UNIVERSITY OF MISKOLC
FACULTY OF ECONOMICS

SAFA SEN

Predicting Bank Failures: A Comparative Analysis of Traditional and Machine Learning Models in the Post-2008 Financial Landscape

Theses of the PhD dissertation

Miskolc, 2025

UNIVERSITY OF MISKOLC
FACULTY OF ECONOMICS
HANTOS ELEMÉR DOCTORAL SCHOOL OF BUSINESS, MANAGEMENT AND REGIONAL SCIENCES

SAFA SEN

Predicting Bank Failures: A Comparative Analysis of Traditional and Machine Learning Models in the Post-2008 Financial Landscape

Theses of the PhD dissertation

Supervisor: Dr. Levente Kovacs

Head of the doctoral school: Prof. Dr. Tóth Géza

Miskolc,2025

Contents

1.Introduction

1.1 Background of the Study

The stability of the banking sector is indispensable to the efficient functioning of modern economies. Banks act as critical financial intermediaries, channeling funds from savers to borrowers, supporting investment and consumption, and maintaining the liquidity necessary for sustainable economic development. Given this centrality, the failure of banks can have wide-ranging and severe repercussions, including the disruption of credit markets, erosion of depositors and investor confidence, and the potential triggering of systemic crises. The 2007–2009 Global Financial Crisis (GFC) served as a watershed moment, underscoring the fragility of financial institutions and the inadequacy of prevailing risk monitoring frameworks. The failure of several large banking institutions not only undermined global financial stability but also led to deep and prolonged macroeconomic contraction across developed and developing economies alike.

In the wake of the GFC, the academic and regulatory communities have renewed their focus on early warning systems for identifying financial distress within banks. Traditional approaches—largely based on financial ratios and linear classification models—have provided valuable insight into institutional performance but have often lacked the predictive precision and adaptability required in dynamic and complex financial environments. In particular, such models have struggled to detect emerging risks in time to support meaningful supervisory intervention, especially in contexts characterized by highly imbalanced datasets where failing banks constitute a statistical minority.

Against this backdrop, the present research investigates the potential of advanced machine learning techniques to enhance the accuracy and reliability of bank failure prediction. The core motivation is rooted in the need for robust, data-driven tools that can complement and enhance traditional risk assessment methods, offering greater sensitivity to emerging risks and the capacity to manage complex, nonlinear relationships within financial datasets. By comparing traditional statistical models with modern machine learning algorithms—including ensemble models and cost-sensitive adaptations, this study aims to identify predictive frameworks that can improve both the timeliness and effectiveness of early regulatory action.

The study is situated at the intersection of financial risk management, predictive analytics, and regulatory technology (RegTech). It addresses pressing questions around the effectiveness, interpretability, and policy applicability of machine learning models for early detection of bank failures. It also acknowledges the evolving role of financial supervision in a post-crisis world, wherein proactive, evidence-based intervention is essential not only for individual institutional oversight but also for preserving broader financial system stability.

1.2 Statement of the Problem

Despite the critical role of banks in facilitating financial intermediation and maintaining macroeconomic stability, accurately predicting bank failures remains a persistent and unresolved challenge. The limitations of traditional statistical models—such as logistic regression and linear discriminant analysis—have become increasingly evident in the wake of complex financial crises. These models often rely on linear assumptions, are sensitive to multicollinearity, and tend to perform inadequately when applied to highly imbalanced datasets where failing banks represent a small minority of cases.

A fundamental shortcoming of many existing prediction models lies in their insufficient ability to detect early warning signals of financial distress. In particular, traditional models tend to produce a high rate of Type II errors (false negatives), where at-risk banks are mistakenly classified as stable. This failure to identify distress in a timely manner undermines the effectiveness of regulatory oversight and delays crucial intervention, thereby increasing the likelihood of systemic disruptions.

Moreover, while more advanced predictive techniques—particularly machine learning algorithms—have shown promise in addressing nonlinearity, class imbalance, and high-dimensional data, their adoption in financial regulation has been limited. This is largely due to concerns regarding model complexity, lack of interpretability, and uncertainty about how such models balance performance with regulatory transparency.

This research seeks to address this multifaceted problem by systematically comparing the predictive performance of traditional statistical models and modern machine learning approaches—including Random Forest, Cost-Sensitive Forest, and Regularized Random Forest—within the domain of bank failure prediction. The study explores how these models perform under the constraints of imbalanced data, evaluates their capacity to reduce Type II errors, and assesses their practical suitability for integration into supervisory frameworks. In doing so, it responds to the broader need for reliable, transparent, and scalable early warning systems that can help regulatory bodies and financial institutions prevent the escalation of financial distress into systemic crises.

1.3 Research Objectives

The main objective of this study is to evaluate and compare the performance of traditional statistical models and advanced machine learning techniques in the prediction of bank failures, with a particular focus on managing class imbalance, minimizing classification errors (especially Type II errors), and understanding the trade-off between model complexity and interpretability. The research is grounded in post-crisis U.S. banking data and contributes to both academic literature and practical applications in financial supervision and regulatory decision-making.
Specifically, the study aims to test the following five hypotheses:

Hypothesis 1: Model Performance in Classification Accuracy
Research Objective:
To evaluate whether the Random Forest model significantly outperforms other machine learning models (including Support Vector Machines, Logistic Regression, and XGBoost) in terms of classification accuracy for predicting bank failures.
Hypothesis Statement:
H1: The Random Forest model does not significantly outperform other machine learning models (including Support Vector Machines, Logistic Regression, and XGBoost) in terms of classification accuracy for predicting bank failures.
Context and Justification:
This objective examines the performance of widely used models in the literature, benchmarking Random Forest's ability to deliver high accuracy and ROC-AUC scores in comparison to alternative models, using a comprehensive post-crisis dataset from the FDIC. By doing so, the study tests the assumption that Random Forest is among the most reliable models for binary classification problems in financial risk settings.

Hypothesis 2: Handling of Imbalanced Datasets
Research Objective:
To determine whether the CS-Forest model outperforms other ensemble models (e.g., AdaBoost, Gradient Boosting) in reducing Type II errors and managing class imbalance in bank failure prediction.
Hypothesis Statement:
H2: The CS-Forest model does not exhibit superior performance in reducing Type II errors and handling imbalanced datasets, such as those used in predicting bank failures, when compared to other hybrid ensemble models (e.g., AdaBoost, Gradient Boosting).
Context and Justification:
Accurate prediction in imbalanced datasets is critical, especially in early warning systems where missing a bank at risk (false negative) can result in regulatory failure. Cost-sensitive learning approaches are designed to reduce this risk by emphasizing the correct classification of the minority class (failed banks). This hypothesis tests whether CS-Forest, a cost-sensitive ensemble model, can effectively prioritize recall while maintaining acceptable overall accuracy.

Hypothesis 3: Effectiveness of Simplicity versus Complexity in Models
Research Objective:
To investigate whether simpler models (e.g., Naive Bayes) perform significantly worse than more complex models (e.g., MLP, Random Forest, ensemble methods) in the prediction of bank failures.
Hypothesis Statement:
H3: Simpler models, such as Naive Bayes, do not exhibit significantly lower effectiveness compared to more complex models, such as Multilayer Perceptron, Artificial Neural Networks, or ensemble methods, in predicting bank failures.
Context and Justification:
While simpler models are often preferred for their transparency and ease of use, they may be unable to capture the complex, nonlinear relationships found in financial datasets. This hypothesis evaluates the performance gap between simple and complex models and informs the discussion on the trade-off between interpretability and predictive accuracy in practice.

Hypothesis 4: Model Choice Influences
Research Objective:
To examine whether the selection of predictive models for bank failure detection is influenced by the trade-off between model complexity and interpretability, especially in regulatory contexts.
Hypothesis Statement:
H4: The selection of machine learning models for bank failure prediction is not predominantly influenced by the trade-off between model complexity and interpretability, nor by specific task requirements.
Context and Justification:
Regulatory environments require not only accurate predictions but also transparent decision-making processes. This objective investigates how complexity impacts adoption, with particular attention to the suitability of models such as Regularized Random Forest, which aim to balance predictive performance with explainability.

Hypothesis 5: Predictive Power and Model Complexity
Research Objective:
To assess whether a positive correlation exists between model complexity and predictive performance in the context of bank failure prediction.
Hypothesis Statement:
H5: The predictive power of machine learning models in bank failure prediction is not directly correlated with model complexity.
Context and Justification:
This hypothesis challenges the assumption that more complex models inherently yield better performance. By comparing a wide array of models across varying levels of complexity, the study examines whether model sophistication translates into significantly improved predictive outcomes and whether simpler alternatives might remain viable in specific scenarios.

1.4 Significance of the Research

The significance of this research lies in its dual contribution to the academic understanding and practical application of predictive modeling in the context of financial risk management and regulatory oversight. Bank failures are rare yet catastrophic events that can destabilize economies, erode investor and depositor confidence, and necessitate large-scale public intervention. The early and accurate identification of such failures is thus not merely a technical challenge but a systemic necessity. This study directly addresses this necessity by exploring innovative predictive methodologies capable of overcoming the limitations of conventional models.
From a theoretical perspective, the study contributes to the expanding field of financial machine learning by empirically validating the superior performance of ensemble models—such as Random Forest, Cost-Sensitive Forest (CS-Forest), and Regularized Random Forest (RRF)—over traditional statistical classifiers. It demonstrates the critical role of model design in handling imbalanced datasets, capturing complex interactions among financial indicators, and reducing the incidence of Type II errors. In doing so, the research not only supports the methodological advancement of financial distress prediction but also deepens the understanding of how interpretability, complexity, and predictive power interact in high-stakes analytical tasks.
In addition to its theoretical contributions, the research carries substantial practical relevance. Regulators, central banks, and financial institutions are increasingly tasked with implementing forward-looking supervisory frameworks capable of identifying and addressing institutional vulnerabilities before they evolve into systemic risks. However, existing risk assessment tools often fail to balance predictive accuracy with explainability, limiting their utility in real-world policy contexts. This study provides an empirically grounded evaluation of how different models perform in practice, offering clear guidance on their strengths, limitations, and optimal use cases.
Furthermore, the study responds to the growing interest in RegTech and macroprudential policy tools by proposing adaptable, transparent, and scalable predictive models that can be integrated into existing supervisory infrastructures. The insights gained from feature importance analyses—highlighting, for instance, the role of Net Interest Margin, Return on Equity, and Loss Allowance Ratios—equip regulators and risk managers with actionable knowledge to improve monitoring strategies and stress-testing frameworks.

Ultimately, this research contributes to the development of more robust and resilient financial systems by equipping key stakeholders with tools that are both analytically sophisticated and operationally relevant. Its findings are especially timely as global financial institutions navigate an increasingly complex, interconnected, and data-intensive regulatory environment.

2.Literature review

The body of academic research on bank failure prediction has evolved significantly over recent decades, encompassing traditional forecasting models, advanced machine learning (ML) techniques, and hybrid approaches that blend statistical robustness with algorithmic precision. These studies have contributed to both theoretical understanding and the development of practical tools for early warning systems (EWS), particularly in the context of financial crises.

2.1 Traditional Forecasting Models

Torna and DeYoung (2012) investigated the relationship between non-traditional banking activities and bank failure during periods of financial distress. Utilizing multi-period logistic regression, they found that fee-based activities such as securities brokerage and insurance sales were associated with a lower probability of failure, while activities involving venture capital, investment banking, and securitization significantly increased risk. Their exclusion of banks with assets exceeding $100 billion, however, limits the generalizability of their findings to the broader banking sector.

Berger and Bouwman (2012) examined the role of capital in determining bank performance and resilience during crises. Drawing on a comprehensive dataset from 1984 to 2010, and employing Logit survival and OLS regression models, they demonstrated that capital buffers reduce the likelihood of bankruptcy for small banks and enhance performance in medium and large banks during systemic stress. The authors emphasized capital as a dual-purpose tool—both a buffer against insolvency and a strategic asset.

Lu and Whidbee (2013) focused on financial fragility using logistic regression applied to a sample of 6,236 U.S. banks. They established that fragility—manifested through weak capital, asset quality, and liquidity—was a significant predictor of failure. Notably, multi-bank holding companies exhibited greater survival likelihood, underscoring the protective effects of institutional diversification.

Cleary and Hebb (2015) employed discriminant analysis to identify key predictors of bank failures between 2002 and 2009, achieving an accuracy rate of 92%, which remained consistent in subsequent years (2010–2011). Their work confirmed the methodological efficacy of linear classification techniques in financial failure contexts.

Chiaramonte et al. (2016) assessed the predictive utility of the Z-Score for U.S. commercial banks over the 2004–2012 period. Their findings indicated that the Z-Score could accurately predict 76% of failures, but the inclusion of macroeconomic variables did not improve predictive performance. This underscores the primacy of internal financial metrics over external economic factors in assessing bank risk.

2.2 Machine Learning and Hybrid Techniques

A shift towards more computationally intensive methodologies is evidenced in the work of Ekinci and Erdal (2017), who applied logistic regression, decision trees (J48), Voted Perceptron, and various hybrid ensemble models to the Turkish banking sector. Their results demonstrated the superior performance of hybrid models, particularly the RS-B-L ensemble, which achieved an AUC of 91.5%.

Le and Viviani (2017) further advanced the literature by comparing traditional statistical methods (Discriminant Analysis, Logistic Regression) with ML algorithms (Artificial Neural Networks, Support Vector Machines, k-Nearest Neighbors). Their study, based on 3,000 U.S. banks, found that ANN and k-NN offered the highest predictive precision, particularly in capturing non-linearities and complex interactions within the data.

Gogas et al. (2018) implemented a two-step feature selection process in conjunction with SVM to forecast bank failures across 1,443 U.S. institutions. Achieving a 99.22% classification accuracy, their model notably outperformed the Ohlson score, a widely used benchmark, thereby reinforcing the practical utility of ML for high-dimensional, imbalanced datasets.

Carmona et al. (2019) adopted XGBoost to model bank failures between 2001 and 2015. Their findings highlight the critical importance of retained earnings, return on assets, and total risk-based capital ratio in predicting failures. The authors recommended policy reforms focusing on earnings retention and dividend control during economic downturns.

Manthoulis et al. (2020) explored the predictive contribution of diversification variables within a combined statistical–ML framework. Their results revealed that ordinal classification models provided a more nuanced understanding of bank distress compared to binary classifiers, particularly for medium- to long-term forecasting horizons.

Momparler et al. (2020) introduced fsQCA (fuzzy-set Qualitative Comparative Analysis) to the bank failure literature. Their results revealed that configurations involving high non-performing loans, low capitalization, and insufficient risk coverage were consistently linked to failures, thus supporting a configurational view of financial distress.

Petropoulos et al. (2020) assessed the CAMELS framework using multiple methods including Logistic Regression, Linear Discriminant Analysis, SVM, Neural Networks, and Random Forests. They concluded that Random Forests consistently outperformed other models in both accuracy and interpretability, positioning them as effective tools for supervisory agencies.

Beutel et al. (2019) provided a cautionary assessment of ML applications, demonstrating that while ML models achieved high in-sample fit, traditional logit models outperformed them in recursive out-of-sample testing. The authors highlighted the need for hybrid models that balance accuracy with generalizability.

2.3 Time-Dynamic and Early Warning Models

Antunes et al. (2018) contributed to the development of systemic early warning systems by employing dynamic probit models enriched with exuberance indicators. Their work, based on four decades of European banking crises, provided significant improvements in both in-sample and out-of-sample prediction performance.

Plessis (2022) extended this strand of research by developing a Qualitative Vector Autoregression (Qual VAR) model using a Markov Chain Monte Carlo approach. His model, based on 40 years of cross-national banking crisis data, demonstrated superior early warning capability—up to 12 months in advance—compared to conventional probit-based systems. The model highlighted the banking sector variables as primary drivers of systemic risk.

2.4 Owner's Contribution to Literature

Building upon the aforementioned studies, this dissertation makes several key contributions. First, it introduces a novel dataset constructed from FDIC bank-level data (2007–2013), integrating 26 CAMEL-based financial indicators. Second, it evaluates 15 ML models, with a particular focus on ensemble methods such as Random Forest and CS-Forest. Notably, CS-Forest achieved a recall rate of 0.897, thereby outperforming previous benchmarks in minimizing false negatives. Third, it explicitly addresses class imbalance through cost-sensitive learning, aligning with recommendations from Chiaramonte et al. (2016) and Ekinci and Erdal (2017). Finally, the study quantifies the trade-off between model complexity and interpretability, advocating for the use of explainable ML models in supervisory applications, as suggested by Beutel et al. (2019) and Petropoulos et al. (2020).

Author(s) & Year	Focus	Methodology	Major Conclusion
Torna & DeYoung (2012)	Non-traditional income & bank failure	Logistic regression (multi-period)	Non-traditional fee income reduces failure risk; venture capital and securitization increase it
Berger & Bouwman (2012)	Bank capital's role in crises	Logit survival & OLS	Capital reduces bankruptcy risk in small banks; improves performance in larger banks during crises
Lu & Whidbee (2013)	Bank-level fragility & survival	Logistic regression	Financial fragility increases failure risk; multi-bank structures improve resilience
Cleary & Hebb (2015)	Bank failure predictors (2002–09)	Discriminant analysis	Achieved 92–95% prediction accuracy across time periods
Chiaramonte et al. (2016)	Z-Score to predict bank failures	Z-Score analysis	Z-Score predicted 76% of failures; macro factors did not improve forecasts

Author(s) & Year	Focus	Methodology	Major Conclusion
Ekinci & Erdal (2017)	ML prediction in Turkey	Hybrid ensembles, Logistic Regression, J48	Hybrid ensembles provided the highest accuracy (AUC: 91.5%)
Le & Viviani (2017)	Traditional vs ML in bank failure	DA, LR, ANN, SVM, k-NN	Artificial Neural Networks and k-NN were the most accurate methods
Gogas et al. (2018)	ML prediction with SVM	Feature selection + SVM	SVM achieved 99.22% accuracy and outperformed traditional models
Carmona et al. (2019)	XGBoost & failure prediction	XGBoost	Retained earnings, ROA, and capital ratio were key predictors of failure
Manthoulis et al. (2020)	Diversification & classification models	Ordinal classification, ML	Diversification improved long-term prediction; ordinal models offered more nuanced insights
Momparler et al. (2020)	fsQCA & causal configurations	Fuzzy-set QCA	Poor loan quality, low risk coverage, and low capital increased failure likelihood
Petropoulos et al. (2020)	CAMELS + multiple methods	LR, LDA, SVM, NN, RF	Random Forest was the most accurate and interpretable method
Beutel et al. (2019)	ML vs Logit for crises	Logit vs ML	Logit outperformed ML in out-of-sample predictions; ML tended to overfit
Antunes et al. (2018)	Early warning systems for banking crises	Dynamic probit	Exuberance indicators improved the accuracy of crisis forecasts
Plessis (2022)	Qual VAR model for continuous crisis indicator	Qual VAR (MCMC)	Banking sector indicators were most predictive; model gave 12-month early warnings

Source: Own construction

3.Study Sample:

This study uses detailed annual data on bank failures from the FDIC covering 2007–2013, focusing on financial conditions before and after the 2008 crisis. The author systematically collected financial ratios from failed and non-failed banks, meticulously extracting data from the FDIC database. For failed banks, the financial data from their last reporting period before failure was used. Non-failed banks' data from 2009, the peak crisis year, was selected to present the most challenging forecasting scenario. This extensive, manual data extraction yielded 123,786 data points involving 26 CAMEL financial ratios. The final dataset provided a rigorous basis for predictive modeling and was analyzed using WEKA software to perform binary classification of bank health during the crisis.

3.1 Ratios Used in Research:

- Performance Ratios (13 indicators): Yield on earning assets, cost of funding, net interest margin, non-interest income/expenses, credit loss provisions, operating income, returns on assets/equity, efficiency, charge-offs, and assets per employee.
- Condition Ratios (9 indicators): Asset quality, loan loss allowances, loan performance, liquidity, deposits, capital adequacy, and risk-based capital ratios.
- Other Ratios (4 indicators): Average total assets, average earning assets, average equity, and average loans and leases.

Multicollinearity testing led to the removal of redundant indicators (e.g., ROA, ROE), ensuring the final dataset consists only of relevant and distinct variables to enhance model accuracy and interpretability.

4: Methodology

4.1 Introduction to Methodology

This chapter delineates the methodological framework employed to evaluate and compare the predictive capabilities of traditional statistical methods and advanced machine learning (ML) models in forecasting bank failures. Recognizing the severe implications of bank insolvencies, particularly underscored by the 2008 Global Financial Crisis (Mian & Sufi, 2009; Shiller, 2008), this research systematically investigates whether modern ML techniques offer superior performance over conventional predictive tools (Petropoulos et al., 2020). Methodological choices, including research philosophy, data handling, model selection, and performance evaluation, are meticulously aligned with the research objectives to ensure academic rigor and practical relevance.

4.2 Research Philosophy and Approach

The study adopts a positivist philosophical stance, emphasizing objectivity, empirical validation, and quantifiable evidence (Saunders et al., 2019). Employing a deductive approach, the research tests specific hypotheses drawn from existing literature concerning the relative predictive performance and interpretability of various classification models (Cole & White, 2012; Gogas et al., 2018; Carmona et al., 2019). Utilizing a quantitative research design, the study analyzes secondary data obtained from the Federal Deposit Insurance Corporation (FDIC), covering detailed bank financial information from 2007 to 2013. This explanatory and comparative strategy systematically assesses multiple models against clearly defined performance criteria, ensuring robustness and reliability.

4.3 Machine Learning Models in Bank Failure Prediction

Acknowledging the complexity inherent in financial datasets—such as non-linearity, dimensionality, multicollinearity, and class imbalance (Hastie, Tibshirani, & Friedman, 2009)—this study employs a diverse set of ML algorithms categorized as follows:

- Traditional Probabilistic and Statistical Models:
  - Naïve Bayes (NB): Computational efficiency and scalability despite strong independence assumptions (Rish, 2001).
  - Logistic Regression (LR): Transparent, interpretable odds ratios, widely used in regulatory contexts (Hosmer, Lemeshow, & Sturdivant, 2013).
- Decision Tree-based Ensemble Methods:
  - Bagging: Reduces prediction variance via bootstrap sampling (Breiman, 1996).
  - Random Forest (RF): Introduces random feature selection, enhancing stability and generalization (Breiman, 2001).
  - Regularized Random Forest (RRF): Mitigates feature redundancy through regularization (Deng & Runger, 2012).
  - Cost-Sensitive Forest (CS-Forest): Prioritizes accurate prediction of minority class events, critical for financial risk (Ling & Sheng, 2008).
- Boosting Techniques:
  - AdaBoost: Iteratively improves weak learners, suitable for imbalanced data (Freund & Schapire, 1996).
  - LogitBoost: Provides calibrated probabilities essential for risk assessment (Friedman, Hastie, & Tibshirani, 2000).
  - GAMBoost: Captures non-linear relationships while maintaining interpretability (Bühlmann & Hothorn, 2007).
  - GLMBoost: Combines logistic regression clarity with boosting power (Bühlmann & Hothorn, 2007).
  - XGBoost: Superior performance, scalability, and efficient handling of missing data (Chen & Guestrin, 2016).
- Neural Network and Instance-Based Learning Methods:
  - Multilayer Perceptron (MLP): Effective in modeling complex non-linear relationships (Haykin, 1998).
  - KStar: Robust, instance-level similarity measure for irregular, high-dimensional data (Cleary & Trigg, 1995).

4.4 Performance Metrics for Model Evaluation

Given inherent dataset imbalance, a rigorous evaluation framework includes both threshold-dependent and threshold-independent metrics:

- Accuracy: General correctness, but limited in imbalanced contexts (Hand, 2009).
- Precision: Minimizes false alarms, critical in regulatory environments (Powers, 2011).
- Recall (Sensitivity): Prioritized due to high costs associated with missed bank failures (Gogas et al., 2018).
- ROC-AUC: Threshold-independent measure of discrimination power (Fawcett, 2006).
- Confusion Matrix: Comprehensive visualization of prediction errors (Stehman, 1997).
- Cohen’s Kappa: Adjusts accuracy for chance agreement (Cohen, 1960).

These metrics collectively ensure nuanced, comprehensive insights into model effectiveness for financial applications.

4.5 Rationale for Model and Metric Selection

Model and metric selections are driven by research objectives, emphasizing recall and predictive accuracy for minority class instances due to their significant economic implications (Petropoulos et al., 2020). Interpretability and computational efficiency are balanced against predictive performance, ensuring practical applicability in financial decision-making contexts.

5.Results:

5.1 Introduction

Chapter 5 presents empirical findings from evaluating multiple machine learning models in predicting bank failures, focusing specifically on imbalanced data handling, error minimization—especially Type II errors—and trade-offs between model complexity and interpretability. The analysis addresses three core hypotheses: whether ensemble methods outperform traditional algorithms, the effectiveness of cost-sensitive models, and the implications of complexity versus interpretability in regulatory decision-making contexts.

5.2 Overview of Model Performance :

Various models were evaluated based on key metrics including Correctly Classified Instances (CCI), Recall, ROC-AUC, and the Kappa statistic, providing insights into accuracy, sensitivity, and agreement beyond chance. Ensemble methods—particularly CS-Forest, Regularized Random Forest (RRF), and Random Forest—demonstrated superior performance across these metrics. Specifically, Random Forest showed the highest ROC-AUC (99.1%), RRF offered strong accuracy with balanced error rates, and CS-Forest achieved the highest recall (89.7%), emphasizing its sensitivity to bank failures (Petropoulos et al., 2020; Ekinci & Sen, 2024).

5.3 Analysis of Key Performance Metrics :

5.3.1 Correctly Classified Instances (CCI):

Correctly Classified Instances measure overall accuracy but can be misleading in imbalanced datasets where non-failure cases dominate. Despite this limitation, top-performing models (Random Forest, RRF, CS-Forest) maintained high CCI (approximately 97.3%–97.7%), while simultaneously demonstrating effectiveness on more sensitive metrics (Ekinci & Sen, 2024).

5.3.2 Recall (Sensitivity):

Recall is crucial in bank failure prediction, as it specifically evaluates how effectively models detect actual failures. CS-Forest excelled with the highest recall (89.7%), effectively minimizing false negatives (Type II errors), essential for early regulatory interventions. RRF (87.5%) and GAMBoost (88.3%) also performed notably well, confirming the value of cost-sensitive learning in handling minority classes (Ekinci & Sen, 2024).

5.3.3 Receiver Operating Characteristic Area (ROC-AUC):

ROC-AUC evaluates model capability to discriminate between classes without threshold bias. Random Forest displayed exceptional performance with the highest ROC-AUC (99.1%), indicating strong discriminatory power. CS-Forest and RRF both closely followed with ROC-AUC scores of 98.7%. High ROC-AUC values reflect robust risk ranking abilities, crucial for regulatory prioritization of interventions (Petropoulos et al., 2020).

5.3.4 Kappa Statistic:

Kappa assesses accuracy beyond random chance, adjusting for class imbalance. Random Forest had the highest Kappa value (0.863), closely followed by RRF (0.860) and CS-Forest (0.849), reaffirming their balanced effectiveness and reliability in classifying bank statuses (Petropoulos et al., 2020).

5.4 Confusion Matrix Analysis :

5.4.1 Type I Errors (False Positives):

Type I errors occur when models incorrectly predict non-failed banks as failures, potentially leading to unnecessary regulatory actions. Random Forest showed the lowest Type I error rate (45 instances), while RRF and CS-Forest had slightly higher rates (56 and 81 instances respectively). The higher false positives in CS-Forest reflect its cost-sensitive nature, trading increased false positives for reduced missed failures (Ekinci & Sen, 2024).

5.4.2 Type II Errors (False Negatives):

Type II errors represent misclassifying actual failures as healthy, significantly risky in financial contexts. CS-Forest demonstrated the lowest Type II errors (46 instances), emphasizing its suitability for high-stakes regulatory scenarios. Regularized Random Forest and Random Forest followed with 58 and 64 instances, respectively, also demonstrating substantial effectiveness (Ekinci & Sen, 2024).

5.4.3 Balancing Type I and Type II Errors:

CS-Forest prioritizes reducing false negatives, ideal in regulatory contexts where missed failures pose substantial systemic risks. Conversely, Random Forest and RRF offer balanced trade-offs between sensitivity and specificity, suitable where accuracy and justification for regulatory actions must align closely (Ekinci & Sen, 2024).

5.5 Comparative Analysis of Top-Performing Models :

Random Forest exhibited high accuracy (97.71%) and the best ROC-AUC (99.1%), providing excellent discrimination between failed and non-failed banks. CS-Forest excelled in recall, demonstrating outstanding sensitivity in identifying failures, though with higher false positives. Regularized Random Forest balanced sensitivity and specificity effectively through regularization, enhancing generalizability (Petropoulos et al., 2020).

Feature importance analysis identified key financial ratios, such as Net Interest Margin (NIM), Return on Equity (ROE), Efficiency Ratio, Loss Allowance to Loans, and Risk-Based Capital Ratio as consistent predictors of bank failures across the models (Petropoulos et al., 2020). Complexity versus interpretability trade-offs showed CS-Forest as most complex yet sensitive, Random Forest as moderately complex yet interpretable, and RRF as a balanced middle-ground solution.

5.6 Hypothesis Testing :

- H1 (Random Forest Model Accuracy): Rejected; Random Forest significantly outperformed traditional models in accuracy, ROC-AUC, and reliability (Petropoulos et al., 2020).
- H2 (Cost-Sensitive Performance): Rejected; CS-Forest significantly reduced Type II errors, outperforming other ensemble and hybrid approaches, reinforcing cost-sensitive learning effectiveness (Ekinci & Sen, 2024).
- H3 (Simplicity versus Complexity): Rejected; simpler models (Naïve Bayes) demonstrated significantly lower effectiveness compared to complex ensemble methods (Le & Viviani, 2017).
- H4 (Model Choice Influences): Rejected; model selection was significantly influenced by trade-offs between complexity, accuracy, interpretability, and specific regulatory requirements (Chiaramonte et al., 2016).
- H5 (Predictive Power vs Complexity): Rejected; predictive power directly correlated with model complexity, validating advanced ensemble model efficacy over simpler alternatives (Carmona et al., 2019).

5.7 Practical Implications :

Regulatory authorities could benefit from deploying ensemble models, particularly CS-Forest and Random Forest, in early warning systems for bank failures. CS-Forest's high recall and sensitivity make it ideal for minimizing missed failures, while Random Forest offers a practical balance of interpretability and predictive accuracy for regulatory transparency. Financial institutions can utilize these models internally for proactive risk management and strategic planning based on feature importance insights (Petropoulos et al., 2020; Ekinci & Sen, 2024).

5.8 Conclusion :

This comprehensive analysis confirms ensemble and cost-sensitive machine learning models (Random Forest, CS-Forest, RRF) significantly outperform traditional predictive methods in accuracy, recall, and ROC-AUC, effectively handling challenges of imbalanced datasets. CS-Forest excels in minimizing Type II errors, essential in financial regulatory contexts, while Random Forest offers balanced performance and interpretability. The study strongly advocates for strategic implementation of these models in financial risk management, enhancing regulatory effectiveness and systemic financial stability (Petropoulos et al., 2020; Ekinci & Sen, 2024; Carmona et al., 2019).

6.Conclusion

6.1 Introduction

Chapter 6 synthesizes findings, revisits initial research questions and objectives, and demonstrates how they were systematically addressed. It assesses the theoretical and practical implications, highlighting contributions to financial risk modeling and regulatory contexts, particularly emphasizing ensemble machine learning models such as Random Forest, CS-Forest, and Regularized Random Forest (RRF).

6.2 Revisiting the Research Aims and Objectives :

The research successfully addressed all stated objectives by:

- Conducting an extensive literature review, identifying gaps related to handling imbalanced data and prioritizing Type II error minimization (Le & Viviani, 2017; Petropoulos et al., 2020).
- Constructing a comprehensive dataset from U.S. commercial banks during a financially turbulent period (2007–2013), integrating key CAMELS financial indicators.
- Implementing multiple machine learning algorithms, clearly demonstrating the superior predictive capability of ensemble models, specifically Random Forest, CS-Forest, and RRF.
- Analyzing model complexity and interpretability, critically examining their regulatory implications (Chiaramonte et al., 2016).
- Explicitly evaluating Type I and Type II errors, underscoring the strength of CS-Forest in minimizing missed bank failures (Ekinci & Sen, 2024).
- Offering actionable insights for regulators and policymakers on the practical implementation of these predictive models in early warning systems.
-

Research questions were methodically answered, clearly establishing CS-Forest, Random Forest, and RRF as the most effective models for predicting bank failures, providing critical insights into balancing predictive power with interpretability.

6.3 Key Findings and Contributions to Knowledge :

Empirical Findings:

- Ensemble models consistently outperformed traditional statistical methods (Logistic Regression, Naïve Bayes), validating their applicability in financial distress prediction (Petropoulos et al., 2020).
- Cost-sensitive learning (CS-Forest) significantly enhanced recall and minimized Type II errors, critical for timely regulatory interventions (Ekinci & Sen, 2024).
- The study confirmed a meaningful trade-off between model complexity and interpretability, particularly relevant in regulatory environments demanding transparency.
- Financial ratios such as Net Interest Margin (NIM), Return on Equity (ROE), Efficiency Ratio, and Risk-Based Capital Ratio emerged as crucial predictors of bank failures across tested models.
- Hybrid and ensemble models effectively handled class imbalance, demonstrating empirical superiority in predictive performance (Carmona et al., 2019).

Theoretical Contributions:

- Advancement in financial risk prediction using ensemble and cost-sensitive machine learning models.
- Successful integration of cost-sensitive approaches into bank failure prediction frameworks.
- Validation of CAMELS indicators' continued relevance when used within advanced machine learning methodologies.

Practical Contributions:

- Provision of robust tools for enhancing regulatory early warning systems, thus enabling timely interventions to mitigate systemic risks.
- Clear demonstration of the value of advanced models for internal risk management within financial institutions, particularly through feature importance insights.
- Identification of Regularized Random Forest as a balanced model for contexts requiring both predictive accuracy and regulatory transparency.

6.4 Implications for Policy and Practice :

- Advanced ensemble machine learning models, especially CS-Forest and Random Forest, have substantial potential for integration into regulatory Early Warning Systems (EWS), improving detection accuracy and timely regulatory intervention (Ekinci & Sen, 2024; Petropoulos et al., 2020).
- Emphasis on cost-sensitive approaches and metrics such as Recall and Kappa, rather than solely accuracy metrics, addresses challenges posed by imbalanced data, enhancing regulatory oversight effectiveness.
- Integration of these predictive models with existing stress-testing frameworks could better assess bank resilience under adverse scenarios.
- Addressing the complexity-interpretability trade-off is vital, with Regularized Random Forest (RRF) offering a viable middle-ground solution for transparent regulatory decision-making.
- Financial institutions can proactively manage risks by incorporating these models into internal assessments and risk management frameworks.
- Models contribute significantly to macroprudential supervision, aiding in systemic risk identification and mitigation through data-driven interventions.

6.5 Limitations of the Study :

- Temporal and geographic limitations exist, given the exclusive use of U.S. commercial bank data (2007–2013), raising questions about the generalizability across different economic periods and banking environments.
- Challenges associated with inherently imbalanced datasets persist despite advanced modeling techniques; cost-sensitive methods improve but do not entirely eliminate biases or precision-recall trade-offs.
- Interpretability limitations in complex ensemble models hinder transparency and potentially regulatory acceptance despite high predictive accuracy.
- Reliance on CAMELS-based financial ratios may exclude other crucial factors such as qualitative aspects, management competence, and macroeconomic indicators.

- Scope limitations in hypothesis testing limit broader causal inferences, and conclusions remain contextually tied to the chosen dataset and methodology.

6.6 Directions for Future Research :

Several areas for future exploration were identified, including:

- Expansion of studies to cross-country and cross-regional datasets to assess predictive model robustness across varied regulatory and economic environments.
- Incorporation of macroeconomic variables and market-based indicators for real-time risk assessment and improved sensitivity.
- Development of real-time predictive frameworks to enhance regulatory early warning systems and operational intervention capabilities.
- Advancements in model interpretability and explainability using techniques like SHAP and LIME, facilitating regulatory trust and adoption.
- Integration of qualitative and unstructured data through text mining and Natural Language Processing (NLP) for a more comprehensive risk evaluation.
- Longitudinal analysis and adaptability of models over different economic cycles and evolving regulatory conditions, potentially utilizing transfer and online learning methods.
- Integration of machine learning models within supervisory stress-testing frameworks, augmenting scenario analysis and systemic risk assessment.

6.7 Closing Remarks :

The research has demonstrated significant advancements in predicting bank failures using sophisticated machine learning methodologies, effectively addressing challenges of imbalanced datasets, minimizing critical Type II errors, and balancing complexity with interpretability. The findings provide a robust foundation for future research and offer practical recommendations for enhancing regulatory oversight, internal risk management, and systemic financial stability through predictive modeling advancements (Petropoulos et al., 2020; Ekinci & Sen, 2024; Carmona et al., 2019).

7.References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. Machine Learning, 6(1), 37-66.

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and bagging ensemble methods. Decision Support Systems, 45(1), 110-122.

Antunes, A., Bonfim, D., Monteiro, N., & Rodrigues, P. M. M. (2018). Forecasting banking crises with dynamic panel probit models. *International Journal of Forecasting*, 34(2), 249-275.

Ashcraft, A., & Schuermann, T. (2008). Understanding the securitization of subprime mortgage credit. *Foundations and Trends in Finance*, 2(3), 191-309.

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 49(3), 312-329.

Bellman, R. (1957). Dynamic Programming. Princeton University Press.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural Networks: Tricks of the Trade (pp. 437-478). Springer.

Berger, A. N., & Bouwman, C. H. S. (2012). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109(1), 146-176.

Bernanke, B. S. (2010). *The Courage to Act: A Memoir of a Crisis and Its Aftermath*. W. W. Norton & Company.

Beutel, J., List, S., von Schweinitz, G., 2019. "An evaluation of early warning models for systemic banking crises: Does machine learning improve predictions? * IWH Discussion Papers 2/2019, Halle Institute for Economic Research (IWH). *

Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems. 50. 602-613. 10.1016/j.dss.2010.08.008.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25. DOI: 10.1007/s11749-016-0481-7.

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford University Press.

Boyacıoglu, M. A., Kara, Y., & Baykan, Ö. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of Savings Deposit Insurance Fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications*, 36(2), 3355-3366.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Breiman, L. Bagging predictors. Mach Learn 24, 123–140 (1996). https://doi.org/10.1007/BF00058655

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. The Annals of Statistics, 30(4), 927-961.

Bühlmann, P., Hothorn, T. (2007). "Boosting Algorithms: Regularization, Prediction and Model Fitting," Statistical Science, Statist. Sci. 22(4), 477-505, (November 2007)

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36(3), 4626-4636.

Carmona, P., Diranzo, F. J. C., & Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357.

Chen, C. & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. University of California, Berkeley.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Chernobai, A. S., Rachev, S. T., & Fabozzi, F. J. (2007). Operational risk: A guide to Basel II capital requirements, models, and analysis. John Wiley & Sons.

Chiaramonte, L., Liu, H., Poli, F., & Zhou, M. (2016). How accurately can z-score predict bank failure? *Financial Markets, Institutions & Instruments*, 25(5), 333–360.

Cleary, J. G., & Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. Proceedings of the 12th International Conference on Machine Learning, 108-114.

Cleary, S., & Hebb, G. (2015). An assessment of machine learning algorithms for the prediction of credit default. Econometrics, 3(4), 589-613.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Cole, R.A. & White, L.J. (2012). Déjà Vu All Over Again: The Causes of U.S. Commercial Bank Failures This Time Around. Journal of Financial Services Research, 42(1-2), pp.5-29.

Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B, 20(2), 215-242.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784-3797.

Deng, H. (2013). Guided random forest in the RRF package.

Deng, H., & Runger, G. (2012). Feature selection via regularized trees. Proceedings of the 2012 International Joint Conference on Neural Networks, 1-8.

Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7, 3.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In International Workshop on

Multiple Classifier Systems (pp. 1-15). Springer.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29(2), 103-130.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. Statistical Science, 11(2), 89-102.

Ekinci, A., & Erdal, H. I. (2017). Bank failure prediction by hybrid machine learning techniques. *Economics & Management*, 20(1), 29-42.

Ekinci, A., Sen, S. Forecasting Bank Failure in the U.S.: A Cost-Sensitive Approach. Comput Econ (2024). https://doi.org/10.1007/s10614-023-10537-6

Elkan, C. (2001). The foundations of cost-sensitive learning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*.

Erdal, H. I., & Ekinci, C. (2013). A new approach to Turkish banking sector's failure prediction via artificial neural network and support vector machine. *Expert Systems with Applications*, 40(3), 887-897.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

Fayyad, U. M., & Irani, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. IJCAI, 93, 1022-1027.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML)*.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337-407.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. Machine Learning, 29(2), 131-163.

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

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Gorton, G. (2009). *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press.

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

Hand, D. J., & Yu, K. (2001). Idiot's Bayes—Not so stupid after all? International Statistical Review, 69(3), 385-398.

Hand, D.J. Measuring classifier performance: a coherent alternative to the area under the ROC curve. Mach Learn 77, 103–123 (2009). https://doi.org/10.1007/s10994-009-5119-5

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. Statistical Science, 1(3), 297-318.

Hastie, T., & Tibshirani, R. (1990). Generalized Additive Models. Chapman and Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Haykin, S. (1998). Neural Networks: A Comprehensive Foundation. Prentice Hall.

Ho, T. K. (1998). The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832–844.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359-366.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons.

Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., Hofner, B. (2010). Model-based Boosting 2.0. Journal of Machine Learning Research. 11. 2109-2113.

Jarrow, R.A. and Turnbull, S.M. (1995) Pricing Derivatives on Financial Securities Subject to Credit Risk. The Journal of Finance, 1, 53-85.http://dx.doi.org/10.1111/j.1540-6261.1995.tb05167.x

Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, 11, 51.

Kleinbaum, D. G., Klein, M., & Pryor, E. R. (2002). Logistic Regression: A Self-Learning Text. Springer.

Kothari, V. (2008). *Securitization: The Financial Instrument of the Future*. Wiley.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering, 160, 3-24.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operational Research, 180(1), 1-28.

Kuncheva, L. I., & Rodriguez, J. J. (2007). An experimental study on rotation forest ensembles. In *Proceedings of the 5th International Conference on Multiple Classifier Systems*.

Laeven, L., & Valencia, F. (2018). Systemic banking crises revisited. *IMF Working Paper, WP/18/206*. International Monetary Fund.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

Le, L., & Viviani, J. L. (2017). An empirical analysis of bank failure prediction in the United States. *Journal of Financial Research*, 40(2), 257-282.

Lessmann, S., Baesens, B., Seow, H., Thomas, L. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research. (doi:10.1016/j.ejor.2015.05.030). 10.1016/j.ejor.2015.05.030.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18-22.

Ling, C. X., & Sheng, V. S. (2008). Cost-sensitive learning and the class imbalance problem. Encyclopedia of Machine Learning, 2011, 231-235.

Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 31-57.

Lu, W., & Whidbee, D. A. (2013). Bank structure and failure during the financial crisis. *Journal of Financial Economic Policy*, 5(3), 281–299.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765–4774).

Manthoulis, G., Mylonakis, J., & Diacogiannis, G. (2020). Predicting bank failures: A statistical and machine learning approach. *Journal of Financial Risk Management*, 9(4), 305-323.

Martens, D., & Baesens, B. (2010). Building acceptable classification models for financial decision support using robust meta-learning. Decision Support Systems, 48(3), 491–501.

Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, 1(3), 249-276.

McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models (2nd ed.). Chapman and Hall/CRC.

Meinshausen, N. (2007). Relaxed lasso. Computational Statistics & Data Analysis, 52(1), 374-393.

Menard, S. (2000). Applied Logistic Regression Analysis (2nd ed.). Sage.

Meyer, P. A., & Pifer, H. W. (1970). Prediction of bank failures. *The Journal of Finance*, 25(4), 853-868.

Mian, A., & Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *The Quarterly Journal of Economics*, 124(4), 1449-1496.

Miguéis, V. L., Camanho, A. S., & Cardoso, F. J. (2018). Predicting bank failure using survival analysis. European Journal of Operational Research, 274(2), 647-661.

Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. MIT Press.

Mitchell, T. M. (1997), Machine learning , Vol. 1 , McGraw-hill New York .

Momparler, A., Carmona, P., & Redondo, Y. (2020). Fuzzy-set qualitative comparative analysis (fsQCA) in banking failure prediction. *Journal of Business Research*, 118, 253-263.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In Proceedings of the 22nd International Conference on Machine Learning (pp. 625-632).

Nielsen, D. (2016). Tree Boosting with XGBoost - Why does XGBoost win "Every" machine learning competition? NTNU Machine Learning Report.

Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217-222.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. The Journal of Educational Research, 96(1), 3-14.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36, 10.1016/j.ijforecast.2019.11.005.

Phua, C., Lee, V., Smith-Miles, K., Gayler, R. (2010). A Comprehensive Survey of Data Mining-based Fraud Detection Research. CoRR. abs/1009.6119.

Plessis, E. (2022). Dynamic forecasting of banking crises with a Qual VAR. *Journal of Applied Economics*, 25(1), 477-503. DOI: 10.1080/15140326.2020.1816132.

Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21-45.

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.

Provost, F., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the 15th International Conference on Machine Learning (ICML 1998)*.

Rätsch, G., Onoda, T., & Müller, K. R. (2001). Soft margins for AdaBoost. Machine Learning, 42(3), 287-320.

Reinhart, C. M., & Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press. ISBN: 978-0691142166.

Ribeiro, M., Singh, S., Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 1135-1144. 10.1145/2939672.2939778.

Rish, I. (2001). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536.

Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.

Saunders, M.N.K., Lewis, P. and Thornhill, A. (2019) Research Methods for Business Students. 8th Edition, Pearson, New York.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. In Nonlinear Estimation and Classification (pp. 149-171). Springer.

Schapire, R. E., & Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. Machine Learning, 37(3), 297-336.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Transactions on Systems, Man, and Cybernetics-

Part A: Systems and Humans, 40(1), 185-197.

Shiller, R. J. (2008). *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It*. Princeton University Press.

Sinkey, J. F. (1975). A multivariate statistical analysis of the characteristics of problem banks. *Journal of Finance*, 30(1), 21-36.

Skurichina, M., & Duin, R. P. W. (2002). Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis & Applications, 5(2), 121-135.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1), 77-89.

Sun, Y., Kamel, M. S., Wong, A. K., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. Pattern Recognition, 40(12), 3358-3378.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. Management Science, 38(7), 926-947.

Thomas, L.C., Edelman, D.B. and Crook, J.N., 2002. Credit scoring and its applications. Philadelphia: Society for Industrial and Applied Mathematics. (Mathematical Modeling and Computation, MM06). Available at: https://doi.org/10.1137/1.9780898718317

Thomson, J. B. (1991). Predicting bank failures in the 1980s. *Federal Reserve Bank of Cleveland, Economic Review*, Q I, 9-20.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267-288.

Torna, G., & DeYoung, R. (2012). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, 21(3), 397-421.

Tsymbal, A. (2004). The problem of concept drift: Definitions and related work. Computer Science Department, Trinity College Dublin.

West, D. (2000). Neural network credit scoring models. Computers & Operations Research, 27(11-12), 1131-1152.

Witten, I. H., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Zadrozny, B., Langford, J., & Abe, N. (2003). Cost-sensitive learning by cost-proportionate example weighting. In Proceedings of the Third IEEE International Conference on Data Mining (pp. 435-442).

Zhang, H. (2004) The Optimality of Naive Bayes. Proceedings of 17th International Florida Artificial Intelligence Research Society Conference, Menlo Park, 12-14 May 2004, 562-567.

Zhang, Y., Wang, S., & Ji, G. (1999). A rule extraction method from neural networks based on the improved genetic algorithm. Expert Systems with Applications, 36(2), 1970-1976.

Zhou, Z. H., Chen, K., & Li, M. (2014). Ensemble methods. Foundations and Trends® in Machine Learning, 6(1), 1-34.

Zhou, Z.-H. (2012). Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC

List of Publications:

Ekinci, A., Sen, S. Forecasting Bank Failure in the U.S.: A Cost-Sensitive Approach. Comput Econ 64, 3161–3179 (2024). https://doi.org/10.1007/s10614-023-10537-6

Sen, S. (2024). Comparison of boosting and random forest models in forecasting bank failures: Revisiting the 2008 financial crisis from a supervisory perspective. Economy and Finance, 11(3), 258–281. https://doi.org/10.33908/EF.2024.3.2

SEN, S., & Figueiredo, S. A. de. (2021). Predicting Bank Failures with Machine Learning Algorithms: A Comparison of Boosting and Cost-Sensitive Models. Journal of Economics, Finance and Accounting Studies, 3(2), 43-50. https://doi.org/10.32996/jefas.2021.3.2.5

SEN, S., & Figueiredo, S. A. de. (2021). Forecasting Bank Failure with Machine Learning Models: A study on Turkish Banks. Journal of Economics, Finance and Accounting Studies,3(2), 51-59. https://doi.org/10.32996/jefas.2021.3.2.6

SEN, S., & Figueiredo, S. A. de. (2021). Impact of Covid-19 Pandemic on Financial Inclusion and Digital Adoption. Journal of Business and Management Studies, 3(2), 108-119. https://doi.org/10.32996/jbms.2021.3.2.11