



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΔΙΟΙΚΗΣΗ, ΛΟΓΙΣΤΙΚΗ ΚΑΙ
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

**ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΕΦΑΡΜΟΓΕΣ ΣΥΣΤΗΜΑΤΩΝ
ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΑΝΑΓΙΩΤΑΣ ΧΑΤΖΗΣΤΑΥΡΟΥ

Επιβλέπων : Ευστάθιος Γ. Κύρκος
Καθηγητής, ΔΙ.ΠΑ.Ε.

Θεσσαλονίκη, Δεκέμβριος 2022

Η σελίδα αυτή είναι σκόπιμα λευκή.



ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΔΙΟΙΚΗΣΗ, ΛΟΓΙΣΤΙΚΗ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΑ
ΣΥΣΤΗΜΑΤΑ

Επιχειρηματικές Εφαρμογές Συστημάτων Βαθιάς Μάθησης

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΑΝΑΓΙΩΤΑΣ ΧΑΤΖΗΣΤΑΥΡΟΥ

Επιβλέπων : Ευστάθιος Γ. Κύρκος
Καθηγητής, ΔΙ.ΠΑ.Ε.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις XX/XX/XXX.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Όνομα Επώνυμο
Βαθμίδα, Ίδρυμα

.....
Όνομα Επώνυμο
Βαθμίδα, Ίδρυμα

.....
Όνομα Επώνυμο
Βαθμίδα, Ίδρυμα

Θεσσαλονίκη, Δεκέμβριος 2022

(Υπογραφή)

.....

Επώνυμο, Όνομα.

.

© date– Allrightsreserved

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Ευστάθιο Κύρκο για την προτροπή του να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα και για την πολύτιμη υποστήριξη και καθοδήγηση που μου παρείχε κατά την εκπόνηση της διπλωματικής μου εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου για την κατανόηση και την υπομονή που έδειξε κατά τη διάρκεια των μεταπτυχιακών μου σπουδών.

Η σελίδα αυτή είναι σκόπιμα λευκή

Περίληψη

Στις μέρες μας η σημασία της πρόβλεψης της μελλοντικής οικονομικής κατάστασης των επιχειρήσεων αποκτά όλο και μεγαλύτερη σημασία, και είναι χρήσιμη τόσο για τις ίδιες τις επιχειρήσεις όσο και για τα ενδιαφερόμενα μέρη της αγοράς, διότι βοηθάει στην έγκαιρη πληροφόρηση και υποστήριξη των διαδικασιών λήψης αποφάσεων για την επίτευξη των στόχων τους, και την προστασία τους από τις οικονομικές απώλειες του παρελθόντος. Η ανάγκη πρόβλεψης της μελλοντικής οικονομικής κατάστασης των επιχειρήσεων οδήγησε τους ερευνητές να ασχοληθούν με την ανάπτυξη αυτοματοποιημένων συστημάτων πρόβλεψης. Μέχρι σήμερα οι περισσότερες μελέτες εφάρμοσαν παραδοσιακές μεθόδους στατιστικής ανάλυσης και μηχανικής μάθησης, όμως με αρκετά μειονεκτήματα λόγω των περιορισμένων δυνατοτήτων τους. Έτσι η ανάγκη χρήσης πιο εξελιγμένης τεχνολογίας ήταν επιτακτική. Τα τελευταία χρόνια η τεχνολογική πρόοδος και η δυνατότητα διαχείρισης μεγάλου όγκου δεδομένων δημιούργησαν τις προϋποθέσεις για την ανάπτυξη της μεθόδου deep learning. Πρόκειται για μια προηγμένη προσέγγιση μηχανικής μάθησης, η οποία μάλιστα χρησιμοποιείται ήδη σε πλήθος εφαρμογών, καλύπτοντας ένα ευρύ φάσμα επιστημονικών πεδίων με επιτυχία. Αντικείμενο της παρούσας εργασίας είναι η βιβλιογραφική ανασκόπηση δεκαπέντε ερευνητικών άρθρων που έχουν πραγματοποιηθεί τα τελευταία χρόνια για την πρόβλεψη της πιθανότητας απάτης, οικονομικής δυσπραγίας, χρεοκοπίας, οικονομικής αποτυχίας και συνεχιζόμενης δραστηριότητας των επιχειρήσεων. Σκοπός είναι η διερεύνηση της ενδεχόμενης ανωτερότητας και της αποτελεσματικότητας των μοντέλων deep learning έναντι των παραδοσιακών στατιστικών και machine learning μοντέλων, ο εντοπισμός του οφέλους χρήσης τους αλλά και των ευκαιριών που προσφέρουν για περαιτέρω έρευνα. Πράγματι, στο σύνολο των πειραματικών διαδικασιών, τα μοντέλα deep learning σημείωσαν εξαιρετικές αποδόσεις επιβεβαιώνοντας την ανωτερότητα και την αξιοπιστία τους, και υποδεικνύοντας ότι πρόκειται για μια πολλά υποσχόμενη μέθοδο, ικανή να συμβάλλει στην ενίσχυση των διαδικασιών λήψης αποφάσεων όλων των ενδιαφερόμενων μερών, στη βελτίωση και επιτάχυνση των ελεγκτικών διαδικασιών και την αποτελεσματικότερη διοίκηση. Ωστόσο χρήζει περαιτέρω έρευνας δεδομένου ότι μέχρι σήμερα η εφαρμογή της στη χρηματοοικονομική ανάλυση είναι ακόμα περιορισμένη.

Λέξεις Κλειδιά: << βαθιά μάθηση, χρεοκοπία, οικονομική απάτη, οικονομική δυσπραγία, ετήσιες εκθέσεις, εταιρική αποτυχία, συνεχιζόμενη δραστηριότητα >>.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Abstract

Nowadays the importance of predicting the future financial situation of companies is becoming more and more important, and it is useful both for the companies themselves and for the market stakeholders because it helps in timely information provision and supporting decision-making processes, to achieve their goals and protect them from the past financial losses. The need of predicting the future financial situation of companies led researchers to deal with the development of automated predicting systems. Until today most studies have applied traditional statistical analysis and machine learning methods, but with several disadvantages due to their limited capabilities. So, the need of using more advanced technology was imperative. In the last years the technological progress and the ability to manage large amounts of data created the conditions for the development of deep learning methods. It is about an advanced machine learning approach which in fact is already used in many applications covering a wide range of scientific fields with success. The object of this paper is the bibliographic review of fifteen research articles which have taken place in recent years to predict the probability of fraud, financial distress, bankruptcy, financial failure and going concern of companies. The purpose is to assess the superiority and the effectiveness of deep learning models over traditional statistical and machine learning models, to identify the benefit of their use and also the opportunities they offer for future research. Indeed, in all the experimental procedures, the deep learning models showed excellent performance, confirming their superiority and reliability, as well as indicating that it is about a promising method capable to contribute to strengthening the decision-making processes of all stakeholders, improving and accelerating the auditing procedures and more efficient administration. However, it needs further research, given the fact that until today its application in financial analysis is still limited.

Keywords: “ deep learning, bankruptcy, financial fraud, financial distress, annual reports, corporate failure, going concern “.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΕΦΑΡΜΟΓΕΣ ΣΥΣΤΗΜΑΤΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ	i
1 Εισαγωγή.....	10
1.1 Αντικείμενο και σκοπός της διπλωματικής	12
1.2 Διάρθρωση κειμένου	12
2 Προηγούμενες Έρευνες.....	13
2.1 Παραδοσιακές στατιστικές μέθοδοι	13
2.2 Παραδοσιακές μέθοδοι Machine Learning	14
2.3 Μέθοδος Deep Learning.....	14
3 Εισαγωγικές έννοιες	15
3.1 Τεχνητή νοημοσύνη (Artificial Intelligence, AI)	15
3.1.1 Κωδικοποίηση γνώσης και αναπαράσταση δεδομένων	16
3.1.2 Αναπαράσταση κειμένου.....	17
3.2 Deep Learning.....	20
3.2.1 Ιστορική αναδρομή	21
3.2.2 Περιγραφή τεχνητού νευρωνικού δικτύου (artificial neural networks, ANNs)	23
3.2.3 Υπερπαραμέτροι τεχνητών νευρωνικών δικτύων	25
3.2.4 Εκπαίδευση, επικύρωση και έλεγχος νευρωνικού δικτύου	26
3.2.5 Μειονεκτήματα μοντέλων deep learning	26
4 Περιγραφή των ερευνητικών άρθρων	28
4.1 Deep learning for detecting financial statement fraud	30
4.2 An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning.....	35
4.3 Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry.....	43
4.4 A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	48
4.5 Predicting distresses using deep learning of text segments in annual reports.....	57

4.6	<i>Bankruptcy prediction using imaged financial ratios and convolutional neural networks.....</i>	64
4.7	<i>Deep learning models for bankruptcy prediction using textual disclosures</i>	71
4.8	<i>A Deep Dense Neural Network for Bankruptcy Prediction</i>	78
4.9	<i>Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE</i>	81
4.10	<i>Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors.....</i>	87
4.11	<i>Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM</i>	93
4.12	<i>Corporate failure prediction: An evaluation of deep learning vs discrete hazard models.....</i>	97
4.13	<i>Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure</i>	102
4.14	<i>Using Deep Learning Algorithms for CPAs' Going Concern Prediction</i>	108
4.15	<i>Artificial intelligence in corporate sustainability: using LSTM and GRU for going concern prediction.....</i>	112
5	Ανάλυση των ερευνητικών άρθρων	119
5.1	<i>Τύποι νευρωνικών δικτύων.....</i>	119
5.1.1	<i>Recurrent Neural Networks (RNN).....</i>	119
5.1.2	<i>Convolutional Neural Networks (CNN)</i>	123
5.1.3	<i>Multi-Layer Perceptron (MLP).....</i>	124
5.2	<i>Ο σκοπός της πρόβλεψης των ερευνητικών άρθρων</i>	125
5.2.1	<i>Πρόβλεψη απάτης</i>	125
5.2.2	<i>Πρόβλεψη οικονομικής δυσπραγίας</i>	126
5.2.3	<i>Πρόβλεψη χρεοκοπίας.....</i>	127
5.2.4	<i>Πρόβλεψη οικονομικής αποτυχίας.....</i>	129
5.2.5	<i>Πρόβλεψη μη συνεχιζόμενης δραστηριότητας</i>	130
5.3	<i>Τα δεδομένα.....</i>	131
5.4	<i>Εξισορρόπηση του συνόλου δεδομένων.....</i>	134
5.5	<i>Τεχνικές επεξεργασίας φυσικής γλώσσας</i>	135
5.6	<i>Επιλογή χαρακτηριστικών</i>	137
5.7	<i>Επικύρωση (validation)</i>	138
5.8	<i>Ερμηνευσιμότητα</i>	140

5.9	Μοντέλα αναφοράς	141
5.10	Μέτρα απόδοσης	143
5.11	Στατιστική ανάλυση των αποτελεσμάτων.....	148
5.12	Συμπεράσματα	148
6	Επίλογος	150
6.1	Σύνοψη και συμπεράσματα	150
6.2	Μελλοντικές προοπτικές (Περιορισμοί και μελλοντικές επεκτάσεις).....	151
7	Βιβλιογραφία	153

Ευρετήριο Εικόνων

Εικόνα 1: Σχηματική απεικόνιση διαγραμμάτων ροής αναπαράστασης δεδομένων	17
Εικόνα 2: Σχηματική αναπαράστασης των αρχιτεκτονικών CBOW & Skip-gram.....	19
Εικόνα 3: Artificial Intelligence (AI) - Machine Learning (ML) - Deep Learning (DL)	20
Εικόνα 4: McCulloch and Pitts neuron.....	21
Εικόνα 5: Ενεργοποίηση νευρώνα.....	23
Εικόνα 6: Τεχνητό νευρωνικό δίκτυο 3 επιπέδων feed forward.....	24
Εικόνα 7: Κωδικοποίηση εγγράφου με το μοντέλο HAN	31
Εικόνα 8: Η αρχιτεκτονική του μοντέλου HAN.....	32
Εικόνα 9: Κείμενο από την ενότητα MD&A (στα αριστερά), εξαγωγή των "red -flag" προτάσεων (στα δεξιά)	35
Εικόνα 10: Η ροή των διαδικασιών της έρευνας.....	36
Εικόνα 11: Προεπεξεργασία δεδομένων για την αντιμετώπιση της ανισορροπίας των δεδομένων.....	38
Εικόνα 12: Η ροή των διαδικασιών της έρευνας.....	44
Εικόνα 13 : Γράφημα loss function	45
Εικόνα 14: Γράφημα accuracy.....	45
Εικόνα 15: AUC του μοντέλου RNN	46
Εικόνα 16: Γράφημα loss function	47
Εικόνα 17: Γράφημα accuracy.....	47
Εικόνα 18: AUC του μοντέλου RNN	47
Εικόνα 19: Το γενικό πλαίσιο της έρευνας.....	49
Εικόνα 20: Δημιουργία του sentiment lexicon	50
Εικόνα 21: Deep Neural Network (DNN)	50
Εικόνα 22: Multi-Head Attention-based Deep Neural Network (MA-DNN)	51
Εικόνα 23: Bidirectional Long Short-Term Memory (Bi-LSTM).....	51
Εικόνα 24: Ανάλυση συναισθήματος για την πρόβλεψη της οικονομικής δυσπραγίας	53
Εικόνα 25: Κατάταξη sentiment features για (α) T3, (β) T4, (γ) T5 και (δ) T6	56
Εικόνα 26: Η αρχιτεκτονική του δικτύου	59
Εικόνα 27: Μέσος όρος AUC (• αριστερός άξονας) και log score (□ δεξιός άξονας).....	60

Εικόνα 28: Μέσος όρος AUC των τεσσάρων μεγαλύτερων ελεγκτικών εταιρειών (● αριστερός άξονας) και log score (□ δεξιός άξονας).....	63
Εικόνα 29: Οι οικονομικοί δείκτες εκφρασμένοι σε εικόνες χρώματος κλίμακας γκρι	65
Εικόνα 30: Τα αποτελέσματα του μέτρου F-measure	68
Εικόνα 31: Τα αποτελέσματα πρόβλεψης των μοντέλων CART, LDA, SVM, MLP, AdaBoost και της προτεινόμενης μεθόδου.....	70
Εικόνα 32: Η καμπύλη ROC	71
Εικόνα 33: Η ροή των διαδικασιών της έρευνας.....	72
Εικόνα 34: Σύγκριση απόδοσης σε διαφορετικό χρονικό ορίζοντα	76
Εικόνα 35: Σύγκριση των καμπυλών ROC κάθε μοντέλου.....	78
Εικόνα 36: Σχηματική απεικόνιση της ερευνητικής διαδικασίας.....	82
Εικόνα 37: Η δομή του Stacked auto-encoder.....	83
Εικόνα 38: Η επίδραση του αριθμού των hidden layers και των hidden units του SAE.....	84
Εικόνα 39: Η πειραματική διαδικασία.....	89
Εικόνα 40: Σύγκριση των επιπτώσεων των μεταβλητών εισόδου στα μοντέλα LSTM-RNN ...	91
Εικόνα 41: Σύγκριση των αποτελεσμάτων ακρίβειας πρόβλεψης.....	93
Εικόνα 42: Δομή NN (14-940-Tanh-Tanh-2-1).....	95
Εικόνα 43: Confusion matrix του training dataset.....	96
Εικόνα 44: Confusion matrix του testing dataset	96
Εικόνα 45: Γραφική αναπαράσταση του GrNet	99
Εικόνα 46: Training objective trend και Validation mis-classification error με διάσταση subspace 9.....	100
Εικόνα 47: Training objective trend και Validation mis-classification error με διάσταση subspace 30.....	101
Εικόνα 48: Αποτελέσματα μετρήσεων απόδοσης του dataset των Ισπανικών εταιρειών	105
Εικόνα 49: Αποτελέσματα μετρήσεων απόδοσης του dataset των Ταϊβανέζικων εταιρειών...	106
Εικόνα 50: Αποτελέσματα μετρήσεων απόδοσης του dataset των Πολωνικών εταιρειών	106
Εικόνα 51: Οι ερευνητικές διαδικασίες.....	110
Εικόνα 52: Ο σχεδιασμός της έρευνας	113
Εικόνα 53: Loss curve του μοντέλου LSTM.....	116
Εικόνα 54: Accuracy curve του μοντέλου LSTM	116
Εικόνα 55: Loss curve του μοντέλου GRU	117

Εικόνα 56: Accuracy curve του μοντέλου GRU.....	117
Εικόνα 57: Τα μοντέλα RNN, LSTM και GRU	120
Εικόνα 58: Το μοντέλο LSTM.....	122
Εικόνα 59: Το μοντέλο CNN.....	124
Εικόνα 60: Το μοντέλο MLP.....	125
Εικόνα 61: Confusion matrix.....	144

Ευρετήριο Πινάκων

Πίνακας 1: Τα δεκαπέντε ερευνητικά άρθρα.....	29
Πίνακας 2: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση διαφορετικού συνδυασμού δεδομένων.....	34
Πίνακας 3: Τομείς δραστηριότητας εισηγμένων κινέζικων εταιρειών	36
Πίνακας 4: Αριθμός δόλιων και μη δόλιων ετήσιων εκθέσεων εισηγμένων εταιρειών	37
Πίνακας 5: Οι χρηματοοικονομικές μεταβλητές	38
Πίνακας 6: Οι μη χρηματοοικονομικές μεταβλητές.....	39
Πίνακας 7: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση όλων των οικονομικών μεταβλητών.....	40
Πίνακας 8: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση των επιλεγμένων χρηματοοικονομικών μεταβλητών	40
Πίνακας 9: Αποτελέσματα απόδοσης μοντέλων με τη χρήση δεδομένων κειμένου 600 λέξεων	41
Πίνακας 10: Αποτελέσματα απόδοσης μοντέλων με τη χρήση δεδομένων κειμένου 1000 λέξεων.....	41
Πίνακας 11: Αποτελέσματα απόδοσης μοντέλων με τον συνδυασμό χρηματοοικονομικών δεδομένων και δεδομένων κειμένου 600 λέξεων	42
Πίνακας 12: Αποτελέσματα απόδοσης μοντέλων με τον συνδυασμό χρηματοοικονομικών δεδομένων και δεδομένων κειμένου 1000 λέξεων	42
Πίνακας 13: Κλάδοι δραστηριότητας των εταιρειών	43
Πίνακας 14: Οι μεταβλητές της έρευνας	44
Πίνακας 15: Accuracy μοντέλου RNN.....	45
Πίνακας 16: Accuracy μοντέλου LSTM.....	46
Πίνακας 17: Confusion matrix των μοντέλων RNN και LSTM.....	48
Πίνακας 18: Οι παράμετροι των μοντέλων deep learning.....	51
Πίνακας 19: Τα αποτελέσματα απόδοσης των μοντέλων deep learning	52
Πίνακας 20: Παράδειγμα sentiment words του λεξικού CFDSL	52
Πίνακας 21: Τα sentiment features της έρευνας.....	53
Πίνακας 22: Οι παράμετροι των μοντέλων	54
Πίνακας 23: Τα αποτελέσματα απόδοσης των μοντέλων.....	54
Πίνακας 24: Τα σημαντικά χαρακτηριστικά κάθε χρονικού διαστήματος T3,T4, T5 και T6	55
Πίνακας 25: Οι μεταβλητές της έρευνας	58

Πίνακας 26: (a) μέσος όρος AUC, (b) μέσος όρος log score	61
Πίνακας 27: (a) μέσος όρος AUC, (b) μέσος όρος log score των μεγάλων εταιρειών	62
Πίνακας 28: (a) μέσος όρος AUC, (b) μέσος όρος log score με βάση το έτος.....	62
Πίνακας 29: Τα αποτελέσματα του προτεινόμενου μοντέλου με πέντε διαφορετικές δομές δικτύου.....	67
Πίνακας 30: Τα αποτελέσματα του προτεινόμενου μοντέλου με 5640 και 3760 εικόνες συνεχιζόμενων εταιρειών.....	68
Πίνακας 31: Τα αποτελέσματα του προτεινόμενου μοντέλου με 600 εικόνες συνεχιζόμενων εταιρειών.....	69
Πίνακας 32: Η κατανομή των επιχειρήσεων ανά έτος.....	72
Πίνακας 33: Οι αριθμητικές μεταβλητές.....	73
Πίνακας 34: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση μόνο δεδομένων κειμένου	74
Πίνακας 35: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση μόνο αριθμητικών δεδομένων	75
Πίνακας 36: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση αριθμητικών δεδομένων και δεδομένων κειμένου	77
Πίνακας 37: Κλάδοι δραστηριότητας χρεοκοπημένων εταιρειών.....	79
Πίνακας 38: Οι 21 οικονομικές μεταβλητές της έρευνας.....	80
Πίνακας 39: Τα αποτελέσματα απόδοσης του μέτρου AUC.....	81
Πίνακας 40: Τα σύνολα δεδομένων.....	82
Πίνακας 41: Τα αποτελέσματα απόδοσης του μέτρου AUC.....	85
Πίνακας 42: Αποτελέσματα paired t-test ($\alpha=0.05$)	85
Πίνακας 43: Τα αποτελέσματα του training time των μοντέλων (milliseconds).....	86
Πίνακας 44: Σύγκριση με προηγούμενες μελέτες πρόβλεψης χρεοκοπίας για τα Πολωνικά σύνολα δεδομένων αναφοράς.....	86
Πίνακας 45: Οι μεταβλητές της έρευνας	88
Πίνακας 46: Σειρά κατάταξης των τιμών Shapley και των σχετικών τιμών Shapley όλων των μεταβλητών των 3 μοντέλων	90
Πίνακας 47: Η ακρίβεια πρόβλεψης των 3 μοντέλων	92
Πίνακας 48: Σύγκριση μοντέλων με τη χρήση των 18 διαφορετικών ομάδων μεταβλητών εισόδου	92
Πίνακας 49: Στατιστικά χαρακτηριστικά των datasets.....	94

Πίνακας 50: Τα αποτελέσματα απόδοσης των πέντε καλύτερων NN	95
Πίνακας 51: Λογιστικά δεδομένα	98
Πίνακας 52: Τα δεδομένα της χρηματιστηριακής αγοράς	98
Πίνακας 53: Το σύνολο δεδομένων	99
Πίνακας 54: Τα αποτελέσματα απόδοσης των μοντέλων	101
Πίνακας 55: Οι μεταβλητές των Ισπανικών εταιρειών	103
Πίνακας 56: Κατανομή δεδομένων πριν και μετά την εφαρμογή των τεχνικών εξισορρόπησης	104
Πίνακας 57: Οι υπερπαράμετροι των μοντέλων DL	104
Πίνακας 58: Οι υπερπαράμετροι των μοντέλων ensemble	105
Πίνακας 59: Σύγκριση απόδοσης των καλύτερων συνδυασμών μοντέλων και τεχνικών εξισορρόπησης	107
Πίνακας 60: Οι εταιρείες ανά τομέα δραστηριότητας	108
Πίνακας 61: Οι μεταβλητές της έρευνας	109
Πίνακας 62: Οι δέκα σημαντικότερες μεταβλητές που επιλέχθηκαν με τη μέθοδο CART	110
Πίνακας 63: Τα αποτελέσματα απόδοσης των μοντέλων	112
Πίνακας 64: Το confusion matrix των μοντέλων	112
Πίνακας 65: Οι επιχειρήσεις ανά κλάδο δραστηριότητας	114
Πίνακας 66: Οι μεταβλητές της έρευνας	115
Πίνακας 67: Σύγκριση αποτελεσμάτων των μοντέλων LSTM και GRU	117
Πίνακας 68: Τύπος δεδομένων - Βάσεις δεδομένων - Χρονικό διάστημα συλλογής	132
Πίνακας 69: Μέθοδοι εξισορρόπησης δεδομένων	134
Πίνακας 70: Τεχνική επεξεργασίας φυσικής γλώσσας	136
Πίνακας 71: Μέθοδοι επιλογής χαρακτηριστικών	137
Πίνακας 72: Μέθοδοι επικύρωσης	139
Πίνακας 73: Οι 100 σημαντικότερες λέξεις της ενότητας MD&A μετά την εφαρμογή της μεθόδου representation erasure. Οι λέξεις με έντονη γραφή υπάρχουν και στο λεξικό Loughran and McDonald (2011) ενώ οι υπογραμμισμένες στο λεξικό MPQA Subjectivity (Wilson, Wiebe & Hofmann, 2005)	140
Πίνακας 74: Τα μοντέλα αναφοράς	142
Πίνακας 75: Τα μέτρα απόδοσης των ερευνητικών άρθρων	147
Πίνακας 76: Μέθοδοι στατιστικής ανάλυσης	148

1 *Εισαγωγή*

Στις μέρες μας η σημασία της πρόβλεψης της μελλοντικής οικονομικής κατάστασης των επιχειρήσεων αποκτά όλο και μεγαλύτερη σημασία, και είναι χρήσιμη τόσο για τις ίδιες τις επιχειρήσεις όσο και για τα ενδιαφερόμενα μέρη της αγοράς. Για τις επιχειρήσεις είναι σημαντική, διότι τις βοηθά στην έγκαιρη λήψη αποφάσεων και την ανάπτυξη κατάλληλων στρατηγικών, ώστε να προσαρμόζονται στο ολοένα μεταβαλλόμενο οικονομικό και επιχειρηματικό περιβάλλον, με στόχο την επίτευξη κερδοφορίας, την αύξηση της ανταγωνιστικότητας, τη διατήρηση της βιωσιμότητάς τους και την ανταπόκρισή τους στις προσδοκίες των ενδιαφερόμενων μερών (Παπαδάκης, 2016). Για τα ενδιαφερόμενα μέρη της αγοράς, όπως επενδυτές, πιστωτές, προμηθευτές, εργαζόμενοι, ρυθμιστικές αρχές κ.α., είναι επίσης σημαντική, διότι τους βοηθά στη λήψη αποφάσεων σχετικά με τις δραστηριότητες τους με στόχο την κάλυψη των απαιτήσεων και των προσδοκιών τους και την αποφυγή εξαπάτησης (Τσακλάγκανος & Σπαθής, 2017).

Κύρια πηγή πληροφόρησης για την οικονομική θέση και απόδοση μιας επιχείρησης είναι οι ετήσιες χρηματοοικονομικές καταστάσεις και είναι χρήσιμες για τη λήψη αποφάσεων όλων των ενδιαφερόμενων μερών (Τσακλάγκανος & Σπαθής, 2017).

Όμως η ασυμμετρία της πληροφόρησης προκαλεί κινδύνους, καθώς η διοίκηση κάθε επιχείρησης έχει στη διάθεση της όλες τις πληροφορίες για την πραγματική οικονομική της κατάσταση ενώ οι υπόλοιποι όχι (Jan, 2021a). Είναι πιθανό λοιπόν η διοίκηση να αποκρύπτει σημαντικές πληροφορίες για διάφορους λόγους (όπως κακή οικονομική απόδοση, πίεση να ανταποκριθεί στις απαιτήσεις τρίτων, ανάγκη απόκτησης χρηματοδότησης, φορολογικοί λόγοι κ.α.) δεδομένου ότι η διασφάλιση μιας θετικής οικονομικής εικόνας είναι ζωτικής σημασίας (Craja, Kim, & Lessmann, 2020).

Μάλιστα μέχρι σήμερα έχει διαπιστωθεί ότι πολλές από τις ετήσιες χρηματοοικονομικές καταστάσεις των επιχειρήσεων δεν είναι αληθινές αλλά παραποιημένες (Craja κ.ά., 2020; Xiuguo & Shengyong, 2022) με σκοπό την παραπλάνηση και εξαπάτηση των ενδιαφερόμενων μερών προς όφελος των επιχειρήσεων. Θεωρείται ως ένα πολύ σοβαρό φαινόμενο δεδομένου ότι συχνά δεν εμπλέκονται μόνο τα διοικητικά στελέχη αλλά και οι ελεγκτικές εταιρείες. Σύμφωνα με έρευνες διαπιστώθηκε ότι η απάτη των χρηματοοικονομικών καταστάσεων προκαλεί υψηλότερες οικονομικές ζημιές σε σύγκριση με άλλες μορφές απάτης (π.χ. κατάχρηση περιουσιακών στοιχείων και διαφθορά), παρόλο που η συχνότητα εμφάνισής της είναι συγκριτικά μικρότερη (Craja κ.ά., 2020).

Χαρακτηριστικά παραδείγματα απάτης αποτελούν τα οικονομικά σκάνδαλα της ενεργειακής εταιρείας Enron και της εταιρείας τηλεπικοινωνιών WorldCom που ξέσπασαν το 2001 και το 2002 αντίστοιχα στις Ηνωμένες Πολιτείες. Τα εν λόγω σκάνδαλα οδήγησαν σε κρίσεις προκαλώντας σοβαρές οικονομικές απώλειες σε όλα τα ενδιαφερόμενα μέρη. Μάλιστα στάθηκαν η αφορμή να αποκαλυφθούν οι ελλείψεις στους λογιστικούς κανονισμούς και τους κανονισμούς των χρηματιστηρίων. Αποτέλεσμα των παραπάνω ήταν να ψηφιστεί το 2002 ο νόμος Sarbanes-Oxley (SOX) από το Κογκρέσο των Η.Π.Α. Ο νόμος επιβλήθηκε υποχρεωτικά σε όλες τις εισηγμένες επιχειρήσεις με σκοπό τον περιορισμό της πιθανότητας απάτης, την ενίσχυση της διαφάνειας και της ανεξαρτησίας των ελεγκτικών μηχανισμών, την προστασία του επενδυτικού κοινού και της λειτουργίας των κεφαλαιαγορών (Jan, 2021a). Πλέον οι εισηγμένες επιχειρήσεις έχουν την υποχρέωση να δημοσιεύουν τις ετήσιες οικονομικές εκθέσεις για την ενημέρωση κάθε ενδιαφερόμενου μέρους αναφορικά με την οικονομική της κατάσταση, τις ενδεχόμενες απειλές, τις μελλοντικές της προοπτικές και την έκφραση της ανεξάρτητης και ελεύθερης γνώμης του ελεγκτή (Xiuguo & Shengyong, 2022; Τσακλάγκανος & Σπαθής, 2017).

Ένα ακόμη χαρακτηριστικό παράδειγμα αποτελεί η χρηματοπιστωτική κρίση του 2008-2009 που είχε ως αποτέλεσμα να χρεοκοπήσουν πολλές επιχειρήσεις πλήττοντας σοβαρά την οικονομία παγκοσμίως. Πολλοί είναι οι μελετητές που πιστεύουν ότι ένας από τους κύριους λόγους αυτής της κρίσης ήταν η αποτυχία των ορκωτών λογιστών και των ελεγκτών να αξιολογήσουν σωστά τη συνεχιζόμενη δραστηριότητα και τη βιωσιμότητα των επιχειρήσεων (Chi & Chu, 2021; Jan, 2021b). Τα τελευταία χρόνια λοιπόν δίνεται όλο και μεγαλύτερη βαρύτητα στην ορθή αξιολόγηση των επιχειρήσεων από τους ορκωτούς λογιστές (Certified Public Accountants - CPA) και τους ελεγκτές προκειμένου να προσδιορίζουν επιτυχώς ενδεχόμενες αμφιβολίες σχετικά με τη συνεχιζόμενη δραστηριότητα μιας επιχείρησης, δηλαδή τη δυνατότητα συνέχισης των δραστηριοτήτων της για τουλάχιστον 12 μήνες από την ημερομηνία του ισολογισμού (Chi & Chu, 2021; Jan, 2021b).

Ως επακόλουθο των παραπάνω, προέκυψε η ανάγκη πρόβλεψης της μελλοντικής οικονομικής κατάστασης των επιχειρήσεων. Παραδείγματα προβλέψεων της μελλοντικής οικονομικής κατάστασης μιας επιχείρησης αποτελούν η πρόβλεψη για την πιθανότητα χρεοκοπίας, την πιθανότητα οικονομικής δυσπραγίας και αποτυχίας, την πιθανότητα μη συνεχιζόμενης δραστηριότητας αλλά και την πιθανότητα απάτης των χρηματοοικονομικών καταστάσεων δεδομένου ότι σκοπός της είναι η απόκρυψη σημαντικών πληροφοριών. Μάλιστα θεωρείται ότι η απάτη των οικονομικών καταστάσεων είναι πρόδρομος της οικονομικής δυσπραγίας και της χρεοκοπίας μιας επιχείρησης (Jan, 2021a). Η πρόβλεψη των παραπάνω είναι ιδιαίτερα πολύπλοκη διαδικασία, λόγω της ύπαρξης πολυάριθμων παραμέτρων και αιτιών που τις προκαλούν και είναι δύσκολο να εντοπισθούν (Alexandropoulos, Aridas, Kotsiantis, & Vrahatis, 2019; Craja κ.ά., 2020; Jang, Jeong, & Cho, 2021).

Τα παραπάνω οδήγησαν τους ερευνητές να ασχοληθούν με την ανάπτυξη αυτοματοποιημένων συστημάτων για την παροχή έγκαιρων και ερμηνεύσιμων προβλέψεων με απώτερο στόχο την ενίσχυση της λήψης αποφάσεων των υπευθύνων, την επιτάχυνση και βελτίωση των διαδικασιών ελέγχου και την αποτελεσματικότερη διοίκηση, προκειμένου να διασφαλιστεί περαιτέρω η σταθερότητα των επιχειρηματικών αγορών και η οικονομική ανάπτυξη (Alexandropoulos κ.ά., 2019; Craja κ.ά., 2020).

1.1 Αντικείμενο και σκοπός της διπλωματικής

Αντικείμενο της παρούσας εργασίας είναι η βιβλιογραφική ανασκόπηση ερευνών που έχουν διεξαχθεί τα τελευταία χρόνια για την πρόβλεψη της πιθανότητας απάτης, οικονομικής δυσπραγίας, χρεοκοπίας, οικονομικής αποτυχίας και μη συνεχιζόμενης δραστηριότητας των επιχειρήσεων. Σκοπός είναι η διερεύνηση της αποτελεσματικότητας και της αξιοπιστίας των προτεινόμενων μοντέλων deep learning σε σύγκριση με τα μοντέλα αναφοράς, ο εντοπισμός του οφέλους χρήσης της μεθόδου deep learning στην πρόβλεψη καθώς και των ευκαιριών που προσφέρει για περαιτέρω έρευνα.

Συγκεκριμένα, παρουσιάζονται δεκαπέντε ερευνητικά άρθρα που ασχολήθηκαν με την ανάπτυξη καινοτόμων μοντέλων πρόβλεψης deep learning και τα οποία, στα πλαίσια των ερευνητικών διαδικασιών, συγκρίθηκαν με παραδοσιακά στατιστικά μοντέλα και μοντέλα μηχανικής μάθησης προκειμένου να αξιολογηθεί η απόδοση και να εξακριβωθεί η αποτελεσματικότητά τους.

Τα δεκαπέντε ερευνητικά άρθρα συλλέχθηκαν από τις βάσεις δεδομένων Scopus, ResearchGate, Google Scholar και Sci-Hub και δημοσιεύθηκαν από τους εκδοτικούς οίκους Elsevier, Springer, MDPI και Emerald από το 2019 ως το 2022. Αναζητήθηκαν με τις λέξεις κλειδιά: “deep learning”, “bankruptcy”, “financial fraud”, “financial distress”, “annual reports”, “corporate failure”, “going concern” και επιλέχθηκαν μόνο όσα ήταν γραμμένα στην αγγλική γλώσσα.

1.2 Διάρθρωση κειμένου

Το υπόλοιπο της εργασίας διαρθρώνεται ως εξής : Στο **κεφάλαιο 2** γίνεται αναφορά στα παραδοσιακά στατιστικά μοντέλα και τα μοντέλα machine learning που εφαρμόστηκαν σε προηγούμενες μελέτες. Στο **κεφάλαιο 3** εισάγονται ορισμένες βασικές έννοιες που αφορούν την τεχνητή νοημοσύνη και τη μέθοδο deep learning, απαραίτητες για την κατανόηση των επόμενων κεφαλαίων της εργασίας. Στα **κεφάλαια 4** και **5** γίνεται περιγραφή και ανάλυση των προτεινόμενων μοντέλων deep learning των δεκαπέντε ερευνητικών άρθρων που αποτελούν και το αντικείμενο της εργασίας αυτής και τέλος, στο **κεφάλαιο 6** παρουσιάζονται συνοπτικά τα συμπεράσματα της βιβλιογραφικής ανασκόπησης και δίνονται ορισμένες χρήσιμες υποδείξεις για τον προσανατολισμό των μελλοντικών ερευνών σε θέματα που χρήζουν διερεύνηση για την περαιτέρω βελτίωση της μεθόδου deep learning.

2

Προηγούμενες Έρευνες

Μέχρι σήμερα πραγματοποιήθηκαν έρευνες για την πρόβλεψη της απάτης των οικονομικών καταστάσεων, για την πρόβλεψη της οικονομικής δυσπραγίας, χρεοκοπίας, οικονομικής αποτυχίας και μη συνεχιζόμενης δραστηριότητας των επιχειρήσεων, οι περισσότερες εκ των οποίων ανέπτυξαν και εφάρμοσαν παραδοσιακές στατιστικές μεθόδους ανάλυσης ή παραδοσιακές μεθόδους machine learning. Όμως κατά την εφαρμογή τους διαπιστώθηκαν αρκετά μειονεκτήματα που οφείλονταν κυρίως στην αδυναμία (α) ανάπτυξης πολύπλοκων χρηματοοικονομικών μοντέλων, (β) διαχείρισης μη ισορροπημένων δεδομένων και (γ) διαχείρισης big data (Smiti & Soui, 2020). Επίσης τα περισσότερα μοντέλα περιορίστηκαν στην ανάλυση των αριθμητικών δεδομένων των οικονομικών καταστάσεων, δηλαδή μόνο των δομημένων δεδομένων, κάτι που δεν επαρκούσε (Xinguo & Shengyong, 2022). Έτσι η ανάγκη χρήσης μιας πιο εξελιγμένης τεχνολογίας και μεθόδου ανάλυσης ήταν επιτακτική.

Ακολουθεί συνοπτική περιγραφή των μεθόδων πρόβλεψης που έχουν χρησιμοποιηθεί μέχρι σήμερα.

2.1 Παραδοσιακές στατιστικές μέθοδοι

Οι περισσότερες έρευνες χρησιμοποίησαν στατιστικές μεθόδους ανάλυσης, όπως τις μεθόδους regression analysis, discriminant analysis, cluster analysis, and factor analysis (Chi & Chu, 2021). Χαρακτηριστικά παραδείγματα είναι η ανάπτυξη της μεθόδου Multivariate Discriminant Analysis (MDA) από τον Altman το 1968 και αργότερα της μεθόδου Logistic Regression (LR) από τον Ohlson το 1980 με σκοπό την πρόβλεψη της οικονομικής κατάστασης των επιχειρήσεων (Alam, Gao, & Jones, 2021; Aljawazneh, Mora, Garcia-Sanchez, & Castillo-Valdivieso, 2021; Hosaka, 2019; Smiti & Soui, 2020). Όμως τα στατιστικά μοντέλα ήταν απλοϊκά και οι περιορισμοί και οι ελλείψεις τους οδηγούσαν σε λάθη (Chi & Chu, 2021; Jan, 2021b). Ανέλυαν αριθμητικούς οικονομικούς δείκτες οι οποίοι εισέρχονταν στο μοντέλο σε γραμμικό συνδυασμό. Βασίζονταν σε εξαιρετικά περιοριστικές στατιστικές υποθέσεις και δεν είχαν τη δυνατότητα χειρισμού μεγάλου αριθμού μεταβλητών κάνοντας δύσκολη την ανάπτυξη πολύπλοκων χρηματοοικονομικών μοντέλων (Alam κ.ά., 2021; Smiti & Soui, 2020).

2.2 Παραδοσιακές μέθοδοι *Machine Learning*

Αργότερα χρησιμοποιήθηκαν τεχνικές μηχανικής μάθησης και αναπτύχθηκαν μοντέλα πρόβλεψης, όπως τα μοντέλα Decision Tree (DT), Neural Networks (NN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Bayesian Network (BN), Artificial Neural Network (ANN) με ικανοποιητικά ποσοστά ακρίβειας (Kirkos, Spathis, & Manolopoulos, 2007; Xiuguo & Shengyong, 2022). Τα πλεονεκτήματα των μοντέλων machine learning έναντι των στατιστικών παραδοσιακών μοντέλων είναι ότι έχουν την ικανότητα χειρισμού μεγάλου αριθμού μεταβλητών (Alam κ.ά., 2021), δεν βασίζονται στις περιοριστικές υποθέσεις των στατιστικών μεθόδων (Jang κ.ά., 2021), χειρίζονται αποτελεσματικά μη γραμμικά προβλήματα και έχουν μεγαλύτερη ακρίβεια ταξινόμησης και πρόβλεψης (Jan, 2021a). Ωστόσο εμφανίζουν αρκετά μειονεκτήματα όσον αφορά την εξαγωγή χαρακτηριστικών, διότι πραγματοποιείται ανεξάρτητα από τη φάση εκπαίδευσης. Επίσης η διαδικασία εκπαίδευσης είναι αρκετά χρονοβόρα, ο χειρισμός των big data δύσκολος (Smiti & Soui, 2020), ενώ οι περισσότερες παραδοσιακές προσεγγίσεις μηχανικής μάθησης, όπως gradient boosting machines και random forests, δεν έχουν σχεδιαστεί να χειρίζονται δομές δεδομένων πίνακα (Alam κ.ά., 2021).

2.3 Μέθοδος *Deep Learning*

Τα τελευταία χρόνια η πρόοδος της τεχνολογίας επέτρεψε στους ερευνητές να στραφούν στη μέθοδο deep learning. Πρόκειται για μια προηγμένη προσέγγιση της μηχανικής μάθησης που ήδη έχει δείξει την αποτελεσματικότητά της σε πολλούς τομείς, όπως τη ρομποτική, την ανάλυση κειμένου, την επεξεργασία φυσικής γλώσσας, την αναγνώριση ομιλίας, την ιατρική κ.α. (Alam κ.ά., 2021; Smiti & Soui, 2020). Είναι ικανή να διαχειρίζεται με επιτυχία τα big data και να δημιουργεί αυτόματα χρήσιμα χαρακτηριστικά κατά τη διάρκεια της εκπαίδευσης επιτυγχάνοντας υψηλότερη ακρίβεια πρόβλεψης και χαμηλότερα ποσοστά λάθους σε σύγκριση με τις παραδοσιακές μεθόδους machine learning (Smiti & Soui, 2020). Έχει την ικανότητα να διορθώνει ανακριβείς προβλέψεις από μόνη της σε αντίθεση με τις παραδοσιακές μεθόδους μηχανικής μάθησης που απαιτείται η παρέμβαση του ερευνητή. Μάλιστα τα τελευταία χρόνια γίνονται προσπάθειες ανάπτυξης μοντέλων πρόβλεψης deep learning που επικεντρώνονται στην εξαγωγή πληροφοριών και από αδόμητα δεδομένα, όπως το κείμενο από την ενότητα των δηλώσεων του ΔΣ των ετήσιων οικονομικών εκθέσεων (Li, Shi, Wang, & Zhou, 2021; Mai, Tian, Lee, & Ma, 2019; Sun & Vasarhelyi, 2018), δίνοντας χρήσιμες πληροφορίες σε όλα τα ενδιαφερόμενα μέρη της αγοράς και ταυτόχρονα συμβάλλοντας στη βελτίωση της ποιότητας και της χρονικής διάρκειας των ελεγκτικών διαδικασιών. Ωστόσο η χρήση της στη χρηματοοικονομική ανάλυση εξακολουθεί να είναι περιορισμένη.

3

Εισαγωγικές έννοιες

Στο κεφάλαιο αυτό παραθέτονται ορισμένες βασικές έννοιες χρήσιμες για την κατανόηση των επόμενων κεφαλαίων.

3.1 Τεχνητή νοημοσύνη (*Artificial Intelligence, AI*)

Η Τεχνητή νοημοσύνη είναι ένας κλάδος της πληροφορικής με πολλές πρακτικές εφαρμογές και ενεργές ερευνητικές περιοχές, όπως τη μηχανική μάθηση, την επεξεργασία και κατανόηση φυσικής γλώσσας, τη ρομποτική, τη μηχανική όραση και ομιλία κ.α. Στην ουσία καλύπτει πολλά επιστημονικά πεδία, όπως, ενδεικτικά, τα μαθηματικά, τα οικονομικά, τη φιλοσοφία, την ψυχολογία, τη μηχανική, τη νευροεπιστήμη, την ιατρική (Κεραυνού, 2000).

Η Τεχνητή νοημοσύνη μελετά την ανθρώπινη ευφυή συμπεριφορά και ασχολείται με την κατανόηση και αυτοματοποίησή της (Κεραυνού, 2000). Αποτελεί ένα σύνολο ενεργειών που εκτελούνται από υπολογιστές και μιμούνται την ανθρώπινη συμπεριφορά. Ασχολείται με τη σχεδίαση αλγορίθμων για τον υπολογισμό ικανοποιητικών λύσεων σε προβλήματα (Κεραυνού, 2000).

Αρχικά, ασχολήθηκε με την επίλυση προβλημάτων που είναι διανοητικά δύσκολα για τον άνθρωπο με επιτυχία (Goodfellow, Bengio, & Courville, 2016). Η επιτυχία οφείλεται στο γεγονός ότι τα διανοητικά προβλήματα περιγράφονται εύκολα με μαθηματικούς κανόνες οι οποίοι είναι κατανοητοί στους υπολογιστές. Όμως η αληθινή πρόκληση ήταν η επίλυση των διαισθητικών προβλημάτων. Προβλημάτων, δηλαδή, που επιλύονται εύκολα από τον άνθρωπο αλλά δύσκολα από τον υπολογιστή (Goodfellow κ.ά., 2016). Ο άνθρωπος χαρακτηρίζεται από ευφυΐα, εμπειρία, γνώση και ικανότητες, όπως την ικανότητα όρασης, μάθησης, ομιλίας, λήψης αποφάσεων, επιχειρηματολογίας κ.α. (Κεραυνού, 2000). Ένα μεγάλο μέρος της γνώσης του δηλαδή είναι υποκειμενικό και διαισθητικό (Goodfellow κ.ά., 2016).

Ένας υπολογιστής όμως δεν διαθέτει τα παραπάνω χαρακτηριστικά αλλά κατανοεί μόνο μαθηματικούς κανόνες. Τα διαισθητικά προβλήματα είναι δύσκολο να περιγραφούν με κανόνες κατανοητούς για τους υπολογιστές (Goodfellow κ.ά., 2016), διότι η δομή των υπολογιστών είναι διαφορετική από τη δομή του βιολογικού εγκεφάλου (Αργυράκης, 2001). Έτσι μια από τις βασικές προκλήσεις της τεχνητής

νοημοσύνης ήταν πως οι υπολογιστές θα αποκτήσουν αυτή τη γνώση, ώστε να συμπεριφέρονται έξυπνα (Goodfellow κ.ά., 2016).

3.1.1 Κωδικοποίηση γνώσης και αναπαράσταση δεδομένων

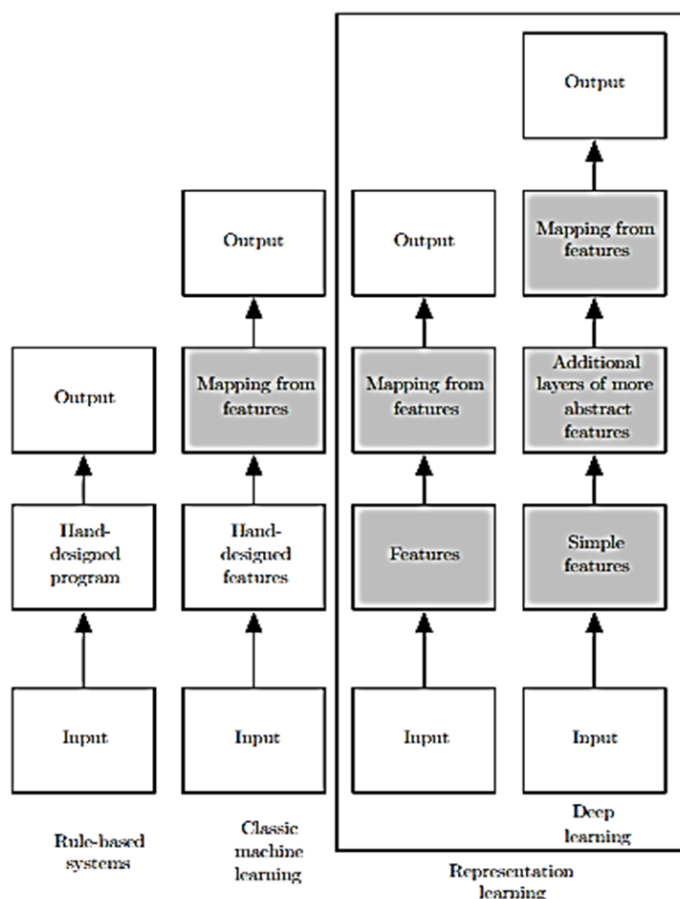
Αρχικά, έγινε προσπάθεια κωδικοποίησης της γνώσης σε κατάλληλες γλώσσες με τη μορφή περίπλοκων κανόνων που εισάγονταν από τον άνθρωπο σε μια μηχανή συμπερασμάτων με σκοπό να συνάγει λογικά συμπεράσματα με τη χρήση των κανόνων αυτών, όμως χωρίς επιτυχία (Goodfellow κ.ά., 2016).

Λύση στο πρόβλημα έδωσε η μηχανική μάθηση όπου με τη χρήση αλγορίθμων μηχανικής μάθησης, οι υπολογιστές μπορούσαν να αναλύουν δεδομένα, να μαθαίνουν και να λαμβάνουν αποφάσεις που φαίνονταν υποκειμενικές (Goodfellow κ.ά., 2016).

Βέβαια, η απόδοση των αλγορίθμων εξαρτάται από την αναπαράσταση των δεδομένων, δηλαδή εξαρτάται από τις πληροφορίες που περιλαμβάνονται στα δεδομένα και είναι γνωστές ως χαρακτηριστικά. Είναι σημαντικό λοιπόν να γίνει επιλογή του συνόλου των χαρακτηριστικών εκείνου που αντιπροσωπεύει καλύτερα τα δεδομένα. Όμως η επιλογή του κατάλληλου συνόλου χαρακτηριστικών με hand-designed representations είναι δύσκολη και χρονοβόρα διαδικασία (Goodfellow κ.ά., 2016).

Η ανάπτυξη μιας πιο εξελιγμένης τεχνικής machine learning που ονομάζεται representation learning (Goodfellow κ.ά., 2016) επέτρεψε τον αυτόματο εντοπισμό των κατάλληλων χαρακτηριστικών από ακατέργαστα δεδομένα σε σύντομο χρόνο. Όμως ο διαχωρισμός των χρήσιμων από τα μη χρήσιμα αφηρημένα χαρακτηριστικά (π.χ. προφορά ενός ομιλητή) εξακολουθούσε να παραμένει δύσκολος και χρειαζόταν πιο εξελιγμένη τεχνολογία. Λύση στο πρόβλημα του representation learning έδωσε το deep learning. Η εξελιγμένη αρχιτεκτονική του deep learning επέτρεψε στον υπολογιστή να δημιουργεί περίπλοκες και πιο αφηρημένες έννοιες από άλλες απλούστερες με επιτυχία (Goodfellow κ.ά., 2016).

Μια σχηματική αναπαράσταση των παραπάνω είναι η ακόλουθη (βλ. εικόνα 1):



Εικόνα 1: Σχηματική απεικόνιση διαγραμμάτων ροής αναπαράστασης δεδομένων

Πηγή: (Goodfellow κ.ά, 2016)

3.1.2 Αναπαράσταση κειμένου

Η αναπαράσταση κειμένου αφορά τη μετατροπή των λέξεων ενός κειμένου, δηλαδή των αδόμητων δεδομένων, σε αριθμητική μορφή, ώστε να είναι κατανοητές από τους υπολογιστές και να είναι δυνατή η αλγοριθμική τους επεξεργασία. Μέχρι σήμερα έχουν αναπτυχθεί πολλές τεχνικές, όπως ενδεικτικά οι τεχνικές Bag Of Words (BOW), Term Frequency – Inverse Document Frequency (TF-IDF), Bidirectional Encoder Representations from Transformers (BERT) και Word2Vec.

3.1.2.1 Bag of Words (BOW)

Η τεχνική Bag Of Words (BOW) χρησιμοποιείται συχνά σε εργασίες ταξινόμησης. Είναι μια απλή τεχνική αναπαράστασης κειμένου όπου τοποθετεί τις λέξεις ενός κειμένου σε μία <<τσάντα λέξεων>> και μετράει τη συχνότητα εμφάνισης κάθε λέξης μέσα στο κείμενο. Το κείμενο μπορεί να είναι μία πρόταση ή ένα έγγραφο. Ουσιαστικά, το βάρος της λέξης είναι η συχνότητα εμφάνισής της στο κείμενο. Το μήκος του διανύσματος είναι σταθερό και ίσο με το μήκος του λεξιλογίου που

δημιουργεί (Chandran, 2020). Τα βασικά μειονεκτήματα είναι ότι δεν λαμβάνει υπόψη τη γραμματική και τη σειρά των λέξεων, οπότε η κατανόηση της σημασίας τους δεν είναι δυνατή. Επίσης, μεταξύ των λέξεων που περιλαμβάνονται στο λεξιλόγιο, υπάρχουν λέξεις που δεν περιέχουν χρήσιμη πληροφορία (π.χ. οι λέξεις “είναι”, “το”, “να” κ.α.), ενώ εστιάζει αποκλειστικά και μόνο στην καταμέτρηση των λέξεων που περιλαμβάνονται στο λεξιλόγιο, αγνοώντας νέες λέξεις που ενδεχομένως να περιέχουν χρήσιμη πληροφορία (Wikipedia, Bag-of-words model, 2022; Chandran, 2020).

3.1.2.2 Term Frequency – Inverse Document Frequency (TF-IDF)

Η τεχνική TF-IDF είναι απλή στη χρήση της και περιλαμβάνει δύο διαδικασίες κατά τις οποίες υπολογίζονται οι όροι TF και IDF. Με τον όρο TF υπολογίζεται η συχνότητα εμφάνισης μιας λέξης σε ένα έγγραφο, ενώ με τον όρο IDF υπολογίζεται η σπανιότητα εμφάνισης της λέξης αυτής στο σύνολο των εγγράφων, δηλαδή σε ολόκληρο το σώμα (corpus) (Capitalone, 2021; Chandran, 2020). Ουσιαστικά το βάρος της λέξης εξαρτάται από τη συχνότητα εμφάνισής της σε ένα έγγραφο και τη σπανιότητα εμφάνισής της στο σώμα και υπολογίζεται πολλαπλασιάζοντας τους δύο όρους TF και IDF (Chandran, 2020):

$$TFIDF = TF(w, d) \times IDF(w)$$

Το πλεονέκτημα είναι ότι εντοπίζονται οι λέξεις υψηλής συχνότητας εμφάνισης που είναι σημαντικές ενώ εκτοπίζονται όσες είναι ασήμαντες αντιμετωπίζοντας έτσι τα προβλήματα θορύβου που δημιουργούν οι ασήμαντες λέξεις. Το βασικό της μειονέκτημα, όπως και στην τεχνική BOW, είναι ότι αγνοεί τη σειρά των λέξεων και δεν είναι δυνατή η κατανόηση της σημασίας των λέξεων (Chandran, 2020).

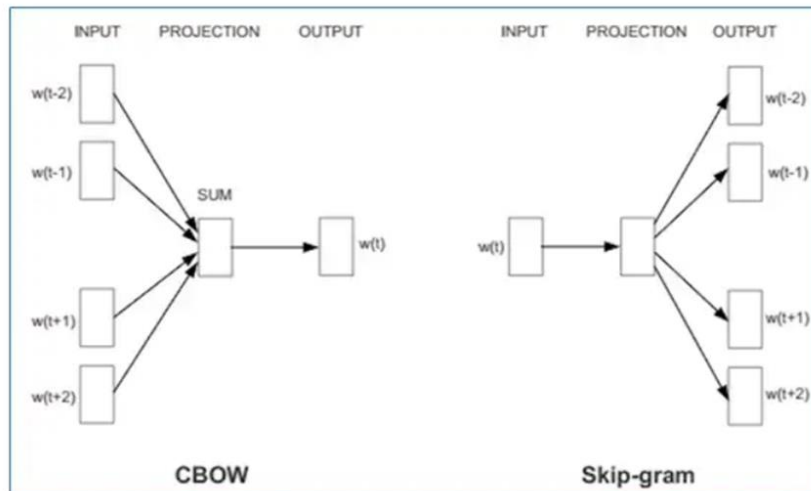
3.1.2.3 Word2Vec

Η τεχνική Word2Vec είναι μια δημοφιλής τεχνική επεξεργασίας φυσικής γλώσσας που αναπτύχθηκε από τον Tomas Mikolov το 2013 στη Google. Είναι ικανή να εντοπίζει σημασιολογικά και συντακτικά παρόμοιες λέξεις και να προτείνει πρόσθετες σε ελλειπείς προτάσεις (Wikipedia, Word2vec, 2022). Πρόκειται για ένα νευρωνικό δίκτυο δύο επιπέδων το οποίο έχει την ικανότητα να δημιουργεί πυκνές ενσωματώσεις λέξεων (Dutta, 2021). Δέχεται ως είσοδο ένα μεγάλο σώμα κειμένου, δημιουργεί ένα διανυσματικό χώρο μεγάλων διαστάσεων όπου κάθε μοναδική λέξη αντιστοιχεί σε ένα διάνυσμα (Wikipedia, Word2vec, 2022). Η ενσωμάτωση λέξεων επιτρέπει παρόμοιες λέξεις να βρίσκονται πιο κοντά στο διανυσματικό χώρο, υποδηλώνοντας έτσι τη σημασιολογική τους ομοιότητα (Dutta, 2021).

Για τη δημιουργία των ενσωματώσεων λέξεων, το Word2Vec χρησιμοποιεί δύο αρχιτεκτονικές, την αρχιτεκτονική Continuous Bag-Of-Words (CBOW) ή την αρχιτεκτονική continuous skip-gram κατά τις οποίες λαμβάνονται υπόψη τόσο οι μεμονωμένες λέξεις όσο και το παράθυρο λέξεων. Το παράθυρο λέξεων είναι ένα σύνολο λέξεων μέσα σε ένα παράθυρο που περιβάλλουν κάθε μεμονωμένη λέξη (Dutta, 2021).

Σύμφωνα με τη μέθοδο CBOW, το μοντέλο προβλέπει τη λέξη-στόχο (τη μεμονωμένη λέξη) λαμβάνοντας ως είσοδο το παράθυρο λέξεων, δηλαδή τις λέξεις που βρίσκονται γύρω από τη λέξη-στόχο. Ενώ με τη μέθοδο skip-gram, το μοντέλο

προβλέπει το παράθυρο λέξεων λαμβάνοντας υπόψη τη λέξη-στόχο (βλ. εικόνα 2) (Dutta, 2021).



Εικόνα 2: Σχηματική αναπαράσταση των αρχιτεκτονικών CBOW & Skip-gram

Πηγή: (Ali, 2019)

Συγκρίνοντας τις δύο μεθόδους, η μέθοδος CBOW είναι πιο γρήγορη και αποτελεσματικότερη στις συχνές λέξεις, ενώ η μέθοδος Skip-gram είναι καλύτερη σε μικρά σύνολα δεδομένων και αποτελεσματικότερη στις σπάνιες λέξεις (Karani, 2018). Η τεχνική Word2Vec είναι απλή στη χρήση της και κατανοητή. Εκπαιδεύεται γρήγορα και είναι κατάλληλη τόσο για μικρά όσο και για μεγάλα σύνολα δεδομένων. Μέχρι σήμερα έχει χρησιμοποιηθεί με επιτυχία για την ανάλυση ερευνητικών δεδομένων και κριτικών πελατών καθώς και σε συστήματα συστάσεων, όπως στις πλατφόρμες Airbnb, Alibaba and Spotify (Devopedia, 2020)

3.1.2.4 BERT

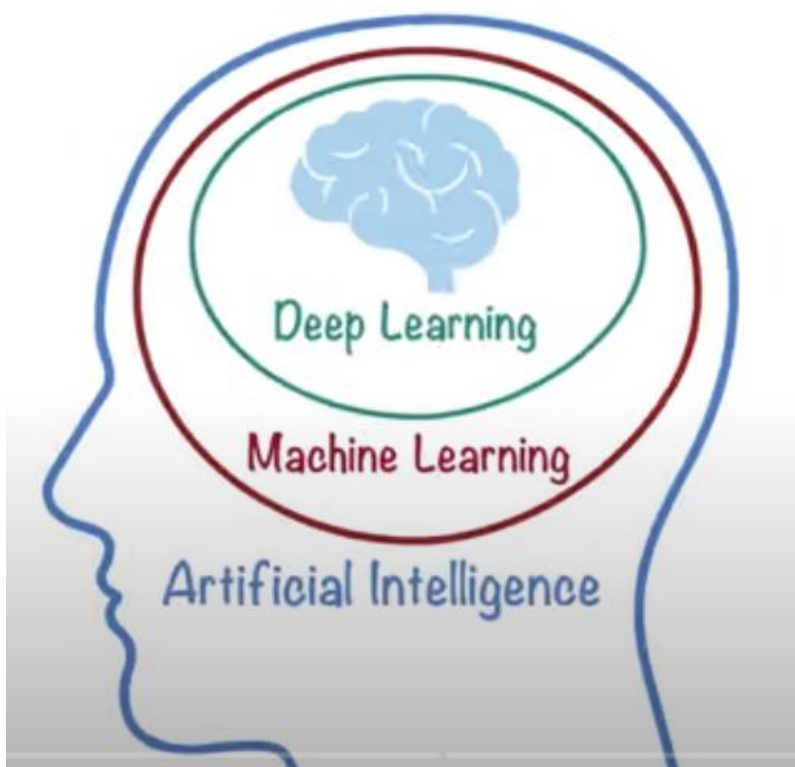
Η τεχνική Bidirectional Encoder Representations from Transformers (BERT) αναπτύχθηκε το 2019 από την Google και βασίζεται στα Transformers (Ankiit, χ.χ.; Wikipedia, BERT (language model), 2022). Είναι ένα μοντέλο νευρωνικού δικτύου που έχει την ικανότητα να μαθαίνει ενσωματώσεις λέξεων και να εντοπίζει λέξεις με βάση τα συμφραζόμενα. Η δυνατότητα ανάγνωσης του κειμένου και προς τις δύο κατευθύνσεις (αμφίδρομα) ταυτόχρονα, επιτρέπει την κατανόηση του πλήρους νοήματος μιας λέξης. Επεξεργάζεται μια δεδομένη λέξη σε σχέση με όλες τις λέξεις μιας πρότασης ή φράσης αντί να επεξεργάζεται μία λέξη κάθε φορά. Δηλαδή αντιστοιχεί περισσότερα διανύσματα για την ίδια λέξη, εφόσον τα συμφραζόμενα είναι διαφορετικά, σε αντίθεση με το word2 που αδυνατεί να αντιμετωπίσει τη συνύπαρξη πολλών πιθανών σημασιών σε μια δεδομένη λέξη ή φράση και δημιουργεί ένα ενιαίο διάνυσμα για την ίδια λέξη (Ankiit, χ.χ.). Θα έλεγε κανείς ότι πρόκειται για ένα μοντέλο ικανό να αναλύει τη γλώσσα πιο κοντά στην ανθρώπινη λογική. Εφαρμόζεται ήδη με επιτυχία στη μηχανή αναζήτησης της Google καθώς και σε εφαρμογές ανάλυσης συναισθήματος, μετάφρασης γλώσσας κ.α. (H2O.a, χ.χ.; Lutkevich, χ.χ.). Ωστόσο το βασικό του μειονέκτημα είναι ότι η εκπαίδευσή του είναι αργή και ακριβή

λόγω του μεγέθους του και των πολλών υπολογισμών που πρέπει να κάνει (ProjectPro, 2022).

3.2 *Deep Learning*

Το Deep Learning (DL) είναι ένα είδος machine learning, ενώ και τα δύο ανήκουν στον τομέα της τεχνητής νοημοσύνης (βλ. εικόνα 3). Το deep learning αποτελεί εξέλιξη του machine learning. Επιτρέπει στα υπολογιστικά συστήματα να βελτιώνονται με εμπειρία και δεδομένα (Goodfellow κ.ά., 2016). Χρησιμοποιεί βαθιά τεχνητά νευρωνικά δίκτυα που μοιάζουν με τον ανθρώπινο εγκέφαλο και επιτρέπει στους υπολογιστές να μαθαίνουν και να ενεργούν μόνοι τους, χωρίς τη βοήθεια του ανθρώπου. Οι υπολογιστές δηλαδή είναι σε θέση να λαμβάνουν έξυπνες αποφάσεις.

Η κύρια διαφορά τους από τα μοντέλα machine learning είναι ότι ένας αλγόριθμος deep learning μπορεί να καθορίσει μόνος του αν η πρόβλεψη είναι σωστή ή όχι και να προβεί σε διορθώσεις χωρίς την ανθρώπινη παρέμβαση σε αντίθεση με ένα μοντέλο machine learning όπου η ανθρώπινη παρέμβαση είναι απαραίτητη (Grieve, 2022).



Εικόνα 3: Artificial Intelligence (AI) - Machine Learning (ML) - Deep Learning (DL)

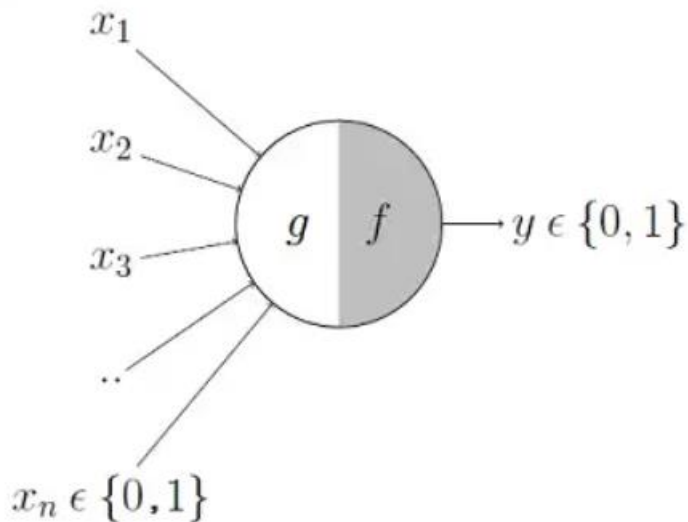
Πηγή: (Simplilearn, 2022)

3.2.1 Ιστορική αναδρομή

Οι περισσότεροι πιστεύουν ότι το deep learning είναι μια νέα τεχνολογία. Στην πραγματικότητα όμως χρονολογείται από το 1940 και μέχρι σήμερα έχει αλλάξει αρκετές ονομασίες. Μόλις τα τελευταία χρόνια μετονομάστηκε με τον όρο “Deep Learning”. Πιο συγκεκριμένα, οι τρεις φάσεις ανάπτυξης του Deep Learning είναι οι εξής (Goodfellow κ.ά., 2016):

Κατά τη **δεκαετία 1940-1960** ήταν γνωστό με τον όρο “**cybernetics**”. Κατά την περίοδο αυτή, αναπτύχθηκαν οι θεωρίες της βιολογικής μάθησης και τα πρώτα μοντέλα ασχολήθηκαν με το πώς γίνεται η μάθηση και τι συμβαίνει στον εγκέφαλο. Δηλαδή προσέγγισε τα νευρωνικά δίκτυα από τη νευροεπιστημονική προοπτική. Τα μοντέλα αυτά ήταν εμπνευσμένα από τον βιολογικό εγκέφαλο και έτσι ένας από τους όρους που επικράτησε μέχρι και σήμερα είναι ο όρος “Artificial Neural Networks” (ANNs).

Ένα πρώιμο μοντέλο νευρωνικού δικτύου αναπτύχθηκε το 1943 από τους **McCulloch και Pitts** και θεωρείται ότι είναι ιστορικά η πρώτη απεικόνιση ενός νευρωνικού δικτύου (βλ. εικόνα 4). Είναι ένα γραμμικό μοντέλο που αποτελείται από έναν νευρώνα και τις διασυνδέσεις του. Δέχεται πολλές δυαδικές εισόδους (δηλαδή εισόδους με τιμές 0 ή 1) αλλά έχει μόνο μία δυαδική έξοδο και τα βάρη ορίζονται με ανθρώπινη παρέμβαση (Goodfellow κ.ά., 2016). Οι τιμές των εισόδων με τα αντίστοιχα βάρη των συνδέσεων τους συναθροίζονται και αν η τιμή υπερβαίνει το όριο του κατωφλιού (threshold) τότε η τιμή εξόδου είναι 1 οπότε ο νευρώνας ενεργοποιείται, ενώ διαφορετικά είναι 0 και ο νευρώνας παραμένει αδρανής (Αργυράκης, 2001).



Εικόνα 4: McCulloch and Pitts neuron

Πηγή: (Chandra, 2018)

Αργότερα, το 1957 ο **F.Rosenblatt** παρουσίασε το γραμμικό μοντέλο **perceptron**. Πρόκειται για ένα μοντέλο feedforward δυαδικής ταξινόμησης το οποίο έχει την ικανότητα να επεξεργάζεται οποιαδήποτε πραγματική τιμή και να μαθαίνει τα βάρη μόνο του από τα παραδείγματα των εισροών που δέχεται (Goodfellow κ.ά., 2016). Αρχικά, η ανάπτυξη αυτού του μοντέλου δημιούργησε μεγάλο ενθουσιασμό όμως γρήγορα διαπιστώθηκε ότι υπήρχαν πολλοί περιορισμοί με αποτέλεσμα η δημοτικότητα του να χαθεί για αρκετά χρόνια (Αργυράκης, 2001).

Κατά τη δεκαετία 1980-1990 ήταν γνωστό με τον όρο “**connectionism**” ή “**parallel distributed processing**” (Goodfellow κ.ά., 2016), λόγω του ομώνυμου έργου που δημοσιεύτηκε από τους McClelland και Rumelhart (Αργυράκης, 2001). Η εποχή ήταν εμπνευσμένη από τη γνωστική επιστήμη (Goodfellow κ.ά., 2016) που ασχολείται με την κατανόηση του νου και των λειτουργιών του. Μια από τις ιδέες που επικράτησε εκείνη την εποχή ήταν ότι ένα νευρωνικό δίκτυο μπορεί να θεωρηθεί ως ένας παράλληλος επεξεργαστής, ανοίγοντας έτσι τον δρόμο για την εκπαίδευση των νευρωνικών δικτύων (Αργυράκης, 2001). Μάλιστα ένα από τα σημαντικότερα επιτεύγματα της εποχής ήταν η ανάπτυξη της μεθόδου **back-propagation** για την εκπαίδευση των βαθιών νευρωνικών δικτύων (Goodfellow κ.ά., 2016). Επίσης θα πρέπει να αναφερθεί ότι κατά τη διάρκεια της δεκαετίας του 1990 οι επιστήμονες έκαναν σημαντικές προόδους στη μοντελοποίηση ακολουθιών και το 1997 οι Hochreiter και Schmidhuber παρουσίασαν το μοντέλο **Long short-term memory** (LSTM) το οποίο χρησιμοποιείται μέχρι και σήμερα (Goodfellow κ.ά., 2016).

Παρά την πρόοδο που σημειώθηκε σε αυτή τη δεκαετία, οι υπερβολικά φιλόδοξες και παράλογες προσδοκίες οδήγησαν στη μείωση της δημοτικότητας των νευρωνικών δικτύων για μια ακόμη φορά.

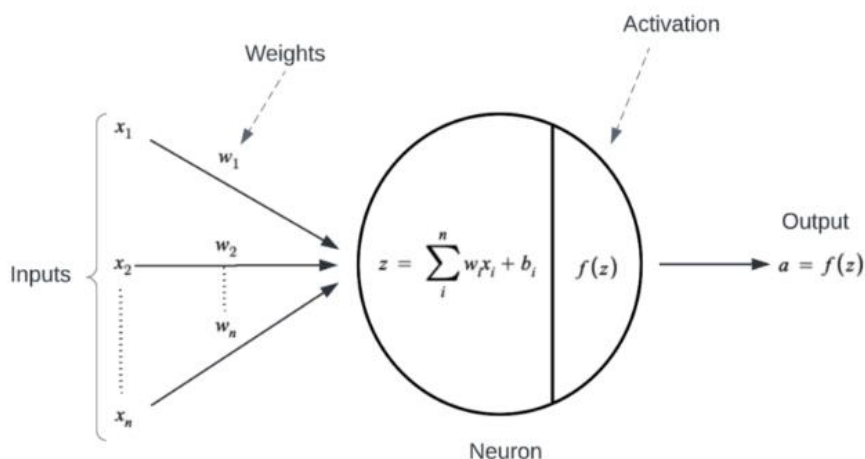
Στις αρχές του **2006 μέχρι και σήμερα** επικράτησε ο όρος “**deep learning**”. Την περίοδο αυτή και συγκεκριμένα το 2006, ο Hinton ανακάλυψε ότι η μέθοδος εκπαίδευσης greedy layer-wise pretraining μπορεί να εκπαιδεύσει αποτελεσματικά το νευρωνικό δίκτυο deep belief network. Παρομοίως και άλλες ερευνητικές ομάδες ανακάλυψαν ότι αυτή η μέθοδος μπορεί να εκπαιδεύσει και άλλους τύπους βαθιών νευρωνικών δικτύων. Πλέον οι ερευνητές είχαν τη δυνατότητα να εκπαιδεύουν βαθύτερα νευρωνικά δίκτυα και γι’ αυτό επικράτησε και ο όρος “**deep learning**” (Goodfellow κ.ά., 2016).

Η δημοτικότητα των νευρωνικών δικτύων πλέον μεγαλώνει και συνεχίζεται μέχρι και σήμερα. Η αύξηση της δημοτικότητάς του οφείλεται στη σημαντική πρόοδο της τεχνολογίας και τη δυνατότητα διαχείρισης μεγάλου όγκου δεδομένων (big data) (Goodfellow κ.ά., 2016). Λόγω της αυξανόμενης ψηφιοποίησης της κοινωνίας, όλες οι δραστηριότητες καταγράφονται μέσω των πολυάριθμων εφαρμογών. Πλέον υπάρχει διαθέσιμος μεγάλος όγκος δεδομένων, κάνοντας εφικτή την εκπαίδευση των μοντέλων deep learning δεδομένου ότι οι αλγόριθμοι DL απαιτούν μεγάλο όγκο δεδομένων για την εκπαίδευσή τους που στο παρελθόν δεν ήταν διαθέσιμος. Επίσης σημαντικό ρόλο, στη δημοτικότητα των νευρωνικών δικτύων, έχει και η αύξηση της υπολογιστικής ισχύος. Σήμερα οι υπολογιστές είναι ισχυρότεροι, ταχύτεροι και διαθέτουν μεγαλύτερη μνήμη. Η ταχύτερη συνδεσιμότητα του δικτύου και τα εξελιγμένα λογισμικά βοήθησαν στην αύξηση του μεγέθους των μοντέλων deep learning (Κύρκος, 2015) καθώς και τη βελτίωση της ακρίβειας και της αξιοπιστίας τους. Σήμερα τα μοντέλα deep learning εφαρμόζονται σε όλο και περισσότερες εφαρμογές καλύπτοντας ένα ευρύ φάσμα επιστημονικών πεδίων. Επίσης η διαθεσιμότητα βιβλιοθηκών λογισμικού ανοικτού κώδικα, όπως Theano, TensorFlow, PyLearn2, Torch, MXNet, Keras, Pytorch κ.α., συνέβαλαν στην υλοποίησή τους αλλά και την πρόοδο του deep learning (Goodfellow κ.ά., 2016; Sun & Vasarhelyi, 2018).

3.2.2 Περιγραφή τεχνητού νευρωνικού δικτύου (artificial neural networks, ANNs)

Τα τεχνητά νευρωνικά δίκτυα είναι εμπνευσμένα από το βιολογικό νευρικό σύστημα (Κύρκος, 2015) και καλούνται να λύσουν κάθε είδους προβλήματα με τη χρήση του ηλεκτρονικού υπολογιστή. Όμως ο τρόπος λειτουργίας και η φιλοσοφία των νευρωνικών δικτύων διαφέρει από τον κλασικό υπολογιστή. Ένα τεχνητό νευρωνικό δίκτυο συνδυάζει τον ανθρώπινο τρόπο σκέψης με τον μαθηματικό τρόπο σκέψης, δηλαδή ένα νευρωνικό δίκτυο μπορεί να μαθαίνει και να εκπαιδεύεται, να θυμάται ή να ξεχνάει μια αριθμητική τιμή αλλά και να χρησιμοποιεί περίπλοκες μαθηματικές συναρτήσεις. Η βασική τους διαφορά από τα βιολογικά νευρωνικά δίκτυα είναι ότι, ενώ μαθαίνουν μέσω της εκπαίδευσης και της εμπειρίας, ακολουθούν προκαθορισμένους κανόνες προκειμένου να επιτελέσουν τις διαδικασίες για τις οποίες έχουν σχεδιαστεί (Αργυράκης, 2001).

Το τεχνητό νευρωνικό δίκτυο αποτελείται από **νευρώνες** που είναι η βασική μονάδα του δικτύου. Οι νευρώνες λαμβάνουν πληροφορίες από άλλους νευρώνες, τις επεξεργάζονται και τις στέλνουν με τη σειρά τους σε άλλους νευρώνες μέσω των **συνδέσεων** τους με τη μορφή ηλεκτρικού σήματος (Αργυράκης, 2001. Κύρκος, 2015). Κάθε σύνδεση έχει μια αριθμητική τιμή που ονομάζεται **βάρος** (w) (Κύρκος, 2015). Το βάρος (w) δείχνει πόσο ισχυρή είναι η σύνδεση μεταξύ των δύο νευρώνων. Όσο πιο μεγάλο το βάρος (w) τόσο πιο ισχυρή και η σύνδεση μεταξύ των δύο νευρώνων (Αργυράκης, 2001). Κάθε νευρώνας λαμβάνει ως τιμές εισόδου, τις τιμές εξόδου των αντίστοιχων συνδεδεμένων νευρώνων πολλαπλασιασμένες με τα βάρη (w) των συνδέσεων τους. Στη συνέχεια, ο νευρώνας αθροίζει τις τιμές εισόδου και με τη βοήθεια μιας συνάρτησης ενεργοποίησης (activation function), προκύπτει η τιμή εξόδου η οποία με τη σειρά της μεταδίδεται μέσω των αντίστοιχων συνδέσεων στους επόμενους νευρώνες πολλαπλασιασμένη με τα αντίστοιχα βάρη (w) των συνδέσεων τους κ.ο.κ. (βλ. εικόνα 5) (Κύρκος, 2015). Επιπλέον μπορεί να υπάρξει η σταθερά bias (πόλωση) που έχει τιμή +1. Η σταθερά bias είναι μια επιπλέον τιμή εισόδου χωρίς εισερχόμενες συνδέσεις και προστίθεται στο άθροισμα των τιμών εισόδου του νευρώνα [$\sum(w * x)+bias$]. Ουσιαστικά εγγυάται την ενεργοποίηση του νευρώνα ακόμα και όταν όλες τιμές εισόδου είναι μηδέν (Κύρκος, 2015).



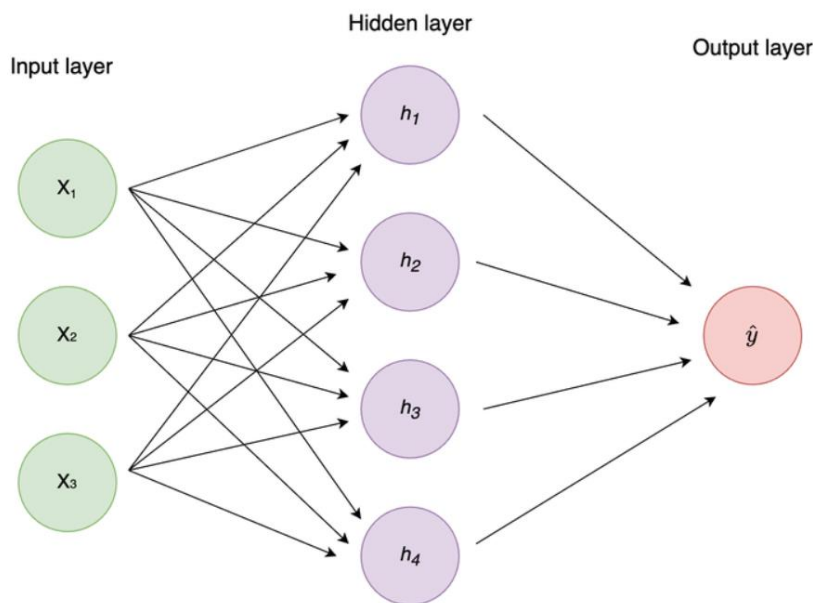
Εικόνα 5: Ενεργοποίηση νευρώνα

Πηγή: (Saxena, 2022)

Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από ένα επίπεδο εισόδου (input layer), ένα επίπεδο εξόδου (output layer) και τα κρυφά επίπεδα (hidden layers). Τα κρυφά επίπεδα βρίσκονται ανάμεσα στο επίπεδο εισόδου και το επίπεδο εξόδου και ποικίλουν στον αριθμό τους. Για παράδειγμα, ένα δίκτυο τριών επιπέδων αποτελείται από ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου. Αναλυτικότερα, **το επίπεδο εισόδου** αποτελείται από νευρώνες που δέχονται απλά τις τιμές εισόδου και τις στέλνουν στους επόμενους συνδεδεμένους νευρώνες, χωρίς δηλαδή να μετασχηματίζουν τις τιμές. **Οι νευρώνες του δεύτερου επιπέδου**, του κρυφού επιπέδου, αποτελούνται από νευρώνες οι οποίοι δέχονται τις τιμές του επιπέδου εισόδου πολλαπλασιασμένες με τα βάρη των αντίστοιχων συνδέσεων, τις αθροίζουν και τις μετασχηματίζουν, με τη βοήθεια της συνάρτησης ενεργοποίησης, σε τιμές εξόδου. Στη συνέχεια, οι τιμές εξόδου αποστέλλονται **στους νευρώνες του τρίτου επιπέδου**, δηλαδή του επιπέδου εξόδου, μέσω των αντίστοιχων συνδέσεων πολλαπλασιασμένες με τα αντίστοιχα βάρη (w) και υπολογίζουν το αποτέλεσμα (Κύρκος, 2015).

Ένας νευρώνας μπορεί να είναι πλήρως συνδεδεμένος, δηλαδή να μεταβιβάζει την τιμή εξόδου σε όλους τους νευρώνες του επόμενου επιπέδου. Επίσης οι συνδέσεις μεταξύ των νευρώνων μπορεί να είναι μονόδρομες ή αμφίδρομες. Στην πρώτη περίπτωση το δίκτυο ονομάζεται feed forward και στη δεύτερη recurrent (Αργυράκης, 2001. Κύρκος, 2015). Επίσης θα πρέπει να σημειωθεί ότι το πλήθος των νευρώνων εισόδου ισούται με το πλήθος των ανεξάρτητων μεταβλητών που λαμβάνονται υπόψη ενώ το πλήθος των νευρώνων εξόδου ισούται με το πλήθος των τιμών κλάσεων (Κύρκος, 2015).

Παρακάτω απεικονίζεται ένα τεχνητό νευρωνικό δίκτυο τριών επιπέδων οι συνδέσεις του οποίου είναι μονόδρομες και οι νευρώνες είναι πλήρως συνδεδεμένοι (βλ. εικόνα 6).



Εικόνα 6: Τεχνητό νευρωνικό δίκτυο 3 επιπέδων feed forward

Πηγή: (Hansen, 2022)

3.2.3 Υπερπαράμετροι τεχνητών νευρωνικών δικτύων

Πριν την εκπαίδευση του νευρωνικού δικτύου γίνεται η επιλογή των υπερπαραμέτρων. Η επιλογή των κατάλληλων υπερπαραμέτρων είναι καθοριστικής σημασίας για την αποτελεσματικότητα του νευρωνικού δικτύου. Οι υπερπαράμετροι αφορούν την επιλογή (α) της κατάλληλης αρχιτεκτονικής του δικτύου και (β) των κατάλληλων μεταβλητών για την εκπαίδευσή του. Πρόκειται για μια δύσκολη διαδικασία διότι δεν υπάρχουν προκαθορισμένοι και τεκμηριωμένοι κανόνες. Η επιλογή τους γίνεται με βάση εμπειρικούς κανόνες ή με βάση τα αποτελέσματα πειραματικών διαδικασιών (Radhakrishnan, 2017).

3.2.3.1 Υπερπαράμετροι αρχιτεκτονικής νευρωνικού δικτύου

Η αρχιτεκτονική των νευρωνικών δικτύων περιλαμβάνει (α) το **πλήθος** των κρυφών επιπέδων και των νευρώνων που επηρεάζουν την ακρίβεια του μοντέλου, (β) τη **συνάρτηση ενεργοποίησης** (activation function) κάθε επιπέδου που επηρεάζει την απόδοσή του (π.χ. Rectified Linear Activation (**ReLU**), Logistic (**Sigmoid**), Hyperbolic Tangent (**Tanh**), **Softmax**) (γ) τη τεχνική **dropout** για την αποφυγή υπερπροσαρμογής και την επίτευξη γενίκευσης του δικτύου, (δ) τη τεχνική **αρχικοποίησης** των βαρών (network weight initialization) για τον καθορισμό των αρχικών τιμών (Brownlee, 2021).

3.2.3.2 Υπερπαράμετροι εκπαίδευσης νευρωνικού δικτύου

Οι υπερπαράμετροι εκπαίδευσης περιλαμβάνουν (α) τον **ρυθμό εκπαίδευσης** (learning rate), (β) την **ορμή** (momentum), (γ) τον **αριθμό εποχών** (epochs) και (δ) το **μέγεθος παρτίδας** (batch size) (Alto, 2019; Radhakrishnan, 2017).

Ο **ρυθμός εκπαίδευσης** (learning rate) είναι ο βαθμός μεταβολής των βαρών και επηρεάζει την ταχύτητα εκπαίδευσης. Μικρή μεταβολή βαρών σημαίνει αργή εκπαίδευση, ενώ μεγάλη μεταβολή γρήγορη εκπαίδευση με κίνδυνο όμως την υπερπροσαρμογή του μοντέλου, δηλαδή τον κίνδυνο το μοντέλο να απομνημονεύσει το σετ εκπαίδευσης αντί να το μάθει (Κύρκος, 2015). Η **ορμή** (momentum) βελτιώνει την ταχύτητα εκπαίδευσης και την ακρίβεια με την εύρεση τιμών για τα βάρη διατηρώντας την κατεύθυνση μεταβολής τους. Ο **αριθμός εποχών** (epochs) είναι ο αριθμός επαναλήψεων εκπαίδευσης όλου του συνόλου εκπαίδευσης (training dataset) και το **μέγεθος παρτίδας** (batch size) είναι ο αριθμός των δειγμάτων που περιέχει κάθε παρτίδα. Δηλαδή το σύνολο δεδομένων χωρίζεται σε παρτίδες και κάθε παρτίδα περιέχει συγκεκριμένο αριθμό δειγμάτων π.χ. αν το σύνολο δεδομένων περιέχει 200 δείγματα και το μέγεθος παρτίδας είναι 5 σημαίνει ότι κάθε παρτίδα θα περιέχει μόνο 5 δείγματα και το πλήθος των παρτίδων θα είναι 40 (200 : 5) (Brownlee, 2022).

3.2.4 Εκπαίδευση, επικύρωση και έλεγχος νευρωνικού δικτύου

Η εκπαίδευση ενός νευρωνικού δικτύου ουσιαστικά αφορά τη ρύθμιση των βαρών (w) των συνδέσεων μεταξύ των νευρώνων. Συνήθως ένα μέρος του συνόλου δεδομένων αποτελεί το **σύνολο εκπαίδευσης** (training dataset) που χρησιμοποιείται για τη ρύθμιση των βαρών των συνδέσεων, ένα δεύτερο μέρος αποτελεί το **σύνολο επικύρωσης** (validation dataset) για την αξιολόγηση της προσαρμογής του μοντέλου στο σύνολο εκπαίδευσης και τη διακοπή της εκπαίδευσης και ένα τρίτο μέρος, το **σύνολο ελέγχου** (test dataset) που χρησιμοποιείται για τον έλεγχο της απόδοσης του τελικού μοντέλου. Συχνά το σύνολο δεδομένων χωρίζεται σε δύο μέρη αντί για τρία, δηλαδή χωρίζεται σε σύνολο εκπαίδευσης και σύνολο ελέγχου. Για το σύνολο επικύρωσης χρησιμοποιείται το σύνολο ελέγχου (Κύρκος, 2015).

Η εκπαίδευση μπορεί να γίνει με δύο τρόπους, την εποπτευόμενη μάθηση και τη μη εποπτευόμενη μάθηση (Κύρκος, 2015).

Στην εποπτευόμενη μάθηση δίνονται κάποια πρότυπα (παραδείγματα) στο δίκτυο με τις τιμές εισόδου και τις τιμές στόχου. Οι τιμές στόχου ουσιαστικά είναι οι τιμές εξόδου που πρέπει να μάθει το δίκτυο. Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο ρυθμίζει τις τιμές των βαρών με σκοπό να διορθώσει το σφάλμα στόχου, δηλαδή να διορθώσει τη διαφορά μεταξύ της τιμής εξόδου που λαμβάνει με την αντίστοιχη τιμή εξόδου του προτύπου. Η διαδικασία επαναλαμβάνεται μέχρι η διαφορά των τιμών να ελαχιστοποιηθεί οπότε και η εκπαίδευση σταματάει (Αργυράκης, 2001. Κύρκος, 2015).

Στη μη εποπτευόμενη μάθηση δίνονται κάποιες πληροφορίες στο δίκτυο χωρίς όμως τις τιμές στόχου. Το δίκτυο εκπαιδεύεται παράγοντας εξόδους που να ταιριάζουν με τις εισόδους. Χρησιμοποιεί κάποιον εσωτερικό έλεγχο για τη ρύθμιση των βαρών του και διορθώνει τα σφάλματα με έναν μηχανισμό ανάδρασης (feedback) γι' αυτό ονομάζεται και αυτό-εποπτευόμενη εκπαίδευση (Αργυράκης, 2001).

Η μέθοδος ρύθμισης των βαρών διαφέρει ανάλογα με τον τύπο του νευρωνικού δικτύου (Κύρκος, 2015). Επίσης κατά τη διάρκεια της εκπαίδευσης, μετά από κάποιον αριθμό επαναλήψεων, η πρόοδος του δικτύου σταματάει. Τότε λέμε ότι το δίκτυο συγκλίνει (convergence), δηλαδή το δίκτυο έχει μάθει να ανταποκρίνεται σωστά στο σύνολο εκπαίδευσης εντός όμως κάποιου περιθωρίου σφάλματος. Αυτό σημαίνει ότι η επιπλέον εκπαίδευση δεν θα βελτιώσει περαιτέρω το δίκτυο (Paperspace, 2020). Ουσιαστικά η εκπαίδευση ολοκληρώνεται όταν το σφάλμα στην έξοδο είναι μηδέν ή κοντά στο μηδέν.

Μετά την ολοκλήρωση της εκπαίδευσης το δίκτυο είναι ικανό να λύνει και άλλα, όμως ίδιας φύσης, προβλήματα στα οποία δεν έχει εκπαιδευτεί και δεν έχει δει ποτέ του. Αυτό συμβαίνει διότι τα κρυφά επίπεδα αναγνωρίζουν τα σημαντικά χαρακτηριστικά, δημιουργούν μια εσωτερική αναπαράσταση των προτύπων που εισέρχονται στο δίκτυο και αναγνωρίζουν εκείνα που δεν έχουν δει (Αργυράκης, 2001).

3.2.5 Μειονεκτήματα μοντέλων deep learning

Παρά την αύξηση της δημοτικότητας των μοντέλων deep learning και της ύπαρξης πολυάριθμων πετυχημένων εφαρμογών, εμφανίζουν κάποια μειονεκτήματα που είναι καλό να αναφερθούν. Τα μοντέλα deep learning χαρακτηρίζονται και ως μοντέλα μαύρου κουτιού, διότι είναι ιδιαίτερα περίπλοκα και δύσκολα ερμηνεύσιμα. Οι άνθρωποι δεν είναι σε θέση να κατανοήσουν το σκεπτικό του μοντέλου και τις

αποφάσεις που λαμβάνει. Παρόλο που η επιλογή των **υπερπαραμέτρων** είναι γνωστή, δεν είναι σαφές πως οι υπολογιστές μαθαίνουν και εξάγουν τα αποτελέσματά τους (Sun & Vasarhelyi, 2018; Κύρκος, 2015). Επίσης η επιλογή των υπερπαραμέτρων γίνεται κυρίως εμπειρικά ή έπειτα από την εφαρμογή πειραματικών διαδικασιών. Ουσιαστικά, δηλαδή, δεν υπάρχουν προκαθορισμένοι κανόνες για την επιλογή τους. Ενώ ο **χρόνος εκπαίδευσης** είναι μεγάλος κυρίως λόγω της αρχιτεκτονικής των δικτύων και του μεγάλου όγκου δεδομένων που απαιτούν (Κύρκος, 2015).

4

Περιγραφή των ερευνητικών άρθρων

Στο παρόν κεφάλαιο παρουσιάζονται τα δεκαπέντε ερευνητικά άρθρα που συλλέχθηκαν και αποτελούν το αντικείμενο αυτής της εργασίας. Στον παρακάτω πίνακα απεικονίζονται οι τίτλοι των ερευνητικών άρθρων, οι συγγραφείς, τα μοντέλα deep learning που αναπτύχθηκαν και ο σκοπός της πρόβλεψης (βλ. πίνακα 1).

Πίνακας 1: Τα δεκαπέντε ερευνητικά άρθρα

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μοντέλο	Σκοπός
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	HAN	Πρόβλεψη απάτης
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiuguo και Shengyong (2022)	LSTM, GRU, TRANSFORMER	Πρόβλεψη απάτης
Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	Jan (2021a)	RNN, LSTM	Πρόβλεψη απάτης
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	Li κ.ά. (2021)	DNN, MA-DNN, Bi-LSTM	Πρόβλεψη οικονομικής δυσπραγίας
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	3 models CNN+RNN	Πρόβλεψη οικονομικής δυσπραγίας
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	CNN	Πρόβλεψη χρεοκοπίας
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	DL-Embedding, CNN	Πρόβλεψη χρεοκοπίας
A Deep Dense Neural Network for Bankruptcy Prediction	Alexandropoulos κ.ά. (2019)	DDMP	Πρόβλεψη χρεοκοπίας
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	DL BSM-SAES	Πρόβλεψη χρεοκοπίας
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	Jang κ.ά. (2021)	LSTM-RNN (χρήση τιμής Shapley)	Πρόβλεψη χρεοκοπίας
Bankruptcy or Success? The effective Prediction of a Company's Financial Development Using LSTM	Vochozka κ.ά. (2020)	NN+LSTM layer	Πρόβλεψη χρεοκοπίας
Corporate failure prediction: An evaluation of deep learning vs discrete hazard models	Alam κ.ά. (2021)	GrNet	Πρόβλεψη οικονομικής αποτυχίας
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	Aljawazneh κ.ά. (2021)	DBN, MLP-6L, LSTM	Πρόβλεψη οικονομικής αποτυχίας
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	Jan (2021b)	CART-DNN, CART-RNN	Πρόβλεψη συνεχιζόμενης δραστηριότητας
Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction	Chi και Chu (2021)	LSTM, GRU	Πρόβλεψη συνεχιζόμενης δραστηριότητας

Ακολουθεί αναλυτική περιγραφή των δεκαπέντε ερευνητικών άρθρων.

4.1 Deep learning for detecting financial statement fraud

Η έρευνα των Craja, Kim και Lessmann (2020) επικεντρώθηκε στην απάτη των χρηματοοικονομικών καταστάσεων με την ανάπτυξη του μοντέλου deep learning **Hierarchical Attention Network (HAN)**, το οποίο συνδυάζει τις πληροφορίες που προέρχονται από τα οικονομικά δεδομένα και την ενότητα των δηλώσεων του Δ.Σ. (MD&A) των ετήσιων οικονομικών εκθέσεων, με σκοπό την παροχή ακριβών και ερμηνεύσιμων προβλέψεων απάτης, χρήσιμων για την υποστήριξη λήψης αποφάσεων όλων των ενδιαφερόμενων μερών (επενδυτές, ελεγκτικές εταιρείες, κρατικές ρυθμιστικές αρχές κλπ).

Η ανάπτυξη αυτού του νέου μοντέλου DL έχει ως στόχο να απαντηθούν τα εξής ερευνητικά ερωτήματα:

- αν ο συνδυασμός των οικονομικών δεδομένων με δεδομένα κειμένου μπορεί να δώσει ικανοποιητικά αποτελέσματα για τον εντοπισμό απάτης,
- αν η νέα μέθοδος DL έχει καλύτερη απόδοση σε σύγκριση με την τεχνική Bag-Of-Words (BOW) και
- αν μπορεί να βοηθήσει στην ερμηνεία του κειμένου και να παρέχει δείκτες κόκκινης σημαίας για την υποστήριξη της λήψης αποφάσεων των ελεγκτών.

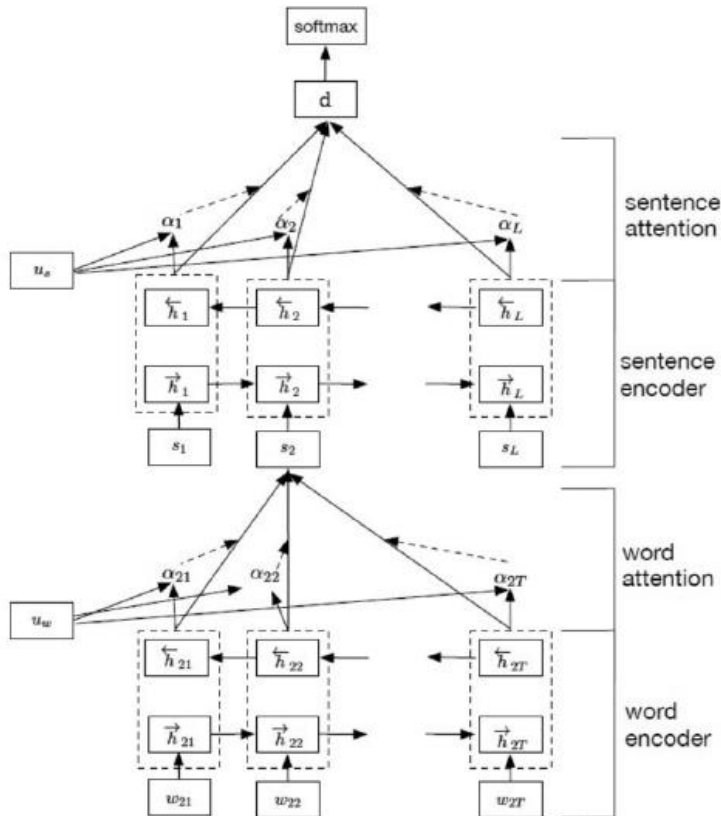
Τα **δεδομένα** της έρευνας αποτελούνται από 7757 εταιρικά έτη με 208 περιπτώσεις απάτης και 7549 περιπτώσεις μη απάτης και προέρχονται από τις ετήσιες οικονομικές εκθέσεις εισηγμένων εταιρειών των ΗΠΑ για το χρονικό διάστημα 1993-2019. Τα δεδομένα κειμένου και οι γλωσσικές μεταβλητές αντλήθηκαν από τη βάση δεδομένων EDGAR του ιστότοπου του SEC και συγκεκριμένα από την ενότητα MD&A. Τα γλωσσικά δεδομένα βασίζονται σε προκαθορισμένες λίστες λέξεων που έχουν σχέση με την απάτη και δηλώνουν αρνητικότητα, αβεβαιότητα, θετικό ή αρνητικό συναίσθημα (L&M word lists by Loughran and McDonald, 2011 όπως αναφέρεται στους Craja κ.ά., 2020) ή έχουν σχέση με αναλογίες αναγνωρισιμότητας, όπως το μήκος, την πολυπλοκότητα ή την ποικιλομορφία των λέξεων (Readability ratios by Humpherys et al., 2011; Li, 2008 όπως αναφέρεται στους Craja κ.ά., 2020). Τα ποσοτικά δεδομένα (47 χρηματοοικονομικές μεταβλητές) αντλήθηκαν από τη βάση δεδομένων COMPUSTAT.

Για την **εξισορρόπηση του συνόλου δεδομένων** εφαρμόστηκε υποδειγματοληψία με αναλογία απάτης προς μη απάτη 1:4. Στο σύνολο δεδομένων λήφθηκαν υπόψη παράμετροι όπως το έτος αναφοράς, ο τομέας δραστηριότητας, οι αλλαγές στη νομοθεσία και τους κανονισμούς και οι οικονομικές συνθήκες που επικρατούσαν, με αποτέλεσμα το **τελικό σύνολο δεδομένων** να αποτελείται από 1163 ετήσιες εκθέσεις από τις οποίες οι 201 είναι δόλιες και οι 962 μη δόλιες.

Οι οικονομικές και οι γλωσσικές μεταβλητές είναι δομημένα δεδομένα πίνακα και δεν απαιτούν προεπεξεργασία. Αντιθέτως, τα δεδομένα κειμένου είναι αδόμητα δεδομένα και απαιτούν προεπεξεργασία προκειμένου να μετατραπούν σε αριθμητική μορφή και να είναι δυνατή η αλγοριθμική τους επεξεργασία. Έτσι το προτεινόμενο μοντέλο HAN κατασκευάστηκε με ενσωματώσεις word2vec με 300 νευρώνες, εκπαιδεύτηκε στο Google News corpus με μέγεθος λεξιλογίου 3 εκατ. λέξεων.

Για την κωδικοποίηση του εγγράφου, το HAN εντοπίζει αρχικά τις λέξεις που είναι σημαντικές μέσα σε μια πρόταση και έπειτα τις προτάσεις που είναι σημαντικές στο έγγραφο λαμβάνοντας ταυτόχρονα υπόψη και τα συμφραζόμενα (βλ. εικόνα 7). Δηλαδή μπορεί να αναγνωρίσει πότε μια λέξη είναι σημαντική σε μια πρόταση και

πότε όχι. Το μοντέλο αποτελείται από (1) ένα **encoder**, το **bidirectional-LSTM**, που δημιουργεί τα contexts και (2) ένα **attention mechanism** που υπολογίζει τα βάρη σημαντικότητας. Οι ίδιοι **αλγόριθμοι** εφαρμόζονται αρχικά σε επίπεδο λέξης και μετά σε επίπεδο πρότασης.

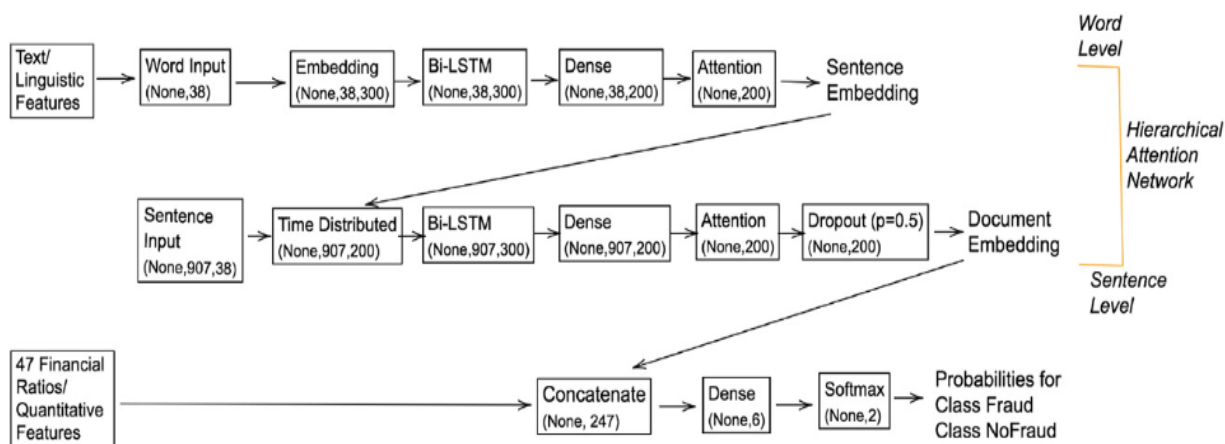


Εικόνα 7: Κωδικοποίηση εγγράφου με το μοντέλο HAN

Πηγή: (Craja κ.ά., 2020)

Η αρχιτεκτονική του μοντέλου HAN για την πρόβλεψη απάτης απεικονίζεται στην εικόνα 8. Αναλυτικότερα, το LSTM layer αποτελείται από 150 νευρώνες, τα HAN dense layer από 200 ενώ το τελευταίο από 6 νευρώνες. Ο συνδυασμός forward-backward LSTM δίνει 300 διαστάσεις για τον σχολιασμό των λέξεων και των προτάσεων. Επίσης στο τελευταίο layer του HAN εφαρμόζεται η τεχνική dropout regularization για την αποφυγή της υπερπροσαρμογής του μοντέλου. Τέλος, τα δεδομένα κειμένου συνενώνονται με τους 47 χρηματοοικονομικούς δείκτες, στη συνέχεια εισάγονται στο dense layer και με τη βοήθεια της συνάρτησης softmax λαμβάνεται ως έξοδος το αποτέλεσμα της πρόβλεψης για την πιθανότητα απάτης ή μη απάτης των οικονομικών καταστάσεων.

Κατά την εκπαίδευση του μοντέλου, το batch size ήταν 32 και οι εποχές 17.



Εικόνα 8: Η αρχιτεκτονική του μοντέλου HAN

Πηγή: (Craja κ.ά., 2020)

Για την αξιολόγηση της απόδοσης του μοντέλου, χρησιμοποιήθηκαν τα μέτρα αξιολόγησης **AUC**, **sensitivity**, **specificity**, **F1-score**, **F2-score** και **Accuracy**. Για τον υπολογισμό των F1-score και F2-score, το όριο κατωφλιού ορίστηκε σε 0,03 που σημαίνει ότι μια ετήσια έκθεση ταξινομείται ως δόλια αν η πιθανότητα απάτης είναι πάνω από 3%.

Κατά τα πειράματα, χρησιμοποιήθηκαν τα μοντέλα **Logistic Regression (LR)**, **Random Forest (RF)**, **Support Vector Machine (SVM)**, **Extreme Gradient Boosting (XGB)** και **Artificial Neural Networks (ANN)** που βασίζονται στην τεχνική BOW και πέντε διαφορετικοί συνδυασμοί δεδομένων. Πιο συγκεκριμένα, οι πέντε συνδυασμοί δεδομένων που εφαρμόστηκαν είναι:

1. οικονομικά δεδομένα (FIN),
2. γλωσσικά δεδομένα (LING),
3. οικονομικά και γλωσσικά δεδομένα (FIN + LING),
4. δεδομένα κειμένου (TXT),
5. οικονομικά δεδομένα και δεδομένα κειμένου (FIN + TXT)

Τα οικονομικά δεδομένα (FIN) αποτελούν το σημείο αναφοράς της έρευνας. Οι δύο τελευταίες στήλες του πίνακα αποτελεσμάτων (βλ. πίνακα 2) δείχνουν τα αποτελέσματα της σύγκρισης κάθε συνδυασμού δεδομένων με τα δεδομένα FIN.

Αρχικά, έγινε σύγκριση των οικονομικών δεδομένων (FIN) με τα γλωσσικά δεδομένα (LING) καθώς και του συνδυασμού τους (FIN+LING).

Στα οικονομικά δεδομένα (FIN), τα μοντέλα **RF** και **XGB** υπερέχουν στην πρόβλεψη απάτης σε σύγκριση με τα υπόλοιπα μοντέλα. Ενώ τα μοντέλα **SVM** και **ANN** έχουν τις χειρότερες αποδόσεις με τη διαφορά το μοντέλο ANN να είναι πιο αποτελεσματικό στο sensitivity. Επίσης λαμβάνοντας υπόψη τις μετρήσεις F1-score, F2-score, το μοντέλο **XGB** υπερέχει έναντι όλων.

Στα γλωσσικά δεδομένα (LING), το μοντέλο **RF** έχει την υψηλότερη απόδοση στα μέτρα AUC, accuracy και F1-score, ενώ ακολουθεί το **SVM** με AUC 74% και accuracy 82%. Όμως συγκρίνοντας τα αποτελέσματα με τον πίνακα FIN, η απόδοση των μοντέλων είναι χειρότερη.

Έπειτα συνδυάστηκαν τα οικονομικά και τα γλωσσικά δεδομένα (FIN + LING). Τα μοντέλα **RF** και **XGB** έχουν την υψηλότερη απόδοση με AUC ~86%. Μάλιστα όσον αφορά το **XGB**, το ποσοστό βελτιώνεται σε σύγκριση με τα δύο προηγούμενα αποτελέσματα, ενώ έχει τη καλύτερη απόδοση στο F2-score (76,87%). Θα πρέπει να

σημειωθεί ότι το AUC όλων των μοντέλων, με εξαίρεση το μοντέλο RF, βελτιώθηκε σε σύγκριση με τα αποτελέσματα (FIN), γεγονός που δείχνει ότι ο συνδυασμός των δεδομένων (FIN + LING) ενισχύει την προγνωστική απόδοση.

Στη συνέχεια, έγινε σύγκριση (α) των δεδομένων κειμένου (TXT) με τα γλωσσικά δεδομένα (LING) και (β) του συνδυασμού δεδομένων (FIN+TXT) με τα (FIN+LING). Για τη σύγκριση προστέθηκαν ακόμη δύο μοντέλα deep learning, το μοντέλο HAN και το μοντέλο GPT-2. Για την αναπαράσταση των δεδομένων, στα παραδοσιακά μοντέλα χρησιμοποιήθηκε η τεχνική TF-IDF, ενώ στα μοντέλα DL pre-trained embeddings.

Σύμφωνα με τα αποτελέσματα, τα δεδομένα κειμένου (TXT) βελτιώνουν όλα τα μοντέλα σε σύγκριση με τα γλωσσικά (LING) και μάλιστα το μοντέλο ANN έχει το μεγαλύτερο delta AUC (0,2). Επίσης αν συγκριθούν τα μοντέλα ANN, GPT-2 και HAN, που όλα αποτελούν αρχιτεκτονικές NN, το μοντέλο ANN έχει την υψηλότερη απόδοση στα μέτρα accuracy (89%), F1-score και F2-score (72% & 89% αντίστοιχα). Επίσης το μοντέλο GPT-2 έχει μία από τις υψηλότερες αποδόσεις στο sensitivity ενώ στο accuracy υστερεί. Επομένως, θα μπορούσε με μια κατάλληλη προσαρμογή του κατωφλιού να ξεπεράσει μοντέλα όπως το LR. Το μοντέλο HAN, παρόλο που έχει χαμηλό accuracy, έχει το υψηλότερο AUC (91,08%) και το δεύτερο καλύτερο sensitivity 80% μετά το μοντέλο SVM καθιστώντας το ένα πολλά υποσχόμενο μοντέλο για τον εντοπισμό απάτης. Γενικά τα αποτελέσματα δείχνουν την υπεροχή των δεδομένων κειμένου έναντι των γλωσσικών (LING) σε όλα τα μοντέλα.

Ακολούθησε η σύγκριση των δεδομένων (FIN+TXT) και (FIN+LING). Σύμφωνα με τα αποτελέσματα, η χρήση των δεδομένων (FIN+TXT) βελτιώνει τις μετρήσεις AUC, specificity και F1-score όλων των μοντέλων και ιδιαίτερα των μοντέλων LR & SVM. Το μοντέλο HAN έχει την καλύτερη απόδοση με AUC 92,64% και ακολουθούν το XGB με 89,83% και το ANN με 89,11%. Επίσης η απόδοση του μοντέλου HAN στο sensitivity είναι η μεγαλύτερη και φθάνει το 90%. Το μοντέλο GPT-2 δεν εμφανίζει σημαντικές διαφορές με την προσθήκη των δεδομένων FIN και μπορεί να οφείλεται στην κακή επιλογή προεκπαιδευμένων ενσωματώσεων. Το μοντέλο ANN έρχεται δεύτερο στη μέτρηση F2-score με 90,55% μετά το μοντέλο XGB και αποτελεί μια καλή εναλλακτική λύση για τις μεθόδους DL.

Πίνακας 2: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση διαφορετικού συνδυασμού δεδομένων

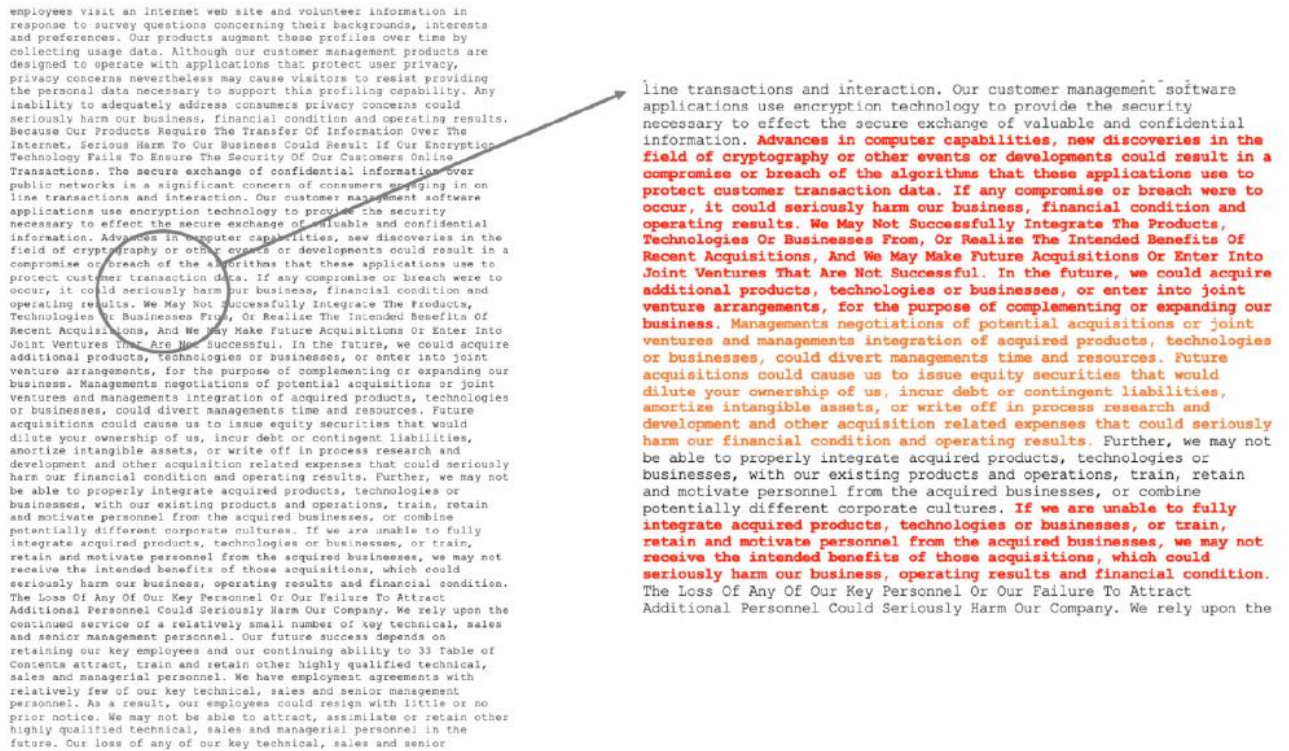
Finance data (FIN)								
	AUC	Sensitivity	Specificity	F1-score	F2-score	Accuracy		
LR	0.7620	0.6833	0.7543	0.4767	0.7480	0.8252		
RF	0.8609	0.7666	0.7889	0.5508	0.7892	0.8653		
SVM	0.7561	0.6166	0.7820	0.4625	0.7595	0.8280		
XGB	0.8470	0.6660	0.8719	0.5839	0.8391	0.8481		
ANN	0.7564	0.7833	0.6574	0.4563	0.6835	0.6790		
Linguistics data (LING)								
	AUC	Sensitivity	Specificity	F1-score	F2-score	Accuracy	Delta AUC	Delta F1
							Comparison to FIN	
LR	0.6719	0.7000	0.6193	0.3962	0.6398	0.8280	-0.0901	-0.0805
RF	0.7713	0.7500	0.7197	0.4839	0.7302	0.8424	-0.0896	-0.0669
SVM	0.7406	0.7000	0.6747	0.4285	0.6857	0.8280	-0.0155	-0.0340
XGB	0.7219	0.3666	0.9446	0.4489	0.8385	0.8338	-0.1251	-0.1350
ANN	0.6782	0.6333	0.6747	0.3958	0.6758	0.6676	-0.0782	-0.0605
Finance data + Linguistics data (FIN + LING)							Comparison to FIN	
LR	0.7682	0.7666	0.6782	0.4623	0.6984	0.8280	0.0062	-0.0144
RF	0.8606	0.7666	0.7543	0.5197	0.7610	0.8567	-0.0003	-0.0311
SVM	0.7973	0.7166	0.7439	0.4858	0.7448	0.8280	0.0567	0.0573
XGB	0.8651	0.8166	0.7543	0.5444	0.7687	0.8653	0.0181	-0.0395
ANN	0.7733	0.8333	0.6228	0.4566	0.6614	0.6590	0.0169	0.0003
Text data, TF-IDF (TXT)								
	AUC	Sensitivity	Specificity	F1-score	F2-score	Accuracy	Delta AUC	Delta F1
							Comparison to LING	
LR	0.8371	0.7333	0.8269	0.5714	0.8145	0.8281	0.1652	0.1752
RF	0.8740	0.7166	0.9377	0.7107	0.8998	0.8681	0.1027	0.2268
SVM	0.8836	0.8382	0.7544	0.5876	0.7731	0.8796	0.1275	0.1251
XGB	0.8785	0.7660	0.8581	0.6258	0.8451	0.8853	0.1566	0.1769
ANN	0.8829	0.7121	0.9434	0.7286	0.8993	0.8990	0.2047	0.3328
HAN	0.9108	0.8000	0.8896	0.5744	0.7982	0.8457		
GPT-2 + Attn	0.7729	0.7619	0.6697	0.4423	0.6905	0.6484		
Finance data + Text data, TF-IDF (FIN + TXT)								
	AUC	Sensitivity	Specificity	F1-score	F2-score	Accuracy	Delta AUC	Delta F1
							Comparison to FIN + LING	
LR	0.8598	0.7833	0.7854	0.5562	0.7890	0.8424	0.0916	-0.0795
RF	0.8797	0.6660	0.9550	0.7079	0.9043	0.8739	0.0191	-0.1571
SVM	0.8902	0.7833	0.8961	0.6861	0.8784	0.8280	0.0929	-0.2576
XGB	0.8983	0.7000	0.9653	0.7500	0.9187	0.9083	0.0332	-0.1661
ANN	0.8911	0.7460	0.9405	0.7401	0.9055	0.9054	0.1178	-0.2838
HAN	0.9264	0.9000	0.8206	0.6506	0.8361	0.8457		
GPT-2 + Attn	0.7776	0.7678	0.6791	0.4455	0.6991	0.6934		

Πηγή: (Craja κ.ά., 2020)

Τέλος, έγινε σύγκριση των λέξεων “red-flags” του παραδοσιακού μοντέλου RF και του προτεινόμενου μοντέλου HAN από την οποία διαπιστώθηκε ότι 15 λέξεις είναι κοινές και σημαντικές για την ένδειξη απάτης. Δεδομένου όμως ότι οι λέξεις μπορεί να χειραγωγηθούν από τη διοίκηση, προτείνεται η χρήση των δεικτών σε επίπεδο πρότασης παρά σε επίπεδο λέξης.

Σε επίπεδο πρότασης, εξάγονται τα attention weights 200 εκθέσεων απάτης και φιλτράρονται οι δέκα σημαντικότερες προτάσεις ανά έκθεση. Το μέσο βάρος μιας πρότασης “red-flag” είναι 0,05 με μέγιστο το 0,61. Οι προτάσεις χωρίστηκαν σε τρεις ομάδες ανάλογα με τα βάρη τους. Έτσι προτάσεις με βάρη μεγαλύτερα από 0,67 χαρακτηρίζονται ως “extra important”, με βάρη από 0,04-0,67 ως “important” και μεταξύ 0,022-0,04 ως “noteworthy” και επισημαίνονται με διαφορετικό χρώμα για

την οπτική καθοδήγηση και διευκόλυνση των ελεγκτών κατά τη διενέργεια των διαδικασιών ελέγχου (βλ. εικόνα 9).



Εικόνα 9: Κείμενο από την ενότητα MD&A (στα αριστερά), εξαγωγή των "red -flag" προτάσεων (στα δεξιά)

Πηγή: (Craja κ.ά., 2020)

Τα αποτελέσματα έδειξαν ότι οι πληροφορίες κειμένου που εξάγονται μέσω του μοντέλου HAN ενισχύσουν την προγνωστική ακρίβεια, το νέο μοντέλο HAN είναι κατάλληλο και ικανό να εντοπίσει σωστά τις περιπτώσεις απάτης σε σχέση με τα μοντέλα που χρησιμοποιούν την τεχνική BOW. Επίσης οι δείκτες "red-flag" σε επίπεδο πρότασης βελτιώνουν τον χρόνο και την ακρίβεια των διαδικασιών ελέγχου των ελεγκτών και υποστηρίζουν τη διαδικασία λήψης αποφάσεων των ενδιαφερόμενων μερών.

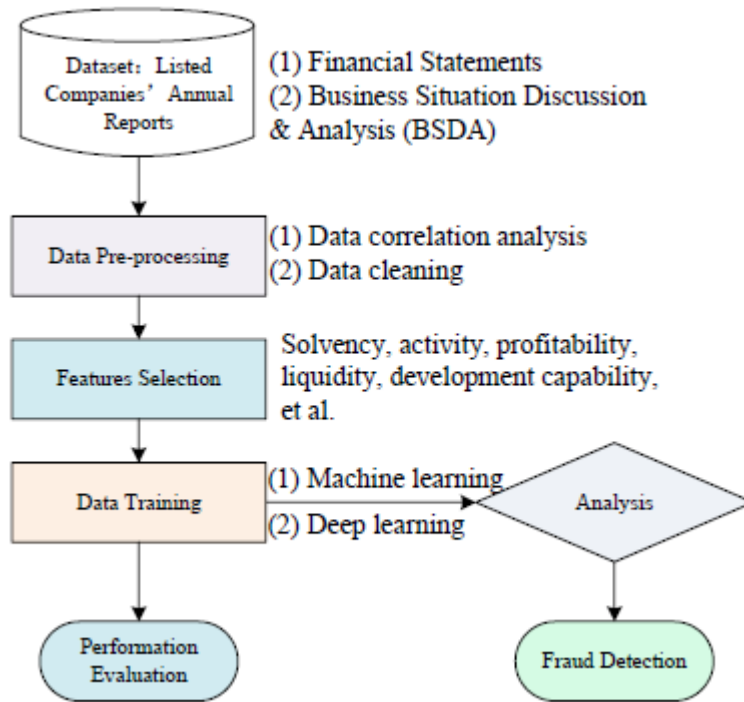
Οι ερευνητές επισημαίνουν ότι οι μελλοντικές έρευνες θα πρέπει να επικεντρωθούν στην ανάπτυξη μεθόδων που συνδυάζουν την ανάλυση οικονομικών δεδομένων και δεδομένων κειμένου για τον εντοπισμό απάτης στις οικονομικές καταστάσεις δεδομένου ότι ελάχιστη έρευνα έχει διεξαχθεί μέχρι σήμερα.

4.2 An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning

Οι Xiuguo και Shengyong (2022) ασχολήθηκαν με την ανάπτυξη ενός συστήματος ανίχνευσης χρηματοοικονομικής απάτης με τη χρήση μοντέλων deep

learning συνδυάζοντας αριθμητικά δεδομένα και δεδομένα κειμένου των ετήσιων εκθέσεων κινέζικων εισηγμένων εταιρειών.

Η ροή των διαδικασιών της έρευνας που ακολουθήθηκε απεικονίζεται παρακάτω (βλ. εικόνα 10):



Εικόνα 10: Η ροή των διαδικασιών της έρευνας

Πηγή: (Xiuguo & Shengyong, 2022)

Τα δεδομένα συλλέχθηκαν από τη βάση δεδομένων China Stock Market & Accounting Research και αφορούν 1068 εισηγμένες κινέζικες εταιρείες που δραστηριοποιούνται σε διάφορους τομείς (βλ. πίνακα 3).

Πίνακας 3: Τομείς δραστηριότητας εισηγμένων κινέζικων εταιρειών

Sector	Number of Companies
Industrial Goods & Services	646
Wholesale and retail	105
Information technology	87
Construction & materials	50
Health care	6
Education	14
Financial service	70
Accommodation & catering	18
Water & power	20
Science	16
Comprehensive	36
Total	1068

Πηγή: (Xiuguo & Shengyong, 2022)

Το σύνολο δεδομένων αντλήθηκε από τις οικονομικές καταστάσεις και την ενότητα MD&A των ετήσιων εκθέσεων. Συνολικά αποτελείται από 5130 ετήσιες εκθέσεις από τις οποίες οι 4886 είναι μη δόλιες (normal) και οι υπόλοιπες 244 δόλιες (special treat – ST) και αφορά το χρονικό διάστημα 2016-2020 (βλ. πίνακα 4).

Πίνακας 4: Αριθμός δόλιων και μη δόλιων ετήσιων εκθέσεων εισηγμένων εταιρειών

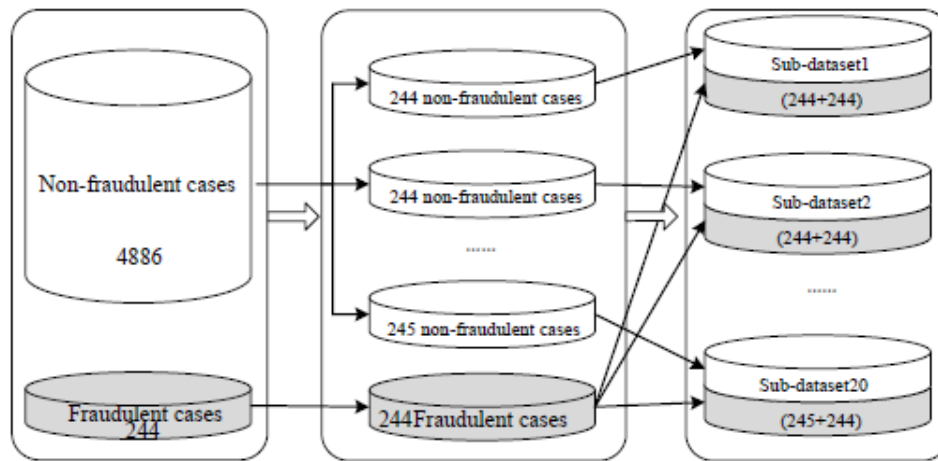
Item	Number of Training	Number of Testing	Total
Number of ST Companies	206	38	244
Number of Normal Companies	4126	760	4886
Total	4332	798	5130

Πηγή: (Xiuguo & Shengyong, 2022)

Ακολούθησε η **προεπεξεργασία** των αριθμητικών δεδομένων για την αντιμετώπιση των προβλημάτων της υψηλής διάστασης, της επανάληψης, του θορύβου και των χαμένων τιμών και εφαρμόστηκαν μέθοδοι κανονικοποίησης, ώστε να διασφαλιστεί η αξιοπιστία των δεδομένων ανάλυσης.

Επίσης τα κινέζικα δεδομένα κειμένου είναι αδόμητα δεδομένα και δεν μπορούν να χρησιμοποιηθούν απευθείας ως είσοδος στο μοντέλο. Έτσι μετατράπηκαν σε κατάλληλη μορφή, ώστε να διευκολυνθεί η αλγοριθμική τους επεξεργασία. Για τον σκοπό αυτό, εφαρμόστηκαν δύο διαδικασίες (α) η τμηματοποίηση των κινέζικων λέξεων κατά την οποία χρησιμοποιήθηκε το Jieba package και (2) ο υπολογισμός του word vector, ώστε οι λέξεις να πάρουν αριθμητική μορφή και να τροφοδοτηθούν ως είσοδος στα μοντέλα deep learning. Το word vector model σχεδιάστηκε από το Institute of Chinese Scientific Space και το εκπαιδευτικό του εργαλείο είναι το Word2Vec.

Επίσης για την αντιμετώπιση του προβλήματος της ανισορροπίας των δεδομένων, το σύνολο δεδομένων των 4886 μη δόλιων ετήσιων εκθέσεων χωρίστηκε σε 20 ομάδες όπου κάθε μία περιείχε 244 (ή 245) περιπτώσεις. Έπειτα κάθε ομάδα συνδυάστηκε με το σύνολο δεδομένων των 244 δόλιων ετήσιων εκθέσεων και δημιουργήθηκαν τελικά 20 ισορροπημένα υποσύνολα δεδομένων με 488 (ή 489) περιπτώσεις η κάθε μία. Στη συνέχεια, πραγματοποιήθηκαν πολλαπλά πειράματα με τα υποσύνολα δεδομένων, χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα, ώστε να υπολογιστούν τα πιο αξιόπιστα αποτελέσματα. Τα αποτελέσματα έδειξαν ότι με αυτό τον τρόπο δειγματοληψίας, η συνολική απόδοση αυξήθηκε (βλ. εικόνα 11). Για τη βελτιστοποίηση των παραμέτρων ελέγχου χρησιμοποιήθηκε η συνάρτηση απώλειας Focal Loss.



Εικόνα 11: Προεπεξεργασία δεδομένων για την αντιμετώπιση της ανισοροπίας των δεδομένων

Πηγή: (Xiuguo & Shengyong, 2022)

Επίσης έγινε επιλογή χαρακτηριστικών και επιλέχθηκαν συνολικά (α) 58 χρηματοοικονομικές μεταβλητές που αφορούν τη φερεγγυότητα, δραστηριότητα, κερδοφορία, EVA, ρευστότητα, ικανότητα ανάπτυξης, επίπεδο κινδύνου, αναλογία δομής, δείκτες ανά μετοχή και αγοραία αξία κάθε επιχείρησης (βλ. πίνακα 5)

Πίνακας 5: Οι χρηματοοικονομικές μεταβλητές

Dimension	Second-level variables	Dimension	Second-level variables
Solvency	X1: Current ratio	EVA	X30: The sale profit ratio
	X2: Quick ratio		X31: Working capital turnover ratio
	X3: Total debt	Liquidity	X32: Net Earnings and cash flows
	X4: The logarithm of total debt		X33: Working capital
	X5: Equity		X34: Working Capital/total assets
	X6: Debt to equity		X35: Current assets/current liabilities
	X7: Total debt/total assets		X36: Current assets/total assets
	X8: Long term debt/total assets		X37: Cash and deposits/total assets
	X9: Short term debt/total assets		X38: Quick assets/current liabilities
Activity	X10: Account receivable/sales	X39: Cash flow/total debt	
	X11: Inventory/sales	X40: Cash flow/current debt	
	X12: Inventory/total assets	X41: Cash flow/cash dividend	
	X13: Inventory turnover	X42: Cash flow/equity	
	X14: Sales growth	Development	X43: Capital accumulation ratio
	X15: Sales		X44: Asset inflation and incremental ratio
X16: Inventory	X45: Revenue growth rate		
Profitability	X46: Profit growth rate		

	X18: Assets return ratio		X47: Total assets growth rate
	X19: Net margins of total assets	Risk	X48: Financial leverage
	X20: Gross profit/total assets		X49: Operating leverage
	X21: Net profit/total assets		X50: Liquidity ratios
	X22: Net profit/sales	Structure	X51: Cash asset ratio
	X23: Net income/fixed assets		X52: Equity/total assets
	X24: Earnings before interest and taxes		X53: Cash dividends per share
	X25: Ebit/total assets	per share	X54: Total income per share
	X26: Z-score		X55: Revenues per share
	X27: Net profit after tax		X56: Price-earnings ratio
	X28: Cash and deposit/current assets	Market value	X57: Price-book ratio
	X29: Return on invested capital		X58: Tangible assets ratio

Πηγή: (Xiuguo & Shengyong, 2022)

και (β) 16 μη χρηματοοικονομικές μεταβλητές (γνωστές και ως μεταβλητές εταιρικής διακυβέρνησης) που αφορούν την ιδιοκτησιακή και διαχειριστική δομή καθώς και τη γνώμη του ελεγκτή (βλ. πίνακα 6).

Πίνακας 6: Οι μη χρηματοοικονομικές μεταβλητές

Dimension	Second-level indices	Dimension	Second-level indices
Ownership structure	Y1: Ownership concentration index CR1	Management structure	Y9: The number of employees
	Y2: Ownership concentration index CR5		Y10: Supervisors' size
	Y3: Index Z		Y11: The number of senior supervisors
	Y4: Index S		Y12: Total annual salary of director, supervisors and senior supervisor
	Y5: Relations among top 10 shareholder		Y13: Total annual salary of top 3 director, supervisors and senior supervisor
Management structure	Y6: Chairman and general manager holding a concurrent post		Y14: Total annual salary of top 3 directors
	Y7: Number of directors (chairman included)	Y15: Total annual salary of top 3 senior supervisors	
	Y8: The proportion of the independent board member	Audit opinion	Y16: A standard and unqualified auditor's report

Πηγή: (Xiuguo & Shengyong, 2022)

Έπειτα το σύνολο δεδομένων **χωρίστηκε τυχαία** σε training dataset (70%) και test dataset (30%) και πραγματοποιήθηκαν έξι πειράματα κατά τα οποία τα δεδομένα χρησιμοποιήθηκαν ως είσοδος σε μοντέλα deep learning και machine learning. Για την αξιολόγηση της απόδοσης των μοντέλων, λήφθηκαν υπόψη τα μέτρα AUC (Area Under Curve), Sensitivity, Specificity, F1-score, F2-score και Accuracy.

Στα δύο πρώτα πειράματα έγινε χρήση μόνο των οικονομικών μεταβλητών και εφαρμόστηκαν τα μοντέλα CNN, Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boosting (XGB), Artificial Neural Network (ANN) και LR.

Συγκεκριμένα, κατά το **πρώτο πείραμα**, χρησιμοποιήθηκαν όλες οι οικονομικές μεταβλητές χωρίς την επιλογή χαρακτηριστικών. Σύμφωνα με τον Πίνακα 7, τα αποτελέσματα έδειξαν ότι τα μοντέλα **CNN** και **RF** υπερέχουν σε σύγκριση με τα υπόλοιπα μοντέλα. Μάλιστα το μοντέλο **CNN** είχε τις μεγαλύτερες αποδόσεις στα μέτρα AUC (0,8851), specificity (0,8936), F1-score (0,6057), F2-score (0,8369), ενώ το RF στο sensitivity (0,7748) και accuracy (0,8774).

Πίνακας 7: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση όλων των οικονομικών μεταβλητών

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
LR	0.7251	0.6925	0.7591	0.4873	0.7594	0.8109
RF	0.8106	0.7748	0.8006	0.5862	0.7922	0.8774
SVM	0.7754	0.5932	0.7765	0.4781	0.7741	0.8392
XGB	0.8096	0.6963	0.8695	0.6132	0.8263	0.8418
ANN	0.7364	0.7654	0.7125	0.5451	0.7125	0.7456
CNN	0.8851	0.7852	0.8936	0.6057	0.8369	0.8419

Πηγή: (Xiuguo & Shengyong, 2022)

Στο **δεύτερο πείραμα**, χρησιμοποιήθηκαν οι οικονομικές μεταβλητές που προέκυψαν μετά την επιλογή χαρακτηριστικών (βλ. πίνακες 5 και 6). Τα αποτελέσματα του πειράματος έδειξαν ότι υπάρχει μια μικρή βελτίωση σε σχέση με το προηγούμενο πείραμα, επομένως εξάγεται το συμπέρασμα ότι η παράλειψη χρηματοοικονομικών μεταβλητών από το αρχικό σύνολο δεδομένων δεν έχει αρνητική επίδραση στην ικανότητα πρόβλεψης των μοντέλων. Αναλυτικότερα, το μοντέλο **CNN** είχε τις μεγαλύτερες αποδόσεις στο AUC (0,8920), specificity (0,9082), F1-score (0,6094), F2-score (0,8464), ενώ το **RF** στο sensitivity (0,7864) και accuracy (0,8784) (βλ. πίνακα 8).

Πίνακας 8: Τα αποτελέσματα απόδοσης των μοντέλων με τη χρήση των επιλεγμένων χρηματοοικονομικών μεταβλητών

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
LR	0.7816	0.7024	0.7683	0.4804	0.7624	0.8351
RF	0.8812	0.7864	0.8012	0.5603	0.8091	0.8784
SVM	0.7831	0.6288	0.7808	0.4836	0.7834	0.8681
XGB	0.8235	0.6924	0.8908	0.6067	0.8394	0.8472
ANN	0.7654	0.7684	0.7002	0.5091	0.7035	0.7012
CNN	0.8920	0.7946	0.9082	0.6094	0.8464	0.8647

Πηγή: (Xiuguo & Shengyong, 2022)

Στα **δύο επόμενα πειράματα** έγινε χρήση μόνο των δεδομένων κειμένου. Το μήκος των κινέζικων δεδομένων κειμένου ορίστηκε σε 600 και 1000 λέξεις αντίστοιχα. Έτσι κείμενα που ξεπερνούσαν το όριο των λέξεων περικόπηκαν, ενώ εκείνα που είχαν λιγότερες λέξεις συμπληρώθηκαν. Τα μοντέλα που χρησιμοποιήθηκαν ήταν τα CNN, LSTM, GRU, Transformer, SVM, XGB και ANN. Τα αποτελέσματα των πειραμάτων απεικονίζονται στους πίνακες 9 και 10.

Σύμφωνα με τα αποτελέσματα του πίνακα 9 όπου χρησιμοποιήθηκαν οι 600 λέξεις κειμένου, τα μοντέλα deep learning υπερσχύουν σε σύγκριση με τα μοντέλα machine learning σημειώνοντας καλύτερες αποδόσεις. Όμως σε σύγκριση με τα προηγούμενα πειράματα (που έκαναν χρήση μόνο των χρηματοοικονομικών μεταβλητών), η απόδοση είναι χειρότερη σε όλα τα μοντέλα που σημαίνει ότι η πρόβλεψη απάτης με τη χρήση μόνο δεδομένων κειμένου δεν έχει καλή απόδοση και είναι ανεξάρτητη από το μοντέλο.

Πίνακας 9: Αποτελέσματα απόδοσης μοντέλων με τη χρήση δεδομένων κειμένου 600 λέξεων

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
SVM	0.6123	0.6004	0.7245	0.4128	0.7125	0.7109
XGB	0.6584	0.6122	0.7341	0.5147	0.7249	0.7006
ANN	0.7006	0.6832	0.7429	0.4682	0.7136	0.7331
CNN	0.8021	0.7009	0.7940	0.5739	0.7364	0.7425
LSTM(128)	0.8351	0.7154	0.8001	0.6034	0.7458	0.7354
LSTM(256)	0.8465	0.7266	0.7946	0.6551	0.7631	0.7369
GRU(128)	0.8157	0.7241	0.8001	0.6187	0.7862	0.7547
GRU(256)	0.8584	0.7457	0.8124	0.6544	0.7952	0.7538
Transformer	0.8321	0.7654	0.8117	0.6420	0.8012	0.7855

Πηγή: (Xiuguo & Shengyong, 2022)

Επίσης σύμφωνα με τα αποτελέσματα του πίνακα 10 όπου χρησιμοποιήθηκαν οι 1000 λέξεις κειμένου, το μοντέλο **Transformer** σημείωσε τις καλύτερες αποδόσεις με το μοντέλο **GRU** (256) να ακολουθεί, ενώ το **LSTM** (128 & 256) είχε χαμηλές επιδόσεις. Επίσης σε σύγκριση με τα αποτελέσματα του προηγούμενου πειράματος (βλ. πίνακα 9), τα αποτελέσματα είναι βελτιωμένα.

Πίνακας 10: Αποτελέσματα απόδοσης μοντέλων με τη χρήση δεδομένων κειμένου 1000 λέξεων

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
SVM	0.6846	0.6154	0.7261	0.5124	0.7296	0.7315
XGB	0.7015	0.6351	0.7352	0.5241	0.7364	0.7415
ANN	0.7674	0.7141	0.7424	0.4957	0.7347	0.7345
CNN	0.8125	0.7251	0.8014	0.5748	0.7651	0.7498
LSTM(128)	0.8241	0.7554	0.8136	0.5941	0.7716	0.7598
LSTM(256)	0.8398	0.8014	0.8247	0.6049	0.7808	0.7629
GRU(128)	0.8256	0.8157	0.8198	0.6365	0.8041	0.7937
GRU(256)	0.8549	0.8396	0.8584	0.6541	0.8165	0.7848
Transformer	0.8664	0.8487	0.8597	0.6874	0.8084	0.7921

Πηγή: (Xiuguo & Shengyong, 2022)

Τέλος, στα δύο τελευταία πειράματα έγινε χρήση τόσο των οικονομικών δεδομένων όσο και των δεδομένων κειμένου 600 και 1000 λέξεων αντίστοιχα και εφαρμόστηκαν τα μοντέλα CNN, LSTM, GRU, Transformer, SVM, XGB και ANN.

Αναλυτικότερα, σύμφωνα με τα αποτελέσματα του πίνακα 11, υπάρχει βελτιωμένη απόδοση σε σχέση με τα πειράματα που χρησιμοποίησαν αποκλειστικά και μόνο οικονομικά δεδομένα ή δεδομένα κειμένου και μάλιστα ιδιαίτερα στα μοντέλα **SVM** και **GRU**. Το **GRU(256)** έχει την καλύτερη απόδοση με AUC 94,49% και ακολουθούν τα **LSTM(256)** και **Transformer** με AUC 93,98% και 93,64% αντίστοιχα που σημαίνει ότι τα **LSTM** και **GRU** έχουν την ικανότητά να ενσωματώνουν και να εξάγουν τιμές από διαφοροποιημένα δεδομένα. Επίσης αν συγκρίνει κανείς τα μοντέλα **LSTM (256)**, **GRU (256)** και **Transformer** διαπιστώνει ότι το Transformer δεν εμφανίζει πολύ μεγάλη βελτίωση που σημαίνει ότι οι τεχνολογίες αιχμής δεν εγγυώνται πάντα τις ανώτερες αποδόσεις.

Πίνακας 11: Αποτελέσματα απόδοσης μοντέλων με τον συνδυασμό χρηματοοικονομικών δεδομένων και δεδομένων κειμένου 600 λέξεων

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
SVM	0.8536	0.6843	0.8261	0.5645	0.8254	0.8699
XGB	0.8751	0.7045	0.9087	0.6582	0.8669	0.8963
ANN	0.8067	0.7741	0.8824	0.6615	0.7916	0.7743
CNN	0.9071	0.8080	0.9234	0.7864	0.8608	0.8924
LSTM(128)	0.9218	0.8454	0.9236	0.8041	0.8216	0.9098
LSTM(256)	0.9398	0.9014	0.9147	0.8749	0.8408	0.9329
GRU(128)	0.9356	0.8857	0.9098	0.8865	0.8641	0.9237
GRU(256)	0.9449	0.8996	0.9184	0.8941	0.9165	0.9448
Transformer	0.9364	0.8887	0.9297	0.9374	0.9084	0.9421

Πηγή: (Xiuguo & Shengyong, 2022)

Επίσης ο πίνακας 12 εμφανίζει βελτίωση σε σύγκριση με τον πίνακα 11, επομένως η αύξηση του μεγέθους του κειμένου ενισχύει την απόδοση πρόβλεψης. Το **GRU** με 256 νευρώνες είχε τις υψηλότερες επιδόσεις στα μέτρα AUC (0,9581), specificity (0,9634) και F1-score (0,9345) καθιστώντας το ένα πολλά υποσχόμενο μοντέλο για τον εντοπισμό απάτης, ενώ το μοντέλο **LSTM** (256) είχε τις υψηλότερες επιδόσεις στα sensitivity (0,8921) και accuracy (0,9498).

Πίνακας 12: Αποτελέσματα απόδοσης μοντέλων με τον συνδυασμό χρηματοοικονομικών δεδομένων και δεδομένων κειμένου 1000 λέξεων

Classifier	AUC	Sensitivity	Specificity	F1-Score	F2-Score	Accuracy
SVM	0.8902	0.7856	0.9012	0.65019	0.8582	0.8109
XGB	0.8836	0.6994	0.9041	0.7102	0.8784	0.9006
ANN	0.8806	0.7610	0.9113	0.6808	0.8947	0.9031
CNN	0.9312	0.8304	0.9351	0.8077	0.9029	0.9165
LSTM(128)	0.9344	0.8528	0.9346	0.8315	0.8307	0.9254
LSTM(256)	0.9367	0.8921	0.9324	0.8689	0.8505	0.9498
GRU(128)	0.9418	0.8471	0.9548	0.9326	0.9418	0.9317
GRU(256)	0.9581	0.8595	0.9634	0.9345	0.9287	0.9462
Transformer	0.9501	0.8854	0.9418	0.9236	0.9148	0.9445

Πηγή: (Xiuguo & Shengyong, 2022)

Σύμφωνα με τα παραπάνω διαπιστώνεται ότι οι πληροφορίες της ενότητας MD&A σε συνδυασμό με τα οικονομικά δεδομένα των ετήσιων εκθέσεων, βελτιώνουν σημαντικά την απόδοση και την προβλεπτική ικανότητα των μοντέλων deep learning σε σχέση με τις υπόλοιπες μεθόδους με τις οποίες συγκρίθηκαν, υποστηρίζοντας με αποτελεσματικότητα τις διαδικασίες λήψης αποφάσεων όλων των ενδιαφερόμενων μερών.

Στο μέλλον προτείνεται (α) η διεύρυνση της χρονικής περιόδου συλλογής των δεδομένων, (β) η χρήση δεδομένων και από άλλες πηγές (π.χ. ανακοινώσεις εισηγμένων εταιρειών), (γ) η συλλογή δεδομένων για όλες τις επιχειρήσεις και όχι αποκλειστικά και μόνο των εισηγμένων και επίσης (δ) η βελτιστοποίηση των αλγορίθμων εξόρυξης κειμένου για την ανάλυση συναισθήματος με σκοπό την αποτελεσματικότερη πρόβλεψη απάτης.

4.3 *Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry*

Ο Jan (2021a) ασχολήθηκε με την ανάπτυξη ενός μοντέλου deep learning για την αποτελεσματική ανίχνευση της απάτης των οικονομικών καταστάσεων με σκοπό την διατήρηση της βιώσιμης ανάπτυξης των κεφαλαιαγορών και τη μείωση των ζημιών που μπορεί να προκληθούν από μια ενδεχόμενη απάτη. Για την κατασκευή του μοντέλου, έγινε χρήση των αλγορίθμων RNN και LSTM οι οποίοι έχουν τη δυνατότητα επεξεργασίας μεγάλου όγκου δεδομένων.

Τα δεδομένα συλλέχθηκαν από το Taiwan Economic Journal (TEJ) και προέρχονται από συνολικά 153 εισηγμένες εταιρίες στο χρηματιστήριο της Ταϊβάν για το χρονικό διάστημα 2001-2019. Από τις 153 εταιρίες, οι 51 αναφέρουν απάτη στις οικονομικές καταστάσεις και οι 102 δεν αναφέρουν. Η αναλογία των εταιρειών είναι 1:2 και η κατανομή τους ανά κλάδο απεικονίζεται στον παρακάτω πίνακα (βλ. πίνακα 13):

Πίνακας 13: Κλάδοι δραστηριότητας των εταιρειών

Industry	FSF	Non-FSF	Total
Food	3	6	9
Textile	3	6	9
Electric machinery	3	6	9
Appliances and electric cables	3	6	9
Medicine and biotech	2	4	6
Steel	2	4	6
Semiconductor	5	10	15
Computers and peripherals	2	4	6
Optical electronics	6	12	18
Telecommunication and networking	2	4	6
Electronic components	4	8	12
Consumer electronics channels	2	4	6
Information services	5	10	15
Other electronic sectors	4	8	12
Building materials and construction	1	2	3
Cultural and creative industry	1	2	3
Others	3	6	9
TOTAL	51	102	153

Πηγή: (Jan, 2021a)

Κατά τη μελέτη, εξετάστηκαν συνολικά 18 μεταβλητές από τις οποίες οι 14 είναι χρηματοοικονομικές και οι 4 μη χρηματοοικονομικές (ή μεταβλητές εταιρικής διακυβέρνησης) και συνοψίζονται στον κάτωθι πίνακα (βλ. πίνακα 14):

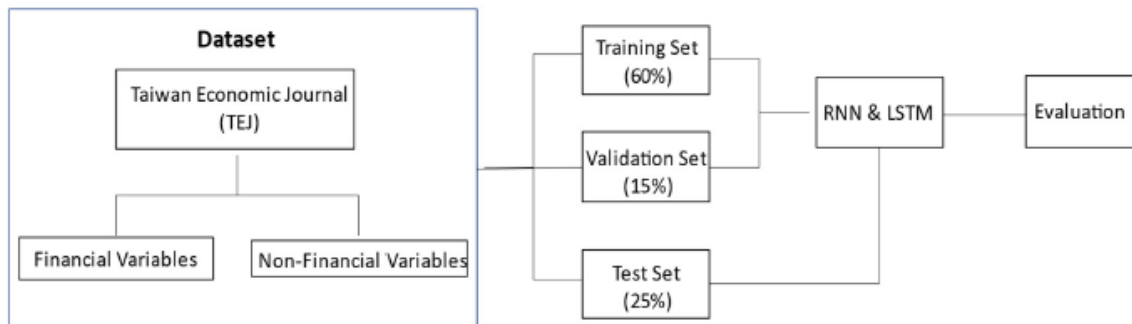
Πίνακας 14: Οι μεταβλητές της έρευνας

Code	Variable Name	Variable Definition or Calculation
X01	Liability (natural logarithm)	$\text{Ln}(\text{total liabilities})$
X02	Current ratio	$\text{Current assets} \div \text{current liabilities}$
X03	Quick ratio	$\text{Quick assets} \div \text{current liabilities}$
X04	Liability/equity ratio	$\text{Liabilities} \div \text{shareholders' equity}$
X05	Liability ratio	$\text{Liabilities} \div \text{assets}$
X06	Return on equity (ROE)	$\text{Net income} \div \text{average shareholders' equity}$
X07	Return on assets (ROA)	$[\text{Net income} + \text{interest expense} \times (1 - \text{tax rate})] \div \text{average total assets}$
X08	Long-term capital adequacy rate	$(\text{Long-term liabilities} + \text{shareholders' equity}) \div \text{Plant, property, and equipment}$
X09	Cash as a percentage of assets	$\text{Cash} \div \text{assets}$
X10	Net income/assets	$\text{Net income} \div \text{assets}$
X11	Gross margin	$\text{Gross profit} \div \text{revenue}$
X12	Non-current liabilities/assets	$\text{Non-current liabilities} \div \text{total assets}$
X13	Net cash flows from operating activities/revenue	$\text{Net cash flows from operating activities} \div \text{revenue}$
X14	Net loss or not	1 for net loss and 0 for net income
X15	Percentage of managing directors	$\text{No. of managing directors} \div \text{No. of board seats}$
X16	Percentage of shares held by major shareholders	$\text{No. of shares held by major shareholders} \div \text{No. of ordinary shares outstanding at the end of the period}$
X17	Percentage of shares held by directors and supervisors	$\text{No. of shares held by directors and supervisors} \div \text{No. of ordinary shares outstanding at the end of the period}$
X18	Percentage of shares pledged by directors and supervisors	$\text{No. of shares pledged by directors and supervisors} \div \text{No. of shares held by directors and supervisors}$

Πηγή: (Jan, 2021a)

Για την κατασκευή των μοντέλων, έγινε χρήση του λογισμικού TensorFlow και το σύνολο δεδομένων χωρίστηκε τυχαία. Το 60% των δεδομένων ήταν το training dataset, το 15% το validation dataset και το υπόλοιπο 25% το test dataset.

Η ροή των διαδικασιών της έρευνας είναι ως εξής (βλ. εικόνα 12):



Εικόνα 12: Η ροή των διαδικασιών της έρευνας

Πηγή: (Jan, 2021a)

Οι μεταβλητές κανονικοποιούνται και τυποποιούνται, ώστε όλες οι τιμές να είναι μεταξύ 0 και 1 προκειμένου να επιταχυνθεί το convergence speed και να βελτιωθεί η ακρίβεια ανίχνευσης του μοντέλου.

Για την προσαρμογή των παραμέτρων, εφαρμόστηκε ο βελτιστοποιητής Adaptive Moment Estimation (Adam). Οι βέλτιστες παράμετροι που ορίστηκαν είναι: learning rate = 0.001, beta 1 = 0.9, beta 2 = 0.999, epsilon = 1×10^{-7} , batch size = 3, epochs = 250, validation_split = 0.2, dropout = 0.25, activation = Sigmoid and ReLU.

Τέλος, για τη βελτιστοποίηση του μοντέλου χρησιμοποιήθηκε η συνάρτηση loss function, ενώ για την αξιολόγηση της απόδοσης χρησιμοποιήθηκαν τα μέτρα accuracy,

precision, sensitivity (recall), specificity, F1 score, error rates Type I and Type II και η καμπύλη ROC/AUC.

Πιο συγκεκριμένα, για το **μοντέλο RNN**, η εκπαίδευση των 250 εποχών είχε διάρκεια μόνο 0,3 δευτερόλεπτα.

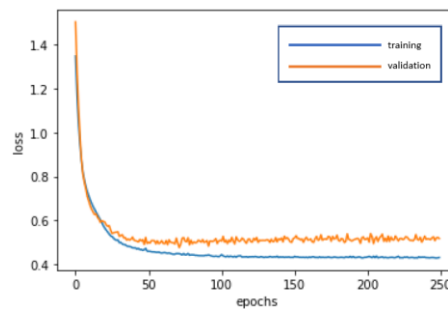
Σύμφωνα με τα αποτελέσματα του πίνακα 15, η ακρίβεια στο training dataset είναι 87,10%, στο validation dataset 85,71% και στο test dataset 87,18%. Η διαφορά των ποσοστών του training dataset και validation dataset είναι μικρή, επομένως το μοντέλο θεωρείται ότι είναι σταθερό και έχει καλή προσαρμογή.

Πίνακας 15: Accuracy μοντέλου RNN

Model	Training Dataset	Validation Dataset	Test Dataset	Average
RNN	87.10%	85.71%	87.18%	86.67%

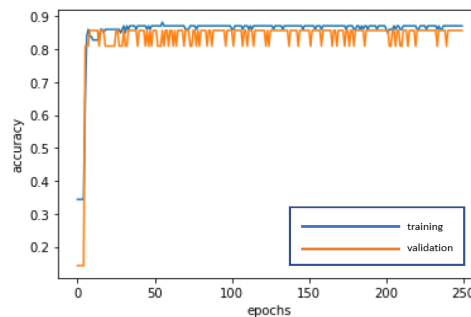
Πηγή: (Jan, 2021a)

Επίσης, στις εικόνες 13 και 14 απεικονίζονται τα γραφήματα loss function και accuracy. Τέλος, τα ποσοστά σφάλματος I και II του test dataset είναι 5,13% και 7,69% αντίστοιχα και δεν θεωρούνται υψηλά.



Εικόνα 13 : Γράφημα loss function

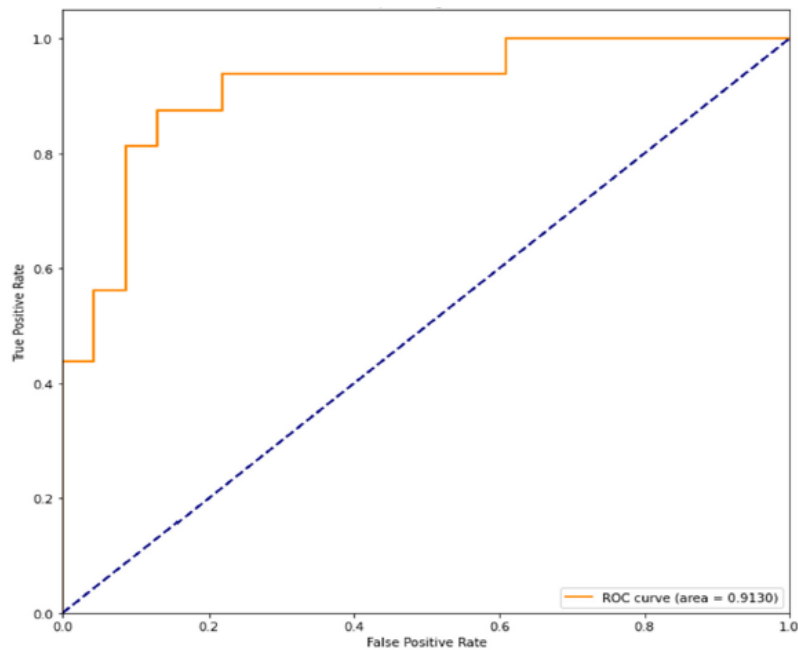
Πηγή: (Jan, 2021a)



Εικόνα 14: Γράφημα accuracy

Πηγή: (Jan, 2021a)

Η τιμή AUC είναι 0,913 και δείχνει υψηλή ακρίβεια, αφού η τιμή της είναι πολύ κοντά στην τιμή 1 (βλ. εικόνα 15).



Εικόνα 15: AUC του μοντέλου RNN

Πηγή: (Jan, 2021a)

Επίσης, για το **μοντέλο LSTM**, η εκπαίδευση των 250 εποχών είχε διάρκεια 1,5 δευτερόλεπτα.

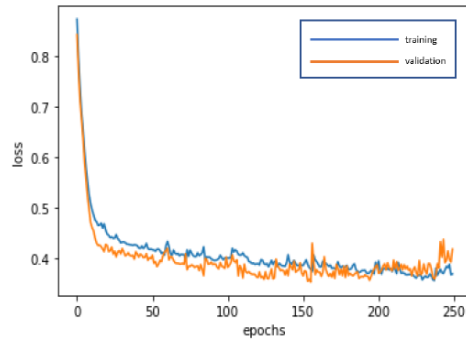
Σύμφωνα με τα αποτελέσματα του πίνακα 16, η ακρίβεια στο training dataset είναι 88,17%, στο validation dataset 85,71% και στο test dataset 94,88%, επομένως το μοντέλο είναι πολύ σταθερό και έχει καλή προσαρμογή.

Πίνακας 16: Accuracy μοντέλου LSTM

Model	Training Dataset	Validation Dataset	Test Dataset	Average
LSTM	88.17%	85.71%	94.88%	89.59%

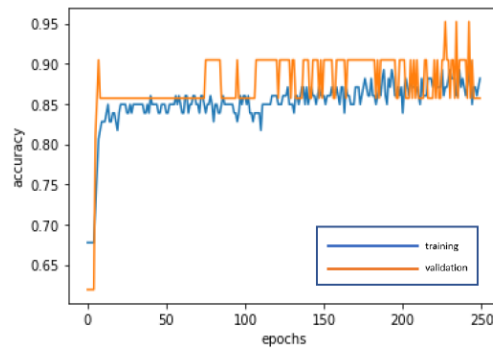
Πηγή: (Jan, 2021a)

Στις εικόνες 16 και 17 απεικονίζονται τα γραφήματα loss function και accuracy κατά τη διαδικασία μοντελοποίησης. Επίσης, τα ποσοστά σφάλματος τύπου I και II του test dataset είναι 2,56% και θεωρούνται αρκετά χαμηλά.



Εικόνα 16: Γράφημα loss function

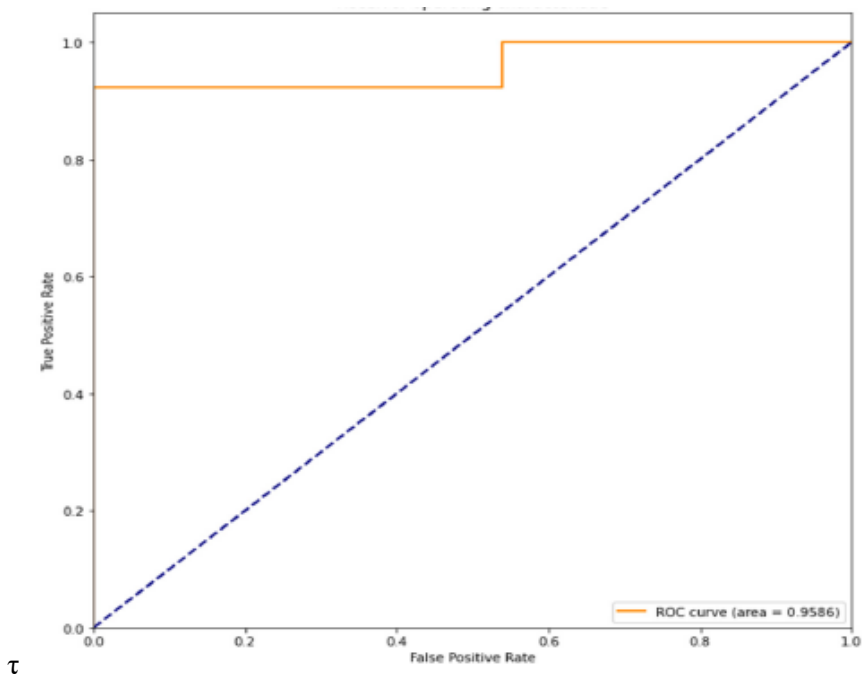
Πηγή: (Jan, 2021a)



Εικόνα 17: Γράφημα accuracy

Πηγή: (Jan, 2021a)

Η τιμή AUC είναι 0,9586 και δείχνει υψηλή ακρίβεια, αφού είναι πολύ κοντά στην τιμή 1 (εικόνα 18).



Εικόνα 18: AUC του μοντέλου RNN

Πηγή: (Jan, 2021a)

Τέλος, τα ποσοστά του confusion matrix και των δύο μοντέλων συνοψίζονται στον πίνακα 17. Συγκεκριμένα, για το **μοντέλο RNN** είναι: accuracy = 87,18%, precision = 86,67%, sensitivity (recall) = 81,25%, specificity = 91,30%, and F1-score = 83,87% και δείχνουν καλή απόδοση του μοντέλου. Ενώ για το **μοντέλο LSTM** είναι : accuracy = 94,88%, precision = 92,31%, sensitivity (recall) = 92,31%, specificity = 96,15%, and F1-score = 92,31% που δείχνουν εξαιρετική απόδοση του μοντέλου.

Πίνακας 17: Confusion matrix των μοντέλων RNN και LSTM

Model	Accuracy	Type I Error	Type II Error	Precision	Sensitivity (Recall)	F1-Score	Specificity	AUC	Training Time
RNN	87.18%	5.13%	7.69%	86.67%	81.25%	83.87%	91.30%	91.30%	0.3s
LSTM	94.88%	2.56%	2.56%	92.31%	92.31%	92.31%	96.15%	95.86%	1.5s

Πηγή: (Jan, 2021a)

Τα αποτελέσματα έδειξαν ότι και τα δύο μοντέλα επιτυγχάνουν υψηλή απόδοση και επομένως είναι κατάλληλα για τον εντοπισμό απάτης στις οικονομικές καταστάσεις. Μάλιστα η ακρίβεια τους είναι πάνω από 85%. Όμως μεταξύ των δύο, το μοντέλο LSTM επιτυγχάνει τη μεγαλύτερη ακρίβεια (94,88%) και υπερτερεί σε όλα τα μέτρα απόδοσης. Όσον αφορά τα ποσοστά σφάλματος I και II, το μοντέλο LSTM έχει τα χαμηλότερα ποσοστά. Επίσης οι τιμές AUC είναι πολύ κοντά στο 1, επομένως και τα δύο μοντέλα έχουν καλή απόδοση αλλά το μοντέλο LSTM έχει την καλύτερη.

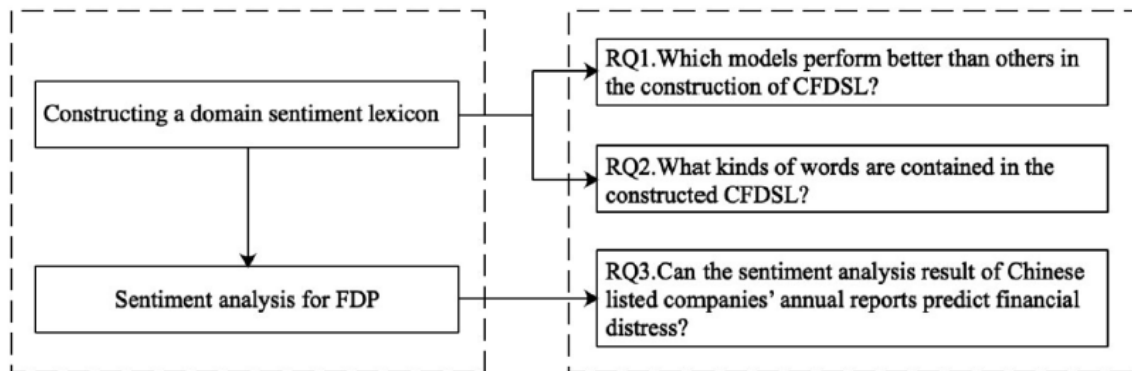
Για τη μελλοντική έρευνα προτείνεται να χρησιμοποιηθούν και άλλοι αλγόριθμοι deep learning, όπως τα μοντέλα DNN, DBN, CNN, CDBN και GRU, να ληφθούν υπόψη και άλλες μεταβλητές (π.χ. δείκτες οικονομικής ανάπτυξης/ύφεσης), να γίνει χρήση και άλλων οικονομετρικών μοντέλων (π.χ. QARDL-ECM) και γενικά οι χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές που θα υιοθετούνται να προσαρμόζονται ανάλογα με τους κανονισμούς και τις οικονομικές συνθήκες που επικρατούν σε κάθε χώρα.

4.4 A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction

Οι Li, Shi, Wang και Zhou (2021) ασχολήθηκαν με (1) την ανάπτυξη μιας μεθόδου deep learning για τη δημιουργία ενός κινέζικου λεξικού συναισθήματος του χρηματοοικονομικού τομέα (Chinese financial domain sentiment lexicon – CFDSL) και (2) την ανάλυση συναισθήματος για την πρόβλεψη της οικονομικής δυσπραγίας των εταιρειών (Financial distress prediction – FDP) . Η έρευνα είχε ως στόχο να απαντήσει στα εξής τρία διερευνητικά ερωτήματα:

- ποια μοντέλα είναι καταλληλότερα για την κατασκευή του λεξικού,
- τι είδους λέξεις θα πρέπει να περιλαμβάνονται στο λεξικό και

- κατά πόσο η ανάλυση συναισθήματος των ετήσιων οικονομικών καταστάσεων μπορεί να προβλέψει το financial distress.
- Η εικόνα 19 απεικονίζει σχηματικά το γενικό πλαίσιο της έρευνας.



Εικόνα 19: Το γενικό πλαίσιο της έρευνας

Πηγή: (Li κ.ά., 2021)

Για την κατασκευή του λεξικού και την ανάλυση συναισθήματος, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python.

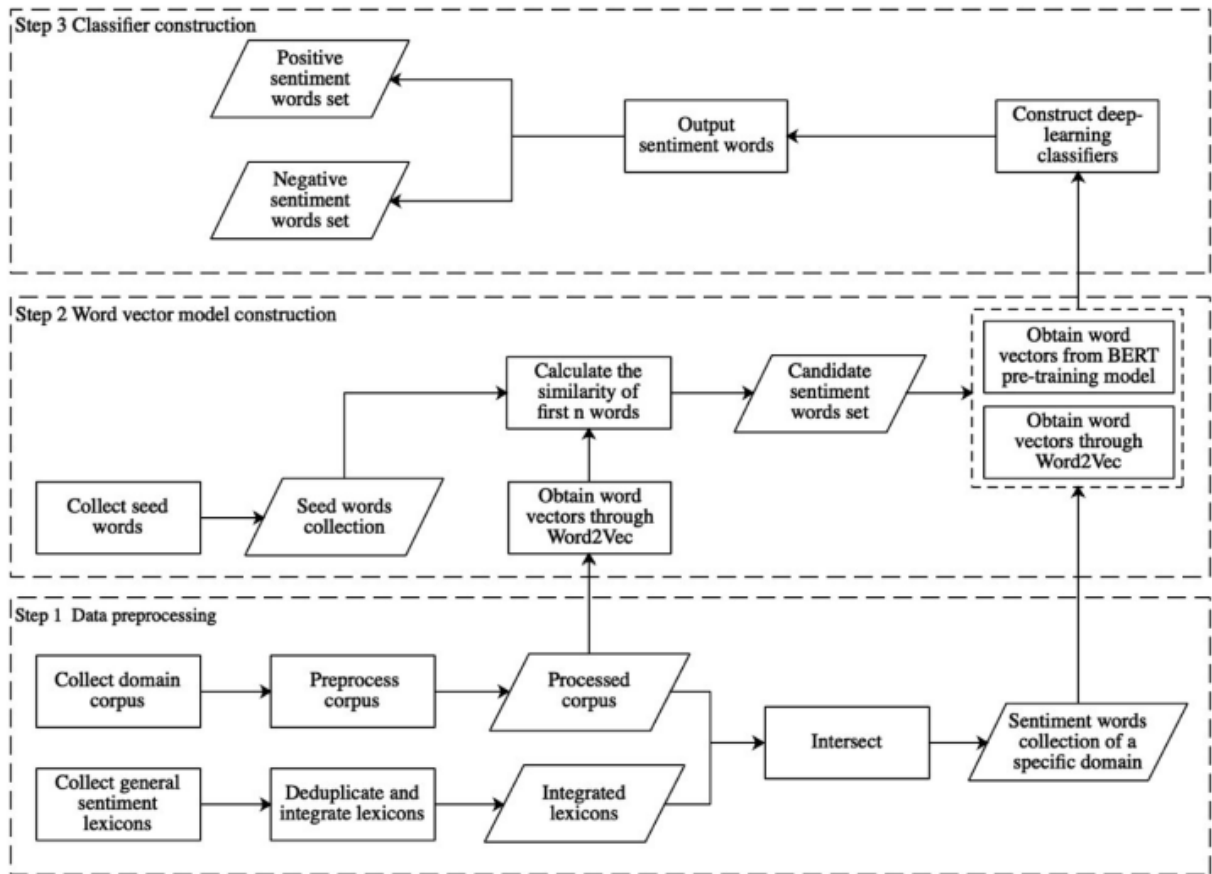
Τα δεδομένα συλλέχθηκαν από 214 κινέζικες εισηγμένες εταιρίες που είχαν χαρακτηριστεί με οικονομική δυσπραγία για το χρονικό διάστημα 2012-2018 και 214 υγιείς εταιρίες για το έτος 2018. Επίσης, για να ελεγχθεί η προγνωστική απόδοση του λεξικού συναισθήματος για το FDP, συλλέχθηκαν ετήσιες εκθέσεις 3, 4, 5 και 6 ετών πριν οι εταιρίες χαρακτηριστούν με οικονομική δυσπραγία.

Για τη δημιουργία του λεξικού, ακολουθήθηκαν 3 βήματα που αφορούν την προεπεξεργασία των δεδομένων, την κατασκευή του μοντέλου word vector και την κατασκευή του κατηγοριοποιητή (βλ. εικόνα 20).

Η συλλογή των λέξεων έγινε με βάση τα γενικά λεξικά συναισθήματος και το domain corpus για τα οποία χρησιμοποιήθηκαν αντίστοιχα τα γενικά κινέζικα λεξικά NTUSD, HowNet, TSING και DUTIR και οι ετήσιες καταστάσεις των κινέζικων εισηγμένων εταιρειών από το 2016 και μετά.

Έπειτα, αφού προηγήθηκαν οι διαδικασίες λήψης των seed words και υπολογισμού της ομοιότητάς τους για τη δημιουργία των candidate sentiment words, έγινε η λήψη των word vectors με τη χρήση των μοντέλων Word2Vec και pre-training BERT.

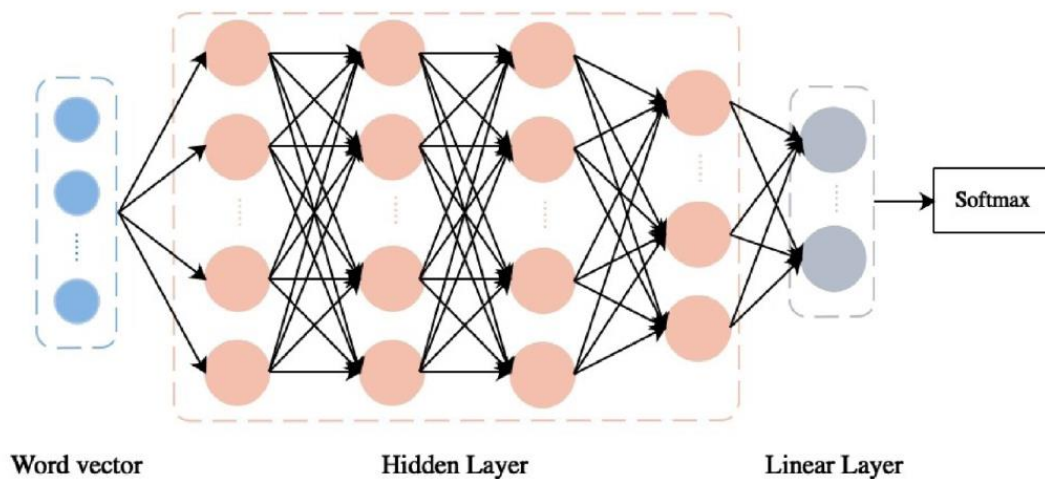
Στη συνέχεια, τα word vectors αποτέλεσαν την είσοδο για την κατασκευή των μοντέλων. Υιοθετήθηκε η τεχνική 10-fold cross-validation για τον διαχωρισμό του training και testing set. Το training set χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων ενώ το testing set για την αξιολόγηση της αποτελεσματικότητάς τους.



Εικόνα 20: Δημιουργία του sentiment lexicon

Πηγή: (Li κ.ά., 2021)

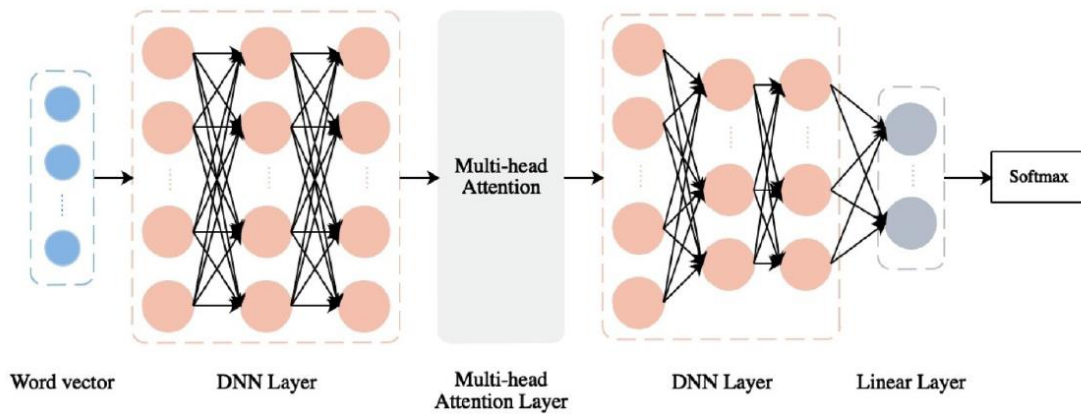
Κατά την έρευνα αναπτύχθηκαν τα εξής τρία μοντέλα:
 (1) το πλήρως συνδεδεμένο Deep Neural Network (DNN) με τέσσερα κρυφά layers (βλ. εικόνα 21).



Εικόνα 21: Deep Neural Network (DNN)

Πηγή: (Li κ.ά., 2021)

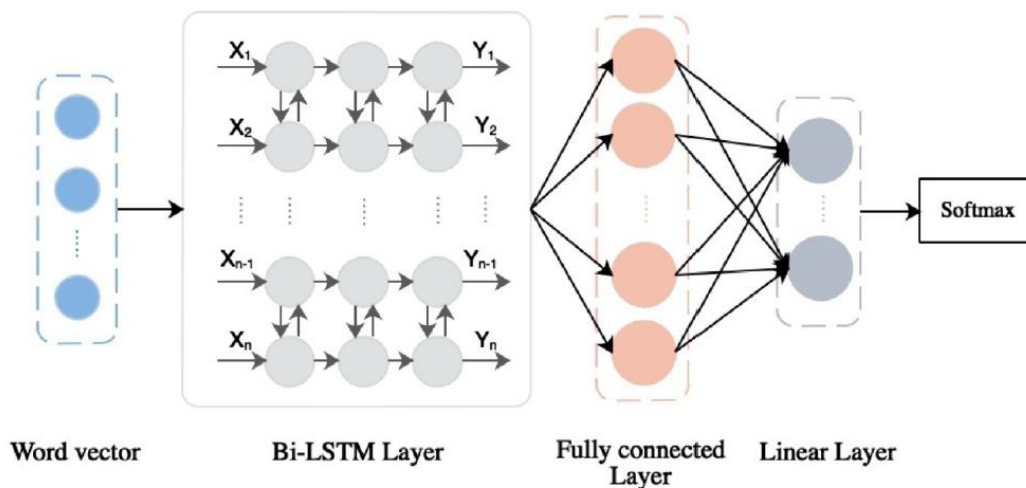
(2) το Multi-head Attention-based DNN (MA-DNN) (βλ. εικόνα 22)



Εικόνα 22: Multi-Head Attention-based Deep Neural Network (MA-DNN)

Πηγή: (Li κ.ά., 2021)

(3) το Bidirectional Long Short-Term Memory (Bi-LSMT) (βλ. εικόνα 23).



Εικόνα 23: Bidirectional Long Short-Term Memory (Bi-LSTM)

Πηγή: (Li κ.ά., 2021)

Μετά από επαναλαμβανόμενα πειράματα, προσδιορίστηκαν οι παράμετροι (συνάρτηση ενεργοποίησης, loss function, batch size, βελτιστοποιητής) κάθε μοντέλου ως ακολούθως (βλ. πίνακα 18):

Πίνακας 18: Οι παράμετροι των μοντέλων deep learning

Models	Parameter
DNN	activation function = softmax; loss = binary_crossentropy; optimizer = adam; batch_size = 64
MA-DNN	activation function = softmax; loss = binary_crossentropy; optimizer = adam; batch_size = 1024
Bi-LSTM	activation function = softmax; loss = binary_crossentropy; optimizer = adam; batch_size = 64

Πηγή: (Li κ.ά., 2021)

Επίσης, για την αξιολόγηση της απόδοσης των παραπάνω μοντέλων, λήφθηκαν υπόψη τα μέτρα αξιολόγησης accuracy, precision, recall και F1-measure.

Σύμφωνα με τα αποτελέσματα του πίνακα 19, τα μοντέλα που χρησιμοποιήσαν ως είσοδο τα BERT-based word vectors είχαν καλύτερη απόδοση. Επίσης το μοντέλο MA-DNN είχε τα καλύτερα αποτελέσματα τόσο στα Word2Vec-based word vectors όσο και στα BERT-based word vectors.

Πίνακας 19: Τα αποτελέσματα απόδοσης των μοντέλων deep learning

Inputs	Classifiers	Accuracy/%	Precision/%	Recall/%	F1/%
Word2Vec-based word vectors	DNN	81.34	82.61	81.13	81.86
	MA-DNN	85.68	85.92	85.71	85.81
	Bi-LSTM	72.52	73.61	72.92	73.26
BERT-based word vectors	DNN	87.76	89.63	88.90	89.26
	MA-DNN	90.12	89.56	89.72	89.64
	Bi-LSTM	79.13	78.93	78.57	78.75

Πηγή: (Li κ.ά., 2021)

Αναλυτικότερα, στα Word2Vec-based word vectors, το **μοντέλο MA-DNN** είχε τα υψηλότερα ποσοστά σε όλες τις μετρήσεις με accuracy 85,68%, precision 85,92%, recall 85,71% και F1-measure 85,81% ενώ στα BERT-based word vectors είχε τα υψηλότερα ποσοστά στο accuracy (90,12%), recall (89,72%) και F1-measure (89,64%) ενώ το μοντέλο DNN το υψηλότερο precision (89,63%). Επίσης το μοντέλο **Bi-LSTM** είχε τις χαμηλότερες επιδόσεις είτε με τη χρήση των BERT-based word vectors είτε των Word2Vec-based word vectors.

Σύμφωνα με τα παραπάνω, η μελέτη υιοθέτησε το μοντέλο **MA-DNN** με είσοδο τα BERT-based word vectors για την ταξινόμηση του συνόλου των candidate sentiment words και **δημιούργησε το λεξικό CFDSL** με 3734 λέξεις από τις οποίες οι 2517 είναι θετικές και οι 1217 αρνητικές. Οι λέξεις έχουν σχέση με την κεφαλαιαγορά, τους χρηματιστηριακούς όρους, την επιχειρηματική κατάσταση και την πολιτική των εταιρειών (βλ. πίνακα 20).

Πίνακας 20: Παράδειγμα sentiment words του λεξικού CFDSL

Positive words	上涨(rise), 涨幅(rate of increase), 涨势(upward trend), 增持(increase holdings), 保持平衡(maintain balance), 快速增长(rapid growth), 持续增长(consistent growth), 持续上升(rise steadily), 走高(go up), 飙涨(skYROCKETING), 冲高(soar), 止跌企稳(stop falling and stabilize), 复苏(recovery), 好转(to become better), 大涨(rise sharply), 牛市(bull market), 薄利多销(small profit but quick turnover), 发明专利(invention patent), 利好(positive), 妥善处理(careful and skillful handling), 猛涨(leap upwards), 增长势头(growth momentum), 创新性(innovative), 鼓励性(encouragement), 升势(upswing), 回升(rise again), 盈利(profit), ...
Negative words	熊市(bear market), 走低(go down), 高开低走(open high and go low), 震荡(seesaw), 触底(bottoming), 减持(reduce holdings), 低潮(low ebb), 大跌(plunge), 疲态(weakness), 积压产品(backlog products), 不振(in low spirits), 违章操作(illegal operation), 低位(low), 解除合同(terminate the contract), 止步不前(cease to advance), 恶性循环(vicious spiral), 售假(selling fake), 跌幅(rate of decline), 急跌(slump), 利空(negative), 萎缩(shrink), 下降(decline), 动荡不安(turbulent), 故障(fault), 安全漏洞(security vulnerabilities), 地区冲突(regional conflicts), 急转直下(a sudden turn for the worse), ...

Πηγή: (Li κ.ά., 2021)

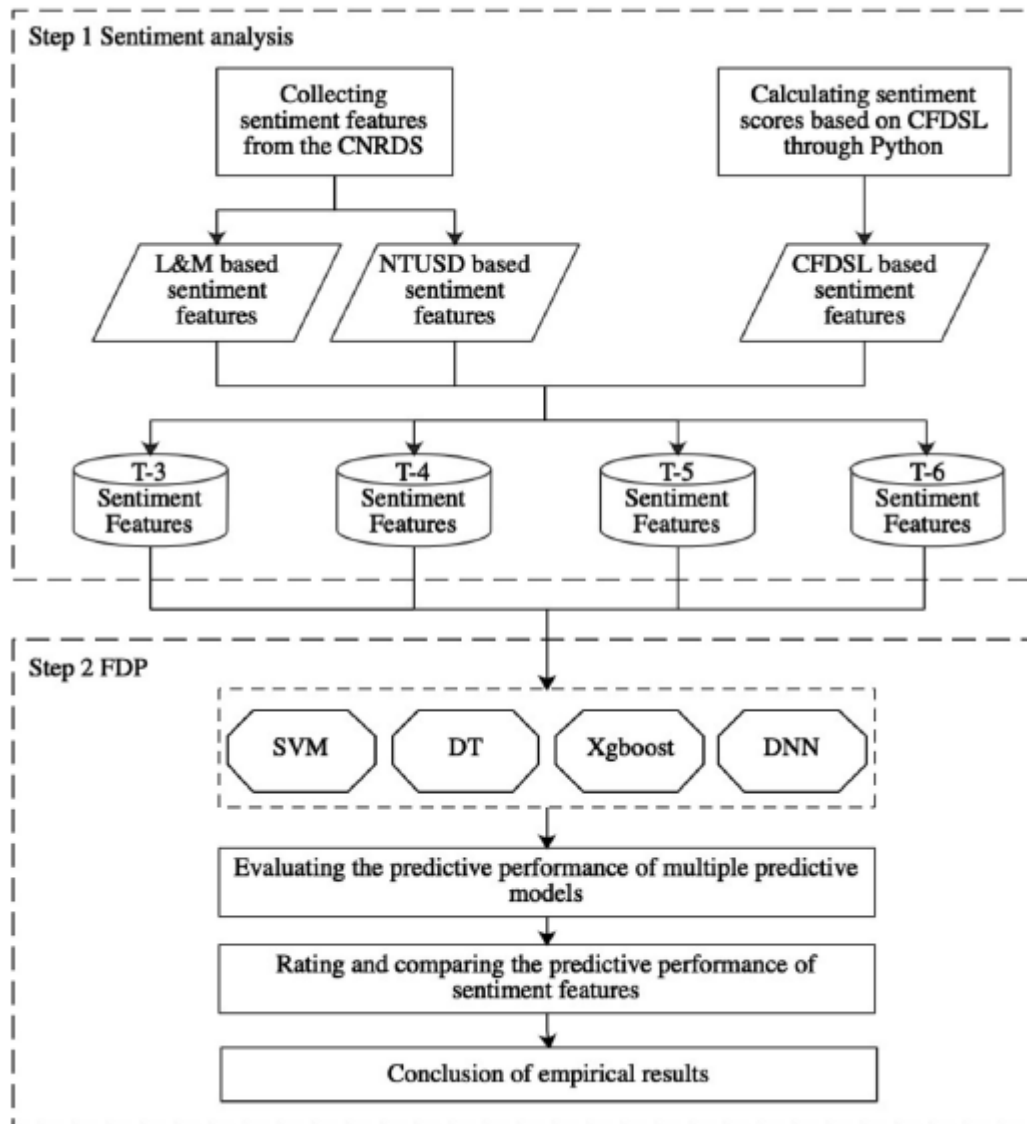
Στη συνέχεια, ακολούθησε η ανάλυση συναισθήματος για την πρόβλεψη της οικονομικής δυσπραγίας. Για την ανάλυση χρησιμοποιήθηκε το λεξικό CFDSL που δημιουργήθηκε προηγουμένως και συγκρίθηκε με τα λεξικά L&M, NTUSD για τα οποία υπολογίστηκαν τα sentiment features (βλ. πίνακα 21).

Πίνακας 21: Τα sentiment features της έρευνας

Features	Definition
L&M_pos	Number of positive words in annual report, based on the L&M dictionary
L&M_neg	Number of negative words in annual report, based on the L&M dictionary
L&M_senti	$(L\&M_pos - L\&M_neg)/(L\&M_pos + L\&M_neg)$
NTUSD_pos	Number of positive words in annual report, based on the NTUSD
NTUSD_neg	Number of negative words in annual report, based on the NTUSD
NTUSD_senti	$(NTUSD_pos - NTUSD_neg)/(NTUSD_pos + NTUSD_neg)$
CFDSL_pos	Number of positive words in annual report, based on the CFDSL
CFDSL_neg	Number of negative words in annual report, based on the CFDSL
CFDSL_senti1	$(CFDSL_pos - CFDSL_neg)/(CFDSL_pos + CFDSL_neg)$
CFDSL_senti2	Sentiment score calculated through Python programming, based on CFDSL

Πηγή: (Li κ.ά., 2021)

Αφού ολοκληρώθηκαν οι υπολογισμοί, τα δεδομένα αποθηκεύτηκαν στις βάσεις δεδομένων T-3, T-4, T-5, T-6 που αντιστοιχούν σε 3, 4, 5, και 6 έτη πριν χαρακτηριστούν οι εταιρείες με οικονομική δυσπραγία (βλ. εικόνα 24).



Εικόνα 24: Ανάλυση συναισθήματος για την πρόβλεψη της οικονομικής δυσπραγίας

Πηγή: (Li κ.ά., 2021)

Έπειτα, εφαρμόστηκαν τα μοντέλα SVM, DT, Xgboost και DNN για να συγκριθεί η απόδοση πρόβλεψης στα χρονικά διαστήματα (T-3, T-4, T-5, T-6). Για τα μοντέλα SVM, DT και Xgboost χρησιμοποιήθηκε η βιβλιοθήκη Python Scikit-learn και για το μοντέλο DNN η βιβλιοθήκη Python Keras. Επίσης, εφαρμόστηκε η τεχνική 10-fold cross validation και το σύνολο δεδομένων χωρίστηκε τυχαία σε 10 ομάδες όπου οι 9 αποτέλεσαν το training dataset και η μία το test dataset. Ολόκληρο το σύνολο των βημάτων επαναλήφθηκε 10 φορές και ο μέσος όρος αυτών καταγράφηκε ως αποτέλεσμα. Επίσης η εκπαίδευση των μοντέλων έγινε με τη χρήση των κάτωθι παραμέτρων (βλ. πίνακα 22).

Πίνακας 22: Οι παράμετροι των μοντέλων

Method	Parameters
SVM	Kernel type used in the algorithm=linear; penalty parameter C of the error term=1; class weight=balanced
DT	Parameters = {'max_depth': range (1, 20), 'criterion': np.array (['gini', 'entropy'])}
Xgboost	Number of trees to fit=100; maximum tree depth for base learners=5; boosting learning rate=0.1
DNN	Number of hidden layers=3; activation function=sigmoid; batch size=16; epoch=100; the optimizer of the model=adam

Πηγή: (Li κ.ά., 2021)

Για την αξιολόγηση της απόδοσης των μοντέλων, χρησιμοποιήθηκαν τα μέτρα accuracy, precision, recall και F1-measure.

Σύμφωνα με τον πίνακα 23, τα μοντέλα εμφανίζουν τις καλύτερες αποδόσεις στο dataset T4 και ακολουθούν τα datasets T3, T6 και T5.

Μάλιστα όσον αφορά το dataset **T4**, το μοντέλο DNN εμφανίζει τα μεγαλύτερα ποσοστά σε accuracy (85,71%), recall (85,53%) και F1 (85,22%) ενώ το μοντέλο Xgboost το υψηλότερο precision (85,03%). Όσον αφορά το dataset **T3**, το μοντέλο DNN εμφανίζει τα υψηλότερα ποσοστά σε accuracy (84,41%), recall (84,78%) και F1 (84,06%) και το μοντέλο Xgboost το υψηλότερο precision (83,44%). Ενώ τα μοντέλα DT και SVM σημειώνουν τις χαμηλότερες αποδόσεις.

Πίνακας 23: Τα αποτελέσματα απόδοσης των μοντέλων

	Method	Accuracy/%	Precision/%	Recall/%	F1/%
T-3	DT	73.09	74.81	73.09	73.94
	SVM	72.89	73.64	73.17	73.40
	Xgboost	82.23	83.44	82.97	83.20
	DNN	84.41	83.36	84.78	84.06
T-4	DT	75.25	74.42	75.25	74.83
	SVM	73.97	74.09	75.76	74.92
	Xgboost	84.22	85.03	84.95	84.99
	DNN	85.71	84.91	85.53	85.22
T-5	DT	69.97	70.28	69.97	70.12
	SVM	69.63	70.45	69.27	69.86
	Xgboost	80.36	81.22	81.60	81.41
	DNN	82.63	82.97	83.08	83.02
T-6	DT	70.79	70.32	70.79	70.55
	SVM	70.45	70.01	70.61	70.31
	Xgboost	81.67	82.34	82.68	82.51
	DNN	83.50	83.06	83.63	83.34

Πηγή: (Li κ.ά., 2021)

Από τα αποτελέσματα, λοιπόν, προκύπτει ότι τα μοντέλα DT και SVM, που είναι single classifiers, εμφανίζουν μειονεκτήματα σε σύγκριση με τα μοντέλα Xgboost (ensemble classifier) και DNN (deep learning model). Επίσης, τα sentiment features

έχουν καλή προγνωστική απόδοση στα datasets T4 και T3 που αντιστοιχούν στα 4 και 3 έτη πριν χαρακτηριστούν οι εταιρείες με οικονομική δυσπραγία.

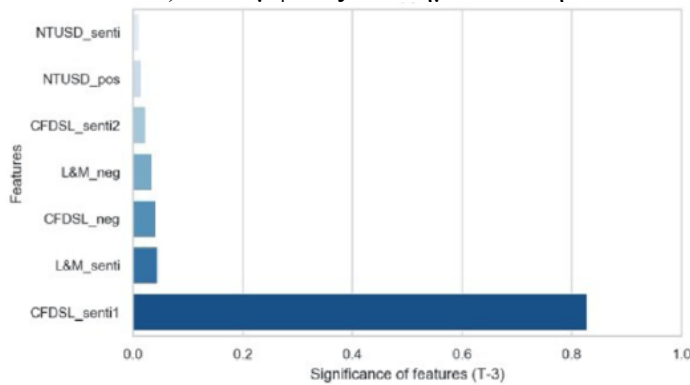
Στη συνέχεια, έγινε εξαγωγή των σημαντικών χαρακτηριστικών για κάθε χρονικό διάστημα με βάση την ερμηνευσιμότητα του μοντέλου DT (βλ. πίνακα 24) και βαθμολογήθηκαν ανάλογα με τη σημαντικότητα προκειμένου να εκτιμηθεί η εγκυρότητά τους.

Πίνακας 24: Τα σημαντικά χαρακτηριστικά κάθε χρονικού διαστήματος T3,T4, T5 και T6

	L&M based features	NTUSD based features	CFDSL based features
T-3	L&M_neg, L&M_senti	NTUSD_pos, NTUSD_senti	CFDSL_neg, CFDSL_senti1, CFDSL_senti2
T-4	L&M_senti	NTUSD_pos, NTUSD_neg, NTUSD_senti	CFDSL_senti1, CFDSL_senti2
T-5	L&M_pos, L&M_neg, L&M_senti	N/A	CFDSL_senti1, CFDSL_senti2
T-6	L&M_neg	NTUSD_pos, NTUSD_neg	CFDSL_pos, CFDSL_neg, CFDSL_senti1, CFDSL_senti2

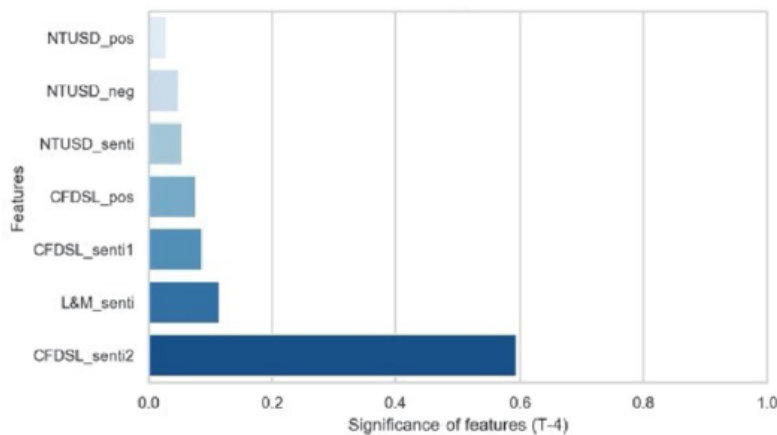
Πηγή: (Li κ.ά., 2021)

Έτσι σύμφωνα με την εικόνα 25α, στο χρονικό διάστημα T3 το χαρακτηριστικό CFDSL_senti1 εμφανίζει σημαντικότητα μεγαλύτερη από 0,8 σε σύγκριση με τα υπόλοιπα (L&M_neg, L&M_senti, NTUSD_pos, NTUSD_senti, CFDSL_neg, CFDSL_senti2) που εμφανίζουν σημαντικότητα κάτω από 0,1.



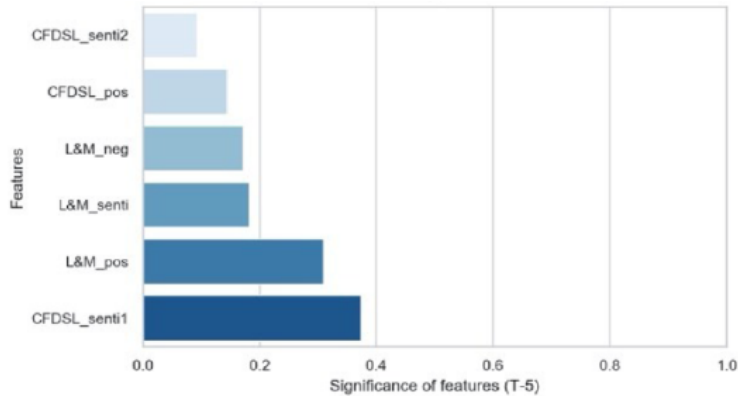
(α): T3

Στο χρονικό διάστημα T4, το χαρακτηριστικό CFDSL_senti2 εμφανίζει τη μεγαλύτερη σημαντικότητα (~0,6) και ακολουθούν τα L&M_senti και CFDSL_senti1 (βλ. εικόνα 25β).



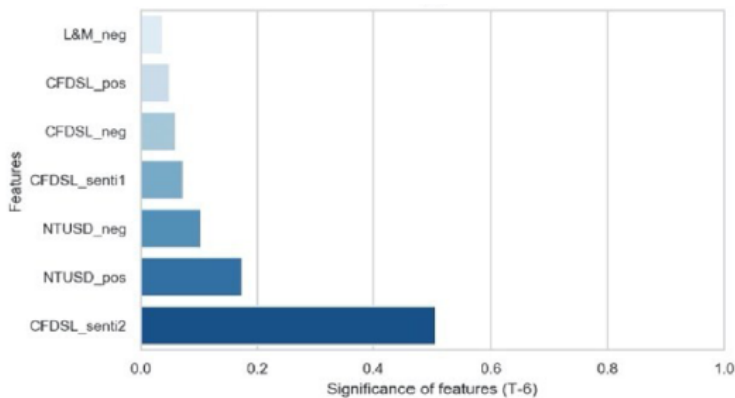
(β): T4

Ενώ στο χρονικό διάστημα **T5**, όλα τα χαρακτηριστικά εμφανίζουν σημαντικότητα μικρότερη από 0,4 με το CFDSL_senti1 να έχει τη μεγαλύτερη (βλ. σχήμα 25γ).



(γ): T5

Τέλος, στο χρονικό διάστημα **T6**, το CFDSL_senti2 έχει τη μεγαλύτερη σημαντικότητα (~0,5) και ακολουθούν τα NTUSD_pos και NTUSD_neg (βλ. σχήμα 25δ).



(δ): T6

Εικόνα 25: Κατάταξη sentiment features για (α) T3, (β) T4, (γ) T5 και (δ) T6

Πηγή: (Li κ.ά., 2021)

Συμπερασματικά, τα χαρακτηριστικά CFDSL έχουν πλεονέκτημα σε όλα τα παραπάνω χρονικά διαστήματα, επομένως μπορούν να επιτύχουν καλύτερη προγνωστική απόδοση στο FDL σε σχέση με τα χαρακτηριστικά L&M και NTUSD.

Απαντώντας στα διερευνητικά ερωτήματα της μελέτης, (1) από τα τρία μοντέλα deep learning που αναπτύχθηκαν για την κατασκευή του λεξικού, το καταλληλότερο είναι το μοντέλο MA-DNN ενώ το μοντέλο DNN, που είναι το απλό μοντέλο deep learning, είχε επίσης ικανοποιητικά αποτελέσματα, (2) οι λέξεις που περιλαμβάνονται στο λεξικό έχουν σχέση με τη χρηματοδότηση που αποτελεί ένα ισχυρό

χαρακτηριστικό του χρηματοοικονομικού τομέα και (γ) τα αποτελέσματα έδειξαν ότι τα sentiment features των ετήσιων εκθέσεων μπορούν να αποκαλύψουν την οικονομική δυσπραγία σε προγενέστερο χρόνο. Επίσης όσον αφορά τα σημαντικά χαρακτηριστικά, αυτά που βασίζονται στο λεξικό CFDSL έχουν καλύτερη απόδοση στο FDP σε σύγκριση με τα λεξικά L&M και NTUSD.

Οι ερευνητές αναφέρουν ότι οι μελλοντικές έρευνες, θα πρέπει να συλλέξουν δεδομένα για μεγαλύτερο χρονικό για την κατασκευή του λεξικού, να συνδυάσουν το μοντέλο deep learning Bi-LSTM με έναν μηχανισμό προσοχής για να επιτευχθεί μεγαλύτερη ακρίβεια δεδομένου ότι δεν είχε καλή απόδοση στην παρούσα έρευνα. Επίσης να συνδυάσουν το ερμηνεύσιμο μοντέλο DT με μοντέλα νευρωνικών δικτύων, ώστε να διατηρήσουν την ερμηνευσιμότητα και να επιτύχουν καλύτερη προγνωστική απόδοση.

4.5 Predicting distresses using deep learning of text segments in annual reports

Στην παρούσα έρευνα οι Matin, Hansen, Hansen και Molgaard (2019) ανέπτυξαν τρία μοντέλα deep learning το κάθε ένα από τα οποία αποτελείται από ένα convolutional και ένα recurrent neural network. Χρησιμοποίησαν αδόμητα δεδομένα κειμένου σε συνδυασμό με αριθμητικές χρηματοοικονομικές μεταβλητές με σκοπό την πρόβλεψη της οικονομικής δυσπραγίας. Το ένα μοντέλο χρησιμοποίησε ως αδόμητα δεδομένα κειμένου τις εκθέσεις των ελεγκτών, το δεύτερο τις δηλώσεις των διευθυντικών στελεχών και το τρίτο τον συνδυασμό και των δύο.

Το σύνολο δεδομένων βασίστηκε στα δεδομένα που χρησιμοποιήθηκαν από τους Christoffersen, Martin, & Molgaard (2018) και αποτελείται από ετήσιες εκθέσεις εταιρειών της Δανίας για το χρονικό διάστημα 2013-2016. Επιπλέον εμπλουτίστηκε με χαρακτηριστικά τα οποία συλλέχθηκαν από το Danish Central Business Register και αφορούν την ηλικία, τον τομέα και το νομικό καθεστώς κάθε εταιρείας. Το σύνολο δεδομένων περιλαμβάνει 278.047 εταιρικά έτη, 112974 εταιρείες και 8033 distresses με συνολικά 50 χρηματοοικονομικές μεταβλητές από τις οποίες οι 44 ήταν αριθμητικές και οι 6 κατηγορικές (βλ. πίνακα 25). Μάλιστα για τις **αριθμητικές μεταβλητές** εφαρμόστηκε η μέθοδος winsorized 5% και 95% για βελτιωμένη απόδοση. Τα δεδομένα κειμένου συλλέχθηκαν από τις **εκθέσεις των ελεγκτών και τις δηλώσεις του ΔΣ**. Το 25%, 50% και 75% των δηλώσεων του ΔΣ αντιστοιχεί σε 37, 54 και 83 λέξεις και των εκθέσεων των ελεγκτών σε 187, 205 και 219 λέξεις.

Πίνακας 25: Οι μεταβλητές της έρευνας

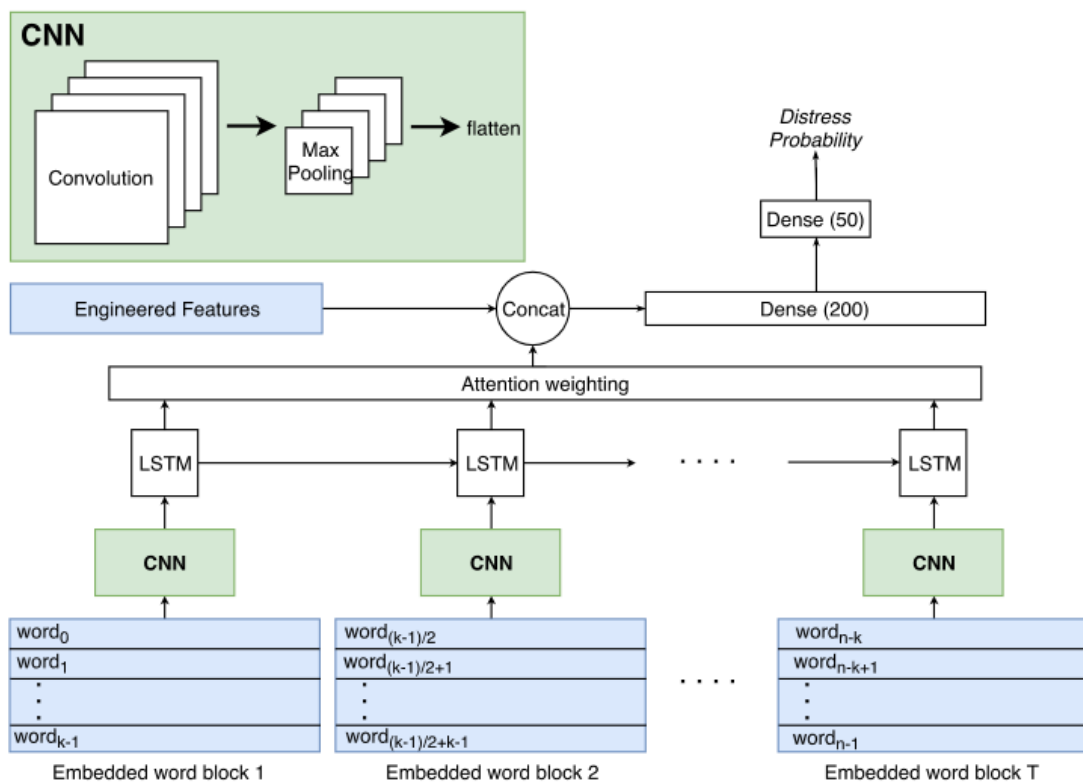
Type	Input variable
Continuous	Accounts payable*
	Accounts receivable*
	Change in log size
	Corporation tax*
	Current assets*
	Deferred tax*
	Depreciation*
	EBIT*
	Equity/invested capital
	Equity*
	Expected dividends*
	Financial assets*
	Financial income*
	Financing costs*
	Fixed costs*
	Ind. EW avg. net profit*
	Interest coverage ratio
	Inventory*
	Invested capital*
	Land and buildings*
	Liquid assets*
	log(age)
	log(size)
	Long-term bank debt*
	Long-term debt*
	Long-term mortgage debt*
	Net profit*
	Other operating expenses*
	Other receivables*
	Other short debts*
	Personnel costs*
	Prepayments*
	Provisions*
	Quick ratio
	Receivables from related parties*
	Relative debt change
	Retained earnings*
	Return on equity (%)
	Short-term bank debt*
	Short-term mortgage debt*
Tangible fixed assets*	
Tax expenses*	
Total receivables*	
Categorical	Has prior distress
	Is private limited (Danish "Anpartsselskab")
	Large debt change
	Negative equity
	Sector

Πηγή: (Matin κ.ά, 2019)

Για τα δεδομένα κειμένου, εφαρμόστηκαν **τεχνικές προεπεξεργασίας**, ώστε να μειωθεί η μεταβλητότητα μεταξύ των κειμένων, να αποφευχθεί η υπερβολική προσαρμογή και να επιτευχθεί η γενίκευση των μοντέλων. Επίσης, λέξεις με μικρή συχνότητα εμφάνισής (κάτω από 25 φορές) αγνοήθηκαν.

Έπειτα ακολούθησαν οι διαδικασίες **Parttern extraction** και **recognition** με τη χρήση ενός convolutional recurrent neural network. Αναλυτικότερα, η **αρχιτεκτονική** του δικτύου, είτε ενσωμάτωνε μόνο τις εκθέσεις των ελεγκτών είτε τις δηλώσεις των διευθυντικών στελεχών ή τον συνδυασμό αυτών, είναι ως εξής (βλ. εικόνα 26):

Αρχικά, χρησιμοποιήθηκαν **word embeddings** για την αντιστοίχιση κάθε λέξης κειμένου σε ένα **dense vector representation** όπου οι παρόμοιες σημασιολογικά λέξεις είναι κοντά η μία με την άλλη. Έγινε χρήση του μοντέλου word2vec και συγκεκριμένα του μοντέλου skip-gram όπου στόχος του είναι κάθε λέξη να προβλέψει τις γύρω της λέξεις. Έτσι το κείμενο χωρίστηκε σε **half-overlapping blocks of words**, κάθε ένα από τα οποία έγινε είσοδος σε ένα **single-layer convolutional neural network** (CNN) αποτελούμενο από ένα convolution και ένα max-pooling step, ώστε να εξαχθούν τα **patterns** από κάθε block. Στη συνέχεια, τα pattern εισήχθησαν σε ένα recurrent neural network (RNN) και συγκεκριμένα σε ένα **Long Short-Term Memory** (LSTM) network και η τελική αναπαράσταση κειμένου υπολογίστηκε ως ένα attention-weighted sum κάθε εξόδου RNN. Η χρήση του attention επιτρέπει το μοντέλο να εστιάσει την προσοχή του στα λιγότερα και πιο σημαντικά τμήματα του κειμένου, ώστε να μάθει μια καλύτερη περιγραφική αναπαράσταση. Έπειτα το attention-weighted sum ενώθηκε με τις αριθμητικές μεταβλητές του πίνακα 25 και τροφοδοτήθηκε σε δύο fully-connected layers μεγέθους 200 και 50 τα οποία χρησιμοποίησαν τη συνάρτηση ενεργοποίησης rectified linear unit (ReLU) για την πρόβλεψη της πιθανότητας distress.



Εικόνα 26: Η αρχιτεκτονική του δικτύου

Πηγή: (Matin κ.ά, 2019)

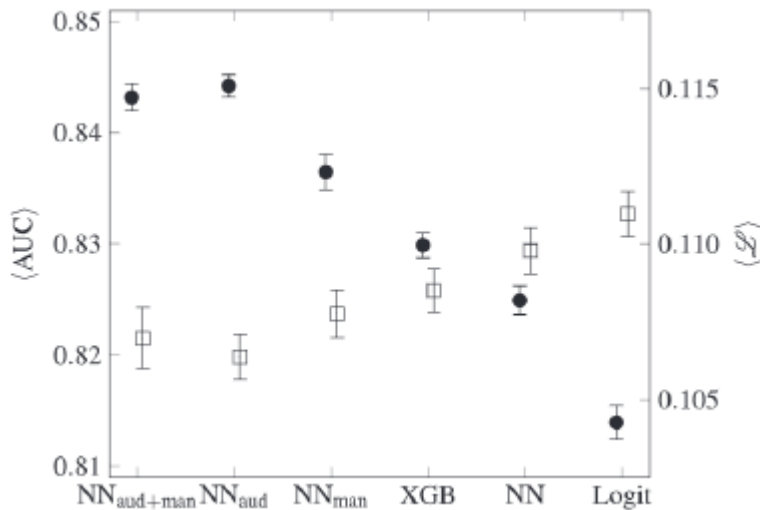
Για την εφαρμογή των μοντέλων έγινε χρήση της βιβλιοθήκης Tensorflow, για την εκπαίδευση χρησιμοποιήθηκε ο βελτιστοποιητής Adam, η συνάρτηση απώλειας binary cross-entropy ενώ για την αποφυγή της υπερπροσαρμογής και την έγκαιρη διακοπή της εκπαίδευσης (early stopping) ορίστηκε ως validation set το 10% του training set.

Για τη **σύγκριση** των προτεινόμενων μοντέλων, χρησιμοποιήθηκαν τα μοντέλα (1) XGBoost, (2) logistic regression (logit) και (3) ένα neural network ίδιας αρχιτεκτονικής με τα προτεινόμενα για τα οποία έγινε χρήση μόνο αριθμητικών μεταβλητών.

Η αξιολόγηση των αποτελεσμάτων έγινε με τη χρήση των μέτρων **AUC** και **log score**. Όσο πιο κοντά στο 1 είναι το AUC και όσο πιο μικρό το log score τόσο καλύτερη η απόδοση του μοντέλου.

Κατά τα πειράματα, τα νευρωνικά δίκτυα που δεν περιείχαν δεδομένα κειμένου συμβολίστηκαν ως NN ενώ όσα περιείχαν συμβολίστηκαν ως εξής : (α) NN_{aud} όταν περιείχαν τις εκθέσεις των ελεγκτών, (β) NN_{man} όταν περιείχαν τις δηλώσεις ΔΣ και (γ) $NN_{aud+man}$ όταν περιείχαν και τα δύο είδη κειμένου.

Ο μέσος όρος AUC και log με ένα standard error για κάθε μοντέλο υπολογίστηκαν με την τεχνική 10-fold cross validation με βάση τις εταιρείες του δείγματος. Σύμφωνα με τα αποτελέσματα της εικόνας 27, διαπιστώθηκε ότι τα μοντέλα που περιέχουν τα δεδομένα κειμένου έχουν μεγαλύτερη τιμή AUC και χαμηλότερο log score σε σύγκριση με τα υπόλοιπα μοντέλα.



Εικόνα 27: Μέσος όρος AUC (● αριστερός άξονας) και log score (□ δεξιός άξονας)

Πηγή: (Matin κ.ά, 2019)

Ο πίνακας 26 δείχνει τα p-values που υπολογίστηκαν με τη μέθοδο paired two-tailed t-test για τη σύγκριση των αποτελεσμάτων και των έξι μοντέλων. Σύμφωνα με τα αποτελέσματα, το AUC αυξάνεται όταν τα μοντέλα περιέχουν κείμενο. Μάλιστα σε σύγκριση με το μοντέλο NN, η αύξηση για το μοντέλο NN_{aud} είναι 1,9, για το $NN_{aud+man}$ 1,8 και για το NN_{man} 1,1. Επίσης το log score του μοντέλου NN_{aud} εμφανίζει το καλύτερο αποτέλεσμα σε σχέση με τα υπόλοιπα μοντέλα.

Πίνακας 26: (a) μέσος όρος AUC, (b) μέσος όρος log score

Model	$\langle \text{AUC} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.843	-	-	-
NN_{aud}	0.844	0.233	-	-
NN_{man}	0.836	0.000	0.000	-
XGB	0.830	0.000	0.000	0.003
NN	0.825	0.000	0.000	0.000
Logit	0.814	0.000	0.000	0.000
(a) AUC				
Model	$\langle \mathcal{L} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.1070	-	-	-
NN_{aud}	0.1064	0.4263	-	-
NN_{man}	0.1078	0.1471	0.0032	-
XGB	0.1085	0.0643	0.0001	0.0372
NN	0.1098	0.0005	0.0000	0.0001
Logit	0.1110	0.0001	0.0000	0.0000
(b) Log score \mathcal{L}				

Πηγή: (Matin κ.ά, 2019)

Σύμφωνα με τα παραπάνω, τα μοντέλα που περιέχουν κείμενο βελτιώνουν την προγνωστική ισχύ. Όμως το μοντέλο (NN_{aud}), που περιέχει μόνο τις εκθέσεις ελεγκτών, δίνει την καλύτερη πρόβλεψη distress και αυτό γιατί οι εκθέσεις των ελεγκτών είναι αντικειμενικές και εκφράζουν την επαγγελματική τους γνώμη αμερόληπτα για την οικονομική κατάσταση μιας εταιρείας σε αντίθεση με τις δηλώσεις των διευθυντών που δίνουν μια ευνοϊκότερη εικόνα της εταιρείας για το μέλλον.

Στη συνέχεια, έγινε επανάληψη του πειράματος λαμβάνοντας υπόψη αυτή τη φορά ως δείγμα μόνο τις μεγάλες εταιρείες (άνω των 5 million DKK) που αντιστοιχεί στο 35,4% του συνολικού δείγματος αλλά στο 95,4% του συνολικού χρέους. Σύμφωνα με τον πίνακα 27, το AUC και το log score έχουν καλύτερα αποτελέσματα σε όλα τα μοντέλα σε σύγκριση με το προηγούμενο πείραμα. Μάλιστα αυτή τη φορά η αύξηση του AUC στο NN_{aud} είναι 2,6 ενώ η διαφορά μεταξύ NN_{man} και XGB δεν είναι σημαντική.

Η αύξηση αυτή πιθανόν να οφείλεται στο γεγονός ότι οι εκθέσεις των ελεγκτών είναι καλύτερης ποιότητας και περιέχουν περισσότερες πληροφορίες από τις οποίες το μοντέλο μπορεί να εξάγει καλύτερα αποτελέσματα. Αντίθετα, οι δηλώσεις των ΔΣ δεν είναι τόσο σημαντικές ίσως λόγω του μικρότερου μεγέθους του δείγματος. Επομένως, η ανάλυση των εκθέσεων ελεγκτών είναι χρήσιμη ιδιαίτερα για τις μεγάλες εταιρείες.

Πίνακας 27: (a) μέσος όρος AUC, (b) μέσος όρος log score των μεγάλων εταιρειών

Model	$\langle \text{AUC} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.877	-	-	-
NN_{aud}	0.879	0.562	-	-
NN_{man}	0.864	0.013	0.004	-
XGB	0.860	0.000	0.000	0.290
NN	0.853	0.000	0.000	0.002
Logit	0.834	0.000	0.000	0.000
(a) AUC				
Model	$\langle \mathcal{L} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.0611	-	-	-
NN_{aud}	0.0611	0.9815	-	-
NN_{man}	0.0627	0.0551	0.0095	-
XGB	0.0629	0.0085	0.0127	0.6588
NN	0.0640	0.0036	0.0001	0.0046
Logit	0.0657	0.0000	0.0001	0.0002
(b) Log score \mathcal{L}				

Πηγή: (Matin κ.ά, 2019)

Επίσης πραγματοποιήθηκε ένα ακόμα πείραμα με σκοπό να επιβεβαιωθεί ότι τα παραπάνω αποτελέσματα δεν είναι τυχαία. Ο μέσος όρος AUC και log υπολογίστηκαν με την τεχνική 4-fold validation με βάση τα έτη έκδοσης των ετήσιων εκθέσεων. Όπως φαίνεται στον πίνακα 28, τα αποτελέσματα εμφανίζουν την ίδια τάση με τον πίνακα 26. Τα λίγο υψηλότερα p-values είναι λόγω του μικρότερου αριθμού των folds του πειράματος. Σύμφωνα με τα αποτελέσματα, επιβεβαιώθηκε η εγκυρότητα των αποτελεσμάτων των πινάκων 26 & 27.

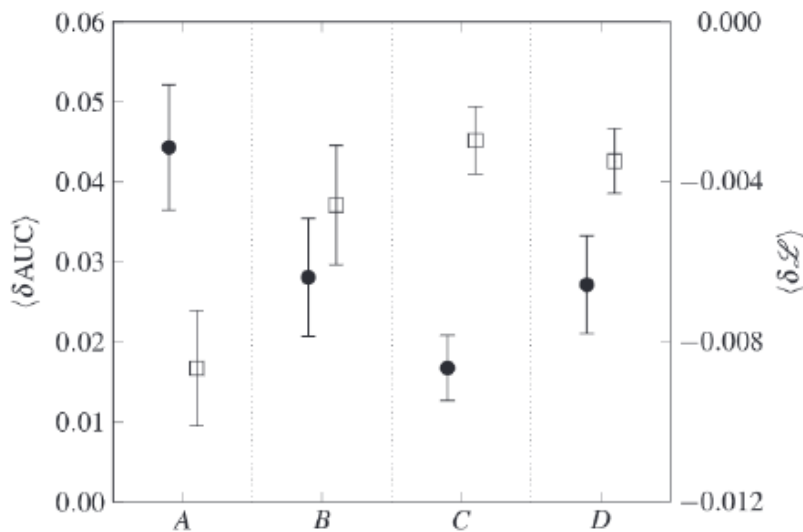
Πίνακας 28: (a) μέσος όρος AUC, (b) μέσος όρος log score με βάση το έτος

Model	$\langle \text{AUC} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.843	-	-	-
NN_{aud}	0.842	0.299	-	-
NN_{man}	0.830	0.003	0.014	-
XGB	0.826	0.001	0.006	0.175
NN	0.822	0.001	0.004	0.054
Logit	0.814	0.000	0.001	0.009
(a) AUC				
Model	$\langle \mathcal{L} \rangle$	$p_{\text{aud+man}}$	p_{aud}	p_{man}
$\text{NN}_{\text{aud+man}}$	0.1090	-	-	-
NN_{aud}	0.1095	0.4130	-	-
NN_{man}	0.1114	0.0312	0.1289	-
XGB	0.1109	0.0128	0.0627	0.3484
NN	0.1122	0.0112	0.0166	0.2098
Logit	0.1127	0.0081	0.0005	0.1892
(b) Log score \mathcal{L}				

Πηγή: (Matin κ.ά, 2019)

Επίσης εξετάστηκε η βελτίωση της πρόβλεψης των εκθέσεων που εκδίδονται από τις τέσσερις μεγαλύτερες ελεγκτικές εταιρείες σε αριθμό πελατών και συμβολίστηκαν με A, B, C & D αντίστοιχα. Κατά την εξέταση λήφθηκε υπόψη όλο το σύνολο δεδομένων και χρησιμοποιήθηκαν μόνο τα μοντέλα NN και NNaud.

Ο μέσος όρος δAUC και $\delta \log$ των δύο μοντέλων με ένα standard error υπολογίστηκαν με βάση το 10-fold cross validation και τα αποτελέσματα φαίνονται στην εικόνα 28. Το θετικό δAUC και το αρνητικό $\delta \log$ σημαίνει βελτίωση και η απόδοση κατάταξης και των τεσσάρων εταιρειών βελτιώνεται όταν χρησιμοποιούνται οι εκθέσεις των ελεγκτών. Επίσης η απόδοση της ελεγκτικής εταιρίας A υπερτερεί έναντι των άλλων γεγονός που δείχνει ότι οι εκθέσεις της συγκεκριμένης ελεγκτικής εταιρείας περιέχουν περισσότερες χρήσιμες πληροφορίες.



Εικόνα 28: Μέσος όρος AUC των τεσσάρων μεγαλύτερων ελεγκτικών εταιρειών (• αριστερός άξονας) και log score (□ δεξιός άξονας)

Πηγή: (Matin κ.ά, 2019)

Τα παραπάνω αποτελέσματα έδειξαν ότι η ικανότητα πρόβλεψης των μοντέλων βελτιώνεται με τη χρήση αδόμητων δεδομένων και ιδιαίτερα με τη χρήση των εκθέσεων των ελεγκτών.

Σύμφωνα με τις υποδείξεις των Matin κ.ά. (2019), οι έρευνες θα πρέπει να επεκταθούν σε μεγαλύτερες οικονομίες δεδομένου ότι η Δανία είναι μια μικρή σχετικά οικονομία και το σύνολο δεδομένων θα πρέπει να περιλαμβάνει περισσότερα έτη παρατηρήσεων. Επίσης θα πρέπει να βελτιωθούν οι αρχιτεκτονικές των νευρωνικών δικτύων και να συλλέγονται δεδομένα κειμένου και από άλλες πηγές (π.χ. ειδήσεις) με απώτερο στόχο την αύξηση της απόδοσης πρόβλεψης των μοντέλων.

4.6 Bankruptcy prediction using imaged financial ratios and convolutional neural networks

Η παρούσα έρευνα (Hosaka, 2019) έχει ως **στόχο** την εφαρμογή του **μοντέλου Convolutional Neural Network (CNN)** για την πρόβλεψη της εταιρικής χρεοκοπίας χρησιμοποιώντας ως δεδομένα αριθμητικούς οικονομικούς δείκτες μετασηματισμένους σε εικόνες.

Τα **δεδομένα** προέρχονται από τις εταιρικές οικονομικές καταστάσεις και συγκεκριμένα τους ισολογισμούς και τις καταστάσεις αποτελεσμάτων. Συλλέχθηκαν από τη βάση δεδομένων Nikkei NEEDS Financial QUEST και αφορούν εισηγμένες επιχειρήσεις σε Ιαπωνικά χρηματιστήρια. Το σύνολο δεδομένων αποτελείται από 102 χρεοκοπημένες και 2062 υγιείς επιχειρήσεις για το χρονικό διάστημα 4 περιόδων (2012-2016).

Οι **λογιστικές μεταβλητές** που λήφθηκαν υπόψη είναι 14 και είναι οι εξής:

1. Current assets
2. Fixed assets
3. Current liabilities
4. Fixed liabilities
5. Net assets
6. Shareholders' equity
7. Retained earnings
8. Sales volume
9. Cost of sales
10. Gross profit
11. Operating profit
12. Ordinary profit
13. Net profit before taxes
14. Net profit

Η **εκπαίδευση** του μοντέλου CNN απαιτεί μεγάλο όγκο δεδομένων όμως ο αριθμός των εγγραφών των χρεοκοπημένων επιχειρήσεων ήταν μόνο 408 (102 επιχειρήσεις x 4 περιόδους = 408). Έτσι το σύνολο δεδομένων και των δύο κατηγοριών (χρεοκοπημένες-συνεχιζόμενες εταιρείες) αυξήθηκε με τη χρήση των τεχνικών interpolation και extrapolation και τελικά δημιουργήθηκαν 90 συνθετικά δεδομένα για κάθε εταιρεία.

Στη συνέχεια, το σύνολο δεδομένων, τόσο των πραγματικών δεδομένων των 4 περιόδων όσο και των συνθετικών, μετασηματίστηκε σε **εικόνες** της κλίμακας χρώματος γκρι. Έτσι κάθε οικονομικός δείκτης αντιστοιχίστηκε σε ένα συγκεκριμένο pixel, η φωτεινότητα του οποίου εξαρτιόταν από την τιμή του αντίστοιχου οικονομικού δείκτη.

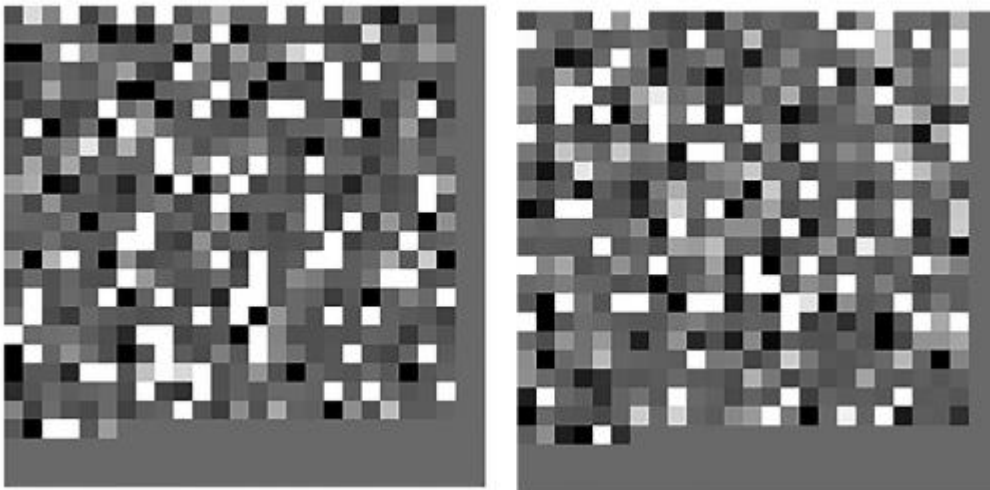
Για την αντιστοίχιση κάθε οικονομικού δείκτη σε ένα συγκεκριμένο pixel εφαρμόστηκαν δύο μέθοδοι, η μέθοδος Random και η μέθοδος Correlated. Σύμφωνα με τη **μέθοδο Random** η αντιστοίχιση των οικονομικών δεικτών σε θέσεις pixel γίνεται τυχαία, ενώ, αντίθετα, στη **μέθοδο Correlated** η αντιστοίχιση γίνεται ανάλογα με τον βαθμό συσχέτισης των οικονομικών δεικτών μεταξύ τους, δηλαδή όσο πιο σχετικοί είναι οι οικονομικοί δείκτες μεταξύ τους τόσο σε κοντινότερες θέσεις pixel τοποθετούνται.

Η φωτεινότητα ενός pixel i υπολογίστηκε με την παρακάτω εξίσωση:

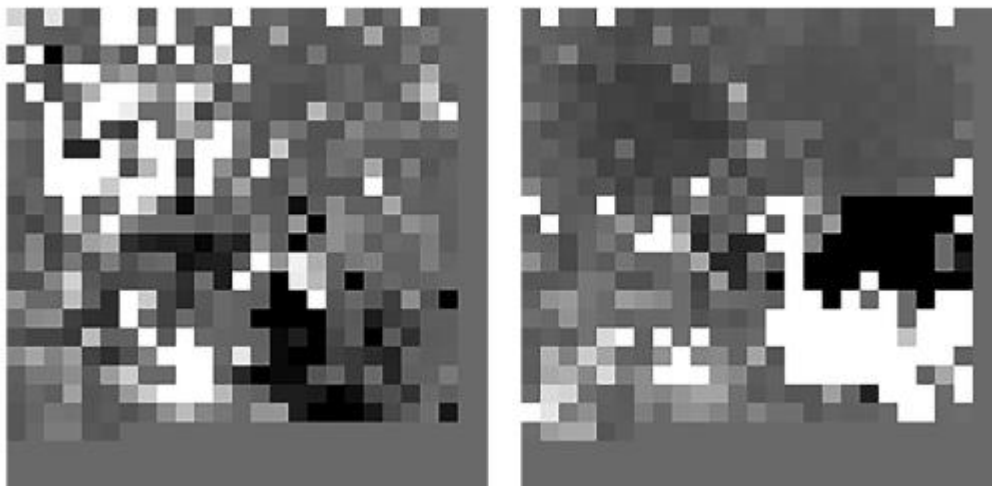
$$\frac{v[R(i)] - m[R(i)]}{\sigma[R(i)]} \times 100 + 128,$$

όπου $v[R(i)]$ η τιμή του οικονομικού δείκτη $R(i)$, $m[R(i)]$ ο μέσος όρος και $\sigma[R(i)]$ η τυπική απόκλιση. Το ανώτερο όριο φωτεινότητας ήταν 255 και το κατώτερο 0 ενώ για τις χαμένες τιμές η τιμή φωτεινότητας ορίστηκε σε 128. Έτσι αν ο οικονομικός δείκτης ήταν μεγαλύτερος του μέσου όρου του συνόλου δεδομένων τότε το αντίστοιχο pixel χρωματίζονταν σε απόχρωση πιο κοντά στο άσπρο και αν ήταν μικρότερος σε απόχρωση πιο κοντά στο μαύρο.

Οι δημιουργούμενες εικόνες (βλ. εικόνα 29) μεγεθύνθηκαν σε 256x256 pixels με τη χρήση της μεθόδου nearest neighbor, ώστε να είναι κατάλληλες ως είσοδοι στο GoogLeNet. Κάθε υποπεριοχή μεγέθους 10x10 pixel αντιστοιχούσε σε έναν συγκεκριμένο οικονομικό δείκτη. Η εικόνα 29α δείχνει την αντιστοιχία με βάση τη μέθοδο Random και η εικόνα 29β με βάση τη μέθοδο correlated. Οι δύο εικόνες στα αριστερά αφορούν μία χρεοκοπημένη εταιρεία και στα δεξιά μία συνεχιζόμενη εταιρεία. Ενώ στις κάτω και δεξιά άκρες κάθε εικόνας φαίνονται οι τιμές φωτεινότητας που είναι ίσες με 128 στις οποίες δεν αντιστοιχούν συγκεκριμένοι οικονομικοί δείκτες.



29α : Μέθοδος ‘Random’



29β : Μέθοδος ‘Correlated’

Εικόνα 29: Οι οικονομικοί δείκτες εκφρασμένοι σε εικόνες χρώματος κλίμακας γκρι

Πηγή: (Hosaka, 2019)

Στη συνέχεια, οι εικόνες που δημιουργήθηκαν, χρησιμοποιήθηκαν ως training dataset και testing dataset για την πρόβλεψη χρεοκοπίας.

Όμως ο αριθμός των χρεοκοπημένων εταιρειών είναι κατά πολύ μικρότερος σε σχέση με τον αριθμό των συνεχιζόμενων εταιρειών. Έτσι το μέγεθος του training data καθορίστηκε με βάση τις χρεοκοπημένες επιχειρήσεις και οι περισσότερες συνεχιζόμενες επιχειρήσεις δεν περιλήφθηκαν στο training dataset παρά μόνο στα test dataset, ενώ οι εικόνες που δημιουργήθηκαν με βάση το συνθετικό dataset χρησιμοποιήθηκαν μόνο για την εκπαίδευση. Επίσης εφαρμόστηκε το five-fold cross validation.

Αναλυτικότερα, από τις **102 χρεοκοπημένες επιχειρήσεις** επιλέχθηκαν τυχαία 20. Η διαδικασία επαναλήφθηκε μέχρι να δημιουργηθούν 5 υποσύνολα (A - E) με κάθε εταιρεία να ανήκει αποκλειστικά και μόνο σε ένα υποσύνολο ενώ οι 2 που περίσσευαν δεν ανήκαν σε κανένα υποσύνολο.

Αντίστοιχα, από τις **2062 συνεχιζόμενες επιχειρήσεις** επιλέχθηκαν τυχαία 20. Η διαδικασία επαναλήφθηκε μέχρι να δημιουργηθούν 5 υποσύνολα με κάθε εταιρεία να ανήκει αποκλειστικά και μόνο σε ένα υποσύνολο ενώ οι υπόλοιπες 1962 εταιρείες δεν ανήκαν σε κανένα. Τα δημιουργούμενα υποσύνολα προστέθηκαν στα υποσύνολα A - E.

Από τα πέντε υποσύνολα, τα τέσσερα χρησιμοποιήθηκαν ως training dataset και το ένα ως test dataset. Κάθε training dataset περιείχε 7520 εικόνες [$80 \text{ εταιρείες} * (4 \text{ περιόδους} + 90 \text{ συνθετικά δεδομένα})$] ενώ το test dataset 88 εικόνες των χρεοκοπημένων εταιρειών ($22 * 4 \text{ περιόδους}$) και 7928 εικόνες των συνεχιζόμενων [$(1962+20) * 4 \text{ περιόδους}$].

Η **εκπαίδευση και η αξιολόγηση** του μοντέλου CNN βασίστηκε στο GoogLeNet και υιοθετήθηκαν πέντε διαφορετικές δομές δικτύου με 27, 23, 17, 11 και 6 επίπεδα (layers). Ο αριθμός των εποχών εκπαίδευσης και το μέγεθος του mini-batch ορίστηκαν εμπειρικά σε 100 και 32, αντίστοιχα.

Στον πίνακα 29 φαίνονται τα αποτελέσματα κάθε dataset (A-E) στις πέντε διαφορετικές δομές δικτύου με τη μέθοδο Random και τη μέθοδο correlated. Παρατηρεί κανείς ότι οι συνεχιζόμενες επιχειρήσεις εμφανίζουν υψηλότερα ποσοστά και τα ποσοστά όλων βελτιώνεται όσο αυξάνεται ο αριθμός των layers.

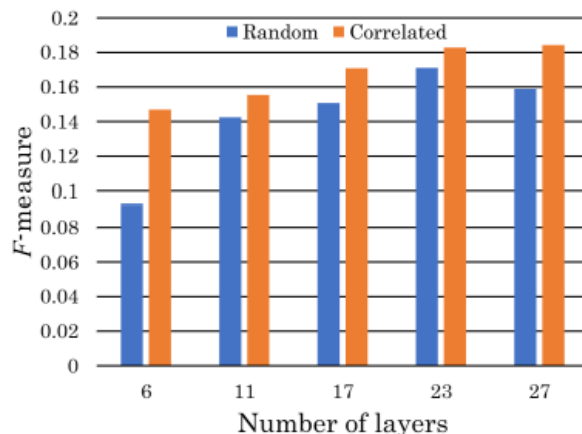
Πίνακας 29: Τα αποτελέσματα του προτεινόμενου μοντέλου με πέντε διαφορετικές δομές δικτύου

(a) Correspondence between financial ratios and pixels for the 'Random' method						
Dataset not used for learning	True class	Total number of layers				
		6	11	17	23	27
Dataset A	Bankrupt	0.830	0.852	0.852	0.830	0.773
	Continuing	0.807	0.885	0.914	0.925	0.920
Dataset B	Bankrupt	0.784	0.920	0.898	0.920	0.909
	Continuing	0.862	0.837	0.891	0.899	0.883
Dataset C	Bankrupt	0.727	0.909	0.920	0.830	0.875
	Continuing	0.839	0.872	0.897	0.905	0.907
Dataset D	Bankrupt	0.682	0.852	0.909	0.830	0.875
	Continuing	0.834	0.919	0.875	0.907	0.898
Dataset E	Bankrupt	0.830	0.920	0.909	0.898	0.898
	Continuing	0.838	0.897	0.868	0.909	0.890
Average	Bankrupt	0.770	0.891	0.898	0.861	0.866
	Continuing	0.836	0.882	0.889	0.909	0.900

(b) Correspondence between financial ratios and pixels for the 'Correlated' method						
Dataset not used for learning	True class	Total number of layers				
		6	11	17	23	27
Dataset A	Bankrupt	0.727	0.818	0.795	0.830	0.795
	Continuing	0.922	0.921	0.927	0.929	0.953
Dataset B	Bankrupt	0.830	0.864	0.886	0.932	0.909
	Continuing	0.882	0.879	0.900	0.884	0.903
Dataset C	Bankrupt	0.761	0.727	0.727	0.761	0.864
	Continuing	0.899	0.917	0.906	0.925	0.930
Dataset D	Bankrupt	0.716	0.886	0.898	0.875	0.875
	Continuing	0.897	0.923	0.930	0.949	0.925
Dataset E	Bankrupt	0.818	0.864	0.886	0.898	0.909
	Continuing	0.916	0.868	0.894	0.895	0.868
Average	Bankrupt	0.770	0.832	0.839	0.859	0.870
	Continuing	0.903	0.902	0.911	0.916	0.916

Πηγή: (Hosaka, 2019)

Επίσης έγινε αξιολόγηση της απόδοσης του μοντέλου με το μέτρο **F-measures** όπου οι χρεοκοπημένες εταιρίες θεωρούνται ως θετικές περιπτώσεις και οι συνεχιζόμενες ως αρνητικές, αφού στόχος της έρευνας ήταν ο εντοπισμός των χρεοκοπημένων εταιρειών. Σύμφωνα με το εικόνα 30, η μέθοδος Correlated επιτυγχάνει καλύτερα αποτελέσματα από τη μέθοδο Random άρα η απόδοση βελτιώνεται όταν η κατανομή των pixel γίνεται με υψηλά συσχετισμένους οικονομικούς δείκτες. Επίσης η διαφορά απόδοσης των δύο μεθόδων μειώνεται όταν ο αριθμός των layers είναι πάνω από 11 ενώ η απόδοση και των δύο μεθόδων αυξάνεται όσο αυξάνεται ο αριθμός των layers.



Εικόνα 30: Τα αποτελέσματα του μέτρου F-measure

Πηγή: (Hosaka, 2019)

Έπειτα έγινε επανάληψη του πειράματος όπου, αυτή τη φορά, χρησιμοποιήθηκαν μόνο οι εικόνες που δημιουργήθηκαν με τη μέθοδο correlated και ένα δίκτυο 23 επιπέδων. Ο αριθμός των συνεχιζόμενων εταιρειών μειώθηκε σε 15 και 10 εταιρείες με αποτέλεσμα το training dataset να αποτελείται από 5640 εικόνες [15 εταιρείες x 4 σετ x (4 περίοδοι + 90 συνθετικές τιμές)] και 3760 εικόνες [10 εταιρείες x 4 σετ x (4 περίοδοι + 90 συνθετικές τιμές)] αντίστοιχα. Οι εταιρείες που αφαιρέθηκαν από τα training dataset προστέθηκαν στο test dataset. Σύμφωνα λοιπόν με τον πίνακα 30, όταν ο αριθμός των συνεχιζόμενων εταιρειών μειώθηκε, η απόδοση πρόβλεψης των χρεοκοπημένων εταιρειών αυξήθηκε ενώ για τις συνεχιζόμενες επιχειρήσεις μειώθηκε. Μάλιστα η διαφορά των ποσοστών μεταξύ χρεοκοπημένων και συνεχιζόμενων εταιρειών είναι μικρότερη όταν ο αριθμός των συνεχιζόμενων επιχειρήσεων είναι 15.

Πίνακας 30: Τα αποτελέσματα του προτεινόμενου μοντέλου με 5640 και 3760 εικόνες συνεχιζόμενων εταιρειών

Dataset not used for learning	True class	Number of training data points for the continuing class	
		5640	3760
Dataset A	Bankrupt	0.784	0.886
	Continuing	0.915	0.837
Dataset B	Bankrupt	0.920	0.955
	Continuing	0.853	0.789
Dataset C	Bankrupt	0.875	0.898
	Continuing	0.870	0.817
Dataset D	Bankrupt	0.852	0.943
	Continuing	0.909	0.801
Dataset E	Bankrupt	0.920	0.909
	Continuing	0.863	0.822
Average	Bankrupt	0.870	0.918
	Continuing	0.882	0.813

Πηγή: (Hosaka, 2019)

Επίσης έγινε αξιολόγηση της απόδοσης με το μέτρο F-measures και για τις 15 συνεχιζόμενες εταιρείες ήταν 0,138 ενώ για τις 10 συνεχιζόμενες εταιρείες ήταν 0,096.

Σε επόμενο πείραμα επιλέχθηκαν τυχαία **150 συνεχιζόμενες επιχειρήσεις** για κάθε μία από τις ομάδες Α–Ε και ενσωματώθηκαν μόνο τα πραγματικά δεδομένα των 4

περιόδων. Ενώ τα δεδομένα των χρεοκοπημένων εταιρειών παρέμειναν αμετάβλητα. Επίσης τα training datasets των συνεχιζόμενων εταιρειών δεν περιείχαν συνθετικά δεδομένα. Χρησιμοποιήθηκαν οι εικόνες που δημιουργήθηκαν με τη μέθοδο correlated και το δίκτυο με 23 layers. Ο πίνακας 31 δείχνει ότι τα αποτελέσματα πρόβλεψης των συνεχιζόμενων εταιρειών είναι σχεδόν ίδια με αυτά του πίνακα 29 ενώ τα αποτελέσματα των χρεοκοπημένων εταιρειών βελτιώνονται σημαντικά.

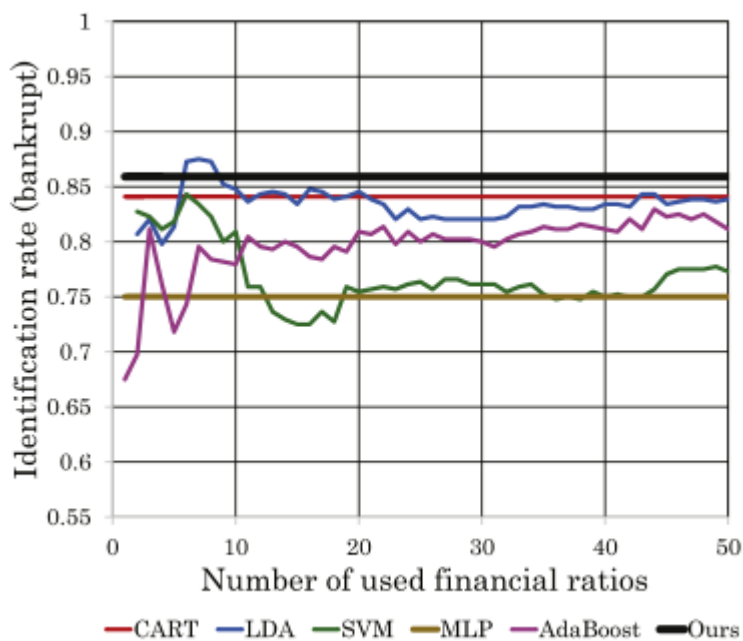
Πίνακας 31: Τα αποτελέσματα του προτεινόμενου μοντέλου με 600 εικόνες συνεχιζόμενων εταιρειών

Dataset not used for learning	True class	Number of training data points for the continuing class
		600
Dataset A	Bankrupt	0.841
	Continuing	0.932
Dataset B	Bankrupt	0.943
	Continuing	0.879
Dataset C	Bankrupt	0.920
	Continuing	0.918
Dataset D	Bankrupt	0.841
	Continuing	0.929
Dataset E	Bankrupt	0.898
	Continuing	0.910
Average	Bankrupt	0.889
	Continuing	0.914

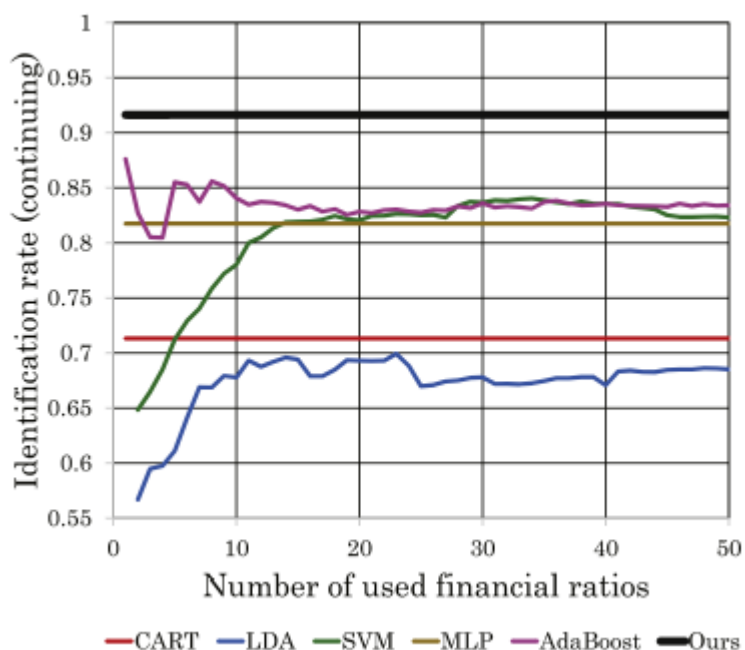
Πηγή: (Hosaka, 2019)

Επίσης το μέτρο **F-measure** για τις 150 εταιρείες βελτιώθηκε και έγινε 0,189.

Έπειτα έγινε σύγκριση της προτεινόμενης μεθόδου με τα μοντέλα **Classification And Regression Trees (CART)**, **Linear Discriminant Analysis (LDA)**, **Support Vector Machine (SVM)**, **Multi-Layer Perceptrons (MLP)** και **AdaBoost**. Χρησιμοποιήθηκαν οι εικόνες που δημιουργήθηκαν με τη μέθοδο correlated και το δίκτυο με 23 layers. Στην εικόνα 31 απεικονίζονται τα αποτελέσματα κάθε μεθόδου τόσο για τις χρεοκοπημένες όσο και τις συνεχιζόμενες εταιρείες. Ο κάθετος άξονας απεικονίζει τα ποσοστά πρόβλεψης στα 5 test datasets και ο οριζόντιος τον αριθμό των οικονομικών δεικτών. Οι μέθοδοι CART, MLP και η προτεινόμενη μέθοδος απεικονίζονται με ευθεία γραμμή, διότι δεν περιλαμβάνουν τη διαδικασία επιλογής χαρακτηριστικών. Σύμφωνα με τα αποτελέσματα, η προτεινόμενη μέθοδος υπερτερεί έναντι των υπολοίπων όσον αφορά τις συνεχιζόμενες εταιρείες ενώ για τις χρεοκοπημένες η μέθοδος LDA υπερτερεί σε σχέση με την προτεινόμενη σε ορισμένη περιοχή.



(α) Αποτελέσματα χρεοκοπημένων εταιρειών



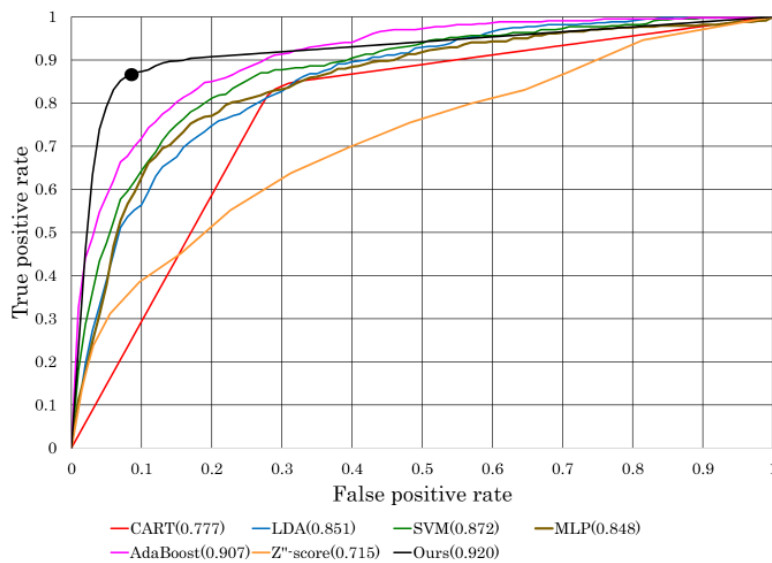
(β) Αποτελέσματα συνεχιζόμενων εταιρειών

Εικόνα 31: Τα αποτελέσματα πρόβλεψης των μοντέλων CART, LDA, SVM, MLP, AdaBoost και της προτεινόμενης μεθόδου

Πηγή: (Hosaka, 2019)

Τέλος, έγινε σύγκριση των μοντέλων με την **καμπύλη ROC** (βλ. εικόνα 32). Σύμφωνα με το διάγραμμα, η προτεινόμενη μέθοδος υπερτερεί δεδομένου ότι η καμπύλη διέρχεται από το πάνω αριστερό τμήμα του διαγράμματος που αντιστοιχεί σε υψηλό “true positive rate” και χαμηλό “false positive rate”. Επιπλέον λήφθηκε

υπόψη το Z-score του Altman το οποίο όμως εμφανίζει χαμηλότερη απόδοση σε σύγκριση με τα υπόλοιπα μοντέλα.



Εικόνα 32: Η καμπύλη ROC

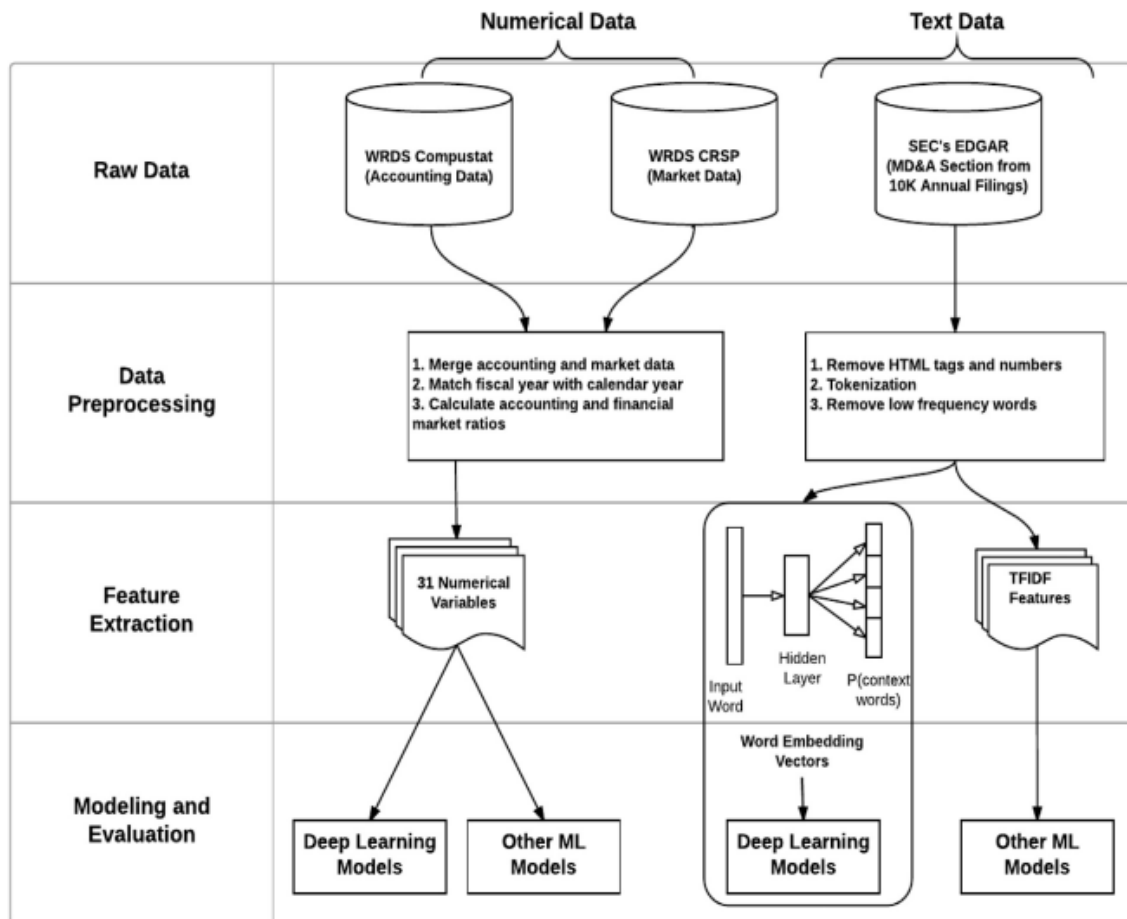
Πηγή: (Hosaka, 2019)

Σύμφωνα με τα παραπάνω, τα πειράματα έδειξαν ότι η προγνωστική ακρίβεια βελτιώνεται με τη χρήση της μεθόδου correlated και επίσης ότι η προτεινόμενη μέθοδος υπερτερεί έναντι των υπολοίπων παραδοσιακών μεθόδων με τα οποία συγκρίθηκε. Παρ' όλα αυτά ο Hosaka (2019) επισημαίνει ότι υπάρχουν σημεία της έρευνας που θα πρέπει να διερευνηθούν περαιτέρω, όπως να διερευνηθεί ποιοι οικονομικοί δείκτες είναι ισχυρότεροι για την πρόβλεψη χρεοκοπίας, ποια άλλα δίκτυα θα μπορούσαν να χρησιμοποιηθούν για την αποτελεσματικότερη πρόβλεψη, οι αναλογίες των πραγματικών και συνθετικών δεδομένων και επίσης να επαληθευτεί αν η απόδοση του μοντέλου βελτιώνεται και με τη χρήση άλλων μεθόδων αντιστοίχισης των οικονομικών δεικτών σε pixel εκτός της μεθόδου correlated.

4.7 *Deep learning models for bankruptcy prediction using textual disclosures*

Στην έρευνα τους οι Mai, Tian, Lee και Ma (2019) συγκρίνουν την απόδοση των μοντέλων average embedding model (DL-Embedding) και convolutional neural network (CNN) για την πρόβλεψη της εταιρικής χρεοκοπίας με τη χρήση δεδομένων κειμένου (textual disclosures) και εξετάζουν ποιες αρχιτεκτονικές δικτύων βελτιώνουν την απόδοση της πρόβλεψης συνδυάζοντας διαφορετικά neural network layers.

Η ροή των διαδικασιών της έρευνας απεικονίζεται στην εικόνα 33:



Εικόνα 33: Η ροή των διαδικασιών της έρευνας

Πηγή: (Mai κ.ά., 2019)

Τα **δεδομένα** συλλέχθηκαν από τρεις πηγές δεδομένων: (1) τα λογιστικά δεδομένα συλλέχθηκαν από το Compustat North America, (2) τα δεδομένα συναλλαγών μετοχών από το Center for Research in Security Prices (CRSP) και (3) τα δεδομένα κειμένου από τις ετήσιες εκθέσεις που κατατέθηκαν στο Securities Exchange Commission (SEC). Το σύνολο δεδομένων αφορά το χρονικό διάστημα 1994-2014 και περιλαμβάνει συνολικά 11.827 εταιρείες, 94.994 εταιρικά έτη και 477 καταθέσεις πτώχευσης. Ο Πίνακας 32 δείχνει την κατανομή των εταιρειών ανά έτος.

Πίνακας 32: Η κατανομή των επιχειρήσεων ανά έτος

Year	Total Firms	Bankrupted Firms	Year	Total Firms	Bankrupted Firms
1994	1783	6	2005	4475	5
1995	3334	22	2006	4451	10
1996	6117	52	2007	4417	38
1997	6183	52	2008	4209	48
1998	5989	57	2009	4036	16
1999	5846	22	2010	3930	12
2000	5567	18	2011	3841	7
2001	5158	36	2012	3799	8
2002	4857	30	2013	3865	8
2003	4579	14	2014	4019	3
2004	4539	13			

Πηγή: (Mai κ.ά., 2019)

Στη συνέχεια, ακολούθησε η **προεπεξεργασία** των δεδομένων. Πιο συγκεκριμένα, κατά την προεπεξεργασία των αριθμητικών δεδομένων, έγινε (1) συγχώνευση των λογιστικών δεδομένων και των δεδομένων της αγοράς, (2) αντιστοίχιση του οικονομικού έτους με το ημερολογιακό και (3) υπολογισμός των λογιστικών και χρηματοοικονομικών δεικτών. Οι ακραίες τιμές και τα σφάλματα αντιμετωπίστηκαν με τη μέθοδο winsorize 1% & 99%, δηλαδή οι τιμές που ήταν χαμηλότερες από το 1% και υψηλότερες από το 99% αντικαταστάθηκαν από το 1^ο και 99^ο εκατοστημόριο. Οι 36 αριθμητικές μεταβλητές που λήφθηκαν υπόψη απεικονίζονται στον παρακάτω πίνακα (βλ. πίνακα 33).

Πίνακας 33: Οι αριθμητικές μεταβλητές

Variable	Description	Variable	Description
ACTLCT	Current Assets/Current Liabilities	LTMTA	Total Liabilities/(Market Equity + Total Liabilities)
APSALE	Accounts Payable/Sales	LOG(AT)	Log(Total Assets)
CASHAT	Cash and Short-term Investment/Total Assets	LOG(SALE)	Log(Sale)
CASHMTA	Cash and Short-term Investment/(Market Equity + Total Liabilities)	MB	Market-to-Book Ratio
CHAT	Cash/Total Assets	NIAT	Net Income/Total Asset
CHLCT	Cash/Current Liabilities	NIMTA	Net Income/(Market Equity + Total Liabilities)
(EBIT+DP)/AT	(Earnings before Interest and Tax + Amortization and Depreciation)/Total Asset	NISALE	Net Income/Sales
EBITAT	Earnings before Interest and Tax/Total Asset	OIADPAT	Operating Income/Total Asset
EBITSALE	Earnings before Interest and Tax/Sales	OIADPSALE	Operating Income/Sales
EXCESS RETURN	Excess Return Over S&P 500 Index	PRICE	Log(Price)
FAT	Total Debts/Total Assets	QALCT	Quick Assets/Current Liabilities
INVCHINVT	Growth of Inventories /Inventories	REAT	Retained Earnings/Total Asset
INVSALE	Inventories/Sales	RELCT	Retained Earnings/Current Liabilities
(LCT-CH)/AT	(Current Liabilities - Cash)/Total Asset	RSIZE	Log(Market Capitalization)
LCTAT	Current Liabilities/Total Asset	SALEAT	Sales/Total Assets
LCTLT	Current Liabilities/Total Liabilities	SEQAT	Equity/Total Asset
LCTSALE	Current Liabilities/Sales	SIGMA	Stock Volatility
LTAT	Total Liabilities/Total Assets	WCAPAT	Working Capital/Total Assets

Πηγή: (Mai κ.ά., 2019)

Τα **δεδομένα κειμένου** αντλήθηκαν από την ενότητα Management Discussion and Analysis (MD&A) των ετήσιων εκθέσεων με τη χρήση των Perl scripts και (1) αφαιρέθηκαν τα HTML tags και τα νούμερα, (2) εφαρμόστηκε η διαδικασία tokenization και (3) αφαιρέθηκαν οι λέξεις χαμηλής συχνότητας και τελικά λήφθηκαν υπόψη οι 20.000 πιο συχνές λέξεις.

Στο επόμενο στάδιο, έγινε η εξαγωγή των χαρακτηριστικών προκειμένου να μετατραπούν οι λέξεις σε διανύσματα πραγματικής αξίας. Για τον υπολογισμό των διανυσμάτων ενσωμάτωσης λέξεων, χρησιμοποιήθηκε το μοντέλο skip-gram με τον αλγόριθμο negative sampling και εφαρμόστηκε η κανονικοποίηση padding, ώστε τα MD&A να έχουν το ίδιο μήκος (7500). Στη συνέχεια, τα διανύσματα ενσωμάτωσης χρησιμοποιήθηκαν ως είσοδος στα μοντέλα deep learning (A) average embedding model (DL-Embedding) και convolutional neural network (CNN).

Για την **υλοποίηση των μοντέλων**, χρησιμοποιήθηκε το πακέτο Keras 2.0 που τρέχει σε TensorFlow backend. Για την εκπαίδευση χρησιμοποιήθηκαν ο αλγόριθμος backpropagation, η συνάρτηση απώλειας cross-entropy και ο βελτιστοποιητής Stochastic gradient descent. Το batch size ορίστηκε σε 32 και για την αποφυγή του overfitting εφαρμόστηκαν η κανονικοποίηση L2 regularization για τη ρύθμιση των βαρών, η τεχνική dropout για την τυχαία παράλειψη ενός υποσυνόλου σε κάθε επανάληψη και η τεχνική early stopping για την παρακολούθηση της απόδοσης του validation set και την έγκαιρη διακοπή της εκπαίδευσης.

Για την **αξιολόγηση των μοντέλων**, τα δεδομένα χωρίστηκαν σε training data set και testing dataset με δύο τρόπους. Σύμφωνα με τον **πρώτο τρόπο**, το σύνολο δεδομένων χωρίστηκε τυχαία όπου το 80% των δεδομένων ήταν το training dataset και

το υπόλοιπο 20% το testing dataset. Ενώ σύμφωνα με τον **δεύτερο τρόπο**, τα δεδομένα πριν το έτος 2007 αποτέλεσαν το training dataset και τα δεδομένα μετά το έτος 2008 το testing dataset.

Για τη μέτρηση της ικανότητας πρόβλεψης των μοντέλων, χρησιμοποιήθηκαν οι μετρήσεις AUC, accuracy ratio και cumulative decile-ranking.

Αρχικά, έγιναν δύο πειράματα για την απόδοση πρόβλεψης για ένα έτος πριν τη χρεοκοπία.

Στο πρώτο πείραμα λήφθηκαν υπόψη μόνο οι ενότητες **MD&A** των ετήσιων εκθέσεων και έγινε σύγκριση των μοντέλων deep learning DL-Embedding και DL-CNN με τα μοντέλα Logistic regression, SVM και Random forest.

Σύμφωνα με τα αποτελέσματα του πίνακα 34, οι τιμές AUC είναι σταθερά πάνω από 0,7 σε όλα τα μοντέλα. Το μοντέλο DL-Embedding έχει την καλύτερη απόδοση σε όλες τις μετρήσεις και με τις δύο μεθόδους διαχωρισμού των δεδομένων. Αναλυτικότερα, σύμφωνα με τον τυχαίο διαχωρισμό των δεδομένων, το μοντέλο **DL-Embedding** έχει τη μεγαλύτερη απόδοση στο AUC (0,784) και είναι το μόνο με accuracy ratio πάνω από 0,5. Επίσης μπορεί να προβλέψει σωστά το 35,7% των μελλοντικών πτωχεύσεων στο top decile. Ενώ η απόδοση του μοντέλου DL-CNN είναι παρόμοια με την απόδοση των υπολοίπων μοντέλων. Σύμφωνα με τον δεύτερο τρόπο διαχωρισμού των δεδομένων, το μοντέλο **DL-Embedding** εμφανίζει επίσης την υψηλότερη τιμή AUC (0,760) ενώ το μοντέλο **DL-CNN** έχει τη χειρότερη τιμή AUC σε σύγκριση με τα υπόλοιπα. Συμπερασματικά, το μοντέλο DL-Embedding είναι η καλύτερη αρχιτεκτονική deep learning για την πρόβλεψη χρεοκοπίας.

Επίσης για την επιβεβαίωση των αποτελεσμάτων, χρησιμοποιήθηκε το Salzberg binomial test προκειμένου να γίνει έλεγχος της διαφοράς απόδοσης των δύο καλύτερων μοντέλων του πειράματος, δηλαδή του μοντέλου DL-Embedding και του Logistic regression. Τα αποτελέσματα του binomial test επιβεβαίωσαν τα αποτελέσματα του πίνακα 34.

Πίνακας 34: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση μόνο δεδομένων κειμένου

Panel A: Random Split					
	DL-Embedding	DL-CNN	Logistic Regression	SVM	Random Forest
Accuracy ratio	0.568	0.428	0.434	0.422	0.433
AUC	0.784	0.714	0.717	0.711	0.716
1	0.357	0.250	0.297	0.297	0.321
2	0.559	0.440	0.487	0.499	0.464
3	0.714	0.559	0.594	0.570	0.595
4	0.821	0.738	0.736	0.724	0.690
5	0.881	0.809	0.807	0.795	0.833
6-10	1	1	1	1	1
Panel B: Split by Year					
	DL-Embedding	DL-CNN	Logistic Regression	SVM	Random Forest
Accuracy ratio	0.521	0.403	0.434	0.432	0.419
AUC	0.760	0.701	0.717	0.716	0.710
1	0.424	0.326	0.315	0.315	0.380
2	0.565	0.478	0.457	0.457	0.489
3	0.728	0.609	0.587	0.587	0.609
4	0.783	0.685	0.685	0.685	0.696
5	0.837	0.739	0.783	0.783	0.707
6-10	1.000	1.000	1.000	1.000	1.000

Πηγή: (Mai κ.ά., 2019)

Στο δεύτερο πείραμα λήφθηκαν υπόψη μόνο οι **αριθμητικές μεταβλητές** και έγινε σύγκριση των μοντέλων deep learning: (α) το DL-1 layer model με ένα κρυφό επίπεδο 4 νευρώνων, (β) DL-Deep model με 4 επίπεδα κρυφών μονάδων όπου κάθε επίπεδο αποτελείται από 4 νευρώνες και το (γ) DL-Wide model με ένα κρυφό επίπεδο 16 νευρώνων με τα μοντέλα Logistic regression, SVM και Random forest.

Σύμφωνα με τα αποτελέσματα του πίνακα 35 και τον τυχαίο διαχωρισμό των δεδομένων, διαπιστώνει κανείς ότι μεταξύ των τριών μοντέλων deep learning, το μοντέλο **DL-1** έχει τις υψηλότερες αποδόσεις με AUC (0,817), accuracy ratio (0,633) και decile 1 & 2 (0,405 και 0,655 αντίστοιχα). Ενώ κατά τη σύγκριση όλων των μοντέλων μεταξύ τους, το **Random forest** έχει την καλύτερη απόδοση με AUC (0,818) και accuracy rate (0,636) ενώ το μοντέλο **SVM** το καλύτερο decile 1 (0,429). Σύμφωνα με τον δεύτερο τρόπο διαχωρισμού των δεδομένων, το **DL-1** έχει την καλύτερη απόδοση στο decile 1 (0,587) αλλά το **Random forest** έχει τις υψηλότερες αποδόσεις σε AUC (0,814) και accuracy rate (0,629).

Επίσης εφαρμόστηκε το Salzberg binomial test και επιβεβαιώθηκαν τα αποτελέσματα του πειράματος. Επομένως, με τη χρήση μόνο αριθμητικών δεδομένων, τα παραδοσιακά μοντέλα Random forest και SVM είναι εξίσου ικανά για την πρόβλεψη και δεν υπάρχει λόγος χρήσης των μοντέλων deep learning.

Πίνακας 35: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση μόνο αριθμητικών δεδομένων

Panel A: Random Split						
	DL-1 Layer	DL-Deep	DL-Wide	Logistic Regression	SVM	Random Forest
Accuracy Ratio	0.633	0.603	0.597	0.616	0.619	0.636
AUC (%)	0.817	0.802	0.798	0.808	0.810	0.818
1	0.405	0.345	0.238	0.369	0.429	0.393
2	0.655	0.547	0.559	0.583	0.643	0.679
3	0.798	0.785	0.797	0.809	0.798	0.810
4	0.917	0.880	0.916	0.892	0.858	0.846
5	0.953	0.963	0.952	0.952	0.941	0.929
6-10	1	1	1	1	1	1

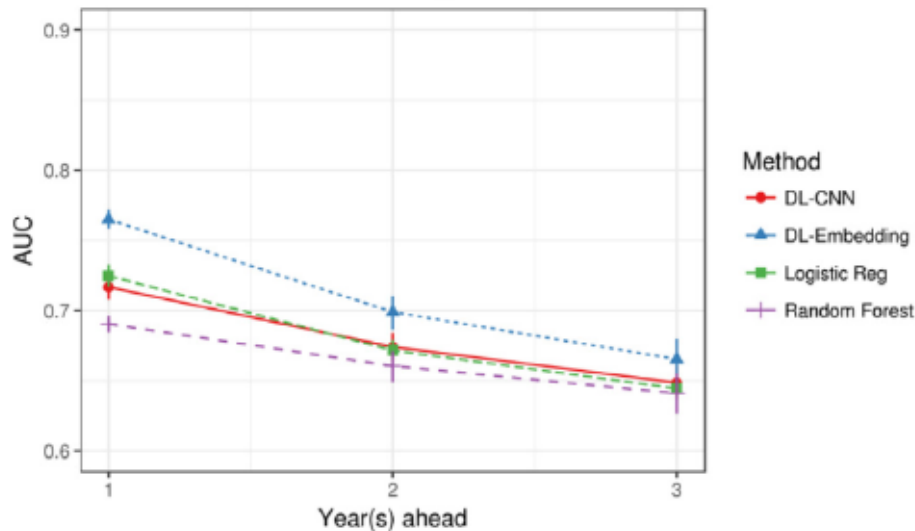
Panel B: Split by Year						
	DL-1 Layer	DL-Deep	DL-Wide	Logistic Regression	SVM	Random Forest
Accuracy Ratio	0.614	0.554	0.623	0.542	0.602	0.629
AUC (%)	0.807	0.777	0.811	0.771	0.801	0.814
1	0.587	0.413	0.457	0.467	0.533	0.467
2	0.717	0.576	0.707	0.685	0.750	0.728
3	0.804	0.837	0.804	0.750	0.783	0.783
4	0.848	0.891	0.870	0.804	0.815	0.815
5	0.880	0.935	0.891	0.859	0.870	0.826
6-10	1	1	1	1	1	1

Πηγή: (Mai κ.ά., 2019)

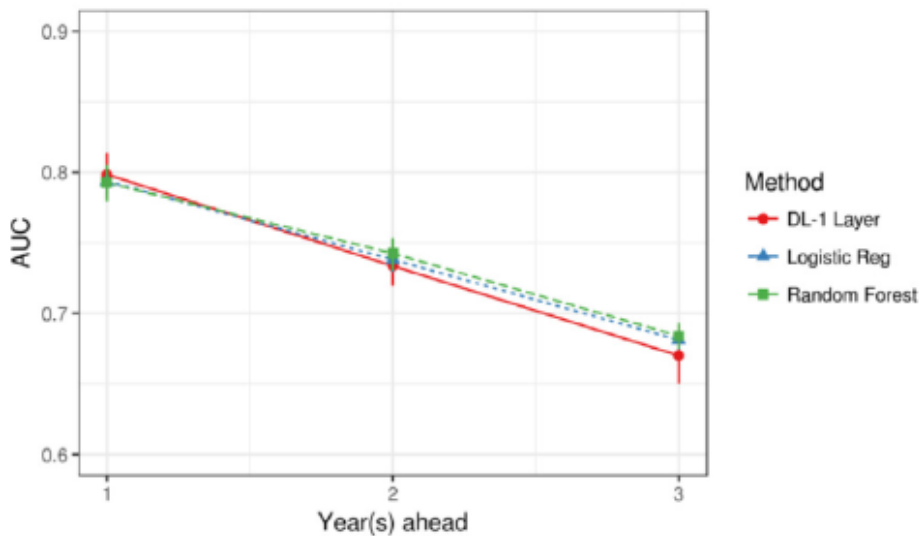
Στη συνέχεια, πραγματοποιήθηκαν πειράματα για την εξέταση της απόδοσης των μοντέλων για 2 και 3 έτη πριν τη χρεοκοπία. Έτσι δημιουργήθηκαν νέα datasets που περιλαμβάνουν δεδομένα για 2 και 3 έτη και έγινε χρήση της τεχνικής 10x10 cross-validation για την επικύρωση των μοντέλων. Σύμφωνα με τα αποτελέσματα (βλ. εικόνα 34), για τα δεδομένα κειμένου, το μοντέλο **DL-Embedding** συνεχίζει να έχει την

καλύτερη απόδοση σε σύγκριση με τα μοντέλα Logistic regression και Random forest ενώ, για τα αριθμητικά δεδομένα, τα **παραδοσιακά μοντέλα** είναι περισσότερο ικανά. Επίσης η απόδοση της πρόβλεψης με τη χρήση των αριθμητικών δεδομένων, για ένα και δύο έτη πριν, είναι καλύτερη σε σχέση με αυτή των δεδομένων κειμένου. Τέλος παρατηρείται ότι όσο αυξάνεται ο χρονικός ορίζοντας, τόσο μειώνεται η απόδοση του μέτρου AUC καθώς και το χάσμα απόδοσης μεταξύ των μοντέλων.

(α) Χρήση δεδομένων κειμένου



(β) Χρήση αριθμητικών δεδομένων



Εικόνα 34: Σύγκριση απόδοσης σε διαφορετικό χρονικό ορίζοντα

Πηγή: (Mai κ.ά., 2019)

Στη συνέχεια, έγινε ένα ακόμα πείραμα για τη διερεύνηση της απόδοσης πρόβλεψης των μοντέλων με τη χρήση τόσο δεδομένων κειμένου όσο και αριθμητικών δεδομένων. Έγινε σύνδεση του μοντέλου average embedding model με το μοντέλο DL-1 και δημιουργήθηκε μια νέα αρχιτεκτονική deep learning (DL-Embedding+DL-1 layer) και συγκρίθηκε με τα μοντέλα Logistic regression και Random forest.

Σύμφωνα με τα αποτελέσματα του πίνακα 36, διαπιστώνεται ότι το **νέο μοντέλο** deep learning σημειώνει τις καλύτερες αποδόσεις. Πιο αναλυτικά, με τον τυχαίο διαχωρισμό δεδομένων, η τιμή AUC είναι 0,856 ενώ με τη δεύτερη μέθοδο διαχωρισμού δεδομένων είναι 0,842. Αντίθετα, για το μοντέλο **Logistic regression**, οι τιμές AUC μειώνονται και μάλιστα είναι χειρότερες σε σύγκριση με τα αποτελέσματα του πίνακα 35 όπου έγινε χρήση μόνο αριθμητικών δεδομένων γεγονός που δείχνει ότι για τις μικτές εισόδους δεδομένων το μοντέλο δεν μπορεί να επιλέξει τα πιο σχετικά χαρακτηριστικά από μόνο του.

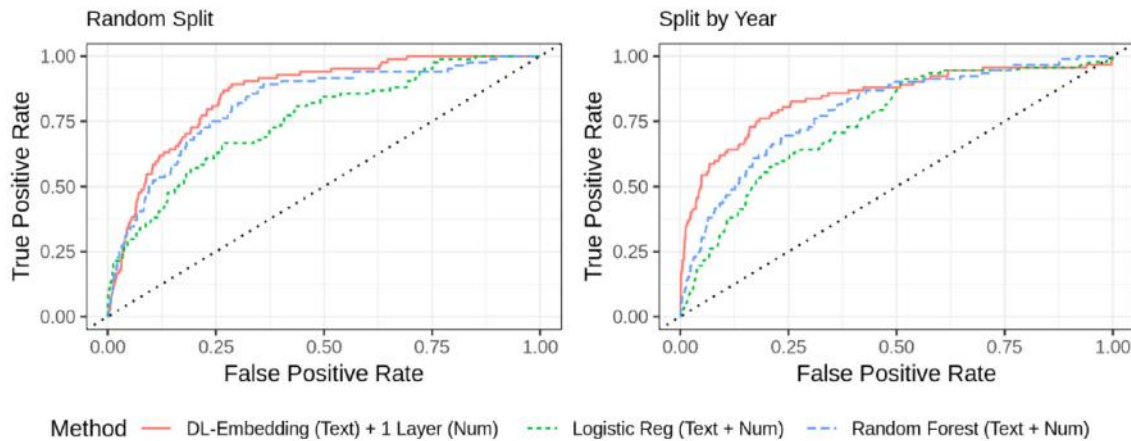
Πίνακας 36: Τα αποτελέσματα απόδοσης ένα χρόνο πριν τη χρεοκοπία με τη χρήση αριθμητικών δεδομένων και δεδομένων κειμένου

Panel A: Random split			
	DL-Embedding + DL-1 layer	Logistic regression	Random forest
Accuracy Ratio	0.712	0.507	0.639
AUC (%)	0.856	0.753	0.819
1	0.547	0.369	0.511
2	0.725	0.571	0.677
3	0.891	0.666	0.807
4	0.926	0.725	0.890
5	0.937	0.832	0.913
6-10	1	1	1

Panel B: Split by year			
	DL-Embedding + DL-1 Layer	Logistic regression	Random forest
Accuracy Ratio	0.685	0.491	0.585
AUC (%)	0.842	0.745	0.793
1	0.587	0.326	0.446
2	0.750	0.554	0.62
3	0.826	0.641	0.728
4	0.870	0.728	0.837
5	0.880	0.848	0.902
6-10	1	1	1

Πηγή: (Mai κ.ά., 2019)

Επίσης στην εικόνα 35 έγινε σύγκριση των καμπυλών ROC κάθε μοντέλου (DL-Embedding+DL-1 Layer, Logistic regression και Random forest) όπου επιβεβαιώθηκε η ανωτερότητα του μοντέλου DL-Embedding+DL-1 Layer έναντι των υπολοίπων.



Εικόνα 35: Σύγκριση των καμπυλών ROC κάθε μοντέλου

Πηγή: (Mai κ.ά., 2019)

Επίσης εφαρμόστηκε το Salzberg binomial test και τα αποτελέσματά του επιβεβαίωσαν τα αποτελέσματα του πειράματος.

Σύμφωνα με τα παραπάνω διαπιστώνεται ότι η μέθοδος deep learning μπορεί να επιτύχει πολύ μεγαλύτερη ακρίβεια πρόβλεψης με την ενσωμάτωση δεδομένων κειμένου και αριθμητικών δεδομένων σε σχέση με τα μοντέλα που χρησιμοποιούν μόνο έναν τύπο δεδομένων. Όμως οι ερευνητές επισημαίνουν την ύπαρξη αρκετών περιορισμών που θα πρέπει να ερευνηθούν μελλοντικά. Πιο συγκεκριμένα, προτείνουν να διερευνηθεί περισσότερο η αποτελεσματικότερη ερμηνεία των αδόμητων δεδομένων και η απόδοση και άλλων μοντέλων deep learning (π.χ. Long Short Term Memory). Επίσης προτείνουν να γίνει συλλογή δεδομένων και από άλλες πηγές δεδομένων (π.χ. ρεπορτάζ ειδήσεων, περιεχόμενο που δημιουργείται από τους χρήστες) και να ληφθούν υπόψη και άλλα μέτρα αξιολόγησης της απόδοσης (όπως H-measure και Kolmogorov-Smirnov goodness-of-fit test statistics).

4.8 A Deep Dense Neural Network for Bankruptcy Prediction

Οι Alexandropoulos, Aridas, Kotsiantis και Vrahatis (2019) ασχολήθηκαν με τη μελέτη της αποτελεσματικότητας ενός dense deep neural network και συγκεκριμένα του Deep Dense Multilayer Perceptron (DDMP) για την πρόβλεψη της πιθανότητας χρεοκοπίας.

Τα δεδομένα που χρησιμοποιήθηκαν συλλέχθηκαν από την Εθνική τράπεζα της Ελλάδος και τη βάση δεδομένων της ICAP. Συγκεκριμένα λήφθηκαν υπόψη οι πτωχευτικές καταθέσεις των ετών 2003 και 2004 και τα οικονομικά δεδομένα που συλλέχθηκαν αφορούσαν χρονική περίοδο 3 ετών όπου ως έτος **0** ορίστηκε το έτος χρεοκοπίας, ως έτος **1** το πρώτο έτος πριν τη χρεοκοπία και ως έτος **2** τα δύο έτη πριν τη χρεοκοπία. Το τελικό σύνολο δεδομένων περιείχε 450 εταιρικά έτη παρατηρήσεων και αφορούσε συνολικά 150 εταιρείες από τις οποίες οι 50 είναι χρεοκοπημένες και οι 100 υγιείς. Οι τομείς δραστηριότητας των χρεοκοπημένων εταιρειών συνοψίζονται στον κάτωθι πίνακα :

Πίνακας 37: Κλάδοι δραστηριότητας χρεοκοπημένων εταιρειών

Industry	Year 2003	Year 2004	Total
Advertisement	1	2	3
Agriculture and Farming	1	0	1
Clothing	2	2	4
Constructions	2	0	2
Electronics Equipment	0	1	1
Food	0	2	2
Freight Forwarding	1	0	1
Health Services	0	1	1
Industrial Minerals	0	1	1
Information Technology	0	1	1
Logistics	0	1	1
Machinery	0	2	2
Metal Products	1	0	1
Motor Vehicle Trade & Maintenance	1	0	1
Other Services	0	1	1
Plastic and Rubber	0	1	1
Private Education	1	0	1
Publishing & Printing	1	0	1
Restaurants	0	1	1
Retail Trade	3	7	10
Supermarkets	0	1	1
Telecommunications	0	2	2
Textiles	3	1	4
Wholesale Trade	2	4	6
Total	19	31	50

Πηγή: (Alexandropoulos κ.ά., 2019)

Επίσης χρησιμοποιήθηκαν 21 οικονομικές μεταβλητές εισόδου (βλ. πίνακα 38):

Πίνακας 38: Οι 21 οικονομικές μεταβλητές της έρευνας

Class	Variables	Short description
Profitability	OPIMAR	Operating income divided by net sales
	NIMAR	Net income divided by sales
	GIMAR	Gross income divided by sales
	ROCE	Net income pre tax divided by capital employed
	ROE	Net income pre tax divided by shareholder's equity capital
Liquidity-Leverage	EQ/CE	Shareholder's equity to capital employed
	CE/NFA	Capital employed to net fixed assets
	TD/EQ	Total debt to shareholder's equity capital
	CA/CL	Current assets to current liabilities
	QA/CL	Quick assets to current liabilities
	WC/TA	Working capital divided by total assets
Efficiency	COLPER	Average collection period for receivables
	INVTURN	Average turnover period for inventories
	PAYPER	Average payment period to creditors
	S/EQ	Sales divided by Shareholder's equity capital
	S/CE	Sales divided by capital employed
	S/TA	Sales divided by total assets
Growth	GRTA	Growth rate of total assets $(TAt - TAt - 1) / (ABS(TAt) + ABS(TAt - 1))$
	GRNI	Growth rate of net income
	GRNI	Growth rate of net sales
Size	SIZE	Size of firm is the $\ln(\text{Total assets}/\text{GDP price index})$

Πηγή: (Alexandropoulos κ.ά., 2019)

Για την πρόβλεψη της χρεοκοπίας χρησιμοποιήθηκε το μοντέλο Deep Dense Multilayer Perceptron (DDMP) με 2 κρυφά επίπεδα. Στο πρώτο κρυφό επίπεδο, ο αριθμός των νευρώνων καθορίστηκε στα 2/3 του αριθμού των χαρακτηριστικών εισόδου, ενώ στο δεύτερο κρυφό επίπεδο καθορίστηκε στο 1/3 του αριθμού των χαρακτηριστικών εισόδου προκειμένου να επιτευχθεί η γενίκευση του δικτύου, να αποφευχθεί η υπερπροσαρμογή και ο μεγάλος χρόνος εκπαίδευσης. Επίσης χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLU, η τεχνική Drop-out (10%) και η συνάρτηση απώλειας LOSSBinaryXENT.

Στη συνέχεια, έγινε σύγκριση του προτεινόμενου μοντέλου με τα μοντέλα Logistic Regression (LR), simple Multilayer Perceptron model with one hidden layer (MP), Naive Bayes (NB) και Cart όπου χρησιμοποιήθηκε η βιβλιοθήκη Keras και τα διαθέσιμα implementations της βιβλιοθήκης Scikit-learn.

Δεδομένου ότι το σύνολο δεδομένων είναι ανισόρροπο, για τη σύγκριση των μοντέλων χρησιμοποιήθηκε το μέτρο απόδοσης AUC και τα αποτελέσματα απεικονίζονται στον παρακάτω πίνακα :

Πίνακας 39: Τα αποτελέσματα απόδοσης του μέτρου AUC

	Cart	NB	LR	MP	DDMP
2 years before	0.532	0.579	0.586	0.584	0.627
1 year before	0.539	0.588	0.643	0.605	0.664
Last year	0.671	0.647	0.646	0.648	0.732

Πηγή: (Alexandropoulos κ.ά., 2019)

Από τα αποτελέσματα εξάγεται το συμπέρασμα ότι το προτεινόμενο μοντέλο **DDMP** υπερτερεί όλων των υπολοίπων μοντέλων σε όλα τα έτη. Αναλυτικότερα, η τιμή AUC για τις χρονικές περιόδους “2 έτη πριν τη χρεοκοπία” και “ένα έτος πριν τη χρεοκοπία” είναι 0,627 και 0,664 αντίστοιχα, με το μοντέλο **LR** να έρχεται δεύτερο με τιμές 0,586 και 0,643 αντίστοιχα. Επίσης η τιμή AUC του μοντέλου **DDMP** για τη χρονική περίοδο “τελευταίο έτος” είναι 0,732 και εξακολουθεί να είναι το μεγαλύτερο σε σύγκριση με όλα τα υπόλοιπα ενώ το μοντέλο **Cart** έρχεται δεύτερο με 0,671.

Τέλος, δεδομένου ότι η παρούσα έρευνα έλαβε υπόψη της μόνο χρηματοοικονομικούς δείκτες, οι ερευνητές προτείνουν οι μελλοντικές έρευνες να χρησιμοποιήσουν περισσότερα ποσοτικά χαρακτηριστικά στο σύνολο δεδομένων, ώστε να βελτιωθεί ακόμη περισσότερο η ακρίβεια της πρόβλεψης του προτεινόμενου μοντέλου.

4.9 Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE

Στην παρούσα έρευνα, οι Smiti και Soui (2020), ασχολήθηκαν με την ανάπτυξη του μοντέλου deep learning BSM-SAES το οποίο συνδυάζει την τεχνική υπερδειγματοληψίας Borderline Synthetic Minority (BSM) με τους κατηγοριοποιητές Stacked AutoEncoder (SAE) και Softmax. Στόχος της έρευνας είναι η ανάπτυξη ενός μοντέλου DL για την πρόβλεψη χρεοκοπίας που να περιλαμβάνει τη διαδικασία εξαγωγής χαρακτηριστικών με σκοπό τη μείωση του χρόνου ταξινόμησης και την απλοποίηση των διαδικασιών λήψης αποφάσεων.

Τα δεδομένα συλλέχθηκαν από το University of California Irvine (UCI) Machine Learning Repository και συνολικά χρησιμοποιήθηκαν πέντε σύνολα δεδομένων που αφορούν πολωνικές εταιρείες (βλ. πίνακα 40):

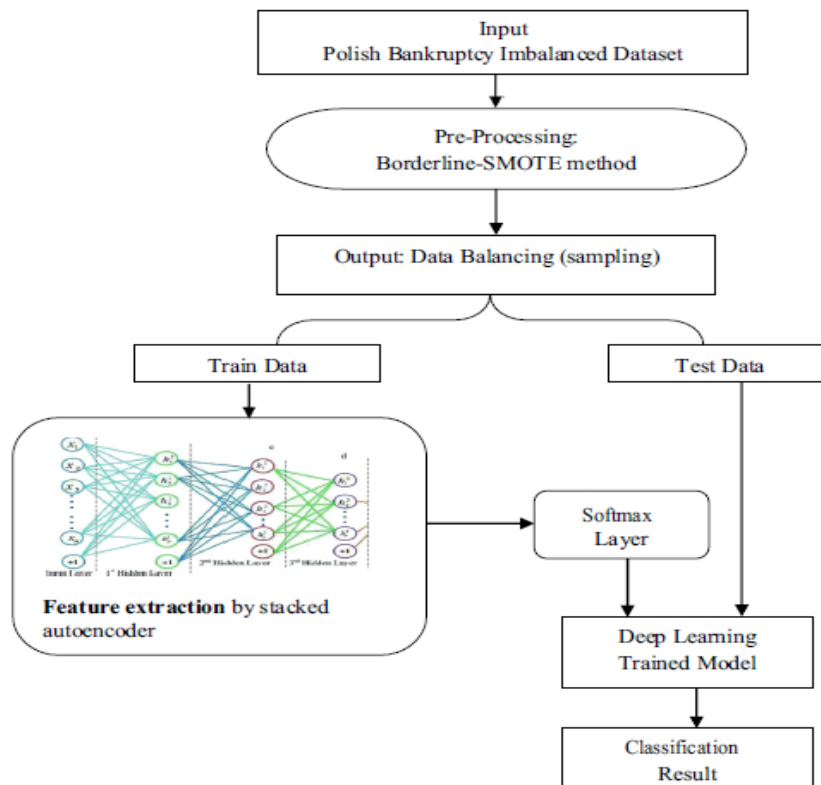
Πίνακας 40: Τα σύνολα δεδομένων

Data set	Bankruptcy after	Bankrupt		Not bankrupt	
		Number	Rate	Number	Rate
1st year	5 years	271	3,86	6756	96,14
2nd year	4 years	400	3,94	9773	96,06
3rd year	3 years	495	4,71	10,008	95,29
4th year	2 years	515	5,26	9277	94,74
5th year	1 years	410	6,94	5500	93,06

Πηγή: (Smiti & Soui, 2020)

Ως χαρακτηριστικά λήφθηκαν υπόψη 64 οικονομικοί δείκτες.

Η ερευνητική διαδικασία περιλαμβάνει τρία στάδια: (α) αρχικά αντιμετωπίζεται το πρόβλημα της ανισορροπίας του συνόλου δεδομένων με την τεχνική της **υπερδειγματοληψίας Borderline SMOTE**, (β) στη συνέχεια ακολουθεί η εξαγωγή των πιο σημαντικών χαρακτηριστικών με την εφαρμογή της **τεχνικής Stacked auto-encoder (SAE)** και (γ) τέλος γίνεται η ταξινόμηση των επιχειρήσεων σε χρεοκοπημένες και μη χρεοκοπημένες με τη χρήση της **συνάρτησης ενεργοποίησης softmax** (βλ. εικόνα 36). Δεδομένου ότι η υπερδειγματοληψία μπορεί να προκαλέσει την υπερπροσαρμογή του μοντέλου, εφαρμόστηκε η μέθοδος linear interpolation. Για την υλοποίηση των πειραμάτων χρησιμοποιήθηκε η Πλατφόρμα Matlab 2017b.

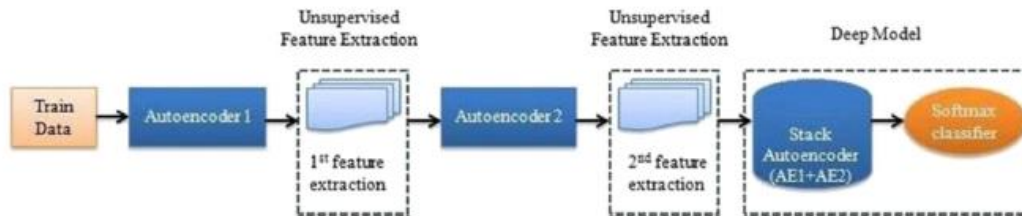


Εικόνα 36: Σχηματική απεικόνιση της ερευνητικής διαδικασίας

Πηγή: (Smiti & Soui, 2020)

Το σύνολο δεδομένων χωρίστηκε τυχαία σε training dataset (70%) και test dataset (30%).

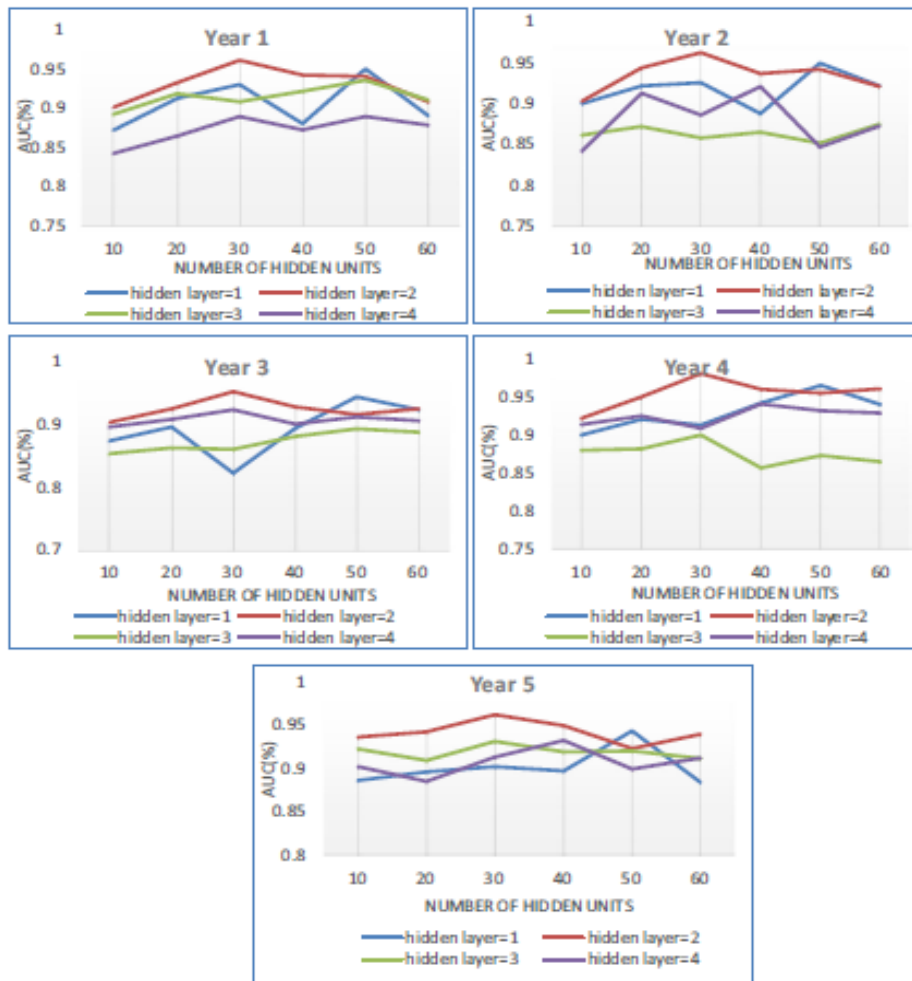
Έπειτα, κατασκευάστηκε το Stacked auto-encoder (SAE) που αποτελείται από 2 hidden layers και ένα softmax layer (βλ. εικόνα 37) με σκοπό τη μείωση της διάστασης των χαρακτηριστικών και τη βελτίωση της ακρίβειας του μοντέλου.



Εικόνα 37: Η δομή του Stacked auto-encoder

Πηγή: (Smiti & Soui, 2020)

Για τη βελτίωση του μοντέλου, πραγματοποιήθηκαν πειράματα σε κάθε dataset και διαπιστώθηκε ότι το SAE έχει υψηλό AUC με δύο hidden layers και με αριθμό νευρώνων 50 στο πρώτο layer και 30 στο δεύτερο. Επίσης η παράμετρος Sparse Regularization, που ελέγχει την τιμή κανονικοποίησης του objective function, ορίστηκε σε 4 ενώ η παράμετρος L2 Regularization, που χρησιμοποιείται για τη μείωση της υπερπροσαρμογής, σε 0,01. Η εικόνα 38 απεικονίζει την επίδραση του αριθμού των επιπέδων και των νευρώνων στην απόδοση του μέτρου AUC.



Εικόνα 38: Η επίδραση του αριθμού των hidden layers και των hidden units του SAE

Πηγή: (Smiti & Soui, 2020)

Το μοντέλο συγκρίθηκε με άλλα μοντέλα machine learning, όπως τα μοντέλα K-Nearest Neighbor, Decision Tree, Support Vector Machine, Artificial Neural Network, Random Forest και C5.0 τα οποία έχουν χρησιμοποιηθεί στο παρελθόν τόσο στην εξόρυξη δεδομένων όσο και στην πρόβλεψη χρεοκοπίας.

Για την αξιολόγηση της απόδοσης του μοντέλου, χρησιμοποιήθηκε το μέτρο Area under the ROC (AUC) καθώς επίσης και το training time.

Για τη μέτρηση της απόδοσης του μέτρου AUC διενεργήθηκαν δύο πειράματα. Στο πρώτο πείραμα δεν εφαρμόστηκε η τεχνική της υπερδευματοληψίας BSM σε καμία από τις μεθόδους ενώ στο δεύτερο εφαρμόστηκε σε όλες.

Σύμφωνα με τα αποτελέσματα του πίνακα 41 και όσον αφορά την πρώτη περίπτωση, παρατηρεί κανείς ότι το **προτεινόμενο μοντέλο** εμφανίζει μεγαλύτερη τιμή AUC σε όλα τα dataset σε σύγκριση με τα μοντέλα αναφοράς. Επίσης συγκρίνοντας μόνο τα μοντέλα ML μεταξύ τους, διαπιστώνεται ότι το μοντέλο **C5.0** υπερτερεί όλων των υπολοίπων.

Στη δεύτερη περίπτωση, όπου εφαρμόστηκε η μέθοδος υπερδευματοληψίας BSM, διαπιστώνεται ότι τα αποτελέσματα βελτιώνονται σε όλα τα μοντέλα σε σύγκριση με το προηγούμενο πείραμα με το προτεινόμενο μοντέλο να υπερτερεί.

Πίνακας 41: Τα αποτελέσματα απόδοσης του μέτρου AUC

Oversampling method	Classifier	1st year	2nd year	3rd year	4th year	5th year
None	KNN	0.480	0.485	0.479	0.476	0.637
	DT	0.674	0.577	0.593	0.631	0.736
	SVM-Linear	0.480	0.485	0.479	0.492	0.500
	SVM-Gaussian	0.647	0.486	0.499	0.620	0.604
	ANN	0.520	0.480	0.481	0.499	0.510
	RF	0.770	0.700	0.666	0.710	0.774
	C5.0	0.845	0.890	0.881	0.90	0.882
	SAE + Softmax	0.864	0.894	0.890	0.902	0.905
BSM	KNN	0.859	0.723	0.862	0.826	0.856
	DT	0.840	0.863	0.834	0.858	0.908
	SVM-Linear	0.824	0.798	0.813	0.836	0.845
	SVM-Gaussian	0.806	0.786	0.757	0.751	0.833
	ANN	0.685	0.735	0.677	0.744	0.734
	RF	0.910	0.921	0.894	0.914	0.912
	C5.0	0.965	0.950	0.968	0.969	0.950
	SAE + Softmax	0.961	0.962	0.950	0.980	0.962

Πηγή: (Smiti & Soui, 2020)

Επιπλέον προκειμένου να ελεγχθεί στατιστικά η υπεροχή του προτεινόμενου μοντέλου, εφαρμόστηκε η μέθοδος paired t-test. Σύμφωνα με τα αποτελέσματα του πίνακα 42, επιβεβαιώνεται ότι το προτεινόμενο μοντέλο BSM-SAES είναι σημαντικά καλύτερο από τις μεθόδους machine learning.

Πίνακας 42: Αποτελέσματα paired t-test ($\alpha=0.05$)

Method A	Method B	AUC	
		t	p value
BSM-SAES	BSM-KNN	4.986	0.0076
	BSM-DT	8.009	0.0013
	BSM-SVM-Linear	18.527	0.0001
	BSM-SVM-Gaussian	10.406	0.0005
	BSM-ANN	22.668	<0.0001
	BSM-RF	12.905	0.0002
	BSM-C5.0	11.302	0.0003

Πηγή: (Smiti & Soui, 2020)

Επιπλέον αξιολογήθηκε το training time των μοντέλων. Σύμφωνα με τα αποτελέσματα (βλ. πίνακα 43), το προτεινόμενο μοντέλο έχει τη χειρότερη απόδοση σε σύγκριση με τα μοντέλα ML.

Πίνακας 43: Τα αποτελέσματα του training time των μοντέλων (milliseconds)

Oversampling method	Classifier	1st year	2nd year	3rd year	4th year	5th year
None	KNN	0.098	0.101	0.108	0.107	0.089
	DT	0.238	0.392	0.412	0.386	0.219
	SVM-Linear	0.051	0.041	0.058	0.042	0.014
	SVM-Gaussian	0.065	0.117	0.172	0.161	0.039
	ANN	0.640	0.488	1.234	0.730	0.797
	RF	5.360	8.299	8.496	8.125	6.074
	C5.0	0.528	0.548	0.452	0.989	0.479
BSM	KNN	2.108	2.765	1.289	1.390	0.435
	DT	1.444	1.646	1.858	1.771	1.264
	SVM-Linear	45.514	48.426	37.586	35.291	30.750
	SVM-Gaussian	14.167	21.132	17.559	16.442	11.023
	ANN	1.023	0.776	1.261	2.554	1.856
	RF	14.465	14.095	16.750	14.888	20.189
	C5.0	1.102	1.723	2.864	1.339	0.905
BSM	SAE + Softmax	60.419	96.994	100.030	85.177	52.633

Πηγή: (Smiti & Soui, 2020)

Τέλος, έγινε σύγκριση του προτεινόμενου μοντέλου με μοντέλα άλλων παρόμοιων προηγούμενων ερευνών που χρησιμοποίησαν τα μοντέλα EXGB, XGB και Isolation Forest για την πρόβλεψη χρεοκοπίας των πολωνικών εταιρειών. Σύμφωνα με τα αποτελέσματα (βλ. πίνακα 44), το προτεινόμενο μοντέλο υπερτερεί των υπολοίπων σημειώνοντας την υψηλότερη τιμή AUC.

Πίνακας 44: Σύγκριση με προηγούμενες μελέτες πρόβλεψης χρεοκοπίας για τα Πολωνικά σύνολα δεδομένων αναφοράς

Authors	Method	AUC				
		1st year	2nd year	3rd year	4th year	5th year
(Zieba et al. 2016)	EXGB	0.959	0.944	0.940	0.941	0.955
(Broelemann and Kasneci 2018)	XGB	0.944	0.903	0.902	0.925	0.944
(Fan et al. 2018)	Isolation Forest	0.930	0.950	0.940	0.950	0.960
Proposed method	BSM-SAES	0.961	0.962	0.950	0.980	0.962

Πηγή: (Smiti & Soui, 2020)

Από τα παραπάνω συμπεραίνει κανείς ότι το μοντέλο BSM-SAES είναι ένα αποτελεσματικό μοντέλο για την πρόβλεψη της χρεοκοπίας με υψηλή ικανότητα ορθής ταξινόμησης των χρεοκοπημένων και μη χρεοκοπημένων επιχειρήσεων. Έχει τη δυνατότητα να εξαγάγει αυτόματα τα χρήσιμα χαρακτηριστικά κατά τη διάρκεια της εκπαίδευσης γεγονός που το καθιστά ως ένα υψηλής ακρίβειας μοντέλο σε σύγκριση με τις υπόλοιπες μεθόδους κατά τις οποίες η εξαγωγή των χαρακτηριστικών γίνεται

ανεξάρτητα από τη φάση εκπαίδευσης. Ωστόσο το προτεινόμενο μοντέλο έχει τη χειρότερη απόδοση στο training time, λόγω του χρόνου που αφιερώνει για την εξαγωγή των σημαντικών χαρακτηριστικών.

Στο μέλλον οι ερευνητές σχεδιάζουν την ανάπτυξη ενός μοντέλου πρόβλεψης χρεοκοπίας που θα βασίζεται σε IF-THEN rules.

4.10 Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors

Στην παρούσα έρευνα, οι Jang, Jeong και Cho (2021) χρησιμοποίησαν το μοντέλο **Long short Term memory - Recurrent Neural Network (LSTM-RNN)** το οποίο αναπτύχθηκε σε προηγούμενη μελέτη (Jang κ.ά., 2019b, Jang κ.ά., 2020 όπως αναφέρεται στους Jang κ.ά., 2021). Όμως παρόλο που το μοντέλο είχε επιτύχει την ανώτερη απόδοση σε σύγκριση με άλλα μοντέλα πρόβλεψης, οι κρυφές διαδικασίες δεν ήταν εύκολα ερμηνεύσιμες δεδομένου ότι πρόκειται για ένα μοντέλο μαύρου κουτιού (black box).

Έτσι η παρούσα έρευνα έχει ως **στόχο** να προσδιορίσει την επίδραση που έχουν οι μεταβλητές εισόδου στο μοντέλο πρόβλεψης χρεοκοπίας **LSTM-RNN** με τη χρήση της τιμής Sharpley προκειμένου να μετρηθεί η επίδραση των μεταβλητών εισόδου και να εντοπιστεί η μεταβλητή εκείνη που επηρεάζει περισσότερο την ακρίβεια της πρόβλεψης. Συγκεκριμένα, αναπτύχθηκαν τρία μοντέλα LSTM-RNN για την πιθανότητα πρόβλεψης χρεοκοπίας πριν από 1, 2 και 3 χρόνια.

Τα **δεδομένα** που συλλέχθηκαν αφορούν εργολάβους κατασκευών των ΗΠΑ εισηγμένους στα Χρηματιστήρια New York Stock Exchange, NASDAQ και America Exchange για το χρονικό διάστημα 1980-2016. Το σύνολο δεδομένων αποτελείται από 1378 συνολικές παρατηρήσεις εκ των οποίων οι 1336 αφορούν υγιή εταιρικά έτη και οι 42 χρεοκοπημένα εταιρικά έτη.

Οι **μεταβλητές** εισόδου που χρησιμοποιήθηκαν είναι συνολικά 18 από τις οποίες οι 12 είναι λογιστικές, οι 2 μακροοικονομικές και οι υπόλοιπες 3 της κατασκευαστικής αγοράς (βλ. πίνακα 45).

Πίνακας 45: Οι μεταβλητές της έρευνας

Input variable	Abbr	Brief description	Sources
<i>Accounting variables</i>			
(1) Return on asset	ROA	Net income/total assets	1,2,3,4,5,6,7,8,9,10,11
(2) Return on equity	ROE	Net income/shareholder equity	1,2,3,4,5,7,9,10,11
(3) Return on sales	ROS	Net income/net sales	1,3,4,5,6,7,10,11
(4) Current ratio	CR	Current assets/current liabilities	1,3,4,5,7,9,10,11
(5) Current assets to net assets	CANA	Current assets/(total assets-current liabilities)	1,3,4,5,10,11
(6) Working capital to total assets	WCTA	(current asset-current liabilities)/total assets	1,2,3,4,5,7,8,10,11
(7) Total liabilities to net worth	TLNW	Total liabilities/shareholder equity	1,2,4,5,10,11
(8) Retained earnings to sales	RES	Retained earnings/net sales	1,2,4,5,10,11
(9) Debt ratio	DR	Total liabilities/total assets	1,2,3,4,5,9,10,11
(10) Working capital turnover	WCT	Net sales/(current assets-current liabilities)	1,3,4,5,10,11
(11) Equity turnover	ET	Net sales/shareholder equity	1,2,3,4,5,10,11
(12) Total asset turnover	TAT	Net sales/total assets	1,2,3,4,5,8,9,10,11
<i>Construction market variables</i>			
(13) Construction spending	CS	A measure of the value of new construction activities, including residential, nonresidential, and public projects	12,13
(14) House starts	HS	The number of new privately-owned housing units that have begun construction in a given period	12,13
(15) Employment in construction	EC	The number of employees on payrolls in construction, which is a useful measure for representing the labor force in the construction sector of the economy	12,13
<i>Macroeconomic variables</i>			
(16) Consumer per index	CPI	A measure of the price level of a representative basket of goods and services purchased by urban consumers	12,13,14
(17) Gross domestic product	GDP	A measure of the total value of goods and services that are produced in a country in a given period	12,13,14
(18) Federal funds rate	FFR	The interest rate at which banks and other depository institutions charge each other for loans	12,13,14
<p>Note(s): 1 = Cheng and Hoang (2015); 2 = Tserng <i>et al.</i> (2011a); 3 = Tserng <i>et al.</i> (2012); 4 = Tserng <i>et al.</i> (2014); 5 = Tserng <i>et al.</i> (2015); 6 = Chen (2012); 7 = Horta and Camanho (2013); 8 = Heo and Yang (2014); 9 = Bal <i>et al.</i> (2013); 10 = Cheng <i>et al.</i> (2014); 11 = Tsai <i>et al.</i> (2012); 12 = Ashuri <i>et al.</i> (2012); 13 = Shahandashti and Ashuri (2013); 14 = Nouri and Soltani (2016)</p>			

Πηγή: (Jang κ.ά., 2021)

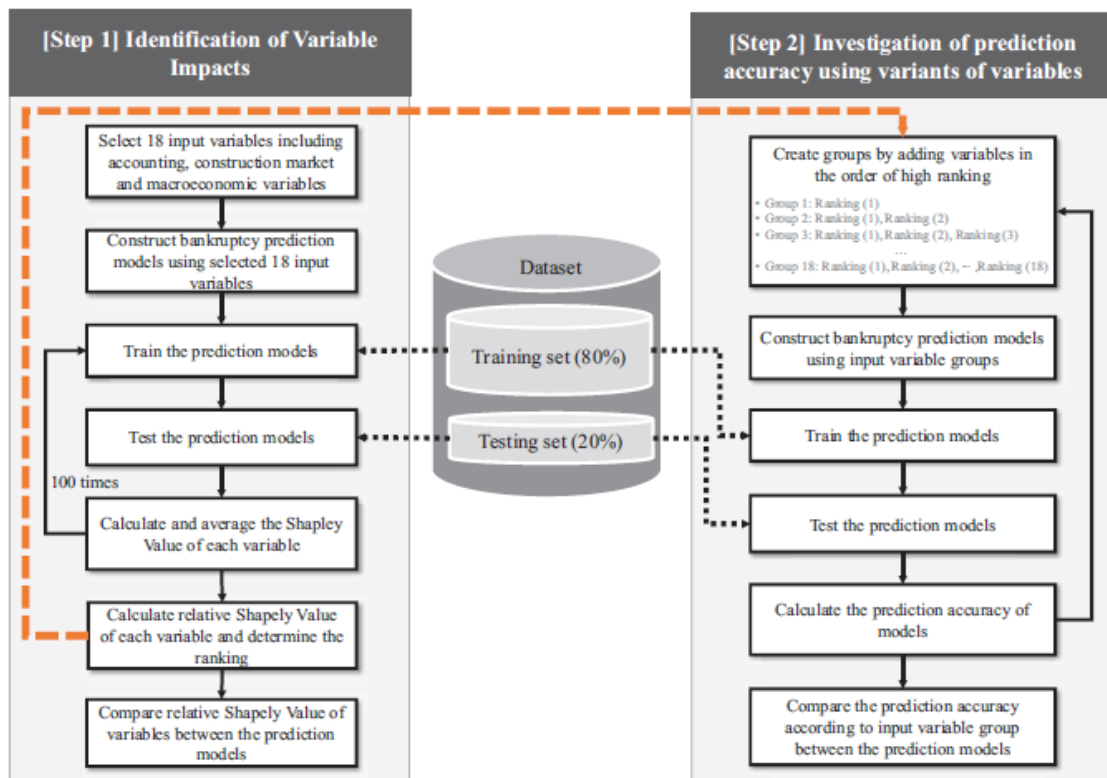
Τα **λογιστικά δεδομένα** συλλέχθηκαν από τη βάση δεδομένων Standard & Poor's COMPUSTAT ενώ τα **δεδομένα της κατασκευαστικής αγοράς και της μακροοικονομίας** από τους οργανισμούς U.S. Bureau of Census, U.S. Bureau of Labor Statistics, U.S. Bureau of Economic Analysis and Board of Governors of the Federal Reserve Systems.

Για την αντιμετώπιση της ανισορροπίας των δεδομένων εφαρμόστηκε η τεχνική SMOTE + Tomek link με σκοπό την εκ νέου δειγματοληψία και την εξισορρόπηση του dataset.

Αναλυτικότερα, αρχικά χρησιμοποιήθηκε η τεχνική SMOTE για τη δημιουργία των συνθετικών δεδομένων της μειοψηφικής τάξης και στη συνέχεια εφαρμόστηκε η τεχνική Tomek link για την εξισορρόπηση του dataset.

Για τη **μέτρηση της επίδρασης** των μεταβλητών χρησιμοποιήθηκε η **τιμή Shapley** η οποία ανήκει στη θεωρία παιγνίων και αφορά τη δίκαιη κατανομή της απόδοσης της συνεισφοράς των παικτών σε ένα παιχνίδι συνεργασίας. Στην παρούσα έρευνα μετρά την επίδραση κάθε μεταβλητής στην απόδοση των μοντέλων όταν χρησιμοποιούνται όλες οι μεταβλητές εισόδου.

Η διαδικασία των πειραμάτων αποτελείται από δύο μέρη (βλ. εικόνα 39) και σκοπός είναι (1) η σύγκριση των επιπτώσεων των μεταβλητών εισόδου στα τρία μοντέλα LSTM-RNN και (2) η διερεύνηση της ακρίβειας της πρόβλεψης εισάγοντας κάθε φορά έναν διαφορετικό συνδυασμό μεταβλητών.



Εικόνα 39: Η πειραματική διαδικασία

Πηγή: (Jang κ.ά., 2021)

Πρώτο μέρος πειραμάτων

Για την κατασκευή του καλύτερου μοντέλου πρόβλεψης LSTM-RNN, χρησιμοποιήθηκε η τεχνική grid search για τον υπολογισμό των βέλτιστων τιμών των υπερπαραμέτρων και συγκεκριμένα τον αριθμό των κρυφών επιπέδων και τον αριθμό των νευρώνων στο κρυφό επίπεδο. Με την εφαρμογή των 18 μεταβλητών διαπιστώθηκε ότι ο βέλτιστος αριθμός των υπερπαραμέτρων είναι δύο κρυφά επίπεδα και 250 νευρώνες.

Το σύνολο δεδομένων χωρίστηκε τυχαία και για την **εκπαίδευση** των μοντέλων χρησιμοποιήθηκε το 80% του dataset ως training dataset. Τα μοντέλα εκπαιδεύτηκαν για 50.000 εποχές και dropout 0,5 ενώ το learning rate μειώθηκε εκθετικά από 0,001 σε 0,00001. Στη συνέχεια, για τον **έλεγχο** των εκπαιδευμένων μοντέλων χρησιμοποιήθηκε

το υπόλοιπο 20% του dataset ως testing dataset και υπολογίστηκαν οι τιμές Shapley κάθε μεταβλητής.

Οι διαδικασίες εκπαίδευσης και ελέγχου επαναλήφθηκαν 100 φορές και υπολογίστηκε ο μέσος όρος των 100 τιμών Shapley κάθε μεταβλητής, ώστε να ληφθεί η σχετική τιμή Shapley. Τέλος, οι σχετικές τιμές Shapley των μεταβλητών συγκρίθηκαν μεταξύ των τριών μοντέλων.

Αφού υπολογίστηκαν οι τιμές Shapley και οι σχετικές τιμές Shapley κάθε μοντέλου, καθορίστηκε η σειρά κατάταξης όλων των μεταβλητών ως ακολούθως (βλ. πίνακα 46):

Πίνακας 46: Σειρά κατάταξης των τιμών Shapley και των σχετικών τιμών Shapley όλων των μεταβλητών των 3 μοντέλων

Input variables	1-year prediction model			2-year prediction model			3-year prediction model		
	Ranking	Shapley value	Relative Shapley value	Ranking	Shapley value	Relative Shapley value	Ranking	Shapley value	Relative Shapley value
ROA	(1)	0.072	0.074	(3)	0.064	0.067	(5)	0.060	0.064
HS	(2)	0.067	0.068	(1)	0.071	0.074	(1)	0.067	0.072
FFR	(3)	0.063	0.064	(2)	0.065	0.068	(3)	0.062	0.066
DR	(4)	0.060	0.061	(7)	0.057	0.060	(7)	0.056	0.060
CPI	(5)	0.059	0.060	(4)	0.061	0.064	(2)	0.062	0.067
CR	(6)	0.058	0.059	(6)	0.058	0.061	(6)	0.056	0.061
GDP	(7)	0.057	0.058	(5)	0.058	0.061	(4)	0.060	0.065
TAT	(8)	0.056	0.057	(8)	0.054	0.057	(8)	0.053	0.057
ROE	(9)	0.053	0.055	(9)	0.050	0.053	(10)	0.048	0.052
ET	(10)	0.053	0.054	(10)	0.050	0.053	(9)	0.049	0.053
TLNW	(11)	0.051	0.052	(11)	0.048	0.051	(11)	0.047	0.051
CP	(12)	0.050	0.051	(12)	0.047	0.050	(13)	0.046	0.049
EC	(13)	0.048	0.049	(15)	0.045	0.047	(18)	0.043	0.046
WCT	(14)	0.048	0.049	(16)	0.045	0.047	(16)	0.044	0.047
WCTA	(15)	0.047	0.048	(13)	0.046	0.049	(12)	0.046	0.049
RES	(16)	0.047	0.048	(14)	0.046	0.048	(14)	0.045	0.049
CANA	(17)	0.045	0.046	(17)	0.044	0.046	(17)	0.043	0.046
ROS	(18)	0.045	0.046	(18)	0.043	0.046	(15)	0.044	0.047
Total	-	0.980	1.000	-	0.953	1.000	-	0.930	1.000

Πηγή: (Jang κ.ά., 2021)

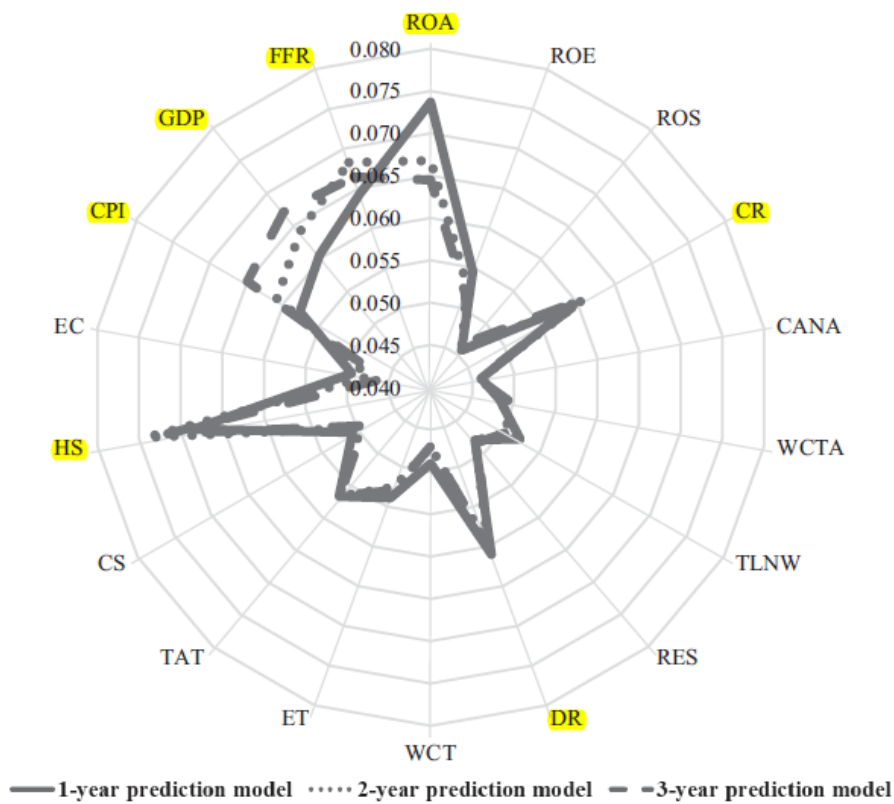
Σύμφωνα με τον παραπάνω πίνακα, όσον αφορά το **πρώτο μοντέλο**, η λογιστική μεταβλητή **Return on asset (ROA)** έχει τη μεγαλύτερη τιμή Shapley (0,072) και ακολουθούν οι μεταβλητές της κατασκευαστικής αγοράς **House starts (HS)** με 0,067 και της μακροοικονομικής μεταβλητής **Federal funds rate (FFR)** με 0,063. Επίσης η τιμή ROA είναι κατά 1,6 φορές υψηλότερη σε σύγκριση με τη μεταβλητή **Return on sales (ROS)** που είναι τελευταία στη σειρά κατάταξης.

Ενώ στο **δεύτερο μοντέλο**, η μεταβλητή της κατασκευαστικής αγοράς **HS** έχει τη μεγαλύτερη τιμή Shapley (0,071) και ακολουθούν οι μεταβλητές της μακροοικονομικής **FFR** (0,065) και της λογιστικής **ROA** (0,064). Επίσης παρατηρείται ότι οι μεταβλητές **HS** (κατασκευαστικής αγοράς) και **FFR**, **CPI** και **GDP** (μακροοικονομικές) ανέβηκαν σειρά κατάταξης σε σύγκριση με το πρώτο μοντέλο. Επίσης η τιμή HS είναι κατά 1,65 φορές υψηλότερη σε σύγκριση με τη μεταβλητή **ROS** που είναι τελευταία στη σειρά κατάταξης.

Στο **τρίτο μοντέλο**, η μεταβλητή **HS** έχει τιμή Shapley (0,067) που είναι και η μεγαλύτερη ενώ ακολουθούν οι μακροοικονομικές μεταβλητές **Consumer per index (CPI)** με τιμή 0,062 και **FFR** με 0,062. Επίσης η τιμή HS είναι κατά 1,55 φορές υψηλότερη σε σύγκριση με τη μεταβλητή **Employment in construction (EC)** που είναι τελευταία στη σειρά κατάταξης.

Αξίζει να σημειωθεί ότι στις 10 πρώτες θέσεις κατάταξης όλων των μοντέλων περιλαμβάνονται 6 λογιστικές μεταβλητές (ROA, DR, CR, TAT, ROE, ET), τρεις μακροοικονομικές (FFR, CPI, GDP) και μία της κατασκευαστικής αγοράς (HS).

Επίσης, συγκρίνοντας τις **σχετικές τιμές Sharpley** (βλ. πίνακα 46), διαπιστώνεται ότι η μεταβλητή **HS** επηρεάζει σημαντικά την ακρίβεια πρόβλεψης. Επίσης οι σχετικές τιμές Sharpley των μεταβλητών **CPI, GDP** και **FFR** αυξάνονται κατά τη διάρκεια των ετών ενώ αντίθετα της μεταβλητής **ROA** μειώνεται, δηλαδή η επίδραση της μεταβλητής ROA στην ακρίβεια πρόβλεψης μειώνεται όσο η περίοδος πρόβλεψης αυξάνεται. Τέλος, οι σχετικές τιμές Sharpley των λογιστικών μεταβλητών **CR** και **DR** ήταν παρόμοιες κατά τη διάρκεια των ετών, επομένως θεωρείται ότι η επίπτωσή τους στην ακρίβεια πρόβλεψης είναι σημαντική σε σύγκριση με τις υπόλοιπες λογιστικές μεταβλητές. Η σύγκριση των σχετικών τιμών Sharpley των μεταβλητών απεικονίζεται παρακάτω (βλ. εικόνα 40):



Εικόνα 40: Σύγκριση των επιπτώσεων των μεταβλητών εισόδου στα μοντέλα LSTM-RNN

Πηγή: (Jang κ.ά., 2021)

Δεύτερο μέρος πειραμάτων

Αρχικά, έγινε κατάταξη των μεταβλητών κατά φθίνουσα σειρά σύμφωνα με τις τιμές Sharpley που υπολογίστηκαν προηγουμένως. Έπειτα δημιουργήθηκαν 18 ομάδες μεταβλητών εισόδου ως εξής: η **πρώτη ομάδα** περιείχε μόνο μία μεταβλητή, τη μεταβλητή με τη μεγαλύτερη τιμή Sharpley. Η **δεύτερη ομάδα** περιείχε δύο μεταβλητές, τη μεταβλητή της πρώτης ομάδας και τη μεταβλητή με τη δεύτερη μεγαλύτερη τιμή Sharpley κ.ο.κ.

Στη συνέχεια, τα τρία μοντέλα εκπαιδεύτηκαν και ελέγχθηκαν χρησιμοποιώντας ως είσοδο όλες τις μεταβλητές.

Για την απόδοση των τριών μοντέλων, υπολογίστηκε το μέτρο απόδοσης **accuracy** λαμβάνοντας υπόψη και τις 18 μεταβλητές. Διαπιστώθηκε ότι η ακρίβεια των μοντέλων είναι 0,980, 0,953 και 0,930 αντίστοιχα (βλ. πίνακα 47).

Πίνακας 47: Η ακρίβεια πρόβλεψης των 3 μοντέλων

Model	Accuracy
1-year prediction model	0.980
2-year prediction model	0.953
3-year prediction model	0.930

Πηγή: (Jang κ.ά., 2021)

Σύμφωνα με τον παραπάνω πίνακα, παρατηρεί κανείς ότι η ακρίβεια της πρόβλεψης μειώνεται καθώς αυξάνεται η χρονική περίοδος.

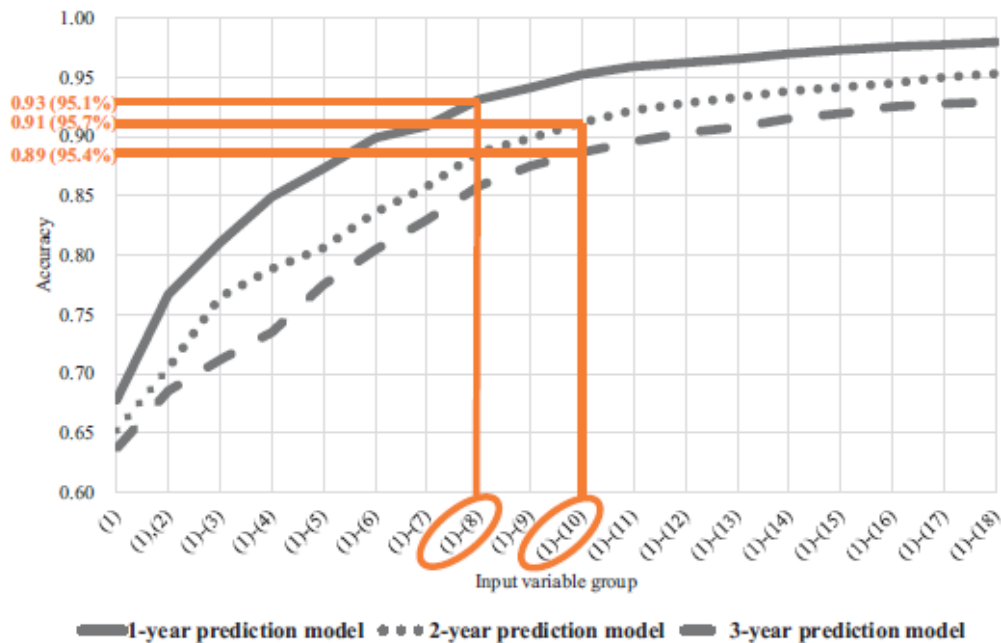
Τέλος, έγινε σύγκριση της ακρίβειας πρόβλεψης λαμβάνοντας υπόψη τις 18 διαφορετικές ομάδες μεταβλητών εισόδου που δημιουργήθηκαν (βλ. πίνακα 48).

Πίνακας 48: Σύγκριση μοντέλων με τη χρήση των 18 διαφορετικών ομάδων μεταβλητών εισόδου

Input variable group	1-year prediction model		2-year prediction model		3-year prediction model	
	Accuracy	Percentage	Accuracy	Percentage	Accuracy	Percentage
Ranking (1)	0.678	69.2	0.652	68.4	0.637	68.6
Ranking (1), (2)	0.767	78.3	0.705	74.0	0.686	73.8
Ranking (1)–(3)	0.811	82.8	0.765	80.2	0.712	76.6
Ranking (1)–(4)	0.849	86.7	0.789	82.7	0.735	79.1
Ranking (1)–(5)	0.873	89.1	0.806	84.6	0.775	83.4
Ranking (1)–(6)	0.899	91.8	0.836	87.7	0.805	86.6
Ranking (1)–(7)	0.909	92.8	0.858	90.0	0.830	89.3
Ranking (1)–(8)	0.931	95.1	0.887	93.0	0.858	92.3
Ranking (1)–(9)	0.941	96.1	0.899	94.3	0.875	94.2
Ranking (1)–(10)	0.952	97.2	0.912	95.7	0.887	95.4
Ranking (1)–(11)	0.959	97.9	0.922	96.8	0.896	96.4
Ranking (1)–(12)	0.962	98.2	0.928	97.4	0.904	97.2
Ranking (1)–(13)	0.966	98.6	0.933	97.9	0.908	97.7
Ranking (1)–(14)	0.970	99.0	0.939	98.5	0.915	98.5
Ranking (1)–(15)	0.973	99.3	0.942	98.8	0.920	98.9
Ranking (1)–(16)	0.976	99.6	0.945	99.1	0.925	99.5
Ranking (1)–(17)	0.978	99.8	0.950	99.7	0.928	99.8
Ranking (1)–(18)	0.980	100.0	0.953	100.0	0.930	100.0

Πηγή: (Jang κ.ά., 2021)

Σύμφωνα με τον παραπάνω πίνακα διαπιστώνεται ότι σε κάθε μοντέλο ξεχωριστά η ακρίβεια της πρόβλεψης αυξάνεται όσο αυξάνεται και ο αριθμός των μεταβλητών. Επίσης όσο αυξάνεται η χρονική περίοδος πρόβλεψης τόσο περισσότερες μεταβλητές απαιτούνται για την αύξηση της ακρίβειας πρόβλεψης.



Εικόνα 41: Σύγκριση των αποτελεσμάτων ακρίβειας πρόβλεψης

Πηγή: (Jang κ.ά., 2021)

Επίσης σύμφωνα με την εικόνα 41, το **πρώτο μοντέλο πρόβλεψης με τη χρήση 8** μόνο μεταβλητών (ROA, HS, FFR, DR, CPI, CR, ΑΕΠ και TAT) έχει ακρίβεια 0,931, το **δεύτερο μοντέλο** με τη χρήση 10 μεταβλητών (HS, FFR, ROA, CPI, GDP, CR, DR, TAT, ROE και ET) έχει 0,912 και το **τρίτο μοντέλο** με τη χρήση επίσης μόνο 10 μεταβλητών (HS, CPI, FFR, GDP, ROA, CR, DR, TAT, ET και ROE) 0,887. Τα παραπάνω ποσοστά αντιπροσωπεύουν το 95,1%, το 95,7% και το 95,4% της ακρίβειας του μοντέλου με τη χρήση και των 18 μεταβλητών, δηλαδή του συνόλου των μεταβλητών. Από τα παραπάνω συμπεραίνει κανείς ότι παρόμοια ακρίβεια πρόβλεψης μπορεί να επιτευχθεί με τη χρήση λιγότερων από 10 μεταβλητών.

Για τη βελτίωση της ακρίβειας πρόβλεψης, οι ερευνητές προτείνουν οι **μελλοντικές έρευνες** να μελετήσουν περισσότερες κατηγορίες μεταβλητών και να εξετάσουν (α) τον αντίκτυπο κάθε μεταβλητής ανάλογα με το μέγεθος, τον τύπο της επιχείρησης ή ακόμα και τον τόπο εγκατάστασής της, καθώς επίσης και (β) τη αλληλεπίδραση των τιμών Sharpley μεταξύ των μεταβλητών, ώστε να βελτιστοποιηθεί η ομάδα μεταβλητών εισόδου, δεδομένου ότι στην παρούσα μελέτη οι ομάδες μεταβλητών δημιουργήθηκαν μόνο κατά τη σειρά κατάταξης που πιθανόν να μην είναι η βέλτιστη επιλογή.

4.11 Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM

Οι Vochozka, Vrbka και Suler (2020) ασχολήθηκαν με την ανάπτυξη ενός μοντέλου τεχνητού νευρωνικού δικτύου (NN) με ένα τουλάχιστον επίπεδο LSTM με

στόχο να προσδιοριστεί η καταλληλότητα του μοντέλου για την πρόβλεψη της πιθανότητας χρεοκοπίας των επιχειρήσεων.

Αρχικά, έγινε επιλογή και προεπεξεργασία των δεδομένων που χρησιμοποιήθηκαν. Τα δεδομένα συλλέχθηκαν από τη βάση δεδομένων Albertina και αφορά συνολικά 5500 βιομηχανικές εταιρείες που δραστηριοποιούνται στην Τσέχικη Δημοκρατία. Το σύνολο δεδομένων καλύπτει το χρονικό διάστημα 2014-2018 και περιλαμβάνει **ενεργές εταιρείες**, δηλαδή εταιρείες οι οποίες δεν αντιμετωπίζουν οικονομική δυσπραγία και μπορούν να επιβιώσουν, και **εταιρείες υπό εκκαθάριση**. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από τις οικονομικές καταστάσεις των εταιρειών. Το σύνολο δεδομένων χωρίστηκε σε δύο μέρη, το training dataset για την εκπαίδευση του μοντέλου και το testing dataset για την επικύρωση.

Τα στατιστικά χαρακτηριστικά των datasets απεικονίζονται στο κάτωθι πίνακα (βλ. πίνακα 49):

Πίνακας 49: Στατιστικά χαρακτηριστικά των datasets

Item	Training				Testing			
	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation
Total Assets	0	6,963,568	101,794.3	421,183.6	0	62,924,684	193,629.3	2,402,017
Fixed Assets	-19,065	4,558,816	49,269.92	251,519.4	0	30,832,576	90,484.69	1,204,674
Current Assets	0	2,391,855	51,801.29	182,982.9	0	32,066,562	102,624.7	1,222,082
Short Term Receivables	-127	1,947,964	24,162.22	106,237.4	0	27,683,668	65,334.36	1,039,648
Equity	-9662	1,923,390	50,993.7	186,972.2	-206,208	53,318,744	131,475.4	2,004,015
Borrowed Capital	0	5,024,089	50,269.76	256,551.2	-21	9,605,513	61,351.15	469,600.1
Short Term Liabilities	0	4,906,382	31,905.76	204,347.7	-96	7,970,386	41,665.4	366,137.2
Performance	-40	4,180,449	97,214.31	335,148.6	-246	33,887,311	142,956.2	1,342,696
Services	0	569,454	14,468.57	48,754.07	0	3,038,338	19,348.75	133,230.2
Added Value	-7788	824,345	26,734.39	76,957.12	-20,284	3,457,583	30,516.86	167,441.7
Personnel Costs	0	463,685	17,940.92	46,808.62	-60	2,665,233	19,125.81	111,284.5
Operating Result	-150,310	852,991	7316.556	44,552.34	-847,530	914,063	4213.27	63,193.46
Interest Payable	0	187,218	744.4554	7525.506	0	52,798	494.3265	2912.81
Economic Result for Accounting Period (+/-)	-245,135	700,092	6354.041	43,196.58	-645,493	673,179	5049.786	53,302.32

Πηγή: (Vochozka κ.ά., 2020)

Για τη δημιουργία του μοντέλου χρησιμοποιήθηκε το λογισμικό Wolfram's Mathematica (έκδοση 13) όπου πραγματοποιήθηκε πείραμα για τον προσδιορισμό της δομής του δικτύου NN. Κάθε επίπεδο του νευρωνικού δικτύου αποτελείται από τα ακόλουθα στοιχεία:

Το επίπεδο εισόδου: είναι ένας πίνακας μίας στήλης που περιλαμβάνει 14 συνεχείς μεταβλητές εισόδου (πίνακας 1x14): total assets, fixed assets, current assets, short term, equity, borrowed capital, short term liabilities, performance, services, added value, personnel costs, operating results, interest payable, economic result.

Το 1ο κρυφό στρώμα: είναι ένα LSTM layer, η έξοδος του οποίου είναι ένας πίνακας 1×n, όπου n είναι ο αριθμός των στοιχείων μήτρας από 5 ως 2000 στοιχεία.

Το 2ο και 3ο κρυφό στρώμα: είναι elementwise layers με στόχο να προστεθεί ένας ορισμένος βαθμός μη γραμμικότητας στο δίκτυο NN. Έγιναν πειράματα κατά τα οποία δοκιμάστηκε η καταλληλότητα των συναρτήσεων Hyperbolic tangent (Tanh), Sinus (Sin), Ramp (ReLU), και Logistic function (logistic sigmoid).

Το 4ο κρυφό στρώμα: είναι ένα LSTM layer, η έξοδος του οποίου είναι ένας πίνακας 1×2, δηλαδή το μέγεθος του προσδιορίστηκε από τον αριθμό των πιθανών αποτελεσμάτων πρόβλεψης ('active' ή 'in liquidation').

Το επίπεδο εξόδου: είναι δύο νευρώνες που αντιπροσωπεύουν ένα διάνυσμα με δύο στοιχεία το οποίο αποκωδικοποιείται σε «active company» ή «company in liquidation».

Κατά το πείραμα δημιουργήθηκαν 1000 neural networks που διέφεραν ως προς το μέγεθος του vector στο πρώτο επίπεδο LSTM (1^ο κρυφό επίπεδο) και τη συνάρτηση ενεργοποίησης του 2^{ου} και 3^{ου} κρυφού επιπέδου του NN.

Για παράδειγμα, η δομή του πρώτου NN είναι 14-940-Tanh-Tanh-2-1 και απεικονίζεται ως εξής (βλ. εικόνα 42):



Εικόνα 42: Δομή NN (14-940-Tanh-Tanh-2-1)

Πηγή: (Vochozka κ.ά., 2020)

Η αξιολόγηση της απόδοσης των μοντέλων NN έγινε με τη χρήση του confusion matrix και στα δύο σύνολα δεδομένων (training dataset και testing dataset). Ο πίνακας 50 δείχνει την απόδοση των πέντε καλύτερων μοντέλων NN και στα δύο σύνολα δεδομένων.

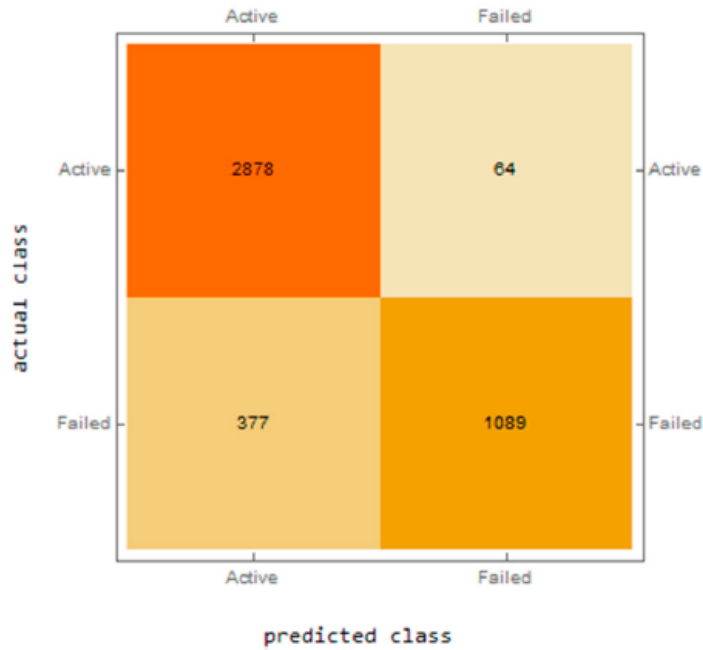
Πίνακας 50: Τα αποτελέσματα απόδοσης των πέντε καλύτερων NN

ID NN	Neural Network	Training Performance			Test Performance		
		Active	Failed	Total	Active	Failed	Total
1.	14-940-Tanh-Tanh-2-1	0.978246	0.742838	0.899955	0.971253	0.752066	0.898491
2.	14-1970-Ramp-Tanh-2-1	0.978586	0.722374	0.893376	0.967199	0.750331	0.899863
3.	14-1980-Ramp-Sin-2-1	0.976547	0.718963	0.89088	0.973306	0.747934	0.898491
4.	14-1990-Sin-Sin-2-1	0.973148	0.736016	0.894283	0.975359	0.747934	0.899863
5.	14-1010-Tanh-Sin-2-1	0.980625	0.734652	0.89882	0.967146	0.772727	0.902606

Πηγή: (Vochozka κ.ά., 2020)

Σύμφωνα με τον παραπάνω πίνακα, το πρώτο μοντέλο NN είναι το καλύτερο στην πρόβλεψη των ενεργών εταιρειών επιτυγχάνοντας απόδοση 0,978 στο training dataset και 0,971 στο testing dataset. Όμως η απόδοση στην πρόβλεψη των χρεοκοπημένων επιχειρήσεων είναι χαμηλότερη και συγκεκριμένα 0,743 για το training dataset και 0,752 για το testing dataset. Αυτό οφείλεται στον θόρυβο των δεδομένων, διότι η απόφαση για την έναρξη της εκκαθάρισης μιας επιχείρησης δεν είναι πάντοτε ορθολογική. Μπορεί να συμβεί οποιαδήποτε στιγμή αποφασίσει η διοίκηση ακόμα και όταν δεν συντρέχουν οικονομικοί λόγοι. Παρόλο αυτά το μοντέλο δείχνει να είναι επιτυχημένο δίνοντας ένα εξαιρετικό αποτέλεσμα πρόβλεψης.

Επίσης για το training dataset, το confusion matrix (βλ. εικόνα 43) δείχνει ότι από τις 2942 ενεργές εταιρείες η πρόβλεψη ταξινόμησης ήταν σωστή σε 2878 περιπτώσεις και εσφαλμένη σε 64, ενώ για την πρόβλεψη χρεοκοπίας από τις 1466 περιπτώσεις, ήταν σωστή σε 1089 περιπτώσεις και εσφαλμένη σε 377.



Εικόνα 43: Confusion matrix του training dataset

Πηγή: (Vochozka κ.ά., 2020)

Τέλος, για το testing dataset, το confusion matrix (βλ. εικόνα 44) δείχνει ότι από τις 487 ενεργές εταιρείες, το μοντέλο εντόπισε 473 να είναι ικανές να επιβιώσουν από πιθανές οικονομικές δυσκολίες, ενώ για την πρόβλεψη χρεοκοπίας από τις 262 περιπτώσεις το μοντέλο εντόπισε τις 182.



Εικόνα 44: Confusion matrix του testing dataset

Πηγή: (Vochozka κ.ά., 2020)

Από τα παραπάνω γίνεται κατανοητό ότι το μοντέλο που αναπτύχθηκε εκπληρώνει τους στόχους του και μπορεί να εφαρμοστεί στην πράξη, διότι είναι σε θέση να προβλέψει με ακρίβεια τη χρεοκοπία. Είναι ευέλικτο, μπορεί να εκπαιδευτεί

σε διαφορετικά σύνολα δεδομένων, να εφαρμοστεί σε πλήθος εταιρειών και να χρησιμοποιηθεί από χρηματοπιστωτικά ιδρύματα, επενδυτές και ελεγκτές για την αξιολόγηση της οικονομικής κατάστασης των επιχειρήσεων.

Όμως παρόλο που τα αποτελέσματα που δίνει το μοντέλο είναι σαφή και ερμηνεύσιμα δεν μπορεί να υπολογιστεί εκ νέου ή να προγραμματιστεί σε διαφορετικό περιβάλλον εκτός από το λογισμικό Wolfram's Mathematica. Επομένως παραμένει αρκετά περίπλοκο για όσους δεν διαθέτουν γνώσεις στις τεχνολογίες πληροφοριών και επικοινωνίας (ΤΠΕ), οπότε και η εφαρμογή του περιορίζεται. Γι' αυτό και οι ερευνητές αναφέρουν ότι η μελλοντική έρευνα θα πρέπει να εστιάσει όχι μόνο στην περαιτέρω βελτίωση της απόδοσης του μοντέλου αλλά και στην ανάπτυξη μιας εφαρμογής φιλικής προς τους χρήστες, ώστε να είναι εφικτή η πρακτική εφαρμογή του.

4.12 Corporate failure prediction: An evaluation of deep learning vs discrete hazard models

Οι Alam, Gao και Jones (2021) ερεύνησαν το καινοτόμο μοντέλο deep learning **Deep Grassmannian Network (GrNet)** το οποίο χρησιμοποιεί δομή panel data για την πρόβλεψη της εταιρικής αποτυχίας. Πρόκειται για μία πλήρως συνδεδεμένη αρχιτεκτονική deep neural network που χρησιμοποιεί Grassmannian-valued data για την καταγραφή της δομής του panel data.

Τα δεδομένα προέρχονται από τις βάσεις δεδομένων (1) **COMPUSTAT** από όπου συλλέχθηκαν τα λογιστικά δεδομένα (βλ. πίνακα 51) και (2) **Center for Research in Security Prices (CRSP)** από όπου συλλέχθηκαν τα δεδομένα της χρηματιστηριακής αγοράς (βλ. πίνακα 52).

Πίνακας 51: Λογιστικά δεδομένα

(1) Liquidity and solvency items/ratios;			
ACT:	Current Assets	ACT-INV-	Current Assets and Inventories to Current Liabilities
		LCT:	
AP:	Accounts payable	AP-Sale:	Accounts Payable to Sale
AT:	Total Assets	CHE:	Cash and Short-term Investments
CH-AT:	Cash to Total Assets	CH-LC:	Cash to Current Liabilities
INVT-Sale:	Inventories to Sale	INVT:	Inventories
WCAP:	Working Capital	WCAP- AT:	Working Capital to Total Assets
IVST:	Short-Term Investments	LCT:	Current Liabilities
LCT-AGT:	Current Liabilities to Current Assets	LCT-AT:	Current Liabilities to Total Assets
LCT-CH- AT:	Current Liabilities less Cash to Total Assets	LCT-LT:	Current Liabilities to Liabilities
LCT-Sale:	Current Liabilities to Sales	LT:	Total Liabilities
TotalDebt:	Total Debt		
(2) Earnings and profitability items/ratios;			
DVPSP:	Dividends per Share	EPSFX:	Earnings Per Share (Diluted) -Excluding Extraordinary items
OIADP- LT:	Operating Income After Depreciation to Total Liabilities	OIADP-	Operating Income After Depreciation to Finance Division
		FINXINT:	Interest Expense
NI:	Net Income	OIADP- Sale:	Operating Income After Depreciation to Total Sales
NI-Sale:	Net Income to Total Sales	NI-AT:	Net Income to Total Assets
OIADP- AT:	Operating Income After Depreciation to Total Assets	OIADP:	Operating Income After Depreciation
RE-AT:	Retained Earnings to Total Sales	RE:	Retained Earnings
REVT:	Total Revenue	RE-LCT:	Retained Earnings to Total Current Liabilities
(3) Cash flow activities;			
FINCF:	Financing Activities – Net Cash Flow	FINXINT:	Finance Division Interest Expense
OIADP:	Operating Activities – Net Cash Flow	IVNCF:	Investing Activities Net Cash Flow
(4) Growth or change in items;			
INVT-G:	Growth in Inventories	XRD-G:	Growth in Research and Development Expense
NI-G:	Growth in Net Income	WCAP-G:	Growth in Working Capital
REVT-G:	Growth in Revenue		
(5) Leverage and capital structure measures;			
DLC-LT:	Debt in Current Liabilities to Total Liabilities	CHE-AT:	Cash and short-term investments to total assets
CHE-LCT:	Cash and short-term investments to current liabilities	DLTT:	Debt in Current Liabilities (DLC) Long-Term Debt
LCT- CHE-AT:	Current Liabilities less Cash and Short-Term Investments to Total Assets	LT-AT:	Total Liabilities to Total Assets
LT-TEQ:	TotalLiabilities to Total Stockholders' Equity	TotalDebt- AT:	Total Debt to Total Assets
TotalDebt- TEQ:	Total Debt to Total Stockholders' Equity	SEQ:	Stockholders Equity Parent Index Fundamental
TEQ:	Total Stockholders' Equity	TEQ-AT:	Total Stockholders' Equity to Total Assets
(6) Activity items/ratios;			
RECT:	Total Receivables	RECT- Sale:	Total Receivables to Total Sales
Sale-AT:	Total Sales to Total Assets		
(7) Investment in future growth (capital expenditure);			
IVLT:	Total Long-term Investments	XRD:	Research and Development Expense
(8) Intangible items/ratios;			
GDWL:	Goodwill (net)	GDWL-AT:	Goodwill to Total Assets
INTAN:	Intangible Assets	INTAN-AT:	Intangible Assets to Total Assets
INTAN- GDWL- AT:	Intangible assets and Goodwill to Total Assets		
(9) Additional Items;			
WDA:	Total Write-downs After-tax	WDA-AT:	Total Write-downs After-tax to Total Assets

Πηγή: (Alam κ.ά., 2021)

Πίνακας 52: Τα δεδομένα της χρηματιστηριακής αγοράς

EROI:	Excess return over S&P 500	BAS:	Bid Ask Spread
RETX:	Returns Without Dividends	Log-price:	Log of Price
VOL:	Volatility in returns	RSIZE:	Relative Size
MKVALT:	Total Market Value	CHE- MKVALT-	Cash and Short-term Investment to Total Market Equity and
		LT:	Total Liabilities
NI- MKVALT- LT:	Net Income to Total Market Equity and Total Liabilities	LT- MKVALT-	Total Liabilities to Total Market Equity and Total Liabilities
IVLT- IVST- MKVALT- LT:	Total Investment to Total Market Equity and Total Liabilities		

Πηγή: (Alam κ.ά., 2021)

Το σύνολο δεδομένων αποτελείται από 641.667 μηνιαίες παρατηρήσεις 6481 εισηγμένων εταιρειών της Β. Αμερικής κατά το χρονικό διάστημα 2001 ως 2018 από

τις οποίες οι 503.595 παρατηρήσεις προέρχονται από 3901 υγιείς εταιρείες και οι υπόλοιπες 138.072 παρατηρήσεις από 2580 αποτυχημένες και χρεοκοπημένες εταιρείες (βλ. πίνακα 53).

Πίνακας 53: Το σύνολο δεδομένων

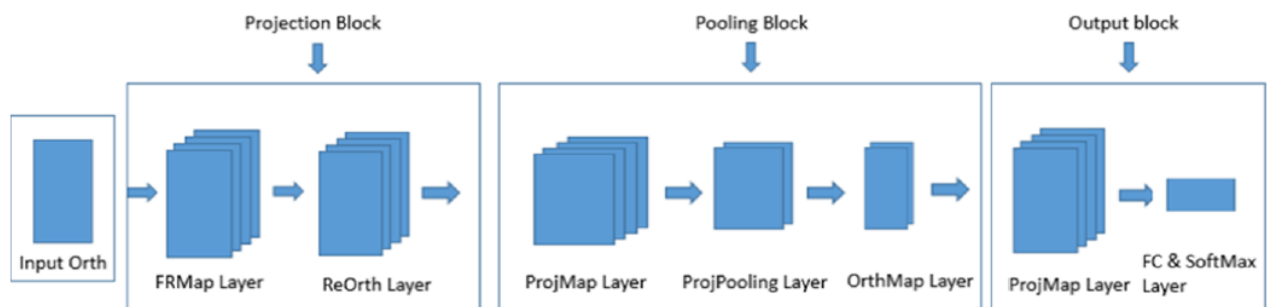
	Firm month observations	Unique firms
Distressed and bankrupt	138,072	2580
Non-bankrupt	503,595	3901
Total	641,667	6481

Πηγή: (Alam κ.ά., 2021)

Για την αντιμετώπιση των ακραίων τιμών, εφαρμόστηκε η μέθοδος winsorized στο 1^ο και 99^ο εκατοστημόριο και για τις χαμένες τιμές έγινε χρήση της τεχνικής singular value decomposition (SVD).

Το σύνολο δεδομένων είναι σε δομή panel. Πρόκειται για ένα μη ισορροπημένο data panel, καθώς η χρονική διάρκεια ζωής των 6481 εταιρειών είναι διαφορετική, δεδομένου ότι οι εταιρείες χρεοκοπούν σε διαφορετικές χρονικές περιόδους και σε διαφορετική φάση του κύκλου ζωής τους. Έτσι τα panel data αντιμετωπίστηκαν ως ένα set of vectors κατά μήκος της διάστασης του χρόνου. Έπειτα το subspace (υποχώρος) που δημιουργήθηκε από το set of vectors χρησιμοποιήθηκε για την αναπαράσταση των panel data. Αφού επιλέχθηκε η κατάλληλη διάσταση για όλα τα subspaces, αυτά έγιναν Grassmann points. Δηλαδή τα panel data αναπαραστάθηκαν ως Grassmann points με βάση τα οποία εξετάστηκε η γενική πληροφορία του panel data και όχι η πληροφορία για κάθε χρονική στιγμή ξεχωριστά.

Γενικά ένα μοντέλο **Deep Grassmannian Network (GrNet)** αποτελείται από ένα επίπεδο εισόδου, ένα επίπεδο εξόδου και μερικά κρυφά επίπεδα στα οποία συμπεριλαμβάνονται convolutional layers, pooling layers και normalization layers. Τα τρία βασικά δομικά στοιχεία του GrNet είναι το **Projection block**, το **Pooling block** και το **Output block** (βλ. εικόνα 45).



Εικόνα 45: Γραφική αναπαράσταση του GrNet

Πηγή: (Alam κ.ά., 2021)

Το **Projection block** αποτελείται από δύο επίπεδα, το Full Rank Mapping layer (FRMap) και το Re-orthonormalisation layer (ReOrth). Το FRMap layer λαμβάνει τις εισόδους των Grassmannian points τα οποία, στη συνέχεια, κανονικοποιούνται μέσω της διαδικασίας reorthonormalisation στο ReOrth layer για τη λήψη ενός ορθογώνιου πίνακα.

Έπειτα η έξοδος από το ReOrth layer περνά στο **Pooling block** το οποίο αποτελείται από τρία επίπεδα, το ProjMap layer, το ProjPooling layer και το OrthMap layer. Τα δεδομένα λοιπόν εισέρχονται στο ProjMap layer για τη διατήρηση του

Grassmann περιεχομένου και τη μεταφορά σε Euclidian data, ώστε στο ProjPooling layer να επιτευχθεί η μείωση διαστάσεων και η συγχώνευση πληροφοριών. Στη συνέχεια, εισάγονται στο OrthMap layer για κανονικοποίηση.

Οι παραπάνω διαδικασίες μπορούν να επαναληφθούν αρκετές φορές πριν συνδεθούν με το τελικό output block όπου εφαρμόζεται ένα ακόμα ProMap layer και στη συνέχεια εφαρμόζονται τα κλασικά Euclidean layers, όπως τα FC (fully connected layers) ή η συνάρτηση Softmax.

Στην παρούσα έρευνα, η δομή του μοντέλου **Deep Grassmannian Network** που χρησιμοποιήθηκε αποτελείται από 2 Projection-Pooling block(s) και ένα output block.

Για την εκπαίδευση του προτεινόμενου μοντέλου, το σύνολο δεδομένων των 6481 εταιρειών χωρίστηκε σε training dataset με 5833 εταιρείες (90%) και test dataset με 648 εταιρείες (10%).

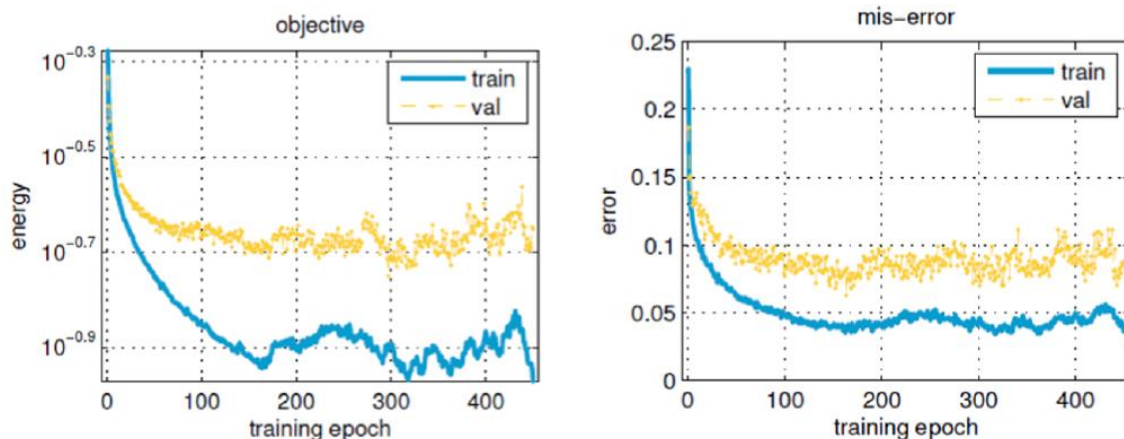
Οι παράμετροι ορίστηκαν ως εξής: Μέγεθος πρώτου κρυφού επίπεδου 80 με δύο pooling μεγέθους 40 και 20 αντίστοιχα, το learning rate σε 0,1 και το μέγεθος παρτίδας 30. Επίσης ο αριθμός των FRMap matrices ανά επίπεδο ορίστηκε σε 8 και για όλα τα ProjPooling layers ο αριθμός των περιπτώσεων σε 4. Η διάσταση των subspaces ορίστηκε σε 9 δεδομένου ότι η συντομότερη περίοδος συλλογής δεδομένων είναι οι 9 μήνες.

Επίσης κατά τα πειράματα χρησιμοποιήθηκαν δύο τύποι δεδομένων. Στη μία περίπτωση χρησιμοποιήθηκαν δεδομένα τα οποία κανονικοποιήθηκαν πριν μετατραπεί το panel σε αναπαράσταση Grassmann και στην άλλη χρησιμοποιήθηκαν μη κανονικοποιημένα δεδομένα.

Η εκτέλεση του GrNet έγινε στις 500 εποχές με ένα πλήρως αυτοματοποιημένο αλγόριθμο βαθιάς μάθησης με ένα τελικό SoftMax layer. Το καλύτερο μοντέλο εντοπίστηκε στις 450 εποχές και χρησιμοποιήθηκαν κανονικοποιημένα και μη κανονικοποιημένα δεδομένα. Σύμφωνα με τα αποτελέσματα του πειράματος, το accuracy του προτεινόμενου μοντέλου είναι 91,2%, το sensitivity 93,71% και το Specificity 90,28% με κανονικοποίηση.

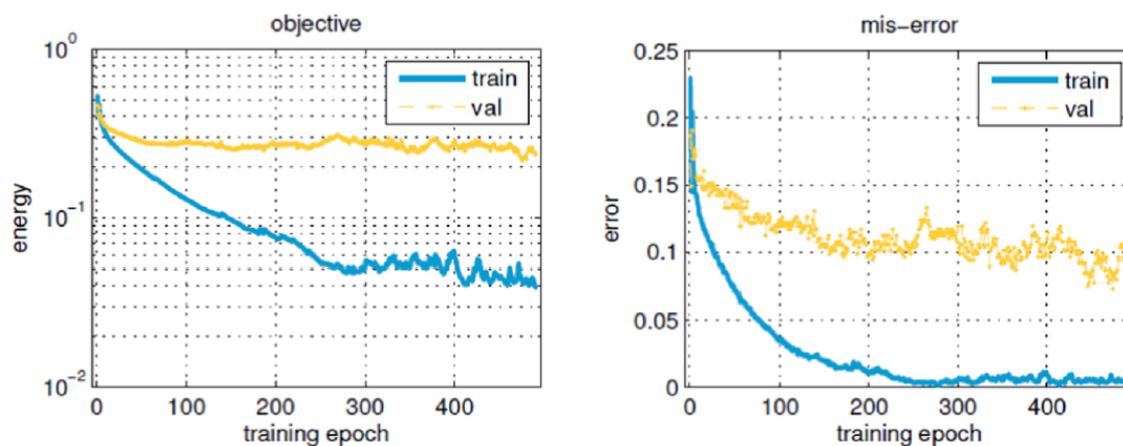
Έγινε επανάληψη του πειράματος με τη διαφορά ότι αυτή τη φορά η διάσταση των subspaces ορίστηκε σε 30 αντί για 9, δηλαδή η συντομότερη περίοδος συλλογής των δεδομένων ήταν 30 μήνες, επομένως το σύνολο δεδομένων μειώθηκε. Στις 500 εποχές, το accuracy του προτεινόμενου μοντέλου ήταν 91,3%, το sensitivity 93,52% και το Specificity 86,71% με κανονικοποίηση.

Παρακάτω απεικονίζονται το training objective trend και το validation mis-classification error των δύο παραπάνω πειραμάτων (βλ. εικόνες 46 και 47) :



Εικόνα 46: Training objective trend και Validation mis-classification error με διάσταση subspace 9

Πηγή: (Alam κ.ά., 2021)



Εικόνα 47: Training objective trend και Validation mis-classification error με διάσταση subspace 30

Πηγή: (Alam κ.ά., 2021)

Από τα αποτελέσματα των πειραμάτων προκύπτει ότι η απόδοση του μοντέλου δεν ενισχύεται σημαντικά από τη χρήση υψηλότερης διάστασης subspace. Έτσι συνίσταται η χρήση χαμηλότερης διάστασης subspace για την αναπαράσταση των panel data, ώστε να εκπαιδεύονται περισσότερα δεδομένα και να είναι δυνατή η εφαρμογή του μοντέλου και σε επιχειρήσεις με μικρότερα ιστορικά δεδομένα.

Τέλος, έγινε σύγκριση της απόδοσης του προτεινόμενου μοντέλου με το discrete-time hazard model όπου χρησιμοποιήθηκαν τα ίδια ακριβώς χαρακτηριστικά και για τα δύο μοντέλα, οι μεταβλητές κανονικοποιήθηκαν και λήφθηκαν υπόψη επιχειρήσεις με παρατηρήσεις για χρονικό διάστημα μεγαλύτερο από 8 μήνες. Το σύνολο δεδομένων και των δύο μοντέλων χωρίστηκε σε training dataset (90%) και testing dataset (10%).

Σύμφωνα με τα αποτελέσματα, το μοντέλο hazard εμφανίζει μεγαλύτερα ποσοστά στα μέτρα απόδοσης accuracy (93,3%) και specificity (93,53%) σε σύγκριση με το προτεινόμενο. Αντιθέτως, το προτεινόμενο μοντέλο εμφανίζει σημαντικά μεγαλύτερο sensitivity (93,71%) σε σχέση με το μοντέλο hazard (86,95%) και θεωρείται πολύ σημαντικό, διότι έχει διαπιστωθεί ότι ένα σφάλμα τύπου I είναι κατά 36 φορές περισσότερο δαπανηρό σε σύγκριση με ένα σφάλμα τύπου II (βλ. πίνακα 54).

Πίνακας 54: Τα αποτελέσματα απόδοσης των μοντέλων

	Discrete hazard model	Deep learning model
Overall accuracy	93.3%	91.2%
Sensitivity	86.95%	93.71%
Specificity	93.53%	90.28%

Πηγή: (Alam κ.ά., 2021)

Από τα παραπάνω διαπιστώνεται ότι το μοντέλο που αναπτύχθηκε είναι ικανό να χειρίζεται μεγάλα σύνολα δεδομένων με πολλές μεταβλητές και μη ισορροπημένα μεγέθη και μάλιστα έχει εξαιρετική απόδοση στη μέτρηση του sensitivity που αφορά την ακρίβεια ταξινόμησης σφάλματος τύπου I.

Οι ερευνητές επισημαίνουν ότι στο μέλλον θα μπορούσε να διερευνηθεί η πιθανότητα εφαρμογής του μοντέλου και σε άλλα προβλήματα ταξινόμησης (π.χ.

συγκριτικές και εξαγορές εταιρειών) καθώς επίσης και αξιολογήσεις συνεχιζόμενης δραστηριότητας από τις ελεγκτικές εταιρείες.

4.13 Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure

Οι Aljawazneh, Mora, Garcia-Sanchez και Castillo-Valdivieso (2021) εξέτασαν τρία προηγμένα μοντέλα Deep Learning (DL) και συγκεκριμένα τα μοντέλα (α) **Deep Belief Network** (DBN) που είναι ένα stochastic neural network, (β) **Multilayer Perceptron** με 6 Layers (MLP-6L) που είναι ένα feed-forward neural network και (γ) **Long-Short Term Memory** (LSTM) που είναι ένα recurrent neural network με στόχο την πρόβλεψη της αποτυχίας των επιχειρήσεων. Ο λόγος που επιλέχθηκαν τα μοντέλα αυτά είναι διότι το καθένα χρησιμοποιεί διαφορετικό τύπο νευρωνικού δικτύου καλύπτοντας έτσι ένα ευρύ φάσμα μεθόδων DL.

Κατά την έρευνα συλλέχθηκαν δεδομένα από Ισπανικές, Ταϊβανέζικες και Πολωνικές εταιρείες.

Το σύνολο δεδομένων των **Ισπανικών εταιρειών** συλλέχθηκε από τη βάση δεδομένων Infotel για το χρονικό διάστημα 1998-2003. Περιλαμβάνει συνολικά 2859 δείγματα εκ των οποίων τα 2797 αφορούν φερέγγυες εταιρείες και τα 62 χρεοκοπημένες. Συνολικά χρησιμοποιήθηκαν 33 μεταβλητές από τις οποίες οι 16 είναι οικονομικές και οι 17 μη χρηματοοικονομικές (βλ. πίνακα 55).

Πίνακας 55: Οι μεταβλητές των Ισπανικών εταιρειών

Financial Variables	Description	Type
Debt Structure	Long-Term Liabilities / Current Liabilities	Real
Debt Cost	Interest Cost / Total Liabilities	Real
Debt Paying Ability	Operating Cash Flow / Total Liabilities	Real
Debt Ratio	Total Assets / Total Liabilities	Real
Working Capital	Working Capital / Total Assets	Real
Warranty	Financial Warrant	Real
Operating Income Margin	Operating Income / Net Sales	Real
Return on Operating Assets	Operating Income / Average Operating Assets	Real
Return on Equity	Net Income / Average Total Equity	Real
Return on Assets	Net Income / Average Total Assets	Real
Stock Turnover	Cost of Sales / Average Inventory	Real
Asset Turnover	Net Sales / Average Total Assets	Real
Receivable Turnover	Net Sales / Average Receivables	Real
Asset Rotation	Asset allocation decisions	Real
Financial Solvency	Current Assets / Current Liabilities	Real
Acid Test	(Cash Equivalent + Marketable Securities + Net receivables) / Current Liabilities	Real
Non-financial Variables	Description	Type
Year	Corresponding to the sample	Integer
Size	Small Medium Large	Categorical
Number of employees		Integer
Age of the company		Integer
Type of company	Public Company Limited Liability Company Others	Categorical
Linked to a group	If the company is part of a group holding	Binary
Number of partners		Integer
Province code	Code of the location where the company is set	Categorical
Number of changes of location		Integer
Delay	If the company has submitted its annual accounts on time	Binary
Historic number of Judicial incidences	Since the company was created	Integer
Number of judicial incidences	Last year	Integer
Historic amount of money spent on judicial incidences	Since the company was created	Real
Amount of money spent on judicial incidences	Last year	Real
Historic number of serious incidences	Such as strikes, accidents...	Integer
Audited	If the company has been audited	Binary
Auditor's opinion	Favourable Exceptions Unfavourable	Categorical

Πηγή: (Aljawazneh κ.ά., 2021)

Το σύνολο δεδομένων των **Ταϊβανέζικων εταιρειών** συλλέχθηκε από το Taiwan Economic Journal και αφορά το χρονικό διάστημα 1999-2009. Περιλαμβάνει συνολικά 6819 δείγματα εκ των οποίων τα 6599 αφορούν φερέγγυες εταιρείες και τα 220 χρεοκοπημένες και αποτελείται από 95 οικονομικές μεταβλητές.

Το σύνολο δεδομένων των **Πολωνικών εταιρειών** συλλέχθηκε από το Emerging Markets Information Service (EMIS) και αφορά το χρονικό διάστημα 2007-2013 για τις χρεοκοπημένες εταιρείες και το χρονικό διάστημα 2007-2012 για τις φερέγγυες. Περιλαμβάνει συνολικά 10.000 δείγματα εκ των οποίων τα 9797 αφορούν τις φερέγγυες εταιρείες και τα 203 τις χρεοκοπημένες και αποτελείται από 64 αριθμητικές μεταβλητές.

Όμως δεδομένου ότι τα σύνολα δεδομένων είναι εξαιρετικά ανισόρροπα, κατά την προεπεξεργασία των δεδομένων, εφαρμόστηκαν τρεις διαφορετικές διαδικασίες εξισορρόπησης : (1) **Oversampling** που περιλαμβάνει τις τεχνικές : SMOTE, Borderline SMOTE, SMOTE-NC, SVM-SMOTE και ADAYSN, (2) **Hybrid Oversampling-Undersampling** που περιλαμβάνει τις τεχνικές : SMOTE-ENN και SMOTE-Tomek και (3) **Clustering-based balancing** που περιλαμβάνει την τεχνική : K-means SMOTE. Ο στόχος εφαρμογής διαφορετικών τεχνικών εξισορρόπησης είναι

για να διαπιστωθεί ποια τεχνική είναι η καταλληλότερη για τη βελτίωση της απόδοσης των μοντέλων DL.

Ο πίνακας 56 δείχνει την κατανομή των τάξεων πλειοψηφίας και μειοψηφίας των τριών datasets πριν και μετά την εφαρμογή των παραπάνω τεχνικών εξισορρόπησης δεδομένων. Η τεχνική SMOTE-NC δεν εφαρμόστηκε στα datasets των Ταϊβανέζικων και των Πολωνικών εταιρειών διότι δεν περιείχαν κατηγορικές μεταβλητές.

Πίνακας 56: Κατανομή δεδομένων πριν και μετά την εφαρμογή των τεχνικών εξισορρόπησης

Balancing techniques	Spanish companies' data		Taiwanese companies' data		Polish companies' data		
	Bankrupt	Solvent	Bankrupt	Solvent	Bankrupt	Solvent	
Original dataset	62	2797	220	6599	203	9797	
Oversampling	SMOTE	2797	2797	6599	6599	9797	9797
	BL-SMOTE	2797	2797	6599	6599	9797	9797
	SVM-SMOTE	1346	2797	2308	6599	4708	9797
	ADASYN	2806	2797	6523	6599	9860	9797
	SMOTE-NC	2797	2797	-	-	-	-
Oversampling-undersampling	SMOTE-Tomek	2765	2765	6565	6565	9794	9794
	SMOTE-ENN	2651	2494	6260	5433	9757	8546
Clustering based Oversampling	K-means SMOTE	2797	2797	6599	6599	9797	9797

Πηγή: (Aljawazneh κ.ά., 2021)

Σύμφωνα με τις πληροφορίες του πίνακα, οι τεχνικές SMOTE, Borderline SMOTE, SMOTE-NC, SMOTE-Tomek και K-means SMOTE δημιούργησαν ισορροπημένα dataset όπου κάθε τάξη περιέχει το ίδιο ποσοστό δεδομένων, η τεχνική ADASYN δημιούργησε περισσότερα δεδομένα για τη μειοψηφική τάξη ενώ η τεχνική SVM-SMOTE δημιούργησε λιγότερο ισορροπημένα datasets.

Κατά την έρευνα, τα τρία μοντέλα DL συγκρίθηκαν τόσο μεταξύ τους όσο και με τα μοντέλα: (α) **Random Forest (RF)**, **Support Vector Machine (SVM)** και **K-Nearest Neighbor (KNN)** που ανήκουν στις μεθόδους bagging based ensemble και (β): **Adaptive Boosting (AdaBoost)** και **EXtreme Gradient Boosting (XGBoost)** που ανήκουν στις μεθόδους boosting based ensemble. Ο σκοπός ήταν να ερευνηθεί ποια μέθοδος είναι η καλύτερη.

Για την εφαρμογή των μεθόδων DL χρησιμοποιήθηκε το λογισμικό Tensorflow, για τα μοντέλα RF, SVM, KNN και AdaBoost εφαρμόστηκε το Scikit-learn module ενώ για το XGBoost το Xgboost package.

Επίσης χρησιμοποιήθηκε η τεχνική 10-fold cross-validation όπου τα δεδομένα χωρίστηκαν τυχαία σε 10 ομάδες από τις οποίες οι εννέα αποτελούσαν το training dataset και η δέκατη το test dataset. Εκτελέστηκαν δέκα επαναλήψεις και σε κάθε επανάληψη χρησιμοποιήθηκε διαφορετική ομάδα ως test dataset.

Για τον προσδιορισμό των βέλτιστων υπερπαραμέτρων έγιναν πειράματα, οι τιμές των οποίων απεικονίζονται στους παρακάτω πίνακες (βλ. πίνακες 57 και 58).

Πίνακας 57: Οι υπερπαραμέτροι των μοντέλων DL

DL methods	Learning rate	Adam optimizer	Batch size	Epochs	Sigmoid	Softmax	ReLU	Categorical cross-entropy	Dropout
DBN	0.005, 0.1	-	40	150	✓	-	-	-	0.2
LSTM	-	✓	15	200	-	✓	-	✓	-
MLP-6L	-	✓	15	200	-	✓	✓	✓	0.2

Πηγή: (Aljawazneh κ.ά., 2021)

Πίνακας 58: Οι υπερπαράμετροι των μοντέλων ensemble

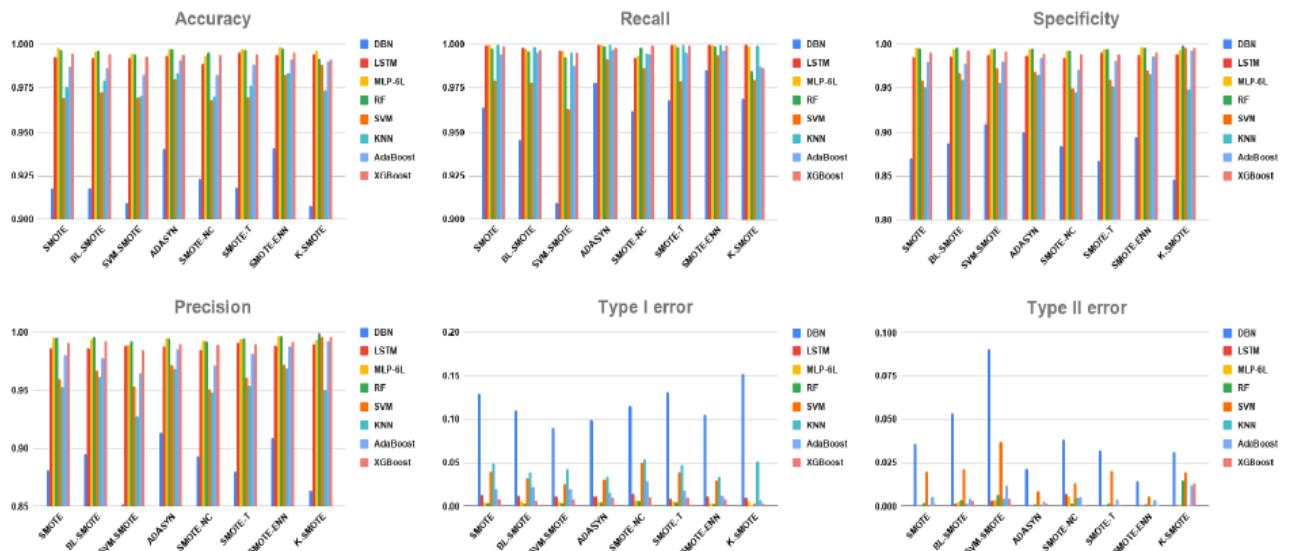
Ensemble methods	Kernel Function	N_estimators	Max_Samples	Max_Features	K	BallTree algorithm	Minkowski distance	Entropy
RF	-	100	1.0	1.0	-	-	-	✓
SVM	RBF	40	1.0	1.0	-	-	-	-
KNN	-	40	1.0	1.0	5	✓	✓	-
AdaBoost	-	50	1.0	1.0	-	-	-	✓
XGBoost	-	50	1.0	1.0	-	-	-	✓

Πηγή: (Aljawazneh κ.ά., 2021)

Συνολικά εκτελέστηκαν οκτώ πειράματα, δηλαδή ένα πείραμα για κάθε ένα από τα οκτώ μοντέλα και κάθε αλγόριθμος δοκιμάστηκε στα οκτώ balanced datasets που δημιουργήθηκαν με την εφαρμογή των τεχνικών εξισορρόπησης.

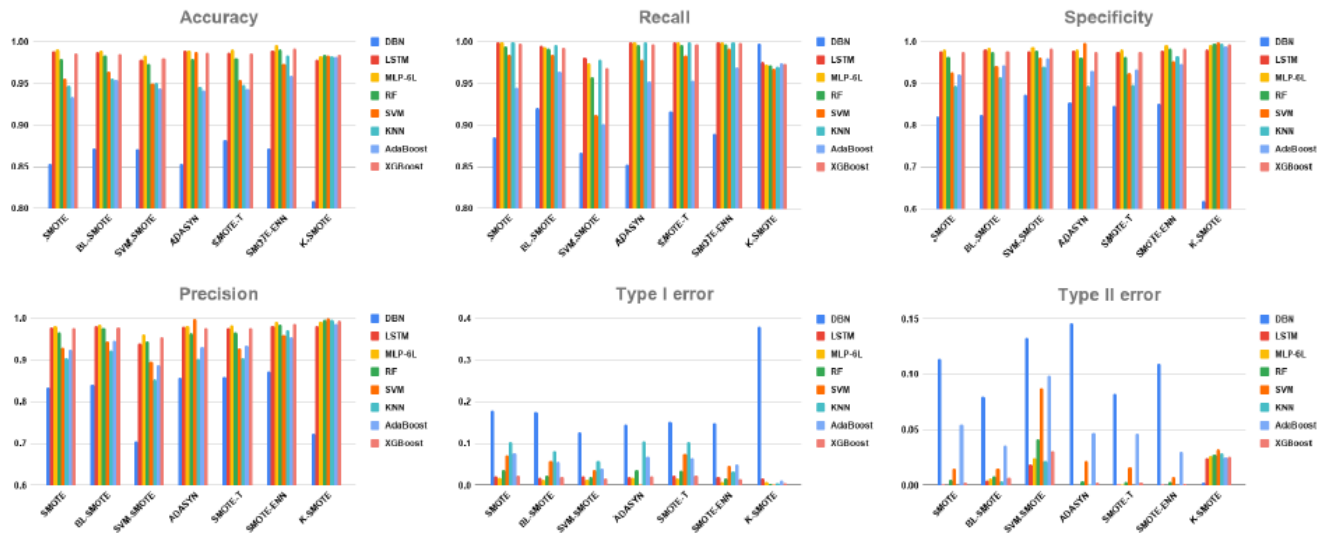
Για την αξιολόγηση της απόδοσης, χρησιμοποιήθηκαν τα μέτρα αξιολόγησης accuracy, recall (sensitivity), specificity, precision, type I error και type II error.

Σύμφωνα με τις εικόνες 48, 49 και 50, διαπιστώνεται ότι όλα τα μοντέλα είχαν υψηλή απόδοση στις περισσότερες μετρήσεις απόδοσης και χαμηλές τιμές στα σφάλματα type I και II.



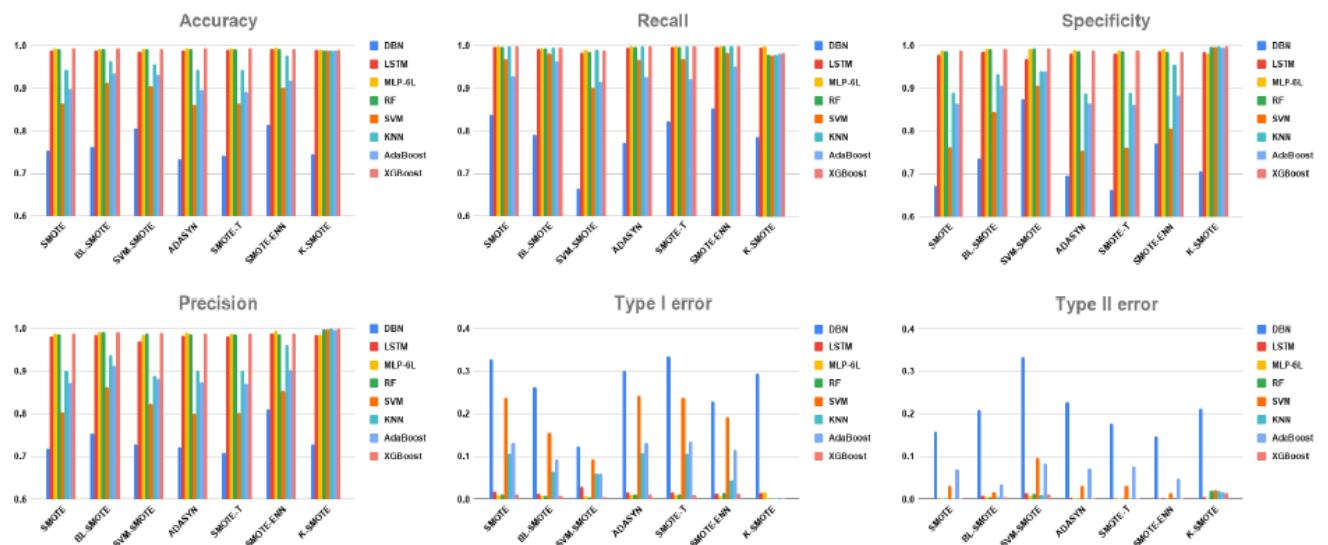
Εικόνα 48: Αποτελέσματα μετρήσεων απόδοσης του dataset των Ισπανικών εταιρειών

Πηγή: (Aljawazneh κ.ά., 2021)



Εικόνα 49: Αποτελέσματα μετρήσεων απόδοσης του dataset των Ταϊβανέζικων εταιρειών

Πηγή: (Aljawazneh κ.ά., 2021)



Εικόνα 50: Αποτελέσματα μετρήσεων απόδοσης του dataset των Πολωνικών εταιρειών

Πηγή: (Aljawazneh κ.ά., 2021)

Αναλυτικότερα, τα μοντέλα **MLP-6L**, **LSTM**, **RF** και **XGBoost** επιτυγχάνουν υψηλή απόδοση σε κάθε τεχνική εξισορρόπησης. Για παράδειγμα, στο σύνολο δεδομένων των Ισπανικών εταιρειών, η απόδοση των μέτρων accuracy και recall ξεπερνάει το 0,99 ενώ το type I error είναι κοντά στο 0,01 και το type II error μηδέν που σημαίνει ότι αυτές οι μέθοδοι έχουν χαμηλή εσφαλμένη ταξινόμηση των εταιρειών. Η χαμηλότερη απόδοση διαπιστώνεται στο μοντέλο **DBN** και αυτό αιτιολογείται από το γεγονός ότι έχει σχεδιαστεί για την επεξεργασία δεδομένων εικόνας. Επίσης τα μοντέλα **SVM**, **KNN** και **AdaBoost** επιτυγχάνουν επίσης καλή απόδοση. Το **SVM** έχει accuracy και recall πάνω από 0,96 ενώ το **KNN** recall περίπου

1,00 αλλά χαμηλότερο specificity σε σύγκριση με τα υπόλοιπα μοντέλα. Τέλος, το μοντέλο **AdaBoost** επιτυγχάνει accuracy και recall πάνω από 0,98.

Στη συνέχεια, έγινε σύγκριση των καλύτερων συνδυασμών μοντέλων και τεχνικών εξισορρόπησης που επιτεύχθηκαν κατά τα πειράματα για κάθε σύνολο δεδομένων (βλ. πίνακα 59).

Πίνακας 59: Σύγκριση απόδοσης των καλύτερων συνδυασμών μοντέλων και τεχνικών εξισορρόπησης

Dataset	Combinations	Accuracy	Recall	Specificity	Precision	Type I error	Type II error
Spanish companies' dataset	DBN + SMOTE-ENN	0.9414	0.9854	0.8946	0.9092	0.1054	0.0146
	LSTM + SMOTE-Tomek	0.9955	1.0	0.9908	0.9914	0.0092	0.0
	MLP-6L + SMOTE-ENN	0.9986	1.0	0.9972	0.9974	0.0028	0.0
	RF + SMOTE-ENN	0.9979	0.9989	0.9968	0.9970	0.0032	0.0011
	SVM + SMOTE-ENN	0.9825	0.9940	0.9703	0.9729	0.0297	0.0060
	KNN + SMOTE-ENN	0.9835	1.0	0.9659	0.9691	0.0341	0.0
	AdaBoost + ADASYN	0.9911	0.9970	0.9847	0.9859	0.0153	0.0030
	XGBoost + SMOTE-ENN	0.9955	0.9996	0.9912	0.9918	0.0088	0.0004
Taiwanese companies' dataset	DBN + SMOTE-ENN	0.8809	0.9465	0.8153	0.8367	0.1847	0.0535
	LSTM + ADASYN	0.9897	1.0	0.9795	0.9797	0.0205	0.0
	MLP + SMOTE-ENN	0.9963	1.0	0.9921	0.9931	0.0079	0.0
	RF + SMOTE-ENN	0.9906	0.9970	0.9834	0.9857	0.0166	0.0030
	SVM + SMOTE-ENN	0.9741	0.9921	0.9535	0.9605	0.0465	0.0079
	KNN + SMOTE-ENN	0.9843	1.0	0.9665	0.9714	0.0335	0.0
	AdaBoost + K-means SMOTE	0.9817	0.9750	0.9883	0.9882	0.0117	0.0250
	XGBoost + SMOTE-ENN	0.9923	0.9991	0.9846	0.9867	0.0154	0.0009
Polish companies' dataset	DBN + SMOTE-ENN	0.8151	0.8531	0.7718	0.8115	0.2282	0.1469
	LSTM + SMOTE-ENN	0.9931	0.9988	0.9866	0.9884	0.0134	0.0012
	MLP-6L + SMOTE-ENN	0.9967	0.9999	0.9930	0.9939	0.007	0.0001
	RF + SMOTE-ENN	0.9926	0.9992	0.9852	0.9871	0.0148	0.0008
	SVM + K-means SMOTE	0.9885	0.9784	0.9987	0.9986	0.0013	0.0216
	KNN + SMOTE-ENN	0.9788	0.9999	0.9546	0.9618	0.0454	0.0001
	AdaBoost + K-means SMOTE	0.9895	0.9822	0.9968	0.9968	0.0032	0.0178
	XGBoost + SMOTE-ENN	0.9936	0.9998	0.9864	0.9883	0.0136	0.0002

Πηγή: (Aljawazneh κ.ά., 2021)

Όσον αφορά τις **Ισπανικές εταιρείες**, ο συνδυασμός MLP-6L + SMOTE-ENN εμφανίζει τα καλύτερα αποτελέσματα και ακολουθεί ο συνδυασμός RF + SMOTE-ENN. Ο συνδυασμός MLP-6L + SMOTE-ENN δείχνει υψηλή απόδοση πρόβλεψης τόσο στις χρεοκοπημένες όσο και στις φερέγγυες εταιρείες.

Όσον αφορά το σύνολο δεδομένων των **Ταϊβανέζικων εταιρειών**, ο συνδυασμός MLP-6L + SMOTE-ENN εμφανίζει επίσης τα καλύτερα αποτελέσματα σε όλες τις μετρήσεις επιτυγχάνοντας την καλύτερη πρόβλεψη τόσο των χρεοκοπημένων όσο και των φερέγγυων εταιρειών. Οι συνδυασμοί LSTM + ADASYN και KNN + SMOTE-ENN επιτυγχάνουν τις ίδιες τιμές στο recall και Type II error, έχουν δηλαδή πολύ καλή απόδοση στην πρόβλεψη των χρεοκοπημένων εταιρειών.

Τέλος, στο σύνολο δεδομένων των **Πολωνικών εταιρειών**, που είναι και το πολυπλοκότερο όλων, ο συνδυασμός MLP-6L + SMOTE-ENN εμφανίζει τα καλύτερα αποτελέσματα στα μέτρα accuracy, recall και type I error, δηλαδή επιτυγχάνει καλύτερη πρόβλεψη των χρεοκοπημένων εταιρειών, ενώ ακολουθεί ο συνδυασμός SVM + K-means SMOTE με καλύτερα αποτελέσματα στα μέτρα specificity, precision και Type I error, δηλαδή επιτυγχάνει καλύτερη πρόβλεψη των φερέγγυων εταιρειών.

Συμπερασματικά, στα σύνολα δεδομένων των Ισπανικών και Ταϊβανέζικων εταιρειών, ο συνδυασμός **MLP-6L + SMOTE-ENN** επιτυγχάνει την καλύτερη πρόβλεψη των χρεοκοπημένων και των φερέγγυων εταιρειών (accuracy, recall, Type II error και specificity, precision, Type I error). Στο σύνολο δεδομένων των Πολωνικών εταιρειών, ο συνδυασμός **MLP-6L + SMOTE-ENN** επιτυγχάνει την καλύτερη πρόβλεψη των χρεοκοπημένων εταιρειών (accuracy, recall, Type II error) ενώ ο συνδυασμός **SVM + K-means SMOTE** την καλύτερη πρόβλεψη των φερέγγυων

(specificity, precision, Type I error). Επίσης η τεχνική εξισορρόπησης **SMOTE-ENN** έχει την καλύτερη απόδοση στα περισσότερα μοντέλα ανεξαρτήτως της πολυπλοκότητας των δεδομένων.

Από τα παραπάνω συμπεραίνεται η υπεροχή του MLP-6L + SMOTE-ENN στην πρόβλεψη της οικονομικής αποτυχίας (χρεοκοπίας) των εταιρειών σε σύγκριση με τις υπόλοιπες προσεγγίσεις.

Στο μέλλον οι ερευνητές προτείνουν (α) τη βελτίωση των μεθόδων DL με την εφαρμογή πιο εξελιγμένων αλγορίθμων, (β) την εξέταση πιο σύνθετων συνόλων δεδομένων και (γ) την ανάπτυξη νέων τεχνικών εξισορρόπησης για την αποτελεσματικότερη αντιμετώπιση της ανισοροπίας των δεδομένων.

4.14 Using Deep Learning Algorithms for CPAs' Going Concern Prediction

Στην παρούσα έρευνα, ο Jan (2021b) μελέτησε την ανάπτυξη μοντέλων πρόβλεψης για τη συνεχιζόμενη δραστηριότητα των επιχειρήσεων με σκοπό να βοηθήσει τους ελεγκτές και τους ορκωτούς λογιστές να κάνουν ορθές αξιολογήσεις και να εκδίδουν αξιόπιστες εκθέσεις ελέγχου. Για την ανάπτυξη των μοντέλων έγινε χρήση των μεθόδων **Deep Neural Network (DNN)** και **Recurrent Neural Network (RNN)**. Τα μοντέλα κατασκευάστηκαν με δύο τρόπους προκειμένου να εξεταστεί πιο μοντέλο επιτυγχάνει τη βέλτιστη απόδοση πρόβλεψης. Στη μία περίπτωση χρησιμοποιήθηκαν όλες οι επιλεγμένες μεταβλητές ενώ στη δεύτερη μόνο οι σημαντικότερες αυτών με τη βοήθεια της μεθόδου **classification and regression tree (CART)**.

Τα λογισμικά που χρησιμοποιήθηκαν για τη συλλογή των δεδομένων είναι τα TEJ Plus και TEJ Pro, για τη μοντελοποίηση και τον έλεγχο το Tensorflow.Keras και για το confusion matrix το Excel.

Τα δεδομένα συλλέχθηκαν από τη βάση δεδομένων Taiwan Economic Journal (TEJ) και αφορούν συνολικά 352 εταιρείες από τις οποίες οι 88 έλαβαν γνωμοδότηση <<με αμφιβολία συνεχιζόμενης δραστηριότητας>> ενώ οι 264 δεν έλαβαν. Η αναλογία του δείγματος είναι 1:3 και αφορά το χρονικό διάστημα 2002-2019. Στον πίνακα 60 απεικονίζονται οι εταιρείες ανά τομέα δραστηριότητας.

Πίνακας 60: Οι εταιρείες ανά τομέα δραστηριότητας

Industry Classification	Number of Companies with Going Concern Doubt	Number of Companies with No Going Concern Doubt
Food	1	3
Textile fiber	5	15
Electric machinery	4	12
Electric cable	3	9
Steel	5	15
Building materials and construction	5	15
Biotechnology and medical treatment	3	9
Semiconductors	7	21
Computers and peripheral equipment	4	12
Optoelectronics	15	45

Electronic components	12	36
Communication networks	1	3
Information service industry	3	9
Other electronic industries	4	12
Cultural and creative industries	4	12
Tourism	4	12
Others	8	24
Total	88	264

Πηγή: (Jan, 2021b)

Επίσης χρησιμοποιήθηκαν συνολικά 19 μεταβλητές από τις οποίες οι 16 είναι οικονομικές και οι 3 μη χρηματοοικονομικές (βλ. πίνακα 61).

Πίνακας 61: Οι μεταβλητές της έρευνας

No.	Variable	Description by Definition or Formula
X01	Debt ratio	Total liabilities ÷ Total assets
X02	Quick ratio	Quick assets ÷ Current liabilities
X03	Current ratio	Current assets ÷ Current liabilities
X04	D/E ratio	Total liabilities ÷ Total equity
X05	Current liabilities ratio	Current liabilities ÷ Total liabilities
X06	Ratio of current assets to total liabilities	Current assets ÷ Total liabilities
X07	Ratio of long-term funds to fixed assets	(Stockholders' equity + long-term liabilities) ÷ fixed assets
X08	Interest coverage ratio	EBIT ÷ Interest expense
X09	ROA	[Net income + interest expense × (1 – tax rate)] ÷ Average total assets
X10	ROE	Net income ÷ Average total equity
X11	Total assets turnover	Net Sales ÷ Total assets
X12	Accounts receivable turnover	Net sales ÷ Average accounts receivable
X13	Inventory turnover	Cost of goods sold ÷ Average inventory
X14	EPS	Net income ÷ Shares of common stock
X15	Gross margin	Gross profit ÷ Net sales
X16	Operating income ratio	Operating income ÷ Net sales
X17	Stockholding ratio of major shareholders	Stockholding ratio of major shareholders ÷ Shares of common stock
X18	Pledge ratio of directors and supervisors	Pledge ratio of directors and supervisors ÷ Shares of common stock
X19	Audited by BIG4 (the big four CPA firms) or not	1 for companies audited by BIG4, otherwise is 0

Πηγή: (Jan, 2021b)

Κατά τη μελέτη ακολουθήθηκαν δύο ερευνητικές διαδικασίες παράλληλα (βλ. εικόνα 51).



Εικόνα 51: Οι ερευνητικές διαδικασίες

Πηγή: (Jan, 2021b)

Αναλυτικότερα, κατά την πρώτη ερευνητική διαδικασία επιλέχθηκαν οι 10 σημαντικότερες μεταβλητές με τη βοήθεια της μεθόδου CART προκειμένου να χρησιμοποιηθούν στη συνέχεια από τους αλγόριθμους DNN και RNN για την κατασκευή του καλύτερου μοντέλου, ενώ κατά τη δεύτερη ερευνητική διαδικασία χρησιμοποιήθηκαν όλες οι μεταβλητές, δηλαδή και οι 19.

Οι 10 σημαντικότερες μεταβλητές που προέκυψαν με τη μέθοδο CART είναι 9 οικονομικές και 1 μη χρηματοοικονομική μεταβλητή (βλ. πίνακα 62).

Πίνακας 62: Οι δέκα σημαντικότερες μεταβλητές που επιλέχθηκαν με τη μέθοδο CART

No.	Variable	Variable Importance
X01	Debt ratio	0.2623
X10	ROE	0.1820
X04	D/E ratio	0.1431
X09	ROA	0.0964
X14	EPS	0.0834
X06	Ratio of current assets to total liabilities	0.0719
X15	Gross margin	0.0583
X18	Pledge ratio of directors and supervisors	0.0281
X03	Current ratio	0.0162
X02	Quick ratio	0.0159

Πηγή: (Jan, 2021b)

Το σύνολο δεδομένων χωρίστηκε τυχαία με τη μέθοδο holdout validation. Έτσι για τη διαδικασία μοντελοποίησης, χρησιμοποιήθηκε το 70% του συνόλου δεδομένων εκ του οποίου το 52,5% χρησιμοποιήθηκε ως training dataset και το υπόλοιπο 17,5% ως validation dataset. Ενώ το υπόλοιπο 30% του συνόλου δεδομένων χρησιμοποιήθηκε ως test dataset. Για τη βελτιστοποίηση των μοντέλων χρησιμοποιήθηκε η συνάρτηση loss function. Επίσης χρησιμοποιήθηκε ο βελτιστοποιητής Adam και οι βέλτιστες παράμετροι καθορίστηκαν ως εξής : learning rate = 0,001, beta 1 = 0,9, beta 2 = 0,999,

$\epsilon = 10^{-7}$, batch size = 4, epochs = 150, activation = ReLU and Sigmoid. Η επιλογή των δεδομένων έγινε με τη μεθόδου random sampling without replacement.

Για την αξιολόγηση της απόδοσης των μοντέλων, χρησιμοποιήθηκε το matrix confusion με τα μέτρα απόδοσης accuracy, precision, sensitivity (recall), specificity, and F1 score ενώ κατά την αξιολόγηση δόθηκε ιδιαίτερη βαρύτητα στο μέτρο accuracy.

Στη συνέχεια έγιναν τέσσερα πειράματα. Στα δύο πρώτα έγινε χρήση μόνο των 10 σημαντικότερων μεταβλητών, ενώ στα δύο επόμενα έγινε χρήση όλων των μεταβλητών.

Αναλυτικότερα, στο **πρώτο πείραμα** κατασκευάστηκε το μοντέλο **CART-DNN** όπου οι 10 σημαντικότερες μεταβλητές, χρησιμοποιήθηκαν από το αλγόριθμο DNN για την εκπαίδευση του καλύτερου μοντέλου το οποίο προέκυψε στις 150 εποχές με χρόνο εκπαίδευσης 500 microseconds. Τα ποσοστά ακρίβειας του training data set είναι 97,83%, του validation dataset 91,94% και του test dataset 93,40%. Τα αποτελέσματα δείχνουν ότι το μοντέλο είναι αρκετά σταθερό. Επίσης τα ποσοστά σφάλματος τύπου I είναι 3,77% και τύπου II 2,83% (βλ. πίνακα 63). Επίσης στο confusion matrix, τα μέτρα απόδοσης είναι υψηλά με accuracy = 93,40%, precision = 85,71%, sensitivity (recall) = 88,89%, specificity = 94,94%, F1-score = 87,27% (βλ. πίνακα 64), επομένως **η απόδοση του μοντέλου είναι αρκετά καλή.**

Στο **δεύτερο πείραμα** κατασκευάστηκε το μοντέλο **CART-RNN** όπου και πάλι χρησιμοποιήθηκαν οι 10 σημαντικότερες μεταβλητές, από το αλγόριθμο RNN για την εκπαίδευση του καλύτερου μοντέλου το οποίο προέκυψε στις 150 εποχές με χρόνο εκπαίδευσης 500 microseconds. Τα ποσοστά ακρίβειας του training και validation dataset είναι 92,93% και 93,55% αντίστοιχα, ενώ στο test dataset 95,28%. Επομένως το μοντέλο είναι αρκετά σταθερό. Τα ποσοστά σφάλματος τύπου I και τύπου II είναι 2,83% και 1,89% αντίστοιχα (βλ. πίνακα 63). Στο confusion matrix, τα μέτρα απόδοσης είναι επίσης υψηλά με accuracy = 95,28%, precision = 88,46%, sensitivity (recall) = 92,00%, specificity = 96,30%, and F1-score = 90,19% (βλ. πίνακα 64), άρα **η απόδοση του μοντέλου θεωρείται ότι είναι πολύ καλή.**

Στο **τρίτο πείραμα** κατασκευάστηκε το μοντέλο **DNN** όπου χρησιμοποιήθηκαν και οι 19 μεταβλητές απευθείας από το αλγόριθμο DNN για την εκπαίδευση του καλύτερου μοντέλου που προέκυψε στις 150 εποχές με χρόνο εκπαίδευσης 500 microseconds. Τα ποσοστά ακρίβειας του training, validation και test dataset είναι 96,74%, 91,94% και 88,68% αντίστοιχα και συμπεραίνεται ότι το μοντέλο είναι αρκετά σταθερό. Τα ποσοστά σφάλματος τύπου I και τύπου II είναι 3,77% και 7,55% αντίστοιχα (βλ. πίνακα 63). Στο confusion matrix, τα μέτρα απόδοσης είναι accuracy = 88,68%, precision = 86,21%, sensitivity (recall) = 75,76%, specificity = 94,52%, and F1-score = 80,64% (βλ. πίνακα 64), επομένως **η απόδοση του μοντέλου είναι καλή.**

Τέλος, στο **τέταρτο πείραμα** κατασκευάστηκε το μοντέλο **RNN** όπου χρησιμοποιήθηκαν οι 19 μεταβλητές απευθείας από το αλγόριθμο RNN για την εκπαίδευση του καλύτερου μοντέλου που προέκυψε στις 150 εποχές με χρόνο εκπαίδευσης 500 microseconds. Τα ποσοστά ακρίβειας των training, validation και test dataset είναι 94,57%, 88,71% και 90,57% αντίστοιχα, επομένως το μοντέλο είναι αρκετά σταθερό. Τα ποσοστά σφάλματος τύπου I είναι 2,83% και τύπου II 6,60% (βλ. πίνακα 63). Τα μέτρα απόδοσης του confusion matrix είναι accuracy = 90,57%, precision = 89,29%, sensitivity (recall) = 78,13%, specificity = 95,95%, and F1-score = 83,33% (βλ. πίνακας 64), άρα **η απόδοση του μοντέλου είναι καλή.**

Πίνακας 63: Τα αποτελέσματα απόδοσης των μοντέλων

Model	Training Dataset	Validation Dataset	Test Dataset	Average	Type I Error	Type II Error
CART-DNN	97.83%	91.94%	93.40%	94.39%	3.77%	2.83%
CART-RNN	92.93%	93.55%	95.28%	93.92%	2.83%	1.89%
DNN	96.74%	91.94%	88.68%	92.45%	3.77%	7.55%
RNN	94.57%	88.71%	90.57%	91.28%	2.83%	6.60%

Πηγή: (Jan, 2021b)

Πίνακας 64: Το confusion matrix των μοντέλων

Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1 Score	Training Time
CART-DNN	93.40%	85.71%	88.89%	94.94%	87.27%	500 μs
CART-RNN	95.28%	88.46%	92.00%	96.30%	90.19%	500 μs
DNN	88.68%	86.21%	75.76%	94.52%	80.64%	500 μs
RNN	90.57%	89.29%	78.13%	95.95%	83.33%	500 μs

Πηγή: (Jan, 2021b)

Από τα παραπάνω διαπιστώνεται ότι το μοντέλο CART-RNN έχει την υψηλότερη ακρίβεια σε σχέση με τα υπόλοιπα μοντέλα, ακολουθεί το μοντέλο CART-DNN, τρίτο στη σειρά είναι το μοντέλο RNN και τέλος το DNN. Γενικά τα ποσοστά ακρίβειας και των τεσσάρων μοντέλων είναι πάνω από 88% και θεωρείται ότι είναι υψηλά. Επίσης διαπιστώνεται ότι το μοντέλο RNN έχει καλύτερη προβλεπτική απόδοση σε σύγκριση με το DNN και μάλιστα η επιλογή των σημαντικότερων μεταβλητών με τη μέθοδο CART βελτιώνει την ακρίβεια πρόβλεψης και των δύο μοντέλων.

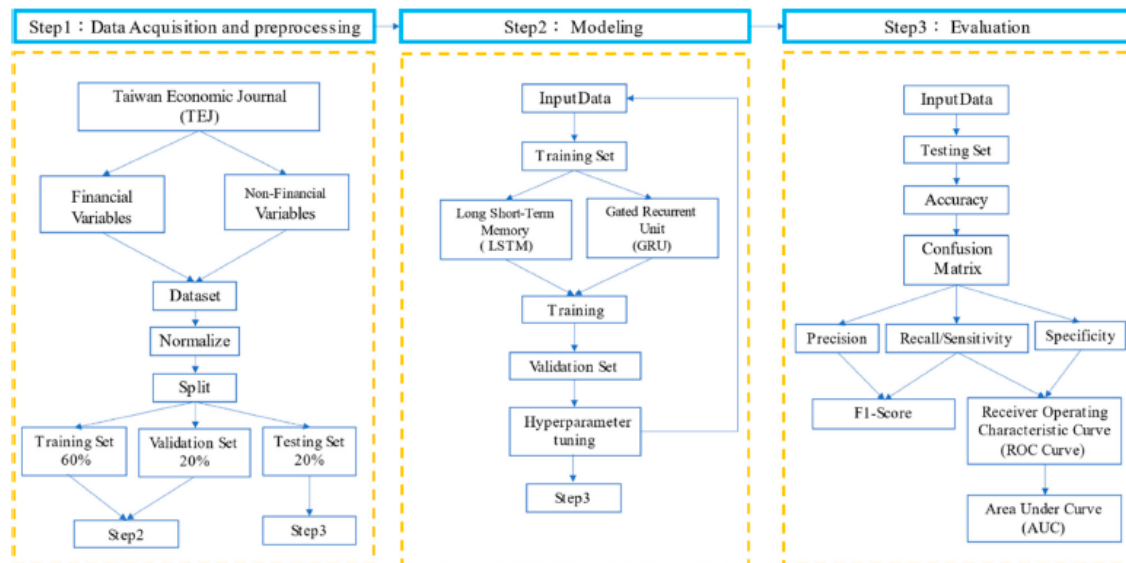
Ο Jan (2021b) επισημαίνει ότι η μελλοντική έρευνα θα μπορούσε να εστιάσει (α) στην επιλογή άλλων αλγόριθμων βαθιάς μάθησης (π.χ. deep belief networks, convolutional neural networks, convolutional deep belief networks, long short-term memory και non-iterative ANN), ώστε να μελετηθεί η συμπεριφορά και η προβλεπτική τους ικανότητα, καθώς επίσης και (2) στην επιλογή άλλων μεταβλητών καταλληλότερων για κάθε χώρα ξεχωριστά ανάλογα με το οικονομικό περιβάλλον, το χρηματοπιστωτικό σύστημα και τους κανονισμούς ελέγχου που τη διέπουν.

4.15 Artificial intelligence in corporate sustainability: using LSTM and GRU for going concern prediction

Στην έρευνα των Chi και Chu (2021) έγινε μια προσπάθεια κατασκευής δύο υψηλής ακρίβειας μοντέλων με τη χρήση των αλγορίθμων deep learning Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) για την πρόβλεψη της συνεχιζόμενης δραστηριότητας των επιχειρήσεων. Στόχος της έρευνας ήταν να βοηθήσει τους ορκωτούς λογιστές (CPA) και τους ελεγκτές στην έκδοση ακριβέστερων και ορθότερων εκθέσεων ελέγχου.

Ο σχεδιασμός της έρευνας αποτελείται από τα εξής **3 στάδια** (βλ. εικόνα 52):

- Συλλογή και προεπεξεργασία δεδομένων
- Μοντελοποίηση
- Αξιολόγηση



Εικόνα 52: Ο σχεδιασμός της έρευνας

Πηγή: (Chi & Chu, 2021)

Το **σύνολο δεδομένων** συλλέχθηκε από το Taiwan Economic Journal (TEJ) και αφορά συνολικά 258 εισηγμένες επιχειρήσεις στο Taiwan Stock Exchange Corporation (TWSE) και το Ταϊπέι Exchange (TPEX) για το χρονικό διάστημα 2004-2019. Από το σύνολο των εισηγμένων επιχειρήσεων, οι 86 είχαν λάβει αμφιβολία συνεχιζόμενης δραστηριότητας και οι 172 δεν είχαν λάβει. Η κατανομή των επιχειρήσεων ανά κλάδο απεικονίζεται στον παρακάτω πίνακα (βλ. πίνακα 65):

Πίνακας 65: Οι επιχειρήσεις ανά κλάδο δραστηριότητας

Industry	Going Concern Doubt	No Going Concern Doubt	Total
Semiconductor	6	12	18
Optoelectronics	13	26	39
Electric machinery	3	6	9
Electronic components	11	22	33
Electric appliances and cables	2	4	6
Computers and peripherals	4	8	12
Communication and networking	1	2	3
Information services	2	4	6
Other electronics	5	10	15
Biotech and medicare	3	6	9
Steel	4	8	12
Building materials and construction	5	10	15
Textiles	5	10	15
Food	1	2	3
Cultural and creative industry	4	8	12
Trade and department stores	2	4	6
Shipping	2	4	6
Oil, electricity, and gas	1	2	3
Tourism	4	8	12
Others	8	16	24
Total	86	172	258

Πηγή: (Chi & Chu, 2021)

Συνολικά χρησιμοποιήθηκαν 20 μεταβλητές και συγκεκριμένα 16 χρηματοοικονομικές και 4 μη χρηματοοικονομικές (ή μεταβλητές εταιρικής διακυβέρνησης) οι οποίες συνοψίζονται στον κάτωθι πίνακα (βλ. πίνακα 66):

Πίνακας 66: Οι μεταβλητές της έρευνας

Code	Variable	Variable Definition or Equation (The Year before the Year of Going Concern Doubt)
X01	Non-current liabilities/total assets	Non-current liabilities/total assets
X02	Working capital/total assets	Working capital/total assets
X03	Retained earnings/total assets	Retained earnings/total assets
X04	Fixed assets/total assets	Fixed assets/total assets
X05	Liability ratio	Total liabilities/total assets
X06	Liability to equity ratio	Total liabilities/shareholders' equity
X07	Current liabilities/revenues	Current liabilities/revenues
X08	Working capital/revenues	Working capital/revenues
X09	Cash flows from operating activities/revenues	Cash flows from operating activities/revenues
X10	Operating cash flow ratio	Cash flows from operating activities/current liabilities
X11	Cash flow adequacy ratio	Operating cash flows over the past five years/(capital expenditures + increase in inventory + cash dividends over the past five years)
X12	Inventory turnover	Cost of goods sold/average inventory
X13	Current ratio	Current assets/current liabilities
X14	Return on equity	Net income/average shareholders' equity
X15	Growth in asset	(Total assets at the end of the current year—total assets at the end of the previous year/total assets at the end of the previous year
X16	Growth in revenues	(Revenues during the current year—revenues during the previous year)/revenues during the previous year
X17	No. of independent directors	No. of independent directors
X18	Percentage of shareholdings by directors and supervisors	No. of shares held by directors and supervisors/No. of ordinary shares outstanding at the end of the period
X19	Percentage of shares pledged by directors and supervisors	No. of shares pledged by directors and supervisors/No. of shares held by directors and supervisors
X20	CEO/Chairperson duality	1 if Chairperson is also CEO and 0 if not

Πηγή: (Chi & Chu, 2021)

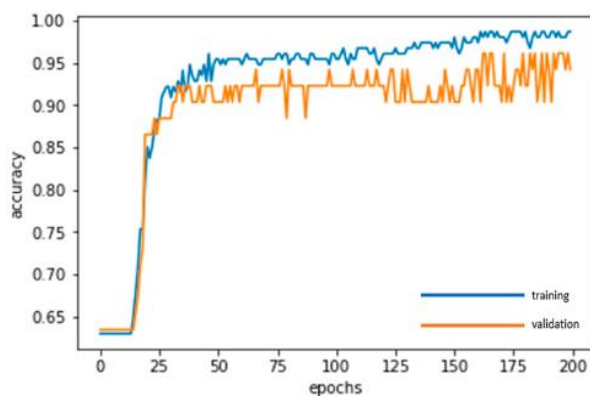
Αρχικά, τα ακατέργαστα δεδομένα κανονικοποιήθηκαν και τυποποιήθηκαν, ώστε οι τιμές των δεδομένων να είναι μεταξύ 0 και 1 με σκοπό να διευκολυνθεί η σύγκριση και η ανάλυση των δεδομένων και να βελτιωθεί η ακρίβεια των αλγορίθμων.

Στη συνέχεια, το σύνολο δεδομένων χωρίστηκε τυχαία σε training dataset (60%), validation dataset (20%) και test dataset (20%).

Για τη μοντελοποίηση χρησιμοποιήθηκε το λογισμικό TensorFlow. Επίσης χρησιμοποιήθηκαν ο βελτιστοποιητής Adam, η Sigmoid συνάρτηση ενεργοποίησης και η συνάρτηση απώλειας (loss function) binary cross-entropy.

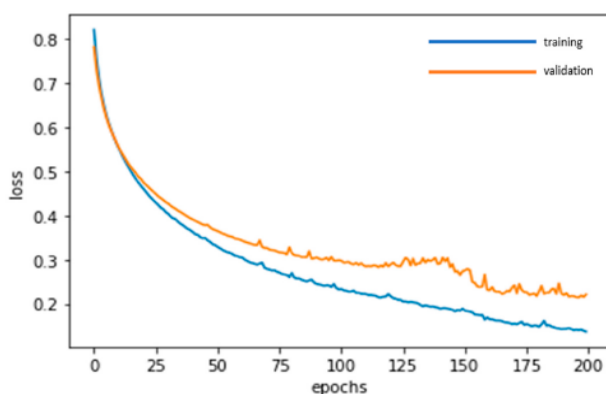
Για την αξιολόγηση της απόδοσης των μοντέλων, χρησιμοποιήθηκαν το confusion matrix με τα μέτρα απόδοσης accuracy, precision, recall/sensitivity, specificity και F1-score καθώς επίσης τα error rates Type I και Type II.

Για το **μοντέλο LSTM** εφαρμόστηκαν οι διαδικασίες εκπαίδευσης και επικύρωσης επαναλαμβανόμενα μέχρι που προέκυψε ένα σταθερό μοντέλο. Σύμφωνα με τα παρακάτω διαγράμματα (βλ. εικόνες 53 και 54), διαπιστώνεται ότι η συνάρτηση απώλειας και η ακρίβεια συγκλίνουν σταδιακά και σταθεροποιούνται μετά από 200 εποχές και χρόνο 1,4 δευτερόλεπτα. Το training dataset έχει ακρίβεια 98,70% και το validation dataset 94,23%.



Εικόνα 53: Loss curve του μοντέλου LSTM

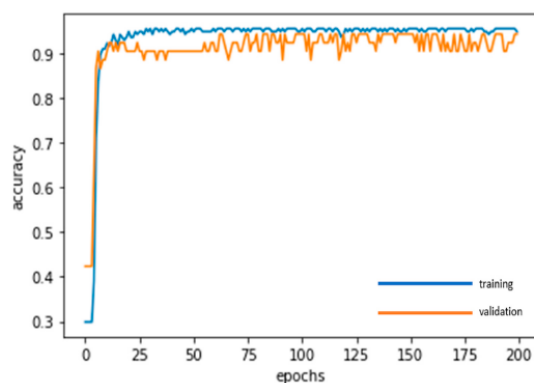
Πηγή: (Chi & Chu, 2021)



Εικόνα 54: Accuracy curve του μοντέλου LSTM

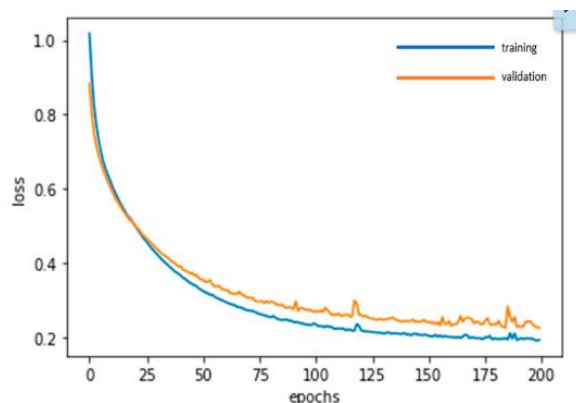
Πηγή: (Chi & Chu, 2021)

Οι διαδικασίες εκπαίδευσης και επικύρωσης επαναλήφθηκαν και για το **μοντέλο GRU** μέχρι να προκύψει ένα σταθερό μοντέλο. Σύμφωνα με τα παρακάτω διαγράμματα (βλ. εικόνες 55 και 56), διαπιστώνεται ότι η συνάρτηση απώλειας και η ακρίβεια συγκλίνουν σταδιακά και σταθεροποιούνται μετά από 200 εποχές και χρόνο 1,4 δευτερόλεπτα. Το training dataset έχει ακρίβεια 94,81% και το validation dataset 94,23%.



Εικόνα 55: Loss curve του μοντέλου GRU

Πηγή: (Chi & Chu, 2021)



Εικόνα 56: Accuracy curve του μοντέλου GRU

Πηγή: (Chi & Chu, 2021)

Μετά ακολούθησε η αξιολόγηση των μοντέλων με τη χρήση του test dataset.

Για το **μοντέλο LSTM**, τα μέτρα απόδοσης του confusion matrix είναι: accuracy 96,15%, precision 90%, recall/sensitivity 90%, specificity 97,62% και F1-score 90%. Επίσης τα error rates Type I και Type II είναι 1,92%. Από τα παραπάνω διαπιστώνει κανείς ότι το μοντέλο έχει καλή απόδοση και το ποσοστό των σφαλμάτων είναι πολύ χαμηλό (βλ. πίνακα 67).

Για το **μοντέλο GRU**, τα μέτρα απόδοσης του confusion matrix είναι: accuracy 94,23%, precision 94,12%, recall/sensitivity 88,89%, specificity 97,06% και F1-score 91,43%. Επίσης τα error rates Type I και Type II είναι 1,92% και 3,85% αντίστοιχα. Επομένως το μοντέλο έχει καλή απόδοση και το ποσοστό των σφαλμάτων είναι επίσης πολύ χαμηλό (βλ. πίνακα 67).

Πίνακας 67: Σύγκριση αποτελεσμάτων των μοντέλων LSTM και GRU

Model	Accuracy	Precision	Recall/Sensitivity	Specificity	F1-Score	Type I Error Rate	Type II Error Rate
LSTM	96.15%	90.00%	90.00%	97.62%	90.00%	1.92%	1.92%
GRU	94.23%	94.12%	88.89%	97.06%	91.43%	1.92%	3.85%
Outperformer	LSTM	GRU	LSTM	LSTM	GRU	Equivalent performance	LSTM

Πηγή: (Chi & Chu, 2021)

Από τα παραπάνω γίνεται κατανοητό ότι τα μοντέλα LSTM και GRU είναι αξιόπιστα και αποτελεσματικά, διότι επιτυγχάνουν υψηλή ακρίβεια ενώ τα ποσοστά σφαλμάτων type I και Type II είναι εξαιρετικά χαμηλά. Το χαμηλό ποσοστό σφαλμάτων σημαίνει χαμηλό ποσοστό εσφαλμένης πρόβλεψης. Επομένως η χρήση τους μπορεί να μειώσει τους κινδύνους και τα κόστη που μπορεί να προκύψουν από λάθη και παραλείψεις κατά τις διαδικασίες ελέγχου παρέχοντας σημαντική βοήθεια όχι μόνο στους ορκωτούς λογιστές και τους ελεγκτές αλλά και σε αναλυτές, συμβούλους επιχειρήσεων, οργανισμούς αξιολόγησης πιστοληπτικής ικανότητας και στη διοίκηση των επιχειρήσεων για θέματα εταιρικής βιωσιμότητας και διαχείρισης κινδύνου.

Οι ερευνητές επισημαίνουν ότι οι μελλοντικές έρευνες θα πρέπει να προσανατολιστούν στη μελέτη και άλλων μοντέλων deep learning (π.χ. ANN, CNN, DNN, RNN κ.α.). Επίσης επισημαίνουν ότι οι επιλεγμένες μεταβλητές θα πρέπει να προσαρμόζονται στο προφίλ των επιχειρήσεων, το περιβάλλον, τα πρότυπα και τους κανονισμούς που ισχύουν και επίσης να περιλαμβάνονται και μη χρηματοοικονομικές μεταβλητές, ώστε να αναπτύσσονται μοντέλα αξιόπιστα και αποτελεσματικά, ικανά να προβλέψουν τη συνεχιζόμενη δραστηριότητα των επιχειρήσεων.

5

Ανάλυση των ερευνητικών άρθρων

Στο κεφάλαιο αυτό γίνεται ανάλυση των μοντέλων deep learning που αναπτύχθηκαν κατά τις έρευνες, αναφορικά με του τύπους των νευρωνικών δικτύων που επιλέχθηκαν, τον σκοπό της πρόβλεψης κάθε μοντέλου, τους τύπους δεδομένων που συλλέχθηκαν, τις τεχνικές εξισορρόπησης δεδομένων, τις μεθόδους επεξεργασίας φυσικής γλώσσας, τις τεχνικές επιλογής χαρακτηριστικών και τις τεχνικές επικύρωσης που εφαρμόστηκαν, καθώς επίσης και τα μέτρα απόδοσης που λήφθηκαν υπόψη για την αξιολόγηση της απόδοσής τους και τις μεθόδους στατιστικής ανάλυσης που εφαρμόστηκαν για τη στατιστική επιβεβαίωση των αποτελεσμάτων τους.

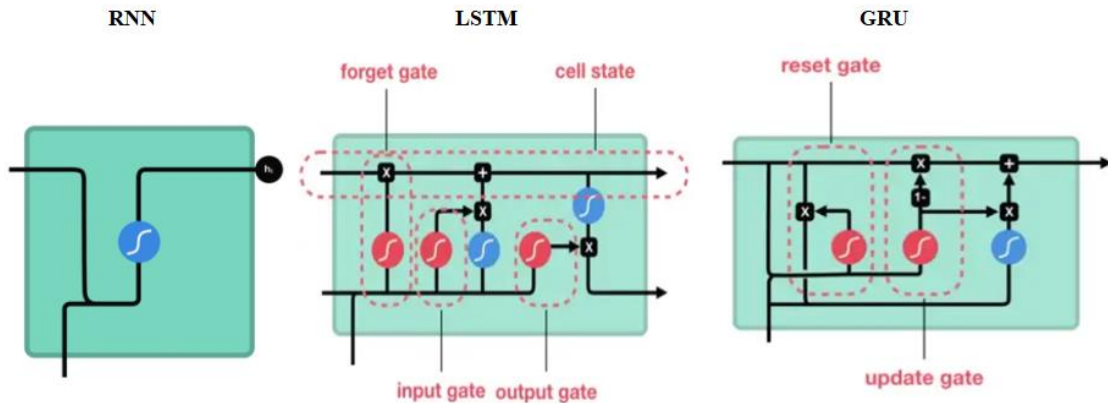
5.1 Τύποι νευρωνικών δικτύων

Οι περισσότερες έρευνες επικεντρώθηκαν στην ανάπτυξη μοντέλων που βασίζονται στους αλγόριθμους Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN) και Multi-Layer Perceptrons (MLP) (βλ. πίνακα 1).

5.1.1 Recurrent Neural Networks (RNN)

Το Recurrent Neural Networks (**RNN**) είναι ένα μοντέλο ακολουθίας (βλ. εικόνα 57) και έχει την ικανότητα καταγραφής της χρονικής εξάρτησης διαδοχικών δεδομένων με ακρίβεια. Εφαρμόζεται ήδη με επιτυχία στην ανάλυση χρονικών διαδοχικών δεδομένων που αφορούν κείμενα και βίντεο (Jan, 2021a). Το μειονέκτημά του είναι ότι αδυνατεί να διατηρήσει μακροπρόθεσμη μνήμη, λόγω του vanishing gradient problem που αποτρέπει την ενημέρωση των βαρών σε μικρότερα επίπεδα κατά το backpropagation. Πρόκειται λοιπόν για ένα μοντέλο βραχυπρόθεσμης μνήμης (Chi & Chu, 2021; Jan, 2021a; Jan, 2021b; Xiuguo & Shengyong, 2022).

Οι αλγόριθμοι Long Short-Term Memory (LSTM) και Gated Recurrent Unit (GRU) είναι τύποι RNN και αναπτύχθηκαν για να αντιμετωπιστούν τα ελαττώματα του αλγορίθμου RNN (Chi & Chu, 2021; Jan, 2021a).



Εικόνα 57: Τα μοντέλα RNN, LSTM και GRU

Πηγή: (Phi, 2018)

5.1.1.1 Long Short-Term Memory (LSTM)

Ο αλγόριθμος LSTM είναι ένα επαναλαμβανόμενο τεχνητό νευρωνικό δίκτυο που προτάθηκε το 1997 από τους Hochreiter και Schmidhuber (Chi & Chu, 2021; Jan, 2021a). Πρόκειται για ένα μοντέλο μακροπρόθεσμης μνήμης και χρησιμοποιείται για την επεξεργασία δεδομένων χρονοσειρών, την πρόβλεψη και την ταξινόμηση. Αναπτύχθηκε για να αντιμετωπιστεί το vanishing gradient problem του μοντέλου RNN. Έχει την ικανότητα να μαθαίνει μακροπρόθεσμες εξαρτήσεις για την επίλυση προβλημάτων ακολουθίας και να ελέγχει τη ροή της μνήμης διατηρώντας τις απαραίτητες πληροφορίες και απορρίπτοντας τις άσχετες (Chi & Chu, 2021; Craja κ.ά., 2020; Jan, 2021a; Xiuguo & Shengyong, 2022; Wikipedia, Long short-term memory, 2022).

Το δίκτυο LSTM αποτελείται από ένα κελί μνήμης, που ονομάζεται cell state και από τις τρεις πύλες forget gate, input gate και output gate (Saxena S. , 2021; Singhal, 2020). Επίσης τα $h(t-1)$ και $h(t)$ απεικονίζουν το hidden state της προηγούμενης και της τρέχουσας χρονικής ακολουθίας που είναι γνωστό ως Short term memory ενώ τα $C(t-1)$ και $C(t)$ απεικονίζουν το cell state της προηγούμενης και της τρέχουσας χρονικής ακολουθίας που είναι γνωστό ως Long term memory (βλ. εικόνα 58) (Saxena S. , 2021).

Το cell state απεικονίζεται με μια οριζόντια γραμμή στην κορυφή του διαγράμματος (βλ. εικόνα 58) στην οποία οι πληροφορίες ρέουν αμετάβλητες. Όσον αφορά τις πύλες, το LSTM χωρίζεται σε τρία μέρη όπου το πρώτο μέρος αντιστοιχεί στο forget gate, το δεύτερο στο input gate και το τρίτο στο output gate (βλ. εικόνα 58) (Saxena S. , 2021). Ουσιαστικά, οι πύλες ρυθμίζουν τη ροή των πληροφοριών μέσα και έξω από το cell state, δηλαδή καθορίζουν αν οι πληροφορίες θα διατηρηθούν, θα

λησμονηθούν ή θα ενημερωθούν (Chi & Chu, 2021; Craja κ.ά., 2020; Jan, 2021a; Xiuguo & Shengyong, 2022).

Αναλυτικότερα, το **forget gate** καθορίζει αν οι πληροφορίες που προέρχονται από την προηγούμενη χρονική ακολουθία, πρέπει να απομνημονευτούν ή να λησμονηθούν. Για τον υπολογισμό του forget gate (f_t) λαμβάνονται υπόψη η τρέχουσα είσοδος x_t , το βάρος της εισόδου U_f , η κρυφή κατάσταση $h_{(t-1)}$, ο πίνακας βάρους της κρυφής κατάστασης W_f και η σιγμοειδής συνάρτηση (σ). Για τον υπολογισμό του forget gate ισχύει η εξίσωση (Saxena S. , 2021):

$$f_t = \sigma(x_t \times U_f + h_{(t-1)} \times W_f)$$

Το αποτέλεσμα του f_t θα είναι ίσο με μια τιμή μεταξύ 0 και 1.

Στη συνέχεια, το αποτέλεσμα της εξίσωσης πολλαπλασιάζεται με το $C_{(t-1)}$. Έτσι όταν το αποτέλεσμα του f_t είναι ίσο με το μηδέν, τότε η πληροφορία θα λησμονηθεί, $C_{(t-1)} \times f_t = 0$. Ενώ αν η τιμή είναι ίση με το 1 (ή κοντά στο 1), η πληροφορία θα απομνημονευτεί, $C_{(t-1)} \times f_t = C_{(t-1)}$ (Saxena S. , 2021).

Το **input gate** καθορίζει ποιες πληροφορίες θα ενημερωθούν. Για τον υπολογισμό του input gate, αρχικά ποσοτικοποιείται η σημαντικότητα της νέας πληροφορίας (i_t) και λαμβάνονται υπόψη η είσοδος x_t , το βάρος της εισόδου U_i , η κρυφή κατάσταση $h_{(t-1)}$, ο πίνακας βάρους της κρυφής κατάστασης W_i και η σιγμοειδής συνάρτηση (σ). Για τον υπολογισμό του input gate ισχύει η εξίσωση (Saxena S. , 2021):

$$i_t = \sigma(x_t \times U_i + h_{(t-1)} \times W_i)$$

Το αποτέλεσμα του I_t θα είναι ίσο με μια τιμή μεταξύ 0 και 1.

Επίσης για τις νέες πληροφορίες (N_t) που πρέπει να περάσουν στο cell state, λαμβάνονται υπόψη η είσοδος x_t , το βάρος της εισόδου U_c , η κρυφή κατάσταση $h_{(t-1)}$, ο πίνακας βάρους της κρυφής κατάστασης W_c , η συνάρτηση ενεργοποίησης \tanh και ισχύει η εξίσωση (Saxena S. , 2021):

$$N_t = \tanh(x_t \times U_c + h_{(t-1)} \times W_c)$$

Οι τιμές i_t και N_t των παραπάνω εξισώσεων πολλαπλασιάζονται και προστίθενται στο αποτέλεσμα που προέκυψε στο προηγούμενο βήμα (βλ. forget gate) δίνοντας το αποτέλεσμα του νέου ενημερωμένου cell state (C_t) (Saxena S. , 2021):

$$C_t = f_t \times C_{(t-1)} + i_t \times N_t$$

Έτσι η νέα τιμή του cell state (C_t) μεταβιβάζεται στην επόμενη χρονική ακολουθία (Singhal, 2020).

Το **output gate** καθορίζει την τιμή της κρυφής κατάστασης (H_t). Το αποτέλεσμα του νέου ενημερωμένου cell state (βλ. προηγούμενο βήμα) περνά από τη συνάρτηση \tanh , ενώ οι τιμές της προηγούμενης κρυφής κατάστασης (H_{t-1}) και της τρέχουσας εισόδου (x_t) περνούν από τη σιγμοειδή συνάρτηση και τα αποτελέσματα πολλαπλασιάζονται, ώστε να προκύψει η νέα τιμή του hidden state (h_t) (βλ. εικόνα 58) (Singhal, 2020). Οι υπολογισμοί γίνονται με τις ακόλουθες εξισώσεις (Saxena S., 2021):

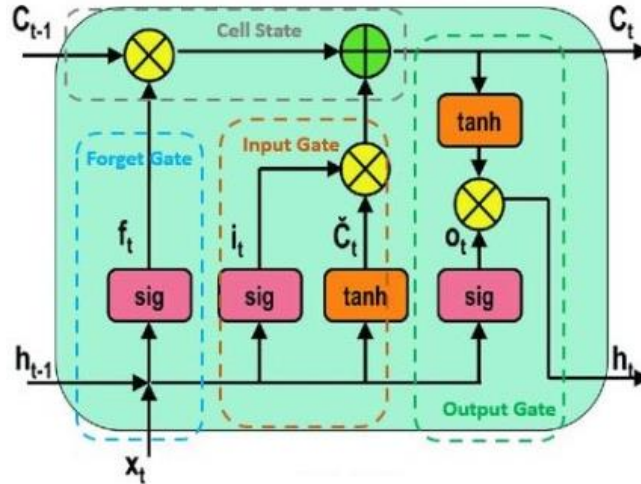
$$o_t = \sigma(x_t \times U_o + h_{(t-1)} \times W_o)$$

όπου x_t η τρέχουσα είσοδος, U_o το βάρος εισόδου, $h_{(t-1)}$ η προηγούμενη κρυφή κατάσταση, W_o ο πίνακας βάρους προηγούμενης κρυφής κατάστασης, σ η σιγμοειδής συνάρτηση ενεργοποίησης

και

$$h_t = o_t \times \tanh(C_t)$$

όπου h_t η τρέχουσα κρυφή κατάσταση, o_t το output gate, \tanh η συνάρτηση ενεργοποίησης και C_t η νέα τιμή cell state.
Έτσι η νέα τιμή του hidden state (h_t) μεταβιβάζεται στην επόμενη χρονική ακολουθία (Singhal, 2020).



Εικόνα 58: Το μοντέλο LSTM

Πηγή: (Singhal, 2020)

Μέχρι σήμερα το μοντέλο LSTM χρησιμοποιείται με επιτυχία στον τομέα της βαθιάς μάθησης σε πολλές εφαρμογές, όπως την αναγνώριση ομιλίας, τη μηχανική μετάφραση, τον έλεγχο ρομπότ, την πρόβλεψη χρονοσειρών, τα βιντεοπαιχνίδια κ.α. (Wikipedia, Long short-term memory, 2022). Είναι ένα μοντέλο εξαιρετικά ακριβές σε μεγάλες χρονικές ακολουθίες, ωστόσο είναι επιρρεπές στην υπερπροσαρμογή και οι απαιτήσεις σε πόρους και χρόνο εκπαίδευσης είναι υψηλές.

5.1.1.2 Gated Recurrent Unit (GRU)

Ο αλγόριθμος **GRU** αποτελεί παραλλαγή του αλγορίθμου LSTM (βλ. εικόνα 57). Η διαφορά του με τον αλγόριθμο LSTM είναι ότι έχει μόνο δύο πύλες, το reset gate και το update gate και δεν διαθέτει memory cell. Το πλεονέκτημά του είναι ότι η μείωση του αριθμού των πυλών, μειώνει τις παραμέτρους εκπαίδευσης και αυξάνει την ταχύτητα εκπαίδευσης (Chi & Chu, 2021; Xiuguo & Shengyong, 2022).

5.1.2 Convolutional Neural Networks (CNN)

Το Convolutional Neural Network (CNN) είναι ένα τεχνητό νευρωνικό δίκτυο (βλ. εικόνα 59) που έχει σχεδιαστεί για την επεξεργασία δεδομένων pixel και χρησιμοποιείται ευρέως για την ανάλυση οπτικών εικόνων. Είναι εμπνευσμένο από τον οπτικό φλοιό των ζώων οργανισμών, όπου κάθε νευρώνας αντιλαμβάνεται ερεθίσματα μιας περιορισμένης περιοχής του οπτικού πεδίου που λέγεται δεκτικό πεδίο. Τα δεκτικά πεδία διαφορετικών νευρώνων επικαλύπτονται, ώστε να καλύπτουν ολόκληρο το οπτικό πεδίο (Wikipedia, Convolutional neural network, 2022).

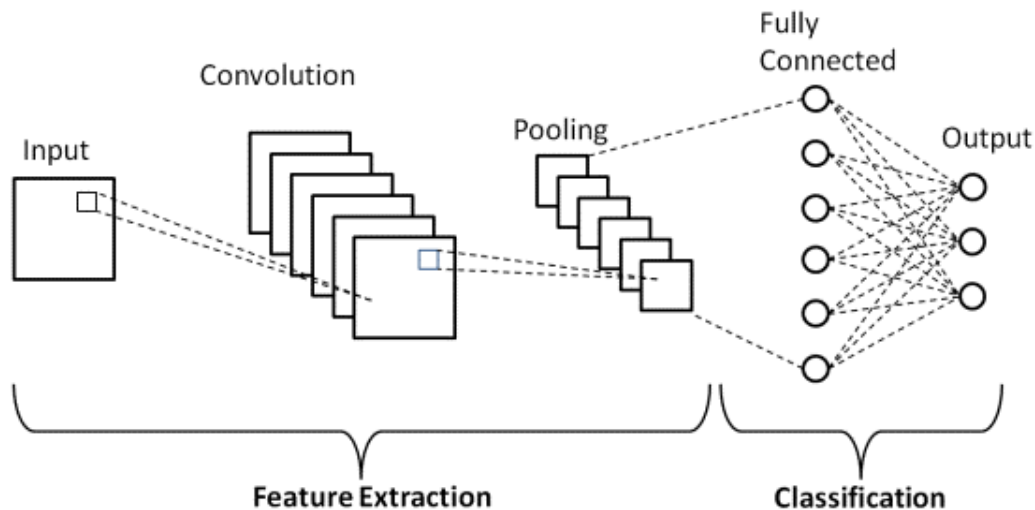
Η αρχιτεκτονική των convolutional neural networks είναι διαφορετική από τα συνήθη νευρωνικά δίκτυα. Το δίκτυο αποτελείται από ένα επίπεδο εισόδου, τα κρυφά επίπεδα και ένα επίπεδο εξόδου. Τα επίπεδα εισόδου και τα κρυφά επίπεδα είναι οργανωμένα στις 3 διαστάσεις πλάτος, ύψος και βάθος ενώ το επίπεδο εξόδου είναι οργανωμένο σε μία διάσταση (freeCodeCamp, 2019). Το CNN αποτελείται από δύο τμήματα. Στο πρώτο τμήμα, γίνεται η εξαγωγή των χαρακτηριστικών και στο δεύτερο τμήμα γίνεται η ταξινόμηση των χαρακτηριστικών (βλ. εικόνα 59).

Αναλυτικότερα, στο πρώτο τμήμα, οι νευρώνες κάθε επιπέδου δεν συνδέονται με όλους τους νευρώνες του επόμενου επιπέδου, αλλά συνδέονται σε περιορισμένη περιοχή. Στο τμήμα αυτό πραγματοποιούνται μια σειρά από συνελίξεις (convolution) και ομαδοποιήσεις (pooling). Κατά τη διαδικασία της συνέλιξης, χρησιμοποιείται ένα φίλτρο που ονομάζεται Kernel. Το φίλτρο ολισθαίνει στα δεδομένα εισόδου, ώστε να προκύψει ο χάρτης χαρακτηριστικών. Θα πρέπει να σημειωθεί ότι στο στάδιο αυτό πραγματοποιούνται πολλές συνελίξεις στα δεδομένα εισόδου και κάθε συνέλιξη χρησιμοποιεί διαφορετικό φίλτρο με αποτέλεσμα να δημιουργούνται διαφορετικοί χάρτες χαρακτηριστικών οι οποίοι τοποθετούνται μαζί στην έξοδο του επιπέδου συνέλιξης (convolution layer) (freeCodeCamp, 2019).

Μετά από κάθε επίπεδο συνέλιξης, τοποθετείται ένα επίπεδο pooling, προκειμένου να μειωθεί διάσταση. Η μείωση διάστασης αφορά το ύψος και το πλάτος, ενώ το βάθος παραμένει αμετάβλητο (Dertat, 2017). Κατά τη διαδικασία αυτή, μειώνεται ο αριθμός των παραμέτρων και των υπολογισμών του δικτύου και κατά συνέπεια μειώνεται ο χρόνος εκπαίδευσης και ο κίνδυνος υπερπροσαρμογής του δικτύου (freeCodeCamp, 2019).

Στο δεύτερο τμήμα του CNN, τα δεδομένα μετατρέπονται από δεδομένα τριών διαστάσεων σε δεδομένα μίας διάστασης και τροφοδοτούνται σε πλήρως συνδεδεμένα επίπεδα για την ταξινόμηση των χαρακτηριστικών, ώστε να προκύψει το αποτέλεσμα (freeCodeCamp, 2019).

Οι κυριότερες υπερπαραμέτροι ενός δικτύου CNN είναι τέσσερες (α) το μέγεθος του φίλτρου που είναι ίσο με το δεκτικό πεδίο και (β) ο αριθμός των φίλτρων που θα χρησιμοποιηθούν, (γ) το stride που αφορά το μέγεθος του βήματος με το οποίο ολισθαίνει το φίλτρο στα δεδομένα εισόδου (αν το μέγεθος του βήματος είναι 1 τότε το φίλτρο ολισθαίνει κατά ένα pixel τη φορά) και (δ) το padding που εφαρμόζεται προκειμένου να αποφευχθεί η συρρίκνωση του χάρτη χαρακτηριστικών δεδομένου ότι το μέγεθός του είναι πάντα μικρότερο από την είσοδο. Έτσι τα δεδομένα εισόδου περιβάλλονται από pixel μηδενικής τιμής, ώστε το μέγεθος του χάρτη χαρακτηριστικών να διατηρηθεί και να μην συρρικνωθεί (freeCodeCamp, 2019).



Εικόνα 59: Το μοντέλο CNN

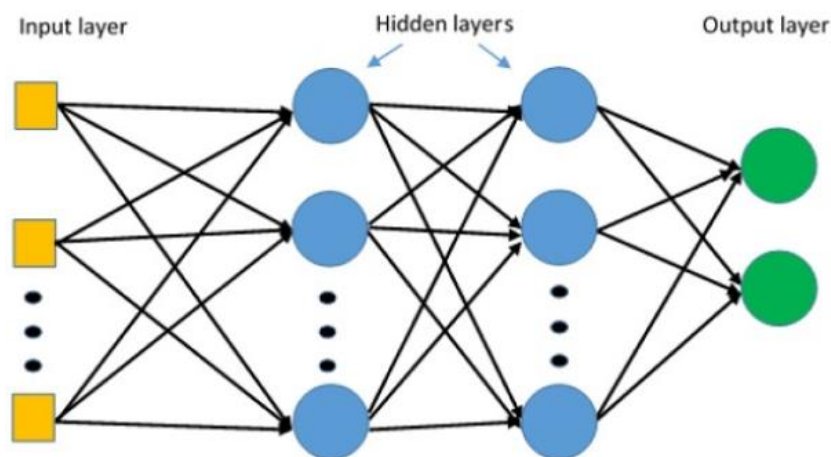
Πηγή: (Gurucharan, 2022)

Τα κυριότερα πλεονεκτήματα του δικτύου CNN είναι ότι είναι υψηλής ακρίβειας και έχει την ικανότητα να εντοπίζει τα σημαντικά χαρακτηριστικά μόνο του χωρίς την ανθρώπινη παρέμβαση.

Μέχρι σήμερα, εφαρμόζεται με επιτυχία σε πλήθος εφαρμογών, όπως σε εφαρμογές οπτικής αναγνώρισης προσώπων και αντικειμένων, ταξινόμησης εικόνων, επεξεργασίας φυσικής γλώσσας κ.α. (Wikipedia, Convolutional neural network, 2022).

5.1.3 Multi-Layer Perceptron (MLP)

Το μοντέλο Multi-Layer Perceptron (MLP) είναι ένα feed forward τεχνικό νευρωνικό δίκτυο (βλ. εικόνα 60) που αποτελείται από τρία ή περισσότερα επίπεδα πλήρως συνδεδεμένα (ένα επίπεδο εισόδου – ένα ή περισσότερα κρυφά επίπεδα – ένα επίπεδο εξόδου). Είναι το βασικότερο deep neural network και για την εκπαίδευσή του χρησιμοποιείται ο αλγόριθμος backpropagation (Κύρκος, 2015). Είναι κατάλληλο για την ταξινόμηση προβλημάτων μη γραμμικών σχέσεων και χρησιμοποιείται σε εφαρμογές αναγνώρισης ομιλίας, αναγνώρισης εικόνας και αυτόματης μετάφρασης (Wikipedia, Multilayer perceptron, 2022).



Εικόνα 60: Το μοντέλο MLP

Πηγή: (Free time learning, χ.χ.)

5.2 Ο σκοπός της πρόβλεψης των ερευνητικών άρθρων

Λαμβάνοντας υπόψη τον σκοπό της πρόβλεψης κάθε ερευνητικού άρθρου, τρεις έρευνες επικεντρώθηκαν στην πρόβλεψη απάτης, δύο έρευνες στην πρόβλεψη οικονομικής δυσπραγίας, έξι στην πρόβλεψη της χρεοκοπίας, δύο ασχολήθηκαν με την πρόβλεψη της εταιρικής αποτυχίας και, τέλος, δύο έρευνες με την πρόβλεψη της συνεχιζόμενης δραστηριότητας των επιχειρήσεων (βλ. πίνακα 1).

5.2.1 Πρόβλεψη απάτης

Η έρευνα των Craja κ.ά. (2020) ανέπτυξε το μοντέλο **Hierarchical Attention Network (HAN)** το οποίο συνδυάζει τις πληροφορίες που προέρχονται από τα οικονομικά δεδομένα και τις δηλώσεις του Δ.Σ (MD&A) των ετήσιων οικονομικών καταστάσεων. Η αρχιτεκτονική του μοντέλου ενσωματώνει δύο μηχανισμούς προσοχής σε επίπεδο λέξης και πρότασης για την εξαγωγή χαρακτηριστικών κειμένου από την ενότητα MD&A που του επιτρέπει να αναγνωρίζει πότε μια λέξη είναι σημαντική μέσα σε μια πρόταση και πότε όχι (βλ. εικόνα 6). Η αρχιτεκτονική του δικτύου, δηλαδή επιτρέπει τη διαφοροποίηση του περιεχομένου ως προς τη σημασία του κατά τη διαδικασία αναπαράστασης του εγγράφου. Τα πειράματα έδειξαν ότι πρόκειται για μια ολοκληρωμένη προσέγγιση μοντέλου DL, διότι σημείωσε την καλύτερη απόδοση με AUC 92,64% (βλ. πίνακα 2) και είναι ικανό και κατάλληλο να εντοπίσει σωστά τις περιπτώσεις απάτης, ιδιαίτερα με τη χρήση του συνδυασμού δεδομένων FIN+TXT. Επίσης οι δείκτες “red flags” (βλ. εικόνα 8) είναι χρήσιμοι για την ερμηνεία του κειμένου και παρέχουν υποστήριξη στις διαδικασίες λήψης αποφάσεων όλων των ενδιαφερόμενων μερών.

Επίσης οι Xiuguo και Shengyong (2022) ασχολήθηκαν με την ανάπτυξη ενός συστήματος ανίχνευσης χρηματοοικονομικής απάτης με τη χρήση μοντέλων deep learning συνδυάζοντας αριθμητικά δεδομένα και δεδομένα κειμένου από ετήσιες εκθέσεις κινέζικων εισηγμένων εταιρειών. Η έρευνά τους έχει ομοιότητες με την έρευνα των Craja κ.ά. (2020), αφού εξέτασαν την απόδοση των μοντέλων εισάγοντας διαφορετικούς τύπους δεδομένων κατά τα πειράματα. Η διαφορά τους είναι ότι εφαρμόσαν ένα κινέζικο πρότυπο εξόρυξης κειμένου, ώστε να μην χαθούν πολύτιμες πληροφορίες που θα επηρέαζαν την απόδοση των μοντέλων. Στα πειράματα που χρησιμοποιήθηκαν μόνο οικονομικά δεδομένα, τα αποτελέσματα έδειξαν ότι η επιλογή χαρακτηριστικών από το αρχικό σύνολο δεδομένων δεν επηρεάζει αρνητικά την ικανότητα πρόβλεψης των μοντέλων (βλ. πίνακες 7 & 8). Επίσης στα πειράματα όπου έγινε χρήση μόνο δεδομένων κειμένου (βλ. πίνακες 9 & 10), η απόδοση των μοντέλων DL είναι καλύτερη σε σύγκριση με τα παραδοσιακά μοντέλα, όμως η απόδοση όλων των μοντέλων είναι χειρότερη σε σύγκριση με την απόδοση που είχαν με τη χρήση μόνο οικονομικών δεδομένων. Αυτό σημαίνει ότι η πρόβλεψη απάτης με τη χρήση μόνο δεδομένων κειμένου δεν έχει καλή απόδοση και είναι ανεξάρτητη από το μοντέλο. Ενώ στα πειράματα όπου χρησιμοποιήθηκαν τόσο οικονομικά όσο και δεδομένα κειμένου (βλ. πίνακες 11 & 12), η απόδοση είναι βελτιωμένη και μάλιστα τα μοντέλα GRU και LSTM, που βασίζονται στο RNN, έχουν τις καλύτερες αποδόσεις που σημαίνει ότι έχουν την ικανότητα να ενσωματώνουν και να εξάγουν αποτελέσματα από διαφοροποιημένα δεδομένα. Επομένως ο συνδυασμός διαφοροποιημένων δεδομένων ενισχύει την απόδοση των μοντέλων DL και μπορεί να υποστηρίξει τις διαδικασίες αποφάσεων όλων των ενδιαφερόμενων μερών. Επίσης θα πρέπει να αναφερθεί ότι κατά τα πειράματα, χρησιμοποιήσαν κείμενα 600 και 1000 λέξεων και διαπιστώθηκε ότι η απόδοση των μοντέλων βελτιώνεται με τη χρήση κειμένων 1000 λέξεων. Αυτό συμβαίνει διότι, όταν ένα μέρος του κειμένου αφαιρείται, είναι πιθανό να χάνονται σημαντικές πληροφορίες. Βέβαια, αν το μήκος του κειμένου αυξηθεί σημαντικά μπορεί να επηρεαστεί αρνητικά η απόδοση του μοντέλου, λόγω του κινδύνου της υπερβολικής προσαρμογής.

Τέλος, ο Jan (2021a) ασχολήθηκε με την ανάπτυξη ενός μοντέλου deep learning με τη χρήση των αλγορίθμων RNN και LSTM. Σκοπός της έρευνας ήταν η μείωση των ζημιών που μπορεί να προκύψουν από μια ενδεχόμενη απάτη και η διατήρηση της βιώσιμης ανάπτυξης των κεφαλαιαγορών. Τα πειράματα έδειξαν ότι τα μοντέλα RNN και LSTM είναι ικανά να επεξεργαστούν μεγάλα σύνολα δεδομένων και επιτυγχάνουν ακρίβεια πάνω από 85% (βλ. πίνακα 17). Μεταξύ των δύο μοντέλων, το μοντέλο LSTM υπερτερεί σε όλα τα μέτρα απόδοσης και μάλιστα έχει τα χαμηλότερα ποσοστά σφάλματος τύπου I και τύπου II. Το πλεονέκτημα του μοντέλου LSTM είναι ότι αντιμετωπίζει τα μειονεκτήματα του αλγορίθμου RNN που αφορούν την αδυναμία διατήρησης μακροπρόθεσμης μνήμης και της εξαφάνισης κλίσης που αποτρέπει την ενημέρωση βαρών των νευρωνικών δικτύων σε μικρότερα επίπεδα κατά το backpropagation.

5.2.2 Πρόβλεψη οικονομικής δυσπραγίας

Κατά την έρευνα τους, οι Li κ.ά. (2021) ασχολήθηκαν με τη δημιουργία ενός κινέζικου λεξικού συναισθήματος του χρηματοοικονομικού τομέα (Chinese financial domain sentiment lexicon – CFDSL) και την εφαρμογή του για την πρόβλεψη της οικονομικής δυσπραγίας. Η δημιουργία του κινέζικου λεξικού ενσωματώνει δύο

προσεγγίσεις (α) την dictionary based και (β) την corpus-based προσέγγιση. Κατά την έρευνα εξετάστηκαν τα μοντέλα DNN, MA-DNN και Bi-LSTM και τα αποτελέσματα έδειξαν την ανωτερότητα του μοντέλου MA-DNN (βλ. πίνακα 19) το οποίο και χρησιμοποιήθηκε για τη δημιουργία του λεξικού. Έπειτα ακολούθησε η ανάλυση συναισθήματος κατά την οποία υπολογίστηκαν τα sentiment features του λεξικού CFDSL και τα πειράματα έδειξαν ότι τα sentiment features (βλ. πίνακα 23) έχουν την ικανότητα να αποκαλύπτουν την οικονομική δυσπραγία σε προγενέστερο χρόνο (4 και 3 έτη πριν), παρά την απουσία των οικονομικών δεδομένων παρέχοντας έγκαιρη και χρήσιμη ενημέρωση στα ενδιαφερόμενα μέρη. Μάλιστα οι λέξεις που περιλαμβάνονται στο λεξικό έχουν άμεση σχέση με τον χρηματοοικονομικό τομέα και μπορεί να χρησιμοποιηθεί για την πρόβλεψη και άλλων οικονομικών ζητημάτων, όπως την πρόβλεψη χρεοκοπίας και την πρόβλεψη απάτης.

Επίσης η έρευνα των Matin κ.ά. (2019) ανέπτυξε **τρία μοντέλα deep learning**, η αρχιτεκτονική των οποίων αποτελείται από ένα convolutional neural network και ένα recurrent neural network (συγκεκριμένα το LSTM) με έναν μηχανισμό προσοχής που του επιτρέπει να εστιάζει στα λιγότερα και πιο σημαντικά τμήματα του κειμένου. Στο ένα μοντέλο χρησιμοποιήθηκαν οι εκθέσεις των ελεγκτών, στο δεύτερο οι δηλώσεις των διευθυντικών στελεχών και στο τρίτο ο συνδυασμός των δύο. Στη συνέχεια, το κείμενο κάθε μοντέλου συνενώθηκε με τις αριθμητικές μεταβλητές, ώστε να προκύψει το αποτέλεσμα της πρόβλεψης (βλ. εικόνα 25). Σύμφωνα με τα πειράματα, η απόδοση των μοντέλων βελτιώθηκε με τη χρήση των δεδομένων κειμένου και ιδιαίτερα εκείνου όπου χρησιμοποιήθηκαν οι εκθέσεις των ελεγκτών (βλ. εικόνα 26 και πίνακα 26). Μάλιστα διαπιστώθηκε ότι όσο καλύτερης ποιότητας είναι μια έκθεση ελέγχου τόσο καλύτερη και η απόδοση του μοντέλου (βλ. πίνακα 27). Αυτό οφείλεται στο γεγονός ότι οι εκθέσεις των ελεγκτών είναι αντικειμενικές και αποτυπώνουν την οικονομική κατάσταση μιας επιχείρησης αμερόληπτα σε αντίθεση με τις δηλώσεις του ΔΣ (ενότητα MD&A) που δίνουν πάντα μια ευνοϊκότερη εικόνα της επιχείρησης. Τα παραπάνω ευρήματα υποδεικνύουν ότι είναι σημαντικότερη η συμπερίληψη αδόμητων δεδομένων στην πρόβλεψη και μάλιστα δεδομένων από περισσότερες πηγές (όπως τις εκθέσεις των ελεγκτών), παρά η δημιουργία πολύπλοκων μοντέλων στα ίδια δεδομένα.

5.2.3 Πρόβλεψη χρεοκοπίας

Η έρευνα του Hosaka (2019) έχει ως στόχο την εφαρμογή του μοντέλου **Convolutional Neural Network (CNN)** χρησιμοποιώντας ως δεδομένα αριθμητικούς οικονομικούς δείκτες μετασχηματισμένους σε εικόνες για την πρόβλεψη χρεοκοπίας η οποία στις περισσότερες περιπτώσεις αντιμετωπίζεται ως πρόβλημα ταξινόμησης. Παρόλο που τα μοντέλα CNN εφαρμόζονται με επιτυχία σε εφαρμογές αναγνώρισης εικόνας, οι εφαρμογές τους σε οικονομικές αναλύσεις είναι περιορισμένη και αυτό διότι οι εφαρμογές τους σε εικόνες είναι καταλληλότερη σε σύγκριση με την εφαρμογή τους σε αριθμητικά δεδομένα. Γι' αυτό και στην παρούσα έρευνα τα αριθμητικά δεδομένα μετασχηματίστηκαν σε εικόνες. Σύμφωνα με τα πειράματα, η ακρίβεια της πρόβλεψης του μοντέλου βελτιώνεται με τη μέθοδο correlated και ιδιαίτερα όσο αυξάνεται ο αριθμός των layers του μοντέλου (βλ. πίνακα 29). Μάλιστα κατά τη σύγκριση της απόδοσής του με άλλα μοντέλα διαπιστώθηκε η υπεροχή του (βλ. εικόνες 30 & 31). Το πλεονέκτημα της προτεινόμενης μεθόδου είναι ότι μπορεί να εφαρμοστεί σε γενικά αριθμητικά δεδομένα για την πρόβλεψη διαφόρων οικονομικών ζητημάτων και όχι μόνο για την πρόβλεψη χρεοκοπίας. Ενώ το μειονέκτημά της είναι ότι είναι

δύσκολο να εντοπιστούν οι χρηματοοικονομικοί δείκτες που επηρεάζουν την πρόβλεψη χρεοκοπίας, ώστε να διερευνηθούν τα αίτια που την προκαλούν.

Οι Mai κ.ά. (2019) σύγκριναν την απόδοση μοντέλων DL και παραδοσιακών με τη χρήση αριθμητικών δεδομένων, δεδομένων κειμένου και του συνδυασμού τους και εξέτασαν ποιες αρχιτεκτονικές δικτύων βελτιώνουν την απόδοση της πρόβλεψης συνδυάζοντας διαφορετικά neural network layers. Τα αποτελέσματα της έρευνας έδειξαν ότι, με τη χρήση μόνο αριθμητικών δεδομένων (βλ. πίνακα 35), τα μοντέλα SVM και RF είναι εξίσου ικανά για την πρόβλεψη και δεν υπάρχει λόγος χρήσης των μοντέλων DL. Αντίθετα, με τη χρήση μόνο δεδομένων κειμένου (βλ. πίνακα 34), τα μοντέλα deep learning είναι καταλληλότερα στην πρόβλεψη χρεοκοπίας σε σύγκριση με τα παραδοσιακά. Μάλιστα μεταξύ των δύο μοντέλων DL που χρησιμοποιήθηκαν, το απλούστερο μοντέλο **average embedding model** (DL-Embedding) είναι αποτελεσματικότερο σε σύγκριση με το **convolutional neural network** (CNN). Όμως, όταν ο χρονικός ορίζοντας της πρόβλεψης αυξάνει τότε η προγνωστική ικανότητα των δεδομένων κειμένου είναι συγκρίσιμη με αυτή των αριθμητικών (βλ. εικόνα 33). Επίσης η ακρίβεια της πρόβλεψης του εξεταζόμενου μοντέλου DL αυξάνεται όταν τα δεδομένα κειμένου συνδυάζονται με αριθμητικά δεδομένα (βλ. πίνακα 36). Επομένως ο συνδυασμός των δύο τύπων δεδομένων μπορεί να προσφέρει αξία στην πρόβλεψη.

Οι Alexandropoulos κ.ά. (2019) μελέτησαν την αποτελεσματικότητα του μοντέλου **Deep Dense Multilayer Perceptron** (DDMP). Η επιλογή του αριθμού των κρυφών επιπέδων και του αριθμού των νευρώνων στα κρυφά επίπεδα είναι σημαντική προκειμένου να επιτευχθεί η γενίκευση του δικτύου, να αποφευχθεί η υπερπροσαρμογή και ο μεγάλος χρόνος εκπαίδευσης. Στη μελέτη, η αρχιτεκτονική του μοντέλου που επιλέχθηκε ήταν δύο κρυφά επίπεδα όπου στο πρώτο κρυφό επίπεδο, ο αριθμός των νευρώνων καθορίστηκε στα 2/3 του αριθμού των χαρακτηριστικών εισόδου, ενώ στο δεύτερο κρυφό επίπεδο, στο 1/3 του αριθμού των χαρακτηριστικών εισόδου. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη αρχιτεκτονική επιτυγχάνει τα καλύτερα αποτελέσματα σε σύγκριση με τα μοντέλα αναφοράς (βλ. πίνακα 39).

Οι Smiti και Soui (2020) ασχολήθηκαν με την ανάπτυξη του μοντέλου deep learning **BSM-SAES** το οποίο συνδυάζει την τεχνική υπερδειγματοληψίας Borderline Synthetic Minority (BSM) με τους κατηγοριοποιητές Stacked AutoEncoder (SAE) και Softmax. Στόχος της έρευνας ήταν η ανάπτυξη ενός μοντέλου DL που να περιλαμβάνει τη διαδικασία εξαγωγής χαρακτηριστικών με σκοπό τη μείωση του χρόνου ταξινόμησης και την απλοποίηση των διαδικασιών λήψης αποφάσεων. Το μοντέλο **συγκρίθηκε** με άλλα μοντέλα machine learning και διενεργήθηκαν δύο πειράματα όπου στο πρώτο δεν εφαρμόστηκε η τεχνική της υπερδειγματοληψίας BSM σε κανένα μοντέλο ενώ στο δεύτερο εφαρμόστηκε σε όλα (βλ. πίνακα 41). Σύμφωνα με τα αποτελέσματα, η τεχνική BSM βελτίωσε τα αποτελέσματα όλων των μοντέλων, ενώ το προτεινόμενο είχε την υψηλότερη απόδοση η οποία βελτιώθηκε περισσότερο με την εφαρμογή της τεχνικής BSB. Ο λόγος είναι ότι το προτεινόμενο μοντέλο είναι σε θέση να εξάγει αυτόματα τα χρήσιμα χαρακτηριστικά κατά τη διάρκεια της διαδικασίας εκπαίδευσης ενώ για τα άλλα μοντέλα machine learning η εξαγωγή των χαρακτηριστικών γίνεται ανεξάρτητα από τη φάση εκπαίδευσης. Όμως, το προτεινόμενο μοντέλο είχε τη χειρότερη απόδοση στο training time (βλ. πίνακα 43) που οφείλεται στο χρόνο που αφιερώνει το μοντέλο για τη μείωση της διάστασης των χαρακτηριστικών.

Η έρευνα των Jang κ.ά. (2021) είχε ως στόχο να προσδιορίσει την επίδραση που έχουν οι μεταβλητές εισόδου στο μοντέλο πρόβλεψης χρεοκοπίας **LSTM-RNN** με τη χρήση της τιμής Shapley προκειμένου να μετρηθεί η επίδραση των μεταβλητών εισόδου και να εντοπιστεί η μεταβλητή εκείνη που επηρεάζει περισσότερο την

ακρίβεια της πρόβλεψης. Κατά την έρευνα, χρησιμοποίησαν το μοντέλο Long short Term memory - Recurrent Neural Network (LSTM-RNN) το οποίο αναπτύχθηκε σε προηγούμενη μελέτη (Jang κ.ά., 2019b, Jang κ.ά., 2020 όπως αναφέρεται στο Jang κ.ά., 2021) και το εφάρμοσαν για την πρόβλεψη της πιθανότητα χρεοκοπίας πριν από 1, 2 και 3 χρόνια. Με τα πειράματα διαπιστώθηκε ότι η μεταβλητή της κατασκευαστικής αγοράς House starts (HS) καθώς και οι λογιστικές μεταβλητές Current ratio (CR) και Debt ratio (DR) επηρέασαν σημαντικά την ακρίβεια της πρόβλεψης (βλ. πίνακα 46). Επίσης, κατά τη διάρκεια των ετών, η επίδραση των μακροοικονομικών μεταβλητών Consumer per index (CPI), Gross domestic product (GDP) και Federal funds rate (FFR) αυξάνεται, ενώ αντίθετα της λογιστικής μεταβλητής Return on asset (ROA) μειώνεται. Επίσης παρατηρήθηκε ότι επιτυγχάνεται παρόμοια ακρίβεια πρόβλεψης και με την επιλογή λιγότερων μεταβλητών εισόδου (βλ. εικόνα 40). Η μεταβλητή HS αφορά τον αριθμό των νέων κατασκευαστικών έργων που έχουν ξεκινήσει σε μια περίοδο και θεωρείται πολύ σημαντική. Επίσης δεδομένου ότι ο κατασκευαστικός κλάδος είναι ευαίσθητος σε μακροοικονομικές αλλαγές, οι διακυμάνσεις των μακροοικονομικών μεταβλητών επηρεάζουν το επενδυτικό κλίμα και τη ζήτηση των κατασκευαστικών έργων ενώ οι λογιστικές μεταβλητές είναι χρήσιμες για την αξιολόγηση των οικονομικών δυνατοτήτων μιας επιχείρησης. Η μεταβλητή ROA αφορά την αποτελεσματικότητα χρήσης των περιουσιακών στοιχείων μιας εταιρείας για τη δημιουργία εισοδήματος. Επομένως όσο υψηλότερη είναι η μεταβλητή, τόσο χαμηλότερος είναι ο κίνδυνος χρεοκοπίας. Επίσης όσο χαμηλότερη η μεταβλητή CR και υψηλότερη η μεταβλητή DR, τόσο μεγαλύτερος ο κίνδυνος χρεοκοπίας. Επομένως οι παραπάνω μεταβλητές είναι χρήσιμες και μπορούν να χρησιμοποιηθούν για την πρόβλεψη της χρεοκοπίας. Ο εντοπισμός των μεταβλητών εισόδου που επηρεάζουν το αποτέλεσμα της πρόβλεψης βοηθάει στην κατανόηση των εσωτερικών διαδικασιών του μοντέλου, στην επιλογή των κατάλληλων μεταβλητών και στην παροχή πληροφοριών για τον τρόπο βελτίωσης των μοντέλων. Επίσης αυτές οι μεταβλητές μπορούν να χρησιμοποιηθούν ως δείκτες διαχείρισης κινδύνου ενώ η επιλογή μόνο των σημαντικών μεταβλητών προσφέρει σημαντικά οφέλη, διότι βοηθά στη μείωση του χρόνου εκπαίδευσης και χρήσης του μοντέλου, τη μείωση των απαιτήσεων μέτρησης και αποθήκευσης αλλά και τη βελτίωση της απόδοσής του.

Οι Vochozka κ.ά. (2020) ασχολήθηκαν με την ανάπτυξη ενός μοντέλου τεχνητού νευρωνικού δικτύου (NN) με ένα τουλάχιστον επίπεδο LSTM με στόχο να προσδιοριστεί η καταλληλότητα του μοντέλου για την πρόβλεψη της πιθανότητας χρεοκοπίας των επιχειρήσεων. Τα αποτελέσματα της έρευνας (βλ. πίνακα 50 και εικόνες 42 & 43) έδειξαν ότι το προτεινόμενο μοντέλο μπορεί να εφαρμοστεί για την πρόβλεψη χρεοκοπίας, να εκπαιδευτεί σε διαφορετικά σύνολα δεδομένων, να εφαρμοστεί σε πλήθος εταιρειών και να χρησιμοποιηθεί από χρηματοπιστωτικά ιδρύματα, επενδυτές και ελεγκτές για την αξιολόγηση της οικονομικής κατάστασης των επιχειρήσεων. Όμως, η πρακτική εφαρμογή του είναι δύσκολη, διότι, λόγω της πολύπλοκης δομής του, είναι δύσκολο να υπολογιστεί εκ νέου ή να προγραμματιστεί σε διαφορετικό περιβάλλον εκτός από το λογισμικό Wolfram's Mathematica. Επομένως παραμένει περίπλοκο για όσους δεν διαθέτουν γνώσεις στις τεχνολογίες πληροφοριών και επικοινωνίας.

5.2.4 Πρόβλεψη οικονομικής αποτυχίας

Οι Alam κ.ά. (2021) ερεύνησαν το μοντέλο **Deep Grassmannian Network** (GrNet) το οποίο χρησιμοποιεί δομή panel data. Πρόκειται για μία πλήρως

συνδεδεμένη αρχιτεκτονική deep neural network που χρησιμοποιεί Grassmannian-valued data για την καταγραφή της δομής των panel data. Είναι ένα καινοτόμο μοντέλο και μπορεί να χρησιμοποιηθεί και σε άλλες εφαρμογές ταξινόμησης με δομή panel. Έχει την ικανότητα να καταγράφει τις χρονικές εξαρτήσεις στο σύνολο δεδομένων σε αντίθεση με τα παραδοσιακά μοντέλα μηχανικής μάθησης (π.χ. gradient boosting machines, random forests) και επιπλέον είναι ικανό να χειρίζεται ένα μεγάλο δείγμα εταιρειών και ένα μεγάλο αριθμό μεταβλητών. Κατά τα πειράματα, έγινε σύγκριση του προτεινόμενου μοντέλου με το discrete-time hazard model (βλ. πίνακα 54) το οποίο έχει επίσης σχεδιαστεί για την ανάλυση δεδομένων πάνελ και είναι ικανό να εκμεταλλευτεί τις χρονικές εξαρτήσεις. Τα αποτελέσματα έδειξαν ότι το προτεινόμενο μοντέλο DL έχει accuracy πάνω από 90% ενώ υπερτερεί στο sensitivity που αφορά την ακρίβεια ταξινόμησης σφάλματος I που θεωρείται πολύ σημαντική.

Οι Aljawazneh κ.ά. (2021) εξέτασαν τρία προηγμένα μοντέλα Deep Learning και συγκεκριμένα τα μοντέλα (α) **Deep Belief Network** (DBN) που είναι ένα stochastic neural network, (β) **Multilayer Perceptron με 6 Layers** (MLP-6L) που είναι ένα feed-forward neural network και (γ) **Long-Short Term Memory** (LSTM) που είναι ένα recurrent neural network. Η απόδοση των μοντέλων συγκρίθηκε και με τα μοντέλα (α) Random Forest (RF), Support Vector Machine (SVM) και K-Nearest Neighbor (KNN) που είναι μέθοδοι bagging based ensemble και με τα μοντέλα (β) Adaptive Boosting (AdaBoost) και EXtreme Gradient Boosting (XGBoost) που είναι μέθοδοι boosting based ensemble. Ο σκοπός της μελέτης ήταν να συγκριθούν μοντέλα διαφορετικών τύπων νευρωνικών δικτύων, ώστε να διερευνηθεί ποιο μοντέλο είναι το καλύτερο για την πρόβλεψη της οικονομικής αποτυχίας. Κατά την έρευνα, χρησιμοποιήθηκαν τρία εξαιρετικά ανισόρροπα σύνολα δεδομένων που διέφεραν μεταξύ τους ως προς το επίπεδο πολυπλοκότητας και ποικιλομορφίας και για τα οποία εφαρμόστηκαν οκτώ τεχνικές εξισορρόπησης τριών διαφορετικών μεθόδων (Oversampling, Hybrid Oversampling-Undersampling και Clustering-based balancing) με σκοπό να διερευνηθεί ποια είναι η καλύτερη για τη βελτίωση της απόδοσης των μοντέλων. Μετά από πειράματα, διαπιστώθηκε ότι η τεχνική εξισορρόπησης SMOTE-ENN είχε την καλύτερη απόδοση στα περισσότερα μοντέλα ανεξαρτήτως της πολυπλοκότητας των δεδομένων (βλ. πίνακα 59). Επίσης διαπιστώθηκε ότι το feed-forward neural network MLP-6L σε συνδυασμό με την τεχνική εξισορρόπησης SMOTE-ENN (Hybrid Oversampling-Undersampling) είχε την καλύτερη απόδοση (βλ. πίνακα 59). Επίσης έγινε σύγκριση του μοντέλου και με αλγόριθμους προηγούμενων ερευνών και τη χρήση των ίδιων συνόλων δεδομένων όπου διαπιστώθηκε και πάλι η υπεροχή του.

5.2.5 Πρόβλεψη μη συνεχιζόμενης δραστηριότητας

Η έρευνα του Jan (2021b) χρησιμοποίησε τις μεθόδους **Deep Neural Network** (DNN) και **Recurrent Neural Network** (RNN) για την ανάπτυξη των μοντέλων. Σκοπός της έρευνας ήταν να βοηθήσει τους ελεγκτές και τους ορκωτούς λογιστές να κάνουν ορθές αξιολογήσεις και να εκδίδουν αξιόπιστες εκθέσεις ελέγχου. Τα μοντέλα κατασκευάστηκαν με δύο τρόπους προκειμένου να εξεταστεί πιο μοντέλο επιτυγχάνει τη βέλτιστη απόδοση πρόβλεψης. Στη μία περίπτωση χρησιμοποιήθηκαν όλες οι επιλεγμένες μεταβλητές ενώ στη δεύτερη μόνο οι σημαντικότερες αυτών με τη βοήθεια του αλγορίθμου classification and regression tree (CART). Ο αλγόριθμος CART έχει υψηλή ικανότητα ταξινόμησης, οι κανόνες ταξινόμησης είναι απλοί, οι discrete και continuous τιμές αντιμετωπίζονται χωριστά και δημιουργεί μόνο δυαδικά δέντρα. Τα πειράματα έδειξαν ότι η ακρίβεια όλων των μοντέλων ήταν πάνω από 88% και μάλιστα

το μοντέλο RNN είχε καλύτερη απόδοση σε σύγκριση με το DNN. Όμως η απόδοση των μοντέλων βελτιώθηκε περισσότερο με τη χρήση του αλγορίθμου CART (βλ. πίνακα 64).

Η έρευνα των Chi και Chu (2021) ασχολήθηκε με την ανάπτυξη μοντέλων πρόβλεψης με τη χρήση των αλγορίθμων deep learning **Long Short-Term Memory (LSTM)** and **Gated Recurrent Unit (GRU)**. Στόχος της έρευνας είναι να βοηθήσει τους ορκωτούς λογιστές (CPA) και τους ελεγκτές στην έκδοση ακριβέστερων και ορθότερων εκθέσεων ελέγχου. Με τα πειράματα διαπιστώθηκε ότι τα μοντέλα LSTM και GRU έχουν την ικανότητα να επεξεργάζονται μεγάλα σύνολα δεδομένων με ταχύτητα και επίσης πέτυχαν πολύ υψηλή ακρίβεια ταξινόμησης και χαμηλά ποσοστά σφάλματος τύπου I και II (βλ. πίνακα 67). Αυτό σημαίνει ότι η χρήση τους μπορεί να μειώσει τους κινδύνους και τα κόστη που προέρχονται από λάθη και παραλείψεις των διαδικασιών ελέγχου, παρέχοντας σημαντική υποστήριξη όχι μόνο στους ορκωτούς λογιστές και ελεγκτές αλλά και σε αναλυτές, συμβούλους επιχειρήσεων, οργανισμούς αξιολόγησης πιστοληπτικής ικανότητας και στη διοίκηση των επιχειρήσεων για θέματα εταιρικής βιωσιμότητας και διαχείρισης κινδύνου. Επομένως θεωρούνται αποτελεσματικά και κατάλληλα μοντέλα πρόβλεψης.

5.3 Τα δεδομένα

Μέχρι σήμερα οι περισσότερες έρευνες περιορίζονταν στη χρήση οικονομικών δεδομένων και μόνο λίγες είναι εκείνες που έχουν ασχοληθεί με τη χρήση και άλλων τύπων δεδομένων. Τα δεκαπέντε ερευνητικά άρθρα χρησιμοποίησαν διάφορους τύπους δεδομένων. Κάποιες χρησιμοποίησαν μόνο οικονομικά δεδομένα ενώ άλλες συνδύασαν διαφορετικούς τύπους δεδομένων προκειμένου να εξετάσουν τη συμπεριφορά των προτεινόμενων μοντέλων.

Τα δεδομένα προέρχονται κυρίως από τις ετήσιες οικονομικές εκθέσεις των επιχειρήσεων και συλλέχθηκαν από βάσεις δεδομένων, όπως Securities Exchange Commission (SEC), Emerging Markets Information Service (EMIS), Compustat, Nikkei NEEDS Financial QUEST, Center for Research in Security Prices (CRSP), Taipei Exchange (TPEX), Taiwan Economic Journal (TEJ), China Stock Market & Accounting Research, Infotel, Εθνική τράπεζα της Ελλάδος, ICAP, University of California Irvine (UCI) Machine Learning Repository, U.S. Bureau of Census, U.S. Bureau of Labor Statistics, U.S. Bureau of Economic Analysis και Board of Governors of the Federal Reserve Systems.

Ο όγκος των συνόλων δεδομένων που χρησιμοποιήθηκε καθώς και το χρονικό διάστημα συλλογής τους ποικίλει (βλ. πίνακα 68).

Πίνακας 68: Τύπος δεδομένων - Βάσεις δεδομένων - Χρονικό διάστημα συλλογής

Τίτλος ερευνητικού άρθρου	Δεδομένα	Χρονικό διάστημα	Βάση δεδομένων
Deep learning for detecting financial statement fraud	οικονομικά δεδομένα και δεδομένα κειμένου	1993 - 2019	SEC, L&M word lists by Loughran and McDonald (2011), Readability ratios by Humpherys et al. (2011) and Li (2008)
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές, δεδομένα κειμένου	2016 - 2020	China Stock Market & Accounting Research
Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές	2001 - 2019	Taiwan Economic Journal (TEJ)
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	δεδομένα κειμένου και κινέζικα λεξικά	2012 - 2018	κινέζικα λεξικά NTUSD, HowNet, TSING και DUTIR
Predicting distresses using deep learning of text segments in annual reports	χρηματοοικονομικές μεταβλητές και δεδομένα κειμένου	2013 - 2016	
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	οικονομικά δεδομένα	2012 - 2016	Nikkei NEEDS Financial QUEST
Deep learning models for bankruptcy prediction using textual disclosures	λογιστικών δεδομένα, δεδομένων χρηματιστηριακής αγοράς και δεδομένα κειμένου	1994 - 2014	Compustat North America, CRSP και SEC
A Deep Dense Neural Network for Bankruptcy Prediction	οικονομικά δεδομένα	2003 - 2004	Εθνική τράπεζα της Ελλάδος και την βάση δεδομένων της ICAP
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	οικονομικά δεδομένα		University of California Irvine (UCI) Machine Learning Repository
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	λογιστικές μεταβλητές, μεταβλητές κατασκευαστικής αγοράς και μακροοικονομικές μεταβλητές	1980 - 2016	COMPUSTAT, U.S. Bureau of Census, U.S. Bureau of Labor Statistics, U.S. Bureau of Economic Analysis and Board of Governors of the Federal Reserve Systems.
Bankruptcy or Success? The effective Prediction of a Company's Financial Development Using LSTM	οικονομικά δεδομένα	2014 - 2018	Albertina
Corporate failure prediction: An evaluation of deep learning vs discrete hazard models	λογιστικά δεδομένα και δεδομένα χρηματιστηριακής αγοράς	2001 - 2018	COMPUSTAT και CRSP
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές	1998 - 2003, 1999 - 2009, 2007 - 2013	Infotel, Taiwan Economic Journal, EMIS
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές	2002 - 2019	Taiwan Economic Journal (TEJ)
Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction	χρηματοοικονομικές και μη χρηματοοικονομικές μεταβλητές	2004 - 2019	Taiwan Economic Journal (TEJ), Taipei Exchange (TPEX)

Αναλυτικότερα, οι περισσότερες έρευνες χρησιμοποίησαν οικονομικά δεδομένα τα οποία αντικατοπτρίζουν την οικονομική κατάσταση και τα λειτουργικά αποτελέσματα μιας επιχείρησης και έχει αποδειχθεί, από προηγούμενες έρευνες, ότι είναι ισχυροί δείκτες για την αξιολόγηση των οικονομικών δυνατοτήτων κάθε επιχείρησης (Alam κ.ά., 2021).

Επίσης οι έρευνες των Aljawazneh κ.ά. (2021), Chi και Chu (2021), Jan (2021a), Jan (2021b), Xiuguo και Shengyong (2022) χρησιμοποίησαν επιπλέον και μη χρηματοοικονομικές μεταβλητές που είναι γνωστές και ως μεταβλητές της εταιρικής διακυβέρνησης και σχετίζονται με την ιδιοκτησιακή δομή, τη δομή διαχείρισης και τη γνώμη του ελεγκτή. Οι μεταβλητές αυτές είναι χρήσιμες για την αξιολόγηση της οργάνωσης, του τρόπου διοίκησης και της διαφάνειας των διαδικασιών που ακολουθεί

μια επιχείρηση για την προώθηση των συμφερόντων κάθε ενδιαφερόμενου μέρους και κατά επέκταση την ομαλή ανάπτυξη της κεφαλαιαγοράς (Xiuguo & Shengyong, 2022; Chi & Chu, 2021)

Ενώ η έρευνα των Jang κ.ά. (2021), που ασχολήθηκε με την πρόβλεψη της χρεοκοπίας του κατασκευαστικού κλάδου, χρησιμοποίησε εκτός από τις λογιστικές μεταβλητές και μεταβλητές της κατασκευαστικής αγοράς και της μακροοικονομίας. Ο λόγος που επιλέχθηκαν ήταν διότι (α) οι μεταβλητές της κατασκευαστικής αγοράς επηρεάζουν σημαντικά τις επιχειρηματικές αποφάσεις και δραστηριότητες των κατασκευαστικών εταιρειών και (β) οποιεσδήποτε διακυμάνσεις στις μακροοικονομικές μεταβλητές επηρεάζουν άμεσα τόσο την οικονομική κατάσταση των εργολάβων όσο και τη δυνατότητα χρηματοδότησης των καταναλωτών.

Οι έρευνες των Alam κ.ά. (2021) και των Mai κ.ά. (2019) χρησιμοποίησαν έναν μεγάλο αριθμό λογιστικών δεδομένων και δεδομένων της χρηματιστηριακής αγοράς. Θεωρείται ότι τα δεδομένα της χρηματιστηριακής αγοράς αποτυπώνουν την οικονομική εικόνα μιας επιχείρησης ταχύτερα, αντικατοπτρίζοντας ακόμα και πολύ μικρές αλλαγές της οικονομικής της κατάστασης, σε αντίθεση με τις οικονομικές μεταβλητές οι οποίες βασίζονται σε ιστορικά δεδομένα, δημοσιεύονται στις ετήσιες οικονομικές εκθέσεις μεταγενέστερα και επιπλέον μπορεί να χειραγωγηθούν από τη διοίκηση.

Αξιοσημείωτη είναι η έρευνα του Hosaka (2019) που χρησιμοποίησε ως δεδομένα αριθμητικούς οικονομικούς δείκτες από τις εταιρικές οικονομικές καταστάσεις τους οποίους μετασχημάτισε σε εικόνες της κλίμακας χρώματος γκρι. Κάθε οικονομικός δείκτης αντιστοιχίστηκε σε ένα συγκεκριμένο pixel, η φωτεινότητα του οποίου καθορίστηκε από την τιμή του αντίστοιχου οικονομικού δείκτη.

Όσον αφορά τα δεδομένα κειμένου, αρκετές έρευνες συνδύασαν αριθμητικά δεδομένα με δεδομένα κειμένου κυρίως από την ενότητα MD&A προκειμένου να εξετάσουν κατά πόσο επηρεάζεται η απόδοση της πρόβλεψης των μοντέλων, θεωρώντας ότι η γλώσσα που χρησιμοποιείται στην ενότητα MD&A μπορεί να αποκαλύψει πολύ χρήσιμες πληροφορίες για την πρόβλεψη.

Πιο συγκεκριμένα, οι έρευνες των Craja κ.ά., (2020), των Mai κ.ά. (2019) και των Xiuguo και Shengyong (2022) χρησιμοποίησαν δεδομένα κειμένου από την ενότητα MD&A των ετήσιων οικονομικών εκθέσεων εισηγμένων επιχειρήσεων σε συνδυασμό με ποσοτικά δεδομένα. Ενώ η έρευνα των Matin κ.ά. (2019) συνδύασε ποσοτικά δεδομένα με δεδομένα κειμένου τόσο από την ενότητα MD&A όσο και από την έκθεση των ελεγκτών των ετήσιων οικονομικών εκθέσεων. Αξιοσημείωτη είναι και η έρευνα των Li κ.ά. (2021) οι οποίοι προτείνουν μια νέα προσέγγιση για την κατασκευή ενός κινέζικου λεξικού συναισθήματος για τον χρηματοοικονομικό τομέα, χρησιμοποιώντας δεδομένα κειμένου κινέζικων ετήσιων οικονομικών εκθέσεων σε συνδυασμό με κινέζικα γενικά λεξικά.

Τα πειράματα έδειξαν ότι η απόδοση των μοντέλων διαφοροποιείται ανάλογα με τον τύπο δεδομένων που χρησιμοποιείται κάθε φορά και η μέγιστη απόδοση επιτυγχάνεται με τον συνδυασμό οικονομικών δεδομένων και δεδομένων κειμένου. Επομένως η χρήση μοντέλων DL σε συνδυασμό με διαφορετικούς τύπους δεδομένων προσφέρει βελτιωμένη απόδοση πρόβλεψης και αποτελεσματικότητα στις διαδικασίες λήψης αποφάσεων.

5.4 Εξισορρόπηση του συνόλου δεδομένων

Τα σύνολα δεδομένων συνήθως δεν είναι ισορροπημένα, δηλαδή το πλήθος των δεδομένων της πλειοψηφικής τάξης είναι πολύ μεγαλύτερος από το πλήθος της μειοψηφικής τάξης. Αυτό συμβαίνει διότι στην πραγματικότητα ο αριθμός των χρεοκοπημένων, αποτυχημένων επιχειρήσεων ή των επιχειρήσεων που διαπράττουν απάτη είναι μικρότερος σε σχέση με τον αριθμό των υγιών με αποτέλεσμα τα σύνολα δεδομένων που συλλέγονται να είναι εξαιρετικά ανισόροπα.

Το πρόβλημα της ανισορροπίας των δεδομένων επηρεάζει την απόδοση πρόβλεψης των μοντέλων δεδομένου ότι αυτά τείνουν να προβλέπουν την πλειοψηφική τάξη και να αγνοούν τη μειοψηφική (Aljawazneh κ.ά., 2021).

Κατά τις έρευνες, το πρόβλημα της ανισορροπίας των συνόλων δεδομένων αντιμετωπίστηκε με διάφορες μεθόδους (βλ. πίνακα 69).

Πίνακας 69: Μέθοδοι εξισορρόπησης δεδομένων

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μοντέλο	Τεχνική εξισορρόπησης δεδομένων
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	HAN	Υποδειγματοληψίας
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiuguo και Shengyong (2022)	LSTM, GRU, TRANSFORMER	20 ισορροπημένα υποσύνολα δεδομένων
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	CNN	Interpolation και extrapolation
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti and Soui (2020)	DL BSM-SAES	Borderline SMOTE (BSM)
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	Jang κ.ά. (2021)	LSTM-RNN (χρήση τιμής Shapley)	SMOTE-Tomek link
Corporate failure prediction: An evaluation of deep learning vs discrete hazard models	Alam κ.ά. (2021)	GrNet	Grassmann points
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	Aljawazneh κ.ά. (2021)	DBN, MLP-6L, LSTM	SMOTE, Borderline SMOTE, SMOTE-NC, SVM-SMOTE, ADASYN SMOTE-ENN, SMOTE-Tomek και K-means)

Πιο συγκεκριμένα, οι Xiuguo και Shengyong (2022) επέλεξαν να δημιουργήσουν 20 ισορροπημένα υποσύνολα δεδομένων όπου το κάθε ένα περιείχε 244 περιπτώσεις μη δόλιων εκθέσεων και 244 δόλιων θεωρώντας ότι οι άλλες τεχνικές έχουν μειονεκτήματα, καθώς με την τεχνική υπερδειγματοληψίας δημιουργούνται πολλά επιπλέον αντίγραφα, με την τεχνική υποδειγματοληψίας χρησιμοποιείται μόνο ένα μέρος των δεδομένων πλειοψηφίας, ενώ με την τεχνική SMOTE χρησιμοποιούνται συνθετικά δεδομένα μειοψηφίας.

Ο Hosaka (2019) έκανε χρήση των τεχνικών interpolation και extrapolation. Ως αποτέλεσμα δημιουργήθηκαν 90 συνθετικά δεδομένα για κάθε εταιρεία τα οποία όμως χρησιμοποιήθηκαν μόνο ως training data set, διότι θεωρήθηκαν ακατάλληλα για τη χρήση τους ως test dataset. Τα πειράματα έδειξαν ότι η χρήση συνθετικών δεδομένων δεν εξασφαλίζει το ίδιο αποτέλεσμα με τη χρήση των πραγματικών δεδομένων.

Επίσης οι Craja κ.ά. (2020) εφάρμοσαν την τεχνική υποδειγματοληψίας λαμβάνοντας επίσης υπόψη το έτος και τον τομέα δραστηριότητας των επιχειρήσεων,

ώστε να υπολογιστούν και οι διαφορετικές οικονομικές συνθήκες που επικρατούσαν σε κάθε τομέα.

Ενώ οι Smiti και Soui (2020) εφάρμοσαν τη μέθοδο Borderline SMOTE (BSM), μια βελτιωμένη έκδοση του αλγορίθμου SMOTE η οποία δημιουργεί συνθετικά δεδομένα μόνο για τις περιπτώσεις μειοψηφίας που βρίσκονται κοντά στα όρια των δύο περιπτώσεων ταξινόμησης, καθώς αυτές είναι οι σημαντικότερες για την ταξινόμηση. Σε αντίθεση με την τεχνική SMOTE η οποία δημιουργεί συνθετικά δεδομένα για κάθε περίπτωση μειοψηφίας. Κατά τα πειράματα χρησιμοποιήθηκαν δεδομένα στα οποία εφαρμόστηκε η μέθοδος BSM και δεδομένα όπου δεν χρησιμοποιήθηκε. Τα αποτελέσματα των πειραμάτων έδειξαν ότι η απόδοση όλων των μοντέλων βελτιώθηκε με τη μέθοδο BSM και ιδιαίτερα του προτεινόμενου (βλ. πίνακα 41).

Επίσης οι Jang κ.ά. (2021) χρησιμοποίησαν την τεχνική SMOTE-Tomek link όπου αρχικά χρησιμοποιήθηκε η τεχνική SMOTE για τη δημιουργία των συνθετικών δεδομένων της μειοψηφικής τάξης και στη συνέχεια εφαρμόστηκε η τεχνική Tomek link στο σύνολο όλων των δεδομένων, δηλαδή των πρωτότυπων και των νέων συνθετικών, για την εξισορρόπηση του dataset.

Οι Aljawazneh κ.ά. (2021) έκαναν χρήση πέντε τεχνικών εξισορρόπησης υπερδειγματοληψίας (SMOTE, Borderline SMOTE, SMOTE-NC, SVM-SMOTE, ADASYN), δύο υβριδικών (SMOTE-ENN, SMOTE-Tomek) και μίας clustering-based τεχνικής (K-means), τις οποίες εφάρμοσαν σε τρία διαφορετικά και εξαιρετικά ανισόροπα σύνολα δεδομένων, με σκοπό να εντοπίσουν την καταλληλότερη. Μετά από πειράματα, διαπιστώθηκε ότι η υβριδική τεχνική Oversampling-Undersampling SMOTE-ENN βελτιώνει την απόδοση των περισσότερων μοντέλων ανεξαρτήτως της πολυπλοκότητας των δεδομένων (βλ. πίνακα 59).

Στην έρευνα των Alam κ.ά. (2021) το σύνολο δεδομένων ήταν ένα μη ισορροπημένο data panel, καθώς η χρονική διάρκεια ζωής των εταιρειών είναι διαφορετική. Οι εταιρείες δηλαδή χρεοκοπούν σε διαφορετικές χρονικές περιόδους και σε διαφορετική φάση του κύκλου ζωής τους. Για την αντιμετώπιση του προβλήματος, τα panel data αντιμετωπίστηκαν ως ένα set of vectors κατά μήκος της διάστασης του χρόνου και ο υποχώρος, που δημιουργήθηκε από τα set of vectors, χρησιμοποιήθηκε για την αναπαράσταση των panel data ως Grassmann points με βάση τα οποία εξετάστηκε η γενική πληροφορία του panel data και όχι η πληροφορία για κάθε χρονική στιγμή ξεχωριστά.

5.5 *Τεχνικές επεξεργασίας φυσικής γλώσσας*

Οι έρευνες που χρησιμοποίησαν δεδομένα κειμένου (Craja κ.ά., 2020; Li κ.ά., 2021; Mai κ.ά., 2019; Matin, Hansen, & Molgaard, 2019; Xiuguo & Shengyong, 2022) εφάρμοσαν την τεχνική επεξεργασίας φυσικής γλώσσας Word2Vec προκειμένου τα δεδομένα κειμένου να μετατραπούν σε αριθμητική μορφή και να είναι δυνατή η αλγοριθμική τους επεξεργασία (βλ. πίνακα 70).

Πίνακας 70: Τεχνική επεξεργασίας φυσικής γλώσσας

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μοντέλο	Τεχνική επεξεργασίας φυσικής γλώσσας
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	HAN	Word2Vec
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiuguo και Shengyong (2022)	LSTM, GRU, TRANSFORMER	Word2Vec
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	Li κ.ά. (2021)	DNN, MA-DNN, Bi-LSTM	Word2Vec και pre-training BERT
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	3 models CNN+RNN	Word2Vec (skip-gram)
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	DL-Embedding, CNN	Word2Vec (skip-gram)

Αναλυτικότερα, οι Craja κ.ά. (2020) κατασκεύασαν ένα δίκτυο ιεραρχικής προσοχής (HAN) με τον αλγόριθμο LSTM και ενσωματώσεις word2vec για την εξαγωγή των χαρακτηριστικών κειμένου από την ενότητα MD&A των ετήσιων οικονομικών εκθέσεων. Κατά την έρευνα, διαπιστώθηκε ότι η τεχνική Word2Vec υπερέχει της μεθόδου BOW, καθώς επιτρέπει στις λέξεις με παρόμοια σημασία να έχουν παρόμοιες διανυσματικές αναπαραστάσεις και να καταγράφουν τις συντακτικές και σημασιολογικές ομοιότητες, δηλαδή επιτρέπει στο μοντέλο να κατανοεί τη σημασία των λέξεων. Αντίθετα, η μέθοδος BOW λαμβάνει υπόψη τη συχνότητα εμφάνισης των λέξεων αγνοώντας τη σειρά των λέξεων και τη γραμματική (Craja κ.ά., 2020; Mai κ.ά., 2019).

Η έρευνα των Matin κ.ά. (2019) χρησιμοποίησε το μοντέλο Word2Vec skip-gram και η αρχιτεκτονική του δικτύου ενσωμάτωσε ένα μηχανισμό προσοχής, ώστε το μοντέλο να εστιάζει στα λιγότερα και πιο σημαντικά τμήματα κειμένου.

Επίσης η έρευνα των Mai κ.ά. (2019) χρησιμοποίησε το μοντέλο Word2Vec skip-gram με τον αλγόριθμο negative sampling και εφαρμόστηκε η κανονικοποίηση padding, ώστε τα κείμενα MD&A να έχουν το ίδιο μήκος.

Ενώ οι Xiuguo και Shengyong (2022), λόγω της ιδιαιτερότητας της κινέζικης γλώσσας, εφάρμοσαν ένα κινέζικο πρότυπο εξόρυξης κειμένου που περιλάμβανε δύο διαδικασίες (α) την τμηματοποίηση των κινέζικων λέξεων και (β) τον υπολογισμό των word vector με την τεχνική Word2Vec, ώστε να μη χαθούν σημαντικές πληροφορίες που θα επηρέαζαν την ακρίβεια πρόβλεψης του μοντέλου.

Κατά την έρευνα τους οι Li κ.ά. (2021), έκαναν λήψη των word vectors με τη χρήση των μοντέλων επεξεργασίας φυσικής γλώσσας Word2Vec και pre-training BERT. Σύμφωνα με τα αποτελέσματα των πειραμάτων, διαπιστώθηκε ότι τα μοντέλα που χρησιμοποίησαν τα BERT-based word vectors είχαν καλύτερη απόδοση. Στη συνέχεια, η εξαγωγή των σημαντικών χαρακτηριστικών συναισθήματος από το λεξικό CFDSL, που δημιούργησαν, έγινε με βάση την ερμηνευσιμότητα του μοντέλου decision tree και τα αποτελέσματα των πειραμάτων έδειξαν ότι τα σημαντικά χαρακτηριστικά του λεξικού CFDSL έχουν καλύτερη απόδοση στην πρόβλεψη σε σύγκριση με τα λεξικά L&M και NTUSD.

5.6 Επιλογή χαρακτηριστικών

Η επιλογή χαρακτηριστικών είναι μια διαδικασία απομόνωσης των σημαντικών χαρακτηριστικών από τα υπόλοιπα χαρακτηριστικά του συνόλου δεδομένων που περιέχουν πληροφορίες άσχετες και περιττές. Σημαντικά είναι τα χαρακτηριστικά εκείνα που περιέχουν χρήσιμες πληροφορίες για το μοντέλο και ουσιαστικά επηρεάζουν την απόδοση πρόβλεψης. Ο βασικός στόχος της διαδικασίας είναι η μείωση του υπολογιστικού κόστους, η μείωση του χρόνου εκπαίδευσης του μοντέλου και η βελτίωση της απόδοσης πρόβλεψης (Κύρκος, 2015).

Από το σύνολο των ερευνητικών άρθρων, τρεις έρευνες εφάρμοσαν μεθόδους επιλογής χαρακτηριστικών και μελέτησαν τη συμπεριφορά των μοντέλων (βλ. πίνακα 71).

Πίνακας 71: Μέθοδοι επιλογής χαρακτηριστικών

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μοντέλο	Επιλογή χαρακτηριστικών
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	Jan (2021b)	CART-DNN, CART-RNN	CART
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	DL BSM-SAES	Stacked auto-encoder
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	CNN	Correlated

Στην έρευνα των Jan κ.ά. (2021b) η επιλογή των δέκα σημαντικότερων μεταβλητών έγινε με τη χρήση του αλγορίθμου CART. Στη συνέχεια, συγκρίθηκε η απόδοση των μοντέλων με τη χρήση τόσο ολόκληρου του συνόλου δεδομένων όσο και μόνο των επιλεγμένων μεταβλητών και διαπιστώθηκε ότι τα μοντέλα που χρησιμοποίησαν τις επιλεγμένες μεταβλητές είχαν βελτιωμένη απόδοση.

Στην έρευνα των Smiti και Soui (2020) η εξαγωγή των πιο σημαντικών χαρακτηριστικών έγινε με την εφαρμογή της τεχνικής Stacked auto-encoder (SAE). Με τη διαδικασία αυτή, το μοντέλο DL ήταν εξαιρετικά ακριβές σε σχέση με τα συγκρινόμενα μοντέλα machine learning που πραγματοποίησαν τη διαδικασία εξαγωγής χαρακτηριστικών ανεξάρτητα από τη φάση εκπαίδευσης. Όμως θα πρέπει να σημειωθεί ότι το προτεινόμενο μοντέλο DL είχε τη χειρότερη απόδοση στο χρόνο εκπαίδευσης που οφείλεται στον χρόνο που αφιερώθηκε για την εξαγωγή αυτών των σημαντικών χαρακτηριστικών.

Η έρευνα του Hosaka (2019) χρησιμοποίησε ως δεδομένα αριθμητικούς οικονομικούς δείκτες από τις εταιρικές οικονομικές καταστάσεις τους οποίους μετασχημάτισε σε εικόνες της κλίμακας χρώματος γκρι. Κάθε οικονομικός δείκτης αντιστοιχίστηκε σε ένα συγκεκριμένο pixel, η φωτεινότητα του οποίου καθορίστηκε από την τιμή του αντίστοιχου οικονομικού δείκτη. Για την αντιστοίχιση των οικονομικών δεικτών σε ένα συγκεκριμένο pixel εφάρμοσε (α) τη μέθοδο Random όπου η αντιστοίχιση των οικονομικών δεικτών σε θέσεις pixel έγινε τυχαία και (β) τη μέθοδο Correlated όπου η αντιστοίχιση έγινε ανάλογα με τον βαθμό συσχέτισης των οικονομικών δεικτών μεταξύ τους, δηλαδή όσο πιο σχετικοί ήταν οι οικονομικοί

δείκτες μεταξύ τους τόσο σε κοντινότερες θέσεις pixel τοποθετήθηκαν. Στη συνέχεια, οι εικόνες που δημιουργήθηκαν χρησιμοποιήθηκαν ως δεδομένα εκπαίδευσης. Τα πειράματα έδειξαν ότι η μέθοδος correlated είναι καταλληλότερη, καθώς τα αποτελέσματα της πρόβλεψης ήταν βελτιωμένα.

5.7 *Επικύρωση (validation)*

Η επικύρωση είναι μία σημαντική διαδικασία για την επιβεβαίωση της ακρίβειας ενός μοντέλου και την επιτυχή εφαρμογή του σε θέματα λήψης αποφάσεων. Μέσω της διαδικασίας επικύρωσης διερευνάται η ικανότητα σωστής ταξινόμησης άγνωστων δεδομένων του μοντέλου, δηλαδή δεδομένων που δεν χρησιμοποιήθηκαν κατά την εκπαίδευσή του. Στην ουσία διερευνάται η ικανότητα γενίκευσης των μοντέλων (Κύρκος, 2015). Για να επιτευχθεί η διαδικασία επικύρωσης γίνεται διαχωρισμός του συνόλου δεδομένων με διάφορες μεθόδους.

Σύμφωνα με τα εξεταζόμενα ερευνητικά άρθρα (βλ. πίνακα 72), αρκετές έρευνες χρησιμοποίησαν τη μέθοδο επικύρωσης cross validation κατά την οποία το σύνολο δεδομένων διαιρέθηκε σε τμήματα (folds), ένα εκ των οποίων χρησιμοποιήθηκε ως σύνολο επικύρωσης και τα υπόλοιπα ως σύνολα εκπαίδευσης ενώ άλλες τη μέθοδο hold out (Κύρκος, 2015).

Συγκεκριμένα, οι έρευνες των Aljawazneh κ.ά. (2021), Li κ.ά. (2021), Matin κ.ά. (2019) χρησιμοποίησαν τη μέθοδο 10-fold cross validation και η έρευνα του Hosaka (2019) τη μέθοδο 5-fold cross validation, δηλαδή τα σύνολα δεδομένων χωρίστηκαν σε 10 και 5 τμήματα αντίστοιχα. Ενώ η έρευνα των Mai κ.ά. (2019) χρησιμοποίησε τη μέθοδο 10*10-fold cross validation κατά την οποία η παραπάνω διαδικασία επαναλήφθηκε 10 φορές.

Επίσης άλλες έρευνες (Alam κ.ά., 2021; Craja κ.ά., 2020; Jang κ.ά., 2021; Mai κ.ά., 2019; Smiti & Soui, 2020; Vochozka κ.ά., 2020; Xiuguo & Shengyong, 2022) χρησιμοποίησαν τη μέθοδο hold out όπου το σύνολο δεδομένων χωρίστηκε τυχαία σε training dataset και testing dataset ενώ σε τρεις έρευνες (Chi & Chu, 2021; Jan, 2021a; Jan, 2021b) το σύνολο δεδομένων χωρίστηκε σε training dataset, validation dataset και testing dataset.

Πίνακας 72: Μέθοδοι επικύρωσης

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Validation
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	Hold out (training & testing dataset)
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiuguao και Shengyong (2022)	Hold out (training & testing dataset)
Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	Jan (2021a)	Hold out (training, validation & testing dataset)
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	Li κ.ά. (2021)	10-fold cross-validation
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	10-fold cross-validation
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	5-fold cross-validation
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	10*10-fold cross-validation
A Deep Dense Neural Network for Bankruptcy Prediction	Alexandropoulos κ.ά. (2019)	
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	Jang κ.ά. (2021)	Hold out (training & testing dataset)
Bankruptcy or Success? The effective Prediction of a Company's Financial Development Using LSTM	Vochozka κ.ά. (2020)	Hold out (training & testing dataset)
Corporate failure preiction: An evaluation of deep learning vs discrete hazard models	Alam κ.ά. (2021)	Hold out (training & testing dataset)
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	Aljawazneh κ.ά. (2021)	10-fold cross-validation
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	Jan (2021b)	Hold out (training, validation & testing dataset)
Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction	Chi και Chu (2021)	Hold out (training, validation & testing dataset)

5.8 Ερμηνευσιμότητα

Τα μοντέλα βαθιάς μάθησης θεωρούνται μοντέλα μαύρου κουτιού (black box), διότι δεν είναι κατανοητή η αλληλεπίδραση των νευρώνων και είναι δύσκολο να ερμηνευθεί το αποτέλεσμα της πρόβλεψης (Κύρκος, 2015).

Η έρευνα των Jang κ.ά. (2021) ασχολήθηκε με τη μέτρηση των επιπτώσεων που έχουν οι μεταβλητές εισόδου στην ακρίβεια πρόβλεψης με τη χρήση της τιμής Sharpley και τον εντοπισμό της μεταβλητής εκείνης που επηρεάζει περισσότερο την ακρίβεια πρόβλεψης, εισάγοντας κάθε φορά διαφορετικό συνδυασμό μεταβλητών με φθίνουσα σειρά κατάταξης των τιμών Sharpley. Ο εντοπισμός των επιπτώσεων των μεταβλητών εισόδου είναι σημαντικός για να γίνουν περισσότερο κατανοητές οι εσωτερικές διεργασίες του μοντέλου και να βρεθούν τρόποι βελτίωσής του.

Επίσης οι Mai κ.ά. (2019) εφάρμοσαν τη μέθοδο representation erasure, ώστε να εντοπίσουν τις λέξεις που είναι σημαντικές με σκοπό να ερμηνευθούν οι αποφάσεις του μοντέλου. Έτσι διαγράφοντας μεμονωμένες λέξεις κάθε φορά από το σώμα εισόδου παρατήρησαν πως διαφοροποιείται η απόδοση του μοντέλου με βάση το μέτρο AUC. Έτσι η μείωση της τιμής AUC οδηγούσε στο συμπέρασμα ότι η συγκεκριμένη λέξη είναι σημαντική για το μοντέλο. Στη συνέχεια, οι λέξεις που συγκεντρώθηκαν, συγκρίθηκαν με τις λέξεις δύο λεξικών συναισθήματος (Loughran & McDonald, 2011; Wilson, Wiebe & Hofmann, 2005 όπως αναφέρεται στους Mai κ.ά., 2019) (βλ. πίνακα 73) και βρέθηκε ότι πολλές από αυτές δεν υπήρχαν σε κανένα από τα δύο λεξικά υποδεικνύοντας ότι το μοντέλο deep learning θεωρεί άλλες λέξεις σημαντικότερες από αυτές των λεξικών συναισθήματος. Μάλιστα, εκτός από τις λέξεις που αφορούν την απόδοση μιας επιχείρησης, υπάρχουν λέξεις οι οποίες σχετίζονται με τη διάρθρωση κεφαλαίου (repurchase, dividend, tranche), τη στρατηγική (international, exit, focus) και τα ενδιαφερόμενα μέρη (compensation, wages, costs, suppliers).

Πίνακας 73: Οι 100 σημαντικότερες λέξεις της ενότητας MD&A μετά την εφαρμογή της μεθόδου representation erasure. Οι λέξεις με έντονη γραφή υπάρχουν και στο λεξικό Loughran and McDonald (2011) ενώ οι υπογραμμισμένες στο λεξικό MPQA Subjectivity (Wilson, Wiebe & Hofmann, 2005)

Non-bankruptcy Firms	Bankruptcy Firms
income, increase, increased, future, revenues, rate, intangible, <u>profit</u> , compensation, growth, tax, percentage, <u>goodwill</u> , cash, <u>value</u> , investment, improved , term, compared, economic, products, intangibles, changes, revenue, repurchases, <u>outstanding</u> , invested, repurchased, marketable, rates, repurchase, electronic, strong , expenditures, construction, maturity, imaging, credit, accounts, dividend, latest, excluding, international, bank, ebitda, holdings, suppliers, <u>well</u> , partners, long	loss , services, trust, initial, <u>decrease</u> , fees, sale, public, extraordinary, ended, structure, managed, measurements, inception, recourse, inventory, room, accordingly, expenses, serviced, approval, prime, restated , incurred, stores, indebtedness, secured, certificates, discontinued , affiliate, convertible, exit, tranche, servicing, focus, backed, announced, disposal, mortgage, joint, reset, aggregate, conversion, generated, production, received, costs, receivables, selling, wages

Πηγή: (Mai κ.ά., 2019)

5.9 Μοντέλα αναφοράς

Για τη σύγκριση της απόδοσης των προτεινόμενων μοντέλων χρησιμοποιήθηκαν διαφορετικά μοντέλα αναφοράς. Τα πιο συχνά χρησιμοποιούμενα μοντέλα είναι τα Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting και (XGB) Support Vector Machine (SVM) (βλ. πίνακα 74). Τα παραπάνω μοντέλα αναφοράς έχουν χρησιμοποιηθεί σε προηγούμενες έρευνες σημειώνοντας πολύ καλές επιδόσεις στην πρόβλεψη.

Πίνακας 74: Τα μοντέλα αναφοράς

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μοντέλα αναφοράς
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	LR, RF, SVM, XGB και ANN
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiuguo και Shengyong (2022)	CNN, RF, SVM, XGB, ANN και LR
Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	Jan (2021a)	
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	Li κ.ά. (2021)	
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	XGB, Logit και NN
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	CART, LDA, SVM, MLP και AdaBoost
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	LR, SVM και RF
A Deep Dense Neural Network for Bankruptcy Prediction	Alexandropoulos κ.ά. (2019)	LR, MP, NB και Cart
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	KNN, DT, SVM, ANN, RF και C5.0
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	Jang κ.ά. (2021)	
Bankruptcy or Success? The effective Prediction of a Company's Financial Development Using LSTM	Vochozka κ.ά. (2020)	
Corporate failure preiction: An evaluation of deep learning vs discrete hazard models	Alam κ.ά. (2021)	Discrete-time hazard model
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	Aljawazneh κ.ά. (2021)	RF, SVM, KNN, AdaBoost και XGBoost
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	Jan (2021b)	
Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction	Chi και Chu (2021)	

Το **Logistic Regression (LR)** είναι ένα δημοφιλές στατιστικό μοντέλο παλινδρόμησης που χρησιμοποιείται συχνά για την πρόβλεψη και αποτελεί ένα από τα

παραδοσιακά εργαλεία οικονομικής ανάλυσης. Είναι απλό στη χρήση του και με καλή απόδοση στα αποτελέσματα. Το κύριο μειονέκτημά του είναι ότι βασίζεται σε αυθαίρετες υποθέσεις (Κύρκος, 2015). Χρησιμοποιήθηκε στις έρευνες των Craja κ.ά. (2020) και των Xiuguo και Shengyong (2022) για την πρόβλεψη απάτης, των Matin κ.ά. (2019) για την πρόβλεψη οικονομικής δυσπραγίας, των Alexandropoulos κ.ά. (2019) και των Mai κ.ά. (2019) για την πρόβλεψη χρεοκοπίας.

Το **Random Forest** (RF) είναι μια μέθοδος ταξινόμησης συνόλου Machine Learning, βασίζεται στη δημιουργία διαφορετικών δέντρων απόφασης από διαφορετικά σύνολα του αρχικού συνόλου δεδομένων με τη μέθοδο bootstrapping και το αποτέλεσμα ταξινόμησης προκύπτει με βάση την πλειοψηφία των δέντρων απόφασης. Έχει τη δυνατότητα χειρισμού μεγάλου αριθμού μεταβλητών, την ικανότητα γενίκευσης και υψηλή ακρίβεια ταξινόμησης. Όμως απαιτεί μεγάλη υπολογιστική ισχύ, μεγάλο χρόνο εκπαίδευσης και η ερμηνευσιμότητα του είναι χαμηλότερη σε σύγκριση με ένα decision tree. Χρησιμοποιήθηκε στις έρευνες των Craja κ.ά. (2020) και των Xiuguo και Shengyong (2022) για την πρόβλεψη απάτης, των Aljawazneh κ.ά. (2021) για την πρόβλεψη της οικονομικής αποτυχίας και των Mai κ.ά. (2019) και Smiti και Soui (2020) για την πρόβλεψη χρεοκοπίας.

Το **Extreme Gradient Boosting** είναι μια μέθοδος ταξινόμησης συνόλου Machine Learning που βασίζεται σε δέντρα απόφασης. Ανήκει στην οικογένεια των gradient boosting machines, έχει τη δυνατότητα χειρισμού μεγάλων συνόλων δεδομένων, είναι υψηλής ακρίβειας αλλά δεν αποδίδει τόσο καλά σε αραιά και αδόμητα δεδομένα. Μάλιστα σε θορυβώδη δεδομένα υπάρχει κίνδυνος υπερπροσαρμογής. Είναι ένα μοντέλο μη ερμηνεύσιμο και συχνά χαρακτηρίζεται ως μοντέλο “μαύρου κουτιού”. Χρησιμοποιήθηκε στις έρευνες των Craja κ.ά. (2020), Li κ.ά. (2021) και Xiuguo και Shengyong (2022) για την πρόβλεψη απάτης, στην έρευνα των Aljawazneh κ.ά. (2021) για την πρόβλεψη της οικονομικής αποτυχίας και των Hosaka (2019), Mai κ.ά. (2019) και Smiti και Soui (2020) για την πρόβλεψη χρεοκοπίας.

Το **Support Vector Machine** (SVM) είναι ένα πολύ δημοφιλές μοντέλο Machine Learning και χρησιμοποιείται σε προβλήματα δυαδικής ταξινόμησης με επιτυχία, λόγω της ισχυρής του απόδοσης. Είναι ανθεκτικό στην υπερπροσαρμογή και έχει την ικανότητα γενίκευσης. Όμως απαιτείται μεγάλος χρόνος για την εκπαίδευση, μεγάλη υπολογιστική ισχύ και τα αποτελέσματά του δεν είναι ερμηνεύσιμα (Κύρκος, 2015). Χρησιμοποιήθηκε στις έρευνες των Craja κ.ά. (2020), Li κ.ά. (2021) και Xiuguo και Shengyong (2022) για την πρόβλεψη απάτης, των Aljawazneh κ.ά. (2021) για την πρόβλεψη της οικονομικής αποτυχίας και των Hosaka (2019), Mai κ.ά. (2019) και Smiti και Soui (2020) για την πρόβλεψη χρεοκοπίας.

5.10 Μέτρα απόδοσης

Η πρόβλεψη αντιμετωπίζεται ως ένα πρόβλημα δυαδικής ταξινόμησης με τέσσερα πιθανά αποτελέσματα (Κύρκος, 2015):

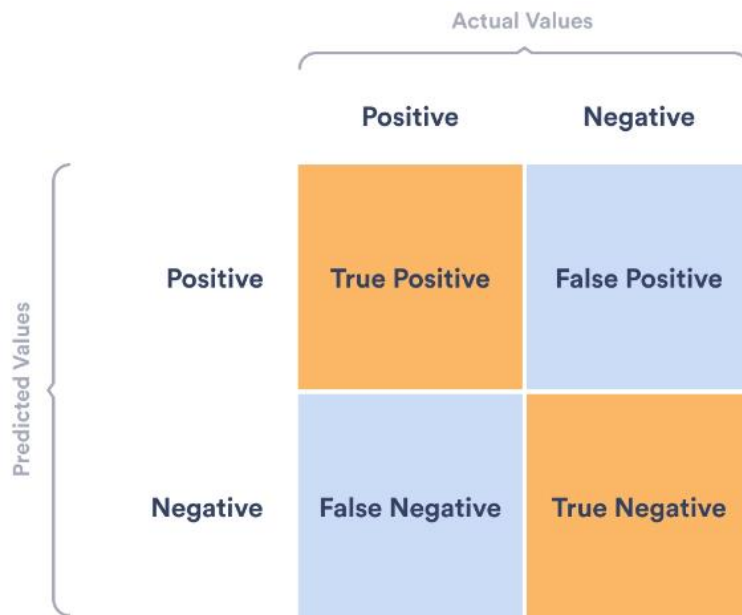
(α) **True positive** (TP) είναι ο αριθμός των θετικών προβλέψεων που στην πραγματικότητα είναι θετικές, δηλαδή η ταξινόμηση είναι σωστή,

(β) **False positive** (FP) είναι ο αριθμός των θετικών προβλέψεων που στην πραγματικότητα είναι αρνητικές, δηλαδή η ταξινόμηση είναι εσφαλμένη.

(γ) **False negative** (FN) είναι ο αριθμός των αρνητικών προβλέψεων που στην πραγματικότητα είναι θετικές, δηλαδή η ταξινόμηση είναι εσφαλμένη.

(δ) **True negative** (TN) είναι ο αριθμός των αρνητικών προβλέψεων που στην πραγματικότητα είναι αρνητικές, δηλαδή η ταξινόμηση είναι σωστή,

Τα παραπάνω αποτελέσματα συνοψίζονται στο confusion matrix ο οποίος είναι ένας πίνακας διαστάσεων 2x2 και βοηθάει στην κατανόηση των αποτελεσμάτων ταξινόμησης (βλ. εικόνα 61).



Εικόνα 61: Confusion matrix

Πηγή: (Seraydarian, 2022)

Ένα μοντέλο θεωρείται αξιόπιστο όταν έχει υψηλά ποσοστά True Positive (TP) και True Negative (TN) και χαμηλά False Positive (FP) και False Negative (FN), δηλαδή όταν τα ποσοστά των σωστών ταξινομήσεων είναι υψηλά και των εσφαλμένων χαμηλά. Η χρήση του confusion matrix είναι κατάλληλη για μη ισορροπημένα σύνολα δεδομένων, διότι δίνει μια καλύτερη εικόνα της απόδοσης ενός μοντέλου απεικονίζοντας τόσο τις σωστές όσο και τις εσφαλμένες περιπτώσεις ταξινόμησης. Χρησιμοποιείται για τη μέτρηση της απόδοσης των μοντέλων μέσω του υπολογισμού των μέτρων accuracy, precision, recall (sensitivity), specificity, F-score, error type I, error type II και της καμπύλης ROC (AUC).

Αναλυτικότερα, το μέτρο απόδοσης **Accuracy** αξιολογεί την ακρίβεια πρόβλεψης των μοντέλων. Μετρά το ποσοστό των σωστά ταξινομημένων περιπτώσεων στο σύνολο των δεδομένων. Είναι κατάλληλο για ισορροπημένα σύνολα δεδομένων και όχι ανισόρροπα.

$$\text{Υπολογίζεται ως εξής: } Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Το **Precision** μετρά το ποσοστό των σωστά ταξινομημένων θετικών προβλέψεων (TP) σε σχέση με το σύνολο όλων των θετικών προβλέψεων (TP+FP).

Υπολογίζεται ως εξής: $Precision = \frac{TP}{TP + FP}$

Το **Recall** (sensitivity) μετρά το ποσοστό των σωστά ταξινομημένων θετικών προβλέψεων (TP) σε σχέση με το σύνολο όλων των πραγματικών θετικών (TP+FN).

Υπολογίζεται ως εξής: $Recall = \frac{TP}{TP + FN}$

Το **Specificity** μετρά το ποσοστό των σωστά ταξινομημένων αρνητικών προβλέψεων (TN) σε σχέση με το σύνολο όλων των πραγματικών αρνητικών.

Υπολογίζεται ως εξής: $Specificity = \frac{TN}{TN + FP}$

Το **F-score** ή **F-measure** συνδυάζει τα μέτρα απόδοσης recall και precision και χρησιμοποιείται για την αξιολόγηση της ακρίβειας του μοντέλου υπολογίζοντας το ποσοστό των εσφαλμένων ταξινομήσεων. Είναι κατάλληλο για ανισόρροπα δεδομένα και υπολογίζεται ως εξής:

$$F - score = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

Το **error type I** (FPR) είναι το ποσοστό των θετικών προβλέψεων που στην πραγματικότητα είναι αρνητικές, δηλαδή η ταξινόμηση είναι εσφαλμένη.

Υπολογίζεται ως εξής: $error_type_I = \frac{FP}{TN + FP} = 1 - specificity$

Το **error type II** (FNR) είναι το ποσοστό των αρνητικών προβλέψεων που στην πραγματικότητα είναι θετικές, δηλαδή η ταξινόμηση είναι εσφαλμένη.

Υπολογίζεται ως εξής: $error_type_II = \frac{FN}{TP + FN} = 1 - recall$

Θα πρέπει να σημειωθεί ότι η εσφαλμένη ταξινόμηση μιας χρεοκοπημένης εταιρείας στις υγιείς προκαλεί διαφορετικό κόστος σε σύγκριση με την εσφαλμένη ταξινόμηση μιας υγιούς εταιρείας στις χρεοκοπημένες. Στην πρώτη περίπτωση μπορεί να προκαλέσει σοβαρές οικονομικές απώλειες στα ενδιαφερόμενα μέρη, όπως στους επενδυτές, δανειστές, προμηθευτές, εργαζόμενους κ.α., ενώ στη δεύτερη να προκαλέσει κόστος στη φήμη μιας επιχείρησης, υψηλό κόστος κεφαλαίου κ.α. (Aljawazneh κ.ά., 2021; Chi & Chu, 2021; Jan, 2021b).

Τέλος, η **καμπύλη ROC** (Receiver Operating Characteristics) είναι μια γραφική παράσταση όπου ο οριζόντιος άξονας εκφράζει το False positive rate (*False positive rate*= $1 - specificity$), δηλαδή το ποσοστό των αρνητικών περιπτώσεων που ταξινομήθηκαν εσφαλμένα και ο κάθετος άξονας αντιστοιχεί στο True positive rate (*True positive rate*=*sensitivity*), δηλαδή το ποσοστό των θετικών περιπτώσεων που ταξινομήθηκαν σωστά και δείχνει την ικανότητα πρόβλεψης ενός μοντέλου (Κύρκος, 2015).

Για τη σύγκριση της απόδοσης των μοντέλων χρησιμοποιείται το μέτρο **AUC** (Area Under ROC Curve) που είναι κατάλληλο για μη ισορροπημένα σύνολα δεδομένων και γι' αυτό χρησιμοποιείται συχνά στην πρόβλεψη δυαδικής ταξινόμησης.

Οι τιμές που παίρνει είναι μεταξύ 0 και 1 και όσο πιο κοντά στο 1 είναι η τιμή τόσο ορθότερα ταξινομεί τα αποτελέσματα της πρόβλεψης το μοντέλο.

Υπολογίζεται ως εξής (Smiti & Soui, 2020): $AUC = \frac{Sensitivity + Specificity}{2}$.

Παρατηρώντας τα μέτρα απόδοσης που υιοθετήθηκαν από τα ερευνητικά άρθρα (βλ. πίνακα 75), διαπιστώνει κανείς ότι τα πιο συχνά χρησιμοποιούμενα μέτρα είναι τα **accuracy, recall (sensitivity), specificity, F-measures, type I error, type II error και η καμπύλη ROC (AUC)**. Πιο συγκεκριμένα, δέκα έρευνες έλαβαν υπόψη τους το accuracy, οκτώ το recall (sensitivity), επτά έρευνες τα μέτρα απόδοσης specificity, F-measures και την καμπύλη ROC (AUC) και τρεις τα precision, type I error και type II error.

Λαμβάνοντας υπόψη το σκοπό της κάθε έρευνας, διαπιστώνεται ότι το σύνολο των ερευνών που επικεντρώθηκε στην πρόβλεψη **απάτης** (Craja κ.ά., 2020; Jan, 2021a; Xiuguo & Shengyong, 2022) χρησιμοποίησαν τέσσερα κοινά μέτρα απόδοσης και συγκεκριμένα τα μέτρα accuracy, recall, specificity, F-measures και AUC. Οι έρευνες που επικεντρώθηκαν στην πρόβλεψη **συνεχιζόμενης δραστηριότητας** (Chi & Chu, 2021; Jan, 2021b) πέντε κοινά μέτρα απόδοσης που ήταν τα accuracy, recall, specificity, F-measure και precision. Οι έρευνες που ασχολήθηκαν με την πρόβλεψη της **οικονομικής αποτυχίας** (Alam κ.ά., 2021; Aljawazneh κ.ά., 2021) υιοθέτησαν τρία κοινά μέτρα απόδοσης τα accuracy, recall και specificity. Από τις έξι έρευνες που ασχολήθηκαν με την πρόβλεψη της **χρεοκοπίας**, οι τέσσερις (Alexandropoulos κ.ά., 2019; Hosaka, 2019; Mai κ.ά., 2019; Smiti & Soui, 2020) χρησιμοποίησαν το μέτρο AUC ενώ οι έρευνες που επικεντρώθηκαν στην πρόβλεψη της **οικονομικής δυσπραγίας** (Li κ.ά., 2021; Matin κ.ά., 2019) χρησιμοποίησαν διαφορετικά μέτρα απόδοσης η κάθε μία.

Με βάση τα παραπάνω, εξάγεται το συμπέρασμα ότι η αξιολόγηση των ερευνών εστιάζει κυρίως (α) στο ποσοστό των σωστά ταξινομημένων περιπτώσεων του συνόλου των δεδομένων (accuracy), (β) στο ποσοστό των σωστά ταξινομημένων θετικών/αρνητικών προβλέψεων σε σχέση με το σύνολο των πραγματικών θετικών/αρνητικών προβλέψεων (recall και specificity αντίστοιχα) και (γ) στην ικανότητα ορθής ταξινόμησης πρόβλεψης των μοντέλων (καμπύλη ROC (AUC)). Επίσης ορισμένες έρευνες (Chi & Chu, 2021; Craja κ.ά., 2020; Hosaka, 2019; Jan, 2021a; Jan, 2021b; Li κ.ά., 2021; Xiuguo & Shengyong, 2022) υιοθέτησαν τα F-measures θέλοντας να αξιολογήσουν την ακρίβεια των μοντέλων ως προς το ποσοστό των εσφαλμένων ταξινομήσεων. Ενώ άλλες έρευνες (Aljawazneh κ.ά., 2021; Chi & Chu, 2021; Jan, 2021a) χρησιμοποίησαν τα error rates type I & II, λόγω του κόστους που προκαλούν και όσο πιο χαμηλά είναι τα ποσοστά τους τόσο καλύτερη και η απόδοση του μοντέλου.

Πίνακας 75: Τα μέτρα απόδοσης των ερευνητικών άρθρων

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Μέτρα απόδοσης
Deep learning for detecting financial statement fraud	Craja κ.ά. (2020)	Accuracy, Recall (Sensitivity), Specificity, F1-score, F2-score, AUC
An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning	Xiugao και Shengyong (2022)	Accuracy, Recall (Sensitivity), Specificity, F1-score, F2-score, AUC
Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry	Jan (2021a)	Accuracy, Recall (Sensitivity), Specificity, F1-score, Precision, Type I error, Type II error, Training time, AUC
A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction	Li κ.ά. (2021)	Accuracy, Recall (Sensitivity), Precision, F1-score
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	AUC, log-score
Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka (2019)	Correct estimation rates for each class, F-measure, καμπύλη ROC
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	Accuracy, AUC, Cumulative decile-ranking
A Deep Dense Neural Network for Bankruptcy Prediction	Alexandropoulos κ.ά. (2019)	AUC
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	AUC, training time
Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors	Jang κ.ά. (2021)	τιμή Shapley, Accuracy
Bankruptcy or Success? The effective Prediction of a Company's Financial Development Using LSTM	Vochozka κ.ά. (2020)	Confusion matrix
Corporate failure prediction: An evaluation of deep learning vs discrete hazard models	Alam κ.ά. (2021)	Accuracy, Recall (Sensitivity), Specificity
Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure	Aljawazneh κ.ά. (2021)	Accuracy, Recall (Sensitivity), Specificity, Precision, Type I error, Type II error
Using Deep Learning Algorithms for CPAs' Going Concern Prediction	Jan (2021b)	Accuracy, Recall (Sensitivity), Specificity, F1-score, Precision
Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction	Chi και Chu (2021)	Accuracy, Recall (Sensitivity), Specificity, F1-score, Precision, Type I error, Type II error

5.11 Στατιστική ανάλυση των αποτελεσμάτων

Για τον έλεγχο των αποτελεσμάτων πρόβλεψης συχνά εφαρμόζονται στατιστικές αναλύσεις, ώστε να ελεγχθούν στατιστικά οι διαφορές μεταξύ των αποτελεσμάτων και να επιβεβαιωθεί αν είναι στατιστικά σημαντικές (Kirkos, 2012).

Μόνο τρεις έρευνες έκαναν στατιστική ανάλυση των αποτελεσμάτων τους για τον υπολογισμό των p-values (βλ. πίνακα 76). Συγκεκριμένα, δύο έρευνες (Matin κ.ά., 2019; Smiti & Soui, 2020) εφάρμοσαν τη μέθοδο paired t-test ενώ η έρευνα των Mai κ.ά. (2019) χρησιμοποίησε τη μέθοδο Salzberg binomial test. Τα αποτελέσματα της ανάλυσης των παραπάνω επιβεβαίωσαν την υπεροχή των προτεινόμενων μοντέλων DL.

Πίνακας 76: Μέθοδοι στατιστικής ανάλυσης

Τίτλος ερευνητικού άρθρου	Συγγραφείς	Στατιστική ανάλυση
Predicting distresses using deep learning of text segments in annual reports	Matin κ.ά. (2019)	Paired t-test
Deep learning models for bankruptcy prediction using textual disclosures	Mai κ.ά. (2019)	Salzberg binomial test
Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE	Smiti και Soui (2020)	Paired t-test

5.12 Συμπεράσματα

Σύμφωνα με τα παραπάνω διαπιστώνεται ότι κάθε έρευνα ανέπτυξε διαφορετικά μοντέλα DL βασισμένα κυρίως στους τύπους νευρωνικών δικτύων MLP, CNN και RNN.

Επίσης διαπιστώθηκε ότι τα προτεινόμενα μοντέλα είναι κατάλληλα για τον σκοπό για τον οποίο αναπτύχθηκαν και ότι υπερτερούν σε σχέση με τα μοντέλα αναφοράς με τα οποία συγκρίθηκαν.

Οι έρευνες υιοθέτησαν διαφορετικούς τύπους δεδομένων τα οποία συλλέχθηκαν από διαφορετικές βάσεις δεδομένων και ποικίλουν ως προς το πλήθος και το χρονικό διάστημα συλλογής τους. Επιπλέον θα πρέπει να σημειωθεί ότι τα δεδομένα προέρχονται από διαφορετικές χώρες όπου επικρατούν διαφορετικές οικονομικές συνθήκες και κανονισμοί.

Ορισμένες έρευνες εφάρμοσαν τεχνικές εξισορρόπησης δεδομένων και επιλογής χαρακτηριστικών διαφορετικών μεθόδων, δίνοντας ικανοποιητικά αποτελέσματα στην

απόδοση των μοντέλων, ενώ οι έρευνες που έκαναν χρήση δεδομένων κειμένου εφάρμοσαν κυρίως την τεχνική επεξεργασίας φυσικής γλώσσας Word2Vec.

Η επικύρωση, για την επιβεβαίωση της ικανότητας σωστής ταξινόμησης άγνωστων δεδομένων των προτεινόμενων μοντέλων DL, έγινε κυρίως με τις μεθόδους cross validation και holdout.

Επίσης δύο έρευνες ασχολήθηκαν (Jang κ.ά., 2021; Xiuguo & Shengyong, 2022) με τον εντοπισμό των επιπτώσεων των μεταβλητών εισόδου στην πρόβλεψη με τη χρήση της τιμής Shapley και της μεθόδου representation erasure με σκοπό την κατανόηση των εσωτερικών διεργασιών των μοντέλων και την ερμηνεία των αποφάσεων που λαμβάνουν.

Τα μοντέλα αναφοράς και τα μέτρα απόδοσης, που λήφθηκαν υπόψη για την αξιολόγηση της απόδοσης των προτεινόμενων μοντέλων DL, διαφέρουν μεταξύ τους ενώ μόνο τρεις (Mai κ.ά., 2019; Matin κ.ά., 2019; Smiti & Soui, 2020) πραγματοποίησαν στατιστική ανάλυση για την επιβεβαίωση των αποτελεσμάτων τους.

6

Επίλογος

6.1 Σύνοψη και συμπεράσματα

Τα μοντέλα deep learning που περιγράφηκαν στα προηγούμενα κεφάλαια αποτελούν μεθόδους ταξινόμησης με σκοπό την πρόβλεψη της πιθανότητας απάτης, οικονομικής δυσπραγίας, χρεοκοπίας, οικονομικής αποτυχίας και μη συνεχιζόμενης δραστηριότητας των επιχειρήσεων. Απώτερος στόχος των ερευνών ήταν να επιβεβαιωθεί ότι τα προτεινόμενα μοντέλα αποτελούν βελτιωμένες μεθόδους ταξινόμησης, τα αποτελέσματα των προβλέψεων είναι υψηλής ακρίβειας και αξιοπιστίας και ότι υπερτερούν έναντι των παραδοσιακών μεθόδων με τα οποία συγκρίθηκαν.

Κατά τα πειράματα, παρατηρήθηκε ότι η απόδοση των μοντέλων DL διαφοροποιείται ανάλογα με τον τύπο δεδομένων που χρησιμοποιείται κάθε φορά και ότι η μέγιστη απόδοσή τους επιτυγχάνεται με τον συνδυασμό οικονομικών δεδομένων και δεδομένων κειμένου. Επομένως ενισχύεται η άποψη ότι η γλώσσα που χρησιμοποιείται στην ενότητα MD&A και τις εκθέσεις των ελεγκτών μπορεί να αποκαλύψει χρήσιμες πληροφορίες για την πρόβλεψη.

Ειδικότερα, τα πειράματα έδειξαν ότι τα μοντέλα DL έχουν την ικανότητα να επεξεργάζονται μεγάλα σύνολα δεδομένων, να ενσωματώνουν και να εξάγουν αποτελέσματα από διαφοροποιημένα δεδομένα και να επιτυγχάνουν υψηλές αποδόσεις ταξινόμησης και χαμηλά ποσοστά σφάλματος τύπου I και II. Αντιθέτως, πειράματα με τη χρήση μόνο αριθμητικών δεδομένων (Μαί κ.ά., 2019) έδειξαν ότι τα μοντέλα SVM και RF είναι εξίσου ικανά για την πρόβλεψη και δεν υπάρχει λόγος χρήσης των μοντέλων DL.

Επίσης διαπιστώθηκε ότι η εφαρμογή μεθόδων εξισορρόπησης δεδομένων, επιλογής χαρακτηριστικών και επεξεργασίας φυσικής γλώσσας βελτιώνουν την αποτελεσματικότητα και την προβλεπτική ικανότητα των μοντέλων.

Ο εντοπισμός των μεταβλητών εισόδου που επηρεάζουν το αποτέλεσμα της πρόβλεψης είναι ουσιαστικής σημασίας, καθώς βοηθάει στην κατανόηση των εσωτερικών διαδικασιών του μοντέλου, στην παροχή πληροφοριών για τον τρόπο βελτίωσης των μοντέλων και ουσιαστικά στην αντιμετώπιση του προβλήματος της ερμηνευσιμότητας. Επίσης η δυνατότητα επιλογής μόνο των σημαντικών μεταβλητών προσφέρει σημαντικά οφέλη, καθώς βοηθάει στη μείωση του χρόνου εκπαίδευσης και

χρήσης των μοντέλων, τη μείωση των απαιτήσεων μέτρησης και αποθήκευσης αλλά και τη βελτίωση της απόδοσής τους.

Σύμφωνα με τα παραπάνω, επιβεβαιώνεται η υπεροχή των μοντέλων DL έναντι των παραδοσιακών μοντέλων αναφοράς με τα οποία συγκρίθηκαν. Τα προτεινόμενα μοντέλα αποτελούν ισχυρά εργαλεία πρόβλεψης και μπορούν να χρησιμοποιηθούν για τον σκοπό για τον οποίο αναπτύχθηκαν με επιτυχία. Είναι υψηλής ακρίβειας, αξιόπιστα και ικανά να παρέχουν υποστήριξη στις διαδικασίες λήψης αποφάσεων των υπευθύνων της διοίκησης και κάθε ενδιαφερόμενου μέρους για την επίτευξη των στόχων τους.

6.2 Μελλοντικές προοπτικές (Περιορισμοί και μελλοντικές επεκτάσεις)

Παρόλο που η μέθοδος deep learning αποτελεί μια πολλά υποσχόμενη μέθοδο, όλες οι έρευνες συγκλίνουν στην άποψη ότι υπάρχουν αρκετοί περιορισμοί που χρήζουν περαιτέρω διερεύνησης. Τα ζητήματα στα οποία κυρίως αναφέρονται αφορούν τα δεδομένα, τις μεταβλητές, την αρχιτεκτονική των μοντέλων DL, τα πεδία χρήσης τους καθώς και την πρακτική εφαρμογή τους.

Αναλυτικότερα, αυτό που προτείνουν οι ερευνητές είναι να μελετηθεί η απόδοση της πρόβλεψης των μοντέλων με την επιλογή και τον συνδυασμό περισσότερων τύπων δεδομένων, δηλαδή τόσο ποσοτικών δεδομένων όσο και δεδομένων κειμένου και μάλιστα προερχόμενων από διαφορετικές πηγές πληροφόρησης π.χ. από χρηματιστήρια, στατιστικές υπηρεσίες, ανακοινώσεις εισηγμένων εταιρειών, ρεπορτάζ ειδήσεων ή από το περιεχόμενο που δημιουργείται από τους χρήστες και όχι μόνο από τις ετήσιες οικονομικές εκθέσεις.

Ειδικότερα, αρκετοί ερευνητές αναγνωρίζουν τη σημασία συμπερίληψης δεδομένων κειμένου στην απόδοση των μοντέλων και αναφέρουν (Aljawazneh κ.ά., 2021; Craja κ.ά., 2020; Mai κ.ά., 2019; Matin κ.ά., 2019; Xiuguo & Shengyong, 2022) ότι μέχρι σήμερα, ελάχιστες έρευνες μελέτησαν την απόδοση της πρόβλεψης με τον συνδυασμό διαφορετικών τύπων δεδομένων οι οποίες μάλιστα έδωσαν ικανοποιητικά αποτελέσματα.

Επίσης, όσον αφορά τα ποσοτικά δεδομένα, υποστηρίζουν ότι, εκτός από τις χρηματοοικονομικές μεταβλητές, θα πρέπει να λαμβάνονται υπόψη οπωσδήποτε και μη χρηματοοικονομικές μεταβλητές, δηλαδή μεταβλητές που σχετίζονται με την εταιρική διακυβέρνηση καθώς επίσης και μακροοικονομικοί δείκτες π.χ. δείκτες που σχετίζονται με την οικονομική ανάπτυξη και ύφεση (ΑΕΠ) (Alexandropoulos κ.ά., 2019; Chi & Chu, 2021; Jan, 2021a; Jang κ.ά., 2021).

Οι περισσότεροι ερευνητές επισημαίνουν ότι ο χρονικός ορίζοντας συλλογής των δεδομένων θα πρέπει να διευρυνθεί, καθώς θα συμβάλει στην περαιτέρω βελτίωση της πρόβλεψης των μοντέλων (Li κ.ά., 2021; Xiuguo & Shengyong, 2022).

Επιπλέον αναφέρονται στο ζήτημα της αντιμετώπισης του προβλήματος των ανισόροπων δεδομένων και προτείνουν την ανάπτυξη νέων μεθόδων εξισορρόπησης δεδομένων και τη διερεύνηση των ιδανικών αναλογιών μεταξύ πραγματικών και συνθετικών δεδομένων για την ορθότερη και αποτελεσματικότερη πρόβλεψη (Aljawazneh κ.ά., 2021; Hosaka, 2019).

Επίσης, όσον αφορά τις μεταβλητές των δεδομένων, θα πρέπει αυτές να είναι κατάλληλες και να προσαρμόζονται ανάλογα με το προφίλ των επιχειρήσεων, το

περιβάλλον, τα πρότυπα, τους κανονισμούς και τις οικονομικές συνθήκες που επικρατούν σε κάθε χώρα ξεχωριστά, ώστε να αναπτύσσονται μοντέλα πιο αξιόπιστα (Alam κ.ά., 2021; Chi & Chu, 2021; Jan, 2021a; Jan, 2021b; Jang κ.ά., 2021).

Επίσης σχολιάζουν τη σημασία μελέτης του βαθμού επιρροής κάθε μεταβλητής στο αποτέλεσμα της πρόβλεψης ανάλογα με το μέγεθος, τον τύπο της επιχείρησης ή ακόμα και τον τόπο εγκατάστασής της (Hosaka, 2019; Jang κ.ά., 2021) και τον εντοπισμό των ισχυρότερων δεδομένου ότι ο εντοπισμός τους θα φέρει πολλαπλά οφέλη τόσο στην κατασκευή και τη βελτίωση των μοντέλων όσο και στην κατανόηση των εσωτερικών διεργασιών των μοντέλων και την ερμηνεία των αποφάσεων που λαμβάνουν (Hosaka, 2019).

Επίσης προτείνουν οι μελλοντικές έρευνες να ασχοληθούν με τη βελτίωση των αρχιτεκτονικών των προτεινόμενων μοντέλων, να μελετήσουν τη συμπεριφορά και την απόδοση και άλλων αλγορίθμων DL και να στραφούν στην ανάπτυξη υβριδικών μοντέλων συνδυάζοντας διαφορετικούς τύπους νευρωνικών δικτύων, ώστε να εκμεταλλευτούν τα πλεονεκτήματα κάθε τύπου νευρωνικού δικτύου μετριάζοντας ταυτόχρονα τα μειονεκτήματά τους, προκειμένου να αναπτυχθούν βελτιωμένα μοντέλα ικανά να αντιμετωπίσουν αποτελεσματικά οποιοδήποτε πρόβλημα ταξινόμησης.

Τέλος, αναφέρονται στη διερεύνηση της πιθανότητας εφαρμογής των μοντέλων και σε άλλα οικονομικά ζητήματα, όπως συγχωνεύσεις, εταιρικές εξαγορές που απασχολούν συχνά τα ενδιαφερόμενα μέρη (Alam κ.ά., 2021) και επισημαίνουν την αναγκαιότητα ανάπτυξης εφαρμογών φιλικών προς τους χρήστες, ώστε να είναι εφικτή η πρακτική εφαρμογή τους χωρίς την απαίτηση εξειδικευμένων γνώσεων (Vochozka, Vrbka, & Suler, 2020).

7

Βιβλιογραφία

- Alam, N., Gao, J., & Jones, S. (2021). Corporate failure prediction: An evaluation of deep learning vs discrete hazard models. *Journal of International Financial Markets, Institutions & Money*, 75.
- Alexandropoulos, S.-A. N., Aridas, C. K., Kotsiantis, S. B., & Vrahatis, M. N. (2019). A Deep Dense Neural Network for Bankruptcy Prediction. *Engineering Applications of Neural Networks. EANN 2019. CCIS, 1000*, σσ. 435-444.
- Ali, Z. (2019, Ιανουάριος 6). *A simple Word2vec tutorial*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://medium.com/@zafaralibagh6/a-simple-word2vec-tutorial-61e64e38a6a1>
- Aljawazneh, H., Mora, A. M., Garcia-Sanchez, P., & Castillo-Valdivieso, P. A. (2021). Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure. *IEEE Access*, 9, σσ. 97010-97038.
- Alto, V. (2019, Ιούλιος 5). *Neural Networks: parameters, hyperparameters and optimization strategies*. Ανάκτηση Αύγουστος 23, 2022, από <https://towardsdatascience.com/neural-networks-parameters-hyperparameters-and-optimization-strategies-3f0842fac0a5>
- Ankiit. (n.d.). *Word2vec vs BERT*. Ανάκτηση Δεκέμβριος 27, 2022, από <https://medium.com/@ankiit/word2vec-vs-bert-d04ab3ade4c9>
- Ankiit. (χ.χ.). *Word2vec vs BERT*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://medium.com/@ankiit/word2vec-vs-bert-d04ab3ade4c9>
- Brownlee, J. (2021, Ιανουάριος 22). *How to Choose an Activation Function for Deep Learning*. Ανάκτηση Νοέμβριος 6, 2022, από <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>

- Brownlee, J. (2022, Αύγουστος 15). *Difference Between a Batch and an Epoch in a Neural Network*. Ανάκτηση Νοέμβριος 5, 2022, από <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>
- Capitalone. (2021, Οκτώβριος 6). *Understanding TF-IDF for Machine Learning*. Ανάκτηση Δεκέμβριος 22, 2022, από <https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/>
- Chandra, A. (2018, Ιούλιος 24). *McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron*. Ανάκτηση Αύγουστος 10, 2022, από <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>
- Chandran, S. (2020, Ιούλιος 12). *Introduction to Text Representations for Language Processing — Part 1*. Ανάκτηση Δεκέμβριος 21, 2022, από <https://towardsdatascience.com/introduction-to-text-representations-for-language-processing-part-1-dc6e8068b8a4>
- Chi, D.-J., & Chu, C.-C. (2021). Artificial Intelligence in Corporate Sustainability: Using LSTM and GRU for Going Concern Prediction. *Sustainability (Switzerland)*, 13(21).
- Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, 139.
- Dertat, A. (2017, Νοέμβριος 8). *Applied Deep Learning - Part 4: Convolutional Neural Networks*. Ανάκτηση Δεκέμβριος 28, 2022, από <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>
- Devopedia. (2020, Σεπτέμβριος 5). *Word2vec*. Ανάκτηση Δεκέμβριος 27, 2022, από <https://devopedia.org/word2vec>
- Dutta, M. (2021, Ιούλιος 13). *Word2Vec For Word Embeddings -A Beginner's Guide*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://www.analyticsvidhya.com/blog/2021/07/word2vec-for-word-embeddings-a-beginners-guide/>
- Free time learning. (n.d.). *Deep Learning Interview Questions*. Ανάκτηση Δεκέμβριος 4, 2022, από <https://www.freetimelearning.com/software-interview-questions-and-answers.php?What-Is-a-Multi-layer-Perceptron%28MLP%29?&id=4161>
- freeCodeCamp. (2019, Απρίλιος 18). *An intuitive guide to Convolutional Neural Networks*. Ανάκτηση Δεκέμβριος 28, 2022, από <https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. London: MIT Press.

- Grieve, P. (2022, Μάρτιος 02). *Deep learning vs machine learning: What's the difference?* Ανάκτηση Αύγουστος 2, 2022, από <https://www.zendesk.com/blog/machine-learning-and-deep-learning/>
- Gurucharan, M. (2022, Ιούλιος 28). *Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network*. Ανάκτηση Δεκέμβριος 2, 2022, από <https://www.upgrad.com/blog/basic-cnn-architecture/>
- H2O.a. (χ.χ.). *BERT*. Ανάκτηση Δεκέμβριος 27, 2022, από 1: <https://h2o.ai/wiki/bert/>
- Hansen, L. (2022, Σεπτέμβριος 1). *Neural Networks - step by step*. Ανάκτηση Νοέμβριος 6, 2022, από <https://lassehansen.me/post/neural-networks-step-by-step/>
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems With Applications*, 117, σσ. 287-299.
- Jan, C.-L. (2021a). Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry. *Sustainability (Switzerland)*, 13(17).
- Jan, C.-L. (2021b). Using Deep Learning Algorithms for CPAs' Going Concern Prediction. *Information (Switzerland)*, 12(2), σσ. 1-22.
- Jang, Y., Jeong, I., & Cho, Y. K. (2021). Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors. *Engineering, Construction and Architectural Management*, 28(10), σσ. 3282-3298.
- Karani, D. (2018, Σεπτέμβριος 1). *Introduction to Word Embedding and Word2Vec*. Ανάκτηση Δεκέμβριος 23, 2022, από <https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>
- Kirkos, E. (2012). Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review*, 43, σσ. 83–123.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), σσ. 995-1003.
- Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: A case study in financial distress prediction. *Information Processing and Management*, 58(5).
- Lutkevich, B. (χ.χ.). *BERT language model*. Ανάκτηση Δεκέμβριος 27, 2022, από <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model>
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274, σσ. 743-758.

- Matin, R., Hansen , C., Hansen, C., & Molgaard, P. (2019). Predicting distresses using deep learning of text segments in annual reports. *Expert Systems With Applications*, 132, σσ. 199-208.
- Paperspace. (2020). *Convergence*. Ανάκτηση Νοέμβριος 6, 2022, από <https://machine-learning.paperspace.com/wiki/convergence>
- Phi, M. (2018, Σεπτέμβριος 24). *Illustrated Guide to LSTM's and GRU's: A step by step explanation*. Ανάκτηση Νοέμβριος 29, 2022, από <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>
- ProjectPro. (2022, Νοέμβριος 24). *BERT NLP Model Explained for Complete Beginners*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://www.projectpro.io/article/bert-nlp-model-explained/558>
- Radhakrishnan, P. (2017, Αύγουστος 9). *What are Hyperparameters ? and How to tune the Hyperparameters in a Deep Neural Network?* Ανάκτηση Νοέμβριος 6, 2022, από <https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a>
- Saxena, S. (2021, Μάρτιος 16). *Introduction to Long Short Term Memory (LSTM)*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/>
- Saxena, S. (2022, Οκτώβριος 14). *Activation Function in Neural Networks*. Ανάκτηση Νοέμβριος 6, 2022, από <https://towardsai.net/p/l/activation-function-in-neural-networks>
- Seraydarian, L. (2022, Αύγουστος 18). *What Is a Confusion Matrix in Machine Learning?* Ανάκτηση Νοέμβριος 29, 2022, από <https://plat.ai/blog/confusion-matrix-in-machine-learning/>
- Simplilearn. (2022). *Deep Learning In 5 Minutes*. Ανάκτηση Αύγουστος 2, 2022, από <https://www.youtube.com/watch?v=6M5VXKLF4D4>
- Singhal, G. (2020, Σεπτέμβριος 9). *Introduction to LSTM Units in RNN*. Ανάκτηση Δεκέμβριος 26, 2022, από <https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn>
- Smiti, S., & Soui, M. (2020). Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE. *Information Systems Frontiers*, 22(5), σσ. 1067-1083.
- Sun, T., & Vasarhelyi, M. A. (2018). Embracing Textual Data Analytics in Auditing with Deep Learning. *International Journal of Digital Accounting Research*, 18, σσ. 49-67.

- Vochozka, M., Vrbka, J., & Suler, P. (2020). Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM. *Sustainability (Switzerland)*, 12(18).
- Wikipedia. (2022, Νοέμβριος 3). *Bag-of-words model*. Ανάκτηση Δεκέμβριος 21, 2022, από https://en.wikipedia.org/wiki/Bag-of-words_model
- Wikipedia. (2022, Δεκέμβριος 25). *BERT (language model)*. Ανάκτηση Δεκέμβριος 28, 2022, από [https://en.wikipedia.org/wiki/BERT_\(language_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))
- Wikipedia. (2022, Νοέμβριος 25). *Convolutional neural network*. Ανάκτηση Νοέμβριος 29, 2022, από https://en.wikipedia.org/wiki/Convolutional_neural_network
- Wikipedia. (2022, Δεκέμβριος 8). *Long short-term memory*. Ανάκτηση Δεκέμβριος 26, 2022, από https://en.wikipedia.org/wiki/Long_short-term_memory
- Wikipedia. (2022, Σεπτέμβριος 8). *Multilayer perceptron*. Ανάκτηση Δεκέμβριος 4, 2022, από https://en.wikipedia.org/wiki/Multilayer_perceptron
- Wikipedia. (2022, Δεκέμβριος 24). *Word2vec*. Ανάκτηση Δεκέμβριος 27, 2022, από <https://en.wikipedia.org/wiki/Word2vec>
- Xiuguo, W., & Shengyong, D. (2022). An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies using Deep Learning. *IEEE Access*, 10, σσ. 22516-22532.
- Αργυράκης, Π. (2001). *Νευρωνικά Δίκτυα και Εφαρμογές* (Τόμ. Β'). Πάτρα: ΕΑΠ.
- Κερανού, Ε. (2000). *Τεχνητή Νοημοσύνη και Έμπειρα Συστήματα* (Τόμ. Α'). Πάτρα: ΕΑΠ.
- Κύρκος, Ε. Γ. (2015). *Επιχειρηματική Ευφροσύνη & Εξόρυξη Δεδομένων*. Kallipos, Open Academic Editions.
- Παπαδάκης, Β. Μ. (2016). *Στρατηγική των Επιχειρήσεων: Ελληνική και Διεθνής εμπειρία* (7η εκδ., Τόμ. ΤΟΜΟΣ Α' ΘΕΩΡΙΑ). Αθήνα: Εκδόσεις Ε. ΜΠΕΝΟΥ.
- Τσακλάγκανος, Α. Α., & Σπαθής, Χ. Θ. (2017). *Ελεγκτική* (γ' εκδ.). Θεσσαλονίκη: Αφοι Κυριακίδη Εκδόσεις Α.Ε.