Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD37672

Taiwanese Bankruptcy Prediction with LASSO and Default-Put-Inspired Random Forests

Zhuo Ren

Faculty of Science, National University of Singapore, Singapore, Singapore
e0952293@u.nus.edu

Abstract. Bankruptcy prediction is a critical but unusually demanding classification problem because failed firms are rare and the cost of a missing is often larger than the cost of a false alarm. This study examines 6,819 Taiwanese firm observations described by 95 financial indicators, of which 220 (3.23%) are labeled bankrupt. Using a stratified 60%/20%/20% split, this study applies LASSO to select 25 accounting variables and compares supervised, ensemble, and anomaly-detection models under a recall-focused measure framework. A repeated LASSO filtering is applied and takes five additional structural-credit features into consideration, including an American default-put value approximated by a neural-network surrogate trained on least-squares Monte Carlo prices and an analytical terminal-shortfall probability; 25 total predictors are again retained. Among baseline models, random forest provides the strongest AUC-precision balance (AUC 0.9548; precision 0.3971; recall 0.6136). The refitted credit proxy-augmented random forest without SMOTE raises average precision from 0.3451 to 0.3790 without changing recall. Its SMOTE variant produces an AUC of 0.9574, average precision of 0.4198, and recall of 0.6364. In conclusion, findings of this paper demonstrate a superior performance for ensemble models (e.g. Random Forest) and an efficient feature engineering framework which can be further extended and integrated under structural credit theory.

Keywords: Bankruptcy prediction, Imbalanced learning, Random forest, LASSO, Structural credit risk

1. Introduction

As a typical type of default events, bankruptcy can bring negative effects extending beyond the firm itself. Creditors face write-downs, employees and suppliers lose income, and investors must revise their assessment of related firms in cases of bankruptcy. Additionally, with development of financial markets, bankruptcy prediction has been playing an increasingly crucial role in early default detection, credit risk measurement and financial losses reduction [1, 2]. Therefore, attaining a precise and prompt forecasting model will help avoid financial frauds, protect interests of stakeholders and eventually develop an efficient market with indicative financial signals and metrics. To be more specific, a useful early-warning model has to do more than classify the average company correctly. It must identify a meaningful share of the distressed minority while keeping false alarms within a range that a lender, auditor, or regulator can act on.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD37672

Early bankruptcy prediction research can be traced back to FitzPatrick, with comparisons between financial ratios of successful industrial firms and those of failed companies [3]. Altman's discriminant model established the primary motivation that a small set of accounting ratios could provide a practical distress signal [4]. With rapid development of artificial intelligence and computing power, more attention was transitioned from original financial models to machine learning methods. Recent comparisons showed that random forests, bagging, and boosting could capture nonlinearities and interactions that were difficult to represent in a single linear score [5]. The gain in flexibility, nevertheless, creates a second problem. Financial statements often contain several versions of the same economic quantity, so a high-dimensional classifier can become unstable or opaque even when its test accuracy appears strong.

The dataset used in this paper drives above problems and requirements more urgent. Containing 6,819 firm observations and 95 financial indicators but only 220 bankrupt observations, the Taiwanese bankruptcy prediction data from Taiwanese economic journal dating from 1999 to 2009 collected in UCI Machine Learning Repository demonstrate an extreme trend of imbalance. A classifier that predicts every firm as healthy will therefore be more than 96% accurate. At the same time, several groups of profitability, leverage, and liquidity ratios are almost duplicates. Consequently, this paper initially aims at measuring and comparing performances of traditional and trending machine learning models with suitable statistical metrics. Additionally, this paper applies LASSO feature selection and SMOTE processing methods for underlying dataset in order to further improve model performances under a more comprehensive and robust evaluation metric framework. Finally, this paper tries to simulate and approximate effective proxy features to further raise model performances and provide a possible measure for companies' credit-adjusted valuation.

The rest of the paper is organized as follows: Section 2 takes the relevant literature review. Section 3 outlines methodologies, including selected supervised learning models, unsupervised models, feature selection method details and concrete feature engineering process. Section 4 demonstrates results, analyzes findings, states limitations and further extensions for improvement and eventually draws conclusions.

2. Related work

2.1. Imbalance-aware bankruptcy prediction

Work on the same Taiwanese dataset increasingly treated bankruptcy as an imbalanced statistical problem. Wang and Liu compared different undersampling strategies and introduced the F2 measure in research to give missed bankruptcies greater weights [6]. Sun combined oversampling with Principal Component Analysis method, reporting that the random forest outperformed baseline support vector machines and boosting models [7]. A more recent study combined random forests with hybrid sampling methods [1], while another cost-sensitive ensemble study argued that the loss function should reflect the asymmetric consequences of the two error types [8].

Research on bankruptcy cases from other markets was also consistent on the importance of imbalance, but not on a single remedy. Garcia demonstrated that synthetic sampling could essentially improve recall and AUC in a U.S. bankruptcy sample [9]. Gnip et al. compared 45 imbalance-learning methods across 15 datasets and found that ensemble and sampling combinations tended to demonstrate a larger probability of stronger statistical performances [10]. In contrast, Amirshahi and Lahmiri showed that unsuitable oversampling could possibly corrupt the performance of some optimized ensembles [11]. These results make resampling an empirical choice

DOI: 10.54254/2754-1169/2026.LD37672

rather than an unconscious casual improvement. They also justify evaluating the original and SMOTE-based models on the same fixed set of test records.

2.2. L1-regularized feature selection

Due to high dimensionality of Taiwanese dataset structure and conceptual overlap of specific financial ratios, problems like multicollinearity and overfitting arise. Feature selection was proved necessary and efficient in order to achieve better prediction performance, reduce computational cost and improve explainability [12]. Recent evidence based on the same dataset also showed that feature selection could improve performance of bankruptcy classifiers [13]. More directly, Yang et al. demonstrated that an L1-penalised selection procedure could reduce the original 95 ratios to a small set of economically interpretable predictors while maintaining, and in their comparison improving, predictive performance [14]. Consequently, this paper includes LASSO as a sparse and interpretable specification to filter redundant features and preserve critical information.

2.3. Feature engineering and model interpretation

Recently, more studies began to expand original dataset information through diversified feature engineering techniques. Wang et al. transformed skewed financial ratios before classification [6], while Li et al. demonstrated that nonlinear unary and pairwise transformations could produce more informative representations of original bankruptcy data [15]. Subsequently, Krimpas et al. Aggregated mutually correlated financial ratios into more interpretable concepts involving liquidity pressure, leverage stress, and profitability deterioration, and propagated them into a firm-level risk index framework.

Meanwhile, Merton et al. had put forward another intriguing perspective which systematically interpreted corporate default as an option-theoretic event and suggested a possible way to construct new features reflecting credit-adjusted value of individual companies through hypothetical European Default Put Options deriving from credit valuation of underlying firms [16]. Inspired by this theory, Charalambous et al. raised a novel neuro-structural framework which approximated firm asset value and volatility for subsequent Black-Scholes-Merton bankruptcy probability estimation [17]. Similar to this direction with structural credit risk analysis but different in details, this paper combines feature engineering with traditional credit risk neutral pricing system. Through constructing reasonable new proxies, this paper eventually aims at further leveling up classification performance of bankruptcy prediction and reasonably reflecting credit-adjusted value of individual companies through hypothetical American Default Put Options deriving from credit valuation of underlying firms.

3. Methodology

3.1. Evaluation metric framework

Traditional machine learning methods adopt accuracy as one of core metrics for model performance evaluation. Nevertheless, imbalanced data can attain high accuracy dominated by the majority class by simply driving models predict most firms as non-bankrupt while failing to detect minority bankrupt cases. Consider definitions of individual elements of a confusion matrix in bankruptcy prediction, which are listed in Table 1:

DOI: 10.54254/2754-1169/2026.LD37672

Table 1. Confusion matrix definitions

Confusion Matrix Statsitics	Detailed Description
True Positive	Bankrupt firms classified as bankrupt
False Positive	Non-bankrupt firms classified as bankrupt
False Negative	Bankrupt firms classified as non-bankrupt
True Negative	Non-bankrupt firms classified as non-bankrupt

From statistical definitions above, potential bankrupt cases which models fail to detect timely can bring drastic shock and damage to stake holders. Accordingly, this paper decides to introduce in multiple evaluation standards with more attention leaning on concepts regarding False Negative Rate and Recall Rate.

To expand in details, accuracy is reported for completeness, but it is not used as the primary decision rule as stated. Precision measures how many flagged firms are actually bankrupt. Recall measures how many bankrupt firms are detected. The F2 score combines the two while weighting recall four times as heavily as precision. ROC-AUC summarizes ranking over all classification thresholds, whereas average precision concentrates on the precision-recall curve and is more informative when the positive class is rare. The six score-producing supervised baselines are tuned with stratified 10-fold cross-validation on the training data; the hard-voting ensemble combines fitted class labels, and the anomaly-detection models remain fixed benchmarks. For each model with probabilities or decision scores, an F2-guided validation grid is ranked primarily by F2, with recall, balanced accuracy, and precision used as tie-breakers. The reported threshold is the mean of the three leading candidates. The resulting classifier is evaluated against the test labels, which are not used for model or threshold selection.

3.2. Feature selection method

High dimensional features (e.g. 95 in this data set) tend to introduce in multicolinearity and overfitting problems as stated before, which can easily cause imprecise statistical results and further torture coefficient directions of certain highly-correlated features. This paper conducts exploratory VIF diagnosis by calculating diagonal values of pseudo-inverse of correlation matrix (see equations (1)-(5)), which was proved equivalent to and under some cases more robust than original VIF statistics mathematically by Marquardt, D. W. [18]:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{1}$$

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2} \tag{2}$$

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{3}$$

DOI: 10.54254/2754-1169/2026.LD37672

$$R = \frac{1}{n} Z^T Z = \begin{pmatrix} 1 & r_j^T \\ r_j & R_{-j} \end{pmatrix} \tag{4}$$

$$VIF_j = \frac{1}{1-\left(reg_coef_j\right)^2} = \frac{1}{1-r_j^T (R_{-j})^{-1} r_j} = \left(R^{-1}\right)_{jj} \tag{5}$$

where $\overline{x}_j$ refers to the sample mean, x_{ij} refers to specific sample value, n refers to sample volume, s_j refers to the sample standard deviation, z_{ij} refers to the adjusted sample value after standardization, Z refers to the standardized sample feature vector, R refers to the correlation matrix of the sample feature vector, r_j refers to the correlation coefficient vector between the j -th feature and other features, R_{-j} refers to the remaining correlation matrix excluding the j -th feature, reg_coef_j refers to the coefficient of determination in auxiliary regression of the j -th feature against other features and VIF_j refers to the VIF statistics of the j -th feature.

The highest VIF for research features achieves $5.123e^{+10}$, which demonstrates obvious multicolinearity in specific variables. To handle the problem, this paper applies LASSO for regularization and feature selection in equations (6)-(8):

$$\widetilde{\beta} = \underset{\beta}{argmin} \left(\|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right) \tag{6}$$

$$\|X\|_1 = \sum_{i=1}^{n} |x_i| \tag{7}$$

$$\|X\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{8}$$

where $\widetilde{\beta}$ refers to the feature coefficient after regularization shrinkage, β refers to the regression coefficient vector, Y refers to the binary bankruptcy target label, X refers to standardized design matrix, x_i refers to the individual element of the corresponding vector, λ refers to the imposed punishment tensity and $\|X\|_1$, $\|X\|_2$ refer to absolute-value and Euclidean norms of the vector respectively. Tuning punishment preserves different amounts of features with non-0 coefficients. With punishment increasing, the feature number firstly declines rapidly and gradually slows down at a relatively stable speed. Table 2 shows relationships between punishment and remaining feature numbers, and lists part of selected features most correlated with bankruptcy label for reference.

Table 2. Punishment-Feature relationships & part of selected features

Punishment	No. of Features	Name of Partly Selected Features
0.001331	51	Net Income to Total Assets

DOI: 10.54254/2754-1169/2026.LD37672

Table 2. (continued)

0.002403	48	Debt ratio %
0.003137	36	Borrowing dependency
0.004817	24	Total expense/Assets
0.007395	14	Liability-Assets Flag
0.011355	8	Current Liability to Current Assets
0.017433	5	Cash Flow to Equity
0.026766	3	Non-industry income and expenditure/revenue
0.041095	1	Cash/Current Liability
0.063096	0	Net Worth Turnover Rate (times)

3.3. Classification model structure

This paper adopts a range of supervised and unsupervised models to tackle the classification task for comparison. To optimize model performance, the complete data set is randomly split into train (60%), validation (20%) and test (20%) sets. One constant predictor, Net Income Flag, is removed, leaving 94 candidate accounting variables. Imputation and scaling remain inside the model pipelines even though the supplied data have no missing values. Subsequently, a conservative outlier rule is estimated from the training partition: a record is flagged and removed only when it falls below the 0.01 percentile or above the 99.99 percentile in at least five variables. Additionally, Grid Search method is used to tune parameters on validation set, targeting at achieving highest F2 score as this paper focuses on recall rate. To expand in details, this paper deploys a closed-endpoint grid of 10 nodes for the regularization penalty parameter in logistic regression and supervised Support Vector Machine. Additionally, the supervised SVC and logistic regression deploy a Gamma type between scaling and auto, and a penalty type between l1 and l2 respectively. For further specifications, Table 3 lists core grid search parameters among main statistical models. Figure 1 illustrates overall model settings and data flow structure among models.

Table 3. Grid-search parameter settings for the evaluated models

Model	Max Depth	Min Samples Leaf	No. of Estimators	Subsample	Colsample
Decision Tree	[3, 5, 8, None]	[10, 15, 20, 25, 30, 35, 40, 45, 50]	—	—	—
Random Forest (RF)	[2, 4, 6, 8, 10, None]	[10, 15, 20, 25, 30]	[100, 400] (10 points)	—	—
Bagging Decision Tree	—	—	[100, 400] (10 points)	—	—
XGBoost	[2, 4, 6, 8, 10, None]	—	[100, 400] (10 points)	[0.5, 0.6, 0.7, 0.8]	[0.5, 1.0] (10 points)
New Feature + RF	[2, 4, 6, 8, 10, None]	[10, 15, 20, 25, 30, 35, 40, 45, 50]	[100, 400] (10 points)	—	—
New Feature + RF + SMOTE	[2, 4, 6, 8, 10, None]	[10, 15, 20, 25, 30, 35, 40, 45, 50]	[100, 400] (10 points)	—	—

Notice:
Each column is one independently tuned parameter.
'—' means that the parameter is not included in that model's grid.
'[a, b] (n points)' denotes the closed-endpoint grid [a, b] for a list of given number n of the underlying parameter.
'None' denotes that there are no specific limitations for the underlying parameter.

DOI: 10.54254/2754-1169/2026.LD37672

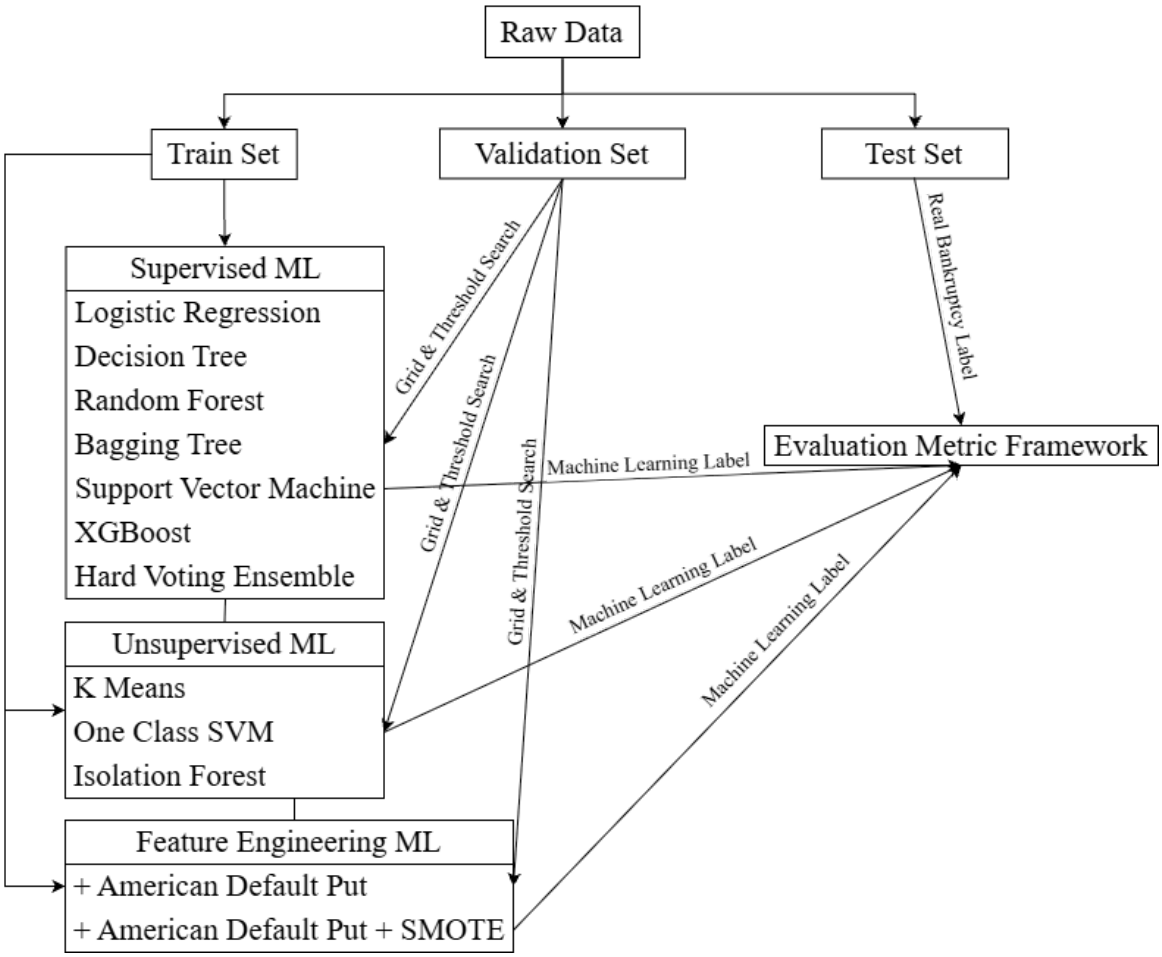


Figure 1. Main pipeline structure (picture credit: original)

3.4. American default put-inspired feature engineering

Structural credit theory relates default risk to the distance between firm asset value and a debt boundary [16]. Accordingly, this section aims at performing synthetic combination of financial features and providing a feasible methodological bridge between structural pricing and bankruptcy prediction. To expand in details, Merton's structural credit theory denoted company's value as tradable assets following Geometric Brownian Motion (GBM) diffusion, see equation (9):

$$\frac{dV_t}{V_t} = rdt + \sigma_V dW_t^{\mathbb{Q}} \tag{9}$$

where V_t refers to tradable underlying asset value, r refers to return rate, σ_V refers to asset volatility, $dW_t^{\mathbb{Q}} \sim N\left(0, dt\right)$ refers to corresponding random walk increment and $\mathbb{Q}$ refers to risk neutral probability measure. Assume the company's liability could be viewed as zero-coupon bond. For equity holders, investing the company could be viewed as holding an American-style default put option, with wealth functional of which stated as equation (10):

DOI: 10.54254/2754-1169/2026.LD37672

$$P_{American}(V_0,\ K,\ \sigma,\ r,\ T) = \sup_{\tau \epsilon \Gamma} E^{\mathbb{Q}}[e^{-r\tau}(K - V_\tau)^+] \tag{10}$$

where $P_{American}(V_0,\ K,\ \sigma,\ r,\ T)$ refers to the corresponding American default put value, K refers to the value of the zero-coupon bond, T refers to time horizon, Γ refers to the simulation paths and τ refers to discrete time point at a specific path. At each time step, equity holders should compare instant default return with continuation option value, stated as equations (11)-(13):

$$H_t = (K - V_t)^+ \tag{11}$$

$$C_t = E^{\mathbb{Q}}\left[e^{-rt} P_{t+\Delta t} \middle| V_t\right] \tag{12}$$

$$H_t \leq C_t \ for\ continuous\ holding;\ H_t > C_t \ for\ executing \tag{13}$$

where H_t refers to the instant option return, C_t refers to the continuation value calculated by the discounted conditional risk neutral expectation of future value which is denoted as $P_{t+\Delta t}$. Investors stipulated an optimal strategy and decided whether to continue holding or exerting the option by comparisons between the two statistics at each time point along the whole time horizon. To simulate the scenario, this paper combines financial indicators within aspects of profitability, liquidity, leverage and growth instability, stated as follows in equations (14)-(16):

$$logV_{0,i} = 1 + 0.35P_i + 0.25L_i - 0.15G_i \tag{14}$$

$$logK_i = 1 + 0.45G_i - 0.10L_i \tag{15}$$

$$\sigma_i = clip(0.18 + 0.18\,|U_i| + 0.08\,|P_i|,\ 0.05,\ 0.90) \tag{16}$$

where P_i, L_i, G_i and U_i refer to corresponding profitability, liquidity, leverage and growth instability indicators respectively, i refers to the specific firm case and $clip(\cdot)$ refers to the truncation function. Nevertheless, it is extremely time-consuming to simulate a bunch of asset paths for each company. This paper solves this problem by using parameter grids to simulate the option price grid and applying a 5-64-32-1 Full-Connected Neural Network to surrogate the option price function according to Universal Approximation Theorem from Leshno et al. [19], which are in all stated as equations (17)-(19) below:

DOI: 10.54254/2754-1169/2026.LD37672

$$V_{t+\Delta t} = V_t exp\left[\left(r - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}Z_t\right] \tag{17}$$

$$\widetilde{C_t} = \widetilde{\beta_0} + \widetilde{\beta_1}V_t + \widetilde{\beta_2}V_t^2 \tag{18}$$

$$f_\theta(V_0,\ K,\ \sigma,\ r,\ T) \approx log(1 + P_{American}(V_0,\ K,\ \sigma,\ r,\ T)) \tag{19}$$

where Z_t refers to specific simulated discrete increment, $\widetilde{\beta}$ refers to regression coefficients and $f_\theta(\cdot)$ refers to the corresponding Deep Neural Network approximation surrogate function.

4. Research results

The classification results are listed in Table 4 as follows:

Table 4. Comparative evaluation metric framework results

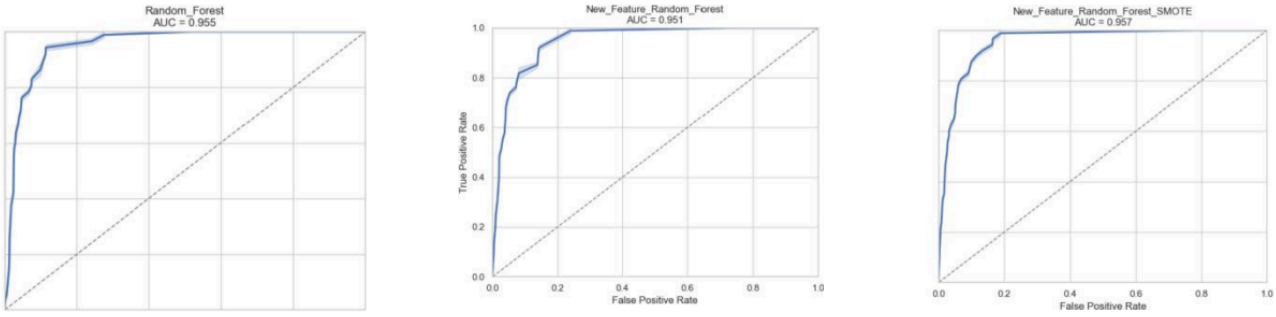
Model	AUC	F2	Precision	Recall	Avg. precision	Accuracy
Logistic Regression	0.9498	0.5577	0.3452	0.6591	0.3592	0.9487
Decision Tree	0.9124	0.5054	0.2772	0.6364	0.2309	0.9348
Random Forest	0.9548	0.5533	0.3971	0.6136	0.3451	0.9575
Bagging Tree	0.9503	0.5556	0.3191	0.6818	0.3569	0.9428
Support Vector Machine	0.9459	0.5020	0.3425	0.5682	0.3485	0.9509
XGBoost	0.9435	0.4506	0.3684	0.4773	0.3729	0.9567
Hard-voting Ensemble	-	0.5833	0.2823	0.7955	-	0.9282
K-means	-	0.0000	0.0000	0.0000	-	0.9677
Isolation Forest	0.7652	0.1570	0.1489	0.1591	0.1765	0.9435
One-class SVM	0.6197	0.1055	0.0820	0.1136	0.1053	0.9304
Proxy-augmented RF	0.9508	0.5336	0.3506	0.6136	0.3790	0.9509
Proxy + SMOTE RF	0.9574	0.5534	0.3636	0.6364	0.4198	0.9523

From results, this paper concludes that Original Random Forest and corresponding two improved versions rank top three considering their excellent and robust performances in a comprehensive evaluation metric framework. The Random Forest is selected as the anchor model because it offers the strongest baseline AUC-precision combination: AUC is 0.9548, precision is 0.3971, and recall remains 0.6136. The choice is therefore based on balance and on the ability to produce feature importance, but not on a claim that random forest wins every column. The cluster and anomaly results show why accuracy can mislead. K-means records 0.9677 accuracy by assigning every test case to the non-bankrupt class, leaving recall at zero. Though, it is noteworthy that Hard-voting Ensemble achieves an extraordinary recall rate, which is better than single Random Forest and all its variations. However, the excellent performance builds partly on the sacrifice of accuracy and precision rate and is not sufficiently stable or interpretable within a wider range of a more balanced evaluation metric framework rather than a framework only focusing on the recall rate. Holding a

DOI: 10.54254/2754-1169/2026.LD37672

comprehensive and synthetic horizon, this paper eventually decides to choose Random Forest family as the baseline model for subsequent feature construction and oversampling.

After LASSO is rerun on the expanded pool with the random forest retuned, the resulting 25-feature specification without SMOTE changes the metric profile rather than improving it uniformly. Average precision rises from 0.3451 to 0.3790 and recall remains 0.6136, but AUC falls from 0.9548 to 0.9508 and precision falls from 0.3971 to 0.3506. Because the selected accounting variables and hyperparameters can also change, this is a comparison of complete specifications rather than an isolated effect of the two retained proxies. The proxy-augmented specification with SMOTE has the strongest ranking results among the three random-forest variants. AUC reaches 0.9574, average precision reaches 0.4198, and recall rises to 0.6364. Relative to the original random forest, average precision increases by 0.0747 and recall by 2.27 percentage points. Individual Receiver Operating Characteristic Curves of the 3 models are plotted in Figure 2.



(a) Original Random Forest (b) Proxy-augmented RF (c) Proxy + SMOTE RF

Figure 2. ROC curves of outperformed models (picture credit: original)

To measure more detailed influence of the constructed American Default Put Price, the distribution curves by bankruptcy categories are plotted in Figure 3, and the corresponding feature importance matrix is depicted in Table 5.

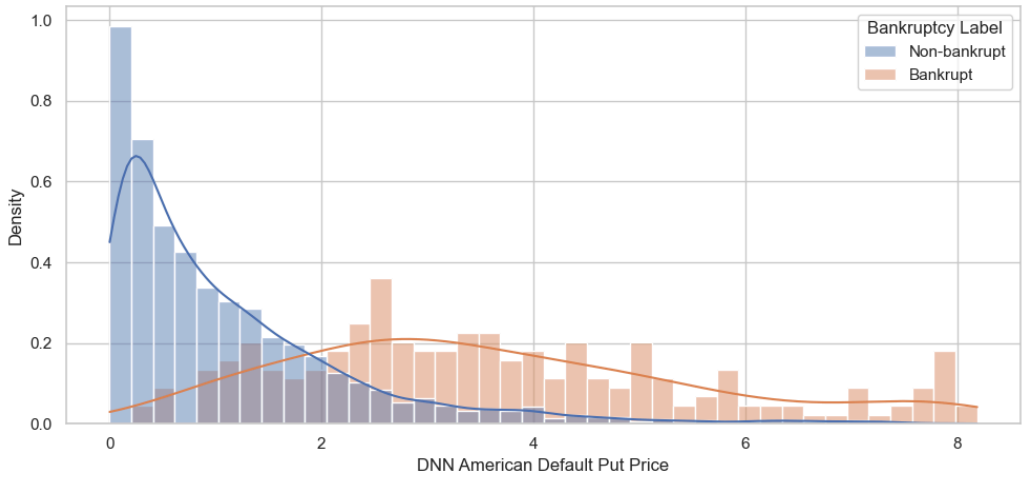


Figure 3. Diversification of distribution curves for American Default Put (picture credit: original)

DOI: 10.54254/2754-1169/2026.LD37672

Table 5. Feature importance matrix of structural credit + Random Forest

Feature Importance Coefficient	Name of Partly Selected Features
0.1823	American Default Put Default Probability
0.1352	American Default Put Price
0.1041	Net Income to Total Assets
0.1027	Non-industry income and expenditure/revenue
0.0946	Net Income to Stockholder's Equity
0.0731	Total income/Total expense
0.0499	Current Liability to Current Assets
0.0366	Cash/Current Liability
0.0293	Operating Profit Per Share (Yuan ¥)
0.0257	Equity to Long-term Liability

Diversification of distribution curves, along with high feature importance, prominently supports the view that the constructed put price in a structured credit view plays a crucial role in bankruptcy prediction and improves model performance by preserving a current precision level, remarkably raising average precision and achieving a higher recall rate.

5. Conclusion

This study is meanwhile subject to several limitations. Firstly, the data set of this paper is cross sectional from 1999 to 2009, thus not able to reflect time-varying changes for asset paths. Moreover, the study assumes a constant interest rate and ignores possible variations throughout the time period. Finally, the study directly calculates credit risk proxies without referring to equity option market for calibration. To further extend and explore detailed influence of the new feature, future research can include in panel data and extend the static classification framework to a structural credit survival model to provide dynamic views of bankruptcy probabilities.

In aggregate, findings of this paper indicate that more efforts should be put in ensemble models, e.g. Random Forest, in Taiwanese bankruptcy prediction cases. SMOTE plays a substantial role in raising recall and average precision rates as an effective over-sampling synthesized method. American default put price feature also proves to provide a possibility for credit-adjusted bankruptcy risk valuation from a novel horizon of structural credit analysis.

References

[1] Pham, H. V., Chu, T., Le, T. M., Tran, H. M., Tran, H. T. K., Yen, K. N., & Dao, S. V. T. (2025). Comprehensive evaluation of bankruptcy prediction in Taiwanese firms using multiple machine learning models. *International Journal of Technology, 16*(1), 289-309. https://doi.org/10.14716/ijtech.v16i1.7227

[2] Altman, E. I. (2018). Applications of distress prediction models: What have we learned after 50 years from the Z-score models? *International Journal of Financial Studies, 6*(3), Article 70. https://doi.org/10.3390/ijfs6030070

[3] FitzPatrick, P. J. (1932). A comparison of the ratios of successful industrial enterprises with those of failed companies. *The Certified Public Accountant, 12*, 598-605, 656-662, 727-731.

[4] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589-609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

[5] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006

[6] Wang, H., & Liu, X. (2021). Undersampling bankruptcy prediction: Taiwan bankruptcy data. *PLOS ONE, 16*(7), e0254030. https://doi.org/10.1371/journal.pone.0254030

DOI: 10.54254/2754-1169/2026.LD37672

[7] Sun, C. (2023). Company bankruptcy prediction with machine learning techniques. In G. Grigoras & P. Lorenz (Eds.), *Advances in artificial intelligence, big data and algorithms: Proceedings of the 3rd International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA 2023), 16-18 June 2023, Zhengzhou, China (Frontiers in Artificial Intelligence and Applications, Vol. 373*, pp. 425-437). IOS Press. https://doi.org/10.3233/FAIA230838

[8] Dam, T. T., & Dang, X. T. (2026). Cost-sensitive ensemble learning for bankruptcy prediction under extreme class imbalance. *Discover Artificial Intelligence, 6*(1), Article 473. https://doi.org/10.1007/s44163-026-01266-4

[9] Garcia, J. (2022). Bankruptcy prediction using synthetic sampling. *Machine Learning with Applications, 9*, Article 100343. https://doi.org/10.1016/j.mlwa.2022.100343

[10] Gnip, P., Kanász, R., Zoričak, M., & Drotár, P. (2025). An experimental survey of imbalanced learning algorithms for bankruptcy prediction. *Artificial Intelligence Review, 58*(4), Article 104. https://doi.org/10.1007/s10462-025-11107-y

[11] Amirshahi, B., & Lahmiri, S. (2024). Bankruptcy prediction using optimal ensemble models under balanced and imbalanced data. *Expert Systems, 41*(8), e13599. https://doi.org/10.1111/exsy.13599

[12] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182.

[13] Pulat, M., & Kocakoç, İ. D. (2024). Classification with machine learning algorithms after hybrid feature selection in imbalanced data sets. *Operations Research and Decisions, 34*(4), 157-183. https://doi.org/10.37190/ord240410

[14] Yang, C.-C., Deng, J., Deng, L.-Y., & Lu, H. H.-S. (2026). Random integrated subdata ensemble method for key variable selection in rare event setting. *Journal of Forecasting, 45*(4), 1878-1889. https://doi.org/10.1002/for.70120

[15] Li, L., Wang, H., Zha, L., Huang, Q., Wu, S., Chen, G., & Zhao, J. (2023). Learning a data-driven policy network for pre-training automated feature engineering. In *The Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 1-5 May 2023. https://openreview.net/forum?id=688hNNMigVX

[16] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance, 29*(2), 449-470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x

[17] Charalambous, C., Martzoukos, S. H., & Taoushianis, Z. (2023). A neuro-structural framework for bankruptcy prediction. *Quantitative Finance, 23*(10), 1445-1464. https://doi.org/10.1080/14697688.2023.2230241

[18] Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics, 12*(3), 591-612. https://doi.org/10.1080/00401706.1970.10488699

[19] Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks, 6*(6), 861-867. https://doi.org/10.1016/S0893-6080(05)80131-5