



Πανεπιστήμιο Μακεδονίας
Σχολή Επιστημών Πληροφορίας
Τμήμα Εφαρμοσμένης Πληροφορικής-Πληροφοριακά Συστήματα

Πτυχιακή Εργασία

**ΠΡΟΒΛΕΨΗ ΠΤΩΧΕΥΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΤΗΣ Ε.Ε ΜΕ
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ**

της

Μαυριδόγλου Μαρία

ΑΜ: iis22060

Επιβλέπων Καθηγητής: Απόστολος Δασίλας

Θεσσαλονίκη, Μάιος 2025

Ευχαριστίες

Θα ήθελα ευχαριστήσω τον επιβλέποντα καθηγητή μου Απόστολο Δασίλα για την πολύτιμη καθοδήγηση, τις χρήσιμες παρατηρήσεις και τη βοήθεια του καθ' όλη τη διάρκεια της εκπόνησης της παρούσας πτυχιακής εργασίας.

Επίσης, εκφράζω την ευγνωμοσύνη μου στην οικογένειά μου και τους φίλους μου για την ηθική και ψυχολογική υποστήριξη τους όλα αυτά τα χρόνια.

Περίληψη

Η πρόβλεψη της χρηματοοικονομικής αποτυχίας θεωρείται ένα από τα σημαντικότερα και κρίσιμα οικονομικά προβλήματα που έχει απασχολήσει πληθώρα ερευνητών. Η εμφάνιση της συνεπάγεται επιπτώσεις όχι μόνο για την επιχειρηματική βιωσιμότητα, αλλά και για τα ενδιαφερόμενα μέλη και την οικονομία στο σύνολο της.

Η παρούσα μελέτη στοχεύει στην ανάπτυξη και την αξιολόγηση μοντέλων της Μηχανικής Μάθησης, με στόχο την πρόβλεψη πτώχευσης ευρωπαϊκών επιχειρήσεων που δραστηριοποιούνται στους τομείς της τεχνολογίας, των τηλεπικοινωνιών και των υπολογιστών. Για τις ανάγκες της ανάλυσης, αξιοποιήθηκαν ιστορικά, οικονομικά στοιχεία για 180 επιχειρήσεις, εκ των οποίων οι 170 ήταν υγιείς, ενώ 10 πτωχευμένες, καθιστώντας το σύνολο ιδιαίτερα μη ισορροπημένο. Για το λόγο αυτό, δημιουργήθηκε η ανάγκη να χρησιμοποιηθούν οι τεχνικές SMOTE και Random Under-Sampling, ώστε να εξισορροπηθεί το σύνολο δεδομένων πριν την εκπαίδευση των μοντέλων.

Στο πλαίσιο της μελέτης εφαρμόστηκαν οι εξής επτά αλγόριθμοι: Random Forest, XGBoost, Logistic Regression, SVC, Naive Bayes, Decision Tree και Multi-Layer Perceptron, η αξιολόγηση των οποίων έγινε βάσει δεικτών, όπως η ισορροπημένη ακρίβεια (Balanced Accuracy), η ευαισθησία (Sensitivity), η ειδικότητα (Specificity) και το AUC. Με βάση τα δεδομένα που χρησιμοποιήθηκαν, η τεχνική Decision Tree σημείωσε την υψηλότερη ακρίβεια πρόβλεψης, της τάξης 95%, καταφέροντας να ταξινομήσει σωστά όλες τις πτωχευμένες επιχειρήσεις και το 91% των υγιών επιχειρήσεων. Ακολουθούν οι αλγόριθμοι Random Forest και XGBoost με ακρίβεια ίση με 94%. Την χειρότερη επίδοση συγκέντρωσε ο SVC, ο οποίος παρότι κατάφερε να αναγνωρίσει σωστά όλες τις πτωχευμένες επιχειρήσεις, δεν κατάφερε να ταξινομήσει ορθά το σύνολο των υγιών επιχειρήσεων, παρουσιάζοντας ειδικότητα (specificity) μόλις 38%.

Τα ευρήματα αποδεικνύουν πως τα μοντέλα μηχανικής μάθησης μπορούν να παρέχουν σημαντικά και ακριβείς συμπεράσματα, εφόσον συνοδεύονται από κατάλληλη επεξεργασία και εξισορρόπηση δεδομένων. Η μελέτη συμβάλλει στην σχετική βιβλιογραφία και υπογραμμίζει την σημασία της Μηχανικής Μάθησης στην πρόβλεψη και πρόληψη της οικονομικής αποτυχίας των επιχειρήσεων.

Λέξεις-κλειδιά: Πτώχευση, Χρηματοοικονομικός Κίνδυνος, Μηχανική Μάθηση, Random Forest, XGBoost, Logistic Regression, Decision Tree, SVC, Naive Bayes, Multi-Layer Perceptron, Ανισόρροπα Δεδομένα.

Abstract

The prediction of financial failure is considered one of the most important and critical economic challenges, having attracted the interest of a lot of researchers over the years. Its occurrence carries important consequences not only for financial sustainability but also for stakeholders and the economy as a whole.

This thesis aims to develop and evaluate models of machine learning for the prediction of bankruptcy in European companies operating in the fields of technology, telecommunications and computing. For the needs of the analysis, historical financial data for 190 companies were used, of which 180 were healthy and 10 were bankrupt, rendering the dataset highly imbalanced. To solve this issue, the SMOTE and Random Under-Sampling techniques were applied to balance the dataset before the training of the models.

The following seven machine learning algorithms were applied in the study: Random Forest, XGBoost, Logistic Regression, SVC, Naive Bayes, Decision Tree και Multi-Layer Perceptron, which were evaluate based on Balanced Accuracy, Sensitivity, Specificity and AUC. According to results, the Decision Tree achieved the highest balanced accuracy (95%), classified successfully all bankrupt companies and 91% of healthy. The Random Forest and XGBoost followed closely with a balanced accuracy of 94%. The worst performance was recorded by the SVC, which, although it correctly identified all bankrupt companies, failed to classify the majority of solvent firms, achieving a specificity of 38%.

The findings demonstrate that machine learning models can provide valuable and accurate insights when supported by appropriate preprocessing and data balanced techniques. This study contributes to existing literature and highlights the significance of machine learning in the prediction and prevention of business failure.

Keywords: Bankruptcy, Financial Risk, Machine Learning, Random Forest, XGBoost, Logistic Regression, Decision Tree, SVC, Naive Bayes, Multi-Layer Perceptron, Imbalanced Data.

Πίνακας Περιεχομένων

Κεφάλαιο 1.....	9
Εισαγωγή	9
1.1 Γενικό Υπόβαθρο	9
1.2 Σκοπός.....	10
1.3 Δομή.....	10
Κεφάλαιο 2.....	12
Βιβλιογραφική Ανασκόπηση.....	12
2.1 Η έννοια της χρεοκοπίας.....	12
2.2 Η πτώχευση στην Ευρωπαϊκή Ένωση.....	13
2.3 Πρόγνωση της Χρηματοοικονομικής Αποτυχίας.....	16
2.4 Μοντέλα Πρόβλεψης Πτώχευσης	17
2.4.1 Στατιστικές Μεθοδολογίες Πρόβλεψης.....	17
2.4.2 Σύγχρονες Μεθοδολογίες Πρόβλεψης	19
2.5 Εμπειρικές Μελέτες	21
Κεφάλαιο 3.....	29
Μεθοδολογία και Δεδομένα.....	29
3.1 Συλλογή Δεδομένων.....	29
3.2 Προ Επεξεργασία και Καθαρισμός Δεδομένων	31
3.3 Αντιμετώπιση Ανισορροπίας Δεδομένων	31
3.3.1 Random Under-Sampling	32
3.3.2 SMOTE.....	32
3.3.3 Εφαρμογή Τεχνικών Εξισορρόπησης στο Σύνολο Εκπαίδευσης.....	33
3.4 Επιλογή και Περιγραφή Αλγορίθμων	34
3.4.1 Random Forest.....	34
3.4.2 Decision Tree.....	34
3.4.3 Logistic Regression	35
3.4.4 XGBoost	35
3.4.5 Support Vector Classifier (SVC).....	35
3.4.6 Naive Bayes	36
3.4.7 Multi-layer Perceptron.....	36

3.5 Διαχωρισμός Train και Test Split	37
3.6 Μετρικές Αξιολόγησης	37
Κεφάλαιο 4.....	39
Αποτελέσματα	39
4.1 Παρουσίαση Αποτελεσμάτων	39
4.1.1 Random Forest.....	39
4.1.2 Decision Tree.....	40
4.1.3 Logistic Regression	42
4.1.4 XGBoost	43
4.1.5 SVC	45
4.1.6 Multi-Layer Perceptron	46
4.1.7 Naive Bayes	48
4.2 Επιλογή Βέλτιστου Αλγορίθμου	49
Κεφάλαιο 5.....	51
Συμπεράσματα, περιορισμοί και προτάσεις για μελλοντική έρευνα	51
5.1 Συμπεράσματα	51
5.2 Περιορισμοί και προτάσεις για μελλοντική έρευνα.....	52
Βιβλιογραφία	53
Παράρτημα Α: Πηγαίος Κώδικας Υλοποίησης	56

Κατάλογος Σχημάτων

Σχήμα 1: Οι δηλώσεις πτώχευσης στην Ε.Ε.....	13
Σχήμα 2: Δηλώσεις πτώχευσης ανά κράτος-μέλος.....	14
Σχήμα 3: Δηλώσεις πτώχευσης ανά τομέα δραστηριότητας.	15
Σχήμα 4: Κατανομή επιχειρήσεων με βάση την ενεργή ή ανενεργή κατάστασή τους.....	31
Σχήμα 5: Κατανομή δεδομένων πριν και μετά τις τεχνικές εξισορρόπησης.	33
Σχήμα 6: Παράδειγμα Confusion Matrix.....	37
Σχήμα 7: Confusion Matrix για τον Random Forest.	40
Σχήμα 8: Η καμπύλη ROC για τον Random Forest.....	40
Σχήμα 9: Confusion Matrix για τον Decision Tree.	41
Σχήμα 10: Η καμπύλη ROC για τον Random Forest.....	42
Σχήμα 11: Confusion Matrix για τον Logistic Regression.	43
Σχήμα 12: Η καμπύλη ROC για τον Logistic Regression.	43
Σχήμα 13: Confusion Matrix για τον XGBoost.....	44
Σχήμα 14: Η καμπύλη ROC για τον XGBoost.	45
Σχήμα 15: Confusion Matrix για τον SVC.	46
Σχήμα 16: Η καμπύλη ROC για τον SVC.	46
Σχήμα 17: Confusion Matrix για τον Multi-Layer Perceptron.	47
Σχήμα 18: Η καμπύλη ROC για τον Multi-Layer Perceptron.	48
Σχήμα 19: Confusion Matrix για τον Naive Bayes.....	49
Σχήμα 20: Η καμπύλη ROC για τον Naive Bayes.....	49

Κατάλογος Πινάκων

Πίνακας 1: Περιγραφή μεταβλητών.	30
Πίνακας 2: Αποτελέσματα απόδοσης Random Forest.....	39
Πίνακας 3: Αποτελέσματα απόδοσης Decision Tree.....	41
Πίνακας 4: Αποτελέσματα απόδοσης Logistic Regression.	42
Πίνακας 5: Αποτελέσματα απόδοσης XGBoost.	44
Πίνακας 6: Αποτελέσματα απόδοσης για τον SVC.	45
Πίνακας 7: Αποτελέσματα απόδοσης για τον Multi-Layer Perceptron.	47
Πίνακας 8: Αποτελέσματα απόδοσης για τον Naive Bayes.	48
Πίνακας 9: Συγκεντρωτικά αποτελέσματα απόδοσης.	50

Κεφάλαιο 1

Εισαγωγή

1.1 Γενικό Υπόβαθρο

Η οικονομική σταθερότητα και βιωσιμότητα των επιχειρήσεων μιας χώρας αποτελεί έναν από τους βασικότερους παράγοντες που επηρεάζουν την ανάπτυξη και την εμφάνισή της στην αγορά. Το περιβάλλον μέσα στο οποίο δραστηριοποιούνται μεταβάλλεται συνεχώς ανεξάρτητα από τον κλάδο, το μέγεθος ή την τοποθεσία τους, με αποτέλεσμα να αναζητούν συνεχώς τρόπους προσαρμογής προκειμένου να παραμείνουν ενεργές, υγιείς αλλά και ανταγωνιστικές. Η συνεχής πρόοδος της τεχνολογίας, οι μεταβαλλόμενες οικονομικές συνθήκες και η εντεινόμενη παγκοσμιοποίηση, ενισχύουν την πολυπλοκότητα του επιχειρηματικού περιβάλλοντος και δημιουργούν καταστάσεις που απειλούν την επιβίωση των οικονομικών οντοτήτων.

Στον παγκόσμιο επιχειρηματικό κόσμο, η χρηματοοικονομική αποτυχία μιας επιχείρησης είναι ένα σοβαρό γεγονός με συνέπειες όχι μόνο για την ίδια την εταιρεία και το εσωτερικό της, μετόχους και ανθρώπινο δυναμικό, αλλά και για το ευρύτερο οικονομικό σύστημα, όπως τις τράπεζες, τους προμηθευτές, το αγοραστικό κοινό, το κράτος και γενικότερα τον τομέα στον οποίο δραστηριοποιείται (Mbat & Eyo, 2013). Στο πλαίσιο αυτό, η ικανότητα πρόβλεψης μιας επικείμενης χρεοκοπίας αποτελεί σημαντικό εργαλείο στα χέρια των επενδυτών και των χρηματοπιστωτικών ιδρυμάτων, καθώς δίνει τη δυνατότητα τόσο για άμεσο εντοπισμό και αξιολόγηση του χρηματοοικονομικού κινδύνου, όσο και για την λήψη έγκαιρων και στρατηγικών κινήσεων για την αποφυγή ή την αντιμετώπισή του.

Ιστορικά, η πρόβλεψη της πτώχευσης έχει βασιστεί κυρίως σε παραδοσιακές μεθόδους όπως η ανάλυση χρηματοοικονομικών δεικτών, η ποσοτική αξιολόγηση Ισολογισμών και τα λογιστικά μοντέλα. Χαρακτηριστικό παράδειγμα αποτελεί το Z-score του Altman, το οποίο αποτελεί για δεκαετίες το βασικό εργαλείο για την αξιολόγηση της χρηματοοικονομικής κατάστασης της εξεταζόμενης οντότητας (Altman, 1968). Παρότι η χρήση των μοντέλων αυτών προσφέρει ακόμη και σήμερα σημαντικά και χρήσιμα συμπεράσματα, αδυνατούν να προσαρμοστούν σε δυναμικά περιβάλλοντα, όπως και να ανιχνεύσουν περίπλοκες και μη γραμμικές σχέσεις μεταξύ των μεταβλητών που χρησιμοποιούνται για την πρόβλεψη (Aziz & Dar, 2006).

Με την έλευση της ψηφιακής εποχής και την ραγδαία αύξηση της διαθεσιμότητας οικονομικών δεδομένων, άρχισαν να αναπτύσσονται μοντέλα Μηχανικής Μάθησης όπου η αξιοποίηση τους έρχεται να ενισχύσει σημαντικά την ικανότητα των οργανισμών να αναγνωρίζουν καταστάσεις αυξημένου κινδύνου χρεοκοπίας (Barboza et al., 2017). Η ικανότητα των τεχνικών αυτών να εντοπίζουν μη γραμμικές σχέσεις μεταξύ μεταβλητών, να προσαρμόζονται σε μεγάλα και πολύπλοκα σύνολα δεδομένων και να επιτυγχάνουν

αυξημένα ποσοστά ακρίβειας όχι μόνο συμβάλλει στην καλύτερη αξιολόγηση του χρηματοπιστωτικού κινδύνου, αλλά και στην ενίσχυση της διαφάνειας των αγορών, ενώ παράλληλα συνεισφέρει στην αποτροπή ευρύτερων οικονομικών κρίσεων (Geng et al., 2015).

Ως εκ τούτου, η ενσωμάτωση της μηχανικής μάθησης στη διαδικασία πρόβλεψης της εταιρικής αποτυχίας αποτελεί όχι μόνο μια τεχνολογική καινοτομία, αλλά και μια αναγκαία εξέλιξη στη σύγχρονη χρηματοοικονομική ανάλυση, με σκοπό να ενισχύσει και να προστατέψει την αγορά από μελλοντικές αστάθειες (Lahmiri & Bekiros, 2024). Η κατανόηση και αξιοποίηση αυτών των εργαλείων λειτουργεί ως κινητήρια δύναμη για τους οργανισμούς που επιθυμούν να παραμείνουν αλώβητες και ανταγωνιστικές στο απαιτητικό και αβέβαιο οικονομικό περιβάλλον του σήμερα.

1.2 Σκοπός

Η παρούσα μελέτη επικεντρώνεται στην αξιολόγηση της ικανότητας μοντέλων της μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν, ώστε να προβλέψουν μια πιθανή πτώχευση. Πιο συγκεκριμένα, η εργασία αξιοποιεί οικονομικά, ιστορικά στοιχεία από πληθώρα επιχειρήσεων της Ευρωπαϊκής Ένωσης, με δραστηριοποίηση στους τομείς των τηλεπικοινωνιών, των υπολογιστών και της τεχνολογίας μεταξύ των χρονολογιών 2014 έως 2023. Το εν λόγω σύνολο δεδομένων αποτελείται τόσο από υγιείς επιχειρήσεις, όσο και από ένα μικρό ποσοστό επιχειρήσεων που έχουν οδηγηθεί σε χρεοκοπία. Χρησιμοποιώντας αυτά τα δεδομένα και αλγορίθμους μηχανικής μάθησης γίνεται εκπαίδευση των μοντέλων πρόβλεψης, ώστε να χαρακτηρίσουν μια επιχείρηση είτε ως πτωχευμένη είτε ως υγιή.

Στόχος είναι η αξιολόγηση και η σύγκριση των αποτελεσμάτων από διαφορετικούς αλγορίθμους Μηχανικής Μάθησης ικανών να προβλέψουν με ακρίβεια την πιθανότητα πτώχευσης μια επιχείρησης βάσει των οικονομικών της δεικτών, με σκοπό τον προσδιορισμό του περισσότερο αποτελεσματικού για την εν λόγω λειτουργία. Πέρα όμως από αυτό, μέσα από την έρευνα δίνονται απαντήσεις σε ερωτήσεις, όπως ποιοι χρηματοοικονομικοί δείκτες βοηθάνε περισσότερο στην πρόβλεψη ενός τέτοιου σκοπού, τι προβλήματα μπορούν να προκύψουν και πως μπορούν να αντιμετωπιστούν.

1.3 Δομή

Στο κεφάλαιο 2 εξετάζεται η έννοια της πτώχευσης, αναλύονται οι βασικοί παράγοντες οδηγούν σε αυτή, καθώς και οι επιπτώσεις που επιφέρει τόσο σε μικροοικονομικό, όσο και σε μακροοικονομικό επίπεδο. Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα επίσημα στατιστικά στοιχεία που δημοσίευσε η Eurostat, σχετικά με τον αριθμό των πτωχεύσεων συνολικά στην Ευρωπαϊκή Ένωση, καθώς και ανά χώρα και τομέα δραστηριότητας. Ακολουθώντας, αναφέρεται η σημασία της πρόβλεψης της χρηματοοικονομικής αποτυχίας, ενώ εξετάζονται οι μέθοδοι που συμβάλλουν στο εγχείρημα αυτό. Τέλος, παρατίθενται εμπειρικές μελέτες

που έχουν πραγματοποιηθεί με γνωστικό αντικείμενο την πρόβλεψη της πτώχευσης, εστιάζοντας τόσο στην σύγκριση παραδοσιακών και σύγχρονων τεχνολογιών, όσο και στην ανάπτυξη νέων προγνωστικών μοντέλων.

Το κεφάλαιο 3 παρέχει μια αναλυτική παρουσίαση της μεθοδολογία που ακολουθήθηκε για την εκπόνηση της παρούσας μελέτης. Παρουσιάζονται στοιχεία σχετικά με την συλλογή, την προεπεξεργασία και την ανάλυση των δεδομένων, καθώς και την επιλογή των αλγορίθμων που χρησιμοποιήθηκαν για την πρόβλεψη της χρεοκοπίας. Ακόμη, παρουσιάζονται οι μετρικές αξιολόγησης που επιλέχθηκαν, όπως επίσης και οι τεχνικές που χρησιμοποιήθηκαν για το πρόβλημα της ανισορροπίας των δύο κλάσεων (πτώχευμένες και υγιείς επιχειρήσεις) που κυριαρχούσε στο σύνολο δεδομένων.

Στο κεφάλαιο 4 παρουσιάζονται ενδελεχώς τα αποτελέσματα της έρευνας. Συγκεκριμένα, για κάθε αλγόριθμο που εφαρμόστηκε παρατίθενται οι επιδόσεις τους στην ταξινόμηση τόσο των υγιών επιχειρήσεων, όσο και των πτωχευμένων, ενώ αξιολογείται συνολικά η ακρίβεια και η διακριτική του ικανότητα. Ακολουθεί συγκριτική αξιολόγηση των αποτελεσμάτων, με σκοπό να αναδειχθεί το μοντέλο με την καλύτερη και πιο ισορροπημένη απόδοση.

Τέλος, στο κεφάλαιο 5 γίνεται μια συνολική αποτίμηση της παρούσας μελέτης. Αφενός εξετάζεται η χρησιμότητά και η συνδρομή των αποτελεσμάτων στον επιχειρηματικό και χρηματοοικονομικό τομέα, κυρίως για τη λήψη διορθωτικών μέτρων για τη διαχείριση του χρηματοοικονομικού κινδύνου. Αφετέρου, επισημαίνονται ορισμένες αδυναμίες που εμφανίζει η μελέτη και διατυπώνονται προτάσεις για περαιτέρω έρευνα, με στόχο την ενίσχυση της υφιστάμενης προσέγγισης.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

2.1 Η έννοια της χρεοκοπίας

Ο όρος χρεοκοπία είναι συχνά συνώνυμη της πτώχευσης και αφορά την αδυναμία ενός φυσικού ή νομικού προσώπου να καλύψει τις χρηματοοικονομικές του υποχρεώσεις και να εξυπηρετήσει τις δανειακές του οφειλές προς τρίτους, δηλαδή τους πιστωτές του. Συγκεκριμένα, σύμφωνα με το Άρθρο 3, Παρ. 1 του Πτωχευτικού Κώδικα:

«Σε πτώχευση κηρύσσεται ο οφειλέτης που αδυνατεί να εκπληρώνει τις ληξιπρόθεσμες χρηματικές υποχρεώσεις του κατά τρόπο γενικό και νόμιμο (παύση πληρωμών). Δεν αποτελούν εκπλήρωση των υποχρεώσεων οι πληρωμές που γίνονται με δόλια ή καταστρεπτικά μέσα».

Η πτώχευση αποτελεί μία από τις πιο σημαντικές και πολυσύνθετες καταστάσεις που μπορεί να κληθεί να αντιμετωπίσει μια επιχείρηση. Δεν είναι ξαφνικό ή στιγμιαίο γεγονός, αλλά αποτελεί το τελικό στάδιο μιας διαδικασίας όπου συνήθως ξεκινάει με οικονομική αστάθεια, αδυναμία διαχείρισης των ταμειακών ροών και διαρκής υποβάθμιση της χρηματοοικονομικής της θέσης.

Η εμφάνιση της δεν εξαρτάται από ένα και μόνο παράγοντα, αλλά από την συνδυαστική επίδραση τόσο εσωτερικών, όσο και εξωτερικών αιτιών. Οι κυριότερες από αυτές είναι (Bradley & Cowdery, 2004):

1. Οικονομική κρίση: Οι δυσμενείς μακροοικονομικές συνθήκες, όπως η ανεργία, η ύφεση, η απουσία κρατικής χρηματοδότησης και η μείωση της αγοραστικής δύναμης επιδεινώνουν την οικονομική κατάσταση μιας επιχείρησης.
2. Αναποτελεσματική Διοίκηση: Η διοικητική ανεπάρκεια αποτελεί αποτέλεσμα της λήψης λανθασμένων στρατηγικών αποφάσεων, της απουσίας σαφούς οράματος και της έλλειψης εσωτερικής οργάνωσης, οδηγώντας σε αποτυχία του επιχειρηματικού μοντέλου .
3. Αδυναμία κάλυψης βραχυπρόθεσμων υποχρεώσεων: Η βιωσιμότητα μιας επιχείρησης μειώνεται σταδιακά, καθώς αδυνατεί να ανταπεξέλθει στις άμεσες υποχρεώσεις της, όπως η πληρωμή προμηθευτών, μισθοδοσίας, λογαριασμών.
4. Εσκεμμένη παύση πληρωμών και ανακριβής εμφάνιση οικονομικών στοιχείων της επιχείρησης: Η καθυστέρηση ή ακόμη και η άρνηση της επιχείρησης να τακτοποιήσει τις οικονομικές οφειλές, με σκοπό της αποφυγής νομικών συνεπειών, καθώς επίσης και η παραποίηση των οικονομικών καταστάσεων και του Ισολογισμού της μπορεί να οδηγήσουν σε λανθασμένες αποφάσεις και στην έλλειψη εμπιστοσύνης από οικονομικούς οργανισμούς.

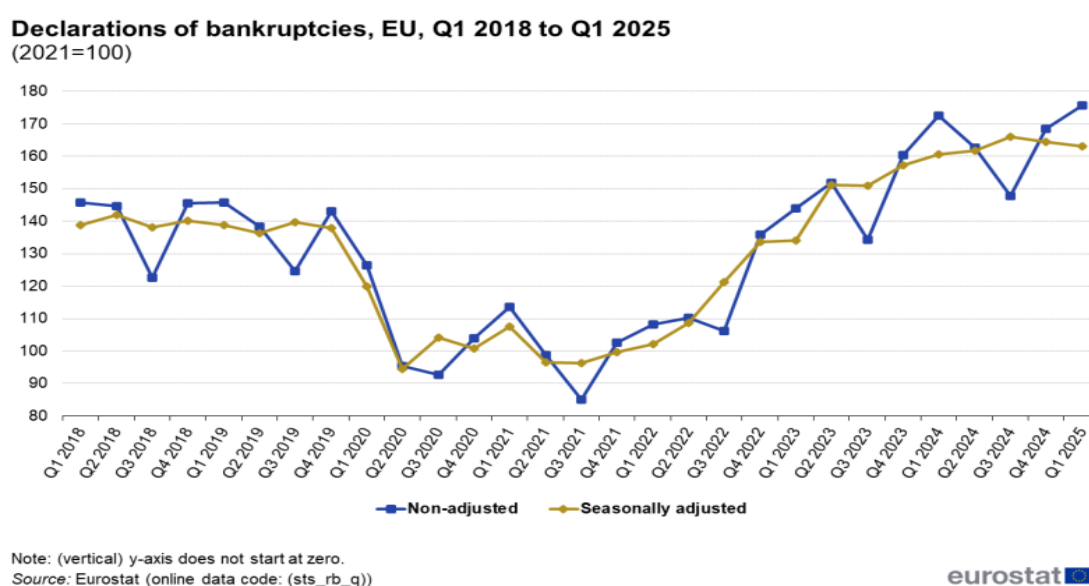
5. Συμβάντα τα οποία δύνανται να επηρεάσουν αρνητικά την οικονομική πορεία ενός κλάδου: Η αδυναμία της επιχείρησης να προσαρμοστεί γρήγορα σε φαινόμενα που αφορούν την είσοδο νέων ανταγωνιστών στην αγορά και την διαρκή τεχνολογική εξέλιξη μπορεί να προκαλέσει την οικονομική της καταστροφή.

Όπως είναι αναμενόμενο, οι αιτίες αυτές οδηγούν σε σοβαρές οικονομικές και κοινωνικές συνέπειες, όχι μόνο για την ίδια την επιχείρηση, αλλά και για το ευρύτερο οικονομικό περιβάλλον (Mbat & Eyo, 2013). Συγκεκριμένα, η χρεοκοπία μιας επιχείρησης σημαίνει το οριστικό της κλείσιμο, με αποτέλεσμα το ανθρώπινο δυναμικό, δηλαδή οι εργαζόμενοι, να χάσουν τις θέσεις εργασίας τους και ταυτόχρονα να αυξηθεί το ποσοστό της ανεργίας. Παράλληλα, επηρεάζονται αρνητικά οι δανειστές και οι επενδυτές της επιχείρησης, καθώς ενδέχεται να αντιμετωπίσουν σοβαρές οικονομικές ζημιές και να μην ανακτήσουν τα οφειλόμενα ποσά. Εκτός αυτών, η εξέλιξη μιας επιχείρησης αποτελεί δομικό στοιχείο της εγχώριας αγοράς, με αποτέλεσμα η παύση της λειτουργίας της να έχει άμεσο αντίκτυπο και στο κράτος, αφού μειώνονται τα φορολογικά έσοδα. Έτσι, αποδυναμώνεται η οικονομική δραστηριότητα της χώρας και κατ' επέκταση η αναπτυξιακή της πορεία.

2.2 Η πτώχευση στην Ευρωπαϊκή Ένωση

Σύμφωνα με τα επίσημα δεδομένα της Eurostat (2025), τα οποία καλύπτουν την περίοδο μεταξύ του πρώτου τριμήνου του 2018 έως και του πρώτου τριμήνου του 2025, παρατηρούνται σημαντικές διακυμάνσεις στον αριθμό χρεοκοπημένων επιχειρήσεων στις χώρες της Ευρωπαϊκής Ένωσης, ιδιαίτερα κατά την περίοδο της πανδημίας COVID-19.

Όπως φαίνεται στο Σχήμα 2.1 που ακολουθεί, οι δηλώσεις πτωχεύσεων από το 2018 έως το 2019 στο σύνολο της Ε.Ε., κινούνται χωρίς σαφές μοτίβο, παρουσιάζοντας ήπιες διακυμάνσεις.



Σχήμα 1: Οι δηλώσεις πτώχευσης στην Ε.Ε.

Ωστόσο, μεταξύ του πρώτου και δεύτερου τριμήνου του 2020 παρατηρείται σημαντική μείωση των πτωχεύσεων, της τάξης του 32,3%, γεγονός που εξηγείται από την κυβερνητική ενίσχυση του επιχειρηματικού τομέα κατά την διάρκεια της πανδημίας του COVID-19.

Από το τρίτο τρίμηνο του 2020 έως και το τρίτο τρίμηνο του 2021 ο αριθμός των δηλώσεων πτώχευσης παραμένει σε χαμηλότερα επίπεδα από την προ πανδημίας περίοδο, δηλαδή μεταξύ του 2018 έως το 2019.

Κατόπιν, από το τέταρτο τρίμηνο του 2021 μέχρι και το τρίτο τρίμηνο του 2024 οι πτωχευμένες επιχειρήσεις αυξάνονται κατά 69% περίπου, σημειώνοντας την μεγαλύτερη αύξηση από το 2018, κυρίως εξαιτίας της απόσυρσης των κρατικών μέτρων στήριξης.

Μεταξύ των κρατών-μελών της Ε.Ε, για τις οποίες υπάρχουν διαθέσιμα οικονομικά στοιχεία, η χώρα που σημείωσε σύμφωνα με το Σχήμα 2.2 την μεγαλύτερη αύξηση στον αριθμό των χρεοκοπημένων επιχειρήσεων από το τέταρτο τρίμηνο του 2024 έως το πρώτο τρίμηνο του 2025 είναι η Ελλάδα, με αύξηση 35,9%, ακολουθούμενη από τη Σουηδία, με 22,7% και την Εσθονία, με 20,4%. Αντίθετα, η Κύπρος, η Μάλτα και η Λετονία παρουσιάζουν μείωση στις επιχειρήσεις που έχουν διακόψει τη λειτουργία τους, με ποσοστά 70%, 66,6% και 21% αντίστοιχα.

Registrations of businesses by country*, Q1 2018 to Q1 2025
(% change compared with the previous quarter)

	Q1 2018	Q2 2018	Q3 2018	Q4 2018	Q1 2019	Q2 2019	Q3 2019	Q4 2019	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022	Q2 2022	Q3 2022	Q4 2022	Q1 2023	Q2 2023	Q3 2023	Q4 2023	Q1 2024	Q2 2024	Q3 2024	Q4 2024	Q1 2025
EU	2.5	0.0	-1.6	3.9	4.6	-4.2	0.7	-0.4	-8.7	-22.6	43.7	-2.6	1.0	4.5	-2.2	2.8	-2.9	0.4	2.1	-0.9	2.1	1.1	-0.3	0.1	2.1	-2.3	0.9	2.6	-5.1
Euro area	2.5	-0.6	-1.6	3.8	5.8	-4.4	1.7	0.7	-9.8	-23.2	45.9	-1.7	-0.4	5.2	-2.6	2.4	-3.0	-0.4	0.9	1.3	2.1	3.8	-0.7	-1.4	3.0	-2.1	1.7	2.2	-5.5
Belgium	33.6	-23.0	-3.3	35.7	1.5	-6.6	5.7	-0.4	-4.8	-22.7	55.8	-8.7	3.3	2.9	-3.3	0.9	-15.8	10.7	6.8	-17.7	6.5	-4.1	-2.0	3.4	23.9	-6.1	-3.1	8.0	-10.1
Bulgaria	-6.4	3.0	-8.2	12.9	-6.5	-4.7	-0.8	-7.0	-11.1	-25.2	31.1	-5.8	13.0	-1.3	0.4	-3.8	-2.9	7.8	2.9	-0.6	6.3	-9.7	-1.7	3.2	1.1	0.5	3.1	-0.2	-3.2
Czechia	0.9	0.4	-5.6	0.9	-9.7	1.3	-0.6	-5.9	-6.7	-5.4	17.7	-13.5	10.6	9.7	0.3	5.6	0.1	10.4	-4.4	3.2	-0.2	-5.8	-1.4	11.2	6.1	-3.0	6.4	-4.6	2.9
Denmark	3.2	3.7	-3.1	8.5	3.1	-9.0	-3.8	-3.4	-7.4	-11.4	27.5	-8.4	-3.3	-1.4	-1.4	-2.2	-10.5	-6.0	4.8	-2.7	4.3	-7.9	8.7	-10.6	8.0	1.2	-9.0	1.3	-1.7
Germany	-0.9	-0.6	-2.8	1.1	3.2	-7.2	3.2	-0.4	-7.9	-18.2	23.4	-1.2	-3.8	0.0	0.5	6.4	-5.0	-3.1	1.3	3.0	20.3	12.9	-10.9	-13.0	-0.8	-3.1	2.3	12.7	3.0
Estonia	2.2	-0.7	0.6	-1.1	0.6	3.1	3.0	3.4	1.8	-16.9	18.8	0.4	0.4	1.5	0.4	28.0	-29.4	-2.8	-1.1	4.8	2.3	0.6	1.6	-2.8	4.7	0.4	-0.3	2.6	-2.9
Ireland	1.7	1.3	-0.6	1.6	8.4	-11.9	0.9	-4.0	-5.0	-41.2	71.0	1.1	-7.7	13.6	-5.5	-0.4	-3.8	-4.9	0.3	2.3	-64.9	20.2	1.3	-0.5	8.1	-7.8	-0.6	-5.2	-10.3
Greece														5.1	1.3	4.1	-0.7	5.6	6.5	2.0	0.1	-0.2	0.8	1.0	8.8	-2.8	-0.5	-0.1	-2.4
Spain	5.6	0.1	-3.1	1.1	3.9	-6.9	0.9	2.9	-10.7	-43.6	100.4	-1.5	3.0	9.1	-2.7	-1.7	0.2	-1.6	-3.3	6.6	9.1	-2.0	4.9	3.6	6.6	0.5	2.8	7.4	-16.1
France	4.5	3.0	0.0	3.8	9.3	-0.7	3.5	2.7	-9.7	-15.1	55.4	-0.8	-0.2	3.1	-6.2	1.2	1.7	-2.2	4.7	1.4	-8.2	1.5	7.9	0.2	2.3	-0.9	-1.0	0.4	-1.2
Croatia	4.3	0.6	0.6	1.2	3.5	3.9	0.4	3.3	-25.2	-35.6	66.4	6.3	-2.6	1.4	9.6	1.0	0.0	5.7	-0.1	-1.0	6.3	0.0	5.1	5.8	-0.8	-0.8	-0.4	-1.4	4.2
Italy	0.4	-2.9	-3.1	3.2	6.3	-1.9	-1.9	2.0	-13.2	-30.4	59.9	-8.8	3.9	4.8	-3.8	5.6	-6.4	-2.1	-3.4	0.9	5.5	-4.1	1.3	0.7	6.6	-5.7	3.4	-2.0	3.8
Cyprus														-3.4	7.5	-1.6	6.3	4.5	-7.0	12.3	-16.0	2.8	-8.3	2.5	1.3	0.7	9.7	-11.5	9.8
Latvia	-8.2	0.5	0.1	2.8	1.4	-8.2	-1.3	6.8	-2.5	-38.3	53.4	-8.9	0.2	4.5	-1.8	-1.6	-10.1	2.8	11.4	0.9	0.3	-6.2	-9.5	9.4	3.2	2.4	2.9	6.3	-0.5
Lithuania	-2.9	-0.9	-1.1	6.0	-2.3	5.2	4.2	-2.6	-7.9	8.5	15.3	0.9	3.7	0.5	-13.4	10.1	-1.4	-3.2	1.0	-4.8	10.0	3.6	2.7	-3.3	1.6	-2.2	2.0	0.2	8.9
Luxembourg	5.4	-9.9	11.3	-0.2	4.4	-0.9	-5.4	-4.9	4.1	-29.5	47.9	-0.1	-0.8	3.2	-5.2	1.6	-4.7	-1.7	-3.3	8.9	8.1	-5.2	-3.0	4.2	-5.5	3.2	6.6	-7.3	4.1
Hungary														-0.9	-0.2	2.1	-3.9	-0.8	16.2	-19.4	-0.7	-3.0	-0.6	-2.7	-2.7	-0.6	2.0	3.7	1.0
Malta	3.7	5.3	2.9	-6.2	-3.9	-9.9	1.6	-4.7	6.7	-26.9	39.3	1.1	-0.5	8.9	-6.7	-11.5	-3.6	1.0	3.9	-3.6	-0.6	4.1	-11.1	2.7	-0.8	-3.8	4.7	5.8	-8.9
Netherlands	-0.9	0.9	5.4	1.1	11.2	-4.9	3.6	-1.3	-6.1	9.8	-12.7	10.2	6.7	2.2	4.7	-1.1	0.9	-3.6	5.3	1.5	0.2	4.2	-1.7	-0.5	-2.6	-5.9	6.1	1.4	-38.4
Austria						0.2	-4.6	-5.8	-0.8	-15.8	38.3	-3.0	21.1	-12.5	-12.7	1.4	-6.6	16.8	-4.7	-1.3	-3.3	4.7	-5.5	5.8	-6.4	5.1	-4.2	2.4	-9.1
Poland	4.5	5.8	-5.1	-1.4	5.2	-4.1	-2.0	-6.6	0.6	-25.8	42.8	-9.1	3.9	6.4	-3.1	8.9	-1.9	-2.2	5.6	-3.9	4.4	-10.7	2.3	7.5	-2.6	-7.5	2.4	0.5	-4.0
Portugal	6.6	4.0	-4.8	13.1	8.0	-12.0	4.4	0.7	-17.2	-40.0	83.5	-6.8	-19.5	32.3	1.3	4.9	-2.7	4.9	4.9	0.8	1.4	-1.6	3.8	-0.6	7.1	-4.4	-4.0	0.2	
Romania	4.8	-10.6	22.7	25.6	3.0	-12.6	-8.6	-2.7	-10.5	-18.4	31.3	18.5	8.0	-8.1	14.5	-5.1	-2.7	19.0	16.1	-22.7	1.7	-8.8	7.9	5.3	-7.2	0.2	-34.0	51.3	-32.3
Slovenia	-0.3	2.5	-1.2	-2.0	4.0	0.4	-3.3	-0.6	-3.5	-30.6	66.8	-17.3	-5.4	23.0	3.6	1.3	-2.6	4.4	-2.4	0.4	1.9	2.7	-2.9	3.8	3.1	-5.1	-1.9	0.5	1.6
Slovakia	-7.5	1.2	3.9	1.3	-1.3	0.6	5.6	-10.2	-2.3	-24.7	42.3	27.0	-26.6	18.5	-8.1	-3.4	-12.3	6.7	-13.6	0.8	-37.2	95.2	17.6	-8.9	-0.8	-4.7	16.0	-13.4	2.3
Finland										-10.3	17.4	4.9	0.7	4.5	-6.1	1.9	-4.7	6.4	-1.7	-2.4	0.7	2.9	0.0	2.0	4.4	1.5	-3.4	-2.8	-1.0
Sweden														-0.3	-8.1	5.6	-5.5	-3.7	5.9	-5.1	0.0	-5.8	-3.3	-0.9	2.9	3.8	0.6	-2.8	1.7
Iceland	-8.2	1.0	1.1	0.8	2.4	-7.1	-5.0	2.4	-5.2	-0.8	45.7	-0.2	8.3	-3.1	0.3	1.2	-5.6	-0.6	7.8	-6.5	1.7	6.3	-2.0	1.0	-4.2	-1.0	-0.6	-0.8	2.0
Norway	0.1	3.3	-5.2	2.6	-0.2	3.1	1.7	6.4	-2.7	-6.3	4.8	1.6	5.4	-7.2	-0.4	1.4	-10.2	-19.6	29.5	3.1	6.9	-9.6	4.7	-9.4	-4.0	7.4	-3.3	-2.3	-3.8

* Seasonally adjusted data.

: Data not available.

Source: Eurostat (online data code: sts_rb_q)

eurostat

Σχήμα 2: Δηλώσεις πτώχευσης ανά κράτος-μέλος.

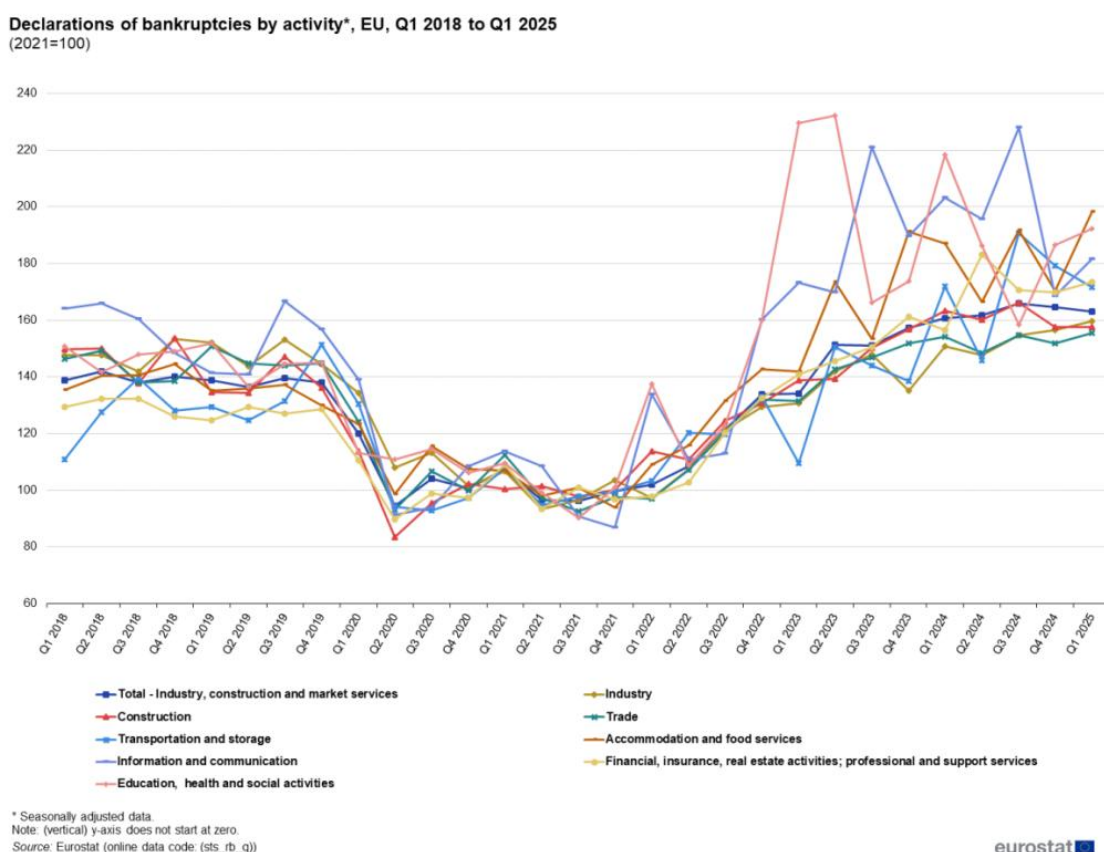
Αντίστοιχα, σε επίπεδο τομέα δραστηριοποίησης, αναλύοντας το Σχήμα 2.3 μπορούμε να κατανοήσουμε ότι το ποσοστό των πτωχευμένων επιχειρήσεων σε όλους τους τομείς στο πρώτο τρίμηνο του 2025 είναι αρκετά υψηλό σε σχέση με αυτό πριν την έξαρση της πανδημίας.

Πιο συγκεκριμένα, στα δύο πρώτα τρίμηνα του 2020 υπάρχει σημαντική μείωση στις δηλώσεις πτώχευσης σε όλους τους τομείς της οικονομίας, λόγω της στήριξης των επιχειρήσεων από το κράτος στη διάρκεια της πανδημίας. Ακολουθεί μια γενική ανοδική πορεία το τρίτο τρίμηνο του 2020, με εξαίρεση στον τομέα των μεταφορών, η οποία σταθεροποιείται μέχρι το δεύτερο τρίμηνο του 2022.

Από το τρίτο τρίμηνο του 2022 μέχρι και το πρώτο του 2024 υπάρχει έξαρση της πτώχευσης, με έντονες διακυμάνσεις στους τομείς της εκπαίδευσης και κοινωνικών υπηρεσιών, της πληροφόρησης και επικοινωνίας, της διαμονής και εστίασης και των μεταφορών.

Οι δηλώσεις πτώχευσης μειώθηκαν κατά το τέταρτο τρίμηνο του 2024, ειδικότερα στους τομείς της πληροφόρησης και επικοινωνίας, της διαμονής και εστίασης, ενώ αύξηση σημειώθηκε μόνο στους τομείς της εκπαίδευσης και κοινωνικών υπηρεσιών, καθώς και στη βιομηχανία. Αυτό συνεχίστηκε μέχρι τις αρχές του 2025, όπου αυξάνονται οι πτωχεύσεις σε όλους τους τομείς, εκτός από αυτούς των μεταφορών και των κατασκευών.

Επομένως, συμπεραίνουμε ότι οι τομείς που εμφανίζουν μεγαλύτερη έκθεση στον πιστωτικό κίνδυνο κατά την εξεταζόμενη περίοδο είναι αυτοί της εκπαίδευσης και κοινωνικών επιστημών, της πληροφορικής και επικοινωνιών και τέλος της διαμονής και της εστίασης.



Σχήμα 3: Δηλώσεις πτώχευσης ανά τομέα δραστηριότητας.

2.3 Πρόγνωση της Χρηματοοικονομικής Αποτυχίας

Όπως αναφέρθηκε προηγουμένως, η χρηματοοικονομική αποτυχία δεν επέρχεται αμέσως μετά την εμφάνιση των πρώτων δυσχερειών. Αντιθέτως, αποτελεί καταλυτικό στάδιο της μακροχρόνιας και σταδιακής χειροτέρευσης της οικονομικής θέσης της επιχείρησης (Aziz & Dar, 2006). Συνεπώς, ο έγκυρος εντοπισμός των ενδείξεων πτώχευσης μπορεί να οδηγήσει στη λήψη αποφάσεων για την αποφυγή της κατάρρευσης.

Η εκτίμηση των επιδόσεων και της βιωσιμότητας της επιχείρησης αποτελεί σημαντικό κομμάτι τόσο για τους άμεσα εμπλεκόμενους στη λειτουργία της, όπως οι ιδιοκτήτες-μέτοχοι και οι εργαζόμενοι, όσο και για τους έμμεσα, όπως οι τράπεζες, οι προμηθευτές, το κράτος (Mbat & Eyo, 2013). Το ενδιαφέρον καθενός, βέβαια, διαφέρει ανάλογα με τα συμφέροντα του. Οι εργαζόμενοι, για παράδειγμα, ενδιαφέρονται για τη διατήρηση της θέσης εργασίας τους, ενώ οι τράπεζες για την επιστροφή των χρημάτων που έδωσαν σε περίπτωση δανειοδότησης ή επένδυσης.

Η ανάπτυξη μοντέλων, ικανών να προβλέψουν το χρηματοπιστωτικό κίνδυνο που αναμένεται να αντιμετωπίσει η επιχείρηση δεν αποτελεί απλώς μια ακαδημαϊκή ή θεωρητική προσέγγιση, αλλά ένα ουσιαστικό και πρακτικό εργαλείο (Geng et al., 2015). Αφενός, μπορούν να λειτουργήσουν ως σύστημα «έγκαιρης προειδοποίησης» για τα διοικητικά στελέχη, ώστε να λάβουν τα κατάλληλα στρατηγικά και διορθωτικά μέτρα για να αποτρέψουν μια δυσμενή εξέλιξη ή σε περίπτωση που αυτή είναι αναπόφευκτη, τη μείωση του συνολικού κόστους που συνεπάγεται. Αφετέρου, μπορούν να χρησιμοποιηθούν από τα χρηματοπιστωτικά ιδρύματα και τους επενδυτικούς φορείς, προκειμένου να χαρακτηρίζουν τις επιχειρήσεις προς δανειοδότηση ή επένδυση και κατ' επέκταση τον κίνδυνο που συνεπάγεται η πραγμάτωση μια τέτοιας ενέργειας (Olson et al., 2012).

Η αξιοπιστία και η χρησιμότητα ενός τέτοιου υποδείγματος εξαρτάται σε μεγάλο βαθμό από το κόστος που προκύπτει από τις λανθασμένες εκτιμήσεις και ταξινομήσεις των επιχειρήσεων, καθώς μπορεί να προκύψουν δύο τύποι σφαλμάτων (Weiss & Carkun, 2005):

Σφάλμα Τύπου I: Όταν η επιχείρηση θεωρείται υγιής, ενώ στην πραγματικότητα πρόκειται να χρεοκοπήσει.

Σφάλμα Τύπου II: Όταν η επιχείρηση θεωρείται λανθασμένα πως θα πτωχεύσει, ενώ στη πραγματικότητα είναι υγιής και βιώσιμη.

Και οι δύο τύποι σφαλμάτων που παρουσιάστηκαν παραπάνω συνδέονται με κάποιο κόστος, το οποίο διαφοροποιείται ανάλογα με την περίπτωση και την οπτική γωνία που αντιμετωπίζει κανείς το πρόβλημα (Jones et al., 2017). Για παράδειγμα, για ένα επενδυτή ή δανειστή το κόστος που ενδέχεται να αντιμετωπίσει στη περίπτωση του σφάλματος Τύπου I είναι η απώλεια του συνολικού κεφαλαίου που πρόκειται να διαθέσει σε μια προβληματική επιχείρηση, σε αντίθεση με τη περίπτωση του σφάλματος Τύπου II, όπου το κόστος αφορά την απώλεια μιας σημαντικής, κερδοφόρας ευκαιρίας.

Από την πλευρά της ίδιας της επιχείρησης που αποτελεί αντικείμενο αξιολόγησης, η σημασία του κόστους αλλάζει. Πιο συγκεκριμένα, η περίπτωση εμφάνισης σφάλματος Τύπου Ι κατά τη διάρκεια εσωτερικού ελέγχου από τα διοικητικά στελέχη, εγκυμονεί την επιδείνωση της κατάστασης, καθώς δε θα μπορέσουν να αναγνωρίσουν και να αντιμετωπίσουν έγκαιρα το πρόβλημα (Beaver, 1966). Βεβαία, αναλύοντας το από διαφορετική οπτική γωνία, ένα τέτοιο σφάλμα μπορεί να λειτουργήσει υπέρ της εξεταζόμενης επιχείρησης, καθώς διατηρεί την εικόνα της σταθερότητας και την εμπιστοσύνη των συναλλασσόμενων με αυτή, καταφέροντας να οδηγήσει ακόμη και στην ανάκαμψη της (Altman, 1968). Ωστόσο, στη περίπτωση που μια υγιής και δυναμική επιχείρηση υπόκειται σε σφάλμα Τύπου ΙΙ, δηλαδή θεωρείται πως είναι προβληματική, το κόστος που αντιμετωπίζει είναι ιδιαίτερα βαρύ. Συγκεκριμένα, το κόστος αυτό συνδέεται με την απώλεια της αξιοπιστίας της στην αγορά, της εμπιστοσύνης της από τρίτους και την δυσκολία πρόσβασης σε κεφάλαια, τα οποία μπορούν να οδηγήσουν σε πραγματικά δυσχερή θέση (Shetty et al., 2022).

2.4 Μοντέλα Πρόβλεψης Πτώχευσης

Η χρηματοοικονομική θεωρία έχει διαχρονικά επιδιώξει με κάθε τρόπο να συμβάλλει στην αποφυγή των δυσάρεστων συνεπειών που συνεπάγεται μια πιθανή χρεοκοπία. Ο σκοπός αυτής της προσπάθειας είναι η έγκαιρη και έγκυρη αναγνώριση των ενδείξεων μιας τέτοιας κατάστασης και η λήψη διορθωτικών μέτρων, με σκοπό την αποτροπή της (Dasilas & Rigani, 2024).

Στο πλαίσιο αυτό, έχουν πραγματοποιηθεί πολυάριθμες ερευνητικές προσπάθειες, τόσο από την ακαδημαϊκή κοινότητα, όσο και από επαγγελματίες του χρηματοοικονομικού τομέα. Οι έρευνες αυτές επικεντρώθηκαν κυρίως στην ανάπτυξη και την αξιολόγηση μοντέλων πρόβλεψης, τα οποία χρησιμοποιούν στατιστικές μεθοδολογίες και χρηματοοικονομικούς δείκτες, με πολλά από αυτά να είναι ιδιαίτερα αποτελεσματικά και να παρουσιάζουν μεγάλη ακρίβεια προβλέψεων (Geng et al., 2015).

Τα μοντέλα αυτά ομαδοποιούνται σε δύο κύριες κατηγορίες, στις στατιστικές τεχνικές και στα υποδείγματα τεχνητής νοημοσύνης. Τα τελευταία χρόνια, βέβαια, έχουν αναπτυχθεί τα λεγόμενα υβριδικά μοντέλα, τα οποία αποτελούν συνδυασμό των υποδειγμάτων τεχνητής νοημοσύνης (Sun et al., 2014).

2.4.1 Στατιστικές Μεθοδολογίες Πρόβλεψης

Οι πρώτες προσπάθειες πρόβλεψης πτώχευσης βασίστηκαν σε παραδοσιακές στατιστικές μεθόδους και στη χρήση χρηματοοικονομικών δεικτών που προέρχονται από τις οικονομικές καταστάσεις των επιχειρήσεων. Οι κυριότερες κατηγορίες στατιστικών μεθόδων που έχουν αναπτυχθεί είναι οι εξής (Jones et al., 2017):

1. Μονομεταβλητή Ανάλυση (Univariate Analysis)
2. Πολυμεταβλητή Ανάλυση Διαφοροποίησης (Multivariate Discriminant Analysis)

3. Γραμμικά Υποδείγματα Πιθανότητας υπό συνθήκη (Linear Probability Models)

Η Μονομεταβλητή Ανάλυση αποτέλεσε την πρώτη μέθοδο για την πρόβλεψη της εταιρικής αποτυχίας, λαμβάνοντας υπόψη τον κάθε χρηματοοικονομικό δείκτη ξεχωριστά. Ο Beaver (1966), θεμελιωτής αυτής της προσέγγισης, όρισε την πτώχευση ως την αδυναμία της επιχείρησης να καλύψει τις ληξιπρόθεσμες οφειλές της, ενώ παράλληλα εξέτασε την διακριτική ικανότητα ενός δείκτη κάθε φορά. Πιο συγκεκριμένα, χρησιμοποίησε δείκτες, όπως:

- η Δανειακή Επιβάρυνση (Debt Ratio),
- οι Ταμειακές Ροές προς Σύνολο Υποχρεώσεων (Cash Flow/ Total Liabilities),
- το Καθαρό Κεφάλαιο Κίνησης (Net Working Capital),
- η Αποδοτικότητα Ιδίων Κεφαλαίων (Return On Equity),
- η Αποδοτικότητα Ενεργητικού (Return On Assets) και
- η Γενική Ρευστότητα (Current Ratio).

Βέβαια, αξίζει να αναφέρουμε πως η τεχνική αυτή δέχθηκε πολλές κριτικές, κυρίως επειδή δεν επιτρέπει την ταυτόχρονη αξιολόγηση της προβλεπτικής ικανότητας του συνόλου των χρηματοοικονομικών δεικτών (Altman, 1968).

Σε αντίθεση με την Μονομεταβλητή Ανάλυση, στην Πολυμεταβλητή Ανάλυση χρησιμοποιούνται περισσότερες από μία μεταβλητές κάθε φορά, με σκοπό της κατηγοριοποίησης των παρατηρήσεων σε δύο κλάσεις, οι οποίες εμφανίζουν κοινά γνωρίσματα (Alaka et al., 2018). Η ταξινόμηση αυτή στηρίζεται στην εύρεση ενός γραμμικού συνδυασμού μεταξύ των χρησιμοποιούμενων μεταβλητών.

Ο πρώτος που εφάρμοσε την εν λόγω τεχνική είναι ο Edward Altman (1968), ο οποίος χρησιμοποιώντας ένα δείγμα 66 επιχειρήσεων, εκ των οποίων οι 33 είναι υγιείς και οι 33 πτωχευμένες, καθώς επίσης και 5 χρηματοοικονομικούς δείκτες κατέληξε στη σχέση:

$$Z = 0,021X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,999X_5$$

όπου:

X_1 :Κεφάλαιο Κίνησης προς Σύνολο Ενεργητικού

X_2 : Παρακρατηθέντα Κέρδη προς Σύνολο Ενεργητικού

X_3 : Κέρδη προ τόκων και φόρων προς Σύνολο Ενεργητικό

X_4 : Τρέχουσα Αξία Μετοχών προς Λογιστική Αξία Συνολικών Υποχρεώσεων

X_5 : Πωλήσεις προς Σύνολο Ενεργητικού

Ο τελικός βαθμός (Z-score) αποτελεί σταθμικό μέσο των μεταβλητών X_1 , X_2 , X_3 , X_4 , X_5 . Συγκεκριμένα,

- Εάν $Z > 2,67$, τότε η επιχείρηση δεν κινδυνεύει άμεσα με πτώχευση εντός του τρέχοντος έτους.

- Εάν $1.81 < Z < 2.67$, τότε δεν είναι δυνατό να γίνει σαφής εκτίμηση.
- Εάν $Z < 1.81$, τότε η επιχείρηση αντιμετωπίζει κίνδυνο αποτυχίας εντός του τρέχοντος έτους.

Τα Γραμμικά Υποδείγματα Πιθανότητας χρησιμοποιούνται για την εκτίμηση της πιθανότητας να συμβεί ένα δυαδικό γεγονός, δηλαδή την πιθανότητα εμφάνισης του κινδύνου πτώχευσης μια επιχείρησης ως συνάρτηση διαφόρων χρηματοοικονομικών μεταβλητών. Συγκεκριμένα, το υπόδειγμα έχει τη μορφή:

$$P(Y=1|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

όπου:

$P(Y=1|X)$: η πιθανότητα πτώχευσης της επιχείρησης

X_i : οι ανεξάρτητες χρηματοοικονομικές μεταβλητές

β_i : οι συντελεστές παλινδρόμησης

Ανεξάρτητα από την απλότητα και την ευχρηστία του συγκεκριμένου μοντέλου, παρουσιάζει ορισμένες αδυναμίες. Πιο συγκεκριμένα, οι προβλεπόμενες πιθανότητες ενδέχεται να βρεθούν εκτός του διαστήματος $[0,1]$, με αποτέλεσμα να προκύπτουν μη ρεαλιστικές και λανθασμένες εκτιμήσεις. Για την αντιμετώπιση του γεγονότος αυτού, αναπτύχθηκαν μη γραμμικά υποδείγματα, όπως η λογιστική παλινδρόμηση (Logit) και το υπόδειγμα Probit, τα οποία εξασφαλίζουν πιο αξιόπιστα αποτελέσματα, καθώς οι εκτιμήσεις βρίσκονται εντός των αποδεκτών ορίων (Bernhardsen, 2001).

Παρά τα ικανοποιητικά αποτελέσματα των παραδοσιακών στατιστικών μεθόδων, παρουσιάζουν περιορισμούς ως προς την ευαισθησία στις ακραίες τιμές, την υπόθεση της γραμμικότητας, καθώς και η αδυναμία τους να αποτυπώνουν την πολυπλοκότητα και την αβεβαιότητα που κυριαρχεί πλέον στον επιχειρηματικό τομέα (Alaka et al., 2018). Για το σκοπό αυτό, οι ερευνητές στράφηκαν προς την αναζήτηση νέων, πιο εξελιγμένων μεθόδων, που δεν υιοθετούν τους περιορισμούς των παραδοσιακών μοντέλων και παρέχουν σημαντικά βελτιωμένη ακρίβεια των προβλέψεων (Tsai, 2014).

2.4.2 Σύγχρονες Μεθοδολογίες Πρόβλεψης

Η ραγδαία και συνεχής ανάπτυξη της τεχνολογίας, έδωσαν την δυνατότητα στην Τεχνητή Νοημοσύνη (Artificial Intelligence) και στη Μηχανική Μάθηση (Machine Learning) να διεισδύσουν στον χρηματοοικονομικό τομέα και να δημιουργήσουν νέες προοπτικές για τη διαδικασία αναγνώρισης οικονομικών δυσχερειών. Η ικανότητα των μοντέλων Τεχνητής Νοημοσύνης να διαχειρίζονται μεγάλο όγκο δεδομένων και να ανιχνεύουν πολύπλοκες, μη γραμμικές σχέσεις μεταξύ των μεταβλητών με υψηλή ακρίβεια τα καθιστά ισχυρά εργαλεία έναντι των παραδοσιακών μεθόδων που στηρίζονται κυρίως σε γραμμικές σχέσεις μεταξύ χρηματοοικονομικών δεικτών (Barboza et al., 2017).

Ιδιαίτερα σημαντικό ρόλο διαδραματίζουν τα νευρωνικά δίκτυα τα οποία είναι εμπνευσμένα από τη δομή του ανθρώπινου εγκεφάλου. Ουσιαστικά πρόκειται για υπολογιστικά μοντέλα που έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά στη πρόβλεψη εταιρικής αποτυχίας, καθώς μπορούν να αναγνωρίσουν μη προφανείς σχέσεις μεταξύ των μεταβλητών και να διαχειριστούν πολυδιάστατα δεδομένα (Brédart, 2014). Ειδικά σε περιπτώσεις μεγάλου όγκου και σύνθετων δεδομένων ξεπερνούν σε απόδοση τις κλασσικές μεθόδους, όπως το Z-score του Altman.

Παράλληλα, αλγόριθμοι της Μηχανικής Μάθησης, όπως τα Δέντρα Απόφασης, (Decision Tree), ο Random Forest, οι Μηχανές Υποστήριξης Διανυσμάτων (SVM), καθώς και άλλες τεχνικές, εκπαιδεύονται από τα δεδομένα και έχουν την ικανότητα να προσαρμόζονται σε αυτά (Shetty et al., 2022). Η δυνατότητά τους να ανακαλύπτουν σχέσεις και να χρησιμοποιούν ταυτόχρονα πολλές μεταβλητές χωρίς εμφανή σχέση μεταξύ τους, τα έχει καταστήσει ευρέως χρησιμοποιήσιμα (Fallahpour et al., 2017).

Τα τελευταία χρόνια παρατηρείται αυξημένη τάση για χρήση υβριδικών μοντέλων, δηλαδή συνδυασμών πολλαπλών και διαφορετικών αλγορίθμων της Μηχανικής Μάθησης (Dasilas & Rigani, 2024). Συνδυάζοντας, λοιπόν, διαφορετικούς αλγορίθμους ή και τύπους δεδομένων (χρηματοοικονομικά και μη), τα υβριδικά μοντέλα παρουσιάζουν υψηλότερη ακρίβεια και σταθερότητα των προβλέψεων, μειώνοντας παράλληλα το κίνδυνο υπερπροσαρμογής (overfitting).

Επίσης αξίζει να αναφέρουμε πως στις σύγχρονες μεθόδους δε λαμβάνονται υπόψη μόνο τα οικονομικά δεδομένα μιας επιχείρησης, αλλά και διάφοροι μη χρηματοοικονομικοί αλλά εξίσου κρίσιμοι παράγοντες, όπως η γεωγραφική τοποθεσία, η φήμη στην αγορά και η ποιότητα διοίκησης, προσφέροντας πιο ρεαλιστικά και ακριβή αποτελέσματα (Dasilas & Rigani, 2024).

Η αξιολόγηση των μοντέλων βασίζεται σε πολλαπλές μετρικές, ώστε να υπάρχει μια πιο σφαιρική εικόνα της απόδοσης τους. Οι πιο διαδεδομένοι δείκτες είναι η Ακρίβεια (Accuracy), η Ευαισθησία (Sensitivity), η οποία μετράει την αναγνώριση των πτωχευμένων επιχειρήσεων, η Ειδικότητα (Specificity), όπου εκφράζει τη σωστή ταξινόμηση των υγιών επιχειρήσεων, το F1-score και το AUC (Jones et al., 2017). Η χρήση τους επιτρέπει την σταδιακά μετάβαση της αξιολόγησης από μια μονοδιάστατη προσέγγιση προς μια πιο ολοκληρωμένη.

Ανεξάρτητα από τα σημαντικά οφέλη που παρέχουν, τα σύγχρονα μοντέλα αντιμετωπίζουν αρκετές σημαντικές προκλήσεις. Ένα από τα βασικότερα προβλήματα είναι η ανισορροπία μεταξύ των δύο κλάσεων που κυριαρχεί στα περισσότερα σύνολα δεδομένων, καθώς οι υγιείς επιχειρήσεις συνήθως υπερσχύουν αριθμητικά έναντι των πτωχευμένων, δυσχεραίνοντας την διαδικασία της πρόβλεψης. Για το σκοπό αυτό, εφαρμόζονται τεχνικές εξισορρόπησης όπως η υπερδειγματοληψία (oversampling), η υποδειγματοληψία (undersampling) ή και συνδυασμός αυτών (He & Garcia, 2009).

Εξίσου σημαντική πρόκληση αποτελεί η ανάγκη για συνεχώς επισημοποιημένα και καθαρά δεδομένα, χωρίς ελλειπείς ή ακραίες τιμές, όπως και η απαίτηση για αυξημένη υπολογιστική ισχύ. Ακόμη, η περιορισμένη κατανόηση των σύνθετων μοντέλων, όπως είναι τα Deep Learning δίκτυα που λειτουργούν ως «μαύρα κουτιά», δυσκολεύει την αποδοχή τους από ρυθμιστικές αρχές και επαγγελματίες (Meese & Viken, 2019).

2.5 Εμπειρικές Μελέτες

Ένας από τους ανθρώπους που εφάρμοσαν στατιστικές μεθόδους, αξιοποιώντας χρηματοοικονομικούς δείκτες ως βασικές μεταβλητές για την πρόβλεψη της εταιρικής αποτυχίας ήταν ο Altman (1968). Ο συνδυασμός αυτών των δύο προσεγγίσεων είχε ως αποτέλεσμα την ανάπτυξη του γνωστού μοντέλου Z-score, το οποίο χρησιμοποιείται ακόμη και σήμερα. Για την κατασκευή του εν λόγω μοντέλου, χρησιμοποίησε δείγμα 66 επιχειρήσεων με ίδιο μέγεθος και κλάδο δραστηριότητας, εκ των οποίων οι 33 ήταν χρεοκοπημένες, ενώ οι υπόλοιπες υγιείς. Κατόπιν ανάλυσης των οικονομικών καταστάσεων, επιλέχθηκαν 22 αριθμοδείκτες, που αντιπροσωπεύουν το βαθμό της κερδοφορίας, της φερεγγυότητας, της ρευστότητας, της μόχλευσης και της αποτελεσματικότητας. Στην συνέχεια, εφάρμοσε τη στατιστική μέθοδο της πολλαπλής διακριτικής ανάλυσης, η οποία κατηγοριοποιεί τις επιχειρήσεις σε δύο ομάδες, ενεργές και ανενεργές, με βάση τις τιμές των ανεξάρτητων μεταβλητών. Το μοντέλο παρουσίασε εξαιρετική ακρίβεια, καθώς ταξινόμησε ορθά το 95% των περιπτώσεων. Όσον αφορά τη μελέτη του χρονικού ορίζοντα, η έρευνα έδειξε πως το μοντέλο μπορούσε να προβλέψει την πτώχευση με αξιοπιστία έως και δύο έτη πριν την πραγματοποίησή της, με την ακρίβεια να μειώνεται ραγδαία πέρα αυτού του χρονικού ορίζοντα.

Οι Aziz & Dar (2006), με σκοπό να παρουσιάσουν μια από τις πιο ολοκληρωμένες ανασκοπήσεις της βιβλιογραφίας στην πρόβλεψη της εταιρικής αποτυχίας, σύγκριναν 89 εμπειρικές μελέτες από 46 δημοσιευμένα άρθρα. Η ανάλυση στηρίχθηκε κυρίως στα στατιστικά μοντέλα, στα μοντέλα τεχνητής νοημοσύνης και στα θεωρητικά μοντέλα. Τα αποτελέσματα αναδεικνύουν τα τεχνητά νευρωνικά δίκτυα ως την ελαφρώς αποδοτικότερη τεχνική, με ακρίβεια 88%, έναντι των στατιστικών (85%) και των θεωρητικών τεχνικών (86%). Αναλυτικότερα, η πολυμεταβλητή ανάλυση και το λογιστικό μοντέλο αποτελούν τα πιο χρησιμοποιημένα στατιστικά εργαλεία, κυρίως λόγω της χαμηλής τυπικής απόκλισης στην απόδοση τους. Όσον αφορά τα μοντέλα τεχνητής νοημοσύνης, τα νευρωνικά δίκτυα παρουσίασαν πολύ καλή απόδοση, με ακρίβεια 87,4%. Στην κατηγορία των θεωρητικών προσεγγίσεων, ιδιαίτερο ενδιαφέρον παρουσιάζουν τα Balance Sheet Decomposition Models, που πέτυχαν ακρίβεια έως και 94%, αν και υπήρχαν λιγότερα δεδομένα. Εκτός αυτών, οι ερευνητές επισημαίνουν πως οι περισσότερες μελέτες βασίστηκαν κυρίως σε χρηματοοικονομικούς δείκτες, ενώ υπάρχει περιορισμένη χρήση των μεταβλητών που αφορούν την εταιρική διοίκηση ή φήμη.

Ο Chen (2011) θέλοντας να ερευνήσει την αποτελεσματικότητα διαφόρων στατιστικών και τεχνητών μεθόδων νοημοσύνης στην πρόβλεψη οικονομικής δυσχέρειας, χρησιμοποίησε δείγμα 200 επιχειρήσεων, εισηγμένες στο Χρηματιστήριο της Ταϊβάν. Η ανάλυση βασίστηκε σε 42 μεταβλητές, εκ των οποίων οι 33 είναι χρηματοοικονομικές, οι 8 είναι μη χρηματοοικονομικές και η 1 αποτελεί μακροοικονομικό δείκτη. Οι μέθοδοι που εξετάστηκαν περιλαμβάνουν παραδοσιακά στατιστικά μοντέλα, όπως η γραμμική διακριτική ανάλυση και η λογιστική παλινδρόμηση, καθώς και μοντέλα μηχανικής μάθησης, όπως τα δέντρα απόφασης (C5.0, CART), τα τεχνητά νευρωνικά δίκτυα (SOM, LVQ) και τις μηχανές υποστήριξης διανυσμάτων συνδυασμένες με το γενετικό αλγόριθμο και τη βελτιστοποίηση με σμήνη σωματιδίων (GA-SVM, PSO-SVM). Τα αποτελέσματα της έρευνας έδειξαν πως το PSO-SVM εμφάνισαν μακροπρόθεσμα την υψηλότερη ακρίβεια της τάξης 95%, παρουσιάζοντας παράλληλα ισορροπία μεταξύ της ευαισθησίας και της ειδικότητας. Αντίθετα, η ακρίβεια των δέντρων απόφασης είναι υψηλή για σύντομο χρονικό διάστημα, δηλαδή για έως 2 τρίμηνα πριν την πτώχευση, ενώ τα παραδοσιακά μοντέλα αποδείχθηκαν λιγότερο αποτελεσματικά.

Οι Olson, Delen and Meng (2012) προχώρησαν σε σύγκριση των εργαλείων της μηχανικής μάθησης, ώστε να συγκρίνουν την απόδοση τους στην πρόβλεψη χρεοκοπίας. Για την έρευνα τους χρησιμοποίησαν οικονομικά δεδομένα και χρηματοοικονομικούς δείκτες που αφορούν χρεοκοπημένες και μη επιχειρήσεις, εισηγμένες στο αμερικανικό χρηματιστήριο. Τα ευρήματα ανέδειξαν τα Δέντρα Απόφασης ως την πιο αποδοτική τεχνική για την ταξινόμηση των επιχειρήσεων στις δύο κατηγορίες. Πιο συγκεκριμένα, τα Δέντρα Απόφασης C5 και CART σημείωσαν την υψηλότερη ακρίβεια, αγγίζοντας το 93,7% ο καθένας. Από την άλλη, η Λογιστική Παλινδρόμηση αποδείχθηκε λιγότερο αποτελεσματική, ενώ ο SVM κατέγραψε τη χαμηλότερη ακρίβεια, μόλις 66,1%. Όσον αφορά τα νευρωνικά δίκτυα, οι ερευνητές ανέφεραν πως θα μπορούσαν να επιτύχουν εξίσου υψηλή ακρίβεια, εφόσον υπήρχε ο κατάλληλος υπολογιστικός χρόνος και γινόντουσαν οι απαραίτητες τροποποιήσεις των παραμέτρων.

Την ίδια χρονιά, οι Koyuncugil and Ozgulbas (2012) θέλησαν να αναπτύξουν ένα μοντέλο έγκαιρης προειδοποίησης για μικρομεσαίες επιχειρήσεις, ώστε να μπορούν να ανιχνεύσουν τους κινδύνους που ενδέχεται να επηρεάσουν αρνητικά τη βιωσιμότητα τους και να οδηγήσει ακόμη και σε πτώχευση. Για το σκοπό αυτό, χρησιμοποιήθηκαν χρηματοοικονομικά δεδομένα και αριθμοδείκτες για 7.853 μικρομεσαίες επιχειρήσεις της Τουρκίας που αποτυπώνουν την οικονομική τους επίδοση. Προκειμένου να κατηγοριοποιηθούν οι εταιρείες με βάση το προφίλ τους σε υγιείς και πτωχευμένες, χρησιμοποιήθηκε το δέντρο απόφασης CHAID, ο οποίος αναδεικνύει τις πολύπλοκες σχέσεις και τις υποομάδες στο σύνολο δεδομένων με τη δημιουργία πολλαπλών κόμβων, σε αντίθεση με τα συνηθισμένα Decision Trees. Έτσι, δημιουργήθηκαν 31 διαφορετικά προφίλ σε όρους χρηματοοικονομικής επίδοσης, όπου αναγνωρίζουν ποιες εταιρείες αντιμετωπίζουν οικονομικές δυσχέρειες και σε τι βαθμό. Τα αποτελέσματα έδειξαν πως το 31,4% των εξεταζόμενων επιχειρήσεων αντιμετώπιζαν οικονομικές δυσκολίες. Παρόλο που ο διαχωρισμός των προφίλ βασίστηκε σε 15 αριθμοδείκτες, τα τελικά αποτελέσματα έδειξαν πως μόνο δύο εξ αυτών μπορούν να

αποτελέσουν σημάδι κινδύνου. Συγκεκριμένα, όταν οι δείκτες Κέρδη προ φόρων/ Ίδια Κεφάλαια και Απόδοση Ιδίων Κεφαλαίων (ROE) είναι μικρότεροι ή ίσοι του μηδέν (0) είναι σχεδόν αναπόφευκτη η εμφάνιση οικονομικής δυσκολίας. Εκτός αυτού, προκειμένου οι αδύναμες οντότητες να μπορέσουν να ανακάμψουν και να διαχειριστούν το κίνδυνο, οι ερευνητές κατασκεύασαν τέσσερις οικονομικούς χάρτες (financial road map) βάση των τεσσάρων πιο επιτυχημένων επιχειρήσεων, περιλαμβάνοντας τιμές-στόχους για διάφορους δείκτες. Βέβαια, θα πρέπει να σημειωθεί ότι, πέρα τα ενθαρρυντικά αποτελέσματα του μοντέλου, θα πρέπει να λαμβάνονται υπόψη κι άλλοι παράγοντες, όπως ο λειτουργικός κίνδυνος και τα στρατηγικά λάθη, που μπορούν να βλάψουν τη βιωσιμότητα των επιχειρήσεων.

Η μελέτη της Shirata (2012), προκειμένου να δημιουργήσει ένα μοντέλο πρόβλεψης πτώχευσης που να χαρακτηρίζεται από εγκυρότητα και να εφαρμόζεται ανεξάρτητα από τον κλάδο ή το μέγεθος της επιχείρησης, αξιοποίησε τα οικονομικά δεδομένα για επιχειρήσεις της Ιαπωνίας. Αναλυτικότερα, το δείγμα αποτελείται από 686 πτωχευμένες και 300 υγιείς επιχειρήσεις διαφορετικών κλάδων, για τις οποίες περιλαμβάνονται 42 χρηματοοικονομικοί δείκτες. Για την επιλογή των σημαντικότερων μεταβλητών πραγματοποιήθηκε στατιστική ανάλυση σε συνδυασμό με την τεχνητή εξόρυξη γνώσης με το δέντρο απόφασης CART. Οι δείκτες που επιλέχθηκαν ήταν: Διατηρούμενα κέρδη προς σύνολο ενεργητικού, Ρυθμός αύξησης κεφαλαίου, Χρηματοοικονομικά έξοδα προς υποχρεώσεις, Λογαριασμοί πληρωτέοι προς πωλήσεις. Το τελικό μοντέλο πρόβλεψης βασίστηκε στη γραμμική διακριτική ανάλυση και παρουσίασε ποσοστό ακριβείας 86,14%. Αξίζει να σημειωθεί πως η πτωχευτική συμπεριφορά των επιχειρήσεων μπορούσε να εντοπιστεί για τουλάχιστον δύο χρόνια πριν την πτώχευση. Επιπλέον, η ερευνήτρια επισημαίνει πως οι χρηματοοικονομικοί δείκτες θα πρέπει να αξιολογούνται όχι μόνο προς την στατιστική ισχύ τους αλλά και ως προς τη λογιστική τους ερμηνεία, ώστε να αποφευχθούν λανθασμένα ή μη ερμηνεύσιμα αποτελέσματα.

Ο Xavier Brédart (2014) προκειμένου να αναπτύξει ένα μοντέλο πρόβλεψης εταιρικής χρεοκοπίας, χρησιμοποίησε τρεις χρηματοοικονομικούς δείκτες για 3.728 μικρομεσαίες επιχειρήσεις του Βελγίου, εκ των οποίων οι 1.864 κηρύχθηκαν σε πτώχευση κατά την περίοδο 2002 έως 2012. Πιο συγκεκριμένα, επέλεξε να υλοποιήσει ένα ρηχό νευρωνικό δίκτυο με τρεις ανεξάρτητες μεταβλητές, οι οποίες χαρακτηρίζουν την φερεγγυότητα (σχέση ιδίων κεφαλαίων προς σύνολο ενεργητικού), την κερδοφορία (καθαρά κέρδη προς σύνολο ενεργητικού) και τη ρευστότητα (τρέχων δείκτης), όπως καταγράφονται στο τελευταίο οικονομικό έτος πριν την χρεοκοπία. Επίσης, χρησιμοποιήθηκε μια εξαρτημένη μεταβλητή «Χρεοκοπία», η οποία λαμβάνει την τιμή 1, εάν η εξεταζόμενη επιχείρηση έχει πτωχεύσει, ενώ στην αντίθετη περίπτωση λαμβάνει την τιμή 0. Τα αποτελέσματα της έρευνας έδειξαν ότι τα νευρωνικά δίκτυα σε συνδυασμό με τους τρεις χρηματοοικονομικούς δείκτες αποτελούν καλή τεχνική πρόβλεψης, καθώς κατάφεραν να ταξινομήσουν τις επιχειρήσεις με ακρίβεια 80%.

Η μελέτη των Tsai, Hsu και Yen (2014) εξετάζει την απόδοση διαφόρων μοντέλων ταξινόμησης στην πρόβλεψη πτώχευσης επιχειρήσεων. Ειδικότερα, αξιολογούνται τρεις

αλγόριθμοι μηχανικής μάθησης, τα Νευρωνικά Δίκτυα, οι Μηχανές Υποστήριξης Διανυσμάτων (SVM) και τα Δέντρα Απόφασης, μέσω bagging και boosting. Τα αποτελέσματα της έρευνας βασίστηκαν σε τρία ξεχωριστά σύνολα δεδομένων που το καθένα περιλαμβάνει οικονομικά δεδομένα από επιχειρήσεις της Αυστραλίας, της Γερμανίας και της Ιαπωνίας, αντίστοιχα. Το βασικό συμπέρασμα της μελέτης είναι πως οι συνδυασμοί δέντρων απόφασης με boosting παρουσιάζουν την υψηλότερη ακρίβεια σε σύγκριση με άλλες μεθόδους. Συγκεκριμένα, η εν λόγω μέθοδος οδήγησε σε ακρίβεια 87,2% στο σύνολο των αυστραλιανών επιχειρήσεων και 88,4% στο σύνολο δεδομένων της Ιαπωνίας.

Δύο άλλοι ερευνητές που επιχείρησαν να αναπτύξουν κάποιο μοντέλο που να ανιχνεύει το χρηματοοικονομικό κίνδυνο ήταν οι Geng, Bose and Chen (2015). Για την επίτευξη του εγχειρήματός τους χρησιμοποίησαν οικονομικά δεδομένα για κινέζικες επιχειρήσεις που είχαν χαρακτηριστεί ως «ειδικής μεταχείρισης», λόγω σοβαρών οικονομικών προβλημάτων, αλλά και υγιείς επιχειρήσεις. Το μοντέλο στηρίχθηκε σε 31 χρηματοοικονομικούς δείκτες και η απόδοσή του αξιολογήθηκε για τρεις χρονικές περιόδους, δηλαδή για τρία, τέσσερα και πέντε έτη πριν την κρίση, με σκοπό να διαπιστωθεί ο χρονικός ορίζοντας που είναι δυνατόν να προβλεφθεί αποτελεσματικά η κατάρρευση μιας εταιρείας. Μεταξύ των Δέντρων Απόφασης, των Μηχανών Διανυσμάτων Υποστήριξης (SVM) και των Τεχνητών Νευρωνικών Δικτύων, ο αλγόριθμος που παρουσίασε ανώτερη προβλεπτική ικανότητα ανεξαρτήτου χρονικού ορίζοντα ήταν τα Τεχνητά Νευρωνικά Δίκτυα, για διαφορετικά train και test set. Από την πλευρά των χρονικών περιόδων, τα χρηματοοικονομικά στοιχεία των πιο πρόσφατων ετών παρείχαν καλύτερα αποτελέσματα, αν και τα στοιχεία της πενταετίας θα μπορούσαν να συμβάλλουν στην πρώιμη πρόβλεψη της πτώχευσης. Επίσης, οι δείκτες που αποδείχθηκαν στενά συνδεδεμένοι με την εταιρική βιωσιμότητα είναι οι Καθαρές Ταμειακές Ροές ανά Μετοχή, Καθαρά Κέρδη ανά Μετοχή και Απόδοση Ενεργητικού.

Οι Mansouri, Nazari & Ramazani (2016) με σκοπό να διερευνήσουν την πτώχευση εταιρικής πτώχευσης με Λογιστική Παλινδρόμηση και Τεχνητά Νευρωνικά Δίκτυα, χρησιμοποίησαν οικονομικά δεδομένα για 190 επιχειρήσεις εισηγμένες στο Χρηματιστήριο της Τεχεράνης, εκ των οποίων οι 90 είναι ενεργές και οι υπόλοιπες 90 ανενεργές λόγω πτώχευσης, με τα δεδομένα να καλύπτουν έως και τρία χρόνια πριν την κήρυξη της. Επίσης, χρησιμοποιήθηκαν 7 οικονομικοί δείκτες ως ανεξάρτητες μεταβλητές, ενώ υπήρχε και μια εξαρτημένη μεταβλητή, που λαμβάνει την τιμή 1 σε περίπτωση χρεοκοπίας και την τιμή 0 σε αντίθετη περίπτωση. Η αξιολόγηση πραγματοποιήθηκε βάσει της Ακρίβειας (Accuracy) και απέδειξε πως τα Τεχνητά Νευρωνικά Δίκτυα υπερτερούν την Λογιστικής Παλινδρόμησης, καθώς αναγνωρίζουν πολύπλοκες μη γραμμικές σχέσεις. Συγκεκριμένα, ο πρώτος αλγόριθμος σημείωσε ακρίβεια από 88,9% έως 93,3%, σε αντίθεση με τη Λογιστική Παλινδρόμηση, που κινήθηκε από 86,7% έως 91,1%, με την απόδοση να αυξάνεται καθώς πλησιάζει το έτος πτώχευσης.

Με σκοπό να ελέγξουν την απόδοση διαφόρων μεθόδων της μηχανικής μάθησης, οι Barboza, Kimura & Altman (2017) χρησιμοποίησαν χρηματοοικονομικά δεδομένα για αμερικανικές και καναδέζικες οντότητες του επιχειρηματικού τομέα από το 1985 μέχρι το 2013. Αναλυτικότερα, η συγκριτική αυτή μελέτη περιλαμβάνει τους επτά αλγορίθμους Random

Forest, Bagging, Boosting, SVM (SVM-Lin, SVM-RBF), Τεχνητά Νευρωνικά Δίκτυα, Λογιστική Παλινδρόμηση και Διακριτική Ανάλυση. Οι βασικοί δείκτες απόδοσης που χρησιμοποιήσαν βασίστηκαν στο μοντέλο του Altman και συγκεκριμένα είναι η ρευστότητα, η κερδοφορία, η παραγωγικότητα, η μόχλευση και η κυκλοφοριακή ταχύτητα ενεργητικού. Επιπλέον, επιλέχθηκαν οι 6 επιπλέον δείκτες: ανάπτυξη ενεργητικού, ανάπτυξη πωλήσεων, ανάπτυξη προσωπικού, λειτουργικό περιθώριο, μεταβολή απόδοσης ιδίων κεφαλαίων και μεταβολή δείκτη τιμής προς λογιστική αξία. Το εκπαιδευτικό δείγμα περιλάμβανε 449 υγιείς και 449 πτωχευμένες επιχειρήσεις. Αντίθετα, το δείγμα επικύρωσης αποτελούνταν από 133 πτωχευμένες και 13.167 μη πτωχευμένες επιχειρήσεις. Έπειτα από την εκπαίδευση των μοντέλων, τα αποτελέσματα έδειξαν πως ο Random Forest είναι ο πιο αποδοτικός αλγόριθμος, καθώς πέτυχε την υψηλότερη ακρίβεια της τάξης 87%, ενώ τη χειρότερη θέση είχαν η λογιστική παλινδρόμηση και η γραμμική διακριτική ανάλυση με ποσοστό ακρίβειας 69% και 50% αντίστοιχα.

Οι Veganzones & Séverin (2018), θέλοντας να διερευνήσουν την ακρίβεια πρόβλεψης της εταιρικής αποτυχίας με ανισόρροπο σύνολο δεδομένων, χρησιμοποίησαν 4 διαφορετικά δείγματα με οικονομικά δεδομένα για γαλλικές επιχειρήσεις που δραστηριοποιούνται αντίστοιχα στον τομέα των υπηρεσιών, των κατασκευών, το λιανικό εμπόριο, καθώς και ένα μεικτό δείγμα από όλους τους κλάδους. Το μέγεθος κάθε συνόλου εκπαίδευσης αποτελούνταν από 1500 εγγραφές, με την αναλογία των πτωχευμένων επιχειρήσεων προς τις μη πτωχευμένες να κυμαίνεται μεταξύ του 50/50 (εξαιρετικά ισορροπημένο σύνολο) έως 95/5 (εξαιρετικά ανισόρροπο σύνολο), ενώ στο σύνολο δοκιμής η αναλογία ήταν 95/5. Για την πρόβλεψη της πτώχευσης χρησιμοποιήθηκαν πέντε αλγόριθμοι ταξινόμησης, όπως η Γραμμική Διακριτική Ανάλυση, η Λογιστική Παλινδρόμηση, τα Τεχνητά Νευρωνικά Δίκτυα, οι Μηχανές Διανυσμάτων Υποστήριξης και ο Random Forest. Τα αποτελέσματα έδειξαν πως καθώς αυξάνεται η ανισορροπία μειώνεται η ικανότητα του μοντέλου να ανιχνεύει τις πτωχευμένες επιχειρήσεις, ενώ η εξειδίκευση για τις υγιείς αυξήθηκε, δείχνοντας τη μεροληψία του μοντέλου προς την πλειονοτική τάξη. Συγκεκριμένα, η ευαισθησία μειώθηκε από 81% που επικρατούσε στο ισορροπημένο σύνολο σε 15% στο ακραία ανισόρροπο δείγμα. Για την αντιμετώπιση του προβλήματος αυτού, οι τεχνικές που εφαρμόστηκαν ήταν η τυχαία υπερδειγματοληψία, η τυχαία υποδειγματοληψία, το SMOTE και το EasyEsample. Η χρήση των τεχνικών αυτών αύξησαν σημαντικά την απόδοση του μοντέλου, με το SMOTE να αποτελεί την καλύτερη επιλογή, επιτυγχάνοντας ανάκτηση απόδοσης έως και 54,7%. Επιπλέον, τα Τεχνητά Νευρωνικά Δίκτυα πέτυχαν την υψηλότερη ακρίβεια της τάξης του 57%, σε αντίθεση με τα SVM που είναι ευαίσθητα στα ανισορροπία.

Οι Meese & Viken (2019) κατά τη διάρκεια εκπόνησης της μεταπτυχιακής τους διατριβής, επικεντρώθηκαν στην εταιρική πρόβλεψη χρεοκοπίας, χρησιμοποιώντας ένα εκτενές σύνολο δεδομένων αποτελούμενο από 4,6 εκατομμύρια επιχειρήσεις, από το 1991 έως το 2016. Για κάθε οντότητα, καταγράφηκαν 231 χαρακτηριστικά, τα οποία καλύπτουν χρηματοοικονομικά, διοικητικά και κλαδικά στοιχεία. Πριν την ανάλυση, πραγματοποιήθηκε ενδελεχής καθαρισμός των δεδομένων. Διαγράφηκαν κρατικές και μη κερδοσκοπικές οντότητες, καθώς και παρατηρήσεις με λογιστικές ασυνέπειες ή ελλειπείς τιμές. Ακόμη,

εξαιρέθηκαν δεδομένα πριν το 1999, λόγω της αλλαγής στο λογιστικό πρότυπο. Εν τέλει, το τελικό σύνολο δεδομένων αποτελείται από 3.327.405 παρατηρήσεις με 159 μεταβλητές. Κατά την ανάλυση, δοκιμάστηκαν και συγκρίθηκαν συνολικά εννέα αλγόριθμοι μηχανικής μάθησης, οι Generalized Linear Model, Generalized Additive Model, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Mixture Discriminant Analysis, K-Nearest Neighbors, Support Vector Machines, Neural Network και Random Forest με αξιολόγηση βάσει των δεικτών Accuracy, AUC, Confusion Matrix, Recall, Precision και TPR/TNR. Τα αποτελέσματα ανέδειξαν τον Random Forest ως το πιο αποδοτικό αλγόριθμο, επιτυγχάνοντας ακρίβεια περίπου 78%, ενώ οι υπόλοιποι αλγόριθμοι παρουσίασαν κατώτερες επιδόσεις. Τα ευρήματα αυτά αναδεικνύουν την πολυπλοκότητα των μεγάλων συνόλων δεδομένων, όσο και την ικανότητα του Random Forest να αποδίδει αξιόπιστα σε τέτοιες συνθήκες.

Στο ίδιο πλαίσιο κινήθηκαν και οι Narvekar & Guha (2021) αξιοποιώντας δείγμα 21.114 αμερικανικών επιχειρήσεων διαφορετικών κλάδων για τη περίοδο 1970-2020 με 75 χρηματοοικονομικούς δείκτες για κάθε μία. Από το δείγμα αυτό, οι 19.902 εγγραφές ανήκουν στις υγιείς επιχειρήσεις και οι 1.212 στις πτωχευμένες, ενώ μερικοί δείκτες παρουσίαζαν ελλειπείς τιμές. Έπειτα από τη διαγραφή αυτών των δεικτών και την εξισορρόπηση του συνόλου δεδομένου με τις τεχνικές υπερδειγματοληψίας SMOTE, Borderline-SVM SMOTE και ADASYN, το τελικό σύνολο αποτελείται από 20.517 χρεοκοπημένες και 19.902 υγιείς, καθώς και από 57 χρηματοοικονομικούς δείκτες. Η μελέτη βασίστηκε στους αλγορίθμους Random Forest, SVM και XGBoost, οι οποίοι εκπαιδεύτηκαν να προβλέπουν χρεοκοπίες εντός 30, 90 και 180 ημερών. Τα αποτελέσματα έδειξαν πως ο XGBoost αποτελεί τον βέλτιστο αλγόριθμο για τα συγκεκριμένα δεδομένα, με ακρίβεια 99,68% περίπου και για τις τρεις χρονικές περιόδους.

Ομοίως, ο Abdullah (2021) με σκοπό να διερευνήσει την απόδοση αλγορίθμων μηχανικής μάθησης στην πρόβλεψη της χρηματοοικονομικής αποτυχίας, χρησιμοποίησε οικονομικά δεδομένα για 244 εισηγμένες επιχειρήσεις στο Χρηματιστήριο της Ντάκα για την περίοδο 2015-2019. Οι επιχειρήσεις ταξινομήθηκαν σε τρεις κατηγορίες, Secure (καλή χρηματοοικονομική κατάσταση), Healthy (πιθανή μελλοντική αφερεγγυότητα) και Insolvent (κατάσταση οικονομικής δυσχέρειας), μέσω του δείκτη Z-score του Altman. Από τα συνολικά δεδομένα δημιουργήθηκε ένα σύνολο εκπαίδευσης για τα έτη 2015 έως 2017 και ένα σύνολο δοκιμής για τα έτη 2018-2019. Το πλήθος των χρηματοοικονομικών δεικτών που χρησιμοποιήθηκαν μέσω της συσχέτισης τους σε προηγούμενη βιβλιογραφία ήταν 11, μερικές εκ των οποίων είναι το κεφάλαιο κίνησης προς σύνολο ενεργητικού, τα καθαρά κέρδη προς πωλήσεις, τα ίδια κεφάλαια προς ενεργητικό, η αποδοτικότητα ενεργητικού (ROA) και η αποδοτικότητα ιδίων κεφαλαίων (ROE). Τα μοντέλα που χρησιμοποιήθηκαν ήταν τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), ο Naïve Bayes, ο K-nearest neighbors και ο Ταξινομητής Συνόλου, αξιολόγηση των οποίων έγινε με βάση τις εξής μετρικές: ακρίβεια (accuracy), AUC, ευαισθησία (sensitivity), εξειδίκευση (specificity), καθώς και η απώλεια λογαρίθμου (logLoss) και ο δείκτης kappa. Η μελέτη παρουσίασε πως τα Τεχνητά Νευρωνικά Δίκτυα πέτυχαν ποσοστό ακρίβειας 94% στο

σύνολο εκπαίδευσης και 88% στο σύνολο δοκιμής, υπερτερώντας μεταξύ των άλλων αλγορίθμων. Αντίθετα, ο αλγόριθμος Naive Bayes εμφάνισε την χαμηλότερη απόδοση, με ακρίβεια 60% στο σύνολο εκπαίδευσης και 74% στο σύνολο δοκιμής.

Οι Shekar Shetty, Mohamed Musa and Xavier Brédart (2022), με σκοπό να εξετάσουν ποιο εργαλείο μηχανικής μάθησης είναι περισσότερο ακριβές στη πρόβλεψη πτώχευσης, αξιοποίησαν τα ίδια δεδομένα με τον Brédart (2014) που παρουσιάστηκε προηγουμένως. Συγκεκριμένα, χρησιμοποίησαν τρεις χρηματοοικονομικούς δείκτες, την απόδοση ενεργητικού (Return on Assets), το δείκτη ρευστότητας (Current Ratio) και το δείκτη φερεγγυότητας (Solvency Ratio), για 3.728 μικρομεσαίες επιχειρήσεις του Βελγίου κατά την περίοδο 2002 έως 2012. Ο λόγος που χρησιμοποίησαν αυτούς τους δείκτες είναι για την απλότητα και την ευκολία συλλογή τους, ώστε το μοντέλο τους να είναι πρακτικό και εφαρμόσιμο. Για την ταξινόμηση των επιχειρήσεων σε υγιής και μη επέλεξαν να εφαρμόσουν ένα βαθύ νευρωνικό δίκτυο με έξι επίπεδα, τον αλγόριθμο SVM και τον XGBoost. Η προσέγγιση τους κατάφερε απόδοση μεταξύ 82% και 83%, σημειώνοντας βελτίωση από αυτή του Brédart που βασίζεται σε ένα ρηχό νευρωνικό δίκτυο.

Οι Domicián Mát, Hassan Raza and Ishtiaq Ahmad (2023) προχώρησαν σε σύγκριση της απόδοσης διαφορετικών αλγορίθμων μηχανικής μάθησης με σκοπό τον εντοπισμό του πιο αποτελεσματικού στην πρόβλεψη πτώχευσης. Στην έρευνα χρησιμοποιήθηκαν οικονομικά δεδομένα και χρηματοοικονομικοί δείκτες για 385 χρεοκοπημένες και μη επιχειρήσεις του Πακιστανικού Χρηματιστηρίου, καλύπτοντας την περίοδο 2016-2021. Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι οι AdaBoost, Gradient Boosting, Decision Tree, Random Forest, SVM, Naive Bayes και Logistic Regression. Προκειμένου να επιτευχθεί η μέγιστη δυνατή απόδοση των μοντέλων, οι ερευνητές επέλεξαν τα χαρακτηριστικά βάσει στατιστικών ελέγχων και εφάρμοσαν την τεχνική SMOTE για εξισορρόπηση των δεδομένων. Επίσης, πραγματοποιήθηκε μείωση διαστάσεων με PCA, επιλογή βέλτιστων υπερπαραμέτρων μέσω grid search και διασταυρούμενης επικύρωσης (cross-validation). Εκτός αυτών, ενσωματώθηκαν στρατηγικές τακτοποίησης (regularization) και pruning για την αποφυγή υπερεκπαίδευσης, ενώ χρησιμοποιήθηκε ακόμη το early stopping για τα επαναληπτικά μοντέλα boosting. Με βάση αυτά τα δεδομένα, οι αλγόριθμοι Decision Tree, AdaBoost και Gradient Boosting σημείωσαν ακρίβεια 100% στην ταξινόμηση των επιχειρήσεων, ακολουθούμενοι από τους SVM και Logistic Regression με ακρίβεια μεταξύ 89% και 99%, ανάλογα με την επιλογή των χαρακτηριστικών. Αντίθετα, ο Naive Bayes είχε τη χαμηλότερη ακρίβεια, κινούμενη μεταξύ 58% και 70%, καθιστώντας τον λιγότερο αξιόπιστο για τα συγκεκριμένα δεδομένα.

Την επόμενη χρονιά πραγματοποίησαν παρόμοια έρευνα οι Lahmiri & Bekiros (2024), αξιοποιώντας ένα δείγμα 250 επιχειρήσεων, εκ των οποίων οι 107 είναι πτωχευμένες και οι 143 μη πτωχευμένες. Κάθε παρατήρηση, περιγράφεται από 6 ποιοτικά χαρακτηριστικά, τα οποία μετατρέπονται σε ποσοτικά κατά το στάδιο της προεπεξεργασίας. Συγκρίθηκαν πέντε αλγόριθμοι, οι Generalized Regression Neural Network, Multi-layer Back-Propagation Neural Network, Probabilistic Neural Network, Radial Basis Function Neural Network και Regression Trees, η αξιολόγηση των οποίων έγινε με τις μετρικές Ακρίβεια, Ευαισθησία και

Ειδικότητα. Τα αποτελέσματα της έρευνας ανέδειξαν τον Generalized Regression Neural Network ως τον πλέον αποτελεσματικό αλγόριθμο, καθώς αναγνώρισε σχεδόν όλες τις χρεοκοπημένες (ευαισθησία=99,91%) και μη (ειδικότητα=100%), σημειώνοντας συνολική ακρίβεια ίση με 99,96%. Στην συνέχεια, ακολουθούν οι υπόλοιπες κατηγορίες νευρωνικών δικτύων, καταδεικνύοντας την αποτελεσματικότητά τους ακόμη και σε μικρά σύνολα δεδομένων.

Μια από τις πιο πρόσφατες και εμπεριστατωμένες εμπειρικές ανασκοπήσεις στο χώρο της χρηματοοικονομικής αποτυχίας των επιχειρήσεων αποτελεί η μελέτη των Dasilas & Rigani (2024). Ακολουθώντας τη μεθοδολογία PRISMA, ανέλυσαν 207 εμπειρικές μελέτες της περιόδου 2012 έως 2023, με στόχο την αξιολόγηση των εφαρμογών της τεχνητής νοημοσύνης και της μηχανικής μάθησης στην πρόβλεψη πτώχευσης, καταλήγοντας σε κρίσιμα ευρήματα. Αρχικά, οι πιο αποτελεσματικές και διαδεδομένες προσεγγίσεις στην πρόβλεψη της εταιρικής αποτυχίας είναι τα Τεχνητά Νευρωνικά Δίκτυα, καθώς και το μοντέλο XGBoost, ενώ επίσης χρησιμοποιούνται όλο και περισσότερο υβριδικά και ensemble μοντέλα, όπως το SMOTEBoost-C4.5, καθώς παρουσιάζουν υψηλή ακρίβεια. Επιπλέον, οι ερευνητές επισημαίνουν πως υπάρχει μια αυξανόμενη τάση για ενσωμάτωση μη χρηματοοικονομικών μεταβλητών, όπως η φήμη στην αγορά και η ποιότητα διοίκησης, αποδεικνύοντας την σημαντική πρόοδο προς την ενσωμάτωση πολυδιάστατων μοντέλων. Η ανασκόπηση καταγράφει ακόμη τις κυριότερες τεχνικές διαχείρισης της ανισορροπίας των δύο κλάσεων, δηλαδή των πτωχευμένων και υγιών επιχειρήσεων, αναδεικνύοντας ως συχνότερες τις τεχνικές SMOTE, cost-sensitive learning και undersampling. Όσον αφορά την αξιολόγηση των μοντέλων, οι πιο διαδεδομένες μετρικές είναι οι accuracy, sensitivity, specificity, AUC και F1-score. Τέλος, οι συγγραφείς τονίζουν την ανάγκη για ενίσχυση της ερμηνείας και της κατανόησης των πιο πολύπλοκων μοντέλων, όπως το deep learning, ώστε να γίνουν περισσότερο αποδεκτά από επιχειρησιακούς φορείς.

Το 2025 οι Narvekar & Guha αξιοποίησαν ένα σύνολο δεδομένων για πολωνικές επιχειρήσεις, χρεοκοπημένες και μη, κατά την περίοδο 2000-2013, αποτελούμενο από 64 χρηματοοικονομικούς δείκτες και μία μεταβλητή-στόχο. Μετά την προεπεξεργασία των δεδομένων και τη διαχείριση των ελλειπών τιμών, χρησιμοποιήθηκε η μέθοδος επιλογής χαρακτηριστικών με Random Forest, από την οποία βρέθηκαν 26 ισχυρά συσχετιζόμενες μεταβλητές και αφαιρέθηκαν. Επίσης, λόγω της μεγάλης ανισορροπίας μεταξύ των δύο κλάσεων, ενεργές και ανενεργές επιχειρήσεις, εφαρμόστηκε η τεχνική BORDERLINE-SMOTE για την αντιμετώπισή της. Στο τελικό δείγμα χρησιμοποιήθηκαν 5 αλγόριθμοι: Random Forest, Decision Tree, XGBoost, AdaBoost και CatBoost. Για την αξιολόγησή τους χρησιμοποιήθηκαν οι μετρικές Ακρίβεια (Accuracy), Ευαισθησία (Recall), Ειδικότητα (Specificity) και ο Γεωμετρικός Μέσος Όρος (Geometric Mean, GM). Τα αποτελέσματα της έρευνας έδειξαν πως, με βάση τα υπάρχοντα δεδομένα, ο Random Forest αποτελεί την καταλληλότερη μέθοδο για την εν λόγω πρόβλεψη, επιτυγχάνοντας ακρίβεια 94%, ευαισθησία 91%, ειδικότητα 93% και GM 91%. Ακολουθούν οι αλγόριθμοι XGBoost και CatBoost με εξίσου ικανοποιητικές επιδόσεις, ενώ την τελευταία θέση κατέκτησε ο Decision Tree, καθώς εμφάνισε τη χαμηλότερη απόδοση από το σύνολο των αλγορίθμων.

Κεφάλαιο 3

Μεθοδολογία και Δεδομένα

3.1 Συλλογή Δεδομένων

Για την πραγματοποίηση της έρευνας χρησιμοποιήθηκαν ιστορικά στοιχεία των οικονομικών καταστάσεων επιχειρήσεων της Ε.Ε που δραστηριοποιούνταν στους τομείς των τηλεπικοινωνιών, των υπολογιστών και της τεχνολογίας, καλύπτοντας την χρονική περίοδο 2014 έως 2023.

Αναλυτικότερα, χρησιμοποιήθηκαν 13 μεταβλητές (ανεξάρτητες) για κάθε χρονική περίοδο, εκ των οποίων οι 12 είναι ποσοτικές και μπορούν να πάρουν μόνο αριθμητικές τιμές. Επίσης, δημιουργήθηκε μία ακόμη μεταβλητή στόχος (εξαρτημένη), με σκοπό να παίρνει δύο τιμές, την τιμή 1 και 0. Ουσιαστικά η μεταβλητή αυτή απαντάει στο ερώτημα εάν η εξεταζόμενη εταιρεία είναι πτωχευμένη ή όχι, αντίστοιχα και εξαρτάται από τις τιμές που λαμβάνουν οι ανεξάρτητες μεταβλητές.

Η επιλογή των μεταβλητών έγινε με τέτοιο τρόπο, ώστε να χρησιμοποιηθούν εκείνες που αποτυπώνουν την ικανότητα της επιχείρησης να είναι κερδοφόρα, να διαχειρίζεται τις υποχρεώσεις της, όπως και τη χρηματοοικονομική δύναμη που έχει ώστε να προσαρμόζεται σε αλλαγές και να αναπτύσσεται. Έτσι, όλες οι βασικές ποσοτικές μεταβλητές ανήκουν σε μία από τις τέσσερις κατηγορίες αριθμοδεικτών (Titman et al., 2015):

1. Δείκτης Δραστηριότητας: μετράει τον βαθμό αποτελεσματικότητα της επιχείρησης να διαχειρίζεται τους διαθέσιμους πόρους της, δηλαδή τα περιουσιακά στοιχεία
2. Δείκτης Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας: εκφράζει την ικανότητα της επιχείρησης να ανταποκρίνεται στις υποχρεώσεις της μακροχρόνια (π.χ. αποπληρωμή δανείων) και στο βαθμό προστασίας που απολαμβάνουν οι πιστωτές της (π.χ. τράπεζες, δανειολήπτες).
3. Δείκτης Αποδοτικότητας και Κερδοφορίας: εκφράζει την ικανότητα κερδοφορίας διοίκησης της επιχείρησης, καθώς και το βαθμό επιτυχίας της.
4. Δείκτης Ρευστότητας: δείχνει τη δυνατότητα της επιχείρησης να ανταποκρίνεται στις βραχυπρόθεσμες υποχρεώσεις της (π.χ. προμηθευτές, μισθοί εργαζομένων).

Ο Πίνακας 3.1 παρουσιάζει και περιγράφει τις μεταβλητές που χρησιμοποιήθηκαν προκειμένου να υλοποιηθεί η παρούσα εργασία.

A/A	Όνομα	Τύπος	Περιγραφή
1	Πάγια Στοιχεία (Non-current assets)	Ποσοτική	Τα πάγια περιουσιακά στοιχεία της επιχείρησης (κτίρια, εξοπλισμός)
2	Κυκλοφορούντα στοιχεία (Current assets)	Ποσοτική	Τα βραχυπρόθεσμα περιουσιακά στοιχεία (μετρητά, αποθέματα)
3	Σύνολο περιουσιακών στοιχείων (Total assets)	Ποσοτική	Το άθροισμα του πάγιου και του κυκλοφορούντος ενεργητικού
4	Μετοχικό κεφάλαιο (Capital)	Ποσοτική	Τα κεφάλαια που έχουν επενδυθεί στην επιχείρηση από τους μετόχους και τους ιδιοκτήτες.
5	Δάνεια και βραχυπρόθεσμες υποχρεώσεις (Loans & short-term debt)	Ποσοτική	Το σύνολο των δανείων και των βραχυπρόθεσμων υποχρεώσεων που οφείλει η επιχείρηση.
6	Προμηθευτές (Creditors)	Ποσοτική	Οι βραχυπρόθεσμες υποχρεώσεις της επιχείρησης προς τους προμηθευτές και άλλους πιστωτές.
7	Other current liabilities	Ποσοτική	Άλλες βραχυπρόθεσμες υποχρεώσεις που δεν κατατάσσονται στις προηγούμενες κατηγορίες.
8	Βραχυπρόθεσμες υποχρεώσεις (Current liabilities)	Ποσοτική	Το σύνολο των βραχυπρόθεσμων χρεών της επιχείρησης.
9	Working capital	Ποσοτική	Η διαφορά μεταξύ των κυκλοφορούντων ενεργητικών στοιχείων και των βραχυπρόθεσμων υποχρεώσεων
10	Profit after tax [PAT]	Ποσοτική	Το καθαρό κέρδος της επιχείρησης αφού αφαιρεθούν οι φόροι.
11	Λειτουργικά έσοδα (Operating revenue-Turnover)	Ποσοτική	Τα έσοδα που προκύπτουν από τις κύριες δραστηριότητες της επιχείρησης.
12	ΚΠΤΦ (EBIT)	Ποσοτική	Τα κέρδη πριν από τόκους και φόρους
13	Bankruptcy	Ποιοτική	0-Υγιείς Επιχείρηση, 1-Χρεοκοπημένη Επιχείρηση

Πίνακας 1: Περιγραφή μεταβλητών.

3.2 Προ Επεξεργασία και Καθαρισμός Δεδομένων

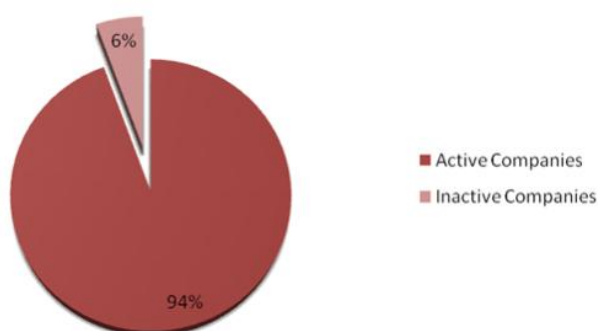
Αρχικά, πραγματοποιήθηκε ενοποίηση των επιμέρους αρχείων σε ένα ενιαίο σύνολο δεδομένων με τη μορφή πινάκων, αποτελούμενο από 3.102 εγγραφές, όπου η κάθε γραμμή αντιστοιχεί σε μια εταιρεία, ενώ η κάθε στήλη στις τιμές των οικονομικών μεταβλητών που την περιγράφει για κάθε οικονομικό έτος.

Ακολούθησε ο έλεγχος για διπλότυπες εγγραφές, για ασυνέπειες στου τύπους των δεδομένων, όπως αριθμητικά πεδία αποθηκευμένα ως συμβολοσειρές, καθώς και για συμβολισμούς όπως «N/A» που δυσχεραίνουν τη διαδικασία της ανάλυσης και αφαιρέθηκαν, εξασφαλίζοντας με αυτό το τρόπο την αξιοπιστία και την ποιότητα του συνόλου.

Παράλληλα, ανιχνεύθηκαν αρκετές παρατηρήσεις με ελλιπείς τιμές σε βασικά οικονομικά πεδία. Καθώς τα πεδία αυτά είναι κρίσιμα για την πρόβλεψη επιλέχθηκε η πλήρης αφαίρεση των αντίστοιχων εγγραφών. Με αυτό το τρόπο κατέστη δυνατή η χρησιμοποίησή τους, με το μειονέκτημα βέβαια της δημιουργίας μικρού και ανισόρροπου συνόλου δεδομένων, αποτελούμενο από 180 επιχειρήσεις.

3.3 Αντιμετώπιση Ανισορροπίας Δεδομένων

Σημαντική πρόκληση προκάλεσε η ανισορροπία που επικρατούσε στο σύνολο δεδομένων μεταξύ των δύο κλάσεων, πτωχευμένη και μη πτωχευμένη επιχείρηση. Ο αριθμός των υγιών (μη πτωχευμένων) επιχειρήσεων ανέρχεται στις 170 καθιστώντας τον υπερβολικά μεγάλο έναντι αυτού των πτωχευμένων, οι οποίες είναι μόλις 10, αντιπροσωπεύοντας το 6% περίπου του συνολικού δείγματος, όπως φαίνεται στην Εικόνα 1. Το γεγονός αυτό θα είχε ως αποτέλεσμα την κακή απόδοση του μοντέλου, καθώς θα είχε εκπαιδευτεί να αναγνωρίζει και να προβλέπει μόνο την πλειοψηφική κατηγορία, δηλαδή τις εταιρείες που θεωρούνται υγιείς, αγνοώντας τις περιπτώσεις της μειοψηφίας.



Σχήμα 4: Κατανομή επιχειρήσεων με βάση την ενεργή ή ανενεργή κατάστασή τους.

Για την αντιμετώπιση αυτής της ανισορροπίας και την ενίσχυση της αποτελεσματικότητας και της ικανότητας των αλγορίθμων να πραγματοποιούν με ακρίβεια προβλέψεις αξιοποιήθηκαν δύο συμπληρωματικές τεχνικές: η SMOTE (Synthetic Minority Over-sampling Technique) και η Random Under-Sampling.

3.3.1 Random Under-Sampling

Η τεχνική Random Under-Sampling (RUS) είναι μια από τις πιο απλές αλλά και ευρέως χρησιμοποιούμενες προσεγγίσεις για την αντιμετώπιση προβλημάτων με ανισόρροπες κλάσεις, ιδιαίτερα όταν η πλειονοτική κατηγορία υπερισχύει σημαντικά έναντι της μειονοτικής (Drummond & Holte, 2003; He & Garcia, 2009). Η βασική της αρχή έγκειται στην τυχαία αφαίρεση παρατηρήσεων από την πλειοψηφία, με σκοπό τη μείωση του αριθμητικού χάσματος μεταξύ των δύο τάξεων.

Σε αντίθεση με τις τεχνικές υπερδειγματοληψίας, που προσθέτουν συνθετικά δεδομένα στη μειονοτική τάξη, η RUS στοχεύει στην απλοποίηση του προβλήματος, περιορίζοντας την υπεροχή της πλειονοτικής, καθιστώντας το σύνολο περισσότερο ισορροπημένο. Σημαντικό πλεονέκτημα αυτής της μείωσης είναι ο λιγότερος χρόνος που απαιτείται για την εκπαίδευση του μοντέλου και η αποφυγή της προκατάληψης προς την μεγαλύτερη κατηγορία (He & Garcia, 2009; Fernández et al., 2018).

Ωστόσο, η απλότητα της συγκεκριμένης τεχνικής εγκυμονεί τον κίνδυνο απώλειας σημαντικής πληροφορίας. Καθώς οι παρατηρήσεις αφαιρούνται τυχαία, ενδέχεται να απομακρυνθούν κρίσιμες ή αντιπροσωπευτικές εγγραφές της μεγαλύτερης τάξης, οδηγώντας στη μείωση της γενικευτικής ικανότητας του μοντέλου ή σε υπεραπλουστευμένες αποφάσεις ειδικά σε περιβάλλοντα υψηλής πολυπλοκότητας (Yen & Lee, 2006).

Αξίζει να αναφερθεί πως είναι προτεινόμενη η χρήση της τεχνικής σε συνδυασμό με άλλες μεθόδους εξισορρόπησης ή τεχνικές καθαρισμού (filtering), ώστε να περιορίζονται οι αρνητικές επιπτώσεις της τυχαίας δειγματοληψίας και να διατηρείται η πληρότητα του πληροφοριακού περιεχομένου (Lemaître et al., 2017; Kubat & Matwin, 1997). Έτσι, η RUS μπορεί να λειτουργήσει ως ένα ευέλικτο εργαλείο για τη διαχείριση της ανισορροπίας, χωρίς να υποβαθμίζεται η προβλεπτική ισχύς του μοντέλου (Fernández et al., 2018).

3.3.2 SMOTE

Η τεχνική SMOTE αποτελεί μια από τις πιο διαδεδομένες τεχνικές υπερδειγματοληψίας της μειονοτικής κατηγορίας σε ανισόρροπα σύνολα δεδομένων, στη περίπτωση μας των πτωχευμένων επιχειρήσεων (Chawla et al., 2002). Σε αντίθεση με τις απλούστερες μεθόδους oversampling που απλώς αντιγράφουν τις υπάρχουσες παρατηρήσεις, το SMOTE παράγει συνθετικά σημεία δεδομένων δημιουργώντας νέες παρατηρήσεις, υπολογίζοντας ενδιάμεσες τιμές ανάμεσα σε παρόμοια υπάρχουσα δείγματα της μειοψηφίας, προκειμένου να ενισχύσει την εκπαίδευση του μοντέλου (Fernández et al., 2018).

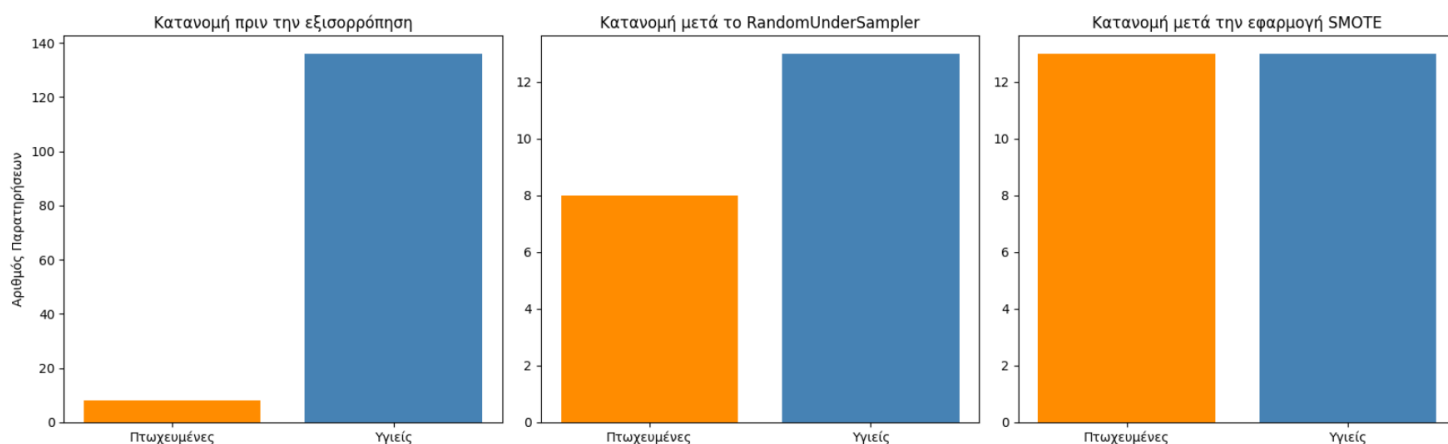
Αναλυτικότερα, η φιλοσοφία που επικρατεί για την συγκεκριμένη τεχνική βασίζεται στην υπόθεση ότι η περιοχή γύρω από τα υπάρχοντα σημεία της μειοψηφικής τάξης έχει σημασιολογική αξία και η επέκταση αυτής με συνθετικά δείγματα βελτιώνει τη γενίκευση του ταξινομητή (Blagus & Lusa, 2013). Για κάθε τέτοιο δείγμα επιλέγονται k κοντινοί γείτονες, συνήθως 5, και από το ευθύγραμμο τμήμα μεταξύ τους παράγεται μια νέα παρατήρηση, αυξάνοντας το υπάρχον σύνολο δεδομένων και καθιστώντας το λιγότερο επιρρεπές στην υπερπροσαρμογή (overfitting) (Chawla et al., 2002; Fernández et al., 2018).

Η εν λόγω τεχνική αποτελεί σημαντικό εργαλείο, καθώς ανταποκρίνεται αποτελεσματικά σε χρηματοοικονομικά περιβάλλοντα όπου τα σπάνια γεγονότα, όπως η πτώχευση, παρουσιάζουν μεγάλες διαφορές μεταξύ τους, λόγω χάρη διαφορετικά χαρακτηριστικά και αιτίες (Fernández et al., 2018). Εφαρμόζοντας το SMOTE στην παρούσα μελέτη καταφέραμε να βελτιώσουμε την παρουσία των πτωχευμένων επιχειρήσεων κατά την διάρκεια εκπαίδευσης των αλγορίθμων Μηχανικής Μάθησης, καθώς τα επέτρεψε να μάθουν και να αναγνωρίζουν ουσιαστικά εκείνα τα χαρακτηριστικά που επηρεάζουν την πιθανότητα πρόβλεψης.

Ωστόσο, να σημειωθεί σε αυτό το σημείο πως η εφαρμογή του SMOTE ενδέχεται να εισάγει «θόρυβο» ή να δημιουργήσει ακραίες παρατηρήσεις. Για τον σκοπό αυτό είναι κρίσιμο να συνδυάζεται και με τεχνικές φιλτραρίσματος ή δειγματοληψίας της πλειονότητας (Kubat & Matwin, 1997), με σκοπό τη διασφάλιση της ποιότητας των δεδομένων.

3.3.3 Εφαρμογή Τεχνικών Εξισορρόπησης στο Σύνολο Εκπαίδευσης

Η διαδικασία εξισορρόπησης του συνόλου εκπαίδευσης περιλαμβάνει την αρχική παρατήρηση της κατανομής, την εφαρμογή του Random UnderSampling και τέλος την εφαρμογή του SMOTE στο υποδειγματοληφθέν σύνολο. Η μεταβολή αυτή αποτυπώνεται στο Σχήμα 3.2, παρουσιάζοντας την εξέλιξη της ισορροπίας μεταξύ των δύο τάξεων.



Σχήμα 5: Κατανομή δεδομένων πριν και μετά τις τεχνικές εξισορρόπησης.

Παρατηρώντας την αρχική κατάσταση του συνόλου παρατηρούμε μια άνιση κατανομή με τις υγιείς επιχειρήσεις να υπερτερούν σημαντικά αριθμητικά από τις πτωχευμένες. Η εικόνα αυτή τροποποιήθηκε με τη χρήση της τεχνικής Random UnderSampling, η οποία στοχεύει στην μείωση της πλειοψηφικής τάξης. Συγκεκριμένα, από τις 136 υγιείς επιχειρήσεις διατηρήθηκε με τυχαία δειγματοληψία ένα υποσύνολο 13 εγγραφών, διατηρώντας παράλληλα το σύνολο των πτωχευμένων επιχειρήσεων σταθερό στις 10 εγγραφές.

Στο επόμενο στάδιο έχουμε την εφαρμογή του SMOTE στο ήδη τροποποιημένο σύνολο, στοχεύοντας την ενίσχυση της μειονοτικής τάξης με τη δημιουργία συνθετικών παρατηρήσεων βασισμένες στα υπάρχοντα δείγματα των ανενεργών εταιρειών. Η προσέγγιση αυτή είχε ως αποτέλεσμα τη βελτίωση του συνόλου εκπαίδευσης όπου αποτελείται πλέον από 13 υγιείς και 13 χρεοκοπημένες επιχειρήσεις. Έτσι, οι δύο κλάσεις είναι ισορροπημένες, αποφεύγοντας την μεροληψία υπέρ της πλειοψηφίας.

3.4 Επιλογή και Περιγραφή Αλγορίθμων

Για την υλοποίηση του μοντέλου πρόβλεψης, επιλέχθηκαν και εφαρμόστηκαν διάφοροι αλγόριθμοι ταξινόμησης, προκειμένου να συγκριθεί η απόδοσή τους υπό συνθήκες ανισόρροπου συνόλου δεδομένων. Συγκεκριμένα, χρησιμοποιήθηκαν επτά αλγόριθμοι, οι οποίοι περιλαμβάνουν τόσο απλές όσο και πιο σύνθετες τεχνικές Μηχανικής Μάθησης, ώστε να καλυφθεί ένα ευρύ φάσμα υπολογιστικών προσεγγίσεων.

3.4.1 Random Forest

Ο αλγόριθμος Τυχαίου Δάσους αποτελεί έναν ισχυρό αλγόριθμο βασισμένο σε σύνολα αποφάσεων, ο οποίος δημιουργεί πολλαπλά δέντρα απόφασης κατά τη διάρκεια της εκπαίδευσης και προβλέπει την τελική κλάση μέσω πλειοψηφίας (Breiman, 2001). Η βασική αρχή του συγκεκριμένου αλγορίθμου είναι η τυχαία δημιουργία αυτών των δέντρων με διαφορετικά υποσύνολα των χαρακτηριστικών που έχουν τα δεδομένα, επιφέροντας βελτίωση της ακρίβειας και μείωση του κινδύνου υπερπροσαρμογής του μοντέλου.

Η δυνατότητα του Random Forest να δουλεύει με μεγάλο και ανισόρροπο όγκο δεδομένων και να παράγει σταθερές και αξιόπιστες προβλέψεις τον καθιστά ιδανικό για εφαρμογές σε χρηματοοικονομικά δεδομένα. Παράλληλα, το γεγονός ότι μπορεί να αναγνωρίσει πολύπλοκες σχέσεις και να εξετάσει όλα τα χαρακτηριστικά και τη σημασία τους ταυτόχρονα, ενισχύει την ικανότητά του να εντοπίζει σημαντικούς παράγοντες που οδηγούν στην πτώχευση (Zhou et al., 2014).

3.4.2 Decision Tree

Ο αλγόριθμος Decision Tree είναι ένα δημοφιλές και εύχρηστο εργαλείο μηχανικής μάθησης για προβλήματα ταξινόμησης και παλινδρόμησης. Τα Δέντρα Απόφασης λειτουργούν με

βάση τη διάσπαση των δεδομένων σε επαναλαμβανόμενους κόμβους, οι οποίοι καταλήγουν σε τελικές προβλέψεις, ανάλογα με τις τιμές των χαρακτηριστικών κάθε παρατήρησης. Για το βέλτιστο διαχωρισμό σε κάθε κόμβο και τη δημιουργία του τελικού δέντρου, το οποίο είναι εύκολα ερμηνεύσιμο και παρέχει οπτική αναπαράσταση της λογικής ταξινόμησης, ο αλγόριθμος χρησιμοποιεί μετρικές όπως το Gini Index ή η Entropy (Quinlan, 1986).

Να επισημανθεί σε αυτό το σημείο πως τα Δέντρα Απόφασης είναι ιδιαίτερα επιρρεπή στην υπερπροσαρμογή, ειδικά όταν τα δεδομένα είναι πολύπλοκα ή περιέχουν θόρυβο. Ωστόσο, η απλότητα και η ερμηνευσιμότητα τους τα καθιστά εξαιρετικά χρήσιμα σε περιβάλλοντα όπου απαιτείται η κατανόηση της διαδικασίας λήψης αποφάσεων, καθώς παρέχουν με ξεκάθαρο τρόπο τις παραμέτρους που οδηγούν σε αυτή (Yeh & Lien, 2009).

3.4.3 Logistic Regression

Η Λογιστική Παλινδρόμηση αποτελεί ένα από τα πιο απλά και ευρέως χρησιμοποιημένα μοντέλα δυαδικής ταξινόμησης, ιδιαίτερα στο χώρο των χρηματοοικονομικών. Στοχεύει στην εκτίμηση της πιθανότητας υπαγωγής μιας παρατήρησης σε μία από τις δύο κατηγορίες με βάση έναν γραμμικό συνδυασμό των γνωρισμάτων (Hosmer et al., 2013). Το αποτέλεσμα ερμηνεύεται μέσω της συνάρτησης «sigmoid» που περιορίζει την έξοδο στο διάστημα $[0, 1]$.

Αν και εστιάζει μόνο στη μελέτη γραμμικών αλληλεπιδράσεων μεταξύ των μεταβλητών, το μοντέλο είναι στατιστικά ισχυρό, γρήγορο και ερμηνεύσιμο, γεγονός που το καθιστά αξιόπιστο εργαλείο.

3.4.4 XGBoost

Ο XGBoost είναι ένας ενισχυτικός αλγόριθμος μάθησης (boosting), που συνδυάζει πολλαπλά αδύναμα μοντέλα, συνήθως δέντρα απόφασης, σε ένα ισχυρότερο, διορθώνοντας τα σφάλματα των προηγούμενων μοντέλων μέσω της διαδικασίας gradient descent. Για να μειώσει την υπερπροσαρμογή χρησιμοποιεί κανονικοποίηση (regularization), καθιστώντας τον πιο σταθερό και αποδοτικό έναντι άλλων αλγορίθμων boosting (Chen & Guestrin, 2016).

Ο εν λόγω αλγόριθμος είναι ιδιαίτερα αποτελεσματικός, καθώς χειρίζεται δεδομένα με ελλείψεις ή θόρυβο, διατηρώντας παράλληλα υψηλή ακρίβεια. Εκτός αυτού, προσφέρει υψηλή ταχύτητα εκπαίδευσης και μπορεί να εκπαιδευτεί σε μεγάλες κλίμακες δεδομένων, γεγονός κρίσιμο για το χρηματοοικονομικό περιβάλλον.

3.4.5 Support Vector Classifier (SVC)

Ο SVC είναι ένας από τους ισχυρότερους αλγορίθμους Μηχανικής Μάθησης που χρησιμοποιείται σε προβλήματα ταξινόμησης δεδομένων. Στηρίζεται στην εύρεση ενός ορίου-υπεριπέδου (hyperplane) που διαχωρίζει τις παρατηρήσεις διαφορετικών κατηγοριών,

διασφαλίζοντας τη μέγιστη δυνατή απόσταση μεταξύ των κοντινότερων σημείων κάθε κλάσης. Στην περίπτωση που τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, ο αλγόριθμος χρησιμοποιεί τη μέθοδο των πυρήνων (kernel trick), για να τα διαχωρίσει σε ένα μεγαλύτερο χώρο διάστασης (Cortes & Vapnik, 1995).

Χαρακτηριστικό γνώρισμα του SVC είναι η υψηλή του απόδοση σε προβλήματα με πολύπλοκες κατανομές. Ωστόσο, κατά τη διαχείριση μεγάλου όγκου δεδομένων ενδέχεται να προβεί σε υπερβολική υπολογιστική δαπάνη, μειώνοντας σημαντικά με αυτό το τρόπο την απόδοση του.

3.4.6 Naive Bayes

Ο αλγόριθμος Naive Bayes είναι βασισμένος στη θεωρία των πιθανοτήτων και χρησιμοποιεί το θεώρημα του Bayes για την ταξινόμηση δεδομένων. Η βασική του λειτουργία υπολογίζει την πιθανότητα κάθε κλάσης για μια νέα παρατήρηση και την ταξινομεί βάσει της μέγιστης πιθανοφάνειας.

Παρά την απλότητά του και τη μειωμένη απόδοση σε πολύπλοκες μη γραμμικές σχέσεις, ο αλγόριθμος αποδεικνύεται αποτελεσματικός σε αρκετές περιπτώσεις, ιδιαίτερα όταν τα χαρακτηριστικά των δεδομένων είναι ανεξάρτητα μεταξύ τους (Rish, 2001). Η ταχύτητα εκπαίδευσης του τον καθιστά ως έναν καλό αλγόριθμο για απλές εκτιμήσεις του κινδύνου πτώχευσης, ιδανικά όταν οι πληροφορίες είναι περιορισμένες ή το μοντέλο απαιτεί υψηλή επεξεργαστική αποδοτικότητα.

3.4.7 Multi-layer Perceptron

Ο Multi-layer Perceptron (MLP) είναι ένας τύπος τεχνητού νευρωνικού δικτύου που αποτελείται από πολλά επίπεδα (layers) νευρώνων που συνδέονται με βάρη. Η εκπαίδευση του MLP πραγματοποιείται μέσω της διαδικασίας της οπισθοδιάδοσης σφαλμάτων (backpropagation), με στόχο την ελαχιστοποίηση του σφάλματος εξόδου. Η ικανότητα του να μαθαίνει μη γραμμικές σχέσεις τον καθιστά πολύ ισχυρό σε προβλήματα, όπως η πρόβλεψη χρεοκοπίας (Haykin, 1998).

Η μεγαλύτερη πρόκληση του είναι η απαιτητική υπολογιστική διαδικασία εκπαίδευσης, καθώς ο αριθμός των παραμέτρων που πρέπει να ρυθμιστούν είναι μεγάλος (Yeh et al., 2011). Εκτός αυτού, η ερμηνευσιμότητα των αποτελεσμάτων είναι περιορισμένη, καθώς ο MLP ανήκει στην κατηγορία των «μαύρων κουτιών», καθιστώντας δύσκολη την κατανόηση των εσωτερικών του αποφάσεων.

3.5 Διαχωρισμός Train και Test Split

Για την εκπαίδευση και την αξιολόγηση των μοντέλων μηχανικής μάθησης που χρησιμοποιήθηκαν στην παρούσα μελέτη, το τελικό σύνολο δεδομένων χωρίστηκε σε σύνολο εκπαίδευσης (training set) και σύνολο δοκιμής (test set) με αναλογία 80% προς 20%. Η επιλογή αυτής της αναλογίας έγινε ώστε να διατηρηθεί ένα επαρκές πλήθος παρατηρήσεων για την εκπαίδευση των αλγορίθμων, χωρίς να θυσιάζεται η αξιοπιστία της τελικής αξιολόγησης.

Ο διαχωρισμός πραγματοποιήθηκε με τυχαία δειγματοληψία, ενώ για τη διασφάλιση της στατιστικής συνέπειας εφαρμόστηκε η μέθοδος διαστρωμάτωσης (stratified sampling), για να διατηρηθεί η αρχική αναλογία μεταξύ των δύο κατηγοριών, πτωχευμένων και υγιών επιχειρήσεων, τόσο στο σύνολο εκπαίδευσης, όσο και αυτό της δοκιμής. Παράλληλα χρησιμοποιήθηκε σταθερή τιμή τυχαίου σπόρου `random_state=42` για τη δειγματοληψία, στοχεύοντας στην επαναληψιμότητα και την επαλήθευση των αποτελεσμάτων σε μελλοντικές εφαρμογές.

Σε αυτό το σημείο αξίζει να σημειωθεί πως οι τεχνικές εξισορρόπησης που αναλύθηκαν σε προηγούμενη ενότητα εφαρμόστηκαν αποκλειστικά στο σύνολο εκπαίδευσης. Στόχος αυτής της ενέργειας είναι να διασφαλίσει ότι η αξιολόγηση των μοντέλων πραγματοποιείται σε πραγματικά και μη επεξεργασμένα δεδομένα, αποφεύγοντας το φαινόμενο της διαρροής πληροφορίας (data leakage) και ενισχύοντας την αξιοπιστία των αποτελεσμάτων.

3.6 Μετρικές Αξιολόγησης

Η αξιολόγηση της απόδοσης των αλγορίθμων επιτεύχθηκε με τη χρήση πολλαπλών μετρικών, με στόχο την ολοκληρωμένη κατανόηση της συμπεριφοράς καθενός στα δεδομένα της παρούσας μελέτης. Ειδικότερα, χρησιμοποιήθηκαν οι μετρικές ευαισθησία (sensitivity), ειδικότητα (specificity), ισορροπημένη ακρίβεια (balanced accuracy), καμπύλη ROC, δείκτης AUC και πίνακας σύγχυσης (confusion matrix).

Για την ανάλυση των προβλέψεων χρησιμοποιήθηκε ο πίνακας σύγχυσης (Confusion Matrix), ο οποίος καταγράφει τέσσερις βασικούς τύπους αποτελεσμάτων:

		Predicted Values	
		Positive	Negative
Actual Values	Positive	TP	FN
	Negative	FP	TN

Σχήμα 6: Παράδειγμα Confusion Matrix.

- (a) Αληθώς Θετικά (True Positive-TP): περιπτώσεις όπου το μοντέλο προέβλεψε πτώχευση και πτώχευσε όντως.
- (b) Ψευδώς Θετικά (False Positive-NP): περιπτώσεις όπου το μοντέλο προέβλεψε πτώχευση ενώ στην πραγματικότητα δεν ισχύει.
- (c) Αληθώς Αρνητικά (True Negative-TN): περιπτώσεις όπου το μοντέλο προέβλεψε χρεοκοπία και αυτό επιβεβαιώθηκε.
- (d) Ψευδώς Αρνητικά (False Negative-FN): περιπτώσεις όπου το μοντέλο προέβλεψε ότι μια επιχείρηση θα πτωχεύσει και τελικά δεν πτώχευσε.

Η ευαισθησία μετράει την ικανότητα του μοντέλου να εντοπίζει σωστά τις επιχειρήσεις που πτώχευσαν, ελαχιστοποιώντας τα ψευδώς αρνητικά αποτελέσματα. Υπολογίζεται ως το ποσοστό των πραγματικά πτωχευμένων που προβλέφθηκαν σωστά.

$$Sensitivity = \frac{TP}{TP + NP}$$

Η ειδικότητα, αντίστοιχα, μετράει την ακρίβεια του μοντέλου να αναγνωρίζει την αρνητική τάξη, δηλαδή τις υγιείς επιχειρήσεις. Υπολογίζεται ως το ποσοστό των πραγματικά ενεργών επιχειρήσεων που προβλέφθηκαν σωστά.

$$Specificity = \frac{TN}{TN + FN}$$

Η μετρική Balanced Accuracy είναι ο μέσος όρος μεταξύ της ευαισθησίας και της ειδικότητας, ιδιαίτερα χρήσιμος σε περιπτώσεις μη ισορροπημένων συνόλων δεδομένων. Παρέχει μια πιο δίκαιη εκτίμηση της απόδοσης και για τις δύο κλάσεις, σε αντίθεση με την συνολική ακρίβεια (Accuracy) όπου μπορεί να εμφανίσει υψηλές τιμές ακόμα και όταν αγνοείται η μειονοτική τάξη.

$$Balanced Accuracy = \frac{Sensitivity + Specificity}{2}$$

Η καμπύλη ROC (Receiver Operating Characteristic) αποτελεί εργαλείο οπτικής αξιολόγησης της ικανότητας ενός δυαδικού ταξινομητή να διακρίνει μεταξύ των δύο κατηγοριών, απεικονίζοντας τη σχέση μεταξύ της ευαισθησίας και της ειδικότητας. Όσο πιο πάνω και αριστερά βρίσκεται η καμπύλη, τόσο καλύτερη είναι η επίδοση του μοντέλου.

Ο δείκτης AUC (Area Under the Curve) συνοψίζει την καμπύλη ROC σε μια αριθμητική τιμή που κυμαίνεται μεταξύ του διαστήματος 0 και 1. Αντιπροσωπεύει την πιθανότητα του μοντέλου να ταξινομεί σωστά ένα τυχαίο ζευγάρι δειγμάτων, ένα θετικό και ένα αρνητικό. Τιμή AUC ίση με 0,5 υποδηλώνει τυχαία πρόβλεψη, ενώ τιμή κοντά στο 1 φανερώνει εξαιρετικά διακριτική ικανότητα.

Κεφάλαιο 4

Αποτελέσματα

4.1 Παρουσίαση Αποτελεσμάτων

Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά τα αποτελέσματα πρόβλεψης για κάθε αλγόριθμο Μηχανικής Μάθησης που χρησιμοποιήθηκε, σύμφωνα με τις μετρικές αξιολόγησης που αναλύθηκαν σε προηγούμενη ενότητα. Πέρα από την απλή παρουσίαση των τιμών, υπάρχει σχολιασμός της ικανότητας του εκάστοτε αλγορίθμου να αναγνωρίζει τις καταστάσεις χρεοκοπίας, δίνοντας μεγαλύτερη βαρύτητα στον δείκτη ευαισθησίας.

4.1.1 Random Forest

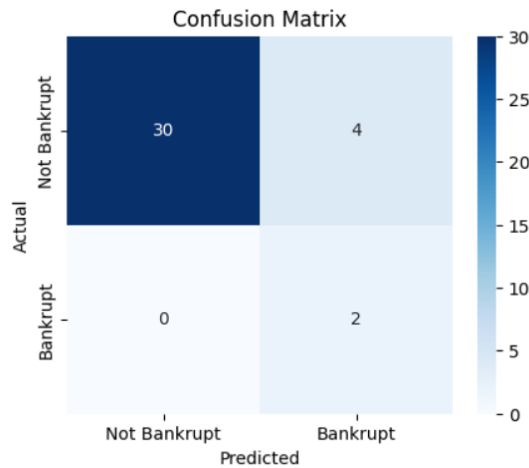
Ο πρώτος αλγόριθμος που εφαρμόστηκε στο εξισορροπημένο πλέον σύνολο δεδομένων είναι ο Random Forest. Αφού προηγήθηκε η εκπαίδευση του, ακολούθησε η πρόβλεψη χρησιμοποιώντας το σύνολο δοκιμής (test set) με τα εξής αποτελέσματα:

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	1.0
Ειδικότητα (Specificity)	0,88
Balanced Accuracy	0,94
AUC	1,0

Πίνακας 2: Αποτελέσματα απόδοσης Random Forest.

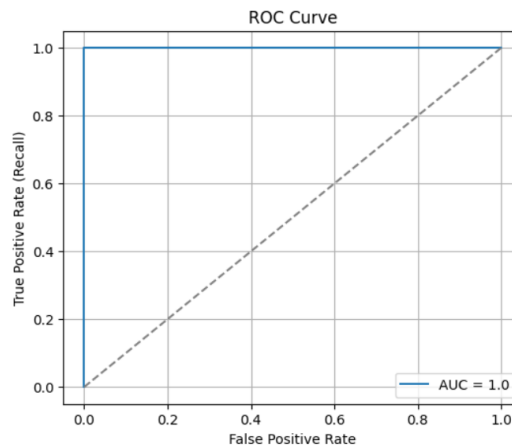
Το Balanced Accuracy που ανέρχεται στο 0,94 είναι αρκετά ικανοποιητικό, καθώς το μοντέλο παρουσιάζει καλή επίδοση για τον εντοπισμό και των δύο κατηγοριών. Η τιμή της ευαισθησίας ίση με 1 δείχνει την εξαιρετική ικανότητα του αλγορίθμου να εντοπίζει όλες τις επιχειρήσεις που είναι πτωχευμένες, ενώ η ειδικότητα που ισούται με 0,88 φανερώνει ότι αναγνώρισε σωστά μόνο το 88% των ενεργών επιχειρήσεων.

Πιο συγκεκριμένα αν παρατηρήσουμε το Σχήμα 4.1 μπορούμε να διακρίνουμε ότι από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 30, ενώ και οι 2 ανενεργές που είχε προβλέφθηκαν ορθά.



Σχήμα 7: Confusion Matrix για τον Random Forest.

Ο δείκτης AUC ανήλθε στο 1 δείχνοντας εξαιρετική διακριτική ικανότητα του μοντέλου μεταξύ των πτωχευμένων και μη επιχειρήσεων, γεγονός που επιβεβαιώνεται και από την καμπύλη ROC στο Σχήμα 4.2.



Σχήμα 8: Η καμπύλη ROC για τον Random Forest.

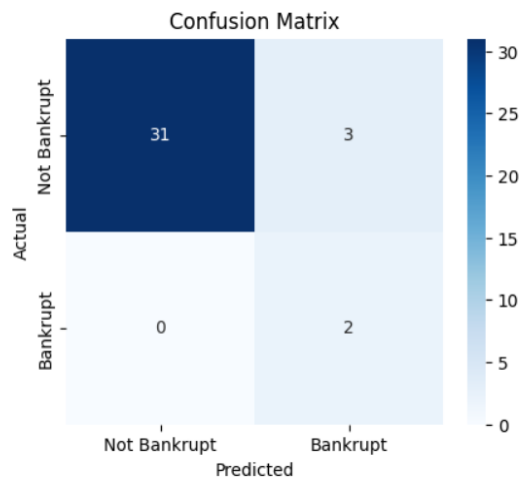
4.1.2 Decision Tree

Ο αλγόριθμος Decision Tree παρουσιάζει αρκετά ικανοποιητική απόδοση στην πρόβλεψη τόσο των ανενεργών όσο και των ενεργών επιχειρήσεων, γεγονός που αποδεικνύεται από την τιμή της ευαισθησίας, ίση με 1, και την τιμή της ειδικότητας, ίση με 0,91. Το Balanced Accuracy έφτασε μέχρι 0,95, υποδηλώνοντας ισορροπημένη απόδοση και στις δύο κατηγορίες.

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	1,0
Ειδικότητα (Specificity)	0,91
Balanced Accuracy	0,95
AUC	0,95

Πίνακας 3: Αποτελέσματα απόδοσης Decision Tree.

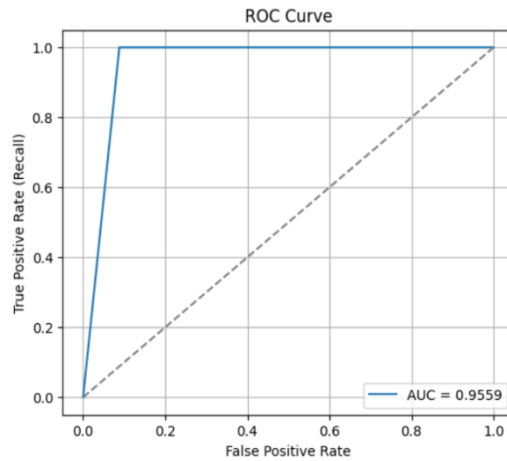
Στο Σχήμα 4.3 μπορούμε να διακρίνουμε ότι από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 31, ενώ και οι 2 ανενεργές που υπάρχουν προβλέφθηκαν ορθά.



Σχήμα 9: Confusion Matrix για τον Decision Tree.

Ο δείκτης AUC ανήλθε στο 0,95 δείχνοντας ιδιαίτερα ικανοποιητική ικανότητα του μοντέλου να διακρίνει με ακρίβεια μεταξύ των πτωχευμένων και μη επιχειρήσεων.

Στο Σχήμα 4.4, η καμπύλη ROC πλησιάζει την επάνω αριστερή γωνία, επιβεβαιώνοντας υψηλή ευαισθησία και χαμηλό ποσοστό ψευδών θετικών προβλέψεων.



Σχήμα 10: Η καμπύλη ROC για τον Random Forest.

4.1.3 Logistic Regression

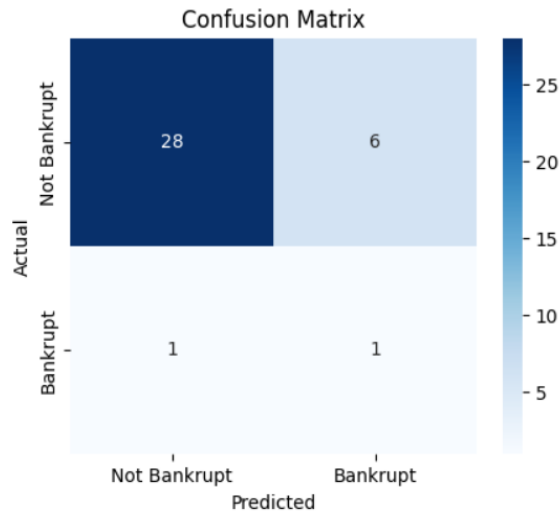
Τα αποτελέσματα από την εφαρμογή της Λογιστικής Παλινδρόμησης στο σύνολο δοκιμής είναι τα ακόλουθα:

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	0,5
Ειδικότητα (Specificity)	0,82
Balanced Accuracy	0,66
AUC	0,86

Πίνακας 4: Αποτελέσματα απόδοσης Logistic Regression.

Το Balanced Accuracy ανέρχεται στο 0.66 υποδεικνύοντας τη μέτρια απόδοση του εν λόγω αλγορίθμου. Ειδικότερα, το μοντέλο ανιχνεύει μόνο το 50% των επιχειρήσεων που πτώχευσαν και το 82% των επιχειρήσεων που βρίσκονται σε λειτουργία, υποδηλώνοντας περιθώριο βελτίωσης.

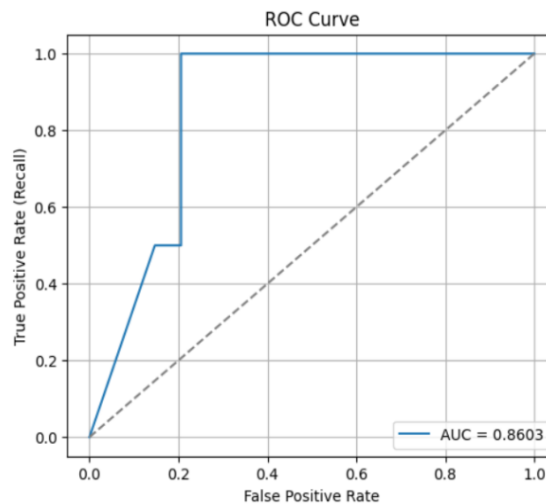
Από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 28, σε αντίθεση με τις ανενεργές επιχειρήσεις όπου έχουμε μόνο 1 σωστή πρόβλεψη, όπως φαίνεται στο Σχήμα 4.5.



Σχήμα 11: Confusion Matrix για τον Logistic Regression.

Επίσης το μοντέλο έχει AUC ίσο με 0,86 δείχνοντας μια καλή ικανότητα διάκριση μεταξύ θετικών και αρνητικών παρατηρήσεων.

Τέλος, στο Σχήμα 4.6 φαίνεται η καμπύλη ROC, η οποία στην αρχή ανεβαίνει πολύ γρήγορα αλλά μετά γίνεται πιο επίπεδη, γεγονός που δείχνει ότι το μοντέλο αναγνωρίζει καλά τις πτωχευμένες επιχειρήσεις, αλλά κάνει αρκετές ψευδείς προβλέψεις όσον αφορά τις υγιείς.



Σχήμα 12: Η καμπύλη ROC για τον Logistic Regression.

4.1.4 XGBoost

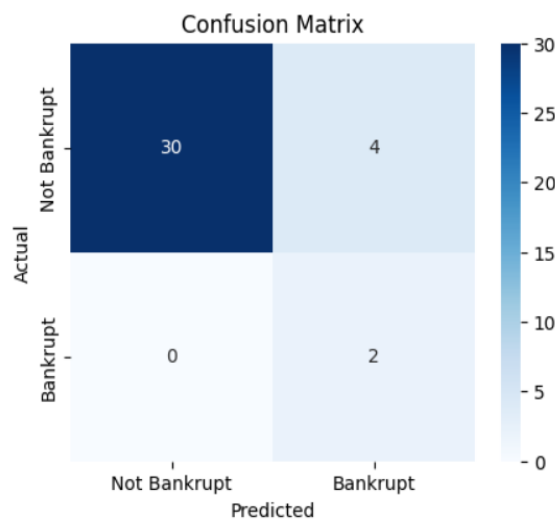
Σύμφωνα με τον πίνακα που παρουσιάζεται παρακάτω, ο XGBoost έχει ικανοποιητική διακριτική απόδοση και για τις δύο κατηγορίες.

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	1,0
Ειδικότητα (Specificity)	0,88
Balanced Accuracy	0,94
AUC	0,98

Πίνακας 5: Αποτελέσματα απόδοσης XGBoost.

Το Balanced Accuracy που ανέρχεται στο 0,94 είναι αρκετά ικανοποιητικό, καθώς το μοντέλο παρουσιάζει καλή επίδοση για τον εντοπισμό και των δύο κατηγοριών. Η τιμή της ευαισθησίας ίση με 1 δείχνει την εξαιρετική ικανότητα του αλγορίθμου να εντοπίζει όλες τις επιχειρήσεις που είναι πτωχευμένες, ενώ η ειδικότητα που ισούται με 0,88 φανερώνει ότι αναγνώρισε σωστά το 88% των ενεργών επιχειρήσεων.

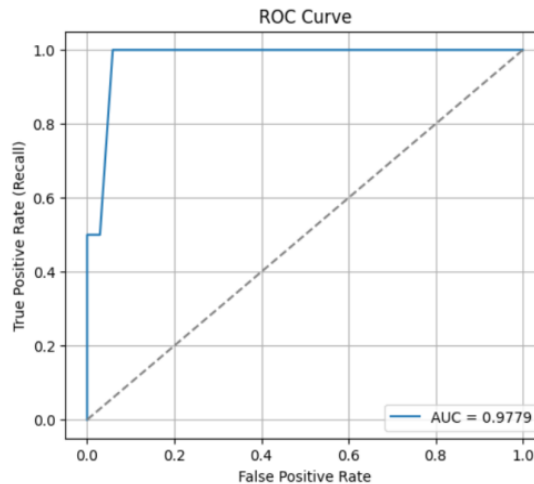
Πιο συγκεκριμένα αν παρατηρήσουμε το Σχήμα 4.7 μπορούμε να διακρίνουμε ότι από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 30, ενώ και οι 2 ανενεργές προβλέφθηκαν ορθά.



Σχήμα 13: Confusion Matrix για τον XGBoost.

Ο δείκτης AUC ανήλθε στο 0,98 δείχνοντας εξαιρετική διακριτική ικανότητα του μοντέλου να ξεχωρίζει με ακρίβεια μεταξύ των πτωχευμένων και μη επιχειρήσεων.

Η καμπύλη ROC έχει πολύ απότομη άνοδο προς τα πάνω και στη συνέχεια κινείται προς το άνω άκρο, ανιχνεύοντας όλες τις παρατηρήσεις πτώχευσης και κάνοντας ελάχιστες λάθος προβλέψεις για αυτές των μη πτωχευμένων.



Σχήμα 14: Η καμπύλη ROC για τον XGBoost.

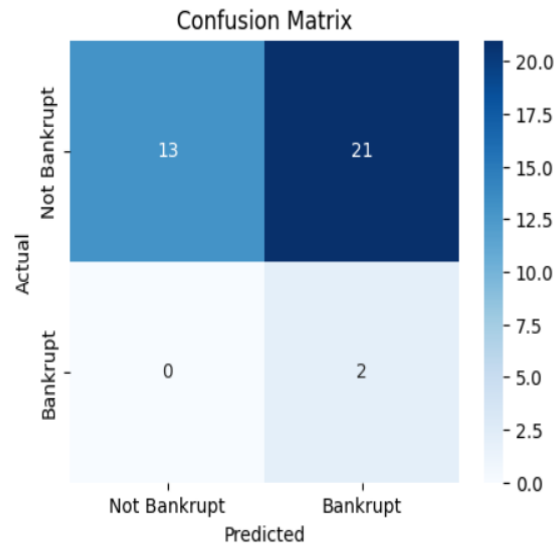
4.1.5 SVC

Ακολουθούν τα αποτελέσματα απόδοσης του SVC. Αναλυτικότερα, το Balanced Accuracy του αλγορίθμου ανέρχεται στο 0,69, φανερώνοντας την μέτρια διακριτική ικανότητα του μοντέλου. Παρ' όλο που η ευαισθησία ισούται με 1, δηλαδή εντοπίζονται όλες οι πτωχευμένες επιχειρήσεις, ο SVC παρουσιάζει ειδικότητα ίση με 0,38. Αυτό σημαίνει πως αρκετές υγιείς επιχειρήσεις χαρακτηρίζονται ως χρεοκοπημένες.

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	1,0
Ειδικότητα (Specificity)	0,38
Balanced Accuracy	0,69
AUC	1,0

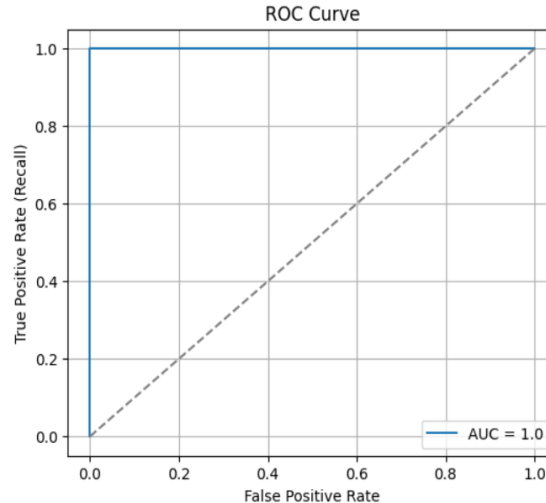
Πίνακας 6: Αποτελέσματα απόδοσης για τον SVC.

Εάν παρατηρήσουμε το Σχήμα 4.9 μπορούμε να διακρίνουμε ότι οι 2 ανενεργές επιχειρήσεις προβλέφθηκαν ορθά, όμως από το σύνολο των 34 ενεργών επιχειρήσεων ανιχνεύθηκαν σωστά μόνο οι 13.



Σχήμα 15: Confusion Matrix για τον SVC.

Ωστόσο, αν και η καμπύλη ROC που φαίνεται στο Σχήμα 4.10, όπως και ο δείκτης AUC που ισούται με 1, δείχνουν ότι το μοντέλο μπορεί να διακρίνει σε θεωρητικό επίπεδο εξαιρετικά και τις δύο κατηγορίες, παρότι η πραγματική του απόδοση είναι μέτρια, όπως αποδείχθηκε παραπάνω.



Σχήμα 16: Η καμπύλη ROC για τον SVC.

4.1.6 Multi-Layer Perceptron

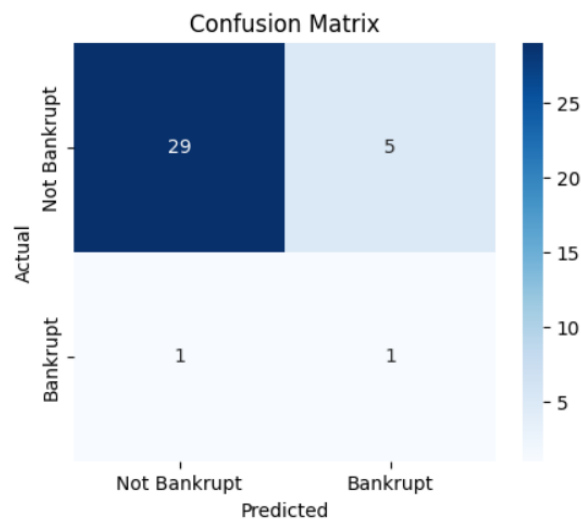
Όσον αφορά τον Multi-Layer Perceptron, τα αποτελέσματα αποτυπώνουν μια μέτρια διαχωριστική ικανότητα, γεγονός που επιβεβαιώνεται από το Balanced Accuracy που ισούται με 0,68.

Μετρική Αξιολόγησης	Τιμή
Εναισθησία (Sensitivity)	0,5
Ειδικότητα (Specificity)	0,85
Balanced Accuracy	0,68
AUC	0,68

Πίνακας 7: Αποτελέσματα απόδοσης για τον Multi-Layer Perceptron.

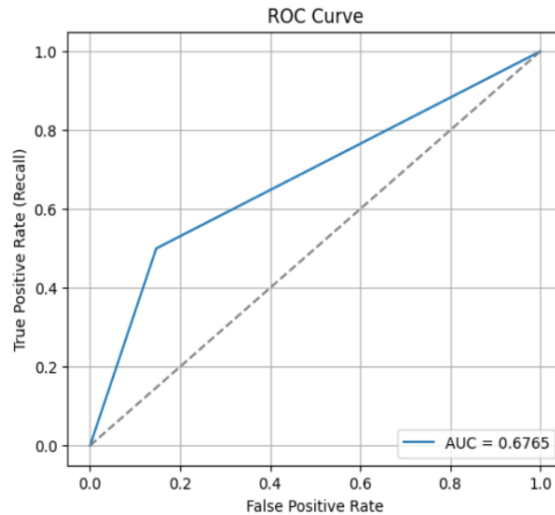
Αναλυτικότερα, το μοντέλο ανιχνεύει μόνο το 50% των επιχειρήσεων που πτώχευσαν και το 85% των επιχειρήσεων που βρίσκονται σε λειτουργία, υποδηλώνοντας μεγάλο περιθώριο βελτίωσης.

Από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 29, σε αντίθεση με τις ανενεργές επιχειρήσεις όπου έχουμε μόνο 1 σωστή πρόβλεψη, όπως φαίνεται στο Σχήμα 4.5.



Σχήμα 17: Confusion Matrix για τον Multi-Layer Perceptron.

Η καμπύλη ROC που απέχει αισθητά από το ιδανικό σημείο (0,1), όπως φαίνεται στο Σχήμα 4.12 και ο δείκτης AUC που ισούται με 0,68 επιβεβαιώνουν την μέτρια απόδοση του μοντέλου να διακρίνει τις δύο περιπτώσεις.



Σχήμα 18: Η καμπύλη ROC για τον Multi-Layer Perceptron.

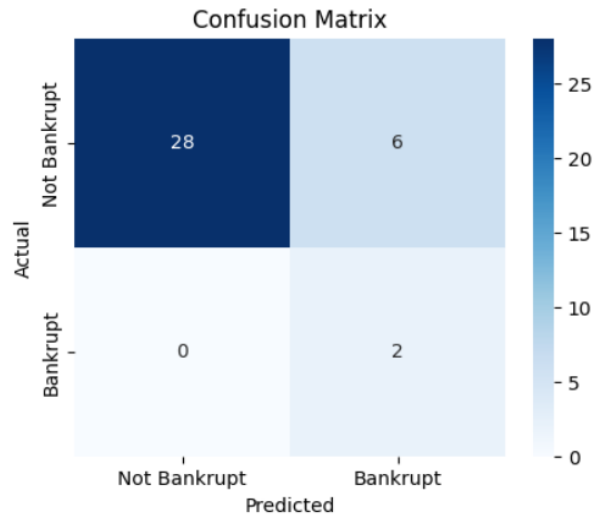
4.1.7 Naive Bayes

Ο αλγόριθμος Naive Bayes παρουσιάζει ικανοποιητική απόδοση να διακρίνει τις παρατηρήσεις και των δύο κατηγοριών, με το Balanced Accuracy να ανέρχεται στο 0,91. Συγκεκριμένα, αναγνωρίζει όλες τις πτωχευμένες παρατηρήσεις, καθώς πετυχαίνει ευαισθησία ίση με 1, ενώ από την κατηγορία των πτωχευμένων επιχειρήσεων χαρακτηρίζει ορθά το 82% αυτών.

Μετρική Αξιολόγησης	Τιμή
Ευαισθησία (Sensitivity)	1,0
Ειδικότητα (Specificity)	0,82
Balanced Accuracy	0,91
AUC	0,91

Πίνακας 8: Αποτελέσματα απόδοσης για τον Naive Bayes.

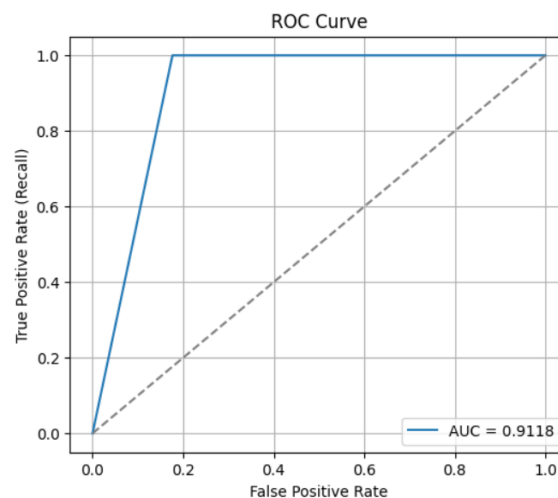
Από το Σχήμα 4.13 μπορούμε να διακρίνουμε ότι από τις 34 ενεργές επιχειρήσεις που περιλαμβάνει το σύνολο δοκιμής ανιχνεύθηκαν σωστά 28, ενώ και οι 2 ανενεργές προβλέφθηκαν ορθά.



Σχήμα 19: Confusion Matrix για τον Naive Bayes.

Ο δείκτης AUC ανήλθε στο 0,91 δείχνοντας ικανοποιητική ικανότητα του μοντέλου να διακρίνει τις πτωχευμένες και μη επιχειρήσεις.

Στο Σχήμα 4.14, η καμπύλη ROC πλησιάζει την επάνω αριστερή γωνία, επιβεβαιώνοντας υψηλή ευαισθησία και χαμηλό ποσοστό ψευδών θετικών προβλέψεων.



Σχήμα 20: Η καμπύλη ROC για τον Naive Bayes.

4.2 Επιλογή Βέλτιστου Αλγορίθμου

Αναλύοντας τα αποτελέσματα του Πίνακα 4.8 που παρατίθεται παρακάτω, εξάγονται σημαντικά και κρίσιμα συμπεράσματα σχετικά με την απόδοση των αλγορίθμων που εφαρμόστηκαν, δίνοντας μεγαλύτερη βαρύτητα στην ευαισθησία, λόγω της σημασίας της έγκαιρης αναγνώρισης ενδεχόμενης αποτυχίας.

Μετρική Αξιολόγησης	Random Forest	Decision Tree	Logistic Regression	XGBoost	SVC	Multi-Layer Perceptron	Naive Bayes
Ευαισθησία (Sensitivity)	1,0	1,0	0,5	1,0	1,0	0,5	1,0
Ειδικότητα (Specificity)	0,88	0,91	0,82	0,88	0,38	0,85	0,82
Balanced Accuracy	0,94	0,95	0,66	0,94	0,69	0,68	0,91
AUC	1,0	0,95	0,86	0,98	1,0	0,68	0,91

Πίνακας 9: Συγκεντρωτικά αποτελέσματα απόδοσης.

Αναλυτικότερα, οι αλγόριθμοι Random Forest, Decision Tree, XGBoost, SVC και Naive Bayes παρουσίασαν άριστη προγνωστική ικανότητα, καθώς πέτυχαν ευαισθησία ίση με 1,0, αναγνωρίζοντας ορθά το σύνολο των πτωχευμένων επιχειρήσεων. Αντίθετα, οι Logistic Regression και Multi-Layer Perceptron ταξινόμησαν σωστά τις μισές επιχειρήσεις που χρεοκόπησαν, πετυχαίνοντας ευαισθησία ίση με 0,5.

Ωστόσο, υπάρχουν σαφείς διαφοροποιήσεις μεταξύ των αλγορίθμων σχετικά με την ικανότητα τους να ταξινομούν σωστά τις υγιείς επιχειρήσεις. Συγκεκριμένα, μόνο ο Decision Tree κατάφερε να πετύχει υψηλή ειδικότητα, αναγνωρίζοντας σωστά το 91% των ενεργών εταιρειών, ακολουθώντας οι Random Forest και XGBoost με ειδικότητα ίση με 0,88, πετυχαίνοντας ισορροπία στην ταξινόμηση και των δύο κατηγοριών. Αντίθετα, ο SVC παρότι πέτυχε AUC ίσο με 1, η ειδικότητά του ανέρχεται μόλις στο 0.38, δείχνοντας πως είναι ιδιαίτερα αδύναμος, καθώς χαρακτηρίζει τις περισσότερες υγιείς επιχειρήσεις ως χρεοκοπημένες.

Η ισορροπημένη ακρίβεια (Balanced Accuracy), η οποία λαμβάνει υπόψη την ευαισθησία και την ειδικότητα, αναδεικνύει τον Decision Tree ως τον αλγόριθμο με την καλύτερη συνολική επίδοση, με τιμή 0,95. Ικανοποιητικά αποτελέσματα παρουσιάζουν και οι Random Forest με 0,94, σε αντίθεση με τους SVC (0,69), MLP (0,68) και Logistic Regression (0,66), επιβεβαιώνοντας την περιορισμένη ικανότητά τους να διαχειρίζονται σύνθετα σύνολα δεδομένων.

Συνοψίζοντας, βέλτιστη επιλογή για το σύνολο δεδομένων που αξιοποιείται στην παρούσα μελέτη αποτελεί ο Decision Tree, καθώς πετυχαίνει υψηλή απόδοση τόσο στην αναγνώριση των υγιών, όσο και των πτωχευμένων επιχειρήσεων. Η σταθερότητα και η αυξημένη ερμηνεία του τον καθιστούν κατάλληλο εργαλείο, υπερτερώντας έναντι πιο σύνθετων και ισχυρών αλγορίθμων.

Κεφάλαιο 5

Συμπεράσματα, περιορισμοί και προτάσεις για μελλοντική έρευνα

5.1 Συμπεράσματα

Η πτώχευση αποτελεί ένα πολυδιάστατο φαινόμενο που υφίσταται μια οικονομική οντότητα, η οποία δεν απορρέει μόνο από την επίδραση ενός παράγοντα, αλλά από τον συνδυασμό ενδογενών και εξωγενών μεταβλητών. Οι επιπτώσεις της επηρεάζουν όχι μόνο τους άμεσα εμπλεκόμενους, αλλά και την συνολική οικονομία, καθιστώντας τη δυνατότητα έγκαιρης πρόβλεψης της ένα στρατηγικό εργαλείο για την αποτροπή οικονομικών κρίσεων.

Η ανάλυση των στατιστικών δεδομένων της Eurostat για τις επιχειρήσεις της Ευρωπαϊκής Ένωσης την περίοδο 2018 έως 2025 παρουσίασε ενδιαφέροντα συμπεράσματα. Συγκεκριμένα, παρατηρείται υψηλή διακύμανση στον αριθμό των πτωχεύσεων, ο οποίος ελαχιστοποιείται κατά την περίοδο της πανδημίας Covid-19, λόγω των κρατικών ενισχύσεων. Ωστόσο, από το 2022 όπου πραγματοποιήθηκε άρση των οικονομικών μέτρων, παρατηρείται δραματική αύξηση των πτωχεύσεων, ιδιαίτερα στους τομείς της πληροφορικής, της εκπαίδευσης και της εστίασης.

Στο πλαίσιο αυτό, έχουν αναπτυχθεί και αξιοποιηθεί ποικίλες μεθοδολογίες πρόβλεψης, που περιλαμβάνουν τόσο παραδοσιακές στατιστικές τεχνικές, όπως η διακριτική ανάλυση, η πολυμεταβλητή ανάλυση και το γραμμικό υπόδειγμα πιθανότητας, όσο και σύγχρονα μοντέλα μηχανικής μάθησης, όπως ο Random Forest, ο Decision Tree και ο SVC. Τα τελευταία χρόνια η βιβλιογραφία δείχνει μια σαφή στροφή προς τα υβριδικά μοντέλα, τα οποία συνδυάζουν πολλαπλούς αλγορίθμους με στόχο τη βελτίωση της ακρίβειας των προβλέψεων. Παράλληλα, δίνεται ιδιαίτερη έμφαση στη συμπερίληψη μη χρηματοοικονομικών παραγόντων, όπως η φήμη της επιχείρησης, η ποιότητα διοίκησης και η στρατηγική κατεύθυνσης, οι οποίες μπορούν να επηρεάσουν σημαντικά την πιθανότητα οικονομικής κατάρρευσης.

Η παρούσα μελέτη πραγματοποίησε μια συγκριτική ανάλυση μεταξύ διαφόρων αλγορίθμων μηχανικής μάθησης. Αναλυτικότερα, χρησιμοποιήθηκαν οικονομικά δεδομένα για 190 ευρωπαϊκές επιχειρήσεις από τους τομείς των τηλεπικοινωνιών, των υπολογιστών και της τεχνολογίας για την περίοδο 2014 έως 2023. Το εν λόγω δείγμα περιλαμβάνει 180 ενεργές επιχειρήσεις, καθώς και 10 που έχουν υποστεί πτώχευση, για το οποίο αξιοποιήθηκαν 13 χρηματοοικονομικοί δείκτες. Το σύνολο δεδομένων είναι μη ισορροπημένο, με τις ενεργές επιχειρήσεις να υπερτερούν αριθμητικά έναντι των πτωχευμένων. Το φαινόμενο αυτό αντιμετωπίστηκε με τεχνικές εξισορρόπησης, όπως η SMOTE και η Random Under-Sampling.

Στο ερευνητικό σκέλος, εφαρμόστηκαν επτά αλγόριθμοι: Random Forest, XGBoost, Logistic Regression, SVC, Naive Bayes, Decision Tree και Multi-Layer Perceptron. Η αξιολόγηση τους έγινε με βάση πολυδιάστατες μετρικές απόδοσης, όπως η ισορροπημένη ακρίβεια (Balanced Accuracy), η ευαισθησία (Sensitivity), η ειδικότητα (Specificity) και το AUC.

Από την ανάλυση των αποτελεσμάτων, ο Decision Tree αποδείχθηκε ως το βέλτιστο μοντέλο για την πρόβλεψη της χρηματοοικονομικής αποτυχίας με βάση τα συγκεκριμένα δεδομένα της εργασίας, καθώς πέτυχε την υψηλότερη απόδοση από το σύνολο των αλγορίθμων. Συγκεκριμένα, ο αλγόριθμος αυτός κατάφερε να ταξινομήσει με απόλυτη ακρίβεια το σύνολο των πτωχευμένων επιχειρήσεων, καθώς και το 91% των ενεργών, επιτυγχάνοντας συνολική ισορροπημένη ακρίβεια 95%.

Αντιθέτως, τη χειρότερη απόδοση κατέγραψε ο SVC, ο οποίος μπόρεσε να αναγνωρίσει μόνο το 38% από τις ενεργές επιχειρήσεις και το 50% των πτωχευμένων. Το εύρημα αυτό αποδεικνύει την ανάγκη να αξιολογούνται τα μοντέλα όχι μόνο προς την συνολική ακρίβεια, αλλά και ως προς την ικανότητά τους να αναγνωρίζουν ορθά και τις δύο κατηγορίες.

5.2 Περιορισμοί και προτάσεις για μελλοντική έρευνα

Ανεξάρτητα από τα θετικά ευρήματα της έρευνας, υπάρχουν κάποιες αδυναμίες, που περιορίζουν την εμβέλεια και την γενικευσιμότητα των συμπερασμάτων. Η αναγνώριση αυτών των περιορισμών μπορούν να αποτελέσουν αντικείμενο για περαιτέρω μελέτη.

Καταρχήν, ένας βασικός περιορισμός της μελέτης είναι η μεγάλη ανισορροπία που επικρατεί στο σύνολο δεδομένων, καθώς η αναλογία των πτωχευμένων επιχειρήσεων σε σχέση με τις ενεργές είναι ιδιαίτερα μικρή. Παρόλο που χρησιμοποιήθηκαν τεχνικές, όπως οι SMOTE και Random Under-Sampling, η περιορισμένη εκπροσώπηση της συγκεκριμένης τάξης ενδέχεται να έχει επηρεάσει αρνητικά την αξιοπιστία των μοντέλων. Για το σκοπό αυτό, μελλοντικές έρευνες προτείνεται να διευρύνουν το δείγμα αξιοποίησης με περισσότερες επιχειρήσεις, διατηρώντας μεγαλύτερη ισορροπία μεταξύ των δύο κατηγοριών.

Επιπλέον, επιπρόσθετο ενδιαφέρον θα παρουσίαζε η επιλογή επιπλέον κρίσιμων μεταβλητών που θα μπορούσαν να επηρεάσουν σε ένα βαθμό την οικονομική κατάσταση της επιχείρησης, καθώς στη μελέτη αξιοποιούνται αποκλειστικά χρηματοοικονομικές μεταβλητές. Έτσι, παράγοντες όπως η εταιρική φήμη, η ποιότητα διοίκησης ή και η ανταγωνιστική θέση στην αγορά θα μπορούσαν να συνεισφέρουν στην προβλεπτική ισχύ των μοντέλων. Επομένως, συνίσταται η ενσωμάτωση μη χρηματοοικονομικών δεικτών σε μελλοντικές μελέτες, για μια πιο ολοκληρωμένη εκτίμηση του κινδύνου πτώχευσης.

Καταλήγοντας, γίνεται κατανοητό πως η βιωσιμότητα και η μακροχρόνια ανάπτυξη μιας οικονομικής οντότητας εξαρτάται από πλήθος παραγόντων. Διαρκή πρόκληση αποτελεί ο εντοπισμός κάθε πιθανού παράγοντα που μπορεί να επηρεάσει την κατάσταση της επιχείρησης, καθώς και η εξέλιξη των προσεγγίσεων πρόβλεψης της εταιρικής πτώχευσης.

Βιβλιογραφία

- Abdullah, M. (2021). *The implication of machine learning for financial solvency prediction: An empirical analysis on public listed companies of Bangladesh*. Journal of Asian Business and Economic Studies, 28(4), 303–320.
- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Ajayi, A. O., Bilal, M., Akinade, O. O., & Muhammad, M. (2018). *Systematic review of bankruptcy prediction models: Towards a framework for tool selection*. Expert Systems with Applications, 94, 164–184.
- Altman, E. I. (1968). *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. The Journal of Finance, 23(4), 589–609.
- Aziz, M. A., & Dar, H. A. (2006). *Predicting corporate bankruptcy: Where we stand?* Corporate Governance: The International Journal of Business in Society, 6(1), 18–33.
- Barboza, F., Kimura, H., & Altman, E. (2017). *Machine learning models and bankruptcy prediction*. Expert Systems with Applications, 83, 405–417.
- Beaver, W. H. (1966). *Financial Ratios as Predictors of Failure*. Journal of Accounting Research, 4, 71–111.
- Bernhardsen, E. (2001). *A model of bankruptcy prediction [Working paper]*. Norwegian School of Economics and Business Administration.
- Blagus, R., & Lusa, L. (2013). *SMOTE for high-dimensional class-imbalanced data*. BMC Bioinformatics, 14(1), 106.
- Bradley, D. B., III, & Cowdery, C. (2004). *Small business: Causes of bankruptcy*. University of Central Arkansas.
- Brédart, X. (2014). *Bankruptcy prediction model using neural networks*. Accounting and Finance Research, 3(2), 124–128.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321–357.
- Chen, M.-Y. (2011). *Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches*. Computers & Mathematics with Applications, 62(12), 4514–4524.
- Dasilas, A., & Rigani, A. (2024). *Machine learning techniques in bankruptcy prediction: A systematic literature review*. Expert Systems with Applications, 255, 124761.
- Drummond, C., & Holte, R. C. (2003). *C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling*. Workshop on Learning from Imbalanced Datasets II.
- Eurostat. (2025). *Quarterly registrations of new businesses and declarations of bankruptcies - statistics*.
- Fallahpour, S., Najafi, M., & Rahmani, D. (2017). *Bankruptcy prediction using SVM and feature selection*. International Journal of Finance & Economics, 22(2), 124–135.

- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from Imbalanced Data Sets*. Springer.
- Geng, R., Bose, I., & Chen, X. (2015). *Prediction of financial distress: An empirical study of listed Chinese companies using data mining*. *European Journal of Operational Research*, 241(1), 236–247.
- He, H., & Garcia, E. A. (2009). *Learning from imbalanced data*. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- Jones, S., Johnstone, D., & Wilson, R. (2017). *Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks*. *Journal of Business Finance & Accounting*, 44(1–2), 3–34.
- Koyuncugil, A. S., & Ozgulbas, N. (2012). *Financial early warning system model and data mining application for risk detection*. *Expert Systems with Applications*, 39(6), 6238–6253.
- Kubat, M., & Matwin, S. (1997). *Addressing the curse of imbalanced training sets: One-sided selection*. In *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 179–186). Morgan Kaufmann.
- Lahmiri, S., & Bekiros, S. (2024). *Can machine learning approaches predict corporate bankruptcy? Evidence from a qualitative experimental design*. *Quantitative Finance*.
- Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). *Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning*. *Journal of Machine Learning Research*, 18(17), 1–5.
- Mansouri, A., Nazari, A., & Ramazani, M. (2016). *A comparison of artificial neural network model and logistics regression in prediction of companies' bankruptcy (A case study of Tehran Stock Exchange)*. *International Journal of Advanced Computer Research*, 6(24), 81–89.
- Máté, D., Raza, H., & Ahmad, I. (2023). *Comparative analysis of machine learning models for bankruptcy prediction in the context of Pakistani companies*. *Risks*, 11(10), 176.
- Mbat, D. O., & Eyo, E. I. (2013). *Corporate Failure: Causes and Remedies*. *Business and Management Research*, 2(4), 19–24.
- Meese, E. N., & Viken, T. (2019). *Machine Learning in Bankruptcy Prediction*. Master Thesis, Norwegian School of Economics.
- Narvekar, A., & Guha, D. (2021). *Bankruptcy prediction using machine learning and an application to the case of the COVID-19 recession*. *Data Science in Finance and Economics*, 1(2), 180–195.
- Narvekar, A., & Guha, D. (2025). *Intelligent corporate bankruptcy prediction: A data-driven machine learning approach*.
- Olson, D. L., Delen, D., & Meng, Y. (2012). *Comparative analysis of data mining methods for bankruptcy prediction*. *Decision Support Systems*, 52, 464–473.
- Shetty, S., Musa, M., & Brédart, X. (2022). *Bankruptcy prediction using machine learning techniques*. *Journal of Risk and Financial Management*, 15(1), 35.

- Shirata, C. Y. (2012). *Financial ratios as predictors of bankruptcy in Japan: An empirical research*. Tokyo International University.
- Titman, S., Keown, A. J., & Martin, J. D. (2015). *Χρηματοοικονομική διαχείριση: Αρχές και εφαρμογές (12η έκδ.)*. Αθήνα: Εκδόσεις Παπαζήση.
- Tsai, C.-F., Hsu, Y.-F., & Yen, D. C. (2014). *A comparative study of classifier ensembles for bankruptcy prediction*. *Applied Soft Computing*, 24, 977–984.
- Veganzones, D., & Séverin, E. (2018). *An investigation of bankruptcy prediction in imbalanced datasets*. *Decision Support Systems*, 112, 111–124.
- Weiss, L. A., & Capkun, V. (2005). *The impact of incorporating the cost of errors into bankruptcy prediction models*. SSRN.
- White, M. J. (2016). *Small business bankruptcy*. *Annual Review of Financial Economics*, 8, 317–336.
- Yen, S. J., & Lee, Y. S. (2006). *Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset*. *Intelligent Control and Automation*, 6(1), 1–9.

Παράρτημα Α: Πηγαίος Κώδικας Υλοποίησης

Στο παράρτημα αυτό παρατίθενται ο κώδικας που αναπτύχθηκε για τις ανάγκες της παρούσας μελέτης. Η υλοποίηση του πραγματοποιήθηκε στο περιβάλλον Google Colab με τη γλώσσα προγραμματισμού Python, λόγω των πλούσιων βιβλιοθηκών που προσφέρει για ανάλυση δεδομένων και τεχνικές μηχανικής μάθησης.

Ο κώδικας περιλαμβάνει όλα τα απαραίτητα στάδια, από την εισαγωγή και προεπεξεργασία των δεδομένων, έως την εκπαίδευση των μοντέλων και την αξιολόγηση τους. Για κάθε βήμα παρέχονται σχόλια, ώστε να γίνεται κατανοητή η διαδικασία που ακολουθείται.

1. Δήλωση Βιβλιοθηκών

```
#Εισαγωγή βιβλιοθηκών
import pandas as pd
import numpy as np
import xgboost as xgb
from google.colab import files
from collections import Counter
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from imblearn.pipeline import Pipeline
from imblearn.under_sampling import RandomUnderSampler
from imblearn.combine import SMOTEENN
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from imblearn.over_sampling import SMOTE
from sklearn.naive_bayes import GaussianNB
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import (
    confusion_matrix,
    roc_auc_score,
    roc_curve
)
```


Σχόλια: Στο στάδιο αυτό εισάγονται οι απαραίτητες βιβλιοθήκες για την ανάλυση των δεδομένων, την χρήση των αλγορίθμων μηχανικής μάθησης και της οπτικοποίησης της απόδοσης τους.

2.Δημιουργία μεθόδων

```
#Δημιουργία μεθόδου για ανέβασμα του αρχείου
#Επιτρέπει στον χρήστη να ανεβάσει ένα αρχείο και επιστρέφει το όνομα του αρχείου που
#ανέβηκε, το οποίο στη συνέχεια θα επεξεργαστεί.
def upload_dataset():
    data_upload = files.upload()
    filename = list(data_upload.keys())[0] # Παίρνουμε το όνομα του αρχείου όπως είναι
    return filename

#Δημιουργία μεθόδου για διαχείριση της ανισορροπίας των δεδομένων του dataset
def handle_class_imbalance(X, y):
    # Εφαρμόζουμε RandomUnderSampler
    X_under, y_under = RandomUnderSampler(sampling_strategy=0.6,
    random_state=42).fit_resample(X, y)

    # Εφαρμόζουμε SMOTE για να δημιουργήσουμε τεχνητά δείγματα
    X_resampled, y_resampled = SMOTE(random_state=42).fit_resample(X_under, y_under)

    return X_resampled, y_resampled

#Δημιουργία μεθόδου για εκπαίδευση του μοντέλου και πρόβλεψη
def train_and_predict(model, X_resampled, y_resampled, X_test):
    model.fit(X_resampled, y_resampled) #Εκπαίδευση του μοντέλου με το μετασχηματισμένο
    train set
    y_pred = model.predict(X_test) #Πρόβλεψη χρησιμοποιώντας το test set
    return y_pred

#Δημιουργία μεθόδου για την εμφάνιση των μετρικών αξιολόγησης του μοντέλου
def evaluation_metrics(y_test, y_pred, y_prob=None):
    #Δημιουργία του Confusion Matrix
    cm = confusion_matrix(y_test, y_pred)
    TN, FP, FN, TP = cm.ravel()

    #Υπολογισμός μετρικών
    sensitivity = TP / (TP + FN) #Το ποσοστό των πτωχευμένων επιχειρήσεων που
    προβλέφθηκαν σωστά από το μοντέλο.
```

```
specificity = TN / (TN + FP) #Το ποσοστό των υγιών επιχειρήσεων που αναγνωρίστηκαν σωστά από το μοντέλο.
```

```
balanced_acc = (sensitivity + specificity) / 2 #Ο μέσος όρος της Sensitivity και της Specificity.
```

```
#Υπολογισμός AUC
```

```
auc = None
```

```
if y_prob is not None:
```

```
    auc = roc_auc_score(y_test, y_prob)
```

```
#Εκτύπωση Μετρικών
```

```
print("Sensitivity (Recall):", round(sensitivity, 4))
```

```
print("Specificity:", round(specificity, 4))
```

```
print("Balanced Accuracy:", round(balanced_acc, 4))
```

```
if auc is not None:
```

```
    print("AUC:", round(auc, 4))
```

```
print("Confusion Matrix:\n", cm)
```

```
#Εμφάνιση του Confusion Matrix σε γράφημα
```

```
plt.figure(figsize=(5,4))
```

```
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=["Not Bankrupt",  
"Bankrupt"], yticklabels=["Not Bankrupt", "Bankrupt"])
```

```
plt.title("Confusion Matrix")
```

```
plt.xlabel("Predicted")
```

```
plt.ylabel("Actual")
```

```
plt.show()
```

```
#Δημιουργία και εμφάνιση διαγράμματος ROC
```

```
if y_prob is not None:
```

```
    fpr, tpr, thresholds = roc_curve(y_test, y_prob)
```

```
    plt.figure(figsize=(6,5))
```

```
    plt.plot(fpr, tpr, label=f"AUC = {round(auc, 4)}")
```

```
    plt.plot([0, 1], [0, 1], linestyle='--', color='gray')
```

```
    plt.xlabel("False Positive Rate")
```

```
    plt.ylabel("True Positive Rate (Recall)")
```

```
    plt.title("ROC Curve")
```

```
    plt.legend(loc="lower right")
```

```
    plt.grid()
```

```
    plt.show()
```

Σχόλια: Στο σημείο αυτό υπάρχουν συγκεντρωμένες όλες οι κρίσιμες μέθοδοι, προσφέροντας ευελιξία στο πρόγραμμα, καθώς σε περίπτωση μελλοντικής τροποποίησης σε κάποια δραστηριότητα δε χρειάζεται να επέμβουμε σε ολόκληρο το κώδικα. Συγκεκριμένα, εδώ περιλαμβάνονται οι μέθοδοι για το ανέβασμα του αρχείου, το διαχωρισμού και την εξισορρόπηση του συνόλου δεδομένων, καθώς για την εμφάνιση των μετρικών αξιολόγησης.

3.Ανέβασμα του αρχείου excel με τα δεδομένα

```
# Ανέβασμα του αρχείου
filename = upload_dataset()

# Ανάγνωση του Excel αρχείου
companies = pd.read_excel(filename, na_values=["n.a."]) # Μετατρέπει τα "n.a." σε NaN
κατά την ανάγνωση

# Διαγραφή όλων των γραμμών που περιέχουν έστω και ένα NaN
companies.dropna(inplace=True)

companies
```

Σχόλια: Εδώ γίνεται η ανάγνωση του αρχείου, η επεξεργασία του, ώστε να διαγραφούν τυχόν τιμές τύπου n.a και η εμφάνιση του νέου τροποποιημένου αρχείου.

4.Διαχωρισμός Δεδομένων σε Χαρακτηριστικά (Features) και Ετικέτες (Labels)

```
X = companies.drop(columns=['Bankruptcy (0=active 1=inactive)']) #Η μεταβλητή x
περιλαμβάνει όλες τις στήλες εκτός της Bankruptcy (Features)
y = companies['Bankruptcy (0=active 1=inactive)'] #Η μεταβλητή y περιλαμβάνει τη στήλη
Bankruptcy (Labels)

#Το συγκεκριμένο κομμάτι κώδικα εκτελείται μόνο στην περίπτωση που χρησιμοποιήσουμε
τον αλγόριθμο XGBOOST. Αφαιρεί τους ειδικούς χαρακτήρες.
X.columns = X.columns.astype(str).str.replace(r'[\[\]<>]', "", regex=True)
```

Σχόλια: Στο σημείο γίνεται ο διαχωρισμός των δεδομένων σε χαρακτηριστικά και σε ετικέτες. Με άλλα λόγια, διαχωρίζουμε το σύνολο δεδομένων σε δύο νέα σύνολα, όπου το ένα περιλαμβάνει την εξαρτημένη μεταβλητή που δηλώνει αν μια επιχείρηση είναι πτωχευμένη (τιμή 1) ή όχι (τιμή 0) και ένα άλλο σύνολο που περιλαμβάνει τις εξαρτημένες μεταβλητές, δηλαδή τους χρηματοοικονομικούς δείκτες για κάθε επιχείρηση.

5. Διαχωρισμός Δεδομένων σε Σύνολα Εκπαίδευσης (train test) και Δοκιμής (test set)

```
X_train, X_test, y_train, y_test = train_test_split(X, #Features
                                                    y, #Labels
                                                    test_size=0.2, stratify=y, #Ορίζει ότι το 20% των δεδομένων
                                                    θα χρησιμοποιηθεί για το test set
                                                    random_state=42)
# - X_train: Τα χαρακτηριστικά για το σύνολο εκπαίδευσης
# - X_test: Τα χαρακτηριστικά για το σύνολο δοκιμής
# - y_train: Οι ετικέτες για το σύνολο εκπαίδευσης
# - y_test: Οι ετικέτες για το σύνολο δοκιμής
```

Σχόλια: Τα σύνολα δεδομένων X και y που δημιουργήθηκαν παραπάνω διαχωρίζονται σε train και test set. Το train set περιλαμβάνει το 80% του συνόλου των επιχειρήσεων, ενώ το υπόλοιπο 20% θα χρησιμοποιηθεί στο test set. Ο διαχωρισμός γίνεται με τυχαία δειγματοληψία.

6. Random Forest

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model = RandomForestClassifier(random_state=42, class_weight='balanced')#Δημιουργία
μοντέλου Random Forest
#To class_weight='balanced' διορθώνει την
ανισορροπία κατηγοριών δίνοντας μεγαλύτερη βαρύτητα στη μειοψηφία

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης
```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή Random Forest. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η

οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

7. *Decision Tree*

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model=DecisionTreeClassifier(random_state=42, class_weight='balanced') #Δημιουργία
μοντέλου Decision Tree

#To class_weight='balanced' διορθώνει την
ανισορροπία κατηγοριών δίνοντας μεγαλύτερη βαρύτητα στη μειοψηφία

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης
```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή Decision Tree. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

8. *Logistic Regression*

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model=LogisticRegression(random_state=42, class_weight='balanced') #Δημιουργία
μοντέλου Logistic Regression

#To class_weight='balanced' διορθώνει την
ανισορροπία κατηγοριών δίνοντας μεγαλύτερη βαρύτητα στη μειοψηφία
```

```
y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης
```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή Logistic Regression. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

9.XGBoost

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model=xgb.XGBClassifier(reg_alpha=1, reg_lambda=1) #Δημιουργία μοντέλου XGBoost

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης
```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή XGBoost. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

10.SVC

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας
```

```

model=SVC(random_state=42,probability=True, class_weight='balanced') #Δημιουργία
μοντέλου SVC

#To class_weight='balanced' διορθώνει την
ανισορροπία κατηγοριών δίνοντας μεγαλύτερη βαρύτητα στη μειοψηφία

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης

```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή SVC. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

11. Multi-Layer Perceptron

```

X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model=MLPClassifier(hidden_layer_sizes=(100), random_state=42, max_iter=1000)
#Δημιουργία μοντέλου Multi-Layer Perceptron

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης

```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή Multi-Layer Perceptron. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.

12.Naive Bayes

```
X_resampled, y_resampled=handle_class_imbalance(X_train, y_train) #Καλεί την
συνάρτηση handle_class_imbalance για διαχείριση της ανισορροπίας

model=GaussianNB() #Δημιουργία μοντέλου Naive Bayes

y_pred=train_and_predict(model, X_resampled, y_resampled, X_test) #Καλεί την
συνάρτηση train_and_predict για εκπαίδευση του μοντέλου και πρόβλεψη
y_proba = model.predict_proba(X_test)[:, 1]

evaluation_metrics(y_test, y_pred, y_proba) #Καλεί την συνάρτηση evaluation_metrics για
εμφάνιση των μετρικών αξιολόγησης
```

Σχόλια: Στο σημείο αυτό δημιουργούμε έναν ταξινομητή Naive Bayes. Αρχικά, καλούμε τη μέθοδο για την εξισορρόπηση μεταξύ των δύο κατηγοριών (πτωχευμένες και ενεργές), η οποία εφαρμόζεται μόνο στο σύνολο εκπαίδευσης. Ακολουθεί η εκπαίδευση του αλγορίθμου με βάση το σύνολο εκπαίδευσης, που μόλις ολοκληρωθεί εκτελείται η πρόβλεψη αξιοποιώντας το test set. Στο τέλος, εμφανίζονται οι μετρικές αξιολόγησης και τα αντίστοιχα διαγράμματα.