iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IX **Month of publication:** September 2018

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

An Expert System for Credit Risk Stratification using Data Mining

Deepika Tyagi[1], Sanjiv Sharma[2]
Department of Computer Science & Engineering, Madhav Institute of Technology and Science, Gwalior (M.P.), India

Abstract: In the current era, Credit business is one of the functions related to the financial institution that has been increasing rapidly. Previously credit risk assessment is done by the human judgements but now days, it is done by using data mining techniques which will improve the result in terms of accuracy. The speed for making the decision of granting loan is also improved as machines consume less time. This paper discusses the different data mining techniques to assess the credit risk assessment. After speculating the importance of different data mining techniques, came to the conclusion that combination of different algorithms can enhance the result of the expert system that will help the financial institution to stratify the risk that whether the customer will pay back the amount or not. The goal of this proposal is to develop the model that will help to assess the risk of defaulters by evaluating the combination of feature selection and classification algorithm.
Keywords: Credit risk, Data mining, Naïve bayes, Genetic algorithm, Bagging

I. INTRODUCTION

It is well known that money is needed for the development. Credit facilities play a vital role in ensuring the growth of developing countries. These credit facilities are provided by the financial institution such as banks. In the current era, credit business is increasing rapidly. Numerous financial institutions are set up to provide their services to the customer. The granting of loan by the financial institution is one of the key areas concerning decision problems that requires acquires subtle care. The advantage of information technology (IT) such as data mining are in used to solve such problems and to become more profitable. Predicting behavior and outcome plays vital role for the retaining or the sustainability of financial institution. In ensuring stability and profitability of a financial institution, proper management of granting loan facilities is vital. There is demand of efficient methods for handling loan granting process due to the rapid development in a business environment, emerging competition in the lending strategies, marketing strategies of financial institution and the borrowing patterns adopted by the customers.

The problem of the financial institution is to predict defaulters and non defaulters exactly as they are in actual so that financial institute will not suffer due to the wrong prediction of borrowers. This study is to classify the customers either as good payer or bad payer. Good payer customers are those who will repay the amount which is granted by the financial institution as loan or credit. While bad payers are those customers who will not repay the amount to the financial institution at the time which is given by the financial institution.

Thus, the financial institution faces credit risk to carry out its business. The credit scoring model has been developed to support traditional judgmental methods to correctly classify the credit or loan granting application. It becomes a challenging task to deal with large amount of data. Only financial institutions are stable and successful, which are able to cope up with that large amount of data and extracting the valuable knowledge from them to make the decision of granting loan.

Credit risk prediction is considered as classification problem in the field of data mining.

There is need of systems that can learn from the experience, history of the applicant and efficiently take the decision of granting loan is worthy or not. To decrease the cost and to improve the effectiveness and productivity, there are lots of tools and techniques are contributed in the financial domain which can ensure quick and quality delivery of service to the customer.

By using better tools and techniques time can be reduced which will spend on making the decision of granting. With the continuous development of the financial and banking institutions, problems associated with credit risk has also raised. Thus, a systematic methodology and accurate predictor are acquired to prepare accurate and realistic model as this type of model is useful for the financial institutions.

The model building process has to be systemized and the risk consolidation of the system is a computational challenge as it acquires a lot of information and calculation. In the context of this, new model is proposed. The motivation of this study is to understand the potential of data mining technologies to solve the problem of credit risk analysis of financial institution.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

A. Data Mining Techniques

In earlier days, various efforts are made to find solutions so that loans can be granted to the borrower efficiently and the borrower as well as the lender can earn profit. Researchers use data mining algorithms for the risk prediction accurately. Data mining is the process of extracting or mining the valuable information from the huge amount of data generated. Data mining techniques are association, classification, clustering, prediction. Various data mining that are used in the credit risk evaluation domain [1],[2]. Various statistical method such as LDA, decision tree (DT) [3],[4] and artificial intelligence techniques such as artificial neural network (ANN) [5],[6],[7],[8], support vector machine (SVM) [9],[11], K-nearest neighbor (K-NN) are utilized to analyze the defaulter risk in the financial institution. Fuzzy models are also beneficial for credit risk analysis model as fuzzy models have the ability to describe the behavior of systems by using linguistic rules [10]. Ensemble learning such as bagging, boosting etc., are used to classify the defaulters and non defaulters by using the combination of different classification algorithms. Probability of default is estimated and a threshold value is set to determine the potential borrower. Many solutions are obtained from the previous studies or experiments, but objective is to find the optimal solution, which can be find by using genetic algorithms (GA), principal search optimization (PSO) and many more.

II. LITERATURE REVIEW

Over the past decade, various researches have done in the credit risk assessment. Various statistical methods (such as LDA, decision tree), Artificial Intelligence methods (such as Artificial Neural Network, Support Vector Machine, K-Nearest Neighbor), ensemble methods (such as bagging, boosting) has been applied to the credit risk analysis domain to prevent the financial institution from granting loans or credit to the default customer or bad customer.

1) [11] states that simple linear model performs better than nonlinear kernel as nonlinear model over-fits the data. SVM with linear or Gaussian model performs best yielding the highest AUC. SVM with polynomial kernel and KNN standout poor. SVM weights and LR coefficient estimation indicates the contribution each feature to the risk of default. Large number of support vector are required to achieve better performance. This corroborates the finding of previous other researchers.
2) [12] presents the new procedure to build credal decision trees. Bagging scheme on this credal system outperforms the results obtained by previously presented methods for the credit scoring and bankruptcy prediction. According to this research work, B-CDT [bagging credal] is better than RSCDT [random subspace CDT]. In this way ensembling learning model or classifier is used for credit scoring.
3) Gradient boosting and random forest were focused in the context of credit scoring in dealing with sample where a large class imbalance was present [13]. It also states that the use of linear kernel LS-SVM would not be beneficial in the scoring of datasets where large class imbalance exists.
4) [14] proposed the mixture-of-experts and radial basis function neural network models must consider for credit scoring models.
5) The combinations of ANN and the rules of CART decision tree reduce the model error and increases the forecast accuracy of credit granting system [15]. Customers demographic data and payment related statistics were analyzed to identify feature variables and applied ANN to predict customers pattern of consumption, payment and default and utilized decision tree to develop set of principle for credit sanction.
6) CRED based knowledge discovery method that covers both classification and rule based extraction for CRA of Turkish SME customer portfolio [16]
7) [17] dataset is taken from the january2005 to april 2009 from customer transactions and credit bureau. Machine learning techniques are applied to construct nonlinear nonparametric forecasting models of consumer credit risk. It significantly improve the the classification rates of credit delinquencies and defaults with linear regression.
8) [18] Kohonen network is used to generate clusters of credit delinquents of the department store for the credit prediction and then used Cox's proportional hazard model to analyze the patterns of credit recovery for each delinquent segment. This study is to analyze the delinquent customers who had recovered from credit delinquency state to good credit state.
9) Bagging and random subspace ensemble learning with decision tree as base learner is used on the dual strategy for the credit scoring. RS bagging DT and bagging RS DT provides improved accuracy in the comparison of four ensemble learning includes bagging DT, random subspace DT, random forest, rotation forest [19].
10) Bagging, boosting and stacking with SMO, J48, logistic as base learner is implemented over the German and Australian dataset and find that bagging perform better than boosting and gives better results in terms of accuracy, type 1 error and type 2 error [20].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

III. PROPOSED METHODOLOGY

The emphasis of this paper is on evaluating the performance of credit risk analysis model using bagging as ensemble learning algorithm and naïve bayes as base machine learning algorithm.

A. Simulation Tool

Weka is a tool which is used to perform data mining tasks such as data preprocessing, classification, clustering, association, visualization, regression, because this tool consists of machine learning algorithm. Weka tool is used to create models for classifying the credit as good or bad. Classification algorithm are used in this study over the weka tool.

B. Methodology

1) Data collection- Collecting dataset for this study.
2) Feature selection- Selecting the important features of the database by using genetic search which are then potentially useful for credit risk evaluation.
3) Training- Train model using bagging ensemble learning algorithm and naïve bayes as base learner.
4) Cross validation is applied.
5) Testing- Test dataset is applied on the trained model so that it can predict the risk of the borrowers who applied for the credit or loan.
6) Result analysis-This study shows good results in terms of prediction.

IV. DATASET DESCRIPTION

The German credit dataset used for this study was taken from the UCI Machine learning database. It contains 1000 instances with 20 attributes where 13 attributes are categorical or qualitative and 7 numerical attributes. These instances are divided into two classes as good credit and bad credit.

Table 1: Dataset attributes

S.No.	Attributes
1.	Checking status
2.	Duration in month
3.	Credit history
4.	Purpose
5.	Credit amount
6.	Savings status
7.	Employment
8.	Installment commitment
9.	Personal status and sex
10.	Other debtors / guarantors
11.	Present residence since
12.	Property
13.	Age in years
14.	Other installment plans
15.	Housing
16.	Number of existing credits at this bank
17.	Job
18.	Number of dependents
19.	Telephone
20.	Foreign worker

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

V. DATA MINING ALGORITHMS

A. Data Preprocessing

Data preprocessing is the essential step. Data preprocessing is essential because data have different scales which would make the results biased. In the data preprocessing phase, resampling instance filter is utilized and split the dataset instances into 80% as training dataset and 20% as test dataset.

1) *Feature Selection:* Features subset selection is the process of identifying and removing irrelevant and redundant data information as much as possible. Feature selection is one of the important task to determine appropriate features for the credit model that increase the performance of classification [21]. Thus feature selection is the process of determining the subset of attributes which are more essential in the knowledge discovery from large number of existing set of attributes. This is also known as dimensionality reduction. The training phase is more difficult if there is irrelevant, redundant and noisy information is available in the data. The objective of feature selection are as follows:

a) Reduces the dimensionality of the feature space
b) Obtain features which are more relevant to classification problem
c) Speeds up the process
d) Reduces the cost of learning

Genetic algorithm is a heuristic search algorithm for solving the optimizing problems. GA algorithm works on the principle of natural selection and survival of fittest [22]. Genetic algorithm is used to narrow the search space as it is used for the optimal feature subset selection. Genetic algorithm works on potential solutions simultaneously called generations. Each individual in the generation is represented by chromosome. By using genetic operator such as selection, crossover and mutation operator, new batch of chromosome is generated from the chromosomes of previous generation in each iteration. Fitness score is used to represent the goodness of the corresponding solution. Fitness function is used to incorporate different optimization criteria and provide a numerical value to the goodness of chromosomes. Chromosomes are ranked and chosen as parents, according to fitness score. Chromosomes having best fitness score are selected as parent chromosome for the next generation. The genetic algorithm stops if in the further iteration, fitness score of the best chromosomes did not improve. In this study, genetic search with classifier subset evaluator applied on the dataset for attribute selection to select optimal 16 attributes which are important in the consideration of the credit risk analysis domain. Genetic algorithm works in following steps:

i) For each chromosome, fitness score is calculated.
ii) On the basis of fitness scores, parent chromosomes are selected.
iii) New chromosomes are created by combining selected parent chromosomes using crossover operator.
iv) Mutate some of the chromosomes.

Table 2: Genetic Parameters

Parameter	Value
Size of population	20
Mutation rate	0.033
Crossover rate	0.6
Number of generation	20

Table 3: Selected Attributes

S. No.	Attributes
1.	Checking Status
2.	Duration in month
3.	Credit history
4.	Purpose
5.	Credit amount
6.	Savings status
7.	Employment
8.	Installment commitment
9.	Personal status and sex
10.	Other debtors / guarantors
11.	Present residence since
12.	Job
13.	Number of dependents
14.	Telephone
15.	Foreign worker

2) Ensemble Learning

Ensemble learning is the process by which strategically multiple classifiers are generated and combined. The general principle of ensemble method is to construct a combination of some model fitting method or statistical learning, instead of using single learning method. Ensemble learning methods like bagging and boosting have recently been used in various domains owing to the improvement in computational efficiency. However, with the advent of wide variety of applications of machine learning techniques to solve classification problems, further focus is needed to improve the performance measures such as specificity (TNR) and sensitivity (TPR). There are various kinds of ensemble learning methods (bagging, boosting) which are considered as sub category of hybrid intelligent system ensemble learning is a method to combine multiple classifiers to solve the machine learning problem. It is used to improve the predictive performance of model. There are two types of multi classifier system. One of them is serial multi classifier system and the other one is parallel multi classifier system. Ensemble learning is used to classify the borrowers as good or bad (i.e. defaulter or non-defaulter). In this study bagging is used to classify and predict the class or category of borrower either as defaulter or as non-defaulter. Naïve bayes is used as base learner in the study.

3) Bagging

Bagging stands for the bootstrap aggregating. It is the simplest ensemble learning that gives good result [19],[20]. From the training dataset, different training data subsets are randomly drawn with replacement. When available data is of limited size bagging is useful. Bagging is a variance reduction technique for the base procedure [19], [20]. It has attracted very much attraction, probably due to its implementation simplicity and bootstrap technology. Bagging is a smoothing operation which turns out to be very advantageous when aim is to improve the predictive performance of classification problem. It also reduces the mean squared error.

4) Naïve bayes

The Naïve Bayes algorithm is the probabilistic classifier that works on the basis of Bayes theorem from probability theory [23]. Naïve Bayes has the independent feature capability. Naïve bayes classifier is a special case of Bayesian theorem. Naïve bayes algorithm incorporates decision rule. Advantages of naïve bayes is that it is easy to implement and training and classification low computational cost naïve bayes. Disadvantage is the low classification quality.

$$P(h|D) = \left(\frac{P(D|h)P(h)}{P(D)}\right)$$

Where P(h): Prior probability of hypothesis h (prior)
P(D): prior probability of training data D (evidence)
P(D|h): conditional probability of D on given h (likelihood)
P(h|D): conditional probability of h on given D (posterior probability)

VI. PROPOSED FRAMEWORK

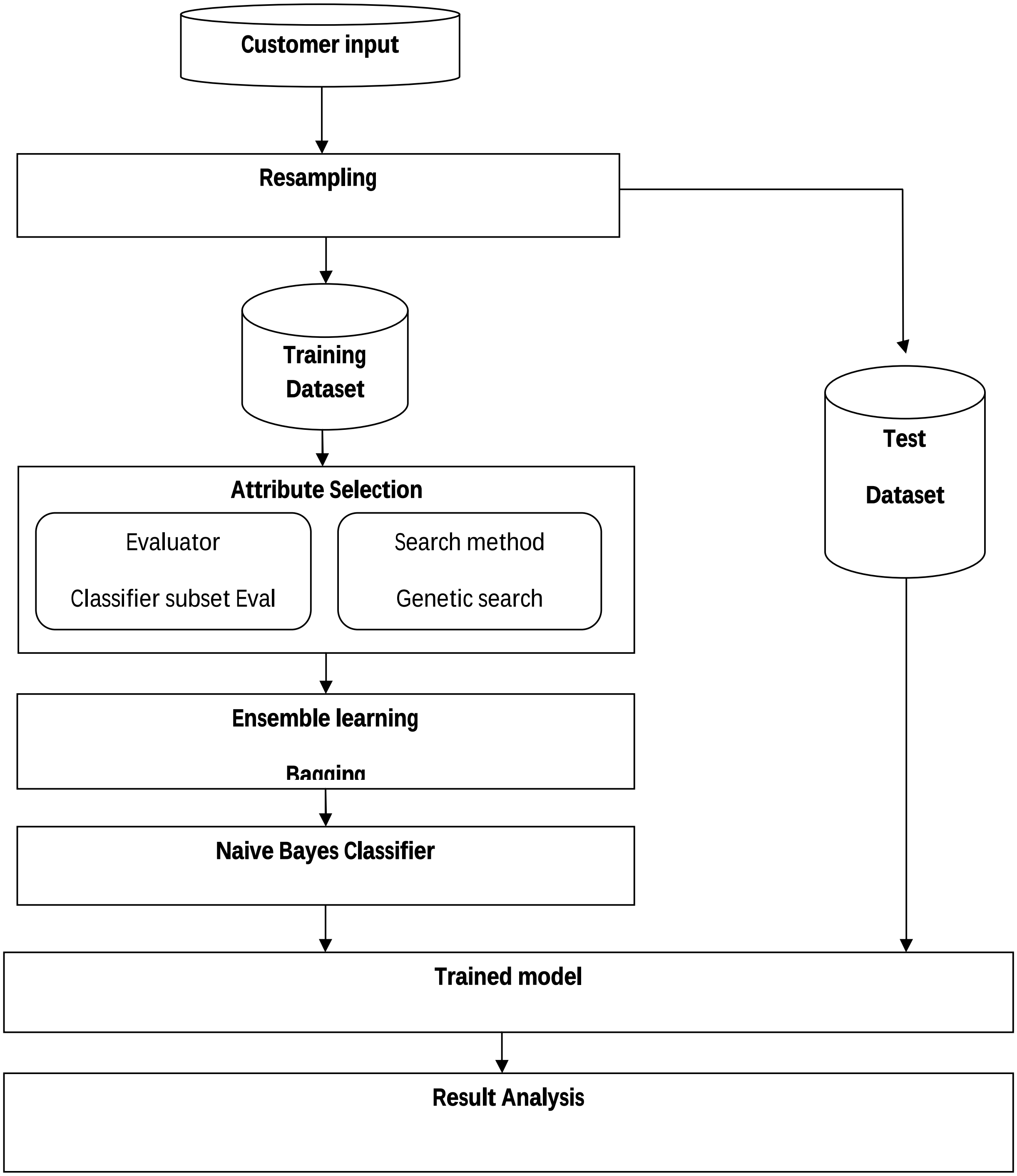


Fig. 1 Proposed Framework

Example

1) *Non Defaulter (Good Customer)*

P(good)= 0.72125

P(d|h) is calculated as following:

$$(P(d|h) = P(attribute\_1 \mid class) * P(attribute\_2 \mid class) * P(attribute\_n \mid class))$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

P(d|h)=3.0455616
P(h|d) is calculated as following:

$$P(h|d) = \left(\frac{P(d|h)P(h)}{P(d)}\right)$$

P(h|d)=3.0455616*0.72125

P(h|d)=2.1928

2) Defaulter (Bad customer)
P(bad)=0.27875
Similarly, P(d|h) is calculated here as following:

$$(\,P(d|h) = P(attribute\_1 \,|\, class) * P(attribute\_2 \,|\, class) * P(attribute\_n \,|\, class)\,)$$

P(d|h)=1.1513042

$$P(h|d) = \left(\frac{P(d|h)P(h)}{P(d)}\right)$$

P(h|d)=1.1513042*0.27875
P(h|d)=3.22365
MAP(H)=MAX(P(d|h)*P(h))
It results that customer will be defaulter.

VII. EXPERIMENTAL RESULT

The proposed model is a combination of different algorithm such as Classifier subset evaluator and genetic algorithm for feature selection and bagging with naïve bayes as a base learner for classifying the credit risk assessment. The following figure represents results of the proposed model on WEKA 3.8 in terms of different performance measure such as accuracy.

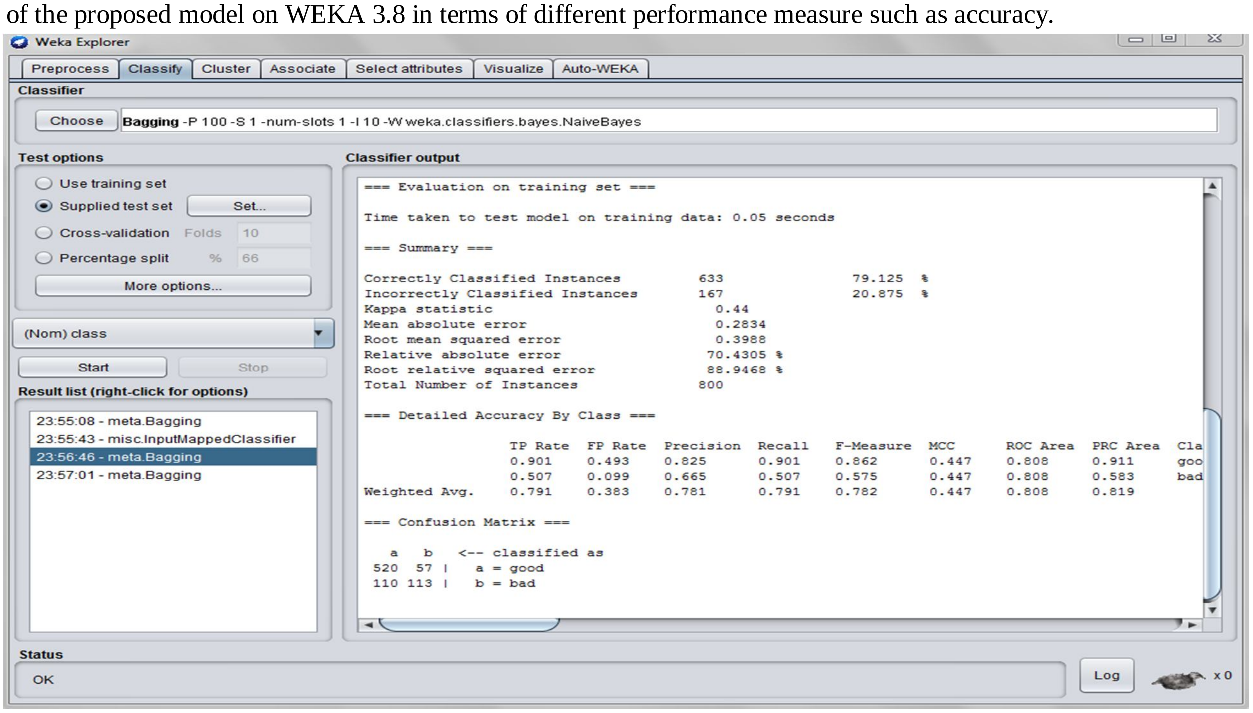


Fig. 2: Proposed Model Performance on Weka

A. Performance Measure

The confusion matrix is in the form of table contains rows and columns. Rows define the actual class and columns represent the predicted class. TP refers to the number of customers which are actually good and predicted as good, TN refers to the number of customers which are actually bad and predicted also as bad. FP defines the number of customers which are actually bad and this proposed model predicts them as good customers. FN defines the number of customers which are good and predicted as bad.

Table 4: Confusion Matrix

Customers		Predicted	
		Good	Bad
Actual	Good	520	57
	Bad	110	113

True Positive (TP) = 520
True Negative (TN) =113
False Positive (FP) =110
False Negative (FN) = 57

Accuracy is one of the various performance measures that are included in this proposed work. An estimation of classification accuracy on new instances is the most common performance evaluation criteria to evaluate the performance of any model. In this study, accuracy is the primary evaluation criteria for experiments using feature selection with classification algorithm. If the accuracy of learning algorithm improves by reducing the dimensionality of the data then feature selection is considered useful. This study analyzes that accuracy is improved. In this way it can be said that this model provides better result. The financial institution has to suffer in two cases i.e. if good customer is predicted as bad and when bad customers are predicted good. The aim is to reduce the FP. As compared to other model this proposed model reduces the FP value. Because if good customer is predicted as bad then financial institution will not grant credit or loan to the customer and losses the chances to earn profit in the form of interest which they can earn if those good customers are predicted as good. While if bad customer are predicted as good customer then bank or financial institution grant loan to them and bad customers will never repay the credit amount and financial institution has to suffer a lot.

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN}\right)$$

$$Acuracy = \left(\frac{520 + 113}{520 + 113 + 110 + 57}\right) = 0.79125$$

Now analyze the result in detail that are presented below based on the evaluation. The results are arranged in the following order. First, the table the accuracy is depicted and the graph of accuracy is plotted. Here, the following tables give the summary of the results obtained from the instances of the dataset. This graph is very helpful in visualizing the improvement in the proposed model that can be utilized for the prediction task. The X-axis of the graph has the name of the classification algorithm and the Y- axis has the accuracy value of the specific classifier algorithm. The following table and graph shows that genetic search and bagging with naïve bayes provides high accuracy in comparison to other combination of classification algorithm for classifying defaulter and non defaulter.

Table 5: Results

Algorithm	Accuracy
Bagging NB	76.125
Bagging NB-GS	79.125
Bagging LRA	75.625
Bagging J48	75.875
Bagging SMO	76.375

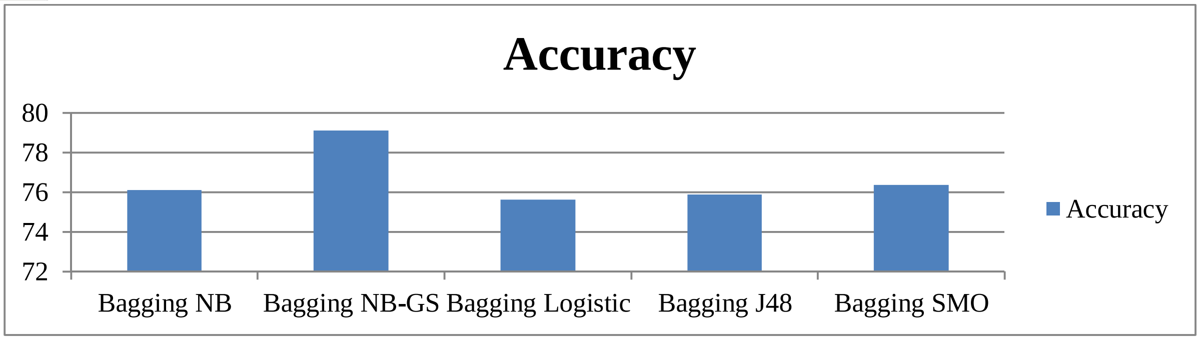


Fig. 3: Accuracy of Different Classification Algorithm

The graph shows the TP, FP, FN and FP of various classification algorithm. This graph represents that lowest false positive (FP) by applying genetic search and bagging with naïve bayes. As if the bad customer is predicted as good customer by the credit risk assessment system and the financial institution grants loan to the bad customer and who will never repay the payment back thus, the aim should be to obtain the less FP by the credit risk assessment system and this proposed methodology provides less FP.

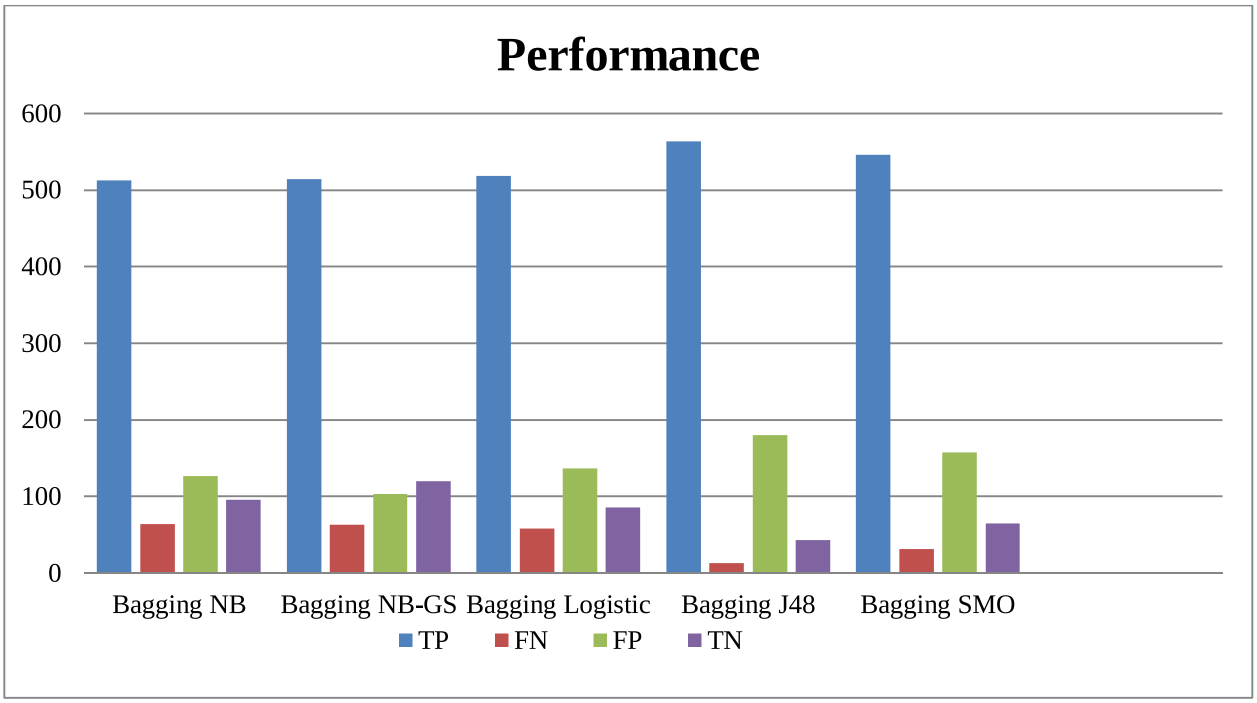


Fig. 4: Performance Comparison Of Different Classification Algorithm

The following graph is plotted which represents the TPR and FPR calculated by the above estimated information. Here, the bagging with naïve bayes applied on 20 attributes and in other case firstly feature selection is applied which selects the subset of attributes from the given set of attributes and after that bagging with naïve bayes is applied. The graph depicts the higher TPR and lesser FPR in the second case which justifies that the proposed methodology provides better performance.

$$TPR = \left(\frac{TP}{TP + FN}\right)$$

$$TPR = \left(\frac{520}{520 + 57}\right) = 0.90121$$

$$FPR = \left(\frac{FP}{FP + TN}\right)$$

$$FPR = \left(\frac{110}{110 + 113}\right) = 0.49327$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

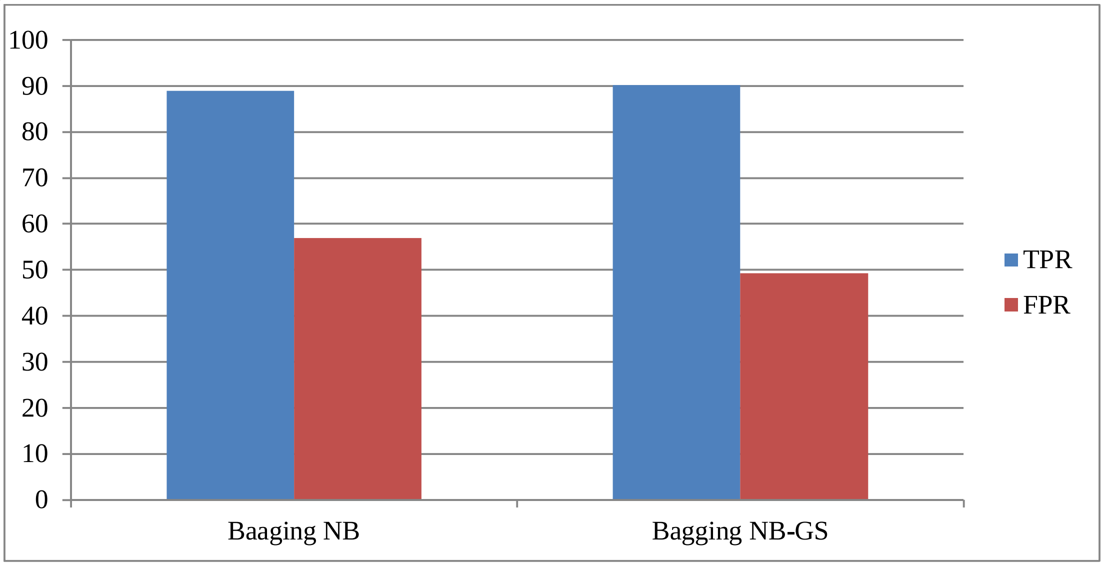


Fig. 5: TPR and FPR Rate Representation

The bar graph is plotted to compare the correctly classified instances and incorrectly classified instances through the Naïve bayes only, bagging with naïve bayes as base learner and feature selection with genetic search followed by bagging bayes with naïve bayes as base learner. This comparison shows that the third case i.e. Genetic search followed by bagging naïve bayes, represents more correctly classified instances and less incorrectly classified instances.

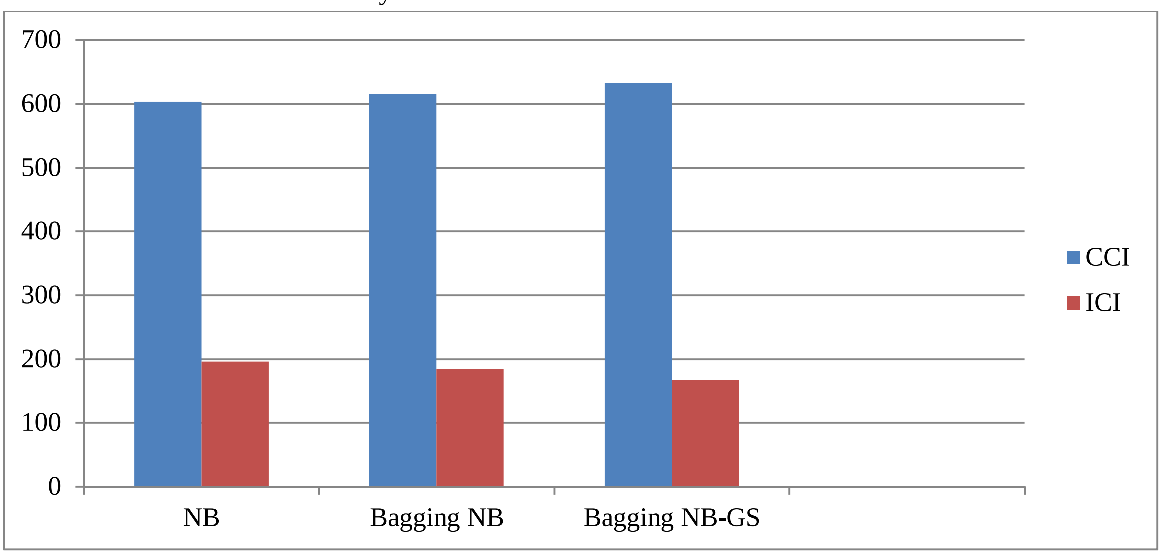


Fig. 6: Graphical Representation of correctly and incorrectly classified instances

VIII. CONCLUSION

This study aims to propose model that predicts the risk of granting credit to the borrower who apply in the financial institution. This proposed methodology utilizes the data mining techniques in the domain of data classification. This paper demonstrates that there are different classification algorithms are implemented in this field so that misclassification of the customers can be reduced and prediction of customer should be accurate so that customer and financial institution both can received benefit without incurring loss to each other. This paper concludes that the performance of the proposed model gives better result in terms of accuracy for the credit risk stratification. The next step in this line of research is to extend the analysis for other risk factors.

IX. ACKNOWLEDGEMENTS

I would like to thank Madhav Institute of Technology and Science (MITS) Gwalior and the faculty member of this institute for their support in this proposed work.

REFERENCES

[1] Bhekisipho Twala, "Multiple classifier application to credit risk assessment", Expert system with applications (37) 3326-3336, 2010.

[2] Deepika Tyagi, Sanjiv Sharma, "Credit risk assessment using data mining approaches: A review", National conference on smart materials: energy and environment for smart cities, ISSN: 2454-5155, 43 February 2018.

[3] Antonakis, A., Sfakianakis, M., "Assessing naïve bayes as a method for screening credit applicants", Journal of applied statistics 36(5), 537-545, 2009.

[4] Bahnsen, A., Aouada, D., Ottersten, B., "Example dependent cost sensitive decision trees", Expert system with application 42 (19) 6609-6619, 2015.

[5] Abdou, H., Pointon, J., El-Masry, A., "Neural nets versus conventional techniques in credit scoring in Egyptian banking", Expert system with applications 35 (3), 1275-1292, 2008.

[6] Akkoc, S., "An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system model for credit scoring analysis: the case of Turkish credit card data", European journal of operational research 222 (1), 168-178, 2012.

[7] Baesens, B., Van Gestel, T., Stepanova, M., Van Den Poel, D.,Vanthienen, J., " Neural network survival analysis for personal loan data", Journal of the operational research society 56 (9), 1089-1098, 2005.

[8] Bekhet, H., Eletter, S., "Credit risk assessment model for Jordanian commercial banks: Neural scoring approach", Review of development finance 4 (1), 20-28, 2014.

[9] Harris, T., "Credit scoring using the clustered support vector machine", Expert system with applications 42 (2), 741-750, 2015.

[10] A. Lahsasna, R.N. Ainon, T. Ying Wah, "Intelligent credit risk evaluation system using evolutionary neuro- fuzzy scheme", IEEE 1-4244-1355, 2007.

[11] Tony Belloti, Jonathan Crook, "Support vector machines for the credit scoring and discovery of significant features", Expert system with applications (36) 3302-3308, 2009.

[12] Joaquin Abellan, Carlos J Mantas, "Improving experimental studies about ensembles of classifier for bankruptcy prediction and credit scoring", Expert system with applications (41) 3825-3830, 2014.

[13] Iain Brown, Christophe Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring datasets", Expert system with applications 3446-3453, 2012.

[14] Tsai, C.-F. and J.-W. Wu., "Using neural network ensembles for bankruptcy prediction and credit scoring", Expert Systems with Applications, 34(4), 2639-2649, 2008.

[15] S.C. Chen, M.Y. Huang, "Constructing credit auditing and control & management model with data mining technique", Expert system with applications, (38) 5359-5365, 2011.

[16] Gulnur Derelioglu, Fikret Gurgen, "Knowledge discovery using neural approach for SME's credit risk analysis problem in Turkey", Expert system with applications (38) 9313-9318, 2011.

[17] Amir E, Kandani, Adlar J. Kim, Andrew W. Lo, "Consumer credit risk models via machine learning algorithms", Journal of Banking & finance, 34, 2767-2787, 2010.

[18] Sung Ho H., "Behavioral assessment of recoverable credit of retailer's customers", Information Sciences, 180, 3703-3717, 2010.

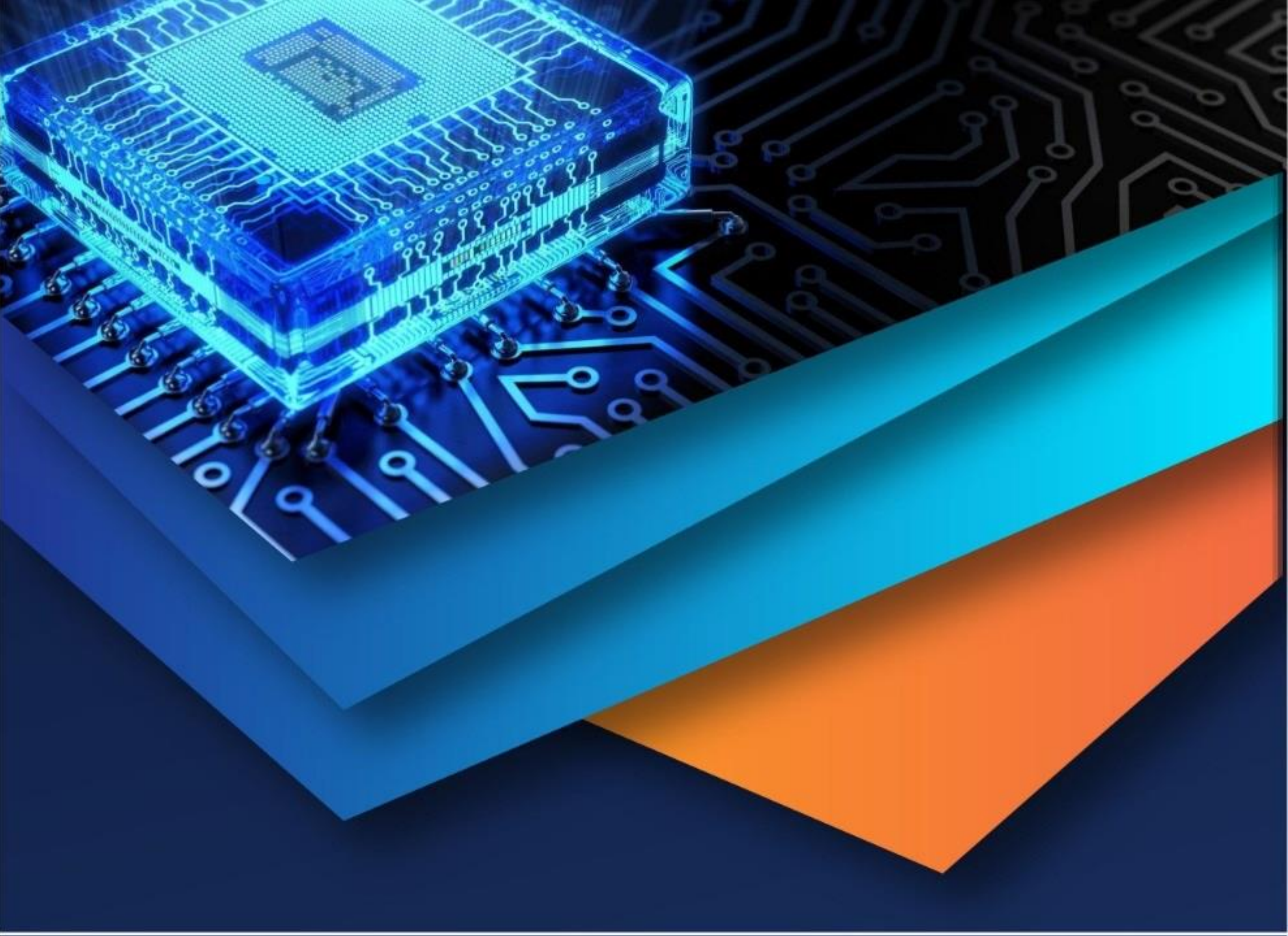
[19] Wang, G., Jian M., Lihua H., Kaiquan X., "Two credit scoring models based on dual strategy ensemble trees. Knowledge Based Systems, 26(0), 61-68, 2012.

[20] G. Wang, Jinxing H., Jian M., Hongbing J., " A comparative assessment of ensemble learning for credit scoring tree", Expert system with applications (38) 223-230, 2011.

[21] Aryuni, M., Madyatmadja, E., "Feature selection in credit scoring model for credit card applicants in XYZ bank: A comparative study", International journal of multimedia and ubiquitous engineering, 10 (5), 17-24, 2015.

[22] Pedro G.Espejo, Sebastian Ventura, and F. Herrera, "A survey on the application of genetic programming to classification", IEEE transactions on systems, 40 (2), March 2010.

[23] Antonakis, A., Sfakianakis, M., "Assessing naïve bayes as a method for screening credit applicants", Journal of applied statistics 36(5), 537-545, 2009.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY