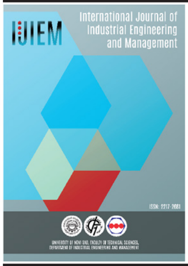
ISSN 2683-345X
journal homepage: http://ijiemjournal.uns.ac.rs/



International Journal of Industrial Engineering and Management

Volume 12 / No 1 / March 2021 / 1 - 13

Original research article

Predicting Delinquency on Mortgage Loans: An Exhaustive Parametric Comparison of Machine Learning Techniques

S. E. Azhar Ali [a*], S. S. H. Rizvi [b], F. Lai [a], R. Faizan Ali [c], A. Ali Jan [a]

[a] *Department of Management & Humanities, Universiti Teknologi PETRONAS, Tronoh, Malaysia;*

[b] *Department of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science & Technology, Karachi, Pakistan;*

[c] *Department of Computer Science, Universiti Teknologi PETRONAS, Tronoh, Malaysia*

ABSTRACT

This paper explores the potential of 19 machine learning techniques to model and forecasts the risk of delinquency on mortgage loans. These techniques include variants of artificial neural networks (ANN), ensemble classifiers, support vector machine, K-nearest neighbors, and decision trees. ensemble classifiers variants. Our dataset comprises 14,062 mortgage loans that have been approved by bank underwriters in the US. We find that Multi-Layer Perceptron (MLP), a variant of ANN, outperforms all other techniques in training time and the precision for testing and training. We have also compared Artificial Neural Network-Multilayer Perceptron (ANN-MLP) results with the traditional binary logistic regression technique's findings. The comparison shows that the ANN-MLP behaves better than the binary logistic regression technique. The study suggests that ANN-MLP could be a valuable extension towards developing the existing toolkit, banks and regulators have to predict delinquency risk on mortgage loans.

ARTICLE INFO

Article history:

Received October 12, 2020
Revised December 23, 2020
Accepted December 24, 2020
Published online January 27, 2021

Keywords:
Delinquency risk prediction;
Mortgage loan;
Machine learning techniques;
Artificial neural networks;
Multi-layer perceptron;
Binary logistics regression

*Corresponding author:
Syed Emad Azhar Ali
emadazhar2018@gmail.com

1. Introduction

The key challenges that financial institutions face are delinquency risks (i.e., failure to pay the debt required), and the sudden increase in this risk in previous years has been at the center of all crises in financial institutions. Consequently, delinquency risk management is a core constituent of the risk management practices of financial institutions. In recent decades, risk management methods for customer loans have become increasingly model-driven. Financial institutions generally use credit rating models to determine the probability of consumer loan delinquency or default, such as mortgages, car loans, credit cards, or personal loans [1]. Today, in the financial sector, most consumer loan decisions are automated based on model performance. These models are employed by financial institutions to evaluate the creditworthiness of borrowers and to decide if a loan should be issued to the borrower and, if so, how to develop loan contracts. With that in mind, the delinquency risk models are meticulously associated with the bank's lending decisions as they have an apparent operational profitability effect on the portfolios to which the models are applied. Thus, developing the proper delinquency assessment and refining of

Published by the University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia.

DOI: http://doi.org/10.24867/IJIEM-2021-1-272

these models has become a critical risk management feature in all financial institutions. However, even a small improved predictive capacity of such models can dramatically increase the profitability of a financial institution. A small improvement in the output models can be related to a significant change in the business strategy of a financial institution.

It is not surprising given this context that many scholars have demonstrated significant interest in these models. Logistic regression is a well-established technique for such modeling. Starting in the 1990s, scholars have begun to employ machine learning (ML) techniques to predict consumer credit risk [2-9]. Financial institutions are already looking forward to using ML techniques. They have been experimenting and have begun to incorporate ML approaches for organizational decisions in certain situations. In the case of the mortgage market, the outstanding balance is much higher compared to other loan categories, and the mortgage loans are a large part of financial institution operations. Nevertheless, few studies document delinquency models for loans than in the credit literature for credit cards or personal lending.

This paper attempts to advance the literature on mortgage loans through exhaustive parametric comparison of various ML techniques. Few studies discussed using ML techniques on mortgage models, such as [10] and [11]. We contribute in two ways to literature. First, we have explored the performance of 19 ML techniques in this paper to predict delinquency. Compared to articles that employ ML approaches to predict mortgage delinquency in current literature, such as[10-12], our investigation is more comprehensive, exploring the maximum number of ML techniques. In literature, the comprehensive performance evaluation of well-known machine learning classifiers on mortgage loans was not conclusive. Thus, in this paper, a response of well set up machine learning classifiers on the said mortgage loans dataset is given. We have analyzed variants of ML algorithms, namely artificial neural networks, Ensemble Classifiers, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble Classifiers. Under neural networks, we have used a multi-layer perceptron (MLP), which has an excellent capability for learning problems and adapting to unseen situations as per the previous learning. This is achieved through a backpropagation algorithm [13]. Furthermore, we have used three variants of the Decision Tree Algorithm, four variants of the SVM Algorithm, six variants of KNN, and five ensemble classifiers variants.

Second, we have performed an exhaustive analysis of ML techniques, which was not explored previously in the literature. Using MATLAB 2018, we have explored the operational outcomes under each technique, such as training precision, testing precision, training time, and prediction speed. All these outcomes have significant implications for academicians and managers.

Third, we have examined and compared statistical and neural network methods for forecasting. A statistical model is based on consumer behavior seen in historical data. However, such behavior may change in the future, and the historical transformation of a sub-population may not be generalized at such times. In contrast to the statistical models, neural networks may propose a stimulating alternative due to their general estimation characteristic and because no standard hypothesis is required [14]. Therefore, we have also compared the outcomes obtained by neural networks (i.e., through MATLAB 2018) with the results of those statistical methods such as binary logistic regression (i.e., through SPSS 25). Binary logistic regression is one advocated technique for foretelling the possible delinquency concerning loans. These methods have gained popularity due to their ease of use, user-friendly interface, and ease of understanding. Regardless of the user-friendliness feature of statistical software, it is known that the backpropagation MLP can perform non-linear regressions [15]. By using the mortgage loan dataset in this study, it has been demonstrated that the MLP outclasses all other machine learning techniques for classification or prediction and including those using statistical techniques [16].

The remainder of the paper is structured as follows. Section 2 discusses the relevant literature. Whereas, Section 3 describes the dataset and the methodology of training and testing the dataset. Section 4 presents the empirical results. The last section presents a summary of the overall paper.

2. Literature Review

Many researchers have studied the impacts of different factors on financial stress measures, such as delinquency rate and bankruptcy, and most have used mathematical methods to build their models. For example, the logistic regression model [17-20]. However, these conventional approaches also put strong assumptions on the mechanism of producing data and on the linear relationship between output and input variables. They also have disadvantages when it comes to outliers, non-linear relations, and variable choice. They hardly speak about situations under which independent variables are highly correlated, leading to significant variances and substantial predictive errors for the model [21].

Studies have shown that machine learning methods can help with these methodological issues. Reference [22] claims that machine learning can usually enhance a variety of econometric methods by choosing different methods, in particular with datasets with several variables. Besides, machine learning can enhance the inference and adjust with a less targeted approach to the current statistical system for rough survey answers and variables, including missing observations [23]. To assess the causal effect, [24] create a causal forest to estimate heterogeneous treatment effects based on a random forest algorithm, a machine learning technique. In particular, machine learning approaches have shown that they are flexible with some common data issues to boost prediction over traditional methods in the sense of financial crisis prediction. Reference [25] reviews the US subprime mortgage crisis literature and suggests that analytical solutions must be built to help predict and interpret defaults and crises, including machine learning methods.

Reference [26] employs a unique neural network approach to classify commercial bank failures in the United States and show that their model outperforms conventional bankruptcy prediction models at a 96% accuracy rate. However, research findings on the credit risk prediction of mortgage loans are scarce when it comes to secured lending, notwithstanding the element that they are among the major asset groups on the balance sheets of European banks. This paper aims to determine, using data on the level of real-world mortgage loans, whether the selection of these newer methods can provide improved predictive efficiency over more developed methods such as logistic regression (LR).

A range of goals are achieved by assessing and comparing the efficacy of various mortgage default prediction techniques. First, financial institutions are interested in identifying a lender's attributes as excellent and poor for profitability and credit risk management purposes. This is the fundamental focus of credit risk assessment. [27]. Second, to allow banks to cope with unexpected losses over their planned losses, adequate regulatory capital buffers are needed. To determine regulatory capital requirements, precise estimation of the probability or possibility of default on mortgage loans is crucial. The Probability of Default (PD) models built for this reason are generally set in time (one year) for retail credit risk groups such as mortgages and have so far been traditionally modeled using logistic regression; being able to build more precise models will allow more acceptable levels of capital to be set.

A variety of credit risk applications have led to advance in statistical and machine learning methods of classification. Various modeling techniques and analytical experiments can be witnessed from studies [5, 10, 28-31]. Some of their findings indicate that novel methods such as ensemble classifiers provide some enhancement over logistic regression in a predictive capacity that could prove useful for credit risk management. The proposed performance improvement, however, is not guaranteed; novel techniques could not dramatically boost predictive performance on specific datasets [32]. This implies that to decide whether and where this is the case, empirical work is required. We explore the performance of 19 ML techniques in this study to assess delinquency prediction accuracy on mortgage loans. We have employed Multi-Layer Perceptron (MLP) as a variant of Artificial Neural Networks (ANN) [33]. Support Vector Machine (SVM) [34] method was used with four variants: coarse gaussian, medium gaussian, fine gaussian SVM, and cubic SVM. K-Nearest Neighbors (KNN) [35] with six variants: weighted KNN, cubic KNN, cosine KNN, coarse KNN, medium KNN, and fine KNN. Decision Tree was used with three variants: coarse tree, fine tree and meduim tree. For ensembled classifiers, we investigate five homogenous ensemble methods– RUSBoosted Trees, subspace KNN, subspace discriminant, bagged trees, and boosted trees. It was found out the MLP as a variant of ANN was the best candidate for predicting delinquency on mortgage loans, and it outperforms all other techniques of ML. Lastly, the results from the most optimum ML method, that is, ANN-MLP, is matched with the results of traditional BLR to analyze the difference in performance.

The studies closely linked to this paper are [11, 12, 36] all predicting mortgage delinquency through ML methods. Reference [12] investigated early models for delinquency. On the other hand, [11] examine default rates of one-year. However, both studies were concentrated on ordering ranks without taking predictive precision into account and risk classification. Mortgage risk management models for a seasoned mortgage portfolio were also employed by [36] to apply for reserve forecasting, capital measurement, stress testing. Information on past loan performance is among the risk drivers in the [36] models, and the prediction of loan status in the following months can significantly improve this information. Nevertheless, information on past loan performance is not available at some levels of credit risk management, For example, at credit origination. Our analysis is based on a publicly available dataset provided by the University of Tennessee [37, 38]. To the best of our knowledge, the current study conducted an exhaustive parametric comparison of ML techniques that had not been performed in previous studies. We also argue that the findings of the techniques used in this study are more convincing than those examined in previous studies.

3. Data and Methodology

3.1 Data description

The current study has used the data of 14,062 mortgage loans approved by Bank underwriters from the US. The data has been extracted from the datasets provided by the University of Tennessee [37, 38] for data mining and business intelligence. Few other studies have also used a similar data source for their data analysis [39, 40]. From Table 1, our dataset has 11 variables in total: borrower's age, original loan amount, the ratio of loan to the home purchase price, borrower's credit score, first time home buyer, borrower's monthly debt expense, borrower's monthly income, the purchase price of a house, the appraised value of a home at origination and, finally, the outcome variable, that is, loan as delinquent or non-delinquent. The input variables in our dataset has been extensively used in previous studies to predict the outcome of mortgage loans. Our outcome variable is equal to '1', if default, and equal to '0' otherwise.

Although nonparametric approaches are more multifunctional than parametric in terms of explanatory variable forms, they are still based on a set of pre-defined variables. The benefit of ML techniques over conventional nonparametric techniques is their capability to learn multifaceted structures and relations that have not previously been specified. In this article, we used a supervised ML algorithm. This algorithm consists of a goal/result variable from which a specific predictor (or dependent variable) is predicted (independent variables). We create a function that maps inputs to the output layer using this set of variables. The training process will continue until the model correctly collects the training data. In this paper, we have focused on supervised learning of ML algorithms: Decision Tree, Random Forest, KNN, Ensemble Classifier, and Artificial Neural Networks.

3.2 Training the Dataset

We use MATLAB 2018 to train the ML techniques under investigation in this paper. For neural networks, we use the MLP Classifier () function, and the hyper-parameters are layer number, node size, and learning rate. To train the linear SVM model, we use the Linear SVC () function, and the main tuning parameter is the penalty term. We use the KNeighbors Classifier () function for K-Nearest Neighbors to find the optimum number of neighbors and distance

Table 1. Comparison of the characteristics and performances achieved by the two developed predictive models

Variable	Explanation	Type
Bo_Age	Borrower age	Continuous
Ln_Orig	Value of loan	Continuous
Orig_LTV_Ratio_Pct	The ratio of loan to the home purchase price	Continuous
Credit_score	Borrower's credit score	Continuous
First_home	First time home buyer? (Yes=1, No=0)	Categorical
Tot_mthly_debt_exp	Borrower's total monthly debt expense	Continuous
Tot_mthly_incm	Borrower's total monthly income	Continuous
orig_apprd_val_amt	The appraised value of a home at origination	Continuous
pur_prc_amt	The purchase price for a house	Continuous
DTI_ratio	Borrower debt to income ratio (Tot_mthly_debt_exp/Tot_mthly_incm)	Continuous
Deilinquency_outcome	Binary version of "outcome" (either default=1 or non-default=0)	Categorical

metrics. In comparison, ensemble classifier learning generates a population of essential learners from the training data first; it then integrates them into a composite model [21].

In-sample forecasting, where forecasting precision is calculated using the same data to fit the model, results in overly optimistic estimates of accuracy. When a model collects much noisy information, known as overfitting, it gives inaccurate predictive results for other information. Thus, our study was based on out-of-sample rather than sample predictions break the original dataset into a collection of "training data" to match the model and then apply the trained model to the "test data" in particular to assess its predictive accuracy. We randomly allocate training findings (70%) and test data to the training (70%) (30 percent of the observations).

For a binary variable, each model's objective is to generate a loan-level prediction; Y=1 means default, and Y=0 implies no default. N observations of training data with p predictor variables are used to make this prediction. A predictor vector (x_i) and a related response (y_i = 1 or 0) consist of each observation (x_i,y_i), i=1,...,n). A combination of continuous and categorical variables are predictive factors. We define default as greater than 90 days arrears.

We measure predictive accuracy utilizing measures derived from the confusion matrix, accounting for false positive (C_FP) and false negative (C_FN) non-negative costs. We assume that zero costs are true-positive and true-negative. The threshold for positive or negative classification is defined as the threshold: (C_FP x Neg) / (C_FN x Pos + C_FP x Neg), where the actual number of negatives and positives from the training sample is Neg and Pos. The threshold decreases in the same sample as the cost of false-negative (C-FN) becomes higher relative to false positive (C_FP). Two measures are then described as follows.

$$Precision = \frac{\#truepositives}{\#true\ positives + \#false\ positives}$$

$$Recall = \frac{\#true\ positives}{\#true\ positives + \#false\ negatives}$$

Higher accuracy is correlated with a low number of false positives (negative) (recall). In other words, the 'precision' typically composed how well the model can capture the true positives, and the 'recall' measure calculates how well the model captures the actual negatives. Total false classification is the third measure from the confusion matrix, defined as TFC = FP + FN *c, where c = C_FN/C_FP. Given a specific set of C_FP and C_FN values, this calculation measures how many incorrect classifications the system produces. The higher this metric is, the more a system yields incorrect classifications.

Among all the machine learning techniques, neural networks are advocated as one of the most efficient machine learning techniques. An artificial neural network is an influential data modeling instrument capable of capturing and symbolizing problematic input/output associations. A Neural Network is a beneficial technique for processing information as it works similarly to a biological nervous system, and also as an information processing technique as it works like a human brain and processes the information accordingly [41]. It can be employed for issues that do not have an algorithmic solution. For all-purposes, neural networks are well suited to problems that humans are good at solving but for which computers usually are not. These issues comprise pattern recognition and forecasting, which require the identification of tendencies in data. Our results exhibited that MLP as a variant of ANN was most efficient and most influential in predicting delinquency on mortgage loans. Therefore, the next section will elaborate in detail, the training and testing process of MLP.

3.2.1 Training Multi-layer Perceptron through Back Propagation Network

The MLP network is the most famous variant of the Artificial Neural Network (ANN), and it is trained using an error backpropagation algorithm [13]. This kind of neural network is famous as a direct network as it needs a targeted yield to be trained. This kind of system aims to form a prototype that appropriately plans the input to the output by using the past data. It is also used to predict the output when applying past data so that the formed function can then be used to produce the output when the anticipated output is not known. Such a type of perceptron has several hidden levels of computational neurons. The input levels are propagated in a feed-forward direction on a level by level basis. In an MLP, the backpropagation learning method is most commonly employed, which has two stages. In the first stage, a prototype for the training input is offered to the network input level. So far, the output prototype is produced by the output level, and the network will keep transmitting the input prototype from level to level.

An inaccuracy will be calculated if this prototype is different from the required output. The prototype will be progressed backward through the network from the output level to the input level. Due to this inaccuracy, the weights will be modified. The public figure for the working of an MLP is shown below.

In Figure 1, $x_1, x_2, \ldots, x_n$ represent the input signal. Feed-forward is used to send the input to the hidden layer with a link weight of w_{ij}. This input and invisible layer weight help compute the output y_j at the hidden layer using Equation 1. In the next step of the feed-forward, the hidden layer output y_j is combined with the output layer link weight of w_{jk}, which forms the output layer using Equation 2.

$$y_j(p) = Sigmoid\left[\sum_{i=0}^{n} x_i(p).w_{ij}(p) - \theta_j\right] \quad (1)$$

In Equation 1, y_j is the output of hidden layer of pth neuron, sigmoid is the activation function, x_i is the input of the p^{th} neuron of the input layer, w_{ij} is the link weight of the hidden layer, θ_j is the threshold for the hidden layer neuron.

$$y_{k(p)} = \text{sigmoid}\left[\sum_{j=1}^{m} x_{jk}(p).w_{jk}(p) - \theta_k\right] \quad (2)$$

Replacing w_{jk} with y_j in Equation 2 in which y_k is the output of the output layer of p^{th} neuron, w_{jk} is the link weight of the output layer and θ_k is the threshold for the hidden layer neuron. The above discussion refers to the feed-forward phase, which gives the output y_k. The next step of ANN training is the backpropagation. In backpropagation, the error signal back propagates to the input layer to update the hidden layer and output layer weight. In the subsequent section, a brief description of the backpropagation learning method is presented.

3.2.1.1 Backpropagation Algorithm

Back Propagation is an algorithm that trains an MLP and many other neural networks. The input data is recurrently offered to the neural network using backpropagation. With each exhibition, the neural network's output is compared to the anticipated output, and an error is calculated. Feedback is provided to the neural network, which is applied to regulate the weights in order to reduce the errors with each training cycle, and the neural model becomes nearer and nearer to forming the anticipated output in a process called “training.”

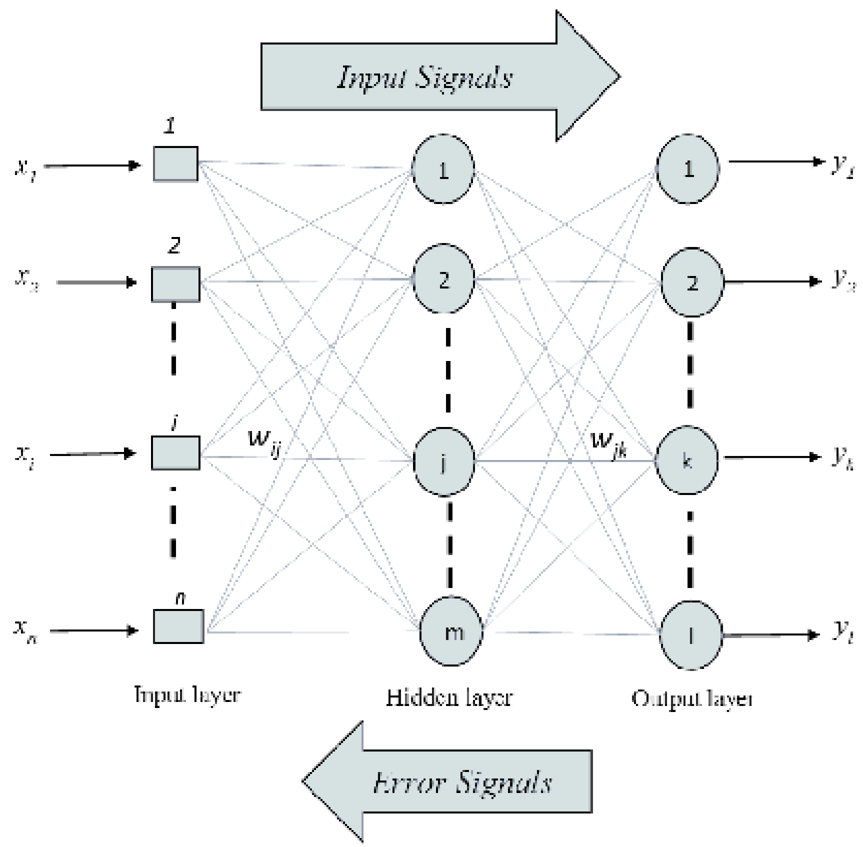


Figure 1. MLP with one hidden layer and backpropagation learning method

The weights of the backpropagation network are updated, which will backward progress the errors associated with the output neurons. Therefore, the slope of the error is computed for the neurons in the output level using Equation 3:

$$e_k(p) = y_{d.k}(p) - y_k(p) \tag{3}$$

Where e_k (p) is the error of p_{th} neuron, y_{dk} (n) is the required output of p^{th} neuron on the output layer, and y_k is the actual output of the ANN. The weight corrections are then calculated, and the newly updated weights applied using Equation 4 and Equation 5, respectively.

$$\Delta w_{jk}(p) = \alpha . y_i(p)\delta_k(p) \tag{4}$$

and

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \tag{5}$$

Where w_{jk} (p+1) is the new updated weight, w_{kj} (P) is the previous weight, alpha = learning rate, sigma is the error gradient of the output layer. Similarly, in the next stage of backpropagation, the error gradient and weight correction are calculated and newly updated for the neurons in the hidden level using Equation 6, Equation 7, and Equation 8, respectively.

$$\delta_i(p) = y(p).[1 - y_i(p)].\Sigma_{k=1}^{i} \delta_k(p) w_{jk}(p) \tag{6}$$

$$\Delta w_{ij}(p) = \alpha . x_i(p).\delta_i(p) \tag{7}$$

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \tag{8}$$

Where w_{ij} (p+1) is the new updated weight, w_{ik} (p) is the previous weight and alpha as the learning rate, and sigma is the error gradient of the hidden layer.

In order to have a thorough analysis of delinquency prediction on the given dataset, we have compared the results attained through ANN-BPN with variants of Ensemble Classifiers, Decision Tree (SVM), (KNN), and Ensemble Classifiers. In this work, a comprehensive empirical comparison of 19 well-known classifiers for loan delinquency prediction has been demonstrated. Moreover, each algorithm's operation was evaluated as a part of training & testing accuracy, prediction speed, and training time. Also, the class-wise performance of each algorithm is presented as a separate confusion matrix.

3.3 Binary Logistic Regression

Like the ordinary least square (OLS), binary logistic regression (BLR) is a classical forecasting technique. Nevertheless, BLR forecasts a dichotomous outcome that violates the homoscedasticity principle of OLS. In logistic regression, an algebraic transformation is necessary to arrive at the traditional linear regression function. Mathematically, logistic regression represents the natural logarithm of an odds ratio (i.e., the logit).

$$Logit(Y) = \ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k \tag{9}$$

$$P = \frac{e^{\alpha + \beta_1 X_1 + \ldots + \beta_k X_k}}{1 + e^{\alpha + \beta_1 X_1 + \ldots + \beta_k X_k}} \tag{10}$$

Where P is the probability of delinquency, α is the Y-intercept, βs are regression coefficients, and X_n is a set of predictors. Unlike ML techniques discussed in this paper, the results from BLR are extracted from the use of statistical software (i.e., SPSS).

4. Simulation Results and Analysis

The current section exhibits the results revealed after the application of different machine learning methods. The performance of each method has been recorded as a function of training time, prediction speed, training, and testing precision. Besides, the class-wise performance of each algorithm is presented as a separate confusion matrix in Table 2. Lastly, the results of the most efficient ML method are compared with the results of BLR, which is one of the most common techniques in social sciences to predict a dichotomous outcome.

The detailed results from confusion matrices are tabulated in Table 2, whereas the graphical results are illustrated in Figures 2, 3, and 4. Table 2 presents a much-compounded view of simulation results demonstrating the performance of each classifier. The first scale of valuation is the training accuracy. From Figure 2, most of the machine learning variants exhibit higher training accuracy of over 97% other than Boosted Trees and RUS Boosted Trees.

4.1 Confusion matrix results

However, the simulation results are quite different for our second scale, that is, testing accuracy. Again, from Figure 2, ANN, Subspace KNN, and Cosine KNN were the only variants exhibiting an accuracy level above 90%. Besides, the accuracy level for ANN is substantially ahead of all, that is, 97%.Our third measure was prediction speed (observations/sec). It is a well-known fact that the prediction speed of any classifiers highly depends on the configuration of computer system resources like processing speed and RAM. In this study, all the simulations were carried on a Core2Quard System @ 2. 5GhZ with 6GB of RAM. In terms of prediction speed, the variants of RUSBoosted Trees, Boosted Trees, demonstrates a higher prediction speed followed by ANN-BPN (Figure 3). for a high-performance computing machine. Therefore, at this point, accuracy is It is explicitly mentioned that the prediction speed could be ascended by opting the key parameter. ANN-BPN depicts a lower prediction speed (700,000 observations/second) compared to RUSBoosted Trees (1400,000 observations/second) and Boosted Trees (10,00,000 observations/second). However, testing and training accuracy for ANN-BPN was observed as 97%, which is much higher compared to 40% accuracy for variants of RUSBoosted Tree and Boosted Trees.

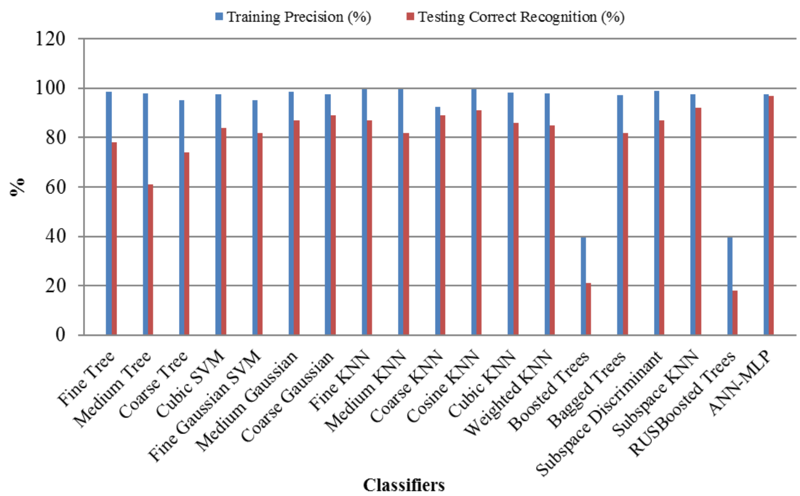


Figure 2. Training precision and correct testing recognition

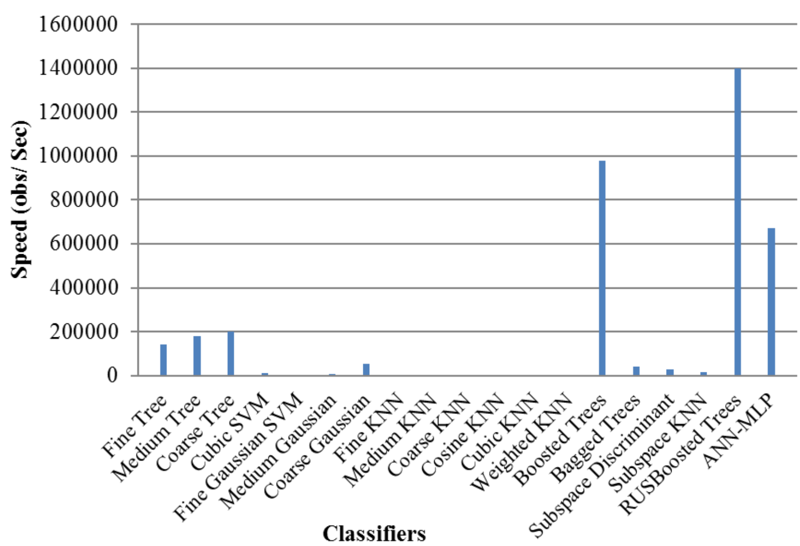


Figure 3. Prediction Speed (obs/sec) for each ML technique

Table 2. Confusion matrix for variants of ML techniques

	Algo	Confusion Matrix		True Class 0	True Class 1	Training Precision (%)	Testing Correct Recognition (%)	Prediction Speed (obs/sec)	Training Time (machine cycles)
Decision Tree	Fine Tree	Predicted Class	0	98	2	98.5	78	140000	40.844
			1	1	99				
	Medium Tree	Predicted Class	0	96	4	98.1	61	180000	35.598
			1	0	100				
	Coarse Tree	Predicted Class	0	92	8	95.3	74	200000	62.241
			1	2	98				
Support Vector Machine	Cubic SVM	Predicted Class	0	96	4	97.5	84	13000	200.897
			1	2	98				
	Fine Gaussian SVM	Predicted Class	0	93	7	95.2	82	2400	189.83
			1	3	97				
	Medium Gauss-ian	Predicted Class	0	98	2	98.5	87	5700	189.17
			1	1	99				
	Coarse Gauss-ian	Predicted Class	0	97	3	97.5	89	55000	195.36
			1	2	98				
KNN	Fine KNN	Predicted Class	0	99	1	99.5	87	3000	194.96
			1	0	100				
	Medium KNN	Predicted Class	0	100	0	99.5	82	2900	194.42
			1	1	99				
	Coarse KNN	Predicted Class	0	100	0	92.5	89	2200	195.5
			1	6	94				
	Cosine KNN	Predicted Class	0	100	0	99.6	91	2800	195.26
			1	1	99				
	Cubic KNN	Predicted Class	0	98	2	98.2	86	92	501.46
			1	1	99				
	Weighted KNN	Predicted Class	0	98	2	98.1	85	3000	253.41
			1	2	98				
Ensemble Classifiers	Boosted Trees	Predicted Class	0	0	100	39.5	21	980000	252.78
			1	21	79				
	Bagged Trees	Predicted Class	0	97	3	97.2	82	43000	251.03
			1	2	98				
	Subspace Discrimi-nant	Predicted Class	0	99	1	99.1	87	28000	250.54
			1	1	99				
	Sub-space KNN	Predicted Class	0	98	2	97.5	92	15000	250.26
			1	3	97				
	RUS Boosted Trees	Predicted Class	0	0	100	39.5	18	1400000	249.48
			1	21	79				
Artificial Neural Network	ANN-MLP	Predicted Class	0	96	4	97.5	97	671000	27.124
			1	1	99				

The outstanding results for ANN-BPN were also witnessed in terms of the Training Time (Figure 4). Therefore, we can advocate that ANN-BPN is the most optimum classifier for predicting delinquency on mortgage loans as the simulation results depict a higher accuracy for training (97.5%) and testing (97%) along with minimum utilization of training time. The next section will further discuss the results attained through ANN-BPN and a comparison of results through logistic regression by statistical methods.

4.2 Results with Multi-layer Perceptron

Considering the outstanding results observed from ANN-MLP, we have explored an in-depth analysis of those results. Table 3 portrays the accuracy prediction percentage of selected and unselected cases. Some 70 % of the dataset had selected instances that came out with a 96.9 % correct prediction, while 30 % of the dataset was unselected cases predicted as 96.5 % as the right outcome.

The data was then supplied to the Multi-layer Perceptron Neural Network (MLPNN). The network consisted of 11 input nodes and one output node. The hidden layer consisted of 25 nodes. The activation function was log sigmoid. The learning rate was set to 0.1. The learning was performed on 70 % of the data for 50 epochs. The weights of the MLPNN were updated using the Cost Function. The trained MLPNN was then tested with the remaining 30 % data observations new to the MLPNN. The MLPNN succeeded in achieving 98% accuracy in predicting delinquency. Figure 1 shows that the error was minimized to approximately 0 after 15 Iterations. This shows that ANN-MLP can be used with confidence to predict delinquency based on the data collected in the loan application. Figure 5 depicts that the first iteration had 75 errors, which reduced gradually over the subsequent iterations. This occurred due to the continuous training provided and the learning gained thereof by the said training. As a result, the error rate declined gradually by the 50th training cycle.

4.3 Comparison of MLP Results with Binary Logistic Regression

The results below demonstrate the comparison between the performance evaluation of an ANN-MLP and BLR. The outputs for BLR were produced by applying the Statistical Package for the Social Sci-

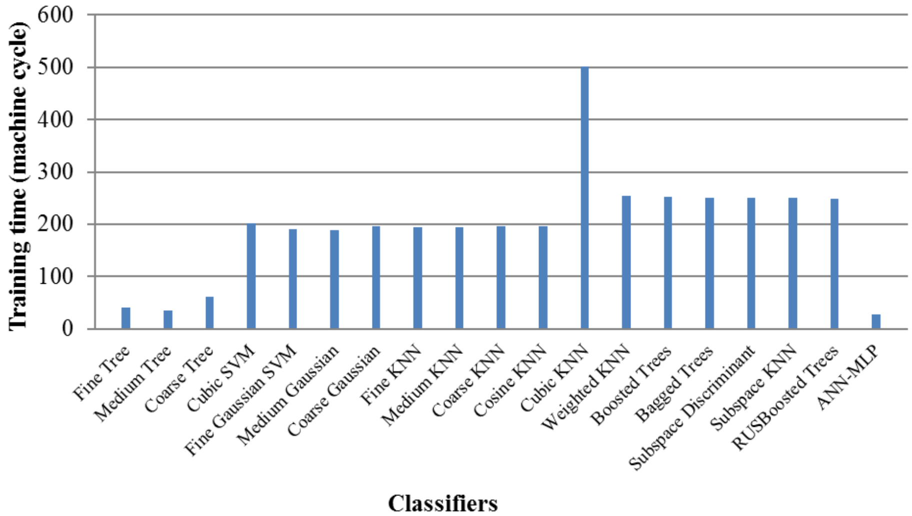


Figure 4. Training time for each ML technique

Table 3. Classification Table

	Predicted (selected cases)			Predicted (unselected cases)		
Outcome	Delinquency	Non-delinquency	Percentage correct	Delinquency	Non-delinquency	Percentage correct
Delinquency	1838	265	87.4 %	763	134	85.1 %
Non-delinquency	36	7686	99.5 %	15	3325	99.6 %
Overall Percentage			96.9 %			96.5 %

ences (SPSS 25). At the same time, the results for ANN-MLP were produced through MATLAB 2018.

A detailed analysis of the results in Table 4 revealed that the variables ‘borrower’s age,’ ‘original loan amount,’ and ‘purchase price’ were declared ‘insignificant variables’ by SPSS with a significance value higher than 0.05. This was because the data might have been built so that logistic analysis has misled to this conclusion. However, it was evident that ‘age’ was one of the critical factors in lending. Similarly, ‘original loan amount’ and ‘purchase price’ were also factors that must not be ignored when determining the lending amount. Therefore, the accuracy of prediction may be considered flawed. The Table 5 depicts a quick comparison of results obtained through both methods. That is ANN-MLP through MATLAB 2018 and BLR through SPSS 25.

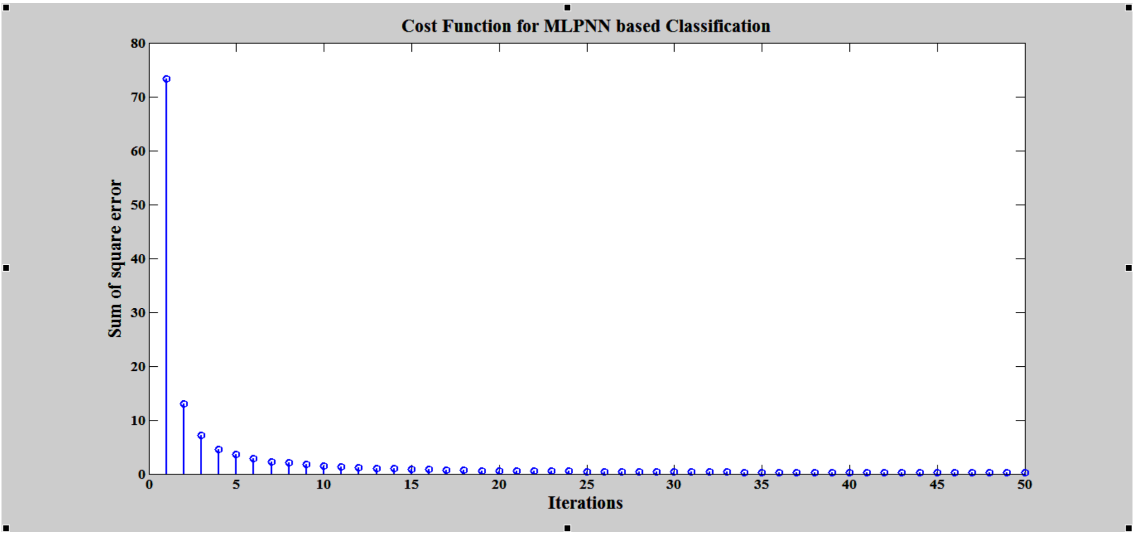


Figure 5. Error plot obtained by applying ANN-MLP

Table 4. Detailed results of BLR obtained through SPPS

	B	SE.	Wald	df	Sig.	Exp(B)
Borrower’s age	-0.007	0.005	1.952	1	0.162	0.993
Value of loan, USD	0.000	0.000	3.498	1	0.061	1.000
Ratio of loan to home purchase price	-0.026	0.007	16.021	1	0.000*	0.974
Borrower's credit score	0.009	0.001	125.007	1	0.000*	1.009
First time home buyer? (Y/N)	-0.260	0.121	4.601	1	0.032*	0.771
Borrower's total monthly debt expense	-0.003	0.000	1281.501	1	0.000*	0.997
Borrower's total monthly income	0.001	0.000	692.230	1	0.000*	1.001
Appraised value of home at origination	0.000	0.000	5.747	1	0.017*	1.000
Purchase price for house	0.000	0.000	0.008	1	0.930	1.000
Borrower debt to income ratio	19.522	0.388	2532.798	1	0.000*	3.008E8
Current loan status	-6.482	0.913	50.369	1	0.000*	0.002

Note: * indicates statistical significance at the 5 % level

Table 5. Comparison of techniques applied

Techniques	Accuracy	Ignored Variables
1. Binary Logistic Regression analysis	95%	i) Borrower’s age ii) Original Loan Amount iii) Purchase Price of Home
2. ANN-MLP based prediction	98 %	None

However, a detailed analysis of the results in Table 2 revealed that the variables 'borrower's age,' 'original loan amount,' and 'purchase price' were declared 'insignificant variables' by SPSS. with a significance value higher than 0.05. This was because the data might have been built so that logistic analysis has misled to this conclusion. However, it was evident that 'age' was one of the critical factors in lending. Similarly, 'original loan amount' and 'purchase price' were also factors that must not be ignored when determining the lending amount. Therefore, the accuracy of prediction may be considered flawed.

The comparison in Table 5 shows that the ANN-MLP performed better than the Binary Logistic Regression. Moreover, the MLPNN technique could predict accurately with 98% confidence. The results suggest that the MLPNN technique could be developed and used by the lending agencies to predict delinquency among the loan applicants.

5. Summary and Discussion

This study has performed an exhaustive parametric comparison of different machine learning techniques to predict delinquency on mortgage loans. Compared to existing literature papers that apply ML methods to cross-sectional model mortgage delinquency, such as [10, 11], our investigation is extensive as we have applied, in total, 19 different techniques to predict delinquency on mortgage loans. In this paper, we find that ANN-MLP outperforms other default prediction techniques in terms of training time, training, and testing precision. The ANN-MLP. has been trained and updated using the backpropagation algorithm. The non-linear regression capability of the ANN-MLP showed 98% accuracy in predicting delinquency. The ANN-MLP behaved better than other machine learning techniques and traditional binary logistic regression techniques and is very promising for such an analysis. Nevertheless, the prediction speed was slower for ANN-MLP compared to the boosted trees technique, but the issue can be mitigated by using a machine with a better configuration. It was found that ANN-MLP has been the most efficient and effective technique to predict our outcome. The MLPN.N. has been trained and updated using the backpropagation algorithm. The non-linear regression capability of the ANN-MLP showed 98% accuracy in predicting delinquency. The ANN-MLP behaved better than other machine learning techniques and traditional binary logistic regression techniques and is very promising for such an analysis. The study, therefore, contributes to the literature by advocating that a model based on ANN predict loan delinquency with higher efficiency and effectiveness. Likewise, similar ANNs can be applied to solve various other business problems.

This research also suggests that financial institutions should regularly revise their credit risk assessment frameworks, especially in the case of developing countries, as they face a more unstable macroeconomic environment as compared to developed countries [42]. Moreover, it might be thought-provoking to match the credit scoring framework of financial institutions during different growth phases.

The current study can provide meaningful insight for future studies by integrating other delinquency elements into the ANN methodology for predicting the probability of delinquency on a mortgage. For instance, researchers in the future can incorporate collateral determinants and bank-borrower relationships while predicting the delinquency on bank loans. The integration of these elements into the delinquency prediction model can significantly contribute to methodology and practice. This study would help establish a more improved classifier for the dataset by providing the best classical candidate. Moreover, the same inference could also be validated on some other benchmark datasets of loans when available in the future.

Acknowledgments

The authors would also like to thank the reviewers for their valuable suggestions to enhance the manuscript. We would also like to thank the Department of Management & Humanities, Universiti Teknologi PETRONAS (UTP), Centre of Social Innovation (UTP) and Shaheed Zulfiqar Ali Bhutto Institute of Science &Technology, Karachi to facilitate this research study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] L. Thomas, J. Crook, and D. Edelman: 'Credit scoring and its applications' (SIAM, 2017. 2017)

[2] V. David, C. Fraboul, J. Rousselot, and P. Siron, Partitioning and mapping communication graphs on a modular reconfigurable parallel architecture, Parallel Processing: CONPAR 92—VAPP V, (Springer, 1992), pp. 43-48.

[3] V.S. Desai, J.N. Crook, and G.A. Overstreet Jr, A

comparison of neural networks and linear scoring models in the credit union environment, European journal of Operational Research 95 (1) (1996) pp. 24-37.
[4] J. Galindo, and P. Tamayo, Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications, Computational Economics 15 (1-2) (2000) pp. 107-143.
[5] J.N. Crook, D.B. Edelman, and L.C. Thomas, Recent developments in consumer credit risk assessment, European Journal of Operational Research 183 (3) (2007) pp. 1447-1465.
[6] A.E. Khandani, A.J. Kim, and A.W. Lo, Consumer credit-risk models via machine-learning algorithms, Journal of Banking Finance 34 (11) (2010) pp. 2767-2787.
[7] P.M. Addo, D. Guegan, and B. Hassani, Credit risk analysis using machine and deep learning models, Risks 6 (2) (2018) pp. 38.
[8] J. Sirignano, A. Sadhwani, and K. Giesecke, Deep learning for mortgage risk, arXiv preprint arXiv:.02470(2016)
[9] Y. Li, X. Wang, B. Djehiche, and X. Hu, Credit Scoring by Incorporating Dynamic Networked Information, European Journal of Operational Research(2020)
[10] S. Lessmann, B. Baesens, H.-V. Seow, and L.C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, European Journal of Operational Research 247 (1) (2015) pp. 124-136.
[11] T. Fitzpatrick, and C. Mues, An empirical comparison of classification algorithms for mortgage default prediction: Evidence from a distressed mortgage market, European Journal of Operational Research 249 (2) (2016) pp. 427-439.
[12] S. Chen, Z. Guo, and X. Zhao, Predicting mortgage early delinquency with machine learning methods, European Journal of Operational Research In press (August) (2020)
[13] A.A. Heidari, H. Faris, S. Mirjalili, I. Aljarah, and M. Mafarja, Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks, Nature-Inspired Optimizers, (Springer, 2020), pp. 23-46.
[14] D. Zhu, X. Cheng, F. Zhang, X. Yao, Y. Gao, and Y. Liu, Spatial interpolation using conditional generative adversarial neural networks, International Journal of Geographical In formation Science 34 (4) (2020) pp. 735-758.
[15] A.-K. Seghouane, and N. Shokouhi, Adaptive learning for robust radial basis function networks, Ieee Transactions on Cybernetics (November) (2019) pp. 1-10.
[16] A.A. Heidari, H. Faris, I. Aljarah, and S. Mirjalili, An efficient hybrid multilayer perceptron neural network with grasshopper optimization, Soft Computing 23 (17) (2019) pp. 7941-7958.
[17] Z. Khemais, D. Nesrine, and M. Mohamed, Credit scoring and default risk prediction: A comparative study between discriminant analysis & logistic regression, International Journal of Economics and Finance 8 (4) (2016) pp. 39.
[18] A.C. Bahnsen, D. Aouada, and B. Ottersten: 'Example-dependent cost-sensitive logistic regression for credit scoring', in Editor (Ed.)^(Eds.): 'Book Example-dependent cost-sensitive logistic regression for credit scoring' (IEEE, 2014, edn.), pp. 263-269
[19] A.A. Jan, M. Tahir, F.-W. Lai, A. Jan, M. Mehreen, and S. Hamad, Bankruptcy profile of the islamic banking industry: Evidence from Pakistan, Business Management and Strategy 10 (2) (2019) pp. 265-284.
[20] J. Papula, and J. Volna: 'Intellectual capital as value adding element in knowledge management', in Editor (Ed.)^(Eds.): 'Book Intellectual capital as value adding element in knowledge management' (2011, edn.), pp.
[21] J. Friedman, T. Hastie, and R. Tibshirani: 'The Elements of Statistical Learning' (Springer-Verlag New York, 2001.2001)
[22] S. Mullainathan, and J. Spiess, Machine learning: An applied econometric approach, Journal of Economic Perspectives 31 (2) (2017) pp. 87-106.
[23] J. Ifft, R. Kuhns, and K. Patrick, Can machine learning improve prediction–an application with farm survey data, International Food and Agribusiness Management Review 21 (1030-2019-611) (2018) pp. 1083-1098.
[24] S. Wager, and S. Athey, Estimation and inference of heterogeneous treatment effects using random forests, Journal of the American Statistical Association 113 (523) (2018) pp. 1228-1242.
[25] Y. Demyanyk, and I. Hasan, Financial crises and bank failures: A review of prediction methods, Omega 38 (5) (2010) pp. 315-324.
[26] F.J.L. Iturriaga, and I.P. Sanz, Bankruptcy visualization and prediction using neural networks: A study of US commercial banks, Expert Systems with Applications 42 (6) (2015) pp. 2857-2869.
[27] R.A. McDonald, A. Matuszyk, and L.C. Thomas, Application of survival analysis to cash flow modelling for mortgage products, OR Insight 23 (1) (2010) pp. 1-14.
[28] K. Kennedy, B.M. Namee, and S.J. Delany, Using semi-supervised classifiers for credit scoring, Journal of the Operational Research Society 64 (4) (2013) pp. 513-529.
[29] I. Brown, and C. Mues, An experimental comparison of classification algorithms for imbalanced credit scoring data sets, Expert Systems with Applications 39 (3) (2012) pp. 3446-3453.
[30] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, Benchmarking state-of-the-art classification algorithms for credit scoring, Journal of the Operational Research Society 54 (6) (2003) pp. 627-635.
[31] T. Bellotti, and J. Crook, Credit scoring with macroeconomic variables using survival analysis, Journal of the Operational Research Society 60 (12) (2009) pp. 1699-1707.
[32] D.J. Hand, Classifier technology and the illusion of progress, Statistical Science 21 (1) (2006) pp. 1-14.
[33] C.M. Bishop: 'Neural networks for pattern recognition' (Oxford University Press, 1995. 1995)
[34] C. Cortes, and V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) pp. 273-297.
[35] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician 46 (3) (1992) pp. 175-185.
[36] J. Sirignano, A. Sadhwani, and K. Giesecke, Deep learning for mortgage risk, Working Papers in Linguistics(2018)
[37] U.o. Tennessee: 'Data Mining for Business Intelligence', in Editor (Ed.)^(Eds.): 'Book Data Mining for Business Intelligence' (The University of Tennessee, 2011, edn.), pp.
[38] G. Shmueli, N.R. Patel, and P.C. Bruce: 'Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner' (John Wiley and Sons, 2011. 2011)
[39] P.J. Heptonstall, and R.J. Gross, A systematic review of the costs and impacts of integrating variable renewables into power grids, Nature Energy (November) (2020) pp. 1-12.
[40] S. Few, and P. Edge: 'Solutions to the Problem of Over-plotting in Graphs', in Editor (Ed.)^(Eds.): 'Book Solutions to the Problem of Over-plotting in Graphs' (Perceptual Edge, 2008, edn.), pp.
[41] P. del Hougne, M.F. Imani, A.V. Diebold, R. Horstmeyer, and D.R.J.A.S. Smith, Learned integrated sensing pipeline: Reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network 7 (3) (2020) pp. 1901913.
[42] S.E.A. Ali, and S. Khurram, Impact of demographic and health factors on GDP growth of South Asian Countries, International Journal of Academic Research in Business and Social Sciences 7 (3) (2017) pp. 2222-6990.