Journal of the Nigerian Association of Mathematical Physics
Volume 57(**June – July** *2020 Issue*), **pp143 – 150**
© J. of NAMP

ON MACHINE LEARNING AND CREDIT RISK MANAGEMENT

A.U. Rufai and O. Fatayo

Department of Computer Science, Faculty of Science, University of Lagos

Abstract

Corporate Risk Management (CRM) is central to the growth of the economy of nation-states. With the advent of various financial organizations making credit available, the greatest risk is that of default in repayment of these funds. This can lead to a domino effect on the country's economy. This problem informed the need to have a mechanism to predict the risk of default beforehand to prevent or have minimal defaults on loans.

Machine Learning techniques have been adapted in the prediction in other domains. The various algorithms such as Adaptive boosting, ANN (Artificial Neural Networks, SVM (Support Vector Machine), Gaussian Processes etc. can with high degree of accuracy identify patterns that manual or human analyst cannot easily identify in a large volume of data. This work studied the probability of default from datasets that contains different features that can cause default, the features were recommended by industry experts due the historical patter to defaulters. The datasets is partitioned, with one partition for training the model and the other used to test the model. The result was able to predict the level of probability that a certain credit facility seeker will default in its repayment of the facility granted. A combination of different machine learning techniques was applied to the model and based on these techniques; the best fit score was used for the prediction. Some features were identified as an important feature that really need attention to every financial organization that grants credit facilities. The predictions ranges from 0 to 1 with probabilities greater that 0.5 having the probability of default while probabilities lesser than 0.5 are most likely to conform with repayment schedule.

1. Introduction

Corporate bankruptcy can impact a country's economy in a devastating way. Recently, there has been an increase in a number of companies trying to expand in order to grow their business. In the event of many multinational companies going into bankruptcy, the nation's economy will be greatly impacted. With rigorous qualitative (e.g. brand) and quantitative (e.g. econometric factors) analyzes, the financial credit risk that a company is being exposed to can be projected.

The advent of machine learning led to understanding complex models. Knowing trends in high-precision data which are not apparent to a human analyst requires the use of different methods such as Gaussian processes, artificial neural networks, adaptive boosting and support vector machines.

This paper focused on corporate financial company bankruptcy using machine learning techniques. Using datasets available to the public, many of the above listed machine learning approaches were used to learn about the correlation between the present state of the organization and its predicted prospects. Results showed that predictions with accuracy exceeding 95 per cent could be achieved using any machine learning technique when informative features such as expert evaluation were used. Nonetheless, when only financial variables are used to evaluate if a company will go bankrupt, the accuracy is not as high.

The prediction of bankruptcy goes all the way back to over two centuries ago, when most evaluations were carried out qualitatively [1]. More objective (and less subjective) methods became popular in the 20th century; some examples include Beaver's pioneering univariate analysis work [2] and Altman's numerous discriminant theoretical function in the 1960s [3]. This intelligence is an advantage not just to investors, compliance officers, shareholders, senior executives, etc., since it can influence them directly, but including other stakeholders, such as vendors plus workers [4].

The aim of this paper is to apply machine learning using rigorous and qualitative analysis done in an objective manner to expose the potential financial credit risk that a firm will likely face. The objectives include:

i. Review existing problem in credit risk management
ii. Perform findings on current application of machine learning in relation to credit risk management
iii. To use machine learning algorithms to diagnose the credit risk / financial health of companies.

Corresponding Author**:** Rufai A.U., Email: arufai@unilag.edu.ng, Tel: +2348034336246, +2348035364763 (OF)

iv. Train models in forecasting financial credit risk using both the qualitative evaluations from financial experts and quantitative econometric variables.

2. Related Works

The state of the financials in a particular organization using machine learning algorithms was identified [5]. The research focused extensively on the manufacturing industry. Aston Jacky however did not consider the actual financial sectors that constitute the strength of the economy, from his research, it can be concluded that machine learning has a very powerful capability to predict and detect patterns in a data, this makes it a very powerful tool. It compromises of different validation methodologies that aids its decision-making processes.

In [6], the focus was on FinTech's promotion of responsible lending in the economy. Unsurprisingly, there were other significant supporting conditions for cautious borrowing that went beyond the reach of this paper, such as sound regulation and supervisory regimes, complexity of the financial system and high financial literacy of borrowers. The research had strong suits for Machine Learning to ease the access of credit facility assessment for small borrower which happens to be more feasible and economical, his research identifies how machine learning can make meaning out of soft information available to financial institutions and how nonlinearities can be captured using machine learning. The drawbacks in [6] were machine Learning-based lending bears risks of financial exclusion, machine Learning-based credit rating could cause consumer protection, ethical, and data privacy issues and lastly it may not address structural changes.

The ROE and ROA was used in the determination of credit risk of some selected banks in Nigeria [7]. It tackles the traditional financial risk pertaining to the impact of antecedents of loan and advance loss provision on ROE and ROA. They relied on the output of non-automated processes; their research did not look at the inclusion of technology in the analysis of the data obtained for the purpose of credit risk management.

Machine learning was employed to solved statistical problem like regression, classifications, and clustering [8]. The first two can be solved by supervised learning and the other by unsupervised learning. They also looked on how to solve some financial related issues like, credit risk and revenue modelling, fraud and surveillance of conduct breaches in trading. This article did not discuss various algorithms that are applicable to the objective of the article, it was more of theoretical analysis of the application of the various machine learning algorithms. With such improved automation, it enabled financial institutions to gain good insight in various business processes. Secondly, Simpler approach combining non-linear analysis with simplicity to improve high auditability.

A binary classifier based on deep and machine learning models which tells the potential performance of a creditor was built [9]. With the recent advancement in the technology for computing power, big data and data availability most financial institutions are making modifications to business models possible. The major keys to decision-making are credit predictions, monitoring, reliability, and loan processing. This work was able to identify the various models for credit risk management, it also stated the need for checks on data quality to avoid bias for a class. There was also the need for the regulators/policy makers to come up with quick decisions in the use of data science techniques to boost performance. They stated the need to combine pools of models to match the data and business problem. There was however the need for data integrity, as they did not state the possible scientific solution to avoid data manipulation and contamination. It was concluded that the algorithms based on artificial neural network does not implicitly provide best performance, which infers that there was the need to provide regulations that ensures data integrity and transparency with respect to the decision algorithms. This was essential as it aimed to avoid discrimination and consequently reduce the negative impact on the host and global economy.

Machine-learning techniques were reviewed and assessed in respect of finding gaps or issues relating to risk management that were insufficiently addressed and had possible areas for more study [10]. The review found that the application of machine learning in banking risk management such as credit risk, market risk, operational risk and liquidity risk needed more exploration [10]. Though their research did not seem directly associated with the present level of emphasis on both risk management and machine learning in the industry. There were lots of areas in the bank's risk management that could generally benefit from the research of how machine learning can be applied to address pressing risk issues. Consequently, the potential of machine learning in the banking and financial sector was well understood, and risk management was expected to also strive by applying machine learning strategies to improve their capabilities.

Asurvey of different classifiers for credit risk evaluation was carried out in [11]. This paper was able to survey various classifiers that could be applicable for credit risk evaluation to satisfy that loan applicants have good credit history and also had the highest probability of returning the borrowed money back to the bank. There was more work to be done with respect to credit risk using machine learning classifier that could evaluate nosily dataset and make adequate comparison with advanced classifier which were available in the financial field. From the analysis and comparison of the accuracies using different types classifiers, it was found that the ELM classifier gave better accuracies compared to other classifiers. ELM gave 96.33%in German dataset and 96.32%in Australian dataset.

3. Experiments/Techniques

Various techniques for forecasting bankruptcy have been suggested. Several research articles have tried to group them into different methods statistically, smart structures, data extraction techniques and machine training techniques. Thus, all information-driven learning methods for cumulative and distinct outputs are known merely as machine learning techniques. Pre-processing, reduction of dimensionality (main component analysis), learning (learning parameters), model choice (validation), and testing (predictive accuracy assessment) are all involved in machine learning pipeline as shown in Figure1.

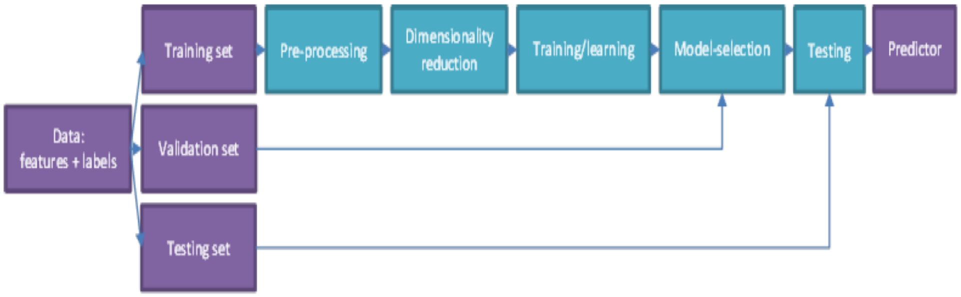

Figure 1: Machine learning process

It is worthwhile to observe patterns and progress in the past provided a succinct overview of credit risk prediction machine learning, highlighting several important research efforts over the last fifty years.

3.1 Linear Regression and Logistic Regression

Least-squares regression / estimation / adjustment is one of the most common statistical, engineering and econometric mathematical tools. Because of a set of N observations ỹ (i.e. dependent variable or regressand) and some characteristic map ϕ of explanatory variables χ (i.e. regressor), the best set of weights ŵ, which can be estimated to eliminate square residual errors (i.e. L 2-norm) The objective function for such linear models can be expressed mathematically by Equation 1. It can further be shown that this gives the Best Linear Unbiased Estimate (BLUE) of the unknown weights [12]. If the residuals, $\vec{e}$, obey a Gaussian probability distribution, the least-square solution can be shown to agree with the MLE (Maximum Likelihood Estimate).

$$\underset{w}{min}\vec{e}^T\vec{e} = \underset{w}{min}\left(\vec{y} - \vec{\phi}(\vec{x})\vec{w}\right)^T\left(\vec{y} - \vec{\phi}(\vec{x})\vec{w}\right) \dots\dots\dots\dots\dots\dots\dots..1$$

3.2 K-Dimensional Tree

K-Dimensional (K-D) Trees are one of the most popular nearest neighbour classification algorithms. In low dimensions D, (i.e. D < 20; in other words when only a handful of financial factors were used for bankruptcy prediction), a K-D Tree seems be a very efficient algorithm even when training the classifier with large number of companies (N > million).

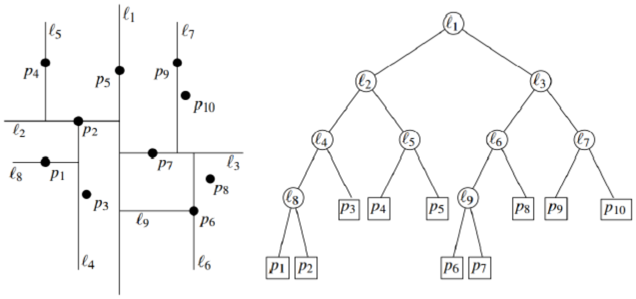

Figure 2: Graphical representation of a 2D K-D tree. The left side shows the spatial partitioning (with hyperplanes ℓ1,ℓ2…ℓ9 drawn) and the right side shows the corresponding tree structure [13].

3.3 Support Vector Machine

SVM (Support vector machine) is extensively applied compared to other algorithms in classification activities among the various Kernel Machines. The mathematical description can be quite complicated, but it can be quite easy to imagine and describe from a geometrical point of view. In a two-dimensional function space, if there are two linearly separable groups (O and X), there could be many possible solutions that all provide great accuracy (alternatively, minimized errors) on the training data (Figure 3).

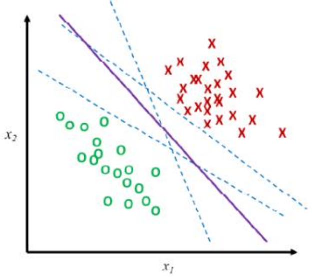

Figure 3: Binary separation of **O** and **X** using SVM in 2D.

For the Support Vector Machine to be trained from the dataset, there is need to convert the primal problem in Equation 2 to equation 3 which is a dual of equation 2. For this scenario, the labels $y^{\rightarrow}$ were {-1, 1} rather than {0, 1} as used in the formulation of logistic regression.

$$\min_{w} \vec{w}^T\vec{w} + C\sum_{i}^{N} \max\left(0{,}1 - y_i\left(\vec{w}^T\vec{\Phi}(x_i) + b\right)\right) \qquad \ldots\ldots\ldots\ldots\ldots 2$$

$$\max_{\alpha_i \geq 0} \sum_{i} \alpha_i - \frac{1}{2}\sum_{jk} \alpha_j\alpha_k y_j y_k \Phi(x_j)^T \Phi(x_k) \qquad \ldots\ldots\ldots\ldots..\ldots..3.1$$

$$s.t.\ 0 \leq \alpha_i \leq C\ \forall i \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots.\ldots\ldots\ldots\ldots\ldots 3.2$$

$$\sum_{i} \alpha_i y_i = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots.3.3$$

3.4 Decision Trees

A Decision Tree just like K-D tree, it includes binary tree structure and a decision-making logic which is easily interpreted.

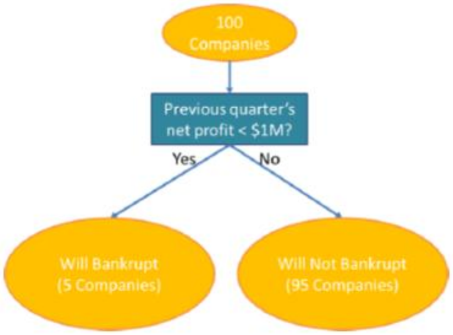

Figure 4: Decision tree

3.5 Ada Boost

This is based on the metaphor of a committee in which better conclusions are arrived at by the committee than a single individual. In this technique the decisions of many learners are combined to form a consensus that will lead to better learning.

3.6 Artificial Neural Network

This is based on the metaphor of the brain. The architecture of a two-layer neural network referred as multi-layer perceptron is illustrated in Figure 5.

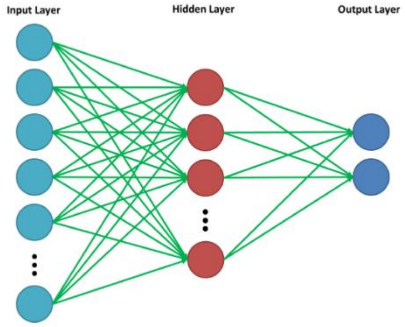

Figure 5: Artificial Neural Network

3.7 Gaussian Processes

Gaussian process is an attractive form of machine learning, which has gained popularity in the prediction of bankruptcy in recent years. Due to its high learning intricacy (Equation 4), which is O(N³), which is why it has not received much reception. But that bottleneck is slowly disappearing with recent computers and better numerical estimates. In addition to it being a technique that can without overfitting be endlessly versatile to suit any data which is a Bayesian non-parametric regression and classification technique, Gaussian method allocates acceptable possibility measures to its outputs. Therefore, when used with a particular kernel it is equivalent to a single layer of neural network with an infinite number of neurons (Equation 4)[14]

$$\min_{\theta} \frac{1}{2}\vec{w}^T K(\theta,\vec{x})^{-1}\vec{w} + \sum_{i=1}^{N} \log\left(sigm(y_i w_i)\right) + \frac{1}{2}\log\left|I + \nabla\nabla\log\left(sigm(y_i w_i)\right)^{\frac{1}{2}} K(\theta,\vec{x})\nabla\nabla\log\left(sigm(y_i w_i)\right)^{\frac{1}{2}}\right|$$

$$K(x,x^1)=\frac{2}{\pi}\sin^{-1}\left[\frac{2\vec{x}^T\sum\overrightarrow{x^1}}{\sqrt{(1+2\vec{x}^T\sum\vec{x})(1+2\overrightarrow{x^1}^T\sum\overrightarrow{x^1})}}\right]\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.4$$

4. Results and Discussions

For greater accuracy, two data sets were used, one of the datasets for the purpose of training and another for testing. The testing data was only used in the final accuracy assessment stage.

a. Pre-processing stage

The data set for training and testing the models had to go through various pre-processing stages. The below stages itemize the various processing taken.

i. **The visualization of data** - The raw data head is processed for the removal of special characters from the headers of the data set. This was achieved with a replace function withing python.

ii. **Missing Values** - Data gotten from different sources are hardly clean and incomplete, they are usually inconsistent, and it is a very critical task of the data scientist to process the data by cleaning it and filling the missing values. The missing values were handled by inputting the missing values based on any of these options;

a. Using the central tendency values of mean, median or mode for the column

b. Use estimated values obtained from a predictive model

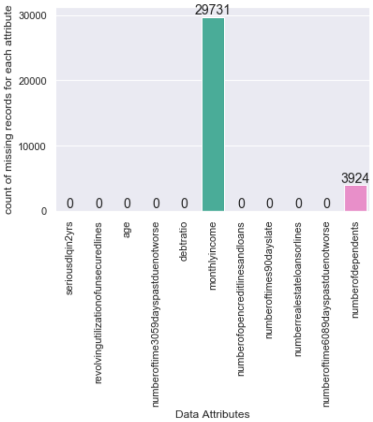

Figure 6: Figure show the graph of the count of missing records plotted against the different data attributes

The data set attributes monthly income and number of dependencies had more missing values compared to other data attributes.

iii. **Correlation** -If the outcome has a high correlation, this can be interpreted as two or more variables having strong relationship with each other and the other hand the outcome if the variables are hardly related, this is called weak correlation

iv. **Outliers Detection** - Outliers seem to be irrational values that differ from other data observations, they may imply a measurement variability, experimental errors, or newness. In other words, an outlier is an observation that differs from a pattern overall on a sample.

Outliers are of two types; Univariate and Multivariate. Univariate is usually seen in a single feature space when focus is on a distribution of values from the same single space. Multivariate is from multi-dimensional space. The detection from n-dimensional space is usually done by a trained model and it can be very difficult for the human brain. From the dataset used, Outlier detection was carried out in the dataset using three methods

a. Percentile based Outliers

b. Median Based Outliers

c. Standard deviation Outliers

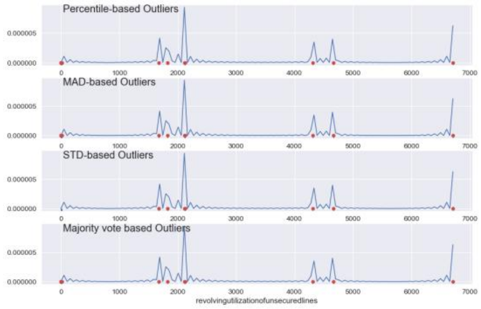

Figure 7: This figure shows the outlier detection done on the most important feature in the data set.

X-axis is the deviation from the Revolving Utilization Of Unsecured Lines in relation total number of Revolving Utilization Of Unsecured Lines analysed in Y-axis.

From the Figure 7all four outliers detection methods produced similar outputs detecting spikes in the same range of dataset analysed.

v. **Handling the Outliers** - In general, Outliers belong to two categories, the outlier can be due to a mistake in the data or it is a true outlier. The mistake in the data can be due to omission or transposition which can result to a big deviation when the data is being analysed. A true outlier would be something like finding an exceptional feature in the date.

There are four approaches to handling these outliers;

a. Try Transformation
b. Assign new value
c. Drop the outlier records
d. Cap your outlier data

vi. **Feature Importance** - Many times in data science there are millions of features and there is need to create a model the involves only important features. These has various benefits, it makes the model simpler and easily interpreted, it also reduces the variance in the model and finally the computational cost of training the model is reduced.

4.2. Train and Build Baseline Model

Training a model requires applying different machine learning processes to a base line model and subsequently use this model to perform prediction. In this project, different machine learning algorithms were applied to the model the following algorithms were applied.;

a. k-nearest neighbours
b. Logistic regression
c. AdaBoost Classifier
d. gradient boosting classifier
e. Random Forest Classifier

The Table 1 shows the score for each of the classifiers;

Table 1: The Receiver Operating Characteristic score.

Name of Classifier	Classifier Score	Multi-class(roc_auc_score)
k-nearest neighbours	0.9311733333333333	0.590447087413346
Logistic regression	0.9361333333333334	0.846805671097765
AdaBoost Classifier	0.9349866666666666	0.8571030443639817
gradient boosting classifier	0.9361333333333334	0.8629606221838481
Random Forest Classifier	0.93056	0.7777068858261436

A python function called the roc_auc_score function is used calculates the area under ROC curve, the ROC is termed Receiver Operating Characteristic, the ROC is also known or denoted by AUC or AUROC. Theoutput of the roc_auc_score is a summary of the curve's information in a single number.

4.3 Cross Validation

On the completion of the training stage of the model, it cannot be assumed that the model will work as planned on a new and unfamiliar data, it cannot be ascertained that the result will give the desired precision and variance in a new environment. There is need for a kind of confirmation that the predictions to be done by the model will be correct. For this there is need to carry out a validation. One of the methods of checking the effectiveness of a learning model is known as Cross Validation (CV). The output of the cross validation is as shown below;

Table 2: Cross Validation score (Mean and Standard Deviation).

Classifier	CV score mean	CV score STD
KNeighborsClassifier	0.5952570076118191	0.0023729926242542316
LogisticRegression	0.8494611439416649	0.0035935995416355084
AdaBoostClassifier	0.8586370125547675	0.0020946319753293264
GradientBoostingClassifier	0.8639041851581906	0.0026163153747133846
RandomForestClassifier	0.7789363956353903	0.0021570069056101324

4.4 Hyper-parameter optimization using Randomized search

Hyper-parameters are parameters indicated, which can influence the action of a machine learning algorithm by tuning. Hyperparameters are adjusted by selecting the optimal values of the parameter for greater precision. This mechanism can be tedious, but methods are available to facilitate such as Random Search and Grid Search process. Random Search, as its title implies, uses hyperparameter combinations at random. This implies that not all parameter values are attempted, and parameters with fixed number of iterations given by n_iter will be sampled instead. Two algorithms were tried namely Adaboost and Gradient Boosting.

Table 3: Estimators for the models.

Name	n_estimators	best_score_
AdaBoost	50	0.8575042244214536
GradientBoosting	272	0.8628721792424396

Train models with help of new hyper-parameter was done with the best estimator, below is the new CV output.

Table 4: Cross Validation for the models.

Classifier	CV score mean	CV score STD
GradientBoostingClassifier	0.8639753667550968	0.0026057280159451084
AdaBoostClassifier	0.8577636690935092	0.0028861895301146943

4.5 Testing on Real Dataset

The second dataset that was not used at all during the training of the model will now be used to test the model and to predict the probability of default. The first activity carried out on the dataset is to remove the special characters in the headers and populate the missing values with the median.

The 'predict_proba' method was used to predict the probability of default using the models earlier trained. The output of the probability was exported to a csv file called predictions.csv.

In the Table 5, the sample data from the input data set is displayed and the sample output from the probability.csv is also shown below.

Table 5: Sample Output

ID	probability	prediction
0	0.205195	Conform
1	0.188722	Conform
2	0.169358	Conform
3	0.210448	Conform
4	0.257336	Conform
5	0.180162	Conform
6	0.182105	Conform
7	0.177888	Conform
8	0.161514	Conform
9	0.200298	Conform
10	0.165593	Conform
11	0.167957	Conform
12	0.166519	Conform
13	0.206198	Conform
14	0.192904	Conform
15	0.168679	Conform
16	0.178837	Conform
17	0.170706	Conform
18	0.316757	Conform
19	0.533143	Default
20	0.556547	Default
21	0.560927	Default
22	0.593209	Default
23	0.615306	Default
24	0.704969	Default
25	0.676878	Default
26	0.528168	Default
27	0.550362	Default
28	0.531007	Default

4. Conclusion and Future Work

Two datasets of companies in the financial sector were analysed and compared to other researchers' results. The first dataset was used to train the date model to be able to learn how to handle data from similar dataset and predict a certain probability. The second dataset uses 64 quantitative financial features for assessing the likelihood of a certain lender to default in repayment. Results from this work indicated that all machine learning algorithms applied to the test dataset yielded a probability that is enough for the decision makers of a financial firm. This can be attributed to the fact that the model had been trained and having identified the important features of the model. In future data extraction, only the important features could be mined and used for analysis, this should reduce the size and time of data compilation. As demonstrated in this study, multiple machine learning techniques were used to train the model, relying on a single control metric can be biased. In general, ML has a very powerful capability to predict and detect patterns in a data, this makes it a very powerful tool. It compromises of different validation methodologies that aids its decision-making processes. Telling which technique is superior to another is very difficult. It is therefore essential to harness the strengths of various methods and combine their strengths to reach better decisions and judgments as advised by experts.

References.

[1] Andrés, Javier &Landajo, Manuel & Lorca, Pedro. (2012). Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. Knowledge-Based Systems. 30. 67–77. 10.1016/j.knosys.2011.11.005.)

[2] Beaver, W. (1966). Financial Ratios as Predictors of Failure. Journal of Accounting Research, 4, 71-111. doi:10.2307/2490171

[3] Altman, & I, Edward. (2012). The prediction of corporate bankruptcy: a discriminant analysis.

[4] Boritz, J. & Kennedy, Duane & Albuquerque, Miranda. (1995). Predicting Corporate Failure Using a Neural Network Approach. Intelligent Systems in Accounting Finance & Management. 4. 1995. 10.1002/j.1099-1174.1995.tb00083.x.

[5] Chow, J. C. K. (2017). *Analysis of Financial Credit Risk Using Machine Learning*. *April*. https://doi.org/10.13140/RG.2.2.30242.53449

[6] Bazarbash, M. (2019). FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. *IMF Working Papers*, *19*(109), 1. https://doi.org/10.5089/9781498314428.001

[7] Adekunle, O., Alalade, S. Y., &Agbatogun, T. (2015). Credit Risk Management and Financial Performance of Selected Commercial Banks in Nigeria. *Journal of Economic & Financial Studies*, *3*(01), 01. https://doi.org/10.18533/jefs.v3i01.73

[8] Bart, V. L. (2017). Machine learning: A revolution in risk management and compliance? *Journal of Financial Transformation*, *45*, 60–67. https://ideas.repec.org/a/ris/jofitr/1592.html

[9] Addo, P. M., Guegan, D., &Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks*, *6*(2), 1–20. https://doi.org/10.3390/risks6020038

[10] Leo, M., Sharma, S., &Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1). https://doi.org/10.3390/risks7010029

[11] Pandey, T. N., Jagadev, A. K., Mohapatra, S. K., &Dehuri, S. (2018). Credit risk analysis using machine learning classifiers. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017*, *August*, 1850–1854. https://doi.org/10.1109/ICECDS.2017.8389769

[12] Förstner, Wolfgang & Wrobel, Bernhard. (2016). Photogrammetric Computer Vision. 10.1007/978-3-319-11550-4.[13] Berg, M.T. & Kreveld, M.J. & Overmars, M.H.. (2008). Computational Geometry: Algorithms and Applications. 10.1007/978-3-540-77974-2.

[14] Rasmussen, C. & Williams, C. (2006). Gaussian Process for Machine Learning.

[15] Bishop, Christopher. (2006). Pattern Recognition and Machine Learning. 10.1117/1.2819119.

[16] Freund, Y. & Schapire, R. (1997), 'A decision theoretic generalization of on-line learning and an application to boosting', Journal of Computer and System Sciences 55(1), 119–139.

[17] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, *34*(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001