Evaluating Consumer Loans Using Neural Networks Ensembles

Maher Alaraj, Maysam Abbod, and Ziad Hunaiti

***Abstract*---** Banks should take high care on their loan granting policies. Banks rely on credit scoring systems when it comes to granting loans to customers to reduce any potential losses. Neural networks are considered as mostly wide used statistical tool in finance and business applications. Recent studies emphasize using ensemble models or multiple classifiers over single ones to solve credit scoring problems. This study focuses on 2 parts: (1) developing a group of neural network ensemble models to assist in predicting the probability of default; (2) exploring the importance of each of the attributes of the dataset in the overall performance of the classifiers. The datasets used in this study are real world datasets.

***Keywords*---** Credit Scoring; Neural networks, Ensemble Technique, Feature selection.

I. Introduction

FOR banks or any financial institution, credit lending activities are the principal of their business. However, good lending action will lead to high profits otherwise loss will take place. In order to minimize risk and to choose where the money should be granted, a critical evaluation of loan applications should be carried out to come to a reliable and effective decision. According to Basel Committee for Banking supervision [1], all banks should adopt rigorous credit scoring systems to help them in estimating degrees of credit risk and different risk exposures, as well as improving capital allocation and credit pricing. However, and due to the substantial growth in the consumer credit industry [2] and the promoting of wide range of products and financial service to the clients [3] this resulted in an increasing demand on consumer loans and having high competitive market, along with this, the chances of defaulting risk will increase [4], and from here embracing of a credit scoring systems is an essential tool in evaluating the credibility of applicants.

Credit Scoring has been a challenging topic and widely discussed in the field of credit management and in literature, and a lot of attempts have considered developing scoring Models to solve such problem, see [5]-[7], for example

Credit scoring objective is to determine if loan applicant should belong to either ‘Good’ or ‘Bad group of applicants [8], [9]. ‘Good’ applicants are who could fulfill their obligations while ‘Bad’ could not. Credit scoring systems can be classified in to two major approaches [4], [10]: Judgmental approach and statistical approach, the first one relies solely on guidelines of the bank in addition to the intuition and the experience of the credit analyst. This approach is not efficient as it is subject to human error or bias and it is time consuming [6]. [11] Argued that the results of judgmental approach are inefficient, inconsistency and un-uniform. However, due to the fast growth of consumer credit industry and enormous amount of data, banks and financial institutions tend to have objective, fast, consistent and uniform methods to replace the traditional ways.

Supplement the judgmental [12]. The second approach is used to overcome the former one; it is based on building scoring models using statistical techniques which are known in its ability to differentiate between two populations such as linear discriminant analysis (LDA) and logistic regression (LR) as traditional methods and advanced methods such as neural networks (ANN) and support vector machines (SVM), and decision tree (DT). Credit scoring models are developed in a way that can identify the financial variables that have discriminatory power to classify applicants in two groups. The advantages of having a consistent credit scoring model systems are reducing of credit analysis costs, allow faster decisions, guarantee credit collectivity and reduce any exposure to risks [5], [13]. Accuracy is a major issue when building credit scoring models, as it affects banks' profits in case of default. Yet, increasing the accuracy of the scoring systems, even of 1% could save the banks from great losses, especially for bad applicants [14]. Many credit scoring models have been developed by researches and practitioners.

The most common employed models are LDA and LR [15]-[17], [5], [18]. In credit scoring both LDA and LR tends to look for the best linear combination between variables, but the weaknesses occur when the data are non-linearly related, and for that reason they are considered to have lack of accuracy [19]. The evolution of new technologies allowed for building and developing more robust and rigors models [20]. To overcome the shortcomings of LDA and LR, new advanced statistical techniques were revealed as alternatives to traditional methods in developing credit scoring models and other areas of finance, such as ANN [4], [21], [22]- [16], SVM [10], [23], [16], GP [8], and DT [24]. These methods are called machine learning method as they can learn from the data and extract knowledge from it, and build properties that are relevant to data target. However, these methods showed that

Maher Alaraj is with Electrical and Computer Engineering Department, Brunel University, UB8 3PH, London, UK (Tel: +447450601162; e-mail: maher.ala’raj@brunel.ac.uk).

Maysam Abbod is with Electrical and Computer Engineering Department, Brunel University, UB8 3PH, London, UK (e-mail: maysam.abbod@ brunel.ac.uk).

Ziad Hunaiti is with Electrical and Computer Engineering Department, Brunel University, UB8 3PH, London, UK (e-mail: ziad.hunaiti@ brunel.ac.uk).

they are very superior in dealing with credit scoring problems, especially for nonlinear relationships between variables and for its generalization abilities and it are superior to LDA and LR [22], [25]. Although, of the promising results achieved by related studies, and its ability to hold classification problems, there was a trend by scholars for having more developed models with improved accuracy. [18] Pointed that more customized designs must be developed in credit scoring.

Still, researchers have been developing scoring models based on new techniques, such hybridization and ensemble techniques which show higher accuracy than individual models. Hybrid approaches is mainly based on combining two different classifiers or machine learning models together. When the two models are combined, one is used to do data per-processing (un-supervised learner) and then the second classifier (supervised learner) will learn from the pre-processed data, see [26] for explore different ways to build hybrid models. The motivation behind building hybrid models is the data pre-processing or (variable selection) of significant variables. Variable selection or feature selection is an important problem in credit scoring, it is based on reducing the variables that is irrelevant, redundant or causing noise, which it may affect classification accuracy, although most developed scoring models are built without any modifications, variable selection has showed its effectiveness in increasing accuracy as well as having less complex and low dimensional model [27]. Using variable selection does not only increase model accuracy but it gives better initial solution for the model used [14], several studies have proposed effective hybrid models in credit scoring using different classifiers [13], [17], [28], [14], [29], [26], [30], all related studies that concluded that hybrid models is the best for credit scoring in terms capability, misclassification costs, making profits and generating business revenues when compared to single classifiers.

The other technique which was inspired by hybridization approach is the ensemble techniques, which are based on training several classifiers to solve the same problem, and then their outputs, are combined or aggregated by a method(s) to give an enhanced output. As the ensemble approaches are the main focus of this paper, some literature will be covered, however all related studies showed that ensemble method is efficient and gives higher classification accuracies than that of single models.

However, there is no overall, best classification technique used in building credit scoring model, selecting a model that could discriminate between two groups, depends on the nature of the problem, data structure, variables used and the market and environment it will be developed [12], [31], [24]. Hereafter, there can be a right model for the right purpose, and any credit scoring model when it is developed, must take in consideration the problem and how to fit it based the data availability, computational capacity and the model should be easy and understandable on how it works, the model should be adaptable with all conditions and through time and it should also extends to be profitable and detect problems. The rest of paper is divided as the follows: section 2 reviews some related work on ensemble techniques applied in credit scoring. Section 3 describes NN and its architecture and an overview of ensemble strategies. Section 4 states the experimental design that will be conducted. Section 5 reports the results and analysis of the experiments. Section 6 draws conclusions and future direction of research.

II. Related Work

Building efficient credit scoring model is quite a big challenge. The number of credit scoring models being developed by researchers is increasing larger and larger. Their research motivation is improving models accuracy that reduces risk, increase profit and helps in efficient decision making. The research trend is going towards hybrid and ensemble learning, because of its effectiveness in handling the credit scoring. Therefore, models are getting more difficult and complex, in addition to higher costs of implementation.

Many studies have highlighted the use of several ensembles classifiers in building credit scoring models. [7] tested the performance of single NN classifier and compare it with multiple and diversified NN classifiers, the results showed that there is no exactly better classifier. [32] They used three NN ensemble strategies (cross-validation, bagging and boosting), the results showed that NN ensembles are more accurate, robust and superior than single NN. [33] Investigated the performance of several ensemble classifiers, results showed superiority of ensemble in terms of classification accuracy. [34] Investigated five classifiers with different noise levels of attributes, and try to improve the accuracy their ensembles, results revealed that ensemble classifiers are more accurate at different noise levels. Also a study by [35] were they added a new group beside 'good' and bad', which is 'borderline' this is based on initial information of applicants, they performed a class-wise classification as pre-processing for data, instead of using ensemble classifier without pre-processing, this will lead to an efficient ensemble classifiers, results shows better accuracy than hybrid and single classifiers. A study by [36] was they built an ensembles of LS-SVM, each SVM agents is given a weight using three ensemble strategies, results showed that ensemble strategies improve accuracy and it is a good tool or building credit scoring models. [24] Carried out a comparative study on three classifiers based on three ensemble strategies namely bagging, boosting and bagging, results shows that bagging performs boosting all the time. Another study conducted by [37], they applied a new bagging procedure called 'poly-bagging' based on combining predictors over a succession resampling, results showed that this approach may approve the accuracy of scoring models. A recent study by [38] used DT in developing the credit scoring models bas on dual strategy ensemble trees, based on Random space- Bagging DT and Bagging-Random Space DT, the random subspace bagging are used to reduce noisy and redundant data which affects DT, the results showed that the proposed methods outperform many other classifiers used in the study. Accordingly, all the above studies afford evidence that ensemble techniques can be useful in building credit scoring models.

III. Overview on Artificial Neural Network and Ensemble Learning Strategies

3.1 Artificial Neural Network (ANN)

As ANN will be our base classifier to construct the ensemble models, an explanation of ANN will be given first in terms of: what it is? its purposes, its mechanism, advantages and critiques and its application on other disciplines other than credit scoring. ANN is basically is an advanced soft statistical technique based on the concept of artificial intelligent on how to function like human brain. It has been extensively used in building credit scoring models as an alternative to the traditional techniques (LDA, LR). Neural networks are defined as “an artificial intelligence problem solving computer program that learns through a training process of trial and error” [39] p.147. A neural network model contains are made up of 3 layers input, output and one or more hidden layers between the input and output [14], [40], [5]. Neural networks model process information through the connections of a large number of processing units. The structure of the neural network model for credit scoring starts from the input layer where it takes inputs (customers’ characteristics) to process them, then the inputs are passed to the hidden layer where the input values are further processed, then the values are sent to the output layer which gives the final answer for problem either good or bad loan. See [5] p.87 for more details of the training process of the NN layers. The topology of the neural network model is show in Figure 1.

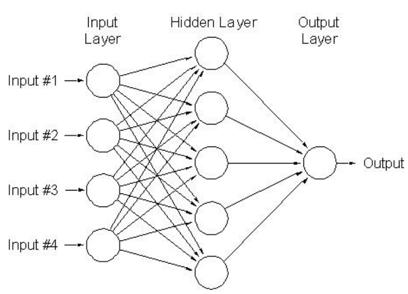

Fig. 1 A three layer backpropagation neural networks [5]

Building a neural network involves training procedure for the variables in order to differentiate between that to get a better decision and results. However, it the results are improper, the estimated values will be changed by the neural network model until they become proper and acceptable [41]. The neural networks tend to find the relationship between a customers’ probability of default and their given characteristics, Filter them and choose among the most important prediction variables. The major advantages of neural network models that it can handle incomplete, missing or noisy data, requires no prior assumptions about the distribution of the variables; it can recognize complex patterns between variables [42]. Conversely, neural networks is critiqued that it lacks explanatory capability, such that it cannot give an explanation or justification on how customer is chosen as ‘good’ or ‘bad’, accepted or rejected, another shortcoming in neural networks is the selection of parameters, as there is no official method to select the appropriate parameters for the model in which it might affect its prediction accuracy [42], [27]. Moreover, the decision of network topology which is important for accuracy and the problem long training process are also criticized. Some authors reported that neural networks outperform traditional techniques in prediction accuracy [5], [16], [18], [4], [43]-[44], [19]. However, neural networks are widely used in other disciplines such as stock markets, medicine, engineering, education [13]. In the study a feedforward backpropagation neural network model will be employed along with ensemble classifiers with different ensemble techniques.

3.2 Ensemble learning strategies

Ensemble learning is a machine learning paradigm where multiple classifiers are trained to solve same problem, in compare single classifiers only learn and trained on one set of data, but ensemble classifiers learn and train on diversified sets of data generated from the original dataset, which in a result will built a set of hypothesis from the data trained, which will lead to a better accuracy. In order to get single output from the ensemble models, results are combined using certain strategies for instance, majority-vote, weighted-average or average mean. In this study, the mean will be used since this study is one part of further research, when other strategies will be examined.

According to [33], the focal idea for these methods is to modify the training data, in a way to guarantee that the classifiers are trained on diversified data (i.e. each model should be trained on different data from the other model), hence, different hypothesis will constructed and different results will come out. It is worth to mention that generalization ability of ensemble classifier is stronger than single classifier, as the weakness in ensemble learner can be recompensed by other members of the ensemble [45]. Two important things should be considered to achieve good ensemble learner: accuracy and diversity [46].

To ensure diversity of data training among classifiers, several ensemble strategies can be used such as cross-validation, bagging, boosting and stacking. Due to space restriction, only the adopted methods will be explained.

3.2.1 Cross-validation

This method of sample variation is the simplest and most widely used method of training data used in the literature [47]. Its main objective is to divide the dataset into several k-fold partitions (k = 1, 2, 3, …, n) and each partition uses random data generated from the original data set. Each k-fold acts like an independent validation set for the model trained by the k-1folds. The benefits of this method are the data dependency is reduced and the consistency of results can be improved [48].

3.2.2 Bagging

It is a shortcut for bootstrap aggregation; it is the one of the first ensemble learning algorithms [49]. It is easy to implement it and it gives good performance results. The diversity of data trained is achieved by creating different number of bags (n-bags) each bag is filled by data randomly generated from training dataset, with replication. For instance each bag could have data selected more than once or not selected at all, and each bag is trained by only one classifier. In order to ensure the most diversification of data to be trained the next two proposed methods to be explained are might be introduced for the first time in literature of credit scoring, to the best of our knowledge.

3.2.3 CV-Bagging

This method is based on combing the cross validation technique with bagging technique. The way it works is the following; first, the training data is divided into k partitions, each partition randomly generate data from the training set, according to the standard CV technique described above. Then the bagging technique and n-bags are created randomly by selecting data from each partition. Thus, from the data in each bag, in each partition, a different NN model will be trained. For example, if there are 10 partitions and 100 bags in each partition, in total there will be 10x100 = 1000 neural networks trained.

3.2.4 Bagging-CV

This method is similar to the previous one, but the order by which CV and bagging are applied is reversed. Initially, n bags are created by randomly sampling the training set data with replication. Next, the data in each bag is used for CV, which essentially provides k partitions for each one of the bags. At the end, k × n neural networks will be trained, one from the data in each partition of each of the bags. It will be demonstrated that these proposed hybrids of the well-established ensemble methods result in higher diversification of classifiers. Their performance is at least equivalent to the standard ensemble methods and in some cases even higher. At the same time, the standard deviation of the results is greatly reduced, which hints at more stable classifiers.

IV. Experimental Design

4.1 Real work credit datasets

Two real world credit datasets will be used to evaluate the performance of the predictive accuracy of the four NN ensemble methods along with the single NN. German and Australian datasets are available publicly online from UCI machine learning depository [50]. These credit sets have been the pillars of developing credit scoring models among researchers. Summary of the datasets is given in Table I.

TABLE I
DESCRIPTION OF THE DATASETS USED IN THE STUDY

Credit set	Loan cases	Good/Bad	Attributes No.	Variables: Categorical / continuous
German	1000	700/300	20	13/7
Australian	690	307/382	14	6/8

4.2 Data normalization

The datasets contains inputs that hold values that will be fed to the NN. Each attribute in the dataset contains values that vary in range. In order to avoid bias and feed the network with data within the same interval, data should be transformed from different scale of values to a common scale values. To achieve this dataset attributes are normalized to values in the range between 0 and 1. These transformations are done by taking the maximum value within each attribute and then divide all the values in the attributes with its maximum value.

4.3 NN ensemble models

In this paper NN will be adopted as the base classifier for the ensemble models in this study, since 80% of business application were developed using NN [42], [47]. In NN it is important to decide the architecture of the network topology. Among the different network topologies test, the one selected for this study consists of one input layer, two hidden layers and one output neuron. According to [16] one hidden neuron is sufficient to have model with the desired accuracy, [18] stated that if to avoid long training time, another hidden layer can be added, which it is in our study. Some trials and error [51] were conducted to determine the best number of neurons in the hidden layer, a range of 20-40 in both hidden layers have been tested. The NN stopped learning or training after reaching the pre-determined number of epochs. The optimal topology was selected based on: the lowest (Mean Square Error), and the highest accuracy, especially for bad loans. In the experiment the datasets were divided in two sets, Training (80%) to train the NN, and Testing (20%) to test the trained NN.

4.4 Exploring significant variables

The features or variables of a used dataset may have an important effect on the performance of credit scoring models [52]. Referring to [53] and [13], they stated that NN might not easily recognize the relative importance of the inputs variables. For this reason, in this study it is intended to explore the importance and significance of the inputs variables in both datasets. After calculating the accuracies of the ensemble models, two tests will be carried out on the input variables to investigate how effective they are and whether there are significant variables that could affect the accuracy by increasing or decreasing it. The first test (1) consists of keeping the value of each variable fixed and then performing the training and classification to see whether this has any significant effect on the accuracy of our classifiers. All the variables are considered one at a time; each time the values of the variable is fixed for all contracts to be the mean/median value for all contracts. Then, the training and classification is conducted and the impact this variable on the accuracy results is recorded. The mean square error (MSE) is then compared to the MSE of the classifier when all variables are taken into account. If the error rate of the model increases when a given variable is fixed then that variable is deemed to be significant, otherwise it is considered insignificant.

In the second test, all the variables are tested again, but this time all the rest are fixed to their mean/median values and only this one is allowed to vary. A range of 10 testing values is created, equally spaced between the minimum and maximum values for this specific variable and then run the classifier 10 times, once for each value of the variable in the testing range. It is then assessed how much sensitive the accuracy of the classifier is to the value of this variable. The greater the sensitivity, the more important this variable is for our classifier.

V. Experiment Results and Analysis

The whole experiments in this this study were executed using Matlab 2013b version, on a PC with 3.4 GHz, Intel CORE i7 and 8 GB RAM, using Microsoft Windows 7 operating system.

5.1 Experiment I results

Our aim in this experiment is to show how ensembles methods perform against single model, and also to show how modifying training data in different and not complicated way, could give results comparable or better than other models proposed in the literature. Results of NN ensembles along with the single NN are compared with each other, the testing set results will be shown with their standard deviation calculated to explore how stable the results were at each group of ensemble. The ensemble model results are based on 100 simulations. The credit scoring results of the resting set are summarized in Tables II and III for German and Australian datasets respectively.

TABLE II
CLASSIFICATION RESULTS FOR THE NN ENSEMBLES (GERMAN DATASET)

NN Models	Average classification accuracy (%)	Standard deviation (SD)
Single	73.73	0.0386
Cross Validation (CV)	78.00	0.0141
Bagging	77.65	0.0106
Bagging-CV	77.35	0.0073
CV-Bagging	78.38	0.0052

As it can be seen from Table II, Cross-Bagging has achieved the highest accuracy and the lowest SD 78.38% and 0.52% respectively, followed by CV with accuracy of 78% and SD of 1.4%, and then comes Bagging with accuracy and SD of 77.65%, 1% respectively. Bagging-Cross got 77.35% accuracy and 0.073% SD. Last is the Single NN with lowest accuracy and SD of 73.73% and 3.86% respectively, as expected. Although the proposed Bagging-Cross got accuracy slightly lower than CV and Bagging, but the interesting thing that it's SD is lower than both, which gives indication that it was more stable in its accuracy over the 30 runs carried out, and the other maybe were stochastic over the 30 runs.

TABLE III
CLASSIFICATION RESULTS FOR THE NN ENSEMBLES (AUSTRALIAN DATASET)

NN Model	Average classification accuracy (%)	Standard deviation (SD)
Single	83.24	0.0249
Cross Validation (CV)	85.45	0.0144
Bagging	83.74	0.0065
Bagging-CV	84.13	0.0055
CV-Bagging	82.90	0.0056

From the results in Table IV, it can be observe that the best accuracy rate is CV with 85.45% and the worst is proposed Cross-Bagging method with 82.9%. Bagging-Cross achieved accuracy of 84.13%, followed by 83.74% and 83.24% for Bagging and Single NN respectively. The SD for the proposed methods is still more stable than the other ensembles, although their accuracy is lower than CV. In summary, two new ensemble techniques for classifying credit loans were proposed. The proposed methodologies have been shown to be producing comparable or better results than other approaches in the literature with the added advantage of having a much smaller variability in the accuracy results. This is a significant advantage for the industry, as it increases the robustness of the classifier.

To validate the results, the achieved results were compared of the results obtained by scholars reported in the literature. The results are summarized in Table IV.

TABLE IV
RESULTS FOR RELATED CREDIT SCORING MODELS

Study	Dataset	Proposed model results
Wang et al., 2012 (ANN)	Ger/Aus	71.43, 75.56, 73.3, 75.97/83.28, 85.01, 83.65, 86.57
Wang et al., 2011	Ger/Aus	78.36/88.17
Chen and Li, 2010	Ger/Aus	76.7/85.10
Zhou et al., 2010	Ger	76.36
Lee and Chen 2005	Aus	84.71
West et al., 2005	Ger/Aus	75.82, 74.87, 73.53 74.69/86.28, 87.21, 85.24, 86.78
West, 2000 (ANN)	Ger/Aus	77.57/87.8

5.2 Experiment II results

Figure 2 shows the German dataset MSE when one attribute is fixed to its median value. It can be seen than there is no single attribute stands out from the crowd. The small differences are not statistically significant. On the other hand, Figure 3 shows that there is a single attribute in the Australian dataset; attribute 8 which represent a categorical value, its removal would greatly impact on the MSE. Therefore, it can be concluded that this is a significant attribute which contributes to the overall accuracy of the classifiers.

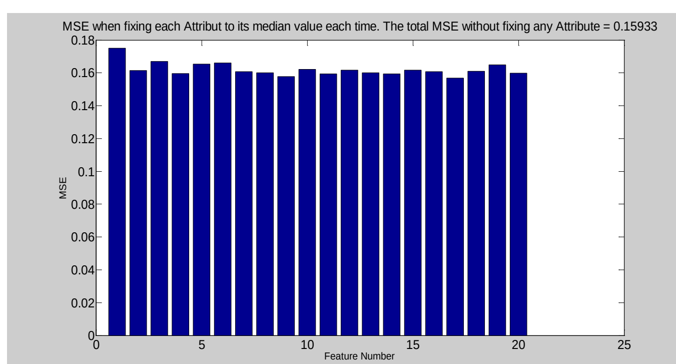

Fig. 2 First significant test of variables in German dataset

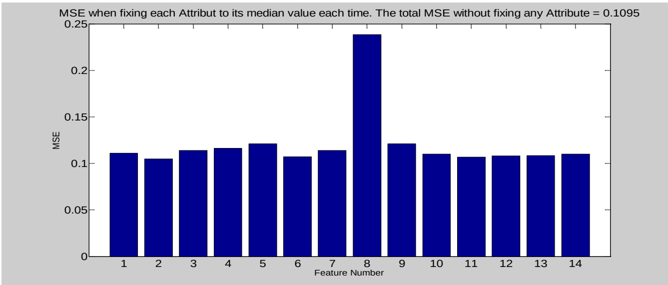

Fig. 3 First significant test of variables in Australian dataset

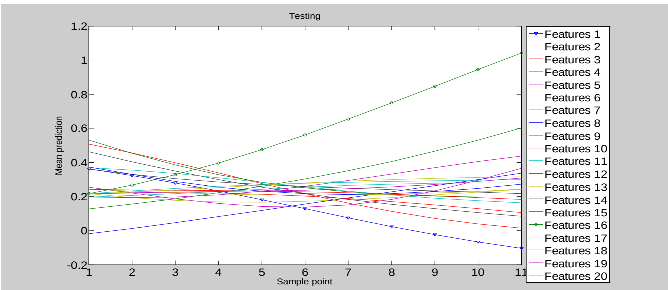

Fig. 4 Second significant test of variables in German dataset

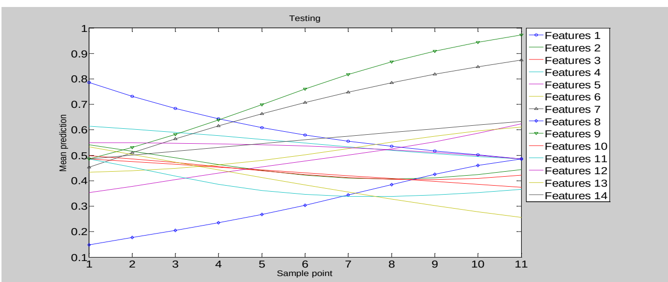

Fig. 5 Second significant test of variables in Australian dataset

Figures 3 and 4 show the sensitivity of the classifier to different values of each attribute, when all others are fixed. For most of the attributes, it appears the prediction of the classifier is relatively stable. However, for the German dataset, the accuracy of the classifier seems to change significantly for the values of attributes 1 and 16 respectively. On the other hand it can be seen that four attributes of the Australian dataset are sensitive to the classifier. The attributes are 1, 7, 8 and 9 respectively.

VI. Conclusion and Future Research

In this study, two new ensemble methods for classification of credit loan customers based on neural networks were proposed. It can be shown that the proposed methodologies provide similar or better accuracy results compared to the literature as well as to single classifier, and they have sown to be more stable than the other methods. In addition, the sensitivity of these classifiers with respect to each attributes has been examined for both datasets. Two data testing results using single NN were tested, and ensemble methods which showed that the models are stable and have no serious effect on the classifiers accuracy.

Future work could be extended by (1) using different credit datasets with different sizes and attributes for extra validation, (2) using different ensemble methods and strategies to learn more diversified sets (3) using different machine learning methods such as SVM, DT.

References

[1] Basel committee on banking supervision 2000 - "Credit Ratings and Complementary Sources of Credit Quality Information." Bank for International Settlements, Basel Committee on Banking Supervision, Working Paper 3. Available online from: http://www.bis. org/publ/bcbs72a.pdf.

[2] D. J. Hand, "Modelling consumer credit risk," IMA Journal of Management Mathematics, vol. 12, pp. 139-155, 2001. http://dx.doi.org/10.1093/imaman/12.2.139

[3] R. Anderson, The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press, 2007.

[4] R. Malhotra and D. Malhotra, "Evaluating consumer loans using neural networks," Omega, vol. 31, pp. 83-96, 2003. http://dx.doi.org/10.1016/S0305-0483(03)00016-1

[5] D. West, "Neural network credit scoring models," Comput. Oper. Res., vol. 27, pp. 1131-1152, 2000. http://dx.doi.org/10.1016/S0305-0548(99)00149-5

[6] M. Handzic, F. Tjandrawibawa and J. Yeo, "How neural networks can help loan officers to make better informed application decisions," Informing Science, vol. 6, pp. 97-109, 2003.

[7] C. Tsai and J. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," Expert Syst. Appl., vol. 34, pp. 2639-2649, 2008. http://dx.doi.org/10.1016/j.eswa.2007.05.019

[8] C. Ong, J. Huang and G. Tzeng, "Building credit scoring models using genetic programming," Expert Syst. Appl., vol. 29, pp. 41-47, 2005. http://dx.doi.org/10.1016/j.eswa.2005.01.003

[9] B. W. Yap, S. H. Ong and N. H. M. Husain, "Using data mining to improve assessment of credit worthiness via credit scoring models," Expert Syst. Appl., vol. 38, pp. 13274-13283, 2011. http://dx.doi.org/10.1016/j.eswa.2011.04.147

[10] S. Li, W. Shiue and M. Huang, "The evaluation of consumer loans using support vector machines," Expert Syst. Appl., vol. 30, pp. 772-782, 2006. http://dx.doi.org/10.1016/j.eswa.2005.07.041

[11] V. Limsombunchai, C. Gan and M. Lee, "An analysis of credit scoring for agricultural loans in Thailand," American Journal of Applied Sciences, vol. 2, pp. 1198, 2005. http://dx.doi.org/10.3844/ajassp.2005.1198.1205

[12] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: a review," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 160, pp. 523-541, 1997. http://dx.doi.org/10.1111/j.1467-985X.1997.00078.x

[13] T. Lee, C. Chiu, C. Lu and I. Chen, "Credit scoring using the hybrid neural discriminant technique," Expert Syst. Appl., vol. 23, pp. 245-254, 2002. http://dx.doi.org/10.1016/S0957-4174(02)00044-1

[14] Y. Chen, R. Guo and R. Huang, "Two stages credit evaluation in bank loan appraisal," Econ. Model., vol. 26, pp. 63-70, 2009. http://dx.doi.org/10.1016/j.econmod.2008.05.008

[15] D. Durand, "Risk elements in consumer instalment financing," 1941.

[16] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," J. Oper. Res. Soc., vol. 54, pp. 627-635, 2003. http://dx.doi.org/10.1057/palgrave.jors.2601545

[17] T. Lee and I. Chen, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines," Expert Syst. Appl., vol. 28, pp. 743-752, 2005. http://dx.doi.org/10.1016/j.eswa.2004.12.031

[18] 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," Eur. J. Oper. Res., vol. 95, pp. 24-37, 1996. http://dx.doi.org/10.1016/0377-2217(95)00246-4

[19] M. Šušteršič, D. Mramor and J. Zupan, "Consumer credit scoring models with limited data," Expert Syst. Appl., vol. 36, pp. 4736-4744, 2009.

http://dx.doi.org/10.1016/j.eswa.2008.06.016

[20] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," Int. J. Forecast., vol. 16, pp. 149-172, 2000. http://dx.doi.org/10.1016/S0169-2070(00)00034-0

[21] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," Neural Networks, IEEE Transactions on, vol. 12, pp. 929-935, 2001. http://dx.doi.org/10.1109/72.935101

[22] Z. Huang, H. Chen, C. Hsu, W. Chen and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," Decis. Support Syst., vol. 37, pp. 543-558, 2004. http://dx.doi.org/10.1016/S0167-9236(03)00086-1

[23] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," Expert Syst. Appl., vol. 36, pp. 3302-3308, 2009. http://dx.doi.org/10.1016/j.eswa.2008.01.005

[24] G. Wang, J. Hao, J. Ma and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," Expert Syst. Appl., vol. 38, pp. 223-230, 2011. http://dx.doi.org/10.1016/j.eswa.2010.06.048

[25] V. S. Desai, D. G. Conway, J. N. Crook and G. A. Overstreet, "Credit-scoring models in the credit-union environment using neural networks and genetic algorithms," IMA Journal of Management Mathematics, vol. 8, pp. 323-346, 1997. http://dx.doi.org/10.1093/imaman/8.4.323

[26] C. Tsai and M. Chen, "Credit rating by hybrid machine learning techniques," Applied Soft Computing, vol. 10, pp. 374-380, 2010. http://dx.doi.org/10.1016/j.asoc.2009.08.003

[27] A. Lahsasna, R. N. Ainon and T. Y. Wah, "Credit Scoring Models Using Soft Computing Methods: A Survey." Int.Arab J.Inf.Technol., vol. 7, pp. 115-123, 2010.

[28] C. Huang, M. Chen and C. Wang, "Credit scoring with a data mining approach based on support vector machines," Expert Syst. Appl., vol. 33, pp. 847-856, 2007. http://dx.doi.org/10.1016/j.eswa.2006.07.007

[29] F. Chen and F. Li, "Combination of feature selection approaches with SVM in credit scoring," Expert Syst. Appl., vol. 37, pp. 4902-4909, 2010. http://dx.doi.org/10.1016/j.eswa.2009.12.025

[30] C. Chuang and S. Huang, "A hybrid neural network approach for credit scoring," Expert Syst., vol. 28, pp. 185-196, 2011. http://dx.doi.org/10.1111/j.1468-0394.2010.00565.x

[31] H. A. Abdou and J. Pointon, "Credit scoring and decision making in Egyptian public sector banks," International Journal of Managerial Finance, vol. 5, pp. 391-406, 2009. http://dx.doi.org/10.1108/17439130910987549

[32] D. West, S. Dellana and J. Qian, "Neural network ensemble strategies for financial decision applications," Comput. Oper. Res., vol. 32, pp. 2543-2559, 2005. http://dx.doi.org/10.1016/j.cor.2004.03.017

[33] L. Nanni and A. Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring," Expert Syst. Appl., vol. 36, pp. 3028-3033, 2009. http://dx.doi.org/10.1016/j.eswa.2008.01.018

[34] B. Twala, "Multiple classifier application to credit risk assessment," Expert Syst. Appl., vol. 37, pp. 3326-3336, 2010. http://dx.doi.org/10.1016/j.eswa.2009.10.018

[35] N. Hsieh and L. Hung, "A data driven ensemble classifier for credit scoring analysis," Expert Syst. Appl., vol. 37, pp. 534-545, 2010. http://dx.doi.org/10.1016/j.eswa.2009.05.059

[36] L. Zhou, K. K. Lai and L. Yu, "Least squares support vector machines ensemble models for credit scoring," Expert Syst. Appl., vol. 37, pp. 127-133, 2010. http://dx.doi.org/10.1016/j.eswa.2009.05.024

[37] F. Louzada, O. Anacleto-Junior, C. Candolo and J. Mazucheli, "Poly-bagging predictors for classification modelling for credit scoring," Expert Syst. Appl., vol. 38, pp. 12717-12720, 2011. http://dx.doi.org/10.1016/j.eswa.2011.04.059

[38] G. Wang, J. Ma, L. Huang and K. Xu, "Two credit scoring models based on dual strategy ensemble trees," Knowledge-Based Syst., vol. 26, pp. 61-68, 2012. http://dx.doi.org/10.1016/j.knosys.2011.06.020

[39] E. Gately, Neural Networks for Financial Forecasting. John Wiley & Sons, Inc., 1995.

[40] H. L. Jensen, "Using neural networks for credit scoring," Managerial Finance, vol. 18, pp. 15-26, 1992.

[41] H. A. Abdou and J. Pointon, "Credit scoring, statistical techniques and evaluation criteria: a review of the literature," Intelligent Systems in Accounting, Finance and Management, vol. 18, pp. 59-88, 2011. http://dx.doi.org/10.1002/isaf.325

[42] A. Vellido, P. J. Lisboa and J. Vaughan, "Neural networks in business: a survey of applications (1992–1998)," Expert Syst. Appl., vol. 17, pp. 51-70, 1999. http://dx.doi.org/10.1016/S0957-4174(99)00016-0

[43] H. Abdou, J. Pointon and A. El-Masry, "Neural nets versus conventional techniques in credit scoring in Egyptian banking," Expert Syst. Appl., vol. 35, pp. 1275-1292, 2008. http://dx.doi.org/10.1016/j.eswa.2007.08.030

[44] G. Zhang, M. Y Hu, B. Eddy Patuwo and D. C Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," Eur. J. Oper. Res., vol. 116, pp. 16-32, 1999. http://dx.doi.org/10.1016/S0377-2217(98)00051-4

[45] A. Ghodselahi, "A hybrid support vector machine ensemble model for credit scoring," International Journal of Computer Applications, vol. 17, pp. 1-5, 2011.

[46] L. K. Hansen and P. Salamon, "Neural network ensembles," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 12, pp. 993-1001, 1990. http://dx.doi.org/10.1109/34.58871

[47] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques: A review of applications," Expert Syst. Appl., vol. 36, pp. 2-17, 2009. http://dx.doi.org/10.1016/j.eswa.2007.10.005

[48] S. L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," Data Mining and Knowledge Discovery, vol. 1, pp. 317-328, 1997. http://dx.doi.org/10.1023/A:1009752403260

[49] L. Breiman, "Bagging predictors," Mach. Learning, vol. 24, pp. 123-140, 1996. http://dx.doi.org/10.1007/BF00058655

[50] A. Asuncion and D. Newman, UCI Machine Learning Repository, 2007.

[51] F. S. Wong, "Time series forecasting using backpropagation neural networks," Neurocomputing, vol. 2, pp. 147-159, 1991. http://dx.doi.org/10.1016/0925-2312(91)90045-D

[52] Y. Liu and M. Schumann, "Data mining feature selection for credit scoring models," J. Oper. Res. Soc., vol. 56, pp. 1099-1108, 2005. http://dx.doi.org/10.1057/palgrave.jors.2601976

[53] M. W. Craven and J. W. Shavlik, "Using neural networks for data mining," Future Generation Comput. Syst., vol. 13, pp. 211-229, 1997. http://dx.doi.org/10.1016/S0167-739X(97)00022-8