U.S Bank Failure Predictions – Testing Probabilistic Neural Networks Across Feature Selection Algorithms

Wyatt Schwanbeck

Davenport University College of Technology

CSCI 794: Masters Project

Dr. Samer Hanoudi

December 15th, 2023

Abstract

U.S Bank failures impact local and world economies systemically and early warning signs within filed quarterly FDIC reporting can be extracted and analyzed using machine learning algorithms to determine the likelihood of failure. The probabilistic neural network is unique in that it can provide a standard deviation with which predictions are generated. As far as we are aware, the PNN has yet to be tested to predict bank failures within published research. This work tests the capability of multiple machine learning algorithms and feature selection algorithms to predict the likelihood of bank failures within 6 months of published quarterly reports. The probabilistic neural network is tested alongside an artificial neural network, random forest regressor, gradient boosting regressor, and k nearest neighbors regressor. Random forest Boruta (RFB) and two embedded recursive feature elimination algorithms are used with a gradient boosting regressor (RFE-GBR) and random forest regressor (RFE-RFR). All machine learning algorithms were then trained on datasets compiled from three feature selection algorithms to predict the likelihood of failure using all failed and non-failed U.S FDIC insured state banks from December 1992 to June 2022 and tested to predict bank failures occurring between December 2022 and July 2023. RFB and RFE-RFR selected features produced the lowest combined mean absolute error using the PNN for non-failing and failing banks. Classification metrics are also measured for test and train datasets using an imbalanced ratio of 10 non-failed banks to 1 failed bank across 25 samples, the PNN trained using an ensemble dataset of 37 features produced a training recall of 98%, train precision of 31%, and train f-measure of 47%. Further, the PNN produced a test recall of 100%, test precision of 34%, and an F-Measure of 51%. With lower precision being more acceptable given the disproportionate costs associated with false negatives in comparison to false positives.

Introduction

During a bank failure, insured accounts may only experience illiquidity, while uninsured accounts and shareholders of institutions may lose substantial amounts of wealth. Further, the contraction in the supply of bank credit has significant and permanent effects on real economic activity (Ashcraft, 2003). The bankruptcy of one bank can even propagate to other banks in the financial system due to pervasive and cumulative effects to lender confidence (Anari et al, 2005). Banks are subjected to both financial and non-financial risk. Non-financial risk may be present in the form of operational risks, security risks, and reputational risks. The focus of this research pertains to the estimation of financial risk across FDIC insured state-chartered banks using probabilistic neural networks. Information regarding financial risk is plentiful, publicly available, and measurable through liquidity, credit, and market sensitivity metrics. These metrics can be used to identify institutions susceptible to sudden types of failure where there may be insufficient cash for customer withdrawals as well as gradual failures of non-profitable institutions.

Prior Research

Due to associated risks from illiquidity and loss of uninsured wealth, it is reasonable for individuals and companies to scrutinize the stability of the banking institutions in which their wealth resides. Most recently, Silicon Valley Bank's failure propagated to Signature bank (failing March 12th, 2023) and eventually First Republic bank (failing May, 2023). Although the FDIC covered all deposits over the standard $250k deposit insurance limit, these failures alone amounted to an estimated total cost of $34 billion dollars due to the subsequent loss of shareholder capital and government intervention. The United States regulates commercial and savings banks through the Federal Deposit Insurance Corporation using composite scores which are not made available publicly. The focus of this research tests classification and regression-based feature selection tools for bank failure determinants across multiple machine learning models. Feature selection methods include Random Forest Boruta (RFB), a classification-based feature selection model, and recursive feature elimination (RFE) which is regression based. RFB is a feature selection wrapper over random forest classification (Kursa & Rudnicki, 2010). RFE is an embedded feature selection estimator that recursively considers smaller sets of features to prune important features over a machine learning model. RFE is tested with gradient boosting regression (RFE-GBR) and random forest regression (RFE-RFR) using SciKitLearn packaged machine learning models. Selected features from these methods are then used to train machine learning algorithms using a probabilistic neural network, random forest classification model, and an artificial neural network.

Liu et al synthesized existing research in the context of predicting bank distress using machine learning and found an average accuracy rate of 95% with approximately half of existing research in the field utilizing data from the U.S. Further, several studies have tested and confirmed the usefulness of variables associated with the bank rating system used by the U.S Federal Reserve, CAMELS. This rating system consists of metrics associated with capital adequacy, asset quality, management ability, earnings liquidity, and market sensitivity. Although CAMELS ratings generated by the FDIC are not available to the public, researchers have proved the predictive power of related variables using probit models (Kerstein & Kozberg, 2013), and multivariate logistic regression (Cole & White, 2012). One of the earliest models for predicting corporate bankruptcy likelihood was Ohlson's O-score which used 11 features as an input to a formula with fixed coefficients (Ohlson, 1980).

Gogas et al (2018) demonstrated SVMs surpass the accuracy of Ohlson's O-score across 481 failed U.S banks. Le & Vivani (2018) also found machine learning techniques, artificial neural networks & k nearest neighbors' regression, to excel against SVM and traditional statistical techniques when analyzing 1438 bank failures from 3000 U.S banks. Le & Vivani (2018) used 31 financial ratios selected manually based on the CAMELS framework with a 1 to 1 ratio of failed to non-failed training and test examples. Gogas et al (2018) used a two-step procedure for feature selection by applying a local learning-based feature selection algorithm followed by a final selection process of iterative shrinking to select 2 ratios. Using only 2 ratios: Tier 1 risk-based capital over assets and interest expense over interest income, Gogas et al produced an accuracy of 98% in test samples. However, it is important to note their results do not cover liquidity failures like Signature and Silicon Valley bank, but they did implement a more realistic test dataset balance with 10 solvent banks for each insolvent bank.

Probabilistic neural networks (PNNs) have already been tested with the task of predicting bankruptcy across 38 U.S oil company financial reports (Yang et al, 1999). PNNs were found to produce the highest classification accuracy at the cost of classifying bankrupt companies as stable. The output of the probabilistic neural network is a distribution with a mean and standard deviation which can further be used to determine confidence intervals of predictions. Yang et al found linear discriminant analysis ultimately produced more accurate results in classifying bankrupt companies while balancing type 1 and type 2 error. Understandably, their research limited predictions to only 6 features across 38 samples as PNNs operate with all training records in the activations of their hidden layers and require a significant amount of computational power (Kusy & Kowalski, 2022). There is further opportunity to test the PNN's ability to predict banking industry failures across larger available datasets and feature selection algorithms as computer hardware has improved significantly.

Method

With a greater amount of training data and publicly available from the FDIC's Financial reporting API, there is an opportunity to train and test machine learning algorithms across several datasets compiled from various feature selection algorithms. For example, Random Forest Boruta (RBF), has not been tested in the context of predicting bank failures as far as we are aware within public literature. However, RBF has been used to identify the optimal set of important predictors for improving results on regression problems like evapotranspiration forecasting (Jayasinghe et al, 2021) and risk-based capital forecasting for life insurance companies (Park & Shin, 2023).

2200 features across all available quarterly banks reports from the US FDIC were retrieved for over 9000 individual banks since 1992. Ratio features were extracted based on the YAML formula specification of the FDIC API. Regression prediction targets for RFE-RFR and RFE-GBR are calculated by the difference of days from the reporting date to the failure date divided by 6 months. As the quarterly report date approaches the failure date, the target (Y) approaches 1. Classification prediction targets are calculated as a 1 or 0 based on whether the respective bank would fail within 6 months of a given quarterly financial report or not. Rank 1 variables from each feature selection algorithm are merged into the next available set of variables for testing until all variables have been tested. Table 1 shows the total selected features by feature selection algorithm.

Feature Selection Algorithm	**Total Selected Features**
Random Forest Boruta (RFB)	65
Recursive Feature Elimination - Random Forest Regressor (RFE-RFG)	72
Recursive Feature Elimination - Gradient Boosted Regressor (RFE-GBR)	26

Table 1- Count of selected features by feature selection algorithm out of 736 available financial ratios across all U.S based FDIC Insured banks between December 1992 and June 2023.

The Probabilistic Neural Network (PNN) model architecture is a feed-forward neural network reconfigured using a Keras Tensorflow-Probability example. PNN architecture was reconfigured for the regression problem of estimating the likelihood of bank failures and further captures aleatoric uncertainty present within training inputs since it generates a distribution prediction. With the generated distribution predictions, the standard deviation can be calculated for how certain a prediction is. A Laplacian prior distribution is initially specified over the weights of the network since the Laplacian distribution has tails heavier than the Normal distribution, making it more suitable for the modeling of financial returns (Vinod & Rao, 2019). During training, the model's posterior weights, a multivariate normal triangular distribution, is learned from the inputs. The multivariate triangular distribution allows for multiple correlations between weights and feature inputs using a single fully connected layer of size 128.

Results

For comparison against the PNN across all feature selection algorithms, a random forest regressor (RFR), gradient boosting regressor (GBR), and a densely connected artificial neural network (ANN) are all hyper-parameter tuned and trained using up to 50 stratified folds. The PNN was further trained using a fold sample of 10 to 1 for non-failing banks to failing banks over up to 50 folds. Stratification of training samples produced less accurate results for the PNN but all other machine learning algorithm performance was lesser using un-balanced training data. Further, the PNN and ANN used standard scaler pre-processing for each reporting quarter; RFR and GBR produced more accurate results without scaling. Figure 1 summarizes the test results of actual bank failures between September 2022 and March 2023. Only 5 failed banks exist within the test period compared to 3471 non-failing bank state charter bank peers. Due to this imbalance, feature selection datasets and machine learning models are tested across failing and non-failing banks. The sum MAE for both actual bank failures and non-failing banks is used as the performance metric with lower combined MAE indicating better overall performance for the PNN using RFB and RFE-RFR.

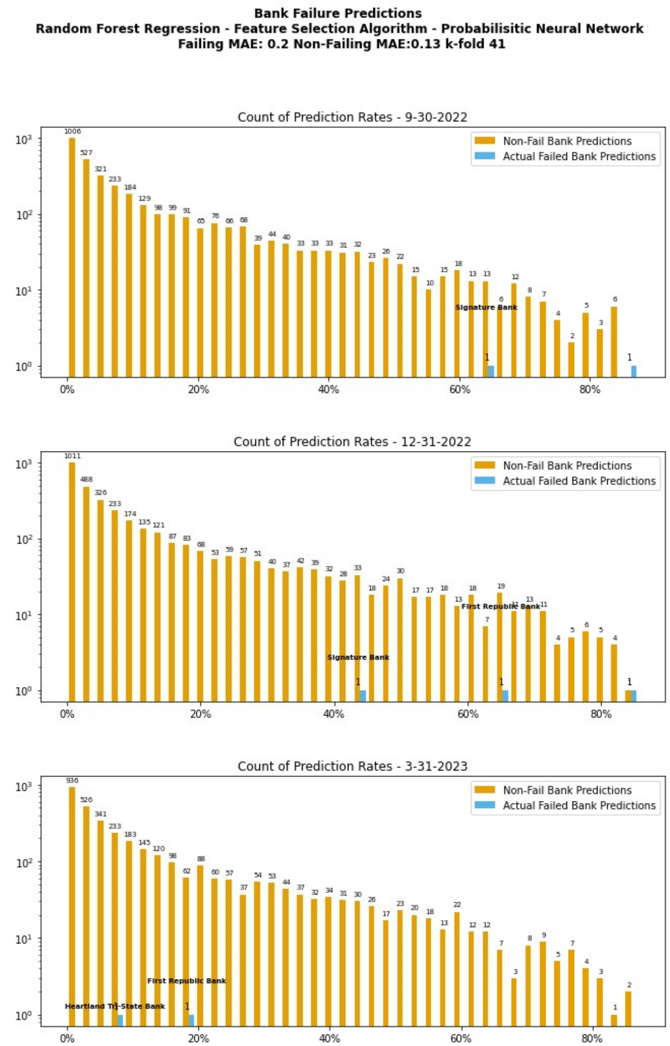
Machine Learning Model Performance by Feature Selection Algorithm Dataset
9/30/2022 – 3/31/2023

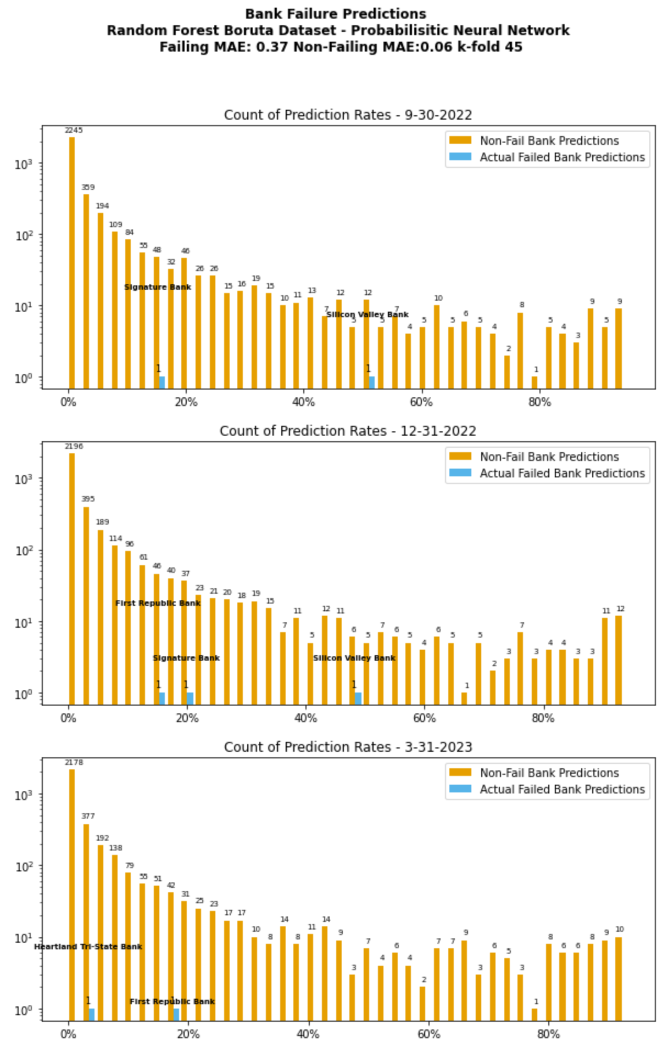
Mean Absolute Error for Failing Banks					
Feature Dataset	*Probabilistic Neural Network (PNN)*	*Artificial Neural Network (ANN)*	*Gradient Boosting Regressor (GBR)*	*Random Forest Regressor (RFR)*	*K Neighbors Regressor (KNR)*
Recursive Feature Eliminator – Gradient Boosted Regressor	29%	48%	56%	57%	54%
Recursive Feature Eliminator – Random Forest Regressor	20%	53%	56%	56%	45%
Random Forest Boruta	37%	48%	56%	56%	53%

Mean Absolute Error for Non-Failing Banks					
Feature Dataset	*Probabilistic Neural Network (PNN)*	*Artificial Neural Network (ANN)*	*Gradient Boosting Regressor (GBR)*	*Random Forest Regressor (RFR)*	*K Neighbors Regressor (KNR)*
Recursive Feature Eliminator – Gradient Boosted Regressor	17%	8%	2%	2%	1%
Recursive Feature Eliminator – Random Forest Regressor	13%	5%	3%	3%	13%
Random Forest Boruta	6%	10%	2%	3%	8%

Sum of Mean Absolute Error for Non-Failing and Failing Banks					
Feature Dataset	*Probabilistic Neural Network (PNN)*	*Artificial Neural Network (ANN)*	*Gradient Boosting Regressor (GBR)*	*Random Forest Regressor (RFR)*	*K Neighbors Regressor (KNR)*
Recursive Feature Eliminator – Gradient Boosted Regressor	46%	56%	58%	59%	55%
Recursive Feature Eliminator – Random Forest Regressor	33%	58%	59%	59%	58%
Random Forest Boruta	43%	58%	58%	59%	61%

Tables 2, 3, 4 - Mean Absolute Error for Failing Banks (top) and non-failing banks (middle) and combined total error (bottom) by machine learning algorithm and dataset. The probabilistic neural network produces the most accurate results based on the sum of group MAE compared to all other machine learning model algorithms.





Figures 1 & 2 - RFE-RFR (Left) and RFB (Right) datasets used to train PNN produced the lowest cumulative MAE. 36 Similar features from these feature selection algorithms were used to further refine descriptive power in predictions as demonstrated in figure 3.

Based on these results, an ensemble dataset using classification target predictions was generated using 36 similar features selected by RFE-RFR and RFB feature selection algorithms. Figure 3 shows the resulting overall MAE is higher but also shows the PNN trained using these ensemble features was found to produce more discriminant failure predictions against truly failing banks in comparison to peers. Inclusion of the regression output to a binary classification is based on whether the predicted mean failure is more than 50% and is selected to limit false negative classifications of failed banks as this error is significantly more costly than mislabeling a non-failed bank (Viviani & Le, 2018). To produce classification metrics comparable to Le & Viviani, 100 samples using a 10 to 1 ratio of non-failing bank predictions to each failing bank prediction are used to calculate the precision, recall, and f-measure of classifications. Where precision measures the ratio of predicted positive failures that are truly positive, recall is the fraction of truly failing banks that were correctly predicted positive. F-measure is a combined metric and is calculated as the harmonic mean. The PNN trained using the ensemble produced the best classification results between September 2022 and March 2023 which even includes Heartland Tri-State Bank (HTSB). This bank in particular may be considered an outlier in the test data as its failure was caused by an isolated event in which they were swindled by a single nefarious party (McCoy, 2023). Despite an imbalanced ratio of 10 to 1 non-failing to failed banks in classification prediction metrics, the overall recall was greater in test and train results at the expense of precision due to the disproportional costs associated with false negatives.

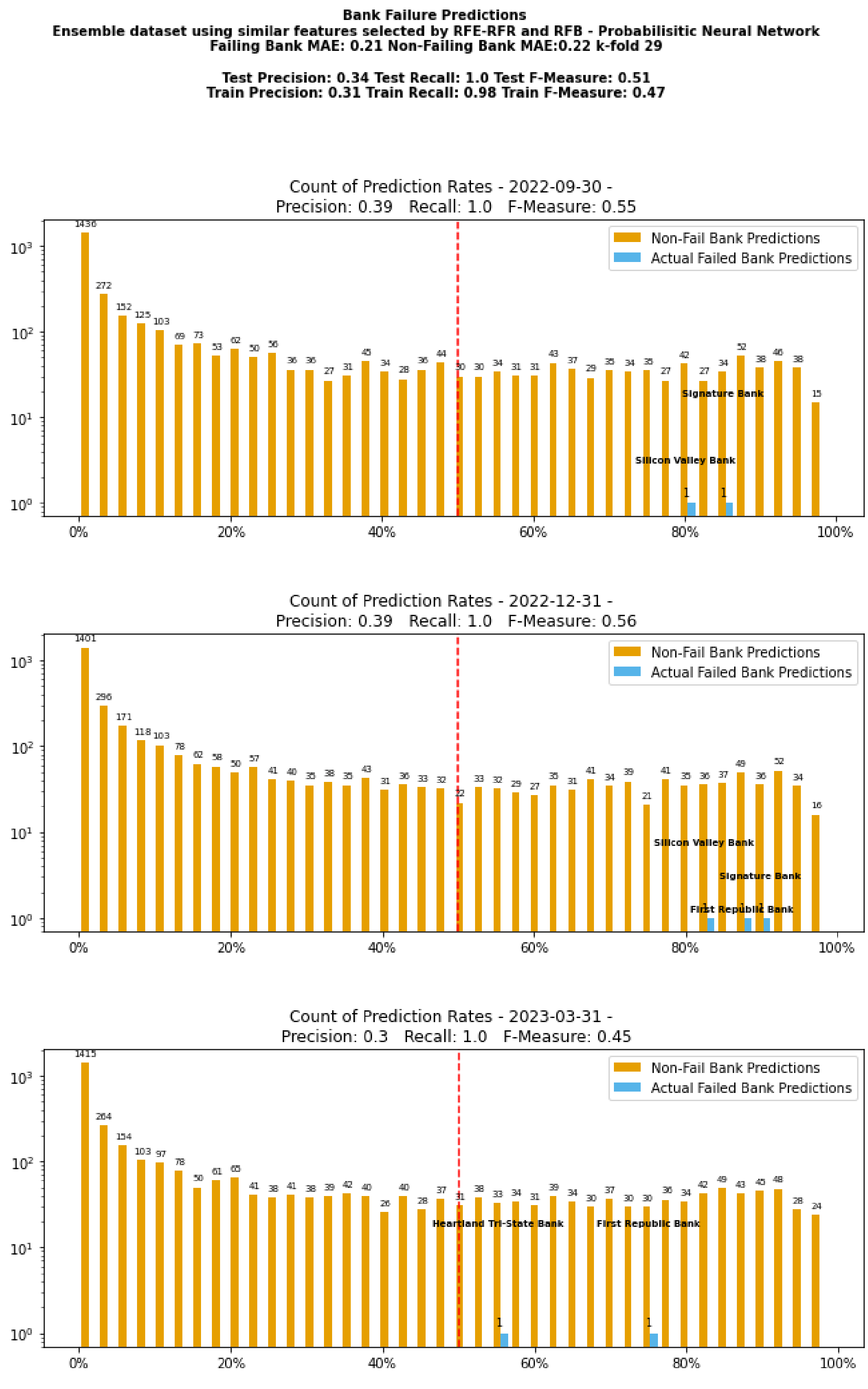


Figure 3 - PNN trained using the final compiled ensemble dataset with associated regression and classification metrics. Regression prediction mean absolute error is calculated holistically across each reporting quarter while classification metrics precision, recall, and f-measure are calculated over 25 samples with a 10 to 1 ratio of non-failed banks to failing banks.

Conclusion

The results highlight the effectiveness of probabilistic neural networks with respect to tested bank failures occurring between September 2022 and March 2023. The performance of PNNs trained on datasets compiled from feature selection algorithms, RFB and RFE-RFR, demonstrates a high degree of accuracy in unbalanced datasets for regression tasks. Datasets compiled using RFB and RFE-RFR produced the lowest cumulative MAE between 33% and 43% across non-failing banks and failing banks. Similar features selected between these datasets lead to the creation of a final ensemble dataset used to train the PNN. Results in Figure 3 demonstrate greater discriminant power against truly failing banks and show classification results using a more realistic balance of 10 to 1 between non-failing and failing banks. Although classification results sacrifice precision, false negatives are minimized with a 98% recall in training and 100% recall in the test period. Future work could test additional feature selection algorithms and may further improve precision and granularity of PNN predictions with respect to the type of predicted failure, failure by solvency and liquidity.

References

Ashcraft, A. B. (2003). Are banks really special? new evidence from the FDIC-induced failure of healthy banks. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.497102

Anari, A., Kolari, J. W., & Mason, J. R. (2005). Bank asset liquidation and the propagation of the U.S. Great Depression. *Journal of Money, Credit, and Banking*, *37*(4), 753–773. https://doi.org/10.1353/mcb.2005.0038

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, *36*(11). https://doi.org/10.18637/jss.v036.i11

Liu, L. X., Liu, S., & Sathye, M. (2021). Predicting bank failures: A synthesis of literature and directions for future research. Journal of Risk and Financial Management, 14(10), 474. https://doi.org/10.3390/jrfm14100474

Kerstein, J., & Kozberg, A. (2013). Using accounting proxies of proprietary FDIC ratings to predict bank failures and enforcement actions during the recent financial crisis. Journal of Accounting, Auditing; Finance, 28(2), 128–151. https://doi.org/10.1177/0148558x13478846

Cole, R. A., & White, L. J. (2011). Déjà vu all over again: The causes of U.S. commercial bank failures this time around. *Journal of Financial Services Research*, *42*(1–2), 5–29. https://doi.org/10.1007/s10693-011-0116-9

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109. https://doi.org/10.2307/2490395

Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2017). Forecasting bank failures and stress testing: A machine learning approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3039358

Le, H. H., & Viviani, J.-L. (2018). Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance*, *44*, 16–25. https://doi.org/10.1016/j.ribaf.2017.07.104

Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research*, *44*(2), 67–74. https://doi.org/10.1016/s0148-2963(97)00242-7

Kusy, M., & Kowalski, P. A. (2022). Architecture reduction of a probabilistic neural network by merging K-means and K-nearest neighbour algorithms. *Applied Soft Computing*, *128*, 109387. https://doi.org/10.1016/j.asoc.2022.109387

Jayasinghe, W. J., Deo, R. C., Ghahramani, A., Ghimire, S., & Raj, N. (2021). Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the boruta-random forest algorithm. *IEEE Access*, *9*, 166695–166708. https://doi.org/10.1109/access.2021.3135362

Park, J., & Shin, M. (2022). An approach for variable selection and prediction model for estimating the risk-based capital (RBC) based on machine learning algorithms. Risks, 10(1), 13. https://doi.org/10.3390/risks10010013

Vinod, H. D., & Rao, C. R. (2019a). 3.3 Multivariate Laplace. In Handbook of Statistics. essay, Elsevier. https://www.sciencedirect.com/topics/mathematics/laplace-distribution

Max McCoy, K. R. S. 24. (2023, September 24). *What took down this Kansas Bank? it's a secret, but what saved depositors isn't.* Kansas Reflector. https://kansasreflector.com/2023/09/24/what-took-down-this-kansas-bank-its-a-secret-but-what-saved-depositors-isnt/