

# **Weighted Random Forests for Evaluating Financial Credit Risk**

Tzu-Tsung Wong<sup>\*</sup>, Shang-Jung Yeh

Institute of Information Management, National Cheng Kung University, Tainan, Taiwan

Received 13 May 2019; received in revised form 12 June 2019; accepted 06 July 2019

## **Abstract**

Credit evaluation of customers is a critical issue in financial organizations. Classification algorithms have been proposed for credit evaluation in recent years, and the class distribution in the financial data for those studies are not skewed. However, only a small proportion of customers will be the cases for bad credit. Financial records should be considered as an imbalanced data set for analyzing credit risk. Ensemble algorithms that make predictions by group decisions generally have relatively high accuracy than the ones inducing only one model from data. This study introduces a mechanism based on the weighted random forest to improve the prediction accuracy on the records with bad credit. This mechanism is tested on two financial data sets to demonstrate that it can achieve relatively high performance in evaluating credit risk and that the number of decision trees in a forest is not helpful. Critical attributes are also identified to provide practical meanings for credit risk analysis.

**Keywords:** credit risk analysis, decision tree, imbalanced data, random forest

## **1. Introduction**

Credit risk evaluation is getting important since the global financial crisis. The credit evaluation of customers is critical to financial organizations. It will be extremely helpful if the data of customers can be used to build models for supporting decisions. Support vector machine and neural network are two popular classification algorithms for credit evaluation [1-11]. Feature selection is a possible way to improve the performance of financial risk analysis [12-14]. Ensemble algorithms induce several prediction models to make group decisions, and hence they generally have a better performance than the ones that learn only one model from data. For example, the random forest is an ensemble algorithm proposed by Breiman [15], that grows multiple decision trees to take a majority vote for class prediction. Studies have demonstrated that a random forest is a proper tool for credit risk analysis [16-20].

Only a small proportion of the customers in a financial organization will have troubles in paying their loans. This implies that the data for credit risk analysis will be imbalanced, while none of the previous studies considered this to process data and to interpret the learning results. There are generally two class values in an imbalanced data set, and the one with fewer instances is called positive, and another is called negative. When over 90% of the instances in a data set are negative, the prediction accuracy of a classification model will be larger than 0.90 if it predicts a negative class value for all instances. This indicates that pursuing higher prediction accuracy is not appropriate for processing imbalanced data sets, while most classification algorithms are designed to have high accuracy.

The learning procedure of traditional classification algorithms is generally designed to achieve a high probability of correct prediction. There are two popular ways in processing imbalanced data sets: change the class distribution in data sets, or modify the learning procedure of a classification algorithm [21]. Resampling is a popular approach to adjust the class

---

<sup>\*</sup> Corresponding author. E-mail address: tzutsung@mail.ncku.edu.tw

Tel.: +886-6-2757575; Fax: +886-6-2362162

distribution in a data set [22-23]. For example, SMOTE is an over-sampling method for generating synthetic instances to balance positive and negative ones in a data set [24]. It is also possible to apply under-sampling methods for removing negative instances [25]. However, the model induced from the revised data set can be overfitting and difficult to interpret [26-27]. Barandela et al. [28] consider weights to calculate distances for k-nearest neighbors in processing imbalanced data. An ensemble-based classifier that collects the predictions from multiple classification models to make group decisions is also an effective way for improving the performance on imbalanced data sets [29-30]. This study will employ the weighted random forest proposed by Chen et al. [31] to analyze customer credit and propose a way to rank attribute for credit evaluation. The experimental results will demonstrate that the weighted random forest is more effective and suitable for discovering important attributes in credit risk analysis.

The purpose of this paper is to consider the data for credit evaluation as an imbalanced set. Ensemble method weighted random forest will be employed to analyze this kind of data sets not only for improving the prediction accuracy on positive instances but also for discovering critical attributes in evaluating the credit of customers. The data set after discretization will also be used to test ordinary random forest for demonstrating that the weighted random forest is a better tool for imbalanced data analysis. The implications of the critical attributes will be discussed. The whole research procedure is shown in Fig. 1.

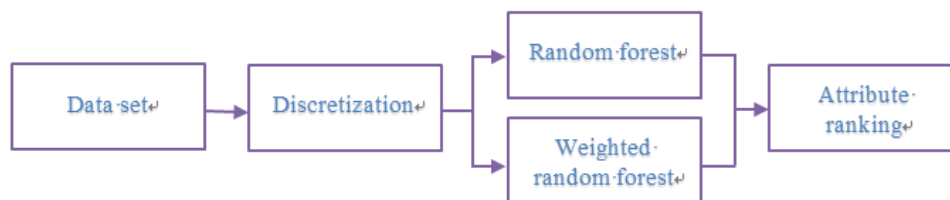


Fig. 1 The research procedure

## 2. Weighted Random Forest

A forest is composed of a set of decision trees, and every decision tree has two kinds of nodes: internal nodes and leaf nodes. Every internal node is marked by an attribute to indicate the paths for branching, and there is a class value inside every leaf node for prediction. Training data will be used to calculate impurity measure, such as information value and Gini index, to determine the branching attributes in internal nodes.

The random forest proposed by Breiman [15] iteratively performs two steps to find a decision tree until the number of decision trees reaches a pre-specified threshold. Let a data set have  $n$  instances composed of  $m$  attributes. The first step is to randomly select  $n$  instances with replacement from the data set. The instances that are not chosen for training are called out-of-bag data. They will be used to validate the decision tree built from the training data. When  $n$  is large, about 36.8% of the instances in the data set will play the role of validation. After the training set is determined, the second step is to randomly select  $\sqrt{m}$  attributes for growing a decision tree; i.e. only the chosen attributes can be considered in growing a decision tree. Since the instances and the attributes for growing a decision tree are both random, the decision trees built in this way will be diverse for making group decisions in classifying new instances. Ordinary decision tree learning has a pruning step to avoid the occurrence of overfitting. Group decisions are not sensitive to noise, and hence every decision tree in a forest need not be pruned.

The Gini index of a candidate attributes  $A$  is calculated as  $Gini(A) = 1 - \sum_i p_i^2$ , where  $p_i$  is the proportion of the instances with a class value  $i$ . The one with the largest Gini index will be the branching attribute. This implies that the misclassification cost of all instances are the same regardless of their class values. A decision tree grown by this way is unlikely to have a high prediction accuracy on the positive instances in an imbalanced data set. This is the main reason why the random forest is not a proper tool for classifying imbalanced data sets. The weighted random forest proposed by Chen et al. [31] considers the

misclassification costs of various class values to determine branching attributes. Let  $c_+$  and  $c_-$  be the misclassification cost of positive and negative instances, respectively. Then the weight of positive class value is  $w = c_+ / (c_+ + c_-)$ , and the weighted Gini index of attribute  $A$  is calculated as  $Gini(A) = 1 - wp_+^2 - (1-w)p_-^2$ .

Weight  $w$  is also considered in classifying new instances. Let  $y_+$  and  $y_-$  be the respective numbers of positive and negative training instances reaching the leaf node for classing a new instance. Then the weighted votes for positive and negative class values from this decision tree are  $wy_+$  and  $(1-w)y_-$ , respectively. These weighted votes are aggregated over decision trees to determine the predicted class value of the new instance. Weighted random forest is, therefore, more suitable for processing imbalanced data sets.

### 3. Performance Evaluation

True positives and false negatives represent the numbers of correct and wrong predictions on positive instances, respectively, and similarly for true negatives and false positives on negative instances. Let  $TP$ ,  $FP$ ,  $TN$ , and  $FN$  be true positives, false positives, true negatives, and false negatives, respectively in a confusion matrix as shown in Table 1. Then accuracy is calculated as  $(TP+TN)/N$ , where  $N = TP+FP+TN+FN$  is the number of testing instances. Since most of the instances are negative,  $FP+TN$  is far larger than  $TP+FN$ . Accuracy will be large when a negative value is assigned to all instances. That is the reason why accuracy is an improper measure for performance evaluation when data sets are imbalanced. *F-measure* is a harmonic mean of  $recall = TP/(TP+FN)$  and  $precision = TP/(TP+FP)$ , and it can be expressed as Eq. (1):

$$F - measure = 2 \times recall \times precision / (recall + precision) \tag{1}$$

Recall shows the percentage of the positive instances that are correctly predicted, and precision indicates the accuracy of positive predictions. When the values of recall and precision are largely different, F-measure will be close to the smaller one. An algorithm that has a high value on only one of recall and precision cannot achieve a large F-measure.

Table 1 The confusion matrix for an imbalanced data set

		Predicted class value	
		Yes	No
Actual class value	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

AUC represents the area under the ROC curve determined by the value of recall and  $FP-rate = FP/(FP+TN)$ . The ROC curve represents the achievement of positive instances with respect to the number of misclassifications on negative instances. The maximal possible value of AUC is one when an algorithm makes correct predictions on all positive and negative instances. Both F-measure and AUC can provide information about the prediction accuracy on positive instances, and hence they are popular measures in analyzing imbalanced data sets.

Knowing the importance of attributes in credit evaluation is extremely helpful for decision makers. It is easy to interpret the result implied by a single decision tree. However, every prediction of weighted random forest is made by the weighted majority vote of its induced decision trees. To rank attributes about their importance becomes relatively complex. This study proposes the following method composed of three steps to evaluate the critical level of each attribute.

- Step 1. Calculate the precision of each decision tree from its out-of-bag data, denoted as  $q_1$ .
- Step 2. Consider attributing  $A$  as noise to recalculate the precision of each decision tree from its out-of-bag data, denoted as  $q_2$ .  
the way of changing attribute  $A$  into a noisy one is to randomly permute the values of this attribute.
- Step 3. Calculate difference  $q_1 - q_2$  that represents the benefits of introducing attribute  $A$  for this decision tree, and average the differences of all decision trees to be the critical level of this attribute.

A larger critical level for an attribute means that this attribute is more useful in identifying positive instances. Attributes are thus sorted in a descending order based on their critical levels to filter important ones for credit evaluation.

#### 4. Experimental Study

Two data sets for credit risk evaluation are download from the internet to test whether weighted random forest can have a better performance on imbalanced data sets than random forest. The basic characteristics of the two data sets are given in Table 2. Note that the instances with missing values have been removed from the data sets. Calculating the Gini index is relatively straightforward for discrete attributes. The entropy-based discretization proposed by Fayyad and Irani [32] is employed to transform continuous attributes into discrete ones. This discretization method recursively bipartitions an interval into two subintervals based on information gain. The maximal description length principle is employed to derive a threshold for stopping the bipartition. The evaluation method is  $k$ -fold cross-validation with  $k = 5$ . This kind of cross-validation method first divides a data set into  $k$  folds. Every fold will be in turn used to test the learning results induced from the other  $k-1$  folds. The values of  $TP$ ,  $FN$ ,  $FP$ , and  $TN$  are aggregated over the testing instances for computing F-measure and AUC in each iteration of  $k$ -fold cross-validation. The mean values of the metrics over the  $k$  folds will be compared to investigate the effectiveness of weighted random forest. The number of trees in both classification algorithms will be ranged from 100 to 500 with step size 100 to investigate the impact of forest size.

Table 2 Basic characteristics of the two data sets for credit evaluation

Data set	Credit card scoring	Give me some credit
Source	<a href="https://github.com/gastonstat/CreditScoring">https://github.com/gastonstat/CreditScoring</a>	<a href="https://www.kaggle.com/c/GiveMeSomeCredit">https://www.kaggle.com/c/GiveMeSomeCredit</a>
Number of positives	1,217	8,357
Number of negative	3,160	111,912
Class	Credit status	Serious past-due delinquency
Attributes	Job seniority	Revolving utilization of unsecured lines
	Type of home ownership	Age
	Time of requested loan	Number of times 30-59 days past due
	Client's age	Debt ratio
	Marital status	Monthly income
	Existence of records	Number of open credit lines and loans
	Type of job	Number of times 90 days late
	Amount of expenses	Number real estate loans or lines
	Amount of income	Number of times 60-89 days past due
	Amount of assets	Number of dependents
	Amount of debt	-
	Amount requested of loan	-
Price of good	-	

##### 4.1. Data set "credit card scoring"

This data set has 4,454 instances, and every instance has 13 attributes. The proportion of the instances with bad credit is less than 30%. The prediction accuracy is higher than 0.7 when all instances are classified as negatives. It is not a typical imbalanced data set, while can still be used to test the applicability of weighted random forest. The respective misclassification costs for positive and negative class values are 3160/1217 and 1 based on their proportions in the data set. The testing results of random forest and weighted random forest are summarized in Table 3. Note that all values in Table 3 are the averages over five folds. Consider the case for the random forest with 100 decision trees. The five pairs of *recall* and *precision* for folds 1 through 5 are (0.3770, 0.6835), (0.4052, 0.6225), (0.4783, 0.6395), (0.4570, 0.7222), and (0.4575, 0.6848). The F-measures calculated from the five pairs are 0.4859, 0.4909, 0.5473, 0.5598, and 0.5485, and their average is 0.5265 as shown in the second row of Table 3(a). In this case, the mean recall and mean precision are 0.4350 and 0.6705, respectively, and the harmonic mean of these two values is 0.5277 that is not equal to 0.5265.

For the sake of clarity, the values of *recall*, *precision*, and *F-measure* for algorithms Random Forest (RF) and Weighted Random Forest (WRF) are depicted in Fig. 2. Regardless of the number of trees, precision is larger than recall. This suggests that increasing the value of *recall* will have a larger chance to improve F-measure. The precisions for random forest and weighted random forest are close for any given forest size. Since weighted random forest has a larger recall, it also has a larger F-measure. This data set is not highly skew, and hence the improvement on F-measure is not large. Though every AUC in Table 3(b) is larger than its corresponding value in Table 3(a), their difference is less than 0.01 in all cases. Weighted random forest almost has no improvement on AUC for this data set. Table 3 and Fig. 2 also show that increasing the number of trees is not an effective way to improve the performance of both classification algorithms.

Table 3 The experimental results of (a) random forest and (b) weighted random forest for data set Credit card scoring

(a) No. of trees	Recall	Precision	F-measure	AUC
100	0.4350	0.6705	0.5265	0.8291
200	0.4412	0.6772	0.5335	0.8316
300	0.4415	0.6810	0.5344	0.8331
400	0.4443	0.6819	0.5370	0.8336
500	0.4514	0.6814	0.5416	0.8336
(b) No. of trees	Recall	Precision	F-measure	AUC
100	0.4634	0.6594	0.5430	0.8322
200	0.4753	0.6677	0.5534	0.8338
300	0.4711	0.6642	0.5499	0.8337
400	0.4852	0.6749	0.5630	0.8339
500	0.4812	0.6661	0.5572	0.8343

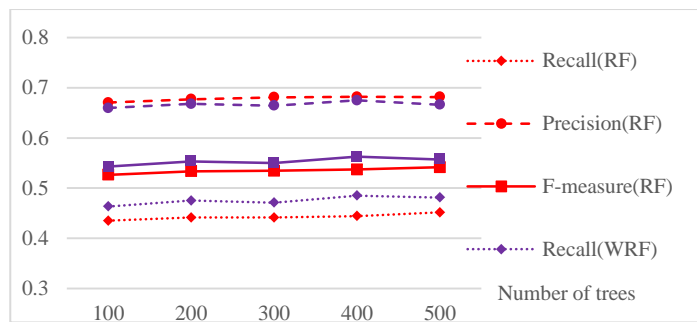


Fig. 2 The line charts of recall, precision, and F-measure for data set credit card scoring

#### 4.2. Data set “give me some credit”

There are 120269 instances in this data set, and every instance has 8 attributes. The proportion of the instances with bad credit is approximately 7%. The prediction accuracy is higher than 0.9 when all instances are classified as negatives, and hence it is an imbalanced data set. The respective misclassification costs for positive and negative class values are 111912/8357 and 1 based on their proportions in the data set. The testing results of random forest and weighted random forest are summarized in Table 4 and depicted in Fig. 3. Similar to the previous data set, every precision is larger than its corresponding recall. Every recall is less than 0.2, which implies less than 20% of positive instances can be predicted correctly. This suggests that F-measure can be greatly improved by increasing recall. This is the reason why weighted random forest that has a slightly smaller precision can still achieve a higher F-measure than random forest.

Similar to the previous data set, weighted random forest achieves a slightly larger AUC than a random forest, while their difference is even smaller. This suggests that AUC may not be a proper measure for evaluating the performance of classification algorithms on financial credit data. Again, the performance of both algorithms is not sensitive to the increasing forest size. Since computational cost is higher for larger forest size, increasing the number of decision trees should not be necessary. The experimental results of these two data sets suggest that increasing the number of decision trees is not beneficial for improving the performance of the two classification algorithms.

Table 4 The experimental results of (a) random forest and (b) weighted random forest for data set “give me some credit”

(a) No. of trees	Recall	Precision	F-measure	AUC
100	0.1407	0.5604	0.2247	0.8250
200	0.1422	0.5649	0.2269	0.8307
300	0.1405	0.5616	0.2245	0.8324
400	0.1407	0.5605	0.2247	0.8335
500	0.1408	0.5670	0.2252	0.8327
(b) No. of trees	Recall	Precision	F-measure	AUC
100	0.1796	0.5507	0.2705	0.8289
200	0.1769	0.5464	0.2670	0.8335
300	0.1828	0.5606	0.2753	0.8353
400	0.1819	0.5536	0.2736	0.8377
500	0.1800	0.5544	0.2715	0.8373

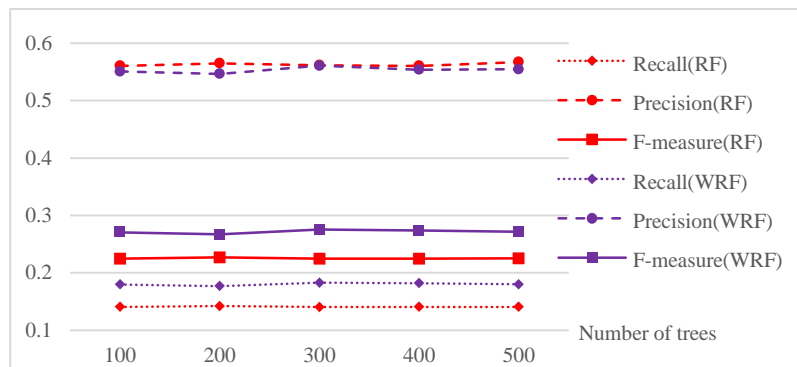


Fig.3 The line charts of recall, precision, and F-measure for data set “give me some credit”

#### 4.3. Attribute ranks

By using the critical level introduced in Section 3 to rank attributes, the results are summarized in Table 5. In data set “credit card scoring,” the top three important attributes filtered by weighted random forest are “amount of income,” “job seniority,” and “price of good.” All of these three attributes can imply whether customers can consistently pay their bills. Marital status is the most useless attribute in this case. It is surprising that the rank of attribute “amount of debt” is 12. This implies that the amount of debts is not very helpful in evaluating the financial status of a customer.

Table 5 Attribute ranks for both data sets

Rank	Credit card scoring	Give me some credit
1	Amount of income	Revolving utilization of unsecured lines
2	Job seniority	Debt ratio
3	Price of good	Monthly income
4	Amount requested of loan	Age
5	Client’s age	Number of times 90 days late
6	Amount of assets	Number of open credit lines and loans
7	Amount of expenses	Number of times 30-59 days past due
8	Existence of records	Number of time 60-89 days past due
9	Type of job	Number of dependents
10	Time of requested loan	Number real estate loans or lines
11	Type of home ownership	-
12	Amount of debt	-
13	Marital status	-

The top three important attributes filtered by weighted random forest for data set “Given me some credit” is “revolving utilization of unsecured lines,” “debt ratio,” and “monthly income.” It is surprising that the three attributes (number of times 30-59, 60-89, and 90 days past due) relevant to whether a customer has a late payment are not critical to the evaluation of bad credit. This is not reasonable and could be the major reason why most positive instances cannot be corrected classified, as indicated by the small recalls given in Table 4.

Attribute “amount of income” has a high rank in both data sets, and hence it is absolutely a necessary attribute in evaluating customer credit. The age of a customer should also be considered in approving a loan. The job of a customer is relatively important to the information about his/her family. The results of these two data sets are not consistent with the attributes relevant to debt. This can be an interesting question to explore further. Since data sets for credit risk analysis may be collected from various resources, feature selection should be helpful for improving the performance of weighted random forest.

## 5. Conclusions

Most data sets for credit evaluation are imbalanced, and hence random forest may not be a proper tool for this purpose. In this paper, the weighted random forest is introduced to analyze this kind of imbalanced data sets, and a way for filtering critical attributes from the learning results of weighted random forest is proposed. The testing results of two credit data sets show that weighted random forest outperforms random forest in terms of F-measure because the misclassification costs considered in the weighted random forest can increase the probability of correctly identifying positive instances. Area under the ROC curve may not be proper measure in determining superior classification algorithms for processing imbalanced data sets. More decision trees are generally not helpful for performance improvement, while more computational effort is needed. The critical attributes filtered from the two data sets also provide useful implications for understanding the credit of customers. Those implications can provide practical suggestions in reviewing the applications of loans.

The experimental results on two data sets for credit evaluation indicate that identifying positive instances is generally more difficult than having a high prediction precision on positive instances. This suggests that a classification algorithm will be a better tool for credit risk analysis if its learning procedure can be adjusted to improve the capability in identifying positive instances. This can shed some lights on designing the learning mechanism of classification algorithms for processing imbalanced data sets.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgement

This research was supported by the Center for Innovative Fintech Business Models of National Cheng Kung University and the Ministry of Science and Technology in Taiwan under Grant No. 107-2410-H-006-045-MY3.

## References

- [1] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, “Benchmarking state-of-the-art classification algorithms for credit scoring,” *Journal of the Operational Research Society*, vol. 54, no. 6, pp. 627-635, June 2003.
- [2] C. L. Huang, M. C. Chen, and C. J. Wang, “Credit scoring with a data mining approach based on support vector machines,” *Expert Systems with Applications*, vol. 33, no. 4, pp. 847-856, November 2007.
- [3] P. Danenas and G. Garsva, “Selection of support vector machines based classifiers for credit risk domain,” *Expert Systems with Applications*, vol. 42, no. 6, pp. 3194-3204, April 2015.
- [4] N. Mohammadi and M. Zangeneh, “Customer credit risk assessment using artificial neural networks,” *International Journal of Information Technology and Computer Science*, vol. 8, pp. 58-66, March 2016.
- [5] J. Yao and C. Lian, “A new ensemble model based support vector machine for credit assessing,” *International Journal of Grid and Distributed Computing*, vol. 9, no. 6, pp. 159-168, 2016.
- [6] A. AghaeiRad, N. Chen, and B. Ribeiro, “Improve credit scoring using transfer of learned knowledge from self-organizing map,” *Neural Computing and Applications*, vol. 28, no. 6, pp. 1329-1342, June 2017.

- [7] Q. Zhang, J. Wang, A. Lu, S. Wang, and J. Ma, "An improved SMO algorithm for financial credit risk assessment evidence from China's banking," *Neurocomputing*, vol. 272, pp. 314-325, January 2018.
- [8] D. K. Gupta and S. Goyal, "Credit risk prediction using artificial neural network algorithm," *International Journal of Modern Education and Computer Science*, vol. 10, no. 5, pp. 9-16, May 2018.
- [9] S. Ben Jabeur, A. Sadaoui, A. Sghaier, and R. Aloui, "Machine learning models and cost-sensitive decision trees for bond rating prediction," *Journal of the Operational Research Society*, DOI: 10.1080/01605682.2019.1581405, April 2019.
- [10] L. Munkhdalai, T. Munkhdalai, O. E. Namsrai, J. Y. Lee, and K. H. Ryu, "An empirical comparison of machine-learning methods on bank client credit assessments," *Sustainability*, vol. 11, pp. 1-23, January 2019.
- [11] M. Z. Abedin, C. Guotai, Fahmida-E-Moula, A. S. M. S. Azad, and M. S. U. Khan, "Topological applications of multilayer perceptrons and support vector machines in financial decision support systems," *International Journal of Finance and Economics*, vol. 24, pp.474-507, January 2019.
- [12] D. Liang, C. F. Tsai, and H. T. Wu, "The effect of feature selection on financial distress prediction," *Knowledge-Based Systems*, vol. 73, pp. 289-297, January 2015.
- [13] S. Dahiya, S. S. Handa, and N. P. Singh, "A rank aggregation algorithm for ensemble of multiple feature selection techniques in credit risk evaluation," *International Journal of Advanced Research in Artificial Intelligence*, vol. 5, pp. 1-8, October 2016.
- [14] F. N. Koutanaei, H. Sajedi, and M. Khanbabaeei, "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring," *Journal of Retailing and Consumer Services*, vol. 27, pp. 11-23, November 2015.
- [15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, October 2001.
- [16] N. Ghatasheh, "Business analytics using random forest trees for credit risk prediction: a comparison study," *International Journal of Advanced Science and Technology*, vol. 72, pp. 19-30, 2014.
- [17] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621-4631, June 2015.
- [18] V. Garcia, A. I. Marques, and J. S. Sanchez, "Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction," *Information Fusion*, vol. 47, pp. 88-101, May 2019.
- [19] A. Coser, M. M. Maer-matei, and C. Albu, "Predictive models for loan default risk assessment," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 53, pp. 149-165, June 2019.
- [20] M. Oskarsdottir, C. Bravo, C. Sarraute, J. Vanthienen, and B. Baesens, "The value of big data for credit scoring: enhancing financial inclusion using mobile phone data and social network analytics," *Applied Soft Computing*, vol. 74, pp. 26-39, January 2019.
- [21] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp. 113-141, November 2013.
- [22] G. E. Batista, R. C. Prati, and M. C. Monard "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations*, vol. 6, pp. 20-29, 2004.
- [23] A. Estabrooks, T. Jo, and N. Japkowicz, "A multiple resampling method for learning from imbalanced data sets," *Computational Intelligence*, vol. 20, pp. 18-36, 2004.
- [24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [25] M. A. Tahir, J. Kittler, K. Mikolajczyk, and F. Yan, "A multiple expert approach to the class imbalance problem using inverse random under sampling," *Proceedings of the Eighth International Workshop on Multiple Classifier Systems*, pp. 82-91, 2009.
- [26] Y. Li, H. Guo, X. Liu, Y. Li, and J. Li, "Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data," *Knowledge-Based Systems*, vol. 94, pp. 88-104, February 2016.
- [27] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, "A novel ensemble method for classifying imbalanced data," *Pattern Recognition*, vol. 48, no. 5, pp.1623-1637, May 2015.
- [28] R. Barandela, J. S. Sánchez, V. García, and E. Rangel, "Strategies for learning in class imbalance problems," *Pattern Recognition*, vol. 36, pp. 849-851, 2003.
- [29] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, pp. 221-232, 2016.



- [30] L. I. Kuncheva and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles," *Knowledge and Information Systems*, vol. 38, no. 2, pp. 259-275, February 2014.
- [31] C. Chen, "Using random forest to learn imbalanced data," Ph.D. Thesis, Department of Statistics, University of California, Berkeley, July 2004.
- [32] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous valued attributes for classification learning," *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1993, pp. 1022-1027.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).