Credit Risk Prediction Using Artificial Neural Network Algorithm

Deepak Kumar Gupta
UCD Michael Smurfit Graduate Business School, Dublin, Ireland
Email: deepak.gupta@ucdconnect.ie

Shruti Goyal
UCD Michael Smurfit Graduate Business School, Dublin, Ireland
Email: shruti.goyal@ucdconnect.ie

Abstract—Artificial neural network is an information processing system which is influenced by the human brain and works on the same principles of the biological nervous system. They possess the ability to extract meaning from complex and intricate data, by detecting trends and extracting patterns from it. This paper illustrates the ability of neural network model and linear regression model constructed to predict the creditworthiness of an application accurately and precisely with minimal false predictions and errors. The results are shown to be similar for both the models, thus, models are efficient to use depending on the type of application and attributes.

Index Terms—Credit Risk, Artificial Neural Network, Linear Regression, Credit Risk Analysis, Credit Rating, Credit Rating

I. INTRODUCTION

Credit risk or credit default indicates the probability of non-repayment of bank financial services that have been given to the customers [1]. Credit risk has always been an extensively studied area in bank lending decisions. Credit risk plays a crucial role for banks and financial institutions, especially for commercial banks and it is always difficult to interpret and manage [2]. Due to the advancements in technology, banks have managed to reduce the costs, to develop robust and sophisticated systems and models to predict and manage credit risk. The objective of credit risk models is to evaluate the risk portfolio of the borrower and then assign a probability of default [3]. Therefore, there has been a discussion on classification and discrimination problems for solving credit risk models [4].

Banks evaluate loan applications based on a subjective assessment made by the borrower [5]. This assessment can lead to inefficient and inconsistent applications. Banks will be successful if they are able to reduce the credit risk and have a significant effect on economic growth of the country. To discriminate between good customers and bad customers, banks developed a need for a model-based approach that can predict credit default accurately [6]. The model-based approach provides better credit default management and efficiently allocate capital [7].

To predict the credit default, several methods have been created and proposed. The use of method depends on the complexity of banks and financial institutions, size and type of the loan [8]. The commonly used method has been discrimination analysis [9]. This method uses a score function that helps in decision making whereas some researchers have stated doubts on the validity of discriminates analysis because of its restrictive assumptions; normality and independence among variables [10]. Artificial neural network models have created to overcome the shortcomings of other inefficient credit default models [11].

The objective of this paper is to study the ability of neural network algorithms to tackle the problem of predicting credit default, that measures the creditworthiness of the loan application over a time [12]. Feed forward neural network algorithm is applied to a small dataset of residential mortgages applications of a bank to predict the credit default [13]. The output of the model will generate a binary value that can be used as a classifier that will help banks to identify whether the borrower will default or not default. This paper will follow an empirical approach which will discuss two based on neural network models and experimental results will be reported by training and validating the models on residential mortgage loan applications [14]. As the final step in the direction, linear regression method is also performed on the dataset. Results will provide a comparison between the efficiency and accuracy of the neural network and linear regression methods [15]. As the paper follows an empirical approach, this paper will show structured experimental approach to the design of models.

II. LITERATURE REVIEW

The primary objective of credit evaluation process is to compare characteristics of an applicant with other previous candidates who have repaid the loan amount.

Bank will check candidate's profile with earlier candidates, if a profile is very much similar, then they will check if an applicant has repaid the loan on time [16]. If a claimant did not default then the loan can be granted, if not then loan application will be rejected. Two techniques for credit evaluation: Credit Scoring and Officials Subjective Assessment. Traditional judgement assessment method is entirely dependent on evaluator's experience and knowledge [17]. Subjective assessment is subjective and inconsistent, but on the other hand it can be successful, creditor's experience can be qualitative that helps in taking successful credit decisions [18].

While in credit scoring method, creditors use their knowledge and historical information of the loan applications to form an evaluation model to determine creditworthiness [19]. Credit scoring methods are consistent, and self-operated that includes quantitative measurements of applicant's credit score subjected to predictor variables such as employment duration or credit history. Also, credit scoring method provides an advantage to a bank to keep their good credit customers intact and to improve customer service [20]. Consequently, this process has been criticized because data that has been used consists of some assumptions to evolve model statistically.

A. Advantages and Disadvantages of Credit Scoring

Credit scoring process does not require too much information because the process the model has been statistically developed for a set of variables; on the other hand, subjective assessment does not have any variable reduction method because of no statistical importance [21]. Credit scoring method reduces bias by inspecting rejected applications; it will keep score how rejected applicants would have behaved if they have given the loan. It considered both good and bad credit players and built a model on many applications compared to traditional methods [22]. Scoring models also contain a significant number of relevant variables that show a correlation between variables and payment behavior. A great advantage of this approach is its re-usability; the process can be used multiple time over the same data set with accuracy. Scoring models reduce processing cost and time with efficiency and ease decision-making process [23].

But, at times credit scoring model can inaccurately predict the creditworthiness of an applicant because of misclassification error. Due to its variable reduction technique, a model can miss out important variables to evaluate application which can be necessary. There may be chances that an applicant can repay the loan on time but based on the historical data or any missing information; a model can predict the wrong result. Also, these models cannot be standardized as each industry can have different credit scoring models. Historical data can play a disadvantage as due to advancements in technology and rapid changes in economic factors, credit score model prediction can be inaccurate. Models are standardized and need to update as per the economic factors, that can cost much, and the process is not easy.

B. Is credit scoring process optimal

Despite so much criticism on credit scoring models' performance, credit scoring models are in use. But, there are some open questions which have left unanswered: Optimal evaluation of an applicant, relevant variables to evolve a model, information needed to enhance decision making, best measures that can predict loan accuracy, extent to which an applicant can be classified as defaulter.

Open questions to credit scoring process [24]: How to choose appropriate technique to perform classification? Are there any other better classification methods better than credit scoring method? Is predicted value of the credit scoring model efficient than other methods? How to find out appropriate factors that influence credit scoring?

As mentioned above that credit risk majorly enhances bank's quality despite economic and environmental changes. So, banks need to have suitable methods to evaluate credit risk. A good system should be able to correctly classify between good and bad credit customers because bad credit could cause some severe issues to the bank. Our work will discuss few techniques that can be used to evaluate credit risk by determining a probability of default and classification of chances of default. Also, our work will try to find out techniques that can enhance the assessment and analysis process of the credit.

C. Diffrrerent Technology in Credit Risk

Logistic Regression allows one to build to simple model using a dependent and two or more predictor data points, and it is being used in credit scoring models as the two class problems can be represented using a dummy variable [25]. A Poisson regression can be used to classify cases where customer tends to partial repayments, and these payments can represent as a Poisson count in the model. Credit analysts can promptly analyses using linear regression credit model to investigate customer factor such as past payments record, credit guarantees and default, etc. against a predefined cut-off credit score. If new applicant credit score is higher than cut-off score, then credit is granted.

Discriminant Analysis: In credit scoring models, a statistical analysis method called Discriminant Analysis is regularly used by the researcher to rapidly build a prototype model when there are two or more categorical dependent variables for analysis. Multiple Discriminant Analysis(MDA) utilized in various studies and business verticals for the variety of applications since its inception in 1930's [26]. In 1941, Discriminant analysis used for modelling a scoring system that gives a prediction about loan repayment. Many researchers agreed that the MDA is the best use to classify a group of categorical variables into two or more predictor or classes. For example, Credit Analyst can build a scoring system using MDA to categorize a new loan application into Default or Non-Default category, and this will help banks to avoid those applicants who have potential to default in repayment sooner or later. MDA is used by developing a scoring model based on five financial ratios by analyzing financial statements to select eight variables for

predicting financial bankruptcy in Corporates. But problem associated with Discriminant Analysis such as reduction in dimensionality, improper estimation of classification error, using linear functions instead of quadratic functions, etc. Despite these limitations in MDA, it is still one of the techniques which are often used by credit analyst in building credit scoring system.

Logistic Regression has resemblance with Linear regression and it is also most commonly used statistical technique for building credit scoring system. Dichotomous nature of logistic regression outcome probability (good credit or bad credit) makes it different from linear regression [27]. By using two or more independent variables, one can build the simple logistic regression model. However, logistic regressions with more than one independent variables use the maximum likelihood method to build credit scoring model [28]. Logistic regression has been widely used in building credit scoring system in financial domain.

Neural Networks in machine learning or data mining is modelling system, which is based on the human brain and nervous system. A Neural network consists of several neurons(nodes) connected to determine the functionality of the network [29]. Several experiments to measure the performance five different types of the neural network for credit scoring. While conducting experiments, it was observed that Logistic regression is slightly more accurate in prediction in comparison to neural networks. This Research also noted that CART and k-Nearest Neighbor results are not par with logistic regression. The neural network requires being trained on a dataset to predict the outcome of decision variables correctly Applications of using the neural network in financials domains such as fraud detection in credit card transactions, forecasting company bankruptcy, classifying bad or good loan application and other areas where neural networks are successful [30]. The Performance of a neural network compared with logistic regression and found that neural network able to correctly predict loan portfolio when the measure of success is accurately classifying bad loans only.

III. ARTIFICIAL NEURAL NETWORK

Discriminant analysis method has been the most common method to build credit default or credit score models. Although, this linear methodology has been criticized by the researchers because of its assumptions on the categorical data and unequal covariance matrices of good and bad customer loans.

An artificial neural network [31] is a nonlinear approach that provides a new alternative to linear methods, especially in the situations where the dataset possesses complex relationships between the independence of the nonlinear variables [32]. Artificial neural network is a learning system that models a relationship between inputs and outputs, considering the relationship is nonlinear. They are also known as black box systems, in which extraction of information from internal system is impossible.

Artificial Neural networks are machine learning system that simulates structure and function like a biological neuron. ANNs (Artificial Neural Networks) perform a task by changing its parameters, the same way a neuron changes its states to perform a cognitive task. A network is composed of a set of neurons structured in a specified topology. Neurons are connected by links with associated weights which determines information flow intensity; weights are the functions that represent behavior of the neural network.
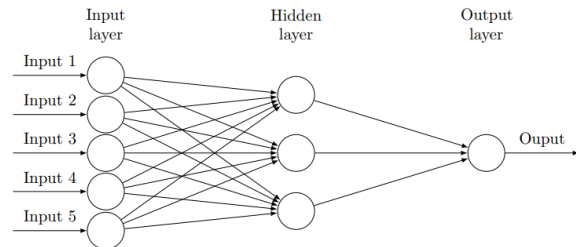

Fig.1. Basic Structure of Artificial Neural Network

Fig. 1, ANN has three layers, input layer, a hidden layer and output layer, input layer represents neurons receiving input stimulus. Then the information is transferred to next level of layer known as a hidden layer. Information is weighed before sending to next level of layers depending upon the size of the connections amongst neurons. Information is sized as per the processing unit or a transfer function represented in Fig. 2.
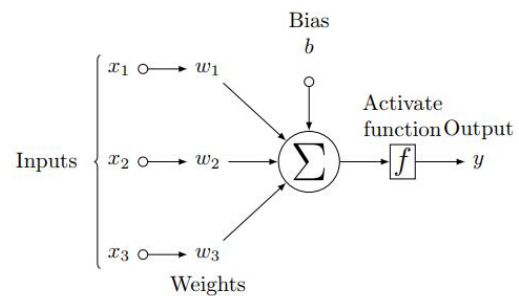

Fig.2. Mathematical Equation of Artificial Neural Network

Each neuron is characterized by a minimum value that activates a neuron (threshold value) and a transfer function. A Hidden layer can consist of several layers and performs the summation of input neurons and multiplies the weights with the summation to generate output neurons. Output generation is a two-step process: first, each input is multiplied by the weight on corresponding connection and then all valued are summed together; second, activation function is applied to summation of the inputs [33].

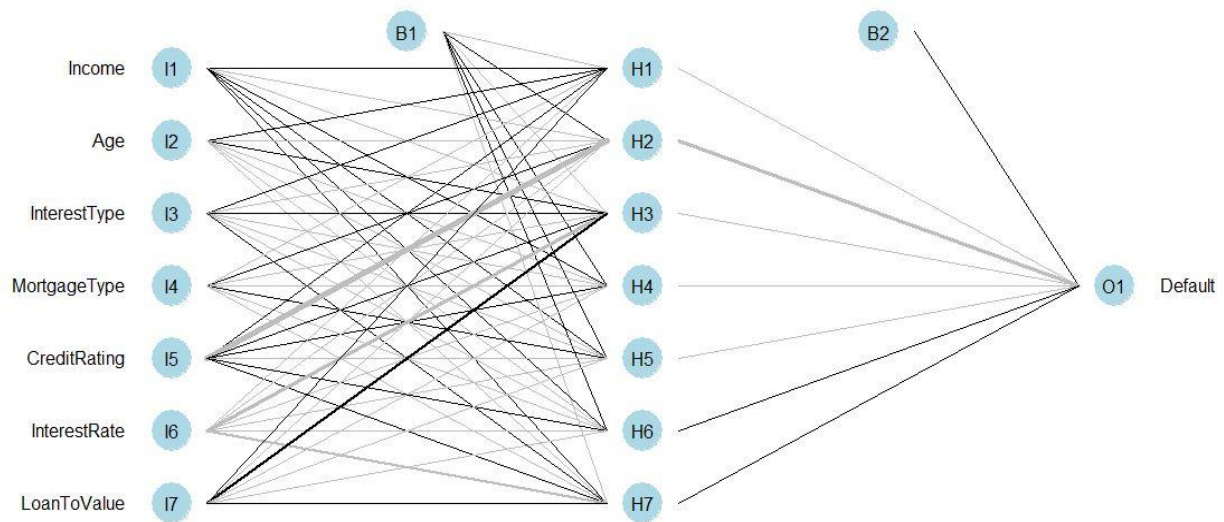$$y_i = \sum_{j \in I} W_{j,i}\, a_j \qquad (1)$$

$$a_i = g(y) \qquad (2)$$

Fig.3. Neural Network Plot of the Credit Default Model.

As (1) represents evaluation of input and (2) represents activation function; where W*j,i* represents weights on the connection between *j* and *i* and a*j* is activation function of neuron *j*.

For a neural network to work efficiently, weights should be tuned accurately. This task can be achieved by using a learning algorithm, which trains the network and modifies weights until verified. Mostly, these algorithms stop when there occurs an error between output generated by network falls under threshold and expected output. There are three types of learning algorithms for artificial neural networks i.e. supervised Learning, unsupervised Learning and reinforced Learning.

In supervised learning, a training set of correct examples is being used to train the network model. It consists of pairs of several inputs and expected outputs. Weights will be tuned based on the errors generated in the network. Most common example of supervised learning is classification, where the network has to learn to generalize relations between corresponding input and output variables. In this paper, we will be dealing with the typical classification problem to predict the credit default.

IV. METHODOLOGY

A. About Data Source

Data has been collected from kaggle.com (lending club loan data) that consists of more than 1 million records. A random sample data of 60,000 records have been pulled out from the dataset and appropriate attribute selection has been done from 80 attributes. Attribute selection includes numeric and integer attributes along with some factor attribute relevant to the problem this paper is dealing with. Dataset consists of combination of variables as follows:

Dependent Variable: loan_status (0 and 1); this paper aims to predict the creditworthiness of the borrower in near future. In this context, if the borrower will default then the investment will be bad and if the borrower will not default then he or she will be able to repay the full loan amount. So, to differentiate in neural network 0 indicates borrower will default and 1 indicates borrower will not default.

Independent Variable: Following variables are considered as an independent variable:

- *loan_amnt*: Amount of the loan that has been given to the borrower
- *funded amnt*: Amount that has been asked by the borrower to loan at beginning
- *funded_amnt_inv*: Amount that has been committed by the investors
- term: Total period that has been agreed upon to repay the loan
- *int_rate*: Interest rate on which loan has been given
- *instalment*: Amount agreed to repay the loan monthly
- *grade*: Credit rating assigned to the loan application
- *emp_length*: Total number of years the borrower has been employed
- *annual_inc*: Annual income reported by the borrower while filing an application
- *issue_d*: Issued date of the loan application
- *application_type*: Whether the application is individual or joint

B. Model

In this study, a classic feed-forward neural network has been used. The feedforward network consists of an input

layer with 10 input variables, 7 hidden layers and an output layer with one neuron that represents a classifier. The network is trained by using a supervised learning algorithm (back propagation algorithm [34]). The algorithm optimizes the neuron weights be minimizing the error between actual and desired output. Error is $error_i = D_i - A_i$ for neuron i. Weights will be updated by formula $W_{i,k} = W_{i,k} + \phi I_k error_i g'(in_i)$ where φ be the learning coefficient and Ik is the output from hidden layer. Algorithm will work until a stopping criterion is found [35].

Fig.3. It is necessary to carefully choose the parameters, such as the value of φ and several neurons and number of hidden layers, for the neural network algorithm. Fig. 3 connections are represented by black lines between every layer and weights and the blue line shows the bias (intercept of the model) in every step. The network is a black box and training algorithm is ready to use as it is converged. Also, a random sample has been created from the extracted dataset for the network algorithm. Then a training and test dataset is created used to train the model and to validate the performance of the model respectively.

Also, to keep in mind that useless fields will be erased from the dataset, such that neural network will only work on numeric and integer variables. Next, data normalization is performed to feed the neural network with data that range in the same interval. Min and max linear transformation function have been used to normalize the data before splitting the dataset into training and test datasets.

V. EXPERIMENTS AND RESULTS

Once the dataset was trained, it was tested on the test dataset. To compute the output based on the other inputs, compute function has been used. Severn hidden layers were added to the network and model was created.
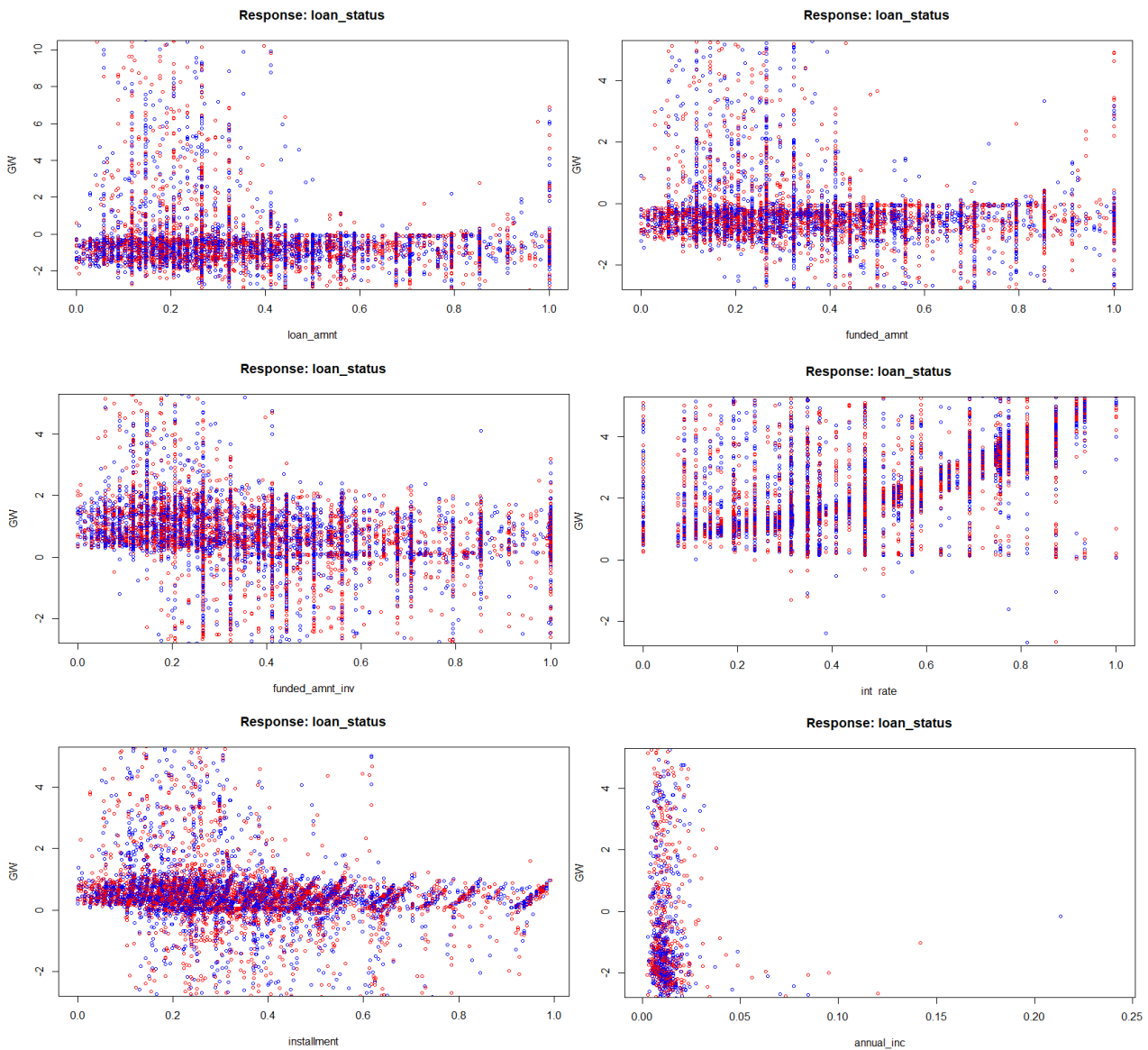

Fig.4. Generalized Weights of the Input

Fig.4. generalized weights of all the covariates are shown. Distribution has shown that all the attributes have a nonlinear effect on loan status since all weights have the generalized weight of greater than 1.

Fig.5 there are 10 normalized variables have been fed as the input to the network arranged in an order. The output of the network is a classifier that results in 0 and 1. At first, data has been checked for missing data point value, no data was missing; there was no need to fix the dataset.
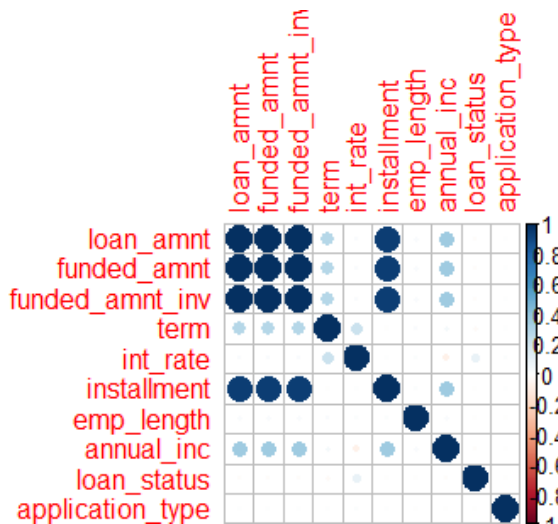

Fig.5. Correlation Plot of the Input Dataset


```
           Reference
Prediction     0      1
         0  46145   1079
         1      0      0

            Accuracy : 0.9771514
              95% CI : (0.9757635, 0.9784803)
 No Information Rate : 0.9771514
 P-Value [Acc > NIR] : 0.5080963
```

Fig.6. Confusion Matrix and Statistics of linear Regression

Total 6765 steps were needed until all derivatives of the error function was smaller than default threshold (0.01) and model took 480 seconds. Initial error was 322.833 which subsequently reduced to 0.09 after running for 6765 steps. After implementing classic feed forward algorithm, another model has been implemented by using back propagation algorithm with 0.01 learning rate. Classic process and back propagation process have almost same error rate. Thus, classic model fit is less satisfied than back propagation algorithm. Following result matrix has been generated by the network:

Table 1.Result Matrix for Classic Feed Forward Neural Network

Attribute	Value
Error	322.833
Reached Threshold	0.0998
Steps	6765

Last, linear regression [36] have been applied to the dataset to compare the accuracy of both the algorithms. glm() function has been used to fit the linear regression model. For regression a probability greater than 0.5 has been assigned, if predicted values in the regression are greater than 0.5, then the value is 1 else 0. Fig.6 Accuracy has been calculated by incorporating misclassification error and confusion matrix has also calculated.

Table 2. Mean Square error of both the processes

MSE neural network	MSE linear regression
0.0220449	0.0227334

To highlight the comparison, mean square error of both linear regression and the neural network has been calculated as shown in table 2. As can be seen in the table mean square error of both the process is approximately same and thus both the process is doing same work. It is necessary to know that deviation in MSE depends on the training and test split.

VI. CONCLUSION AND FUTURE WORK

This paper has studied artificial neural network and linear regression models to predict credit default. Both the system has been trained on the loan lending data provided by kaggle.com. Results of both the system have shown an equal effect on the dataset and thus are very effective with the accuracy of 97.67575% by artificial neural network and 97.69609%. The system classifies the output variable correctly with a very low error. So, both process can be used to identify credit default with equal accuracy. Also, the neural network represents a black box method such that it is difficult to explain the outcome compared to the linear regression model. Therefore, which model to use depends on the application one must use. Moreover, while fitting a model using neural network process user needs to take extra care of the attributes and data normalization to improve the performance. To conclude, neural network provides strong evidence to efficiently predict the credit default for a loan application.

Neural network algorithms have a wide range of applications that are not only essential for residential mortgages. Other applications can be rating bonds issued by companies commonly known as bond rating, rating short term investments that can last up to 1 year, long-term and short-term ratings of local and foreign currencies, sovereign, or country ratings. The prediction system can be further enhanced to assign a credit rating to an application by using appropriate algorithms and technologies.

REFERENCES

[1] Bielecki, T.R. and Rutkowski, M., 2013. *Credit risk: modeling, valuation and hedging*. Springer Science & Business Media.

[2] Imbierowicz, B. and Rauch, C., 2014. The relationship between liquidity risk and credit risk in banks. *Journal of Banking & Finance*, *40*, pp.242-256.

[3] Acharya, V., Davydenko, S.A. and Strebulaev, I.A., 2012. Cash holdings and credit risk. *The Review of Financial Studies*, 25(12), pp.3572-3609.

[4] Cole, S., Kanz, M. and Klapper, L., 2015. Incentivizing calculated risk-taking: evidence from an experiment with commercial bank loan officers. *The Journal of Finance*, 70(2), pp.537-575.

[5] Pacelli, V. and Azzollini, M., 2011. An artificial neural network approach for credit risk management. *Journal of Intelligent Learning Systems and Applications*, 3(02), p.103.

[6] Kozeny, V., 2015. Genetic algorithms for credit scoring: Alternative fitness function performance comparison. *Expert Systems with applications*, 42(6), pp.2998-3004.

[7] Angelini, E., di Tollo, G. and Roli, A., 2008. A neural network approach for credit risk evaluation. *The quarterly review of economics and finance*, 48(4), pp.733-755.

[8] Laeven, M.L., Ratnovski, L. and Tong, H., 2014. *Bank size and systemic risk* (No. 14). International Monetary Fund.

[9] Zopounidis, C. and Doumpos, M., 2002. Multicriteria classification and sorting methods: a literature review. *European Journal of Operational Research*, 138(2), pp.229-246.

[10] Eisenbeis, R.A., 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. *The Journal of Finance*, 32(3), pp.875-900.

[11] Khemakhem, S. and Boujelbene, Y., 2015. Credit risk prediction: A comparative study between discriminant analysis and the neural network approach. *Accounting and Management Information Systems*, 14(1), p.60.

[12] Lessmann, S., Baesens, B., Seow, H.V. and Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), pp.124-136.

[13] Thomas, L., Crook, J. and Edelman, D., 2017. *Credit scoring and its applications* (Vol. 2). Siam.

[14] Halper, S.C., Wilson, C.A. and Hourigan, S.M., Interthinx Inc, 2013. *Automated loan risk assessment system and method*. U.S. Patent 8,458,082.

[15] Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I. and de Mendonça, A., 2011. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, 4(1), p.299.

[16] Hotz, S., Kelly, J., Srinivasan, K. and Jindia, A.K., Compucredit Intellectual Property Holdings Corp II, 2011. *Method and system for rapid loan approval*. U.S. Patent 7,933,833.

[17] Mohammadi, N. and Zangeneh, M., 2016. Customer Credit Risk Assessment using Artificial Neural Networks. *IJ Information Technology and Computer Science*, 8(3), pp.58-66.

[18] Brown, K. and Moles, P., 2014. Credit risk management. *K. Brown & P. Moles, Credit Risk Management*, p.16.

[19] Mester, L.J., 1997. What's the point of credit scoring? *Business review*, 3(Sep/Oct), pp.3-16.

[20] Bornhofen, B., Byrne, L., Bray, D., Elder, R., Feight, R., Pinnola, K.H., Shimshi, F., Quinlan, R. and Cheeseman, M., Citibank NA, 2018. *Method and system for the issuance of instant credit*. U.S. Patent 9,898,780.

[21] Halvaiee, N.S. and Akbari, M.K., 2014. A novel model for credit card fraud detection using Artificial Immune Systems. *Applied Soft Computing*, 24, pp.40-49.

[22] Bunn, D. and Wright, G., 1991. Interaction of judgemental and statistical forecasting methods: issues & analysis. *Management science*, 37(5), pp.501-518.

[23] Bazmara, A. and Donighi, S.S., 2014. Bank customer credit scoring by using fuzzy expert system. *International Journal of Intelligent Systems and Applications*, 6(11), p.29.

[24] Rahman, M.M., Ahmed, S. and Shuvo, M.H., 2014. Nearest Neighbor Classifier Method for Making Loan Decision in Commercial Bank. *International Journal of Intelligent Systems and Applications*, 6(8), p.60.

[25] Lee, T.S. and Chen, I.F., 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4), pp.743-752.

[26] Harris, T., 2015. Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42(2), pp.741-750.

[27] Hosmer, D.W., Jovanovic, B. and Lemeshow, S., 1989. Best subsets logistic regression. *Biometrics*, pp.1265-1270.

[28] Altland, H.W., 1999. Regression analysis: statistical modeling of a response variable.

[29] Chittineni, S. and Bhogapathi, R.B., 2012. Determining contribution of features in clustering multidimensional data using neural network. *IJ Information Technology and Computer Science*, 10, pp.29-36.

[30] Ghosh, S. and Reilly, D.L., 1994, January. Credit card fraud detection with a neural-network. In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on* (Vol. 3, pp. 621-630). IEEE.

[31] Laitinen, E.K., 1999. Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International review of financial analysis*, 8(2), pp.97-121.

[32] Pang, S.L., Wang, Y.M. and Bai, Y.H., 2002, November. Credit scoring model based on neural network. In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on* (Vol. 4, pp. 1742-1746). IEEE.

[33] Atiya, A.F., 2001. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on neural networks*, 12(4), pp.929-935.

[34] Hornik, K., Stinchcombe, M. and White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), pp.359-366.

[35] Pineda, F.J., 1987. Generalization of back-propagation to recurrent neural networks. *Physical review letters*, 59(19), p.2229.

[36] Hand, D.J. and Henley, W.E., 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), pp.523-541.

Authors' Profiles

Deepak Kumar Gupta has degrees in Master of Science in Business Analytics (2017) from UCD Michael Smurfit Business School, and Bachelor of Technology in Computer Science & Engineering (2013) from Guru Gobind Singh Indraprastha University Delhi, India. He has over 4 years

of experience in Digital Analytics and Business Intelligence. He is Co-Founder of Reckon Analytics (http://www.reckonanalytics.com), a marketing automation platform based in Ireland.

Shruti Goyal has degrees in MSc Business Analytics (2017) from UCD Michael Smurfit Graduate Business School, Dublin, Ireland and B.Tech in Instrumentation & Control Engineering (2013) from Guru Gobind Singh Indraprastha University Delhi, India. She has over 4 years of experience in data analytics and performance management consulting, previous she has worked IBM Labs and Tata Consultancy Services. Currently, she is working as Co-Founder of Reckon Analytics.