International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 4, August 2023, pp. 4683~4691
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i4.pp4683-4691

Bankruptcy prediction model using cost-sensitive extreme gradient boosting in the context of imbalanced datasets

Wirot Yotsawat[1], Kanyalag Phodong[1], Thawatchai Promrat[1], Pakaket Wattuya[2,3]
[1]Department of Science, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University, Phranakhon Si Ayutthaya, Thailand
[2]Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand
[3]Center of Artificial Intelligence Innovation for Healthtech, Faculty of Science, Kasetsart University, Bangkok, Thailand

Article Info

Article history:

Received Oct 7, 2022
Revised Dec 13, 2022
Accepted Jan 14, 2023

Keywords:

Bankruptcy
Cost-sensitive learning
Extreme gradient boosting
Imbalanced dataset
Prediction model

ABSTRACT

In the process of bankruptcy prediction models, a class imbalanced problem has occurred which limits the performance of the models. Most prior research addressed the problem by applying resampling methods such as the synthetic minority oversampling technique (SMOTE). However, resampling methods lead to other issues, e.g., increasing noisy data and training time during the process. To improve the bankruptcy prediction model, we propose cost-sensitive extreme gradient boosting (CS-XGB) to address the class imbalanced problem without requiring any resampling method. The proposed method's effectiveness is evaluated on six real-world datasets, i.e., the LendingClub, and five Polish companies' bankruptcy. This research compares the performance of CS-XGB with other ensemble methods, including SMOTE-XGB which applies SMOTE to the training set before the learning process. The experimental results show that i) based on LendingClub, the CS-XGB improves the performance of XGBoost and SMOTE-XGB by more than 50% and 33% on bankruptcy detection rate (BDR) and geometric mean (GM), respectively, and ii) the CS-XGB model outperforms random forest (RF), Bagging, AdaBoost, XGBoost, and SMOTE-XGB in terms of BDR, GM, and the area under a receiver operating characteristic curve (AUC) based on the five Polish datasets. Besides, the CS-XGB model achieves good overall prediction results.

Corresponding Author:

Wirot Yotsawat
Department of Science, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University
96 Pridi Banomyong Road, Pratuchai, Phra Nakhon Si Ayutthaya, 13000, Thailand
Email: ywirot@aru.ac.th

1. INTRODUCTION

The bankruptcy prediction model has become an important tool because it has an impact on many players, such as fund managers, financial market planners, investors, stockholders, employees, customers, and the nation. The purpose of bankruptcy prediction is to learn a classification model to predict whether an enterprise will go bankrupt or not in the future [1]. The bankruptcy prediction model has been used to minimize the suffering of bankruptcy. For example, i) when a firm requests a loan from a lender, the lender needs to know that the borrower can either repay the loan or go bankrupt and ii) before making an investment in the stock of a firm, investors always worry about the bankruptcy of the company, which may cause a loss of all investments [2].

There are different challenge issues in the bankruptcy prediction task. Firstly, a class imbalanced problem often occurs in the task. It has received great attention in recent research on machine learning.

Resampling techniques, i.e., oversampling and under-sampling, have been utilized to deal with the class imbalanced problem. The resampling techniques change the nature and size of the datasets before the learning process started. Most of the research tackled the class imbalanced problem by applying resampling techniques. For example, He *et al.* [3] and Sun *et al.* [4] tackled a class imbalanced problem in credit scoring by using different imbalance ratios for resampling data. Namvar *et al.* [5] and Moscato *et al.* [6] applied the random under sampling (RUS) technique to form a training set for constructing a credit scoring model. That research achieved good prediction results. However, using resampling techniques leads to several inevitable major limitations that adversely degrade the performance of classification models. For example, model over-fitting, the problem of information loss after under-sampling, and computational cost during the resampling process.

Secondly, prior research considered the improvement of model accuracy without thoroughly considering the detection of minority data. For example, Dželihodžić *et al.* [7] proposed a bagging neural network model, Guo *et al.* [8] improved the linear discriminant analysis model, Goh *et al.* [9] optimized the hyper-parameters of random forest (RF) to form a model, and Garcia *et al.* [10] proposed a bagging decision tree model. The prior research outperformed on overall accuracy and the area under the ROC curve (AUC). However, they were unsuccessful in predicting the minority data, which is more economically important [11], [12]. Smiti and Soui [13] proposed a bankruptcy prediction model by using an auto-encoder and the synthetic minority oversampling technique (SMOTE)-based technique. However, the studies only used AUC to evaluate the model's performance, which is insufficient because other indicators, such as the bankruptcy detection rate (BDR) and geometric mean (GM), may be low, as in other recent studies [3], [10]. Thus, the minority data should be given more consideration, and the models should be measured by suitable indicators.

Extreme gradient boosting (XGBoost) is a powerful algorithm for business failure prediction as well as other domains. Pawełek [14] improved company bankruptcy prediction by applying quantile range outliner removal and XGBoost. Yotsawat *et al.* [15] improved the performance of a credit scoring model by using XGBoost with Bayesian hyper-parameters' optimization. However, those studies ignored the class imbalanced problem. Jabeur *et al.* [16] found that the combination of a feature selection method and the XGBoost algorithm can improve the performance of bankruptcy prediction models. However, the research adopted resampling methods to address the imbalanced issue, which led to other problems as described above.

Cost-sensitive learning is one of the approaches to dealing with a class imbalanced problem. There are only a few articles discussing the cost-sensitive learning in bankruptcy prediction. Ghatasheh *et al.* [11] proposed a bankruptcy model and addressed a highly imbalanced data distribution of firms in Spain by a cost-sensitive ensemble. Siers and Islam [17] implemented a cost-sensitive learning by manipulating predictive threshold's moving manner for software defect prediction. Different from Ghatasheh *et al.* [11] and Siers and Islam [17] works, this study implements cost-sensitive learning by minimizing the overall misclassification error through the training process, which still needs further investigation.

Based on the above considerations, this study proposes a bankruptcy prediction model by using the cost-sensitive XGBoost (CS-XGB) in the context of a class imbalance problem. The performance of the proposed model is measured by accuracy (Acc), AUC, specificity (Spec), BDR, GM, and F1. The contributions of this study are as: i) the proposed model solves a class imbalanced problem in bankruptcy datasets without requiring any complex resampling processes. Experimental results illustrate that the CS-XGB can improve the BDR; ii) compared with the SMOTE-XGB, which employs SMOTE to address the imbalanced problem before creating the XGBoost model, the CS-XGB not only avoids the training time and space during the oversampling and training process, but also provides high performance; iii) this study compares the proposed approach with the existing state-of-the-art techniques used for bankruptcy prediction to confirm the performance of the proposed approach, e.g., RF, bagging, AdaBoost, XGBoost, and SMOTE-XGB. As justified by the empirical study, the proposed approach achieves high predictive performance and provides the best results in terms of BDR and GM.

The rest of the paper is organized as: section 2 describes the theoretical background used in the study. Research methodology is illustrated in section 3. The experimental results are discussed by comparing models in section 4. The paper is concluded in section 5.

2. THEORETICAL BACKGROUND

2.1. Classification methods

Classification is a major task for predicting the health of firms. We evaluated the performance of the proposed CS-XGB with other five ensemble methods, including RF, bagging decision tree (Bagging-DT), Bagging neural network (Bagging-NN), AdaBoost decision tree (AdaBoost-DT), and XGBoost. In addition, the classic and popular oversampling method, namely SMOTE, is applied to the training set for constructing

SMOTE-XGB. The experiment's main objective is to determine the most effective classification model for bankruptcy prediction. These ensemble models used for evaluation are based on four ensemble mechanisms: bagging, random forest, AdaBoost, and XGBoost.

Bagging: different weak based learners are produced from various bootstrap samples taken from the training dataset. A bootstrap sample, also known as sampling with replacement, is a sample taken from the training dataset in which a sample may occur more than once. The majority vote prediction for the classes across all of the predictions provided by the based learners is used to make the final predictions [18]. Decision tree (DT) is a popular algorithm used for based learners in Bagging because it is sensitive to different training sets [4].

RF: from bootstrap samples taken from the training dataset, RF creates a large number of decision trees. In contrast to bagging, RF also selects a subset of input attributes at each point when the trees are split. By this way, each decision tree in the ensemble is forced to be more different [19].

AdaBoost: AdaBoost combines a number of weak classifiers to get a stronger one. To do this, a model is created using the weighted training data, and the instances that are difficult to classify are given more weight than the instances that can be correctly classified. Subsequently, a second model is created that tries to correct the error of the first model. The weak learners' predictions are voted on by a majority, weighted by each learner's accuracy, to determine the final predictions [20].

XGBoost: XGBoost method is based on the gradient boosting framework, which combines DT and gradient boosting [21]. The residual of a base DT classifier is used in the next DT classifier at each step of the training process to enhance the objective loss function. The XGBoost technique decreases modeling complexity and avoids the issues associated with overfitting problem. Finally, the aggregate of all trees generates the final output. Recently, the XGBoost method is a decision tree ensemble method that has been used in a variety of classification and regression problems.

2.2. Cost-sensitive XGBoost

The main concept of the CS-XGB is that the error generated by a positive sample (Bankrupt) of a misclassified class (predicted as healthy) is given a bigger weight in the loss function of the algorithm, so that positive samples receive more attention throughout the learning process [22]. Table 1 illustrates a cost matrix based on two-class classification tasks. True positive (TP) and true negative (TN) represent the correctly classified samples by the model, while false positive (FP) and false negative (FN) represent the test results that are wrongly classified by the model.

In general, the extra loss is on the FN due to it can cause a vast damage. In the cost-sensitive manner, only misclassification cases are considered. Let $c_{00}=c_{11}=0$, $c_{01}=1$, and $c_{10}=a(a>1)$, the (1) illustrates the loss function with a cost-sensitive factor:

$$L_a = -\sum_{i=1}^{n}[a y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (1)$$

where L represents the cross-entropy function which is the default loss function of XGBoost, a represents the cost-sensitive factor, y_i is the true label, and $\hat{y}_i$ represents the output of raw prediction after applied the sigmoid activation.

Table 1. Cost matrix for two-class classification

		Actual class	
		Bankrupt (Positive)	Healthy (Negative)
Predicted class	Bankrupt (Positive)	TP (cost=c_{11})	FP (cost=c_{01})
	Healthy (Negative)	FN (cost=c_{10})	TN (cost=c_{00})

2.3. Assessment criteria

Six criteria were adopted to evaluate the quality of the models: Acc, Spec, BDR, AUC, GM, and F1. We focus on the improvement of BDR with competitive overall performance because the higher BDR reflected in the model can prevent bankruptcy situations. The six criteria are derived from TP, TN, FP, and FN.

The Acc indicates the overall true predicted class across entire classes. Due to the class imbalanced issue, the Acc value by itself, however, cannot reflect model performance. It simply reflects the dataset's total prediction accuracy, which can be dominated by majority samples. The Spec and BDR measure the percentage of TN and TP that the classifier properly classifies as negative and positive, respectively. The GM is a comprehensive evaluation method constructed by Spec and BDR. The higher GM demonstrates an acceptable and effective performance between the classes in the binary classification model. F1 is the weighted harmonic mean of the precision and BDR. The AUC score is utilized to evaluate the quality of the models due to the imbalanced nature of training data. The AUC is calculated by the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate (TPR) at the y-axis and false positive rate (FPR) at the x-axis [23]. The (2)-(6) is used to calculate the Acc, Spec, BDR, GM, and F1, respectively.

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$Spec = \frac{TN}{TN+FP} \tag{3}$$

$$BDR = \frac{TP}{TP+FN} \tag{4}$$

$$GM = \sqrt{Spec \times BDR} \tag{5}$$

$$F1 = \frac{(1+\beta)^2 TP}{(1+\beta)^2 TP+\beta^2 FN+FP} \tag{6}$$

To calculate the F1, denote that the value for β is 1 when the weights for BDR and precision are equal [24].

3. RESEARCH METHOD

3.1. Experimental setup and research framework

In this study, the bankruptcy prediction models were established using Python version 3.7.9 along with other associated libraries. XGBoost library version 1.3.1 was used to create XGBoost-based models. Scikit-learn library version 0.24.0 was used to construct other ensemble models, i.e., RF, AdaBoost, and Bagging. Hyperopt library version 0.2.5 was used for turning the hyper-parameters of classification algorithms. The experiment was performed on a 64-bit platform with an Intel® Core™ i7 7500 CPU and 8 GB of RAM.

The framework of the experimental design is shown in Figure 1. In order to verify the effectiveness of the proposed CS-XGB model, 6 datasets are utilized. To reduce the bias caused by random sampling of the training and test sets, the experiment was set to five-fold cross-validation, which ensured that all samples were chosen for both the training and test sets. The average of the performance indicators will be presented based on the five-fold experiment.

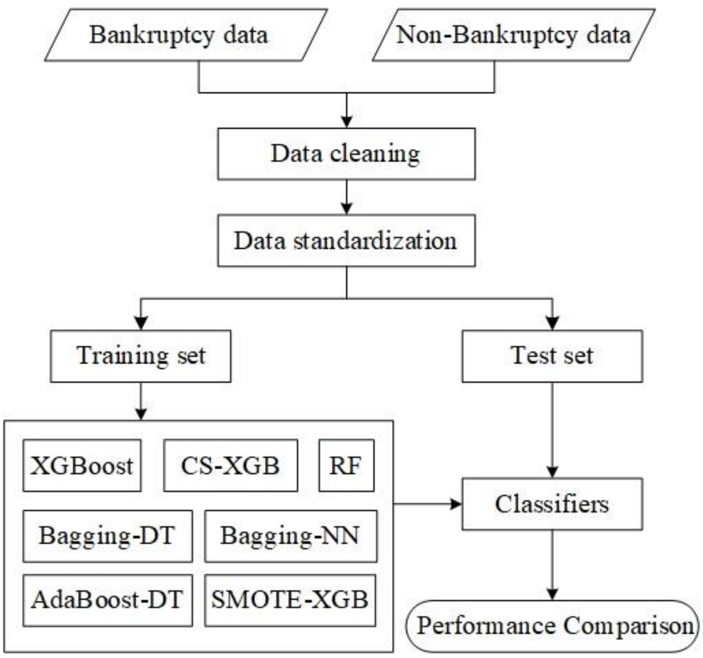

Figure 1. The framework of the experimental design

3.2. Datasets

The performance of the proposed approach is evaluated based on six real-world datasets, consisting of the LendingClub and five Polish datasets. The Polish datasets are obtained from the University of California, Irvine (UCI) machine learning repository, while another dataset is obtained from the previous study [12], which was collected from Kaggle. The Polish datasets present 64 features (financial ratios). The LendingClub dataset consists of 22 features. The dataset's description is described in detail in Table 2.

Table 2. Dataset's description

Datasets	Features	Instances	Negative	Positive	IR	Source
Polish1	64	7,027	6,756	271	24.93	UCI
Polish2	64	10,173	9,773	400	24.43	UCI
Polish3	64	10,503	10,008	495	20.22	UCI
Polish4	64	9,792	9,277	515	18.01	UCI
Polish5	64	5,910	5,500	410	13.41	UCI
LendingClub	22	80,000	64,377	15,623	4.12	[12]

4. RESULTS AND DISCUSSION

In this study, seven ensemble classifiers (e.g., RF, AdaBoost-DT, Bagging-DT, Bagging-NN, XGBoost, SMOTE-XGB, and CS-XGB) were evaluated based on six datasets (e.g., LendingClub and five Polish companies' bankruptcy). To enhance the reliability of the estimates and avoid overfitting problem, the experiment was performed with five-fold cross-validation. The obtained results were based on the average of the five iterations.

4.1. Performance analysis

Table 3 presents the comparison of the seven models' performance based on the LendingClub dataset. XGBoost, SMOTE-XGB, RF, and Bagging failed to detect bankruptcy companies. The proposed CS-XGB performs the best on two evaluated indicators consisting of BDR and GM, while there is no sufficient difference on the best AUC generated by Bagging-NN. Although the Bagging-NN model outperforms other models on Acc, AUC, and F1, it fails on bankruptcy detection, which leads to more damage. Besides, except for CS-XGB, the other ensemble models are unsuccessful in detecting bankrupt firms. Biased bankruptcy models will be produced if the imbalanced problem is ignored. The models were dominated by a majority class and led to non-performing loan (NPL).

Table 3. Comparative results between the proposed CS-XGB and other methods over LendingClub dataset

Methods	Acc	AUC	Spec	BDR	GM	F1
RF	80.47	69.69	**98.48**	5.54	23.34	89.05
AdaBoost-DT	70.64	54.96	80.70	29.23	48.55	81.56
Bagging-DT	77.34	64.27	91.22	20.12	42.84	86.63
Bagging-NN	**80.67**	70.83	98.28	8.10	28.14	**89.11**
XGBoost	80.28	69.72	97.41	9.70	30.71	88.83
SMOTE-XGB	80.43	70.35	95.69	9.57	30.55	88.92
CS-XGB	66.78	**70.87**	68.27	**60.64**	**64.34**	76.78

Table 4 illustrates the comparative results between the proposed CS-XGB and other ensemble methods over five Polish datasets. Cost-insensitive XGBoost outperformed on overall accuracy over the five datasets. However, Acc is not a suitable indicator when datasets are imbalanced. AUC, GM, and F1 are better measurements for identifying the ability of models. The best AUC scores generated by seven models over each dataset were obtained from XGBoost and CS-XGB, which were not significantly different. Three of them were obtained by CS-XGB.

The CS-XGB also achieved the best GM which is a wildly used measure when datasets are imbalanced. It means the ratio of classification over bankrupted and non-bankrupted firms can be correctly classified better than other models. Furthermore, the best F1 scores on four of the five datasets were generated by the CS-XGB approach.

Over the five Polish datasets, all models achieved very high Spec scores, ranging between 96.31 and 100 percentage scores. However, RF, AdaBoost-DT, Bagging-DT, and Bagging-NN were not successful in detecting the bankrupted firms, while CS-XGB outperformed in the bankrupted firms' detection. Thus, CS-XGB showed the best BDR based on the five datasets.

According to [25], bankruptcy detection rates were very low, and the RF model was dominated by a majority class. In this study, the results illustrated in Table 4 conformed to [25] work and showed very high

Acc, AUC, and Spec. However, CS-XGB achieved over RF on BDR, GM and F1. The proposed CS-XGB also outperformed SMOTE-XGB on the compared criteria (Acc, AUC, Spec, BDR, GM, and F1). Besides, SMOTE-XGB takes up more training time during the synthetic oversampling process and makes the model more complex than the proposed CS-XGB.

According to the experimental findings and analyses presented above, in general, the proposed CS-XGB model outperformed the other compared models on BDR and GM. It is critical for avoiding bankruptcy situations. Therefore, it can be considered an effective method to solve bankruptcy prediction problems from the perspective of imbalanced datasets.

Table 4. Comparative results between the proposed CS-XGB and other methods over Polish datasets

Datasets	Methods	Acc	AUC	Spec	BDR	GM	F1
Polish1	RF	97.64	89.66	**99.84**	42.83	64.08	56.44
	AdaBoost-DT	95.80	75.67	97.48	53.86	71.67	49.46
	Bagging-DT	97.71	85.29	99.70	47.98	67.79	60.06
	Bagging-NN	96.94	87.64	99.56	31.77	54.46	43.11
	XGBoost	**98.11**	95.26	99.72	57.94	75.53	69.52
	SMOTE-XGB	96.95	93.89	98.26	64.19	79.11	61.83
	CS-XGB	97.88	**95.54**	99.22	**64.56**	**79.99**	**70.17**
Polish2	RF	97.30	88.39	99.98	31.75	52.74	44.85
	AdaBoost-DT	95.30	72.86	97.22	48.50	66.66	43.61
	Bagging-DT	97.13	82.40	99.58	37.50	58.45	48.27
	Bagging-NN	96.07	75.85	**100.00**	0.00	0.00	0.00
	XGBoost	**97.67**	92.66	99.94	42.25	62.82	56.02
	SMOTE-XGB	95.70	91.13	97.21	57.90	75.25	51.60
	CS-XGB	97.20	**92.71**	98.79	**58.25**	**75.26**	**61.31**
Polish3	RF	96.45	89.06	99.91	26.55	46.61	38.59
	AdaBoost-DT	94.11	70.68	96.54	44.82	65.62	41.63
	Bagging-DT	96.30	87.70	**99.98**	21.89	41.31	33.67
	Bagging-NN	95.43	71.81	100.00	2.04	14.29	4.00
	XGBoost	**97.02**	**93.68**	99.85	39.82	62.96	55.61
	SMOTE-XGB	95.13	91.44	96.97	57.99	74.82	52.83
	CS-XGB	95.99	92.91	97.73	**60.84**	**76.95**	**58.88**
Polish4	RF	95.92	89.80	**99.94**	23.50	48.03	37.36
	AdaBoost-DT	94.25	72.37	96.82	47.93	67.92	46.79
	Bagging-DT	96.25	85.28	99.61	35.70	59.36	49.75
	Bagging-NN	94.62	80.95	99.29	10.52	31.68	16.87
	XGBoost	**96.99**	93.00	99.80	46.40	67.79	61.44
	SMOTE-XGB	94.89	91.33	96.60	60.93	76.26	52.22
	CS-XGB	95.93	**93.82**	97.62	**65.40**	**79.77**	**62.78**
Polish5	RF	95.26	92.95	99.16	42.93	64.97	55.45
	AdaBoost-DT	94.53	79.80	96.91	62.68	77.82	61.38
	Bagging-DT	95.47	93.55	**99.73**	38.29	61.28	53.32
	Bagging-NN	93.93	87.81	98.87	27.56	51.46	37.95
	XGBoost	**96.68**	95.68	99.42	60.00	76.98	**71.11**
	SMOTE-XGB	94.53	95.17	96.07	73.90	84.21	65.25
	CS-XGB	94.92	**95.69**	96.31	**76.34**	**85.64**	67.66

4.2. Performance comparison

Tables 5 and 6 show the comparison results of the proposed CS-XGB and prior works on the LendingClub and Polish datasets, respectively. In Table 5, each research study randomly selected the LendingClub data on different ratios, attributes, and periods based on the different sizes of the data. However, the goal of the studies is to improve the performance of a predictive model. As shown in both tables, the bold numbers represent the best score for each indicator, and the italic numbers represent the worst score for indicators, which the models illustrated contrast with the best scores.

As we can see in Table 5, although EBCA+PSO [3] shows the best AUC and F1 scores, it provides only 1.64 on GM. It means EBCA+PSO can correctly predict only one class. CatBoost model proposed by [26] shows very high Acc. However, it shows the three lowest AUC scores. The best model for the LendingClub dataset is RF-RUS [6]. However, it shows slightly better than the proposed CS-XGB which is not sufficiently different.

In Table 6, CS-XGB outperforms other methods on AUC, BDR, and GM based on five Polish datasets. In the context of imbalanced datasets, the high AUC, BDR, and GM are very important indicators for bankruptcy prediction models. The number of NPLs will be lower when the BDR is high. NPLs, on the other hand, will rise as the BDR decreases. CS-XGB also provides the best F1 score on three of five Polish datasets. However, CS-XGB gives slightly lower results than other methods in Acc and Spec.

Table 5. Comparative results between the proposed CS-XGB and prior works over LendingClub dataset

Techniques	Year	Instances	Acc	AUC	Spec	BDR	GM	F1
RF [25]	2015	68,000	78.00	71.00	**88.00**	*31.00*	-	-
EBCA+PSO [3]	2018	95,633	-	**73.07**	-	-	*1.64*	**99.38**
RF-RUS [5]	2018	66,376	69.20	69.00	71.70	58.20	65.00	-
CatBoost [26]	2019	11,467	**79.59**	*63.33*	-	-	-	-
[26]	2019	26,288	77.16	*62.66*	-	-	-	-
[26]	2019	26,384	75.23	*61.62*	-	-	-	-
DM–ACME [27]	2020	70,860	72.31	66.97	76.78	46.07	60.09	-
XGBoot [28]	2020	1,347,681	63.60	67.40	64.50	60.20	63.60	73.50
RF-RUS [6]	2021	462,378	64.00	71.70	68.00	**63.00**	65.60	-
CS-XGB	2023	88,890	66.78	70.87	68.27	60.64	64.34	76.78

Table 6. Comparative results between the proposed CS-XGB and prior works over Polish datasets

Datasets	Techniques	Year	Acc	AUC	Spec	BDR	GM	F1
Polish1	IFNA+backflow XGB [29]	2019	97.46	91.98	-	-	-	53.65
	Bag-C4.5 [10]	2019	-	92.30	99.00	*5.02*	-	-
	ABoost(C4.5) [10]	2019	-	87.00	99.70	53.10	-	-
	V-GANs [30]	2019	-	73.79	-	52.05	-	48.83
	BLOF-RF [31]	2021	98.10	94.88	-	-	-	71.09
	XGB-BO [15]	2021	98.11	95.32	99.78	56.48	74.61	-
	RGA-XGBoost [32]	2022	97.58	80.97	98.98	62.96	-	-
	CS-XGB	2023	97.88	**95.54**	99.22	**64.56**	**79.99**	70.17
Polish2	Bag-C4.5 [10]	2019	-	87.9	99.9	40.4	-	-
	RotF(C4.5) [10]	2019	-	88.8	100.0	30.8	-	-
	V-GANs [30]	2019	-	70.7	-	46.08	-	44.01
	BLOF-LightGBM [31]	2021	96.94	87.41	-	-	-	52.98
	RGA-XGBoost [32]	2022	97.38	78.50	99.01	58.00	-	-
	CS-XGB	2023	97.20	**92.71**	98.79	**58.25**	75.26	**61.31**
Polish3	Bag-C4.5 [10]	2019	-	90.0	99.6	34.6	-	-
	RF(C4.5) [10]	2019	-	88.8	99.8	*16.6*	-	-
	V-GANs [30]	2019	-	73.51	-	51.22	-	49.17
	BLOF-XGB [31]	2021	97.00	91.58	-	-	-	59.76
	RGA-XGBoost [32]	2022	95.60	77.15	97.53	56.77	-	-
	CS-XGB	2023	95.99	**92.91**	97.73	**60.84**	76.95	58.88
Polish4	Bag-C4.5 [10]	2019	-	90.1	99.7	37.9	-	-
	RF(C4.5) [10]	2019	-	89.1	99.7	*13.2*	-	-
	BLOF-AdaBoost [31]	2021	96.25	93.03	-	-	-	52.81
	RGA-XGBoost [32]	2022	95.39	77.68	97.49	57.86	-	-
	CS-XGB	2023	95.93	**93.82**	97.62	**65.40**	79.77	**62.78**
Polish5	Bag-C4.5 [10]	2019	-	93.6	99.3	58.8	-	-
	RF(C4.5) [10]	2019	-	93.3	99.4	34.6	-	-
	V-GANs [30]	2019	-	74.18	-	52.2	-	50.95
	BLOF-LightGBM [31]	2021	95.77	94.90	-	-	-	66.27
	RGA-XGBoost [32]	2022	95.25	82.90	97.26	68.54	-	-
	CS-XGB	2023	94.92	**95.69**	96.31	**76.34**	85.64	**67.66**

5. CONCLUSION

Bankruptcy prediction is very important for the sustainable development of companies as a financial early warning tool. In many tasks, including bankruptcy prediction, a class imbalanced problem has been occurring, which limits the performance of the predictive models. Biased models will be produced if the class imbalanced problem is ignored. The prior research addressed the class imbalanced problem by applying resampling methods such as under-sampling and oversampling. SMOTE is a classic and popular oversampling technique used to solve the class imbalanced problem. However, SMOTE and other resampling methods lead to other problems such as increasing noisy data if applying oversampling techniques, information loss if applying under-sampling techniques, and the training time during the resampling process. We propose cost-sensitive extreme gradient boosting (CS-XGB) to improve the bankruptcy prediction model while ensuring the model's efficiency. The proposed approach addresses the class imbalanced problem without requiring any complex resampling method. The proposed method's effectiveness is evaluated on six widely used real-world datasets, i.e., LendingClub and five Polish companies' bankruptcy. This research compares the performance of the proposed CS-XGB with other ensemble methods, i.e., RF, Bagging, Boosting, XGBoost, and SMOTE-XGB. The experimental results revealed that the proposed CS-XGB approach improves the predictive performance over RF, Bagging, AdaBoost, XGBoost and SMOTE-XGB, especially on the BDR and GM with competitive overall prediction results.

ACKNOWLEDGEMENTS

This research was supported by the Research and Development Institute, Phranakhon Si Ayutthaya Rajabhat University, Thailand.

REFERENCES

[1] N. Chen, B. Ribeiro, and A. Chen, "Financial credit risk assessment: a recent review," *Artificial Intelligence Review*, vol. 45, no. 1, pp. 1–23, Oct. 2015, doi: 10.1007/s10462-015-9434-x.
[2] D. Ogachi, R. Ndege, P. Gaturu, and Z. Zoltan, "Corporate bankruptcy prediction model, a special focus on listed companies in Kenya," *Journal of Risk and Financial Management*, vol. 13, no. 3, Mar. 2020, doi: 10.3390/jrfm13030047.
[3] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: Adaption of different imbalance ratios," *Expert Systems with Applications*, vol. 98, pp. 105–117, May 2018, doi: 10.1016/j.eswa.2018.01.012.
[4] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Information Sciences*, vol. 425, pp. 76–91, Jan. 2018, doi: 10.1016/j.ins.2017.10.017.
[5] A. Namvar, M. Siami, F. Rabhi, and M. Naderpour, "Credit risk prediction in an imbalanced social lending environment," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, 2018, doi: 10.2991/ijcis.11.1.70.
[6] V. Moscato, A. Picariello, and G. Sperlí, "A benchmark of machine learning approaches for credit score prediction," *Expert Systems with Applications*, vol. 165, Mar. 2021, doi: 10.1016/j.eswa.2020.113986.
[7] A. Dželihodžić, D. Đonko, and J. Kevrić, "Improved credit scoring model based on bagging neural network," *International Journal of Information Technology and Decision Making*, vol. 17, no. 6, pp. 1725–1741, Nov. 2018, doi: 10.1142/S0219622018500293.
[8] Y. Guo, J. He, L. Xu, and W. Liu, "A novel multi-objective particle swarm optimization for comprehensible credit scoring," *Soft Computing*, vol. 23, no. 18, pp. 9009–9023, Sep. 2019, doi: 10.1007/s00500-018-3509-y.
[9] R. Y. Goh, L. S. Lee, H.-V. Seow, and K. Gopal, "Hybrid harmony search-artificial intelligence models in credit scoring," *Entropy*, vol. 22, no. 9, Sep. 2020, doi: 10.3390/e22090989.
[10] V. García, A. I. Marqués, and J. S. Sánchez, "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, doi: 10.1016/j.inffus.2018.07.004.
[11] N. Ghatasheh *et al.*, "Cost-sensitive ensemble methods for bankruptcy prediction in a highly imbalanced data distribution: a real case from the Spanish market," *Progress in Artificial Intelligence*, vol. 9, no. 4, pp. 361–375, Dec. 2020, doi: 10.1007/s13748-020-00219-x.
[12] W. Yotsawat, P. Wattuya, and A. Srivihok, "A novel method for credit scoring based on cost-sensitive neural network ensemble," *IEEE Access*, vol. 9, pp. 78521–78537, 2021, doi: 10.1109/ACCESS.2021.3083490.
[13] S. Smiti and M. Soui, "Bankruptcy prediction using deep learning approach based on borderline SMOTE," *Information Systems Frontiers*, vol. 22, no. 5, pp. 1067–1083, Oct. 2020, doi: 10.1007/s10796-020-10031-6.
[14] B. Pawełek, "Extreme gradient boosting method in the prediction of company bankruptcy," *Statistics in Transition New Series*, vol. 20, no. 2, pp. 155–171, Jun. 2019, doi: 10.21307/stattrans-2019-020.
[15] W. Yotsawat, P. Wattuya, and A. Srivihok, "Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5477–5487, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5477-5487.
[16] S. B. Jabeur, N. Stef, and P. Carmona, "Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering," *Computational Economics*, Jan. 2022, doi: 10.1007/s10614-021-10227-1.
[17] M. J. Siers and M. Z. Islam, "Software defect prediction using a cost sensitive decision forest and voting, and a potential solution to the class imbalance problem," *Information Systems*, vol. 51, pp. 62–71, Jul. 2015, doi: 10.1016/j.is.2015.02.006.
[18] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.
[19] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
[20] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52.
[21] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[22] S. He, B. Li, H. Peng, J. Xin, and E. Zhang, "An effective cost-sensitive XGBoost method for malicious URLs detection in imbalanced dataset," *IEEE Access*, vol. 9, pp. 93089–93096, 2021, doi: 10.1109/ACCESS.2021.3093094.
[23] J. Abellán and J. G. Castellano, "A comparative study on base classifiers in ensemble methods for credit scoring," *Expert Systems with Applications*, vol. 73, pp. 1–10, May 2017, doi: 10.1016/j.eswa.2016.12.020.
[24] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Dec. 2020, doi: 10.1186/s12864-019-6413-7.
[25] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, Jun. 2015, doi: 10.1016/j.eswa.2015.02.001.
[26] Y. Xia, L. He, Y. Li, N. Liu, and Y. Ding, "Predicting loan default in peer-to-peer lending using narrative data," *Journal of Forecasting*, vol. 39, no. 2, pp. 260–280, Mar. 2020, doi: 10.1002/for.2625.
[27] Y. Song, Y. Wang, X. Ye, D. Wang, Y. Yin, and Y. Wang, "Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending," *Information Sciences*, vol. 525, pp. 182–204, Jul. 2020, doi: 10.1016/j.ins.2020.03.027.
[28] M. J. Ariza-Garzon, J. Arroyo, A. Caparrini, and M.-J. Segovia-Vargas, "Explainability of a machine learning granting scoring model in peer-to-peer lending," *IEEE Access*, vol. 8, pp. 64873–64890, 2020, doi: 10.1109/ACCESS.2020.2984412.
[29] S. Wei, D. Yang, W. Zhang, and S. Zhang, "A novel noise-adapted two-layer ensemble model for credit scoring based on backflow learning," *IEEE Access*, vol. 7, pp. 99217–99230, 2019, doi: 10.1109/ACCESS.2019.2930332.
[30] H. Mansourifar, L. Chen, and W. Shi, "Virtual big data for GAN based data augmentation," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 1478–1487, doi: 10.1109/BigData47090.2019.9006268.
[31] W. Zhang, D. Yang, S. Zhang, J. H. Ablanedo-Rosas, X. Wu, and Y. Lou, "A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring," *Expert Systems with Applications*, vol. 165, Mar. 2021, doi: 10.1016/j.eswa.2020.113872.

[32] K. Borowska and J. Stepaniuk, "Rough-granular approach in imbalanced bankruptcy data analysis," *Procedia Computer Science*, vol. 207, pp. 1832–1841, 2022, doi: 10.1016/j.procs.2022.09.241.

BIOGRAPHIES OF AUTHORS

Wirot Yotsawat received the B.Sc. degree in Computer Science from the School of Informatics, Walailak University, Thailand in 2007. He obtained the M.Sc. and Ph.D. degree in Computer Science from the Faculty of Science, Kasetsart University, Thailand in 2014 and 2021, respectively. Currently, he is a lecturer at the Division of Computer Science, Department of Science, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University, Thailand. His research interests include data mining, machine learning, and internet of things. He can be contacted at email: ywirot@aru.ac.th.

Kanyalag Phodong graduated with Bachelor's Degree in Computer Science from Naresuan University, Thailand in 2008. She obtained the M.Sc. and Ph.D. Degree in Computer Science at the Department of Computer Science, Thammasat University, Thailand in 2011 and 2021, respectively. Currently, she is a lecturer at the Division of Computer Science, Department of Science, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University, Thailand. Her researches are in fields of artificial intelligence, natural language processing, learning analytics, and information system. She can be contacted at email: pkanyalag@aru.ac.th.

Thawatchai Promrat received the B.Sc. degree in Computer Science from Naresuan University, Thailand in 2010. He obtained the M.Sc. degree in Software Engineering from the Faculty of Information Technology, King Mongkut's University of Technology Thonburi (KMUTT), Thailand in 2020. Currently, he is a lecturer at the Division of Information Technology, Department of Science, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University, Thailand. His research interests include computer vision, deep learning, machine learning, and internet of things. He can be contacted at email: thawatchai@aru.ac.th.

Pakaket Wattuya received the B.Sc. degree in computer science from the Faculty of Science and the M.Eng. in computer engineering from the Faculty of Engineering, Kasetsart University, Thailand in 2000 and 2004, respectively. She received the Dr.rer.nat. degree in Computer Science from Westfälische Wilhelms-Universität Muenster, Germany in 2010. She is currently an Assistant Professor with the Department of Computer Science, Kasetsart University. Her research interests include image processing, computer vision, machine learning, and deep learning. She can be contacted at email: fscipkw@ku.ac.th.