



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

**ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ & ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**

ΤΙΤΛΟΣ ΕΡΓΑΣΙΑΣ:
**Ανάλυση δεδομένων τραπεζικού τομέα
και τεχνικές μηχανικής μάθησης.**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΤΟΥ ΦΟΙΤΗΤΗ:
MARTIN PEKA

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΔΡ ΙΩΣΗΦ ΒΙΚΤΩΡΑΤΟΣ

ΘΕΣΣΑΛΟΝΙΚΗ, ΜΑΪΟΣ 2023

**Ανάλυση δεδομένων τραπεζικού τομέα
και τεχνικές μηχανικής μάθησης.**

Η πτυχιακή εγκρίθηκε από:

1. ΔΡ ΒΙΚΤΩΡΑΤΟΣ ΙΩΣΗΦ
2. ΔΡ ΒΑΣΙΛΕΙΑΔΗΣ ΑΧΙΛΛΕΑΣ
3. ΔΡ ΝΟΤΟΠΟΥΛΟΣ ΠΑΝΑΓΙΩΤΗΣ

Υπεύθυνη Δήλωση

Με το παρόν κείμενο βεβαιώνω ότι ο κάτωθι υπογράφων είμαι συγγραφέας της παρούσης πτυχιακής εργασίας, η οποία εκπονήθηκε στο πλαίσιο των απαιτήσεων του προγράμματος σπουδών του Τμήματος Λογιστικής & Πληροφοριακών συστημάτων του Διεθνούς Πανεπιστημίου Ελλάδος και η οποία παραδόθηκε, μετά από έγκριση του επιβλέποντα καθηγητή μου, σε έντυπη και ψηφιακή μορφή στη Γραμματεία του Τμήματος. Επίσης δηλώνω πως οι πηγές που χρησιμοποίησα (βιβλιογραφία, αρθρογραφία, ή πηγές από το διαδίκτυο), για την υποστήριξη των υποθέσεων της μελέτης και της ερευνάς μου, είναι πλήρως συμβατή με τα ακολουθούμενα επιστημονικά πρότυπα και, επιπλέον, αναφέρεται, υπό μορφή αναφοράς-παραπομπής, στο κείμενο της παρούσης εργασίας καθώς επίσης κάθε άλλη πηγή που χρησιμοποιήθηκε (έστω και μόνο για κατανόηση) αναφέρεται μαζί με τις υπόλοιπες πηγές στο κεφάλαιο της βιβλιογραφίας. Το αυτό ισχύει για τη χρήση δευτερογενών δεδομένων (πινάκων, διαγραμμάτων και εικόνων), ιδεών και λέξεων, τα οποία και αναφέρονται είτε ακριβώς όπως υπάρχουν στις πηγές είτε μεθερμηνεύονται από εμένα.

ΕΠΩΝΥΜΟ	ΡΕΚΑ
ΟΝΟΜΑ	MARTIN
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ	144/19
ΗΜΕΡΟΜΗΝΙΑ	27/06/2023

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Ιωσήφ Βικτωράτο, αναπληρωτής καθηγητής στο Διεθνές Πανεπιστήμιο Ελλάδος – τμήμα Λογιστικής & Πληροφοριακών Συστημάτων, για την καθοδήγηση που μου προσέφερε και το χρόνο που διέθεσε δίνοντάς μου χρήσιμες συμβουλές και οδηγίες για την ολοκλήρωση της πτυχιακής μου εργασίας. Στο ίδιο πλαίσιο ευγνωμοσύνης, θα ήθελα να ευχαριστήσω τους καθηγητές της τριμελούς επιτροπής Αχιλλέα Βασιλειάδη και Νοτόπουλο Παναγιώτη καθώς και όλους τους καθηγητές του Τμήματος Λογιστικής & Πληροφοριακών Συστημάτων για τη συμβολή τους στην επιστημονική μου συγκρότηση στα χρόνια της φοίτησής μου στο Τμήμα. Τέλος, η εργασία είναι αφιερωμένη στην μητέρα μου.

Περίληψη

Η παρούσα ερευνητική πτυχιακή εργασία αφορά τον πειραματισμό με ένα σύνολο τραπεζικών **δεδομένων** για την πρόβλεψη του όσο το δυνατόν μικρότερου σφάλματος στην ανάλυσή τους, ώστε να γίνει η καλύτερη δυνατή πρόβλεψη δανείου. Αρχικά αναλύεται ο τραπεζικός τομέας και τα είδη **τραπεζών** καθώς και ο ρόλος τους στην κοινωνία και στην ευρύτερη οικονομία. Στην συνέχεια παρουσιάζονται τα είδη των **δανείων** που οι τράπεζες προσφέρουν και οι τρόποι με τους οποίους αυτά λειτουργούν. Σημαντικό κομμάτι της εργασίας καθώς και ο λόγος για τον οποίο διεξάγεται η έρευνα είναι η αντιμετώπιση του προβλήματος των **κινδύνων** που οι τράπεζες αντιμετωπίζουν, τα είδη των οποίων αναλύονται στην συνέχεια. Ύστερα, γίνεται λόγος για την **διαχείριση** του κινδύνου, των τεχνικών και της διαδικασίας της, καθώς επίσης γίνεται σύνδεση της αντιμετώπισης κινδύνων με την ανάλυση δεδομένων. Έτσι, επεξηγείται η **ανάλυση** δεδομένων και μερικές από τις πιο χρησιμοποιημένες τεχνικές με τις οποίες αυτή διεξάγεται. Έπειτα, παρουσιάζονται ερευνητικά άρθρα παρόμοιας φύσης με την **έρευνα** της εργασίας και γίνεται σύγκριση μεταξύ τους. Τέλος γίνεται επεξήγηση της διαδικασίας της έρευνας, των αποτελεσμάτων της και σχολιασμός αυτών, ολοκληρώνοντας με τα συμπεράσματα και τις πιθανές μελλοντικές βελτιώσεις.

Λέξεις Κλειδιά:

1. Δεδομένα
2. Τράπεζες
3. Δάνεια
4. Κίνδυνοι
5. Διαχείριση
6. Ανάλυση
7. Έρευνα

Abstract

The present research thesis concerns the experimentation with a set of banking data for predicting the smallest possible error in their analysis in order to achieve the best possible loan prediction. Initially, the banking sector is analyzed, including the types of banks and their role in society and the broader economy. Then, the types of loans offered by banks and the ways in which they operate are presented. A significant part of the work, as well as the reason for conducting the research, is addressing the problem of risks that banks face, which are further analyzed. Subsequently, the focus is on risk management, its techniques, and the process involved, as well as the connection between risk management and data analysis. Thus, data analysis is explained, along with some of the most commonly used techniques for conducting it. Then, research articles of similar nature to the thesis research are presented, and a comparison is made among them. Finally, the research process, its results, and their discussion are explained, concluding with the findings and possible future improvements.

Keywords:

1. Data
2. Banks
3. Loans
4. Risks
5. Management
6. Analysis
7. Research

Περιεχόμενα

Εισαγωγή:	13
1. Ο τραπεζικός τομέας:	15
1.1. Είδη τραπεζών.	15
1.2. Παροχή δανείων.	20
1.3. Είδη δανείων.....	21
1.4. Θετικές και αρνητικές επιπτώσεις του δανεισμού.	23
2. Οι κίνδυνοι και ο τραπεζικός τομέας.	25
2.1. Γενικά για τα είδη κινδύνων.....	25
2.2. Οι κίνδυνοι του τραπεζικού τομέα.....	26
2.2.1. Πιστωτικός κίνδυνος.	26
2.2.2. Κίνδυνος επιτοκίου.	27
2.2.3. Κίνδυνος ρευστότητας.....	27
2.2.4. Κίνδυνος Αγοράς.	28
2.2.5. Λειτουργικός κίνδυνος.	28
2.2.6. Άλλα είδη κινδύνου.....	28
3. Διαχείριση κινδύνων:	30
3.1. Διαδικασία διαχείρισης κινδύνου.	31
3.2. Τεχνικές διαχείρισης κινδύνου.	32
3.2.1. Η μέθοδος Value at Risk.	32
3.2.2. Ιστορική προσομοίωση.....	33
3.2.3. Η μέθοδος Monte Carlo.....	33
3.2.4. Back testing & Scaling.	34
4. Τεχνικές μηχανικής μάθησης και τραπεζικός τομέας:	35
4.1. Η συμβολή της Ανάλυσης Δεδομένων.....	35
4.2. Σωστή διαχείριση της Ανάλυσης Δεδομένων.....	37
4.3. Τεχνικές μηχανικής μάθησης.	39
4.3.1. Γραμμική Παλινδρόμηση (Linear Regression).....	39
4.3.2. Παλινδρόμηση με Δένδρα Αποφάσεων (Decision Trees).	40
4.3.3. Παλινδρόμηση Τυχαίου Δάσους (Random Forest Regression).	42
4.3.4. Παλινδρόμηση XG Boost & Light G Boost.	42
4.3.5. Παλινδρόμηση Cat Boost.	44
4.3.6. Νευρωνικά Δίκτυα (Neural Networks).....	44
4.3.7. Παλινδρόμηση με GRU & LSTM.....	46
5. Βιβλιογραφική επισκόπηση:	49
5.1. Murthy et al.	49

5.2.	Farjana & Mamun.....	49
5.3.	Tejaswini et al.....	50
5.4.	Kadam et al.....	51
5.5.	Gautam & Nidhi.	51
5.6.	Dutta P.....	52
5.7.	Sharma H.....	52
5.8.	Anand, Velu & Whig.	53
5.9.	Majumdar et al.....	54
5.10.	Supriya et al.....	54
6.	Ερευνητική μέθοδος & Αποτελέσματα Έρευνας.....	56
6.1.	Διαδικασία και μέθοδος εξαγωγής της έρευνας.	56
6.2.	Αποτελέσματα έρευνας & Σχολιασμός.....	57
6.2.1.	Προετοιμασία των δεδομένων του Training Data Set.....	57
6.2.2.	Προετοιμασία των δεδομένων του testing data set.	61
6.2.3.	Γραμμική Παλινδρόμηση.	65
6.2.4.	Παλινδρόμηση Τυχαίου Δάσους.....	67
6.2.5.	Παλινδρόμηση XG Boost & Light G Boost.	68
6.2.6.	Παλινδρόμηση με δέντρα αποφάσεων.	70
6.2.7.	Παλινδρόμηση Cat Boost.	71
6.2.8.	Παλινδρόμηση με νευρωνικά δίκτυα.	73
6.2.9.	Παλινδρόμηση με GRU & LSTM.....	75
6.3.	Σύνοψη και σύγκριση των αποτελεσμάτων από τους κώδικες.....	77
	Σχολιασμός & Συμπέρασμα:.....	79
	Ελληνική Βιβλιογραφία:.....	81
	Ξένη Βιβλιογραφία:.....	81

Λίστα Πινάκων:

Αριθμός πίνακα	Τίτλος Πίνακα	Σελίδα
6.2.1.1.	Προετοιμασία των δεδομένων Training	57 - 58
6.2.1.2.	Αποτελέσματα training code	59 - 60
6.2.2.1.	Προετοιμασία των δεδομένων testing	61
6.2.2.2.	Αποτελέσματα testing code	62 - 64
6.2.3.1.	Linear Regression	65
6.2.4.1.	Random Forest Regression	66
6.2.5.1.	XG Boost Regression	67
6.2.5.2.	Light G Boost Regression	68 - 69
6.2.6.1.	Decision Trees Regression	70
6.2.7.1.	Cat Boost Regression	71
6.2.8.1.	Neural Networks Regression	69 - 70
6.2.9.1.	GRU & LSTM Regression	74 - 75
6.3.1.1.	Αποτελέσματα	76

Λίστα Εικόνων:

Αριθμός εικόνας	Τίτλος Εικόνας	Σελίδα
1.2.1.	FISCO (Tamplin, 2023)	20
4.3.1.1.	Γραμμική Παλινδρόμηση, (Kanade, 2023).	39
4.3.2.1.	Δένδρα Αποφάσεων (Decision Trees).	40
4.3.2.2.	Παλινδρόμηση με δένδρα αποφάσεων (Steorts R., 2017).	40
4.3.3.1.	Random Forest Regression (Chakure A, 2023).	41
4.3.4.1.	LG Boost (Amazon.com, 2023)	42
4.3.4.2.	LGBM (Datascience.eu, 2019).	42
4.3.5.1.	CatBoost (catboost.ai, 2018)	43
4.3.6.1.	Neural Network Layers (Chen J., 2023).	44
4.3.6.2.	Συνολικό Σήμα Εισόδου (Κύρκος Ε., 2015).	45
4.3.6.3.	Σιγμοειδής Συνάρτηση (Κύρκος Ε., 2015).	45
4.3.7.1.	LSTM & GRU (Phi M., 2018).	46
6.1.1.1.	Διάγραμμα Gantt.	56
6.2.8.1.	MAE & EPOCHS – NEURAL NETWORKS	74
6.2.9.1.	MAE & EPOCHS – LSTM & GRU	76

Συντομογραφίες:

Συντομογραφία	Έννοια
DTI	Debt - to - income ratio (Αναλογία χρέους προς εισόδημα)
VaR	Value at Risk
XG Boost	Extreme Gradient Boosting
LGBM ή LG Boost	Lingt Gradient Boosting Machine
GRU	Long Short – Term Memory Networks – Δίκτυα Μακροπρόθεσμης και Βραχυπρόθεσμης μνήμης
LSTM	Gated Recurrent Unit ή Μονάδα Επαναλαμβανόμενης Θύρας
ML	Machine Learning ή Μηχανική Μάθηση.
DM	Data Mining ή Εξόρυξη Δεδομένων

Εισαγωγή:

Γίνεται λόγος για μία ερευνητική πτυχιακή εργασία χαρακτηριζόμενη ως ποσοτική έρευνα. Τα δεδομένα της έρευνας παρέχονται από την διαδικτυακή βιβλιοθήκη συνόλων δεδομένων, την Kaggle.com. Το θέμα μελέτης αφορά την πρόβλεψη δανείου μέσω της ανάλυσης των συνόλων δεδομένων με τεχνικές μηχανικής μάθησης. Σκοπός της εργασίας είναι η περεταίρω έρευνα και κατανόηση του τρόπου με τον οποίο ο τραπεζικός τομέας μπορεί να αυτοματοποιήσει τις διάφορες εργασίες του και να μπορέσει να κάνει σωστή διαχείριση των δεδομένων που έχει στην κατοχή του. Το εργαλείο που χρησιμοποιείται για την ανάλυση είναι η γλώσσα προγραμματισμού της Python.

Το ερευνητικό ερώτημα της εργασίας είναι η εύρεση τεχνικής μηχανικής μάθησης με το ελάχιστο σφάλμα και η θεωρητική κατανόηση της σημασίας αυτού. Αρχικός περιορισμός της πτυχιακής ήταν το περιορισμένο χρονικό περιθώριο εφόσον η εργασία άρχισε τον Φεβρουάριο του 2023 λόγω απουσίας του φοιτητή σε πρόγραμμα Erasmus + και ολοκληρώθηκε τον Ιούνιο του ίδιου έτους. Βέβαια, αυτό με την σωστή διαχείριση των εργασιών αντιμετωπίστηκε σε αρκετά ικανοποιητικό βαθμό με αποτέλεσμα την μελέτη και τον πειραματισμό με διάφορες τεχνικές μηχανικής μάθησης. Αδυναμία της εργασίας είναι το γεγονός ότι ήταν η πρώτη επαφή του φοιτητή – συγγραφέα με τεχνικές μηχανικής μάθησης και το προγραμματιστικό περιβάλλον της Python, κάτι που αντιμετωπίστηκε μέσω της συστηματικής θεωρητικής έρευνας και της καθοδήγησης και στήριξης εκ μέρους του επιβλέποντα καθηγητή.

Γενικά, η έρευνα αποτελεί ικανοποιητικά αναλυτική, όμως υπάρχουν πολλά κομμάτια της τα οποία μπορούν να βελτιωθούν. Για παράδειγμα η χρήση περισσότερων τεχνικών μηχανικής μάθησης ή η πιο λεπτομερής και μαθηματική επεξήγηση όλων των κεφαλαίων, όπως του τρόπου που οι τεχνικές λειτουργούν στο περιβάλλον της Python ή η μαθηματική έκφραση του τρόπου λειτουργίας του τραπεζικού συστήματος και των στατιστικών μεθόδων που αυτό χρησιμοποιεί. Σημαντικό είναι να αναφέρουμε ως αδυναμία πως και ο τομέας της οικονομίας αλλά και ο τομέας της τεχνολογίας που συνδυάζονται στην παρούσα εργασία, αποτελούν διαρκώς μεταβαλλόμενα και ραγδαία εξελισσόμενα μεγέθη, γεγονός που καθιστά την ερευνητική εργασία ευάλωτη στο πέρασμα του χρόνου. Έτσι, καθίσταται απαραίτητη η διαρκής ενημέρωση μίας τέτοιου είδους έρευνας.

Η πρόοδος της τεχνολογίας και οι αλλαγές στην παγκόσμια κοινωνία έχουν καταστήσει απαραίτητη την ύπαρξη ορισμένων χρηματοπιστωτικών ιδρυμάτων προκειμένου να πραγματοποιήσουν τις πολλές συναλλαγματικές διαδικασίες που σχετίζονται με τη διαχείριση, τη διανομή, την έκδοση και την αποθήκευση χρημάτων. Η λειτουργία του δανείου, στην οποία θα βασίσουμε την ανάλυση δεδομένων σε αυτή τη μελέτη, αποτελεί βασική λειτουργία των τραπεζών. Η τράπεζα χρεώνει συνήθως ένα συγκεκριμένο ποσό - επιτόκιο στα δάνεια, μέσω του οποίου βγάζει κέρδος. Οι λειτουργίες των τραπεζικών συστημάτων και οι κοινωνικές απαιτήσεις που προκύπτουν χρησιμοποιούνται για την κατηγοριοποίησή τους σε διαφορετικές ομάδες που αναλύονται στην συνέχεια.

Οι τράπεζες πολλές φορές για να αυτοματοποιήσουν τις εργασίες τους και να μειώσουν τον κίνδυνο λαθών χρησιμοποιούν την ανάλυση των δεδομένων –

πληροφοριών που έχουν στην διάθεσή τους από τους πελάτες τους. Χωρίς τα κατάλληλα εργαλεία, όπως τεχνικές μηχανικής μάθησης, δεν μπορεί να πραγματοποιηθεί ανάλυση δεδομένων. Αυτές οι τεχνικές είναι χρήσιμες για τον καθαρισμό και την οργάνωση δεδομένων, τα οποία με τη σειρά τους επιτρέπουν την καλή διαχείριση και την πρόβλεψη κινδύνου.

Στο πρώτο κεφάλαιο της εργασίας γίνεται εισαγωγή στην έννοια του χρήματος και προσδιορίζεται ο ρόλος του τραπεζικού τομέα στην οικονομία. Τα χρηματοπιστωτικά ιδρύματα – τράπεζες χωρίζονται σε μερικά από τα πιο συνήθη και συχνά εμφανιζόμενα στην οικονομία είδη τους, η λειτουργία των οποίων επεξηγείται αναλυτικά. Τονίζεται η έννοια των δανείων ως βασική υπηρεσία - παροχή από τις τράπεζες και ο τρόπος χορήγησής τους. Τα είδη των δανείων και οι επιπτώσεις τους αναλύονται λεπτομερώς ολοκληρώνοντας το κεφάλαιο.

Το δεύτερο κεφάλαιο εστιάζει στην έννοια του κινδύνου στον τραπεζικό τομέα. Αρχικά εξηγείται η ο όρος η γενική περιγραφή των κατηγοριών. Στην συνέχεια, παρουσιάζονται οι πιο συχνόι τύποι κινδύνων που καλούνται να αντιμετωπίσουν οι τράπεζες και τα χαρακτηριστικά τους. Το τρίτο κεφάλαιο συνδέεται στενά με το δεύτερο καθώς παρουσιάζει την έννοια της διαχείρισης των κινδύνων. Μετά την γενική περιγραφή της διαχείρισης αναφέρεται η διαδικασία που οι τράπεζες ακολουθούν για να προστατευτούν.

Η σημασία της ανάλυσης δεδομένων μέσω τεχνικών μηχανικής μάθησης για την διαχείριση των κινδύνων είναι ο βασικός πυρήνας του τετάρτου κεφαλαίου. Η σημασία της ανάλυσης δεδομένων στις τράπεζες αποτελεί την αρχή του κεφαλαίου που βοηθά στην κατανόηση της σχέσης μεταξύ μηχανικής μάθησης και τραπεζικών διαδικασιών. Μάλιστα τονίζεται η σωστή διαχείριση των δεδομένων, καθώς και τα βήματα που συνήθως ακολουθούνται για μία τέτοια έρευνα. Ολοκληρώνοντας, γίνεται η θεωρητική ανάλυση και μαθηματική επεξήγηση των τεχνικών που χρησιμοποιούνται στην έρευνα για την ανάλυση των συνόλων δεδομένων.

Το πέμπτο κεφάλαιο προετοιμάζει τον αναγνώστη για την φύση της εργασίας, παρουσιάζοντας παρόμοια έργα. Συγκεκριμένα παρουσιάζονται δέκα επιστημονικά άρθρα και η γενική περιγραφή της έρευνάς τους. Με αυτό το έναυσμα, γίνεται σύγκριση και συνένωση των θετικών που η παρούσα έρευνα παρέχει ως επιπρόσθετο υλικό σε αυτά. Μετά από την είσοδο στον τρόπο σκέψης της έρευνας ακολουθεί το έκτο κεφάλαιο, το οποίο περιέχει την ουσία της εργασίας. Αρχικά, περιγράφεται αναλυτικά η διαδικασία για την οποία έγινε εν μέρει αναφορά στην αρχή της εισαγωγής. Στην συνέχεια παρουσιάζεται η λειτουργία του κώδικα και των τεχνικών μηχανικής μάθησης, καθώς και τα αποτελέσματα τα οποία έδωσαν. Στο τέλος του κεφαλαίου, γίνεται σύνοψη και σύγκριση των αποτελεσμάτων.

Κλείνοντας την εργασία παρουσιάζονται τα συμπεράσματα που προκύπτουν από την συγγραφή. Στο τέλος παρατίθενται οι βιβλιογραφικές αναφορές και οι τιμές που εισάχθηκαν στους κώδικες για να γίνουν πειράματα και να δώσουν τα εκάστοτε αποτελέσματα.

1. Ο τραπεζικός τομέας.

Η ανάγκη των ανθρώπων για να μπορέσουν να διεκπεραιώσουν τις μεταξύ τους συναλλαγές οδήγησε στην δημιουργία χρήματος, το οποίο αποτελεί οτιδήποτε είναι κοινά αποδεκτό ως μέσο ανταλλαγής αγαθών ή υπηρεσιών. Με την εξέλιξη της τεχνολογίας και του τρόπου που λειτουργούν οι κοινωνίες παγκοσμίως, το χρήμα πλέον έχει εξελιχθεί σε διάφορες μορφές όπως είναι οι καταθέσεις όψεως στις τράπεζες και οτιδήποτε άλλο που έχει παρόμοια λειτουργία με το χρήμα. Οπότε, η διαχείριση, διανομή, έκδοση και αποταμίευση του χρήματος, απαιτεί την ύπαρξη ορισμένων χρηματοπιστωτικών ιδρυμάτων για την εκτέλεση αυτών των διαδικασιών τα οποία ονομάζονται τράπεζες.

Ο τραπεζικός τομέας αποτελεί βασικό μέσο συναλλαγών στην σημερινή οικονομία όπως η λήψη καταθέσεων και χορήγηση δανείων. Η λειτουργία του δανεισμού, βάση της οποίας θα κάνουμε την ανάλυση δεδομένων στην παρούσα εργασία, συνήθως διεκπεραιώνεται μέσω του έντοκου δανεισμού. Δηλαδή, συνήθως η τράπεζα δανείζει με ένα συγκεκριμένο επιτόκιο. Το ποσό που χρεώνει η τράπεζα σε έναν δανειολήπτη ονομάζεται επιτόκιο και εκφράζεται ως ποσοστό της αρχής ή του ποσού του δανείου. Το ετήσιο επιτόκιο (Annual Percentage Rate - APR), το οποίο αντιπροσωπεύει το επιτόκιο ενός δανείου, αναφέρεται συνήθως σε ετήσια βάση. (Banton, 2023). Το δανειζόμενο ποσό αποτελείται είτε από κεφάλαια της ίδιας της τράπεζας ή από καταθέσεις πελατών της. (Κιόχος & Παπανικολάου, 2011).

1.1. Είδη τραπεζών.

Τα τραπεζικά συστήματα χωρίζονται σε διάφορες κατηγορίες ανάλογα με τις λειτουργίες τους και τις ανάγκες που εμφανίζονται στην κοινωνία. Οι περισσότερες χώρες ορίζουν το κράτος – κυβέρνηση ή την κεντρική τράπεζα ως ελεγκτή των υπολοίπων τραπεζών. Στην παρούσα εργασία θα χωρίσουμε τις τράπεζες στις εξής 8 κατηγορίες:

Κεντρικές τράπεζες. Αναφερόμαστε σε ιδρύματα που διαδραματίζουν πολύ σημαντικό ρόλο για την οικονομία μίας χώρας καθώς διαχειρίζονται την δημιουργία και προσφορά χρήματος σε αυτήν και γενικότερα την νομισματική πολιτική. Μάλιστα, μέσω ρύθμισης των επιτοκίων μπορούν να ελέγχουν άλλα χρηματοπιστωτικά ιδρύματα όπως τις εμπορικές τράπεζες, αλλά επίσης παίζουν πολύ σημαντικό ρόλο στην ισορροπία των τιμών καθώς καλούνται υπεύθυνες για την διατήρησή της. Αυτό σημαίνει πως φροντίζουν για την διαχείριση κινδύνων στον χρηματοπιστωτικό τομέα, όπως το να μην υπάρχουν υπερβολικά επίπεδα πληθωρισμού σε μία οικονομία μέσω διαφόρων τεχνικών όπως η προσαρμογή των επιτοκίων. (Segal, 2022).

Η κεντρική τράπεζα είναι ένας κρατικός οργανισμός που επιβλέπει την προσφορά χρήματος ή τη συνολική ποσότητα χρήματος σε κυκλοφορία, καθώς και τη διαχείριση των νομισμάτων ενός έθνους ή ενός συνόλου των νομισμάτων των εθνών. Ορισμένα κράτη έχουν επίσης νόμους που απαιτούν από τις κεντρικές τράπεζες να ενθαρρύνουν την πλήρη απασχόληση. Ως μέρος της νομισματικής της πολιτικής, το κύριο μέσο οποιασδήποτε κεντρικής τράπεζας είναι ο καθορισμός των επιτοκίων, μερικές φορές γνωστά ως «κόστος χρήματος». Οι εμπορικές τράπεζες δεν είναι κεντρικές τράπεζες. Μια κεντρική τράπεζα δεν μπορεί να χρησιμοποιηθεί από ένα άτομο για να

δημιουργήσει λογαριασμό ή να ζητήσει δάνειο, και ως δημόσιο ίδρυμα, δεν βασίζεται στο κέρδος.

Λειτουργεί ως τράπεζα για τις υπόλοιπες εμπορικές τράπεζες, πράγμα που είναι ο τρόπος με τον οποίο επηρεάζει τη ροή πιστώσεων και χρήματος σε όλη την οικονομία για να διατηρήσει τις τιμές σταθερές, μέσω της αύξησης ή μείωσης των επιτοκίων δανεισμού προς τις εμπορικές. Μια κεντρική τράπεζα είναι μια πηγή πίστωσης για τις εμπορικές τράπεζες, συχνά για να ανταποκριθεί σε εξαιρετικά επείγουσες απαιτήσεις. Οι εμπορικές τράπεζες πρέπει να παρέχουν εξασφαλίσεις, όπως ένα πολύτιμο περιουσιακό στοιχείο, όπως κρατικό ομόλογο ή επιχειρηματικό ομόλογο, στην κεντρική τράπεζα προκειμένου να δανειστούν χρήματα από αυτήν. Η εξασφάλιση χρησιμεύει ως διαβεβαίωση ότι ο δανειολήπτης θα επιστρέψει το δάνειο.

Οι εμπορικές τράπεζες μπορεί να παρέχουν μακροπρόθεσμα δάνεια έναντι βραχυπρόθεσμων καταθέσεων, γεγονός που μπορεί να έχει ως αποτέλεσμα ζητήματα «ρευστότητας» ή αδυναμία ταχείας μετατροπής κεφαλαίων σε μετρητά, ενώ έχουν αρκετά για την εξόφληση των χρεών. Μια κεντρική τράπεζα μπορεί να ενεργήσει σε αυτή την κατάσταση ως «δανειστής έσχατης ανάγκης». Αυτό προάγει τη σταθερότητα του χρηματοπιστωτικού συστήματος. Εκτός από τη νομισματική πολιτική, οι κεντρικές τράπεζες μπορεί να είναι υπεύθυνες για μια ποικιλία καθηκόντων. Διατηρούν συχνά τη σωστή λειτουργία των συστημάτων πληρωμών για τις τράπεζες και τα χρηματοοικονομικά προϊόντα που διαπραγματεύονται, εκτυπώνουν τραπεζογραμμάτια και νομίσματα, διαχειρίζονται συναλλαγματικά αποθέματα και βοηθούν στην εκπαίδευση του κοινού σχετικά με την οικονομία. (European Central Bank, 2015).

Εμπορικές Τράπεζες. Οι εμπορικές τράπεζες είναι τα ιδρύματα τα στα οποία οι πολίτες καταθέτουν τα χρήματά τους. Ο τρόπος με τον οποίο οι εμπορικές τράπεζες κερδίζουν χρήματα είναι μέσω χρεώσεων για τις διάφορες υπηρεσίες που παρέχουν. Μία από αυτές είναι η υπηρεσία των δανείων όπως είναι τα στεγαστικά, τα επιχειρηματικά και τα προσωπικά δάνεια. Για να χορηγήσει μία τράπεζα ένα δάνειο χρησιμοποιεί είτε τα διαθέσιμα από τις καταθέσεις των υπολοίπων πελατών της είτε από τα δικά της διαθέσιμα. Το κέρδος βγαίνει μέσω των τόκων στα δάνεια το οποίο είναι μεγαλύτερο από το επιτόκιο καταθέσεων ώστε να υπάρχει κέρδος. Βέβαια, οι εμπορικές τράπεζες κινούνται πάντα στα πλαίσια που τους θέτουν οι κεντρικές τράπεζες, για παράδειγμα υποχρεούνται να κρατήσουν συγκεκριμένο μέρος των καταθέσεών τους στην κεντρική τράπεζα, σε περίπτωση έκτακτης ανάληψης από την πλειοψηφία των πελατών των εμπορικών. (Kagan, 2023).

Μια εμπορική τράπεζα ή χρηματοπιστωτικό ίδρυμα προσφέρει στους πελάτες της υπηρεσίες όπως δάνεια, λογαριασμούς ταμειοτηρίου, τραπεζικές αναλήψεις κ.λπ. Αυτοί οι οργανισμοί παράγουν έσοδα χορηγώντας δάνεια σε άτομα και εισπράττοντας τόκους για αυτά τα χρέη. Τα επιχειρηματικά δάνεια, τα δάνεια αυτοκινήτων, τα στεγαστικά δάνεια, τα προσωπικά δάνεια και τα φοιτητικά δάνεια είναι μερικά από τα είδη δανείων που μπορούν να ληφθούν από μια εμπορική τράπεζα.

Αυτά τα δάνεια χρηματοδοτούνται από καταθέσεις που γίνονται από τους πελάτες τους σε διάφορα είδη λογαριασμών. Οι καταθέσεις χρησιμεύουν ως κεφάλαιο για δάνεια που γίνονται από αυτούς. Οι εμπορικές τράπεζες διαδραματίζουν κρίσιμο ρόλο στην οικονομία ενός έθνους, καθώς διευκολύνουν τη δημιουργία κεφαλαίου,

πιστώσεων και ρευστότητας της αγοράς. Αυτές οι τράπεζες βρίσκονται συχνά σε αστικές περιοχές, αν και όλο και περισσότερες από αυτές λειτουργούν τώρα στο διαδίκτυο όπως θα αναλύσουμε στην συνέχεια.

Έτσι, λοιπόν, οι εμπορικές τράπεζες έχουν σημαντικό ρόλο στην οικονομία. Παράλληλα με το να προσφέρουν στους καταναλωτές μια κρίσιμη υπηρεσία, συμβάλλουν επίσης στην κεφαλαιοποίηση και τη ρευστότητα της αγοράς. Οι εμπορικές τράπεζες διατηρούν ρευστότητα δανείζοντας τα χρήματα που καταθέτουν οι πελάτες τους στους λογαριασμούς τους. Έτσι, συμβάλλουν στην έκδοση πιστώσεων, οι οποίες αυξάνουν την παραγωγή, την απασχόληση και τις καταναλωτικές δαπάνες ενώ παράλληλα τονώνουν την οικονομία. (Kagan, 2023).

Τράπεζες Λιανικής. Η συγκεκριμένη κατηγορία αφορά τράπεζες οι οποίες παρέχουν ποικίλες χρηματοοικονομικές υπηρεσίες σε φυσικά πρόσωπα ή και νομικά πρόσωπα μικρού μεγέθους. Μερικές από αυτές τις υπηρεσίες είναι η αποταμίευση, ο δανεισμός, οι πιστωτικές κάρτες ή ακόμα και συμβουλές στον χρηματοοικονομικό τομέα. (Majaski, 2021). Οι χρηματοοικονομικές υπηρεσίες που επικεντρώνονται σε μεμονωμένους πελάτες και όχι σε μεγάλες επιχειρήσεις αναφέρονται ως λιανική τραπεζική, συχνά γνωστή ως προσωπική τραπεζική (Personal Banking) ή καταναλωτική τραπεζική (Consumer Banking). Τα υψηλά επίπεδα εξυπηρέτησης πελατών είναι απαραίτητα για την απασχόληση στη λιανική τραπεζική, καθώς αυτά τα ιδρύματα παρέχουν στο ευρύ κοινό αγαθά όπως λογαριασμούς ταμειυτηρίου και χρεωστικές κάρτες.

Τα χρήματα που καταθέτει ένα άτομο σε μια τράπεζα λιανικής χρηματοδοτούν δάνεια και υποθήκες για άλλους πελάτες, καθιστώντας τη λιανική τραπεζική μια 'κοινή προσπάθεια' της κοινωνίας. Οι τράπεζες λιανικής παράγουν χρήματα χρεώνοντας τόκους σε αυτά τα δάνεια. Τέλος, υπάρχει ασφάλεια λόγω δέσμευσης της τράπεζας, που υποχρεούνται να εξασφαλίσει ότι δεν θα χάσει κάποιος πελάτης τα χρήματά του εάν απευθυνθεί στην τράπεζα για να κάνει ανάληψη των αποταμιεύσεών του και η τράπεζα δεν έχει τα χρήματα. (Girardin, 2023).

Ταμειυτήρια και Δανειακές Ενώσεις. Αυτά τα ιδρύματα δέχονται καταθέσεις από τους πελάτες τους και τα χρησιμοποιούν για παροχή δανείων στην αγορά ακινήτων, Ως οργανισμοί ανήκουν στους πελάτες – επενδυτές τους και έχουν κυρίως τοπικό χαρακτήρα. Η ένωση ταμειυτηρίου και δανείων είναι ένας τύπος χρηματοπιστωτικού οργανισμού που συνήθως ελέγχεται και ανήκει στους πελάτες της. Οι οργανισμοί αποταμιεύσεων και δανείων είναι σε θέση να παρέχουν στεγαστικά δάνεια και άλλα χρηματοοικονομικά προϊόντα σε καταναλωτές που ενδέχεται να μην έχουν πρόσβαση σε αυτά χωρίς τη βοήθεια αυτής της συμφωνίας. Έχουν συχνά τοπικά γραφεία και είναι εξουσιοδοτημένοι από το νόμο να επικεντρώνουν το μεγαλύτερο μέρος των δανείων τους στη χρηματοδότηση κατοικιών. (Bankrate, 2022).

Ο κύριος στόχος μιας τέτοιας ένωσης είναι να υποστηρίξει την ανάπτυξη της οικιστικής και κατασκευαστικής βιομηχανίας της γειτονιάς. Τα ταμειυτήρια και οι Δανειακές Ενώσεις είναι παρόμοιες με τις πιστωτικές ενώσεις καθώς επικεντρώνονται στην κοινότητα και διοικούνται και ανήκουν στα άτομα που προορίζονται να εξυπηρετήσουν. Τυπικά μη κερδοσκοπικοί οργανισμοί, αυτά τα ιδρύματα επιτρέπουν στα μέλη να ψηφίζουν και να επηρεάζουν τους διοικητικούς στόχους του οργανισμού.

Ακόμα είναι εφικτό για μια ένωση να στραφεί σε μια δομή που βασίζεται σε μετοχές και να αρχίσει να διαπραγματεύεται δημόσια, κάτι που θα έκανε τα μέλη μέτοχους και θα παραιτηθούν από τη διαχειριστική εξουσία.

Οπότε, οι τράπεζες παρέχουν μια σειρά από χρηματοοικονομικές υπηρεσίες και αγαθά, όπως λογαριασμοί επιταγών και ταμειευτηρίου και πολυάριθμα είδη δανείων. Μια ομάδα με ισχυρή αίσθηση της κοινότητας, μια τέτοια ομάδα, μπορεί να ιδρύσει μια πιστωτική ένωση, η οποία είναι ένας χρηματοοικονομικός οργανισμός που λειτουργεί από κοινού. Οι οργανισμοί αυτοί δίνουν πιο περιορισμένη έμφαση, εστιάζοντας στη χρηματοδότηση για επενδύσεις και ακίνητα. Μπορεί να έχουν λιγότερα οικονομικά αγαθά να προσφέρουν, αλλά πιθανότατα έχουν καλύτερους όρους δανεισμού και επιτόκια. (Bond, 2023).

Επενδυτικές Τράπεζες. Οι επενδυτικές τράπεζες αποτελούν χρηματοπιστωτικά ιδρύματα τα οποία προσφέρουν υπηρεσίες κυρίως σε νομικά πρόσωπα μεγάλου μεγέθους. Προσφέρουν συμβουλευτικές υπηρεσίες όπως συγχωνεύσεις, εξαγορές ή και επενδύσεις, (Csiszar, 2023). Μερικές από τις πιο γνωστές επενδυτικές τράπεζες παγκοσμίως είναι οι: JPMorgan Chase, Goldman Sachs, Morgan Stanley, Citigroup, Bank of America, Credit Suisse, and Deutsche Bank.

Το συμβουλευτικό τμήμα μιας επενδυτικής τράπεζας λαμβάνει αμοιβή για τις υπηρεσίες της. Με βάση το πόσο καλά τα πάει το τμήμα συναλλαγών στην αγορά, κερδίζονται προμήθειες. Πολλοί έχουν επίσης τμήματα λιανικής τραπεζικής που δημιουργούν έσοδα δανείζοντας χρήματα σε ιδιώτες και εταιρείες. Οι υπάλληλοι της επενδυτικής τράπεζας μπορούν να ασκήσουν επαγγέλματα ως οικονομικοί σύμβουλοι, έμποροι ή πωλητές. Αν και ένα επάγγελμα επενδυτικής τραπεζικής μπορεί να είναι ικανοποιητικό, περιλαμβάνει επίσης πολλές ώρες και πολύ άγχος.

Προτού οι τίτλοι προσφερθούν προς πώληση, η επενδυτική τράπεζα είναι υπεύθυνη για την εξέταση των οικονομικών αρχείων μιας εταιρείας ως προς την ορθότητα και τη διάδοση ενός ενημερωτικού δελτίου στους επενδυτές που περιγράφει διεξοδικά την προσφορά. Οι εταιρείες, τα συνταξιοδοτικά ταμεία, άλλα χρηματοπιστωτικά ιδρύματα, οι κυβερνήσεις και τα αμοιβαία κεφάλαια αντιστάθμισης κινδύνου είναι μόνο μερικά παραδείγματα της πελατείας των επενδυτικών τραπεζών. (Hargrave, 2022).

Πιστωτικές Ενώσεις. Σε αυτήν την περίπτωση γίνεται λόγος για χρηματοπιστωτικούς οργανισμούς οι οποίοι δημιουργούνται από ορισμένες κοινότητες (φυσικών ή νομικών προσώπων). Η λειτουργία τους είναι να αποταμιεύουν χρήματα από τα μέλη των κοινοτήτων με σκοπό να προσφέρουν στα ίδια τα μέλη χρηματοοικονομικές υπηρεσίες όπως η παροχή δανείων. Παραδείγματα πιστωτικών ενώσεων αποτελούν οι γερμανικές πιστωτικές εταιρείες Schulze-Delitzsch και Raiffeisen οι οποίες αποτέλεσαν πρωτοπόρους στην δημιουργία τέτοιων ενώσεων. (McKillop & Wilson, 2011). Προκειμένου να διασφαλιστεί η εκπροσώπηση των συμφερόντων τους, τα μέλη εκλέγουν ένα διοικητικό συμβούλιο για να επιβλέπει την πιστωτική ένωση.

Παρέχοντας προϊόντα που είναι ανταγωνιστικά και έχουν χαμηλότερα επιτόκια και προμήθειες από αυτά που προσφέρουν οι κερδοσκοπικές τράπεζες, οι πιστωτικές ενώσεις επιδιώκουν να εξυπηρετήσουν τα μέλη τους. Ενώ οι πιστωτικές ενώσεις

εισπράττουν προμήθειες λογαριασμού και τόκους όπως κάνουν οι τράπεζες, επενδύουν ξανά τα κέρδη τους στα στοιχεία που προσφέρουν αντί να τα δίνουν στους μετόχους τους όπως κάνουν οι τράπεζες. Πρέπει να πληρούνται τα κριτήρια που θέτει η ομάδα για την δημιουργία λογαριασμού, να γίνει αίτηση για πιστωτική κάρτα ή να ληφθεί δάνειο, καθώς μια πιστωτική ένωση ανήκει στα μέλη της. Διάφορες πιστωτικές ενώσεις έχουν διαφορετικά πρότυπα.

Η κύρια διάκριση μεταξύ πιστωτικών ενώσεων και τραπεζών είναι ότι οι πρώτες είναι μη κερδοσκοπικά χρηματοπιστωτικά ιδρύματα μόνο για μέλη, ενώ οι δεύτερες είναι δημόσια προσβάσιμες κερδοσκοπικές επιχειρήσεις. Επιπλέον, οι πιστωτικές ενώσεις είναι γνωστές για την παροχή ανώτερης προσωπικής εξυπηρέτησης πελατών στις φυσικές εγκαταστάσεις των υποκαταστημάτων τους, καθώς ιδρύθηκαν συχνά για να εξυπηρετούν συγκεκριμένες περιοχές, πόλεις ή εταιρείες.

Τα οφέλη των πιστωτικών ενώσεων περιλαμβάνουν:

1. Μειωμένα επιτόκια πιστωτικών καρτών και δανείων.
2. Υψηλότερα επιτόκια καταθέσεων.
3. Μειωμένο κόστος.
4. Εξατομικευμένη φροντίδα πελατών.

Μειονεκτήματα των πιστωτικών ενώσεων:

1. Πρέπει να είστε μέλος.
2. Λιγότερες επιλογές προϊόντων.
3. Περιορισμένη διακλάδωση. (White, 2022).

Τοπικές τράπεζες. Φροντίζουν για την ανάπτυξη εταιριών συγκεκριμένων κοινοτήτων. Συνήθως υποστηρίζουν τοπικές επιχειρήσεις μέσω της παροχής δανείων τα οποία μπορεί να μην δικαζόντουσαν σε άλλα χρηματοπιστωτικά ιδρύματα. (Csiszar, 2023). Γενικά, η φράση «κοινοτική ή τοπική τράπεζα» είναι λέξη καθομιλουμένη χωρίς σαφή ορισμό. Ο όρος συχνά εφαρμόζεται σε τράπεζες με μικρό αριθμό τοποθεσιών που απευθύνονται κυρίως σε γειτονικούς κατοίκους και τοπικές επιχειρήσεις. Συνήθως οι «τοπικές τράπεζες» παρουσιάζονται ως χρηματοπιστωτικά ιδρύματα με ενοποιημένα περιουσιακά στοιχεία μικρότερα από 10 δισεκατομμύρια δολάρια.

Ενώ τεράστιες τράπεζες είναι δημόσια εισηγμένες, οι μικρές τράπεζες είναι πιο πιθανό να έχουν τοπικούς ιδιοκτήτες. Συνεπώς, η διαχείριση της κοινοτικής τράπεζας εξαιρείται από την εξωτερική εποπτεία των μετόχων. Σε αντίθεση με έναν μεγαλύτερο οργανισμό με στενότερους δεσμούς με τις χρηματοπιστωτικές αγορές, οι κοινοτικές τράπεζες μπορεί να αξιολογούν διαφορετικά τα διάφορα συμφέροντα των μετόχων, των πελατών, του προσωπικού και της τοπικής κοινότητας. Συχνά επικεντρώνονται στην εκτέλεση παραδοσιακών καθηκόντων όπως η λήψη καταθέσεων και η εκταμίευση πιστωτικών γραμμών, στεγαστικών δανείων και επιχειρηματικών δανείων. Παρά το γεγονός ότι επικεντρώνονται στην εξυπηρέτηση τοπικών πελατών, ορισμένοι έχουν αναπτύξει λειτουργίες διαδικτυακής τραπεζικής που τους επιτρέπουν να προσεγγίσουν ένα μεγαλύτερο κοινό. (Kurt, 2022).

Online τράπεζες. Τέτοια χρηματοπιστωτικά ιδρύματα δεν έχουν φυσικό κατάστημα και λειτουργούν αποκλειστικά μέσω του ίντερνετ, κάτι που τις επιτρέπει να προσφέρουν μεγαλύτερο επιτόκιο καθώς έχουν λιγότερα έξοδα. Προσφέρουν παρόμοιες υπηρεσίες με τις παραδοσιακές τράπεζες όπως λογαριασμοί καταθέσεων, δάνεια και πιστωτικές κάρτες. Οι πελάτες – χρήστες των ηλεκτρονικών τους υπηρεσιών μπορούν να έχουν πρόσβαση σε αυτές από οπουδήποτε στον κόσμο αρκεί να έχουν πρόσβαση στο διαδίκτυο. (Frankenfield, 2023).

Απαιτείται υπολογιστής, τηλέφωνο ή άλλη συσκευή, καθώς και σύνδεση στο Διαδίκτυο και τράπεζα ή χρεωστική κάρτα για τη χρήση της ηλεκτρονικής τραπεζικής. Πρέπει να δημιουργηθεί ένας προσωπικός κωδικός για εγγραφή στην υπηρεσία της ηλεκτρονικής τραπεζικής της τράπεζας. Οι παρεχόμενες διαδικτυακές τραπεζικές υπηρεσίες διαφέρουν ανά ίδρυμα. Η πλειοψηφία των τραπεζών συχνά παρέχει βασικές υπηρεσίες, όπως μεταφορές και πληρωμή λογαριασμών. Μέσω διαδικτυακών τραπεζικών πλατφόρμων, ορισμένες τράπεζες επιτρέπουν πλέον στους καταναλωτές να υποβάλουν αίτηση για πιστωτικές κάρτες και να δημιουργήσουν νέους λογαριασμούς.

Άλλες εργασίες μπορεί να περιλαμβάνουν την τοποθέτηση διακοπής πληρωμής σε μια επιταγή, την παραγγελία επιταγών ή την αναφορά αλλαγής διεύθυνσης. Ένα από τα κύρια οφέλη της διαδικτυακής τραπεζικής είναι η ευκολία. Απλές τραπεζικές εργασίες, όπως η πληρωμή λογαριασμών και οι μεταφορές χρημάτων μεταξύ λογαριασμών, είναι εύκολο να ολοκληρωθούν κάθε μέρα της εβδομάδας, οποιαδήποτε ώρα της ημέρας ή της νύχτας. Μερικοί άνθρωποι προτιμούν τις αλληλεπιδράσεις πρόσωπο με πρόσωπο με έναν ταμεία, καθώς η χρήση ηλεκτρονικών συστημάτων για πρώτη φορά μπορεί να δημιουργήσει δυσκολίες που απαγορεύουν την εκτέλεση συναλλαγών για έναν αρχάριο πελάτη ηλεκτρονικής τραπεζικής. Τέλος, πολλές από αυτές όπως η Revolut δίνουν την δυνατότητα δημιουργίας ηλεκτρονικής κάρτας μίας χρήσης για ασφαλείς ηλεκτρονικές αγορές. (Frankenfield, 2023).

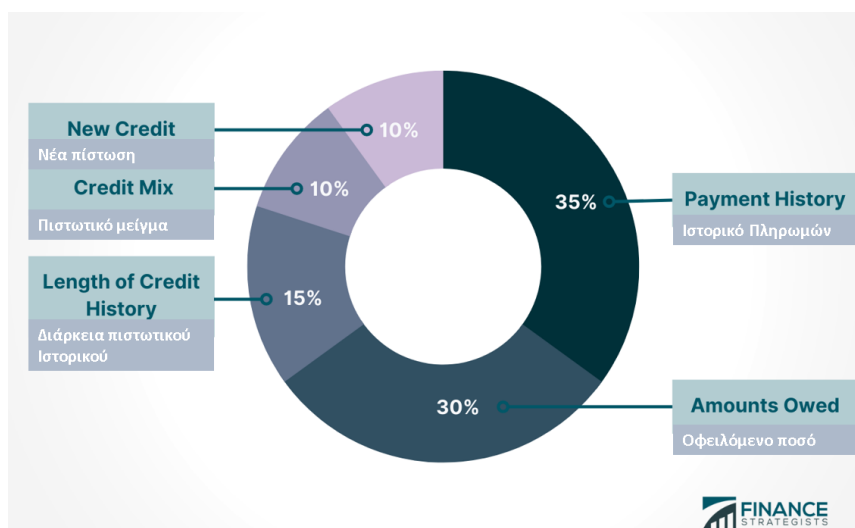
1.2. Παροχή δανείων.

Οι φυσικές καθώς και οι νομικές οντότητες όπως οι επιχειρήσεις κάθε μεγέθους χρησιμοποιούν τα δάνεια ως ένα χρηματοοικονομικό εργαλείο για να πετύχουν στόχους με μεγαλύτερη οικονομική δυνατότητα – κέρδος χάρη σε αυτά. Ένα δάνειο είναι μία συμφωνία μεταξύ του δανειολήπτη και του δανειστή, σύμφωνα με την οποία ο δανειστής παρέχει οικονομική υποστήριξη στον δανειολήπτη. Ο λήπτης του δανείου όμως, υποχρεούνται να αποπληρώσει το δάνειο με ορισμένους τόκους σε μία ορισμένη χρονική περίοδο. Οι λόγοι για τους οποίους μπορεί να λαμβάνει ένα δάνειο μία επιχείρηση μπορεί να διαφέρουν από προσωπικούς όπως η αγορά αυτοκινήτου ή ακινήτου, η κάλυψη δαπανών εκπαίδευσης ή ιατρικής περίθαλψης ή αποπλήρωση ενός χρέους. Πολλές φορές όμως οι λόγοι μπορεί να είναι και επενδυτικοί από επιχειρήσεις κυρίως, όπως η χρηματοδότηση λειτουργιών, η αγορά αποθεμάτων, η αγορά εξοπλισμού ή ακόμα και για επέκταση δραστηριοτήτων. Οι διάφοροι λόγοι οι οποίοι προκύπτουν από τις ανάγκες της κάθε οντότητας, συμβάλουν στην δημιουργία κατηγοριών δανείων.

Για να μπορέσει να γίνει ένα δάνειο αποδεκτό από τον δανειστή απαιτείται ο λήπτης να πληροί ορισμένες προϋποθέσεις. Σημαντική προϋπόθεση θα μπορούσε να είναι η

αναλογία χρέους προς το εισόδημα του δανειολήπτη, όπου βλέπουμε την αναλογία του χρέους προς εισόδημα ή Debt – to – income – ratio (DTI) που αντιπροσωπεύει την διαίρεση των συνολικών μηνιαίων πληρωμών του δανείου με το συνολικό καθαρό μηνιαίο εισόδημα του δανειολήπτη. Οι τράπεζες συνήθως προτιμούν να δανείζουν σε οντότητες οι οποίες έχουν DTI μικρότερο από 36%, αλλά αυτό διαφέρει σε κάθε ίδρυμα. (Treece, 2021).

Πολύ σημαντικό ρόλο για την αποδοχή μίας αίτησης δανείου είναι η πιστωτική ικανότητα ή βαθμολογία ή πιστωτικό σκορ ή Credit Score του εκάστοτε πελάτη της τράπεζας. Το πιστωτικό σκορ είναι μία αριθμητική κλίμακα μεταξύ 330 – 850 και συνήθως σχηματίζεται από πολλά μεγέθη όπως το ιστορικό πληρωμών, το ποσό του ήδη υπάρχοντος χρέους, την διάρκεια του πιστωτικού ιστορικού ή το πιστωτικό μείγμα και την δυνατότητα για νέα πίστωση, που ως παράγοντες διαμορφώνονται από την πολιτική της κάθε τράπεζας. Όσο μεγαλύτερο είναι το συγκεκριμένο σκορ, τόσο μικρότερο είναι το ρίσκο που φέρει ο πελάτης. Το πιο γνωστό μοντέλο για την εύρεση του πιστωτικού σκορ είναι η Βαθμολογία FISCO αλλά υπάρχουν και άλλες όπως το Vantage Score. (Tamplin, 2023). Παρακάτω παρουσιάζεται εικόνα με τους παράγοντες που συμβάλλουν στην τελική βαθμολογία FISCO τους οποίους αναφέραμε, με τα αντίστοιχα βάρη τους σε ποσοστά τους %.



Εικόνα 1.2.1. FISCO (Tamplin, 2023)

Τέλος για να υπάρξει η δυνατότητα έγκρισης ενός δανείου θα πρέπει ο δανειολήπτης να έχει το ανάλογο εισόδημα και απασχόληση. Το μέγεθος του εισοδήματος προσφέρει ασφάλεια στην τράπεζα όσον αφορά για την αποπληρωμή του δανείου. Η δυνατότητα του πελάτη να έχει μία σταθερή απασχόληση αποδεικνύει στην τράπεζα ότι έχει την δυνατότητα να συνεχίσει να αποπληρώνει το δάνειο στο μέλλον. Τα κριτήρια όσον αφορά το εισόδημα διαφέρουν από τράπεζα σε τράπεζα αλλά σίγουρα ένας πελάτης με μεγάλο και σταθερό εισόδημα είναι ευκολότερο να λάβει έγκριση για ένα δάνειο, καθώς αποτελεί μικρότερο ρίσκο για την τράπεζα.

1.3. Είδη δανείων.

Υπάρχουν πολλά είδη δανείων που αντιστοιχούν στους τρόπους με τους οποίους οι τράπεζες διαχειρίζονται τις ανάγκες των πελατών των τραπεζών, ένα από αυτά τα είδη

είναι αυτό των **εξασφαλισμένων δανείων** ή δανείων με εγγύηση (Secured Loans). Τα εξασφαλισμένα δάνεια συμβάλλουν στην απόκτηση αγαθών που μπορούν να κατασχεθούν από την τράπεζα σε περίπτωση που ο δανειολήπτης αδυνατεί να αποπληρώσει το δάνειο, όπως είναι ένα ακίνητο ή όχημα. Στην συγκεκριμένη περίπτωση το ρίσκο που έχει η τράπεζα είναι μικρότερο, οπότε το επιτόκιο του δανείου είναι επίσης μικρότερο.

Άλλη μία κατηγορία είναι εκείνη των **ακάλυπτων δανείων** (Unsecured Loans). Τα συγκεκριμένα δάνεια αφορούν την υποστήριξη των πελατών στην απόκτηση στοιχείων που δε μπορούν να κατασχεθούν από την τράπεζα όπως είναι τα φοιτητικά δάνεια και γενικώς κάθε είδους προσωπικού δανείου της όμοιας κατηγορίας. Οπότε, λόγω αδυναμίας κατάσχεσης του στοιχείου που θέλει να αποκτήσει ο πελάτης χάρη στο δάνειο, υπάρχει μεγάλο ρίσκο για την τράπεζα που σημαίνει ότι επιβάλλει μεγαλύτερα επιτόκια για να εξασφαλίσει την θέση της.

Εκτός από τις παραπάνω δύο μεγάλες κατηγορίες όμως μπορούμε να χωρίσουμε τα δάνεια και σε **Ανανεώσιμα ή Ανακυκλωμένα δάνεια** (Revolving Loans) και σε **Προθεσμιακά Δάνεια** (Term Loans). Η πρώτη κατηγορία που αναφέραμε επιτρέπει στον δανειολήπτη, ανάλογα με το εισόδημά του και την πολιτική της τράπεζας, να δανειζεται μέχρι ένα ορισμένο ποσό. Ο δανειολήπτης μπορεί να μην χρειάζεται όλο το ποσό, οπότε στην ουσία έχει μία πιστωτική γραμμή μέσω της οποίας μπορεί να δανειστεί ανάλογα με τις ανάγκες του μέχρι ένα συγκεκριμένο όριο. Μόλις αποπληρώσει το συγκεκριμένο δάνειο, στην συνέχεια μπορεί να δανειστεί ξανά. Ενώ στην κατηγορία των προθεσμιακών δανείων ο δανειολήπτης δανειζεται ένα ορισμένο ποσό. Συνήθως στην δεύτερη κατηγορία δανείων γίνεται λόγος για απόκτησης κάποιου στοιχείου όπως κάποιο ακίνητο ή αυτοκίνητο ή ακόμα και ένα επιχειρηματικό δάνειο, που σημαίνει ότι η αποπληρωμή εκτελείται με ορισμένο επιτόκιο σε ορισμένο χρονικό διάστημα με τις ανάλογες μηνιαίες δόσεις. (Treece, 2021).

Τα δάνεια χωρίζονται επίσης σε κατηγορίες που αφορούν τις ανάγκες των δανειοληπτών. Παράδειγμα είναι τα **προσωπικά δάνεια**, τα οποία μπορεί να αφορούν καλύψεις ενεργειών όπως ιατρική περίθαλψη, ανακαινίσεις, αγορά υπολογιστών, γάμοι κτλ. Επίσης υπάρχουν δάνεια τα οποία αφορούν αγορά **αυτοκινήτων** αλλά και τα **φοιτητικά δάνεια** τα οποία προορίζονται για να καλύψουν την εκπαίδευση των πελατών. Τα **στεγαστικά δάνεια** είναι άλλη κατηγορία με μεγάλη ζήτηση, τα οποία αφορούν την αγορά ακινήτων και ανάλογα με την περίπτωση υπάρχει ο αντίστοιχος τύπος στεγαστικού δανείου. Σε περίπτωση που μία οντότητα χρειάζεται να αποπληρώσει κάποιο άλλο δάνειο, υπάρχουν τα δάνεια εξυγίανσης οφειλών, τα οποία βοηθούν στο να υπάρχει πλέον μόνο μία πληρωμή και στην εξοικονόμηση χρημάτων μακροπρόθεσμα εάν ο λήπτης πληροί προϋποθέσεις για χαμηλότερες τιμές επιτοκίου.

Ένας άλλος τρόπος δανεισμού είναι τα **πιστωτικά δάνεια** τα οποία συνήθως χρησιμοποιούνται για προσωπικές ανάγκες και τις περισσότερες φορές είναι μικρά και βραχυπρόθεσμα. Σε αυτήν την περίπτωση δανείων ο δανειολήπτης πληρώνει μηνιαίες δόσεις για να λάβει τα χρήματα του δανείου στο τέλος της προκαθορισμένης περιόδου. (Tretina, 2019). Άλλες κατηγορίες είναι τα **δάνεια ημέρας πληρωμής** που διαρκούν μέχρι την ημέρα πληρωμής του επόμενου μισθού και τα δάνεια μικρών επιχειρήσεων για την ανάπτυξή τους. Τα **δάνεια τίτλων** είναι μία αξιοσημείωτη περίπτωση, όπου για

την αγορά τίτλων δεσμεύεται ένα όχημα μέχρι 25% - 50% της αξίας του. Παρόμοια είναι και τα **δάνεια ενεχυροδανειστηρίου** όπου δεσμεύεται ξανά στα ίδια ποσοστά με τα δάνεια τίτλων κάποιο πολύτιμο αντικείμενο αλλά στην συγκεκριμένη περίπτωση χρειάζεται να αποπληρωθεί σε αρκετά σύντομο χρονικό διάστημα, συνήθως μέχρι 30 μέρες. Βέβαια υπάρχουν και άλλα δάνεια όπως **δάνεια σκαφών** για αγορά σκάφους, τα **δάνεια γης** για αγορά εκτάσεων αλλά και τα **δάνεια για αγορά οχημάτων αναψυχής (RV)**. (VanSomeren, 2021). Τέλος, τα **κυβερνητικά δάνεια** είναι μία ακόμη κατηγορία όπου συνήθως μία χώρα δανείζεται χρήματα από μία άλλη και οι τράπεζες διαμεσολαβούν σε αυτήν την συναλλαγή. (Abbas & Pienkowski, 2022).

1.4. Θετικές και αρνητικές επιπτώσεις του δανεισμού.

Η απόκτηση ενός δανείου μπορεί να έχει πολλές θετικές επιπτώσεις για τον πελάτη μίας τράπεζας. Σημαντικό πλεονέκτημα είναι η δυνατότητα του λήπτη να μπορεί να διαχειριστεί την οικονομική του κατάσταση με αρκετά εύκολο τρόπο, καθώς τα επιτόκια και οποιοδήποτε άλλο είδος εξόδου που μπορεί να προκύψει λόγω του δανείου καθορίζεται από την αρχή. Μάλιστα οι πληρωμές για το δάνειο παραμένουν οι ίδιες για τη διάρκεια του δανείου, ιδιαίτερα με δάνεια σταθερού επιτοκίου όπου το επιτόκιο είναι σταθερό για την περίοδο του δανείου. Έτσι, απλοποιείται για τις επιχειρήσεις η διαδικασία των προϋπολογισμών τους και να κάνουν μηνιαίες πληρωμές δανείων. Ακόμα κι αν το δάνειο έχει προσαρμόσιμο επιτόκιο, οι ιδιοκτήτες της εταιρείας μπορούν να χρησιμοποιήσουν ένα απλό υπολογιστικό φύλλο για να εκτιμήσουν τις μελλοντικές πληρωμές σε περίπτωση που τα επιτόκια κυμαίνονται με ασταθή τρόπο.

Στην περίπτωση που ο δανειστής είναι κάποια εταιρία, είναι πολύ σημαντικό το γεγονός ότι η τράπεζα δεν απαιτεί κάποιο κεφάλαιο η τίτλο της εταιρίας ως αντάλλαγμα. Με αυτόν τον τρόπο τα δάνεια αποτελούν πολύ σημαντική πηγή οικονομικής στήριξης για μία εταιρία η οποία αναζητά εξέλιξη και ανάπτυξη μέσω αυτών. Οι τόκοι του δανείου εκπίπτουν της φορολογίας και ο λήπτης στις περισσότερες χώρες, δεν καλείται να πληρώσει φόρους όσον αφορά το δάνειο. (DeMerceau, 2019). Το γνωστό όνομα και η ισχυρή φήμη μιας τράπεζας την βοηθούν να ξεχωρίζει από τις ανταγωνιστικές πηγές δανεισμού. Μια τράπεζα μπορεί να προτιμάται από ορισμένες μικρές επιχειρήσεις λόγω της εδραιωμένης φήμης της και της ασφάλειας που πιστεύουν ότι προσφέρει. (Philps & Campbell, 2023).

Βέβαια, όπως κάθε οικονομική δραστηριότητα τα τραπεζικά δάνεια φέρουν ορισμένα μειονεκτήματα. Η αίτηση συχνά χαρακτηρίζεται από πολύπλοκες διαδικασίες οι οποίες απαιτούν την συμπλήρωση αρκετών εγγράφων, κάτι που μπορεί να αποβεί χρονοβόρο. Οι διαδικασίες όμως δεν αφορούν μόνο την αίτηση, αφορούν και την έγκριση η οποία μπορεί επίσης να καταλάβει αρκετά μεγάλο χρονικό διάστημα. Επίσης, η έγκριση είναι αρκετά δύσκολη και περιέχει πολλές προϋποθέσεις, ιδιαίτερα σε περιόδους κατά τις οποίες η οικονομία μίας χώρας είναι ασταθής και κρύβει κινδύνους για την αποπληρωμή του δανείου.

Οι δόσεις του δανείου ορίζονται ως ένα ορισμένο ποσό που αποπληρώνεται σε ορισμένα τακτά χρονικά διαστήματα βάση της αρχικής συμφωνίας. Οποιαδήποτε αθέτηση αυτής της συμφωνίας μπορεί να οδηγήσει σε παραπάνω έξοδα ή ακόμα και σε κατάσχεση, κάτι που επηρεάζει αρνητικά το πιστωτικό σκορ του δανειολήπτη. Τέλος, ένας αρνητικός παράγοντας που μπορεί να επηρεάσει αρνητικά ειδικά τις μικρές

επιχειρήσεις είναι ότι η αποπληρωμή του δανείου περιλαμβάνει ορισμένους τόκους οι οποίοι μπορεί να είναι πολύ υψηλοί για μικρά ποσά. Αυτό μπορεί να έχει ως αποτέλεσμα μέχρι και την αδυναμία του δανειολήπτη να πετύχει τους επιθυμητούς στόχους μέσω του δανείου. (Tsang, 2023).

2. Οι κίνδυνοι και ο τραπεζικός τομέας.

Σε όλες τις επιχειρηματικές δραστηριότητες υπάρχουν διάφορα είδη κινδύνων ανάλογα με το είδος της επιχείρησης. Υπάρχουν διάφοροι ορισμοί για το τι είναι κίνδυνος όπως η πιθανότητα ζημίας, η αβεβαιότητα για το αποτέλεσμα, η διαφορά του τελικού αποτελέσματος από το αναμενόμενο ή ένα μη αναμενόμενο γεγονός. Γενικότερα ως κίνδυνο ορίζουμε οποιαδήποτε πιθανή και μη προβλέψιμη οικονομική απώλεια που μπορεί να προέλθει από διάφορα ζημιογόνα γεγονότα τα οποία προέρχονται είτε από το εσωτερικό, είτε από το εξωτερικό περιβάλλον της επιχείρησης. Παραδείγματα τέτοιων κινδύνων αποτελούν τα εργατικά ατυχήματα, η καταστροφή εξοπλισμού, η ολική καταστροφή ενός κτηρίου αλλά και μία οικονομική κρίση σε εθνικό ή παγκόσμιο επίπεδο. Βασικό χαρακτηριστικό του κινδύνου είναι η αβεβαιότητα δηλαδή η αδυναμία γνώσης και πρόβλεψης για το τι μπορεί να συμβεί στο μέλλον. Υπάρχουν διάφοροι ορισμοί για το τι είναι κίνδυνος όπως η πιθανότητα ζημίας, η αβεβαιότητα για το αποτέλεσμα, η διαφορά του τελικού αποτελέσματος από το αναμενόμενο ή ένα μη αναμενόμενο γεγονός. (Κιόχος, Παναγόπουλος & Κυρμίζογλου, 2018).

Βάσει της ορολογίας του κινδύνου και όσον αναφέραμε για τον τραπεζικό τομέα, συμπεραίνουμε πως κίνδυνος είναι οποιοδήποτε μη απρόσμενο ενδεχόμενο αρνητικών αποτελεσμάτων. Αυτό αφορά τις επιχειρηματικές ενέργειες μίας τράπεζας που οδήγησαν σε οποιαδήποτε οικονομική ζημιά. Η ζημιά μπορεί να προκύψει από διάφορους παράγοντες, είτε ενδογενείς, είτε εξωγενείς. Πιθανές περιπτώσεις είναι οι απώλειες από δανειακές πράξεις, επενδύσεις, συνθήκες που επικρατούν στην αγορά, αποτυχία ορθής λειτουργίας των υπηρεσιών της τράπεζας και οποιουδήποτε είδους παράγοντες μπορεί να προκαλέσουν οικονομική απώλεια.

Ο τραπεζικός τομέας διαμεσολαβεί ανάμεσα σε πολλές οντότητες με αποτέλεσμα να εκτίθεται σημαντικά σε διάφορα είδη κινδύνων εξαιτίας της φύσης των λειτουργιών που ακολουθεί. Οι κίνδυνοι αυτοί εκτός από επιπτώσεις στην απόδοση των τραπεζών, μπορεί να επηρεάσουν και την εικόνα τους, κάτι που μπορεί να οδηγήσει στην απώλεια πελατών και την αλλοίωση της εικόνας τους. Οι τράπεζες κατέχουν έναν πολύ υπεύθυνο ρόλο στην σημερινή οικονομία καθώς φροντίζουν όχι μόνο για την δικιά τους ορθή λειτουργία και προστασία αλλά επίσης και για την προστασία των πελατών τους. Έτσι, υπάρχει απαίτηση από τους πελάτες για σταθερότητα και για δυνατότητα εκπλήρωσης των οικονομικών ευθυνών με αποτελεσματικό τρόπο, κάτι που μπορεί πολύ εύκολα να αποτύχει λόγω των διαφόρων κινδύνων στους οποίους ο τραπεζικός τομέας εκτίθεται.

2.1. Γενικά για τα είδη κινδύνων.

Αρχικά, είναι αξιοσημείωτο το γεγονός ότι κάθε επιχείρηση αντιμετωπίζει κινδύνους οι οποίοι σχετίζονται με τον κλάδο στον οποίο ανήκει. Οι κίνδυνοι μπορεί να κατηγοριοποιηθούν με βάση το αρνητικό τους αποτέλεσμα σε οποιονδήποτε τομέα και κλάδο, όπως τον πολιτικό, οικονομικό, κοινωνικό, νομικό κτλ. Πριν την λεπτομερή ανάλυση των κινδύνων του τραπεζικού τομέα είναι σημαντικό να γίνουν κατανοητές οι γενικές κατηγορίες με τις οποίες μπορεί να έρθει μία οποιαδήποτε επιχείρηση όπως και οι τράπεζες.

Οι καθαροί κίνδυνοι αποτελούν κατηγορία κατά την οποία μετά την πραγματοποίησή τους έχουν ως αποτέλεσμα ζημία. Παράδειγμα είναι η ολική απώλεια κάποιου υλικού περιουσιακού στοιχείου. Δεν υπάρχει περιθώριο μερικής ζημίας ή περιθώριο για προσδοκία κέρδους από έναν καθαρό κίνδυνο, καθώς επιφέρει είτε μόνο ζημία ή μόνο κέρδος. Οι συγκεκριμένοι κίνδυνοι έχουν την δυνατότητα να υπολογιστούν με την θεωρία των πιθανοτήτων ή με ασφαλιστικά μαθηματικά, άρα μπορούν να ασφαλιστούν. Το αντίθετο αποτελούν οι κερδοσκοπικοί κίνδυνοι, οι οποίοι κρύβουν είτε κέρδος είτε ζημία και δε μπορούν να ασφαλιστούν.

Πολλές φορές λόγω φυσικών καταστροφών ή λόγω ανθρώπινων λαθών μπορεί να προκύψουν **οι στατικοί κίνδυνοι** οι οποίοι είναι στην πλειοψηφία καθαροί και έχουν ζημία. Βέβαια, **οι δυναμικοί κίνδυνοι** που είναι και κερδοσκοπικοί σχετίζονται με μεταβολές κυρίως εξωτερικού περιβάλλοντος όπως είναι οι οικονομικές ή τεχνολογικές. Αναλόγως του επιπέδου του κινδύνου έχουμε τους **γενικούς** που επηρεάζουν μεγάλο κομμάτι της οικονομίας και της κοινωνίας και τους **ειδικούς** που επηρεάζουν ορισμένες οντότητες. (Κιόχος, Παναγόπουλος & Κυρμιζογλου, 2018).

Άλλες κατηγορίες κινδύνων είναι οι **υποκειμενικοί** οι οποίοι βασίζονται σε υποκειμενικές εκτιμήσεις της αβεβαιότητας που μπορεί να κρύβουν ενώ αντιθέτως οι **αντικειμενικοί** προκύπτουν από την σχετική απόκλιση της πραγματικής ζημιάς από την αναμενόμενη. Υπάρχουν πολυάριθμες κατηγορίες κινδύνων οι οποίες μπορεί να ανήκουν σε δύο γενικές κατηγορίες. Οι κατηγορίες αυτές χωρίζονται σε **ενδογενείς** που προκύπτουν από το εσωτερικό περιβάλλον της επιχείρησης, και **εξωγενείς** που είναι αποτέλεσμα γεγονότων που λαμβάνουν χώρα στο εξωτερικό της περιβάλλον. (Ζύγουρας, 2014).

2.2. Οι κίνδυνοι του τραπεζικού τομέα.

Όπως αναφέρθηκε προηγουμένως, ο κάθε τομέας απασχόλησης κρύβει τους δικούς του κινδύνους. Ο τραπεζικός τομέας αποτελεί ζωτικό συστατικό κάθε οικονομίας, καθώς παρέχει υπηρεσίες ως χρηματοοικονομικός διαμεσολαβητής σε ιδιώτες, επιχειρήσεις ή ακόμα και κυβερνήσεις. Όμως, όπως και οι υπόλοιποι τομείς κρύβει ορισμένους κινδύνους. Η αβεβαιότητα είναι αρκετά σημαντικό μέγεθος με το οποίο οι τράπεζες καλούνται να έρθουν αντιμέτωπες καθώς επιτελούν τις διάφορες δανειοδοτικές και επενδυτικές τους δραστηριότητες. Τα είδη των κινδύνων είναι αμέτρητα καθώς τα χαρακτηριστικά τους εξαρτώνται από την κάθε κατάσταση την οποία αντιμετωπίζει η τράπεζα. Στο παρόν υπό κεφάλαιο αναλύονται κάποια από τα βασικά είδη αυτών και τα χαρακτηριστικά τους.

2.2.1. Πιστωτικός κίνδυνος.

Συχνό είδος κινδύνου που αντιμετωπίζουν οι τράπεζες είναι ο πιστωτικός κίνδυνος ο οποίος σχετίζεται με την αποπληρωμή των δανείων. Η συγκεκριμένη κατηγορία προκύπτει διότι υπάρχει αβεβαιότητα για το αν οι δανειολήπτες θα συνεχίσουν να είναι συνεπείς στην αποπληρωμή ή αν θα συμβεί κάτι και θα σταματήσουν να αποπληρώνουν τα δάνειά τους. Στην περίπτωση που οι λήπτες του δανείου αδυνατούν να ανταπεξέλθουν αυτό συνεπάγεται ζημία για την τράπεζα. Έτσι, είναι πολύ σημαντικό οι τράπεζες να μπορούν να υπολογίσουν την πιστοληπτική δυνατότητα των δανειοληπτών και τις πιθανότητες πιθανής ζημίας που μπορεί να προκύψει. Για να γίνει αυτό οι τράπεζες χρησιμοποιούν διάφορα μοντέλα και στατιστικές τεχνικές ώστε να

μπορούν να ελέγξουν την συγκεκριμένη ομάδα κινδύνου. (Caouette, Altman & Narayanan, 1998).

Για την μέτρηση του πιστωτικού κινδύνου στον προσωπικό δανεισμό η κάθε τράπεζα αναλύει δεδομένα βάσει του εισοδήματος σε σχέση με τις υπόλοιπες υποχρεώσεις τους και το πιστωτικό τους ιστορικό. Βέβαια όταν πρόκειται για δάνειο εμπορικού χαρακτήρα, τότε δίνεται βαρύτητα και στον κλάδο δραστηριοποίησης του πελάτη, της οικονομίας καθώς και την παραγωγικότητα της ίδιας της επιχείρησης και της οικονομικής της δυνατότητας. (Peterdy, 2023).

2.2.2. Κίνδυνος επιτοκίου.

Ο κίνδυνος του επιτοκίου οφείλεται στις μεταβολές των τιμών των επιτοκίων. Αυτό συμβαίνει καθώς όταν ανεβαίνουν οι γενικές τιμές των επιτοκίων οι τράπεζες αναγκάζονται να ανεβάσουν κι εκείνες τις τιμές των καταθέσεων και των δανείων. Το πρόβλημα είναι στον άμεσο επηρεασμό των επιτοκίων καταθέσεων αλλά στην σταδιακή αύξηση των επιτοκίων χορηγήσεων. Αυτό σημαίνει ότι βραχυχρόνια η τράπεζα θα δίνει περισσότερα από όσα παίρνει πίσω με αποτέλεσμα να οδηγήσει σε μείωση του κέρδους. (Paul & W.Zhu, 2020).

Σημαντική είναι επίσης η αλληλοεπίδραση ομολόγων και επιτοκίων καθώς εάν αυξηθούν τα επιτόκια, τότε η αξία ενός ομολόγου μειώνεται. Άρα με σκοπό την έκδοση νέων ομολόγων μειώνεται η αξία όσων έχουν ήδη εκδοθεί. Το αντίστροφο συμβαίνει εάν πέσουν οι τιμές των επιτοκίων. Για να υπολογίσουν τον συγκεκριμένο κίνδυνο οι τράπεζες χρησιμοποιούν την διάρκεια ενός τίτλου σταθερού εισοδήματος. Βέβαια το κάθε ομόλογο έχει διαφορετική ανταπόκριση στις μεταβολές καθώς όσο πιο μακροπρόθεσμο, τόσο πιο εύκολα χάνει ή κερδίζει αξία. (Chen, 2022).

2.2.3. Κίνδυνος ρευστότητας.

Γίνεται αναφορά στην αδυναμία της τράπεζας να ανταπεξέλθει στις τρέχουσες ταμειακές υποχρεώσεις της, δηλαδή η έλλειψη ρευστότητας στα ταμειακά της διαθέσιμα. Αυτό μπορεί να συμβεί λόγω της υπερβολικής εξάρτησης από βραχυπρόθεσμες πηγές χρηματοδότησης και από μη ρευστοποιήσιμα περιουσιακά στοιχεία. Με αυτόν τον τρόπο η τράπεζα χάνει την δυνατότητά της να δώσει χρήματα στους πελάτες της. Επίσης μία τράπεζα μπορεί να αντιμετωπίσει αναντιστοιχία μεταξύ ενεργητικού και παθητικού λόγω επένδυσης μεγάλου μέρους του ενεργητικού σε μακροπρόθεσμα δάνεια και υποχρεώσεις. Οπότε, μία τράπεζα καλείται να διατηρήσει ισορροπία μεταξύ ιδίων κεφαλαίων και των υποχρεώσεών της, υπολογίζοντας την αναλογία δανείων προς καταθέσεις που αποτελεί τον τρόπο με τον οποίο οι τράπεζες υπολογίζουν τον κίνδυνο ρευστότητας. (CFI, 2023).

Τα χρηματοπιστωτικά ιδρύματα βασίζονται σε μεγάλο βαθμό σε δανειακά κεφάλαια, επομένως εξετάζεται συχνά για να διαπιστωθεί εάν μπορούν να αποπληρώσουν τα χρέη τους χωρίς να υποστούν σημαντικές απώλειες που μπορεί να είναι καταστροφικές. Ως αποτέλεσμα, τα ιδρύματα πρέπει να τηρούν υψηλά πρότυπα συμμόρφωσης και να υποβάλλονται σε δοκιμασίες ακραίων καταστάσεων για να αξιολογήσουν την οικονομική τους δυνατότητα και την ρευστότητά τους. (Kenton, 2021).

2.2.4. Κίνδυνος Αγοράς.

Ο κίνδυνος αγοράς είναι συστηματικός δηλαδή μπορεί να αντιμετωπιστεί από κάθε χρηματοπιστωτικό ίδρυμα ανάλογα με την οικονομική κατάσταση και δυνατότητα του καθενός, αλλά δεν μπορεί να διαχειριστεί ολοκληρωτικά από την τράπεζα. Η προέλευση του κινδύνου αυτού οφείλεται σε καταστάσεις που επηρεάζουν όλη την χρηματοπιστωτική αγορά και όχι μόνο μία τράπεζα. Παράδειγμα είναι ο πληθωρισμός, τα επιτόκια, πολιτική αναταραχή κτλ. Αντίθετα από τον κίνδυνο αγοράς έχουμε τον ειδικό ή μη συστηματικό κίνδυνο, ο οποίος αφορά συγκεκριμένη εταιρεία ή κλάδο. (Hayes, 2023).

Οι χρηματοπιστωτικές αγορές λειτουργούν ουσιαστικά ασταμάτητα και νέες τιμές δημιουργούνται συνεχώς. Ως εκ τούτου, ο κίνδυνος αγοράς είναι ένας από τους συνηθέστερους οικονομικούς κινδύνους που πρέπει να εξεταστούν λόγω της πληθώρας δεδομένων και της συλλογικής εμπειρίας για την αντιμετώπιση αυτού του κινδύνου. Ο κίνδυνος αγοράς εξακολουθεί να είναι ένα δύσκολο ζήτημα να εκτιμηθεί ποσοτικά. Οι πληροφορίες που χρησιμοποιούνται για τον υπολογισμό πιθανών ζημιών βασίζονται σε προηγούμενες τιμές και επιτόκια και όχι σε μελλοντικές. Τα μοντέλα διαχείρισης κινδύνου δίνουν στον έμπειρο διαχειριστή κινδύνου ένα πλαίσιο μέσα στο οποίο μπορεί να αξιολογήσει αυτή την κρίση και τη δυνατότητα να συνδυάσει αυτά τα ιστορικά δεδομένα με τη δική του μελλοντική κρίση. (CFA Institute, 2023).

2.2.5. Λειτουργικός κίνδυνος.

Λειτουργικός είναι ο κίνδυνος που σχετίζεται με την πιθανότητα ύπαρξης ζημιών που προκαλούνται από την αδυναμία σωστής λειτουργίας ή ακόμα και καταστροφής των μέσων με τα οποία η τράπεζα διεξάγει τις διαδικασίες της. Για παράδειγμα ο υπερβολικός όγκος πληροφοριών ή η δυσλειτουργία των συστημάτων, αποτελούν λειτουργικό κίνδυνο. Πολλές φορές μπορεί να παρουσιαστούν προβλήματα ακόμα και στην διαχείριση περιουσιακών στοιχείων, στην καταγραφή πληροφοριών και στην διεξαγωγή δραστηριοτήτων όπως οι πωλήσεις και οι συναλλαγές.

Η εσωτερική και εξωτερική ασφάλεια παίζουν επίσης πολύ σημαντικό ρόλο στην λειτουργία μίας επιχείρησης. Οι διαδικασίες επιτελούνται βασισμένες στις δεξιότητες των υπαλλήλων, στους ελέγχους και στα διάφορα συστήματα δικτύων επικοινωνίας και στα αυτοματοποιημένα συστήματα. Από την στιγμή που ξεκινάει μία συναλλαγή η τράπεζα εκτίθεται στον κίνδυνο να απειληθεί η ασφάλεια των πληροφοριών που διαθέτει για τους διάφορους πελάτες της. Διάφορες εξωτερικές απειλές μπορεί να διαρρεύσουν τις πληροφορίες αυτές και να αποσπάσουν χρήματα από τα ιδρύματα μέσω των δεδομένων πελατών και τραπεζικών κεφαλαίων. Οπότε κατανοούμε πως η τράπεζες φέρουν ευθύνη στην ορθή λειτουργία των εργασιών τους και στην ασφάλεια των δεδομένων με τα οποία τις εμπιστεύονται οι πελάτες τους. (CFI, 2023).

2.2.6. Άλλα είδη κινδύνου.

Τα χρηματοπιστωτικά ιδρύματα εκτίθενται σε διάφορα είδη κινδύνου τα οποία καταγράφονται ξεχωριστά ανά κάθε περίπτωση. Μερικά από αυτά είναι ο **συναλλαγματικός, ο κίνδυνος εξαπάτησης, ο κίνδυνος χώρας, κίνδυνος απώλειας προσωπικού, ο ρυθμιστικός, ο τεχνολογικός, οι φυσικοί, ο νομικός ή και ο συνεταιρικός.**

Ο **συναλλαγματικός** κίνδυνος συνδέεται με τις συναλλαγματικές ισοτιμίες και την πιθανή αστάθειά τους. Οι τράπεζες χορηγούν δάνεια σε ξένα νομίσματα, αλλά και δέχονται καταθέσεις σε αυτά, οπότε όταν μεταβάλλονται οι ισοτιμίες αυτό επηρεάζει τα έσοδα και τα έξοδά της. Για παράδειγμα όταν μειώνονται οι ισοτιμίες μειώνονται τα έσοδα της επιχείρησης από τα δάνεια αλλά και τα έξοδά της από τις καταθέσεις. Όμως αν η μείωση στα έσοδα είναι μεγαλύτερη αυτής των εξόδων, τότε η επιχείρηση έχει ζημία. (Παπαδάμου, 2009).

Οι υπόλοιποι κίνδυνοι που αναφέρθηκαν είναι επίσης σημαντικοί για την υγιή λειτουργία της τράπεζας ως προς την διαχείρισή τους. Ο κίνδυνος εξαπάτησης είναι σημαντικός διότι χωρίς τους απαραίτητους ελέγχους μπορεί να υπάρχουν αρνητικές επιπτώσεις στα κέρδη της επιχείρησης, είτε από λάθη υπαλλήλων, είτε από εκμετάλλευση της θέσης τους για προσωπικά τους κέρδη. Ο κίνδυνος χώρας αφορά την αδυναμία ή απροθυμία μίας χώρας να αποπληρώσει τα δάνειά της στον οφειλέτη της. Πολλές φορές το προσωπικό της επιχείρησης λόγω διαφόρων παραγόντων μπορεί να μην είναι πλέον διαθέσιμοι για εργασία, οπότε είναι απαραίτητη η εκπαίδευση άλλων μελών για την θέση τους, καθώς υπάρχει ο κίνδυνος να μην μπορούν να εκτελεστούν ορισμένες λειτουργίες λόγω έλλειψης προσωπικού. Οι φυσικοί κίνδυνοι οφείλονται στις φυσικές καταστροφές και οι νομικοί οφείλονται στην απώλεια αξίας ενός συμβολαίου λόγω λανθασμένης διατύπωσης των όρων της. Τέλος ο συνεταιρικός κίνδυνος οφείλεται στην αδυναμία ή άρνηση εκπλήρωσης υποχρεώσεων από έναν εμπορικό συνεταιίρο της τράπεζας για διάφορους προσωπικούς ή γενικούς παράγοντες. (Κιόχος, Παναγόπουλος & Κυρμιζογλου, 2018).

3. Διαχείριση κινδύνων.

Η κάθε τράπεζα χρησιμοποιεί τεχνικές εντοπισμού και αξιολόγησης του κινδύνου βάσει των δικών της δεδομένων. Μετά τον εντοπισμό του κινδύνου είναι σημαντικό να γίνει εύρεση τρόπου αντιμετώπισης αυτού. Ο ορισμός της διαχείρισης των κινδύνων μπορεί να πάρει πολλές εκδοχές, μπορούμε όμως να συμπεράνουμε από κάθε ορισμό πως αποτελεί τον σχεδιασμό και την εφαρμογή ενός λογικού σχεδίου για την αντιμετώπιση πιθανών απωλειών. Η προστασία του κεφαλαίου και της κερδοφορίας μέσω ρύθμισης των υποχρεώσεων είναι ο στόχος της διαδικασίας διαχείρισης κινδύνων. Η αποτελεσματική διαχείριση κινδύνων δίνει στις τράπεζες τη δυνατότητα να μειώσουν τις συνέπειές τους και να αποφασίσουν για τις εμπορικές τους δραστηριότητες με πιο αποτελεσματικό τρόπο. Ο συνδυασμός των διαδικασιών που αναφέρθηκαν αποτελούν την διαχείριση των κινδύνων. (Tursoy, 2018).

Η διαχείριση τραπεζικού κινδύνου είναι μία διαδικασία η οποία περιέχει αρκετά στάδια. Αρχικά απαιτείται ο προσδιορισμός των κινδύνων στους οποίους εκτίθενται οι διάφορες συναλλαγές και επιχειρηματικές ενέργειες. Για να επιτευχθεί αυτό απαιτείται η ανάλυση των λειτουργιών της τράπεζας και του εσωτερικού και εξωτερικού περιβάλλοντός της και να ληφθούν υπόψη οι πιθανές επιπτώσεις αυτών των κινδύνων στην σχέση με τους πελάτες. Οι τράπεζες καλούνται να αξιολογήσουν το μέγεθος και την επίδραση του κινδύνου, αφού έχει εντοπιστεί, προσδιορίζοντας την πιθανότητα και το μέγεθος των μελλοντικών απωλειών. Έτσι, είναι απαραίτητο να αναπτυχθούν μοντέλα τα οποία συντελούν στην εκτέλεση δοκιμών και ανάλυση δεδομένων.

Η χρήση μοντέλων είναι ζωτικής σημασίας για τη διαχείριση του κινδύνου αγοράς. Ένα μοντέλο είναι μια συμπυκνωμένη εκδοχή ενός πραγματικού φαινομένου. Τα βασικά στοιχεία που επηρεάζουν τις τιμές και τις ευαισθησίες στις χρηματοπιστωτικές αγορές αποτυπώνονται από τα χρηματοοικονομικά μοντέλα. Αυτό το επιτυγχάνουν δίνοντας σημαντικές πληροφορίες που απαιτούνται για τον έλεγχο του κινδύνου. Για παράδειγμα, τα μοντέλα επενδυτικού κινδύνου βοηθούν έναν διαχειριστή χαρτοφυλακίου να προσδιορίσει πόσο μια αλλαγή σε ένα συγκεκριμένο στοιχείο κινδύνου αναμένεται να επηρεάσει την αξία του χαρτοφυλακίου. Επιπλέον, ρίχνουν φως στα κέρδη και τις ζημίες που θα μπορούσε να αναμένει το χαρτοφυλάκιο καθώς και στην πιθανότητα να συμβαίνουν συχνά σημαντικές ζημίες. (CFA Institute, 2023).

Μετά την αξιολόγηση, οι τράπεζες απαιτούνται να λάβουν δράση για να θέσουν τους κινδύνους και τις πιθανές συνέπειες που μπορεί να φέρουν υπό τον έλεγχό τους. Αυτό οδηγεί στην δημιουργία διαφόρων πρακτικών για την μείωση των συνεπειών όπως η εφαρμογή κατάλληλων ελέγχων. Βέβαια, καθώς οι αγορές, η τεχνολογία και η παγκόσμια οικονομία εξελίσσεται και μεταβάλλεται συνεχώς, απαιτείται διαρκής παρακολούθηση και αξιολόγηση των καταστάσεων που περιβάλλουν τα χρηματοπιστωτικά ιδρύματα. Με αυτόν τον τρόπο δημιουργούνται τα διάφορα πρωτόκολλα τα οποία οι τράπεζες, συνεπώς και όσοι συναλλάσσονται με αυτές πρέπει να ακολουθούν. Η αποτελεσματική διαχείριση κινδύνων είναι ζωτικής σημασίας για κάθε είδους επιχείρηση αλλά ιδιαίτερα για τις τράπεζες που φέρουν μεγάλη οικονομική ευθύνη ως προς όλους όσους συναλλάσσεται. (Hull, 2018).

3.1. Διαδικασία διαχείρισης κινδύνου.

Οι τράπεζες δημιουργούν έναν συγκεκριμένο τύπο προγράμματος για να εντοπίσουν κάποιον κίνδυνο. Γίνεται εντοπισμός κινδύνων που σχετίζονται με τις εταιρείες και τις αιτίες τέτοιων περιστατικών. Μέσω διαφόρων τεχνικών οι τράπεζες μπορούν να μετριάσουν τον κίνδυνο που τις περιβάλλει. Όπως υποστηρίζει η παρούσα εργασία μία τράπεζα μπορεί να χρησιμοποιήσει αναλυτικές τεχνικές και δεδομένα μηχανικής μάθησης για αυτόματη εκτέλεση των διαδικασιών της και για την παρακολούθησή τους.

Η αυτοματοποίηση των διαδικασιών ελέγχου συμβάλλουν επίσης στην ειδοποίηση των διαχειριστών κινδύνου για οποιεσδήποτε μη φυσιολογικές τιμές στις λειτουργίες της. Έτσι η τράπεζα δίνει εντολή στο προσωπικό διαχείρισης κινδύνου να επικεντρωθούν σε συναλλαγές και λειτουργίες υψηλού κινδύνου, ακολουθώντας συγκεκριμένα βήματα όπως ο προσδιορισμός, η ανάλυση, η αξιολόγηση και η ταξινόμηση, η αντιμετώπιση και τέλος η παρακολούθηση και επανεξέταση του κινδύνου.

Το πρώτο βήμα είναι ο προσδιορισμός του κινδύνου. Όπως αναλύσαμε στο προηγούμενο κεφάλαιο μπορεί να υπάρχουν πολλά είδη κινδύνων για τα οποία η τράπεζα πρέπει να είναι σε διαρκή επαγρύπνηση και να εντοπίσει όσο περισσότερα μπορεί είτε βάσει του κλάδου της, είτε βάσει της δικιάς της δραστηριότητας. Υπάρχουν επίσης διάφορα συστήματα τα οποία μπορούν να ενημερώσουν τους χρήστες σε περίπτωση που οι τιμές δεν είναι φυσιολογικές.

Στην συνέχεια, αφού γίνει ο εντοπισμός, απαιτείται η ανάλυση του κινδύνου. Η ομάδα διαχείρισης καλείται να υπολογίσει την έκταση του κινδύνου και να κάνει κατανομή μεταξύ σχέσης κινδύνου και άλλων οργανωτικών χαρακτηριστικών. Είναι πολύ σημαντικό να γίνεται εξέταση του αριθμού των επιχειρηματικών δραστηριοτήτων που επηρεάζει ο κίνδυνος και των τμημάτων για να μετρηθεί ο βαθμός και ή το μέγεθος των συνεπειών του. Η ανάλυση γίνεται συνήθως σε κάποιο σύστημα που παρέχει τα απαραίτητα εργαλεία, ώστε στην συνέχεια να γίνει και η χαρτογράφηση των κινδύνων και να χρησιμοποιηθούν στις επιχειρηματικές διαδικασίες.

Αφού ολοκληρωθεί η διαδικασία της ανάλυσης γίνεται αξιολόγηση και ταξινόμηση του κινδύνου. Ορίζεται το μέγεθός του και αναλόγως αυτού και του είδους ταξινομείται σε διάφορες ομάδες ανά είδος. Για παράδειγμα οι κίνδυνοι που οδηγούν σε μικρότερες συνέπειες κατατάσσονται χαμηλότερα, με αποτέλεσμα να γίνει σωστή διαχείριση των προτεραιοτήτων για την επιχείρηση. Βέβαια η αξιολόγηση και ταξινόμηση μπορεί να αφορά μετρήσιμους και μη μετρήσιμους κινδύνους. Στην περίπτωση των μη μετρήσιμων αξιολογούμε ποιοτικά ενώ στους μετρήσιμους που είναι συνήθως συνδεδεμένοι με οικονομικές δραστηριότητες τους αξιολογούμε ποσοτικά.

Έχοντας ολοκληρώσει όλα τα παραπάνω βήματα η τράπεζα μπορεί να συγκεντρωθεί στην εφαρμογή των τεχνικών διαχείρισης που αναλύονται στο επόμενο υπό κεφάλαιο για την εξάλειψη ή ελαχιστοποίηση του κινδύνου. Η διαδικασία αντιμετώπισης του κινδύνου στην περίπτωση που δεν εξαλείφεται, δεν τελειώνει σε κάποιο βήμα καθώς απαιτείται η συνεχής παρακολούθηση και επανεξέτασή του. Η παρακολούθηση αυτή συνήθως γίνεται από συστήματα διαχείρισης κινδύνου και για να είναι αποτελεσματική

η χρήση χρειάζεται η ομάδα διαχείρισης κινδύνου να έχει τις κατάλληλες γνώσεις και σχέδια. (Thomas, 2017).

Ο κάθε κίνδυνος αντιμετωπίζεται με ορισμένες τεχνικές που προσαρμόζονται στην φύση του. Ένας από τους κινδύνους που αναφέραμε είναι ο πιστωτικός ο οποίος μπορεί να διαχειριστεί με πολλές τεχνικές. Μία από αυτές είναι η μέθοδος εσωτερικών διαβαθμίσεων (Internal Ratings Based Method) η οποία εξετάζει την ικανότητα του πελάτη να αποπληρώσει το δάνειό του. Άλλη μία μέθοδος είναι η τυποποιημένη (Standardized Method) μέσω της οποίας γίνεται δυνατός ο ορισμός κινδύνου για κάθε περιουσιακό στοιχείο. Στον κίνδυνο ρευστότητας η τράπεζα πρέπει να διατηρήσει σταθερή σχέση μεταξύ των ιδίων κεφαλαίων και των υποχρεώσεών της. (Στεργίου, 2013). Ο λειτουργικός κίνδυνος ο οποίος βασίζεται στις χρηματοοικονομικές εξελίξεις μπορεί να διαχειριστεί μέσω της ανάλυσης δεδομένων του τομέα χρησιμοποιώντας τις τεχνικές που αναφέρθηκαν όπως γίνεται και σε όλα τα άλλα είδη κινδύνου.

Πολύ σημαντικό ρόλο στην διαχείριση του κινδύνου παίζει η διαχείριση χαρτοφυλακίου, το οποίο αποτελεί μια ομαδοποίηση πολλών τύπων χρηματοοικονομικών περιουσιακών στοιχείων, όπως κεφάλαια και τίτλοι όπως μετοχές, ομόλογα, εμπορεύματα, μετρητά και ισοδύναμα μετρητών. Τις περισσότερες φορές παρουσιάζεται ότι ο πυρήνας ενός χαρτοφυλακίου αποτελείται από μετοχές, ομόλογα και μετρητά. Αν και αυτό συμβαίνει συχνά, δεν αποτελείται απαραίτητα μόνο από αυτά. Διάφοροι τύποι περιουσιακών στοιχείων, όπως ιδιωτικές επενδύσεις, ακίνητα και καλές τέχνες, μπορούν να περιλαμβάνονται επίσης σε ένα χαρτοφυλάκιο.

3.2. Τεχνικές διαχείρισης κινδύνου.

Το σημερινό περιβάλλον είναι διαρκώς μεταβαλλόμενο καθώς οι κοινωνίες, οι οικονομίες αλλά και η τεχνολογία μεταβάλλονται διαρκώς, κάτι που κάνει την βάση για την λειτουργία οποιασδήποτε επιχείρησης αρκετά ασταθή. Οι τεχνικές διαχείρισης κινδύνου βοηθούν στη αντιμετώπιση της αστάθειας και των κινδύνων που αυτή κρύβει σε κάθε τομέα αλλά ιδιαίτερα στον χρηματοπιστωτικό που αποτελεί μεσάζων πολλών συναλλαγών. Οι τεχνικές διαχείρισης συμβάλλουν στην εκτέλεση όλων των βημάτων αντιμετώπισης των πιθανών κινδύνων και να λάβουν λογικές αποφάσεις βελτιώνοντας την χρηματοοικονομική τους σταθερότητα.

3.2.1. Η μέθοδος Value at Risk.

Τα χρηματοπιστωτικά ιδρύματα χρειάζονται μια μέθοδο να υπολογίσουν τον κίνδυνο που περιβάλλει τα χρηματοοικονομικά τους στοιχεία. Το χαρτοφυλάκιο μίας τράπεζας περιλαμβάνει όλα εκείνα τα στοιχεία, οπότε η δυνατότητα μέτρησης του κινδύνου του είναι πολύ σημαντική. Η μέτρηση αυτή γίνεται με την μέθοδο Value at Risk ή VaR η οποία υπολογίζει την πιθανότητα ύπαρξης απώλειας σε ένα χαρτοφυλάκιο χρησιμοποιώντας δεδομένα της αγοράς και στατιστικά μοντέλα. Στην ουσία η μέθοδος αυτή υπολογίζει το ύψος της μέγιστης απώλειας (L_p), την οποία πρέπει να αναμένουμε σε διαφορετικές χρονικές περιόδους με βάση κάποια πιθανότητα. (Πιτσέλης, 2006).

Οι διάφορες διακυμάνσεις των τιμών στην αγορά των επιτοκίων μπορεί να προκαλέσουν αναταραχές για μία τράπεζα και η VaR είναι μία μέθοδος υπολογισμού αυτού του κινδύνου. Οι κίνδυνοι στην αγορά μπορεί να προκαλέσουν πτώση της αξίας των περιουσιακών στοιχείων και της εμπορικής θέσης μίας τράπεζας οπότε είναι πολύ σημαντικό να χρησιμοποιείται μία τέτοια τεχνική. Υπολογίζεται η πιθανή απώλεια στα

χαρτοφυλάκια μία τράπεζας και το ποσό έκθεσής της στον κίνδυνο. Βέβαια η συγκεκριμένη μέθοδος έχει τον περιορισμό του χειρότερου σεναρίου. Δηλαδή υπολογίζει μόνο την πιθανότητα απώλειας στο χειρότερο σενάριο καθώς επίσης λαμβάνει υπόψιν ότι το χαρτοφυλάκιο είναι σωστά διαμορφωμένο χωρίς να λαμβάνει υπόψιν άλλα είδη κινδύνων. (Jorion, 2007).

Η τεχνική VaR αποτελείται από τον χρόνο, το επίπεδο εμπιστοσύνης, το χαρτοφυλάκιο και το στατιστικό μοντέλο που εκτιμά τις πιθανές περιπτώσεις ζημιών. Πιο συγκεκριμένα, σύμφωνα με τους Κιόχος, Παναγόπουλος & Κυρμίζογλου, γίνεται λόγος για έναν αριθμό που προκύπτει από την πιθανότητα $(1-\alpha)\%$ ότι μία τράπεζα δε θα χάσει πάνω από X μονάδες στις επόμενες n ημέρες όπου το n ισούται συνήθως με 10 ημέρες. Το χρηματοφυλάκιο μπορεί να περιλαμβάνει μετοχές, νομίσματα, εμπορεύματα, ομόλογα και άλλα παρόμοια μεγέθη που μπορεί να έχει στην κατοχή της η τράπεζα. Έτσι, μετά τον υπολογισμό του αριθμού VaR, τα χρηματοπιστωτικά ιδρύματα έχουν την δυνατότητα να προσδιορίσουν το μέγεθος του κινδύνου και να λάβουν τα απαραίτητα μέτρα για την αντιμετώπισή του.

3.2.2. Ιστορική προσομοίωση.

Με βάση την συγκεκριμένη μέθοδο διαχείρισης κινδύνου ο οργανισμός χρησιμοποιεί ήδη διαθέσιμα ιστορικά δεδομένα για να εξετάσει την πιθανότητα μελλοντικών συμβάντων και απωλειών. Βασική υπόθεση για αυτήν τη μέθοδο αποτελεί η χρήση των πληροφοριών που παρέχουν τα ιστορικά δεδομένα για την πιθανότητα και τη σοβαρότητα των μελλοντικών κινδύνων. Συνήθως ένα μοντέλο προσομοίωσης εκτιμά πιθανά αποτελέσματα διαφορετικών αποτελεσμάτων που μπορεί να δώσει η πραγματοποίηση ενός κινδύνου. Μέσω της εξέτασης των δεδομένων το μοντέλο εντοπίζει πρότυπα και τάσεις που μπορεί να συμβάλουν στην πρόβλεψη μελλοντικών κινδύνων.

Μεγάλο πλεονέκτημα της συγκεκριμένης μεθόδου είναι η δυνατότητα παροχής ακριβής εκτίμησης του κινδύνου σε σύγκριση με άλλα μοντέλα που βασίζονται μόνο σε στατιστικά μοντέλα ή στην εμπειρία της ομάδας διαχείρισης. Τα συμπεράσματα που εκδίδει το μοντέλο βασίζονται σε πραγματικά γεγονότα και μπορούν να βοηθήσουν μία επιχείρηση να αποφύγει κινδύνους. Υπάρχουν όμως ορισμένοι περιορισμοί στην συγκεκριμένη μέθοδο καθώς μπορεί τα δεδομένα να μην είναι καθαρά ή να είναι ανακριβή με αποτέλεσμα η πρόβλεψη να απέχει από ακριβείς προβλέψεις. Μάλιστα, πολλές φορές λόγω του ότι οι προβλέψεις βασίζονται σε δεδομένα προέλευσης από ιστορικά γεγονότα, μπορεί να εκδίδουν υποθέσεις που πλέον λόγω της οικονομικής ή τεχνολογικής ανάπτυξης να μην ισχύουν. Η ιστορική προσομοίωση μπορεί να είναι αρκετά βοηθητική για να μην επαναληφθούν λάθη που μπορεί να έγιναν στο παρελθόν από άλλα ιδρύματα, αλλά χρειάζεται να υπάρχουν σωστά και καθαρά δεδομένα και διαρκής ενημέρωση του μοντέλου με όσο πιο πρόσφατες πληροφορίες. (Hull, 2018).

3.2.3. Η μέθοδος Monte Carlo.

Η μέθοδος διαχείρισης κινδύνου Monte Carlo χρησιμοποιεί τυχαία δείγματα και κατανέμει τις πιθανότητες για να κάνει εκτιμήσεις που οδηγούν σε διαφορετικά σενάρια στα οποία θα οδηγούσε ένας κίνδυνος. Αυτό γίνεται με την δημιουργία ενός μοντέλου που προσομοιώνει διάφορα πιθανά αποτελέσματα τα οποία προέρχονται από ορισμένες μεταβλητές εισόδου και από υποθέσεις. Μέσα από αυτήν την διαδικασία

δημιουργούνται χιλιάδες ή ακόμα και εκατομμύρια προσομοιώσεις με αποτέλεσμα να υπάρχει πιο ακριβής εκτίμηση των πιθανοτήτων πραγματοποίησης του κινδύνου. Μπορούμε να χρησιμοποιήσουμε αυτήν την μέθοδο για την διαχείριση διαφόρων κινδύνων, όπως ο πιστωτικός, της αγοράς, ο λειτουργικός και της ρευστότητας.

Βασικό πλεονέκτημα της συγκεκριμένης μεθόδου είναι το γεγονός ότι εξετάζει διάφορα πολύπλοκα σενάρια τα οποία περιέχουν πολυάριθμες μεταβλητές που αντικατοπτρίζουν την αβεβαιότητα. Με αυτόν τον τρόπο παρουσιάζεται μία ολοκληρωμένη εικόνα των πιθανοτήτων και των επιπτώσεων στην ομάδα διαχείρισης κινδύνου η οποία διαμορφώνει στρατηγικές βάσει αυτών. Η Monte Carlo παρέχει ακριβείς υποθέσεις και πληροφορίες βάσει των δεδομένων τα οποία χρησιμοποιούνται. Άρα, όπως και στην ιστορική προσομοίωση, εάν τα δεδομένα δεν είναι ακριβή τότε παρόλο της ακρίβειας της μεθόδου, υπάρχει μεγάλη πιθανότητα να προβάλλονται ανακριβή δεδομένα για τους πιθανούς κινδύνους. Είναι αναγκαίο, λοιπόν, τα δεδομένα να είναι υψηλής ποιότητας και το μοντέλο να ενημερώνεται τακτικά. (McNeil, Frey & Embrechts, 2005).

3.2.4. Back testing & Scaling.

Εκτός από τις τεχνικές που ήδη αναφέραμε υπάρχουν και άλλες τεχνικές οι οποίες βασίζονται σε αυτές. Για τον έλεγχο και την αξιολόγηση πολλών τεχνικών ταυτόχρονα χρησιμοποιούμε την τεχνική Back testing. Πολύ σημαντικό είναι επίσης το scaling το οποίο αποτελεί μεθοδολογία για διαχείριση κινδύνων περισσότερων από μία περιόδους στις μεθοδολογίες που αναφέραμε. Το scaling είναι πολύ σημαντικό καθώς οι προηγούμενες τεχνικές αφορούσαν μόνο μία συγκεκριμένη περίοδο, όμως πρακτικά αυτό δεν εξυπηρετεί διότι μπορεί να έχουμε νέα δεδομένα καθημερινά και το scaling απλοποιεί την διαδικασία ενημέρωσης του μοντέλου με τα νέα δεδομένα. (McNeil, Frey & Embrechts, 2005).

4. Τεχνικές μηχανικής μάθησης και τραπεζικός τομέας.

Οι τράπεζες όπως παρουσιάστηκε στα προηγούμενα κεφάλαια έρχονται αντιμέτωπες με πολλούς κινδύνους, όμως καλούνται να βρουν τρόπους διαχείρισης αυτών των κινδύνων και αυτό το κάνουν μέσω της ανάλυσης δεδομένων. Η ανάλυση δεδομένων αποτελεί βασικό στοιχείο των τραπεζικών διαδικασιών, καθώς παρέχει ουσιώδεις γνώσεις στην ομάδα διαχειριστών κινδύνου αλλά και σε όλους τους υπαλλήλους της τράπεζας. Τα δεδομένα των τραπεζών προέρχονται είτε από συναλλαγές, καταθέσεις, αναλήψεις, πληρωμές, πιστωτικές κάρτες και διάφορες άλλες επιχειρηματικές δραστηριότητες. Ο όγκος αυτών των δεδομένων είναι τεράστιος και δύσκολος ως προς την διαχείριση και την χρησιμοποίηση για την διεξαγωγή χρήσιμων πληροφοριών που βοηθούν τις τράπεζες να αντιμετωπίσουν τους διάφορους κινδύνους. Ως λύση σε αυτό το πρόβλημα ο τραπεζικός τομέας χρησιμοποιεί την ανάλυση δεδομένων μέσω των τεχνικών μηχανικής μάθησης.

4.1. Η συμβολή της Ανάλυσης Δεδομένων.

Ο όρος δεδομένα είναι τα γεγονότα που αφορούν τις οντότητες και τις καταστάσεις που τις περιβάλλουν και μπορούν να καταγραφούν και να αποθηκευτούν σε αποθήκες δεδομένων σε μορφή βάσεων δεδομένων, όπως είναι τα data sets που χρησιμοποιούνται στην έρευνα της παρούσας εργασίας. (Hoffer, Ramesh & Tori, 2017). Η ανάλυση δεδομένων μπορεί να έχει διάφορα πλεονεκτήματα για μία τράπεζα. Αρχικά, Βελτιώνονται οι υπηρεσίες καθώς γίνεται πρόβλεψη των αναγκών των πελατών. Βασικό μέγεθος που προσφέρει έσοδα σε μία τράπεζα είναι τα δάνεια που χορηγεί σε μία οντότητα, κάτι που μπορεί να κρύβει πιστωτικό κίνδυνο. Για την αντιμετώπιση του ορισμένου κινδύνου γίνονται προβλέψεις της οικονομικής κατάστασης του κάθε πελάτη καθώς εξετάζονται και αναλύονται τα εισοδηματικά τους στοιχεία και όποιο άλλο χρηματοοικονομικό στοιχείο τους που μπορεί να συνδέεται με τον πιστωτικό κίνδυνο.

Η ασφάλεια της τράπεζας απέναντι στις απάτες είναι πολύ σημαντικό στοιχείο της ανάλυσης δεδομένων καθώς μπορεί μέσω συστημάτων ανίχνευσης να γίνει δυνατή η πρόβλεψη και η αποτροπή τους. Ακόμη μία θετική επίπτωση είναι η σωστή διαχείριση των δαπανών. Αναλύοντας δεδομένα από κάθε τμήμα του ιδρύματος μπορεί να γίνει παρακολούθηση των καταναλωτικών συνηθειών των πελατών και εστίαση – επένδυση στις πιο αποτελεσματικές υπηρεσίες. Μέσω ειδικών μοντέλων, μπορεί επίσης να γίνει η ανάλυση του χρηματοπιστωτικού κλάδου και αυτόματα να αξιολογηθεί η θέση της τράπεζας σε αυτόν. Αποτέλεσμα είναι η αύξηση της ανταγωνιστικότητάς της καθώς ανακαλύπτει πως μπορεί να βελτιώσει τα προϊόντα και τις υπηρεσίες της, μέσω ανάλυσης των τάσεων της αγοράς. (Κύρκος, 2015).

Οι τράπεζες, μέσω της ανάλυσης δεδομένων, μπορούν να αυξήσουν την απόδοσή τους, χάρη στην πιο αποτελεσματική διαχείριση και πρόβλεψη κινδύνων που αυτή τους παρέχει. Ιδιαίτερα σε ένα συνεχώς εξελισσόμενο και μεταβαλλόμενο περιβάλλον, η ανάλυση της αγοράς του κλάδου στον οποίο βρίσκεται οποιοδήποτε είδος επιχείρησης, είναι βασικός παράγοντας για την επίτευξη των στόχων τους, που είναι η μεγιστοποίηση του κέρδους με όσο το δυνατόν λιγότερους και φθηνότερους πόρους. Έτσι, η ανάλυση δεδομένων βοηθά στην ανάπτυξη της στρατηγικής που μία τράπεζα

καλείται να ακολουθήσει, ώστε να αντιμετωπίσει αποτελεσματικά τους κινδύνους και τις προκλήσεις που την περιβάλλουν.

Συγκεκριμένα οι αναλυτές τραπεζικών δεδομένων χρησιμοποιούν ιδιαίτερα τις τεχνικές μηχανικής εκμάθησης. Η μηχανική εκμάθηση έχει διευρυμένο αντίκτυπο στις εμπορικές εφαρμογές, με αρκετές λύσεις που χρησιμοποιούνται επί του παρόντος και πολλές άλλες υπό διερεύνηση. Από την παγκόσμια χρηματοπιστωτική κρίση, η διαχείριση κινδύνων βρίσκεται στο επίκεντρο των τραπεζών, με συνεχή προσοχή στον τρόπο αναγνώρισης, αξιολόγησης, αναφοράς και διαχείρισης των κινδύνων. Οι εξελίξεις στον τραπεζικό τομέα και τη διαχείριση κινδύνων, καθώς και τα σημερινά και μελλοντικά προβλήματα έχουν λάβει σημαντική ακαδημαϊκή και βιομηχανική έρευνα. Η μηχανική μάθηση παρέχει την ικανότητα εντοπισμού σημαντικών προτύπων στα δεδομένα και έχει αναδειχθεί ως ένα δημοφιλές εργαλείο για πρακτικά οποιαδήποτε δραστηριότητα που απαιτεί την εξαγωγή σημαντικών δεδομένων από σύνολα δεδομένων. Ένας προγραμματιστής από μόνος του μπορεί να μην είναι σε θέση να προσφέρει σαφείς και ακριβείς προδιαγραφές για τη διαδικασία εκτέλεσης όταν αντιμετωπίζει την ανάγκη εξαγωγής ουσιαστικών πληροφοριών από δεδομένα και την πολυπλοκότητα των προς διερεύνηση προτύπων. Αυτό το πρόβλημα επιλύεται με την παροχή προγραμμάτων με ικανότητα μάθησης και προσαρμογής στη μηχανική μάθηση.

Ο χρηματοοικονομικός τομέας μπορεί τώρα να εξετάσει και να εξορύξει μια τεράστια υποδομή δεδομένων που περιέχει πολλά διαφορετικά είδη μη δομημένων χρηματοοικονομικών δεδομένων για αγορές και πελάτες χάρη σε μια σειρά τεχνικών προόδων. Παρά τους περιορισμούς της μηχανικής μάθησης όσον αφορά τον προσδιορισμό της αιτιώδους συνάφειας, οι οικονομολόγοι τη χρησιμοποιούν όλο και περισσότερο σε συνδυασμό με άλλες μεθόδους και γνώσεις για την αξιολόγηση περίπλοκων σχέσεων. Η υπόσχεση για εξοικονόμηση κόστους, αυξημένη παραγωγικότητα και καλύτερη διαχείριση κινδύνων έχει οδηγήσει στη χρήση της μηχανικής μάθησης. Επιπλέον, νέοι κανόνες ανάγκασαν τις τράπεζες να αυτοματοποιηθούν προκειμένου να διατηρήσουν την αποτελεσματικότητα και την ανταγωνιστικότητά τους. (Leo, Sharma & Maddulety, 2019).

Οι τάσεις εντοπίζονται με τη χρήση τραπεζικών αναλυτικών στοιχείων, τα οποία βοηθούν την διοίκηση στη λήψη αποφάσεων. Οι μελλοντικές τάσεις μπορούν να προβλεφθούν και να χρησιμοποιηθούν ως βάση για τη λήψη αποφάσεων με την αξιολόγηση προηγούμενων δεδομένων. Αυτό επιτρέπει την αναγνώριση των συμπεριφορών. Η τράπεζα μπορεί να κάνει σοφές κρίσεις που μπορούν να μειώσουν τα λάθη και να αυξήσουν την αποτελεσματικότητα, εφαρμόζοντας εργαλεία και διαδικασίες ανάλυσης δεδομένων. Ένα βασικό εργαλείο διαχείρισης, τα τραπεζικά αναλυτικά στοιχεία, επισημαίνουν τομείς που απαιτούν ανάπτυξη. Χωρίς να έχουν πρόσβαση σε μετρήσεις που υποστηρίζονται από δεδομένα και προβλέψεις μελλοντικών αποτελεσμάτων, μπορεί να είναι δύσκολο για τις τράπεζες να καθορίσουν ποιοι τομείς πρέπει να αξιολογηθούν ή να βελτιωθούν. Οι αναλυτές δεδομένων μέσω της χρήσης τεχνικών όπως η κωδικοποίηση, μπορούν να αυτοματοποιήσουν πολλές εργασίες ανάλυσης και να ανακαλύψουν τις τάσεις πιο εύκολα. (Miles & Huberman, 1994). Τα κρυφά ζητήματα απόδοσης μπορούν να βρεθούν και να επιλυθούν χάρη στα τραπεζικά αναλυτικά στοιχεία.

Όπως αναφέραμε, η πώληση δανείων είναι μια σημαντική πηγή εισοδήματος για τις τράπεζες, αντιπροσωπεύοντας το μεγαλύτερο μέρος των εσόδων μιας τράπεζας. Παρά τα πιθανά κέρδη που συνδέονται με αυτά τα δάνεια, μπορεί να εξακολουθεί να είναι μια πολύ επικίνδυνη λειτουργία για τις τράπεζες. Υπάρχει πιθανότητα οι δανειολήπτες να αθετήσουν τα δάνειά τους, κοστίζοντας χρήματα στο χρηματοπιστωτικό ίδρυμα. Είναι κρίσιμο για τις τράπεζες να έχουν μια διαδικασία ελέγχου νέων πελατών προκειμένου να αξιολογήσουν τον κίνδυνο και να επιλέξουν εάν θα τους εγκρίνουν ή όχι. Οι χρηματοπιστωτικοί οργανισμοί ταξινομούν τους πελάτες τους με βάση τον βαθμό κινδύνου που ενέχουν χρησιμοποιώντας τεχνικές τραπεζικής ανάλυσης. Οι πελάτες χαμηλότερου κινδύνου επωφελούνται από τα χαμηλότερα επιτόκια, καθώς αποτελούν λιγότερο κίνδυνο για αυτούς τους οργανισμούς. Οι πελάτες με χειρότερες αξιολογήσεις πιστοληπτικής ικανότητας θα πληρώσουν υψηλότερα επιτόκια, δεδομένου ότι αυτοί θεωρούνται πιο ριψοκίνδυνοι καταναλωτές. Οι πιθανοί δανειολήπτες υψηλού κινδύνου μπορούν να απορρίψουν πλήρως τα αιτήματά τους για δάνεια.

Οι παραδοσιακές τραπεζικές πρακτικές γίνονται λιγότερο δημοφιλείς, καθώς οι άνθρωποι αναμένουν δεδομένα και τεχνολογία πληροφοριών που να είναι απλή στη χρήση. Για να διασφαλιστεί ότι ακολουθούνται οι βέλτιστες πρακτικές δεδομένων και ανάλυσης, είναι σημαντικό να προσληφθεί μια ποικιλία επαγγελματιών, συμπεριλαμβανομένων επιστημόνων δεδομένων, στατιστικολόγων και ατόμων με εμπειρία σε αναλυτικά εργαλεία και λογισμικό. Οι ηγέτες των τμημάτων στον τραπεζικό τομέα πρέπει να ενθαρρύνουν και να βοηθήσουν αυτά τα άτομα. Οι τράπεζες πρέπει να αλλάξουν τις διαδικασίες τους εάν θέλουν να παραμείνουν ανταγωνιστικές και να διατηρήσουν το μερίδιο αγοράς τους. Ένας τρόπος για να γίνει αυτό είναι μέσω της ανάλυσης των τραπεζικών δεδομένων για τη βελτίωση της εξυπηρέτησης πελατών, την αύξηση του μεριδίου αγοράς μέσω πιο επιτυχημένου μάρκετινγκ, την σωστή διαχείριση του κινδύνου στον τραπεζικό τομέα και τη διάκριση από τον ανταγωνισμό. (tibco.com, 2022).

4.2. Σωστή διαχείριση της Ανάλυσης Δεδομένων.

Η ανάλυση δεδομένων αποτελεί αναπόσπαστο κομμάτι των διαδικασιών που ακολουθούν οι τράπεζες, αλλά συνεχίζει να αποτελεί ένα εργαλείο το οποίο, όπως κάθε άλλο εργαλείο, απαιτεί σωστή χρήση. Πολλές φορές υπάρχουν προκλήσεις που η ανάλυση μπορεί να φέρει σε μία τράπεζα καθώς οι τεράστιοι όγκοι δεδομένων περιλαμβάνουν προσωπικά στοιχεία των πελατών τους. Αυτό σημαίνει πως πριν γίνει χρήση μίας τεχνικής ανάλυσης και του συστήματος που την υποστηρίζει, τα χρηματοπιστωτικά ιδρύματα οφείλουν να αναπτύξουν ορισμένες πολιτικές προστασίας και διαφάνειας ως προς τους υπαλλήλους που διαχειρίζονται αυτά τα συστήματα αλλά και ως προς εξωτερικές απειλές όπως επιθέσεις στον κυβερνοχώρο. Μάλιστα απαιτείται διαφάνεια ως προς την αιτία και τον τρόπο χρήσης των προσωπικών δεδομένων προς κάθε πελάτη της τράπεζας.

Σημαντικό κομμάτι της διαδικασίας ανάλυσης είναι η διάθεση σωστά εκπαιδευμένου και ικανού προσωπικού. Τα συστήματα και τα μοντέλα είναι απλά εργαλεία τα οποία χρειάζεται να υποστηρίζονται και να χρησιμοποιούνται από ικανές ομάδες κάθε οργανισμού για να παράγουν χρήσιμες πληροφορίες άρα και χρήσιμη γνώση. Οπότε, η ανάλυση δεδομένων μπορεί να φανεί χρήσιμη στην ανάληψη σωστών αποφάσεων σε

ένα χρηματοπιστωτικό ίδρυμα. Έτσι γίνεται η αντιμετώπιση διαφόρων κινδύνων όπως ο χρηματοπιστωτικός που είναι από τους πιο σημαντικούς, καθώς συμβάλει στην δυνατότητα της τράπεζας να αυξήσει τα έξοδά της επιλέγοντας τους κατάλληλους πελάτες για την χορήγηση δανείων. Για να διεξαχθεί αυτή η διαδικασία αποτελεσματικά, όμως, είναι απαραίτητη η διαμόρφωση κατάλληλων πολιτικών και η πρόσληψη και εκπαίδευση ικανού προσωπικού για την διαχείριση των εξειδικευμένων εργαλείων και τεχνολογιών. (Κύρκος, 2015).

Για να γίνει η ανάλυση των δεδομένων χρησιμοποιούνται εργαλεία μηχανικής μάθησης και ακολουθεί μία διαδικασία ανακάλυψης μοτίβων και σχέσεων μεταξύ των συνόλων δεδομένων. Αυτή η διαδικασία ονομάζεται εξόρυξη δεδομένων (Data Mining - DM). Το DM ενσωματώνει κυρίως τεχνικές και εργαλεία από τους τομείς της μηχανικής μάθησης, των στατιστικών, της τεχνητής νοημοσύνης και των βάσεων δεδομένων. Στόχος του είναι να αποκαλύψει ενδιαφέροντα μοτίβα σε πολλά δεδομένα. Η ανακάλυψη γνώσης βάσης δεδομένων είναι ένα άλλο όνομα για αυτό. (Li, 2018).

Η κωδικοποίηση των δεδομένων είναι ένα εξαιρετικά κρίσιμο βήμα στη διαδικασία ανάλυσής τους. Ένας κώδικας είναι μόνο λίγες λέξεις ή φράσεις που παρέχουν τα δεδομένα που σχετίζονται με το νόημα των αναλυτικών σκοπών του ερευνητή. Η χρήση κώδικα είναι επίσης σημαντικό βήμα της παρούσας εργασίας λαμβάνει χώρα σε πολλές φάσεις για μια σωστή ανάλυση δεδομένων.

Αν και δεν υπάρχει κάποιος συγκεκριμένος αριθμός σταδίων ή κύκλων κωδικοποίησης, μπορούμε να χωρίσουμε την διαδικασία της κωδικοποίησης τουλάχιστον σε τρία:

1. Ένας ερευνητής συχνά δίνει κωδικούς στο πλήρες σύνολο δεδομένων κατά την πρώτη φάση. Αυτό το επίπεδο κωδικοποίησης χρησιμοποιείται για την επισήμανση σημαντικών ισχυρισμών, περιστατικών και προβληματισμών. Ως αποτέλεσμα, το αρχικό επίπεδο κωδικοποίησης εστιάζει κυρίως στην εκκίνηση του συνόλου δεδομένων και στην προσπάθεια ελαχιστοποίησης του μεγέθους του σώματος δεδομένων προσδιορίζοντας αυτούς τους ισχυρισμούς, τις εμπειρίες και τις σκέψεις που είναι κρίσιμες για την ανάλυση.
2. Οι ερευνητές συχνά επιστρέφουν στα αποσπάσματα / τμήματα δεδομένων που κωδικοποίησαν στην πρώτη φάση στη δεύτερη φάση και προσθέτουν νέους κώδικες. Καθώς ένας ερευνητής αρχίζει να συλλογίζεται έννοιες ή / και ιδέες που σχετίζονται πιο άμεσα με το θέμα της έρευνας, οι κώδικες σε αυτό το στάδιο αρχίζουν να κινούνται προς έναν υψηλότερο βαθμό συμπερασμάτων. Το έναυσμα της σύνδεσης δηλώσεων, συμπερασμάτων και προβληματισμών που παρέχονται από τους συμμετέχοντες στην έρευνα με τους αναλυτικούς στόχους της μελέτης είναι ο στόχος της δεύτερης φάσης κωδικοποίησης.
3. Τέλος, όχι πάντα αλλά αρκετά συχνά ολοκληρώνεται ένα τρίτο βήμα κωδικοποίησης, κατά το οποίο γίνονται σαφείς συνδέσεις με τις εννοιολογικές ή / και θεωρητικές ιδέες της μελέτης. Η κωδικοποίηση συχνά επιτυγχάνει το μέγιστο επίπεδο συμπερασμάτων σε αυτό το στάδιο. Αυτοί οι κώδικες συνδέουν αυτούς τους ισχυρισμούς, τις εμπειρίες και τους προβληματισμούς με συγκεκριμένες εννοιολογικές ή θεωρητικές έννοιες, ενώ τα προηγούμενα στάδια κωδικοποίησης στόχευαν στον εντοπισμό του τι συνέβαινε στο σύνολο

δεδομένων προκειμένου να μειωθεί το μέγεθος και η πολυπλοκότητα του συνόλου δεδομένων. Μια δήλωση ή ένα σχόλιο από προηγούμενα πειράματα κωδικοποίησης, για παράδειγμα, μπορεί να επισημανθεί ιδιαίτερα ως απεικόνιση μιας φιλοσοφικής ή θεωρητικής ιδέας. (Lester, Cho & Lochmiller, 2020).

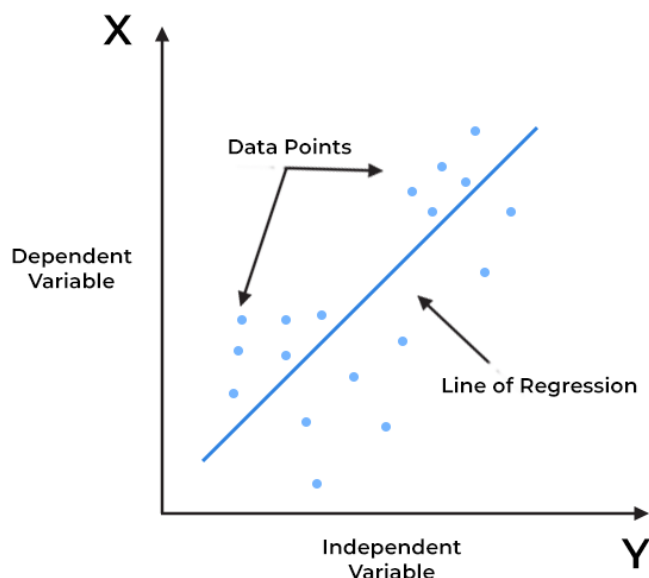
4.3. Τεχνικές μηχανικής μάθησης.

Η ανάλυση των δεδομένων δε μπορεί να λάβει χώρα χωρίς τα κατάλληλα μοντέλα τα οποία εξυπηρετούν τον σκοπό της. Όπως αναφέραμε Οι τεχνικές μηχανικής μάθησης αποτελούν αποτελεσματικά εργαλεία στον καθαρισμό και στην διαχείριση των δεδομένων που στην συνέχεια οδηγεί στην σωστή διαχείριση και πρόβλεψη ενός κινδύνου. Οι τράπεζες λαμβάνουν αιτήσεις για δάνεια από τους πελάτες της σε καθημερινή βάση κάτι που τις θέτει αντιμέτωπες με την διαχείριση του πιστωτικού κινδύνου.

Τα χρηματοπιστωτικά ιδρύματα έχουν ειδικά συστήματα με ενσωματωμένα μοντέλα που χρησιμοποιούν τεχνικές μηχανικής μάθησης και εξετάζουν το πιστωτικό σκορ, το εισόδημα και άλλα κριτήρια των πελατών για να αποφασίσουν για την έγκριση ενός δανείου. Όπως παρουσιάζεται στην συνέχεια της παρούσας εργασίας, θα γίνει χρήση μίας από τις τεχνικές μηχανικής μάθησης σε δεδομένα όπως είναι το εισόδημα και το ποσό δανείου, με σκοπό την ανάλυσή τους. Οι τεχνικές μηχανικής μάθησης είναι πολυάριθμες, αλλά στην παρούσα εργασία αναλύονται 15 από τις πιο γνωστές. (Mamun et al, 2022).

4.3.1. Γραμμική Παλινδρόμηση (Linear Regression).

Μέσω της παλινδρόμησης αντικατοπτρίζεται η αλληλοεπίδραση μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών. Οι εξαρτημένες μεταβλητές αποτελούν εκείνες που επιθυμούμε να βρούμε, ενώ οι ανεξάρτητες είναι εκείνες που χρησιμοποιούνται στην εύρεση των εξαρτημένων. Χάρη στην παλινδρόμηση, λοιπόν, δίνεται η δυνατότητα για εύρεση των τιμών των εξαρτημένων μεταβλητών μέσω των ανεξάρτητων οι οποίες τις επηρεάζουν. Άρα, η γραμμική παλινδρόμηση αποτελεί έναν αλγόριθμο που εκφράζει την σχέση μεταξύ αυτών των μεταβλητών. Ο αλγόριθμος αυτός μπορεί να είναι άλλες φορές απλός και άλλες πιο πολύπλοκος.



Εικόνα 4.3.1.1. Γραμμική Παλινδρόμηση, (Kanade, 2023).

Στο παραπάνω διάγραμμα ο άξονας των μεταβλητών X είναι οι εξαρτημένες μεταβλητές, ενώ αυτός των Y είναι οι ανεξάρτητες μεταβλητές. Η μπλε γραμμή είναι η ευθεία της παλινδρόμησης και οι κουκίδες είναι τα σημεία δεδομένων. Η γραμμική παλινδρόμηση χαρακτηρίζεται ως απλή όταν η εξαρτημένη μεταβλητή εξαρτάται μόνο από μία ανεξάρτητη. Σύμφωνα με τον Κύρκο Ε. παρακάτω παρουσιάζονται οι εξισώσεις απλής και πολλαπλής γραμμικής παλινδρόμησης και οι εξηγήσεις τους. Η εξίσωση της απλής γραμμικής παλινδρόμησης αποτυπώνεται ως εξής:

$$Y = \alpha + b \cdot X$$

Όπου X έχουμε την εξαρτημένη μεταβλητή που είναι ο στόχος, όπου Y έχουμε την ανεξάρτητη μεταβλητή, το α είναι η τιμή του Y όταν το X ισούται με 0 και το b είναι η κλίση της ευθείας. Οι παράμετροι α & b είναι άγνωστες και υπολογίζονται από τα διαθέσιμα δεδομένα με μαθηματικούς τύπους.

Βέβαια υπάρχει το ενδεχόμενο η εξαρτημένη μεταβλητή να εξαρτάται από περισσότερες από μία μεταβλητές. Η εξίσωσή της είναι παρόμοια με της απλής απλά προσθέτουμε και τα στοιχεία $b_n \cdot X^n$, δηλαδή:

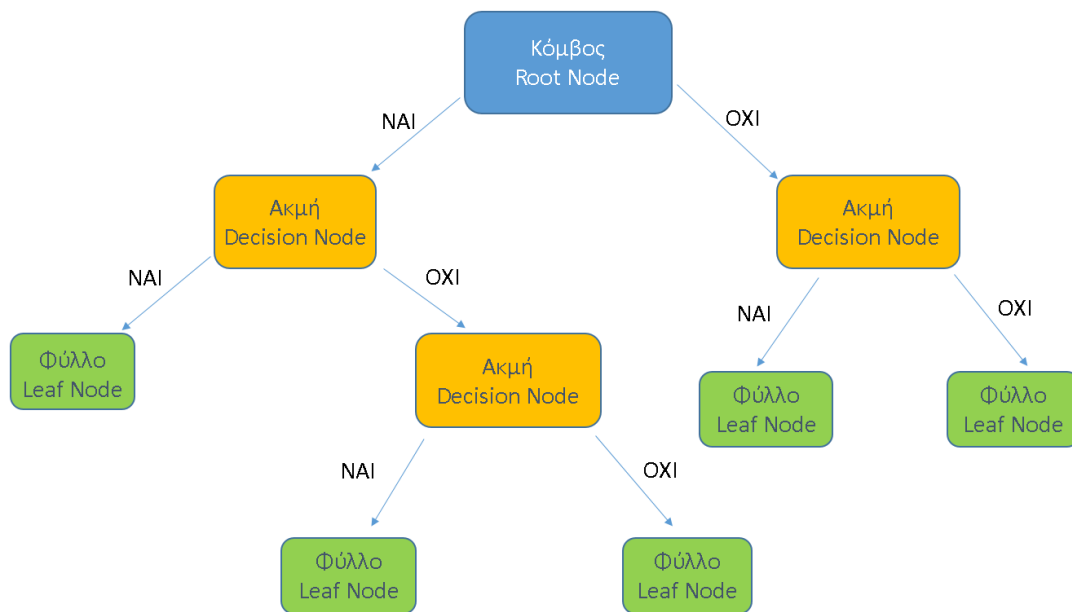
$$Y = \alpha + b_1 \cdot X + b_2 \cdot X^2 + \dots + b_n \cdot X^n$$

Η γραμμική παλινδρόμηση είναι πολύ χρήσιμη τεχνική και για αυτόν τον λόγο χρησιμοποιείται συχνά για την ανάλυση δεδομένων. Χαρακτηρίζεται από ευκολία στην υλοποίηση, χωρίς πολλά έξοδα και εύκολο στην συντήρηση ως μοντέλο. Σε αντίθεση με άλλα μοντέλα είναι πολύ απλή τεχνική και μπορεί να επεκταθεί ως προς τις λειτουργίες τους. Τέλος, χάρη στην απλότητά του μπορεί να εκπαιδευτεί συνεχώς χωρίς να καταλαμβάνει υπερβολικό χώρο. (Kanade, 2023).

4.3.2. Παλινδρόμηση με Δένδρα Αποφάσεων (Decision Trees).

Τα δένδρα αποφάσεων βασίζονται στην διαδικασία διάσπασης ενός συνόλου παρατηρήσεων σε υποσύνολα. Ονομάζονται δένδρα αποφάσεων, διότι ο τρόπος με τον οποίο αναπαρίστανται είναι παρόμοιος με την δομή ενός ανεστραμμένου δένδρου.

Στην κορυφή βρίσκεται ο κόμβος – ρίζα, και στην συνέχεια κατεβαίνοντας παρατηρούνται επιπλέον κόμβοι οι οποίοι συνδέονται με ακμές που θυμίζουν παρακλάδια και φύλλα. Ο αρχικός κόμβος έχει εξερχόμενες ακμές και οι υπόλοιποι έχουν εισερχόμενες ακμές που συνδέονται με τους ανώτερους κόμβους. Οι κόμβοι αποτελούν τον έλεγχο των δεδομένων και ο αρχικός ριζικός κόμβος αποτελεί το σύνολο των δεδομένων, οι ακμές είναι ο έλεγχος της απόφασης και τα φύλλα είναι η απόφαση κατηγοριοποίησης ή αλλιώς το αποτέλεσμα. Στην παρακάτω εικόνα παρουσιάζεται η δεντρική δομή που αναφέρθηκε. (Gurucharan, 2020)



Εικόνα 4.3.2.1. Δένδρα Αποφάσεων (Decision Trees).

Σύμφωνα με την Steorts R. από το πανεπιστήμιο Duke της Βόρειας Carolina, η μαθηματική αναπαράσταση του αλγορίθμου γραμμικής παλινδρόμησης με δένδρα αποφάσεων δίνεται από τον εξής τύπο:

$$\Delta_{g'} = \sum_i (y_i - \hat{y}_{R_m})^2 + \lambda(|T|)$$

Εικόνα 4.3.2.2. Παλινδρόμηση με δένδρα αποφάσεων (Steorts R., 2017).

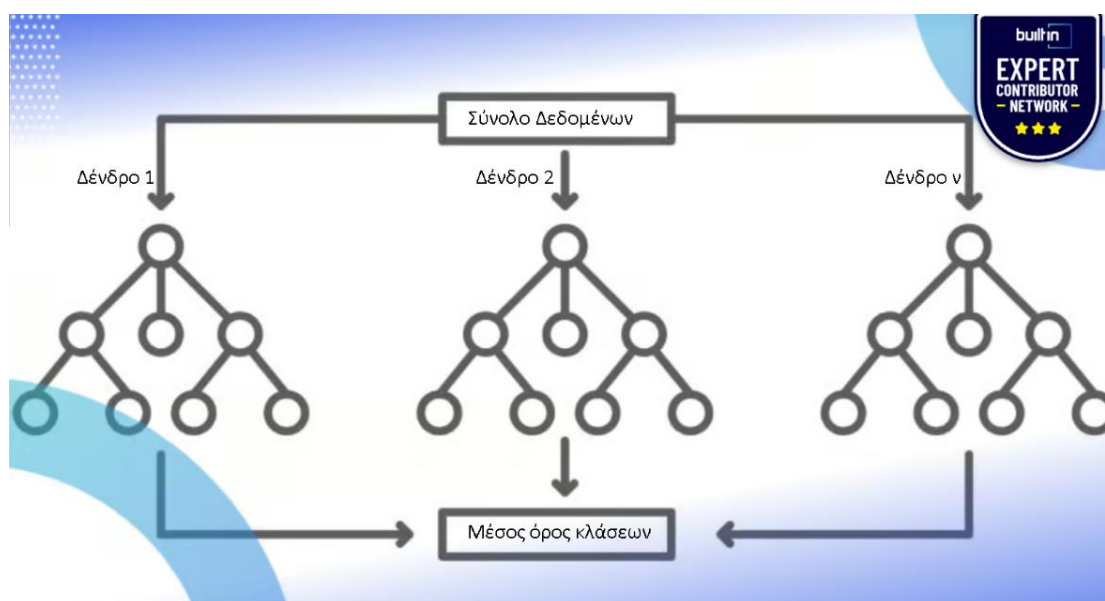
Όπου το $\Delta_{g'}$ αναπαριστά ένα πρόβλημα βελτιστοποίησης, το $|T|$ είναι ο αριθμός των τερματικών κόμβων του δένδρου T , R_m είναι το πλαίσιο στο οποίο τοποθετείται η ανταπόκριση στον τελικό κόμβο ενώ το \hat{y}_{R_m} είναι η ίδια η ανταπόκριση που σχετίζεται με το πλαίσιο και το α είναι ο έλεγχος της αντιστάθμισης μεταξύ πολυπλοκότητας του υπό δένδρου και της εφαρμογής τους στα δεδομένα εκπαίδευσης, η οποία συμβολίζεται με $\lambda|T|$. Τέλος το \sum_i συμβολίζει το άθροισμα του αποτελέσματος της εφαρμογής της εξίσωσης σε όλα τα σημεία του συνόλου δεδομένων.

Τα δένδρα αποφάσεων έχουν το πλεονέκτημα ότι δε κάνουν αυθαίρετες υποθέσεις για τις σχέσεις μεταξύ των μεταβλητών εισόδου και εξόδου, ενώ η αναπαράσταση της γνώσης γίνεται με κατανοητό τρόπο καθιστώντας εύκολη την εξαγωγή κανόνων. Μάλιστα η εκπαίδευση του αλγορίθμου αποτελεί αρκετά γρήγορη και υπάρχει η δυνατότητα διαχείρισης δεδομένων με χαμένες τιμές. Βέβαια, έχουν αρκετά

μειονεκτήματα επίσης αλλά πολύ σημαντικό είναι το χαρακτηριστικό τους ως ευπαθή σε μεταβολές του δείγματος εκπαίδευσης. Μερικά παραδείγματα του αλγορίθμου είναι ο ID3 και ο CART. (Κύρκος Ε., 2015).

4.3.3. Παλινδρόμηση Τυχαίου Δάσους (Random Forest Regression).

Ο αλγόριθμος παλινδρόμησης Random Forest χρησιμοποιεί την μέθοδο εκμάθησης συνόλου. Η μέθοδος αυτή συνδυάζει τις προβλέψεις που προκύπτουν από πολλούς άλλους αλγορίθμους μηχανικής μάθησης, με σκοπό την πιο ακριβή πρόβλεψη. Ο λόγος που ονομάζεται παλινδρόμηση τυχαίου δάσους οφείλεται στο γεγονός ότι συνδυάζει πολλά δένδρα αποφάσεων και τις προβλέψεις τους, τα δένδρα αυτά τρέχουν παράλληλα και στο τέλος υπολογίζεται ο μέσος όρος όλων των προβλέψεων - κλάσεων. (Ronaghan S., 2018). Η εικόνα 4.3.3.1 παρουσιάζει την συγκεκριμένη διαδικασία.



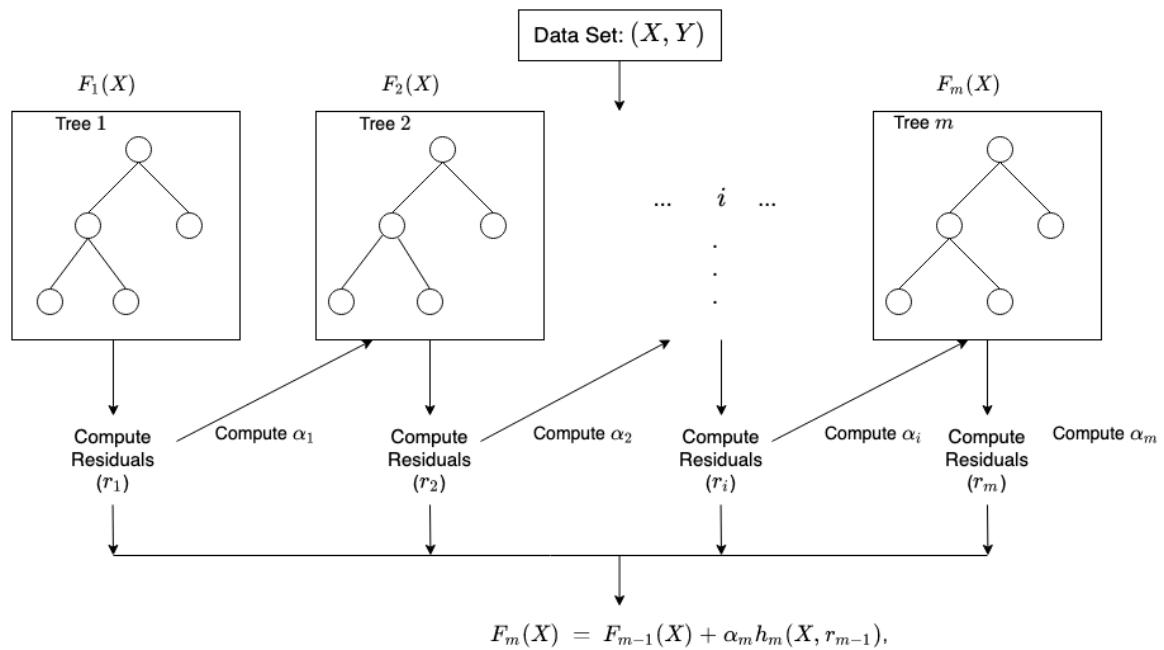
Εικόνα 4.3.3.1. Random Forest Regression (Chakure A, 2023).

Η συγκεκριμένη διαδικασία διάσπασης του αρχικού συνόλου σε υποσύνολα τα οποία αναλύονται παράλληλα με σκοπό τον συνδυασμό των αποτελεσμάτων τους που χρησιμοποιούμε στο Random Forest, ονομάζεται Bagging και είναι διαφορετικό από το Boosting, στο οποίο η επεξεργασία των υποσυνόλων γίνεται διαδοχικά και όχι παράλληλα. Οπότε η τεχνική ονομάζεται τεχνική εκμάθησης συνόλου (Ensemble Learning). Θετικό του αλγορίθμου είναι ότι μπορεί να λύσει πολύπλοκα προβλήματα καθώς δεν εξαρτάται μόνο σε ένα δένδρο αποφάσεων κάτι στο οποίο, όπως αναλύεται σε επόμενο κεφάλαιο, οφείλονται καλύτερα αποτελέσματα ανάλυσης των δεδομένων. Όμως, εφόσον χρησιμοποιούν δένδρα αποφάσεων, είναι επίσης πολύ ευπαθής αλγόριθμος σε μεταβολές του δείγματος εκπαίδευσης. (Saini A., 2021).

4.3.4. Παλινδρόμηση XG Boost & Light G Boost.

Η τεχνική Extreme Gradient Boosting (XG Boost) βασίζεται επίσης στον συνδυασμό ενός συνόλου πιο αδύναμων μοντέλων για την πρόβλεψη μίας μεταβλητής με όσο το δυνατόν μεγαλύτερη ακρίβεια. Είναι ένας αλγόριθμος αρκετά γρήγορος και ακριβής. Τα αδύναμα μοντέλα αποτελούνται από δένδρα παλινδρόμησης και κάθε δένδρο αντιστοιχεί σε ένα σύνολο δεδομένων εισόδου. Η συνάρτηση του αλγορίθμου συνδυάζει την απώλεια μεταξύ της εξόδου στόχου και της προβλεπόμενης εξόδου με

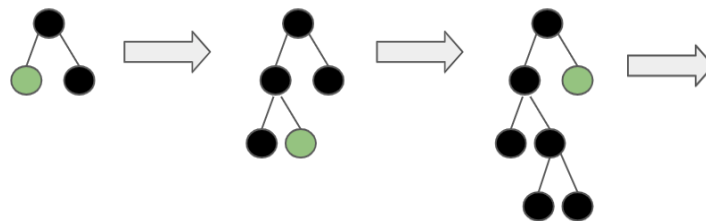
συναρτήσεις δένδρου παλινδρόμησης, μία διαδικασία που επαναλαμβάνεται για έναν ορισμένο αριθμό φορών που θέτει το μοντέλο εκπαίδευσης. (Amazon.com, 2023). Παρακάτω παρουσιάζεται γραφική αναπαράσταση της λειτουργίας του αλγορίθμου:



Εικόνα 4.3.4.1. LG Boost (Amazon.com, 2023)

Όπου α και r είναι οι παράμετροι κανονικοποίησης, δηλαδή οι τεχνικές τις οποίες χρησιμοποιούμε για να μειώσουμε την ενδεχόμενη απώλεια. (Banoula M., 2023). Επίσης, h είναι η λειτουργία της εκπαίδευσης για την πρόβλεψη του r χρησιμοποιώντας το X για το εκάστοτε δένδρο. Η λειτουργία compute Residuals αφορά τον υπολογισμό του r_1 και του α από τον αλγόριθμο ώστε να συνεχιστεί η διαδικασία με όλα τα δένδρα μέχρι το τελευταίο. Τα τετράγωνα στην εικόνα Trees αποτελούν τα δένδρα τα οποία προέρχονται από το σύνολο δεδομένων data set: (X, Y) .

Ο αλγόριθμος Light Gradient Boosting Machine (LGBM) είναι παρόμοιος με τον XG Boost καθώς και αυτός χρησιμοποιεί δένδρα αποφάσεων, αλλά τα διαχωρίζει με βάση τα φύλλα τα οποία προβλέπει ότι θα δώσουν το καλύτερο αποτέλεσμα και όχι με βάση τον κόμβο. Αυτό δίνει την δυνατότητα στον αλγόριθμο να εκτελείται πιο γρήγορα από τους υπολοίπους αλγορίθμους με πολύ μεγάλη απόδοση κάτι που τον καθιστά κατάλληλο για μεγάλα σύνολα δεδομένων. (Kasturi N., 2019). Στην εικόνα 4.3.4.2. αναπαρίσταται η λειτουργία αυτή.

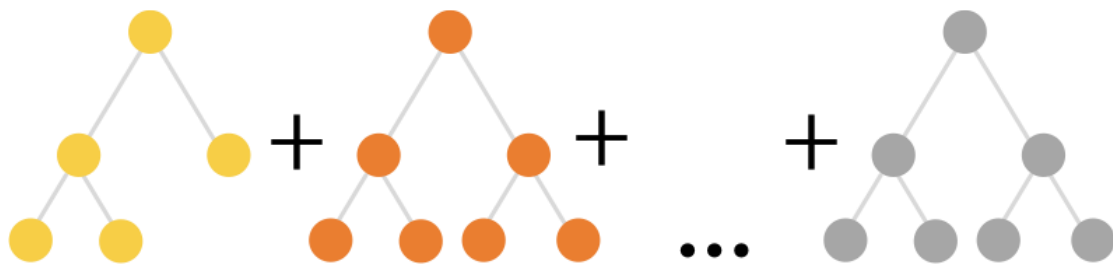


Εικόνα 4.3.4.2. LGBM (Datascience.eu, 2019).

4.3.5. Παλινδρόμηση Cat Boost.

Το μοντέλο Cat Boost Regression αναπτύχθηκε πολύ πρόσφατα, το 2017 και καθιστά δυνατή την ανάλυση ακόμα και μη αριθμητικών δεδομένων όπως εικόνες, ήχο ή κείμενο. Η εκτέλεσή του γίνεται με βάση τον διαδοχικό συνδυασμό πολλών αδύναμων μοντέλων αναζητώντας το μοντέλο που το μοντέλο που δίνει τα καλύτερα αποτελέσματα, διαδικασία η οποία ονομάζεται ενίσχυση κλίσης. Λόγω της διαδοχικότητας που ακολουθεί η διαδικασία ενίσχυσης κλίσης, τα δένδρα προσαρμόζονται μαθαίνοντας από τα λάθη των προηγούμενων μειώνοντας με αυτόν τον τρόπο το σφάλμα. (Thiesen S., 2021).

Το Cat Boost δημιουργεί ισορροπημένα, συμμετρικά δέντρα, τα οποία το ξεχωρίζουν από άλλους αλγόριθμους ενίσχυσης κλίσης, όπως ο XG Boost και ο Light GBM. Αντίστοιχα, το ζεύγος διαχωρισμού χαρακτηριστικών που δίνει τη χαμηλότερη απώλεια σε κάθε φάση επιλέγεται και εφαρμόζεται σε όλους τους κόμβους σε αυτό το επίπεδο. Αυτός ο ισορροπημένος σχεδιασμός προσφέρει μια σειρά από πλεονεκτήματα, όπως αποτελεσματική υλοποίηση CPU, ταχύτερη εφαρμογή μοντέλου, μικρότερους χρόνους πρόβλεψης και τακτοποίηση για την ελαχιστοποίηση του σφάλματος. (Oppermann, 2023). Στην παρακάτω εικόνα αναπαρίσταται η διαδικασία συνδυασμού των διαφόρων αδύναμων μοντέλων μέσω CatBoost.



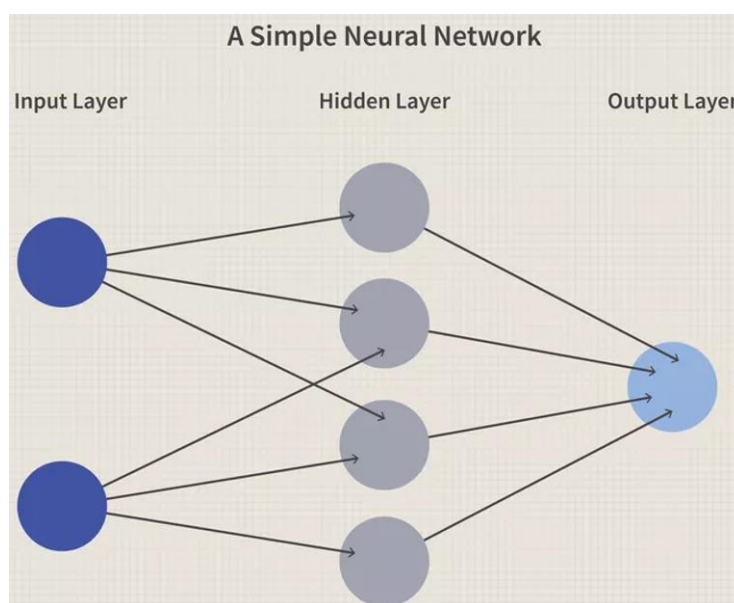
Εικόνα 4.3.5.1. CatBoost (catboost.ai, 2018)

4.3.6. Νευρωνικά Δίκτυα (Neural Networks).

Η γραμμική παλινδρόμηση μπορεί να μάθει μόνο τη γραμμική σχέση μεταξύ των χαρακτηριστικών και του στόχου και δεν μπορεί να μάθει την περίπλοκη μη γραμμική σχέση. Έτσι, για να μελετήσουμε την μη γραμμική σχέση χρησιμοποιούμε τεχνικές όπως τα τεχνητά νευρωνικά δίκτυα τα οποία προτιμώνται για παλινδρόμηση από τη γραμμική παλινδρόμηση σε αυτήν την περίπτωση. Η ονομασία της τεχνικής βασίζεται στην ικανότητά της για βαθιά μάθηση που μιμείται τις δράσεις των νευρώνων στον ανθρώπινο εγκέφαλο. Τα νευρωνικά δίκτυα Vanilla και τα επαναλαμβανόμενα νευρωνικά δίκτυα είναι δύο παραδείγματα τεχνητών νευρωνικών δικτύων. (Srivignesh R., 2021).

Η παλινδρόμηση νευρωνικού δικτύου αποτελείται από πολλά συνδεδεμένα στρώματα τεχνητών νευρώνων, καθένα από τα οποία εφαρμόζει μια μη γραμμική συνάρτηση ενεργοποίησης και διεξάγει έναν σταθμισμένο υπολογισμό για να παρέχει μια έξοδο. Ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου είναι οι παραδοσιακές ομαδοποιήσεις για αυτά τα επίπεδα. Το δίκτυο μπορεί να μάθει περίπλοκες αναπαραστάσεις και συσχετίσεις στα δεδομένα χάρη στα κρυφά επίπεδα.

Το επίπεδο εισόδου του νευρωνικού δικτύου είναι το αρχικό του και λαμβάνει δεδομένα εισόδου, τα οποία μπορεί να είναι χαρακτηριστικά ή χαρακτηριστικά ενός συγκεκριμένου ζητήματος. Τα επίπεδα που έρχονται μετά το επίπεδο εισόδου και πριν από το επίπεδο εξόδου είναι γνωστά ως κρυφά επίπεδα. Επειδή δεν συνδέονται άμεσα με την είσοδο ή την έξοδο, είναι γνωστά ως "κρυμμένα" στοιχεία. Τα δεδομένα εισόδου υποβάλλονται σε επεξεργασία και μετασχηματίζονται από κρυφά επίπεδα, τα οποία μαθαίνουν και εξάγουν σημαντικά μοτίβα ή χαρακτηριστικά. Το επίπεδο εξόδου του νευρωνικού δικτύου είναι το τελευταίο επίπεδο. Με βάση τους υπολογισμούς και τις αλλαγές που έγιναν στα κρυφά επίπεδα, προσφέρει την πρόβλεψη ή την έξοδο του δικτύου. (Chen J., 2023). Σημαντικό για αναφορά είναι η ύπαρξη βαρών στους νευρώνες οι οποίοι αντιπροσωπεύουν τις πραγματικές αξίες κάθε στοιχείου και την βαρύτητα – σημασία που έχουν αυτά για την διαμόρφωση του τελικού αποτελέσματος. Η εικόνα 4.3.6.1. παρουσιάζει παράδειγμα αναπαράστασης ενός νευρωνικού δικτύου.



Εικόνα 4.3.6.1. Neural Network Layers (Chen J., 2023).

Κατευθυνόμενα βέλη ή συνδέσεις συνδέουν τους νευρώνες μεταξύ τους. Ένας νευρώνας λαμβάνει είσοδο από άλλους νευρώνες μέσω συνδέσεων. Μέσω των συνδέσεων, επικοινωνεί επίσης την τιμή της εξόδου του σε γειτονικούς νευρώνες. Το βάρος w είναι ένας αριθμητικός αριθμός που εκχωρείται σε κάθε σύνδεσμο. Το βάρος επηρεάζει τον τρόπο αλληλεπίδρασης των συνδεδεμένων νευρώνων. Εάν η τιμή εξόδου του νευρώνα j , u_j , αποσταλεί στον νευρώνα i , τότε το u_j πολλαπλασιάζεται με την ισχύ της σύνδεσης μεταξύ των δύο νευρώνων, w_{ij} . Δύο στάδια συνθέτουν την επεξεργασία που κάνει ένας νευρώνας i . Οι τιμές εισόδου αθροίζονται στο αρχικό στάδιο. Τα βάρη των σχετικών συνδέσεων πολλαπλασιάζονται με τις τιμές εισόδου, οι οποίες είναι ισοδύναμες με τις τιμές εξόδου των συνδεδεμένων νευρώνων. Χρησιμοποιώντας μια συνάρτηση που αναφέρεται ως συνάρτηση ενεργοποίησης (συνάρτηση ενεργοποίησης) ή συνάρτηση μετασχηματισμού, το σύνολο των τιμών εισόδου μετατρέπεται στο δεύτερο βήμα. (Κύρκος Ε., 2015). Σύμφωνα με την εξίσωση της εικόνας 4.3.6.2., προσδιορίζεται το συνολικό σήμα εισόδου S_i για έναν νευρώνα i που λαμβάνει τιμές εισόδου u_j από n νευρώνες.

$$S_i = \sum_{j=1}^n w_{ij} * u_j$$

Εικόνα 4.3.6.2. Συνολικό Σήμα Εισόδου (Κύρκος Ε., 2015).

Μπορεί να χρησιμοποιηθεί μια ποικιλία συναρτήσεων ως λειτουργίες ενεργοποίησης του νευρωνικού δικτύου. Η συνάρτηση ημιτόνου, συνημίτονου, η συνάρτηση υπερβολικής εφαπτομένης κ.λπ. είναι παραδείγματα αυτών των συναρτήσεων. Ωστόσο, συνήθως χρησιμοποιείται η σιγμοειδής συνάρτηση που παρουσιάζεται στην εικόνα 4.3.6.3.

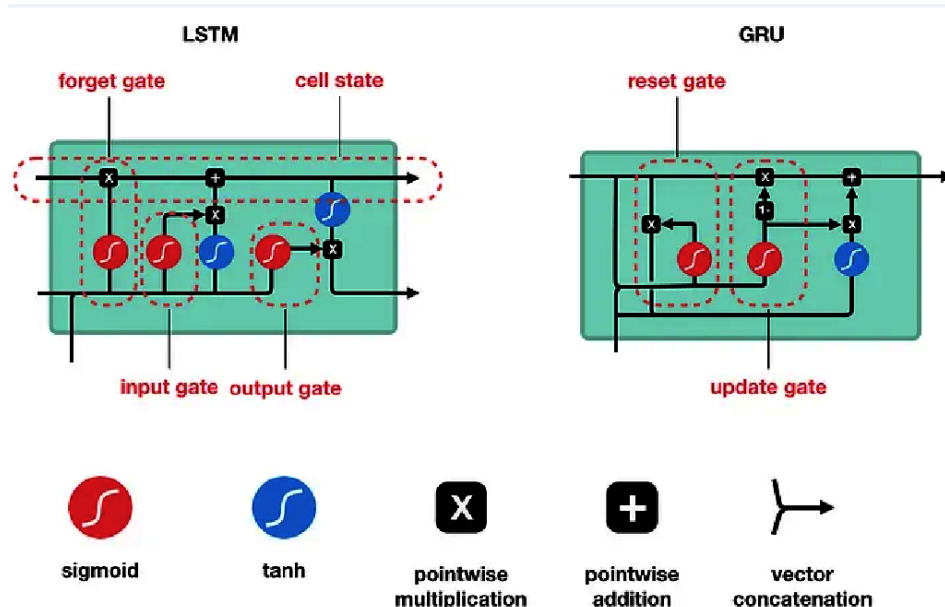
$$f(x) = \frac{1}{1 + e^{-x}}$$

Εικόνα 4.3.6.3. Σιγμοειδής Συνάρτηση (Κύρκος Ε., 2015).

4.3.7. Παλινδρόμηση με GRU & LSTM.

Στο παρόν υποκεφάλαιο αναλύεται η Παλινδρόμηση με Μονάδες Επαναλήψεων με Πύλες (Gated Recurrent Units - GRU) & Μακροπρόθεσμη Μνήμη Μικρού Χρονικού Διαστήματος (Long Short-Term Memory - LSTM). Με την χρήση νευρωνικών δικτύων μπορεί να εμφανιστεί το πρόβλημα της βραχυπρόθεσμης μνήμης, καθώς εάν μία ακολουθία είναι μεγάλη καθίσταται δύσκολη η μετάβαση πληροφοριών από προηγούμενα σε επόμενα βήματα του νευρωνικού δικτύου.

Το GRU & LSTM αντιμετωπίζουν το πρόβλημα της βραχυπρόθεσμης μνήμης με την χρήση εσωτερικών μηχανισμών που ονομάζονται πύλες και ελέγχουν τις πληροφορίες και την ροή τους. Η κατάσταση κυττάρου και οι πολλαπλές πύλες της αποτελούν τη θεμελιώδη ιδέα των LSTM & GRU. Η κατάσταση κυττάρου λειτουργεί ως αυτοκινητόδρομος για τη μεταφορά σχετικών πληροφοριών σε ολόκληρη την αλυσίδα αλληλουχίας. Από αυτό αποτελείται η «μνήμη» του δικτύου. Η κατάσταση κυττάρου μπορεί, κατ' αρχήν, να μεταφέρει σημαντικές πληροφορίες κατά την επεξεργασία της ακολουθίας. Επομένως, τα οφέλη της βραχυπρόθεσμης μνήμης μειώνονται, καθώς ακόμη και η γνώση από προηγούμενα χρονικά βήματα μπορεί να ταξιδέψει σε μεταγενέστερα χρονικά βήματα. (Phi M., 2018).



Εικόνα 4.3.7.1. LSTM & GRU (Phi M., 2018).

Στην παραπάνω εικόνα παρουσιάζονται τα διαγράμματα της διαδικασίας που ακολουθούν οι αλγόριθμοι LSTM (Long Short – Term Memory Networks – Δίκτυα Μακροπρόθεσμης και Βραχυπρόθεσμης μνήμης) & GRU (Gated Recurrent Unit ή Μονάδα Επαναλαμβανόμενης Θύρας) για την ανάλυση δεδομένων. Όπου sigmoid έχουμε την σιγμοειδή συνάρτηση που αναφέρθηκε προηγούμενος.

Όπου pointwise multiplication & pointwise addition έχουμε τον πολλαπλασιασμό και την πρόσθεση δεδομένων και τέλος όπου vector concatenation έχουμε την διανυσματική ένωση όπου ενώνονται όλα τα αποτελέσματα και προχωράνε στο επόμενο στάδιο μέχρι το τελικό.

Τα μοντέλα GRU & LSTM χρησιμοποιούν διάφορες εξισώσεις για το κάθε είδους πύλη που χρησιμοποιούν. Οι πύλες επιτρέπουν στο σύστημα να επεξεργαστεί κατάλληλα τις πληροφορίες κατά την διάρκεια της ροής. Στον αλγόριθμο LSTM 3 υπάρχουν διάφορες πύλες και σημαντικές είναι: η πύλη εισόδου που ρυθμίζει το πόσο νέα πληροφορία εισέρχεται στο σύστημα, η πύλη λήθης που αποφασίζει ποιες πληροφορίες αγνοούνται λόγω μη χρησιμότητας και η πύλη εξόδου που αποφασίζει το ποσό της πληροφορίας που θα χρησιμοποιηθεί. Στον GRU υπάρχει η πύλη ενημέρωσης που αποφασίζει το ποσό χρήσης προηγούμενης πληροφορίας και η πύλη επαναφοράς η οποία εξετάζει ξανά τα δεδομένα της πύλης λήθης και αποφασίζει αν κάποια από αυτά θα ξεχαστούν. (Phi M., 2018). Παρακάτω παρουσιάζονται οι μαθηματικές εξισώσεις της κάθε πύλης:

LSTM:

1. Πύλη εισόδου: $i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t + b_i])$
2. Πύλη λήθης: $f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, x_t + b_f])$
3. Πύλη εξόδου: $o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t + b_o])$

Η κρυφή κατάσταση στο χρονικό διάστημα t αντιπροσωπεύεται από h_t και τα βάρη των νευρώνων αντιπροσωπεύονται από τα W_i , W_f , και W_o . Sigmoid είναι η

σιγμοειδής συνάρτηση και h_{t-1} είναι η έξοδος δεδομένων από το προηγούμενο βήμα. Το x_t είναι η είσοδος δεδομένων σε ορισμένο χρόνο και b_x είναι οι παρεμβολές για τις αντίστοιχες πύλες οι οποίες αποτελούν παραμέτρους που προθέτονται στο άθροισμα των βαρών πριν την ενεργοποίηση. (Thakur, 2018).

GRU:

1. Πύλη ενημέρωσης: $z_t = \text{sigmoid}(W_z \cdot x_t + U_z \cdot [h_{t-1}])$
2. Πύλη επαναφοράς: $r_t = \text{sigmoid}(W_r \cdot x_t + U_r \cdot [h_{t-1}])$

Τέλος, η κρυφή κατάσταση για την οποία γίνεται λόγος αποτελεί το βήμα κατά το οποίο μεταφέρεται πληροφορία από προηγούμενα σε επόμενα βήματα. (Saxena, 2021). Όταν το x_t είναι συνδεδεμένο στη μονάδα δικτύου, το βάρος του $W(z)$ διπλασιάζεται. Το ίδιο ισχύει για το h_{t-1} , το οποίο πολλαπλασιάζεται με το δικό του βάρος $U(z)$ και περιέχει δεδομένα για τις προηγούμενες μονάδες $t-1$. Μαζί, τα δύο αποτελέσματα αθροίζονται και το αποτέλεσμα συμπιέζεται μεταξύ 0 και 1 χρησιμοποιώντας μια σιγμοειδή συνάρτηση ενεργοποίησης. Η ίδια μεθοδολογία ισχύει και για την πύλη επαναφοράς, όπου αντί για z χρησιμοποιείται r (Kostadinov, 2017).

5. Βιβλιογραφική επισκόπηση.

5.1. Murthy et al.

Στο άρθρο που παρουσιάζεται από τους Murthy et al., γίνεται λόγος για τη σημασία της διανομής δανείων για τις τράπεζες και τονίζεται η ζήτηση για ένα αξιόπιστο σύστημα έγκρισης δανείων. Οι παραδοσιακές μέθοδοι επικύρωσης και επαλήθευσης δεν μπορούν να εξασφαλίσουν την ασφάλεια των αιτούντων δανείων. Το σύστημα πρόβλεψης δανείων με χρήση Python παρουσιάζεται ως λύση σε αυτό το πρόβλημα.

Το λογισμικό συστήματος πρόβλεψης δανείου αξιολογεί την ικανότητα ενός πελάτη να αποπληρώσει ένα δάνειο λαμβάνοντας υπόψη μια ποικιλία χαρακτηριστικών, όπως η οικογενειακή κατάσταση, το εισόδημα, οι δαπάνες και άλλα πράγματα. Χρησιμοποιεί ένα εκπαιδευτικό σύνολο δεδομένων για τη δημιουργία ενός μοντέλου, το οποίο στη συνέχεια χρησιμοποιείται σε ένα δοκιμαστικό σύνολο δεδομένων για να εξακριβωθεί εάν ένας καταναλωτής είναι ικανός ή όχι να αποπληρώσει το δάνειο. Για την ανάλυση των δεδομένων χρησιμοποιούνται τεχνικές μηχανικής μάθησης όπως το Decision Tree, Random Forests και K Nearest Neighbor. Δεδομένου του μεγάλου αριθμού αιτούντων για δάνεια στον τραπεζικό κλάδο, αυτοί οι αλγόριθμοι καθιστούν δυνατή την απόκτηση γρήγορων και ακριβών αποτελεσμάτων για τις εγκρίσεις δανείων. Η χρήση της μεθόδου Random Forest για την ανάπτυξη ενός μοντέλου πρόβλεψης δανείων επισημαίνεται στη μελέτη. Επειδή μπορεί να παράγει πιο ακριβή αποτελέσματα από άλλους αλγόριθμους, ευνοείται το Random Forest. Το συμπέρασμα του άρθρου τονίζει τις δυνατότητες του συστήματος για χρήση σε διαδικτυακές εφαρμογές τραπεζικών συναλλαγών και συμβουλεύει τη χρήση άλλων αλγορίθμων για περαιτέρω βελτίωση της ακρίβειας. (Murthy et al., 2020).

Η πιο βασική διαφορά της έρευνας μεταξύ του επιστημονικού άρθρου και της παρούσας εργασίας είναι το γεγονός ότι στο άρθρο χρησιμοποιούνται τεχνικές μηχανικής μάθησης για την λύση προβλήματος ταξινόμησης, ενώ στην εργασία χρησιμοποιούνται με σκοπό την εκτέλεση παλινδρόμησης. Το άρθρο εστιάζει στην έγκριση ή μη των πελατών για την λήψη δανείου, ενώ το παρόν έγγραφο αναλύει δεδομένα με σκοπό την εύρεση όσο το δυνατόν μικρότερου σφάλματος για τον υπολογισμό δανείου. Τέλος, παρόλο των διαφορών που έχει η εργασία με το άρθρο, αυτή συνεισφέρει στην έρευνα του άρθρου καθώς παρουσιάζει ακόμα πιο σύγχρονες τεχνικές μηχανικής μάθησης όπως XG Boost, Light G Boost ή Cat Boost.

5.2. Farjana & Mamun.

Σύμφωνα με τους Farjana & Mamun είναι σημαντικό για τις τράπεζες να αξιολογούν γρήγορα τις αιτήσεις δανείων και να διαπιστώνουν την επιλεξιμότητα των αιτούντων καθώς αυξάνεται η ζήτηση για τραπεζικά δάνεια. Το άρθρο επικεντρώνεται στην εξαγωγή μοτίβων από ένα σύνολο δεδομένων αιτούντων που έχουν εγκριθεί από δάνειο, προκειμένου να προβλεφθούν αξιόλογοι αιτούντες για το δάνειο χρησιμοποιώντας τεχνικές Μηχανικής Μάθησης (ML). Για την εύρεση των πιο σημαντικών χαρακτηριστικών που επηρεάζουν την έξοδο πρόβλεψης, χρησιμοποιούνται διάφορες τεχνικές ML, όπως Random Forest, XGBoost, Adaboost, Light G Boost, Decision tree και K-Nearest Neighbor. Ο πιο αποτελεσματικός αλγόριθμος σύμφωνα στο σύνολο δεδομένων αποδείχθηκε ο Light G boost. Οι αλγόριθμοι μηχανικής μάθησης και η τεχνητή νοημοσύνη (AI – Artificial Inteligence)

έχουν αναπτυχθεί ευνοϊκά ως αποτελεσματικά εργαλεία για την ανάλυση δεδομένων. Οι διαδικασίες για την έγκριση τραπεζικών δανείων διαρκούν πολύ και είναι επιρρεπείς σε λάθη. Προκειμένου οι τράπεζες να μειώσουν τον κίνδυνο και να κάνουν σοφές κρίσεις, είναι σημαντικό να σχεδιάσουν ένα ακριβές και αποτελεσματικό σύστημα πρόβλεψης δανείων που χρησιμοποιεί τεχνικές μηχανικής μάθησης.

Ο κύριος στόχος του άρθρου είναι να προσφέρει μια γρήγορη, εύκολη και αποτελεσματική τεχνική επιλογής των επιλέξιμων αιτούντων δανείου. Όπως στο προηγούμενο έτσι και στο παρόν άρθρο υπολογίζεται η πιθανότητα να εγκριθεί ένα δάνειο εξετάζοντας προηγούμενα δεδομένα πελατών, συμπεριλαμβανομένης της ηλικίας, του τύπου εισοδήματος, του δανείου, του πιστωτικού ιστορικού και της διάρκειας εργασίας. Αυτό δίνει τη δυνατότητα στους εργαζόμενους στις τράπεζες να αξιολογούν και να ιεραρχούν κατάλληλα τις αιτήσεις. Τα αποτελέσματα αυτής της μελέτης δείχνουν πόσο καλά οι αλγόριθμοι ML προβλέπουν την δυνατότητα επιλογής ενός αιτούντος τραπεζικού δανείου. Οι τράπεζες μπορούν να κάνουν σοφές κρίσεις και να διανείμουν τους πόρους τους σωστά χρησιμοποιώντας τεχνικές μηχανικής μάθησης, οι οποίες τελικά θα αυξήσουν την κερδοφορία τους και θα μειώσουν τις ζημιές τους. (Farjana & Mamun, 2022)

Η συγκεκριμένη μελέτη χρησιμοποιεί σύγχρονες τεχνικές μηχανικής μάθησης όπως και η παρούσα εργασία. Μάλιστα γίνεται εκτενείς απεικόνιση των αποτελεσμάτων των αλγορίθμων όπως στην έρευνα της πτυχιακής Βέβια δε γίνεται η χρήση του αλγορίθμου GRU & LSTM στο άρθρο καθώς επίσης στην πτυχιακή οι τεχνικές μηχανικής μάθησης χρησιμοποιούνται για παλινδρόμηση και όχι για ταξινόμηση. Τέλος, η εργασία συνεισφέρει σε θεωρητικό επίπεδο στο άρθρο καθώς βοηθάει τον αναγνώστη να κατανοήσει τον τρόπο λειτουργίας του τραπεζικού τομέα και των τεχνικών μηχανικής μάθησης που χρησιμοποιούνται στο άρθρο.

5.3. Tejaswini et al.

Το άρθρο υποστηρίζει την χρήση της ανάλυσης δεδομένων για την πρόβλεψη των αποτελεσμάτων χρησιμοποιώντας τεχνικές μηχανικής μάθησης. Προκειμένου να προβλεφθεί η αποδοχή δανείου για πελάτες, τρεις αλγόριθμοι μηχανικής μάθησης, Logistic Regression (LR), Decision Tree (DT) και Random Forest (RF), χρησιμοποιήθηκαν σε αυτή τη μελέτη. Σύμφωνα με πειραματικά ευρήματα, ο αλγόριθμος Decision Tree είναι πιο ακριβής από την Logistic Regression και το Random Forest.

Επίσης αναφέρονται και αλγόριθμοι όπως, τα δέντρα απόφασης, το τυχαίο δάσος, η μηχανή υποστήριξης διανυσμάτων, τα γραμμικά μοντέλα, το νευρωνικό δίκτυο και το Adaboost ως μοντέλα ταξινόμησης μηχανικής μάθησης που χρησιμοποιούνται σε αυτή τη μελέτη. Κάθε μοντέλο έχει μοναδικά πλεονεκτήματα και ιδιότητες. Υποστηρίζεται πως το δημιουργημένο σύστημα πρόβλεψης δανείων επιδεικνύει εξαιρετική αποτελεσματικότητα και ικανοποιεί τις ανάγκες των χρηματοπιστωτικών οργανισμών. Προσφέρει ασφάλεια, αξιοπιστία και ικανότητα δυναμικής τροποποίησης της χωρητικότητας βάρους και μπορεί να ενσωματωθεί ομαλά σε μια ποικιλία συστημάτων. Βέβια, οι συγγραφείς αφήνουν το περιθώριο βελτίωσης της έρευνάς τους αφήνοντας ανοιχτό το ενδεχόμενο εμφάνισης καλύτερης τεχνικής από αυτές που χρησιμοποίησαν οι ίδιοι. (Tejaswini et al., 2020).

Το συγκεκριμένο άρθρο αναλύει την σημασία χρήσης των τεχνικών μάθησης στον τραπεζικό τομέα καθώς και παρουσιάζει ορισμένα μοντέλα πάνω σε ένα σύνολο δεδομένων ως παράδειγμα όπως στην παρούσα εργασία. Η συνεισφορά της εργασίας στο άρθρο αφορά αρχικά το θεωρητικό κομμάτι καθώς επεξηγείται και αναλύεται σε βάθος ο τραπεζικός τομέας και ο τρόπος με τον οποίο λειτουργούν οι τεχνικές μηχανικής μάθησης. Τέλος, εκτός από τις τρεις τεχνικές που παρουσιάζονται στο άρθρο η εργασία συνεισφέρει με περισσότερες και πιο πρόσφατες.

5.4. Kadam et al.

Σύμφωνα με το άρθρο των Kadam et al., 2021, η πρόβλεψη των κακοπληρωτών δανείων παίζει καθοριστικό ρόλο στο τραπεζικό σύστημα. Οι τράπεζες μπορούν να μειώσουν τους κινδύνους τους και να κάνουν σοφές κρίσεις σχετικά με τις εγκρίσεις δανείων εντοπίζοντας με ακρίβεια πιθανούς κακοπληρωτές. Η προγνωστική ανάλυση, ιδιαίτερα το μοντέλο Nave Bayes, αποτελεί την προτροπή των συγγραφέων για να κάνουν ανάλυση των δεδομένων καθώς υποστηρίζουν πως λειτουργεί καλά στην πρόβλεψη δανείων, σύμφωνα με προηγούμενη έρευνα σε αυτόν τον τομέα. Οι σημαντικές φάσεις στη χρήση αυτού του παραδείγματος είναι η συλλογή δεδομένων, ο καθαρισμός δεδομένων και η αξιολόγηση της απόδοσης. Αναφέρονται επίσης μοντέλα όπως ταξινόμηση, λογιστική παλινδρόμηση, και δέντρα αποφάσεων.

Γενικά το άρθρο παρουσιάζει θεωρητικά κομμάτια και βγάζει συμπέρασμα πως το μοντέλο Naive Bayes ξεχωρίζει ως αξιόπιστη και επιτυχημένη στρατηγική για την πρόβλεψη δανείων. Η λήψη μορφωμένων κρίσεων σχετικά με τις εγκρίσεις δανείων γίνεται ευκολότερη με τη μείωση των κινδύνων, την αύξηση των κερδών και τη χρήση αυτών των πληροφοριών.

Η πτυχιακή διαφέρει από το άρθρο καθώς παρουσιάζει πρόβλημα παλινδρόμησης και όχι ταξινόμησης. Παρόλα αυτά συνεισφέρει στο άρθρο καθώς αναλύει περισσότερο τον τραπεζικό τομέα και τις τεχνικές μηχανικής μάθησης. Τέλος, στην εργασία παρουσιάζεται παράδειγμα εκτός από θεωρητικό υπόβαθρο το οποίο συμβάλλει στην γενική παρουσίαση των αλγορίθμων των τεχνικών.

5.5. Gautam & Nidhi.

Οι συγγραφείς του άρθρου τονίζουν πως βασικός στόχος των τραπεζών είναι να ελαχιστοποιήσουν τον κίνδυνο αποτυχίας επενδύοντας τα κεφάλαιά τους σε ασφαλείς επιχειρήσεις. Η διαδικασία έγκρισης δανείου εξακολουθεί να είναι δύσκολη και δεν υπάρχει καμία διαβεβαίωση ότι ο δανειολήπτης θα είναι αξιόπιστος ή σε θέση να αποπληρώσει το δάνειο εντός της συμφωνηθείσας προθεσμίας. Σύμφωνα με τις απαιτήσεις του πελάτη, τα δάνεια μπορούν χονδρικά να χωριστούν σε δάνεια αορίστου και κλειστού τύπου. Το μοντέλο μηχανικής μάθησης χρησιμοποιεί δώδεκα παράγοντες, συμπεριλαμβανομένου του φύλου, της οικογενειακής κατάστασης, της εκπαίδευσης και του εισοδήματος. Κατά τη διαδικασία αίτησης δανείου, το μοντέλο περνά από μια φάση εκπαίδευσης με γνωστές περιπτώσεις πριν αξιολογηθεί με νέα δεδομένα αιτούντος. Αυτό δίνει τη δυνατότητα στον αλγόριθμο, χρησιμοποιώντας μια ποικιλία συνόλων δεδομένων εκπαίδευσης, να αξιολογήσει την ασφάλεια της αποδοχής ενός δανείου.

Το σύνολο δεδομένων εκπαίδευσης, το σύνολο δεδομένων δοκιμής, η επεξεργασία δεδομένων και η επιλογή χαρακτηριστικών είναι οι τέσσερις κύριες φάσεις στις οποίες περνά το σύστημα πρόβλεψης δανείων. Προκειμένου να αποφευχθούν λάθη, η προ επεξεργασία δεδομένων περιλαμβάνει τη διαγραφή διπλότυπων, ανεπαρκών τιμών ή τιμών που λείπουν. Το άρθρο παρουσιάζει ανάλυση δεδομένων με δένδρα αποφάσεων και με τυχαίο δάσος. Προκειμένου να βελτιωθεί η αποτελεσματικότητα και η ασφάλεια στον τραπεζικό κλάδο, αυτή η μελέτη εστιάζει στην κατηγοριοποίηση και αξιολόγηση των αιτήσεων δανείων.

Επιπρόσθετα μεγέθη αποτελούν οι υπόλοιποι αλγόριθμοι που αναλύουμε στην παρούσα εργασία. Οι αλγόριθμοι τυχαίου δάσους και δέντρων αποφάσεων αναλύονται ήδη, αλλά υπάρχει συνεισφορά στο άρθρο, μέσω της εργασίας, με πιο σύγχρονες τεχνικές. Το άρθρο αναλύει αρκετά τον τραπεζικό τομέα όπου η εργασία επίσης συνεισφέρει αναλύοντας τα είδη των δανείων και τον κλάδο.

5.6. Dutta P.

Ο τομέας δανεισμού εξαρτάται από την παροχή δανείων σε δανειολήπτες με προσδοκία τόκων και αποπληρωμής συν αποπληρωμή (Dutta, 2021). Ωστόσο, οι δανειστές έχουν ένα σημαντικό πρόβλημα λόγω του κινδύνου αθέτησης των δανείων των καταναλωτών. Η σημασία των αλγορίθμων μηχανικής μάθησης έγκειται στην ικανότητά τους να κατασκευάζουν μοντέλα χρησιμοποιώντας ιστορικά δεδομένα, επιτρέποντας καλύτερες προβλέψεις όχι απαραίτητα μόνο για πιστωτικούς κινδύνους αλλά και για άλλους κινδύνους.

Το έργο χρησιμοποιεί ένα σύνολο δεδομένων που προέρχεται από το Kaggle που χωρίζεται σε δύο σύνολα δεδομένων (εκπαίδευση και δοκιμή) όπως και στην παρούσα εργασία, το καθένα με 13 στήλες που αντικατοπτρίζουν διαφορετικές ιδιότητες πελατών και κατάσταση δανείου. Για την πρόβλεψη δανείων, χρησιμοποιούνται τρεις εποπτευόμενοι αλγόριθμοι ταξινόμησης — Λογιστική Παλινδρόμηση, Δέντρο Αποφάσεων και Τυχαίο Δάσος. Ο συγγραφέας υποστηρίζει πως αν και αυτό το έργο χρησιμοποιεί τις προεπιλεγμένες ρυθμίσεις υπερπαραμέτρων. Ως βέλτιστος αλγόριθμος για την πρόβλεψη δανείων με χρήση μηχανικής μάθησης καθορίζεται ως η λογιστική παλινδρόμηση στην ορισμένη έρευνα. Με βάση τις ιδιότητες των καταναλωτών και τα προηγούμενα δεδομένα, οι δανειστές μπορούν να χρησιμοποιήσουν αυτά τα μοντέλα για να λάβουν ενημερωμένες αποφάσεις, να μειώσουν τους κινδύνους και να αυξήσουν τα ποσοστά αποπληρωμής δανείων. Η πτυχιακή εργασία συνεισφέρει προσθέτοντας την χρήση μοντέλων όπως το XG Boost, το Light G Boost, τους GRU & LSTM και το Cat Boost, καθώς και την πιο εκτεταμένη ανάλυση των τεχνικών και του τρόπου λειτουργίας τους.

5.7. Sharma H.

Το άρθρο υποστηρίζει πως για τις τράπεζες, η κατανομή των δανείων είναι σημαντικός παράγοντας αφού επηρεάζει άμεσα την κερδοφορία τους. Οι τράπεζες πρέπει να επιλέγουν προσεκτικά τους δανειολήπτες βάσει πτυχών, συμπεριλαμβανομένης της ικανότητάς τους να αποπληρώσουν το δάνειο και της οικονομικής κατάστασης για να εξασφαλίσουν μέγιστο επιτόκιο και να μειώσουν τους οικονομικούς κινδύνους. Για την εξαγωγή σχετικών πληροφοριών από τεράστια

σύνολα δεδομένων, χρησιμοποιούνται τεχνικές εξόρυξης δεδομένων και το μοντέλο εκπαιδεύεται χρησιμοποιώντας προηγούμενα αρχεία πελατών. Το μοντέλο πρόβλεψης δανείων δημιουργείται χρησιμοποιώντας τεχνικές ταξινόμησης όπως δέντρα αποφάσεων, νευρωνικά δίκτυα, μηχανές διανυσμάτων υποστήριξης, λογιστική παλινδρόμηση και k-πλησιέστερα γείτονες.

Το συνιστάμενο σύστημα πρόβλεψης δανείων που βασίζεται στη μηχανική μάθηση προσφέρει έναν πιο ακριβή και αποτελεσματικό τρόπο επιλογής υποψηφίων δανείων, υπό το φως όλων των παραπάνω. Η λογιστική παλινδρόμηση είναι η μέθοδος που προβλέπει καλύτερα τα δάνεια σύμφωνα με το άρθρο. Η εφαρμογή αυτής της στρατηγικής θα επιτρέψει στις τράπεζες και τους πελάτες να λάβουν άμεσες αποφάσεις που θα βελτιώσουν τη συνολική διαδικασία έγκρισης δανείων. Η πτυχιακή συμβάλλει προσθέτοντας τεχνικές όπως και η γραμμική παλινδρόμηση καθώς ζητάμε υπολογισμό δανείου και όχι απλά εξέταση των αιτούντων. Τέλος, συμβάλλει μέσω πιο σύγχρονων τεχνικών που εμπλουτίζουν την έρευνα του άρθρου.

5.8. Anand, Velu & Whig.

Το άρθρο τονίζει πως το τμήμα δανεισμού πιστώσεων του τραπεζικού κλάδου γνώρισε ταχεία επέκταση και έντονο ανταγωνισμό από νέες πιστωτικές επιχειρήσεις. Αυτό συνοδεύτηκε από άνοδο των αιτήσεων για δάνεια και του δανεισμού, γεγονός που αύξησε τις ζημιές λόγω κακής πίστωσης. Τα πιστωτικά δάνεια αποπληρώνονται σε ορισμένο χρονικό διάστημα με ή χωρίς τόκους και δίνονται από τράπεζες ή άλλους χρηματοπιστωτικούς οργανισμούς. Οι τράπεζες και τα χρηματοπιστωτικά ιδρύματα πρέπει να δημιουργήσουν αποτελεσματικά μοντέλα χρησιμοποιώντας έρευνα και ήδη διαθέσιμα δεδομένα για τη μείωση του κινδύνου κακής πίστωσης, προκειμένου να αντιμετωπίσουν την αύξηση των αιτήσεων δανείων και την ανταγωνιστικότητα.

Προκειμένου να βρεθούν πρότυπα στα δεδομένα δανεισμού πιστώσεων και να εντοπιστούν παράγοντες που επηρεάζουν τις αθετήσεις των δανείων, πρέπει να αναπτυχθούν μοντέλα μηχανικής μάθησης. Έχουν χρησιμοποιηθεί πολυάριθμες μέθοδοι μηχανικής μάθησης για ταξινόμηση, όπως η λογιστική παλινδρόμηση, τα δέντρα αποφάσεων, τα τυχαία δάση, τα νευρωνικά δίκτυα και οι μηχανές διανυσμάτων υποστήριξης. Σύμφωνα με το άρθρο διαφορετικά μοντέλα αποδίδουν σε διάφορους βαθμούς, το τυχαίο δάσος έχει αποδειχθεί ότι είναι απίστευτα ακριβές στην πρόβλεψη του πιστωτικού κινδύνου.

Στην ενότητα τεχνικής, εξηγείται πώς χρησιμοποιήθηκαν μέθοδοι ταξινόμησης Python για τη συλλογή, την επεξεργασία από προηγούμενο στάδιο και την αξιολόγηση των δεδομένων. Προκειμένου να προβλεφθεί σωστά η αθέτηση τραπεζικών δανείων, αυτή η έρευνα χρησιμοποίησε πολλά συστήματα κατηγοριοποίησης. Για τη μελέτη χρησιμοποιήθηκε Python και υπολογίστηκαν μετρήσεις απόδοσης. Η κατάσταση της εργασίας, το εισόδημα, η αναλογία χρέους προς εισόδημα και η αναλογία πίστωσης προς χρέος ήταν οι πιο κρίσιμοι παράγοντες για την πρόβλεψη της αθέτησης των δανείων. Αυτά τα αποτελέσματα μπορούν να βοηθήσουν τις τράπεζες και άλλους χρηματοπιστωτικούς οργανισμούς να κάνουν καλύτερες κρίσεις και να διαχειριστούν τον πιστωτικό κίνδυνο. Η παρούσα ερευνητική εργασία συνεισφέρει σε αυτήν την έρευνα μέσω της περεταίρω θεωρητικής ανάλυσης του τραπεζικού τομέα, των τεχνικών μηχανικής μάθησης και την μεταξύ τους σχέση. Τέλος, συνεισφέρει μέσω της

δοκιμής αρκετά πρόσφατων τεχνικών μηχανικής μάθησης προσθέτοντάς τες στο γνωστικό περιθώριο του εν λόγω άρθρου.

5.9. Majumdar et al.

Το άρθρο καλύπτει την αξία της πρόβλεψης δανείων στον τραπεζικό τομέα και πώς μπορεί να επιτευχθεί η αυτοματοποίηση της διαδικασίας χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης. Ο κύριος στόχος είναι να αξιολογηθεί η δυνατότητα επιλογής δανείου των αιτούντων εξετάζοντας μια ποικιλία δεδομένων, συμπεριλαμβανομένου του φύλου, της οικογενειακής κατάστασης, του εισοδήματος, της εκπαίδευσης, του πιστωτικού ιστορικού και άλλων. Υπογραμμίζεται η αξία της ακριβούς πρόβλεψης δανείων για τις τράπεζες στον έλεγχο των μη εξυπηρετούμενων περιουσιακών στοιχείων και στη λήψη μορφωμένων επιλογών έγκρισης δανείων και τονίζεται η σημασία της προγνωστικής ανάλυσης και της μηχανικής μάθησης για τη δημιουργία προγνωστικών μοντέλων αθέτησης δανείων.

Οι διαδικασίες που εμπλέκονται στη διαδικασία πρόβλεψης δανείου, όπως η συλλογή δεδομένων, η εκκαθάριση δεδομένων και η αξιολόγηση απόδοσης, εξηγούνται περαιτέρω στο άρθρο. Για την πρόβλεψη δανείων, συζητά τη χρήση πολλών μεθόδων μηχανικής μάθησης όπως Naive Bayes, δέντρα αποφάσεων, λογιστική παλινδρόμηση και K-πλησιέστερους γείτονες (KNN). Τα αποτελέσματα των δοκιμών δείχνουν ότι το μοντέλο Naive Bayes υπερέχει από άλλα μοντέλα όσον αφορά την ακρίβεια πρόβλεψης δανείου. Το άρθρο εξετάζει επίσης τη χρήση της μηχανικής εκμάθησης σε διάφορους τομείς, συμπεριλαμβανομένων των επιχειρηματικών θεμάτων, της αναγνώρισης ομιλίας και του φιλτραρίσματος δεδομένων. Δηλώνει ότι η ουσιαστική γνώση μπορεί να εξαχθεί από ακατέργαστα δεδομένα χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων, όπως ταξινόμηση, ομαδοποίηση, πρόβλεψη και νευρωνικά δίκτυα.

Η σημασία της εξόρυξης δεδομένων στον τραπεζικό τομέα καλύπτεται επίσης στο έγγραφο, με έμφαση στο μάρκετινγκ, τη διαχείριση κινδύνου, τη διαχείριση σχέσεων με τους πελάτες και τον εντοπισμό απάτης. Εφίσταται η προσοχή στον τρόπο με τον οποίο η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τάσεων των πελατών, την ανάλυση της συμπεριφοράς των καταναλωτών και τη διαχείριση του κινδύνου έγκρισης δανείου. Συνολικά, η μελέτη υπογραμμίζει την αξία της πρόβλεψης δανείων για τον τραπεζικό τομέα και προσφέρει πληροφορίες για το πώς μπορούν να χρησιμοποιηθούν αλγόριθμοι μηχανικής μάθησης και τεχνικές εξόρυξης δεδομένων για την αυτοματοποίηση της διαδικασίας έγκρισης δανείου και τη βελτίωση της λήψης αποφάσεων.

Η παρούσα έρευνα συνεισφέρει στο άρθρο μέσω της πιο αναλυτικής περιγραφής του τρόπου με τον οποίο οι τεχνικές μηχανικής μάθησης λειτουργούν σε σχέση με τον τραπεζικό τομέα. Μάλιστα παρατίθενται παραπάνω τεχνικές μηχανικής μάθησης και οι πιο πρόσφατες στην έρευνα παρά στο άρθρο. Έτσι, το άρθρο και η έρευνα συμπληρώνονται μεταξύ τους ως έργα με σκοπό την καλύτερη κατανόηση του αναγνώστη.

5.10. Supriya et al.

Ο στόχος αυτής της έρευνας ήταν να χρησιμοποιήσει τεχνικές μηχανικής μάθησης και να αναλύσει μεγάλα δεδομένα από ιστορικά αρχεία δανείων για να μειώσει τον

κίνδυνο που σχετίζεται με την έκδοση δανείων. Ο στόχος ήταν να καθοριστεί εάν θα ήταν ασφαλές να εκχωρηθεί ένα δάνειο σε ένα συγκεκριμένο άτομο ή όχι. Η συλλογή δεδομένων, η σύγκριση μοντέλων, η εκπαίδευση του συστήματος με τη χρήση του καλύτερου μοντέλου και η δοκιμή ήταν τα τέσσερα τμήματα του έργου. Τα δεδομένα δανείων προβλέφθηκαν χρησιμοποιώντας μια ποικιλία μεθόδων μηχανικής μάθησης, όπως ταξινόμηση, λογιστική παλινδρόμηση, δέντρα αποφάσεων και ενίσχυση κλίσης. Η μέθοδος του δέντρου αποφάσεων, γνωστή για την ακρίβεια και την ικανότητά της να παρέχει ένα στατιστικό μοντέλο σε απλή γλώσσα, ήταν η κύρια έμφαση της έρευνας.

Το προτεινόμενο μοντέλο προσπάθησε να χρησιμοποιήσει προσεγγίσεις δένδρων αποφάσεων για να αντιμετωπίσει την πρόβλεψη του δανείου. Τα δεδομένα συλλέχθηκαν, υποβλήθηκαν σε επεξεργασία για να προσαρμοστούν τιμές και ακραίες τιμές που λείπουν, τα χαρακτηριστικά συσχετίστηκαν για την εύρεση σχετικών παραγόντων, το μοντέλο ταξινόμησης κατασκευάστηκε χρησιμοποιώντας την προσέγγιση του δέντρου αποφάσεων και προβλέφθηκαν τα προεπιλεγμένα αποτελέσματα

Τα πειραματικά ευρήματα έδειξαν ότι ο αλγόριθμος του δέντρου αποφάσεων προέβλεπε γεγονότα αθέτησης δανείων με ποσοστό ακρίβειας 81%. Διαπιστώθηκε ότι οι αιτούντες με υψηλά εισοδήματα που ζητούσαν μικρότερα δάνεια είχαν περισσότερες πιθανότητες να χορηγηθούν και ότι το πιστωτικό ιστορικό ήταν σημαντικός παράγοντας για την αποδοχή του δανείου. Συμπερασματικά, αυτή η μελέτη έδειξε πώς η μηχανική μάθηση, ειδικά η μέθοδος του δέντρου αποφάσεων, μπορεί να χρησιμοποιηθεί για την πρόβλεψη του πόσο ασφαλής είναι η προσφορά δανείων. Οι τράπεζες μπορούν να λάβουν σοφές αποφάσεις, να μειώσουν τους κινδύνους και να αυξήσουν τη διατήρηση πελατών εξετάζοντας ιστορικά δεδομένα.

Συνδυάζοντας την έρευνα του άρθρου και την έρευνα της παρούσας εργασίας ο αναγνώστης λαμβάνει μία αρκετά σφαιρική εικόνα της σημασίας της εφαρμογής τεχνικών μηχανικής μάθησης. Η εργασία συμπληρώνει το άρθρο αναλύοντας και διεξάγοντας πειράματα με περισσότερες τεχνικές. Έτσι, με την συνεισφορά της εργασίας το άρθρο εμπλουτίζεται σε θεωρητικό και ερευνητικό επίπεδο.

6. Ερευνητική μέθοδος & Αποτελέσματα Έρευνας.

6.1. Διαδικασία και μέθοδος εξαγωγής της έρευνας.

Η διαδικασία επιλογής θέματος και η συνομιλία με τον υπεύθυνο καθηγητή ήταν τα πρώτα βήματα στη μεθοδολογία έρευνας και ανάπτυξη της εργασίας με θέμα την ανάλυση τραπεζικών δεδομένων και τεχνικές μηχανικής μάθησης. Ακολούθησαν η συνολική διάταξη του περιεχομένου της εργασίας και η ανακάλυψη του συνόλου δεδομένων (Data Set) στο Kaggle.com. Ο λόγος επιλογής των δύο συγκεκριμένων Data Sets είναι η χρησιμότητά τους σε προβλήματα παλινδρόμησης καθώς επίσης το μέγεθος και το μεγάλο πλήθος των δεδομένων τους. Το θεωρητικό μέρος της διαδικασίας προχώρησε στη μελέτη του τραπεζικού κλάδου, των διαφορετικών ειδών δανείων, της αξίας της χρήσης τεχνικών μηχανικής μάθησης για ανάλυση δεδομένων και της σύνδεσης μεταξύ ανάλυσης δεδομένων και τραπεζικού κλάδου. Επιπρόσθετα, συζητήθηκε η θεωρητική βάση για τις διαδικασίες δημιουργίας των αλγορίθμων.

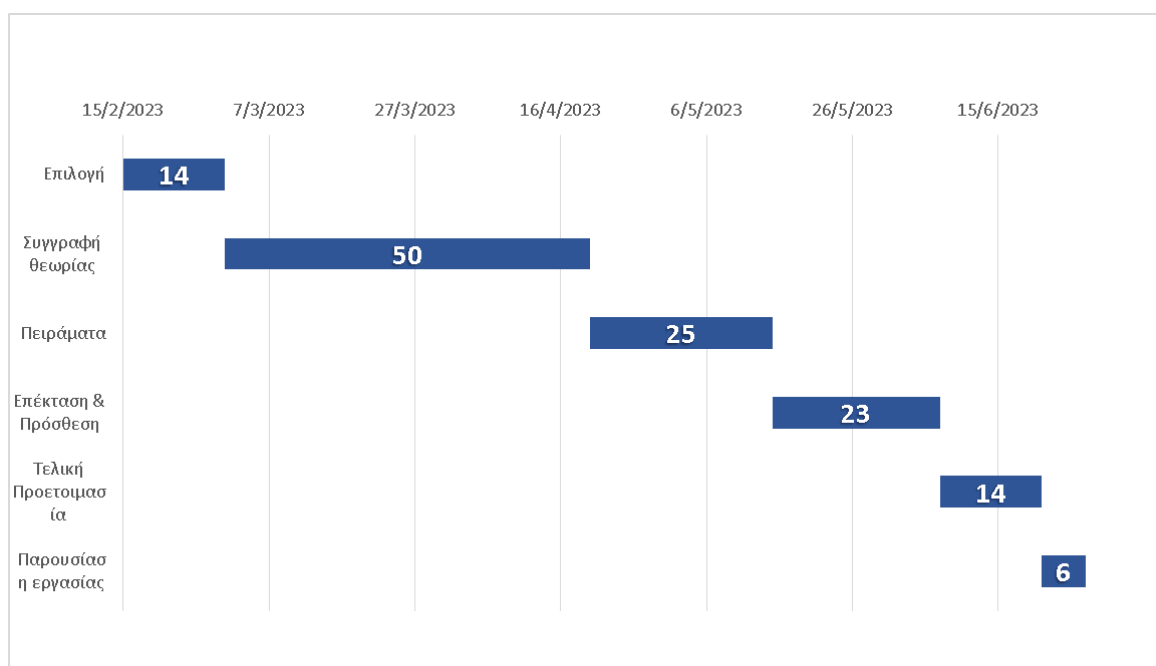
Το στάδιο ανάπτυξης των αλγορίθμων μηχανικής μάθησης ξεκίνησε μετά την ολοκλήρωση της συνολικής διερεύνησης των θεωρητικών θεμελίων και την προετοιμασία των δεδομένων. Τα βήματα αυτής της διαδικασίας ήταν η επιλογή των σωστών μοντέλων μηχανικής εκμάθησης, η προετοιμασία των δεδομένων εκπαίδευσης και δοκιμής και η χρήση των αλγορίθμων στις πληροφορίες.

Το επόμενο στάδιο ήταν η επιλογή του καλύτερου εργαλείου για την ανάπτυξη των αλγορίθμων και αυτό το εργαλείο ήταν το Google Colab και η γλώσσα προγραμματισμού για την λειτουργία του κάθε κώδικα που επιλέχθηκε είναι η Python. Αυτό έγινε αφού δημιουργήθηκε η ενότητα για τις τράπεζες και τα δάνεια και παρουσιάστηκε η σημασία της ανάλυσης δεδομένων στον τραπεζικό κλάδο με τεχνικές μηχανικής μάθησης. Ύστερα, τα δεδομένα καθαρίστηκαν και προετοιμάστηκαν και αναπτύχθηκαν αλγόριθμοι μηχανικής μάθησης οι οποίοι προσαρμόστηκαν στο συγκεκριμένο σύνολο δεδομένων. Το σύνολο δεδομένων εκπαίδευσης (train) και το σύνολο δεδομένων δοκιμής (test) ήταν και τα δύο ξεχωριστά αρχεία CSV που αποτελούσαν το σύνολο δεδομένων.

Προκειμένου να επιτευχθεί ακρίβεια και απόδοση στην πρόβλεψη των τραπεζικών δεδομένων, έγιναν πολλές δοκιμές και βελτιστοποίηση παραμέτρων μοντέλων καθ' όλη τη διάρκεια της δημιουργίας των αλγορίθμων. Σε αυτό το πείραμα, δοκιμάστηκε μια ποικιλία αλγορίθμων παλινδρόμησης όπως γραμμική παλινδρόμηση, τυχαίου δάσους, παλινδρόμηση XG Boost & Light G Boost, δέντρων αποφάσεων, Catboost, παλινδρόμηση νευρωνικών δικτύων και τέλος GRU & LSTM. Η παροχή των αλγορίθμων έγινε από τον υπεύθυνο καθηγητή Βικτωράτο Ι. σε συνδυασμό με την δοκιμή τους από τον φοιτητή Ρέκα Μ. Κατά τη διάρκεια του πειραματισμού, προσαρμόστηκαν οι παράμετροι των αλγορίθμων και πραγματοποιήθηκε η αξιολόγηση της απόδοσής τους με χρήση μετρικών όπως για παράδειγμα η αλλαγή στις τιμές των optimizers στα νευρωνικά δίκτυα. Οι παράμετροι των αλγορίθμων άλλαξαν κατά τη διάρκεια του πειράματος και η απόδοση των αλγορίθμων αξιολογήθηκε χρησιμοποιώντας μετρήσεις όπως, για παράδειγμα, η αλλαγή στις τιμές των βελτιστοποιητών στα νευρωνικά δίκτυα.

Στη συνέχεια έλαβε χώρα η αξιολόγηση της απόδοσης των αλγορίθμων στα δεδομένα ελέγχου αφού προσδιορίσαμε τις καλύτερες τιμές παραμέτρων και εκπαιδεύσαμε τους

αλγόριθμους στο εκπαιδευτικό σύνολο δεδομένων. Τέλος, με βάση τα αποτελέσματα που έδωσαν οι αλγόριθμοι έγινε η μεταξύ τους σύγκριση για την επιλογή εκείνου με το μικρότερο σφάλμα ως πιο αποτελεσματικού. Η περαιτέρω επεξήγηση των τεχνικών και των αλγορίθμων με τους οποίους εκτελούνται παρουσιάζεται στο επόμενο στην συνέχεια της εργασίας. Παρακάτω παρουσιάζεται εικόνα με το διάγραμμα Gantt με τις εργασίες που εξηγήσαμε παραπάνω σε μορφή διάρκειας ημερών σε σχέση με τους οριζόντιους άξονες που είναι οι ημερομηνίες και τους κάθετους που είναι οι εργασίες. Η τελική ημερομηνία παρουσίασης ήταν στις 27/06/2023.



Εικόνα 6.1.1.1. Διάγραμμα Gantt.

6.2. Αποτελέσματα έρευνας & Σχολιασμός.

Σε αυτό το μέρος της εργασίας παρουσιάζεται τελική μορφή των αλγορίθμων και η επεξήγηση των μεθόδων που χρησιμοποιήθηκαν για να μεγιστοποιήσουν την αποτελεσματικότητά τους στην ανάλυση των δεδομένων. Αρχικά, θα γίνει η προετοιμασία και ο καθαρισμός του κάθε ένα από τα δύο σύνολα δεδομένων. Στην συνέχεια αναλύονται τα βήματα που ακολουθήθηκαν για την διεξαγωγή της ανάλυσης και των αλγορίθμων.

6.2.1. Προετοιμασία των δεδομένων του Training Data Set.

Παρακάτω παρατίθεται ο αλγόριθμος που χρησιμοποιήθηκε για να γίνει προετοιμασία και καθαρισμός του πρώτου συνόλου δεδομένων, αυτού της εκπαίδευσης. Αυτός ο αλγόριθμος εκτελεί λειτουργίες προετοιμασίας δεδομένων, όπως καθαρισμό δεδομένων, κωδικοποίηση one-hot encoder, συνένωση κωδικοποιημένων κατηγορικών χαρακτηριστικών με αριθμητικά χαρακτηριστικά και κλιμάκωση των δεδομένων χρησιμοποιώντας πρότυπα. Αυτές οι ενέργειες είναι ζωτικής σημασίας για την προετοιμασία των δεδομένων για μοντελοποίηση και ανάλυση με χρήση μηχανικής μάθησης.

Πίνακας 6.2.1.1. Προετοιμασία των δεδομένων Training.

```
!pip install category_encoders
#!pip install category_encoders
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g.
pd.read_csv)
from datetime import datetime
from category_encoders.m_estimate import
MEstimateEncoder, OrdinalEncoder
from category_encoders.one_hot import OneHotEncoder
import matplotlib.pyplot as plt # Matlab-style plotting
from matplotlib import pyplot
import json

import seaborn as sns
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import regularizers
from keras.optimizers import Adam
from tensorflow.keras.models import Model, Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Dropout,
Flatten, Dense
from sklearn.metrics import mean_absolute_error as mae
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from keras.layers import Lambda
from tensorflow.keras.models import load_model

import math
from tensorflow.keras.layers import Input
from tensorflow.keras.layers import Dense
from scipy import sparse

from keras import backend as K

from tensorflow.keras import layers
from tensorflow.keras import regularizers

import sklearn
from sklearn import datasets
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, MinMaxScaler

from keras.models import Model
from keras.layers import Input, Dense, Layer
#from densenet import DenseTied
from tensorflow.keras import activations

# Read raw data from the CSV
datatr = pd.read_csv('/credit_train.csv')

datatr.head()
datatr=datatr.dropna()
```

```

datatr.describe()
datatr = datatr[datatr['Current Loan Amount'] != 99999999]

# Dropping some features.
y_train = datatr['Current Loan Amount'].values
# Read raw data from the CSV

# Drop columns we don't need
datatr = datatr.drop(['Loan ID', 'Customer ID', 'Current Loan
Amount', 'Loan Status'], axis=1)

# Specify the categorical features to encode
cat_features = ['Term', 'Years in current job', 'Home Ownership',
'Purpose']

# Create the encoder
encoder = OneHotEncoder()

# Fit and transform the data
cat_data = encoder.fit_transform(datatr[cat_features])
cat_data_sparse = sparse.csr_matrix(cat_data)
num_data = datatr.drop(cat_features, axis=1).values
x_train = np.concatenate([num_data, cat_data_sparse.toarray()],
axis=1)

print(x_train.shape)
# Scale the data
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_train

```

Ακολουθεί επεξήγηση των βασικών βημάτων που εμπλέκονται στην δημιουργία του παραπάνω αλγορίθμου:

1. Εισαγωγή βιβλιοθηκών: Εισάγονται οι απαραίτητες βιβλιοθήκες για σκοπούς χειρισμού δεδομένων, οπτικοποίησης, μηχανικής μάθησης και αξιολόγησης. όπως numpy, pandas, datetime, category_encoders, matplotlib, seaborn, tensorflow, keras, και sklearn.
2. Προεπεξεργασία δεδομένων: Διαβάζονται τα δεδομένα από φάκελο της μορφής CSV χρησιμοποιώντας την εντολή 'read_csv()'. Μετά ρίχνεται όποια σειρά που περιέχει ελλείψεις τιμές με την μέθοδο dropna(). Καταργούνται τα ακραία στοιχεία φιλτράροντας τις σειρές όπου το τρέχον ποσό δανείου έχει οριστεί σε 99999999. Στην συνέχεια ορίζεται η μεταβλητή στόχος 'Current Loan Amount' στον πίνακα 'y_train'. Οι στήλες που δε χρειάζονται στην ανάλυση όπως 'Loan ID', 'Customer ID', 'Current Loan Amount', and 'Loan Status', παραλείπονται από το σύνολο δεδομένων χρησιμοποιώντας την μέθοδο 'drop()'
3. Κωδικοποίηση χαρακτηριστικών: Τα χαρακτηριστικά 'Term', 'Years in current job', 'Home Ownership', 'Purpose' επιλέγονται για κωδικοποίηση. Η


```

Requirement already satisfied: six in
/usr/local/lib/python3.10/dist-packages (from patsy>=0.5.1-
>category_encoders) (1.16.0)
Requirement already satisfied: joblib>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.20.0-
>category_encoders) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.20.0-
>category_encoders) (3.1.0)
Requirement already satisfied: packaging>=21.3 in
/usr/local/lib/python3.10/dist-packages (from statsmodels>=0.9.0-
>category_encoders) (23.1)
Installing collected packages: category_encoders
Successfully installed category_encoders-2.6.1
(31505, 44)
array([[ -0.24524012,  -0.29620345,  -0.70064255,  ...,  -0.03739729,
         -0.01126854,  -0.03188643],
       [ -0.27900835,   1.18139908,  -0.05276094,  ...,  -0.03739729,
         -0.01126854,  -0.03188643],
       [ -0.24718829,  -0.82383113,  -0.42205977,  ...,  -0.03739729,
         -0.01126854,  -0.03188643],
       ...,
       [ -0.25368218,  -0.33034407,  -1.08035037,  ...,  -0.03739729,
         -0.01126854,  -0.03188643],
       [ -0.2523834 ,  -0.72694285,  -1.29226821,  ...,  -0.03739729,
         -0.01126854,  -0.03188643],
       [ -0.23744745,  -0.33236555,  -0.99444455,  ...,  -0.03739729,
         -0.01126854,  -0.03188643]])

```

Πλέον, το σύνολο δεδομένων εκπαίδευσης έχει καθαριστεί και ετοιμαστεί για περαιτέρω έρευνα. Η φόρμα των προεπεξεργασμένων δεδομένων (31505, 44) υποδεικνύει ότι το σύνολο δεδομένων περιέχει 31.505 δείγματα με 44 χαρακτηριστικά. Τα δεδομένα έχουν κλιμακωθεί και μετατραπεί και εμφανίζονται ως πίνακας. Κάθε στήλη είναι ένα ξεχωριστό χαρακτηριστικό και κάθε σειρά υποδηλώνει ένα δείγμα. Χρησιμοποιώντας το StandardScaler, οι τιμές του πίνακα έχουν τυποποιηθεί με μέσο όρο 0 και τυπική απόκλιση 1. Τα μοντέλα μηχανικής εκμάθησης μπορούν να εκπαιδευτούν χρησιμοποιώντας αυτά τα τυποποιημένα δεδομένα.

6.2.2. Προετοιμασία των δεδομένων του testing data set.

Με την εξάλειψη των τιμών που λείπουν, τη διαγραφή επιπλέον στηλών, την κωδικοποίηση κατηγορικών χαρακτηριστικών, τη σύνδεση των κωδικοποιημένων κατηγορικών και αριθμητικών δεδομένων και την κλιμάκωση των δεδομένων, ο προτεινόμενος αλγόριθμος παράγει το σύνολο δεδομένων δοκιμής. Αυτές οι διαδικασίες εγγυώνται ότι τα δεδομένα δοκιμών αντιμετωπίζονται με τρόπο συνεπή με τα δεδομένα εκπαίδευσης, προετοιμάζοντάς τα για χρήση στην αξιολόγηση των εκπαιδευμένων μοντέλων μηχανικής εκμάθησης.

Πίνακας 6.2.2.1. Προετοιμασία των δεδομένων Testing

```
# Read raw data from the CSV
datate = pd.read_csv('/credit_test.csv')

datate=datate.dropna()
print(datate)
datate = datate[datate['Current Loan Amount'] != 9999999]

# Dropping some features.
y_test = datate['Current Loan Amount'].values
# Read raw data from the CSV

# Drop columns we don't need
datate = datate.drop(['Loan ID', 'Customer ID', 'Current Loan
Amount'], axis=1)

# Specify the categorical features to encode
cat_features = ['Term', 'Years in current job', 'Home Ownership',
'Purpose']

# Fit and transform the data
cat_data = encoder.transform(datate[cat_features])
cat_data_sparse = sparse.csr_matrix(cat_data)
num_data = datate.drop(cat_features, axis=1).values
x_test = np.concatenate([num_data, cat_data_sparse.toarray()],
axis=1)

print(x_test.shape)
# Scale the data
x_test = scaler.transform(x_test)
```

Τα βήματα που ακολουθούνται στον παραπάνω αλγόριθμο είναι παρόμοια με εκείνα του αλγορίθμου του συνόλου εκπαίδευσης.

1. Ανάγνωση των μη επεξεργασμένων δεδομένων από το CSV: Χρησιμοποιώντας τη μέθοδο `pd.read_csv()`, το σύνολο δεδομένων δοκιμής (testing data set) εισάγεται από ένα αρχείο CSV.
2. Αφαίρεση ελλιπών τιμών: Οι σειρές τιμών που λείπουν από το σύνολο δεδομένων δοκιμής αφαιρούνται χρησιμοποιώντας τη μέθοδο `dropna()`.
3. Αφαίρεση ανεπιθύμητων στηλών: Οι στήλες 'Loan ID', 'Customer ID', and 'Current Loan Amount' παραλείπονται από το σύνολο δεδομένων χρησιμοποιώντας την λειτουργία `drop()`.
4. Καθορισμός κατηγορικών χαρακτηριστικών: Η λίστα `cat_features` περιέχει τα χαρακτηριστικά της κατηγορίας που πρέπει να κωδικοποιηθούν. Αυτά τα χαρακτηριστικά περιλαμβάνουν τα στοιχεία 'Term', 'Years in current job', 'Home Ownership', and 'Purpose'.

5. Κωδικοποίηση των κατηγορικών χαρακτηριστικών: Τα κατηγορικά χαρακτηριστικά του συνόλου δοκιμής κωδικοποιούνται χρησιμοποιώντας την λειτουργία transform().
6. Μετατροπή δεδομένων σε αριθμητική μορφή: Τα κωδικοποιημένα δεδομένα και τα υπολειπόμενα αριθμητικά δεδομένα συνενώνονται με την λειτουργία np.concatenate().
7. Κλιμάκωση των δεδομένων: Η λειτουργία transform() που αναφέρθηκε προηγουμένως χρησιμοποιείται επίσης για την κλιμάκωση των δεδομένων. Αυτό γίνεται για την ομαλοποίηση των δεδομένων και τον μετασχηματισμό τους με βάση τον μέσο όρο και την τυπική απόκλιση που προκύπτει από αυτά.
8. Εκτύπωση του αποτελέσματος: Μέσω της λειτουργίας 'shape' προβάλλονται οι διαστάσεις των επεξεργασμένων δεδομένων.

Το αποτέλεσμα του αλγορίθμου εκτυπώθηκε ως εξής:

Πίνακας 6.2.2.2. Αποτελέσματα testing code.

	Loan ID \	
2	f7744d01- 894b-49c3-8777 -fc6431a2cff1	
3	83721 ffb-b99a-4a0f-aea5-ef472a138b41	
4	08f3789f-5714-4b10-929d -e1527ab5e5a3	
9	d08f3a5e- 93df-40e7 -bdd8-cba59180bddf	
11	663ad9a3-4691-4f8c-90c6-75a5f4f0b34f	
...	...	
9989	f775380b- 57c0-409c-9824 -cafbb9aa475a	
9990	91e26 ead- 810b-44a0-892f -d623e1e444a0	
9994	a2701102- 3cb0-46a6-8658 -e6f20d950182	
9996	bbd3a392- 01b4-4e0e-9c28 -b2a4a39beac7	
9998	0cc8e0e0-1bc6-49d7-ad0f-0598 b647458f	
	Customer ID	Current Loan Amount
Term \		
2	2c60938b-ad2b-4702-804d -eeca43949c52	153494.0
Short Term		
3	12116614-2f3c-4d16-ad34-d92883718806	176242.0
Short Term		
4	39888105-fd5f-4023-860a-30a3e6f5ccb7	321992.0
Short Term		
9	4080a828-a61a-4f04-a627-397f4319500c	266288.0
Long Term		
11	8ec9f388-a275-40ee-9edc-82dec9f978c3	258104.0
Short Term		
...
...		
9989	5af41412-52e0-498d-b68b-5c6ef419d259	110946.0
Short Term		
9990	1e7cbdfc-d3a9-41a4-8416-3a23b08ef27b	260128.0
Long Term		
9994	6dc5fa63-93f4-43a9-8192-2df57548287b	442596.0
Short Term		
9996	6c306306-f5c2-4db5-b74a-af2895123ecb	132550.0
Short Term		
9998	f90cf410-a34b-49e7-8af9-2b405e17b827	99999999.0
Short Term		

	Credit Score	Annual Income	Years in current job	Home Ownership
2	709.0	871112.0	2 years	
3	727.0	780083.0	10+ years	
4	744.0	1761148.0	10+ years	Home
9	683.0	2031518.0	2 years	
11	723.0	1284514.0	7 years	
...	
9989	736.0	632396.0	2 years	
9990	611.0	3182519.0	4 years	Home
9994	739.0	1528968.0	7 years	Home
9996	718.0	763192.0	4 years	Home
9998	721.0	972097.0	10+ years	Home

	Purpose	Monthly Debt	Years of Credit History
2	Debt Consolidation	8391.73	12.5
3	Debt Consolidation	16771.87	16.5
4	Debt Consolidation	39478.77	26.0
9	Debt Consolidation	12443.10	24.4
11	Debt Consolidation	6368.99	14.6
...
9989	Debt Consolidation	13122.16	14.1
9990	Debt Consolidation	79563.07	20.7
9994	Debt Consolidation	19494.38	16.7
9996	Debt Consolidation	12401.87	9.9
9998	Debt Consolidation	12232.20	16.8

	Months since last delinquent	Number of Open Accounts
2	10.0	10.0
3	27.0	16.0
4	44.0	14.0
9	56.0	8.0
11	64.0	12.0
...
9989	48.0	17.0
9990	40.0	47.0
9994	79.0	11.0
9996	20.0	8.0
9998	24.0	8.0

	Number of Credit Problems Open Credit	Current Credit Balance	Maximum
2	0.0	38532.0	
3	1.0	156940.0	
388036.0			
531322.0			

4	0.0	359765.0
468072.0		
9	2.0	31445.0
251130.0		
11	0.0	128402.0
266904.0		
...
...		
9989	0.0	214966.0
324126.0		
9990	0.0	1398932.0
3673032.0		
9994	0.0	419235.0
656876.0		
9996	0.0	74309.0
329692.0		
9998	1.0	184984.0
240658.0		
	Bankruptcies	Tax Liens
2	0.0	0.0
3	1.0	0.0
4	0.0	0.0
9	2.0	0.0
11	0.0	0.0
...
9989	0.0	0.0
9990	0.0	0.0
9994	0.0	0.0
9996	0.0	0.0
9998	0.0	0.0
[3573 rows x 18 columns]		
(3099, 44)		

Το σύνολο δεδομένων που παράγεται έχει 18 στήλες και 3573 σειρές. Ως αποτέλεσμα κώδικας βοηθά στην εξαγωγή ορισμένων δεδομένων από το ευρύτερο σύνολο δεδομένων εστιάζοντας σε σχετικές στήλες και φιλτράροντας σειρές σύμφωνα με τις απαιτήσεις. Η μορφή ή το μέγεθος του τελικού συνόλου δεδομένων υποδεικνύεται από τη σημείωση (3099, 44). Δείχνει ότι το σύνολο δεδομένων σε αυτήν την περίπτωση περιλαμβάνει 3099 σειρές και 44 στήλες. Αυτό υποδηλώνει ότι το φιλτραρισμένο σύνολο δεδομένων έχει υποβληθεί σε περαιτέρω επεξεργασία για να γίνει ένα νέο σύνολο δεδομένων με λιγότερες σειρές και περισσότερες στήλες. Το σύνολο δεδομένων που δημιουργείται έχει 44 στήλες που αντικατοπτρίζουν διάφορα χαρακτηριστικά ή χαρακτηριστικά που σχετίζονται με κάθε στοιχείο και 3099 σειρές.

6.2.3. Γραμμική Παλινδρόμηση.

Αυτός ο αλγόριθμος εκτελεί γραμμική παλινδρόμηση στα δεδομένα. Με βάση τα δεδομένα εκπαίδευσης, εκπαιδύεται ένα μοντέλο γραμμικής παλινδρόμησης. Χρησιμοποιώντας τα δεδομένα της δοκιμής, γίνονται προβλέψεις και υπολογίζονται το μέσο απόλυτο σφάλμα και το μέσο τετράγωνο σφάλμα μεταξύ των προβλεπόμενων και των πραγματικών τιμών. Στη συνέχεια, η απόδοση του μοντέλου γραμμικής

παλινδρόμησης αξιολογείται εκτυπώνοντας τις απόλυτες και τις τετραγωνικές τιμές σφάλματος.

Πίνακας 6.2.3.1. Linear Regressor

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error

# Create a linear regression model
model = LinearRegression()

# Train the model on the training data
model.fit(x_train, y_train)

# Predict on the test data
y_pred = model.predict(x_test)

# Calculate the absolute error
abs_error = mean_absolute_error(y_test, y_pred)
print("Absolute error:", abs_error)

# Calculate the squared error
sq_error = mean_squared_error(y_test, y_pred)
print("Squared error:", sq_error)

Αποτελέσματα:
Absolute error: 111734.19576554342
Squared error: 20744263292.664566
```

Όπου Absolute Error είναι το απόλυτο σφάλμα και όπου Squared Error είναι το τετραγωνισμένο σφάλμα. Το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE) υπολογίζει τη διαφορά μεταξύ των πραγματικών δεδομένων και του απόλυτου μέσου όρου των αναμενόμενων δεδομένων. Η διαφορά του τετραγώνου του μέσου όρου μεταξύ των πραγματικών δεδομένων και των προβλεπόμενων δεδομένων υπολογίζεται με Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE). (Pyne, 2020).

Η διαδικασία που ακολουθεί η γραμμική παλινδρόμηση ορίζεται ως εξής:

1. Εισαγωγή στοιχείων: Εισάγονται οι απαραίτητες για την λειτουργία του μοντέλου ενότητες όπως η 'LinearRegression' από το μοντέλο 'sklearn.linear_model', τα 'mean_absolute_error' και 'mean_squared_error' από τα 'sklearn.metrics'.
2. Δημιουργία του μοντέλου 'LinearRegression' και εισαγωγή του στην μεταβλητή 'model'.
3. Εκπαίδευση του μοντέλου: Εισάγεται η λειτουργία 'fit' για τα δεδομένα (x_train & y_train).
4. Δημιουργία προβλέψεων με την λειτουργία 'predict' για το σύνολο 'x_test' και είσοδος των τιμών των προβλέψεων στην μεταβλητή 'y_pred'.
5. Υπολογισμός του απόλυτου σφάλματος μεταξύ των προβλεπόμενων τιμών 'y_pred' και των πραγματικών 'y_test' μέσω της λειτουργίας 'mean_absolute_error' και αποθήκευση των αποτελεσμάτων στην μεταβλητή 'abs_error'.

- Υπολογισμός και εκτύπωση του απόλυτου σφάλματος χρησιμοποιώντας την λειτουργία 'print' και υπολογισμός του τετραγωνισμένου σφάλματος μέσω της λειτουργίας 'mean_squared_error' μεταξύ των πραγματικών και των προβλεπόμενων τιμών.

6.2.4. Παλινδρόμηση Τυχαίου Δάσους.

Αυτός ο κώδικας δημιουργεί ένα μοντέλο τυχαίας παλινδρόμησης δασών, το εκπαιδεύει στα παρεχόμενα δεδομένα εκπαίδευσης, το χρησιμοποιεί για την πρόβλεψη των αποτελεσμάτων για τα δεδομένα δοκιμής και, στη συνέχεια, αξιολογεί την αποτελεσματικότητα του μοντέλου χρησιμοποιώντας MAE και MSE.

Πίνακας 6.2.4.1. Random Forest Regression.

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error

# Define the model
model = RandomForestRegressor(n_estimators=100,
                             max_depth=100,
                             min_samples_split=2,
                             min_samples_leaf=1,
                             max_features='sqrt')

# Train the model
model.fit(x_train, y_train)

# Predict the output variable for the test set
y_pred = model.predict(x_test)

# Evaluate the model
mael = mean_absolute_error(y_test, y_pred)
print("Mean Absolute Error:", mael)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)

Αποτελέσματα:
Mean Absolute Error: 81224.77592771863
Mean Squared Error: 12343360736.95672
```

Η διαδικασία εκτέλεσης του κώδικα αρχίζει με την εισαγωγή των απαραίτητων ενοτήτων όπως 'RandomForestRegressor' από το 'sklearn.ensemble' και τα 'mean_absolute_error' & 'mean_squared_error' από τα 'sklearn.metrics'. Στη συνέχεια γίνεται ορισμός του μοντέλου: Γίνεται λήψη παλινδρόμησης τυχαίου δάσους μέσω της κλάσης 'RandomForestRegressor'. Στην συνέχεια τίθενται υπερπαράμετροι για το μοντέλο όπως:

- Οι 'n_estimators' που αποφασίζουν τον αριθμό των δένδρων αποφάσεων που θα χρησιμοποιηθούν. Στην περίπτωσή μας μετά από πειράματα ο καλύτερος αριθμός estimators, δηλαδή αυτός που έδινε τα μικρότερα σφάλματα ήταν 100

(δηλαδή με τιμές μεγαλύτερες ή μικρότερες του 100 χαλούσε το αποτέλεσμα δίνοντας μεγαλύτερο σφάλμα).

- Το 'max_depth' που θέτει το μέγιστο βάθος κάθε δέντρου το οποίο επίσης δίνει καλύτερο αποτέλεσμα με την τιμή 100.
- Το 'min_samples_split' το οποίο θέτει τον ελάχιστο αριθμό δειγμάτων που χρειάζονται και δίνει καλύτερα αποτελέσματα για την τιμή 2.
- Το 'min_samples_leaf' που ήταν καλύτερο στην τιμή 1.
- Το 'max_features'.

Ύστερα το μοντέλο εκπαιδεύεται στο σύνολο εκπαίδευσης 'x_train' & 'y_train' μέσω της τεχνικής 'fit'. Γίνονται προβλέψεις μέσω της εντολής 'predict' οι οποίες αποθηκεύονται στην μεταβλητή 'y_pred'. Τέλος, γίνεται αξιολόγηση του μοντέλου χρησιμοποιώντας τα MAE & MSE και η εκτύπωση της αξιολόγησης.

6.2.5. Παλινδρόμηση XG Boost & Light G Boost.

Αρχικά παρουσιάζεται ο κώδικας XG Boost ο οποίος δημιουργεί ένα μοντέλο παλινδρόμησης XG Boost, το εκπαιδεύει στα παρεχόμενα δεδομένα εκπαίδευσης, το χρησιμοποιεί για την πρόβλεψη των αποτελεσμάτων για τα δεδομένα δοκιμής και αξιολογεί την αποτελεσματικότητα του μοντέλου χρησιμοποιώντας MAE και MSE. Οι υπερπαραμέτροι του μοντέλου ορίζονται για να ρυθμίζουν τη μαθησιακή συμπεριφορά και τη δομή του δέντρου.

Πίνακας 6.2.5.1. XG Boost Regression.

```
import xgboost as xgb
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_absolute_error,
mean_squared_error, make_scorer

# Define the model
model = xgb.XGBRegressor(learning_rate=0.09,
                        n_estimators=500,
                        max_depth=20,
                        subsample=1,
                        colsample_bytree=1,
                        objective='reg:squarederror',
                        tree_method='hist')

# Train the model
model.fit(x_train, y_train)

# Predict the output variable for the test set
y_pred = model.predict(x_test)

# Evaluate the model
mae = mean_absolute_error(y_test, y_pred)
print("Mean Absolute Error:", mae)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)

Αποτελέσματα:
Mean Absolute Error: 70415.51053326174
Mean Squared Error: 12598712665.137253
```

Αρχικά γίνεται εισαγωγή των απαραίτητων στοιχείων όπως είναι το 'xgb' για την εισαγωγή του μοντέλου παλινδρόμησης XG Boost, το 'GridSearchCV' από το 'sklearn.model_selection' για τις υπερπαραμέτρους και τα 'mean_absolute_error' & 'mean_squared_error' από το 'sklearn.metrics' για να μπορέσει να γίνει η αξιολόγηση. Ύστερα γίνεται ορισμός του μοντέλου μέσω του 'XGBRegressor' από το 'xgb'. Οι υπερπαραμέτροι τέθηκαν ως εξής:

- Το 'learning_rate' τέθηκε στο 0.09 ως καλύτερη τιμή.
- Οι 'n_estimators' τέθηκαν στα 500 δένδρα.
- Το 'max_depth' στην τιμή του 20.
- Το 'subsample' καθορίζει το κομμάτι δειγμάτων που χρησιμοποιείται για κάθε δένδρο στην τιμή του 1.
- Το 'colsample_bytree' στην τιμή του 1 και καθορίζει το κομμάτι των δειγμάτων που εκπαιδεύουν το κάθε δένδρο.
- Το 'objective' θέτει το squared error ως το σφάλμα που πρέπει να μεγιστοποιηθεί κατά την διάρκεια της παλινδρόμησης.
- Το 'tree_method' το οποίο αποφασίζει τον τρόπο με τον οποίο χτίζονται τα δένδρα αποφάσεων και τίθεται στην τιμή 'hist'.

Στην συνέχεια ακολουθούνται τα βήματα που παρουσιάστηκαν και στους υπολοίπους αλγορίθμους. Εκπαιδεύεται το μοντέλο μέσω της εντολής 'fit', γίνονται προβλέψεις και αποθηκεύονται στην μεταβλητή 'y_pred'. Τέλος, γίνεται αξιολόγηση του μοντέλου μέσω του MAE & MSE, σφάλματα τα οποία εκτυπώνονται στο τέλος τους αλγορίθμου μέσω της εντολής 'print'.

Παρακάτω αναπτύσσεται ο αλγόριθμος Light G Boost ο οποίος εκπαιδεύεται στα δεδομένα εκπαίδευσης, εφαρμόζεται για να προβλέψει τα αποτελέσματα για τα δεδομένα δοκιμής και αξιολογεί την απόδοση του μοντέλου χρησιμοποιώντας MAE και MSE. Η μαθησιακή συμπεριφορά και η δεντρική δομή του μοντέλου ελέγχονται από τις υπερπαραμέτρους του, επιτρέποντας ευέλικτη προσαρμογή σύμφωνα με τις ιδιαίτερες αναλυτικές ανάγκες.

Πίνακας 6.2.5.2. Light G Boost Regression.

```
import lightgbm as lgb
from sklearn.metrics import mean_absolute_error,
mean_squared_error, make_scorer

# Define the model
model = lgb.LGBMRegressor(learning_rate=0.09,
                           n_estimators= 500,
                           max_depth=100,
                           min_child_weight=1,
                           subsample=1,
                           colsample_bytree=0.6,
                           objective='regression')

# Train the model
model.fit(x_train, y_train)

# Predict the output variable for the test set
y_pred = model.predict(x_test)
```

```
# Evaluate the model
mae = mean_absolute_error(y_test, y_pred)
print("Mean Absolute Error:", mae)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

Αποτελέσματα:

Mean Absolute Error: **94416.10678195728**

Mean Squared Error: **14926832692.185143**

Το μοντέλο αρχίζει με την εισαγωγή των απαραίτητων ενότητων 'lgb' για να εισάγει το μοντέλο του Light G Boost, 'mean_absolute_error' & 'mean_squared_error' και του 'make_scorer' από την ενότητα 'sclearn.metrics'. Το μοντέλο ορίζεται από την κλάση 'LGBMRegressor' της ενότητας 'lgb'. Οι υπερπαραμέτροι τέθηκαν ως εξής:

- Το 'learning_rate' απέδιδε καλύτερα στην τιμή 0.09
- Οι 'n_estimators' κατά τα πειράματα αποδείχθηκε πως συνέχιζαν να βελτιώνονται αρκετά μέχρι και 25.000 + δένδρα, αλλά αυτό αργούσε τον κώδικα και δεν ξεπερνούσε τον xgboost ο οποίος το έκανε αυτό με πολύ λιγότερα δένδρα. Οπότε η τιμή των estimators ορίστηκε στα 500.
- Το 'max_depth' ορίζεται στην τιμή του 100
- Το 'min_child_weight' θέτει το ελάχιστο άθροισμα βαρών και ορίζεται στην τιμή του 1.
- Το 'subsample' αποκτά επίσης της τιμή του 1.
- Το 'colsample_bytree' που ορίζει τα στοιχεία που χρησιμοποιούνται για την εκπαίδευση του κάθε δέντρου αποκτά την τιμή 0.6 δηλαδή χρησιμοποιεί το 60% των στοιχείων από τα δεδομένα.
- Το 'objective' έχει τεθεί σε 'regression' δηλαδή σε λειτουργία παλινδρόμησης.

Ακολουθούν η εκπαίδευση του μοντέλου χρησιμοποιώντας την μέθοδο 'fit' στα σύνολα 'x_train' και 'y_train'. Το εκπαιδευμένο μοντέλο χρησιμοποιείται στην δημιουργία προβλέψεων μέσω της μεθόδου 'predict' οι οποίες αποθηκεύονται όπως και στα προηγούμενα μοντέλα στην μεταβλητή 'y_pred'.

6.2.6. Παλινδρόμηση με δέντρα αποφάσεων.

Αυτό το πρόγραμμα δημιουργεί ένα μοντέλο decision tree regressor χρησιμοποιώντας τις υπερπαραμέτρους που επιλέγετε, το εκπαιδεύει στο εκπαιδευτικό σύνολο δεδομένων που παρέχετε, το εφαρμόζει για την πρόβλεψη των αποτελεσμάτων του δοκιμαστικού συνόλου δεδομένων και αξιολογεί την απόδοση του μοντέλου χρησιμοποιώντας MAE και MSE. Με βάση τα δεδομένα εκπαίδευσης, το μοντέλο δέντρου αποφάσεων μαθαίνει μια ιεραρχία συνθηκών if-else, επιτρέποντας μη γραμμικές αλληλεπιδράσεις μεταξύ των χαρακτηριστικών εισόδου και της μεταβλητής στόχου.

Πίνακας 6.2.6.1. Decision Trees Regression.

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error

# Define the model
model = DecisionTreeRegressor(max_depth=200,
                              min_samples_split=2,
                              min_samples_leaf=1,
                              max_features='sqrt')

# Train the model
model.fit(x_train, y_train)

# Predict the output variable for the test set
y_pred = model.predict(x_test)

# Evaluate the model
mael = mean_absolute_error(y_test, y_pred)
print("Mean Absolute Error:", mael)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

Αποτελέσματα:
Mean Absolute Error: 97492.7479832204
Mean Squared Error: 24278824455.05905

Η κλάση ‘DecisionTreeRegressor’ εισάγεται από την ενότητα ‘sklearn.tree’ για να δημιουργήσει παλινδρόμηση με δέντρα αποφάσεων. Επίσης εισάγονται τα ‘mean_absolute_error’ & ‘mean_squared_error’ από την ενότητα ‘sklearn.metrics’. Οι υπερπαραμέτροι ορίστηκαν στις εξής τιμές:

- Το ‘max_depth’ στην τιμή των 200
- Το ‘min_sample_split’ στην τιμή του 2.
- Το ‘max_features’ ορίζεται στο ‘sqrt’ που σημαίνει ότι θα ληφθεί υπόψη η τετραγωνισμένη ρίζα του συνολικού αριθμού του αποτελέσματος.

Το μοντέλο εκπαιδεύεται μέσω ‘fit’ και γίνεται πρόβλεψη και αποθήκευση αυτής στην μεταβλητή ‘y_pred’. Τέλος αξιολογείται το μοντέλο υπολογίζοντας και εκτυπώνοντας το MAE το οποίο αποθηκεύεται στην μεταβλητή ‘mael’ και το MSE το οποίο αποθηκεύεται στην μεταβλητή ‘mse’.

6.2.7. Παλινδρόμηση Cat Boost.

Αυτό το πρόγραμμα δημιουργεί ένα μοντέλο Decision Tree Regression χρησιμοποιώντας τις υπερπαραμέτρους που επιλέγονται, το εκπαιδεύει στο εκπαιδευτικό σύνολο δεδομένων, το εφαρμόζει για την πρόβλεψη των αποτελεσμάτων του δοκιμαστικού συνόλου δεδομένων και αξιολογεί την απόδοση του μοντέλου χρησιμοποιώντας MAE και MSE.

Πίνακας 6.2.7.1. Cat Boost Regression.

```
!pip install catboost
import catboost as cb
from catboost import Pool
from sklearn.metrics import mean_absolute_error, mean_squared_error

# Define the model
model = cb.CatBoostRegressor(learning_rate=0.5,
                             n_estimators=500,
                             subsample=1,
                             colsample_bylevel=1,
                             loss_function='RMSE',
                             random_seed=300)

# Train the model
train_pool = Pool(x_train, y_train)
model.fit(train_pool)

# Predict the output variable for the test set
test_pool = Pool(x_test)
y_pred = model.predict(test_pool)

# Evaluate the model
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)

# Print the results
print("Mean Absolute Error:", mae)
print("Mean Squared Error:", mse)

Αποτελέσματα:
Mean Absolute Error: 94396.3243419289
Mean Squared Error: 15460287271.263735
```

Η βιβλιοθήκη CatBoost εγκαθίσταται μέσω της εντολής ‘!pip install catboost’ στο περιβάλλον της Python. Το ‘catboost’ εισάγεται ως ‘cb’, το ‘pool’ μετατρέπει τα σύνολα σε μορφή κατάλληλη για τον συγκεκριμένο κώδικα. Στη συνέχεια εισάγονται τα ‘mean_absolute_error’ & ‘mean_squared_error’ από το ‘sklearn.metrics’, ενώ το CatBoost εισάγεται μέσω της κλάσης ‘CatBoostRegressor’. Οι πιο αποτελεσματικές υπερπαραμέτροι ορίζονται ως εξής:

- Το ‘learning_rate’ στην τιμή του 0.5.
- Οι ‘n_estimators’ στην τιμή των 500.
- Το ‘subsample’ στην τιμή 1.
- Το ‘colsample_bylevel’ που καθορίζει τα δεδομένα για κάθε δέντρο είναι επίσης στην τιμή 1.
- Το ‘loss_function’ μεγιστοποιεί την ρίζα του τετραγωνισμένου σφάλματος. (Root Mean Squared Error – RMSE).
- Το ‘random_seed’

Μετά τον ορισμό των υπερπαραμέτρων γίνεται η εκπαίδευση του μοντέλου η οποία εισάγεται σε συγκεκριμένη δομή για το CatBoost την 'Pool' η οποία επιτρέπει την σωστή διαχείριση λανθασμένων η ελλιπών τιμών. Η εκπαίδευση γίνεται μέσω της εντολής 'fit'. Στην συνέχεια γίνονται, όπως στους προηγούμενους αλγορίθμους, προβλέψεις οι οποίες εισάγονται στην μεταβλητή 'y_pred'. Η αξιολόγηση του μοντέλου γίνεται μέσω των MAE & MSE τα οποία εκτυπώνονται ως αποτελέσματα.

6.2.8. Παλινδρόμηση με νευρωνικά δίκτυα.

Αυτός ο κώδικας κατασκευάζει ένα μοντέλο νευρωνικού δικτύου πολλαπλών κρυφών επιπέδων, το εκπαιδεύει, αξιολογεί την απόδοσή του με μετρήσεις MAE και MSE και τα εκτυπώνει. Με βάση τα χαρακτηριστικά εισόδου, το μοντέλο επιδιώκει να προβλέψει συνεχείς αριθμητικές τιμές, ελαχιστοποιώντας ταυτόχρονα την απώλεια μέσου τετραγώνου σφάλματος κατά τη διάρκεια της εκπαίδευσης.

Πίνακας 6.2.8.1. Neural Networks Regression.

```
from keras import optimizers

# Define the model (neural net.)
model = Sequential()
model.add(Dense(400,
input_dim=x_train.shape[1],kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l2(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
model.add(Dense(300,kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l2(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
model.add(Dense(200,kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l1(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
model.add(Dense(100,kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l1(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
model.add(Dense(50,kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l1(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
model.add(Dense(1,kernel_initializer='he_normal',
activation='gelu', kernel_regularizer=regularizers.l1(0.01),
activity_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))

opt = optimizers.SGD(lr=0.9, momentum=0.3, decay=0.3,
clipvalue=0.3)

# Compile the model
optimizer = opt
model.compile(loss='mean_squared_error', optimizer=optimizer,
metrics=['mae'])

# Train the model
history= model.fit(x_train, y_train, batch_size = 300, epochs =
200,validation_data=(x_test,y_test))
# Plot the training and validation MAE
plt.plot(history.history['mae'])
plt.plot(history.history['val_mae'])
plt.title('Model MAE')
```

```

plt.ylabel('MAE')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper left')
plt.show()

# Make predictions on the test data
y_pred = model.predict(x_test)

# Calculate absolute error and squared error
abs_error = np.abs(y_test - y_pred.flatten())
squared_error = (y_test - y_pred.flatten())**2

# Print the mean absolute error and mean squared error
print('Mean Absolute Error:', np.mean(abs_error))
print('Mean Squared Error:', np.mean(squared_error))

Αποτελέσματα:
Mean Absolute Error: 104548.57904389521
Mean Squared Error: 18583484933.937614

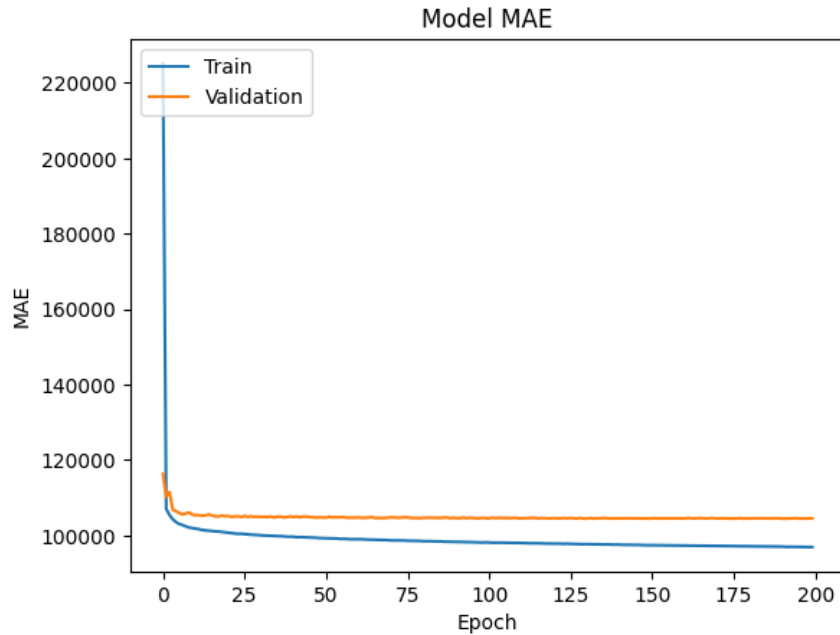
```

Εισάγονται οι απαραίτητες ενότητες όπως ‘keras’ για βαθιά μάθηση, ‘optimizers’, ‘numpy’, ‘matplotlib.pyplot’ και ‘Sequential’. Τα επίπεδα - dense που χρησιμοποιούνται είναι 6 και παίρνουν τις τιμές 400, 300, 200, 100, 50, 1 με σειρά από το πρώτο προς το τελευταίο, ενώ επίσης για την βελτίωση του αποτελέσματος χρησιμοποιούνται και kernel regularizers & activity regularizers. Οι optimizers έχουν τις εξής τιμές:

- Ir = 0.9
- Momentum = 0.3
- Decay = 0.3
- Clipvalue = 0.3

Η εκπαίδευση γίνεται για 200 epochs και batch size των 300. Στην συνέχεια γίνονται προβλέψεις και αξιολόγηση του μοντέλου μέσω των MAE & MSE.

Η σχέση μεταξύ MAE και Epochs μας δείχνει την πορεία του μοντέλου σε σχέση με τα epochs και αναπαρίσταται στο παρακάτω διάγραμμα:



Εικόνα 6.2.8.1. MAE & EPOCHS – NEURAL NETWORKS

6.2.9. Παλινδρόμηση με GRU & LSTM.

Χρησιμοποιώντας τους αλγορίθμους GRU και LSTM, αυτός ο κώδικας δημιουργεί και εκπαιδεύει ένα μοντέλο νευρωνικού δικτύου. Στη συνέχεια, αξιολογεί την απόδοση του μοντέλου σε ένα σύνολο δοκιμών και εμφανίζει τις τιμές MAE και MSE για να παρακολουθεί την ικανότητα του μοντέλου για γενίκευση και μάθηση.

Πίνακας 6.2.9.1. GRU & LSTM Regression.

```
import matplotlib.pyplot as plt

def create_model(input_shape):
    inputs = tf.keras.Input(shape=input_shape)

    # Bidirectional GRU layer
    gru =
tf.keras.layers.Bidirectional(tf.keras.layers.GRU(units=20,
return_sequences=True))(inputs)

    # Bidirectional LSTM layer
    lstm =
tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(units=20))(gru)

    # Dense layers
    dense1 = tf.keras.layers.Dense(300, activation='gelu')(lstm)
    dense1 = tf.keras.layers.Dense(250, activation='gelu')(dense1)
    dense1 = tf.keras.layers.Dense(200, activation='gelu')(dense1)
    dense1 = tf.keras.layers.Dense(150, activation='gelu')(dense1)
    dense1 = tf.keras.layers.Dense(100, activation='gelu')(dense1)
    output = tf.keras.layers.Dense(1, activation='linear')(dense1)

    model = tf.keras.Model(inputs=inputs, outputs=output)

    return model
```

```

input_shape = (44, 1)
model = create_model(input_shape)

# Compile the model
model.compile(
    loss=tf.keras.losses.Huber(delta=0.1),
    optimizer=tf.keras.optimizers.Adam(lr=0.01, beta_1=0.5,
beta_2=0.5, amsgrad=False),
    metrics=["mse", "mae"]
)

history = model.fit(x_train, y_train, validation_data=(x_test,
y_test), epochs=300)

# Evaluate the model
loss = model.evaluate(x_test, y_test)
print("Test loss:", loss)

# Plot training and validation MAE
plt.plot(history.history['mae'])
plt.plot(history.history['val_mae'])
plt.title('Model MAE')
plt.ylabel('MAE')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper right')
plt.show()

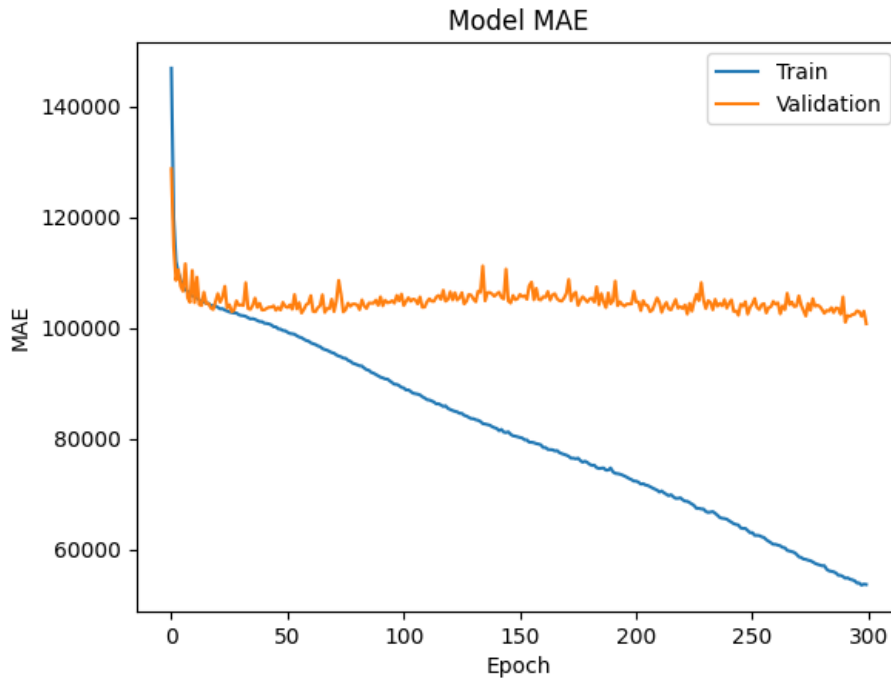
Αποτελέσματα:
loss: 9921.8340 - mse: 19944927232.0000 - mae: 99218.3984

```

Το μοντέλο δημιουργείται από την λειτουργία 'create_model' και χρησιμοποιεί τις λειτουργίες keras στις GRU & LSTM. Η τιμή των units έδινε την καλύτερη τιμή στα 20 ενώ τα επίπεδα - layers έδιναν μικρότερο σφάλμα όταν χωρίζονταν στα 6, ξεκινώντας από τα 300, 250, 200, 150, 100 και 1, από το πρώτο προς το τελευταίο. Τέλος πριν ολοκληρωθεί η εκτέλεση του αλγορίθμου και εκτυπωθούν τα MAE & MSE, παρακάτω δίνονται οι τιμές των optimizers με τις καλύτερες τιμές:

- Ir = 0.01
- Beta_1 = 0.5
- Beta_2 = 0.5
- Amsgrad = False.

Η σχέση MAE & Epochs αναπαρίσταται στο παρακάτω διάγραμμα δείχνοντας την εξέλιξη του σφάλματος ως προς τα epochs.



Εικόνα 6.2.9.1. MAE & EPOCHS – LSTM & GRU

6.3. Σύνοψη και σύγκριση των αποτελεσμάτων από τους κώδικες.

Σύμφωνα με τους αλγορίθμους των τεχνικών που παρουσιάστηκαν παραπάνω, η λειτουργία τους έδωσε τα εξής αποτελέσματα:

Table 6.3.1.1. Αποτελέσματα.

```

LINEAR REGRESSION:
Absolute error: 111734.19576554342
Squared error: 20744263292.664566

RANDOM FOREST:
Mean Absolute Error: 81224.77592771863
Mean Squared Error: 12343360736.95672

XGBOOST:
Mean Absolute Error: 70415.51053326174
Mean Squared Error: 12598712665.137253

Light G boost:
Mean Absolute Error: 94416.10678195728
Mean Squared Error: 14926832692.185143

Decision Trees:
Mean Absolute Error: 97492.7479832204
Mean Squared Error: 24278824455.05905

CatBoost:
Mean Absolute Error: 94396.3243419289
Mean Squared Error: 15460287271.263735

Neural Networks:

```

```
Mean Absolute Error: 104548.57904389521  
Mean Squared Error: 18583484933.937614
```

```
GRU  LSTM
```

```
Mean Absolute Error: 99218.3984  
Mean Squared Error: 19944927232.0000
```

Σύμφωνα με τα παραπάνω αποτελέσματα ο πιο αποτελεσματικός αλγόριθμος όσον αφορά το απόλυτο σφάλμα, το καλύτερο αποτέλεσμα το δίνει η τεχνική XG Boost με αρκετά μεγάλη διαφορά σε σχέση με τις υπόλοιπες. Ενώ όσον αφορά το τετραγωνισμένο σφάλμα, η παλινδρόμηση τυχαίου δάσους έδωσε μικρότερο, άρα καλύτερο σφάλμα από τον XG Boost με διαφορά μερικών μονάδων. Βέβαια η διαφορά που έχει το XG Boost στο απόλυτο σφάλμα σε σχέση με το τυχαίο δάσος είναι πολύ πιο μεγάλο οπότε ως καλύτερη επιλογή τεχνικής τίθεται η XG Boost. Η σειρά από το καλύτερο προς το χειρότερο αποτέλεσμα ακολουθεί ως: XG Boost, Random Forest ή Τυχαίο Δάσος, Catboost & Light G boost είναι περίπου το ίδιο, στην συνέχεια τα Decision Trees ή Δένδρα Αποφάσεων, GRU & LSTM, Νευρωνικά δίκτυα ή Neural Networks, και τελευταία η Γραμμική παλινδρόμηση ή Linear Regression που δίνει τα μεγαλύτερα σφάλματα, άρα τα χειρότερα αποτελέσματα. Οπότε, μέσω του κώδικα X G Boost η τράπεζα μπορεί στην συνέχεια να προχωρήσει στην οπτικοποίηση των δεδομένων και στην λήψη της σωστής απόφασης.

Σχολιασμός & Συμπέρασμα:

Η παρούσα εργασία αποτελεί ένα ολοκληρωμένο project ανάλυσης δεδομένων του τραπεζικού κλάδου. Ο αναγνώστης χάρη σε αυτήν αποκτά βασικές και γενικές γνώσεις πάνω στην ανάλυση δεδομένων και τον τρόπο που αυτή τελείται μέσω των τεχνικών μηχανικής μάθησης. Ο τραπεζικός τομέας είναι ένα πολύ σημαντικό, ζωτικής σημασίας εργαλείο, το οποίο οι σημερινές οικονομίες χρησιμοποιούν για να μπορούν να συναλλάσσονται και να διεκπεραιώνουν τις δραστηριότητές τους. Από την έρευνα που διεξήχθη μελετάται το τεχνικό κομμάτι, δηλαδή αυτό της ανάλυσης. Αυτό σημαίνει πως υπάρχει περιθώριο για μελλοντική έρευνα ως προς το κομμάτι της οπτικοποίησης των δεδομένων της διεξαγωγής γνώσης και της διερμηνεύσης αυτής για την λήψη υποθετικών αποφάσεων από την τράπεζα.

Στην ανάλυση δεδομένων για πρόβλεψη δανείου όπως αυτή διεξάγεται στην πτυχιακή, την πιο πρόσφατη τεχνική μηχανικής μάθησης αποτελεί η GRU & LSTM. Η τεχνική αυτή όπως διαπιστώνεται και από την παρούσα διερεύνηση είναι πιο αποτελεσματική από παραδοσιακές τεχνικές όπως είναι η Linear Regression και η Random Forest. Βέβαια, παρατηρείται πως άλλες τεχνικές όπως δέντρα αποφάσεων, Light G Boost, Decision Trees, XG Boost ή Cat Boost, αποδίδουν καλύτερα στο συγκεκριμένο σύνολο δεδομένων. Αυτό σημαίνει πως το σύνολο δεδομένων μπορεί να έχει κάποια έλλειψη ή κάποια λάθος τιμή. Το γεγονός αυτό δίνει την δυνατότητα σε μελλοντικές έρευνες να χρησιμοποιήσουν τους αλγορίθμους και τις τεχνικές της πτυχιακής πάνω σε ένα άλλο, πιο σωστά δομημένο σύνολο δεδομένων, όπου αυτές μπορεί να δίνουν καλύτερες τιμές.

Η εργασία δε καλύπτει μερικές ακόμα πιο πρόσφατες τεχνικές μηχανικής μάθησης, όπως Convolutional Neural Network, Spy Bot & Attention Model. Οι τεχνικές αυτές είναι αρκετά σύγχρονες και αποτελούν βασικό κομμάτι για κάποια μελλοντική έρευνα η οποία μπορεί να αποτελείται από το ίδιο ή από ένα νέο σύνολο δεδομένων. Γενικό συμπέρασμα είναι πως τα μοντέλα έδωσαν αρκετά καλά αποτελέσματα. Τα αποτελέσματα του πειράματος δείχνουν πόσο καλά αποδίδουν διαφορετικά μοντέλα μηχανικής μάθησης στην πρόβλεψη δεδομένων δανείων.

Σύμφωνα με τα αποτελέσματα, συμπεραίνουμε ότι μεταξύ των μοντέλων που δοκιμάστηκαν, το XG Boost έχει την καλύτερη απόδοση, ακολουθούμενο από το Light GBM και το Cat Boost. Αυτά τα μοντέλα έχουν μειώσει τα μέσα απόλυτα και τα μέσα τετραγωνικά σφάλματα, επιδεικνύοντας ισχυρότερη ικανότητα πρόβλεψης δεδομένων δανείου. Βάσει των συγκριτικά μεγαλύτερων λαθών τους, η γραμμική παλινδρόμηση, τα δέντρα αποφάσεων και τα νευρωνικά δίκτυα μπορεί να μην είναι τα καλύτερα μοντέλα για τη συγκεκριμένη εργασία. Σύμφωνα με τα ευρήματα αυτού του έργου, το XG Boost συνήθως ξεπερνά άλλα μοντέλα όσον αφορά το μέσο απόλυτο σφάλμα και το μέσο τετράγωνο σφάλμα, καθιστώντας το καταλληλότερο μοντέλο για την πρόβλεψη δεδομένων δανείων. Τα ευρήματα δείχνουν ότι το XG Boost κάνει προβλέψεις που είναι πιο ακριβείς, μειώνοντας την απόκλιση μεταξύ των αναμενόμενων και των πραγματικών ποσών δανείων.

Είναι σημαντικό να έχουμε κατά νου ότι οι μοναδικές ιδιότητες του συνόλου δεδομένων που χρησιμοποιούνται σε αυτήν την έρευνα μπορεί να έχουν αντίκτυπο στο πόσο καλά λειτουργούν τα μοντέλα. Το σύνολο δεδομένων μπορεί να περιλαμβάνει

ορισμένα ελαττώματα ή ακραίες τιμές που βλάπτουν την απόδοση ορισμένων μοντέλων και αυξάνουν τα λάθη τους. Η απόδοση αυτών των μοντέλων θα μπορούσε να βελτιωθεί με περισσότερη έρευνα και προσαρμογή του συνόλου δεδομένων. Μάλιστα, τα αποτελέσματα υποδεικνύουν ότι πιο προηγμένες τεχνικές όπως το GRU-LSTM και τα νευρωνικά δίκτυα δεν ξεπερνούν απαραίτητα τα μοντέλα που βασίζονται σε σύνολο όπως το XGBoost, το Light GBM και το CatBoost σε αυτό το συγκεκριμένο πλαίσιο. Υποδηλώνει ότι για το δεδομένο σύνολο δεδομένων δανείου, τα μοντέλα όπως XG Boost ή τυχαίο δάσος προσφέρουν καλύτερες προβλέψεις.

Μελλοντική μελέτη και ανάλυση στον τομέα της πρόβλεψης δανείων μπορεί να καθοδηγηθεί από αυτά τα ευρήματα. Για να αυξηθεί η ακρίβεια της πρόβλεψης, οι ερευνητές μπορεί να επικεντρωθούν στη βελτίωση του συνόλου δεδομένων, στη διερεύνηση στρατηγικών μηχανικής χαρακτηριστικών ή στη δοκιμή διαφόρων επαναλήψεων μοντέλων. Η διερεύνηση των ακριβών στοιχείων που οδηγούν στη βελτιωμένη απόδοση των XG Boost, Light GBM και Cat Boost σε σύγκριση με άλλα μοντέλα θα ήταν επίσης χρήσιμη. Πολύ σημαντική θα ήταν η συμβολή μελέτης βασισμένης στον τρόπο με τον οποίο μπορεί μία διαδικασία ανάλυσης και οπτικοποίησης δεδομένων να μπορούσε να λάβει χώρα με την χρήση τεχνητής νοημοσύνης.

Τα αποτελέσματα αυτού του πειράματος καταδεικνύουν την αποτελεσματικότητα του μοντέλου XG Boost στην ακριβή πρόβλεψη δεδομένων δανείων. Οι χρηματοπιστωτικοί οργανισμοί και οι ερευνητές μπορούν να χρησιμοποιήσουν αυτά τα μοντέλα για να τους βοηθήσουν να λάβουν καλύτερες αποφάσεις και να μειώσουν τους κινδύνους που συνδέονται με τις δανειοδοτικές πράξεις. Για να ανταποκρινόμαστε σε μεταβαλλόμενα μοτίβα δεδομένων και να παρέχουμε ακριβείς προβλέψεις σε πραγματικές συνθήκες, είναι σημαντικό να αξιολογούμε και να βελτιώνουμε τακτικά τα μοντέλα.

Συμπερασματικά, ερευνήθηκε σε βασικό θεωρητικό και αναλυτικό επίπεδο ένα σύνολο των τεχνικών μηχανικής μάθησης. Συγκρίνοντας άρθρα προηγούμενων αλλά και σχετικά πρόσφατων χρόνων με την εργασία, γίνεται αισθητή η ραγδαία εξέλιξη της τεχνολογίας και η εμφάνιση όλο και περισσότερων τεχνικών. Οπότε, όπως αναφέρθηκε στην εισαγωγή της εργασίας, θα υπάρχει πάντα περιθώριο για περαιτέρω μελλοντική έρευνα και εξέλιξη των τεχνικών. Ως αποτέλεσμα της ερευνητικής εργασίας γίνεται κατανοητή η σημασία της χρήσης νέων τεχνολογικών τεχνικών από τις τράπεζες όχι μόνο για την κερδοσκοπία των ίδιων, αλλά και για την ευημερία της κοινωνίας. Άρα, η ανάλυση των τραπεζικών δεδομένων μέσω τεχνικών μηχανικής μάθησης μπορεί να προβλέψει καταστάσεις και να συμβάλλει στην οικονομική και κοινωνική ευημερία μίας χώρας ή ακόμη πιο ιδανικά όλου του κόσμου.

Ελληνική Βιβλιογραφία:

Βιβλία:

- ✚ Κιόχος Π. & Παπανικολάου Γ., 2011, “ΧΡΗΜΑ – ΠΙΣΤΗ - ΤΡΑΠΕΖΕΣ”, Αθήνα, εκδόσεις Κιόχου.
- ✚ Κιόχος Π., Παναγόπουλος Α. & Κυρμιζογλου Π., 2018, “Διαχείριση Κινδύνων & Διαχείριση Χαρτοφυλακίου”, Αθήνα, εκδόσεις Κιόχου.
- ✚ Παπαδάμου Σ., 2009, “Διαχείριση Χαρτοφυλακίου Μία Σύγχρονη Προσέγγιση”, Αθήνα, Εκδόσεις Gutenberg.
- ✚ Πιτσέλης Ζ., 2006, “Τεχνική Ανάλυση & Risk Management”, Θεσσαλονίκη, Εκδόσεις Ζυγός.

Ηλεκτρονική Βιβλιογραφία:

- ✚ Ζύγουρας Γ., 2014. “ΑΣΦΑΛΙΣΗ ΚΙΝΔΥΝΩΝ ΚΑΙ ΕΜΠΕΙΡΙΚΗ ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΙΔΙΩΤΙΚΗΣ ΔΑΠΑΝΗΣ ΥΓΕΙΑΣ”, Πειραιάς, Πανεπιστήμιο Πειραιώς, (διαθέσιμο στις 22/06/2023 στο: dione.lib.unipi.gr).
- ✚ Κύρκος Ε., 2015, “Επιχειρηματική Ευφυΐα & Εξόρυξη Δεδομένων”, Kallipos, Open Academic Editions, (διαθέσιμο στις 22/06/2023 στο: repository.kallipos.gr).
- ✚ Νικολάου Σ., 2013, “Τεχνικές διαχείρισης τραπεζικών κινδύνων και αξιολόγησης των χρηματοπιστωτικών ιδρυμάτων”, Θεσσαλονίκη, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, (διαθέσιμο στις 22/06/2023 στο: ikee.lib.auth.gr).

Ξένη Βιβλιογραφία:

Βιβλία:

- ✚ Caouette J., Altman E. & Narayanan P., 1998, “Managing Credit Risk”, Canada, John Wiley & Sons, Inc.
- ✚ Hoffer J., Ramesh V. & Topi H., 2017, “Modern Database Management”, Pearson Education, Inc.
- ✚ Miles M. & Huberman, 1994, “Qualitative Data Analysis”, London, Sage Publications.
- ✚ Mishkin F. & Serletis A., 2011, “THE ECONOMICS OF MONEY, BANKING, AND FINANCIAL MARKETS”, Toronto, Pearson Canada.

Ηλεκτρονική βιβλιογραφία:

- ✚ Abbas S. & Pienkowski A., 2022, “WHAT IS SOVEREIGN DEBT?” (διαθέσιμο στις 22/06/2023 στο: www.imf.org).
- ✚ Adaloglou N., 2020, “Recurrent Neural Networks: building GRU cells VS LSTM cells in Pytorch”, (theaisummer.com).
- ✚ Amazon Web Services, 2023, “How XGBoost Works”, (διαθέσιμο στις 22/06/2023 στο: docs.aws.amazon.com).
- ✚ Anand M, Velu A. & Whig P., 2022, “Prediction of Loan Behaviour with Machine Learning Models for Secure Banking” (διαθέσιμο στις 22/06/2023 στο: www.icsejournal.com).

- ✚ Bankrate.com, 2022, “What is a savings and loan association (S&L)?”, (διαθέσιμο στις 22/06/2023 στο: www.bankrate.com).
- ✚ Banoula M, 2023, “The Best Guide to Regularization in Machine Learning”, (διαθέσιμο στις 22/06/2023 στο: www.simplilearn.com).
- ✚ Banton C., 2023, “Interest Rates: Different Types and What They Mean to Borrowers” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Biswal A., 2023, “Bagging in Machine Learning: Step to Perform And Its Advantages”, (διαθέσιμο στις 22/06/2023 στο: www.simplilearn.com).
- ✚ Bond C., 2023, “Savings and Loan Association: Is It Right for You?” (διαθέσιμο στις 22/06/2023 στο: www.gobankingrates.com).
- ✚ Catboost.ai, 2018, “CatBoost Enables Fast Gradient Boosting on Decision Trees Using GPUs”, (διαθέσιμο στις 22/06/2023 στο: catboost.ai).
- ✚ CFA Institute, 2023, “Measuring and Managing Market Risk” (διαθέσιμο στις 22/06/2023 στο: www.cfainstitute.org).
- ✚ CFI Team, 2023, “Major Risks for Banks”, (διαθέσιμο στις 22/06/2023 στο: corporatefinanceinstitute.com).
- ✚ Chakure a., 2023, “Random Forest Regression in Python Explained”, (διαθέσιμο στις 22/06/2023 στο: builtin.com).
- ✚ Chen J., 2022, “Interest Rate Risk Definition and Impact on Bond Prices”, (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Chen J., 2023, “What Is a Neural Network?”, (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Csiszar J., 2023, “9 Types of Banks and Their Benefits”, (διαθέσιμο στις 22/06/2023 στο: www.gobankingrates.com).
- ✚ Data Science Team, 2019, “What is Light GBM?” (datascience.eu).
- ✚ Duffie D. & Schaefer S., 2005, “Quantitative Risk Management”, (διαθέσιμο στις 22/06/2023 στο: <http://ndl.ethernet.edu.et>).
- ✚ Dutta P., 2021, “A STUDY ON MACHINE LEARNING ALGORITHM FOR ENHANCEMENT OF LOAN PREDICTION” (διαθέσιμο στις 22/06/2023 στο: www.irjmets.com).
- ✚ Ellis C., 2023, “An introduction to XGBoost regression”, Kaggle.com, (www.kaggle.com).
- ✚ European Central Bank, 2015, “What is a central bank?” (διαθέσιμο στις 22/06/2023 στο: www.ecb.europa.eu).
- ✚ Frankenfield J., 2023, “What Is Online Banking? Definition and How It Works”, (www.investopedia.com).
- ✚ Frankenfield J., 2023, “What Is Online Banking? Definition and How It Works” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Ganesh K., 2020, “What’s The Role Of Weights And Bias In a Neural Network?” (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Gautam & Nidhi, 2021. “Loan Prediction using Decision Tree and Random Forest” (διαθέσιμο στις 22/06/2023 στο: cloudfront.net).
- ✚ Girardin M., 2023, “What Is Retail Banking?” (διαθέσιμο στις 22/06/2023 στο: www.theforage.com).
- ✚ Gurucharan M., 2020, “Machine Learning Basics: Decision Tree Regression”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).

- ✚ Hargrave M., 2022, “Investment Bank: What It Is, How It Works, Major Examples” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Hayes A., 2023, “Market Risk Definition: How to Deal with Systematic Risk”, (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Hull J., 2018, “Risk Management and Financial Institutions”, (διαθέσιμο στις 22/06/2023 στο: dl.rasabourse.com).
- ✚ Jain S. et al., 2022, “Machine Learning based model for Loan Amount Prediction and Distribution”, (διαθέσιμο στις 22/06/2023 στο: ceur-ws.org).
- ✚ Jorion P., 2007, “Value at Risk - The New Benchmark for Managing Financial Risk”, (διαθέσιμο στις 22/06/2023 στο: www.academia.edu).
- ✚ Kadam A., 2021, “Prediction for Loan Approval using Machine Learning Algorithm” (διαθέσιμο στις 22/06/2023 στο: cloudfront.net).
- ✚ Kagan J., 2022, “Federal Savings and Loan (S&L)” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Kagan J., 2023, “How Do Commercial Banks Work, and Why Do They Matter?” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Kagan J., 2023, “How Do Commercial Banks Work, and Why Do They Matter?” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Kaggle.com, 2020, “When to use Mean Absolute Error vs Mean Squared Error?” (διαθέσιμο στις 22/06/2023 στο: www.kaggle.com).
- ✚ Kanade V., 2023, “What Is Linear Regression? Types, Equation, Examples, and Best Practices for 2022”, (διαθέσιμο στις 22/06/2023 στο: www.spiceworks.com).
- ✚ Kata H., 2021, “Understanding Light Gradient Boosting Machine”, (διαθέσιμο στις 22/06/2023 στο: datamahadev.com).
- ✚ Katsuri S., 2019, “XGBOOST vs LightGBM: Which algorithm wins the race !!!”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Kenton W., 2021, “Understanding Liquidity Risk in Banks and Business, With Examples” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Khandelwal Y., 2021, “Ensemble Stacking for Machine Learning and Deep Learning”, (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Kostadiou S., 2017, “Understanding GRU Networks”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Kurt D., 2022, “Community Banks: Meaning, Overview, History and FAQ” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Lancett P., 2012, “Types of Long-Term Debt”, (διαθέσιμο στις 22/06/2023 στο: smallbusiness.chron.com).
- ✚ Leo M., Sharma S. & Maddulety K., 2019, “Machine Learning in Banking Risk Management: A Literature Review” (διαθέσιμο στις 22/06/2023 στο: www.mdpi.com).
- ✚ Lester J., Cho Y. & Lochmiller C., 2020, “Learning to Do Qualitative Data Analysis: A Starting Point”, (διαθέσιμο στις 22/06/2023 στο: <https://journals.sagepub.com>).
- ✚ Li S., 2018, “Research on financial data analysis based on data mining algorithm” (διαθέσιμο στις 22/06/2023 στο: onlinelibrary.wiley.com).

- ✚ Majaski C., 2023, “Retail Banking: What It Is, Different Types, and Common Services”, (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Majumdar A. et al., 2022, “Loan Prediction by using Machine Learning” (διαθέσιμο στις 22/06/2023 στο: <https://ijirt.org>).
- ✚ Mamun M., 2022, “Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis” (διαθέσιμο στις 22/06/2023 στο: ieomsociety.org).
- ✚ McKillop D. & Wilson J., 2011, “Credit Unions: A Theoretical and Empirical Overview”, (διαθέσιμο στις 22/06/2023 στο: onlinelibrary.wiley.com).
- ✚ Murthy P. et al., 2020, “Loan Approval Prediction System Using Machine Learning”, (διαθέσιμο στις 22/06/2023 στο: innovation-journals.org).
- ✚ Opperman A., 2023, “What Is Cat Boost?” (διαθέσιμο στις 22/06/2023 στο: <https://builtin.com>).
- ✚ Paul P. & Zhu S., 2020, “Are Banks Exposed to Interest Rate Risk?” (διαθέσιμο στις 22/06/2023 στο: www.frbsf.org).
- ✚ Peterdy K., 2023, “Credit Risk”, (διαθέσιμο στις 22/06/2023 στο: corporatefinanceinstitute.com).
- ✚ Phi M., 2018, “Illustrated Guide to LSTM’s and GRU’s: A step by step explanation”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Philips R. & Campbell C., 2023, “Advantages and Disadvantages of Bank Loans for Small Businesses” (διαθέσιμο στις 22/06/2023 στο: www.nerdwallet.com).
- ✚ Premanand S., 2021, “Building an End-to-End Logistic Regression Model” (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Ronaghan S., 2018, “The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Saini A., 2022, “An Introduction to Random Forest Algorithm for beginners”, (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Saxena S., 2021, “Introduction to Gated Recurrent Unit (GRU)”, (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Saxena S., 2021, “Learn About Long Short-Term Memory (LSTM) Algorithms”, (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Segal T., 2022, “What Is a Central Bank, and Does the U.S. Have One?” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).
- ✚ Singal G., 2020, “LSTM versus GRU Units in RNN”, (διαθέσιμο στις 22/06/2023 στο: www.pluralsight.com).
- ✚ Srivignesh R., 2021, “A Walk-through of Regression Analysis Using Artificial Neural Networks in Tensorflow”, (διαθέσιμο στις 22/06/2023 στο: www.analyticsvidhya.com).
- ✚ Steorts R., 2017, “Tree Based Methods: Regression Trees”, Duke University (διαθέσιμο στις 22/06/2023 στο: www2.stat.duke.edu).
- ✚ Supriya P. et al, 2019, “Loan Prediction by using Machine Learning Models” (διαθέσιμο στις 22/06/2023 στο: <https://www.semanticscholar.org>).
- ✚ Tamplin T., 2023, “Credit Score”, (www.financestrategists.com).
- ✚ Tardi C., 2023, “Financial Portfolio: What It Is, and How to Create and Manage One” (διαθέσιμο στις 22/06/2023 στο: www.investopedia.com).

- ✚ Tejaswini J. et al., 2020, “ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH” (διαθέσιμο στις 22/06/2023 στο: jespublication.com).
- ✚ Thakur D., 2018, “LSTM and its equations” (διαθέσιμο στις 22/06/2023 στο: <https://medium.com>).
- ✚ Thiesen S., 2021, “CatBoost regression in 6 minutes”, (διαθέσιμο στις 22/06/2023 στο: towardsdatascience.com).
- ✚ Thomas C., 2017, “Five Steps of the Risk Management Process”, (διαθέσιμο στις 22/06/2023 στο: www.360factors.com).
- ✚ Tibco.com, 2022, “What is Banking Analytics?” (διαθέσιμο στις 22/06/2023 στο: www.tibco.com).
- ✚ Treece K., 2021, “What Is A Loan?”, (διαθέσιμο στις 22/06/2023 στο: www.forbes.com).
- ✚ Tretina K., 2019, “How Do Credit-Builder Loans Work?”, (www.experian.com).
- ✚ Tsang B., 2023, “How Bank Loans Work: Advantages, Disadvantages and Things You Need To Know”, (διαθέσιμο στις 22/06/2023 στο: choco-up.com).
- ✚ Turgut T., 2018, “Risk management process in banking industry”, (διαθέσιμο στις 22/06/2023 στο: mpira.ub.uni-muenchen.de).
- ✚ Vadapalli P., 2022, “Bagging vs Boosting in Machine Learning: Difference Between Bagging and Boosting” (διαθέσιμο στις 22/06/2023 στο: www.upgrad.com).
- ✚ VanSomeren L., 2021, “16 Types of Loans to Help You Make Necessary Purchases”, (διαθέσιμο στις 22/06 στο: www.forbes.com).
- ✚ White A., 2022, “What is a credit union?” (διαθέσιμο στις 22/06/2023 στο: www.cNBC.com).

Ειδική Βιβλιογραφία:

- ✚ Begiev Z., 2017, “Bank Loan Status Dataset”, Kaggle.com, (Διαθέσιμο στις 22/06/2023 στο: www.kaggle.com).