Journal of Mathematics and Informatics
Vol. 26, 2024, 1-11
ISSN: 2349-0632 (P), 2349-0640 (online)
Published 22 January 2024
www.researchmathsci.org
DOI:http://dx.doi.org/10.22457/jmi.v26a01239

Journal of Mathematics and Informatics

A Model to Improve Effectiveness and Efficiency of Credit Decision-Making

H.H.M.I. Kumara

Department of Mathematics, University of Colombo
Colombo 3, Sri Lanka
Email: 2017s16234@stu.cmb.ac.lk

Received 10 December 2023; accepted 21 January 2024

Abstract. Credit decisions are important because of the credit risk exposure of lending. When lenders offer mortgages, credit cards, or other types of loans, there is a critical risk that the borrower may not repay the loan. Similarly, if a company offers credit to a customer, there is also a risk that the customer may not pay their invoices. Therefore, it is an important step in the Credit Origination process in which the counterparty assuming a Credit Risk decides to proceed with the granting of credit in accordance with their Internal Governance. This paper, it is clearly points out that modelling non-linear relationships using Artificial Neural Networks to enhance the credit decision-making process of lending institutions by predicting the probability of default of individual borrowers as well it also provides that Introduction to Risk Based Lending by using the Naïve Bayes' Classification Algorithm to maximize the profitability and to minimize the credit risk. Furthermore, in this paper, the author has developed a model to predict the probability of default by using an ANN and introduced a Risk Rating Scale based on qualitative data which will lead to Risk Based Lending using Naïve Bayes' classification algorithm.

Keywords: Artificial Neural Network (ANN); Naïve Bayes classification; Credit Risk; Risk Rating Scale

AMS Mathematics Subject Classification (2010): 91G40

1. Introduction

Credit decisions are important because of the credit risk exposure of lending. When lenders offer mortgages, credit cards, or other types of loans, there is a critical risk that the borrower may not repay the loan. Similarly, if a company offers credit to a customer, there is also a risk that the customer may not pay their invoices. Moreover, the Increasing NPA (Non-Profit Assets) trend creates a negative impact on financial stability. A loan is in arrears when principal or interest payments are late or missed. A loan is in default when the lender considers the loan agreement to be broken and the debtor is unable to meet his obligations. Hence, it is an important step in the Credit Origination process in which the counterparty assuming a Credit Risk decides to proceed with the granting of credit in accordance with their Internal Governance.

From the advancement of modern science, it is very clear that a combination of Mathematics and Machine Learning can be used to boost decision-making in finance and economics. Previous research has been done to predict the probability of default using different techniques such as Regression Models, Linear Discriminant Analysis (LDA), ANN Models (Feed Forward Back Propagation, ANN (Feed Forward Back Propagation) Models have been identified as the most accurate method by previous researchers since ANN Models have been outperformed all other conventional statistical models. A Machine Learning approach to investment indication has been introduced using Multilayer Perceptron, Logistic Regression and Decision Tree [15]. The key finding that is after comparing facts such as accuracy, precision, recall, and F1-score was obtained to support the decision tree provide best classification. Mirete-Ferrer, Pedro & Garcia-Garcia, Alberto & Baixauli-Soler, Juan & Prats, Maria [13] discuss how machine learning methods apply to the assets management by carrying out a critical analysis to discuss the current state-of-the-art and lay down a set of future research. Dowling, Michael & Piepenbrink, Anke & Aziz, Saqib & Hammami, Helmi introduced comprehensive structuring of the literature applying machine learning to finance and economics using a probabilistic topic modeling approach [1]. Forecasting and making predictions based on market values and experience have taken a huge jump as applications of machine learning in finance & economics. Nandi et al. discussed the trend of machine learning by highlighting the challenges associated with finance, such as data quality, model interpretability, and ethical considerations as well as demonstrating the significance of Machine Learning in finance [14].

The application of Machine Learning techniques in the scope of credit risk analysis has uniquely drawn attention. In general, it is clearly noted that the demand for the Identification and introduction of new variables, classifiers, and more assertive methods is constant [5]. In the modern corporate world using Machine learning models have emerged as innovative and essential tools in predicting financial distress. On the other hand, data preprocessing, such as missing data, imbalanced data, feature selection, and outliers are crucial aspects of Machine Learning implementations as they significantly impact the robustness and performance of the models [6]. Analyze large financial databases using machine learning models has shown that the inclusion of historical data, before the analysis, increases the prediction accuracy by the way Support Vector Machines are the most accurate predictor for the prediction of financial distress [17]. Vats and Samdani introduced different trading techniques and their effectiveness in quantitative trading and, finance to generate alphas is observed. These techniques are categorized by their reliance on Neural Networks, Support Vector Machines and other quantitative variables in finance as well as classifications based on supervised and unsupervised techniques and K-Mean clustering are also made [16]. Developing models for optimal asset allocation using the martingale method to assist an investor in selecting an asset that performs better under the conditions of a market information cascade has proven from past research since asymmetric information propagation frequently results in distorted financial markets and is generally a feature of informationally inefficient markets [10]. Komunte, Kasumo and

Masanja proposed the perturbed mathematical model for modelling the portfolio of insurance companies with possibilities of recovery after ruin. In that case return on investment and refinancing are used as approaches for overcoming ruin [11]. Moreover, past studies demonstrating the stock market model for the uncertainty of European stock prices [18] further motivated to do look back into modelling credit risks.

The banking system is the most vulnerable entity in the financial world. Financial risk has to be dealt with in a wide spectrum, from currency transactions to bond issuance and financial policy formulation. However, due to the liquidity and variability of the financial system, it is a problem that credit risk analysis is considered as a basic step in decision-making, but there is no definitive method for it. The interaction between bank capital regulation and financial innovation takes the form of structured finance and trenching of assets as well as the creation of separate structures with different seniority risk and capital charges [9]. Martin and Parigi did an interesting case study on US Prompt Corrective Action (PCA) where they explored the rationale for regulatory rules that prohibit banks from developing some of their natural activities when the capital level is low [12]. Callegaro, Fiorin and Grasselli defined The first recursive quantization-based approach for pricing options in the presence of stochastic volatility which was introduced using the concept of the transition probabilities for the discretized stock process that can be applied to any model for which an Euler scheme is available for the underlying price process and it allows one to price vanillas, as well as exotics [3]. A generalization of the Dynamic Conditional Correlation multivariate GARCH model of Engle and the Asymmetric Dynamic Conditional Correlation model of Cappiello et al. has defined an approach for Dynamic Conditional Correlation model for portfolio risk evaluation [2]. Studies on optimal investment with intermediate consumption in a general semi-martingale model of an incomplete market guarantee that utility maximization theory's key conclusion follows under the assumptions of no unbounded profit with bounded risk and of the finiteness of both primal and dual value functions [4]. Many recent works of proposing credit risk models where default intensities and interest rate are defined as various factors on different scenarios have been provided that coherent and unified approaches to pricing and risk management for a continuous tuning of the model to the actual situation of the economy [7,8].

The motivation behind this study is due to the potential of modeling non-linear relationships using Artificial Neural Networks to enhance the credit decision-making process of lending institutions by predicting the probability of default of individual borrowers. As well as it is also provided with an Introduction to Risk risk-based lending by using the Naïve Bayes' Classification Algorithm to maximize profitability and to minimize credit risk. Hence in this paper, the author attempted to develop a model to predict the probability of default by using an ANN and introducing Risk Based Lending based on Naïve Bayes' Classification Algorithm. The main focus of this research is to introduce a Risk Rating Scale based on qualitative data which will lead to Risk Based Lending.

2. Methodology

According to the "Statlog (German Credit Data)" dataset 5 qualitative attributes (Figure 1) such as credit history, Present employment since, Personal status and sex, Housing, and Job-status as well as 7 quantitative attributes (figure 2) such as Age in years, Present resident since, Duration in month, Credit amount, Installment rate in percentage, existing credits have been chosen to testing the desired model.
For the simplicity, qualitative attributes are labeled as follows.

Attribute 3: Credit history

A30 : no credits taken/ all credits paid back duly
A31 : all credits at this bank paid back duly
A32 : existing credits paid back duly till now
A33 : delay in paying off in the past
A34 : critical account/ other credits existing (not at this bank)

Frequency Table

	Credit History		Present Employment Since		Personal Status & Gender		Housing		Job Status	
Attributes	A30	15	A71	39	A91	30	A151	109	A171	15
	A31	21	A72	102	A92	201	A152	527	A172	144
	A32	361	A73	235	A93	402	A153	64	A173	444
	A33	60	A74	135	A94	67			A174	97
	A34	243	A75	189						
Observations		700		700		700		700		700

Figure 1. Qualitative Attributes

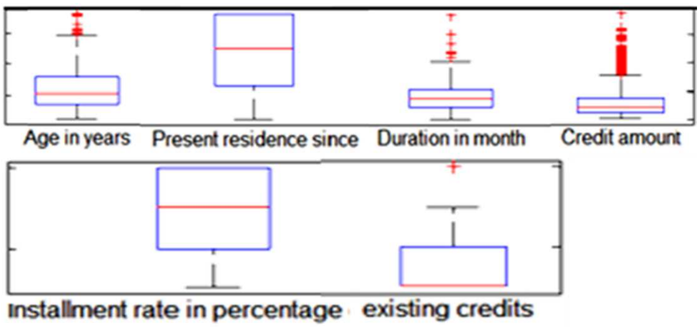


Figure 2. Quantitative Attributes

Attribute 7: Present employment since

A71 : unemployed
A72 :... < 1 year
A73 : 1 <= ... < 4 years

A74 : 4 <= ... < 7 years
A75 :.. >= 7 years

Attribute 9: Personal status and sex
A91 : male: divorced/separated
A92 : female : divorced/separated/married
A93 : male : single
A94 : male: married/widowed
A95 : female : single

Attribute 15: Housing
A151 : rent
A152 : own
A153 : for free

Attribute 17: Job
A171 : unemployed/ unskilled - non-resident
A172 : unskilled - resident
A173 : skilled employee / official
A174 : management/ self-employed/highly qualified employee/ officer

The particular dataset can be divided into two independent sets such as 300 defaulted loans and 700 paid loans. Here Artificial Neural Network model is used for the quantitative attributes while Navi Byes' Classifier Algorithm is used for qualitative attributes (Figure 3).

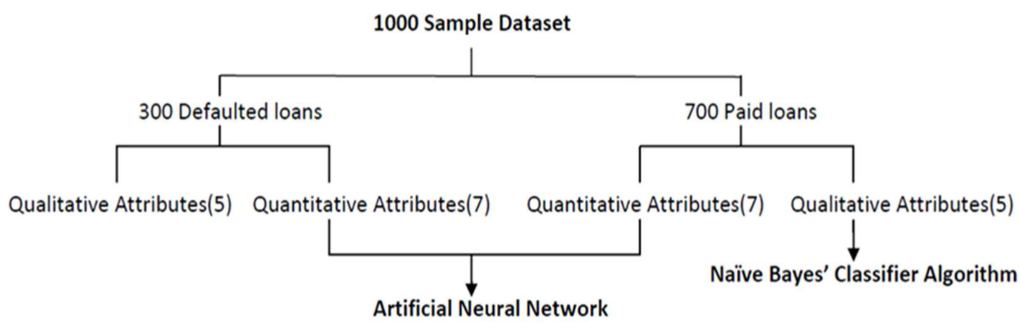


Figure 3. Data Analysis Methods

2.1. Artificial Neural Network for predicting the probability of default of individual borrowers

In this study, we have introduced a Three-layer ANN model with Tanh activation for the intermediate hidden layer with 10 nodes. The Proposed ANN architecture can be shown as follows (Figure 4).

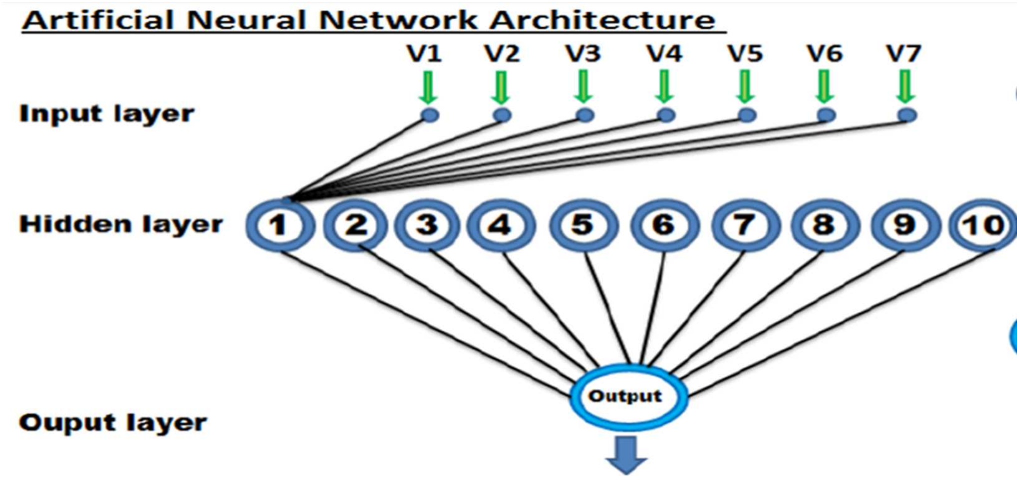


Figure 4. ANN Architecture

The ANN model can be divided into three main steps Feed Forward, Error Calculation, and Back Propagation. Hence following variables are introduced to the model. For simplicity we assumed that bias values for both of hidden layer and output layer are constant.

V_i = Attribute Value
W_i = Weights (Input Layer)
B = Hidden Layer Bias
K_i = Input Value to Hidden layer i^{th} node
F_i = Activated Value for i^{th} node
W_n = Weights (Output Layer)
b = Output Layer Bias
Y_i = Input Value to i^{th} node of Output Layer

1. Feed Forward
   The feed-forward approach addresses all three layers where Linear Regression is used for Inputs and Output Nodes. However, the Tanh activation function has been taken as the activation function for the hidden layer since its values lie between -1 to 1 that's why the mean for the hidden layer comes out be 0 or close to 0, hence tanh functions promote in centring the data by bringing mean close to 0 which makes learning for the next layer much easier than sigmoid function.

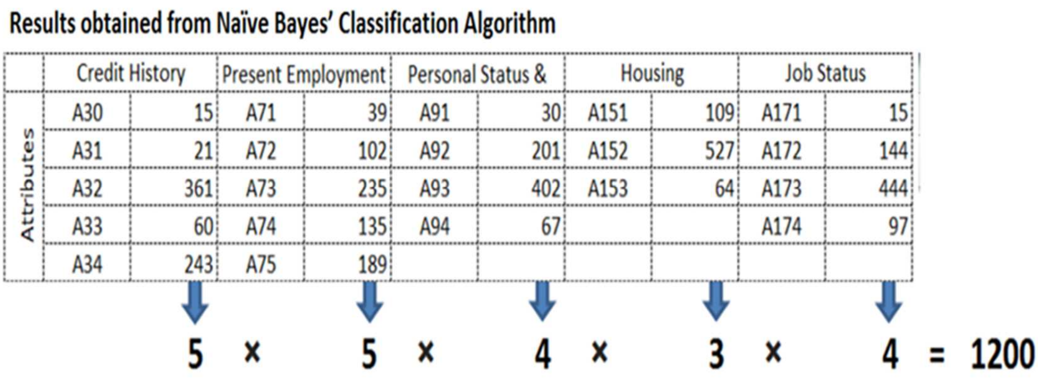
   $K_i = [\sum_{i=1}^{7} V_i\, W_i] + B$; Linear Regression Model for Input Layer
   $F_i = \frac{2}{1+e^{-2K}} - 1$; Tanh Model for Hidden Layer
   $Y_i = [\sum_{n=1}^{10} W_n\, f_n] + b$; Linear Regression Model for Output Layer

2. Error Calculation
   $\text{Error}_{\text{Total}} = \frac{1}{2}[Target\ -\ Output]^2$

3. Back Propagation (Levenberg – Marquardt Algorithm)
   In ANN model, backpropagation is more significant since that is used to test for errors working back from output nodes to input nodes. Because it is a very important mathematical concept for improving the accuracy of the model. On the other hand, this study is proposed on the Levenberg – Marquardt algorithm which is particularly designed to minimize sum-of-square error. Levenberg-Marquardt algorithm gives the best performance in the prediction compared to any other backpropagation.

2.2. Naïve Bayes' classification on introducing risk based lending

Naïve Bayes Classifier is a classification technique based on Bayes' Theorem with an independence assumption among predictors. In this case it is assumed that the presence of a particular feature in a class is unrelated to the presence of any other feature. In this example, the dataset contains with independent set of classes that are credit history, Present employment since, Personal status and sex, Housing, Job status as well as they have collection features. Hence the problem converts to the 1200 collection of observations (Figure 5).


Results obtained from Naïve Bayes' Classification Algorithm

	Credit History		Present Employment		Personal Status &		Housing		Job Status	
Attributes	A30	15	A71	39	A91	30	A151	109	A171	15
	A31	21	A72	102	A92	201	A152	527	A172	144
	A32	361	A73	235	A93	402	A153	64	A173	444
	A33	60	A74	135	A94	67			A174	97
	A34	243	A75	189						

5 × 5 × 4 × 3 × 4 = 1200

Figure 5. How problem converts according to Naïve Bayes Classification

The NBC model can be defined as follows.

Let,

$P(C|x)$ = Posterior Probability

$P(x|C)$ = Likelihood

$P(C)$ = Class Prior Probability

$P(x)$ = Predictor Prior Probability

Then,

$$P(C|x) = \frac{P(x|C)\,P(C)}{P(x)} \quad \text{and}$$

$$P(C|x) = P(x_1|C) * P(x_2|C) * P(x_3|C) * \ldots\ldots * P(x_n|C) * P(C)$$

In this study there are total of record of satisfactory borrowers as training data.

Now,

Probability of each attribute = $\frac{Attribute\ Frequency}{70}$

Note that numerical value 70 has been taken instead of 700 to obtain feasible probabilistic value.

3. Results and discussions

From this study the author aims to developing a model to predict the probability of default of by using an ANN and introducing Risk Based Lending based on Naïve Bayes' Classification Algorithm. In Here, results can be taken to account in two ways where the results from ANN model and Navi Bayes Classifier. In this study MATLAB® R2018a & MS Excel 2013 have been used for implementing models.

3.1. Results obtained from the artificial neural network model

Mean Squar error and R statistic are the key statistics to come up decision on accuracy of fitted model. In here Mean Square Error is the average squared difference between outputs and targets. On the other hand, Regression R values measure the correlation between outputs and targets. As per the results taken from MATLAB (Figure 6), it can be seen that both of Training, Testing, and Validation data lower values of MSE define that the better fitting of the model. However, R values close to 1 describes that the possible correlation of output and target. Moreover, this finding can be also further discussed from the diagrams possibly derived for Validation. Diagram of Mean Square Error against 10 Epochs for validation data (Figure 7) indicates that the accuracy of the model to the best fitting with best validation performance is 0.18888 at epoch 4. Furthermore, Regression R value (R=0.30059) shows a general correlation between output values and target values (Figure 8). All weights and biases generated throughout the training process of Artificial Neural Network can be used to predict the probability of default for new data sets.

Results obtained from Artificial Neural Network

	Samples	MSE	R
Training:	700	1.90489e-1	3.18590e-1
Validation:	150	1.88881e-1	3.34246e-1
Testing:	150	1.95827e-1	1.75071e-1

Figure 6. Results from ANN

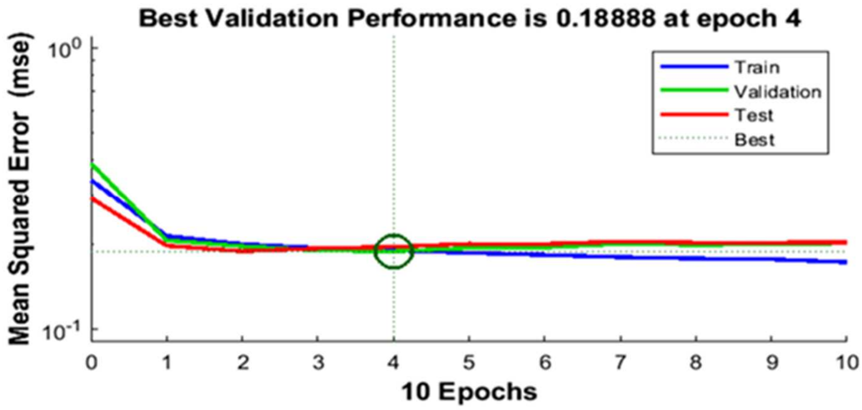


Figure 7. Model validation for training, validation and testing

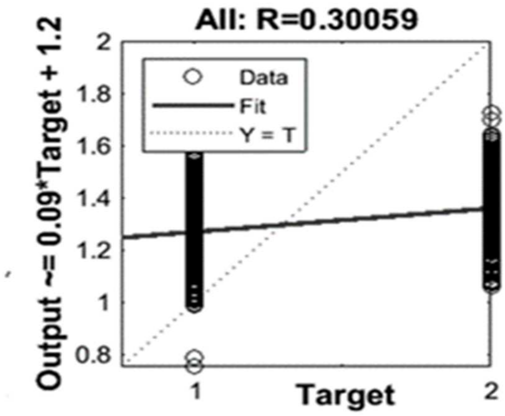


Figure 8. Correlation of output vs. target

3.2. Results obtained from the Naïve Bayes classifier

In the case of introducing a Risk Rating Scale, observations are been split into quartiles. Therefore, two hypotheses (H1: Highest Hypothetical Combination & H2: Lowest Hypothetical Combination) have been made to derive rating scales. In this study it is been appropriate 5 different rating scales for the accurate predictions. For the best decision makings Highest & Lowest Hypothetical Combination as well as observation averages of each quartile are to be using as boundaries of the class of rating scales A, B, C, D and E. According to the results derived from the particular training data it can be clearly identified that H1 takes 4747.932 & H2 takes 0.010 (Figure 9). Furthermore, it is clear that ratings of A, B, and C are good enough to classify as good and Rating D contains an uncertainty of crediting while E can be denoted as bad (Figure 10).

Combinatins Summary

		Q1	Q2	Q3	Q4
	Quartile	Q1	Q2	Q3	Q4
Range	Lower Value	0.010	1.396	5.786	23.957
Range	Higher Vale	1.383	5.777	23.768	4747.934
	Number of obsevations	300	300	300	300
	Observations Average	0.551	3.182	12.591	219.812

Figure 9. Results for Naïve Bayes classifier

Risk Rating

Low	Rating	High
219.812 <	A	<= 4747.934
12.591 <	B	<= 219.812
3.182 <	C	<= 12.591
0.551 <	D	<= 3.182
0.010 <	E	< = 0.551

Figure 10. Risk rating scale

4. Conclusion

As per the results and observations obtained during the course of this research, Machine Learning (ANN and NBC) can be used to improve the credit quality of lending institutions with the identification of the creditworthiness of borrowers. ANN can be used to categorize the satisfactory and unsatisfactory borrowers as the primary filtering and NBC can be used as the secondary filtering to introduce Risk Based Lending. Lending products can be developed to match the Risk-Based Lending concept to optimize the profitability of credit portfolios of lending institutions. The effectiveness and efficiency of the credit decision-making mechanism can be improved by using the results obtained from this research.
Examples:- Automation of Credit, *Approvals *Limit enhancements *Terms and Conditions

Acknowledgments. The author would like to express sincere gratitude to Dr. Buddhini Wijesuriya for assisting with modeling and implementation of this project. Also, the author would like to thank the reviewers who provided useful and detailed comments in improving the paper. The data that support the findings of this study are openly available in (*https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)*). I thank the reviewer for taking the time and effort to review this manuscript and the comments that have improved its quality.

Conflicts of Interest. The authors declare no conflicts of interest.

Authors' contributions. All authors contributed equally to this work.

REFERENCES

1. S.Aziz, M.Dowling, H.Hammami and A.Piepenbrink, Machine learning in finance: A topic modeling approach, *European Financial Management*, 28(3) (2022) 744-770.
2. M.Billio and M.Caporin, A generalized dynamic conditional correlation model for portfolio risk evaluation, *Mathematics and Computers in Simulation*, 79(8) (2009) 2566-2578.
3. G.Callegaro,L.Fiorin and M.Grasselli, Pricing via recursive quantization in stochastic volatility models, *Quantitative Finance*, 17(6) (2017) 855-872.
4. H.N.Chau, A.Cosso, C.Fontana and O.Mostovyi, Optimal investment with intermediate consumption under no unbounded profit with bounded risk, *Journal of Applied Probability*, 54(3) (2017) 710-719.
5. D.L.Duarte and F.L.de Moraes Barboza, Forecasting Financial Distress with Machine Learning–A Review, *Future Studies Research Journal: Trends and Strategies*, 12(3) (2020) 528-574.
6. K.El Madou, S.Marso, M.El Kharrim and M.El Merouani, Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis, *Expert Systems*, (2023) 13485.
7. C.Fontana and W.J.Runggaldier, Credit risk and incomplete information: filtering and EM parameter estimation, *International Journal of Theoretical and Applied Finance*, 13(05) (2010) 683-715.
8. C.Fontana, Credit risk and incomplete information: a filtering framework for pricing and risk management, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, (2012) 193-201.
9. X.Freixas and B.M.Parigi, Banking regulation and prompt corrective action, *21st Australasian Finance and Banking Conference*, June 2009.
10. A.A.Kilange, K.Ananda and P.Ngare, Portfolio Optimization under Informationally Asymmetric Markets, *Journal of Mathematics and Informatics*, 24 (2023) 45-60.
11. M.Komunte, C.Kasumo and V.G.Masanja, Insurance companies portfolio optimization with possibilities of recovery after ruin: a case of exponential utility function, *Journal of Mathematics and Informatics*, 21 (2021) 31-42.
12. A.Martin and B.M.Parigi, Bank capital regulation and structured finance.Journal of Money, *Credit and Banking*, 45(1) (2013) 87-119.
13. P.M.Mirete-Ferrer, A.Garcia-Garcia, J.S.Baixauli-Soler and M.A.Prats, A review on machine learning for asset management, *Risks*, 10(4) (2022) 84.
14. B.Nandi, S.Jana, and K.P.Das, Machine learning-based approaches for financial market prediction: A comprehensive review, *Journal of Applied Math*, 1(2) (2023).
15. A.V.D.Oliveira, M.C.S.Dazzi, A.M.D.R.Fernandes, R.L.S.Dazzi, P.Ferreira and V.R.Q.Leithardt, Decision support using machine learning indication for financial investment, *Future Internet*, 14(11) (2022) 304.
16. P.Vats and K.Samdani, Study on machine learning techniques in financial markets, *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, 1-5.

17. A.S.Vieira, J.Duarte, B.Ribeiro and J.C.Neves, Accurate prediction of financial distress of companies with machine learning algorithms, *In Adaptive and Natural Computing Algorithms: 9th International Conference*, ICANNGA 2009, Kuopio, Finland, 23-25 April 2009, 569-576.
18. X.Yu, A stock model for uncertain European finance market, *Journal of Mathematics and Informatics*, 1 (2014) 13-24.