An Overview of Classification Algorithms and Ensemble Methods in Personal Credit Scoring

[1]Pramod S. Patil, [2]Dr. J. V. Aghav, [3]Vikram Sareen
[1,2]College of Engineering, Pune, Maharashtra, India
[3]BlueBricks Technologies Pune, Maharashtra, India

Abstract
The Indian finance market has been growing steadily at 17 percent in the recent years. At the same time, the loan defaulter's rate is also growing. In the present scenario, it is very important for the banks and the financial institutes to minimize the loan defaults. One of the important strategy is to predict the likely defaulters so that such loans are either not issued or monitored closely after the issuance. This study attempts to build robust data mining models to predict the defaulters using data obtained from one of finance company. In this paper various ensemble algorithms like bagging, boosting and stacking are implemented and their efficiency and accuracy is compared. The robust predictive models are able to predict the default with high degree of accuracy.

Keywords
Prediction, Ensemble Model, Classification, Credit Scoring

I. Introduction
The motivation behind writing this paper is to give an overview of credit scoring system in India and machine learning algorithms used in credit scoring. We have also covered ensemble methods in machine learning to ensemble multiple base classifiers. For every machine learning algorithm, there is some limit beyond that it can't fit that data. So accuracy stops at that point. If still we try to fit it more, it leads to an over-fit problem. Ensemble tries to overcome this problem with the use of multiple models with different ensemble techniques like bagging, boosting or stacking. As each ensemble method is having some advantages and disadvantages over each other. So we have applied basic and ensemble techniques on data set available on https://archive.ics.uci.edu/ml/datasets.html and advantages and disadvantages of these techniques made.

Nowadays banks play a crucial role in the market economy. Before giving a credit loan to borrowers, the bank decides who is bad (defaulter) or good (non-defaulter) borrower. The prediction of borrower status i.e. in future borrower will be defaulter or non-defaulter is a challenging task for the bank. The recent issue of kingfisher airlines brought this topic into focus. The loan defaulter prediction is a binary classification problem. For market and society to function, individuals and companies need access to credit [24]. Credit scoring algorithm, which gives the probability of making default, are the methods bank use to determine whether or not a loan should be granted.

Banks need to struggle a lot to get more business due to high competition. Banks have realized that preventing fraud and retaining good customers has to be the strategy for healthy competition. Indian Banks have been maintaining large amount valuable data since many decades [21]. This data has now unlocked the secrets of money movements, helped to prevent major disasters, thefts and understand customer behaviors. With data mining, we can estimate the probability of individual applicant to default on his/her loan. Support vector machines, neural network, and random forest are the most favorite technique used recently.

Taking these considerations into account, the present paper examines the use of seven well-known classifiers with three effective ensemble methods. The aim of this study is to find out what individual models are suitable for each ensemble strategy in the area of credit scoring. To this end, several experiments on three real credit data sets are carried out and the results are analyzed for statistically significant differences by AUC of ROC curves.

II. Literature Survey
The whole idea of our classification system is to compare the features of an applicant with the other earlier customers, whose took loans from banks and paid back successfully. If characteristics of applicant match with similar kind of customers, then the loan will be accepted otherwise rejected i.e. customer's characteristics are satisfactory like non-defaulted customers then the application is normally granted. Generally, two techniques are used in India for this process "Loan officer's assessment and credit scoring" [5].

Normally in judgmental technique evaluation, each loan application includes essential information of applicant like property of applicant, income of applicant, accounts, which are evaluated by individual decision-makers of a creditor. The success of a judgmental method depends on the common sense and experience of the credit analyst [3]. As an output, these techniques are associated with inconsistency, subjectivity and individual preferences motivating decisions. These methods have good strengths, like considering qualitative features and checking track of record in evaluating past history by utilizing the wealth of the credit analyst's past experience.

Otherwise, in a credit scoring systems, banking analysts usually use their historical experience with the debtors to build a quantitative model for grouping of grantable and non-grantable loan applications. By using a credit scoring system, a loan application is a self-operating process and can be consistently applied to all loan decisions [12]. The scoring system is made up of addition or subtraction of a statistically extracted related points of applicant's score forwarded to the predictor variables, like time of a job or the number of income sources used. As a result, we can say that credit scoring enables banks or financial institutions to assess the repaying ability of applicant quickly. Moreover, credit scoring gives a chance to banks to improve their services and maintaining good customers. By using a statically calculated cut-of score, a banking analyst can surely separate the grantable from the non-grantable loan applicants. On the other hand, credit scoring has been in a disapproving way because of statistical issues with the data used to prepare the model, and assumptions of certain static methods used to calculate point scores. Despite these issues of credit scoring systems, these systems are referred as one of the most successful models used in the fields of business and finance [16].

A. Statistical Techniques used in Credit Scoring
A variety of statistical techniques are used for the construction of scoring models. Most of these statistical, and some of these non-

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

linear, models are already used in the construction of effective and efficient credit scoring systems that are effectively used for predictive purposes. Techniques, like weight of different characteristics, linear regression, discriminant analysis, logistic regression, support vector machines(svm), decision trees, neural networks, genetic algorithms and k-nearest-neighbor are famously used algorithms for the construction of credit scoring systems by researchers, credit analysts, lenders, software developers, and providers.

1. Linear Regression

Linear regression is an essential component of any data analysis, deals with describing the relationship between a target feature and one or more independent features. Linear regression is used in credit scoring applications, as a binary classification problem, represented by a dummy variable. By use of Poisson regression instead of linear regression could be used to cover cases where the customer makes varying degrees of partial repayments. For such the proportionate repayments can be re-expressed as Poisson counts. Factors, like customer's historical payments, default rates, guarantees in a timely manner, is analyzed by credit analysts, with linear regression to set a score for each factor, and then every bank compares with their cut-off score. If a new applicant's score passes the bank's cut-off, the loan will be approved. First time regression analysis is used for commercial loans by Orgler (1970), and that model was only limited to evaluation of existing loans and could be used for review of loan and examinations. Later, Orgler (1971) used a regression approach for evaluation of outstanding consumer loans [16]. He concluded that information which is not there on the application form has a greater predictive power than the information which is there on the original application form, in assessing loan quality. Then Lucas(1992), Henley(1995), Hand & Henley(1997), Hand & Jacka(1998) used the regression analysis with some extension in this applications to improve the quality of loan assessment. But linear regression was unable to get that much accuracy [19].

2. Discriminant Analysis

Discriminant analysis is a simple method in machine learning and statistics to find the combination of features so that forms two or more groups as per characteristics and features. Many credit analyst agreed that discriminant analysis is still one of the most established techniques to classify customers as good or bad credit [3]. This technique is used in many credit scoring applications under various fields. Therefore, a credit model based on based on a discriminant analysis is used for statistical analysis to categories group's variables into two or more categories. The discriminant approach was first presented by Fisher (1936) as a classification discrimination technique [29]. In 1941, Durand first time applied numerous discriminant classification approaches in credit scoring system, for examination of car loan applications. In nineteen-sixties, there was one well known corporate bankruptcy prediction application proposed by Altman (1968). Altman developed 1st working scoring model formed on five financial ratios, taken from eight variables which were extracted from corporate financial statements [10]. He invented a Z-score, which is a linear combination of the financial ratios taken from variables of financial statements. In fact, discriminant analysis approach is a valid method used in developing credit scoring models (Abdou and Pointon, 2009; Sarlija stated, 2004; Hand stated, 1998;Caouette stated., 1998;Hand and Henley, 1997;Desai stated, 1996;)[29].

3. Logistic Regression

Logistic regression, work is same like discriminant analysis, is also one of the most widely used statistical techniques in the field of data mining. Logistic regression is the generalization of linear regression. The difference between a logistic and linear regression model is that the outcome of logistic regression is discrete (0/1 outcome). So as the output of logistic regression is discrete, it cannot be constructed directly from linear regression. This difference between linear regression and logistic regression is reflected both in the choice of a parameter selection and in considering the assumptions. Once this change is accounted for, the methods employed in an analysis using logistic regression follow the same general principles which are used by linear regression (Lemeshow and Hosmer 1989) [11]. Two variable logistic regression model is easily buildable, but extending more number of variables is also gets harder to see multiple observations at each level of each variable. So, as per Freund and William (1998) most of the logistic regressions with more than one variable are implemented using the maximum likelihood method. As per theory, it is supposed that logistic regression is more proper method than linear regression for classification problem like default prediction with two class output, 'good' credit and 'bad' credit said by (Henley and Hand, 1997). Until now logistic regression is broadly used in credit scoring applications (e.g. Abdou stated, 2008; Crook, 2007; Baesens, 2003; Lee and Jung, 2000; Desai, 1996; Lenard, 1995).

4. Decision Tree

Decision tree is powerful classification algorithm that is becoming more popular with the growth of data mining in the field of information systems. In the literature, there are so many popular decision tree algorithms like ID3, C4.5, C5 etc. As its name implies, this technique repeatedly separates input observations in sub-branches to build a tree for the purpose of getting more accurate prediction [12]. By doing so, decision trees internally use mathematical formulas like Gini index, information gain, entropy to find a variable and its corresponding threshold that going to split the input space into two or more subspaces. This step is recursively done at each leaf node of decision tree until the tree is constructed completely. The objective of the splitting at each node is that it maximizes the homogeneity of the resulting two or more input subspace samples. The most commonly used mathematical term for splitting is Entropy which is derived from information gain and used in ID3, C4.5, and C5 decision tree algorithms. CART uses Gini index [15]. Decision tree is another classification technique used widely in developing credit scoring models, also known as classification and regression trees (CARTs) or recursive partitioning (Hand and Henley, 1997). Breiman et al. used CART in credit scoring _rst time at 1984. But, Gleit and Rosen-berg (1994) mentioned at Harvard Business School that the Schlaifer and Rai_a initiated credit scoring by using decision tree in 1961[3]. Gleit and Rosenberg also mentioned at the University of Richmond that, later, David Sparks had developed credit scoring model from decision tree in 1972 [16]. A comparison of recursive partitioning and discriminant analysis was done by Boyle et al. (1992). Other applications of decision trees in credit scoring were mentioned by Coffman (1986),Handand Henley (1996), Jacka and Hand(1998), Hosemann and Fritz(2000),Thomas(2000), Wilk and Stefanowski (2001),Baesens et al. (2003)[15].

5. Neural Networks

Neural Network is a mathematical representation motivated from functioning of the human brain. Neural Network is composed of a number of neurons which works in parallel, without any centralized control. The neurons are usually arranged in layers. A system which weights the connections calculates the flow of through the network. Neural networks have been used extensively in many domains to solve compound real-world problems. [18].

Neural networks started themselves as an alternative to traditional statistical techniques and many researchers proved that neural networks outperform statistical methods with respect to classification problems accuracy. Neural networks are used widely in building credit scoring systems. Feed forward Multi-Layer Perception are trained by Back Propagation are used more than 75% in applications of neural network stated by Vellido. [25] The commercial credit scoring model was compared with MLP model with sample of 125 applications. The accuracy of classification scaled from 76% to 80% on test samples. In [25], a comparison of the predictive accuracy of two statistical techniques: logistic regression and linear discriminant analysis against two neural networks: modular neural network and the multi-layer perceptron for classification of loans into grant-able and not grantable was made and they made conclusion as neural networks outperforms only if measure of accuracy is percentage of not grantable loans correctly classified. If the measure is percentage of both grant-able and non- grantable classified correctly, then neural network accuracy is comparable to those of statistical modeling techniques. (Baldwin and Trinkle, 2007; Thomas and Seow, 2006; Leippold and Blochlinger 2006; Mitchell and Yim, 2005; chen and Lee 2005; Sohn and Kim 2004; Wilk and Stefanowski, 2001) stated that accuracy rates of neural networks with respect to conventional techniques under different sample spaces then neural networks are always better than conventional networks. Meanwhile, comparisons between advanced statistical techniques and traditional have been investigated too (Abdou et al., 2008; pointon and Abdou, 2009; Ong et al., 2005; Chen and Lee 2005;Malhotra and Malhotra, 2003;Hosemann and Fritz,2000; Lee et al., 2002;). Then there discussion is extended to comparison of back-propagation nets and feed-forward nets. [9] Statistical accuracy measures viewed that the models built by neural networks are better description of data than CARTs and logistic regression (ZekicSusac, 2004), where discriminant analysis, in general, has a better ability in terms of classification but worse ability of prediction, whereas logistic regression has a relatively better prediction capability (Liang, 2003). Generally, models built by neural networks have highest AUC as compared to other traditional techniques, such as logistic regression and discriminant analysis, considering the fact that results are too close (e.g. Abdou et al., 2008; Crook et al., 2007; ZekicSusac et al., 2004; Haykin, 1994) [28].

6. Support Vector Machine

Vapnik developed a new powerful machine learning method relied on ideas of the latest advancement in statistical learning in 1998 named as support vector machine. Support vector machines are widely used in classification and regression problems due to its promising performance [16]. Many implementations of credit scoring have used support vector machines with a promising accuracy as per Lee (2006) developed to evaluate loans to identify creditworthiness customers for consumer loans. And as per these implementations, the results revealed that the support vector machine achieved more PCC (percentage of correctly classified) than neural networks in generalization. But as the other machine libraries, the main problem is that Support vector machine is a complex function, so it is unintelligible for the humans. For solving the same problem, [1] stated a cognizance credit scoring system with SVM by pattern extraction method. Patterns are ex-tractable from a trained support vector machine that is understandable to humans while keeping the accuracy as much as possible by SVM. The obtained method showed that this technique has only a small drop of PCC (percentage of correctly classified) as compared to SVM, so this technique ranked as top of cognitive classification techniques as stated by Martens et al. (2007) [2]. Huang applied data mining to get credit scoring model with relatively few features. In this, he used three strategies to build hybrid support vector machine credit scoring model. They evaluated it with two credit data sets. The result showed that the proposed hybrid svm achieved accuracy same like neural network, genetic programming and decision tree classifier [2].

B. Ensemble Models

An ensemble classier (also referred as mixture of experts, committee of learners, multi-classifier system) consists of a set of individually trained base classifiers and the decisions from base classifiers can be combined in many ways, like weighted un-weighted voting, while classifying new record (Kittler, 1998; Kuncheva, 2004) [15]. It is found many times that most of the times ensemble classifiers produce more accurate predictions than the base classier from which ensemble is made (Dietterich, 1997). nevertheless, for an ensemble to achieve much better generalization capability than its members, it is critical that the ensemble consists of highly accurate base classifiers whose decisions must be as diverse as possible (Kuncheva & Whitaker, 2003; Bian & Wang, 2007) [22]. In statistical pattern recognition, a large number of methods have been developed for the construction of ensembles that can be applied to any base classier. In the following sections, the ensemble approaches relevant for this study are briey described.

1. Bagging

As per standard definition, bagging or bootstrap aggregating algorithm proposed by Breiman, 1996 creates M bootstrap samples T1,T2,...,Tm randomly drawn with replacement from the original training set T of size n [15]. Each base classier Ci is trained with bootstrap sample Ti of size n. New observations are predicted by taking the majority vote of the ensemble C built from C1,C2,..,Cm. As bagging phenomenon is to re-sample the training data with replacement, some instances are represented multiple times while others may or may not be even taken into considerations. Since each base classifier is not exposed to the same set of records, they are diverse from each other. So by taking votes of each of these classifiers, bagging tries to reduce the error caused because of the variance of the base classifier [22].

2. Boosting

Like bagging, boosting also creates an ensemble of classifiers by resampling the original data set, which are then merged by majority voting. But, in boosting, resampling is directed to give the most informative instances of training data set to each consecutive classifier. Ada-Boost also known as Adaptive Boosting algorithm is proposed by Schapire and Freund (1996) is the best-known classifier in boosting family [15]. It builds a series of base classifiers C1,C2,..,Cm by giving successive bootstrap samples T1,T2,...,TM as a training input which are produced by weighting each training

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

instances in M consecutive iterations. Ada-Boost initializes all training instances with equal weights, then it adjusts these weights based on the misclassifications made by the resulting base classier. Thus, instances misclassified by model Ci1 are more likely to appear in the next bootstrap sample Ti. The final decision is then obtained through a weighted vote of the base classifiers (the weight wi of each classier Ci is computed according to its performance on the weighted sample Ti it was trained on).

3. Stacking

In stacking, an ensemble of classifiers is first trained using bootstrapped samples of the training data like bagging, creating first level classifiers. Level one classifiers outputs are then used for training a 2nd level classifier (meta- classifier) (Wolpert 1992) [15]. The underlying key point is to learn whether training data is learned properly or not. For example, let us consider a particular classifier learned a certain region of the feature space incorrectly, so consistently misclassifies records coming from that region, then Tier 2 classifier may be able to learn this behavior, and with learned behaviors of other classifiers, meta-classifier can make correction of such improper training entries. Cross-validation is typically used for training the 1st level classifiers: the whole train dataset is divided into T blocks, and each 1st level classifier is firstly trained on (a different set of) T-1 blocks of the training data. And then each classifier is evaluated on the Tth block which is not seen during training by that classifier. The outputs of these 1st level classifiers on their pseudo-training blocks, along with actual labels for those instances are constituted as a training dataset for the 2nd level classifier (Meta classifier).

III. Experiments

In order to test the validity and performance of the method just proposed, several experiments have been carried out. The objective of this paper is to compare the performance of method that is just proposed and the existing ensemble techniques bagging, boosting and stacking. We have used Linear Discriminant Analysis, Logistic Regression, Neural network, support vector machines, Decision Tree, Random Forest, Gradient Boost, and XGBoost etc as base classi_ers for various ensembles. Standard performance evaluation criteria in the _elds of credit scoring include accuracy, error rate, Gini coe_cient, KolmogorovSmirnov statistic, mean
squared error, area under the ROC curve, and speci_city and sensitivity errors. Most credit scoring applications often employ the accuracy as the criterion for performance evaluation. It represents the proportion of the correctly predicted cases (good and bad) on a particular data set. However, empirical and theoretical evidences show that this measure is strongly biased with respect to data imbalance and proportions of correct and incorrect predictions (Provost & Fawcett, 1997). Because credit data are commonly imbalanced, the area under the ROC curve (AUC) has been suggested as an appropriate performance evaluator without regard to class distribution or misclassification costs (Baesens et al., 2003; Lee & Zhu, 2011) [7].

A. Datasets

Three real-world credit data sets have been taken to compare the performance of various ensemble with other classifier ensembles and base classifiers. The widely-used Australian, German and Japanese data sets are from the UCI
Machine Learning Database Repository, (http://archive.ics.uci.edu/ml/).

Table 1: Characteristics of Data sets

Data Set	#Attributes	#Good	#Bad
German	24	700	300
Australian	14	307	383
Japanese	15	296	357

B. Model Evaluation

The main objective of the data mining models in this study is to be able to predict the defaulters and non-defaulters with as much high accuracy as possible. Since there are two categories namely defaulters and non-defaulters, four possibilities can be defined to measure the effectiveness of the models. The first category is called the True Positives (TP) consisting of the defaulters that were correctly predicted as defaulters by the model. The second category is the True Negatives (TN) which consists of those non defaulters who were correctly predicted as non-defaulters [1]. The remaining two categories are False Positives (FP) and False Negatives (FN). FP is those non defaulters who were wrongly classified as defaulters by the model. Similarly, FN is those defaulters who were misclassified by the model as non-defaulters. High percentage of FP imply large number of non-defaulters misclassified as defaulters resulting in more diligent follow-up which, in turn, will lead to unnecessary expenditure to the company [4]. On the other hand, high percentage of FN will result in paying less attention to the potential defaulters and consequently lead to a higher default rate. In the same token, it is important to maximize the TP and TN. Needless to say, maximizing TP and TN will automatically lead to minimizing FP and FN.

C. Feature Significance Tests

Feature selection is an important step for practical commercial data mining which is often characterized by data sets with far too many variables for model building. There are two main approaches to selecting the features (variables) we will use for the analysis: the minimal-optimal feature selection which identifies a small (ideally minimal) set of variables that gives the best possible classification result (for a class of classification models) and the all-relevant feature selection which identifies all variables that are in some circumstances relevant for the classification.
I have used boruta package for this purpose. Boruta is a feature selection algorithm. Precisely, it works as a wrapper algorithm around Random Forest.
German data set: Boruta performed 99 iterations in 3.378572 mins.
11 attributes confirmed important: age, amount, chk acct, duration, employment and 6 more.
6 attributes confirmed unimportant: foreign, num_depend, other_credits, pstatus, telephone and 1 more.
3 tentative attributes left: housing, installrate, job.
Australian data set: Boruta performed 25 iterations in 28.63557 sec.
12 attributes confirmed important: V10, V12, V13, V14, V2 and 7 more.
2 attributes confirmed unimportant: V1, V11.
Japanese data set: Boruta performed 67 iterations in 50.89449 sec.
13 attributes confirmed important: V10, V11, V13, V14, V15 and 8 more.
2 attributes confirmed unimportant: V1, V12.

D. Results

The whole experimental analysis in performed in R statistical programming language. Table 2 shows AUC values of different prediction models. As it can be seen ensembles adabag, random forest, gbm, xgboost, blending(stacking) have performed better than individual algorithms linear discriminant analysis, logistic regression, neural network and support vector machine. That is as expected individual classifiers achieve lowest AUC values than ensemble.

Although differences in AUC may appear to be relatively low, it should be noted that even a small increase in prediction performance can yield substantial cost savings for financial institutions [7].

The highest differences are observed for the German and Australian credit data set, which corresponds to unbalanced data sets with 70% positive entries where as 30% negative entries and 44-56% respectively. For example, stacking ensemble has performed better than Bagging and Boosting algorithms for majority of data sets i.e for german data set blending has achieved 0.0500 than bagging, 0.0138 than boosting.

Table 2: AUC Values for Various Classifiers

	German	Australian	Japanese
LDA	0.6518	0.6180	0.6450
Logistic	0.6227	0.5871	0.6136
1NN	0.6040	0.6486	0.6151
NaiveBayes	0.5727	0.6112	0.7217
SVM	0.6312	0.6172	0.7411
rpart	0.5557	0.5949	0.6675
C5.0	0.5789	0.5852	0.6881
Adabag	0.5904	0.6087	0.6677
RandomForest	0.5918	0.6384	0.6974
GBM	0.6305	0.6507	0.7537
XGBoost	0.6381	0.6513	0.7581
Blending	0.6418	0.6518	0.7718

Blending model: Blending model is ensemble made up of 3 algorithms gbm, rpart and treebag as base classifiers. Another gbm is used as meta learner.

IV. Conclusion

In this work, we have surveyed various machine learning algorithms with respect to credit scoring along with various ensemble techniques. In general, the stacking ensemble has produced the better results in terms of area under curve, which leads to better cost savings in credit scoring applications. Stacking is the optimum ensemble technique. Meta learner in stacking tries to learn what the base classifiers fail. So with stacking we can take advantage of both other bagging which reduces variance and boosting which reduces bias error. So stacking leads better accurate models than bagging and boosting with condition that the first level classifiers should be good enough.

References

[1] Cheng-Lung Huanga, Mu-Chen Chen b, Chieh-JenWang, Credit scoring with a data mining approach based on support vector machines, Expert Systems with Applications 2007.

[2] Van Gestel, Tony, Baesens, Bart, Garcia, Joao, Van Dijcke, Peter,"A support vector machine approach to credit scoring." Bank en Financiewezen, 2003.

[3] Thabiso Peter Mpofu, Macdonald Mukosera,"Credit Scoring Techniques: A Survey", International Journal of Science and Research (IJSR) 2014.

[4] Lieli, White,"The construction of empirical credit scoring rules based on maximization principles", Journal of Econometrics, July 2010.

[5] Lyn C. Thomas, David B. Edelman, Jonathan N. Crook, Credit Scoring and Its Applications", Society for Industrial and Applied Mathematics, 2002.

[6] Utkarsh Srivastava, Santosh Gopalkrishnan,"Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks", 2nd International Symposium on Big Data and Cloud Computing, 2015.

[7] Schreiner M,"Credit Scoring for Microfinance: Can It Work?", Journal of Microfinance, 2000.

[8] Hussein Abdou, Ahmed El-Masry, John Pointon,"On the applicability of credit scoring models in Egyptian banks.

[9] Amira Kamil, Ibrahim Hassan, Ajith Abraham,"Modeling Consumer Loan Default Prediction Using Ensemble Neural Networks."

[10] Hafiz Alaka, Lukumon O. Oyedele, Muhammad Bilal, Olugbenga O Akinade, Hakeem Owolabi, Saheed Ajayi "Bankruptcy Prediction of Construction Businesses: Towards a Big Data Analytics Approach", IEEE 2015.

[11] "Oluwarotimi Odeh, Praveen Koduru, Sanjoy Das, Allen M. Featherstone, Stephen M. Welch, "A multi-objective approach for the prediction of loan defaults", ACM Press, New York, NY, USA, 2011.

[12] Vadlamani Ravi; Sagar Koparkar; N. Prudhvi Raju; S. Sridher, Improving retail banking loans recovery via data mining techniques: A case study from Indian market", International Journal of Electronic Customer Relationship Management (IJECRM), Vol. 9, No. 2/3, 2015.

[13] Lifeng Zhou, Hong Wang, Loan Default Prediction on Large Imbalanced Data Using Random Forests, TELKOMNIKA Indonesian Journal of Electrical Engineering Vol. 10, No. 6, October 2012.

[14] S.M.Sadatrasoul, M.R.Gholamian, M.Siami, Z.Hajimohammadi,"Credit scoring in banks and financial institutions via data mining techniques: A literature review", Journal of AI and Data Mining, Vol. 1, No. 2, 2013.

[15] Sivasree M S, Rekha Sunny T, Loan Credibility Prediction System Based on Decision Tree Algorithm, International Journal of Engineering Research & Technology (IJERT), September 2015.

[16] H.A.Abdou, J.Pointon Credit scoring,"Statistical Techniques and Evaluation Criteria: A review of the literature", Intelligent systems in accounting finance & management, April 2011.

[17] Soni P M, Varghese Paul, M.Sudheep Elayidom,"An efficient data preprocessing framework for loan credibility prediction system", IJRET: International Journal of Research in Engineering and Technology, May-2015.

[18] Adel Lahsasna, Raja Noor Ainon, Teh Ying Wah,"Credit Scoring Models Using Soft Computing Methods: A Survey", The International Arab Journal of Information Technology, Vol. 7, No. 2, April 2010.

[19] Tian-Shyug Lee, I-Fei Chen,"A two-stage hybrid credit scoring model using articial neural networks and multivariate adaptive regression splines, expert Systems with Applications 2005.

[20] Nan-Chen Hsieh, Lun-Ping Hung,"A data driven ensemble classifier for credit scoring analysis, Expert Systems with

Applications 2010.
[21] Iain Brown, Christophe Mues,"An experimental comparison of classification algorithms for imbalanced credit scoring data sets", Expert Systems with Applications 2012.
[22] Loris Nanni, Alessandra Lumin,"An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring Expert Systems with Applications 2009".
[23] Dean Karlan, Sendhil Mullainathan, Omar Robles, "Measuring Personality Traits and Predicting Loan Default with Experiments and Surveys", January 2011.
[24] Ozef Zurada, Robert M. Barker,"Using Memory-Based Reasoning For Predicting Default Rates On Consumer Loans", Review of Business Information Systems First Quarter 2007.
[25] Adel Lahsasna, Raja Noor Ainon, Teh Ying Wah,"Credit Scoring Models Using Soft Computing Methods: A Survey", The International Arab Journal of Information Technology, April 2010.
[26] Dmytro Goriunov, Katerina Venzhyk,"Loan Default Prediction in Ukrainian Retail Banking", Economics Education and Research Consortium.
[27] Rob Gerritsen,"Assessing Loan Risks: A Data Mining Case Study", IEEE 1999.

Pramod Shrikrishna Patil received his B.E. degree in computer science and engineering from Shivaji University, Kolhapur, Maharashtra, India in 2012, and currently he is persuing M.Tech. degree in computer engineering from College Of Engineering, Pune, India. He was a teaching assistant with Department of Computer Engineering and Information Technology, College of Engineering, Pune, India, in 2014 -2016. His research interests include machine learning, data mining, credit scoring, ensemble learning. At present, He is engaged in BI tools and techniques for Indian businesses.