DCU School of Computing
Assignment Submission

Student Name(s): Nitin Yadav
Student Number(s): 16212304
Programme: MCM
Project Title: Credit Score Predictive Modeling
Module code: CA685
Lecturer: Dr. Andrew McCarren
Project Due Date: 22/08/2017

Declaration

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying is a grave and serious offence in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion, or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references.

I have not copied or paraphrased an extract of any length from any source without identifying the source and using quotation marks as appropriate. Any images, audio recordings, video or other materials have likewise been originated and produced by me or are fully acknowledged and identified.

This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. I have read and understood the referencing guidelines found at http://www.library.dcu.ie/citing&refguide08.pdf and/or recommended in the assignment guidelines.

I understand that I may be required to discuss with the module lecturer/s the contents of this submission.

I/me/my incorporates we/us/our in the case of group work, which is signed by all of us.

Signed: Nitin

Print this page

Credit Score Predictive Modeling:
Comparison of parametric, non-parametric, and neural network credit scoring models

Nitin Yadav
Department of Engineering and Computing
Dublin City University
Dublin, Ireland
nitin.yadav2@mail.dcu.ie

Dr. Andrew McCarren
Department of Engineering and Computing
Dublin City University
Dublin, Ireland
andrew.mccarren@dcu.ie

Abstract— **Credit scoring has been extensively used by financial institutions to identify the potential risks associated with loans granted to their customers. Various models have been developed using machine learning and statistical techniques to eliminate the inefficiencies involved in decision making using personal judgement. This paper ranks the credit scoring models based on their prediction power. Two statistical techniques, logistic regression and random forest, and one machine learning algorithm utilizing neural networks are selected for the analysis conducted in this research. Decision tree boosting is also implemented to reduce the discrepancy between train and test data accuracies for the random forest technique. The models are trained and tested using the credit data from a bank in Taiwan obtained from the UCI repository. Exhaustive data analysis is performed to identify the influential variables. Demographics are found to be the least influential in the overall score prediction, excluding education while the most influential are found to be bill and repayment amounts. Results indicate that neural networks achieve the best prediction accuracy of 0.829 while the accuracies for logistic and random forest technique are found to be 0.828 and 0.825. While the train and test accuracies for the logistic regression and neural network model are close enough, a large discrepancy of 7.0% exists between test (0.883) and train (0.825) accuracies for random forest model. The employed boosting technique reduces the discrepancy to 1.45%.**

Keywords—credit scoring, model prediction, statistical techniques, data mining algorithms, credit score prediction.

I. INTRODUCTION

Over the past few decades, financial institutions have encountered numerous challenges across the globe. These challenges are associated with the growing complexity of financial markets and increasing number of financial services offered by the institutions. To account for these aspects, it has become necessary for these institutions to analyze the risks and losses associated with the introduction of the new services. There is a need for appropriate tools and prediction systems to develop scales for measuring risks. Loans are one of the most crucial services provided by the banking institutions. As such, the credit risk assessment is often considered to form a major part of the overall risk assessment in any financial institution. The credit risk assessments are commonly referred to as credit scoring models.

The decision to whether issue a loan or not is usually narrowed down to a binary classification distinguishing good applicants from the bad ones [1]. The financial experts base their decision based on the financial status, demographic details, and the intention of the applicant. Nowadays, the decision is mostly based on the predictions done by credit scoring models using statistical techniques. These techniques utilize the applicant data and previous payment records to predict credit scores which are then used by the financial institutions to make informed decisions [2].

Logistic regression, decision trees, Bayes classifier etc., are some of the widely used statistical techniques to develop credit scoring models [3]. Models employing neural networks have also been implemented in credit risk forecasting [4].

Along with a binary classification of default status of customers, it is equally important to emphasize the fact that each attribute of the model influences the final credit score in a different way [5]. Moreover, some of the attributes might influence the credit score indirectly through existing interlinks with the influential attributes [6].

This paper compares the working efficiencies of a parametric statistical technique, a non-parametric statistical technique, and a machine learning technique using neural networks based on their credit score prediction power. The research conducted draws its conclusions using a dataset obtained from UCI learning repository [7]. A study by Yeh et. al [8] compared the different data mining methods, namely, logistic regression, discriminant analysis, decision trees, Bayesian classifier, K-nearest neighbor, and artificial neural networks using the above mentioned dataset. This work further adds to the research conducted by Yeh et. al by implementing boosting technique to improve the non-parametric statistical techniques. Alongside, the dependency of different attributes has been analyzed to identify the individual influence of each attribute on the response variable, i.e., default status.

The paper is structured as follows: Section II describes a brief literature review of concepts in credit scoring followed by credit score prediction models developed in the past. Section III describes the research methodology implemented to conduct this research, which includes the description of data used in this study and models implemented for score prediction. Section IV lists important results and discussions related to data analysis and model efficiencies. Section V comes out with a conclusion and recommendations for future research.

II. Literature review

This section gives a brief introduction of credit scoring concepts and models implemented so far in the literature for credit score prediction.

A. Credit Scoring

Credit scoring is the standardized process of analyzing and classifying the payment behavior of buildings, i.e. private consumers or companies' as stated by Schwarz and Arminger [9]. The process involves the implementation of statistical techniques for predicting an individual's possibility to repay the credit or settle payments regularly. The assessment helps in the identification of the credit trustworthiness of credit applicants. These statistical techniques result in the development of credit scoring models, which can be used to identify the credit health score of existing clients or to assess the new clients [10]. Furthermore, the model can be used to comprehend the driving factors behind the non-repayment of credit. Overall, these models assist the banks in enhancing their risk assessment procedures.

Another definition by Anderson [11] involves defining the two terms separately; credit and scoring. The authors suggested the demarcation of terms to establish a clear understanding of the aspects involved. Credit scoring can be defined as 'the use of statistical models to transform relevant data into numerical measures to guide credit decisions. It is the industrialization of trust; a logical future development of the subjective credit ratings, first provided by nineteenth century credit bureaus, that has been driven by a need for objective, fast and consistent decisions, and made possible by advances in technology' [11]. Furthermore, 'Credit scoring is the use of statistical models to determine the likelihood that a prospective borrower will default on a loan. Credit scoring models are widely used to evaluate businesses, real estate, and consumer loans' [12]. Also, 'Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit. These techniques decide who will get credit, how much credit they should get, and what operational strategies will enhance the profitability of the borrowers to the lenders' [13].

Credit scoring only makes use of variables that are statistically significant and correlated with repayment performance [14]. Thus requires less information to reach a decision. Credit scoring removes the bias resulting from the judgmental methods that consider only the characteristics of accepted applicants [15]. However, it should be noted that the credit scoring model is just one of the factors in credit risk assessment. The end decision depends on a final review by the credit expert [16]. Furthermore, the data associated with credit scoring is mostly historical and variables are assumed to be static leading to an inaccurate model. This issue can be addressed by frequently updating the model to include any new changes.

B. Models for Credit Scoring Prediction

With the fast growth of the credit industry, credit scoring has emerged as one of the crucial techniques in banks and other financial institutions. Credit scoring models have been widely used by these institutions to issue loans to the good customers and set a concrete demarcation between good and bad ones. Risk assessment through credit scoring models reduces the cost and time associated with the credit process [17]. One of the main goals of such models is to help the development of the credit management process and to provide credit analysts and decision makers with an efficient and effective credit tool to help to determine strengths, weaknesses, opportunities and threats (SWOT), and to help to evaluate credit more precisely [15].

Financial institutions in developed countries have a well-established credit scoring system and often have many customers compared to developing countries due to the vast number of services provided to them. West [18] has stated that credit scoring is widely used by the 'financial industry', mainly to improve the credit collection process and analysis, including a reduction in credit analysts' costs, faster credit decision- making and monitoring of existing customers. According to the study by West [18], credit scoring models are used by 97% of banks in the US for credit card applications. Furthermore, credit institutions and especially mortgage organizations are developing new credit scoring models to support credit decisions to avoid large losses. These losses are usually considerable. For example, West [18] reported that 'in 1991 $1 billion of Chemical Bank's $6.7 billion in real estate loans were delinquent'.

Non-linear statistical techniques form the backbone of credit scoring models. Some of the widely used techniques for building credit scoring models include weight- of- evidence measure, regression analysis, discriminant analysis, probity analysis, logistic regression, linear programming, Cox's proportional hazard model, support vector machines, decision trees, neural networks, K- nearest neighbor (K- NN), genetic algorithms, and genetic programming [19]. Although sophisticated techniques like genetic algorithms and genetic programming are resource and time intensive, such techniques are able to model extremely complex functions and thus form excellent classifiers [20]. Data exploration or data mining also forms a crucial part of the model building process. Companies nowadays try to extract more and more information from the available datasets to assign levels of significance to different attributes [21].

Regression methods, linear or non-linear, help in identifying the relationship between a response variable and one or more independent variables and thus, are crucial in data analysis. Linear regression is quite popular in credit scoring applications because of its simplicity and ease in score prediction. A regression approach was implemented by Orgler [22] in the identification of risks associated with commercial loans. The study was extended to risk identification of consumer loans [23]. An important conclusion drawn from these studies points to the fact that information not included in the applicant's form has more weightage in the assessment of future repayments. Logistic regression is an another widely used statistical technique in this field. The difference between linear and logistic regression lies in the outcome, which is in binary form for the latter. A logistic regression model works with the probability of occurrence of any event where the probability is a function of observed values of the explanatory variables [15]. On theoretical grounds, logistic regression is a better statistical instrument than linear regression, provided the good and bad are described appropriately [24].

Another parametric statistical technique, discriminant analysis, allows to discriminate between two groups. This technique is an alternative to logistic regression and assumes a multivariate normal distribution for the explanatory variables [25]. Although one of the broadly established techniques, several authors have identified statistical loopholes in performing discriminant analysis for credit score prediction. Problems such as group definitions, classification error prediction etc., should be addressed when employing discriminant analysis. Probit analysis works on the same principles as discriminant except that it assumes normal distribution of the threshold values and estimates of coefficients which can be individually tested for their significance [26].

Decision tree or recursive partitioning is a non-parametric method and utilizes dichotomous tree classification of records to evaluate all the possible splits [27]. Overall error rate is used to select the best sub-tree. This technique often results in simple classification rules and can handle non-linearity. However, a specific tree structure is only applicable for a specific application and hence, this technique lacks generalization. An efficient technique using decision trees, known as random forests, combines all the weak models to build a single powerful model for improving the overall accuracy. This technique is an ensemble learning technique, which allows models to be trained quite fast [28]. However, the constructed models are often slow in making predictions and suffer from overfitting issues. Also, this technique doesn't indicate any existing relationships between the variables involved. To solve the overfitting issue, decision tree boosting can be implemented in which a classifier is added at a time such that the next classifier is trained to improve the already trained ensemble [29]. Another statistical technique, K-Nearest Neighbor (KNN), employs classifiers built using analogy between the unknown samples and already established patterns. The most common class is assigned to the unknown sample among all the KNNs. While it is not required to establish a predictive model before classification, KNN doesn't produce any classification probability formula and the accuracy is affected by the measure of distance and the cardinal k of the neighborhood [8].

More sophisticated techniques, like neural networks, utilize computer programs that are trained based on trial and error. Training allows to distinguish variables and implement an enhanced risk assessment. Neural networks have become very popular with financial institutions and are being implemented in applications like detection of credit card frauds, bankruptcy prediction, mortgage applications, and several others [27]. These networks can handle any degree of non-linearity and any number of interdependencies between explanatory variables. However, the model doesn't result in any probabilistic formula of classification.

Genetic programming is one of the recently developed techniques used to develop the credit scoring models. It employs genetic algorithms, which transform a dataset according to the fitness value using genetic operations [30]. A set of competing programs are generated by processes of mutation and crossover resulting in an output in the form of a string. This is one of the widely growing area and the number of applications have increased over the past few decades.

C. *Performance Evaluation Criteria*

Different types of evaluation criteria are used to rate the credit scoring models. Some of these criteria include the confusion matrix, mean square error (MSE), root mean square error (RMSE), the ROC curve and mean absolute error (MSE) [15]. This section describes some of these criteria implemented in this research study.

The most widely used criteria in credit scoring is the confusion matrix [31]. It measures the proportion of correctly classified cases as good and as bad credit in any dataset. The matrix represents the combinations of actual and predicted observations. The only disadvantage associated is that this criterion doesn't take into consideration the different misclassification costs for the actual good predicted bad and bad predicted good observations.

Another widely used criteria is the ROC curve or the Lorentz diagram. The ROC curve is a two dimensional graph representing the proportion of bad cases classified as bad vs the proportion of good cases classified as good at all cut-off score values [24]. This curve illustrates the classifiers' behavior irrespective of misclassification costs or different class distributions. Criteria such as MSE, RMSE etc., rely on a number as the output to evaluate the model performance.

III. Research Methodology

This section describes in detail the research process implemented in this study. Data analysis is exhaustively performed to identify interlinks between different variables. Furthermore, techniques for credit score evaluation are analyzed and rated based on their accuracies.

Python is used to perform the analysis on the dataset and to compare the selected techniques. Python smoothens the overall implementation process. Moreover, there are already existing standard codes in Python to implement the statistical techniques.

A. *Data Description*

This research study utilizes payment data obtained from the UCI Machine Learning Repository [32]. The data was a part of research study published by Yeh et. al and contains the details of credit card payments and demographics of customers in Taiwan. The dataset includes a total of 30,000 observations and a binary response variable (No= 0 or Yes= 1) has been assigned to signify whether a need to clear off a default payment or not. Each customer is linked to 23 variables mostly describing his/her demographic and financial information. There are 22 explanatory variables and one dependent variable, namely, default status.

For the entire 30,000 observations, there are no missing attributes in the dataset. The values taken by the variables appear to be quite random and hence, each variable was considered for the analysis. The variables were also tested for correlation amongst them. The following list describes all the considered attributes in the model building process.

X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
X2: Gender (1 = male; 2 = female).
X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
X4: Marital status (1 = married; 2 = single; 3 = others).
X5: Age (year).
X6–X11: History of past payment. We tracked the past monthly payment records (from April to September 2005) as follows: X6 = the repayment status in September 2005; X7 = the repayment status in August 2005; ...; X11 = the repayment status in April 2005. The measurement scale for the repayment status is: 1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; ...; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
X12–X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September 2005; X13 = amount of bill statement in August 2005; ...; X17 = amount of bill statement in April 2005. X18–X23: Amount of previous payment (NT dollar). X18 = amount paid in September 2005; X19 = amount paid in August 2005; ...; X23 = amount paid in April 2005.

B. *Credit Scoring Model Implementation*

Among the credit scoring techniques that exist in literature, a few of them are tested in their ability to predict the credit scores. One technique from parametric type and one from non-parametric type are chosen to establish the difference between their respective predictive powers. Another technique using neural networks is also implemented to establish its supremacy over the conventional credit scoring techniques. The models are trained with the dataset mentioned in the data description. The dataset is divided into train (60% of the data selected at random) and test (remaining 40% data) data. The implementation is performed in Python and confusion matrix and ROC curves are used to analyze the credit scoring models.

A feature selection process is also run in Python to rank the attributes in the order of their increasing effect on the credit scoring model. The Python source code has been uploaded on Gitlab and can be accessed via link https://gitlab.computing.dcu.ie/yadavn2/Practicum.

IV. RESULTS AND DISCUSSIONS

This section describes the exhaustive data analysis performed to extract detailed information that would aid the process of credit score predictive modeling. Furthermore, credit scoring models are implemented using statistical techniques. It should be stressed that 22.12% of the total customers are defaulters. Statistically, a dataset with 50% would result in a clear demarcation between the customers with defaults and those with no defaults. However, the observed 22.12% of defaulters is still enough to allow firm conclusions to be drawn.

A. *Data Analysis*

Data analysis forms a crucial step of the model building process. For data analysis in this study, programming language Python has been used.

The dataset contains two classes of variables, namely, factor and numeric variables. A factor variable takes a specific value from a pool of values. For instance, gender is a factor variable as it can take either 1 (Male) or 2 (Female). The response variable, the default status of a customer, is also a factor variable. The customer either owes some credit to the bank or he/she doesn't. The other factor variables are education, marriage, and the repayment status. The remaining set of variables are known as numerical variables, take on random

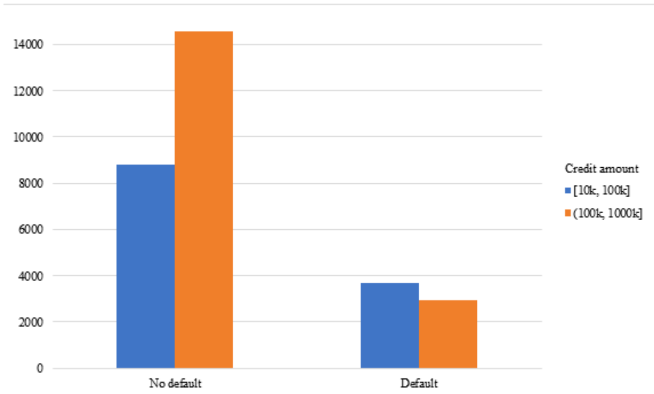

Fig. 1. Number of customers with and without default against two levels of credit amount issued.

numeric values. The numerical variables in the dataset are the loan amounts, bill amounts, and the repayments. To smoothen the process, a separate analysis is carried out for the two sets of variables. The numerical variables are then transformed into character variables or factors to perform a combined analysis of the dataset.

The analysis is initiated with considering single response variable and is carried out further with other variables to analyze the inter-dependencies between the response and explanatory variables. Out of the 30,000 customers, the count of defaulters who did not repay the loan is 6636.To analyze the inter dependency between default rate and credit amount, the entire column of the credit amount is categorized into two levels: between 10k and 100k and between 100k and 1000k. Fig. 1, depicts the number of customers with and without defaults against the amount of credit given. It can be inferred that the people with a default status have a lower credit amount than the people who are paying on time.

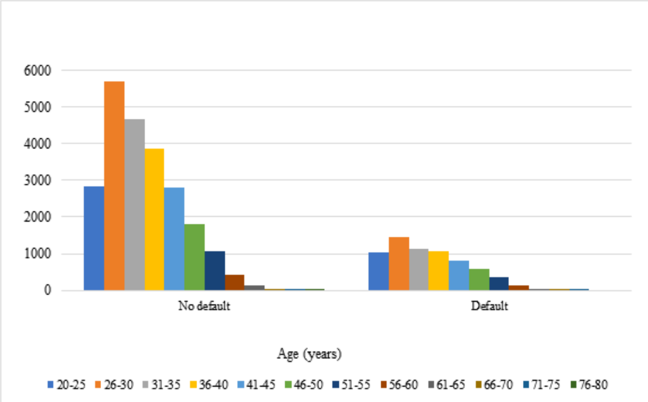

Fig. 2. Number of customers with and without defaults classified based on their age.

Fig. 2, depicts the inter dependency between the ages of customers and their default status. There seems to be not much of a significant difference between the ages and between the number of people who are defaulting, and not defaulting. The bar graph, as in Fig. 2, remains identical for defaulters and non-defaulters.

To further analyze the effect of variables like education on the default status, Chi-square test of Independence is performed as both the concerned variables are categorical variables. Low p-value at a 5% significance level indicates that a relationship exists between the education level of customers and their default status. Furthermore, it is interesting to note that the proportion of customers who are defaulters and pursued a degree at a university or went for higher education is more than the customers who are defaulters and only passed high school, as depicted in Fig. 3.

Furthermore, a gender bias exists in the number of loans granted. The percentage of females (approximately 60% of total) taking loans is significantly higher than males. Also, the loan amount given to females is higher than the amount given to males.

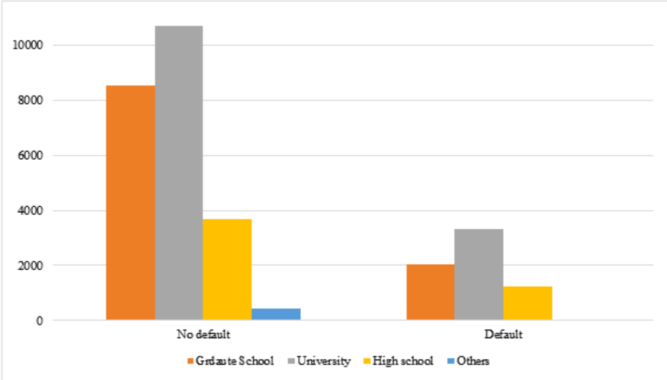

Fig. 3. Number of customers with and without defaults classified based on their education status.

Moreover, females share a higher proportion (approximately 57%) of the total number of default cases compared to males, as shown in Fig. 4.

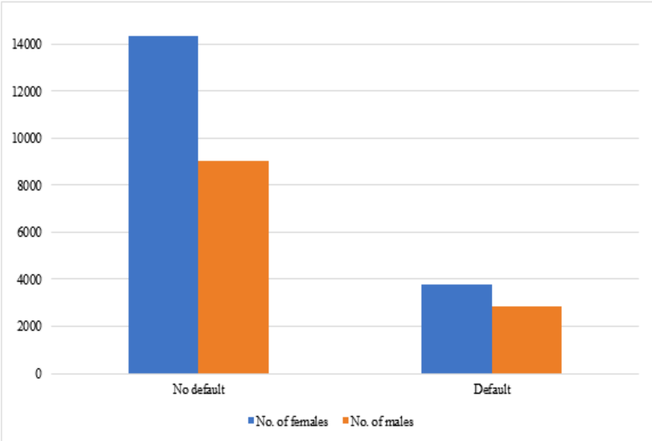

Fig. 4. Number of customers with and without defaults classified based on their sex.

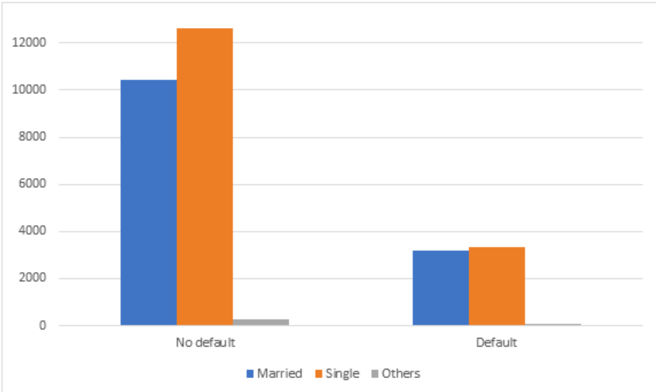

Fig. 5. Number of customers with and without defaults based on their marital status.

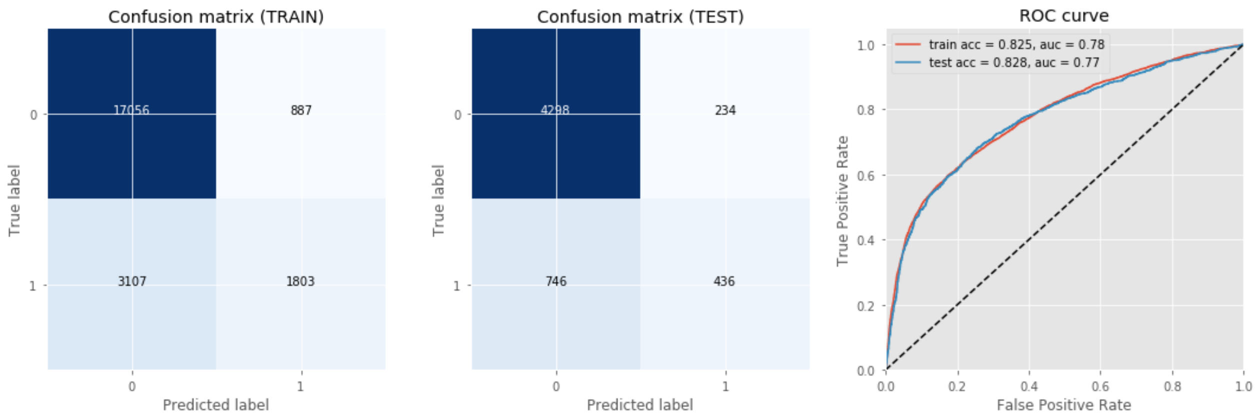

Fig. 6. Confusion matrix for the test and train data for the logistic regression model.

Fig. 5, depicts the number of customers with and without defaults classified based on their marital status. Nothing significant can be inferred from the graph. The number of people who are single tend to apply for loans more than those who are married. The proportion of people in single and married category remains same irrespective of the default status.

B. Validation of credit scoring models

Logistic regression, random forest, decision trees boosting, and feed forward deep neural networks are validated based on the confusion matrix and ROC curves. Among all the attributes, the numerical variables, like bill amounts and the repayment amounts, appear to be the top indicators. Demographics do not strongly influence the credit prediction process. However, education does have a strong effect on the predicted score.

Fig. 6, depicts the confusion matrix and ROC curve of the train and test data for the credit scoring model built using logistic regression technique. To test the model's classification results against the actual observed classification, confusion matrix is calculated. Confusion matrix displays the number of true positives (number of cases predicted as default and there was an actual default) and true negatives (number of cases predicted as no default and the payment status was clear) along the diagonal of the matrix. The off-diagonal elements represent false positives (number of cases predicted as default but the repayment status was clear) and false negatives (number of cases predicted as no default but there was an actual default). Model accuracy is calculated by dividing the total number of true positives and true negatives by the total number of cases. For the training dataset of logistic regression model, the number of true positives and the number of true negatives are found to be 1803 and 17056. The classification accuracies for the train and test data are found to be 0.825 and 0.828. The values for the area under the ROC curve (AUC) are found to be 0.78 and 0.77 for test and train data. The numbers suggest that the logistic regression technique has a moderate predictive power.

For model developed using the random forests technique, the train and test data accuracies are found to be 0.88 and 0.825, which suggest that the model performance is sub-optimal and can be improved by manipulation of control parameters and the predictors used. The AUC value for the train data is 0.97, which is quite high. However, the AUC value for test data drops to 0.78, which is in accordance with the accuracy values found using the confusion matrix. The confusion matrix and ROC curve are depicted in Fig. 7.

The accuracies for the decision tree boosting method are 0.837 and 0.825 for the train and test data respectively. The observed difference in accuracies for this method is less compared to the difference in accuracies using random forests technique. This is because this technique iteratively learns weak classifiers and adds them to the final strong classifier changing the weightage of each one. The AUC values from the ROC curve, shown in Fig. 8, are found to be 0.81 and 0.78

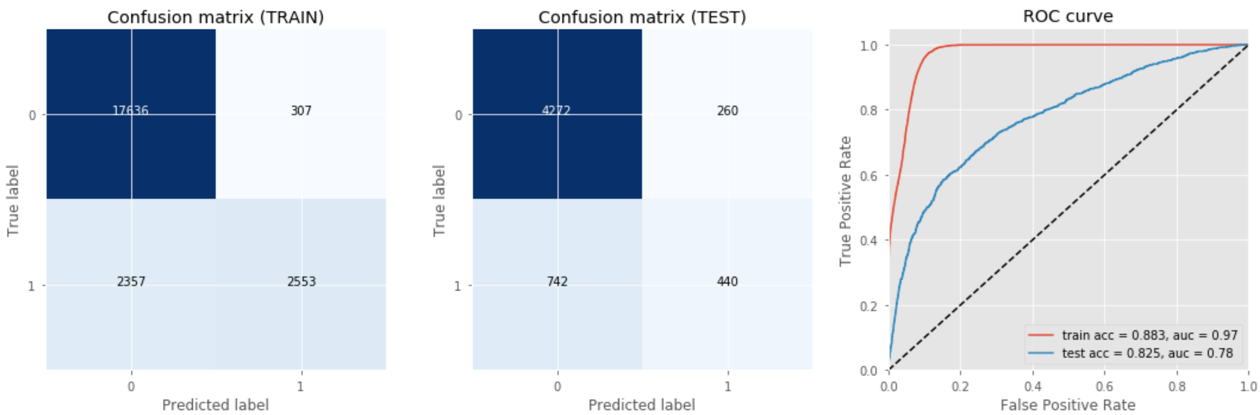

Fig. 7. Confusion matrix for the test and train data for the random forest model.

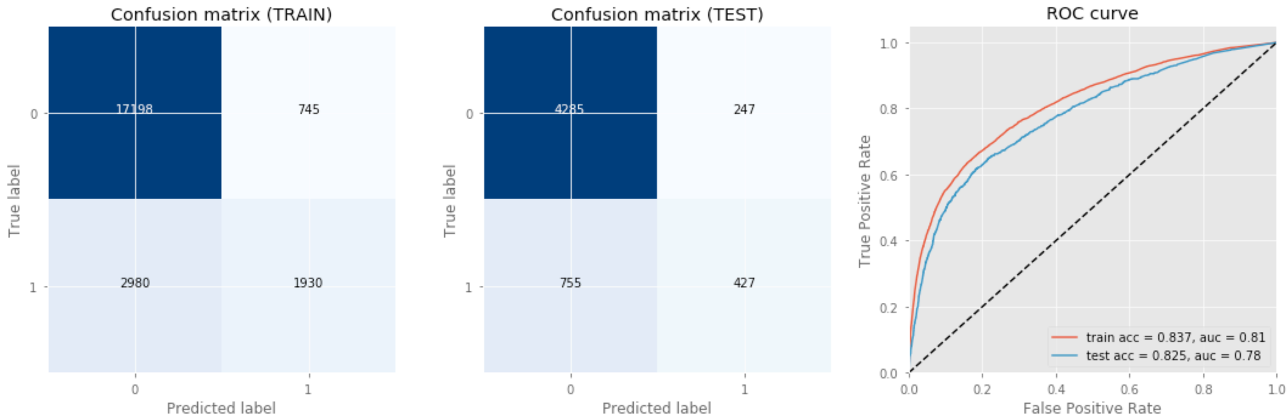

Fig. 8. Confusion matrix of the train and test data for the decision tree boosting model.

for train and test data, which is again in accordance with the values of accuracy.

Neural networks should have the highest accuracy amongst all models according to the established literature. The accuracy for credit scoring model using neural networks has the highest accuracy of 0.829 for both test. The train data accuracy is found to be 0.829. The confusion matrix and ROC curve are depicted in Fig. 9. The AUC values are 0.78 and 0.77 for test and train data.

V. Conclusion and Future Work

This research study assessed the different statistical techniques and established the importance of data analysis in the credit scoring model building process. Furthermore, the reviewed literature illustrates the importance of such credit scoring models and establishes that these models have an edge over personal judgement. Data analysis was carried in this study to identify the influential attributes and disregard the ones that don't have any effect on credit score prediction.

Demographics are found to be the least influential attributes contrasting education when the payments (bill amounts or repayments) are the most influential ones. A comparison of the non-parametric and parametric models reveals that the parametric models like logistic regression seem to have a higher accuracy than the non-parametric ones like random forests. Moreover, the difference in the test and train data accuracies for non-parametric models stresses the sub-optimal performance of such models.

In most of the studies, the interrelated aspects between different attributes that span over different domains is often neglected, for instance, analyzing personal bankruptcy from a social science perspective. Future work could integrate these aspects with existing models to form effective indicators and hence, improve the overall prediction accuracy.

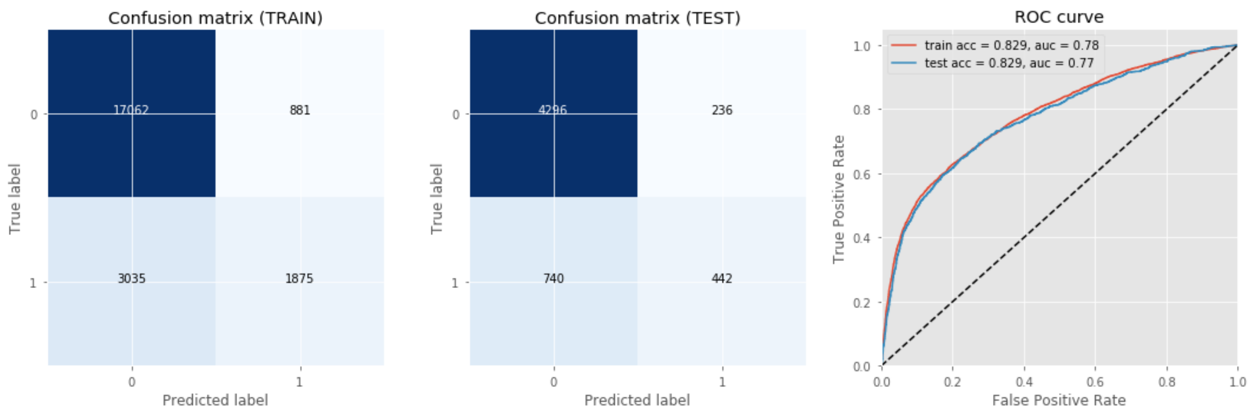

Fig. 9. Confusion matrix for the train and test data for neural networks.

REFERENCES

[1] A. K. Reichert, C.-C. Cho, and G. M. Wagner, "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models," *J. Bus. Econ. Stat.*, vol. 1, no. 2, pp. 101–114, Apr. 1983.

[2] N. C. Hsieh, "Hybrid mining approach in the design of credit scoring models," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 655–665, 2005.

[3] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Syst. Appl.*, vol. 78, pp. 225–241, 2017.

[4] J. Abellán and J. G. Castellano, "A comparative study on base classifiers in ensemble methods for credit scoring," *Expert Syst. Appl.*, vol. 73, pp. 1–10, 2017.

[5] J. Banasik, J. Crook, and L. Thomas, "Sample selection bias in credit scoring models," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 822–832, 2003.

[6] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," *Int. J. Forecast.*, vol. 16, no. 2, pp. 149–172, Apr. 2000.

[7] C. H. Yeh, I. C., & Lien, "UCI Machine Learning Repository: default of credit card clients Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients. [Accessed: 22-Aug-2017].

[8] I. C. Yeh and C. h. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2 PART 1, pp. 2473–2480, 2009.

[9] W. Gaul *et al.*, *Studies in Classification , Data Analysis , and Knowledge Organization Managing Editors*. 2006.

[10] A. Keramati and N. Yousefi, "A proposed classification of data mining techniques in credit scoring," *Proc. 2011 Int. Conf. Ind. Eng. Oper. Manag. Kuala Lumpur, Malaysia*, pp. 416–424, 2011.

[11] T. Credit and S. Toolkit, *The Credit Scoring Toolkit*. 2007.

[12] B. Gup and K. JW, *Commercial Banking: The Management of Risk.* 2005.

[13] L. Thomas, D. Edelman, and J. Crook, *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, 2002.

[14] L. J. Mester, "What ' s the Point of Credit Scoring ?," *Bus. Rev.*, no. January, pp. 3–16, 1997.

[15] H. A. Abdou and J. Pointon, "Credit scoring statistical techniques and evaluation criteria: A review of the literature," *Intell. Syst. Accounting, Financ. Manag.*, vol. 18, no. 2–3, pp. 59–88, 2011.

[16] N. Skantzos and N. Castelein, "Credit scoring Case study in data analytics," *Deloitte*, no. April, 2016.

[17] T. S. Lee, C. C. Chiu, C. J. Lu, and I. F. Chen, "Credit scoring using the hybrid neural discriminant technique," *Expert Syst. Appl.*, vol. 23, no. 3, pp. 245–254, 2002.

[18] D. West, "Neural network credit scoring models," *Comput. Oper. Res.*, vol. 27, no. 11–12, pp. 1131–1152, Sep. 2000.

[19] R. M. Oliver, "Financial performance measures in credit scoring," *EURO J. Decis. Process.*, vol. 1, no. 3, pp. 169–185, 2013.

[20] "NeuralTools: Sophisticated Neural Networks for Excel - Palisade." [Online]. Available: http://www.palisade.com/neuraltools/. [Accessed: 18-Aug-2017].

[21] A. E. Khandani, A. J. Kim, and A. W. Lo, "Consumer Credit Risk Models via Machine-Learning Algorithms," pp. 1–55, 2010.

[22] Y. E. Orgler, "A Credit Scoring Model for Commercial Loans," *J. Money, Credit Bank.*, vol. 2, no. 4, pp. 435–445, 1970.

[23] Y. E. Orgler, "Evaluation of bank consumer loans with credit scoring models," *J. Bank Res.*, pp. 31–37, 1971.

[24] H. A. Abdou, "An evaluation of alternative scoring models in private banking," *J. Risk Financ.*, vol. 10, no. 1, pp. 38–53, Jan. 2012.

[25] R. A. Eisenbeis, "Problems in applying discriminant analysis in credit scoring models," *J. Bank. Financ.*, vol. 2, no. 3, pp. 205–219, 1978.

[26] D. J. Hand and W. E. Henley, "Statistical Classification Methods in Consumer Credit Scoring: a Review," *R. Stat. Soc.*, vol. 160, no. 3, pp. 523–541, Sep. 1997.

[27] M. Zekic-susac, "Small Business Credit Scoring : A Comparison of Logistic Regression , Neural Network , and Decision Tree Models," *Business*, 2004.

[28] S. Bhatia, P. Sharma, R. Burman, S. Hazari, and R. Hande, "Credit Scoring using Machine Learning Techniques," *Int. J. Comput. Appl.*, vol. 161, no. 11, pp. 975–8887, 2017.

[29] A. Niculescu-Mizil and R. Caruana, "Obtaining Calibrated Probabilities from Boosting," *Uai*, pp. 413–420, 2005.

[30] C. Ong, J. Huang, and G. Tzeng, "Building credit scoring models using genetic programming," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 41–47, Jul. 2005.

[31] H. A. Abdou, M. D. D. Tsafack, C. G. Ntim, and R. D. Baker, "Predicting creditworthiness in retail banking with limited scoring data," *Knowledge-Based Syst.*, vol. 103, pp. 89–103, Jul. 2016.

[32] "UCI Machine Learning Repository: Data Sets," *UCI*, 2015. [Online]. Available: https://archive.ics.uci.edu/ml/datasets.html. [Accessed: 16-Aug-2017].