A Rank Aggregation Algorithm for Ensemble of Multiple Feature Selection Techniques in Credit Risk Evaluation

Shashi Dahiya
Deptt. Of Computer Science and Engineering
Manav Rachna International University, (MRIU)
Faridabad, India

S.S Handa
Deptt. Of Computer Science and Engineering
Manav Rachna International University, (MRIU)
Faridabad, India

N.P Singh
Management Development Institute, (MDI)
Gurgaon, India

***Abstract*—In credit risk evaluation the accuracy of a classifier is very significant for classifying the high-risk loan applicants correctly. Feature selection is one way of improving the accuracy of a classifier. It provides the classifier with important and relevant features for model development. This study uses the ensemble of multiple feature ranking techniques for feature selection of credit data. It uses five individual rank based feature selection methods. It proposes a novel rank aggregation algorithm for combining the ranks of the individual feature selection methods of the ensemble. This algorithm uses the rank order along with the rank score of the features in the ranked list of each feature selection method for rank aggregation. The ensemble of multiple feature selection techniques uses the novel rank aggregation algorithm and selects the relevant features using the 80%, 60%, 40% and 20% thresholds from the top of the aggregated ranked list for building the C4.5, MLP, C4.5 based Bagging and MLP based Bagging models. It was observed that the performance of models using the ensemble of multiple feature selection techniques is better than the performance of 5 individual rank based feature selection methods. The average performance of all the models was observed as best for the ensemble of feature selection techniques at 60% threshold. Also, the bagging based models outperformed the individual models most significantly for the 60% threshold. This increase in performance is more significant from the fact that the number of features were reduced by 40% for building the highest performing models. This reduces the data dimensions and hence the overall data size phenomenally for model building. The use of the ensemble of feature selection techniques using the novel aggregation algorithm provided more accurate models which are simpler, faster and easy to interpret.**

Keywords—Classification; Credit Risk; Feature Selection; Ensemble; Rank Aggregation; Bagging

I. Introduction

The data size is increasing regarding records and dimensions both. It presents challenges to the machine learning community which is working on new methods and techniques to fasten the data exploration, analysis, and validation tasks. One way of handling this problem is by using an effective sampling methodology to choose a subset of samples describing the dataset as a whole. This method results in a reduced dataset having less number of instances. Another way of handling this problem is to use an appropriate dimensionality reduction/ feature selection method to reduce the dimensions of the dataset.

In a vital machine learning problem of classification, the accuracy of a classifier plays an important role. The accuracy of the classifier depends on many factors such as – the single, hybrid or an ensemble method used for modelling; the base models used for the ensemble; the learning algorithm used for model training; the feature selection method used for selecting the relevant features; the sampling technique used for sampling the data; the evaluation method used for testing the model and many more.

Feature selection is an important pre-processing step in machine learning and pattern recognition problems. It has been an active area of research since past three decades [1]. Feature selection increases the performance of classification models by eliminating redundant and irrelevant features and thus reducing the dimensionality of datasets [2]. This study uses the feature selection approach for the enhancement of accuracy of credit risk evaluation models.

A. *Credit Risk Evaluation*

Quantifying the credit risk is a typical bank decision problem of classification in which the new loan applicants are to be classified accurately into either a creditworthy or a non-creditworthy category based on the historical dataset of loan applicants. This historical dataset is used for training the classifier, and the new loan applicant’s data is tested on this trained classifier. The Class labels i.e. creditworthy or non-creditworthy are automatically assigned to the new applicants records during testing phase. The credit dataset contains the features mainly describing the financial status, demographic details of the applicant and his personal profile. Some features of the dataset may provide more significant information needed for classifying a new loan applicant than others. While some of the features are not required, some may contain redundant or irrelevant information and don’t provide any additional information during the model development task. They don’t contribute to the accuracy of the model and sometimes even decrease it by slowing down the classifier learning process. The big feature set can make a more complex model whose interpretation also becomes cumbersome. It can make a classifier overfitting the training data [3].

B. Feature Selection for Credit Evaluation

In credit risk evaluation the accuracy of the classifier is very crucial. Even a small increase in model accuracy may result in huge profit for the bank. For performance enhancement of single models, the literature proposed the hybrid and ensemble based models. In credit risk evaluation. Many of the ensemble based and hybrid models are developed using feature selection methods during the initial stage [4]. Feature selection is crucial for the selection of significant and appropriate features for model development. If the number of features is large, more computation is required, and the accuracy and interpretation of the classification model decrease [5], [6]. A large number of features in credit evaluation implies that there are a large number of questions for the loan applicants, which will be time-consuming and confusing. According to [7], exploring a big number of features lead to identifying a relevant subset of features for building the credit model.

The relevance of the features needs to be identified before the model development task so that the undesired, redundant and irrelevant features are not used as input to the model. Supervised feature selection determines relevant features by their relations with the corresponding class labels and discards irrelevant and redundant features. The subset of features identified as important will help in reducing the size of the hypothesis space and allows the algorithms to operate faster and more effectively [8]. This smaller feature subset will help in building simplified models reducing the time and space complexity of the algorithms and hence improving the accuracy with well interpreted results.

The purpose of this paper is the enhancement of classification accuracy of the credit risk evaluation models. This study uses the ensemble of multiple feature selection techniques for ranking and selecting the significant features.

II. Feature Selection Criteria for Filter based Feature Selection

The filter approach to feature selection works independently of learning/Induction algorithm (Fig. 1.). It operates as a pre-processing step and selects and presents the important features to the learning algorithm as input. Filter approach makes use of the complete training data for its operation. It ranks the features in accordance with their importance w.r.t selecting a class. A threshold has to be then defined for selecting the number of most important features from the ranking.

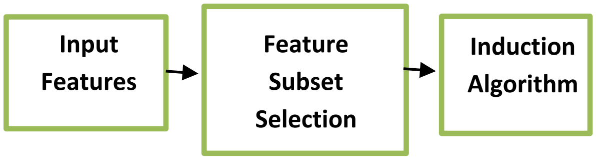

Fig. 1. The Filter based method of Feature Selection

There are several features ranking methods [9] available in the literature, some of them are - correlation based, mutual information based and methods based on decision tree and the distance between probability distributions. Any of the predefined measures such as – the Dependency measures, Information measures, distance measures [10] [11], independent component analysis [12], class separability measure [13], or variable ranking [14] are the basis of these feature ranking methods.

A. Dependency measures

As discussed by [15] and [2], the dependency measures or correlation measures quantify the ability to predict the value of one variable based on the value of the other. The Pearson's correlation coefficient (PCC) is very useful for feature selection [16] [17], as it quantifies the relationship of a feature with its corresponding class label and with other features in the dataset. As per [18], PCC for continuous features is a simple measure but can be effective in a wide variety of feature selection methods.

A uniform manner is used to treat the features and the class, then the feature-class correlation and feature-feature inter-correlations are calculated according to the following equation:

$$\mathrm{CC}\ (X_j,\, c) = \frac{\left[\sum_{i=1}^{m} (x_i^j - \bar{X}^j)(c_i - \bar{c})\right]}{\sigma_{Xj} \cdot \sigma_c}$$

$\bar{X}^j$ and σ_{Xj} are the mean and standard deviation of j^{th} feature and $\bar{c}$ and σ_c are the mean and standard deviation of vector c of class labels). The ranking values are absolute values of CC:

J_{CC} (Xj) = | CC (Xj, c) |

This ranking has a low complexity of the order of *O (mn)* and is very simple to implement for numerical variables.

For, nominal or categorical variables the popular feature selection method used is Pearson's chi-squared (χ 2) test. The numerical variables can also be converted into nominal or categorical types for applying the χ 2 test. First, a contingency table is made by converting the raw data. Then, the independence between each variable and the target variable is measured using the contingency table. χ^2 is defined by :

$$\chi^2 = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency; E_i is the expected theoretical frequency, asserted by the hypothesis of independency and c the number of cells in the contingency table.

Correlation-based feature selection is the base for symmetrical uncertainty (SU) also. It is a symmetric measure and can be used to measure feature-feature correlation. The value of symmetrical uncertainty ranges between 0 and 1. The value of 1 indicates that one variable (either X or Y) completely predicts the other variable [19] .The value of 0 indicates that both variables are completely independent.

B. Information Measures

Information theory has been proved to be very successful in solving many problems [20]. It provides a theoretical framework for measuring the relation between the classes and a feature or more than one feature. Mutual Information (MI) is a filter-based feature selection metric used to find the relevance of features. It works on the principle of information shared by two features using MI [20], the relevance of a feature subset on the output vector C can be quantified. Formally, the MI is defined as follows:

$$I(x;y) = \sum_{i=1}^{n}\sum_{j=1}^{n} p(x(i), y(j)).\log(\frac{p(x(i),y(j))}{p(x(i)) \cdot p(y(j))})$$

Where MI is zero when x and y are statistically independent, i.e., p(x(i), y(j)) = p(x(i))·p(y(j)).

Large values of MI indicate a high correlation between the two features and zero indicates that two features are uncorrelated. Many feature selection methods are proposed based on MI such as [20] [21].

Information Gain (IG) and Gain Ratio (GR) are feature ranking methods based on information measures. IG is the reduction in entropy of the class variable when the value of the independent variable is known. The IG of an attribute X with respect to class variable Y is given by:

$$I(Y;X) = H(Y) - H(Y|X)$$

Where *H(Y)* is the entropy of Y,

H (Y|X) is the uncertainty about Y for a given X

The information gain measure is biased towards tests with many outcomes. Therefore C4.5 uses Gain Ratio (GR) for overcoming this bias and is an extension of IG.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

Where Gain(A) is the encoding information gained by branching on A and SplitInfo(A) is the information got by splitting the dataset into ‘n’ distinct values of the attribute A. The maximum GainRatio attribute is subject to splitting.

C. *Distance Measures*

Distance measures, also known as separability, divergence, or discrimination measures, study the difference between the two-class conditional probabilities in a binary context [15] [22]. In other words, a feature X_j is chosen over another feature X_j’ if it induces a greater difference between the two-class conditional probabilities than X_j’. In the case where the difference is zero then the two features are identical. Relief is one of the most famous feature selection method based on distance measures. Relief algorithm has been given by [23]. It is a multivariate method which is sensitive to interactions [24]. It estimates the features relevance according to how well their values distinguish between the instances of the same and different classes that are near each other. It performs well on small sample size datasets having a large collection of features. Its computational complexity is O (mn), which is linear in comparison to other multivariate methods often having quadratic complexity in the number of features.

D. *Feature Ranking*

Feature ranking uses the above discussed filter based measures to compute a scoring function from the values (x^j_i ; y_i). It is considered that a high score indicates a valuable feature and the features are sorted in decreasing order of the scoring function [25]. It is computationally efficient since it requires only the computation of d scores and sorting them. It is statistically robust against overfitting because it introduces bias, however it may have considerably less variance [26]. Therefore, feature ranking can be preferable than any other feature selection method.

III. Background

In general the feature ranking criteria for filter based feature selection discussed above have one or the other limitation in their performance. The distance based measures like - Relief are good in capturing the relevance of features to the target variable but doesn’t capture the redundancy among the features. The dependency measure such as PCC is not able to capture the correlations that are not linear [2]. The dependency measures and information measures suffer from time complexity issues since they have to evaluate all possible subsets. Therefore they are not practical to deal with high dimensional data.

Due to these limitations of the filter based methods, it is difficult to find out the best criteria for a particular problem.

According to [27] this problem is called the selection trouble. The best approach is to independently apply a combination of the available methods and evaluate the results.

Aggregating the ranked lists from individual rankers into a single better ranking is called as rank aggregation. Rank aggregation method is an Ensemble based feature selection method which is considered as an upcoming important tool for combining information with the purpose of getting higher accuracy.

IV. Ensemble Method for Feature Selection

An ensemble of classifiers is a set of base Classifiers that are individually trained. For classifying new instances, the decisions of these classifiers are combined using weighted or un-weighted majority voting [28] [29]. According to [30], the ensemble model could outperform the single base models when weak/ unstable models are combined. Looking at advantage of ensemble based classifiers over individual ones, the concept of ensemble can be applied for performance enhancement in the feature selection process also.

A. *Ensemble of a Single Feature Ranking Technique*

Ensemble of a single feature ranking technique involves Bagging (Bootstrap Aggregation) or some other Algorithms to generate various bags of data. For each bag the feature ranking is done and the ensemble is formed by combining the individual bag rankings by weighted voting, using linear aggregation [31].

B. *Ensemble of Multiple Feature Ranking Technique*

In this method, multiple feature ranking techniques are used for ranking the features in order of their relevance for building an ensemble. The same training data is used by the ranking methods and the results of these methods i.e. the ranking lists are combined in a certain way to obtain a final ranked list of the features. Thus, multiple feature ranking lists creates a single feature ranking list in the following two steps: First a set of different ranking lists are created using corresponding rankers and secondly these ranking lists are combined using rank ordering of features [32].

Suppose a dataset 'D' has 'I' instances and 'k' features. During the first step a set of n ranking lists {F1, F2, F3…Fn} are obtained (one for each 'n' feature selection methods used).

In the second step, a rank aggregation method R is used for combining the ranks of individual features from n ranking lists obtained in first step. Let f_i^j be the rank of feature i from ranking list j, then the set of rankings of feature i is given by:

$R_i = f_i^1, f_i^2, f_i^3, \ldots\ldots, f_i^n$

The new rank obtained by feature i using the combination method C is

$\breve{F}_i = R\ (f_i^1, f_i^2, f_i^3, \ldots\ldots, f_i^n)$

C. Rank Aggregation

There are different combination or rank aggregation methods used for creating an aggregated feature ranking list from various individual feature ranking lists for the ensembles of multiple feature selection techniques. Recently, there have been studies applying the ensemble concept to the process of feature selection [33]. The results of this technique are more stable and accurate as the different ranking methods explore different important qualities of the data. A combination of these qualities in one ranking scheme will outperform each ranking method.

Research in the field of feature selection proposed some rank aggregation methods such as the sum, mean, median, highest rank or lowest rank aggregation and some are more difficult [33]. Moreover, research is on to give more weight to top ranking features or combining well-known aggregation methods in search of finding the best list which is an optimization problem.

V. METHODOLOGY

In this paper, the ensemble of multiple feature selection methods has been used for the selection of important features for the classifier. For the combination of ranks of individual feature selection methods the ensemble uses the fusion based rank aggregation method. For, the FS ensemble, five individual filter based methods of FS were chosen based on different measures of feature ranking. These were – Chi Square and Symmetrical Uncertainty methods of FS based on Dependency Measures; Information Gain and Gain Ratio FS methods based on Information Theory Measures; and Relief FS method based on Distance Measures.

In the first step, the five filter-based feature selection methods were used for ranking the features by their importance. The result of the first step is five ranked lists from the five individual feature selection methods.

The results of the first step are five ranked lists from the five individual feature selection methods.

The individual feature selection methods used are the Chi-Square, Information Gain, Gain Ratio, ReliefAttributeEval and SymmetricalUncertaintyAttributeEval from the WEKA software environment for knowledge analysis [34]. The study conducts experiments for ranking features using each feature selection method.

The second step proposes a new fusion based Rank Aggregation Algorithm for an ensemble of multiple feature selection techniques. The algorithm is described in Fig. 2. This method makes use of both rank score and ranks order of each feature in the ranked lists for rank aggregation. Fig. 2. describes the rank aggregation algorithm and its operation as follows:

First, the **k** individual feature selection methods rank the **n** features in order of their importance in descending order. Hence, each feature selection method generates a ranked list depicting the rank score (the value of a feature in the ranked list) and a sequence number **m** of each feature in the descending ordered ranked list.

In the second step, most of the rank aggregation methods use a combination of the ranked scores of multiple feature selection techniques in a certain way such as the sum, mean, median or taking the highest or lowest rank scores. But the rank score alone can't depict the importance of a feature in the ranked list. The order of the feature in the ranked list is also crucial for considering the importance of a feature. The proposed novel aggregation algorithm considers both rank scores and the rank orders of the features. This aggregation will give more weight to the features which not only have higher rank scores but also have higher rank orders in the ranked list. Equation (1) computes the rank order of a feature having sequence no. 'm' in a ranked list of 'n' features. Therefore, for a feature having sequence number 1, in a ranked list of 20 features, the aggregation finds the rank order of this feature as 20 by using (1).

$$\textit{rankorder} = n - m + 1 \qquad (1)$$

Algorithm: A Novel Rank Aggregation Algorithm for Ensemble of Multiple Feature Selection Techniques

```
Input:
Dataset m*n containing m instances and n features $f_j$ , where j = 1, 2, -----, n

Initialize Ensemble Rank List E = φ
Suppose $F_1$, $F_2$, -------, $F_k$ be the feature selection techniques used for the ensemble
For each $F_i$, i= 1, 2, ---, k
Calculate rank score of each feature and construct ranked lists $R_i$, i = 1, 2, ----, k
Sort each $R_i$ in descending order of rank scores
Give a sequence number m=1, 2, ------, n; to all the features in each $R_i$ starting from top.
ENDFOR

For each feature $f_j$, j = 1, 2, -----, n
For each sorted ranked list $R_i$, i = 1, 2, ------, k
For Sequence no. m = 1, 2,-------,n;
rankorder = n – m +1

Ensemble rankscore $E_j$ = $\sum_{i=1}^{k}(rankscore_{ji} * rankorder_{ji})$
E = E U $E_j$
ENDFOR
ENDFOR
ENDFOR
Sort the Ensemble rank list E using ensemble rank scores in descending order

Output: A sorted ensemble ranked list E containing features and their corresponding ensemble rank scores.
```

Fig. 2. A Novel Rank Aggregation Algorithm for Ensemble of FS

VI. Experiments

A. Data Used

The data set chosen for this experiment is the German dataset from UCI repository [35]. It is a credit dataset having 1000 loan applicants' records and 20 predictor variables. There is one class variable having two classes - Good and Bad. Most of the features are qualitative, and few are numerical.

B. Feature Selection

For ranking the features in order of their importance, the experiments considers the ensemble of multiple feature ranking techniques and five individual rank based feature selection methods. Those feature selection methods are used which perform better on qualitative data since the data is mostly qualitative. The novel rank aggregation algorithm uses the rank scores and rank orders of the individual rank based feature selection methods. The threshold values of 80%, 60%, 40% and 20% i.e. 16, 12, 8 and four features are used for selecting the features from the top of the sorted, ranked lists. In this way, only the highly ranked features identified as important and relevant by the individual and ensemble feature selection methods have been selected for building the classification models.

The performance of the classifiers is compared to find out the best threshold, best model and the best feature selection method which yielded the highest ROC value. The best threshold value indicates that the features selected using it are the most important ones which best described the dataset.

The best model is the one whose average classification performance across all the feature selection methods is the highest. The best feature selection method is the one which yields best average performance across all models built over the features selected by it.

C. Classifier

For testing the impact of the new rank aggregation algorithm on the accuracy of classifiers, the features selected from the aggregated ranked list are taken as inputs to the classifiers. The individual and ensemble based classifiers are used for model building and performance assessment. The individual classifiers used are the C4.5 and the MLP, while Bagging is used as the ensemble classifier.

The ensemble based bagging technique is used since the use of bootstrapping with replacement in bagging creates diversity within the data being used by the classifier hence impacting the performance of the classifier. The base classifiers used for bagging are the C4.5 and the MLP. These classifiers are considered acceptable to use at the cost of time and complexity of the system, since the focus of the study is the enhancement of classification accuracy of the credit risk evaluation models using the proposed rank aggregation method. For data sampling using bootstrapping, 20 iterations are used, as the classifier didn't show any increase in performance using more iterations. More iterations would rather have slowed down the classification process by increasing data samples and hence time.

D. Accuracy Assessment

The Area under the Receivers Operating Curve (ROC) popularly known as AUC, is used for accuracy assessment. The ROC Curve is a graph of True Positive Rate (TPR) versus False Positive Rate (FPR). The models are built using 70% training and 30% test partitions. A random sampling of 70% of training data is done from the dataset for training the classifier. The classifier uses the remaining 30% of data for testing the classifiers. The correctly classified instances were taken from the test data for classification. A ROC graph was plotted using TPR against the FPR for assessing the accuracy.

VII. Results and Discussion

Four classifier models - C4.5, MLP, C4.5 based Bagging and MLP based Bagging were built on the German credit dataset using a different number of features selected by each FS method. Each model was generated on four different threshold percentages (80%, 60%, 40% and 20%) i.e. (16, 12, 8 and 4) features selected from the sorted, ranked lists of the five individual feature selection methods and an ensemble of multiple FS methods. The performance of the classifiers has been observed using the ROC measure which is considered as a true measure of accuracy. For comparison of accuracy, each model has also been built using all the features. The average performance of six FS methods using four different thresholds across four different classifier models is depicted in Table I.

TABLE I. Average Performance of Rank based Feature Selection Methods

FS Ranking Methods	**80%**	**60%**	**40%**	**20%**	**Ranking Methods Average Performance**
Chi-square	.766	.754	.747	.740	**0.752**
Gain Ratio	.766	.771	.734	.733	**0.748**
Information Gain	.766	.754	.747	.740	**0.752**
Relief-F	.762	.766	.729	.734	**0.748**
Symmetrical Uncertainty	.766	.759	.741	.748	**0.753**
Ensemble	.769	.772	.741	.740	**0.756**

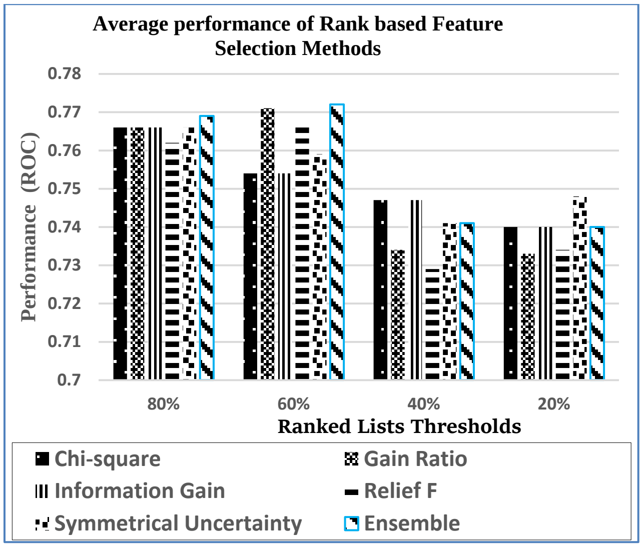

Fig. 3. Average performance of Rank based Feature Selection Methods across all models

The performance of each FS based ranking method is recorded for the four models for all thresholds. An average of performances of all the models on the features selected by the FS methods using a particular threshold is observed. Similarly, the average performances of all FS methods including Ensemble FS method have been calculated across all models using different thresholds. The comparative performance of these FS methods is depicted in Fig. 3.

The graph of Table I. summarizes that the performance of the ensemble of multiple FS methods is higher than all individual FS methods for the thresholds of 80% and 60%, while the performance of FS methods Chi-square and Information gain is higher than others for the 40% threshold. The symmetrical uncertainty method outperforms the others for 20% threshold. It is clearly observed from the graph that, for 40% and 20% thresholds (i.e. small no. of features), the performance of all the FS methods is substantially lower than that for 80% and 60% thresholds.

By looking at the graph, it can also be inferred that the performance of the Ensemble of FS methods is the highest for the 60% threshold followed by 80% threshold. Also, the performance of all FS methods including the ensemble of multiple FS techniques started declining drastically after the 60% threshold.

The individual model performance based on different thresholds using the ensemble of FS method is depicted in Table II. It can be seen across all thresholds, the performance of bagging models based on C4.5 and MLP as the base classifiers is much better than the individual C4.5 and MLP models. Moreover, the average model performance for the bagging model based on MLP as the base classifier is the best. It can also be observed that the average performance of all the models is the best for 60% threshold. The graph depicting the average performance of the individual models in Fig. 4, shows that the performance of bagging based on MLP classifier is the highest followed by bagging based on C4.5 classifier at 60% threshold. While the individual models C4.5 and MLP performed best at 80% threshold, the individual C4.5 model performed the worst of all for all thresholds.

TABLE II. INDIVIDUAL MODEL PERFORMANCE ON DIFFERENT THRESHOLDS USING THE ENSEMBLE BASED ON FS METHOD

Classification Models	**100%**	**80%**	**60%**	**40%**	**20%**	**Avg. Model Performance**
C4.5	.730	.730	.727	.707	.726	**0.728**
Bagging (C4.5)	.773	.773	.787	.774	.752	**0.775**
MLP	.725	.770	.765	.711	.716	**0.743**
Bagging (MLP)	.788	.801	.809	.773	.767	**0.794**
Avg. performance	**.754**	**0.769**	**0.772**	**0.741**	**0.740**	

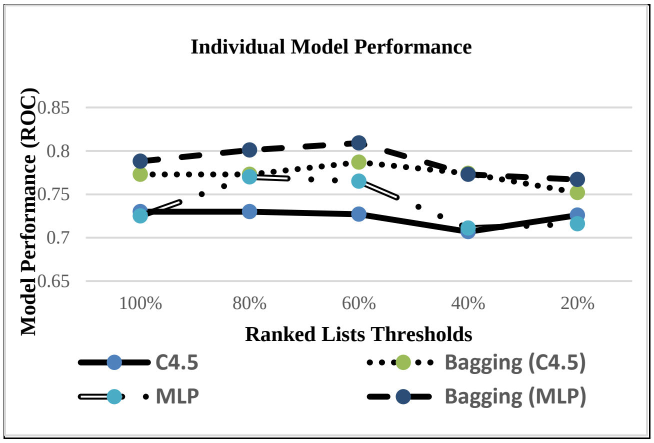

Fig. 4. Average performance of Individual Models using Ensemble based FS method

VIII. CONCLUSION

In credit risk evaluation the accuracy of a classifier is very crucial. Even a small increase in model accuracy may result in huge profit for the bank. For accuracy enhancement, this study uses the ensemble of multiple feature selection techniques for ranking and selecting the important features. A novel rank aggregation algorithm has been proposed using the rank scores and rank orders of the individual rank based feature selection methods. The ensemble of FS technique uses the novel rank aggregation algorithm for ranking the features in order of their importance and relevance. The ranked lists of 5 FS methods and 1 Ensemble based FS method were used to select the top 16, 12, 8 and 4 features. The Ensemble based FS method attained the best performance for the threshold of 12 top features with an average ROC value of .772 followed by the threshold of 16 giving an average ROC value of .769 while the average ROC value for the dataset without FS is .754. Moreover, these ROC values for the ensemble method are higher than all other individual FS methods used. On comparing the ROC values it is inferred that using the Ensemble based FS method, the average performance of the four models increased by a ROC of .018 using the 60% threshold.

The results also concluded that the bagging based models outperformed the individual models using the ensemble of FS methods for all thresholds. The performance of Bagging using MLP as the base classifier is the highest with a ROC of .809 followed by Bagging using C4.5 as the base classifier with a ROC of .787 at 60% threshold, while the individual MLP and C4.5 models performed with an ROC value of .765 and .727 respectively for the same threshold. By using Bagging, there is an average performance enhancement of .044 and .060 respectively for individual MLP and C4.5 models across all thresholds. One more inference drawn from the results is that the average performance of Bagging model with MLP as the base classifier is the best across all thresholds with a ROC of .794 followed by .775 for the Bagging model with C4.5 as the base classifier.

Therefore, the study concluded that, using an ensemble of multiple feature selection techniques with the novel rank aggregation algorithm proposed in the study, a significant enhancement in the performance of credit risk evaluation models is observed. The accuracy of the models is enhanced with the selection of top 80% and 60% features from the ranked list of the ensemble. Although, the accuracy of the models declined with the selection of top 40% and 20% features. It may be attributed to the rejection of many relevant features required for building the accurate model.

By using the ensemble of multiple feature selection techniques, the bagging based models outperformed the individual models for all thresholds but most significantly for the 60% threshold.

This increase in performance is more significant from the fact that the number of features reduces by 40% for building the highest performing models which indicates a phenomenal reduction in the instance size and hence the overall data size. The reduction of irrelevant features simplifies the model building task and hence the time and space complexity of running the models. A simpler and faster model would be helpful for the bankers in a quick and precise overall assessment of the risk involved in granting the loan to a customer. Moreover, the irrelevant features with very low ranks are identified which do not contribute to the model building process. These features can be ignored by the banks in the loan application forms, making them simpler and faster for the applicants to fill in and for the banks to get them verified quickly.

Future studies can focus on testing the novel rank aggregation algorithm on other high dimensional credit datasets collected from the real world. The algorithm may prove to be more useful for such data with a large number of attributes by selecting only a small number of relevant attributes contributing to the accuracy and simplicity of the model. Even a small enhancement in the accuracy of credit risk evaluation models is very beneficial as the financial risk associated with the credit defaulters get assessed accurately on time.

References

[1] H. Liu, H. Motoda, R. Setiono, Z. Zhao, “Feature Selection: An Ever Evolving Frontier in Data Mining”, JMLR: Workshop and Conference Proceedings, vol. 4, Publisher: Citeseer, pp. 4-13, 2010.

[2] L. Yu, H. Liu, “Feature Selection for High-Dimensional Data: A Fast Correlation- Based Filter Solution”, Proceedings of the Twentieth International Conference on Machine Leaning, ICML-03, Washington, D.C., August, 2003, pages 856-863.

[3] M. Hall and G. Holmes, “Benchmarking attribute selection techniques for discrete class data mining,” IEEE Transactions on Knowledge and Data Engineering, vol. 15(3), November/December 2003.

[4] Dahiya, S., Handa, S.S., Singh, N.P. (2015). Credit Evaluation using Ensemble of various classifiers on reduced feature set”, Industrija, Vol.43, No.4.pp 163-174.

[5] Y. Liu, M. Schumann, “Data mining feature selection for credit scoring models,” Journal of Operations Research Society, vol. 56(9), pp. 1099–1108, 2005.

[6] T. Howley, M. G. Madden, M. L. O'Connell, and A. G. Ryder, “The effect of principal component analysis on machine learning accuracy with high dimensional spectral data,” Knowledge Based Systems, vol. 19 (5), pp. 363-370, 2006.

[7] Hand, D. J., Henley, W. E. (1997).Statistical Classification Methods in Consumer Credit Scoring: A Review. Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 160, No. 3, pp. 523-541.

[8] M. A. Hall, “Feature Selection for Discrete and Numeric Class Machine Learning,” In Proceedings of the 17th international conference on Machine Learning (ICML-2000).

[9] W. Duch, “Filter methods,” Feature extraction, foundations and Applications, pp. 89–117. Studies in fuzziness and soft computing, Springer (2006).

[10] T. W. S. Chow, D. Huang, “Using Mutual information for Feature selection with bioinformatics applications,” Neural Networks Applications in Information Technology and Web Engineering, 2005, Borneo Publishing Co.

[11] V. Sindhwani, S. Rakshit, D. Deodhare, D. Erdogmus, P J, Niyogi, “Feature selection in MLPs and SVMs based on maximum output information”. IEEE Trans Neural Netw ork, vol.15(4), pp.937–948, 2004.

[12] MD Plumbley, E Oja, “A nonnegative PCA" algorithm for independent component analysis,” IEEE Transactions on Neural Networks vol. 15 (1), pp. 66-76, 2004.

[13] K.Z. Mao, “Orthogonal forward selection and backward elimination algorithms for feature subset selection,” IEEE Trans Syst Man Cybern B Cybern, vol. 34(1), pp.629–634, 2004.

[14] R. Caruana, De Sa V , “Benefitting from the variables that variable selection discards," Journal of Machine Learning, Res 3, pp. 1245–1264, 2003.

[15] M. Dash and H. Liu, “Consistency-based search in feature selection,” Artificial Intelligence, vol. 151 (1-2), pp. 155-176, 2003.

[16] K. Grabczewski, N. Jankowski, “Mining for complex models comprising feature selection and classification,” Feature extraction, foundations and Applications, pp. 473–489, Studies in fuzziness and soft computing, Springer, 2006.

[17] Guyon, A. Elisseef, “An introduction to variable and feature selection,” Journal of Machine Learning Research pp. 1157–1182, (2003).

[18] Rodriguez-Lujan, R. Huerta, C. Elkan and C.S. Cruz, “Quadratic programming feature selection”. Journal of Machine Learning Research, vol. 11, pp. 1491–1516, 2010.

[19] Ienco, R.G. Pensa, R. Meo, “Context-based Distance Learning for Categorical Data clustering”, IDA 2009, LNCS 5772, Springer, Berlin, pp. 83 – 94, 2009.

[20] Kumar, and K. Kumar, “A novel evaluation function for feature selection based upon information theory”, In Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 395–399, 2011.

[21] Al-Ani, A. and M. Deriche, “An optimal feature selection technique using the concept of mutual information,” In Proceedings of the Sixth International Symposium on Signal Processing and its Applications, pp. 477–480, 2001.

[22] Liu and L. Yu, “Towards integrating feature selection algorithms for classification and clustering,” IEEE Transactions on Knowledge and Data Engineering, vol.17 (4), pp. 491-502, 2005.

[23] K. Kira, and L. Rendell, “A Practical Approach to Feature Selection,” Proceedings of the Ninth International Workshop on Machine Learning, pp.249-256, 1992.

[24] Kononenko, “Estimating Attributes: Analysis and Extensions of RELIEF,” In: Proceedings of the 7th European Conference on Machine Learning (ECML-94), pp. 171–182, New York: Springer, 1994.

[25] W. Bouaguel, “On Feature Selection Methods for Credit Scoring” Doctoral Thesis, ISG University of Tunis, 2015.

[26] T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning,” Springer series in statistics, Springer New York Inc. 2001.

[27] O. Wu, H. Zuo, W. Zhu, M. Hu, J. Gao, and H. Wang, “Rank aggregation based text feature selection,” In: Proceedings of the Web Intelligence, pp. 165-172, 2009.

[28] J. Kittler, “Combining classifiers: A theoretical framework,” Pattern Analysis & Applications, vol. 1 (1), pp. 18-27, 1998.

[29] L.I. Kuncheva, “Combining Pattern Classifiers: Methods and Algorithms,”. John Wiley & Sons, (2004).

[30] T.G. Dietterich, “Ensemble methods in machine learning,” In Proceedings of the First International Workshop on Multiple Classifier Systems, London, UK, pp. 1-15. Springer-Verlag, 2000.

[31] Y. Saeys, T. Abeel, Y. V. Peer, “Robust feature selection using ensemble feature selection techniques”, ECML PKDD 2008, Part II, LNAI 5212, pp. 313–325.

[32] S.H. Vege, “Ensemble of Feature Selection Techniques for High Dimensional Data”. Master’s Thesis, Western Kentucky University, 2012.

[33] Dittman, D. J., T. M. Khoshgoftaar, R. Wald, and A. Napolitano (2013). Classification performance of rank aggregation techniques for ensemble gene selection. In C. Boonthum-Denecke and G. M. Youngblood (Eds.), Proceedings of the International Conference of the Florida Arti_cial Intelligence Research Society (FLAIRS). AAAI Press.

[34] http://www.cs.waikato.ac.nz/, accessed on 6 June’2016.

[35] https://archive.ics.uci.edu/ml/datasets/ accessed on 3, June’2016.