Use weighting methods instead of SMOTE for higher out-of-time performance in imbalanced bank failure prediction models!

Bank Failure Prediction Models Addressing Imbalanced Data and Out-of-Time Performance

Seyma GUNONU
@seyma__gunonu
seymagunonu@gmail.com
Gizem ALTUN
@GizemAltn99
gizemaltn99@gmail.com
Mustafa CAVUS
@mustafa___cavus
mustafacavus@eskisehir.edu.tr
Department of Statistics, Eskisehir Technical University

Introduction

Banks are crucial to the financial system and must operate effectively. Recently, bank failure models have been used to predict the likelihood of bank failures by analyzing relevant metrics. In addition to ensuring high out-of-sample performance, it is important to address the poor out-of-time performance of bank failure prediction models to maintain their utility `(Du Jardin & Séverin, 2011; Manthoulis et al., 2020)`. When using tabular data sets, deep learning models are used more than tree-based models `(Carmona et al., 2019); Petropoulos et al., 2020); Grinsztajn et al., 2022)`. As a difference, it is being examined what kind of results such data sets may encounter with Decision Trees, Random Forests `(Breiman, 2001)`, and Extra Trees `(Geurts et al., 2006)` models.

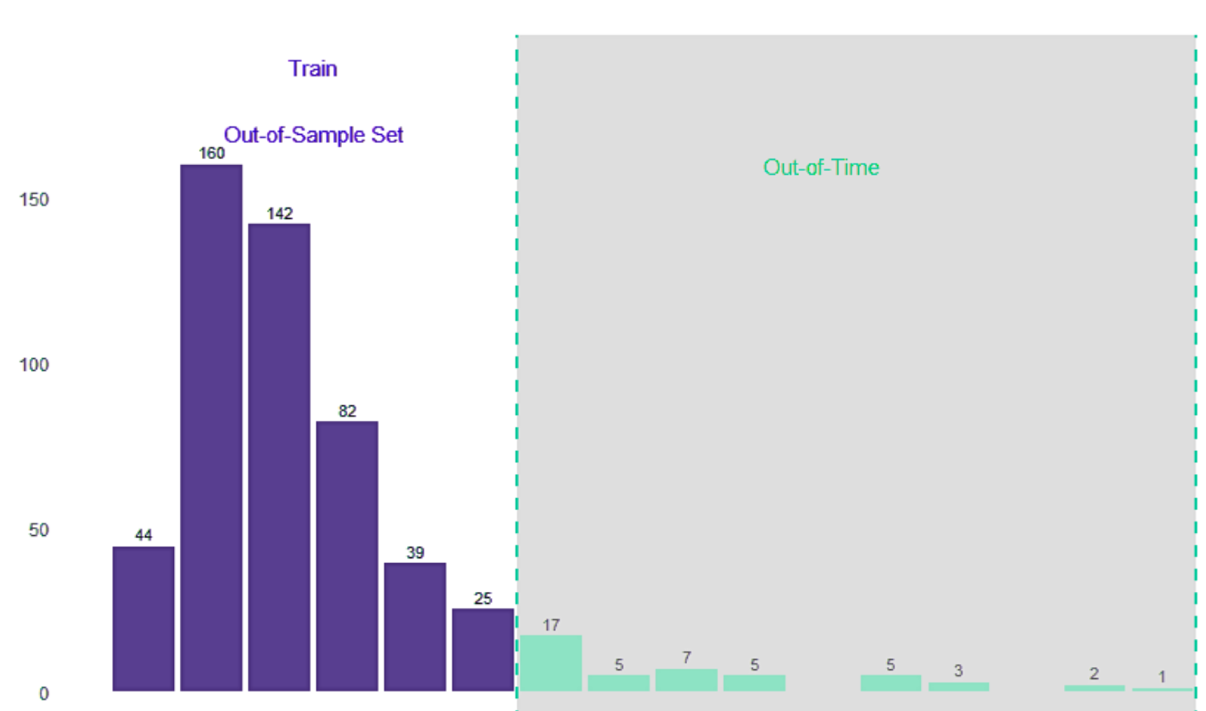

Figure 1: Failed Banks in the U.S. by years

The data has collected from the FDIC database using the `{fdicdata}` package in R `(Dar & Pillmore, 2023)`. It covers information about whether banks were active or closed within 15-year period from 2008 to 2023. When determining the time ranges in the data set, **in-sample** and **out-of-sample** were obtained between **2008-2014**, while models were builded using the **out-of-time** set between **2014-2023**. Figure 1 includes the banks that failed in the U.S. during these time ranges. CAMELS indicators (Capital, Asset Quality, Management Adequacy, Earnings, Liquidity, and Sensitivity to Market Risk) in the first as in `(Gogas et al. 2018)` and second variable groups as in `(Petropoulos et al. 2020)` were used. In the third variable group, different indicators that may be important were added. The variables in the data set are listed in detail in Figure 2 .

Variable Group	Name	Range	Description
1	TICRC	[-0.01, 0.19]	Tier 1 Risk-Based Capital Ratio / Total Assets
	PLLL	[-3, 10]	Provisions for Loan & Lease Losses / Total Interest Income
	TIE	[0, 2.2]	Total Interest Expense / Total Interest Income
	EQR	[-20, 100]	Equity Capital Ratio
2	TICRC	[-0.01, 0.19]	Tier 1 Risk-Based Capital Ratio / Total Assets
	NIMY	[-4, 26]	Net Interest Margin
	INTEXPYQ	[-0.5, 5.5]	Cost of Funding Earning Assets Quarterly
	RBC1AAJ	[-20, 200]	Leverage Ratio
	ROE	[-12000, 1000]	Return on Equity
3	NIMYQ	[-4, 26]	Net Interest Margin Quarterly
	LNATRESR	[0, 26]	Loan Loss Reserve / Gross Loan & Lease
	NONIXAYQ	[-20, 300]	Noninterest Expenses / Average Assets Quarterly
	ROAQ	[-100, 350]	Quarterly Return on Assets

Figure 2: Details of variables used in the data set

Methods

Three different models were employed in this study because they provides varying variance of predictions. When comparing the prediction variances of these three methods, it becomes apparent that the **Decision Trees** yields high-variance predictions, the **Random Forests** provides predictions with moderate variance, and the **Extra Trees** generate predictions with low variance `(Gogas et al., 2018)`. The structures of Decision trees, Random forests and Extra trees models are shown in Figures 3, and 4. Random Forests lies in aggregating predictions generated by multiple decision trees. Breiman improved upon the overfitting-prone CART method by introducing Random Forests, an extension of bagging trees. It differs by using feature subsets for each tree, reducing correlation. This added randomness enhances stability and generalization, making Random Forests valuable across applications `(Breiman, 2000)`.

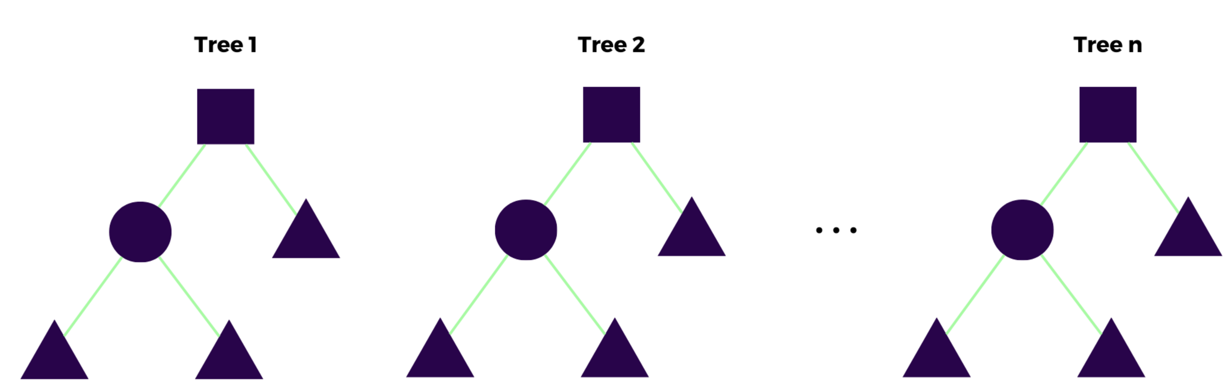

Figure 3: Random Forests Structure

Its two primary distinctions from Random Forests are that it splits nodes by randomly selecting cut-points and that it grows the trees using the entire learning sample `(Geurts et al., 2006)`.

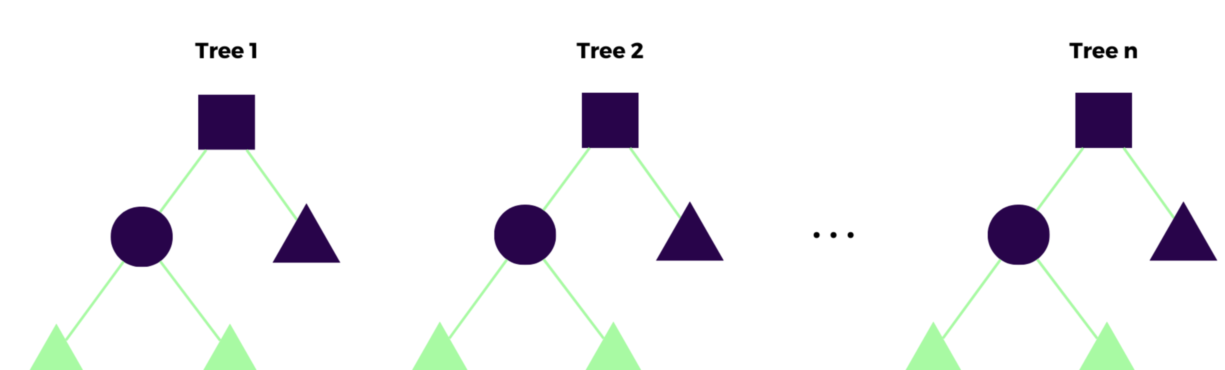

Figure 4: Extra Trees Structure

To address imbalanced data, resampling techniques like **undersampling** reduce majority class samples, **oversampling** increases minority class samples, and **SMOTE** creates synthetic minority samples for better representation in Figure 5.

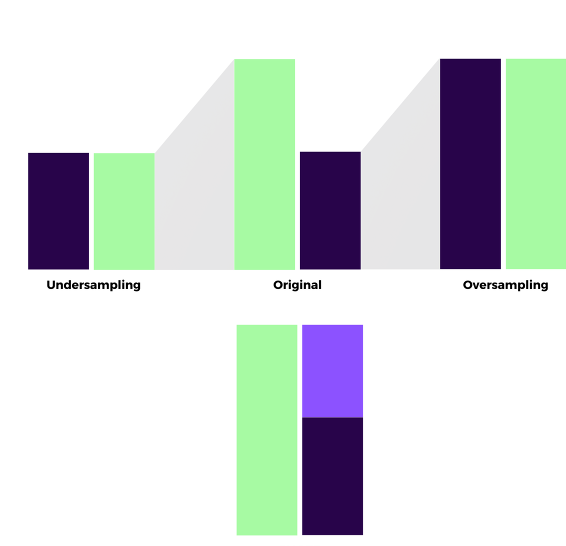

Figure 5: Resampling Techniques' Structures

Results

Three different variable groups were considered, and three different models were applied. The dataset exhibited an imbalance between the classes. So, imbalance in the data was addressed using various resampling methods. **Accuracy** and **F1 scores** were calculated for each variable groups to assess model performance. The results of out-of-sample showed that generally the weighted-based which is cost-sensitive method had the highest accuracy for all variable groups, while SMOTE had the lowest accuracy. Accuracy values were closer between variable groups in the out-of-time strategy, and some results were the same in random forests and extra trees. Variable groups with under-sampling had lower F1 values in the out-of-time strategy. As a result, **the choice of resampling method's effectiveness varied depending on the variable group and model. The out-of-time strategy is important to assess how models perform with changing data over time.**

Model	Variable Group	Accuracy					F1				
		Original	Under-sampling	Over-sampling	SMOTE	Weight-based	Original	Under-sampling	Over-sampling	SMOTE	Weight-based
Decision Trees	1	**0.9784**	0.9558	0.9577	0.9735	0.9705	**0.8981**	0.8207	0.8273	0.8831	0.8672
	2	**0.9852**	0.9823	0.9665	0.9626	0.9813	**0.9282**	0.9158	0.8521	0.8430	0.9132
	3	**0.9577**	0.8928	0.8968	0.8938	0.9381	**0.7860**	0.6472	0.6579	0.6516	0.7449
Random Forests	1	**0.9862**	0.9784	0.9853	0.9803	**0.9862**	**0.9333**	0.9027	0.9282	0.9082	**0.9333**
	2	**0.9872**	0.9803	**0.9872**	0.9823	**0.9872**	**0.9378**	0.9090	**0.9378**	0.9174	0.9383
	3	**0.9626**	0.9479	**0.9626**	0.9518	0.9607	0.8224	0.7871	**0.8303**	0.7967	0.8245
Extra Trees	1	**0.9872**	0.9764	0.9862	0.9813	0.9843	**0.9372**	0.8928	0.9326	0.91163	0.9245
	2	0.9872	0.9842	**0.9882**	0.9823	**0.9882**	0.9372	0.9259	**0.9423**	0.9174	**0.9423**
	3	**0.9607**	0.939	**0.9607**	0.9518	0.9587	0.8148	0.7596	**0.8198**	0.7967	0.8157

Figure 6: Accuracy and F1 values for out-of-sample with three different models for each variable group

Model	Variable Group	Accuracy					F1				
		Original	Under-sampling	Over-sampling	SMOTE	Weight-based	Original	Under-sampling	Over-sampling	SMOTE	Weight-based
Decision Trees	1	0.9936	0.9689	0.97	0.9592	**0.9957**	0.40	0.3255	0.3333	0.2692	**0.75**
	2	**0.9989**	**0.9989**	0.9602	0.9677	**0.9989**	**0.9230**	0.9230	0.2448	0.2857	**0.9230**
	3	0.9946	0.9571	0.9356	0.9635	**0.9968**	0.5454	0.2307	0.1666	0.2608	**0.7692**
Random Forests	1	0.9957	0.9356	**0.9979**	0.9871	0.9968	0.6666	0.1891	**0.8571**	0.5384	0.80
	2	**0.9989**	0.9699	**0.9989**	0.9925	**0.9989**	**0.9230**	0.30	**0.9230**	0.6315	**0.9230**
	3	0.9946	0.9485	**0.9957**	0.9861	**0.9957**	0.5454	0.20	**0.6666**	0.4347	**0.6666**
Extra Trees	1	0.9946	0.9356	0.9345	0.9871	**0.9957**	0.5454	0.1891	0.1866	0.5384	**0.6666**
	2	**0.9989**	0.9828	**0.9989**	0.9925	**0.9989**	**0.9230**	0.4285	**0.9230**	0.6315	**0.9230**
	3	**0.9946**	0.9614	**0.9946**	0.9861	**0.9946**	**0.5454**	0.28	**0.5454**	0.4347	**0.5454**

Figure 7: Accuracy and F1 values for out-of-time with three different models for each variable group

Conclusion

In study focuses on using Decision Trees, Random Forests, and Extra Trees to make bank failure predictions. What makes this study apart is the usage of a **1-year lag (t-1) period** in the dataset. When examining the overall results of the models, **it was observed that Random Forests and Extra Trees yielded similar and high-quality results, making them the most effective models for predicting bank failures.**

References

Breiman, L. (2000). Some infinity theory for predictor ensembles. Technical Report 579, Statistics Dept. UCB.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. Int. Rev. Econ. Finance, 61, 304-323.

Dar, U., & Pillmore, B. (2023). fdicdata: Accessing FDIC Bank Data. R package version 0.1.0.

Du Jardin, P., & Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. Decis Support Syst, 51(3), 701-711.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63, 3-42.

Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach. Int. J. Forecast., 34(3), 440-455.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data?. Adv. Neural Inf. Process, 35, 507-520.

Manthoulis, G., Doumpos, M., Zopounidis, C., & Galariotis, E. (2020). An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks. Eur. J. Oper. Res., 282(2), 786-801.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. Int. J. Forecast., 36(3), 1092-1113.

ESKİŞEHİR TECHNICAL UNIVERSITY

IRSYSC 2023