INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

A Credit Scoring Heterogeneous Ensemble Model Using Stacking and Voting

C J Anil Kumar[1,2]*, B K Raghavendra[3], S Raghavendra[4]

1 Research Scholar, Department of Computer Science and Engineering Research Centre, BGS Institute of Technology, B G Nagara, Mandya, Karnataka, India
2 Associate Professor, Department of Computer Science and Engineering, ATME College of Engineering, Mysuru, Karnataka, India
3 Professor & Head, Department of Computer Science and Engineering, BGS Institute of Technology, B G Nagara, Mandya, Karnataka, India
4 Associate Professor, Department of Computer Science and Engineering, School of Engineering and Technology, CHRIST Deemed to be University, Bengaluru, Karnataka, India

OPEN ACCESS

Received: 15.09.2021
Accepted: 28.01.2022
Published: 23.02.2022

Citation: Anil Kumar CJ, Raghavendra BK, Raghavendra S (2022) A Credit Scoring Heterogeneous Ensemble Model Using Stacking and Voting. Indian Journal of Science and Technology 15(7): 300-308. https://doi.org/10.17485/IJST/v15i7.1715

* **Corresponding author**.

anilkumarcj@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: Recent studies emphasized on using ensemble models over single ones to solve credit scoring problems. The objective of this study is to build a heterogeneous ensemble classifier model with an improved classification accuracy. **Methods:** This study focuses on developing a heterogeneous ensemble classifier using Logistic Regression, K-nearest neighbor, Decision tree, Random Forest, Naïve Base and Support vector machine as base classifiers and Random Forest, Logistic Regression and Support vector machine as meta-classifiers. The proposed model is built using these six base classifiers for ensemble aggregation. A feature selection algorithm based on the random forest technique is used for selecting the best features. A stacking and voting method are used for building ensemble model. **Findings:** The ensemble classifier gives superior predictive performance than single classifiers SVM, DT, RF, NB, KNN and LR with an accuracy of 91.56% for Australian dataset and 84.35% for German dataset. **Novelty:** The proposed model uses stacking and majority voting method for ensemble classification. Initially, stacking is applied to the base classifiers. This is done in two levels. First the training dataset is split into 10 folds for cross validation. The output of each classifier is taken, and the dataset is updated with the meta-features. In the second level, three meta-classifiers (MC), namely LR, SVM and RF are used. Majority voting is applied to the output of these meta-classifiers for the prediction.

Keywords: Credit scoring; ensemble model; SVM; DT; RF; NB; KNN; LR

1 Introduction

A credit scoring model is an analysis tool used to determine the creditworthiness of a loan applicant based on historical data and by estimating the default probability. The performance of the credit scoring model is proven to be more effective by using ensemble modeling. The models are designed by training single base classifiers and the resulting output is integrated by using an ensemble strategy to enhance the performance.

Credit scoring models are used for evaluating financial threats associated with applicant's credit granting process [(1)]. The credit scoring model is used to assess the credit risk of a new applicant [(2)] or to assess the likelihood of a default using information from a previous loan applicant [(3)].

The 2 most commonly and widely used statistical methods in credit scoring are Logistic Regression (LR) and Linear Discriminant Analysis (LDA). Machine learning classification approaches like K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naïve Base (NB), Classification and Regression Tree (CART), Genetic Algorithms (GA), and Artificial Neural Networks (ANN) are extensively used in credit scoring.

Credit industry demands a well-structured and effective credit scoring systems. The research incentive is to have an improved classification accuracy. The research affinity is heading towards building hybrid models and ensemble models.

To improve the machine learning algorithms performance, it is very important to adopt hybridization and ensemble learning approach. The authors [(1)] discusses a credit scoring model built using hybrid ensemble, which consists of 5 feature selection methods, four classification algorithms, 8 ensemble models using soft voting approach and 3 different voting techniques. Generalized linear model (GLM), SVM, NB and DT are the classification methods used. These classifiers are combined into ensemble model in 8 different ways, i.e., GLM+SVM+NB+DT, GLM+SVM+NB, GLM+SVM+DT, GLM+NB+DT, SVM+NB+DT, GLM+DT, GLM+NB, SVM+DT. Among these 8 models, GLM+DT model performed well with higher accuracy.

Handling of imbalanced classification problem is very important and tough job [(4)], a novel heterogeneous credit scoring model is implemented and five performance measures are carried out on four credit datasets. The new model performed better compared to other baseline models. Authors in [(5)] proposes an ensemble weighted SMOTE to improve the robustness. The accuracy of the minority class is improved when compared with its competitors.

In recent years, the research focus is mainly on ensemble strategy. Different approaches are adopted to perform ensemble learning, by using different base classifiers and different consensus methods [(6)]. Ensemble techniques are subjected to training various classifiers to find solution for the same problem. The enhanced single predicted output is obtained by aggregating each classifier prediction into one classifier [(7)].

The authors in [(8)] proposes a novel multistage ensemble model with an improved outlier adaptation. For effective identification of the outliers, a local outlier factor (LOF) - algorithm with bagging approach has been improved. It is then boosted back again into the training set from which a training set based on outlier-adaptation is constructed. This set enhances the base classifiers outlier adaptability. The authors also used an ensemble learning method based on stacking, to further strengthen the prediction. They had tested Ten datasets with 6 performance indicators AU, AUC, Brier score, BA, F-Score and Log loss. The result shows that the new model has superior performance over benchmark models.

The authors in [(9)] proposes a data distribution novel ensemble model based on resampling. This method solves class imbalance problem by using under-sampling technique - data distribution using majority class. The result shows that the new ensemble model has good prediction performance.

The authors in [(10)] performed comparative assessment of 5 ensemble algorithms RF, AdaBoost, Stacking, XGBoost, and LightGBM, and 5 traditional individual learners, that is, Neaural Network, Decision Tree, Support Vector Method and Naïve Base. The credit dataset from lending club in the united states is used for experimentation. The result shows that the ensemble method gives higher performance than individual learners. Authors suggested that LR, Random Forest, LightGBM and XGBoost are the best options for financial institutions. The limitation in this study is that only single parameter tuning method is used.

The authors in [(11)] proposes a boosting ensemble method based on weight adjustment. For feature selection, they have used feature selection method based on rough set. Regression based pre-processing is done to fill in missing values. The authors had conducted experiment with 5 base learners, LogR, KNN, NB, DT and SVM. The ensemble methods used are WABEM, Bagging and Random space. The result shows that the boosted ensemble method proposed is better than the other ensemble and base classifier methods.

The authors in [(12)] proposed a new ensemble model, with a hybrid genetic algorithm to achieve accurate and stable credit prediction. Base classifiers are integrated using stacking approach. The results yield superior performance.

The authors in [(13)] in their work used bagging neural network, which yield in higher performance compared to benchmark models. Diversity of ensemble is not considered and also not focused on asymmetric misclassification problem.

Authors in [14] empirically examined several multiple classifiers methods using new boosting method called Error Trimmed boosting, bagging and boosting. Bagging is a technique which is used to minimize the variance in the prediction, whereas, Boosting is an iterative method, which based on the last classification adjusts the weight of an observation. Stacking is an ensemble ML algorithm which combines the predictions from multiple ML models. Predictions of base classifiers are fused through two-layer ensemble modeling [15], with a satisfying improvement in different performance measures.

A wide series of machine learning algorithms are available in current era and that can be used to build individual models. Each algorithm has its own strength and weaknesses. There is no such algorithm which can achieve best on all available credit datasets. A broad set of different and accurate algorithms can be ensemble to get superior performance [16].

Ensemble method is presently booming in banking and finance industry. Ensemble modelling is the ability of combining different classifiers together to improve the predictive power and the stability of classification model. An effective prediction system will help bankers to assess credit risk when making loans to credit applicants [17]. Predictive models are built using different ensemble methods like stacking, bagging and boosting and their results are compared for achieving better accuracy.

A stacked support vector machine is proposed in [18] which demonstrated higher performance than ML ensemble models. Stacking strategy used in constructing ensemble models demonstrated superiority however, performance of the base classifiers is important in getting the effect of stacking strategy [19].

Hence, this study uses a stacking strategy to construct an ensemble model by integrating base classifiers. In this paper six different classifiers are used for building heterogeneous classifier ensemble. The six base classifiers used are SVM, LR, KNN, RF, NB and DT.

The proposed model is made up of the following steps:

1. Collection of datasets.
2. Splitting the available data sets into training and testing samples.
3. Apply the classification algorithm.
4. Build an ensemble classifier using stacking and voting method.

Using performance evaluation measure, assess the built models.

The result of every individual classifier is compared with heterogeneous classifier. The rest of the paper is structured as follows: Section 2 details on the research methodology. Section 3 carries out discussion on results and its analysis. Finally, section 4 provides the conclusion and future research recommendations.

2 Materials and Methods

2.1 Classification Techniques

2.1.1 Support vector machines

Support vector machine (SVM) is a classification technique and was proposed by Vapnik. It is extensively used in the field of credit risk assessment due to its powerful assessment capabilities. It performs better compared to other algorithms due to the better solution to solve the sparseness problem. In fact, its main idea is to find a hyperplane by projecting input data into a feature space of higher dimension. The hyperplane is supported by support vectors which are used to separate the two classes with the maximum margin.

Let S be a dataset with M observations. (xi, yi) is labelled instance pairs, for i=1,2,….M and xi $\in$ Rn and yi $\in \{1,0\}$. SVM is used to find an optimal separating hyperplane. The constraint is given as

$$\text{yi}\ (\langle\ \text{w. xi}\ \rangle +\ \text{b})\ -1 \geq 0 \tag{1}$$

where w denotes the plane's normal and b denotes the plane's intercept.

To solve the quadratic optimization problem, we can use Lagrange multipliers. The Lagrange function is:

$$L_p(w,b,\alpha) = \frac{1}{2}W^T.W - \sum_{i=1}^{M}(\alpha_i y_i(\langle w.x_i\rangle + b) - 1) \tag{2}$$

Where ai represents Lagrange multipliers.

To obtain an peak saddle point, Lp must be maximized with respect to the non-negative dual variable ai and minimized with reference to variables w and b. Lp is converted to the dual Lagrangian LD(α):

$$\max_\alpha L_D(\alpha) = \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{ij=1}^{M}\alpha_i\alpha_j y_i y_j \langle x_i \cdot x_j\rangle \tag{3}$$

Where $\alpha i \geq 0, i = 1, 2, \ldots\ldots M$ and $\sum_{i=1}^{M} x_i y_i = 0$

if the corresponding $\alpha i > 0$, then xi is called support vector. We can obtain the decision function as,

$$f(x) = \text{sgn}(\langle w, x \rangle + b) = \text{sgn}\left(\sum_{i=1}^{M} \alpha_i y_i \langle x_i, x \rangle b\right) \quad (4)$$

Using the above function in equation 4, SVM classifies samples as class 0 if $\langle w, x \rangle + b < 0$ and as class 1 if $\langle w, x \rangle + b > 0$.

We can predict the label of the new input sample using the features of the support vectors. Many kernels can be selected to map input data to multidimensional feature space in SVMs, such as linear, sigmoid, radial basis function (RBF) and polynomial.

All the four linear kernel, sigmoid kernel, radial basis kernel and polynomial kernel are used in our ensemble aggregation.

2.1.2 Logistic Regression

Logistic regression is a specific form of the linear regression. LR is used to measure correlation between a dependent and one or more independent variables. It is used in the area of social sciences, medical and machine learning. LR model consists of n predictors and one dichotomous output (response) variable. The response variable has only two possible outcomes: 1 (good) and 0 (bad). The equation is as shown below,

$$log \frac{p}{(1-p)} = {}_1 * x_1 + {}_2 * x_2 + {}_3 * x_3 \cdots + {}_k * x_k \quad (5)$$

Where, p represents specific customer's probability. bi, (intercept term) is the coefficient related to predictors xi (i = 1, ..., k). The purpose of logistic regression model is to estimate the conditional probability of a particular instance belonging to a particular class.

2.1.3 Random Forest

Random Forest is a classification technique based on ensemble learning. Each classifier in the ensemble is built using DT classifier. It is a collection of classifiers which forms a forest. Individual decision trees are constructed by using attributes randomly selected at each node. Each individual tree votes during classification. Output is based on votes, where the most voted class is considered.

Bagging or Bootstrap aggregation techniques are applied for random forest during training. Consider a training set $xi \in X$ with class output $yi \in Y$. Bagging fits decision trees to the samples of the training dataset based on repeated selection of a random sample. To classify a new instance, multiple trees are created and based on attributes; each tree assigns a particular class. Based on voting, the algorithm chooses the classification, that is, to which class the new instance belongs to.

RF classifier to select attributes uses Gini index. Attributes impurity with respect to the class is measured by Gini index. The Gini index for selecting one case at random for a given training set is;

$$\sum\sum (f(C_i, T)/(T|)) (f(C_j, T)/|T) \quad (6)$$

Where, f (Ci, T) / |T|) represents the probability of the selected case Ci.

2.1.4 Decision Tree

Decision tree is powerful and very popular classification algorithm with the ability to interpret simple rules with very little user intervention. The most widely used DT algorithms are ID3, C4.5 and C5. Building an optimal decision tree is a key task in decision tree classifier. Many decision trees can be built with the given attributes set. In building decision tree, at each step information gain is used to determine on which feature to spilt. Based on information theory, entropy is given as,

$$H(T) = IE(p_1, p_2, \ldots, p_j) = \sum_{i=1}^{j} p_i \log_2 p_i \quad (7)$$

Where p1, p2, …, pj are fractions. These fractions are added up to 1which represents the percentage of each class present in the node.

2.1.5 KNN

K – nearest neighbor (KNN) classifier is the most frequently used methods for credit scoring. It is a non-parametric classification method. Non-parametric means, the method does not make any assumptions on the underlying data. KNN considers the entire dataset for making decision.

KNN is used as a benchmark for many classifiers. Euclidean distance is adopted in KNN for computing distance between a test sample and training samples. The prediction of a new point in KNN is that, it chooses k nearest neighbors from the training dataset and computes the average of k nearest neighbors. KNN can be easily handled as there is only one parameter 'k' in KNN algorithm.

2.1.6 Naïve Bayes

The Naïve Bayesian (NB) algorithm is based on applying Bayesian theorem with strong independence assumptions amongst the features. To estimate the probability terms that are required for classification a set of training data is used. This performance is measured by the accuracy of the predictable required probability terms.

The naïve Bayes classifier mainly focuses on conditional probability. It assumes that the attributes and features are independent and it is suitable for high dimensional inputs. The assumption is that given the target value of the instance, the probability of noticing the conjunction a1, a2, …, an is the product of the probabilities for the individual attributes.

$$P(a_1, a_2, \ldots, a_n \mid v_j) = \prod_i P(a_i \mid v_j) \tag{8}$$

Where, ai is distinct attribute value and vj is distinct target value.

2.2 Ensemble classifier

Each individual classifier performs differently on different datasets. On a specified dataset, it is difficult to predict which classifier performs best. Ensemble classifier is an ideal classifier for any dataset [20]. Aggregation or combining of various classifiers will give higher classification performance than its base classifiers [21].

Ensemble classification methods are used to solve the same problem by training multiple classifiers. Ensemble learning comprises of a set of base classifiers, which are trained individually. The predicted output of these classifiers are combined using majority voting, weighted voting, bagging, stacking and boosting [22]. When classifying new instance, individual classifiers will learn and train on single set of data, whereas, ensemble classification models learn and train on different data that are created from original dataset. A set of hypothesis is constructed from trained data. This hypothesis leads to better prediction accuracy. Many strategies emerged to diversify the classifiers that create the ensemble [12].

Each classifier is trained on heterogeneous data, which will yield a predicted output. These predictions are combined together in several ways: i) Simple average: In this, for each sample the average of predictions of all the classifiers is calculated to produce the final prediction. ii) Majority voting: Here, predictions of all classifiers are combined together and for each sample the class that has highest number of votes is selected as final output [23].

In this study, a combination of stacking and majority voting model is developed for aggregating the output of six base classifiers to improve the predictive accuracy of credit scoring system.

2.3 Experimental study

2.3.1 Credit Datasets

Two real world datasets, namely German and Australian, taken from UCI machine learning repository are used in the experiments. The details of the datasets are presented in Table 1.

The German credit dataset contains a total of 1000 samples with 700 positive samples and 300 negative ones. Each instance has 7 numerical features, 13 categorical features and a target attribute. Australian dataset consists of 690 samples among them 307 samples are positive and 383 samples are negative. Each instance comprises of 8 numerical features, 6 categorical features and a target attribute.

Table 1. Credit datasets from UCI machine learning repository

Dataset	Number of Instances	Good cases	Bad cases	Categorical features	Numeric features	Total features
German	1000	700	300	13	7	20
Australian	690	307	383	6	8	14

2.3.2 Feature Selection

Feature selection provides cost-effective and faster classifiers to improve the prediction of credit scoring systems. The process of feature selection can be combined with the subset selection. There are three categories for selecting subsets of features namely: wrappers, filters and embedded methods [6].

In this work, we have used random forest method for feature selection – a tree based feature selection method. Most important features are selected and irrelevant features are removed by computing feature importance using random forest method. Random forest is commonly used as a classifier. It also has the capacity to estimate the feature importance and hence can be used as feature selector.

Random forest builds a set of decision trees for its working. Given a difference of the performance for the tree i, denoted by di, the final important feature Aj can be computed as

$$I(A_j) = \sum d_i / (n \, x \, SE_d) \tag{9}$$

where SEd denotes the standard error of di considering all trees (SEd = SDdi$\sqrt{n}$), n represents the number of elements in the dataset and SDd represents the standard deviation of di.

2.3.3 Proposed Method

The proposed model is built using six base classifiers for ensemble aggregation. The Figure 1 shows the proposed framework block diagram.

A feature selection algorithm based on the random forest technique is used for selecting the best features. The credit dataset is divided into training and test sets. The proposed model uses stacking and majority voting method for ensemble classification. Initially, stacking is applied to the base classifiers. This is done in two levels. First the training dataset is split into 10 folds for cross validation. In each iteration 10-1 (9) folds were used to train base classifiers and, remaining one fold is used for output prediction. After 10 iterations, the predicted result for the entire training set is obtained. The output of each classifier is taken and the dataset is updated with the meta-features, i.e., the predictions made by each classifier. In the second level, three meta-classifiers (MC), namely LR, SVM and RF are used. Majority voting is applied to the output of these meta-classifiers for the prediction.

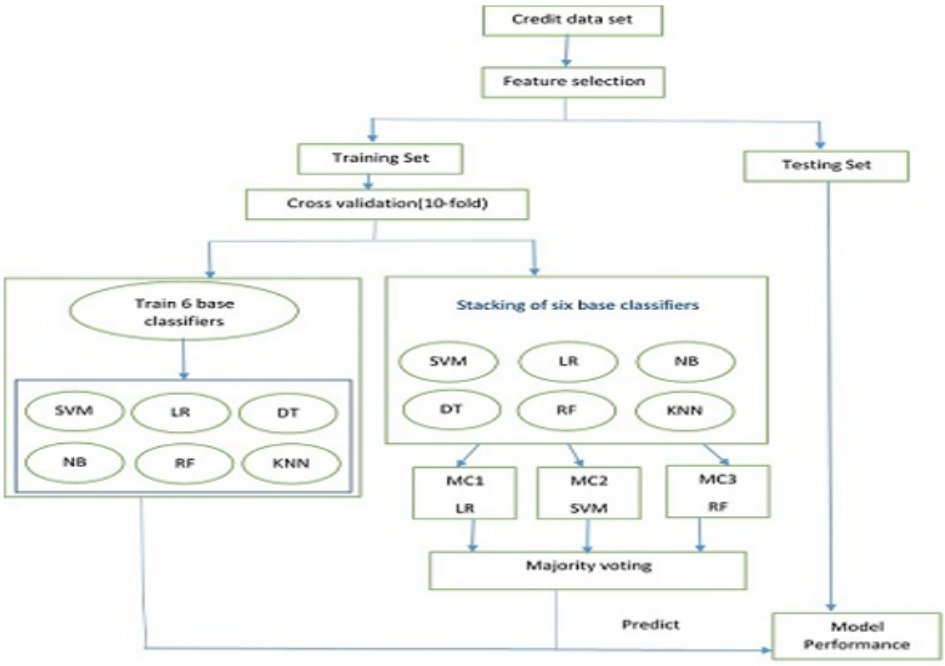

Fig 1. Credit scoring framework using Ensembleaggregation

2.3.4 Evaluation Measures

The selection of evaluation measures is very important for validating the performance of the classification models. Confusion matrix has been considered for various assessment measures in prediction. The confusion matrix is shown in Table 2.

The true positives (TP) are the positive instances which are predicted as positive. The False positives (FP) are negative samples which are predicted as positive. Likewise, false negatives (FN) are positive samples which are predicted as negative, and the true negatives (TN) are negative samples which are predicted as negative. Using confusion matrix, Accuracy, Precision, Recall, F-

Table 2. Confusion Matrix

		Predicted	
		Positive	**Negative**
Real	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

measure, specificity and Area Under ROC Curve (AUC) are expressed as,

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Recall = Sensitivity \frac{TP}{TP+FN} \tag{12}$$

$$F-Measure = \frac{2\ X\ Precision\ X\ Recall}{Precision+Recall} \tag{13}$$

$$Specificity = \frac{TN}{TN+FP} \tag{14}$$

F-measure given in equation (13) is a measure of models accuracy. It is computed as two times the product of recall and precision to the ratio of sum of recall and precision. The precision given in equation (11) is TP (number of correctly classified positive results) divided by the sum of TP and FP. Recall is the number of correctly classified positive results divided by the sum of TP and FN. AUC (Area Under Curve) is measure of two-dimensional area Receiver Operating Characteristic (ROC) curve.

It represents the area under the ROC curve. True-positive rate (sensitivity, equation (12)) is represented along y-axis and the false-positive rate (calculated as 1-specificity (equation (14))) is represented along x-axis. A model with larger AUC indicates better performance.

3 Results and Discussion

Based on the proposed heterogeneous ensemble model and the experimental setup, we compare the predictive performance between the heterogeneous ensemble model and other 6 individual machine learning models, and the final computational results are shown in Table 3 and Table 4.

The dataset is split into two parts, 80% as training dataset and 20% as testing dataset. SVM, LR, KNN, RF, Naive Base, DT are used as base models.

To examine whether the proposed ensemble model is effective in terms of accuracy, the following steps are executed:

- Each single classifier is tested using test dataset and the results are noted.
- The proposed heterogeneous ensemble model is then tested on test dataset.
- Finally, the accuracy of individual and ensemble models is compared for selecting the best model.

The results reported in this section were based on the testing set of both German and Australian datasets. Ten-fold cross validation is applied during training of each classifier.

Table 3 and Table 4 shows the results of single classifiersand also proposed Ensemble classifier with respect to Accuracy, F-measure, AUC and Precision-recall Score.

For Australian dataset, Random Forest classifier shows good performance among individual classifiers. And in German dataset, Logistic Regression performed well among single classifiers. The proposed Ensemble method has achieved the highest accuracy on both Australian and German datasets with 91.56% and 84.35% respectively. We can observe that ensemble classifier gives better classification when compared with single classifiers. The comparison of single base classifiers with the proposed model is shown in Table 3 and Table 4.

The validity of the proposed model is compared with the state-of-the-art models proposed in [24], [15] and [25], the comparison is shown table 5. The comparison result shows that the proposed model performed significantly high.

Table 3. Results of classifiers using Australian dataset

Classifier	Accuracy	F-Measure	AUC	Precision-Recall Score
LR	85.35	78.15	71.36	73.57
DT	88.23	87.42	65.47	77.12
SVM	70.65	69.56	64.35	68.54
KNN	70.95	69.84	66.09	67.46
NB	83.62	74.47	75.64	72.35
RF	89.50	**94.09**	83.54	**93.56**
Proposed Ensemble model	**91.56**	86.45	**86.48**	85.42

Table 4. Results ofclassifiers using German dataset

Classifier	Accuracy	F-Measure	AUC	Precision-Recall Score
LR	79.21	86.26	70.86	68.43
DT	74.52	83.45	66.37	68.18
SVM	74.65	84.62	54.58	**71.23**
KNN	73.76	85.28	65.35	68.18
NB	74.54	83.63	74.24	65.56
RF	77.15	**91.92**	82.35	69.09
Proposed Ensemble model	**84.35**	87.52	**84.25**	70.56

Table 5. Comparison of performancewith other credit scoring models

Method	Evaluation indicator	Australian dataset	German dataset
Our Proposed Model	Accuracy	91.56	84.35
H Van Sang et al.[24]	Accuracy	89.40	76.20
S. Wei et al.[15]	Accuracy	87.92	-
S. Guo et al.[25]	Accuracy	87.4	78.3

4 Conclusion and Future work

In this study we had built six individual credit scoring models and proposed heterogeneous ensemble model. The proposed ensemble model was developed as follows; First, started with collecting datasets (Australian and German), next, the dataset is split into training (80%) and testing (20%) sets, then six individual classifiers and ensemble classifier were built. Each classifier was trained on both Australian and German datasets. Finally, the output of each of the classifier was combined in ensemble using stacking and majority voting to achieve final results. The result shows that the ensemble model has achieved the accuracy 91.56% on Australian dataset and 84.35% on German dataset compared to other individual classifiers.

Future research directions focus more on feature selection and to extend the work by using ensemble of feature selection methods that is combining different feature selection algorithms. Although the developed model resulted in better performance, our future work focuses on building credit scoring system using neural network classifier in ensemble aggregation and also to adopt weighted voting approach.

References

1) Nalić J, Martinović G, Žagar D. New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. *Advanced Engineering Informatics*. 2020;45:101130–101130. Available from: https://dx.doi.org/10.1016/j.aei.2020.101130.
2) Xia Y, Zhao J, He L, Li Y, Niu M. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Systems with Applications*. 2020;159:113615–113615. Available from: https://dx.doi.org/10.1016/j.eswa.2020.113615.
3) Rojarath A, , Songpan W. Probability-Weighted Voting Ensemble Learning for Classification Model. *Journal of Advances in Information Technology*. 2020;11(4):217–227. Available from: https://dx.doi.org/10.12720/jait.11.4.217-227.
4) Zhang T, Chi G. A heterogeneous ensemble credit scoring model based on adaptive classifier selection: An application on imbalanced data. *International Journal of Finance & Economics*. 2021;26(3):4372–4385. Available from: https://dx.doi.org/10.1002/ijfe.2019.
5) Abedin MZ, Guotai C, Hajek P, Zhang T. Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk. *Complex & Intelligent Systems*. 2022. Available from: https://dx.doi.org/10.1007/s40747-021-00614-4.
6) Bao W, Lianju N, Yue K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*. 2019;128:301–315. Available from: https://dx.doi.org/10.1016/j.eswa.2019.02.033.

7) Plawaik P, Moloudabdar, Acharya R. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied soft computing*. 2020. Available from: https://doi.org/10.1016/j.asoc.2019.105740.
8) Zhang W, Yang D, Zhang S, Ablanedo-Rosas JH, Wu X, Lou Y. A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Systems with Applications*. 2021;165:113872–113872. Available from: https://dx.doi.org/10.1016/j.eswa.2020.113872.
9) Kunniu Z, Zhang Y, Liu R, Li. Resampling ensemble model based on data distribution for imbalanced credit risk evaluation in p2p lending. *Information science*. 2020;V(536):124–134. Available from: https://doi.org/10.1016/j.ins.2020.05.040.
10) Li Y, Chen W. A Comparative Performance Assessment of Ensemble Learning for Credit Scoring. *Mathematics*. 1756;8(10):1756–1756. Available from: https://doi.org/10.3390/math8101756.
11) Sivasankar E, Selvi C, Mahalakshmi S. Rough set-based feature selection for credit risk prediction using weight-adjusted boosting ensemble method. *Soft Computing*. 2020;24(6):3975–3988. Available from: https://dx.doi.org/10.1007/s00500-019-04167-0.
12) Jin Y, Zhang W, Wu X, Liu Y, Hu Z. A Novel Multi-Stage Ensemble Model With a Hybrid Genetic Algorithm for Credit Scoring on Imbalanced Data. *IEEE Access*. 2021;9:143593–143607. Available from: https://dx.doi.org/10.1109/access.2021.3120086.
13) Dzelihodz C, Donko D, Kevric J. Improved credit scoring model based on bagging neural network. *International Journal of Information Technology and Decision Making*. 2018;17:1725–1741. Available from: https://doi.org/10.1142/S0219622018500293.
14) Wang Z, Jiang C, Ding Y, Lyu X, Liu Y. A Novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending. *Electronic Commerce Research and Applications*. 2018;27:74–82. Available from: https://dx.doi.org/10.1016/j.elerap.2017.12.006.
15) Wei S, Yang D, Zhang W, Zhang S. A Novel Noise-Adapted Two-Layer Ensemble Model for Credit Scoring Based on Backflow Learning. *IEEE Access*. 2019;7:99217–99230. Available from: https://dx.doi.org/10.1109/access.2019.2930332.
16) He H, Zhang W, Zhang S. A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*. 2018;98(98):105–117. Available from: https://dx.doi.org/10.1016/j.eswa.2018.01.012.
17) Dietterich TG. Machine-learning research: Four current directions. *AI Magazine*. 1997;18:96–136. Available from: https://doi.org/10.1609/aimag.v18i4.1324.
18) Ali L, Niamat A, Khan JA, Golilarz NA, Xingzhong X, Noor A, et al. An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure. *IEEE Access*. 2019;7:54007–54014. Available from: https://dx.doi.org/10.1109/access.2019.2909969.
19) Zhang W, He H, Zhang S. A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications*. 2019;121:221–232. Available from: https://dx.doi.org/10.1016/j.eswa.2018.12.020.
20) Parvin H, MirnabiBaboli M, Alinejad-Rokny H. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Engineering Applications of Artificial Intelligence*. 2015;37:34–42. Available from: https://dx.doi.org/10.1016/j.engappai.2014.08.005.
21) Ala'raj M, Abbod MF. A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications*. 2016;64:36–55. Available from: https://dx.doi.org/10.1016/j.eswa.2016.07.017.
22) Kuncheva LI. Combining Pattern Classifiers: Methods and Algorithms. and others, editor;John Wiley & Sons. 2004. Available from: http://www.books24x7.com/marc.asp?bookid=72712.
23) Marqués AI, García V, Sánchez JS. Exploring the behaviour of base classifiers in credit scoring ensembles. *Expert Systems with Applications*. 2012;39(11):10244–10250. Available from: https://dx.doi.org/10.1016/j.eswa.2012.02.092.
24) Van Sang H, Nam NH, Nhan ND. A Novel Credit Scoring Prediction Model based on Feature Selection Approach and Parallel Random Forest. *Indian Journal of Science and Technology*. 2016;9(20). Available from: https://dx.doi.org/10.17485/ijst/2016/v9i20/92299.
25) Guo S, He H, Huang X. A Multi-Stage Self-Adaptive Classifier Ensemble Model With Application in Credit Scoring. *IEEE Access*. 2019;7:78549–78559. Available from: https://dx.doi.org/10.1109/access.2019.2922676.