost, fraud cases cause many other harmful effects. Major financial scandals (e.g., Enron, Parmalat, WorldCom, Freddie Mac, Lehman Brothers, etc.), lead to the rising unemployment rates wide-ranging operational turmoil (Abdullahi & Mansor, 2015b), conflicting interests, and the collapse of governance standards. The last three difficulties have been the catalyst for the shake-up of stakeholder confidence in the functioning of the financial market and the weakening of its credibility (Abdullahi & Mansor, 2015b; Albeksh, 2016; Bierstaker et al., 2006).

During crisis conditions, companies face complex financial problems such as lack of liquidity and higher possibility of failure, thus increasing the likelihood of resorting to fraudulent activities (Giovanis et al., 2016; Rezaee, Z., and Riley, 2010). Some companies use innovative accounting methods and other actions that are contrary to the rules governing the preparation of financial statements, in an effort to avoid failure and improve their performance and reputation (Pazarskis et al., 2017). According to Amiram *et al.* (2020), over 50% of the offenders in their sample would rationally consider it advantageous to commit misbehavior if the risk of detection is 31% as predicted by (Dyck, Morse, and Zingales, 2010).

In response to the growing concerns about the identification of falsified financial statements, researchers have intensified their efforts to develop effective methods for fraud detection. Studies such as those of Beneish (1999) and Dechow *et al.* (2011), have relied on annual financial statements to extract relevant information for financial fraud detection. According to Jofre and Gerlach (2018), normal auditing procedures frequently fail to detect falsified accounting reports due to audit limitations, thus necessitating the usage of analytical approaches to detect accounting fraud. Many studies focused solely on financial variables (Kirkos et al., 2007; Spathis et al., 2002; Wyrobek, 2020) while recent research has additionally examined linguistic variables (Humpherys et al., 2011;

Kydros et al., 2022) corporate governance information (Birol, 2019; G. Chen et al., 2006), as well as the fraud triangle or fraud diamond theories (Abdullahi & Mansor, 2015b; Sánchez-Aguayo et al., 2021; Skousen et al., 2009).

Falsified reporting fraud is usually performed by management (Goel & Gangolly, 2012; Wells, 2011) and it is difficult to detect because it is mainly carried out by individuals who have the ability to bypass internal and external audit systems (Firth et al., 2011; Wells, 2011). Due to the heavy summarization and consolidation of financial statement data, it is both easier to conceal frauds and almost impossible to detect them using traditional modeling techniques, as demonstrated by Whiting *et al.* (2012). For this reason, researchers trying to change the traditional approaching by understanding the reason why perpetrators appear fraudulent behavior. More specifically, as it is already mentioned, they deal with fraud triangle and fraud diamond theory.

The fraud diamond approach examines four dimensions that lead fraudsters to conduct fraud: (a) pressure/motivation, (b) opportunity, (c) rationalization, and (d) ability. The first and most important factor is intense pressure, which is linked to an individual's motivation and drive to illegal actions. The second aspect, opportunity, refers to the activity that leads to fraudulent actions and the potential to commit them. The third component, rationalization, refers to the idea that individuals may excuse their dishonest behavior, making it appear justified and reasonable (Awang et al., 2020; Wolfe & Hermanson, 2004). Finally, it is considered as a prerequisite that the individual must have the ability to recognize the opportunity and seize it, not just once, but repeatedly (Wolfe and Hermanson, 2004).

Several notorious cases such as WorldCom and Enron and he revealed that senior managers were typically involved in, encouraged, approved of, and aware of the fraudulent operations. According to social psychology studies, managers who want to hide the truth may exhibit certain linguistic clues that make it easier to spot fraud by observing their emotions and thought processes (DePaulo et al., 1982). Consequently, earlier research has highlighted the growing significance of textual analysis of financial reporting. In the framework of annual reporting, textual information, such as the Management Discussion and Analysis section, which tries to give investors an insight into the management's perspectives regarding the organization's prospects, supports quantitative financial information. The language in this section could provide insight into managers' thought processes and point to dishonest behavior (Craja et al., 2020).

2.2 Data Mining Techniques

Fraudulent financial reporting is a crucial social and economic concern (Wei et al., 2017) and fraud detection constitutes is a complex issue due to the involvement of the human factor. Thus, commonly applied methods are not enough to detect falsified statements (Aboud & Robinson, 2020). In the literature on fraud detection based on recognizing anomalies, automated approaches based on analytical methodologies are currently popular (Chimonaki et al., 2019; Sadasivam et al., 2016; West & Bhattacharya, 2016). Finding instances of data that do not follow the expected course of events is a challenge that is addressed by the broad discipline of anomaly detection (Chandola et al., 2009).

Once a researcher decides to select the best anomaly detection technique to use, the availability of labeled data is a crucial factor. Any data set's label for each data instance shows whether the instance is anomalous. The supervised anomaly identification method faces a considerable challenge because the anomalous class is often infrequent in comparison to the normal class. It can be difficult to find precise descriptors that accurately describe the anomalous class. Thus, according to the existence of labelling a technique can be supervised, semi-supervised, or unsupervised.

The assumption made by supervised anomaly detection approaches is that the data collection being used contains categorized instances that belong to either a normal or anomalous class. Many of these techniques offer a normal and abnormal forecasting framework that may be applied to categorize newly acquired data. Whilst the only occurrences in the data set that are labeled, according to semi-supervised anomaly detection approaches, are those that correspond to the normal class. Since it is difficult to acquire data that is typical of all anomalous behavior, there is a relatively small range of approaches and literature developed to work with data sets when only the anomalous class is present. The last approach is the unsupervised anomaly detection in which no labels are required. Unsupervised methods implicitly assume that in the test set of the data, anomalous occurrences are far less common than regular events; otherwise, the false predictions of these methods will be larger than anticipated.

A range of statistical and machine learning methods are used to create financial misconduct detection models (Abbasi et al., 2012; Goel & Gangolly, 2012). According to Zhou *et al.* (2018), most fraud-detection systems use at least one supervised learning method such as artificial neural networks (ANNs), decision trees (DTs), support vector machines (SVMs), accounting regression (LR), random forest (RF), k-nearest neighbor (kNN), ensembles (bagging, adaboost, etc.) and multicriteria methods (for instance UTADIS). However, the most used one is the logistic regression model. For instance, Kim *et al. (*2016) noted that logistic regression models have been utilized in 10 of the 15 financial statement fraud detection studies that were considered in the literature review.

Support vector machines, on the other hand, are a more preferred option in more recent predictive studies because of their excellent classification accuracy (Abbasi et al., 2012; Cecchini et al., 2010; Perols, 2011). Perols (2011) for instance, examines the effectiveness of six well-known statistical and machine learning methods for spotting financial statement fraud under various assumptions on the costs of misclassification and the ratios of fraud firms to non-fraud enterprises. He demonstrates that support vector machines and logistic regression outperform artificial neural networks, bagging, C4.5, and stacking. Dyck *et al.* 2010) demonstrated that when it comes to accurately recognizing fraudulent organizations, ensemble methods outperform other machine learning techniques. On the contrary, Mohammadi *et al.* (2020) found that ANNs perform well in comparison to bayesian networks (BN), discriminant analysis (DA), LR, and SVMs.

Using data from the Greek market, Spathis *et al.* (2002) compared the UTADIS classification approach against two multivariate statistical techniques, namely discriminant and logit analysis. They found that the UTADIS multicriteria method outperforms the two statistical methodologies. Moreover, in the same study they concluded that financial ratios such as total debt to total assets ratio, inventories to sales ratio, are significant “red flags” to detect fraud. Kou *et al.* (2004) claim that classes of models may be developed employing techniques includes LR, DTs, SVMs, ANNs, and naive Bayes to identify new fraudulent cases before human experts are able to observe them. Liou (2008) compared the performance of LR to that of ANNs and DTs, concluding that LR was the most successful of the three data mining techniques.

Gaganis (2009) developed 10 distinct classification models for the detection of FFS using financial and non-financial data, including classical classification methods like logistic regression, discriminant analysis, and the k-nearest neighbors, artificial intelligence methodologies like SVMs and neural networks, as well as multicriteria methodologies. According to Gaganis (2009), LR is easy to use and has very limited computational requirements compared to complex data mining approaches. In contrast, Kotsiantis *et al.* (2006) observed that LR performed poorly when compared to other approaches. The effectiveness of decision trees, neural networks, and Bayesian networks in the identification of fraudulent financial statements are also investigated by *Kirkos et al.* (2007), who demonstrated that the Bayesian network model performed better than the others. Multi-class classification is possible using Bayesian network models, which also directly give class probabilities estimates (important for cost-sensitive learning later) and function well even with severely skewed class distributions (Leong, 2015).

Additionally, recent research has started to focus on linguistic factors taken from texts from corporate reports that could contain deceptive remarks (Kydros et al., 2022). Because financial statement fraud is related with enormous amounts of textual data, text mining has been reported to be very beneficial for fraud detection (West & Bhattacharya, 2016). To compare analysts' reports, Huang *et al.* (2018) employed a topic modeling methodology from computational linguistic research. In the basic idea of fraud detection within texts, machine learning (Sánchez-Aguayo et al., 2021; Xia et al., 2020), sentiment analysis (Kauffmann et al., 2019), or meta-synthesis qualitative approaches have been used. More precisely, Bonsall *et al.* (2017) discover that yearly reports have grown in length over time, the percentage of unstructured language has increased dramatically, and there is still a significant quantity of hidden information to be discovered. Zhang *et al.* (2021) converted texts into digital data and used text analysis in conjunction with machine learning and vector index to find those financial issues that were buried.

By creating a dictionary, the Bag-of-Words (BoW) model depicts files as an N-dimensional vector using words as its unit of measurement. The amount of each position represents the word frequency, while the number of digits of the vector represents the size of the dictionary. Based on the findings of Skillicorn and Purda (2012), they may create the financial industry's BoW vector, screen and keep high-contribution words using random forest, and then use SVM to forecast the likelihood of fraud. In terms of the true positive rate (fraudulent detection), Hajek and Henriques (2017) discovered that ensemble approaches performed better than the other methods. Contrarily, for non-fraudulent cases, the Bayesian belief network (BBN) provided the best performance (true negative rate). This discovery for true positive rate and true negative rate is significant because it could lead to the derivation of interpretable "green flag" values (for which fraud is probably not present), which could help auditors make decisions during client selection or audit preparation.

With a focus on the types of variables utilized as inputs, the evaluated methodologies, and the method that was shown to perform best in each study, Table 1 summarizes indicative studies on the identification of FFS. It seems that a 43% of these papers use different types of neural networks. Additionally, methods which were used in papers with best performance metrics, are the random forest algorithm (Whiting et al., 2012; Wyrobek, 2020), support vector machines (Y. J. Chen et al., 2019; Pai et al., 2011), multicriteria methods (Hooda et al., 2019), ensembles (Hooda et al., 2019; Kotsiantis et al., 2006), Bayesian belief networks (Kirkos et al., 2007) and logistic regression (Pazarskis et al., 2017). The vast majority of academic research use financial variables as inputs (Omar et al., 2017; Wyrobek, 2020; Yang & Jiang, 2020). Several studies have combined financial variables with corporate governance and audit control variables (Kim et al., 2016; C. C. Lin et al., 2015), whereas some other studies use linguistic variables, such as the papers of Craja *et al.* (2020) and Throckmorton *et al.* (2015).

Table 1: List of indicative recent studies on the identification of falsified financial statements

Authors	Variables	Methods	Country	Best method
Zhang et al. (2021)	LNG	BOW, NB, RF, SVM	CHN	SVM
Craja et al. (2020)	FIN, LNG	HAN	USA	HAN
Wyrobek (2020)	FIN	LR, DT, RF, ANN, NB, DA, SVM	USA	RF
Hooda et al. (2019)	FIN	MCTOPE, RF, ABOOST, J48, PROBIT, BN, DT, DS, NB, SVM	IND	MCTOPE
Hajek & Henriques (2017)	FIN	Fuzzy	USA	Fuzzy
Chen et al. (2019)	LNG	SVM, LR, KN, DT	TWN	SVM
Dong et al. (2018)	LNG	SVM, ANN, DT, LR	USA	SVM
Jofre & Gerlach (2018)	FIN	DA, LR, ABOOST, Boosted trees, RF	USA	DA Boosted trees
Omar et al. (2017)	FIN	ANN	MYS	ANN
Pazarskis et al. (2017)	FIN	LR	GRC	LR
Kim et al. (2016)	FIN, CG	LR, SVM, BN	USA	LR
Lin et al. (2015)	FIN, CG, AUD	ANN, DT, LR	TWN	ANN
Throckmorton et al. (2015)	FIN, LNG	GLRT, NB, LR, KN	USA	GLRT
Whiting et al. (2012)	FIN	PROBIT, LR, PAE, STGR, RL	USA	RF
Ravisankar et al. (2011)	FIN, AUD	SVM, GP, PNN, GMDH, LR	CHN	PNN
Pai et al. (2011)	FIN, CG	SVM	TWN	SVM
Krambia-Kapardis et al. (2010)	AUD, CG	ANN	CYP	ANN
Kotsiantis et al. (2006)	FIN	KN, SVM, Boosting	GRC	Boosting
Cecchini et al. (2010)	FIN	SVM	USA	SVM
Gaganis (2009)	FIN, CG, AUD	DA, LR, SVM, PNN, ANN, KN, MHDIS, UTADIS	GRC	PNN UTADIS
Gaganis et al. (2007)	FIN	KN, DA, LR	GBR	KN
Kirkos et al. (2007)	FIN	DT, ANN, BN	GRC	BBN
Spathis et al. (2002)	FIN	UTADIS, DA, LR	GRC	UTADIS
Carcello & Nagy (2004)	FIN, CG, AUD	LR	USA	LR
Fanning & Cogger (1998)	FIN, CG, AUD	LDA, QDA, LR, NN	USA	NN

Notes: Variables' Category: Financial (FIN), Corporate Governance (CG), Audit (AUD), Linguistic (LNG)
Methods: ANN=Artificial Neural Networks, PNN=Probabilistic Neural Networks, MLP=Multi-layer Perceptron classifier, LR=Logistic Regression, SVM=Support Vector Machines, GA=Genetic Algorithms, GP=Genetic Programming, MCTOPE=Multicriteria TOPSIS-based Ensemble, STGR=Stochastic gradient boosting, RL=Rule ensembles, DS=Decision Stump, RF=Random Forest, DT=Decision Trees, GMDH= Group Method of Data Handling, GLRT=Generalized likelihood ratio test, NB=Naïve Bayes, BN=Bayesian Networks, DA=Discriminant Analysis, LDA=Linear Discriminant Analysis, QDA=Quadratic Discriminant Analysis, KN=k-Nearest Neighbors, HAN=hierarchical attention network, PAE=Partially adaptive estimators, BOW=Bag of words

From this brief overview of the existing literature, it is evident that prior studies have relied on binary classification approaches for model development. In contrast, the present study explores the effectiveness of multi-label approaches, which enable the simultaneous consideration of different forms of financial statement falsification, thus providing more guidance on the description of each fraud instance.

2.3 Applied variables

Indicators which suggest the existence of financial fraud, are derived from various sources, such as balance sheets, income and cash flow statements, the report of retained earnings (Ravisankar et al., 2011), restatements (Firth et al., 2011), social media (Dong et al., 2018), online data (Van Der Aalst et al., 2011), auditors' reports (Gray & Debreceny, 2014; Spathis et al., 2002) etc. In general, fraud detection is carried out by (i) using accounting data from financial statements, (ii) using financial ratios extracted from them, and (iii) using non-financial indicators related to managerial decisions or corporate governance practices (W. S. Albrecht et al., 2008; Fernández-Gámez et al., 2016).

Financial variables constitute an important tool to detect financial fraud. They can be divided into several categories such as liquidity, solvency, profitability, and efficiency (Ravisankar et al., 2011). The inputs can typically be further subdivided into firm size, corporate reputation, profitability ratios, asset structure, business scenario, liquidity, leverage, and market value ratios (Hajek & Henriques, 2017).

The study of Mohammadi *et al.* (2020) also assesses the impact of a number of additional potential explanatory factors, including nonfinancial indicators, discretionary accruals measurements, and red flags highlighted in auditing standards (Beneish, 1999; Brazel et al., 2009). The main study on financial reporting frauds by Feroz *et al.* (1991) found that dishonest businesses nearly always inflate their inventories and accounts receivable. In a later analysis of the distinctions between fraudulent and non-fraudulent businesses, Beneish (1999) incorporated the volume of receivables and the length of the debt collection process as distinct variables for the two groups. When Beneish (1999) was constructing his suggested model, he highlighted accruals (capital change in non-working capital plus dis-projection) as a viable fraud detection index along with debt collection prediction, gross profit margin, asset growth index, and sales growth index. Employee productivity, debt to equity, sales to total assets, inventory to sales, return on equity, return on sales, liabilities to interest expenses, and assets to liabilities were chosen as inputs in this Mohammadi's study (Mohammadi et al., 2020).

Accruals quality-related variables have demonstrated discriminating potential, including changes in inventories, industry-level, and organizational context-based metrics (Kim et al., 2016). The same study also found that market indicators like the short-term interest ratio and the firm-efficiency measure are helpful in identifying false statements and intentional fraud. According to Wyrobek (2020), the cash flows from financing activities, terminated operations, investment income, interest income, other equity, short-term investments, unrealized profits and losses, and other factors were higher for fraudulent enterprises. Excise taxes (income statement), interest payments, other assets, restricted cash, pension liability adjustments, total receivables, and plant assets were all lower for fraudulent corporations than for "fair" ones. This shows that unfair organizations have a relatively high focus on financial operations and have less value invested in plant assets.

Pazarskis *et al.* (2017) found that profitability ratios such as profit or loss before taxes/ total assets, net income / stockholders' equity (ROE), and net income / total assets (ROA) are variables that significantly influence fraud detection and non-fraudulent firms show better results in their financial reporting. Moreover, the same situation exists for liquidity ratios, such as current assets / current liabilities and (current assets-stocks) / current liabilities. In the same research, the likelihood of falsification was found to be significantly influenced by the capital structure ratios of shareholders' funds / total assets, long term debt / shareholders' funds, shareholders' funds / (long-term liabilities + risk provisions & expenses), and earnings before interest and taxes / interest expenses and, by activity ratio such as net sales / stocks. Companies that don't falsify their rates perform better in these circumstances than those who do.

Moreover, for Whiting *et al.* (2012) the top three features for the discovery of fraud cases were inventory growth, asset turnover and cash flow earnings difference. If inventory growth exceeds 1.253%, fraud is more plausible. Additionally, since net sales are typically fraudulently increased quickly while revenue fraud is being committed, it would be expected that this ratio's coefficient would be positive Spathis *et al.* (2002).

Furthermore, high positive figures are more suggestive of financial statement manipulation because cash flows are typically unaffected by financial statement manipulations (Whiting et al., 2012).

Non-financial factors have been taken into consideration to enhance the reliability of financial statement fraud indicators (Yu, Gao, Zhang, & Liu, 2021). Since scams are carried out by people, they are intrinsically tied to how people behave. A new viewpoint on scam detection may result from a comprehension of criminal motivations and the psychological and personality traits that drive them to violate moral boundaries (Sayal & Singh, 2020). Sánchez *et al.* (2021) and Skousen *et al.* (2009) studied parameters considering the fraud triangle theory (pressure, opportunity, rationalization), while researchers such as (Ozcelik, 2020; Supri, Rura, & Pontoh, 2018; Sunardi & Nuryatno, 2018), added one more element; the ability of fraudster to commit financial scam, using the fraud diamond theory.

Financial targets, pressure from outside sources, and financial stability were all found to have a strong favorable effect on fraudulent financial reports, according to the findings (Supri et al., 2018). In the same study, the proxy factor of opportunity, namely the business sector, has no substantial effect on fake financial statements, but it leads to a strong negative outcome on fraudulent financial documents for the proxy of supervisory efficacy. Supri *et al.* (2018) found that auditor changes, which are a proxy for rationalization, are strong indicators for falsified financial accounts.

Several other features of corporate governance can be included in the fraud diamond theory presenting satisfactory results in detecting FFS, such as characteristics of boards may be categorized as capability or opportunity. According to results reported by Beasley (1996), the size of the board and a greater degree of accounting fraud are directly and positively associated. In this vein, Klein (2002) and Peasnell *et al.* (2005) reasoned that the greater effectiveness of the board's supervisory position and their greater capacity to distribute the tasks among a larger number of members could be used to explain the existence of a favorable relationship between the size of the board and the tendency to falsify the statements. The activity indicator, which is determined by the frequency of meetings and the concentration of duties in the company's chief director, is also related to the board. For their part, Xie *et al.* (2003) argued that the audit committee's and board's increased activity minimizes manipulative behaviors.

The concurrent holding of the roles of company president and director may have an impact on the veracity of the accounting data on the concentration of power due to their potential influence over the enterprise's control mechanisms (Fernández-Gámez, Garca-Lagos, and Sánchez-Serrano, 2016). The first study to draw this conclusion empirically was presented by (Beasley, 1996). The highest percentage of independent advisors was found in firms free from manipulation. Similar conclusions were reached by Peasnell *et al*. (2005) and Klein (2002), who obtained data indicating a lower level of fabrication due to the presence of independent advisors. More lately, Marra *et al.* (2011) argued that the audit committee's and board's independence is crucial in avoiding manipulative tactics. Their statements support the notion that the qualities of corporate governance are more closely related to the accuracy of the financial reporting.

According to Lin *et al.* (2015), the greatest importance in the fraud triangle dimension is pressure/incentives, followed by opportunity, and the lowest impact is attitude/rationalization. The top five factors in each area are, in descending order of significance: poor performance, the requirement for outside funding, liquidity issues, an absence of board oversight, and competition or market saturation. Moreover, they revealed that the only two dimensions in the top five categories are pressure/incentive and opportunity dimension. This discrepancy should serve as a caution to accountants and recipients of financial statements to pay more attention to the attitude/rationalization dimension, especially if the company frequently produces financial restatements. Additionally, they disclosed that the attitude/rationalization dimension should be given extra attention, particularly when the company manipulates the revenue and has an elevated rate of CFO turnover.

Other researchers deal with detecting fraud using textual analysis approaches. The choice of linguistic elements in the study of Craja *et al.* (2020) was based on earlier research that revealed a number of fraudulent agent behaviors, such as an increased propensity for employing phrases that convey negative (Throckmorton et al., 2015), a lack of issues in managing that implies absence of assurance and results in reporting with less certitude (Larcker & Zakolyukina, 2012), or a mean of three times more positive comments and four times more negative sentiment compared to genuine reports (Goel & Gangolly, 2012). Because previous research indicated that reports created by misleading firms have lower readability, measures of complexity and legibility such as the median sentence length, the percentage of compound words, and the fog index are also included (Humpherys et al., 2011).

2.4 Discussion on literature review

Regulating, commercial, and academic institutions have been focused on the enormous losses brought on by financial fraud for an extended period. Crises such pandemic viruses (COVID-19), wars, environmental disasters and so on, are particularly concerning because it unpredictably shocks the global financial system with substantial increases of prices and accelerates the tendency of falsification the financial reports in order to enhance the economic situation of the companies. Once more, occupational financial fraud is in the foreground, requiring immediate actions with modernized approaches to address the detection of FFS.

Traditional approaches, such as manual auditing and inspections, are expensive, inaccurate, and time-consuming for detecting such bogus statements. Moreover, since fraudsters are highly educated, well aware of financial matters, and capable of creating numerous ways to fabricate the statements, they can outperform auditors in several cases. Thus, auditors trying to analyze several financial statements with the use of intelligent methods. Machine learning technology, a subset of artificial intelligence, has been demonstrated to be one of the most effective methods for detecting fraud. By using data mining techniques, fraud can be identified and quickly dealt with to reduce costs. Through the use of data mining tools, it is possible to look for patterns and identify bogus claims among millions of statement papers.

The classification of financial statements as fraudulent or not is the aim of financial statement fraud detection (FSFD). To anticipate fraudulent remarks, supervised and unsupervised algorithms have been applied. The most common approach to discover false financial statements relies on the development of classification models (West & Bhattacharya, 2016). The majority of FSFD practices uses supervised machine learning techniques, which typically have a two-face design (Albizri et al., 2019; Dutta et al., 2017). In the first stage, a model is trained using a dataset with feature vectors and class labels. Following that, test samples are categorized implementing the trained model in the following stage. The way in which and the extent to which feature vectors are derived from the input data affects how well data mining algorithms perform. When the wrong features are chosen, they may result in ineffective performance and irrelevant or meaningless features.

Moreover, Al-Hashedi and Magalingam (2021) found that SVMs were the most common data mining method, followed by LR while Ashtiani and Raahemi (2021) agreed with the previous paper and added that DTs, BN, ANNs, kNN were also popular algorithms in detecting FFS. In terms of ensemble approaches, RF, bagging, boosting, and stacking had also been frequently used in FSFD (Ashtiani & Raahemi, 2021). Furthermore, contemporary research has revealed data regarding the relative efficiency of several classification systems. ANN outperforms logistic regression and other methods when the dataset is balanced. In more recent research, which evaluated the effectiveness of additional classification algorithms in analogous balancing settings, meta-classifiers were found to have the best classification accuracy (Mohammadi et al., 2020).

Most academic researchers have employed financial ratios as inputs in algorithms which have been found to be useful for FSFD. Nevertheless Mohammadi *et al.* (2020) noted in that many of these ratios used to measure financial statement fraud in the literature are inextricably linked to a particular kind of fraud. For instance, an abnormal increase in revenue is a possible indicator of revenue fraud, whereas an abnormally low allowance for questionable accounts is a possible indicator of expense fraud. These variables may be helpful in revealing information about a particular fraud type, but they are less likely to reveal numerous fraud types. Furthermore, by merging all fraud types into a single binary classification problem, the algorithms for classification concentrate on discovering patterns that are shared by all types. Given the diversity of fraud types, it may be challenging to spot these patterns.

Moreover, Mohammadi *et al.* (2020) mentioned the three issues that make it difficult for data analytics to be useful in predicting fraud. The first challenge with FSFD is finding a needle in a haystack. That is, it is challenging to predict fraud since fraud firms are relatively uncommon compared to organizations that do not engage in fraud actions (Bell & Carcello, 2000). The issue of data dimensionality complicates fraud prediction in a second way (Bellman, 1961). Whiting *et al.* (2012) found that the number of fraud observations in comparison to the vast number of explanatory variables discovered in the documenting fraudulent cases can lead to overfitted prediction models that underperform when forecasting fresh data. Third, earlier research typically views all frauds as uniformly occurring occurrences. Because prediction models must find patterns that are prevalent across several fraud types, this might make fraud prediction more challenging.

Falsifying financial statements can be conducted to make a company appear more profitable than it is, boost stock prices, evade paying taxes, or obtain an external financing. The fraud triangle and diamond theories are paradigms used in auditing to show the driving forces behind someone's choice to commit fraud. According to Gupta and Gill (2012), the likelihood of fraud rises when there are incentives, such as the requirement to attain a result or make up losses. The company will face pressures or temptations to engage in dishonest behavior. Additionally, the absence of inspections or ineffective controls creates a favorable environment for fraud. When a fraudster attempts to rationalize their fraudulent behavior, it may be influenced by other people and circumstances (rationalization). Fraudsters tend to remain in their moral comfort zone, according to Dbouk and Zaarour (2017).

As a result, the fraudster prepares to perform the first deception by internally trying to justify and defend the fraudulent activity. According to the same paper, rationalization happens when the perpetrator of the fraud creates a defense because they do not want to be identified as an offender. Because of this circumstance, scammers can view their predicament as a special exemption rather than as criminal activity. Due to the inclusion of a fourth factor, capacity, the well-known fraud triangle has been transformed into the fraud diamond (Wolfe and Hermanson, 2004). According to the essay, a person's abilities and personality attributes directly affect the likelihood of fraud. "Opportunity opens the door to fraud, and motivation -once for instance, pressure is arisen- and reasoning can lead a person toward it," say Wolfe and Hermanson. The person must, however, be able to recognize an opportunity when a chance appears doorway and be able to take it by passing through it once more (Rudewicz, 2011).

Furthermore, according to the literature study, the challenge of combining financial and linguistic data for FSFD has been taken on by Hajek and Henriques (2017) and Throckmorton *et al.* (2015). It is crucial to employ more powerful text processing techniques given the managerial attempts to hide bad news by using specific phrasing

(Humpherys et al., 2011) and by producing reports that are harder to interpret (Loughran & Mcdonald, 2014). In fact, due to the abundance of textual data comprising administration notes, text mining algorithms have been receiving considerable interest in FSFD and related financial decision-making challenges (Kumar & Ravi, 2016). Because senior management has been involved in the majority of significant financial statement frauds and has the chance, capacity, and incentive to perpetrate fraud, analyzing managerial comments is particularly crucial (Hajek & Henriques, 2017).

Nevertheless, identifying fraud requires comprehensive field expertise, which is why external auditors are increasingly responsible for it, prior research has shown that auditors regularly fail to uncover significant frauds. In addition, reported conflicts of interest have restored the confidence of audit companies. The improved detection capabilities of the systems are particularly important for shareholders (to make better choices), auditing businesses (to conduct both client acceptance and routine audits), and regulatory authorities (to better focus their investigative efforts).

Complex FSFD systems have been developed to aid the stakeholders' decision-making processes. Red flags are early alert indications provided by these systems. The latest technologies must be used to combat fraud because fraudulent activities are constantly evolving and becoming more complicated. Equity markets, bond markets, and bank credit markets, for example, rely heavily on business and financial data as well as extended verification from accounting analysis results to monitor and audit fraud and moreover, all those who detect FFS should be more creative than the fraudsters in order to discover the fraudulent actions.

Chapter 3. Empirical Setting

This section describes the details of the empirical analysis. The presentation starts in subsection 3.1 with the description of the dataset, the discussion of the types of auditors' comments that are used in this study to identify firms with FFS, and the analysis of common patterns between the auditors' comments. Subsection 3.2 presents the financial ratios and the variables chosen from the fraud diamond theory derived all from the issued financial statements. Moreover, in 3.2.1 the descriptive statistics which were applied on selected ratios and corporate variables, are presented. Additionally, this section discusses the details of the adopted muti-label classification approach, starting with a brief discussion of the selected binary classifiers (subsection 3.3.1-3.3.4), then proceeding with the presentation of the muti-label classification methodology (subsection 3.3.5), and closing with the performance metrics used to evaluate the discriminating power of the methods used in the analysis (subsection 3.4). Finally, in subsection 3.5 the results of this thesis are presented.

3.1 The dataset

The dataset used in this study includes information from 133 Greek companies (except for those which are in financial sector such as banks etc.) obtained from the annual financial statements published in the Athens Stock Exchange from 2014 to 2019. Financial statements are records that provide information on an organization's operations and financial performance, including income, costs, profits, loans, conceivable future issues, and managerial observations on operational performance (Glancy & Yadav, 2011). Every company is required to release its financial accounts on a quarterly and annual basis. A company's performance can be determined by examining at its financial statements. Financial reports are used by creditors, market analysts, and investors to research and evaluate a company's financial standing and earning potential. The four sections that make up financial statements are the income statement, balance sheet, cash flow statement, and explanatory notes.

A significant difficulty is that all information was hand-collected reducing the flexibility to changes and being an extremely time-consuming process. Regarding the handling of missing values, they were just excluded from the dataset because the number of them was extremely low. After cleaning the dataset, the sample consists of 58 companies from the trade sector, 51 manufacturing companies, and 24 companies from the services sector. The whole sample is an unbalanced panel of 752 firm-year observations, including 560 non-falsified cases and 192 cases with falsified financial statements (FFS) (Table 2).

Following earlier studies (Kotsiantis et al., 2006; Spathis et al., 2002), the classification of a financial statement as falsified or non-falsified, was based on evidence derived from the opinion of auditors published in the financial statement of each company. More specifically, the following comments by the auditors are taken into consideration to identify cases of falsification:

1. inclusion of serious doubts about the continuation of the company's activity (AC1),
2. negative equity or equity that is less than half of the share capital (AC2),
3. the existence of court proceedings or some other evidence which can alter the financial situation of the company (AC3),
4. re-financing uncertainty (AC4),
5. total short-term liabilities more than total current assets (AC5),
6. lack of necessary notifications or different recording of data, for instance long-term liabilities in short-term debt (AC6).

Financial statements with at least one comment by the auditors (AUCRITER) on the above six points, are considered as falsified.

Tables 2 and 3 present the composition of the sample and provide details on the number of falsified cases by sector and year, whereas Table 4 provides information on the comments of the auditors for the falsified cases, including the frequencies and the Jaccard similarity index (Leskovec et al., 2020) for the associations between each pair of comments. The Jaccard similarity index for a pair of comments $\{\mathrm{AC}_i, \mathrm{AC}_j\}$ is defined as $J(\mathrm{AC}_i, \mathrm{AC}_j) = |\mathrm{AC}_i \cap \mathrm{AC}_j| / |\mathrm{AC}_i \cup \mathrm{AC}_j|$, where $|\mathrm{AC}_i \cap \mathrm{AC}_j|$ is the number of cases having both comments AC_i and AC_j, whereas $|\mathrm{AC}_i \cup \mathrm{AC}_j|$ denotes the number of cases having at least one of the two comments. The analysis of the associations between pairs of comments enables the identification of common patterns between the different types of auditors' comments. The existence of such common patterns facilitates the simplification of the modeling process, making it easier to deploy, update, and interpret the resulting models. For instance, more general models can be developed to describe similar types of auditors' comments, avoiding the construction and use of too specialized models. Thus, the clustering of the FFS cases enables the construction of efficient analytical models through the approaches that will be described in the section 4, given that some categories of auditors' opinions (e.g., AC3, AC4, AC6) are represented by a rather small number of observations in the sample.

Table 2: Non-falsified and falsified companies by year

	2014	2015	2016	2017	2018	2019	Total
No. of nFFS	97	94	93	92	94	90	560
No. of FFS	34	37	38	35	25	23	192
Freq. of FFS	26.0%	28.2%	29.0%	27.6%	21.0%	20.4%	25.5%

Table 3: Sample composition by sector

Sector	Companies	Observations	FFS observations	Freq. of FFS
Manufacturing	51	300	73	24.3%
Trade	58	323	70	21.7%
Services	24	129	49	38.0%
Total	133	752	192	25.5%

Table 4: Frequencies and Jaccard similarity indices for the auditors' comments (in %)

	AC1	AC2	AC3	AC4	AC5	AC6
AC1		62.2	33.0	31.7	53.1	23.0
AC2			37.3	29.3	45.1	28.0
AC3				10.1	16.2	40.0
AC4					49.1	12.4
AC5						16.9
Number	172	107	66	65	99	53
Frequency	22.9%	14.2%	8.8%	8.6%	13.2%	7.0%

The results of Table 2 indicate that the frequency of falsified cases increased during the period 2014-2016, reaching 29% in 2016, but it steadily decreased in the later years, dropping down to 20.4% in 2019. Regarding the sectors (Table 3), falsification is most frequent in the services sector and lowest in the trade sector. As far as the comments of the auditors are concerned (Table 4), in most cases, these relate to the accuracy of the financial accounts (AC1), as well as the equity and the short-term liabilities of the firms (AC2, AC5). Finally, the examination of the associations between different types of auditors' comments in Table 4, reveals a strong connection between AC1 and AC2 (i.e., these types of comments often appear together). Other strong associations with Jaccard similarity index at least 40%, include the pairs {AC1, AC5}, {AC2, AC5}, {AC4, AC5}, and {AC3, AC6}.

The associations between the auditors' comments were further explored by clustering them into groups. The dendrogram derived through the application of agglomerative hierarchical clustering, is shown in Figure 3. According to these results, three main clusters can be identified. The first consists of AC12={AC1, AC2} and can be described as a cluster of concerns regarding the manipulation of material financial accounts and the solvency of the company. The second cluster involves AC45={AC4, AC5}, which relates to concerns about a firm's financing and liquidity. The final cluster includes the pair AC36={AC3, AC6}, which involves concerns of the auditors on uncertainties due to legal proceedings and lack of appropriate notifications on some important aspects of the financial position of a company. Based on these results, the subsequent analysis is based on the distinction between FFS and nFFS cases, with the former group further described through the above three clusters of auditors' opinions.

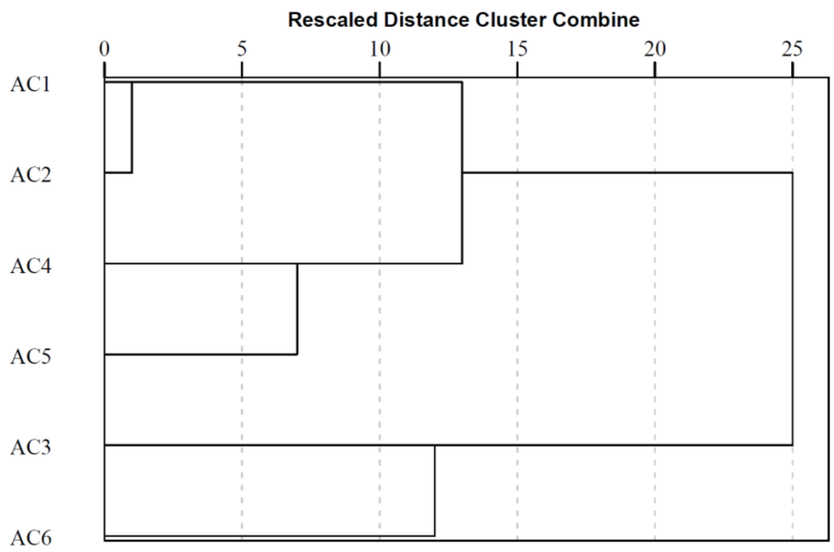

Figure 3: Dendrogram of the hierarchical clustering of the auditors' comments

3.2 Financial ratios and variables from fraud diamond theory

The variables employed in the analysis were selected following the existing literature as well as through various statistical tests (e.g., t-test, Mann-Whitney, χ^2 test, correlation analysis), which helped examine the statistical significance of the variables in describing the falsified and non-falsified statements in the sample. As it was already mentioned in 3.1, the missing values (due to the small number) were excluded from the sample. Moreover, the variables were checked for outliers, which were winsorized at bounds defined through appropriate percentiles of the data (e.g., 1% and 99% percentiles). In accordance with the fraud diamond theory, the selected variables are categorized in four main categories, as shown in Table 5: (a) capability, (b) opportunity, (c) pressure, and (d) rationalization. The following paragraphs elaborate on the selection of the variables in each category. Descriptive statistics are presented and discussed in sub-section 3.2.1.

Table 5: Selected variables

Variable	Description	Category
DIRCHANGE	1 if there the chairman of the board of directors changed in year t, otherwise 0	Capability
SHARE&BOARD	1 if the members of board of directors are also shareholders, otherwise 0	Capability
FAMILY	Share of members of the board of directors with the same surname	Capability
BIG4	1 if the audit company is one of the Big 4, otherwise 0	Opportunity
SECTOR	The sector in which a company belongs to	Opportunity/Pressure
COSTSOLD/AP	Cost of sales / Accounts Payable	Pressure
CA/CL	Current Assets / Current Liabilities	Pressure
P/BV	Stock price / Book value per share	Pressure
AMORT/FA	Amortization / Fixed assets	Pressure
EBIT/TA	Earnings before interest & taxes / Total assets	Pressure
LN(TA)	Natural logarithm of total assets	Pressure
TL/TA	Total liabilities / Total assets	Pressure
SALES/TA	Total Sales / Total assets	Pressure
AUDSWI	1 if the audit company changed in year t, otherwise 0	Rationalization

Capability

The capability dimension refers to the management and governance characteristics of the companies. Various studies (Fanning & Cogger, 1998; Perols, 2011) have examined variables related, for example, to auditor changes, directors' changes, the roles of CEOs and CFOs (Firth et al., 2011), as well as whether a company is family owned or not (Pai et al., 2011). Based on the existing literature and considering the information reported in the financial statements of Greek companies, the following characteristics were employed in this study research:

A. A dummy variable indicating whether there was a change of the chairman of the board of directors in comparison with the previous year (DIRCHANGE). A change of the chairman is considered as a fraud risk factor because it could indicate an attempt to cover up fraudulent activities from the previous management. On the other hand, a long-term stay in this position, may provide the right to postpone several appropriate audits increasing the possibility of a fraudulent behavior.

B. A dummy variable indicative whether there exist members of the board of directors who are also shareholders (SHARE&BOARD). Previous studies, such as the one of Lin *et al.* (2015), have noted that the existence of board members that are also shareholders, may increase the risk of financial fraud, as such board members could act on their own interest and use their power to facilitate the falsification of a company's financial statements in order to misinform investors and gain profits from an increasing share price.

C. The percentage of board members that have the same surname, as a proxy of whether a firm is family owned or not (FAMILY). Family-owned firms may follow questionable audit controls increasing the possibility of engaging in fraudulent actions.

Opportunity

Effective monitoring (e.g., by an independent audit committee) can minimize the risk of fraudulent behavior. Evidence from the literature indicates that audit quality is improved when the audit company is one of the top audit firms (Behn et al., 2008; Verriest, 2014). This is because large audit firms seek to maintain their reputation and employ more experienced auditors. Nowadays, the four major auditing companies are known as "BIG4" (Deloitte, Ernst and Young, KPMG, PricewaterhouseCoopers).

Although there is a significant trust in BIG4 audit companies, there is a possibility that these firms may have the power/knowledge/experts to avoid controls from authorities. An example of this opinion is the involvement of Arthur Andersen (BIG5) in Enron's scandal. Moreover, studying from the history BIG4 can also fail to detect FFS (Lehman Brothers-Ernst and Young).

Moreover, according to ACFE (2022), the frequency and impact of financial fraud events, vary by sector, with real estate having the costliest fraudulent cases, followed by trade, transportation and warehousing, manufacturing, and utilities. In this research, we use the variable SECTOR, which refers to the sector in each the companies operate (services, trade, manufacturing). This indicator is included in the opportunity category, because in every sector there are common patterns of fraudulent practices, such as the same guidelines or suppliers, etc.

Pressure

According to SAS no. 99, managers feel pressure to proceed in fraudulent actions when financial stability and/or profitability is threatened by the economic conjuncture, sector or operating conditions (Skousen et al., 2009). Due to the pressure caused by the business and economic environment, fraud may occur to meet the high earnings expectations by investors and shareholders. Furthermore, executives may use fraudulent financial reporting to meet the requirements of debt contracts. This shows that the probability of fraudulent reports at high debt levels has increased. Moreover, financial and operating targets may increase the pressure on management executives. Ratios, such as COSTSOLD/AP, EBIT/TA, SALES/TA, CA/CL, TL/TA and P/BV are used in the present study to identify anomalies in the financial position of the firms that could be FFS signals.

In addition, companies try to increase the value of their assets, either through real value increases in asset items such as asset acquisition, development of existing assets, or by active growth through valuation of existing assets. Often, the acquisition of new assets and the improvement of existing assets, are reported at values different from the real ones (Ozcelik, 2020). Variables such as LN(TA) or AMORT/FA, can be used for the identification of falsification in asset valuations.

Rationalization

Craswell, (1988) claims that qualified opinions are linked to later auditor changes. Frequent auditor changes may indicate a higher risk of fraudulent reporting. Given that many fraudulent actions are identified years after the fraud occurs, firms that often change auditing companies does not allow auditors to perform thorough audits. Moreover, there is an opinion which states that a long-term stay of an auditor firm may lead to a close relationship with the company, and some omits in appropriate audits. To account for this factor, a dummy indicator is employed regarding change of auditors (AUDSWI).

3.2.1 Descriptive statistics on selected variables

Table 6 presents the means of the variables used in the analysis for each of the two main classes of observations (FFS versus nFFS). It should be noted that for the binary categorical indicators (BIG4, AUDSWI, DIRCHANGE, SHARE&BOARD), the means represent the frequencies in each class. For the sector variable, the means are omitted as they do not apply to this categorical indicator. The table also presents the results (p-values) of statistical tests regarding the discriminating power of the variables. For the categorical indicators, the χ^2 test is employed, whereas for the quantitative attributes we use both the t test as well as the non-parametric Mann-Whitney U test. Tables A1 and A2 in the appendix provide detailed results for all types of auditors' comments, beyond the FFS/nFFS distinction of Table 6.

Table 6: Means of the indicators for the main FFS categories and significance tests (p-values)

Indicator	Category	nFFS	FFS	χ^2	t-test	Mann-Whitney
DIRCHANGE	Capability	0.073	0.031	0.038		
SHARE&BOARD	Capability	0.755	0.750	0.882		
FAMILY	Capability	0.274	0.306		0.025	0.121
BIG4	Opportunity	0.150	0.167	0.581		
SECTOR	Opportunity/Pressure	–	–	0.001		
P/BV	Pressure	5.445	1.997		0.132	0.000
COSTSOLD/AP	Pressure	6.182	3.513		0.000	0.000
CA/CL	Pressure	2.184	0.746		0.000	0.000
AMORT/FA	Pressure	0.077	0.061		0.006	0.000
EBIT/TA	Pressure	0.043	–0.032		0.000	0.000
LN(TA)	Pressure	18.411	17.910		0.000	0.008
TL/TA	Pressure	0.575	1.216		0.000	0.000
SALES/TA	Pressure	0.811	0.521		0.000	0.000
AUDSWI	Rationalization	0.054	0.052	0.937		

The results of Table 6 indicate that the frequency of having BIG4 auditors is higher among FFS cases, even though the difference from the nFFS group is not statistically significant. The detailed results in Tables A1 and A2 of the Appendix show that the frequency of BIG4 auditors is consistently higher for all types of auditors' comments (A1-A6), with the difference being significant (at the 5% level) for AC3-AC6. Moreover, firms from the FFS category change directors less frequently (p-value 0.038) and have more family members in the board of directors (t-test p-value 0.025; Mann-Whitney p-value 0.121). Looking at the results for the different types of auditors' comments (see the tables in the Appendix), it is evident that the change of directors is most significant for identifying cases of serious doubts about the continuation of companies' activity (AC1), whereas the existence of family members in the board is an important factor (at least at the 10% level) for identifying AC3-AC5 cases.

On the one hand, in firms where the chairman of the board does not change frequently (i.e., low DIRCHANGE), and there is a long-term stay of the same person in the management, this may lead to a concentration of high power and lack of appropriate controls and increase the possibility of committing fraud. Moreover, a family firm is more suspect for fraud, because the control and decision-making positions are occupied by a few individuals –mainly family members- and the power is concentrated by them. In these cases, there may be a lack of internal control mechanisms to ensure sound and financial and reporting practices. Finally, although whether the members of board of directors are also shareholders (SHARE&BOARD) does not appear significant in distinguishing between FFS/nFFS, it is strongly significant for identifying two of the auditors' comments (AC3 and AC5; see Tables A1-A2 in the appendix).

Regarding the financial ratios, they all are statistically significant for at least one of the two statistical tests. Overall, the FFS class is characterized by poor financial performance, i.e., lower profitability (e.g., negative EBIT/TA), higher debt burden (TL/TA), low liquidity (CA/TA), low sales turnover (SALES/TA), and lower market valuation (P/BV). These findings are in accordance with the standard theory which postulates that firms in unsound financial position are more likely to engage in delinquent actions due to the pressure that exists to enhance their financial image. Moreover, the firms in the FFS category are smaller in size as measured by the logarithm of total assets. The significance of the financial variables is further confirmed by the detailed results shown in the Appendix when focusing on the different types of auditors' comments. Indeed, all financial variables are found significant for identifying at least one of the six considered types of auditors' comments.

Table 7 shows the correlations for the financial ratios. It is evident that the correlations are generally low. The strongest correlations involve the managerial performance ratios COSTSOLD/AP and SALES/TA (correlation 0.56), as well as the ratios that describe the profitability (EBIT/TA), solvency (TL/TA), and liquidity (CA/CL) of the firms.

Table 7: Correlations between the financial ratios

	P/BV	COSTSOLD/AP	CA/CL	AMORT/FA	EBIT/TA	LN(TA)	TL/TA
COSTSOLD/AP	–0.01						
CA/CL	0.01	0.17					
AMORT/FA	0.01	0.04	0.04				
EBIT/TA	0.07	0.17	0.39	0.13			
LN(TA)	–0.01	0.03	0.08	–0.03	0.24		
TL/TA	–0.01	–0.12	–0.41	0.01	–0.35	–0.08	
SALES/TA	0.02	0.56	–0.04	0.20	0.18	0.06	0.04

3.3 Implementation of supervised machinal learning methods and a multi-label classification approach

Examining the efficacy of various classification algorithms in detecting financial reporting fraud is one of the key goals of this study. Within this context, the approach presented in this study relies on supervised data mining techniques. These techniques are all built around classification algorithms, which are used for the discovery of a model (or function) that distinguishes between predefined classes of data (Han et al., 2006). Training data with labels indicating whether each entity was fraudulent or not are necessary for a supervised learning technique.

In the context of the present study, it is examined the applicability of various machine learning methods for constructing an efficient system for distinguishing between FFS and nFFS cases, as well as for identifying the different comments made by auditors for FFS cases. More specifically, the binary methods used in the analysis include logistic regression (LR, Hosmer et al., 2013), the k-nearest neighbor algorithm (kNN, Cover and Hart, 1967), generalized additive models (GAM, Hastie and Tibshirani, 1990), and the random forest algorithm (RF, Breiman (2001)).

LR is the most popular statistical method for constructing classification models in financial decision-making problems, and it is heavily used in practice due to its simplicity and interpretability. Nevertheless, the linear modeling form assumed in LR has often been found to yield inferior predictive results compared to more complex machine learning models. GAM relaxes the linearity assumption of LR through an additive modeling scheme, which maintains the interpretability of linear models, yet enabling the introduction of non-linear effects. Although GAM approaches have not been considered in prior studies on financial fraud detection, they have been found to provide good predictive results in other similar areas in financial decision-making, such as credit scoring and bankruptcy prediction (Bequé et al., 2017; Lohmann et al., 2022; Lohmann & Ohliger, 2019).

The two other approaches used in the analysis, i.e., kNN and RF, are popular machine learning algorithms with numerous financial applications. The kNN is one of the most straightforward machine learning approaches that is easy to implement. On the other hand, RF is an ensemble learning algorithm that has been shown to provide strong predictive results in various domains. Except for the above algorithms, in the early stages of the analysis, additional methods have also been considered (e.g., ANN and SVM), but they were found to overfit the training data and provide poor generalization results. Thus, to simplify the presentation, we focused on the previous approaches.

Moreover, this study uses an improved method (Multi-Label Approach (ML-kNN) that expands the binary scheme into a more complex one that further identifies the nature of the auditors' comments. Previous research on the identification of FFS detection has concentrated on creating binary classification models to distinguish between FFS and nFFS cases. This results in a multi-label classification scheme, where each observation can belong to more than one class rather than just one.

Given the imbalance of the classes in the dataset, all methods are applied after weighting the training instances, to obtain more balanced results. The weighting scheme is based on the cost-sensitive framework for handling imbalanced data in classification problems, which is incorporated in the model fitting process of the selected learning methods, without requiring to use data processing algorithms such as oversampling or undersampling. More specifically, denoting by M_0 and M_1 the number of training observations from the two classes (e.g., FFS versus nFFS), the corresponding misclassification costs C_0 and C_1 are set such that $C_1/C_0 = M_0/M_1$(Japkowicz & Stephen, 2002; Thabtah et al., 2020). Following this approach, the case weights for the training observations in each class are set equal to $1/M_1$ for the class 1 and $1/M_0$ for the class 0, so that both classes contribute equally to the overall training error, thus resembling a balanced training sample.

The algorithms were all implemented in MATLAB using a Bayesian approach for the optimization of the hyperparameters involved in each method.

3.3.1 Logistic Regression (LR)

A probabilistic model known as logistic regression groups examples according to their likelihood of class membership. Since classification is a probabilistic process, it is advantageous to make sure that, while optimizing the parameters, the anticipated likelihood of the observed class for each training instance is as high as possible. This objective is accomplished by learning the model's parameters using the concept of maximum-likelihood estimation. The probability of the observed labels for each training instance is multiplied by the likelihood of the training data. It is obvious that higher values of this objective function are preferable (Aggarwal, 2018). This value's negative logarithm can be used to get a loss function in minimization form. The sigmoid activation function, which is widely used in neural network construction, can simulate the output layer. Suppose a set of n training pairs $(\bar{X}_1, y_1)$, $(\bar{X}_2, y_2)$, …., $(\bar{X}_n, y_n)$ in which $\bar{X}_i$ includes d-dimensional features and $y_i \in (0, +1)$ is a binary variable. Denoting by $W = (w_1, \dots, w_d)$ the vector of the variables' regression coefficients, LR uses the soft sigmoid function to estimate the probability that y_i is equal to 1:

$$\mathrm{y_i} = \mathrm{P}(\mathrm{y_i} = 1) = \frac{1}{1 + exp(-\overline{\mathrm{W}} \cdot \overline{\mathrm{X}}_\mathrm{i})}$$

According to Persons (1995), financial statement-based logistic regression models perform better than naive methods based on the practical distinction between type I and type II errors. Generally, in the literature review the most popular method for identifying FFS among these models is LR while Yang and Jiang (2020) in their research mentioned that the principal traditional statistical models include probit, discriminant, and logistic regression models. In fact, among 32 classification methods, logistic regression is ranked second in terms of prediction accuracy (Lim et al., 2000).

Spathis *et al.* (2002) employed logistic regression analysis to evaluate which financial parameters are associated to FFS utilizing financial ratios from enterprises as input. (Beasley, 1996) used LR to compare the boards of 75 fraud-prone companies to the boards of 75 non-fraud-prone companies. Using LR, Amara *et al.* (2013) investigated financial statement fraud and discovered that performance pressure on managers is a factor contributing to financial statement fraud. Based on data created by an international public accounting firm, Hansen *et al.* (1996) offered the powerful generalized qualitative-response model EGB2, which primarily uses probit and logistic techniques to detect managerial fraud. Their findings showed that both symmetric and asymmetric cost assumptions have good predictive power.

Lin *et al.* (2003) compared a fuzzy ANN to logistic regression using the same performance metrics, Feroz *et al.* (2000) compared the utility of an ANN model with logistic regression based on Hit-Rate, overall error rate, and estimated relative costs of misclassification (ERC) finding that LR performed better than the others techniques at relative error costs from 1:1 to 1:30. In a Chinese dataset containing 174 enterprises, Yue *et al.* (2009) evaluated public data and suggested a suitable strategy for identifying characteristics connected to false financial statements using LR. In order to find the optimum logistic regression approach parameters, the authors used 21 ratios as potential predictors for financial fraud statements. The authors compared the recommended method's prediction power with that of existing detection techniques. The findings of the created approach showed that its prediction power is around 10% higher than that of other models.

3.3.2 k-Nearest Neighbor (KNN)

KNN is a non-parametric technique that is used for regression and classification. It functions according to the notion that comparable data points typically have comparable output values or belong to the same class. Consequently, the technique uses the class labels of a data point's closest neighbors in the feature space to predict the class label of a new data point. With an instance-level labeled dataset, the k-NN algorithm can be trained without explicit training. Instead, it stores the training set in memory. The algorithm determines the distances between each new data point that is submitted for classification and every existing point in the dataset.

Based on the computed distances, the algorithm chooses the k neighbors who are closest to the new data point. More specifically, using some sort of distance measure, such as local metrics (Short and Fukunaga, 1980), global metrics (Fukunaga and Flick, 1984), the Mahalanobis distance, or the Euclidean distance, the nearest-neighbor rule assigns a class to an object (i.e., a firm) based on its nearest neighbor in the measurement space. The Euclidean distance is one of the most popular ways to calculate the distance (Chandola et al., 2009). Thus, a majority vote among the new data point's k closest neighbors determines the class label for classification tasks. The new data point is allocated the class label with the highest frequency among its neighbors.

Furthermore, 'K' is the option that indicates how many neighbors must be considered. Selecting a suitable value for 'k' is essential since it can greatly influence the model's functionality. The model becomes more sophisticated and sensitive to noise with a decreasing 'k' value, which could result in overfitting. On the other hand, a higher 'k' value could cause underfitting by smoothing out the decision boundaries.

The following paragraphs present some research which applied kNN among the other techniques proving that kNN is a valuable technique for FFS detection.

Gaganis (2009) applied 10 different classification models for FSFD using logit analysis (LR), discriminant analysis (DA), support vector machines (SVM), artificial neural networks (ANN), probabilistic neural networks (PNN), nearest neighbors (kNN), and two multicriteria decision-making methos (UTADIS and MHDIS). Using both financial and non-financial data, various models were created. There are 398 financial statements in the sample, and qualified audit opinions were given on 50 percent of them. To compare these alternative approaches, he employs testing outside of time and outside of sample. The findings are used to draw conclusions about how well the procedures work and to investigate the possibility of creating models that would help auditors spot false financial statements. According to AUC, the best methods are DA, mutlicriteria approaches, SVM and neural networks, followed by LR and kNN.

The effectiveness of machine learning algorithms in identifying businesses that produce falsified financial statements (FFS) is examined by Kotsiantis *et al.* (2006), along with the identification of characteristics that contribute to FFS. In order to do this, several experiments were carried out employing representative learning algorithms that were trained using a data set of 164 fraudulent and non-fraudulent Greek enterprises in the most recent period of 2001–2002. It is a difficult problem to decide the specific strategy to use. Creating a hybrid forecasting system with a variety of potential solution approaches included as components is a viable substitute for selecting only one method (an ensemble of classifiers). To obtain superior results, they developed a hybrid decision support system that combines the representative algorithms (decision trees, kNN, neural networks, Bayesian networks, SVM and rule-learners) with a stacking methodology. Kotsiantis *et al.* (2006) creates a methodology in which the 95.1% of the entire sample, 90.2% of the fraud instances, and 96.7% of the non-fraud cases are accurately classified by the algorithm.

Using the kNN algorithm and the outlier detection method, Malini and Pushpa (2017) examined credit card fraud. The performance results of both strategies were implemented and studied by the authors using a credit card approval system. The empirical results demonstrated the accuracy and efficiency of the kNN approach for identifying credit card fraud. In order to solve the issue of detecting anomalies fraudulent transactions in the prior study that used the HMM model, Heryadi *et al.* (2016) presented a study on debit card fraud transaction recognition. To increase the detection precision of anomaly transactions, the authors used k-Nearest Neighbor and Chi-Square Automatic Interaction Detection (CHAID). The study's findings demonstrated that the CHAID approach enhanced performance for identifying debit card fraud transactions. Awoyemi and Oluwadare (2017) examined the effectiveness of three machine learning models used on credit card transactions to spot fraud. Using LR, NB, and kNN trained on a credit card dataset provided by European cardholders and consisting of 284,807 transactions, the authors conducted a comparative analysis. According to the results of the comparative analysis, K-Nearest Neighbor performed better than LR and NB approaches.

The K-Nearest Neighbor classifier's key benefit is that it is a straightforward approach to implement. The basic drawback of this method is that it requires the algorithm to compute the distance and sort every training sample at each prediction, which can be time-consuming if there are many training examples (Chimonaki et al., 2019).

3.3.3 Generalized Additives Models (GAM)

Generalized additive models (GAM), are a versatile class of statistical models used in classification and regression applications. Through the allowance of non-linear interactions between predictors and the response variable, they expand upon the framework of the linear model. Since a GAM considers nonlinear effects on the predictor and non-monotonous effects on the default probability, it is more sophisticated than a generalized linear model (GLM) (Lohmann & Ohliger, 2019). GAM are especially helpful in situations when there is a complex relationship between the predictors and the response that may be too complex for linear models to fully represent.

The response variable in GAMs is represented as a mixture of other predictors and perhaps linear terms, as well as smooth functions of the predictor variables. The use of smooth functions, which can identify non-linear correlations in the data, is a crucial component of GAMs. Spline functions, like thin plate splines or cubic splines, are commonly used to depict these smooth functions because they offer flexibility for modeling intricate interactions (Hastie & Tibshirani, 1990).

A GAM is composed of three main parts:

Linear Predictor: The linear relationships between predictors and the response variable are captured by the linear predictor component. Like conventional linear models, it has terms for every predictor variable.

Smooth Functions: The non-linear interactions between predictors and the response are modeled by smooth functions. The smooth function of each predictor variable can vary, enabling varying levels of non-linearity.

Link Function: To make sure the predicted values are on the right scale for the response variable, the linear predictor is changed using a link function. The logistic link and the softmax link are common link functions that are used in binary and multiclass classification, respectively.

Maximum likelihood estimation and other optimization approaches are used to estimate the parameters of a GAM, including the coefficients for linear terms and the parameters defining the smooth functions. By fitting spline models to the data, the smooth functions are calculated; regularization parameters or degrees of freedom regulate the smooth function's complexity.

One of the most common generalized linear methods is the LR. Regression diagnostic concepts were expanded by (Landwehr et al., 1984) to situations involving linear logistic regression in which the response variable y is either 0 or 1. Using partial residual plots, nonlinearities in the model were found.

Another technique for estimating models of this form is local likelihood estimation, which is asymptotically equivalent to the local scoring process (Tibshirani & Hastie, 1987). Local scoring has the benefit of being significantly quicker. A nonparametric approach to modeling generalized linear models using spline functions was developed by O'Sullivan *et al.* (1986). They model directly by utilizing high-dimensional splines, which are computationally demanding and challenging to read and display beyond two dimensions. It is preferable to use two-dimensional surfaces to identify connections (Hastie & Tibshirani, 1987).

Although this technique has not been widely used in research into FFS detection, it seems that it is suitable or more complex relationships.

Lohmann *et al.* (2022) in order to find and examine nonlinear interactions between independent variables based on markets and accounting and how these links affected bankruptcy forecasts, this study used generalized additive models. Using extensive data on publicly traded US companies, they demonstrated empirically how statistically, and economically significant nonlinear interactions affected the bankruptcy forecast. The findings showed that numerous statistical validity metrics are greatly improved when these nonlinear interactions are considered. Additionally, they employed a validity metric based on the bankruptcy prediction models' profitability in relation to credit scoring. The results proved that accounting for nonlinear interactions can significantly improve bankruptcy prediction models' discriminatory ability. Moreover, it has been proved that the results from GAM were superior to the ones from generalized linear model (GLM) decreasing the total misclassification costs (Lohmann and Ohliger, 2019; Lohmann *et al.*, 2022).

The interpretability of GAMs is one of their benefits. It is simpler to comprehend how different predictors affect the result when the relationships between the predictors and the response can be visually interpreted thanks to the smooth functions. Furthermore, hypothesis testing can be used to determine key predictors and evaluate the importance of the smooth terms.

Non-linear relationships between predictors and the response are supported by Generalized Additive Models, which provide a versatile and understandable method for classification tasks. GAMs can identify intricate patterns in the data while preserving interpretability by utilizing smooth functions. GAMs can be useful tools for solving a variety of classification problems when they are fitted and evaluated properly.

3.3.4 Random Forest (RF)

Random forests are a collection of tree predictors where each tree depends on the values of a random vector sampled independently and with the same distribution for all the trees in the forest. When the number of trees in a forest increases, the generalization error converges to a limit (Breiman & Friedman, 1985). The strength of each individual tree in the forest and the correlation between them affect the generalization error of a forest of tree classifiers. Each node is split randomly into a set of features, producing error rates that are resilient to noise. Error, strength, and correlation are tracked using internal estimations, which are then utilized to demonstrate how the splitting process responds to an increase in the number of features.

Classification accuracy has been greatly enhanced by growing a collection-ensemble of trees and letting them choose the most preferred class. It is common practice to grow these ensembles by generating random vectors that regulate the growth of individual tree in the ensemble. An essential version is bagging (Breiman, 2001) in which each tree is grown arbitrarily (without replacement) from the training set's instances. Another illustration is the random split selection method (Dietterich, 2000) in which one of the *K* best splits is randomly chosen at each node. Another strategy is to choose the training set at random from a set of weights applied to the training set's examples. The "random subspace" strategy, which selects a subset of features at random to utilize in growing each tree, has been the subject of several articles by (Ho, 1998).

Random forests choose a randomly chosen subset of variables and seeks for the best variable for splitting among them, rather than evaluating all potential predictors while looking for the next splitter. Thus, any two unique nodes in a single random forest tree are likely to be examining different sets of variables, which means that the random forest trees are likely to involve a lot of different splitters. As a result, every tree in the random forest is a type of dynamically built closest neighbor classifier, which generally produces good results (Whiting et al., 2012).

Specifically, RF was used by Liu *et al*. (2015) for multidimensional analysis, partial correlation analysis, detailed feature selection, and the detection of financial fraud techniques using data from listed enterprises from China. Data used in the random forest modeling come from a bootstrap sampling (Liaw & Wiener, 2002). Liu *et al.* (2015) compared two parametric models with two non-parametric models, four other models, and discovered that RF has the highest accuracy. The results showed that RF provides advantages over alternative categorization models in a number of ways (Kirkos et al., 2007; Liaw & Wiener, 2002).

According to (Liu et al., 2015), it is mentioned that firstly, in practically all of its models, it has the highest recognition efficiency of random forests. Second, it handles additional non-normal data effectively by disregarding the assumption of data normality. It may require a significant amount of high-dimensional data processing and a difficult example of co-linear over-fitting. Third, it can assess each variable's significance and effectively remove insignificant ones. Finally, the ideal set of variables can be chosen from which to create models.

3.3.5 Multi-Label Approach

While previous studies on the identification of FFS detection have focused on developing binary classification models to distinguish between FFS and nFFS cases, in this study an enhanced approach is employed that extends this binary scheme into a more elaborate one that further identifies the nature of the auditors' comments. This leads to a multi-label classification scheme, in which an observation may be a member of multiple classes instead of a single category (Tsoumakas & Katakis, 2007). The multi-label classification context is relevant for FFS detection, because auditors provide various comments and concerns on the accuracy of a company's financial reports. A multi-label classification approach enables the distinction between different types of FFS.

In this context, the applicability of various machine learning methods for constructing an efficient multi-label system for distinguishing between FFS and nFFS cases is examined, as well as between the different comments made by auditors for FFS. More specifically, the methods used in the analysis include logistic regression (LR) (Hosmer et al., 2013), the k-nearest neighbor algorithm (kNN) (Cover and Hart, 1967) , generalized additive models (GAM) (Hastie & Tibshirani, 1990), and the random forest algorithm (RF) (Breiman, 2001). These approaches are designed to handle single-label problems. Moreover, a multi-label algorithm is also considered, namely the ML-kNN algorithm, which is extension of kNN to multi-label problems (M. L. Zhang & Zhou, 2007).

For the application of the single-label approaches, the basic setting is the binary relevance scheme (BR, Zhang *et al.*, 2018), which involves the construction of binary classifiers for each pair of classes, through a one-against-all approach. Thus, a model is constructed to distinguish between FFS and nFFS firms. Moreover, three additional models are also developed to differentiate between cases that belong in each one of the FFS groups and the remaining observations (e.g., observations in AC12 versus the rest, observations in AC36 versus the rest, etc.).

A limitation of the BR scheme is that it ignores the relationships between the classes. In the context of the problem considered in this study, it is obvious that AC12, AC36, and AC45, only apply to the FFS class. More specifically, let y^{FFS} be a binary indicator, such that $y_i^{\text{FFS}} = 1$ if and only if observation i belongs in the FFS class. Similarly, let $y^{\text{AC12}}, y^{\text{AC36}}, y^{\text{AC45}}$ be indicators such that $y_i^k = 1$ ($k \in \{\text{AC12}, \text{AC36}, \text{AC45}\}$) if and only if auditors' comments of the type k, apply to observation i. Thus, the true labels for an observation i can be represented by the vector $\mathbf{y}_i = (y_i^{\text{FFS}}, y_i^{\text{AC12}}, y_i^{\text{AC36}}, y_i^{\text{AC45}})$. Denoting by $\hat{\mathbf{y}}_i = (\hat{y}_i^{FFS}, \hat{y}_i^{\text{AC12}}, \hat{y}_i^{\text{AC36}}, \hat{y}_i^{\text{AC45}})$ the vector of output (estimated) labels from a decision model for observation i, the following cases indicate inconsistencies in the obtained results:

$$\hat{y}_i^{\text{FFS}} = 1 \text{ and } \hat{y}_i^{\text{AC12}} + \hat{y}_i^{\text{AC36}} + \hat{y}_i^{\text{AC45}} = 0$$
$$\hat{y}_i^{\text{FFS}} = 0 \text{ and } \hat{y}_i^{\text{AC12}} + \hat{y}_i^{\text{AC36}} + \hat{y}_i^{\text{AC45}} \geq 1$$

The first condition corresponds to the case where a decision model indicates that firm i belongs to the FFS class ($\hat{y}_i^{\text{FFS}} = 1$), yet none of the possible auditors' comments apply ($\hat{y}_i^{\text{AC12}} + \hat{y}_i^{\text{AC36}} + \hat{y}_i^{\text{AC45}} = 0$). The second conditions implies that a model indicates that there is no falsification in the financial statements of company i (i.e., $\hat{y}_i^{\text{FFS}} = 0$), but at the same time, the model shows that at least one of the auditors' comments apply ($\hat{y}_i^{\text{AC12}} + \hat{y}_i^{\text{AC36}} + \hat{y}_i^{\text{AC45}} \geq 1$).

To correct such inconsistent results, two strategies are employed in the analysis. In the first scheme, the label vector $\hat{\mathbf{y}}_i$ for an inconsistent test case i is defined through the class labels $\mathbf{y}_i$ of its nearest neighbor (NN), selected from the training set. The identification of the kNN is based on the posterior class membership probabilities $\mathbf{p}_i = \left(p_i^{FFS}, p_i^{\text{AC12}}, p_i^{\text{AC36}}, p_i^{\text{AC45}}\right)$ that are obtained by a learning algorithm. Thus, the kNN algorithm is used as a meta-learner that is applied only to inconsistent cases, providing results that are compatible with the definition of the classes. Henceforth, this thesis shall refer to this scheme as *case-based inconsistency correction* (CBIC). As an alternative, it also be considered the use of the kNN algorithm to define the label vectors for all test instances (including consistent ones). Similarly, to the CBIC scheme, the kNN algorithm is applied to the posterior class membership probabilities, rather than the original data of the firms. Henceforth, this approach will be referred to as the *kNN-based meta learner* (kNNML). An outline of the two schemes is presented in Figure *4*. The effectiveness of these two schemes is examined in relation to the base learners developed with LR, kNN, GAM, and RF, as well the multi-label ML-kNN algorithm.

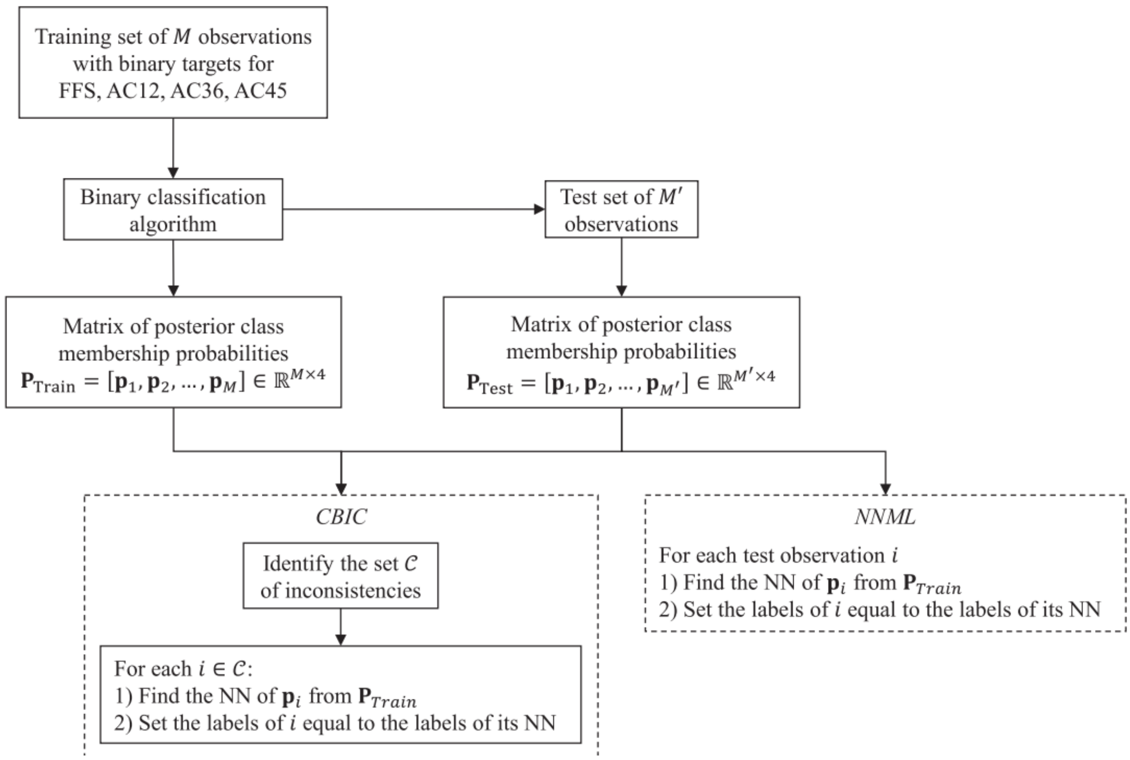

Figure 4: Outline of the procedures for correcting inconsistencies

The effectiveness of these two schemes is examined in relation to the base binary learners developed with LR, kNN, GAM, and RF. Moreover, it is additionally considered a multi-label algorithm, namely the ML-kNN algorithm, which is extension of kNN to multi-label problems (M. L. Zhang & Zhou, 2007). This specific multi-label algorithm was selected to match the structure of the proposed approach, which relies on the use of nearest neighbors to correct inconsistent assignments from the base algorithms described above.

3.4 Performance Metrics

The anomaly detection problem is frequent and sometimes constitutes a classification challenge. Examining the confusion matrix of a classification model is a useful technique to evaluate the performance of decision models in this setting (Hilal et al., 2022). The discrepancy between a data set's ground truth and a model's predictions is described by the confusion matrix. The accuracy of the correct forecasts is the precision. The fraction of positive cases that a classifier accurately detects is known as recall or sensitivity. The amount of overall accurate predictions made by the model is the accuracy. The proportion of successfully categorized negative samples to all negative samples is the classifier's specificity. There is an inverse relationship between the two values, even though one might instinctively strive to increase both recall and precision. A higher precision may be forced at the expense of recall, and vice versa. The precision/recall trade-off is what is referred to as the F-score, or F1-score, which is the which is the harmonic mean of precision and recall, is a superior metric to aim for.

To evaluate the performance of the algorithms used in the analysis, various accuracy metrics are considered (Madjarov et al., 2012):

- Macro-precision: it represents the precision rate (PR), averaged over all labels (i.e., FFS, AC12, AC36, AC45). More specifically, denoting by TP_k the number of true positives (e.g., FFS cases) for label k ($k \in \{\text{FFS}, \text{AC12}, \text{AC36}, \text{AC45}\}$) and by FP_k the number of false positives, the macro-precision index is defined as:

$$\text{Macro PR} = \frac{1}{4}\sum_{k} PR_k = \frac{1}{4}\sum_{k} \frac{TP_k}{TP_k + FP_k}$$

- Macro-recall (macro RC): it represents the recall rate, averaged over all labels:

$$\text{Macro RC} = \frac{1}{4}\sum_{k} RC_k = \frac{1}{4}\sum_{k} \frac{TP_k}{TP_k + FN_k}$$

  where RC_k and FN_k denote the recall and the number of false negatives for label k.

- Macro F_1: it is the harmonic mean of the precision and recall rates for each label:

$$\text{Macro } F_1 = \frac{1}{4}\sum_{k} \frac{2PR_k RC_k}{PR_k + RC_k}$$

- Macro area under the receiver operating characteristic curve (AUC): it represents the average AUC over all labels:

$$\text{Macro AUC} = \frac{1}{4}(\text{AUC}_1 + \cdots + \text{AUC}_4)$$

  where AUC_k is the AUC corresponding to label k.

- Hamming loss (HL): the Hamming loss measures the proportion of incorrect classifications to the total number of classifications:

$$\text{HL} = 1 - \frac{1}{4M'}\sum_{i,k} I\left(y_i^k, \hat{y}_i^k\right)$$

  where M' is the number of test observations, and $I(y_i^k, \hat{y}_i^k)$ is an indicator function, such that $I\left(y_i^k, \hat{y}_i^k\right) = 1$ if and only if $y_i^k = \hat{y}_i^k$, for $k \in \{\text{FFS}, \text{AC12}, \text{AC36}, \text{AC45}\}$.

- Accuracy (AC): accuracy is measured through the Jaccard index between the set of actual classifications $\mathbf{y}_i$ and the estimated ones $\hat{\mathbf{y}}_i$ for each test observation i. The accuracy index is averaged over all test observations:

$$\text{Accuracy} = \frac{1}{M'}\sum_{i} \frac{|\mathbf{y}_i \cap \hat{\mathbf{y}}_i|}{|\mathbf{y}_i \cup \hat{\mathbf{y}}_i|}$$

Except for the above metrics that measure classification performance in a multi-label setting, similar metrics are also employed to evaluate the results in a binary setting for each one of the four independent labels (FFS, AC12, AC36, AC45). The binary metrics include precision, recall, the F_1 index, the AUC, as well as classification accuracy (CA), measured as the ratio of correct predictions to the number of test instances.

To obtain unbiased estimates for the performance of the algorithms, a cross-validation (CV) analysis is employed with 10 folds (i.e., 10-fold CV). Given the panel structure of the dataset, the CV folds are constructed by partitioning the unique firm IDs (133 companies) rather than the firm-year observations (752 observations). Thus, the training and test partitions are independent, in the sense they include observations from different firms.

3.5 Results

Several train-and-test runs are carried out utilizing a moving window technique to get reliable estimates for the algorithms' performance. More specifically, three replications are conducted, each involving a training window of three years and a test period of one year. In the first train-and-test run, the training data covers the period 2014-2016 and the developed models are tested on the 2017 data. For the second run, the period 2015-2017 is used for training the models, which are then tested in 2018. The same process is repeated one more time, with the data for 2016-2018 used for training and the 2019 data for testing the model. The presentation in this section focuses on the test results from these three experiments. I begin with the analysis of the classification performance of the binary and multi-class models (subsection 3.5.1) and then proceed in subsection 3.5.2 with the presentation of the features' importance (i.e., the independent variables).

3.5.1 Classification performance

The analysis begins with the results involving the distinction between FFS and nFFs cases. Table 8 summarizes the performance metrics for all methods and the two corrective schemes that were described in subsection 3.3.5. From Table 8, it is evident that RF and GAM provide the best results in discriminating between the two classes, followed by LR and the MLKNN algorithm, whereas the standard kNN algorithm yields the worst results among the considered methods. The multi-label kNN algorithm (MLKNN) performs slightly better than the LR base model, but LR provides improved results with the two correction schemes, particularly when combined with the NNML scheme.

As far as the comparison of the three modeling schemes is involved (base models, CBIC, NNML), the two correction schemes (CBIC, NNML) provide slightly improved results in terms of the F_1 metric and AUC. The same holds true for the classification accuracy measure, when comparing the base models against the ones developed with the CBIC schemes (i.e., average accuracy 0.851 for the base models versus 0.859 for the CBIC scheme). On the other hand, the classification accuracy obtained with the NNML schemes is significantly higher (0.87, on average), driven by the improved results in terms of precision.

Table 8: Classification results for the distinction between FFS and nFFS cases (averaged over the three tests)

	PR	RC	F_1	CA	AUC
			Base models		
kNN	0.613	0.727	0.657	0.828	0.817
LR	0.593	0.769	0.659	0.818	0.892
GAM	**0.739**	0.767	**0.740**	**0.880**	0.924
RF	0.695	**0.793**	0.735	0.871	**0.939**
MLKNN	0.690	0.708	0.691	0.856	0.872
Average	*0.666*	*0.753*	*0.696*	*0.851*	*0.889*
			CBIC scheme		
NN	0.644	0.727	0.677	0.844	0.816
LR	0.616	**0.805**	0.688	0.835	0.895
GAM	**0.746**	0.757	0.738	**0.880**	0.916
RF	0.716	0.787	**0.743**	0.877	**0.940**
Average	*0.681*	*0.769*	*0.712*	*0.859*	*0.892*
			NNML scheme		
kNN	0.706	0.685	0.694	0.864	0.815
LR	0.681	0.704	0.683	0.852	0.895
GAM	0.761	**0.725**	0.730	0.880	0.925
RF	**0.826**	0.686	**0.738**	**0.894**	**0.937**
Average	*0.744*	*0.700*	*0.711*	*0.872*	*0.893*

Table 9 presents additional results for each one of the three groups of auditors' comments AC12, AC45, and AC36, which were described in subsection 3.1. Similarly to the FFS/nFFS results, the RF algorithm together with GAM provide the best results in most of the cases. On average, the prediction results are similar across the three groups of comments in terms of classification accuracy (CA) and the AUC metric, whereas the F_1 scores are higher for the identification of cases in the AC12 group. Comparing the three schemes, CBIC provides small, yet almost consistent improvements compared to the base models, in terms of the three overall performance metrics (F_1, CA, AUC), with the differences being higher for the AC45 group. The NNML scheme yields improved results in the F_1 score and CA, compared to the base models in the AC12 group, while having similar AUC performance. For the two other groups of comments (AC45, AC36), NNML provides higher classification accuracy, due to its better PR results.

Table 9: Classification results for the FFS groups

	AC12					AC45					AC36				
	PR	RC	F_1	CA	AUC	PR	RC	F_1	CA	AUC	PR	RC	F_1	CA	AUC
	Base models														
kNN	0.653	0.689	0.666	0.856	0.807	0.449	0.764	0.556	0.846	0.824	0.468	0.632	0.497	0.850	0.777
LR	0.574	**0.749**	0.639	0.826	0.894	0.438	0.774	0.549	0.838	0.858	0.390	0.753	0.497	0.819	0.856
GAM	0.663	0.733	0.684	0.862	0.920	0.499	0.800	**0.593**	0.863	0.919	0.496	**0.761**	0.590	0.879	0.907
RF	0.654	0.738	0.684	0.863	**0.934**	0.437	**0.856**	0.573	0.831	**0.923**	**0.615**	0.699	**0.613**	**0.899**	**0.915**
MLKNN	**0.779**	0.706	**0.725**	**0.887**	0.879	**0.564**	0.463	0.440	**0.870**	0.836	0.569	0.410	0.438	0.886	0.827
Average	*0.636*	*0.723*	*0.680*	*0.859*	*0.887*	*0.477*	*0.731*	*0.542*	*0.850*	*0.872*	*0.508*	*0.651*	*0.527*	*0.867*	*0.856*
	CBIC scheme														
kNN	0.709	0.673	0.689	0.877	0.807	0.494	0.683	0.569	0.872	0.821	0.566	0.632	0.577	**0.897**	0.777
LR	0.594	**0.793**	0.667	0.838	0.895	0.502	0.774	0.596	0.866	0.860	0.463	**0.735**	0.560	0.868	0.860
GAM	**0.721**	0.731	0.713	**0.882**	0.921	**0.596**	0.716	0.635	**0.897**	0.921	0.533	0.702	**0.597**	0.891	**0.913**
RF	0.695	0.753	**0.713**	0.880	**0.934**	0.526	**0.828**	**0.641**	0.881	**0.922**	**0.604**	0.665	0.594	0.896	0.912
Average	*0.680*	*0.738*	*0.695*	*0.869*	*0.889*	*0.530*	*0.750*	*0.610*	*0.879*	*0.881*	*0.541*	*0.684*	*0.582*	*0.888*	*0.866*
	NNML scheme														
kNN	0.709	0.673	0.689	0.877	0.807	0.540	0.530	0.534	0.885	0.821	0.664	**0.547**	**0.597**	**0.917**	0.777
LR	0.672	**0.736**	0.694	0.866	0.892	0.514	0.455	0.457	0.868	0.853	0.444	0.308	0.346	0.866	0.859
GAM	0.744	0.717	0.719	0.888	0.922	**0.654**	**0.566**	**0.594**	**0.903**	0.911	0.600	0.520	0.533	0.895	**0.910**
RF	**0.801**	0.705	**0.740**	**0.902**	**0.937**	0.557	0.534	0.543	0.884	**0.919**	**0.767**	0.375	0.477	0.906	0.903
Average	*0.731*	*0.708*	*0.710*	*0.883*	*0.890*	*0.567*	*0.521*	*0.532*	*0.885*	*0.876*	*0.619*	*0.438*	*0.488*	*0.896*	*0.862*

Table 10 summarizes the performance metrics considering the full multi-label setting. Regarding the relative performance of the classification methods, RF is the best performer overall, followed by GAM, whereas the standard kNN algorithm provides the worst results. The multi-label schemes yield significant improvements over the base models, in terms of the Hamming loss and the accuracy metric (Jaccard index). The improvements are more noticeable with the NNML setting. For the F_1 metric, the results obtained with the CBIC scheme are also improved compared to the base models, whereas those of NNML are inferior due to the lower RC. Finally, in terms of AUC, no noticeable differences are observed between the three schemes.

Table 10: Overall classification results

	Macro PR	Macro TC	Macro F_1	Macro AUC	HL	AC
			Base models			
kNN	0.546	0.703	0.594	0.806	0.155	0.758
LR	0.499	0.761	0.586	0.875	0.175	0.734
GAM	0.599	0.765	**0.652**	0.917	0.129	0.791
RF	0.600	**0.771**	0.651	**0.928**	0.134	0.794
MLKNN	**0.651**	0.572	0.573	0.854	**0.125**	**0.801**
Average	*0.579*	*0.715*	*0.611*	*0.876*	*0.144*	*0.776*
			CBIC scheme			
kNN	0.603	0.679	0.628	0.805	0.128	0.809
LR	0.544	**0.777**	0.628	0.878	0.148	0.787
GAM	**0.649**	0.727	0.670	0.918	**0.112**	0.833
RF	0.635	0.758	**0.673**	**0.927**	0.116	**0.834**
Average	*0.608*	*0.735*	*0.650*	*0.882*	*0.126*	*0.816*
			NNML scheme			
kNN	0.655	0.609	0.628	0.805	0.114	0.831
LR	0.578	0.551	0.545	0.875	0.137	0.805
GAM	0.690	**0.632**	**0.644**	0.917	0.108	0.837
RF	**0.738**	0.575	0.624	**0.924**	**0.103**	**0.851**
Average	*0.665*	*0.592*	*0.610*	*0.880*	*0.116*	*0.831*

3.5.2 Features' importance

The interpretability of machine learning models has become a critical issue for their successful application and deployment in business domains. To address this issue in the context of the models developed in this study, Tables 11 and 12 provide results on the role and importance of the explanatory in the models. More specifically, Table 11 presents the LR coefficients for the three-time windows, averaged over the specifications corresponding to the four labels $\{FFS, AC12, AC36, AC45\}$.

Moreover, the same table provides indications on the statistical significance of the variables, i.e., the number of specifications in which the variables are significant at the 1%, 5%, and 10% levels. Regarding the machine learning models, Table 12 shows the mean absolute Shapley values (normalized so they sum up to 1) as an indication of the features' importance. The Shapley value has been proposed by Lundberg and Lee (2017), as a general approach to assess the contribution of explanatory variables in machine learning models. The Shapley value indicates the contribution of an explanatory variable to the outcome of a model for a particular instance, when compared to the outcomes obtained with all possible combinations of variables for that instance. The results shown in Table 12 correspond to Shapley values averaged across all training cases, and over the four modeling specifications (labels).

Table 11: Logistic regression coefficients by year (averaged over all labels) and statistical significance results

	Coefficients			Significance*		
	2014-16	2015-17	2016-18	2014-16	2015-17	2016-18
DIRCHANGE	-3.652	-0.787	-0.481	1 / 1 / 1	0 / 0 / 0	0 / 0 / 0
SHARE&BOARD	-0.983	-0.817	-0.700	1 / 0 / 0	1 / 0 / 1	1 / 0 / 0
FAMILY	-0.245	-0.183	0.477	0 / 2 / 0	0 / 0 / 1	0 / 0 / 1
BIG4	0.922	1.293	1.537	1 / 0 / 0	2 / 0 / 0	2 / 1 / 0
Sector=Trade	-0.315	-0.395	-0.385	0 / 0 / 0	0 / 0 / 1	0 / 0 / 0
Sector=Services	1.005	0.644	0.169	2 / 0 / 0	0 / 0 / 2	0 / 0 / 1
P/BV	0.001	-0.002	-0.007	1 / 1 / 0	0 / 1 / 0	0 / 0 / 0
COSTSOLD/AP	0.076	0.029	-0.013	3 / 1 / 0	2 / 0 / 0	2 / 2 / 0
CA/CL	-1.575	-0.492	-0.544	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
AMORT/FA	-5.360	-6.708	-4.182	0 / 1 / 1	1 / 2 / 0	0 / 2 / 0
EBIT/TA	-11.011	-8.095	-9.021	3 / 0 / 0	2 / 0 / 0	3 / 0 / 0
LN(TA)	-0.376	-0.309	-0.116	2 / 1 / 0	2 / 1 / 0	0 / 1 / 1
TL/TA	3.380	3.903	2.823	3 / 0 / 0	4 / 0 / 0	4 / 0 / 0
SALES/TA	-1.601	-1.465	-1.743	2 / 2 / 0	3 / 1 / 0	3 / 1 / 0
AUDSWI	1.346	1.248	0.360	0 / 1 / 0	2 / 1 / 0	0 / 1 / 0

* Number of specifications where a variable is statistically important at the 1% / 5% / 10% levels.

Table 12: Mean absolute (normalized) Shapley values by year (averaged over all labels)

	kNN			GAM			RF		
	2014-16	2015-17	2016-18	2014-16	2015-17	2016-18	2014-16	2015-17	2016-18
DIRCHANGE	0.034	0.038	0.038	0.021	0.014	0.016	0.005	0.006	0.005
SHARE&BOARD	**0.081**	0.077	0.077	0.037	0.031	0.029	0.011	0.017	0.010
FAMILY	0.080	**0.083**	0.082	0.074	0.043	0.041	0.039	0.044	0.039
BIG4	0.066	0.073	**0.085**	0.029	0.041	0.032	0.005	0.019	0.009
SECTOR	**0.132**	**0.134**	**0.134**	0.036	0.049	0.035	0.019	0.019	0.011
P/BV	0.060	0.072	0.070	0.043	0.050	0.064	0.035	0.044	0.061
COSTSOLD/AP	0.050	0.045	0.045	0.041	0.044	0.048	0.057	0.036	0.044
CA/CL	0.078	0.068	0.072	**0.136**	**0.178**	**0.180**	**0.275**	**0.230**	**0.258**
AMORT/FA	0.062	0.061	0.061	**0.075**	0.065	0.035	0.047	0.049	0.027
EBIT/TA	0.076	0.064	0.066	0.067	0.085	0.070	**0.120**	**0.085**	**0.095**
LN(TA)	0.081	0.083	0.083	0.058	0.068	0.057	0.034	0.052	0.049
TL/TA	**0.097**	**0.086**	**0.091**	**0.290**	**0.208**	**0.302**	**0.287**	**0.309**	**0.323**
SALES/TA	0.057	0.059	0.053	0.069	**0.092**	**0.076**	0.061	0.077	0.064
AUDSWI	0.044	0.056	0.043	0.024	0.030	0.015	0.004	0.011	0.006

Note: The three highest values in each year are shown in bold.

Regarding the results of Table 11 which concerns the role of the independent variables in the LR models, it should be noted that variables with negative coefficients lower the risk of FFS, whereas positive coefficients indicate variables that increase the likelihood of FFS. Going into the details of the results, it is evident that there are some mixed indications regarding the significance of the fraud diamond variables over the three-time windows of the analysis. For instance, DIRCHANGE is found to have a significant negative effect (i.e., lower risk of FFS) in 2014-2016, but its contribution in the next two periods is not significant at the 10% level.

On the other hand, as it is shown in Table 11, AUDSWI is statistically significant in all periods with positive signs. This is expected because a qualified audit opinion often leads a company to change its auditors (Kirkos et al., 2007). Interestingly, the BIG4 variable is found significant over all periods, with positive signs, which indicates FFS events are more often observed for firms that are audited by BIG4 companies. A possible explanation for this result could be that BIG4 auditors are more experience and conduct more thorough audits, thus identifying more cases where firms do not provide correct financial information in their reports. Regarding the sector of the companies, the results indicate that firms in the services sector have higher risk of FFS (positive coefficients).

Considering the financial variables in the LR models presented in Table 11, the variables that are more often significant (at least at the 10% level) are the ratios TL/TA, SALES/TA, and COSTSOLD/AP, followed by EBIT/TA and LN(TA). The solvency indicator TL/TA has a strong positive effect on the likelihood of FFS, whereas the profitability indicators EBIT/TA and SALES/TA both have a strong negative effect (i.e., lower risk of FFS). These results are in accordance with previous studies (Cecchini et al., 2010; Gaganis, 2009; Kirkos et al., 2007; Wyrobek, 2020), thus confirming that firms that face financial difficulties are more prone to report falsified financial statements. Regarding the size of the companies, as measured by the logarithm of total assets, it has a strong negative effect, which indicates that the risk of FFS is lower for larger companies.

The examination of the results shown in Table 12 for the importance of the variables in the three machine learning approaches (NN, GAM, RF), reveals that in these models the significance of the variables differs compared to the LR results. Differences between the methods are also evident, even though GAM and RF share some similar patterns. Nevertheless, the solvency indicator TL/TA is highlighted as a consistently important factor for FFS detection across all methods (including LR) and years. In addition to this ratio, the results of GAM and RF indicate that ratios such as EBIT/TA, SALES/TA, and CA/CL, are also important. The former two ratios were also found significant in the LR models. Regarding the fraud diamond variables, they have higher importance in the kNN models, whereas their contribution in the GAM and RF models appears to be lower.

Overall, the results of all methods highlight a link between the financial soundness of the firms and the likelihood of FFS, whereas the role of non-financial factors, such as those related to the fraud diamond theory appears to be mixed and requires further analysis. Nevertheless, it should be noted that differences between methods should not be a surprise, especially when comparing the statistical results of LR to the feature importance in machine learning models, given that each method relies on different assumptions.

Chapter 4. Conclusion and Future Perspectives

Although the detection of financial fraud is a significant problem for various stakeholders, experts have failed to reduce the number of frauds and the financial and non-financial costs resulting from them. Nowadays, corporate frauds are more complicated, based on elaborate techniques and fraudsters are usually above suspicion, thus causing larger losses (ACFE, 2022). Therefore, it is important to advance the fraud detection procedures beyond the traditional auditing tools and to strengthen these procedures by better understanding fraudsters' motivations and practices.

This thesis proposed the development of fraud detection models based on the fraud diamond theory and examined the performance of models constructed with popular machine learning approaches in a multi-label classification setting using financial statements from 133 Greek companies listed in the Athens Stock Exchange over the period 2014 to 2019. Regarding the applied techniques, RF and GAM provided the best results in discriminating between the two classes, followed by LR and the MLKNN algorithm, whereas the standard kNN algorithm yielded the worst results among the considered methods. Moreover, the multi-label kNN algorithm (MLKNN) performed slightly better than the LR base model, but LR provided improved results with the two correction schemes, particularly when combined with the NNML scheme. Furthermore, the two correction schemes (CBIC, NNML) provided slightly improved results in terms of the F1 metric, CA and AUC.

Considering the features' importance, it was revealed that the significance of the variables differed in the three approaches (kNN, GAM, RF) compared to the LR results. The solvency indicator TL/TA was highlighted as a consistently important factor for FFS detection across all methods (including LR) and years. Moreover, the results of GAM and RF indicate that ratios such as EBIT/TA, SALES/TA, and CA/CL, were also important. The fraud diamond variables had higher importance in the kNN models, whereas their contribution in the GAM and RF models appeared to be lower. Additionally, the results of all methods highlight a link between the financial soundness of the firms and the likelihood of FFS, whereas the role of non-financial factors, such as those related to the fraud diamond theory appears to be mixed and requires further analysis. Differences between methods should not be a surprise, especially when comparing the statistical results of LR to the feature importance in machine learning models, given that each method relies on different assumptions.

From the presented methodology, it had been proved that it is important to advance the fraud detection procedures beyond the traditional auditing tools and to strengthen these procedures by better understanding fraudsters' motivations and practices. This goal can be succeeded by enhancing the accuracy by classifying a firm into more than two classes and accelerating the need of using new theories human-approaching in FFS detection such as Fraud Diamond Theory. Moreover, the two corrective schemes (CBIC and NNML) enhanced the performance metrics of the models, meanwhile the multi-label approach had led to a better understanding of the significance of the auditors' comments. Finally, there was not a clear picture of the contribution of the non-financial variables involved in Fraud Diamond Theory, but this probably happened due to selection of the specific variables.

Undoubtfully, the context of this thesis provides a comprehensive modeling framework that not only enables the identification of financial fraud, but it can also provide insights into the types of fraud that are relevant for each particular case, without having to resort to different models for each type of fraud, which could be difficult to implement, particularly when separate models provide conflicting indications. Moreover, auditors can perform targeted audits focusing on the more relevant types of financial statement fraud succeeding a better performance and spending lesser time.

Furthermore, it provides a detailed examination of all factors that can be red flags to detect FFS considering that the majority of FFS derived from humans in managerial positions being a valuable tool for various receivers. For instance, regulators may use the proposed methodology to design preventive policies and corrective actions to identify and address specific types of falsification, as well as to improve auding practices and the quality of financial reports. Stakeholders may use it to receive optimal decisions for their next moves, or investors may obtain all the appropriate information on the likelihood of falsified financial reporting and insights into the different types of falsification, which have material implications for their investment decisions.

In this context, it is worth acknowledging some limitations of the present study. For instance, due to the difficulty of the hand-collection process to gather the data, the analysis not flexible in changes such as the expansion of years. Thus, this difficulty did not give the chance to study the impact of the current crisis (Covid-19) according to the tendency of FFS. Considering the difficulties derived from hand-collected data, it was not convenient to add new ratios or variables after the initial collection of data which will probably contribute to better results. Finally, this hand-collection process was much more time-consuming spending more time in order to avoid typographical errors.

Additionally, the types of auditors' comments considered for the analysis apply to the Greek context based on data obtained from the firms' financial reports and the comments of the auditors restricting an international approach. Moreover, while the present analysis has provided insights into the classification power of multi-label schemes, there is no consideration of a rich set of different classification algorithms and multi-label approaches.

Based on these limitations, there are various future research directions that could be explored to extend and further enhance the proposed approach. Firstly, additional inputs can be considered to describe the pressure, opportunity, capability, and rationalization dimensions of the fraud diamond theory, such as the directors' wages, the employees' wages, the number of subsidiaries, the experience of the executives, and the dedication to business ethics good practices. Moreover, it would be interesting to investigate the theoretical and empirical relationship between the fraud diamond dimensions with the specific types of auditors' comments to obtain insights into the relevance of these dimensions with different forms of FFS. Furthermore, it will be interesting to consider some external red flags for instance crises (wars, environmental disasters, or inflation etc.) and the tendency to occupational fraud.

Linguistic and textual analysis can also be incorporated in the proposed methodology to derive useful information from the financial reports and the auditors' comments (Kydros et al., 2022). Moreover, it is interesting to examine the applicability of similar approaches for non-listed companies, as well as usage of the proposed methodology to data from other countries. Finally, from the methodological point of view, it would be interesting to consider new multi-label classification approaches and investigate the classification performance of the proposed multi-label schemes (CBIC, NNML) to various datasets, beyond financial fraud detection.

Nevertheless, it should be emphasized that the challenges that have recently emerged in the global business environment, such as geopolitical turmoil, ongoing burden due to climate change, and health crisis (Covid-19), and complexity of data, may have a significant impact on the organizational capital and operation of firms (Karpoff, 2021). This is likely to lead to a further increase in fraud events and a change in the practices that firms follow to manipulate and falsify their financial reports. This increases the need to act as soon as possible, enhancing the detection methods using analytical tools such as the ones developed in this study and to keep them up to date with the dynamic nature of the business environment.

References

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A meta-learning framework for detecting financial fraud. *MIS Quarterly*, *36*(4), 1293–1327.

Abdullahi, R., & Mansor, N. (2015a). Concomitant Debacle of Fraud Incidences in the Nigeria Public Sector: Understanding the power of Fraud Triangle Theory. *International Journal of Academic Research in Business and Social Sciences*, *5*(9), 312–326. https://doi.org/10.6007/ijarbss/v5-i9/1833

Abdullahi, R., & Mansor, N. (2015b). Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent For Future Research. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *5*(4), 30–37. https://doi.org/10.6007/ijarafms/v5-i4/1823

Aboud, A., & Robinson, B. (2020). Fraudulent financial reporting and data analytics: an explanatory study from Ireland. *Accounting Research Journal*, *35*(1), 21–36. https://doi.org/10.1108/ARJ-04-2020-0079

ACFE. (2022). Occupational Fraud 2022: A Report to the nations. *ACFE*, 1–96.

Aggarwal, C. C. (2018). Neural Networks and Deep Learning. In *Springer*. https://doi.org/10.1007/978-3-319-94463-0

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*. https://doi.org/10.1016/j.cosrev.2021.100402

Albeksh, H. M. A. (2016). The Crisis of the Ethics of Audit Profession: Collapse of Enron Company and the Lessons Learned. *OALib*, *03*(11), 1–18. https://doi.org/10.4236/oalib.1103205

Albizri, A., Appelbaum, D., & Rizzotto, N. (2019). Evaluation of financial statements fraud detection research: a multi-disciplinary analysis. *International Journal of Disclosure and Governance*, *16*(4), 206–241. https://doi.org/10.1057/s41310-019-00067-9

Albrecht, W. St., Wernz, G. W., & Williams, T. L. (1995). *Fraud: bringing light to the dark side of business*.

Albrecht, W. S., Albrecht, C., & Albrecht, C. C. (2008). Current trends in fraud and its detection. *Information Security Journal*, *17*(1), 2–12. https://doi.org/10.1080/19393550801934331

Amara, I., Amar, A. and, & Jarboui, B. (2013). Detection of Fraud in Financial Statements: French Companies as a Case Study. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *3*(3), 456–472. https://doi.org/10.6007/ijarafms/v3-i3/34

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent Fraud Detection in Financial Statements using Machine Learning and Data Mining: A Systematic Literature Review. *IEEE Access*, 72504–72525. https://doi.org/10.1109/ACCESS.2021.3096799

Awang, N., Hussin, N. S., Razali, F. A., & Abu Talib, S. L. (2020). Fraud Triangle Theory: Calling for New Factors. *Insight Journal*, *7*, 54–64. https://doi.org/10.24191/ij.v7i0.87

Awoyemi, J. O., & Oluwadare, S. A. (2017). Credit card fraud detection using Machine Learning Techniques : A Comparative Analysis in: Computing Networking and Informatics (ICCNI). *2017 International Conference on, IEEE*, 1–9.

Beasley, M. (1996). An empirical analysis of the relation between board of director composition and financial statement fraud. *The Accounting Review*, *71*(4), 443–465.

Behn, B. K., Choi, J. H., & Rang, T. (2008). Audit quality and properties of analyst earnings forecasts. *Accounting Review*, *83*(2), 327–349. https://doi.org/10.2308/accr.2008.83.2.327

Belkaoui, A., & Karnik, P. G. (1989). Determinants of the Corporate Decision to Disclose Social Information. *Accounting, Auditing & Accountability Journal*, *2*(1), 36–51. https://doi.org/10.1108/09513578910132240

Bell, T. B., & Carcello, J. V. (2000). A Decision Aid for Assessing the Likelihood of Fraudulent Financial Reporting. *Auditing: A Journal of Practice and Theory*, *19*(1), 169–184.

Bellman, R. (1961). *On the reduction of dimensionality for classes of dynamic programming processes*. Princeton, NJ: Princeton University Press.

Beneish, M. D. (1999). The Detection of Earnings Manipulation. *Financial Analysts Journal*, *55*(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296

Bequé, A., Coussement, K., Gayler, R., & Lessmann, S. (2017). Approaches for credit scorecard calibration: An empirical analysis. *Knowledge-Based Systems*, *134*, 213–227. http://dx.doi.org/10.1016/j.knosys.2017.07.034

Bierstaker, J., Brody, R., & Pacini, C. (2006). Accountants' perceptions regarding fraud detection and prevention methods. *Managerial Auditing Journal*, *21*(5), 520–535. https://doi.org/10.1108/02686900610667283

Birol, B. (2019). Corporate Governance and Fraud Detection: a Study From Borsa Istanbul. *Eurasian Journal of Business and Management*, *7*(1), 44–64. https://doi.org/10.15604/ejbm.2019.07.01.005

Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, *63*(2–3), 329–357. https://doi.org/10.1016/j.jacceco.2017.03.002

Brazel, J. F., Jones, K. L., & Zimbelman, M. F. (2009). Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research*, *47*(5). https://doi.org/10.1111/j.1475-679X.2009.00349.x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1109/ICCECE51280.2021.9342376

Breiman, L., & Friedman, J. H. (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association*, *80*, 580–619.

Bressler, M. S., & Bressler, L. A. (2007). A Model for Prevention and Detection of Criminal Activity Impacting Small Business. *The Entrepreneurial Executive*, *12*, 23–37.

Brink, A. G., Lowe, D. J., & Victoravich, L. M. (2013). The effect of evidence strength and internal rewards on intentions to report fraud in the dodd-frank regulatory environment. *Auditing*, *32*(3), 87–104. https://doi.org/10.2308/ajpt-50449

Button, M., Blackbourn, D., Lewis, C., & Shepherd, D. (2015). Uncovering the hidden cost of staff fraud: An assessment of 45 cases in the UK. *Journal of Financial Crime*, *22*(2), 170–183. https://doi.org/10.1108/JFC-11-2013-0070

Carcello, J. V., & Nagy, A. L. (2004). Client size, auditor specialization and fraudulent financial reporting. *Managerial Auditing Journal*, *19*(5), 651–668. https://doi.org/10.1108/02686900410537775

Carnegie, G. D., & Napier, C. J. (2010). Traditional accountants and business professionals: Portraying the accounting profession after Enron. *Accounting, Organizations and Society*, *35*(3), 360–376. https://doi.org/10.1016/j.aos.2009.09.002

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, *56*(7), 1146–1160. https://doi.org/10.1287/mnsc.1100.1174

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, *14*(1), 1–22. https://doi.org/10.1145/1541880.1541882

Chen, G., Firth, M., Gao, D. N., & Rui, O. M. (2006). Ownership structure, corporate governance, and fraud: Evidence from China. *Journal of Corporate Finance*, *12*(3), 424–448. https://doi.org/10.1016/j.jcorpfin.2005.09.002

Chen, Y. J., Liou, W. C., Chen, Y. M., & Wu, J. H. (2019). Fraud detection for financial statements of business groups. *International Journal of Accounting Information Systems*, *32*, 1–23. https://doi.org/10.1016/j.accinf.2018.11.004

Chimonaki, C., Papadakis, S., Vergos, K., & Shahgholian, A. (2019). Identification of Financial Statement Fraud in Greece by Using Computational Intelligence Techniques. *Lecture Notes in Business Information Processing*, *345*, 39–51. https://doi.org/10.1007/978-3-030-19037-8_3

Cottrell, D. M., & Glover, S. M. (1993). Finding auditors liable for fraud: What the jury heard in the Phar Mor case. *CPA Journal*, *63*(11), 64. http://www.nysscpa.org/cpajournal/1997/0797/features/f1.htm

Cover, T. M., & Hart, P. E. (1967a). Approximate Formulas for the Information Transmitted by a Discrete Communication Channel. *IEEE Transactions on Information Theory*, *13*(1), 118–119. https://doi.org/10.1109/TIT.1967.1053945

Cover, T. M., & Hart, P. E. (1967b). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*, 113421. https://doi.org/10.1016/j.dss.2020.113421

Craswell, A. T. (1988). The Association Between Qualified Opinions and Auditor Switches. *Accounting and Business Research*, *19*(73), 23–31. https://doi.org/10.1080/00014788.1988.9728832

Cressey, D. (1973). *Other people's money; a study in the social psychology of embezzlement*. Patterson Smith.

Dbouk, B., & Zaarour, I. (2017). Towards a machine learning approach for earnings manipulation detection. *Asian Journal of Business and Accounting*, *10*(2), 215–251.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting Material Accounting

Misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., & Rieder Green, C. (1982). Actual and Perceived Cues to Deception: A Closer Look at Speech. *Basic and Applied Social Psychology*, *3*(4), 291–312. https://doi.org/10.1207/s15324834basp0304_6

Dietterich, T. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning*, *40*, 139–157.

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging Financial Social Media Data for Corporate Fraud Detection. *Journal of Management Information Systems*, *35*(2), 461–487. https://doi.org/10.1080/07421222.2018.1451954

Dorminey, J., Fleming, S., Kranacher, M. J., & Riley, R. A. (2012). The evolution of fraud theory. *Issues in Accounting Education*, *27*(2), 555–579. https://doi.org/10.2308/iace-50131

Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, *90*, 374–393. https://doi.org/10.1016/j.eswa.2017.08.030

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *Journal of Finance*, *65*(6), 2213–2253. https://doi.org/10.1111/j.1540-6261.2010.01614.x

Fanning, K. M., & Cogger, K. O. (1998). Neural detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management*, *7*(1), 21–41.

Fernández-Gámez, M. A., García-Lagos, F., & Sánchez-Serrano, J. R. (2016). Integrating corporate governance and financial variables for the identification of qualified audit opinions with neural networks. *Neural Computing and Applications*, *27*(5), 1427–1444. https://doi.org/10.1007/s00521-015-1944-6

Feroz, E., Kwon, T., Pastena, V., & Park, K. (2000). The efficacy of red flags in predicting the SEC's targets: An artificial neural networks approach. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *9*(3), 145–157.

Feroz, E., Park, K., & Pastena, V. (1991). The Financial and Market Effects of the SEC's Accounting and Auditing Enforcement Releases. *Journal of Accounting Research*, *29*, 107–148.

Firth, M., Rui, O. M., & Wu, W. (2011). Cooking the books: Recipes and costs of falsified financial statements in China. *Journal of Corporate Finance*, *17*(2), 371–390. https://doi.org/10.1016/j.jcorpfin.2010.09.002

Fukunaga, K., & Flick, T. (1984). An optimal global nearest neighbour metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(3), 314–318.

Gabbioneta, Cl., Greenwood, R., Mazzola, P., & Minoja, M. (2013). The influence of the institutional context on corporate illegality. *Accounting, Organizations and Society*, *38*(6–7), 484–504. https://doi.org/10.1016/j.aos.2012.09.002

Gaganis, C. (2009). Classification techniques for the identification of falsified financial statements: a comparative analysis. *Intelligent Systems in Accounting, Finance & Management*, *16*(3), 207–229. https://doi.org/10.1002/isaf.303

Gaganis, C., Pasiouras, F., Spathis, C., & Zopounidis, C. (2007). A comparison of nearest neighbours, discriminant and logit models for auditing decisions. *Intelligent Systems in*

Accounting, Finance and Management, *15*(1–2), 23–40. https://doi.org/10.1002/isaf.283

Giovanis, N., Pantelidis, P., Pazarskis, M., & Elekidou, F. (2016). Evaluation of the Labour Productivity during the Economic Crisis in Greece : A Financial Accounting Approach at Industrial Listed Firms. *11th MIBES Conference, Crete, Greece*, *June*, 158–168.

Glancy, F. H., & Yadav, S. B. (2011). A computational model for fi nancial reporting fraud detection. *Decision Support Systems*, *50*(3), 595–601. https://doi.org/10.1016/j.dss.2010.08.010

Goel, S., & Gangolly, J. (2012). Beyond the Numbers: Mining the Annual Reports for Hidden Cues Indicative of Financial Statement Fraud. *Intelligent Systems in Accounting, Finance and Management*, *19*(2), 75–89. https://doi.org/10.1002/isaf.1326

Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, *15*(4), 357–380. https://doi.org/10.1016/j.accinf.2014.05.006

Gupta, R., & Singh Gill, N. (2012). A Data Mining Framework for Prevention and Detection of Financial Statement Fraud. *International Journal of Computer Applications*, *50*(8), 7–14. https://doi.org/10.5120/7789-0889

Hajek, P., & 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. https://doi.org/10.1016/j.knosys.2017.05.001

Han, J., Kamber, M., & Pei, J. (2006). Data mining: Data mining concepts and techniques. In *Morgan Kaufmann Publishers* (3rd ed.). https://doi.org/10.1109/ICMIRA.2013.45

Hansen, J. V., Mcdonald, J. B., Messier, W. F. and Bell, T. B. (1996). A generalized qualitative-response model and the analysis of management fraud. *Management Science*, *42*(7), 1022–1032.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*(398), 371–386. https://doi.org/10.1080/01621459.1987.10478440

Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Routledge. https://doi.org/10.1201/9780203753781

Heryadi, Y., Wulandhari, L. A., & Abbas, B. S. (2016). Recognizing Debit Card Fraud Transaction Using CHAID and K-Nearest Neighbor : Indonesian Bank Case. *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Yogyakarta, Indonesia*.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844.

Hooda, N., Bawa, S., & Rana, P. S. (2019). Optimizing Fraudulent Firm Prediction Using Ensemble Machine Learning: A Case Study of an External Audit. *Applied Artificial Intelligence*, *34*(1), 20–30. https://doi.org/10.1080/08839514.2019.1680182

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.).

John Wiley & Sons, Inc. https://doi.org/10.1002/9781118548387

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, *64*(6), 2833–2855. https://doi.org/10.1287/mnsc.2017.2751

Humpherys, S., Moffitt, K., Burns, M., Burgoon, J., & William, F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585–594. https://doi.org/10.1016/j.dss.2010.08.009

Jackson, K., Holland, D., Albrecht, C., & Woolstenhulme, D. R. (2010). Fraud Isn't Just For Big Business: Understanding the Drivers, Consequences, and Prevention of Fraud in Small Business. *Journal of International Management Studies*, *5*(1), 160–164.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*, 429–449.

Jennings, M. (2004). Ethics, incorporating ethics and professionalism into accounting education and research: A discussion of the voids and advocacy for training in seminal works in business ethics. *Issues in Accounting Education*, *17*(1), 7–26.

Jofre, M., & Gerlach, R. H. (2018). Fighting Accounting Fraud Through Forensic Data Analytics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3176288

Jones, M. (2011). The Impact Of Accounting Scandals On Accounting Students. In *Creative Accounting, Fraud and International Accounting Scandals Edited* (pp. 479–491). John Wiley & Sons, Ltd.

Kamarudin, K. A., Ismail, W. A. W., & Mustapha, W. A. H. W. (2012). Aggressive Financial Reporting and Corporate Fraud. *Procedia - Social and Behavioral Sciences*, *65*, 638–643. https://doi.org/10.1016/j.sbspro.2012.11.177

Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance*, *66*(August 2019), 101694. https://doi.org/10.1016/j.jcorpfin.2020.101694

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *November 2018*, 1–15. https://doi.org/10.1016/j.indmarman.2019.08.003

Kazantzis, C. (2006). *Auditing and Internal Control – A Systematic Approach to Concepts, Principles and Standards* (In Greek). Business Plus.

Kim, Y. J., Baik, B., & Cho, S. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. *Expert Systems with Applications*, *62*, 32–43. https://doi.org/10.1016/j.eswa.2016.06.016

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003. https://doi.org/10.1016/j.eswa.2006.02.016

Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics*, *33*(3), 375–400. https://doi.org/10.1016/S0165-4101(02)00059-9

Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Forecasting Fraudulent Financial Statements using Data Mining. *International Journal of Computational*

Intelligence, *3*(2), 104–110.

Kou, Y., Lu, C. T., Sirwongwattana, S., & Huang, Y. P. (2004). Survey of fraud detection techniques. *Conference Proceeding - IEEE International Conference on Networking, Sensing and Control*, 749–754. https://doi.org/10.1109/icnsc.2004.1297040

KPMG. (2013). *Global profiles of the fraudster*. http://www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/global-profiles-of-the-fraudster/Documents/global-profiles-of-the-fraudster

Krambia-Kapardis, M., Christodoulou, C., & Agathocleous, M. (2010). Neural networks: The panacea in fraud detection? *Managerial Auditing Journal*, *25*(7), 659–678. https://doi.org/10.1108/02686901011061342

Krambia-Kapardis, M., & Tsolakis, C. (2011). Financial Crimes in Business: Prevention, Investigation, Deterrence. In *Kritiki Editions* (In Greek).

Kumar, B. S., & Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*, *114*, 128–147. https://doi.org/10.1016/j.knosys.2016.10.003

Kydros, D., Pazarskis, M., & Karakitsiou, A. (2022). A framework for identifying the falsified financial statements using network textual analysis: a general model and the Greek example. *Annals of Operations Research*, *316*, 513–527. https://doi.org/10.1007/s10479-021-04086-0

Landwehr, J. M., Pregibon, D., & Shoemaker, A. (1984). Graphical Methods for Assessing Logistic Regression Modelles. *Journal of the American Statistical Association*, *79*, 61–63.

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting Deceptive Discussions in Conference Calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/j.1475-679X.2012.00450.x

Leong, C. K. (2015). Credit Risk Scoring with Bayesian Network Models. *Computational Economics Volume*, *47*, 423–446.

Leskovec, J., Rajaraman, A., & Ullman, J. (2020). Mining of Massive Datasets. *Cambridge University Press, Cambridge, MA*.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). Comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, *40*(3), 203–228. https://doi.org/10.1023/A:1007608224229

Lin, C. C., Chiu, A. A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, *89*, 459–470. https://doi.org/10.1016/j.knosys.2015.08.011

Lin, J. W., Hwang, M. I., & Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, *18*(8), 657–665. https://doi.org/10.1108/02686900310495151

Liou, F. M. (2008). Fraudulent financial reporting detection and business failure prediction models: A comparison. *Managerial Auditing Journal*, *23*(7), 650–662. https://doi.org/10.1108/02686900810890625

Liu, C., Chan, Y., Alam Kazmi, S. H., & Fu, H. (2015). Financial Fraud Detection Model: Based on Random Forest. *International Journal of Economics and Finance*, *7*(7), 178–188. https://doi.org/10.5539/ijef.v7n7p178

Lohmann, C., Möllenhoff, S., & Ohliger, T. (2022). Nonlinear relationships in bankruptcy prediction and their effect on the profitability of bankruptcy prediction models. *Journal of Business Economics*, *93*(9), 1661–1690. https://doi.org/10.1007/s11573-022-01130-8

Lohmann, C., & Ohliger, T. (2019). The total cost of misclassification in credit scoring: A comparison of generalized linear models and generalized additive models. *Journal of Forecasting*, *38*(5), 375–389. https://doi.org/10.1002/for.2545

Loughran, T., & Mcdonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, *69*(4), 1643–1671. https://doi.org/10.1111/jofi.12162

Loumiotis, V. (2010). *Special Subjects of Applied Accounting and Auditing of Banks* (In greek). S.O.E.L. Training Institute.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4766–4775.

Mackevičius, J., & Giriūnas, L. (2013). Transformational Research of the Fraud Triangle. *Ekonomika*, *92*(4), 150–163. https://doi.org/10.15388/ekon.2013.0.2336

Madjarov, G., Kocev, D., Gjorgjevikj, D., & Džeroski, S. (2012). An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, *45*(9), 3084–3104. https://doi.org/10.1016/j.patcog.2012.03.004

Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification tech- niques based on KNN and outlier detection, in: Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICBN). *2017 Third International Conference IEEE*, 255–258.

Manoledakis, I. (2005). *Criminal Law, Summary of the General Part*. Sakkoula Publications.

Manurung, D. T. H., & Hadian, N. (2013). Detection Fraud of Financial Statement with Fraud Triangle Proceedings of 23rd International Business Research Conference. *Proceedings of 23rd International Business Research Conference*. https://doi.org/10.13140/RG.2.1.1736.6806

Marra, A., Mazzola, P., & Prencipe, A. (2011). Board monitoring and earnings management pre- and post-IFRS. *International Journal of Accounting*, *46*(2), 205–230. https://doi.org/10.1016/j.intacc.2011.04.007

Mayhew, B. W., & Murphy, P. R. (2014). The impact of authority on reporting behavior, rationalization and affect. *Contemporary Accounting Research*, *31*(2), 420–443. https://doi.org/10.1111/1911-3846.12037

Mohammadi, M., Yazdani, S., Khanmohammadi, M., & Maham, K. (2020). Financial Reporting Fraud Detection : An Analysis of Data Mining Algorithms. *International Journal of Finance and Managerial Accounting*, *4*(16), 1–12.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

O'Sullivan, F., Yandell, B., & Raynor, W. (1986). Automatic Smoothing of Regression Functions in Generalized Linear Models. *Journal of the American Statistical Association*, *81*, 96–103.

Omar, N., Johari, Z. A., & Smith, M. (2017). Predicting fraudulent financial reporting using artificial neural network. *Journal of Financial Crime*, *24*(2), 362–387. https://doi.org/10.1108/JFC-11-2015-0061

Ozcelik, H. (2020). An Analysis of Fraudulent Financial Reporting Using the Fraud Diamond Theory Perspective: An Empirical Study on the Manufacturing Sector Companies Listed on the Borsa Istanbul. *Contemporary Studies in Economic and Financial Analysis*, *102*, 131–153. https://doi.org/10.1108/s1569-375920200000102012

Pai, P. F., Hsu, M. F., & Wang, M. C. (2011). A support vector machine-based model for detecting top management fraud. *Knowledge-Based Systems*, *24*(2), 314–321. https://doi.org/10.1016/j.knosys.2010.10.003

Pazarskis, M., Drogalas, G., & Baltzi, K. (2017). Detecting false financial statements: Evidence from Greece in the period of economic crisis. *Investment Management and Financial Innovations*, *14*(3), 102–112. https://doi.org/10.21511/imfi.14(3).2017.10

Peasnell, K. V., Pope, P. F., & Young, S. (2005). Board Monitoring and Earnings Management: Do Outside Directors Influence Abnormal Accruals? *Journal of Business Finance & Accounting*, *32*(7–8), 1311–1346. https://doi.org/10.2139/ssrn.249557

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing*, *30*(2), 19–50. https://doi.org/10.2308/ajpt-50009

Persons, O. S. (1995). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research*, *11*(3), 38–46.

Pitsela, A. (2011). *The criminological approach to economic crimetle*. Sakkoula Publications.

Price, R. A., Sharp, N. Y., & Wood, D. A. (2011). Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures. *Accounting Horizons*, *25*(4), 755–780. https://doi.org/10.2308/acch-50064

Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, *50*(2), 491–500. https://doi.org/10.1016/j.dss.2010.11.006

Reiner, R. (2017). *Political Economy, Crime, and Criminal Justice*. 116–137.

Rezaee, Z., and Riley, R. (2010). *Financial Statement Fraud: Prevention and Detection*. John Wiley & Sons.

Rezaee, Z. (2005). Causes, consequences, and deterence of financial statement fraud. *Critical Perspectives on Accounting*, *16*(3), 277–298. https://doi.org/10.1016/S1045-2354(03)00072-8

Riahi-Belkaoui, A. (2003). Intellectual capital and firm performance of US multinational firms:A study of the resource-based and stakeholder views. *Journal of Intellectual Capital*, *4*(2), 215–226. https://doi.org/10.1108/14691930310472839

Rudewicz, F. (2011). The Fraud Diamond : Use of Investigative Due Diligence to Identify the "Capability Element of Fraud." *CTTMA Newsletter*, *IV*(1), 1–13.

Sadasivam, G. S., Subrahmanyam, M., Himachalam, D., Pinnamaneni, B. P., & Lakshme, S. M.

(2016). Corporate governance fraud detection from annual reports using big data analytics. *International Journal of Big Data Intelligence*, *3*(1), 51. https://doi.org/10.1504/ijbdi.2016.073895

Sánchez-Aguayo, M., Urquiza-Aguiar, L.-, & Estrada-Jiménez, J. (2021). Fraud detection using the fraud triangle theory and data mining techniques: A literature review. *Computers*, *10*(10), 1–22. https://doi.org/10.3390/computers10100121

Short, R. and Fukunaga, K. (1980). A new nearest neighbor distance measure. *In Proceedings of the Fifth IEEE Computer Society Conference on Pattern Recognition and Image Processing. IEEE Computer Society: Silver Spring, MD.*

Skillicorn, D. B., & Purda, L. (2012). Detecting fraud in financial reports. *Proceedings European Intelligence and Security Informatics Conference, EISIC 2012*, 7–13. https://doi.org/10.1109/EISIC.2012.8

Skousen, C. J., Smith, K. R., & Wright, C. J. (2009). Detecting and predicting financial statement fraud: The effectiveness of the fraud triangle and SAS No. 99. *Advances in Financial Economics*, *13*(99), 53–81. https://doi.org/10.1108/S1569-3732(2009)0000013005

Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *European Accounting Review*, *11*(3), 509–535. https://doi.org/10.1080/0963818022000000966

Supri, Z., Rura, Y., & Pontoh, G. T. (2018). Detection of fraudulent financial statements with fraud diamond. *Journal of Research in Business and Management*, *6*(5), 39–45.

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decision Support Systems*, *74*, 78–87. https://doi.org/10.1016/j.dss.2015.04.006

Tibshirani, R., & Hastie, T. (1987). Local Likelihood Estimation. *Journal of the American Stastical Association*, *82*, 559–567.

Tsoumakas, G., & Katakis, I. (2007). Multi-Label Classification : An overview. *International Journal of Data Warehousing and Mining*, *3*(3), 1–13.

Van Der Aalst, W., Van Hee, K., Van Der Werf, J. M., Kumar, A., & Verdonk, M. (2011). Conceptual model for online auditing. *Decision Support Systems*, *50*(3), 636–647. https://doi.org/10.1016/j.dss.2010.08.014

Verriest, A. J. M. (2014). Auditor governance, institutions and analyst forecast properties: International evidence. *Journal of International Accounting Research*, *13*(1), 1–32. https://doi.org/10.2308/jiar-50490

Wei, Y., Chen, J., & Wirth, C. (2017). Detecting fraud in Chinese listed company balance sheets. *Pacific Accounting Review*, *29*(3), 356–379. https://doi.org/10.1108/PAR-04-2016-0044

Wells, J. T. (2011). *Corporate Fraud Handbook: Prevention and Detection* (3rd ed.). John Wiley, Hoboken, NJ.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers and Security*, *57*, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

Whiting, D. G., Hansen, J. V., McDonald, J. B., Albrecht, C., & Albrecht, W. S. (2012). Machine learning methods for detecting patterns of management fraud. *Computational Intelligence*, *28*(4), 505–527. https://doi.org/10.1111/j.1467-8640.2012.00425.x

Wolfe, D. T., & Hermanson, D. R. (2004). The Fraud Diamond : Considering the Four Elements of Fraud: Certified Public Accountant. *The CPA Journal*, *74*(12), 38–42.

Wyrobek, J. (2020). Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture. *Procedia Computer Science*, *176*, 3037–3046. https://doi.org/10.1016/j.procs.2020.09.335

Xia, H., Liu, J., & Zhang, Z. J. (2020). Identifying Fintech risk through machine learning: analyzing the Q&A text of an online loan investment platform. *Annals of Operations Research*, *333*(2), 579–599. https://doi.org/10.1007/s10479-020-03842-y

Xie, B., Davidson, W. N., & Dadalt, P. J. (2003). Earnings management and corporate governance: The role of the board and the audit committee. *Journal of Corporate Finance*, *9*(3), 295–316. https://doi.org/10.1016/S0929-1199(02)00006-8

Yang, R., & Jiang, Q. (2020). Detecting Falsified Financial Statements Using a Hybrid SM-UTADIS Approach: Empirical Analysis of Listed Traditional Chinese Medicine Companies in China. *Discrete Dynamics in Nature and Society*, 1–15. https://doi.org/10.1155/2020/8865489

Yeh, C. C., Chi, D. J., & Hsu, M. F. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, *37*(2), 1535–1541. https://doi.org/10.1016/j.eswa.2009.06.088

Yue, D., Wu, X., Shen, N., & Chu, C. H. (2009). Logistic regression for detecting fraudulent financial statement of listed companies in China. *International Conference on Artificial Intelligence and Computational Intelligence, IEEE*, 104–108.

Zerban, A. M. (2018). Enron of Saudi Arabia: Corporate Accounting and Auditing Failures. *Open Journal of Accounting*, *07*(01), 1–18. https://doi.org/10.4236/ojacct.2018.71001

Zhang, M. L., Li, Y. K., Liu, X. Y., & Geng, X. (2018). Binary relevance for multi-label learning: an overview. *Frontiers of Computer Science*, *12*(2), 191–202. https://doi.org/10.1007/s11704-017-7031-7

Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, *40*(7), 2038–2048. https://doi.org/10.1016/j.patcog.2006.12.019

Zhang, Y., Hu, A., Wang, J., & Zhang, Y. (2021). Detection of fraud statement based on word vector: Evidence from financial companies in China. *Finance Research Letters*, *86*, 102477. https://doi.org/10.1016/j.frl.2021.102477

Zhou, F., Zhang, Z., Yang, J., Su, Y., & An, Y. (2018). Delisting pressure, executive compensation, and corporate fraud: Evidence from China. *Pacific Basin Finance Journal*, *48*, 17–34. https://doi.org/10.1016/j.pacfin.2018.01.003

Appendix

Table A1: Means of the variables for different types of auditors' comments

	AC1		AC2		AC3		AC4		AC5		AC6	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
DIRCHANGE	0.07	0.03	0.06	0.06	0.06	0.06	0.06	0.05	0.07	0.03	0.06	0.11
SHARE&BOARD	0.75	0.76	0.76	0.72	0.77	0.58	0.75	0.80	0.75	0.78	0.77	0.55
FAMILY	0.28	0.29	0.28	0.29	0.28	0.25	0.27	0.39	0.28	0.32	0.28	0.32
BIG4	0.15	0.17	0.14	0.21	0.14	0.27	0.14	0.26	0.13	0.28	0.14	0.28
P/BV	5.33	2.00	4.17	6.97	4.68	3.37	3.80	12.67	4.41	5.59	4.91	0.06
COSTSOLD/AP	6.13	3.37	5.88	3.19	5.70	3.46	5.63	4.11	5.75	3.85	5.73	2.49
CA/CL	2.14	0.73	2.00	0.73	1.93	0.62	1.91	0.87	1.98	0.74	1.91	0.63
AMORT/FA	0.08	0.06	0.07	0.08	0.07	0.07	0.07	0.06	0.07	0.07	0.07	0.07
EBIT/TA	0.04	-0.04	0.03	-0.03	0.03	-0.04	0.03	-0.05	0.03	-0.03	0.03	-0.03
LN(TA)	18.42	17.82	18.32	18.08	18.29	18.21	18.33	17.83	18.35	17.87	18.28	18.36
TL/TA	0.58	1.26	0.62	1.44	0.68	1.30	0.70	1.17	0.67	1.17	0.69	1.38
SALES/TA	0.80	0.54	0.77	0.54	0.75	0.55	0.76	0.53	0.77	0.52	0.75	0.52
AUDSWI	0.05	0.05	0.05	0.06	0.06	0.00	0.05	0.12	0.05	0.08	0.05	0.04

Table A2: Statistical significance of the variables in identifying different types of auditors' comments (p-values from t test / Mann-Whitney U test / χ^2 test)

	AC1	AC2	AC3	AC4	AC5	AC6
DIRCHANGE	- / - / 0.09	- / - / 0.77	- / - / 0.95	- / - / 0.57	- / - / 0.16	- / - / 0.11
SHARE&BOARD	- / - / 0.95	- / - / 0.37	- / - / 0.00	- / - / 0.37	- / - / 0.56	- / - / 0.00
FAMILY	0.24 / 0.34 / -	0.83 / 0.69 / -	0.09 / 0.06 / -	0.00 / 0.00 / -	0.02 / 0.05 / -	0.21 / 0.66 / -
BIG4	- / - / 0.40	- / - / 0.06	- / - / 0.01	- / - / 0.01	- / - / 0.00	- / - / 0.01
SECTOR	- / - / 0.00	- / - / 0.00	- / - / 0.56	- / - / 0.00	- / - / 0.00	- / - / 0.55
P/BV	0.16 / 0.00 / -	0.33 / 0.25 / -	0.71 / 0.61 / -	0.07 / 0.52 / -	0.76 / 0.01 / -	0.08 / 0.11 / -
COSTSOLD/AP	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.10 / 0.01 / -	0.01 / 0.00 / -	0.00 / 0.00 / -
CA/CL	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -
AMORT/FA	0.04 / 0.00 / -	0.47 / 0.64 / -	0.99 / 0.31 / -	0.22 / 0.09 / -	0.49 / 0.12 / -	0.58 / 0.09 / -
EBIT/TA	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -
LN(TA)	0.00 / 0.00 / -	0.12 / 0.79 / -	0.69 / 0.98 / -	0.02 / 0.02 / -	0.01 / 0.01 / -	0.72 / 0.30 / -
TL/TA	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.00 / -
SALES/TA	0.00 / 0.00 / -	0.00 / 0.01 / -	0.00 / 0.10 / -	0.00 / 0.00 / -	0.00 / 0.00 / -	0.00 / 0.02 / -
AUDSWI	- / - / 0.95	- / - / 0.89	- / - / 0.04	- / - / 0.01	- / - / 0.19	- / - / 0.60

Note: The p-values from the t test and the Mann-Whitney U test are reported only for quantitative variables. For categorical variables, the results of the χ^2 are shown.