



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

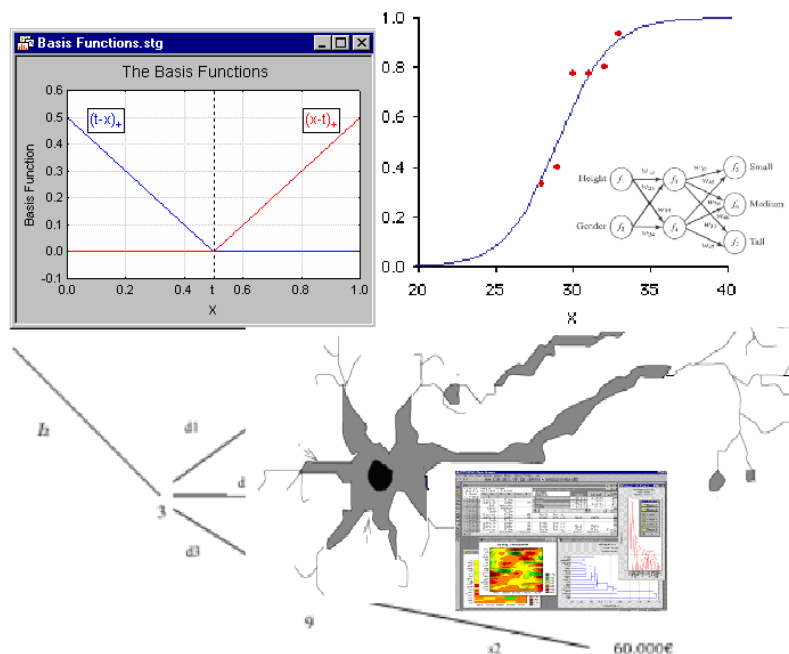
«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ  
και την ΟΙΚΟΝΟΜΙΑ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αλεξάνδρα Γ. Μανωλέσου

e-mail : [amanoless@gmail.com](mailto:amanoless@gmail.com)

ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΜΕΤΡΗΣΗΣ ΠΙΣΤΩΤΙΚΟΥ ΚΙΝΔΥΝΟΥ



ΑΘΗΝΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2017

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην

**«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ  
και την ΟΙΚΟΝΟΜΙΑ»**

που απονέμει η Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του Εθνικού Μετσόβιου Πολυτεχνείου.

Εγκρίθηκε την 13/09/2017 από την εξεταστική επιτροπή :

ΧΡΙΣΤΟΠΟΥΛΟΣ ΑΠΟΣΤΟΛΟΣ (ΕΠΙΒΛΕΠΩΝ)

ΕΠΙΣΚ. ΚΑΘΗΓΗΤΗΣ ΣΤΟ ΣΕΜΦΕ

ΚΟΛΛΙΑΣ ΗΡΑΚΛΗΣ

ΕΠΙΣΚ. ΚΑΘΗΓΗΤΗΣ ΣΤΟ ΣΕΜΦΕ

ΝΤΟΚΑΣ ΙΩΑΝΝΗΣ

ΕΠΙΣΚ. ΚΑΘΗΓΗΤΗΣ ΣΤΟ ΣΕΜΦΕ

## Ευχαριστίες

Η παρούσα εργασία αποτελεί διπλωματική εργασία στα πλαίσια του μεταπτυχιακού προγράμματος «ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ και την ΟΙΚΟΝΟΜΙΑ» της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του Εθνικού Μετσόβιου Πολυτεχνείου.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Χριστόπουλο Απόστολο για την καθοδήγηση και την συνεργασία του, ώστε να ολοκληρώσω τον κύκλο σπουδών μου στο παρόν μεταπτυχιακό πρόγραμμα.

Επίσης, ένα μεγάλο ευχαριστώ στην συμφοιτήτρια, συνάδελφο και πάνω απ' όλα φίλη μου, Κατσιμάρδου Σοφία, για την υποστήριξη της καθ' όλη την διάρκεια φοίτησης μου στο ΠΜΣ, και την ενθάρρυνση της για να ολοκληρώσω την παρούσα εργασία. Σοφάκι η εργασία αυτή είναι αφιερωμένη σε σένα, χωρίς την καθοδήγηση σου, τις σημειώσεις και τις συμβουλές σου, δεν θα είχα καταφέρει να ολοκληρώσω αυτό το πρόγραμμα. Σε ευχαριστώ πολύ για όλα, νιώθω περήφανη που σε έχω γνωρίσει και είσαι φίλη μου. Καλή σταδιοδρομία να έχεις, είμαι σίγουρη πως θα διαπρέψεις σε όπi κάνεις !

## Περίληψη

Η πιστωτική βαθμολόγηση έχει γίνει πολύ σημαντικό ζήτημα λόγω της πρόσφατης ανάπτυξης των χρηματοπιστωτικών ιδρυμάτων, και από το γεγονός ότι η υπερβολική χορήγηση δανειακών κεφαλαίων ήταν ένας από τις κύριες αιτίες της διεθνούς χρηματοπιστωτικής κρίσης. Οι τεράστιες βάσεις δεδομένων των τραπεζικών οργανισμών, είναι αδύνατο να αναλυθούν, τόσο σε οικονομικό όσο και σε ανθρώπινο δυναμικό, χωρίς την χρήση ειδικών τεχνικών εξόρυξης δεδομένων, οι οποίες απλοποιούν σημαντικά την διαδικασία λήψης αποφάσεων, διαχωρίζοντας «καλούς» και «κακούς» δανειολήπτες.

Ο τεράστιος όγκος της βιβλιογραφίας γύρω από τα credit scoring παρουσιάζει πολλές μεθόδους εξόρυξης δεδομένων για τη διαχείριση της πιστωτικής βαθμολόγησης. Ωστόσο, κάθε μέθοδος έχει τα πλεονεκτήματα και τους περιορισμούς της και δεν έχει υπάρξει μια ολοκληρωμένη προσέγγιση στον προσδιορισμό της πιο χρησιμοποιούμενης τεχνικής εξόρυξης δεδομένων.

Σκοπός της παρούσας εργασίας είναι να παρέχει μια βιβλιογραφική έρευνα η οποία θα παρουσιάζει τις ευρέως εφαρμοσμένες μεθόδους μέτρησης του πιστωτικού κινδύνου, όπως είναι οι discriminant analysis, logistic regression, K-nearest neighbour, Bayesian classifier, decision tree, neural network, survival analysis, fuzzy rule-based system, support vector machine, και τα υβριδικά μοντέλα, και μια σύγκριση αυτών αναφορικά με τα πλεονεκτήματα και τα μειονεκτήματα τις κάθε μεθόδου. Συνολικά αναλύονται έντεκα τεχνικές ταξινόμησης και παρουσιάζονται αντίστοιχα άρθρα εφαρμογής της κάθε μεθόδου που έχουν δημοσιευτεί ως επί των πλείστων από το 2000.

Μια τέτοια αναφορά μπορεί να βοηθήσει τους ερευνητές να γνωρίζουν τις πιο ευρέως διαδεδομένες μεθόδους, να βρουν τους περιορισμούς τους, να τους βελτιώσουν και να προτείνουν νέες μεθόδους με καλύτερες δυνατότητες ταξινόμησης.

## **Abstract**

Credit scoring has become an essential component, given the recent growing interest of financial institutions and the fact that excessive lending was one of the main causes of the international financial crisis. It is impossible to analyze the huge databases of banking organizations both in economic and manpower terms, without the use of dedicated data mining techniques, which greatly simplify the decision-making process, distinguishing "good" and "bad" borrowers.

The growing number of academic studies on credit scoring shows a variety of classification methods applied to data mining for credit scoring management. However, each method has its advantages and limitations, and there has not been a comprehensive approach in determining the most utilized data mining technique in the context of credit scoring.

The major goal of this paper is to provide a complete literature survey on applied data mining methods, such as discriminant analysis, logistic regression, K-nearest neighbour, Bayesian classifier, decision tree, neural network, survival analysis, fuzzy rule-based system, support vector machine, and hybrid methods, and a comparison of these with respect to the advantages and disadvantages of each method. In total, eleven classification techniques are presented with corresponding published articles, which have been published for the most part since 2000.

Such a report will assist researchers in realizing the most suitable approach in evaluating credit scores, pinpoint limitations, enhance them, and propose new approaches with improved capabilities.

|                                                                                                     |           |
|-----------------------------------------------------------------------------------------------------|-----------|
| ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ .....                                                  | 1         |
| και την ΟΙΚΟΝΟΜΙΑ» .....                                                                            | 1         |
| Ευχαριστίες.....                                                                                    | 3         |
| Περίληψη .....                                                                                      | 4         |
| Abstract.....                                                                                       | 5         |
| <b>Εισαγωγή .....</b>                                                                               | <b>7</b>  |
| Η έννοια του κινδύνου .....                                                                         | 7         |
| <b>ΚΕΦΑΛΑΙΟ 1<sup>ο</sup> : Βασικές Έννοιες του Κινδύνου και της Πιστοληπτικής Ικανότητας .....</b> | <b>8</b>  |
| Είδη του κινδύνου.....                                                                              | 8         |
| Μη Συστηματικός / Διαφοροποιήσιμος Κίνδυνος (No Systematic risk / Diversifiable risk) ...           | 9         |
| Πιστωτικός Κίνδυνος (Credit Risk) .....                                                             | 9         |
| Είδη Πιστωτικού Κινδύνου .....                                                                      | 10        |
| Παράγοντες που επηρεάζουν την πιστοληπτική ικανότητα.....                                           | 12        |
| Αναγκαιότητα εξέτασης της πιστοληπτικής ικανότητας .....                                            | 13        |
| Κανονιστικό πλαίσιο πιστωτικού κινδύνου .....                                                       | 15        |
| Εταιρική αξιολόγηση πιστοληπτικής ικανότητας .....                                                  | 15        |
| <b>ΚΕΦΑΛΑΙΟ 2<sup>ο</sup> : Τεχνικές &amp; Μέθοδοι μέτρησης Πιστωτικού Κινδύνου .....</b>           | <b>16</b> |
| Εισαγωγή.....                                                                                       | 16        |
| Είδη Βαθμολόγησης.....                                                                              | 18        |
| Εφαρμογές.....                                                                                      | 18        |
| Data mining – Εξόρυξη Δεδομένων .....                                                               | 19        |
| 2.1 Διαχωριστική Ανάλυση (Discriminant analysis) .....                                              | 20        |
| 2.2 Λογιστική Παλινδρόμηση (Logistic regression).....                                               | 22        |
| 2.3 Naïve Bayesian classifier .....                                                                 | 24        |
| 2.4 K-nearest neighbour (KNN) .....                                                                 | 25        |
| 2.5 Multivariate Adaptive Regression Splines (MARS) .....                                           | 26        |
| 2.6 Δέντρα Απόφασης (Decision Trees).....                                                           | 29        |
| 2.7 Survival Analysis (Ανάλυση Επιβίωσης).....                                                      | 30        |
| 2.8 Fuzzy Rule-Based System (FRBS) .....                                                            | 32        |
| 2.9 Νευρωνικά Δίκτυα (Neural Network) .....                                                         | 34        |
| 2.10 Support Vector Machine (SVM) .....                                                             | 36        |
| 2.11 Υβριδικά μοντέλα (Hybrid models) .....                                                         | 37        |
| <b>ΚΕΦΑΛΑΙΟ 3<sup>ο</sup> : Σύγκριση των μεθόδων.....</b>                                           | <b>39</b> |
| <b>Επισκόπηση.....</b>                                                                              | <b>43</b> |
| <b>Παράρτημα Ι – Πρόσθετοι Ορισμοί.....</b>                                                         | <b>44</b> |
| <b>Βιβλιογραφία.....</b>                                                                            | <b>50</b> |

## Εισαγωγή

Ο οικονομολόγος Frank Knight το 1921 στο βιβλίο Risk, Uncertainty and Profit προέβη σε διάκριση μεταξύ των εννοιών **κινδύνου** και **αβεβαιότητας**. Η δεύτερη είναι ριζικά διαφορετική από την έννοια του κινδύνου (ή του ρίσκου). Όπως καταθέτει :

**«Ο κίνδυνος μπορεί να υπαχθεί σε ποσοπική μέτρηση»**, εφόσον μπορεί να υπάρχει (ή να υπολογιστεί) η κατανομή πιθανοτήτων που έχει εφαρμογή. **Αντίθετα η αβεβαιότητα δεν μπορεί να μετρηθεί**. Δεν είναι γνωστή η κατανομή πιθανοτήτων, ούτε μπορούν να περιγραφούν όλα τα πιθανά ενδεχόμενα, ενώ οι πιθανότητες των ενδεχομένων είναι δυνατό να μην αθροίζονται στο 1.

Για πολλά χρόνια, η χρηματοοικονομική επιστήμη βασίστηκε στην υπόθεση ότι οι επενδυτές είναι ορθολογικοί (rational investors). Σε αυτή την περίπτωση, οι ορθολογικές αποφάσεις ανατακλούν τις ορθολογικά διαμορφωμένες προσδοκίες των επενδυτών. Ωστόσο, η συμπεριφορά των ατόμων (Behavioural Economics) διαφέρει ριζικά ανάλογα με το αν αντιμετωπίζουν ρίσκο ή αβεβαιότητα, με αποτέλεσμα οι επενδυτές να μην είναι πάντα ορθολογικοί, αφού η αβεβαιότητα δεν μπορεί να μετρηθεί και οι επενδυτικές επιλογές καθοδηγούνται από την ψυχολογία των ανθρώπων και την υποκειμενική εκτίμηση της αβεβαιότητας.

## Η έννοια του κινδύνου

Η έννοια του κινδύνου (risk) πηγάζει από την άγνοια των επερχόμενων αποτελεσμάτων και είναι άμεσα συνδεδεμένη με την έννοια της μεταβλητότητας (variance) των απροσδιόριστων αποτελεσμάτων, τα οποία θα μπορούσαν να έχουν υλική επίδραση σε μια οικονομική οντότητα.

Σε μια επένδυση με τον όρο κίνδυνος αναφερόμαστε στην άγνοια των μελλοντικών αποδόσεων της επένδυσης. Σε αυτή την περίπτωση η μέτρηση του κινδύνου θα γίνει με τον υπολογισμό της τυπικής απόκλισης (standard deviation) των πιθανών αποδόσεων της επένδυσης ([112] Φιλίππας, 2005).

Η εμφάνιση ενός κινδύνου, καθώς και τα αποτελέσματά του, δεν μπορούν να προβλεφθούν με ακρίβεια, συνεπώς ο κίνδυνος δεν μπορεί να εξαλειφθεί πλήρως, αλλά μπορεί να περιοριστεί μέσω του προσδιορισμού της πιθανότητας πραγματοποίησης του κινδύνου, της κατανόησης των αποτελεσμάτων του κινδύνου, και της γνώσης των παραγόντων που διαμορφώνουν το μέγεθος και την εμφάνιση του ([86] Smith & Merritt, 2002). Όπως θα δούμε στα «Είδη του Κινδύνου» ένα μέρος του κινδύνου (ο συστηματικός κίνδυνος) είναι αυτός που δεν μπορεί να εξαλειφθεί πλήρως, ενώ το μέρος του κινδύνου που προέρχεται από την αλλαγή των τιμών (μη συστηματικός κίνδυνος) μπορεί να περιοριστεί μέσω της διαφοροποίησης.

Σε κάθε περίπτωση, στην σύγχρονη εποχή, κάθε επιχείρηση οφείλει να εξετάζει προσεκτικά τους κινδύνους των επενδύσεων της, και να λαμβάνει υπόψη την ανάλογη σχέση κινδύνου και απόδοσης.

## ΚΕΦΑΛΑΙΟ 1<sup>ο</sup> : Βασικές Έννοιες του Κινδύνου και της Πιστοληπτικής Ικανότητας

### Είδη του κινδύνου

Η κατηγοριοποίηση του κινδύνου δύναται να γίνει με διάφορα κριτήρια. Ένας κίνδυνος μπορεί να χαρακτηριστεί ως **συστηματικός**, ή **μη συστηματικός**.

### Συστηματικός / Μη Διαφοροποιήσιμος Κίνδυνος (Systematic risk / Non-diversifiable risk)

Είναι ο αναπότρεπτος κίνδυνος, μέσω της διαφοροποίησης, ο οποίος μπορεί να επηρεάσει τις αξίες ενός μεγάλου εύρους χρεογράφων, χαρτοφυλακίων και επενδύσεων, ενώ η εμβέλεια του μπορεί να καλύπτει μια συγκεκριμένη αγορά, μια χώρα ή ένα ολόκληρο οικονομικό σύστημα. Τα επιτόκια, οι υφέσεις στην οικονομία και οι πόλεμοι ανηπρωσωπεύουν παράγοντες συστηματικού κινδύνου επειδή επηρεάζουν ολόκληρη την αγορά και δεν μπορούν να αποφευχθούν μέσω διαφοροποίησης χαρτοφυλακίου [1.12].

Ο συστηματικός κίνδυνος, σε αντίθεση με τον μη συστηματικό, μπορεί να μετριαστεί μόνο μέσω της ανπιστάθμισης κινδύνου (hedging) και είναι έμφυτος στην αγορά. Είναι επίσης γνωστός ως «μη διαφοροποιήσιμος κίνδυνος» (non diversifiable risk) ή «κίνδυνος αγοράς» (market risk).

Ο κίνδυνος αυτός είναι περισσότερο ορατός στις δραστηριότητες διαπραγμάτευσης μετοχικών και ομολογιακών τίτλων στην δευτερογενή αγορά ή στο «άνοιγμα» θέσεων σε συνάλλαγμα. Στην χρηματοοικονομική θεωρία ο κίνδυνος αγοράς ορίζεται ως η διασπορά των μη αναμενόμενων αποτελεσμάτων του χαρτοφυλακίου τίτλων που οφείλεται σε αιφνίδιες διακυμάνσεις ορισμένων χρηματοοικονομικών μεταβλητών.

Οι συνηθέστεροι παράγοντες κινδύνου της αγοράς είναι οι ακόλουθοι:

- Ο κίνδυνος του Δείκτη Τιμών Χρημασπλήριου (Equity Index Risk) ή αλλιώς ο χρημασπληριακός κίνδυνος, δηλ. ο κίνδυνος να μεταβληθούν οι τιμές των μετοχών, οπότε εάν για παράδειγμα πέσει όλο το χρημασπλήριο, που σημαίνει ότι δεν υπάρχει το φαινόμενο άλλες μετοχές να ανεβαίνουν και άλλες να πέφτουν, δεν υπάρχει η δυνατότητα επιλογής διαφορετικών μετοχών. Η πτώση της αγορά συμπεριλαμβάνει και τα ομόλογα και τους άλλους τίτλους (παράγωγα κτλ) άρα δεν είναι εφικτή η διαφοροποίηση.
- Ο κίνδυνος του επιτοκίου (Interest rate risk) (βλ. ομώνυμη ενότητα).
- Ο νομισματικός κίνδυνος ή ο κίνδυνος από τις αλλαγές των συναλλαγματικών ισοτιμιών (Foreign Exchange Risk). Με άλλα λόγια οι μεταβολές στις συναλλαγματικές ισοτιμίες επρόκειτο να επηρεάσουν την αξία μιας επένδυσης που πραγματοποιείται σε νόμισμα διαφορετικό από το νόμισμα του επενδυτή αλλά και τις υποχρεώσεις ή τις απαιτήσεις των επιχειρήσεων.



## **Μη Συστηματικός / Διαφοροποιήσιμος Κίνδυνος (No Systematic risk / Diversifiable risk)**

Μη συστηματικός κίνδυνος είναι ουσιαστικά ο κίνδυνος που προέρχεται από την αλλαγή τιμών (ή αδυναμία αποπληρωμής χρέους) εξαιτίας ειδικών περιστάσεων ενός συγκεκριμένου τίτλου (του εκδότη του) σε αντίθεση με τις συνθήκες αγοράς. Σε αυτήν την περίπτωση ο κίνδυνος μπορεί, εν μέρει, να εξαλειφθεί μέσα από ένα χαρτοφυλάκιο μέσω της διαφοροποίησης – διαποικιλιάσης. Εναλλακτικά ο κίνδυνος αυτός ονομάζεται διαφοροποιήσιμος.

Καθώς προσθέτουμε επενδυτικά στοιχεία σε ένα χαρτοφυλάκιο ο συνολικός του κίνδυνος μειώνεται. Η διαδικασία αυτή καλείται διαφοροποίηση του χαρτοφυλακίου. Η μείωση του κινδύνου, προκαλείται ουσιαστικά επειδή δυσάρεστα γεγονότα μιας εταιρείας, νοούνται ως ευχάριστα για μια άλλη. Η μεταβλητότητα της απόδοσης, μειώνεται καθώς αυξάνεται ο αριθμός των αξιογράφων που συμμετέχουν στο χαρτοφυλάκιο. Οι **Evans και Archer (1968) [32]** υπολόγισαν τον κίνδυνο ενός χαρτοφυλακίου το οποίο περιλάμβανε τυχαία επιλεγμένες κοινές μετοχές των οποίων τα ποσοστά συμμετοχής στο χαρτοφυλάκιο ήταν ίσα, καταλήγοντας στο συμπέρασμα ότι ο κίνδυνος ενός χαρτοφυλακίου αποτελούμενου από 15-18 περίπου μετοχές προσεγγίζει τον κίνδυνο του χαρτοφυλακίου της αγοράς. Οι **Wagner και Lau (1971) [99]** έδειξαν ότι ένα χαρτοφυλάκιο αποτελούμενο από περίπου 20 τυχαία επιλεγμένες μετοχές, έχει σχεδόν μόνο συστηματικό κίνδυνο. Ωστόσο, το κέρδος από την διαφοροποίηση του κινδύνου δεν συνεχίζεται αναλογικά όσο προστίθενται αξιόγραφα στο χαρτοφυλάκιο και παρότι ο κίνδυνος μειώνεται, η οριακή του μείωση γίνεται ολοένα και πιο μικρή.

Συνεπώς, ο κίνδυνος που υπάρχει τελικά σε ένα καλά διαφοροποιημένο χαρτοφυλάκιο είναι μόνο ο συστηματικός (**[112] Φιλίππας, 2005**).

Πέραν της γενικής διάκρισης των κινδύνων, υπάρχουν και άλλα είδη ως αποτέλεσμα των δραστηριοτήτων των εταιρειών (ως εταιρείες νοούνται και οι τραπεζικοί οργανισμοί), οι οποίοι παίζουν σημαντικό ρόλο σε μια οικονομική συναλλαγή, γιατί τελικά αυξάνουν ή μειώνουν το τελικό αποτέλεσμα όσον αφορά τον συνολικό κίνδυνο που χαρακτηρίζει ένα περιουσιακό στοιχείο **[1.12]**. Περισσότερα για τα είδη των κινδύνων αναφέρονται στο Παράρτημα της εργασίας.

Ένας από τους μη συστηματικούς κινδύνους είναι ο Πιστωτικός κίνδυνος και είναι αυτός με τον οποίο πραγματεύεται η παρούσα εργασία.

### **Πιστωτικός Κίνδυνος (Credit Risk)**

Είναι ο κίνδυνος που διατρέχει μια επιχείρηση ή ένας οργανισμός να μην εισπράξει έγκαιρα τις απαιτήσεις του ή ακόμα, σε μερικές περιπτώσεις, να μην τις εισπράξει ποτέ.

Συνεπώς, στην έννοια του πιστωτικού κινδύνου εμπίπτουν περιουσιακά στοιχεία και προϊόντα που έχουν να κάνουν με το δανεισμό, όπως είναι τα δάνεια λιανικής τραπεζικής, τα επιχειρηματικά ομόλογα και τα έντοκα γραμμάτια του δημοσίου.

Δεδομένης της παρούσας οικονομικής κατάστασης που διανύουμε ο πιστωτικός κίνδυνος αποτελεί τον σημαντικότερο κίνδυνο που αντιμετωπίζουν σήμερα τα χρηματοπιστωτικά ιδρύματα, καθώς η ιστορία έχει αποδείξει ότι οι μεγαλύτερες απώλειες και καταστροφές των χαρτοφυλακίων οφείλονται σε αυτόν. Η μέτρηση, ανάλυση και εκτίμησή του αποτελούν ένα πολύ ευαίσθητο και σημαντικό τμήμα που αφορά τη χρηματοοικονομική διοίκηση. Η μέτρηση του πιστωτικού κινδύνου έχει ξεκινήσει τις τελευταίες δεκαετίες (βλ. Κανονιστικό Πλαίσιο) με αποτέλεσμα την δημιουργία στατιστικών μοντέλων προκειμένου να έχουμε αναλυτικές

προβλέψεις. Η συλλογή στοιχείων, η επεξεργασία τους και τέλος η εφαρμογή τους στις απαιτήσεις των επιχειρήσεων δεν σταματάει ποτέ, καθώς θα πρέπει κάθε φορά να προσαρμόζεται και να εμπλουρίζεται με νέα δεδομένα. Οι οργανισμοί που δημιουργούν τα μοντέλα θα πρέπει να επιβεβαιώνουν την αποτελεσματικότητά τους και να πραγματοποιούν δοκιμές επαλήθευσης των αποτελεσμάτων. Για αυτό η τεχνική της επαλήθευσης των μαθηματικών μοντέλων είναι πολύ μεγάλης σημασίας και η ανάγκη να μοντελοποιηθεί ο κίνδυνος καθίσταται επιτακτική όχι μόνο λόγω της ανάγκης πρόβλεψης του κινδύνου αλλά και γιατί με αυτόν τον τρόπο θα βελτιωθούν οι αποφάσεις-συμβουλές προς τους πελάτες. Οι τραπεζίτες εστίαζαν την προσοχή τους κυρίως σε πληροφορίες που βασίζονταν σε διάφορα χαρακτηριστικά των δανειζόμενων όπως ο χαρακτήρας (φήμη), το κεφάλαιο (μόχλευση), η μεταβλητότητα των κερδών και τέλος οι εγγυήσεις ώστε να μπορέσουν να καταλήξουν στη σωστή απόφαση για τη χορήγηση ή όχι των πιστώσεων (**Atman & Saunders, 1998 [10]**).

Για την ανάλυση-μέτρηση του πιστωτικού κινδύνου, και ταυτόχρονα για τον υπολογισμό της πιστωτικής απώλειας, θα πρέπει να ληφθούν υπόψη :

- (α) η αδυναμία πληρωμής (**Default**),
- (β) η πιθανότητα αθέτησης υποχρέωσης (**Probability of Default - PD**),
- (γ) οι απώλειες που προέρχονται από την αθέτηση υποχρέωσης (**Loss Given Default - LGD**), και τέλος
- (δ) η έκθεση στον κίνδυνο λόγω της αθέτησης (**Exposure At Default - EAD**).

Από τον ορισμό των παραπάνω διαστάσεων θα προκύψει η μαθηματική σχέση υπολογισμού της αναμενόμενης πιστωτικής απώλειας.

$$\text{Αναμενόμενη Πιστωτική Απώλεια} = \text{PD} * \text{EAD} * \text{LGD}$$

## Είδη Πιστωτικού Κινδύνου

Ο πιστωτικός κίνδυνος θεωρείται ως η συνισταμένη των παρακάτω επιμέρους κινδύνων με τις ακόλουθες μορφές :

- Κίνδυνος Αθέτησης (Default Risk)
- Κίνδυνος Έκθεσης (Exposure Risk)
- Κίνδυνος Ανάκτησης (Recovery Risk)

### Κίνδυνος Αθέτησης (Default Risk)

Ο κίνδυνος αθέτησης περιγράφεται από την δεδομένη χρονική στιγμή κατά την οποία ο δανειολήπτης δεν μπορεί να είναι συνεπής με τις υποχρεώσεις που έχει, με αποτέλεσμα να εμφανίζει δυσκολίες στην αποπληρωμή του χρέους του.

Ο κίνδυνος αυτός μετριέται από την πιθανότητα αθέτησης (Probability of Default - PD) , δηλ. την πιθανότητα να αθετήσει κάποιος την υποχρέωση του για μια συγκεκριμένη πληρωμή.

Συνεπώς, είναι αναγκαίο να καθοριστεί πότε έχουμε αθέτηση υποχρέωσης. Σύμφωνα με την ΠΔΤΕ 2589/20.8.2007 [**I.18**], ως αθέτηση υποχρέωσης ορίζεται η επέλευση ενός ή και των δύο από τα παρακάτω γεγονότα:

- Το πιστωτικό ίδρυμα θεωρεί ότι ο πιστούχος είναι εύλογος πιθανό να μην εκπληρώσει στο σύνολό τους τις υποχρεώσεις πληρωμών προς το πιστωτικό ίδρυμα ή οποιαδήποτε εταιρεία του ομίλου στον οποίο ανήκει το πιστωτικό ίδρυμα.
- Ο πιστούχος έχει εμφανίσει καθυστέρηση άνω των 90 ημερών στην αποπληρωμή κάποιας εκ των υποχρεώσεων του προς το πιστωτικό ίδρυμα ή οποιαδήποτε εταιρεία του ομίλου στον οποίο ανήκει.

Το αποτέλεσμα του μπορεί να είναι είτε μερική είτε ολική απώλεια του οφειλόμενου ποσού. Η πιθανότητα αθέτησης εξαρτάται από τον ορισμό της αθέτησης υποχρέωσης και επειδή ο ακριβής ορισμός αθέτησης υποχρέωσης εξαρτάται από την κάθε περίπτωση χωριστά για αυτόν τον λόγο θα πρέπει να έχουμε σαφή εικόνα για το τι θεωρούμε αθέτηση υποχρέωσης. Εξίσου σημαντικός είναι ο παράγοντας χρονικός ορίζοντας για τον υπολογισμό της πιθανότητας αθέτησης. Επομένως όσο μεγαλύτερος είναι ο χρονικός ορίζοντας τόσο αυξάνει και η πιθανότητα αυτή.

### **Κίνδυνος Έκθεσης (Exposure Risk)**

Ο κίνδυνος έκθεσης δημιουργείται από το ποσό που ο δανειολήπτης αδυνατεί την συγκεκριμένη χρονική στιγμή να πληρώσει και επομένως έχουμε «έκθεση κατά την στιγμή της αθέτησης» (exposure at default - EAD).

Σύμφωνα με την Βασιλεία II (βλ. Κανονιστικό πλαίσιο) μια τράπεζα είναι υποχρεωμένη να δίνει μια εκτίμηση της έκθεσης στον κίνδυνο για κάθε συναλλαγή στα εσωτερικά συστήματα των τραπεζών. Επομένως το μέτρο εκτίμησης της έκθεσης σε κίνδυνο είναι πολύ σημαντικό και για αυτό θα πρέπει να αποφεύγονται λάθη κατά τον υπολογισμό του.

### **Κίνδυνος Ανάκτησης (Recovery Risk)**

Ο κίνδυνος ανάκτησης περιγράφει το ποσό που κατάφερε να ανακτηθεί ως προς την συνολική οφειλή από τον δανειζόμενο κατά την στιγμή της αθέτησης (**Angelo A. & John G. 2001 [7]**).

Το ποσοστό που κατάφερε να ανακτηθεί από την συνολική οφειλή ονομάζεται ποσοστό ανάκτησης (recovery rate) ενώ το ποσό που δεν κατάφερε να ανακτηθεί ονομάζεται «απώλεια δεδομένης της αθέτησης» (loss given default - LGD). Σύμφωνα με την ΠΔΤΕ 2589/20.8.2007 η ποσοστιαία ζημία σε περίπτωση αθέτησης ορίζεται ως «ο λόγος της ζημίας από άνοιγμα, εξαίτιας της αθέτησης υποχρεώσεων από μέρος ενός αντισυμβαλλομένου προς το ποσό που είναι ανεξόφλητο κατά τον χρόνο της αθέτησης».

Το μέγεθος της απώλειας, δηλαδή το LGD, είναι το πηλίκο της ζημίας προς το χρέος. Στην ζημία θα πρέπει να υπολογιστούν το συνολικό χρέος συν τα έξοδα συλλογής του χρέους συν οποιεσδήποτε προεξοφλητικές ζημιές μείον το ποσό που η τράπεζα τελικά θα εισπράξει. Στην περίπτωση του χαρτοφυλακίου το συνολικό ποσό ταυτίζεται με την ονομαστική αξία των στοιχείων του χαρτοφυλακίου.

Εκτός από τον τύπο της αναμενόμενης πιστωτικής απώλειας υπάρχουν και άλλοι τρόποι να μετρήσει κανείς τον πιστωτικό κίνδυνο **ανάλογα και με την φύση των περιουσιακών στοιχείων**, όπως :

- Οι **πιστωτικές διαβαθμίσεις (credit ratings - scoring)** οι οποίες κατατάσσουν την πιστοληπτική ικανότητα σε διαβαθμίσεις εκτιμώντας την ποιότητα του εκδιδόμενου χρέους.
- Τα **πιστωτικά περιθώρια (credit spreads)** που είναι η διαφορά των επιτοκίων χωρίς κίνδυνο και έχουν ως αποτέλεσμα την απαιτούμενη απόδοση των περιουσιακών στοιχείων.

## Παράγοντες που επηρεάζουν την πιστοληπτική ικανότητα

**Πιστοληπτική ικανότητα (Credit rating)** είναι η αξιοπιστία και η φερεγγυότητα ενός ατόμου, μιας επιχείρησης ή ακόμα και μιας χώρας στην αποπληρωμή των χρεών της. Αποκαλύπτει σε ένα δανειστή ή επενδυτή την πιθανότητα να μπορέσει ο δανειολήπτης να ανταποκριθεί στις δανειακές του υποχρεώσεις χωρίς τον κίνδυνο πτώχευσης. **[I.10]**

Με τον όρο **ασυνέπεια (probability of default)** τίθεται ο διαχωρισμός του κάθε δανειολήπτη σε «συνεπή» και «ασυνεπή». Κύρια στοιχεία που προσδιορίζουν την ασυνέπεια είναι (α) η αδυναμία εκπλήρωσης των οφειλών, (β) η εμφάνιση καθυστέρησης οφειλών πέραν των εκάστοτε προκαθορισμένων συνεχόμενων ημερών, (γ) η πτώχευση και οποιοδήποτε γεγονός το οποίο κρίνεται ότι δημιουργεί καθυστέρηση των οφειλών **[I.5], [I.4]**.

Οι παράγοντες που μπορούν να επηρεάσουν την πιστοληπτική ικανότητα του δανειολήπτη και κατ'επέκταση τη δυνατότητα του να δανείζεται χρήματα μέσω χρηματοπιστωτικών ιδρυμάτων είναι:

- Οικονομικό ιστορικό
- Επιτόκιο αποπληρωμής
- Διαθεσιμότητα περιουσιακών στοιχείων
- Ρευστότητα
- Παρούσα οικονομική κατάσταση
- Πιθανό μελλοντικό εισόδημα
- Αποταμιευτική συμπεριφορά
- Καταναλωτική συμπεριφορά
- Ύψος υποχρεώσεων (επιπλέον χρέη και δάνεια)

### **[I.10]**

Στην περίπτωση μιας χώρας, σύμφωνα με ανακοίνωση της Standard and Poor's (S&P), η αξιολόγηση προσδιορίζεται λαμβάνοντας υπόψη πέντε βασικές κατηγορίες παραγόντων, οι οποίοι περιλαμβάνουν τόσο ποιοτικά όσο και ποσοτικά κριτήρια. Ο βασικός στόχος είναι η εκτίμηση της ικανότητας μιας χώρας να εξυπηρετήσει το χρέος της διαχρονικά.

Οι παράγοντες αυτοί είναι:

1. Η αποτελεσματικότητα των θεσμών και της διακυβέρνησης για την εξασφάλιση βιώσιμων δημόσιων οικονομικών, την προώθηση ισόρροπης οικονομικής ανάπτυξης και την αντιμετώπιση οικονομικών ή πολιτικών κρίσεων.
2. Οικονομικοί παράγοντες, όπως είναι το κατά κεφαλήν ΑΕΠ, οι προοπτικές ανάπτυξης της χώρας κτλ.
3. Εξωτερικοί παράγοντες οι οποίοι απεικονίζουν την ικανότητα μιας χώρας να αντλεί χρηματοδότηση από το εξωτερικό για την κάλυψη των υποχρεώσεών της, με ειδική πρόβλεψη για τις χώρες εντός νομισματικών ενώσεων.
4. Δημοσιονομικοί παράγοντες που ανακλούν τη βιωσιμότητα των ελλειμμάτων και του χρέους.
5. Νομισματικοί παράγοντες που εκφράζουν τον βαθμό στον οποίο οι νομισματικές αρχές μπορούν να εκπληρώσουν τον ρόλο τους και να υποστηρίξουν τη βιώσιμη οικονομική ανάπτυξη μέσω της άσκησης ευέλικτης νομισματικής πολιτικής. **([112] Φιλίππας)** .

## Αναγκαιότητα εξέτασης της πιστοληπτικής ικανότητας

Η πιθανότητα ασυνέπειας πρέπει να εκτιμάται το καλύτερο δυνατόν, καθώς σε ειδικό πλαίσιο, μια κακή αξιολόγηση της πιστοληπτικής ικανότητας υποδεικνύει υψηλό κίνδυνο αθέτησης ενός δανείου, και, συνεπώς, οδηγεί σε υψηλά επιτόκια ή ακόμα και την άρνησης χορήγησης δανείου από τον πιστωτή. Αντιθέτως, όσο καλύτερη είναι η πιστοληπτική ικανότητα του δανειολήπτη, τόσο πιο πιθανό είναι μια τράπεζα ή κάποιο άλλο χρηματοπιστωτικό ίδρυμα να επεκτείνει την πίστωση προς αυτόν. [I.10]

Στο γενικό πλαίσιο, μια κάμψη συνολικά της αγοράς μιας χώρας θα επηρεάσει πρωτίστως την οικονομική ευχέρεια κάθε δανειολήπτη εντός της χώρας, αλλά δευτερευόντως την οικονομία των άλλων χωρών, δεδομένου ότι οι χρηματοοικονομικές αγορές είναι στενά συνδεδεμένες μεταξύ τους, βλ. για παράδειγμα την κρίση στις ΗΠΑ που επηρέασε τις αγορές της ΕΕ και πήρε διεθνείς διαστάσεις.

Στις κυριότερες αιτίες αυτής της διεθνούς χρηματοπιστωτικής κρίσης έχουν καταλογιστεί :

- Η αλόγιστη **πυλοποίηση** στεγαστικών δανείων (**securization**). Η τεχνική της πυλοποίησης περιόριζε τον πιστωτικό κίνδυνο στα τραπεζικά χαρτοφυλάκια και συχνά αναφερόταν ως η δεύτερη πιο αποτελεσματική πρακτική μετά τη δημιουργία των *paigow banks*. Η πρακτική επίσης, ήταν συμβατή με τους κανόνες κεφαλαιακής επάρκειας, παράλληλα με τη μετατροπή παγωμένων κεφαλαίων (δάνεια) σε λειτουργικά εργαλεία (τίτλους). Ωστόσο, το 2008 η χρηματοπιστωτική αγορά των ΗΠΑ κλονίστηκε συθέμελα με δραματικές εξελίξεις, καθώς την περίοδο 2003-2007 σημειώθηκε ραγδαία εξάπλωση των στεγαστικών δανείων τα οποία είχαν λανθασμένα βαθμολογηθεί με AAA, ενώ βασίζονταν σε επισφαλείς δανειακές συμβάσεις. Τα στελέχη των τραπεζών προκειμένου να εισπράξουν τα υψηλά Bonus χρησιμοποιούσαν αυτά τα εταιρικά ομόλογα (Mortgage Backed Securities - MBS), τα οποία πωλούνταν από τις Επενδυτικές Τράπεζες (Lehman Brothers) σε Αμοιβαία Κεφάλαια, Επενδυτικές Τράπεζες, Συνταξιοδοτικά Ταμεία (κρατικά και εταιρικά) κα. Καθώς τα εταιρικά αυτά ομόλογα αγοράζονταν από θεσμικούς επενδυτές, προκειμένου αυτοί να διασφαλιστούν, αγόρασαν Συμβόλαια Προστασίας Έναντι Κινδύνων Χρεοκοπίας (Credit Default Swaps), τα οποία προέρχονταν από την ασφαλιστική εταιρεία American International Group (AIG). Η τελευταία λόγω κακής διαχείρισης, όπως λανθασμένη αξιολόγηση του συσσωρευμένου κινδύνου, οδηγήθηκε στο χείλος της κατάρρευσης και διασώθηκε από την αμερικανική κεντρική τράπεζα, η οποία προσέφερε έκτακτη οικονομική ενίσχυση 85 δισεκατομμυρίων δολαρίων. Αίσιο τέλος όμως δεν είχε ο άλλος αμερικανικός κολοσσός, η Lehman Brothers, η οποία κήρυξε πτώχευση, παρασύροντας ένα φαινόμενο domino μεταξύ των εταιρειών και οδηγώντας σε χρεοκοπίες, συγχωνεύσεις, απολύσεις, αύξηση του δείκτη ανεργίας<sup>1</sup>, κα [I.12].
- Η **υπερβολική χορήγηση δανειακών κεφαλαίων** στα νοικοκυριά και στις επιχειρήσεις, η οποία μείωσε υπερβολικά την αποταμίευση με αποτέλεσμα να μην είναι δυνατή η περαιτέρω χρηματοδότηση και οι επιχειρήσεις και τα νοικοκυριά να οδηγηθούν σε χρεοκοπία [I.14]. Ο λόγος Δανειακών προς Ίδια Κεφάλαια από 9:1 σε τραπεζικά ιδρύματα έφθασε σε 19:1 ακόμη και 30:1 σε ενδιάμεσα χρηματοπιστωτικά ιδρύματα,

<sup>1</sup> Το επίπεδο ανεργίας αποτελεί βασική συνιστώσα της οικονομικής και κοινωνικής ευημερίας μιας περιοχής. Η Οικονομία είναι ο κύριος λόγος για την αύξηση των αυτοκτονιών στην χώρα μας, αλλά και παγκοσμίως τα τελευταία χρόνια. Το φαινόμενο έλαβε την ονομασία «αυτοκτονίες του ΔΝΤ» («IMF SUICIDES») λόγω της έξαρσης των αυτοκτονιών το 1998 στη Νότια Κορέα. [I.17]

εκθέτοντας τα τελευταία σε υπερβολικό χρηματοδοτικό κίνδυνο (Financial Leverage Risk). [Νικόλαος Γ. Τραυλός, 2008]

- Η **παγκοσμιοποίηση της οικονομίας**. Το Σεπτέμβριο του 2008, στις ΗΠΑ ο τότε πρόεδρος Τζορτζ Μπους πρότεινε την άμεση ενίσχυση της ρευστότητας των τραπεζών για τη διάσωση του χρηματοπιστωτικού συστήματος. Ο κρατισμός που προέβλεπε το φορολογικό βάρος να πέσει στην πλάτη των πολιτών (Σχέδιο Πόλσον για αγορά των λεγόμενων «τοξικών ομολόγων» με δημόσιο χρήμα) καταψηφίστηκε από το Κογκρέσο [I.11]. Τα χρηματιστήρια αντικατέπησαν την απαισιοδοξία της αγοράς στους δείκτες τους, σημειώνοντας πτώση. Στην ΕΕ ο τότε πρόεδρος της Νικολά Σαρκοζί προανήγγειλε σχέδιο ανπμετώπισης του προβλήματος στα πλαίσια της ΕΕ [I.13] Η κρίση μετά το πέρασμα του Ατλαντικού, πλησίασε όλο και νοτιότερα την Ευρώπη, καθώς μετά την κρατικοποίηση της Bradford and Bingley από τη βρετανική κυβέρνηση πραγματοποιήθηκε μερική κρατικοποίηση του ευρωπαϊκού κολοσσού Fortis από το Βέλγιο, το Λουξεμβούργο και την Ολλανδία [I.9]. Το μεγαλύτερο χτύπημα από την κρίση δέχτηκε η Ισλανδία, με μεγάλη έκθεση των τραπεζών της στα «τοξικά» ομόλογα. Παρά τα μεγάλα χρηματικά αποθέματα της η Βρετανία, ακολούθησε το αμερικανικό μοντέλο και ήταν από τις πρώτες χώρες που επλήγησαν από την κρίση. Πολλές τράπεζες κρατικοποιήθηκαν και άλλες συγχωνεύθηκαν. Η Γαλλία ομοίως είχε υιοθετήσει τέτοια προϊόντα, ενώ ανακατατάξεις παρατηρήθηκαν σε Βέλγιο, Ιταλία, Λουξεμβούργο και Ολλανδία. Οι ελληνικές τράπεζες δηλώνουν πως δεν είχαν προλάβει να επενδύσουν ιδιαίτερα σε CDOs, δε γνωστοποιήθηκε κατοχή Conduits ή SIVs εκ μέρους τους, όμως μετακύλησαν τα υψηλά διατραπεζικά επιτόκια στα παραδοσιακά προϊόντα τους. Τα οικονομικά ιδρύματα που επηρεάστηκαν περισσότερο από την κρίση αφορούσαν εκμετάλλευση ακινήτων (real estate), χωρίς να απουσιάζουν προβλήματα σε τραπεζικές ή ασφαλιστικές υπηρεσίες. Τα πραγματικά προβλήματα της κρίσης διαφάνηκαν στην ελληνική οικονομία, με την άνοδο των επιτοκίων και το συντηρητισμό του χρηματοπιστωτικού συστήματος να πλήπει δανειολήπτες, μικρομεσαίες επιχειρήσεις, την αγοραστική συμπεριφορά των καταναλωτών και λοιπές πτυχές της αγοράς, με άμεσο αντίκτυπο στην οικονομία [I.14], [I.7].
- Η **πλημμελής διαβάθμιση** των επενδυτικών τραπεζικών προϊόντων και των εταιρειών. Οι επενδυτικές τράπεζες που εξέδωσαν τα εταιρικά ομόλογα μειωμένης εξασφάλισης κατόρθωσαν να πείσουν τις εταιρείες διαβάθμισης (Bond Rating Agencies) να βαθμολογήσουν πολύ υψηλά (AAA, AA κ. λπ.) τα ομόλογα που είχαν ως αντίκρισμα τα πλοποιοιμένα δάνεια μειωμένης εξασφάλισης. Ένας από τους λόγους της «λανθασμένης» βαθμολόγησης ήταν ο ηθικός κίνδυνος (Moral Hazard), δηλ. η σχέση εξάρτησης που υπήρχε μεταξύ τους, αφού οι εταιρείες διαβάθμισης πληρώνονται και από τις επενδυτικές τράπεζες που είναι ανάδοχοι των ομολόγων.
- Οι υψηλές αμοιβές των στελεχών του χρηματοοικονομικού τομέα (**Bonus**) [I.14].
- Η **έλλειψη αυστηρού νομοθετικού πλαισίου** και η έως τότε **χαλαρότητα των ελέγχων** στην διακίνηση των κεφαλαίων [I.14].

Οι τρέχουσες εξελίξεις των τραπεζικών συστημάτων σε όλο τον κόσμο δημιούργησε τριγμούς στην φερεγγυότητα των πιστωτικών οργανισμών και αίσθημα ανασφάλειας απέναντι στην αξιοπιστία των τραπεζών στην μέτρηση της πιστοληπτικής ικανότητας των πελατών τους. Διαπιστώνουμε λοιπόν ότι ο πιστωτικός κίνδυνος αποτελεί την σοβαρότερη απειλή των χρηματοπιστωτικών ιδρυμάτων. Επιπρόσθετα, θα πρέπει να πραγματοποιείται μια ορθή αξιολόγηση της πιστοληπτικής ικανότητας, καθώς αυτές βρίσκουν πεδίο εφαρμογής στην

προσαρμογή των ασφαλίσεων, των επιτοκίων αποπληρωμής και των επιτοκίων αποπληρωμής των απαιτούμενων εγγυήσεων για την έκδοση ενός δανείου.

### **Κανονιστικό πλαίσιο πιστωτικού κινδύνου**

Η έλλειψη αυστηρού νομοθετικού πλαισίου και η έως τότε χαλαρότητα των ελέγχων στην διακίνηση των κεφαλαίων ήταν ένας από τους σημαντικούς λόγους που οδήγησαν στην κατάρρευση των αγορών.

Για τον λόγο αυτό, στον τραπεζικό τομέα θεσπίστηκε το σύμφωνο της Επιτροπής της Βασιλείας. Η ονομασία προέρχεται από την ομώνυμη πόλη της Ελβετίας, Basel, όπου έχει την έδρα της η Τράπεζα Διεθνών Διακανονισμών (Bank International Settlements) και βασιζείται στην επιτροπή για την εποπτεία του τραπεζικού συστήματος (Basel Committee on Banking Supervision).

Η Επιτροπή ιδρύθηκε το 1974 με απώτερο σκοπό την θέσπιση ενιαίων κανόνων και εποπτικών αρχών ελέγχου έτσι ώστε να εξασφαλιστεί η σταθερότητα και η ομαλή λειτουργία του τραπεζικού συστήματος, και η διαφάνεια των μεθόδων διαχείρισης και μέτρησης των κινδύνων και της αποτίμησης των χρηματοπιστωτικών μέσων. Το 1988 η Επιτροπή εισήγαγε ένα σύστημα κεφαλαιακής μέτρησης με την ονομασία Basel Capital Accord, το οποίο κάλυπτε μόνο τους πιστωτικούς κινδύνους των τραπεζών, ενώ δέκα χρόνια αργότερα (1998) εκδόθηκε το πλαίσιο εποπτείας του διεθνούς τραπεζικού συστήματος με την ονομασία «**Βασιλεία I**», το οποίο επεκτάθηκε ώστε να περιλαμβάνει και τους κινδύνους αγοράς. Επίσης, η πρόσφατα εφαρμοσθείσα «**Βασιλεία II**» (2007) βελτίωσε τον υπολογισμό του πιστωτικού κινδύνου και συμπεριέλαβε και τον λειτουργικό κίνδυνο. Τέλος, στο πλαίσιο «**Βασιλεία III**», παρουσιάζονται κανονιστικά πρότυπα που αφορούν την κεφαλαιακή επάρκεια και ρευστότητα των τραπεζών [I.2], [I.3].

### **Εταιρική αξιολόγηση πιστοληπτικής ικανότητας**

Η πιστοληπτική διαβάθμιση μιας εταιρείας αποτελεί οικονομικό δείκτη για τους δυνητικούς επενδυτές των χρεογράφων που εκδίδει, όπως τα εμπορικά ομόλογα της. Με βάση αυτή την διαβάθμιση, τα υψηλότερα επίπεδα διοίκησης της εταιρείας θέτουν όρια στα χαμηλότερα επίπεδα, ως προς την ποσότητα και το μέγεθος του κινδύνου που δύναται να αναλάβουν. Η αποτελεσματικότητα των ορίων αυτών εξαρτάται από την ικανότητα των διοικούντων όλων των επιπέδων να μετρούν και να παρακολουθούν την συσσώρευση του κινδύνου.

Η αξιολόγηση της πιστοληπτικής ικανότητας των εταιρειών, των χρεογράφων και των χωρών έχουν ανατεθεί σε διάφορους οίκους αξιολόγησης, με τους τρεις μεγαλύτερους να είναι η Moody's, η Standard & Poor's (S&P) και η Fitch. [I.12]. Ο οίκος αξιολόγησης Standard & Poor's είναι ο αρχαιότερος όλων, καθώς ιδρύθηκε το 1860. Με τη σημερινή όμως μορφή του υπάρχει από το 1941 και μετά τη συγχώνευση του εκδοτικού οίκου Poor's με τη Standard Statistics Bureau. Ο οίκος αξιολόγησης Moody's ιδρύθηκε το 1909 και ο οίκος Fitch το 1913.

Ένας οίκος αξιολόγησης παρακολουθεί τις εξελίξεις (monitoring) και παράλληλα αξιολογεί και διαβαθμίζει (rating) την πιστοληπτική αξιοπιστία ενός Δημοσίου ή ενός Ιδιωτικού φορέα που προχωρά στην έκδοση χρεογράφων (Κρατικά ή Εταιρικά Ομόλογα). Ανάλογα με τις εξελίξεις ένας οίκος αξιολόγησης αναβαθμίζει ή υποβαθμίζει την πιστοληπτική αξιοπιστία ενός εκδότη ομολόγων, γεγονός που έχει άμεση επίπτωση στο διαμορφούμενο ύψος των επιτοκίων δανεισμού που χρεώνεται. Όπως γίνεται έντονα ορατό και στην περίπτωση της παρούσας κρίσης χρέους της Ευρωζώνης. [I.8].

## ΚΕΦΑΛΑΙΟ 2<sup>ο</sup> : Τεχνικές & Μέθοδοι μέτρησης Πιστωτικού Κινδύνου

### Εισαγωγή

Η μέτρηση της πιστοληπτικής ικανότητας έχει λάβει εκτενή έκθεση, λόγω της πρόσφατης αύξησης της πιστωτικής βιομηχανίας και των οικονομικών εξελίξεων, έτσι ώστε το εκάστοτε τραπεζικό πιστωτικό τμήμα να αντιμετωπίζει τον αυξανόμενο πλήθος των πιστωτικών στοιχείων των καταναλωτών.

Ο παραδοσιακός τρόπος μέτρησης του πιστωτικού κινδύνου γινόταν, ή συνεχίζει να γίνεται, με υποκειμενική προσέγγιση (judgmental approach), όπου για την αξιολόγηση του εκάστοτε δανειολήπτη συλλέγονται κάποια προκαθορισμένα ποιοτικά και ποσοτικά χαρακτηριστικά, τα αποκαλούμενα ως τα 3, 4 ή 5 “C” s” (Character, Capital, Capacity, Collateral, Conditions). Αναλυτικότερα :

Φήμη του αιτούντος (borrower character - reputation) ή καλύτερα η συναλλακτική του συμπεριφορά, δηλ. η σχέση του πελάτη με την τράπεζα έχουν καθοριστική σημασία στο αν θα του προχωρηθεί δάνειο ή όχι.

Ο βαθμός δανειακής επιβάρυνσης (capital - leverage), διαφορετικά ο δείκτης δανεισμού ως προς την καθαρά του θέση, δεδομένου ότι επηρεάζει την πιθανότητα της επισφάλειας του, αφού όσο περισσότερο δανείζεται τόσο δυσκολότερη γίνεται η αποπληρωμή.

Μεταβλητότητα των κερδών (capacity – volatility of earnings). Είναι προτιμότερο μια εταιρεία να εμφανίζει σταθερά κέρδη για ένα μεγάλο χρονικό διάστημα από τα να υπάρχει μεγάλη διακύμανση (αστάθεια) στα κέρδη της.

Η ύπαρξη δευτερευουσών εγγυήσεων-εξασφαλίσεων (collaterals) για την εξόφληση οφειλής ή εκπλήρωση υποχρέωσης μειώνει τον κίνδυνο ενός δανείου.

Γενικές συνθήκες τις αγοράς (Conditions), δηλ. παράγοντες οι οποίοι σχετίζονται με την αγορά γενικότερα και με το πώς μπορεί να επηρεαστεί μια απόφαση από τις ευνοϊκές ή τις δυσμενείς συνθήκες στην αγορά του χρήματος και του κεφαλαίου.

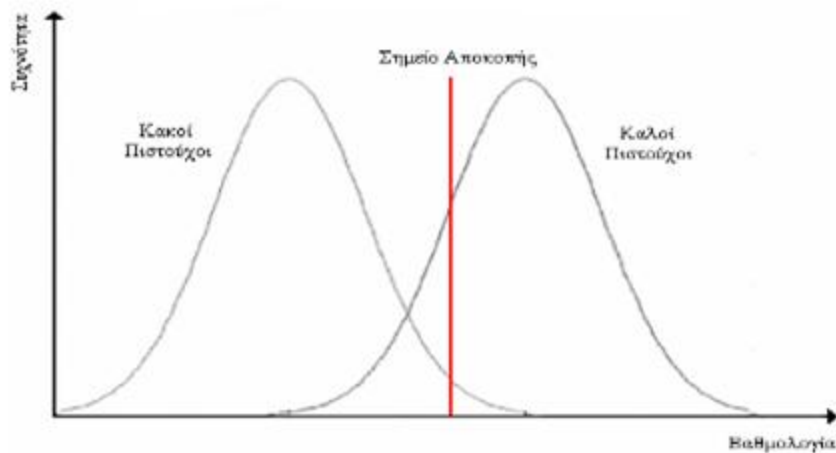
Ωστόσο, για να βελτιωθεί η αξιολόγηση των υποψηφίων οι τράπεζες χρησιμοποιούν τεχνικές credit scoring ή προγνωστικά μοντέλα για να κατατάσσουν τους προσφεύγοντες, μέσω των οποίων εξοικονομείται χρόνος, προσωπικό και ελαχιστοποιούνται τα σφάλματα.

Η μέτρηση του πιστωτικού κινδύνου μέσω των πιστωτικών διαβαθμίσεων (Credit Scoring) έχει λάβει μεγάλη δημοτικότητα λόγω της ικανότητας να λαμβάνει υπόψη «γνώσεις» από βάσεις δεδομένων. Μπορεί να οριστεί ως μια τεχνική που βοηθά τους δανειστές να αποφασίσουν εάν θα χορηγήσουν πίστωση, σε σχέση με τα χαρακτηριστικά τους, όπως η ηλικία, το εισόδημα και η οικογενειακή κατάσταση (**Chen & Huang, 2003**)[21].

Σημασιολογικά, η λέξη «credit» έχει τη ρίζα της στην λατινική γλώσσα και προέρχεται από το ρήμα «credere», που σημαίνει να πιστεύεις ή να εμπιστεύεσαι. Η σχέση αυτή των λέξεων αιτιολογεί και νοηματικά τη λειτουργία και τη φιλοσοφία του Credit Scoring, το οποίο εξετάζει τη πιστοληπτική ικανότητα του δανειολήπτη (ατόμου ή εταιρείας). Ουσιαστικά, η τράπεζα μέσα από αυτή την τεχνική αποφασίζει αν μπορεί να εμπιστευτεί ένα πελάτη της και να του δανείσει χρήματα με την προϋπόθεση να τα επιστρέψει (**Bessis, 2002**)[18].



Οι πρώτες έρευνες σε credit scoring έγιναν από τους **Fisher (1936) [35]** και **Durand (1941) [31]**, οι οποίοι εφάρμοσαν γραμμική και τετραγωνική διαχωριστική ανάλυση (discriminant analysis), αντίστοιχα, για την κατηγοριοποίηση των αιτούντων σε «καλούς» και «κακούς».



Εικόνα 1: Διαχωρισμός «Καλών» και «Κακών» πελατών.

Στην συνέχεια, αναπτύχθηκαν από το 1960 με τις μελέτες των **Beaver (1967) [14]**, **Altman (1968) [6]**, και άλλων, **Zurada, J., and Lonial, S., 2005 [107]**, **Chye Koh, H., Chin Tan, W., and Peng Goh, C., 2006 [24]**, **Kirkos, E., Spathis, C., and Manolopoulos., Y., 2007 [59]**, **Atish P, S., and Huimin, Z., 2008 [9]**, **Yeh, I. C., and Lien, C. h., 2009 [103]**.

Το **1973**, οι **Fischer Black, Myron Scholes και Robert Merton [34]** ανακάλυψαν ένα μαθηματικό τύπο που θα άλλαζε την πορεία των χρηματαγορών. Η εξίσωση των 3 οικονομολόγων ήταν η πρώτη ευρέως γνωστή απόπειρα αποτύπωσης με ένα απλό και παράλληλα αποδοτικό τρόπο της τυχαίας πορείας της τιμής των μετοχών στο μέλλον. Αποδείχτηκε τόσο σημαντική που το 1997 οι 3 οικονομολόγοι βραβεύτηκαν με το Nobel (παρευρέθηκαν μόνο οι 2, αφού ο Black είχε πεθάνει τον προηγούμενο χρόνο). Μέχρι τότε η αποτίμηση των δικαιωμάτων γινόταν με εμπειρικό τρόπο και δεν υπήρχε ένας κοινά αποδεκτός τρόπος επικοινωνίας. Ο **Mester το 1997 [71]** αναφέρει την ευρεία διάδοση των μοντέλων credit scoring. Συγκεκριμένα, το 97% των τραπεζών εμφανίζεται να τα χρησιμοποιεί για να εγκρίνει την έκδοση των πιστωτικών καρτών ενώ το 70% τα χρησιμοποιεί στην δανειοδότηση των μικρών επιχειρήσεων ούτως ώστε να προχωρήσει στην έγκριση ή στην απόρριψη του δανείου ανάλογα με το αποτέλεσμα.

Το 2000 ο **Lyn (2000) [72]** πραγματοποίησε μια ολοκληρωμένη έρευνα για τις πιστωτικές και συμπεριφορικές βαθμολογήσεις (behavioral scoring).

Τα συστήματα Συμπεριφορικής βαθμολόγησης (behavioral scoring) επιτρέπουν στους δανειστές να λαμβάνουν καλύτερες αποφάσεις για τη διαχείριση των υφιστάμενων πελατών μέσα από την πρόβλεψη των μελλοντικών επιδόσεων τους. Οι αποφάσεις που πρέπει να γίνουν περιλαμβάνουν τι πιστωτικό όριο θα εκχωρηθεί, εάν θα αγοραστούν νέα προϊόντα για τους συγκεκριμένους πελάτες, και εάν θα υπάρξει αρνητική έκβαση στην διαχείριση του χρέους. Οι πρόσθετες πληροφορίες στα συστήματα συμπεριφορικής βαθμολόγησης, σε σχέση με τα υποδείγματα credit scoring, είναι η επαναπληρωμή και η ιεραρχική ιστορία του πελάτη. Τα συστήματα αυτά χωρίζονται σε δύο προσεγγίσεις, (α) σε εκείνα που επιδιώκουν την χρήση των υποδειγμάτων credit scoring προσθέτοντας τις νέες πληροφορίες, και (β) σε εκείνα που στοχεύουν στην κατασκευή μοντέλων πιθανότητας της συμπεριφοράς των πελατών. Αυτά δύναται επίσης να κατηγοριοποιηθούν αναλόγως εάν η πληροφορία για την εκτίμηση των παραμέτρων προκύπτει από το δείγμα των προηγούμενων πελατών ή εάν λαμβάνεται με

Bayesian μεθόδους οι οποίες εφοδιάζονται με τις αρχικές πεποιθήσεις της επιχείρησης για την ενδεχόμενη συμπεριφορά του πελάτη. Στην ουσία και οι δύο κατηγορίες αφορούν μοντέλα με αλυσίδες Markov όπου ο πελάτης πραγματοποιεί άλματα από την μια κατάσταση στην άλλη, αναλόγως της συμπεριφοράς του **Lyn (2000) [72]**.

### **Είδη Βαθμολόγησης**

Τα συστήματα αυτά είναι σημαντικά εργαλεία καθώς :

- Παρέχουν μια κοινή γλώσσα επικοινωνίας μεταξύ των ενδιαφερομένων,
- Αποτελούν την βάση της πιστοληπτικής πολιτικής ενός οργανισμού,
- Αποπνέουν στην λήψη αποφάσεων για χορήγηση ή μη δανείων.
- Βελτώνουν την αποτελεσματικότητα της πιστοληπτικής διαδικασίας.

Γενικά, μπορούμε να πούμε ότι υπάρχουν τα εξής είδη βαθμολόγησης (**Paleologo 2010) [79]**:

### **Application Scoring – Βαθμολόγηση Αιτήσεων:**

Πρόκειται για τον βασιλιά της βαθμολόγησης, όπου αποτελείται από την εκτίμηση της πιστοληπτικής ικανότητας των νέων αιτήσεων που ζητούν πίστωση. Εκτιμά τον πιστωτικό κίνδυνο λαμβάνοντας υπόψη κοινωνικές, δημογραφικές και οικονομικές συνθήκες μιας νέας αίτησης, ώστε να αποφασίσει εάν θα χορηγήσει πίστωση ή όχι. Βασικά χαρακτηριστικά που εξετάζονται στα μοντέλα Βαθμολόγησης Αιτήσεων είναι μισθολογικά στοιχεία και γενικά πηγές εισοδήματος, υφιστάμενες καταθέσεις και λοιπά περιουσιακά στοιχεία, οικογενειακή κατάσταση (καθώς και πλήθος παιδιών και προστατευμένων μελών), μηνιαία έξοδα, ύψος δανείου, κ.α.

### **Behavioral Scoring – Βαθμολόγηση Συμπεριφοράς :**

Όπως προαναφέραμε είναι παρόμοιο με το κλασικό υπόδειγμα βαθμολόγησης, με την διαφορά όμως ότι αναφέρεται στους υφιστάμενους πελάτες για την πρόβλεψη των μελλοντικών επιδόσεων τους.

### **Collection Scoring - :**

Η βαθμολόγηση συλλογής ταξινομεί τους πελάτες σε διαφορετικές ομάδες ανάλογα με τα επίπεδα της αφερεγγυότητά τους. Με άλλα λόγια κατηγοριοποιεί τους πελάτες σε αυτούς που χρειάζονται περισσότερο αποφασιστικές ενέργειες από εκείνους που δεν απαιτούν παρακολούθηση κάποιας παραβατικότητας.

### **Fraud Detection – Ανίχνευση Απάτης :**

Κατηγοριοποιεί τους υποψηφίους ανάλογα με την πιθανότητα ο αιτών να είναι ένοχος απάτης.

### **Εφαρμογές**

Τα υποδείγματα Credit Scoring εφαρμόζονται κυρίως (**Ravi, V., 2007) [82]**

- στις πιστωτικές κάρτες
- στα στεγαστικά δάνεια
- στα δάνεια μικρών επιχειρήσεων

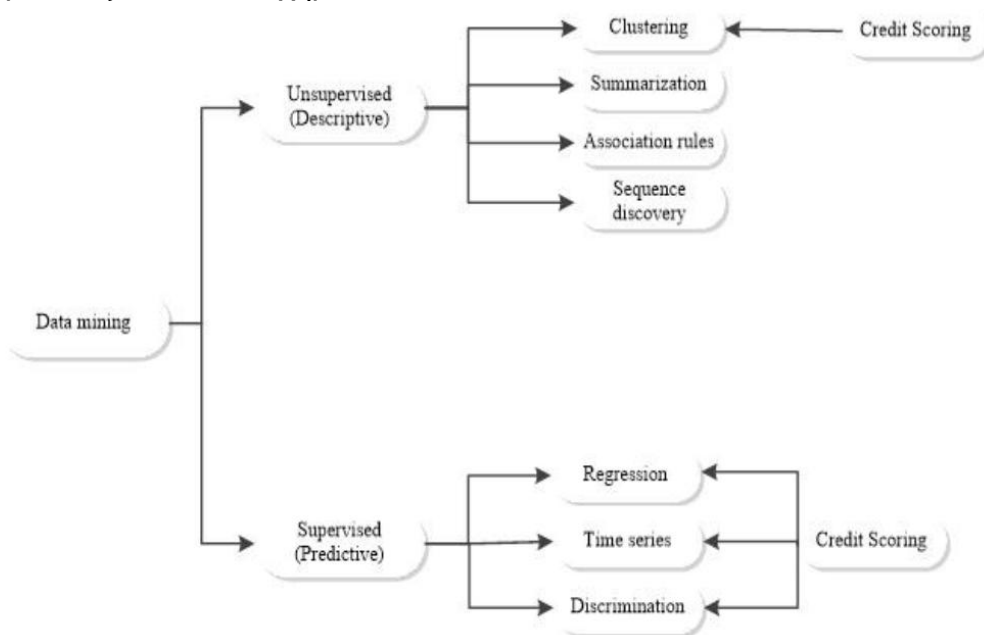
Ωστόσο, σε πιο πρόσφατες μελέτες, τα υποδείγματα Credit Scoring έχουν βρει πεδίο εφαρμογής και στην πρόβλεψη ασφαλιστικών απωλειών (**Golden et al., 2016) [70]**, ενώ, η ραγδαία αύξηση της ηλεκτρονικής κοινωνικής δικτύωσης δεν έχει αφήσει ανεπηρέαστους τους ερευνητές που διερευνούν και την συσχέτιση της απειλής του κοινωνικού αποκλεισμού στην πιστοληπτική ικανότητα. **[102] (Yanhao Wei et al., 2015)**.

## Data mining – Εξόρυξη Δεδομένων

«Η **εξόρυξη δεδομένων (data mining)** είναι η διαδικασία της επιλογής, της εξερεύνησης, και της μοντελοποίησης μεγάλης ποσότητας δεδομένων για να ανακαλυφθούν κανονικότητες ή σχέσεις, με στόχο την λήψη σαφών και χρήσιμων αποτελεσμάτων για τον ιδιοκτήτη της βάσης δεδομένων» (Giudici, 2003) [38].

«Η εξόρυξη δεδομένων είναι ένα χρήσιμο εργαλείο για τη λήψη στρατηγικών αποφάσεων και παίζουν σημαντικό ρόλο στην τμηματοποίηση της αγοράς, την εξυπηρέτηση πελατών, την ανίχνευση της απάτης στις πιστωτικές και την συγκριτική αξιολόγηση» (Giudici, 2001) [39].

Οι πιο διαδεδομένες τεχνικές και μέθοδοι για την εφαρμογή Data mining σε credit scoring, παρουσιάζονται στο σχήμα 2.



Εικόνα 2: Διαδεδομένες Τεχνικές και Μέθοδοι που εφαρμόζονται σε credit scoring.

Οι οικονομετρικές τεχνικές (econometric techniques) αναφέρονται σε στατιστικές μεθόδους όπως :

- η διαχωριστική ανάλυση (discriminant analysis),
  - η πολλαπλή παλινδρόμηση (multiple regression),
  - το λογιστικό μοντέλο (logistic regression),
  - το κανονικό μοντέλο πιθανότητας (probit model),
- με σκοπό να εκτιμηθεί η πιθανότητα χρεοκοπίας (default probability) χρησιμοποιώντας ως ερμηνευτικές μεταβλητές διάφορους χρηματοοικονομικούς δείκτες και γνωρίσματα πελατών.

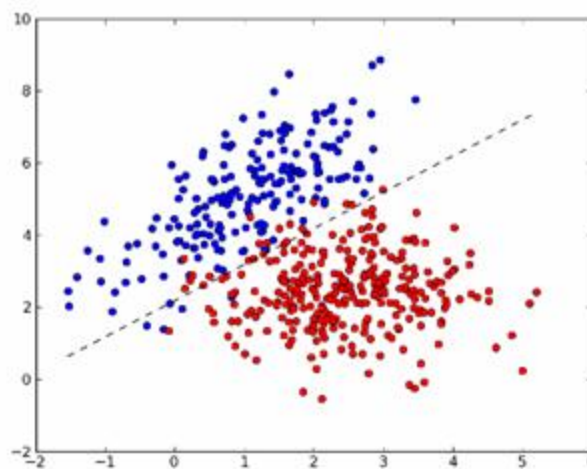
Ανεξάρτητα από την τεχνική που θα επιλεγεί από το πιστωτικό ίδρυμα, θα πρέπει να τηρούνται συγκεκριμένες προϋποθέσεις, οι οποίες είναι (Καλφάογλου (1999)):

- Να είναι πλήρως τεκμηριωμένες από θεωρητική και πρακτική μεριά.
- Να επαληθεύονται τα αποτελέσματα.
- Να περιέχονται στις διαδικασίες λήψης αποφάσεων του πιστωτικού ιδρύματος.

Ας δούμε λοιπόν ποιες είναι οι πιο διαδεδομένες τεχνικές :

## 2.1 Διαχωριστική Ανάλυση (Discriminant analysis)

Η Διαχωριστική ανάλυση, που μελετήθηκε από τον **Fisher το 1936 [35]**, είναι μια εναλλακτική λύση της λογιστικής παλινδρόμησης που υποθέτει ότι οι επεξηγηματικές μεταβλητές ακολουθούν μια πολυμεταβλητή κανονική κατανομή και έχουν κοινό πίνακα διακυμάνσεων - συνδιακυμάνσεων. Σκοπός της μεθόδου είναι η κατάταξη του δείγματος σε  $k$  ομάδες. Όμως, όπως έχουμε ήδη αναφέρει ο διαχωρισμός γίνεται σε δύο κατηγορίες («καλούς» και «κακούς»). Οι **Hand and Henley (1996) [40]** αναφέρουν ότι μια πρώτη προσπάθεια εφαρμογής της μεθόδου έγινε από τον **Durand (1941) [31]** ο οποίος αναφέρει ότι η μέθοδος αυτή εμφανίζει αρκετά καλά αποτελέσματα. Ωστόσο η μέθοδος, αν και χρησιμοποιήθηκε, και χρησιμοποιείται ακόμα, ευρέως, εντούτοις αμφισβητήθηκε από ορισμένους ερευνητές, όπως οι **Rosenberg and Gleit (1994) [83]** οι οποίοι άσκησαν κριτική στην εφαρμογή της μεθόδου και στα αποτελέσματα άλλων ερευνών.



Εικόνα 3: Παράδειγμα διαχωριστικής ανάλυσης όπου φαίνεται ο διαχωρισμός που πραγματοποιεί η συνάρτηση στα δεδομένα.

Η διαχωριστική ανάλυση αποτελεί μια πολυμεταβλητή τεχνική με την οποία αναλύεται μεγάλο σύνολο δεδομένων με σκοπό να μεγιστοποιηθεί η διακύμανση μεταξύ των ομάδων (between group variance) και να ελαχιστοποιηθεί η διακύμανση εντός των ομάδων (within group variance), δηλ. στόχος της μεθόδου είναι η ελαχιστοποίηση της απόστασης εντός των παρατηρήσεων της κάθε ομάδας, και η μεγιστοποίηση της απόστασης μεταξύ των διαφόρων ομάδων χρησιμοποιώντας μια διαχωριστική συνάρτηση. Τα σκορ των δύο ομάδων θα πρέπει να είναι όσο το δυνατόν πιο απομακρυσμένα έτσι ώστε να μπορούμε εύκολα με βάση αυτά τα σκορ να κάνουμε διαχωρισμό και ταξινόμηση των ομάδων. Η τεχνική βασίζεται στις υποθέσεις ότι (α) οι πίνακες συνδιακυμάνσεων κάθε ταξινομημένης ομάδας είναι ίσοι, έτσι χρησιμοποιεί την συνδυασμένη κοινή (pooled) εκτίμηση  $S_p$ , και (β) κάθε ταξινομημένη ομάδα ακολουθεί μια πολλαπλή κανονική κατανομή. Προκύπτει έτσι ότι βασικό μειονέκτημα της μεθόδου είναι οι μη ρεαλιστικές υποθέσεις. Επίσης, οι **(Doumpos & Zopounidis, 2014 [28])** επισημάνουν την ευαισθησία της μεθόδου στην ύπαρξη ακραίων παρατηρήσεων (outliers) **(Doumpos, 2014b [28], Ince, 2009 [46], Kambal, 2013, [50])**.

Εάν υποθέσουμε ότι τα σκορ δίνονται ως  $U_1$  για την 1<sup>η</sup> ομάδα και  $U_2$  για την 2<sup>η</sup> ομάδα, τότε ένα μέτρο που μας δίνει την απόσταση των δύο ομάδων είναι η απόσταση των μέσων τιμών

$(\bar{U}_1 - \bar{U}_2)$ . Ο Fisher πήρε σαν μέτρο απόστασης των δύο ομάδων την ποσότητα  $D = \frac{|\bar{U}_1 - \bar{U}_2|}{S_p}$ ,

$$\text{όπου } S_p = \frac{\sum_{i \in G_1} (U_i - \bar{U}_1)^2 + \sum_i (U_i - \bar{U}_2)^2}{n_1 + n_2 - 2}.$$

Ο δείκτης συνεπάγεται ότι λαμβάνουμε υπόψη τις παρατηρήσεις που ανήκουν στην  $i$  ομάδα. Σκοπός είναι να μεγιστοποιήσουμε την ποσότητα  $D$  ή αντίστοιχα την ποσότητα  $D^2$ . Τα χαρακτηριστικά δίνονται από τον γραμμικό συνδυασμό  $L^T x$  οπότε πρέπει να

$$D^2 = \frac{[L^T (\bar{x}_1 - \bar{x}_2)]^2}{L^T S_p L}$$

μεγιστοποιήσουμε την ποσότητα

$$D^2 = (\bar{x}_1 - \bar{x}_2)^T S_p (\bar{x}_1 - \bar{x}_2)$$

και συνεπώς ο διαχωριστικός κανόνας έχει ως

$$L^T x \geq m (L^T x - m \geq 0)$$

$m = (\bar{U}_1 + \bar{U}_2) / 2 = L^T (\bar{x}_1 + \bar{x}_2) / 2$  όπου αν

τότε κατατάσσουμε στην 1η γενικεύεται και για  $k$  - ομάδες. Ας υποθέσουμε

ομάδα αλλιώς στην δεύτερη. Η μέθοδος φυσικά

ότι έχουμε  $p$  χαρακτηριστικά σε ένα πλήθος  $n$  πελατών ενός τραπεζικού ιδρύματος. Το τυχαίο διάνυσμα  $X = (X_1, X_2, \dots, X_p)$  θα αποτελείται από  $p$  τυχαίες ανεξάρτητες μεταβλητές. Έτσι η γενική μορφή της διαχωριστικής συνάρτησης (discriminant function) είναι :

$$Z = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p$$

Όπου  $a$ - σταθερός όρος και  $b_i, i = 1, 2, \dots, p$  οι συντελεστές των τυχαίων μεταβλητών.

Η πιο γνωστή μέθοδος εφαρμογής της διαχωριστικής ανάλυσης στα μοντέλα βαθμολόγησης πιστοληπτικής ικανότητας είναι το **Z-score του Altman**, το οποίο έχει την μορφή

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

Όπου,

$X_1$  - το κεφάλαιο κίνησης προς το συνολικό κεφάλαιο (working capital / total assets),

$X_2$  - τα αδιανέμητα κέρδη προς το συνολικό κεφάλαιο (retained earnings / total assets),

$X_3$  - τα κέρδη προ τόκων και φόρων προς το συνολικό κεφάλαιο (earnings before interest and taxes/total assets),

$X_4$  - η αγοραία αξία κεφαλαίων προς την λογιστική αξία συνολικού παθητικού (market value equity/book value of total debt), και

$X_5$  - οι πωλήσεις προς το συνολικό κεφάλαιο (sales/total assets).

Με βάση τα παραπάνω, όταν η βαθμολογία είναι μεγάλη συνεπάγεται ότι η επιχείρηση έχει μικρότερη πιθανότητα πτώχευσης. Πιο συγκεκριμένα τιμή του **Z < 1.81** συνεπάγεται ότι η επιχείρηση θα πτωχεύσει («κακή»), ενώ τιμή **Z > 2.99** κατατάσσει την επιχείρηση ως «καλή». Για το διάστημα **[1.81 , 2.99]** οι επιχειρήσεις εντάσσονται στην «ζώνη άγνωστης κατάταξης» (zone of ignorance or gray area) και απαιτούν περισσότερη διερεύνηση. Το υπόδειγμα αυτό εφαρμόζεται και στις περιπτώσεις ομολόγων υπολογίζοντας τις βαθμολογίες για κάθε μία κατηγορία πιστοληπτικής διαβάθμισης. Οι Caouette et al. (1998) αναφέρουν ότι η μέση βαθμολογία ορισμένων ομολόγων με πιστοληπτική διαβάθμιση AAA\* είναι 5, ενώ για πιστοληπτική διαβάθμιση B είναι 1.67.

Παρόλο που η μέθοδος εισαχθεί από το 1936, από το 1974 έως και σήμερα χρησιμοποιείται ευρέως στην πιστοληπτική βαθμονόμηση (credit scoring), είτε αυτούσια είτε συνδυασμένη με νεότερες τεχνικές (βλ. επόμενη παράγραφο για τα υβριδικά μοντέλα), καθώς υπάρχουν και

υποστηρικτές της γραμμικής διαχωριστικής ανάλυσης (LDA), οι οποίοι υποστηρίζουν ότι είναι (α) περισσότερο αποτελεσματική για σκοπούς βαθμολόγησης πιστοληπτικής ικανότητας, (β) πολύ ευκολότερη στην εφαρμογή και την ερμηνεία, και (γ) περισσότερο αποτελεσματική όταν εφαρμόζεται σε μεγάλο δείγμα [5]

## 2.2 Λογιστική Παλινδρόμηση (Logistic regression)

Η λογιστική παλινδρόμηση είναι μια περίπτωση γραμμικής παλινδρόμησης, που προτάθηκε από τον **Berkson (1944)** [17] και αναπτύχθηκε από τον **David Cox (1958)**[25]. Αποβλέπει στην πρόβλεψη ενός δίπμου αποτελέσματος [η εξαρτημένη μεταβλητή είναι κατηγορική και δίπμη  $Y = \{y_1, y_2\}$ ], από ένα σύνολο ανεξάρτητων μεταβλητών  $X = \{X_1, \dots, X_p\}$  που μπορεί να είναι συνεχείς, διακριτές, δίπμες, ή συνδυασμό αυτών. Τα πλεονεκτήματα της μεθόδου είναι ότι δεν απαιτείται γραμμικότητα στην σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης, και δεν απαιτείται κανονικότητα στις κατανομές τους. Κύριο όφελος της μεθόδου είναι ότι μπορεί να δημιουργήσει ένα απλό πιθανολογικό μοντέλο ταξινόμησης [5]. Από την άλλη η αδυναμία του μοντέλου αυτού είναι ότι οι ανεξάρτητες μεταβλητές οφείλουν να σχετίζονται γραμμικά με την link function (logit) της εξαρτημένης μεταβλητής.

Αναλυτικότερα, στην λογιστική παλινδρόμηση εξετάζουμε την πιθανότητα (τα ποσοστά) εμφάνισης των δύο κατηγοριών σε σχέση με τις ανεξάρτητες μεταβλητές – παράγοντες. Επειδή σκοπός είναι να εκτιμηθεί η πιθανότητα εμφάνισης ενός συμβάντος, συνεπάγεται ότι οι πιές που θα πρέπει να προκύπτουν από το γραμμικό υπόδειγμα να περιέχονται στο διάστημα  $[0,1]$ . Για τον λόγο αυτό υποθέτουμε ότι η μεταβλητή ακολουθεί διωνυμική κατανομή και ότι η σύνδεση της πιθανότητας εμφάνισης του γεγονότος  $p$  συνδέεται με το γραμμικό υπόδειγμα μέσω της link function :

$$\log\left(\frac{p_i}{1-p_i}\right) = X^T \beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Όπου  $p_i = P(Y = y_1)$  και  $\log\left(\frac{p_i}{1-p_i}\right)$  ο λογάριθμος του λόγου σχετικής πιθανότητας.

Απολογαριθμίζοντας προκύπτει ότι η πιθανότητα το  $i$ -οστό άτομο να ανήκει στην κατηγορία της εξαρτημένης μεταβλητής  $y_1$ , να είναι :

$$p_i = \frac{\exp\{X_i \beta\}}{1 + \exp\{X_i \beta\}} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

Στο μοντέλο της λογιστικής παλινδρόμησης η εκτίμηση των συντελεστών πραγματοποιείται με την μέθοδο μέγιστης πιθανοφάνειας (maximum likelihood method), αντί της μεθόδου ελαχίστων τετραγώνων.

Η ερμηνεία τους όμως δεν προκύπτει με τον ίδιο τρόπο όπως στην γραμμική παλινδρόμησης, και θα πρέπει να γίνει τροποποίηση, ώστε να εκφραστούν με την κατάλληλη μορφή, δηλ.  $e^{\beta_i}$ . Τα περισσότερα προγράμματα στον πίνακα των συντελεστών εμφανίζουν και την σχέση  $e^{\beta_i}$ . Κάθε συντελεστής εκφράζει τη μεταβολή του λογαρίθμου της σχετικής πιθανότητας  $\log\left(\frac{p_i}{1-p_i}\right)$  για μια μονάδα αύξησης της ανεξάρτητης μεταβλητής.

Έστω ότι έχουμε ένα μοντέλο με μια ανεξάρτητη δίπμη μεταβλητή (π.χ. το φύλο όπου 0=γυναίκες και 1=άνδρες). Οπότε έχουμε :

$$\log\left(\frac{p_A}{1-p_A}\right) = \beta_0 + \beta_1 \Rightarrow \frac{p_A}{1-p_A} = e^{\beta_0 + \beta_1}$$

που είναι η σχετική πιθανότητα πραγματοποίησης του συμβάντος στους άνδρες, και

$$\log\left(\frac{p_\Gamma}{1-p_\Gamma}\right) = \beta_0 \Rightarrow \frac{p_\Gamma}{1-p_\Gamma} = e^{\beta_0}$$

που είναι η σχετική πιθανότητα πραγματοποίησης του συμβάντος στις γυναίκες. Διαιρώντας τα δύο μέλη λαμβάνουμε τον λόγο συμπληρωματικών πιθανοτήτων (odds ratio) πραγματοποίησης του συμβάντος των ανδρών έναντι των γυναικών, δηλ.

$$\frac{\frac{p_A}{1-p_A}}{\frac{p_\Gamma}{1-p_\Gamma}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Όπως και στην γραμμική παλινδρόμηση, ομοίως οι συντελεστές του μοντέλου, οι οποίοι αποτελούν σημειακές εκτιμήσεις, υπόκεινται σε έλεγχο για την σημαντικότητα τους στο μοντέλο. Δηλ. η υπόθεση που εξετάζεται είναι η :

$$H_0 : \hat{\beta}_i = 0$$

$$H_1 : \hat{\beta}_i \neq 0$$

Ο έλεγχος γίνεται σε (1-α)% επίπεδο σημαντικότητας και η στατιστική συνάρτηση υπό την  $H_0$

είναι η  $\frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)}$  η οποία καλείται Wald και ακολουθεί ασυμπτωτικά την κανονική κατανομή, ενώ

το τετράγωνο αυτής ακολουθεί την  $\chi^2$ -κατανομή με 1 βαθμό ελευθερίας. Η υπόθεση απορρίπτεται αν η τιμή του ελέγχου είναι μεγαλύτερη από  $\chi^2_{1;\alpha/2}$ .

Η σύγκριση μοντέλων με σκοπό την επιλογή βέλτιστου μοντέλου, πραγματοποιείται με την βοήθεια της ελεγχουσυνάρτησης Deviance (που στην ουσία είναι ένα likelihood ratio test) η οποία ορίζεται ως :

c

Όπου

$\hat{l}_2$  - η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας του μοντέλου 2,

$\hat{l}_1$  - η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας του μοντέλου 1, και επίσης το μοντέλο 2 είναι υποσύνολο (nested) του μοντέλου 1, και ακολουθεί την  $\chi^2$ -κατανομή με k-p-1 βαθμούς ελευθερίας.

Η λογιστική παλινδρόμηση έχει χρησιμοποιηθεί εκτενώς σε εφαρμογές credit scoring.

Από πρόσφατες μελέτες, οι **Liang, Y., και Xin, H., (2009) [68]** σε μια εμπειρική μελέτη 780 κινεζικών εταιρικών ομολόγων και 72 χρηματοοικονομικών δεικτών χρησιμοποιούν την λογιστική παλινδρόμηση επιλύοντας το πρόβλημα της πολυσυγγραμμικότητας (της μεθόδου), έχοντας αρχικά προεπεξεργαστεί τα οικονομικά δεδομένα με διακριτοποίηση (discretization) πριν την εφαρμογή της μεθόδου. Τα αποτελέσματα δείχνουν ότι με τον τρόπο αυτό θα μπορούσαν να καθιερώσουν την χρήση ενός λογικού λογιστικού μοντέλου το οποίο θα περιέχει, όχι μόνο λιγότερες επεξηγηματικές μεταβλητές, αλλά και υψηλή ακρίβεια πρόβλεψης. Συμπεραίνουν ότι ο αλγόριθμος αυτός της διακριτοποίησης μπορεί να παρέχει έναν αποτελεσματικό τρόπο για την επίλυση των περιορισμών και να παράγει ένα πιο συνοπτικό και αξιόπιστο μοντέλο οικονομικής αξιολόγησης.

Ο **Nguyen, (2015) [77]** παρουσιάζει τον τρόπο εφαρμογής της μεθόδου σε credit scoring χρησιμοποιώντας το αρχείο αυτοκινητιστικών δανείων από μια από τις μεγαλύτερες κατασκευάστριες εταιρείες αυτοκινήτων στην Κίνα. Τα αποτελέσματα της έρευνας δείχνουν ότι η οικογενειακή κατάσταση «παντρεμένος», «ιδιοκτήτη σπιτιού», «φύλο – θηλυκό», η ηλικία, «εργαζόμενος σε δημόσια ιδρύματα, ξένα ή κοινές επιχειρήσεις», το κατώτερο ποσοστό πληρωμών και το ανώτατο όριο σε μήνες για το βιβλίο των τρεχούμενων λογαριασμών, επηρεάζουν αρνητικά την πιθανότητα αθέτησης.

Εν γένει η λογιστική παλινδρόμηση θεωρείται ότι παρέχει το πιο ακριβές μοντέλο μεταξύ των συμβατικών μοντέλων (**Shi, Wang, Qi, & Cheng, 2015) [85]** . Οι **Khemais et al. (2016) [57]**, που στην εργασία τους αποσκοπούν στην ανάπτυξη ενός μοντέλου πρόβλεψης της πιθανότητας αθέτησης μικρών και μεσαίων επιχειρήσεων (SME) από δεδομένα μιας Τυνησιακής εμπορικής τράπεζας, καταλήγουν στο ίδιο συμπέρασμα, χρησιμοποιώντας τις δύο μεθόδους, λογιστική παλινδρόμηση και γραμμική διαχωριστική ανάλυση. Στην εργασία τους γίνεται χρήση μιας βάσης δεδομένων η οποία αποτελείται από 195 πιστωτικά αρχεία που χορηγούνται σε SME της Τυνησίας και χωρίζονται σε πέντε τομείς, βιομηχανίας, γεωργίας, τουρισμού, εμπορίου και υπηρεσιών, για την περίοδο από το 2012 έως το 2014. Τα εμπειρικά αποτελέσματα υποστηρίζουν την ιδέα ότι οι δύο αυτές τεχνικές βαθμολόγησης έχουν στατιστικά σημαντική ισχύ στην πρόβλεψη του κινδύνου αθέτησης των επιχειρήσεων. Καταλήγουν στο ότι η λογιστική παλινδρόμηση κατηγοριοποιεί τις επιχειρήσεις με ποσοστό ορθής ταξινόμησης 76,7%, έναντι 76,4% της γραμμικής διαχωριστικής ανάλυσης, δίνοντας έτσι μια ελαφρά υπεροχή σε αυτή την μέθοδο.

## 2.3 Naïve Bayesian classifier

Ένας μπευζιανός ταξινομητής είναι μια απλή πιθανολογική ταξινόμηση που βασίζεται στο θεώρημα της υπό συνθήκης πιθανότητας του Bayes, και προβλέπει την πιθανότητα ένα νέο άτομο ή αντικείμενο (ένα διάνυσμα χαρακτηριστικών) να ανήκει σε μια από τις προκαθορισμένες κατηγορίες. Αυτού του είδους η κατηγοριοποίηση χαρακτηρίζεται από μεγάλη ταχύτητα εκτέλεσης της διαδικασίας σε μεγάλες Βάσεις δεδομένων. Με την εισαγωγή της υπόθεσης ότι τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους, η μέθοδος αυτή έχει προταθεί με την ονομασία Απλοϊκή (Naïve) κατηγοριοποίηση κατά Bayes και μελετήθηκε εκτενώς την δεκαετία του 1950.

Ο κανόνας του Bayes μας επιτρέπει να προσδιορίσουμε τις εκ των υστέρων πιθανότητες για κάθε παρατήρηση η οποία ανήκει σε μια συγκεκριμένη ομάδα (κλάση). Δηλ. υπολογίζουμε τις πιθανότητες  $P(Y|X)$ , όπου  $Y=(y_1, \dots, y_k)$  μια τυχαία μεταβλητή η οποία κατηγοριοποιείται σε  $k$  κατηγορίες, και  $X=(X_1, \dots, X_p)$  ένα σύνολο από  $p$  επεξηγηματικές μεταβλητές. Η επιλογή της εκ των προτέρων κατανομής πρέπει να γίνεται με προσοχή διότι καθιστά την ανάλυση υποκειμενική (subjective Bayesian analysis), αφού διαφορετική επιλογή θα οδηγήσει σε διαφορετικά αποτελέσματα. Ωστόσο μια «λογική» (μη πληροφοριακή) εκ των προτέρων χάνει την επίδραση της καθώς συγκεντρώνονται όλο και περισσότερα δεδομένα. Έτσι οι πιθανότητες

$$P(Y|X) = \frac{P(X|Y)P(Y)}{\sum P(X|Y)P(Y)}$$

υπολογίζονται από τα δεδομένα.

Θα μπορούσαμε να σχολιάσουμε ότι η συγκεκριμένη μέθοδος υποθέτει ότι η ύπαρξη (ή ανυπαρξία) ενός χαρακτηριστικού μιας κλάσης είναι άσχετη (ανεξάρτητη) με την ύπαρξη (ή ανυπαρξία) από οποιοδήποτε άλλο χαρακτηριστικό. **[87]**



Το κύριο μειονέκτημα αυτής της μεθόδου είναι ότι η προγνωστική ακρίβεια σχετίζεται σε μεγάλο βαθμό με αυτή την υπόθεση (ανεξαρτησία). Σε περίπτωση που αυτή δεν ισχύει τότε δεν δίνει καλά αποτελέσματα. Αντιθέτως, στα πλεονέκτηματά της μεθόδου είναι ότι απαιτεί μια μικρή ποσότητα δεδομένων για εκπαίδευση της εκτίμησης των παραμέτρων (μέση τιμή και διακυμάνσεις των μεταβλητών) που είναι απαραίτητα για την ταξινόμηση, και έχει την δυνατότητα να χειριστεί ελλείπουσα δεδομένα (παραλείποντας ωστόσο τις αντίστοιχες πιθανότητες). [87]

Ορισμένες, από τις πρόσφατες έρευνες όπου εφαρμόζεται η μέθοδος είναι :

Οι **Stefanescu et al. (2009) [89]** αναπτύσσουν ένα μοντέλο το οποίο συλλαμβάνει μοτίβα της ετερογένειας του πελάτη και αξιολογήσεις του βαθμού μεταπήδησης (σε άλλη κατάσταση) διαμέσου μη παρατηρούμενων συστηματικών (απότομων) μακροοικονομικών μεταβολών. Περιγράφουν ένα Μπευζιανό ιεραρχικό πλαίσιο προτυποποίησης από ιστορικά δεδομένα μεταπηδήσεων, και δείχνουν πώς μπορούν να αξιολογηθούν οι προβλεπόμενες αποδόσεις του μοντέλου, ακόμα και με σπάνια δεδομένα. Στο τέλος, γίνεται μια ανάλυση δεδομένων από την Standard and Poor's για την περίοδο από το 1981 έως το 2007, για να δείξουν την εφαρμογή της μεθόδου τους στις σύγχρονες οδηγίες της Βασιλείας II για την τρέχουσα πολιτική στον υπολογισμό των πιθανοτήτων αθέτησης που αγορά χαμηλού κινδύνου στοιχείων του ενεργητικού.

Οι **Αντωνάκης και Σφακιανάκης (2009) [8]**, εξετάζουν την αποτελεσματικότητα του «αφελή» κανόνα Bayes (Naïve Bayes Rule) για την κατασκευή κανόνων ταξινόμησης (πιστωτική scorecards), στο πλαίσιο της αξιολόγησης των πιστωτικών αιτήσεων (credit scoring). Στην μελέτη τους χρησιμοποιούν δύο σύνολα πραγματικών δεδομένων credit scoring, και συγκρίνουν την τεχνική αυτή με τις γραμμική διακριτική ανάλυση, λογιστική παλινδρόμηση, k-πλησιέστερων γειτόνων, δέντρα ταξινόμησης και νευρωνικά δίκτυα. Το πρώτο σύνολο δεδομένων είχε ληφθεί από μεγάλη ελληνική τράπεζα, ενώ το δεύτερο ήταν ένα αυστραλιανό σύνολο δεδομένων έγκρισης πιστώσεων από την UCI Machine Learning Repository. Η προβλεπτική ικανότητα των scorecards μετριέται από το συνολικό ποσοστό των σωστά ταξινομημένων περιπτώσεων, τον συντελεστή Gini και το κακό ποσοστό μεταξύ των αποδεκτών. Σε κάθε ένα από τα σύνολα δεδομένων, ο NBR βρέθηκε να έχει χαμηλότερη προβλεπτική ικανότητα από ό,τι μερικές από τις λοιπές μεθόδους υπό το πλαίσιο όλων των μέτρων που χρησιμοποιήθηκαν. Στην συνέχεια εξετάζονται οι λόγοι που μπορεί να επηρέασαν αρνητικά την προβλεπτική ικανότητα των NBR σε σχέση με εκείνη των εναλλακτικών μεθόδων στο πλαίσιο του credit scoring.

Από την άλλη μεριά του Ατλαντικού, οι **Leonardo et al. (2016) [64]** χρησιμοποιούν Bayesian ταξινομητές έχοντας ως στόχο να δημιουργήσουν ένα μοντέλο για την μέτρηση του κινδύνου των εταιρειών που διαπραγματεύονται με την κυβέρνηση της Βραζιλίας, χρησιμοποιώντας δείκτες οι οποίοι αναφέρονται στον επιχειρηματικό κίνδυνο, την ιστορία κυρώσεων και ευρημάτων, το προφίλ των εταιριών και τους πολιτικούς δεσμούς. Απώτερος στόχος τους η χρήση αυτού του μοντέλου στην βέλπστη επιλογή των συμβάσεων, οι οποίες θα ελέγχονται από την κεντρική μονάδα του συστήματος εσωτερικού ελέγχου της κυβέρνησης της Βραζιλίας, υπεύθυνη για τον έλεγχο περισσότερων από 30.000 δημοσίων συμβάσεων ανά έτος.

## 2.4 K-nearest neighbour (KNN)

Η μέθοδος του K-κοντινότερου γείτονα είναι μία μη παραμετρική μέθοδος ταξινόμησης που βασίζεται στην εκμάθηση από την ομοιότητα. Θεωρείται μακράν ο πιο απλός αλγόριθμος

μεταξύ των αλγορίθμων εκμάθησης, και χρησιμοποιείται και για ταξινόμηση, και για παλινδρόμηση.

Κατά την διαδικασία, συλλέγεται ένα δοκιμαστικό σύνολο δεδομένων, για το οποίο εισάγεται μια συνάρτηση απόστασης μεταξύ της επεξηγηματικής μεταβλητής και των παρατηρήσεων. Για κάθε νέα παρατήρηση, η μέθοδος εξερευνά το μοτίβο στο χώρο των  $K$ - κοντινότερων γειτόνων ώστε να εντοπίσει τους πλησιέστερους γείτονες. Η νέα παρατήρηση τοποθετείται στην κατηγορία που περιέχονται οι περισσότεροι KNN (**Karamizadeh et al. (2013) [51]**).

---

```

Είσοδος:  $T$  // Σύνολο δεδομένων εκπαίδευσης
           $K$  // Αριθμός κοντινότερων γειτόνων
           $t$  // πλειάδα προς κατηγοριοποίηση
Εξόδος:  $c$  // Κλάση όπου θα κατηγοριοποιηθεί η  $t$ 
Αλγόριθμος  $_K\_Κοντινότερων\_Γειτόνων$ 
 $N = \emptyset$ 
Για κάθε  $d \in T$  επανέλαβε
  Αν  $|N| \leq K$  τότε
     $N = N \cup \{d\}$ ;
  Αλλιώς
    Αν  $\exists u \in N$  τέτοιο ώστε  $\text{dist}(t, u) \leq \text{dist}(t, d)$ , τότε
       $N = N - \{u\}$ ;
       $N = N \cup \{d\}$ ;
    Τέλος  $_αν$ 
Τέλος  $_επανάληψης$ 
 $c = κλάση\ όπου\ τα\ περισσότερα\ u \in N$  κατηγοριοποιούνται
Τέλος αλγορίθμου

```

---

Εικόνα 4: Παράδειγμα Αλγορίθμου εφαρμογής της μεθόδου k-NN.

Τα μειονεκτήματα της μεθόδου είναι το γεγονός ότι η KNN δεν παράγει μια απλή φόρμουλα ταξινόμησης και η προγνωστική της ακρίβεια επηρεάζεται σε μεγάλο βαθμό από το μέτρο της απόστασης και το πλήθος  $k$  των γειτόνων.

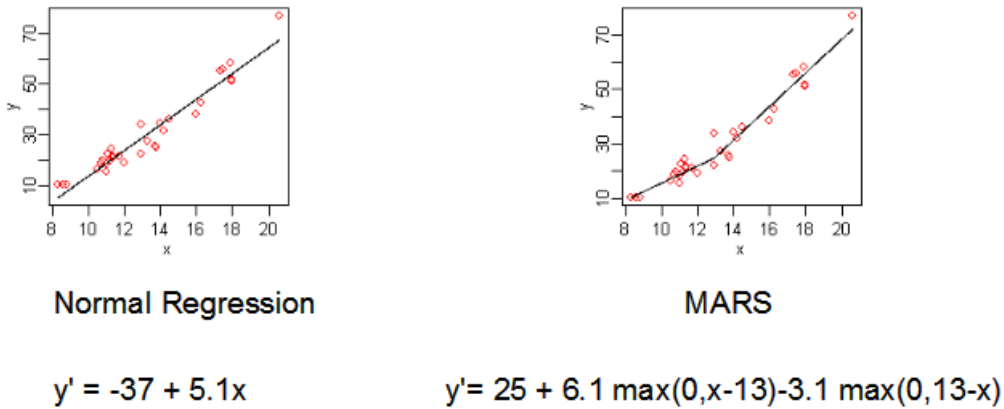
Ωστόσο, στην KNN δεν απαιτείται η καθιέρωση προγνωστικού μοντέλου πριν την ταξινόμηση, ενώ θεωρείται ευνοϊκή, λόγω του ότι (α) επιτρέπει τη μοντελοποίηση των ανπικανοτήτων στη συνάρτηση κινδύνου πάνω από τον χαρακτηριστικό χώρο, (β) είναι ανώτερη από άλλες μη παραμετρικές τεχνικές, όπως είναι οι μέθοδοι πυρήνα (kernel), ειδικά σε περιπτώσεις όπου τα δεδομένα είναι πολυδιάστατα, και (γ) είναι αρκετά έξυπνη και εύκολα κατανοητή από τους διαχειριστές της, οι οποίοι είναι οι υπεύθυνοι για την εφαρμογή της. [5] (**Brown & Mues, 2012; H. Chen & Chen, 2010; Lahsasna, Aïnon, & Teh, 2010**). [19, 20, 61]

## 2.5 Multivariate Adaptive Regression Splines (MARS)

Η Multivariate Adaptive Regression Splines (MARSplines) είναι μια μέθοδος μη παραμετρικής παλινδρόμησης, η οποία προτάθηκε από τον **Friedman (1991) [36]**, με σκοπό την επίλυση προβλημάτων της παλινδρόμησης.

Ένα από τα θέματα που αντιμετωπίζει αυτή η μέθοδος, είναι το γεγονός ότι δεν κάνει υποθέσεις για την σχέση που συνδέει την εξαρτημένη με τις ανεξάρτητες μεταβλητές, σε αντίθεση για παράδειγμα με την γραμμική παλινδρόμηση, που υποθέτει  $a$ -priori γραμμικότητα. Έτσι, η όποια σχέση υπάρχει μεταξύ εξαρτημένης και ανεξάρτητων κατασκευάζεται μέσω των δεδομένων.

Γενικά μπορούμε να πούμε ότι η μέθοδος αυτή, διαχωρίζει τα δεδομένα σε περιοχές, όπου στην κάθε μία εφαρμόζει μια εξίσωση παλινδρόμησης (βλ. εικόνα 4). Το γεγονός αυτό καθιστά την MARSplines ιδιαίτερα κατάλληλη σε περιπτώσεις με υψηλότερες διαστάσεις εισόδου (δηλαδή, με περισσότερες από 2 μεταβλητές), δηλ. σε περιπτώσεις όπου τα πολυδιάστατα δεδομένα, σε άλλες τεχνικές, δημιουργούν προβλήματα **Hastie, Tibshirani, και Friedman (2001) [95]**.

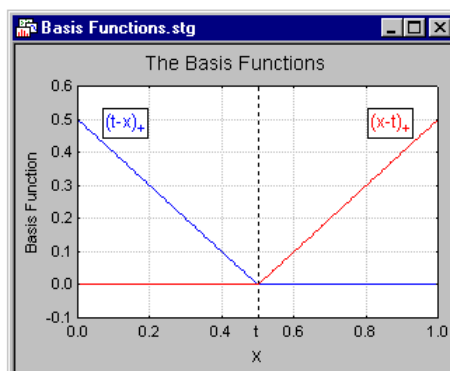


Εικόνα 5: Παράδειγμα εφαρμογής μοντέλου MARS σε σχέση με την απλή γραμμική παλινδρόμηση.

Ο αλγόριθμος MARSplines κατασκευάζει μοντέλα μέσω δύο κολοβών, δύο – όψεων συναρτήσεων βάσης (βλ. εικόνα 5) των ανεξάρτητων μεταβλητών  $x$ , της μορφής (**Hastie, et al., 2001 [95]**):

$$(x-t)_+ = \begin{cases} x-t & x > t \\ 0 & \text{αλλιώς} \end{cases} \quad \text{και} \quad (t-x)_+ = \begin{cases} t-x & x < t \\ 0 & \text{αλλιώς} \end{cases}$$

Η παράμετρος  $t$  είναι ο κόμβος των συναρτήσεων βάσης που καθορίζει τα "κομμάτια" της γραμμικής παλινδρόμησης κατά τμήματα. Αυτοί οι κόμβοι (παράμετροι) καθορίζονται επίσης από τα δεδομένα. Κόμβοι νοούνται τα σημεία όπου συνδέονται δύο πολυωνυμικές συναρτήσεις. Το σύμβολο "+" δίπλα με τους όρους  $(x-t)$  και  $(t-x)$  υποδηλώνει, ότι λαμβάνονται μόνο τα θετικά αποτελέσματα των αντίστοιχων εξισώσεων, διαφορετικά οι αντίστοιχες συναρτήσεις ισούται με μηδέν.



Εικόνα 6 : Συναρτήσεις Βάσεις - Basis Functions.

Οι βασικές συναρτήσεις συνδυάζονται από κοινού με τις παραμέτρους του μοντέλου (οι οποίες εκτιμώνται μέσω των ελάχιστων τετραγώνων) και παράγουν τις προβλέψεις. Η μορφή του γενικού μοντέλου μιας εξαρτημένης μεταβλητής  $y$  και  $M$  όρων, δίνεται από την σχέση :

$$y = f(x) = \beta_0 + \sum_{m=1}^M \beta_m H_{km}(x_{v(km)})$$

όπου το άθροισμα είναι μεγαλύτερο από τους  $M$  όρους του μοντέλου. Το μοντέλο συνεπώς προκύπτει από την συνάρτηση των ανεξάρτητων μεταβλητών  $x$  (και των αλληλεπιδράσεων τους), την διατομή ( $\beta_0$ ) και τα σταθμισμένα (μέσω των  $\beta_m$ ) αθροίσματα μιας ή περισσότερων βασικών συναρτήσεων  $h_m(X)$ . Τα  $\beta_m$  λειτουργούν ως βάρη που ανιππροσωπεύουν την σημαντικότητα κάθε μεταβλητής. Έτσι, επιλέγονται αυτόματα οι ζωτικής σημασίας μεταβλητές. Η συνάρτηση  $H$  ορίζεται ως :

$$H_{km}(x_{v(km)}) = \prod_{k=1}^K h_{km}$$

Όπου  $x_{v(km)}$  είναι η πρόβλεψη της  $k$ -οστής της  $m$ -οστής product. For order of interactions  $K=1$  the model is additive, and for  $K=2$  the model pairwise interactive. Ο αλγόριθμος αναζητά τις θέσεις των κόμβων  $t$  πάνω σε όλο τον χώρο των ανεξάρτητων μεταβλητών και των αλληλεπιδράσεων τους. Κατά την αναζήτηση προστίθεται στο μοντέλο ένα ολοένα και μεγαλύτερο πλήθος βασικών συναρτήσεων προκειμένου να μεγιστοποιηθεί το κριτήριο καλής προσαρμογής για το συνολικό μοντέλο. Έτσι, ο αλγόριθμος αυτόματα προσδιορίζει τις πιο σημαντικές ανεξάρτητες μεταβλητές, καθώς και τις πλέον σημαντικές μεταξύ τους αλληλεπιδράσεις.

Θεωρητικά οι ανεξάρτητες μεταβλητές του μοντέλου μπορούν να είναι είτε κατηγορικές, είτε συνεχείς μεταβλητές, ωστόσο και οι κατηγορικές αντιμετωπίζονται ως συνεχείς, **Friedman (1991) [36]**.

Η μέθοδος χρησιμοποιείται σε προβλήματα ταξινόμησης, διότι είναι σε θέση να χειριστεί εξαρτημένες μεταβλητές με πολλαπλές κατηγορίες. Αρχικά κωδικοποιούμε την κάθε κατηγορία μέσω μια δείκτριας συνάρτησης ( $1=$  η παρατήρηση ανήκει στην κατηγορία  $k$ ,  $0=$  η παρατήρηση δεν ανήκει στην κατηγορία  $k$ ), στην συνέχεια εφαρμόζουμε τον αλγόριθμο MARSplines και υπολογίζουμε την προβλεπόμενη (συνεχή) βαθμολογία. Η παρατήρηση εισάγεται στην κατηγορία που προβλέπεται υψηλότερη βαθμολογία (**Hastie, Tibshirani και Freedman, 2001 [95]**).

Αυτός ο τρόπος εφαρμογής αποφέρει μια ευρετική ταξινόμηση, που ναι μεν λειτουργεί καλά στην πράξη, όμως δεν βασίζεται σε ένα στατιστικό μοντέλο για την εξαγωγή των πιθανοτήτων κατάταξης. Για την καταπολέμηση αυτού του προβλήματος, μετά την εφαρμογή της forward stepwise επιλογής των βασικών συναρτήσεων, γίνεται μια backward διαδικασία κατά την οποία το μοντέλο «κλαδεύεται» (κάπι αντίστοιχο με το κλάδεμα των Δέντρων Αποφάσεων) αφαιρώντας τις βασικές συναρτήσεις που σχετίζονται με την μικρότερη αύξηση (ελαχίστων τετραγώνων) του κριτηρίου καλής προσαρμογής. Στην συνέχεια υπολογίζεται μια συνάρτηση σφάλματος (ελαχίστων τετραγώνων), αντίστροφη από αυτήν της καλής προσαρμογής, η οποία ονομάζεται Generalized Cross Validation error, με σκοπό να μετρήσει όχι μόνο το υπολειπόμενο σφάλμα, αλλά και την πολυπλοκότητα του μοντέλου. Η συνάρτηση δίνεται από

$$\text{την σχέση : } GCV = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{\left(1 - \frac{C}{N}\right)^2}$$

Με  $C=1+cd$ , όπου,  $N$  το πλήθος των περιπτώσεων σε ένα σετ δεδομένων, και  $d$  είναι οι βαθμοί ελευθερίας, οι οποίοι ισούνται με το πλήθος των ανεξάρτητων μεταβλητών. Η ποσότητα  $C$  είναι η ποινή της προσθήκης μιας συνάρτησης βάσης. Πειράματα έχουν δείξει ότι η βέλτιστη τιμή για την  $C$  μπορεί να βρεθεί μεταξύ  $2 < d < 3$  (**Hastie et al., 2001) [95]**.

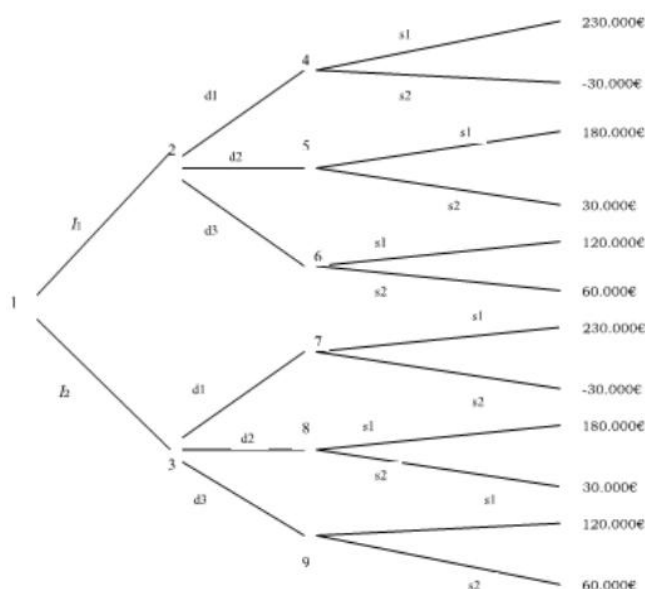
Σε γενικές γραμμές, τα μη παραμετρικά μοντέλα είναι προσαρμοστικά (επιτυγχάνουν μηδενικό σφάλμα στα δεδομένα εκπαίδευσης) και όπως συμβαίνει σε αυτού του είδους τις τεχνικές, εμφανίζεται το φαινόμενο της υπερπροσαρμογής (overfitting). Έτσι, έχουν την τάση να δίνουν κακά αποτελέσματα όταν χρησιμοποιούνται σε νέες παρατηρήσεις (δηλαδή, δεν γενικεύονται στην πρόβλεψη της «νέας» παρατήρησης). Ως τεχνική πάντως είναι πολύ πιο απλή από άλλες μεθόδους όπως τα Νευρωνικά Δίκτυα.

Οι **Lee & Chen (2005) [63]** υπογραμμίζουν αυτά τα συγκεκριμένα πλεονεκτήματα, δηλ. ότι (α) δεν απαιτείται η τοποθέτηση κάποιας αρχικής προϋπόθεσης, οπότε είναι σε θέση να μοντελοποιήσει πολύπλοκες μη γραμμικές σχέσεις μεταξύ των μεταβλητών, (β) επιλέγει αυτόματα τις μεταβλητές που έχουν ζωική σημασία, δηλ. συλλαμβάνει τη σχετική σημασία των ανεξάρτητων μεταβλητών για τις εξαρτημένες, λαμβάνονται υπόψη πολλές πιθανόν σημαντικές ανεξάρτητες μεταβλητές, (γ) δεν υπάρχει μακρά διαδικασία κατάρτισης, και (δ) είναι εύκολο να ερμηνευθεί (**Chuang & Lin, 2009; Li-hua, Jia-shan, & Feng, 2006) [23]**)

## 2.6 Δέντρα Απόφασης (Decision Trees)

Μια άλλη μέθοδος που χρησιμοποιείται για την επίλυση προβλημάτων κατηγοριοποίησης είναι τα **Δέντρα Απόφασης (Decision Trees)**.

Τα δέντρα απόφασης περιγράφονται ως ένα δέντρο που μοιάζει με γραφική παράσταση των αποφάσεων και των πιθανών συνεπειών. Στο υψηλότερο σημείο του δέντρου (κορυφή) είναι ο ριζικός κόμβος από τον οποίο λαμβάνεται (πιθανώς) μια απόφαση. Σε κάθε εσωτερικό κόμβο διεξάγεται μια δοκιμασία / ένα χαρακτηριστικό / ένα πείραμα. Κάθε κλαδί (που απορρέει από τον κόμβο) καταλήγει στο αποτέλεσμα της εκάστοτε δοκιμασίας. Τα δέντρα ταξινόμησης χρησιμοποιούνται όταν η μεταβλητή απόκρισης είναι ποσοτική διακριτή, ή ποιοτική, και με σκοπό να δημιουργηθούν κανόνες εύκολοι στην κατανόηση και την επεξήγηση τους.



Εικόνα 7: Σχηματική απεικόνιση ενός Δέντρου Αποφάσεων.

Στα πλεονέκτημα αυτής της μεθόδου συγκαταλέγονται ότι είναι απλή, κατανοητή, εξηγήσιμη, αποδοτική και μπορεί να χρησιμοποιηθεί για μεγάλες Βάσεις Δεδομένων (καθώς το μέγεθος της βάσης είναι ανεξάρτητο από το μέγεθος του δέντρου). Μειονέκτημα της μεθόδου είναι ότι δεν είναι σε θέση να χειριστεί συνεχή δεδομένα, δεδομένα με ελλείπουσες τιμές και στην

γενίκευση του περιορίζεται, αφού μια σχεδιασμένη δομή για ένα πλαίσιο δεν μπορεί να γενικευτεί σε άλλα πλαίσια (φαινόμενο υπερπροσαρμογής)(**Abbas Keramati et al. (2011) [1]**).

Τα Δέντρα απόφασης έχουν χρησιμοποιηθεί ευρέως από τις τράπεζες για την μέτρηση του πιστωτικού κινδύνου, λόγω της απλότητας και της αμεσότητας τους. Η υπάρχουσα βάση δεδομένων της τράπεζας (δηλ. οι συμπεριφορές των υφιστάμενων πελατών) χρησιμοποιείται στην κατασκευή του δέντρου, έτσι ώστε στην συνέχεια να γίνει ο χαρακτηρισμός και η ταξινόμηση κάθε νέου πελάτη που θα αιτηθεί ένα δάνειο. **Ketaki Chopde et al., (2012) [55]**.

Μια εφαρμογή της μεθόδου αυτή συναντάται στην εργασία των **Qiwei Gan et al., (2009) [81]**, όπου ερευνούν τη διαχείριση των κινδύνων για τα στεγαστικά στην Κίνα. Οι ερευνητές εξετάζοντας ένα μοντέλο συνάφειας της αγοράς ενυπόθηκων δανείων της Κίνας, αποκαλύπτουν ότι με σχετικά σφιχτούς δημοσιονομικούς κανονισμούς από την κινεζική κεντρική τράπεζα, οι εμπορικές τράπεζες εξαρτώνται αποκλειστικά από το πιστωτικό σύστημα για τον έλεγχο πιθανών υποψηφίων με αδυναμία αποπληρωμής του δανείου. Ως εκ τούτου, είναι επείγον για τις εμπορικές τράπεζες να οικοδομήσουν ένα αποτελεσματικό σύστημα ελέγχου για τη μείωση του κινδύνου που εισάγει επισφαλή δάνεια. Η μελέτη βασίζεται σε ένα πραγματικό σύνολο δεδομένων 641.988 παρατηρήσεων μιας κινεζικής εμπορικής τράπεζας. Τα αποτελέσματα δείχνουν ότι το μοντέλο κατάταξης είναι αποτελεσματικό για credit scoring, και φυσικά, ότι βελτίωση της ποιότητας των δεδομένων αποδίδει πιο ακριβή προεπιλεγμένη διαχείριση του κινδύνου.

---

```

Είσοδος:  $D$  // Δεδομένα εκπαίδευσης
Έξοδος:  $T$  // Δένδρο απόφασης
Αλγόριθμος DTBuild
 $T \leftarrow \emptyset$ ;
Καθόρισε το καλύτερο κριτήριο διάσπασης;
 $T \leftarrow$  Δημιούργησε τον κόμβο ρίζα και ονόμασέ τον με το όνομα του
χαρακτηριστικού διάσπασης;
 $T \leftarrow$  Πρόσθεσε τόσα τόξα από τον κόμβο ρίζα όσα και τα κατηγορήματα
διάσπασης (τιμές χαρακτηριστικού) και ονόμασε τα
Για κάθε τόξο επανέλαβε
 $D \leftarrow$  Δεδομένα εκπαίδευσης που παραμένουν εφαρμόζοντας το
κατηγορήμα διάσπασης στο  $D$ 
Αν ικανοποιείται το κριτήριο τερματισμού για αυτό το μονοπάτι τότε
 $T' \leftarrow$  Δημιούργησε έναν κόμβο φύλλο και ονόμασε τον με το
όνομα της κλάσης
Αλλιώς
 $T' \leftarrow$  DTBuild( $D$ );
Τέλος_αν
 $T =$  πρόσθεσε τόξο στο  $T$ 
Τέλος_επανάληψης
Τέλος_αλγορίθμου

```

---

Εικόνα 8: Παράδειγμα αλγορίθμου κατασκευής δέντρου απόφασης.

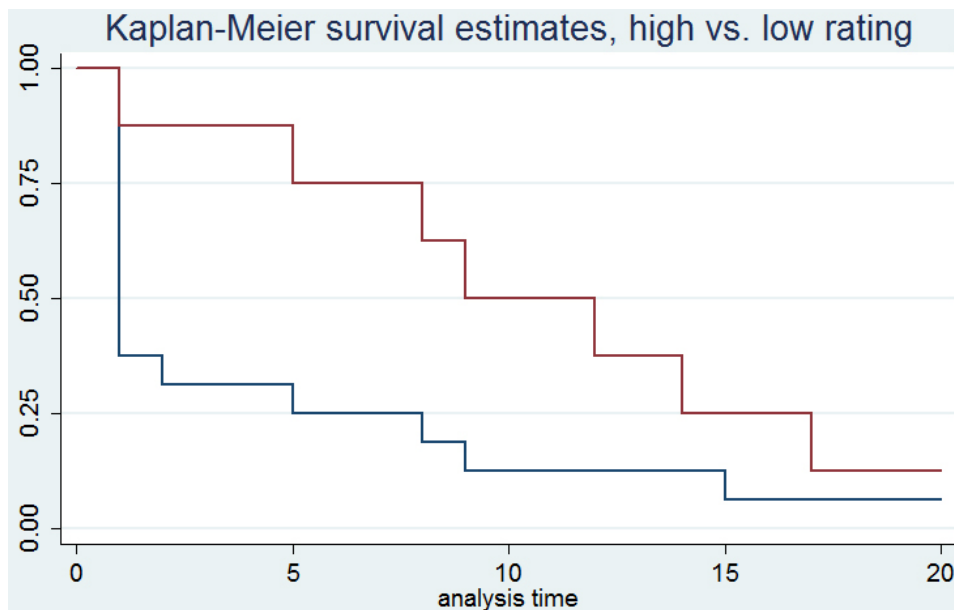
## 2.7 Survival Analysis (Ανάλυση Επιβίωσης)

Η Ανάλυση Επιβίωσης είναι ένα νέο είδος μοντέλου credit scoring. Ενώ, οι συμβατικές μέθοδοι μπορούν να διακρίνουν τους δανειολήπτες σε «καλούς» και «κακούς» κατά τη στιγμή της

αίτησης του δανείου, μέσω της ανάλυσης επιβίωσης δύναται να εκτιμηθεί η **κερδοφορία** από τους πελάτες κατά τη διάρκεια της ζωής τους (ή κατά την διάρκεια μέχρι να φτάσουν σε σημείο αθέτησης πληρωμής) και να δημιουργήσει ένα σύστημα ταξινόμησης του κέρδους (**Baesens, et al., 2005**) [11]. Αυτό διότι, η ανάλυση επιβίωσης μπορεί να προβλέψει το χρόνο μέχρι να συμβεί ένα γεγονός, αντί της πρόβλεψης της πιθανότητας εμφάνισης του γεγονότος που κάνουν οι συμβατικές μέθοδοι. [5]

Η μέθοδος ονομάζεται έτσι επειδή η συγκεκριμένη μεθοδολογία χρησιμοποιήθηκε κυρίως στον χώρο της Βιοστατιστικής για να εξετάσει τον χρόνο μέχρις ότου ένας ασθενής πεθάνει (άρα τον χρόνο που επιβίωσε). Έτσι, αυτό που μετράμε είναι ο χρόνος μέχρις ότου το εκάστοτε άτομο της ομάδας μας, να εμφανίσει το ζητούμενο γεγονός (δηλ. την αθέτηση πληρωμής), είτε τον χρόνο που το άτομο παρέμεινε στην επίβλεψη μας, δηλ.  $T > t$ . Οι τελευταίες περιπτώσεις καλούνται λογοκριμένες (censored), και όταν  $T > t$ , δηλ. το μόνο που γνωρίζουμε είναι ότι ο χρόνος του υπό μελέτη ατόμου υπερβαίνει κάποια τιμή, τότε έχουμε δεξιά λογοκρισία (right censoring).

Ορίζουμε, ως χρόνο λογοκρισίας  $u$  το σημείο στο οποίο αποφασίζουμε να σταματήσουμε τη μελέτη και αναλύουμε τα ευρήματα,  $x_1, x_2, \dots, x_n$  τους χρόνους επιβίωσης μιας κοορτής ανθρώπων, οι οποίοι παρατηρούνται αν και μόνο αν  $x_i \leq u$ . Κοορτή καλείται μια ομάδα ανθρώπων με καλά ορισμένα χαρακτηριστικά. Αν  $x_i > u$ , δηλ. ο χρόνος επιβίωσης είναι μεγαλύτερος από τον χρόνο λογοκρισίας, τότε καταγράφεται μόνο η μέτρηση  $u$ .



Εικόνα 9: Παράδειγμα συνάρτησης «επιβίωσης» με την μη παραμετρική μέθοδο Kaplan-Meier.

Τα δεδομένα θα έχουν την μορφή  $(T_i, \delta_i)$ , όπου  $T_i = \min(x_i, u)$ , είναι η μεταβλητή χρόνος, και

$\delta_i = \begin{cases} 1, & \alpha \nu x_i \leq u \\ 0, & \alpha \nu x_i > u \end{cases}$  μια δίτιμη μεταβλητή όπου  $\delta_i = 1$  όταν εμφανίζεται το γεγονός, και  $\delta_i = 0$

όταν το γεγονός δεν έχει εμφανιστεί. Υποθέτουμε ότι η  $T_i$  είναι μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας  $f(t)$  και αθροιστική συνάρτηση κατανομής  $F(t) = \Pr(T < t)$ . Ενδιαφερόμαστε να υπολογίσουμε την ποσότητα  $\Pr(T > t)$ , η οποία καλείται συνάρτηση επιβίωσης, προκύπτει από την σχέση :

$$S(t) = \Pr(T > t) = 1 - F(t) = \int_t^{\infty} f(u) du$$

και εκφράζει την πιθανότητα ένα άτομο της κοορτής να επιβιώσει πέραν του χρόνου  $t$ .

Οι **Stepanova & Thomas, (2002) [90]** προτάσσουν ότι αυτή η τεχνική είναι σε θέση να μοντελοποιήσει τον χρόνο για την μη πληρωμή (του δανείου). Ακόμα, η τεχνική θεωρείται πλήρης, δεδομένου ότι λαμβάνει υπόψη τη συνολική στάση απέναντι στο credit scoring, όπως επισημάνουν οι **Δούμπος & Ζοπουνίδης, (2014) [28]**.

Υπάρχουν επίσης και άλλες μελέτες που χρησιμοποιούν την ανάλυση επιβίωσης. Οι **Beran & Djaidja, (2007) [16]** κάνουν χρήση στατιστικής μοντελοποίησης για την μέτρηση του πιστωτικού κινδύνου σε πελάτες λιανικής. Λόγω έλλειψης λεπτομερώς ενημερωμένων στοιχείων αναφορικά με πληροφορίες του αντισυμβαλλόμενου, καταλήγουν στην άποψη ότι οι παραδοσιακές προσεγγίσεις, όπως το μοντέλο σταθερής αξίας του Merton, δεν ισχύουν. Αυτό παρακινεί την χρήση ενός στατιστικού μοντέλου που βασίζεται στην ανάλυση επιβίωσης το οποίο θα εξετάζει τον «χρόνο μέχρι την αθέτηση». Το μοντέλο ενσωματώνει τη στοχαστικότητα αυτής της μεταβλητής (του χρόνου αθέτησης) και, η κατανομή που προκύπτει (για τον χρόνο) χαρακτηρίζεται από συνέπεια και ασυμπτωτική κανονικότητα των εκπιητριών μέγιστης πιθανοφάνειας των παραμέτρων. Τα αποτελέσματα τους εφαρμόζονται σε ένα δείγμα 670 υποθηκών.

## 2.8 Fuzzy Rule-Based System (FRBS)

Πρόκειται για μια κατηγορία έμπειρων συστημάτων [**Expert Systems (ES)**]. Γενικά τα έμπειρα συστήματα διακρίνονται σε έμπειρα συστήματα βασισμένα στη γνώση (Knowledge Based ES), στα έμπειρα συστήματα βασισμένα σε κανόνες (Rule-Based ES) και σε έμπειρα συστήματα ασαφούς λογικής (Fuzzy Expert Systems).

Στις αρχές της δεκαετίας του '70 υπήρξε σημαντικό ενδιαφέρον για τη μελέτη της λήψης αποφάσεων από εμπειρογνώμονες (του εκάστοτε αντικείμενου), όπου δεν θα χρησιμοποιούσαν στατιστικά ή άλλα μαθηματικά εργαλεία, και την απεικόνιση αυτής της γνώσης σε ένα υπολογιστή. Με άλλα λόγια η ιδέα είναι να συλλάβουμε την ανθρώπινη εμπειρία και να την εισάγουμε σε έναν υπολογιστή που θα λαμβάνει τις αποφάσεις.

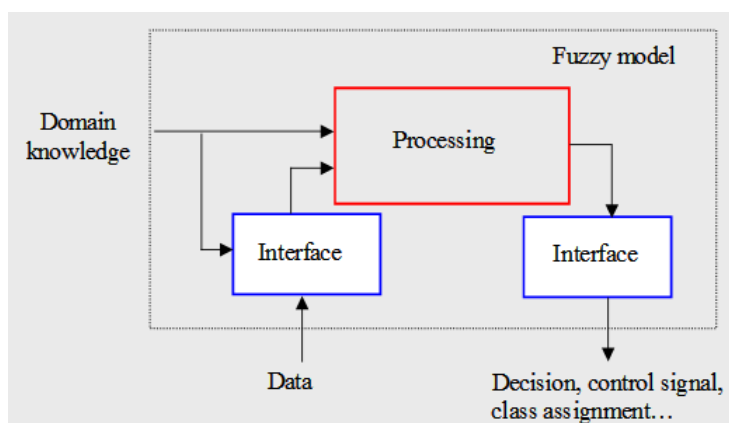
Η ιδέα βασίζεται στο γεγονός ότι η ανθρώπινη λογική (η ανθρώπινη αντίληψη) είναι σε θέση να χειριστεί με κατάλληλο τρόπο ασαφείς έννοιες, την αβεβαιότητα και την ασάφεια από την οποία περιβάλλεται ο κόσμος των πληροφοριών. Τα ασαφή βασιζόμενα σε κανόνες συστήματα (FRBSs) αναπτύχθηκαν από τον **Zadeh (1965) [105]**, ως επέκταση των κλασικών Rule-Based Systems, επειδή χειρίζονται κανόνες κατά προσέγγιση, αντί τους κλασικούς λογικούς κανόνες. Η σημασία των Fuzzy Logic απορρέει από το γεγονός ότι τα περισσότερα είδη της ανθρώπινης συλλογιστικής -και ιδιαίτερα η κοινή λογική- είναι κατά προσέγγιση στη φύση [**105**].

Τα βασικά συστατικά ενός έμπειρου συστήματος παρουσιάζονται στην εικόνα 8, όπου:

- Γνωστική Βάση (Knowledge Base/Domain) περιέχει όλες τις σχετικές πληροφορίες, τα δεδομένα, τους κανόνες τις περιπτώσεις και τις σχέσεις που θα χρησιμοποιηθούν από το έμπειρο σύστημα.
- Κανόνας (Rule) είναι ένας υπό συνθήκη κανόνας που συνδέει δεδομένες συνθήκες σε ενέργειες ή εκβάσεις. Οι κανόνες που περιέχονται σε ένα fuzzy expert system έχουν συνήθως την ακόλουθη μορφή: *If < input variables  $A_i$  > and < input variables  $B_i$  >*



*then* < **output variable X** >, όπου,  $A_i$  και  $B_i$  είναι οι μεταβλητές που εισάγονται (οι πληροφορίες), και X η μεταβλητή που παράγεται [91].



Εικόνα 10: Αρχιτεκτονική δομή ενός Ασαφούς Έμπειρου Συστήματος.

Τα πλεονεκτήματα των έμπειρων συστημάτων είναι η ικανότητα να συλλαμβάνουν και να διατηρούν αναπικατάστατη την ανθρώπινη εμπειρία, η ικανότητα να αναπτύσσουν συστήματα πιο συνεπή από τους ίδιους τους εμπειρογνώμονες, το ότι ελαχιστοποιούν την ανάγκη τοποθέτησης ειδικών εμπειρογνώμωνων σε διάφορες τοποθεσίες και η ταχύτατη παροχή λύσεων. Από αυτό όμως προκύπτει και το βασικό μειονέκτημα τους που είναι η δυσκολία της σύλληψης της ανθρώπινης εμπειρίας, δηλ. η απόκτηση της γνώσης. Ως αποτέλεσμα αυτού του προβλήματος, η έρευνα επικεντρώθηκε σε εργαλεία για την απόκτηση γνώσεων, για να βοηθήσει την αυτοματοποίηση της διαδικασίας σχεδιασμού, εντοπισμού σφαλμάτων, και τη διατήρηση των κανόνων που ορίζονται από τους εμπειρογνώμονες, καθώς επίσης και τη δυνατότητα πρόσβασης σε μεγάλες βάσεις δεδομένων.

Κατά την εφαρμογή της μεθόδου σε credit scoring, η συγκεκριμένη τεχνική βοηθά τους πιστωτές στο σχεδιασμό κανόνων που απορρέουν στην ακρίβεια του πιστωτικού αποτελέσματος με σχετική επεξήγηση, σε αντίθεση με τα περισσότερα μοντέλα βαθμολόγησης που επικεντρώνονται στην εκτίμηση ενός σκορ, χωρίς όμως μιας επεξήγησης για τον τρόπο που τα αποτελέσματα προέκυψαν. Το πλεονέκτημα αυτού του μοντέλου είναι ότι οι ασαφείς κανόνες είναι σε θέση να χειρίζονται τόσο ποσοτικούς, όσο και ποιοτικούς παράγοντες, οπότε αν υπάρχει μεγάλο πλήθος εισροών, τα αποτελέσματα βαθμολόγησης θα είναι λιγότερο ευαίσθητα σε μικρά σφάλματα μέτρησης (Hoffmann, Baesens, Mues, Van Gestel, & Vanthienen, 2007)[105].

Στην βιβλιογραφία διακρίνονται δύο διαφορετικοί τύποι FRBS ανάλογα με τη μορφή των ασαφών κανόνων, και τα είδη των εισροών και εκροών που χρησιμοποιούνται [105]. Έτσι έχουμε τα Mamdani Fuzzy Rule-Based Systems, που προτάθηκαν από τον Mamdani (1974) [74], και τα Takagi-Sugeno-Kang Fuzzy Rule-Based Systems που προτάθηκαν από τους Takagi, Sugeno, and Kang (1985) [91].

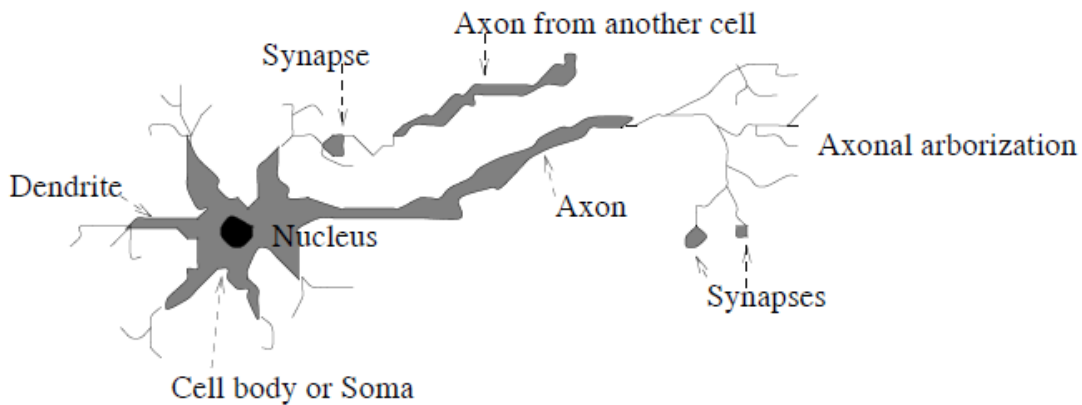
Υπάρχουν πολλές εργασίες με αντικείμενο την επιρροή των ασαφών κανόνων στην μέθοδο (fuzzy rule-based classification) εφαρμοσμένη σε credit scoring (Akkoc, 2012) [4] Σε πιο πρόσφατη στην εργασία, οι Sohn et al, 2016 [87] παρουσιάζουν την fuzzy logistic regression με σκοπό να αντιμετωπίσουν περιπτώσεις που η κλασική στατιστική παλινδρόμηση δεν είναι εφαρμόσιμη, όπως όταν το δείγμα είναι μικρό ή υπάρχει ασάφεια στην σχέση εξαρτημένης και ανεξάρτητων.

## 2.9 Νευρωνικά Δίκτυα (Neural Network)

Τεχνητά νευρωνικά δίκτυα [(ΤΝΔ) - Artificial neural networks (ANN)] είναι μη παραμετρική και μη-γραμμική στατιστική μοντελοποίηση η οποία βασίζεται στην λειτουργία του ανθρώπινου εγκεφάλου. Η δομή του εγκεφάλου είναι τέτοια ώστε να επιτρέπει την παράλληλη επεξεργασία δεδομένων και τη δυνατότητα συνεχούς μάθησης μέσω της αλληλεπίδρασης με το περιβάλλον. Τα δύο αυτά βασικά χαρακτηριστικά συμβάλλουν στην ικανότητα, αφενός, της εκτέλεσης δύσκολων καθηκόντων, όπως η ταχύτατη αναγνώριση μορφών, η ταξινόμηση κ.ά., και αφετέρου, στην αδιάκοπη εξέλιξη, μαθαίνοντας από το περιβάλλον κατά την αλληλεπίδρασή του με αυτό. Είναι ισχυρά εργαλεία όταν η σχέση μοντελοποίησης των δεδομένων είναι άγνωστη.

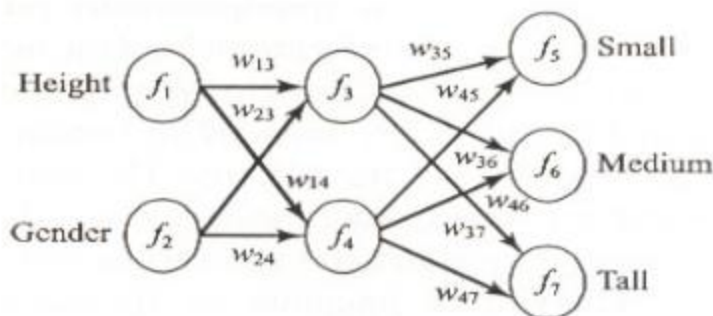
Η ιδέα γεννήθηκε στα τέλη του 1940 από τον ψυχολόγο **Donald Hebb**, αλλά άρχισε να αναπτύσσεται μερικά χρόνια αργότερα, όπου οι **Warren McCulloch and Walter Pitts (1943) [75]** δημιούργησαν ένα υπολογιστικό μοντέλο για νευρωνικά δίκτυα που βασίζονταν σε μαθηματικά και αλγόριθμους. Το μοντέλο αυτό άνοιξε τον δρόμο για την εφαρμογή των νευρωνικών δικτύων.

Στην πραγματικότητα τα νευρωνικά δίκτυα είναι συστήματα επεξεργασίας πληροφορίας που αποτελούνται από έναν γράφο και διάφορους αλγόριθμους που προσπελαίνουν αυτόν τον γράφο. Κάθε κόμβος του γράφου είναι σαν ανεξάρτητοι νευρώνες, ενώ τα τόξα (βλ. σχήμα 6) είναι σύνδεσμοι των νευρώνων.



Εικόνα 11: Παράδειγμα απεικόνισης Νευρώνα.

Κάθε ένας από τους κόμβους είναι στοιχείο επεξεργασίας που λειτουργεί ανεξάρτητα από τους άλλους (π.χ. οικονομικοί δείκτες) και χρησιμοποιεί μόνο τοπικά δεδομένα που καθοδηγούν την επεξεργασία.



Εικόνα 12: Παράδειγμα απεικόνισης Νευρωνικού Δικτύου.

Το επίπεδο εξόδου αποτελείται από έναν αριθμό κόμβων ο οποίος είναι ίσος πάντα με τις κατηγορίες ταξινόμησης. Τα ΤΝΔ είναι σε θέση να αναγνωρίσουν το πολύπλοκες δομές

(pattern) μεταξύ των μεταβλητών εισόδου και εξόδου, και στη συνέχεια, να προβλέψουν την ενδεχόμενη έκβαση από την είσοδο νέων ανεξάρτητων δεδομένων (**Stern, 1996**).

Στα πλεονεκτήματα της χρήσης της μεθόδου, τα νευρωνικά δίκτυα θεωρούνται πιο ισχυρά σε σχέση με τα δέντρα απόφασης, λόγω των βαρών και του ότι βελτιώνονται με την μάθηση, η οποία συνεχίζεται και μετά την εφαρμογή των δεδομένων εκμάθησης. Δίνουν μεγάλη ακρίβεια και δεν επηρεάζονται τόσο από την ύπαρξη θορύβου. Η μέθοδος είναι επίσης επιθυμητή λόγω της ικανότητας γενίκευσης των αποτελεσμάτων (ελλείψη ρητής περιγραφής του προβλήματος), του χειρισμού μεγάλο όγκου δεδομένων, την ύπαρξη λιγότερων στατιστικών υποθέσεων [5]. Από την άλλη όμως, η εφαρμογή τους είναι δύσκολη, και κυρίως δυσνόητη από τους χρήστες (και πάλι σε σχέση με την ευκολία που παρέχουν τα δέντρα απόφασης) αποτελώντας μαύρο κουτί. Βέβαια και αυτά (όπως και τα δέντρα απόφασης) μπορούν να εμφανίσουν το φαινόμενο της υπερπροσαρμογής. Επίσης, απαιτούν μια χρονοβόρα σειρά μαθημάτων κατάρτισης στην ανάπτυξη του βέλτιστου μοντέλου (**Karamizadeh, Abdullah, Zamani, & Kherikah, 2015**)[53]

Σύμφωνα με τους **Θωμαδάκη και Ξανθάκη (2006) [93]** τα κύρια μειονεκτήματα της προσέγγισης αυτής είναι το κόστος εφαρμογής και η συντήρησή τους, η ταχύτητα εξαγωγής αποτελεσμάτων (εξαιτίας του πιθανόν μεγάλου αριθμού συσχεπίσεων), αλλά και η ισχύς των αποτελεσμάτων δεδομένου ότι πολλές φορές τα βήματα που χρησιμοποιούνται δε ερμηνεύονται από την οικονομική θεωρία.

Την τελευταία δεκαετία όλο και περισσότεροι ερευνητές ασχολούνται με τα νευρωνικά δίκτυα :

Ο **West (2000) [100]** στην εργασία του διερευνά την ακρίβεια πέντε μοντέλων νευρωνικών δικτύων: πολυστρωματικό νευρωνικό δίκτυο (multilayer perceptron), μίγμα εμπειρογνομόνων (mixture-of-experts), ακτινική συνάρτηση βάσης (radial basis function), φορέα μάθησης κβαντοποίησης (learning vector quantization), και ασαφή προσαρμοστική συντονισμού (fuzzy adaptive resonance). Τα μοντέλα νευρωνικών δικτύων πιστωτικής βαθμολόγησης ελέγχονται χρησιμοποιώντας 10 αρχεία, τα οποία στη συνέχεια διασταυρώνονται με δύο αρχεία πραγματικών δεδομένων. Τα αποτελέσματα συγκρίνονται με παραδοσιακές μεθόδους ανάλυσης.

Οι **Malhotra et al. (2003) [73]**, αναπτύσσουν μια σειρά από μοντέλα πιστοληπτικής βαθμολόγησης που κατατάσσουν με ακρίβεια τις αιτήσεις των καταναλωτικών δανείων, με σκοπό να βοηθήσουν τις παραδοσιακές μεθόδους απόφασης/κρίσης. Η μελέτη συγκρίνει την απόδοση της διαχωριστικής ανάλυσης και των νευρωνικών δικτύων στην πιθανή λήψη δανείου. Αν και δεν υπάρχει σημαντική βελτίωση από την χρήση των νευρωνικών δικτύων, τα μοντέλα αυτά δίνουν σταθερά καλύτερες επιδόσεις.

Ο **Hsieh (2004) [43]** ασχολήθηκε με τα πολυδιάστατα τραπεζικά δεδομένα, από τα στοιχεία των μηνιαίων λογαριασμών και τα καθημερινά αρχεία συναλλαγών. Στην εργασία του προτείνει ένα ολοκληρωμένο μοντέλο εξόρυξης δεδομένων (data mining) και συμπεριφορικής βαθμολόγησης (behavioral scoring) για να διαχειριστεί τους υπάρχοντες πελάτες πιστωτικών καρτών. Έτσι χρησιμοποιεί ένα αυτό-οργανωμένο χάρτη νευρωνικού δικτύου για να προσδιορίσει τους πελάτες σε τρεις μεγάλες επικερδής ομάδες, βασιζόμενο στην συμπεριφορά αποπληρωμής, στην συχνότητα, και στην νομισματική συμπεριφορά (monetary behavioral scoring predictors). Στην συνέχεια τα χαρακτηριστικά των ομάδων αυτών χρησιμοποιήθηκαν ως εκ των προτέρων γνώση, για να δείξει ότι ο χαρακτηρισμός των πελατών από ένα μοντέλο συμπεριφορικής βαθμολόγησης είναι χρήσιμο και διευκολύνει την ανάπτυξη στρατηγικού μάρκετινγκ.

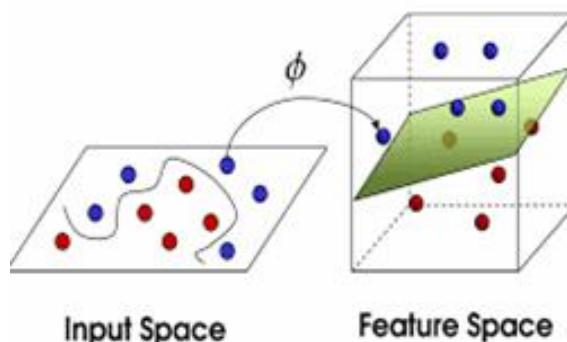
[40] Οι **Yu και Lai (2008) [104]** προτείνουν την εφαρμογή ενός πολυδιάστατου νευρωνικού μοντέλου μάθησης 6 σταδίων, με σκοπό την αξιολόγηση του πιστωτικού κινδύνου. Στο πρώτο στάδιο, χρησιμοποιείται μια δειγματοληπτική προσέγγιση ώστε να δημιουργηθούν διαφορετικά σετ δεδομένων εκπαίδευσης ειδικά για την περίπτωση έλλειψης δεδομένων. Στο δεύτερο στάδιο, από κάθε δείγμα εκπαίδευσης δημιουργούνται τα διάφορα μοντέλα νευρωνικών δικτύων. Αυτά στην συνέχεια (τρίτο στάδιο) εκπαιδεύονται και σύμφωνα με το σκορ ταξινόμησης και την τιμή αξιοπιστίας τους επιλέγεται το κατάλληλο. Στο τέταρτο στάδιο, χρησιμοποιείται ένας αλγόριθμος μεγιστοποίησης αποσυσχέτισης για την επιλογή του κατάλληλου συνόλου δεδομένων. Στο πέμπτο στάδιο, οι τιμές αξιοπιστίας του επιλεγμένου νευρωνικού δικτύου κλιμακώνονται σε ένα μοναδιαίο διάστημα μέσω λογαριθμικού μετασχηματισμού. Στο τελικό στάδιο, το επιλεγμένο νευρωνικό δίκτυο και το σύνολο δεδομένων συνήκονται για να ληφθεί το τελικό αποτέλεσμα της ταξινόμησης.

Ο **Khashman (2010) [56]** εφαρμόζει ένα σύστημα αξιολόγησης πιστωτικού κινδύνου με την χρήση εποπτευόμενων μοντέλων νευρωνικών δικτύων με βάση τον αλγόριθμο της backward μάθησης. Εκπαιδεύει (με πραγματικά δεδομένα 1000 καρτών Γερμανικής βάσης δεδομένων με 24 χαρακτηριστικά – μεταβλητές) και εφαρμόζει τρία νευρωνικά δίκτυα για να αποφασίσει εάν θα δεχτεί ή θα απορρίψει μια αίτηση πίστωσης. Τα πειραματικά αποτελέσματα προτείνουν το κατάλληλο νευρωνικό μοντέλο δικτύου, και υπό το κατάλληλο σύστημα εκμάθησης μπορεί να προταθεί το σύστημα αξιολόγησης πιστωτικού κινδύνου που παρέχει την βέλπστη απόδοση, το οποίο θα χρησιμοποιηθεί αποτελεσματικά και γρήγορα στην αυτόματη επεξεργασία των πιστωτικών αιτήσεων.

Οι **Abeer Hamdy, Walid B. Hussein (2016) [3]**, στην εργασία τους παρέχουν ένα υψηλής ακρίβειας μοντέλο πιστοληπτικού κινδύνου για χρήση σε μικρές και μεγάλες βάσεις δεδομένων χρησιμοποιώντας μια ανάλυση κύριων συνιστωσών (PCA) βασισμένη στην ανάλυση της σημασίας των χαρακτηριστικών που χρησιμοποιούνται συνήθως στα μοντέλα credit scoring. Στο προτεινόμενο μοντέλο credit scoring εφαρμόζεται, αρχικά, PCA για να αποκτηθούν τα βασικά χαρακτηριστικά του credit scoring, και στη συνέχεια, ένας ταξινομητής ANN για να καθορίσει την πιστοληπτική ικανότητα ενός ατόμου αιτούντος. Εξετάζει την απόδοση του προτεινόμενου μοντέλου σε σύγκριση με άλλα μοντέλα όσον αφορά την ακρίβεια και την κατάρτιση του χρόνου. Τα αποτελέσματα του, που βασίζονται σε γερμανική βάση δεδομένων, έδειξαν ότι το προτεινόμενο μοντέλο είναι ανώτερο από άλλα και υπολογιστικά φθηνότερο.

## 2.10 Support Vector Machine (SVM)

Η Support Vector Machine (SVM) [Μηχανή Διανυσμάτων Υποστήριξης (ΜΔΥ)] είναι μια τεχνική ταξινόμησης που προτάθηκε για πρώτη φορά από τον **Vapnik, 1995 [98]**, και στηρίζεται στην Θεωρία Στατιστικής Εκμάθησης (Statistical Learning Theory) και στα νευρωνικά δίκτυα Perceptron (δυναμικός ταξινομητής).



Εικόνα 13: Σχηματική απεικόνιση της μεθόδου Support Vector Machine (SVM).

Σκοπός της μεθόδου είναι να βρεθεί μια μέγιστη υπερ-επιφάνεια (maximum margin hypersurface) που να διαχωρίζει στο χώρο των παραδειγμάτων τα αρνητικά από τα θετικά. [108] Για ένα δοθέν σύνολο εκπαίδευσης έστω  $\{(x_i, y_i)\}$  με  $i = \{1, \dots, n\}$  όπου  $x_i$  είναι το διάνυσμα των επεξηγηματικών μεταβλητών, το  $y_i$  αντιπροσωπεύει την δίτιμη μεταβλητή απόκρισης, και  $n$  το πλήθος των εισαχθέντων διανυσμάτων. Η SVM έχει ως στόχο την εύρεση της βέλτιστης υπερεπιφάνειας κάνοντας το μέσω ενός μη πιθανοθεωρητικού δίτιμου γραμμικού ταξινομητή ο οποίος μπορεί να γραφτεί ως

$$\sum_{i=1}^n w_i x_i + b = 0$$

όπου  $\mathbf{w} = w_1, w_2, \dots, w_n$  είναι the normal of the hyper-plane και  $b$  ένα κατώφλι.

Υποθέτοντας ότι η υπερεπιφάνεια διαχωρίζει λαμβάνοντας υπόψη ότι  $y_i \in \{-1, 1\}$  και με γεωμετρική απόσταση  $\frac{2}{\|\mathbf{w}\|^2}$  η απόσταση μεγιστοποιείται υπό τον περιορισμό ότι

$y_i (\sum_{i=1}^n w_i x_i + b) \geq 1$ . Ευρέως, αυτή η μεγιστοποίηση μπορεί να γίνει μέσω πολλαπλασιαστών Lagrange και γραμμικών, πολυωνυμικών, Gaussian ή σιγμοειδών διαχωριστών.

Η τεχνική αυτή άρχισε να εφαρμόζεται σε μοντέλα credit scoring model μόλις πρόσφατα (Chen et al., 2009). Li et al. (2006); Gestel et al. (2006); Xiao and Fei (2006); Yang (2007); Chuang and Lin (2009); Zhou et al. (2009, 2010); Feng et al. (2010); Hens and Tiwari (2012); Ling et al. (2012). Κατά την εφαρμογή στα credit scoring, η μέθοδος περιλαμβάνει τρία στοιχεία. Μια φόρμουλα βαθμολόγησης η οποία είναι ένας γραμμικός συνδυασμός των χαρακτηριστικών που έχουν επιλεγεί για το πρόβλημα ταξινόμησης, μια αντικειμενική συνάρτηση η οποία θεωρεί τα δύο δείγματα εκπαίδευσης και ελέγχου ότι βελτιστοποιούν την ταξινόμηση των νέων δεδομένων, και ένας αλγόριθμος βελτιστοποίησης για τον προσδιορισμό των βέλτιστων παραμέτρων του δείγματος εκπαίδευσης της αντικειμενικής συνάρτησης. [87]

Τα πλεονεκτήματα της μεθόδου είναι ότι, στην μη-παραμετρική περίπτωση, η SVM δεν απαιτεί υποθέσεις για την δομή των δεδομένων, όπως κανονικότητα και συνέχεια. (Karamizadeh, Abdullah, Halimi, Shayan, & Rajabi, 2014) [52] Μπορεί να εκτελέσει μια μη γραμμική χαρτογράφηση από ένα αρχικό διάστημα εισόδου σε ένα μεγαλύτερου διαστάσεων χαρακτηριστικό χώρο και επίσης είναι ικανή να χειριστεί τόσο συνεχής, όσο και κατηγορικές προβλέψεις. Από την άλλη, οι αδυναμίες της μεθόδου αυτής είναι ότι, είναι δύσκολη η ερμηνεία, (Harris, 2015) [41], εκτός και εάν τα χαρακτηριστικά είναι ερμηνεύσιμα, ή εάν τα τυποποιημένα μοντέλα δεν περιέχουν εξειδικεύσεις εμπορικών περιορισμών (Ravi, 2007) [82].

## 2.11 Υβριδικά μοντέλα (Hybrid models)

Τα Υβριδικά μοντέλα είναι μοντέλα credit scoring που αναπτύσσονται με την ενσωμάτωση δύο ή περισσότερων υφιστάμενων μοντέλων. Το πλεονέκτημα αυτών των μοντέλων είναι ότι επωφελούμαστε από τα πλεονεκτήματα αυτών των δύο ή περισσότερων μοντέλων και επίσης μπορεί να αφαιρέσει την αδυναμία ενός μοντέλου μέσω του συνδυασμού του με ένα άλλο μοντέλο. Ωστόσο αυτές οι μέθοδοι είναι δύσκολο να διατυπωθούν και να εφαρμοστούν (Ravi, 2007) [82].

Σε γενικές γραμμές, ο συνδυασμός αυτός μπορεί να επιτευχθεί με διάφορους τρόπους κατά την εφαρμογή του credit scoring. Οι **Lee et al. (2002) [62]** πρότειναν μια υβριδική μέθοδο που ενσωματώνει τα νευρωνικά δίκτυα αναστροφής διάδοσης με την παραδοσιακή διαχωριστική ανάλυση. Οι **Huang et al. (2007) [44]** πρότειναν μια υβριδική μέθοδο που ενσωματώνει τον γενετικό αλγόριθμο (Genetic Algorithm) και τη Support Vector Machine για να εκτελέσει ταυτόχρονα την επιλογή χαρακτηριστικών και τη βελτιστοποίηση του μοντέλου. Αλλά και πιο πρόσφατα έχουν εφαρμοστεί αρκετά αποτελεσματικά πρωτότυπα υβριδικά μοντέλα credit scoring (**Harris, 2015 [41]; Karamizadeh, Abdullah, & Zamani, 2013 [51]; Oreski, 2014 [78]**), όπου ορισμένοι ερευνητές τα προτείνουν με το επιχείρημα ότι παρέχουν ακριβέστερες επεξηγήσεις.

Οι **Staporl et al. [54]** προτείνουν μια προσέγγιση για την οικοδόμηση ενός μοντέλου credit scoring που βασίζεται στο συνδυασμό της ετεροσκεδαστικής επέκτασης (Loog, Duin, 2002) της κλασικής γραμμικής διαχωριστικής ανάλυσης και έναν αλγόριθμο επιλογής χαρακτηριστικών που διατηρεί επαρκείς πληροφορίες για τον σκοπό της ταξινόμησης. Για να αξιολογηθεί η ακρίβεια του προτεινόμενου μοντέλου βαθμολόγησης πιστοληπτικής ικανότητας χρησιμοποιούν γερμανικά πιστωτικά δεδομένα από τη μελέτη των Chen, Li, 2010. Τα αποτελέσματα της μελέτης τους δείχνουν ότι η υβριδική προσέγγιση είναι μια αποτελεσματική και πολλά υποσχόμενη μέθοδος για την κατασκευή μοντέλων credit scoring.

Πίνακας 1 : Συνδυασμοί Μεθόδων στην δημιουργία Υβριδικών μοντέλων.

| Μέθοδος Ταξινομητή                | Επιλογή Χαρακτηριστικών                             | Συγγραφείς                                             |
|-----------------------------------|-----------------------------------------------------|--------------------------------------------------------|
| ANN                               | MARS                                                | Lee & Chen, 2005 [63]                                  |
| ANN                               | Logistic regression                                 | Lin, 2009 [69]                                         |
| SVM                               | SVM                                                 | Bellotti & Crook, 2009 [94]                            |
| SVM                               | CART + MARS                                         | W. Chen, Ma & Ma, 2009 [22]                            |
| ANN                               | LDA                                                 | Lee, Chiu, Lu, & Chen, 2002 [62]                       |
| MARS                              | ANN                                                 | Lee & Chen, 2005 [63]                                  |
| ANN                               | MARS                                                | Chuang & Lin, 2009 [23]                                |
| Genetic algorithm<br>an algorithm | Fuzzy Rule Based<br>LDA + heteroscedastic extension | Hoffman et. Al, 2007 [42]<br>Staporl et al., 2016 [54] |

## ΚΕΦΑΛΑΙΟ 3<sup>ο</sup> : Σύγκριση των μεθόδων

Στις προηγούμενες ενότητες αναλύθηκαν οι πιο ευρέως χρησιμοποιούμενες μέθοδοι για την εφαρμογή credit scoring. Το Credit Scoring έχει γίνει ένα ουσιαστικό αντικείμενο της πιστωτικής βιομηχανίας, ειδικά στις πιστωτικές υπηρεσίες των τραπεζών που έχουν να αντιμετωπίσουν ένα τεράστιο όγκο πιστωτικών δεδομένων, τα οποία είναι αδύνατον να αναλυθούν με όρους όπως ανθρώπινο δυναμικό και οικονομικά στοιχεία.

Όπως είδαμε οι μέθοδοι αυτές προέρχονται από τεχνικές ταξινόμησης. Η ταξινόμηση μπορεί να περιλαμβάνει οποιοδήποτε πλαίσιο κάτω από το οποίο λαμβάνεται μια απόφαση ή πραγματοποιείται μια πρόβλεψη. Μπορεί να οριστεί ως μια διαδικασία η οποία θα κατηγοριοποιήσει το σύνολο των δεδομένων σε ομάδες με κοινά χαρακτηριστικά. Ξέρουμε ότι η διαδικασία της ταξινόμησης είναι η πλέον κατάλληλη για την εξόρυξη γνώσης (data mining). Η σχεδίαση τεχνικών του σημερινού datamining χρονολογείται γύρω στην δεκαετία του 1950. Κατά την δεκαετία του 1960, η τεχνητή νοημοσύνη και ο κλάδος της στατιστικής εφάρμοσαν νέους αλγορίθμους.

Συνοπτικά οι μέθοδοι που αναλύθηκαν στο προηγούμενο κεφάλαιο ήταν :

1. Διαχωριστική Ανάλυση (Linear Discriminant Analysis - LDA)
2. Λογιστική Παλινδρόμηση (Logistic Regression-LR)
3. Bayesian classifier (BC)
4. K-Nearest Neighbour (KNN)
5. Multivariate Adaptive Regression Splines (MARS)
6. Δέντρα Απόφασης (Decision Trees -DT)
7. Ανάλυση Επιβίωσης (Survival Analysis-SA)
8. Fuzzy Rule-Based System (FRBS)
9. Νευρωνικά Δίκτυα (Neural Network-NN)
10. Support Vector Machine (SVM)
11. Υβριδικά μοντέλα (Hybrid models- HM)

Οι μέθοδοι αυτοί δύναται να ομαδοποιηθούν σε τρεις κατηγορίες :

1. Μέθοδοι που βασίζονται στην παλινδρόμηση (στατιστικές μέθοδοι).  
Σε αυτή εντάσσονται οι **Linear Discriminant Analysis (LDA)**, η **Logistic regression (LR)** και η **MARS**.
2. Μέθοδοι που βασίζονται στην εκμάθηση :  
Σε αυτή εντάσσονται οι : **Bayesian classifier (BC)**, **K-Nearest Neighbour (KNN)**, **Νευρωνικά Δίκτυα (Neural Network-NN)**, **Fuzzy Rule-Based System (FRBS)** και **Support Vector Machine (SVM)**
3. Λοιπές μέθοδοι : **Ανάλυση Επιβίωσης (Survival Analysis-SA)** και **Υβριδικά μοντέλα (Hybrid models- HM)**

Από τις μεθόδους αυτές οι πιο ευρέως διαδεδομένες για την εκτίμηση του πιστωτικού κινδύνου είναι η **Linear Discriminant Analysis (LDA)** και η **Logistic regression (LR)**. Επίσης, η **MARS**, καθώς είναι μέθοδος μη παραμετρικής παλινδρόμησης και άρα προσαρμοστική στα δεδομένα, παρόλο που ξεπερνά την υπόθεση της γραμμικότητας (που απαιτείται στην LDA) εμφανίζει το φαινόμενο της υπερπροσαρμογής. Το ίδιο πρόβλημα αντιμετωπίζουν και τα **νευρωνικά δίκτυα (NN)**, παρόλο που την τελευταία δεκαετία όλο και περισσότεροι ερευνητές ασχολούνται με αυτά. Επίσης σημαντική προσοχή δίνεται στην **Support Vector Machine (SVM)**, καθώς δεν απαιτεί τις υποθέσεις κανονικότητας και συνέχειας, και δε επηρεάζεται από αυτές όπως οι LDA και LR.

Μεταξύ των δύο τελευταίων, η SVM θεωρείται ανώτερη τεχνική στο να αντιμετωπίσει την ταξινόμηση, καθώς τα αποτελέσματα της είναι καλύτερα σε σχέση με τα NN. Συχνά η μεταξύ τους σύγκριση αναφέρει ότι τα SVM σε σχέση με τα NN, έχουν πιο ισχυρό θεωρητικό υπόβαθρο, επηρεάζονται λιγότερο από την υπερπροσαρμογή, χρειάζονται λιγότερη μνήμη για την αποθήκευση του προγνωστικού μοντέλου, και τέλος δίνουν πιο ευανάγνωστα αποτελέσματα και καλύτερη γεωμετρική ερμηνεία.

Βέβαια, καθώς τα ίδια τα δεδομένα είναι αυτά τα οποία θα υποδείξουν ποια τεχνική ταιριάζει καλύτερα σε αυτά, έχουμε ότι στην περίπτωση που έχουμε πολλά σημεία σε ένα χώρο χαμηλών διαστάσεων, τότε τα KNN είναι πιθανώς μια καλή επιλογή, ενώ εάν έχουμε το ανάποδο, δηλ. λίγα σημεία σε ένα χώρο με μεγάλες διαστάσεις, τότε είναι καλύτερα να χρησιμοποιηθεί ένα γραμμικό SVM.

Εάν συγκρίνουμε τα δέντρα αποφάσεων με τα νευρωνικά δίκτυα (και οι δύο αποτελούν σύγχρονες μεθόδους), έχουμε ότι τα δέντρα αποφάσεων εκπαιδεύονται ταχύτερα, καθώς απορρίπτουν στοιχεία που δεν θεωρεί χρήσιμα, οπότε είναι κατάλληλα για πολύ μεγάλες βάσεις δεδομένων (ενώ τα νευρωνικά δίκτυα χρησιμοποιούν όλα τα δεδομένα), τα δέντρα είναι πιο ερμηνεύσιμα, και ότι πρέπει να εξετάζεται πιθανό «κλάδεμα» για την αποφυγή της υπερπροσαρμογής. Από την άλλη τα νευρωνικά σε σχέση με τα δέντρα υπερτερούν στο ότι μπορούν να χειριστούν συνεχόμενα δεδομένα (και όχι κατά ομάδες) και να μοντελοποιήσουν αυθαίρετες συναρτήσεις (μη γραμμικές αλληλεπιδράσεις κλπ.).

Συγκρίνοντας τους ταξινομητές Naïve Bayes, NN και k-NN, έχουμε ότι ως προς την ταχύτητα εκμάθησης τα NN απαιτούν μακρά εκπαίδευση, βς προς την σύγκλιση, τα NN έχουν καλύτερη ακρίβεια, ως προς των χειρισμό ελλείπουσων τιμών τα NN δεν μπορούν να τις χειριστούν, και ως προς την διαφάνεια (transparency) τα NN είναι μαύρο κουτί αφού δεν μπορούν να παρουσιαστούν γραφικά. Εξετάζοντας και τα rule-based systems σε σχέση με τις παραπάνω μεθόδους έχουμε ότι ως προς τον θόρυβο, τα rule-based systems είναι πιο ασταθή στην ύπαρξη θορύβου, δηλ. στην ύπαρξη λανθασμένων δεδομένων, ενώ από την άλλη μεριά η ακρίβεια του Naïve Bayes είναι απίθανο να επηρεαστεί από μικρές ανακρίβειες.

Γενικότερα, όταν πραγματοποιείται μια σύγκριση μεταξύ των μεθόδων, τρία πράγματα είναι αυτά που λαμβάνονται κυρίως υπόψη : η ταχύτητα, η ερμηνευτικότητα και η απόδοση του μοντέλου.

Δεδομένου ότι για τη βελτίωση της απόδοσης του μοντέλου αυτού, είναι απαραίτητη μια διαδικασία **μείωσης της διάστασης του πλήθους των μεταβλητών**, προτείνονται **υβριδικά SVM** μοντέλα που βασίζονται στην παραπάνω λογική. Επιπλέον, τα υβριδικά μοντέλα έχουν ενστερνιστεί την τελευταία δεκαετία, καθώς οι ερευνητές επωφελούνται από τα πλεονεκτήματα των δύο ή περισσότερων μοντέλων που τα συνθέτουν. Τα περισσότερα από τα μοντέλα που προτείνονται, μπορούν να ταξινομήσουν τους δανειολήπτες σε δύο κατηγορίες («καλούς» ή «κακούς»). Όμως, από την οπτική μεριά της διαχείρισης κινδύνου, η πρόβλεψη της πιθανότητας αθέτησης της πίστωσης είναι καλύτερη από ό,τι χρησιμοποιώντας τη δυαδική ταξινόμηση. Και αυτό διότι τα μοντέλα αυτά είναι ικανά να εκτιμήσουν την αναμενόμενη κερδοφορία, ενώ ταυτόχρονα είναι κατανοητά και ερμηνεύσιμα (**Abbas Keramati et al. (2011) [1]**).

Η επιστημονική εξέλιξη της στατιστικής και η ανακάλυψη νέων μεθόδων ταξινόμησης, έθεσαν το ερώτημα, «Ποια μέθοδος ταξινόμησης είναι η καταλληλότερη για την εφαρμογή credit scoring;».



Πίνακας 2: Σύγκριση των τεχνικών εξόρυξης γνώσης από δημοσιεύσεις

|               |                               |                                                                              |                                                                 |
|---------------|-------------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------|
|               | Ακρίβεια / Επάρκεια           | Διαχωριστική Ανάλυση<br>Λογιστική Παλινδρόμηση<br>Probit Παλινδρόμηση<br>SVM | Νευρωνικά Δίκτυα<br>k – κονινότερου γείτονα<br>Δέντρα Αποφάσεων |
| Πλεονεκτήματα | Εύκολα Ερμηνεύσιμο            | Λογιστική Παλινδρόμηση                                                       | Probit Παλινδρόμηση                                             |
|               | Ευπροσάρμοστο σε άλλα μοντέλα | Λογιστική Παλινδρόμηση                                                       | Probit Παλινδρόμηση                                             |
|               | Υλοποιήσιμο                   | Διαχωριστική Ανάλυση<br>Λογιστική Παλινδρόμηση                               | Probit Παλινδρόμηση                                             |
| Μειονεκτήματα | Μακρά Εκπαίδευση              |                                                                              | Νευρωνικά Δίκτυα<br>SVM                                         |
|               | Μαύρο κουτί                   |                                                                              | Νευρωνικά Δίκτυα<br>SVM                                         |
|               | Έλλειψη εφαρμογής σε credit   | Fuzzy Rule-Based System<br>Multivariate Adaptive Regression Splines (MARS)   | Ανάλυση Επιβίωσης                                               |

Για τον λόγο αυτό, αρκετοί ερευνητές πραγματοποίησαν συγκρίσεις των μεθόδων αυτών με σκοπό να δώσουν μια απάντηση. Παραθέτουμε μερικές εξ αυτών κατά χρονολογική σειρά:

Οι **Galindo and Tamayo (2000) [37]** πραγματοποίησαν μια μελέτη η οποία εξετάζει μειονεκτήματα και πλεονεκτήματα διαφόρων τεχνικών ταξινόμησης στην εκτίμηση του πιστωτικού κινδύνου. Για αυτό το σκοπό μελετούν τα μοντέλα Probit, Neural Network, Decision tree και k-NN, καθώς και ορισμένους συνδυασμούς αυτών. Μεταξύ αυτών των μεθόδων η μέθοδος k-NN επιτυγχάνει το καλύτερο αποτέλεσμα.

Την ίδια χρονιά, ο **West (2000) [100]** διερευνά την ακρίβεια πέντε μοντέλων νευρωνικών δικτύων: πολυστρωματικό νευρωνικό δίκτυο (multilayer perceptron), μίγμα εμπειρογνομόνων (mixture-of-experts), ακτινική συνάρτηση βάσης (radial basis function), φορέα μάθησης κβαντοποίησης (learning vector quantization), και ασαφή προσαρμοστική συντονισμού (fuzzy adaptive resonance), συγκρίνοντας τα αποτελέσματα με τις παραδοσιακές μεθόδους, όπως την linear discriminant analysis, την λογιστική παλινδρόμηση, τον k - πλησιέστερου γείτονα (k nearest neighbour), την εκτίμηση πυκνότητας του πυρήνα (kernel density estimation), και τα δέντρα απόφασης. Τα αποτελέσματα έδειξαν ότι το πολυστρωματικό νευρωνικό δίκτυο μπορεί να μην είναι το πιο ακριβές μοντέλο νευρικού δικτύου, και ότι από κοινού τα μοντέλα μίγμα εμπειρογνομόνων και λειτουργίας νευρωνικών δικτύων ακτινικής βάσης δύναται να εφαρμοστούν σε credit scoring. Από τις παραδοσιακές μεθόδους, η λογιστική παλινδρόμηση βρέθηκε ότι είναι η πιο ακριβής.

Οι **Doumpos et al. (2002) [28]** εφάρμοσαν μια μέθοδο πολλαπλής Ιεραρχικής Διαχωριστικής ανάλυσης (Multi-group Hierarchical DIScrimination (M.H.DIS)) με σκοπό την ταξινόμηση του πιστωτικού κινδύνου και τη μέτρηση της αποτελεσματικότητας αυτού του μοντέλου έναντι ορισμένων παραδοσιακών μεθόδων, όπως η διαχωριστική ανάλυση, μοντέλα logit και probit, και καταλήγουν στο συμπέρασμα ότι το προσφερόμενο μοντέλο έχει περισσότερη ικανότητα ταξινόμησης σε σχέση με τις συγκρινόμενες μεθόδους

Οι **Xiao et al. (2006) [101]** αξιολογούν την ικανότητα των ορισμένων μεθόδων ταξινόμησης, όπως την linear discriminant analysis (LDA), logistic regression, neural networks, k-nearest neighbour (k-NN), support vector machines (SVM), classification and regression tree (CART) και multivariate adaptive regression splines (MARS). Με βάση αυτή τη μελέτη, SVM, MARS,

logistic regression και neural networks πραγματοποιούν μια πάρα πολύ καλή κατάταξη, αν και LDA και CART είναι εργαλεία πολύ πιο φιλικά προς τον χρήστη για την οικοδόμηση ενός μοντέλου credit scoring.

**Οι Huysmans et al. (2006) [45]** μέτρησαν την καταλληλότητα των χαρτών αυτο-οργάνωσης (self organization maps - SOMs) στα credit scoring. Επίσης αναφέρθηκαν στα πλεονεκτήματα της εφαρμογής ολοκληρωμένων SOM με κάποια εμποτευόμενο ταξινομητή αντί για την εφαρμογή μόνο της μεθόδου SOMs.

Οι **Kumar et al. (2006) [60]** διαπίστωσαν ότι το μοντέλο νευρωνικών δικτύων (ANN) ξεπέρασε πλήρως το μοντέλο LDA και στο στάδιο της εκπαίδευσης (του μοντέλου) και στις δοκιμαστικές καταμήσεις του συνόλου δεδομένων. Επίσης, το μοντέλο των νευρωνικών δικτύων φέρει μεγαλύτερη αξιοπιστία καθώς διαχειρίζεται πολύ καλύτερα περιπτώσεις ελλειπουσών τιμών, γεγονός που στην διαχωριστική παλινδρόμηση οι παρατηρήσεις θα αφαιρούνταν με αποτέλεσμα την παρατόμηση του συνόλου των δεδομένων.

Οι **Satchidanand et al (2007) [84]** μελέτησαν έντεκα ταξινομητές από πέντε διαφορετικούς τύπους συμπεριλαμβάνοντας την Μπευζιανή θεωρία, τα Νευρωνικά δίκτυα, διάφορα στατιστικά εργαλεία, μεθόδους μηχανικής μάθησης, μοντέλο που βασίζεται σε πυρήνα Kernel. Σκοπός τους να βρεθεί η καλύτερη μέθοδος στην πρόβλεψη της πιθανότητας αθέτησης. Στην εργασία τους αποδεικνύουν ότι ο πυρήνας Kernel που βασίζεται σε RBF νευρωνικά δίκτυα είναι η καλύτερη μέθοδος για τον εντοπισμό των πραγματικά θετικά περιπτώσεων.

Στην συνέχεια, οι **Atish et al. (2008) [9]** συνέκριναν την κερδοφορία επτά τεχνικών ταξινόμησης : Naive Bayes, logistic regression, neural networks, decision table, decision tree, k-nearest neighbour και support vector machine (SVM) με ή χωρίς ενσωμάτωση γνώσης. Για την αξιολόγηση των αποτελεσμάτων χρησιμοποιήθηκαν το κόστος ταξινόμησης και το εμβαδό κάτω από την καμπύλη ROC (Area Under Curve -AUC), όπου προέκυψε ότι η ενσωμάτωση του γνωστικού τομέα βελτώνει την αποτελεσματικότητα ορισμένων μεθόδων data mining.

Ένα χρόνο μετά, οι **Ince et al. (2009) [46]** χρησιμοποίησαν μια μαθηματικού προγραμματισμού (MP) μέθοδο ανάλυσης διακρίτοτητας και συνέκριναν τις επιδόσεις της με την logistic regression, την discriminant analysis, τον k-NN classifier και την SVM. Επίσης, αξιολογούν την ικανότητα των μεθόδων της discriminant analysis, της logistic regression, των neural network και των classification and regression tree (CART), στην πρόβλεψη και στην ταξινόμηση. Σύμφωνα με αυτή την μελέτη, CART και νευρωνικό δίκτυο ξεπέρασαν τις υπόλοιπες μεθόδους. **Οι Zhou et al. (2009) [106]** επανεξετάζουν την εφαρμογή ορισμένων δυαδικών μεθόδων ταξινόμησης στην πρόβλεψη και την ακρίβεια ταξινόμησης, χρησιμοποιώντας δεκατέσσερα σύνολα δεδομένων. Ο **Li (2009) [66]** εξερευνά την ακρίβεια ταξινόμησης της μεθόδου του k-Nearest Neighbor, της Support Vector Machine και των Νευρωνικών Δικτύων, χωρίς την επιλογή χαρακτηριστικών. Τα αποτελέσματα δείχνουν ότι η ενσωμάτωση αυτών των μεθόδων με την αποτελεσματική επιλογή χαρακτηριστικών προσέγγισης, οδηγούν σε ακριβέστερη ταξινόμηση. Ο **Twala (2010) [97]** διερευνά την ικανότητα πέντε ταξινομητών και τα ζεύγη συνόλων ταξινομητή (pairs of classifier ensembles) στο χειρισμό της πρόβλεψης του πιστωτικού κινδύνου. Τα αποτελέσματα δείχνουν ότι συνδυάζοντας τον ατομικό (επιμέρους) ταξινομητή μπορεί να βελτιώσει την ακρίβεια της πρόβλεψης. Οι **Paleologo et al. (2010) [79]** σχεδιάζουν ένα σύνολο credit scoring μοντέλου κατάλληλο να χειριστεί ελλείπουσες τιμές (missing information), μη ισορροπημένα δεδομένα (unbalanced data) και μη ανεξάρτητα και ισόνομα δεδομένα (non iid data).

## Επισκόπηση

Η πιστωτική βαθμολόγηση έχει γίνει πολύ σημαντικό ζήτημα λόγω της πρόσφατης ανάπτυξης της πιστωτικής βιομηχανίας, και το γεγονός ότι η υπερβολική χορήγηση δανειακών κεφαλαίων ήταν ένας από τις κύριες αιτίες της διεθνούς χρηματοπιστωτικής κρίσης. Οι τεράστιες βάσεις δεδομένων των τραπεζικών οργανισμών, είναι αδύνατο να αναλυθούν, τόσο σε οικονομικό όσο και σε ανθρώπινο δυναμικό, χωρίς την χρήση ειδικών τεχνικών εξόρυξης δεδομένων.

Σκοπός της παρούσας εργασίας είναι να παρέχει μια βιβλιογραφική έρευνα η οποία θα παρουσιάζει τις ευρέως εφαρμοσμένες μεθόδους μέτρησης του πιστωτικού κινδύνου και μια σύγκριση αυτών αναφορικά με τα πλεονεκτήματα και τα μειονεκτήματα τις κάθε μεθόδου.

Μια τέτοια αναφορά μπορεί να βοηθήσει τους ερευνητές να γνωρίζουν τα πιο κοινά μεθόδους αξιολόγησης της πιστοληπτικής αξιολόγησης, να βρουν τους περιορισμούς τους, να τους βελτιώσουν και να προτείνουν νέες μεθόδους με καλύτερες ικανότητες.

Βρίσκεται ότι οι πρωταρχικές στατιστικές μεθόδους της διαχωριστικής ανάλυσης και της λογιστικής παλινδρόμησης χρησιμοποιούνται ακόμα, αλλά από κοινού με την χρήση και πιο σύγχρονων μεθόδων (hybrid models). Τα τελευταία χρόνια χρησιμοποιείται ευρέως η SVM (μηχανή φορέα υποστήριξης) και αυτή όμως με τον συνδυασμό άλλων μεθόδων για να βελτιωθεί η απόδοση αυτού του μοντέλου. Τα υβριδικά μοντέλα παρακολουθούνται την τελευταία δεκαετία λόγω των πλεονεκτημάτων των δύο ή περισσότερων μοντέλων που τα απαρτίζουν. Όμως, πολλά από αυτά τα προτεινόμενα μοντέλα μπορούν να ταξινομήσουν τους πελάτες μόνο σε δύο κατηγορίες «καλούς» ή «κακούς».

Στην διαχείριση κινδύνου, η πρόβλεψη της πιθανότητας αθέτησης ενός υποψήφιου που υποβάλλει αίτηση για πίστωση, έχει μεγαλύτερη σημασία από την ταξινόμησή τους σε δύο ομάδες. Για το λόγο αυτό προτείνονται τα μοντέλα που έχουν την ικανότητα εκτίμησης της πιθανότητας αθέτησης και επίσης είναι απλά, και εύκολα να ερμηνευτούν.

## Παράρτημα – Πρόσθετοι Ορισμοί

**Επένδυση** αποτελεί μια δέσμευση χρημάτων ή κεφαλαίων, μέσω χρηματοοικονομικών εργαλείων ή άλλων περιουσιακών στοιχείων με την προσδοκία μιας θετικής απόδοσης στο μέλλον (**Φιλίππας, 2005 [112]**).

**Πιστωτής (Creditor)** είναι ένα πρόσωπο (φυσικό ή νομικό) στο οποίο οφείλονται χρήματα, παρέχει πίστωση, δίνει δηλαδή σε άλλα πρόσωπα τη δυνατότητα να δανειστούν χρήματα με την προϋπόθεση να τα επιστρέψουν σε μελλοντική ημερομηνία **[I.10]**. Μέσω της πίστωσης πραγματοποιείται διευκόλυνση των καταναλωτών στην απόκτηση την παρούσα στιγμή, καταναλωτικών αγαθών που λόγω της αξίας τους θα έπρεπε διαφορετικά να αγοραστούν ύστερα από το πέρασμα αρκετού χρόνου **[I.1]**. Οι πιστωτές μπορούν να διαχωριστούν σε προσωπικούς (άνθρωποι που δανείζουν χρήματα σε φίλους και συγγενείς) και πραγματικούς (όπως είναι οι τράπεζες), οι οποίοι έχουν νομικά συμβόλαια με τους οφειλέτες που εγγυώνται ότι ο δανειστής έχει δικαίωμα διεκδίκησης πάνω στα περιουσιακά στοιχεία του οφειλέτη αν ο τελευταίος αποτύχει να αποπληρώσει το δάνειο του. **[I.10]**

**Πιστοληπτική ικανότητα (Credit rating)** είναι η αξιοπιστία και η φερεγγυότητα ενός ατόμου, μιας επιχείρησης ή ακόμα και μιας χώρας στην αποπληρωμή των χρεών της. Αποκαλύπτει σε ένα δανειστή ή επενδυτή την πιθανότητα να μπορέσει ο δανειολήπτης να ανταποκριθεί στις δανειακές του υποχρεώσεις χωρίς τον κίνδυνο πτώχευσης. **[I.10]** Κάθε πιστωτής προκειμένου να εκτιμήσει την πιθανότητα εμφάνισης «ασυνέπειας» του δανειολήπτη, θέτει έναν **χρονικό ορίζοντα πρόβλεψης**, δηλ. την περίοδο για την οποία γίνεται η εκτίμηση. Ανάλογα με το στόχο, δηλαδή αν η πρόβλεψη πρέπει να είναι βραχυπρόθεσμη ή μεσο-μακροπρόθεσμη, η περίοδος αυτή κυμαίνεται, συνήθως, από 1-5 έτη. Στόχος είναι η βραχυπρόθεσμη πρόβλεψη της εμφάνισης ασυνέπειας. Σύμφωνα με τις κατευθύνσεις του κανονιστικού πλαισίου, ο χρονικός ορίζοντας πρόβλεψης της εμφάνισης ασυνέπειας ορίζεται στους 12 μήνες. **[I.5], [I.4]**. Με τον όρο **ασυνέπεια (probability of default)** τίθεται ο διαχωρισμός του κάθε δανειολήπτη σε «συνεπή» και «ασυνεπή». Κύρια στοιχεία που προσδιορίζουν την ασυνέπεια είναι (α) η αδυναμία εκπλήρωσης των οφειλών, (β) η εμφάνιση καθυστέρησης οφειλών πέραν των εκάστοτε προκαθορισμένων συνεχόμενων ημερών, (γ) η πτώχευση και οποιοδήποτε γεγονός το οποίο κρίνεται ότι δημιουργεί καθυστέρηση των οφειλών **[I.5], [I.4]**.

### Είδη Κινδύνου

Οι κίνδυνοι αυτοί μπορούν να χαρακτηριστούν, αναλόγως το αντικείμενο τους, ως εξής :

- [Λειτουργικός κίνδυνος](#)
- [Κίνδυνος Ρευστότητας](#)
- [Επιτοκιακός κίνδυνος](#)
- [Κεφαλαιακός κίνδυνος](#)
- [Κίνδυνος διακανονισμού](#)
- [Συναλλαγματικός κίνδυνος](#)
- [Κίνδυνος Φερεγγυότητας](#)
- [Κίνδυνος χώρας](#)
- [Νομικός κίνδυνος](#)
- [Κίνδυνος Επανεπένδυσης](#)
- [Πιστωτικός κίνδυνος](#)

**[I.12]** και συχνά τυγχάνει να επικαλύπτονται μεταξύ τους. Ένα διάγραμμα για τον τρόπο που οι κίνδυνοι επικαλύπτονται μεταξύ τους παρουσιάζεται από τον τραπεζικό όμιλο Credit Suisse 2000.

### **Λειτουργικός Κίνδυνος (Operational Risk)**

Είναι ο κίνδυνος που αντιμετωπίζει μία επιχείρηση κατά την διάρκεια της παραγωγικής διαδικασίας της και οφείλεται στην ανεπάρκεια των συστημάτων και των εσωτερικών ελέγχων, διαδικασιών, προσωπικού και πληροφορικών ή επικοινωνιακών συστημάτων, καθώς και λόγω εξωτερικών παραγόντων όπως φυσικές καταστροφές ή τρομοκρατικές επιθέσεις, που θέτουν εκτός λειτουργίας τα συστήματα διακανονισμού των συναλλαγών ή μειώνουν την αξία των περιουσιακών στοιχείων που αποτελούν αντικείμενα της συναλλαγής (π.χ. κίνδυνος κατάρρευσης των τεχνικών συστημάτων μιας οργανωμένης αγοράς, λανθασμένοι χειρισμοί της Διοίκησης μιας εταιρείας με τίλους εισηγμένους στο Χρηματιστήριο). Γενικά ο λειτουργικός κίνδυνος είναι ο κίνδυνος που προέρχεται από ανεπαρκείς ή αποτυγχάνουσες διαδικασίες, ανθρώπους, τεχνολογικά συστήματα ή από εξωτερικά γεγονότα. Στον λειτουργικό κίνδυνο υπάγεται και ο νομικός κίνδυνος. Επιχειρήσεις με μικρότερες απαιτήσεις για ανθρώπινο κεφάλαιο θα έχουν και χαμηλότερο επιχειρησιακό κίνδυνο [I.12].

### **Κίνδυνος Ρευστότητας (Liquidity Risk)**

Συγκαταλέγεται στους χρηματοοικονομικούς κινδύνους, και προκαλείται από τυχόν έλλειψη ρευστότητας στην αγορά ως προς ένα ή και περισσότερα χρηματοπιστωτικά μέσα, π.χ. αβέβαιη έγκαιρη ρευστοποίηση μίας επένδυσης, όπως είναι το ομόλογο, το οποίο αδυνατεί να πωληθεί εγκαίρως ώστε να αποτρέψει ή να μετριάσει μία απώλεια κεφαλαίου. Μπορεί να οφείλεται τόσο στην διαχείριση του ενεργητικού όσο και στην διαχείριση του παθητικού. Από την πλευρά του παθητικού θα πρέπει να γίνει μια διαχείριση, όπως δανεισμός επιπλέον κεφαλαίων, ή πώληση στοιχείων του ενεργητικού προκειμένου να ανταπεξέλθει ο οργανισμός στην απόσυρση των κεφαλαίων. Τα περισσότερα στοιχεία του ενεργητικού δύναται να ρευστοποιηθούν αλλά με πολύ υψηλό κόστος. Εναλλακτικά αντί να ρευστοποιηθούν στοιχεία του ενεργητικού ένας οργανισμός μπορεί να αγοράσει ή να δανεισθεί επιπλέον κεφάλαια. Επομένως συμπεραίνουμε ότι για να ελαχιστοποιηθεί ο κίνδυνος που προέρχεται από κρίσεις ρευστότητας θα πρέπει ο τραπεζικός οργανισμός να μπορεί να απομονώσει τον ισολογισμό του από τον κίνδυνο ρευστότητας με αποτελεσματική διαχείριση των μεταβλητών στοιχείων του ενεργητικού και του παθητικού. Για τον σκοπό αυτό οι τράπεζες λειτουργούν τμήματα-αυτόνομες μονάδες - που καλούνται Μονάδες Διαχείρισης Ενεργητικού / Παθητικού (Asset-Liability Committee) [I.12].

### **Επιτοκιακός Κίνδυνος (Interest rate risk)**

Είναι ο κίνδυνος αλλαγής της αξίας μιας επένδυσης λόγω μεταβολών στο επίπεδο των επιτοκίων. Οι τιμές των ομολόγων έχουν αρνητική συσχέτιση με τα επιτόκια. Καθώς αυξάνονται τα επιτόκια οι τιμές των ομολόγων μειώνονται, και το ανάποδο. Αυτό οφείλεται στο γεγονός ότι καθώς τα επιτόκια αυξάνονται, το κόστος ευκαιρίας διακράτησης ενός ομολόγου αυξάνεται και οι επενδυτές θα στραφούν σε επενδυτικά προϊόντα με υψηλότερες αποδόσεις. Πιο συγκεκριμένα, όταν τα επιτόκια αυξάνονται, παρουσιάζονται στην αγορά νέες εκδόσεις με υψηλότερες αποδόσεις από τις παλαιότερες, κάνοντας τις παλαιές εκδόσεις ομολόγων να αξίζουν λιγότερο. Επομένως οι τιμές τους πέφτουν. Αντίθετα, όταν τα επιτόκια μειώνονται οι νέες εκδόσεις ομολόγων που παρουσιάζονται στην αγορά έχουν χαμηλότερες αποδόσεις από τις παλιότερες εκδόσεις κι άρα κάνουν τις παλιές εκδόσεις να παρουσιάζουν μεγαλύτερη τρέχουσα απόδοση. Επομένως οι τιμές ανεβαίνουν.

Σαν αποτέλεσμα εάν ένας επενδυτής θελήσει να πουλήσει το ομόλογο πριν την λήξη του, αυτό μπορεί να αξίζει περισσότερο ή λιγότερο απ' όπi πλήρωσε για να το αγοράσει. Ο επιτοκιακός κίνδυνος συνήθως επηρεάζει περισσότερο τις τιμές των ομολόγων, από αυτές των μετοχών.

Παραδοσιακά τα επιτόκια αυξάνονται όταν η οικονομία παρουσιάζει ανάπτυξη και μειώνονται όταν η οικονομία παρουσιάζει ύφεση. Με παρόμοιο τρόπο η άνοδος του πληθωρισμού οδηγεί σε άνοδο των επιτοκίων, και η μείωση σε χαμηλότερα επιτόκια.

Τρόποι διαχείρισης της ευαισθησίας του επιτοκίου είναι :

- Το υπόδειγμα ανοίγματος (funding gap model ή repricing model) που χωρίζεται σε στατικό και δυναμικό.
- Το υπόδειγμα του χρόνου ως προς την λήξη (maturity model).
- Το υπόδειγμα μέσης διάρκειας διαμονής (duration gap model).
- Η διαχείριση ενεργητικού-παθητικού συνίσταται στην συντονισμένη παρέμβαση της διοίκησης της εμπορικής τράπεζας έτσι ώστε να αποφευχθούν ή να ελαχιστοποιηθούν οι κίνδυνοι του επιτοκίου.

Τρόποι αντιμετώπισης του επιτοκιακού κινδύνου είναι :

- Η διαφοροποίηση χαρτοφυλακίου (diversification), αγοράζοντας χρεόγραφα με διαφορετικές αποδόσεις και ωριμάνσεις
- Η αντιστάθμιση κινδύνου (hedging), αγοράζοντας προϊόντα όπως το επιτοκιακό swap.

[I.12]

### **Κεφαλαιακός Κίνδυνος (Principal risk)**

Είναι ο κίνδυνος να χαθούν τα χρήματα μιας επένδυσης εξαιτίας χρεοκοπίας ή παράβασης συμβολαίου, καθώς και ο κίνδυνος να μειωθεί η αξία μίας επένδυσης κάτω από την αξία των χρημάτων που επενδύθηκαν σε αυτή. Οι τραπεζικές καταθέσεις δεν έχουν κεφαλαιακό κίνδυνο γιατί η κεντρική τράπεζα τις προστατεύει μέχρι ένα ορισμένο ποσό. Αντιθέτως είναι αυξημένος σε επενδύσεις όπως οι μετοχές και σε μικρότερο βαθμό στα ομόλογα [I.12].

### **Κίνδυνος Διακανονισμού (Settlement risk)**

Είναι ο κίνδυνος ένας αντισυμβαλλόμενος να μην παραδώσει ένα αξιόγραφο ή την ονομαστική αξία του σε μετρητά, ενώ ο άλλος το έχει παραδώσει ήδη σύμφωνα με την μεταξύ τους εμπορική συμφωνία. Ο κίνδυνος διακανονισμού ήταν ένα πρόβλημα στην αγορά συναλλάγματος μέχρι την δημιουργία του συστήματος διακανονισμού αδιάλειπτης σύνδεσης (CLS) που εκμηδενίζει τις χρονικές διαφορές μεταξύ των διακανονισμών, δημιουργώντας μια ασφαλέστερη συναλλαγματική αγορά. Αποκαλείται και «Herstatt risk», παίρνοντας το όνομα του από την παταγώδη αποτυχία της γερμανικής τράπεζας Herstatt. Για την ιστορία, στις 26 Ιουνίου το 1974, η άδεια λειτουργίας της τράπεζας ανακαλέστηκε από τις γερμανικές τραπεζικές ρυθμιστικές αρχές στο τέλος της ημέρας εξαιτίας της έλλειψης κεφαλαίου και αδυναμίας της να καλύψει τις οφειλόμενες υποχρεώσεις. Ωστόσο κάποιες τράπεζες είχαν προχωρήσει σε συναλλαγματικές πληρωμές με την Herstatt δίνοντας της γερμανικά μάρκα με την προσδοκία να εισπράξουν αμερικάνικα δολάρια αργότερα την ίδια ημέρα από την τράπεζα. Λίγο αργότερα όμως η Herstatt σταμάτησε όλες τις πληρωμές σε δολάρια, αφήνοντας τους συμβαλλόμενους με μεγάλες απώλειες. Αυτό το είδος κινδύνου είναι στενά συνδεδεμένο με τον κεφαλαιακό κίνδυνο [I.12].

### **Συναλλαγματικός κίνδυνος (Foreign Exchange Rate Risk)**

Είναι ο κίνδυνος που αναφέρεται στις μεταβολές που προέρχονται από την μεταβολή στις συναλλαγματικές ισοτιμίες και επρόκειτο να επηρεάσουν την αξία μιας επένδυσης που πραγματοποιείται σε «ξένο» νόμισμα, δηλ. σε νόμισμα διαφορετικό από το νόμισμα του επενδυτή αλλά και τις υποχρεώσεις ή τις απαιτήσεις των επιχειρήσεων. Όσο περισσότερα

περιουσιακά στοιχεία και υποχρεώσεις είναι σε ξένο νόμισμα, τόσο μεγαλύτερος (ενδέχεται) να είναι ο συναλλαγματικός κίνδυνος.

Η παγκόσμια αγορά συναλλάγματος (γνωστή με το ακρωνύμιο FOREX) είναι μακράν η μεγαλύτερη σε όγκο αγορά αξιόγραφων παγκοσμίως. Κάθε διεθνής τράπεζα που συμμετέχει στις αγορές FOREX ανακοινώνει καθημερινά δύο τιμές.

1. την ισοτιμία που είναι διατεθειμένη να αγοράσει ένα ξένο συνάλλαγμα (τιμή bid), και
2. την ισοτιμία που είναι διατεθειμένη να πουλήσει ένα ξένο συνάλλαγμα (τιμή ask), και

**Spread = ask price – bid price**

Οι παράγοντες που διαμορφώνουν το ύψος του spread, είναι :

- Το βάθος και η ρευστότητα για την συγκεκριμένη αγορά συναλλάγματος.
- Η αβεβαιότητα για την ισοτιμία του συναλλάγματος (volatility).
- Τα αποθεματικά που έχει μια συγκεκριμένη τράπεζα για το κάθε συνάλλαγμα.

### Κίνδυνος Φερεγγυότητας (Solvency Risk)

Είναι ο κίνδυνος που δημιουργείται από την ανισότητα μεταξύ ενεργητικού και παθητικού μιας εταιρείας, και συνδέεται με την πιθανότητα πτώχευσης μιας επιχείρησης. Συγκεκριμένα, όταν η αξία των στοιχείων του ενεργητικού, είναι μικρότερη από την αξία των στοιχείων του παθητικού, τότε η εταιρεία αντιμετωπίζει πτώχευση, καθώς σε περίπτωση ρευστοποίησης των στοιχείων του ενεργητικού, δεν είναι δυνατή η αποπληρωμή των υποχρεώσεων της. Οι εποπτικές αρχές έχουν εισάγει το θεσμό των Κεφαλαιακών Απαιτήσεων έναντι Πιστωτικού Κινδύνου, με στόχο την Κεφαλαιακή Επάρκεια (Capital Adequacy), η οποία θέτει ένα ελάχιστο επίπεδο Ιδίων Κεφαλαίων που απαιτούνται σε συνάρτηση με τον αναλαμβανόμενο Πιστωτικό Κίνδυνο. Για την μέτρηση του καθιερώθηκε ο συντελεστής του Cokke (**Cokke Ratio, 1988 [I.6] - Investopedia**), ο οποίος στην ελληνική του μορφή (ΠΔ/ΤΕ 2054/1992) ονομάστηκε Συντελεστής Φερεγγυότητας και ορίστηκε ως ο λόγος των Ιδίων Κεφαλαίων προς τα στοιχεία του Ενεργητικού και τα στοιχεία εκτός Ισολογισμού σταθμισμένα σύμφωνα με τον κίνδυνο τους. Από το σύμφωνο της Βασιλείας I η ελάχιστη απαιτούμενη τιμή ορίστηκε στο 8%.

$$\frac{\text{Εποπτιά} \sim \text{Ιδία} \sim \text{Κεφάλαια}}{\text{Πιστωτικός} \sim \text{Κίνδυνος} + \text{Κίνδυνος} \sim \text{Αγοράς} + \text{Λειτουργικός} \sim \text{Κίνδυνος}} \geq 8\%$$

### Κίνδυνος χώρας (Country Risk)

Κίνδυνος χώρας ή αλλιώς διασυνωριακός κίνδυνος είναι ο κίνδυνος που προέρχεται από τις απώλειες λόγω εμπορικών περιορισμών που επιβάλλονται σε διαφορετικές χώρες. Στην ουσία αποτελεί έναν ιδιαίτερο τύπο πιστωτικού κινδύνου που απορρέει από την υψηλή έκθεση μιας τράπεζας σε δανεισμό συγκεκριμένης κυβέρνησης. Αναφέρεται κυρίως στην περίπτωση παύσης πληρωμών σε μια συγκεκριμένη χώρα. Υπάρχουν διάφορες μορφές κινδύνου όπως :

- Αποκήρυξη ή άρνηση εκτέλεσης συμβατικής υποσχέσεως (Repudiation)
- Ρύθμιση (Rescheduling).

Οι μηχανισμοί που έχουν δημιουργηθεί για τον περιορισμό των εμπλεκόμενων είναι :

- Ανταλλαγές χρέους για κεφάλαιο (Debt for Equity Swaps)
- Πολυετής αναδιάρθρωση των δανείων (Multi Year Restructuring Agreements)
- Πώληση των δανείων στην δευτερογενή αγορά και
- Ανταλλαγή χρέους με χρέος (Debt for Debt Swaps).

### Νομικός Κίνδυνος (Legal risk)

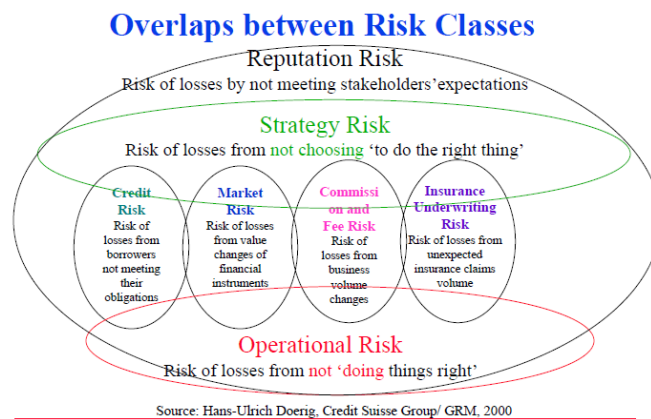
Είναι ο κίνδυνος που αντιμετωπίζουν επιχειρήσεις και επενδυτές να απωλέσουν κεφάλαια ή κέρδη λόγω απουσίας του κατάλληλου νομικού πλαισίου για την υποχρεωτική εκτέλεση συμβάσεων ή λόγω της έκδοσης αντίθετων δικαστικών αποφάσεων σε διαφορετικές χώρες στις οποίες δραστηριοποιούνται [1.12].

**Νομοθετικός Κίνδυνος (Legislative risk)**

Είναι το ρίσκο που παίρνει μια επιχείρηση είτε να εισαχθεί μια νέα νομοθεσία, είτε να αλλάξει μια παλαιότερη, με αποτέλεσμα να επηρεαστούν αρνητικά οι επενδύσεις της [1.12]. Για παράδειγμα το κλείσιμο των ελληνικών τηλεοπτικών σταθμών λόγω της εισαγωγής νέας νομοθεσίας (Νόμος Παππά) για περιορισμένο αριθμό τηλεοπτικών αδειών.

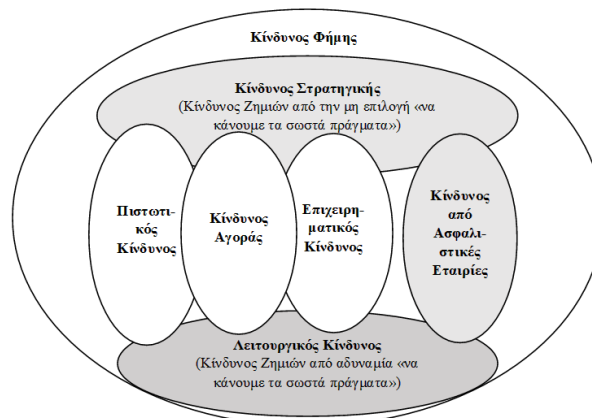
**Κίνδυνος Επανεπένδυσης (Reinvestment rate risk)**

Είναι ο κίνδυνος που προκύπτει για έναν επενδυτή όταν αναγκάζεται να επανεπενδύσει τα κεφάλαια του σε χαμηλότερα επιτόκια, επιτυγχάνοντας μικρότερες αποδόσεις. Ο κίνδυνος επανεπένδυσης είναι ιδιαίτερα υψηλός σε περιόδους πτωτικών επιτοκίων, όταν τα τοκομερίδια επανεπενδύονται σε χαμηλότερα επιτόκια από αυτά που επικρατούσαν όταν αγοράστηκε το επενδυτικό προϊόν. Αυτός ο κίνδυνος είναι ιδιαίτερα αισθητός στα ανακλητά ομόλογα. Ο Όμιλος Credit Suisse Bank (2000) [2,7] χρησιμοποιεί την κατηγοριοποίηση που παρουσιάζεται στο ακόλουθο διάγραμμα, στο οποίο φαίνονται οι επικαλύψεις που υπάρχουν ανάμεσα στις διάφορες κατηγορίες κινδύνων.



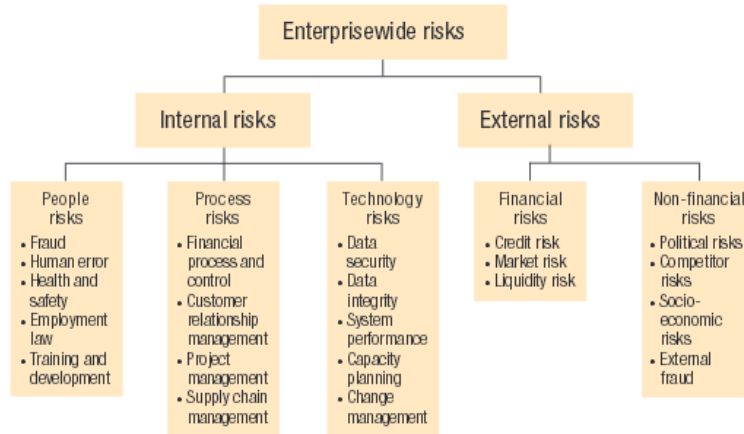
Source: Hans-Ulrich Doerig, Credit Suisse Group/ GRM, 2000

**Εικόνα 14 : Επικαλύψεις ανάμεσα στις κατηγορίες κινδύνων (Πρωτότυπο)  
Πηγή: Όμιλος Τράπεζας Credit Suisse / ΔΔΚΟ 2000**



**Εικόνα 15 : Επικαλύψεις ανάμεσα στις κατηγορίες κινδύνων (Μετάφραση στα Ελληνικά).  
Πηγή: Όμιλος Τράπεζας Credit Suisse / ΔΔΚΟ 2000**





**Εικόνα 16 : Κατηγοριοποίηση κινδύνων μιας επιχείρησης (Πρωτότυπο).**  
 Πηγή: Peyman Mestchian, Risk Intelligence – from compliance to performance, 2005



**Εικόνα 17: Κατηγοριοποίηση κινδύνων μιας επιχείρησης (Μετάφραση στα Ελληνικά).**  
 Πηγή: Peyman Mestchian, Risk Intelligence – from compliance to performance, 2005.

## Βιβλιογραφία

- [1] Abbas Keramati, Niloofar Yousefi, 2011A Proposed Classification of Data Mining Techniques in Credit Scoring, Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, January 22 – 24.
- [2] Abdou, H., Pointon, J., and El-Masry, A., 2008, "Neural nets versus conventional techniques in credit scoring in Egyptian banking." *Expert Systems with Applications* 35(3), 1275-1292.
- [3] Abeer Hamdy, Walid B. Hussein, 2016, Credit Risk Assessment Model Based Using Principal component Analysis And Artificial Neural Network, *MATEC Web of Conferences* 7 02039 (2016)
- [4] Akkoc, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*.
- [5] Alireza Hooman, Mohana Omid, G. Marthandan, Wan Fadzilah Wan Yusoff, Sasan Karamizadeh, 2015, STATISTICAL AND DATA MINING METHODS IN CREDIT SCORING, Proceedings of the Asia Pacific Conference on Business and Social Sciences, Kuala Lumpur (in partnership with The Journal of Developing Areas) ISBN 978-0-9925622-2-9
- [6] Altman, E. I., 1968, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, 23(4), 589–609.
- [7] Angelo Arvanitis & John Gregory, (2001), *Credit- The Complete Guide to Pricing, Hedging and Risk Management*, Risk Books
- [8] Antonakis, A. C., and Sfakianakis, M. E., 2009, "Assessing naive Bayes as a method for screening credit applicants." *Journal of Applied Statistics*, 36(5), 537-545.
- [9] Atish P, S., and Huimin, Z., 2008, "Incorporating domain knowledge into data mining classifiers: An application in indirect lending." *Decision Support Systems* 46(1), 287–299.
- [10] Atman, E. I., & Saunders, A. (1998) "Credit Risk Measurement: Development over the last 20 years", *Journal of banking and finance*: 1721-1742
- [11] Baesens, B., Egmont-Petersen, M., Castelo, R., and Vanthienen, J., 2002, "Learning Bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search" *International Congress on Pattern Recognition*. City: IEEE Computer Society pp. 49-52.
- [12] Baesens, B., Gestel, T. V., Stepanova, M., Poel, D. V. d., and Vanthienen, J., 2005, "Neural Network Survival Analysis for Personal Loan Data." *The Journal of the Operational Research Society*, 56(9), 1089-1098.
- [13] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627-635.

- [14] Beaver, W., 1967, "Financial ratios as predictors of failures." *Empirical Research in Accounting: selected*, 38(1), 63-93.
- [15] Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36(2, Part 2), 3302-3308.
- [16] Beran, J., & Djaïdja, A. Y. K. (2007). Credit risk modeling based on survival analysis with immunes. *Statistical Methodology*, 4(3), 251-276.
- [17] Berkson, J., 1944. Application of the logistic function to bio-assay. *Journal of the American Statistical Association* 39 (227), 357{365.
- [18] Bessis, J., (2002), *Risk Management in baking*, 2nd Revised Edition, United Kingdom: Published by John Wiley and Sons.
- [19] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453.
- [20] Chen, H., & Chen, Y. (2010). A comparative study of discrimination methods for credit scoring. Paper presented at the Computers and Industrial Engineering (CIE), 2010 40th International Conference on.
- [21] Chen, M. C., and Huang, S. H., 2003, "Credit scoring and rejected instances reassigning through evolutionary computation techniques." *Expert Systems with Applications*, 24(4), 433–441.
- [22] Chen, W., Ma, C., & Ma, L. (2009). Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications*, 36(4), 7611-7616.
- [23] Chuang, C.-L., & Lin, R.-H. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications*, 36(2, Part 1), 1685-1694.
- [24] Chye Koh, H., Chin Tan, W., and Peng Goh, C., 2006, "A Two-step Method to Construct Credit Scoring Models with Data Mining Techniques." *Journal of Business and Information*, 1, 96-118.
- [25] Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". *J Roy Stat Soc B*. 20: 215–242. JSTOR 2983890.
- [26] DeLoach James, 2004. *The new risk imperative – an enterprise-wide approach*, *Handbook of Business Strategy*.
- [27] Doerig Hans – Ulrich, *Operational Risk in Financial Services – An old challenge in a new environment*, Όμιλος Credit Suisse, 2003.
- [28] Doumpos, M., & Zopounidis, C. (2014b). *Introduction to Financial Decision Making Multicriteria Analysis in Finance* (pp. 1-10): Springer.
- [29] Doumpos, M., Kosmidou, K., Baourakis, G., and Zopounidis, C., 2002, "Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis." *European Journal of Operational Research*, 138(2), 392–412.

- [30] Dunham, Margaret H. "Data Mining - Introductory and Advances Topics", Prentice hall, 2004
- [31] Durand, D., 1941, Risk elements in consumer instalments financing.
- [32] Evans J. and Archer S. (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis, *Journal of Finance* ,23, December
- [33] Finard J. & Stocks M., A Framework for Corporate Financial Risk Management, included in Klein R. & Lederman J., *Derivatives Risk and Responsibility*, Irwin, USA, 1996
- [34] Fischer Black, Myron Scholes και Robert Merton, 1973.
- [35] Fisher, R., 1936, "The use of multiple measurements in taxonomic problems." *Annals of Eugenics*, 7, 179–188.
- [36] Friedman H. Jerome (1991), *Multivariate Adaptive Regression Splines*, *The Annals of Statistics*, Vol. 19, No 1, 1-141.
- [37] Galindo, J., and Tamayo, P., 2000, "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications." *Basic Methodology and Risk Modeling Applications* 15(1-2), 107–143.
- [38] Giudici, P., 2001, "Bayesian data mining, with application to benchmarking and credit scoring." *Applied Stochastic Models in Business and Society*, 17, 69–81.
- [39] Giudici, P., 2003, "Applied Data Mining: Statistical Methods for Business and Industry". John Wiley & Sons.
- [40] Hand J. D. and Henley E. W (1996), *Statistical Classification Methods in Consumer Credit Scoring: A Review*, *Journal of the Royal Statistical Society, Series A*, Vol. 160, No 3, 523-541.
- [41] Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42(2), 741-750.
- [42] Hoffmann, F., Baesens, B., Mues, C., Van Gestel, T., & Vanthienen, J. (2007). Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research*, 177(1), 540-555.
- [43] Hsieh, N. C., 2004, "An integrated data mining and behavioral scoring model for analyzing bank customers." *Expert Systems with Applications* 27(4), 623-633.
- [44] Huang, C.-L., Chen, M.-C., Wang, C.-J., 2007. Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications* 33 (4), 847-856.
- [45] Huysmans, J., Baesens, B., Vanthienen, J., and Gestel, T. v., 2006, "Failure prediction with self organizing maps." *Expert Systems with Applications* 30(3), 479-487.

- [46] Ince, H., and Aktan, B., 2009, "A Comparison of Data mining Techniques for Credit Scoring in Banking: A managerial Perspective." *Journal of Business Economics and Management*, 10(3), 233–240.
- [47] JH Friedman, Multivariate adaptive regression splines. *The annals of statistics*, 1991 - JSTOR
- [48] Jiawei Han, Micheline Kamber, "Data Mining – Concepts and Techniques", Morgan Kaufmann publishers, 2001
- [49] Kadam, A., and Lenk, P., 2008, "Bayesian inference for issuer heterogeneity in credit ratings migration." *Journal of Banking & Finance* 32 (10), 2267–2274.
- [50] Kambal, E., Osman, I., Taha, M., Mohammed, N., & Mohammed, S. (2013). Credit scoring using data mining techniques with particular reference to Sudanese banks. Paper presented at the Computing, Electrical and Electronics Engineering (ICEEEE), 2013 International Conference on.
- [51] Karamizadeh, S., Abdullah, S. M., & Zamani, M. (2013). An overview of holistic face recognition. *IJCCT*, 2(9), 738-741.
- [52] Karamizadeh, S., Abdullah, S. M., Halimi, M., Shayan, J., & Rajabi, M. J. (2014). Advantage and drawback of support vector machine functionality. Paper presented at the Computer, Communications, and Control Technology (4CT), 2014 International Conference on.
- [53] Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An overview of principal component analysis. *Journal of Signal and Information Processing*, 4(03), 173.
- [54] Katarzyna Stapor, Tomasz Smolarczyk, HETEROSCEDASTIC DISCRIMINANT ANALYSIS COMBINED WITH FEATURE SELECTION FOR CREDIT SCORING Piotr Fabian
- [55] Ketaki Chopde, Pratik Gosar, Paras Kapadia, Niharika Maheshwari, Pramila, M. Chawan, 2012, A Study of Classification Based Credit Risk Analysis Algorithm, *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249–8958, Volume 1, Issue 4, April 2012
- [56] Khashman, A., 2010, "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes." *Expert Systems with Applications*, 37(9), 6233-6239.
- [57] Khemais et al. (2016)
- [58] Kim, Jin Hee Yoon, Technology credit scoring model with fuzzy logistic regression, *Applied Soft Computing* 43 (2016) 150–158
- [59] Kirkos, E., Spathis, C., and Manolopoulos., Y., 2007, "Data Mining techniques for the detection of fraudulent financial statements." *Expert Systems with Applications* 32(4), 995-1003.

- [60] Kumar, K., and Bhattacharya, S., 2006, "Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances." *Review of Accounting and Finance*, 5(3), 216-227.
- [61] Lahsasna, A., Aionon, R. N., & Teh, Y. W. (2010). *Credit Scoring Models Using Soft Computing Methods: A Survey*. *Int. Arab J. Inf. Technol.*, 7(2), 115-123.
- [62] Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23(3), 245-254.
- [63] Lee, T.-S., Chen, I.-F., 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications* 28 (4), 743-752.
- [64] Leonardo J. Sales<sup>1,3</sup> and Rommel N. Carvalho, 2016, *Measuring the Risk of Public Contracts Using Bayesian Classifiers*, BMAW 2016 - Page 13 of 59
- [65] Lhabitant F. & Tinguely T., 2001. *Financial Risk Management: An Introduction*, *Thunderbird International Business Review*, Vol. 43 (3), pages 343-363.
- [66] Li, F. C., 2009a, "Comparison of the Primitive Classifiers without Features Selection in Credit Scoring" *International Conference on Management and Service Science*. City: IEEE.
- [67] Li, X. S., and Guo, Y. H., 2006, "Personal credit scoring models on naive Bayesian classifier." *Computer Engineering and Applications*, 42(1), 197-201.
- [68] Liang, Y., and Xin, H., 2009, "Application of Discretization in the Use of Logistic Financial Rating" *International Conference on Business Intelligence and Financial Engineering*, IEEE
- [69] Lin, S. L. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications*, 36(4), 8333-8341.
- [70] Linda L. Golden, Patrick L. Brockett, Jing Ai & Bruce Kellison, *Empirical Evidence on the Use of Credit Scoring for Predicting Insurance Losses with Psycho-social and Biochemical Explanations*, Pages 233-251 | Published online: 01 Sep 2016
- [71] Loretta J Mester, *What's the point of credit scoring?*, Publication date 1997/9/1
- [72] Lyn C, T., 2000, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers." *International Journal of Forecasting* 16 (2), 149–172.
- [73] Malhotra, R., and Malhotra, D. K., 2003, "Evaluating consumer loans using neural networks." *Omega*, 31(2), 83-96.
- [74] Mamdani, E. H. (1974). Applications of fuzzy algorithm for control a simple dynamic plant. *Proc. Of the IEEE* 121, pp. 1585–1588.
- [75] McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*. 5 (4): 115–133. doi:10.1007/BF02478259.

- [76] McNeil, A. J., and Wendin, J. P., 2007, "Bayesian inference for generalized linear mixed models of portfolio credit risk." *Journal of Empirical Finance* 14(2), 131–149.
- [77] Nguyen, H.-T. (2015). How is credit scoring used to predict default in China? : University of Paris West-Nanterre la Défense, *EconomiX*.
- [78] Oreski, S. (2014). *Hybrid Techniques of Combinatorial Optimization with Application to Retail Credit Risk Assessment*
- [79] Paleologo, G., Elisseeff, A., and Antonini, G., 2010, "Subbagging for credit scoring models." *Journal of Operational Research* 201(2), 490–499.
- [80] Panigrahi, S., Kundu, A., Sural, S., and Majumdar, A. K., 2009, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning." *Information Fusion* 10(4), 354-363.
- [81] QiweiGan, BinjieLuo, Zhangxi Lin, "Risk Management of Residential, Mortgage in China Using Date Mining A Case Study", 2009, *International Conference on New Trends in Information and Service Science*
- [82] Ravi, V., 2007, *advances in banking technology and management: impacts of ICT and CRM: information science reference*.
- [83] Rosenberg and Gleit (1994), *Quantitative Methods in Credit Management : A Survey*. *Operations Research*. 42. 589-613.
- [84] Satchidanand, S. S., and Jay, B. S., 2007, "Empirical evaluation of sampling and algorithm selection for predictive modeling for default risk." *ISIS*.
- [85] Shi, B., Wang, J., Qi, J., & Cheng, Y. (2015). *A Novel Imbalanced Data Classification Approach Based on Logistic Regression and Fisher Discriminant*. *Mathematical Problems in Engineering*, 2015.
- [86] Smith, P.G., Merritt, G.M., 2002. *Proactive Risk Management: Controlling Uncertainty in Product Development*. *Productivity Press, New York*.
- [87] So Young Sohn, Dong Ha Kim, Jin Hee Yoon<sup>1</sup>Yonse, *Technology credit scoring model with fuzzy logistic regression*
- [88] Sohn, S. Y., & Shin, H. (2006). *Reject inference in credit operations based on survival analysis*. *Expert Systems with Applications*, 31(1), 26-29.
- [89] Stefanescu, C., Tunaru, R., and Turnbull, S., 2009, "The credit rating process and estimation of transition probabilities: A Bayesian approach." *Journal of Empirical Finance* 16(2), 216–234.
- [90] Stepanova, M., & Thomas, L. (2002). *Survival analysis methods for personal loan data*. *Operations Research*, 50(2), 277-289.
- [91] Takagi, T. and Sugeno, M. (1985) *Fuzzy identification of systems and its applications of modeling and control*, *IEEE Transactions of Systems. Man and Cybernetics* , USA (pp. 116–132).

- [92] Takagi, T., and Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics* 15(1), pp. 116–132.
- [93] Thomadakis, S. B. & M. D. Xanthakis (2006), *Money and Capital Markets*, Athens: Stamoulis Publishing (in Greek)
- [94] Tony Bellotti and Jonathan Crook, 2009, *Credit Scoring with Macroeconomic Variables using Survival Analysis*
- [95] Trevor Hastie, Jerome Friedman, Robert Tibshirani, *Support Vector Machines and Flexible Discriminants*, The Elements of Statistical Learning, Part of the series Springer Series in Statistics pp 371-409.
- [96] Tsai, M. C., Lin, S. P., Cheng, C. C., and Lin, Y. P., 2009, "The consumer loan default predicting model – An application of DEA–DA and neural network." *Expert Systems with Applications*, 36(9), 11682–11690.
- [97] Twala, B., 2010, "Multiple classifier application to credit risk assessment." *Expert Systems with Applications* 37(4), 3326–3336.
- [98] Vapnik, V., 1995, *the Nature of Statistical Learning Theory*, New York: Springer Vapnik, V., 1998. *Statistical learning theory*.
- [99] Wagner H. and Lau S., (1971). *The Effect of Diversification on Risks*, *Financial Analyst Journal*, November/ December.
- [100] West, D., 2000, "Neural network credit scoring models." *Computers & Operations Research* 27(11-12), 1131-1152.
- [101] Xiao, W., Zhao, Q., and Fei, Q., 2006, "a comparative study of data mining methods in consumer loans credit scoring management." *journal of systems science and systems engineering*, 15(4), 419-435.
- [102] Yanhao Wei, Pinar Yildirim, Christophe Van den Bulte, Chrysanthos Dellarocas, 2015, *Credit Scoring with Social Network Data* Page Range: 234 - 258
- [103] Yeh, I. C., and Lien, C. h., 2009, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." *Expert Systems with Applications* 36(2), 2473–2480.
- [104] Yu, L., Wang, S., and Lai, K. K., 2008a, "Credit risk assessment with a multistage neural network ensemble learning approach." *Expert Systems with Applications* 34(2), 1434-1444.
- [105] Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics* 3, pp. 28–44.
- [106] Zhou, L., and Lai, K. K., 2009, "Benchmarking binary classification models on data sets with different degrees of imbalance." *Frontiers of Computer Science in China*, 3(2), 205-216.



[107] Zurada, J., and Lonial, S., 2005, "Comparison of the Performance of Several Data Mining Methods for Bad Debt Recovery in the Healthcare Industry." the Journal of Applied Business Research, 21(2), 37-53.

[110] Ζοπουνίδης Κ., Λεμονάκης Χ., (2009), «Διαχείριση Πιστωτικού Κινδύνου», Ζοπουνίδης, Λεμονάκης (2009), Διαχείριση Πιστωτικού Κινδύνου, Αθήνα, Εκδόσεις Κλειδάριθμος

[111] Μπαμπινιώτη Γ., «Λεξικό της Νέας Ελληνικής Γλώσσας», αναθεωρημένη έκδοση 2005.

[112] Νικολαός Δ. Φιλίππας, 2005, Αμοιβαία Κεφάλαια : Η ελληνική πραγματικότητα και οι σύγχρονες διεθνείς εξελίξεις, Εκδόσεις Μπάμπαλος Β. – Στυλιανίδης Σ.Ε.Ε., Αθήνα 2010

### Ιστοσελίδες

[I.1] Alpha Bank, Οικονομικό Δελτίο, Τεύχος 85, Μάρτιος 2003

[http://www.alpha.gr/files/infoanalyses/oikon\\_deltio\\_85.pdf](http://www.alpha.gr/files/infoanalyses/oikon_deltio_85.pdf)

[I.2] Bank for International Settlements Press & Communications, (2006), "Basel Committee on Banking Supervision - International Convergence of Capital Measurement and Capital Standards", A Revised Framework Comprehensive Version

[http://www.bankofgreece.gr/BogDocumentDOMLibraryMonthlySubject/%CE%92%CE%B1%CF%83%CE%B9%CE%BB%CE%B5%CE%AF%CE%B1\\_I\\_II\\_III.pdf](http://www.bankofgreece.gr/BogDocumentDOMLibraryMonthlySubject/%CE%92%CE%B1%CF%83%CE%B9%CE%BB%CE%B5%CE%AF%CE%B1_I_II_III.pdf)

[I.3] Basel Committee on Banking Supervision.(February 2006) <http://www.bis.org>

[I.4] Basel Committee on Banking Supervision, 2004, International Convergence of Capital Measurements and Capital Standards, A Revised Framework (paragraph: 452 & 453).

<http://www.bis.org/publ/bcbs107.pdf>

[I.5] ICAP GROUP A.E. – «Μεθοδολογία Απόδοσης Αξιολογήσεων Πιστοληπτικής Ικανότητας» : <http://www.icap.gr>

[I.6] Investopedia: <http://www.investopedia.com/terms/c/cookeratio.asp>

[I.7] Βικιπαίδεια, Διεθνής χρηματοπιστωτική κρίση 2007-2008

[https://el.wikipedia.org/wiki/%CE%94%CE%B9%CE%B5%CE%B8%CE%BD%CE%AE%CF%82\\_%CF%87%CF%81%CE%B7%CE%BC%CE%B1%CF%84%CE%BF%CF%80%CE%B9%CF%83%CF%84%CF%89%CF%84%CE%B9%CE%BA%CE%AE\\_%CE%BA%CF%81%CE%AF%CF%83%CE%B7\\_2007-2008](https://el.wikipedia.org/wiki/%CE%94%CE%B9%CE%B5%CE%B8%CE%BD%CE%AE%CF%82_%CF%87%CF%81%CE%B7%CE%BC%CE%B1%CF%84%CE%BF%CF%80%CE%B9%CF%83%CF%84%CF%89%CF%84%CE%B9%CE%BA%CE%AE_%CE%BA%CF%81%CE%AF%CF%83%CE%B7_2007-2008)

[I.8] Γ. Πρωτονοτάριος, Οι Διεθνής οίκοι αξιολόγησης και η διαμάχη τους με την Ευρωζώνη, Capitalinvest.gr, [http://www.capitalinvest.gr/info.php?product\\_id=227](http://www.capitalinvest.gr/info.php?product_id=227)

[I.9] Ελευθεροτυπία: Ο ιός πέρασε τον Ατλαντικό, τεύχος 2/10/2008

[I.10] Ευρετήριο Οικονομικών Όρων, <https://www.euretirio.com/>

[I.11] Ισοπμία: «Σκότωσαν» το σχέδιο Πόλσον στο Κογκρέσο, τεύχος 30/9/2008

[I.12] Ιστοσελίδα «Ευρετήριο Οικονομικών Όρων», <https://www.euretirio.com/>

[I.13] Καθημερινή: Μέτρα της γαλλικής κυβέρνησης για την ενίσχυση των τραπεζών, τεύχος 2/10/2008

[I.14] Καμπόλης Χ. , Νικόλαος Γ. Τραυλός, «Το management σε καιρούς κρίσης- Τα αίτια της πρόσφατης παγκόσμιας χρηματοπιστωτικής κρίσης», ALBA Graduate Business School.  
<http://www.alba.edu.gr/faculty/Documents/cabolis%20siite.pdf>

[I.15] Μιχάλης Ανθρωπέλος Asset & Liability Management Διάλεξη 5 :Συναλλαγματικός Κίνδυνος <http://web.xrh.unipi.gr/faculty/anthropelos/ALM/Lecture%205.pdf>

[I.16] <http://www.tovima.gr/opinions/article/?aid=551152>

[I.17] Εμμανουήλ Πολυζόπουλου Ψυχίατρου – Ψυχοθεραπευτή  
<http://www.iatropedia.gr/ygeia/se-pies-chores-i-anthropi-aftoktonoun-ti-simveni-stin-ellada/32302/>

[I.18] ΠΔΤΕ 2589/20.8.2007  
<http://www.bankofgreece.gr/Pages/el/Bank/LegalF/Acts.aspx>