

Article

Beyond Predicting Failure: A Cost-Sensitive, Explainable AI Framework for Identifying Actionable Recovery Pathways in the U.S. Small Businesses

Md Majedul Hasan¹, Muhammad Rahim², Naim Rahman Ifti³, Muhammad Shayan Nafees⁴

1. Master of Science in Business Analytics Trine University, USA
2. Master of Science in Computer Science (Data Analytics), Westcliff University
3. Bachelor of Science in Computer Science San Francisco Bay University
4. Bachelor of Science in Computer Science San Francisco Bay University

*Correspondence: majedhpl30195@gmail.com, muhammadrahimrahim68@gmail.com,
naimrahmanifti@gmail.com, shayannafees9586@gmail.com

Abstract: Small companies are a critical component of the U.S. economy, providing jobs, innovations and economic growth. However, many small businesses suffer financial difficulties because they lack access to capital, are not operating efficiently and due to market fluctuations, they default on their loans and businesses fail to succeed. The traditional predictive models are limited in interpretability for the decision-makers and mainly aim to predict failure of the business without taking into account the costs of misclassification. This study introduces a new framework, a Cost-Sensitive, Explainable Artificial Intelligence (AI) Framework, which goes beyond the prediction of failure and also determines actionable pathways for recovery for U.S. small businesses. The proposed framework involves the use of the National Small Business Administration (SBA) Loan Dataset (SBAnational.csv) which contains 899,164 loan records and 27 business and loan related variables to build predictive models for financial distress. The data is extensively preprocessed and engineered, after which various machine learning techniques, such as Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM and CatBoost are tested under a cost-sensitive learning approach to reduce the financial harm of non-symmetric misclassification errors. The use of Explainable AI (XAI) techniques like SHAP and LIME allows for the extraction of insights from the models, thereby helping to identify factors that are crucial for business distress, including loan characteristics, business size, industry classification, and financing conditions. This is a novel approach of translating AI-generated explanations into actionable recovery paths, which enables informed managerial decision making, such as recommendations to enhance liquidity management, financing structures and operational risk management, and long-term business resilience. The accuracy, precision, recall, F1 score, ROC-AUC, Balanced accuracy, Matthews correlation coefficient, and cost-sensitive evaluation metrics are used to evaluate the model performance. The proposed framework will add to the domains of Business Analytics and Explainable AI by combining the three pillars of predictive accuracy, interpretability, and practical decision support in one. The results will help small business owners, financial institutions, policy makers, and the U.S. Small Business Administration to proactively detect financial distress and take evidence-based recovery measures to enhance the sustainability and resilience of small businesses.

Citation: Hasan Md H. Beyond Predicting Failure: A Cost-Sensitive, Explainable AI Framework for Identifying Actionable Recovery Pathways in the U.S. Small Businesses. American Journal of Economics and Business Management 2026, 9(10), 808-905.

Received: 10th Jul 2026
Revised: 11th Aug 2026
Accepted: 11th Sep 2026
Published: 21th Oct 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

Keywords: Cost-Sensitive Machine Learning, Explainable Artificial Intelligence (XAI), Financial Distress Prediction, U.S. Small Businesses, Actionable Recovery Pathways and Business Analytics.

1. Introduction

Background

Small businesses are known to be the lifeblood of the U.S. economy, accounting for a significant share of employment opportunities, innovations, entrepreneurship, and economic growth in the nation. These businesses make up a large percentage of businesses and provide many jobs in a variety of industries [1]. Although they are essential to the economy, many small enterprises face ongoing financial problems with their access to capital, inconsistent demand, business inefficiency, economic instability, and consumer tastes. These difficulties can lead to loan default, economic distress and business collapse, negatively affecting the financial institutions, employees and the economy in general as well as the business owner [2]. With the advent of Artificial Intelligence (AI) and Machine Learning (ML), business analytics has revolutionized, as organizations can make decisions based on the prediction of large-scale financial and operational data. Many studies have created machine learning models based on financial ratios, loan details, and business attributes to predict both the occurrence of bankruptcy and financial distress. There are many conventional prediction models that focus on maximizing classification accuracy, but that do not account for the different economic impacts of various kinds of prediction errors [3]. When a business lender overlooks a business that is high risk, there can be significant losses in business dollars; and when a business lender believes a business is distressed when it's actually healthy, there can be losses of business dollars [4]. As such, it is important to consider asymmetric misclassification costs when building prediction systems instead of just considering traditional prediction measures. Moreover, many of these advanced AI systems are 'black box' and generate highly accurate predictions with little insight into the factors that affect their decision-making [5]. This lack of understanding can limit the trust of the manager and adoption of the concept for financial decisions [6]. This, in turn, suggests the need to combine cost-sensitive learning with Explainable Artificial Intelligence (XAI) to create decision support systems that are transparent, reliable, and actionable, which not only can help identify financial distress businesses but also provide value-added information for business resilience and sustainability.

Importance of Cost-Sensitive and Explainable AI in Small Business Financial Distress Prediction

The growing amount of financial and business information has led to organizations turning to AI in order to forecast a company's financial condition and also help the strategic decision making process [7]. However, many current predictive models focus on the overall classification accuracy only and do not take into account the financial impact of the various kinds of prediction errors. The price of misjudging a business in need of financing instead of one that is truly in trouble is far higher than that of misjudging a healthy business as high risk in small business financing [8]. To overcome this issue, cost-sensitive machine learning incorporates different penalties on false positive and false negative predictions to generate models that more closely align to real-world financial risk and better enhance decision quality. Beyond accuracy, transparency is now a must have part of AI-based decision support. Financial institutions, policy makers and business owners need to understand the results of the models before lending investment or operation [9]. Explanation AI (XAI) methods like SHAP and LIME improve transparency by highlighting which variables are most relevant to the results of problem predictions and how individual company characteristics relate to problems of financial distress. These explanations are valuable for building trust among stakeholders, ensuring compliance with regulations, and fostering ethical implementation of AI. By integrating cost-sensitive learning with Explainable AI, a holistic AI system can be developed that not only improves the accuracy of financial distress predictions but also provides actionable recommendations to enhance business performance [10]. This can help decision-makers recognize and prioritize

corrective measures to address key financial vulnerabilities, allocate resources effectively, and develop evidence-based recovery strategies [11]. Therefore, predictive analytics, explainability, and actionable business intelligence are impressive developments for supporting the sustainability and resilience of U.S. small businesses in the long run.

Problem Statement

The various machine learning business failure prediction models have been developed, but most of them focused mainly on the prediction accuracy without considering the different financial effects of misclassification errors. A number of top-performing AI models are black-box systems, which make them less transparent and can diminish the trust of stakeholders in key financial decisions [12]. More importantly, traditional financial distress prediction models usually end their analysis at a diagnosis of financial distress and fail to give practical advice to help businesses remediate their financial situation [13]. This research overcomes these challenges by developing a cost-sensitive, interpretable AI model that not only forecasts financial distress of small businesses in the United States but also reveals interpretable risk factors and transforms them into actionable recovery pathways to guide managers and financial decision makers in recovery.

Objectives of the Study

The following is the aim of the study:

- To design an ML model that captures financial distress for small businesses in the United States, taking into account their cost.
- To evaluate and compare several machine learning methods with cost-sensitive evaluation metrics [14].
- To determine which of the financial and business attributes are most significant for small business financial distress.
- To improve model transparency with the help of XAI methods such as SHAP and LIME.
- To spot actionable recovery pathways using AI generated explanations for how to make the business more resilient.
- To create an intelligent decision support system for business owners, financial institutions and policymakers.

Research Questions

To address the aforementioned issues, the following research questions are raised:

- What is the predictive performance of cost-sensitive machine learning in financial distress for U.S. small businesses?
- What are the most important financial and business variables to predict financial distress?
- What are the benefits of EAI in making financial distress prediction models more transparent and interpretable?
- What kind of explanations can be produced using AI and how can these be translated into actionable recovery paths for U.S. small businesses that can lead to sustainable decision-making?

Significance of the Study

The current research is among the most recent studies in the areas of Business Analytics, Artificial Intelligence and financial risk management, introducing an integrated framework of combining cost-sensitive machine learning with Explainable AI for the prediction of financial distress in small and medium enterprises from the U.S. market. The proposed framework also considers the uneven cost of misclassification and provides greater transparency and usability of the models, rather than emphasizing classification

accuracy [15]. The framework can be augmented by site-specific explainability techniques which allow business owners, financial institutions, and policy makers to better comprehend the factors that contribute to financial distress and make sound decisions that are informed by evidence. The study also has significant practical application to the sustainability of small business [16]. The proposed framework is not just about finding businesses that might fail, it also provides a framework for the recovery. The way the framework works is by using AI to generate insights, and then to turn those insights into actionable recovery pathways that help businesses to manage their liquidity, optimize financing, reduce risks, and enhance their long-term resilience [17]. All these recommendations go towards preventive measures before it gets out of hand. This study builds on previous work that explores the use of these technologies and their specific interaction by combining cost-sensitive learning, explainable AI and decision-support analytics into one system tailored for U.S. small businesses [18]. The results will serve as a good reference for those who continue to study financial distress prediction, explainable machine learning and intelligent business decision-support systems [19]. This study showcases the potential to translate complex analytics into actionable business insights to help sustain business growth, improve lending decisions and foster greater economic resilience in the U.S. small business ecosystem.

Predicting Financial Distress using Artificial Intelligence and Cost-Sensitive Machine Learning

Financial distress prediction is a crucial research field in business analytics as it allows organizations and financial institutions to take preventive measures before the financial distress of a company occurs. Many statistical models have been applied to the problem of predicting bankruptcy and loan default based on financial ratios and business performance measures, such as logistic regression and discriminant analysis [20]. These techniques produce interpretable results, but can be less effective in modeling complex, nonlinear relationships in large financial data sets. As a result, machine learning algorithms like Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting, XGBoost, LightGBM, and CatBoost have been increasingly used to enhance predictive power in the last few years. These algorithms are able to identify and learn hidden patterns in multidimensional financial data and have proven to be more successful than traditional statistical methods [21]. Despite these developments, there are still many of these prediction models that maximize the accuracy of the overall prediction task and do not take into account the different economic implications of prediction errors [22]. In the outside world, where the business landscape is far more varied and complex, it is possible to misjudge a high-risk business as financially sound and end up losing a lot of money, while it is far less likely to be a bad bet for a healthy business [23]. The challenge is solved in cost-sensitive machine learning by using different penalties for false positive and false negative predictions, so that predictive models can minimize financial loss instead of maximizing prediction accuracy [24]. Recent research has demonstrated how the integration of cost-sensitive learning helps to make better risk decisions, better decision quality and more responsible lending [25]. Few studies, however, have combined cost-sensitive learning specifically in the context of U.S. small business financial distress, offering a chance to derive more practically viable and economically viable predictive models.

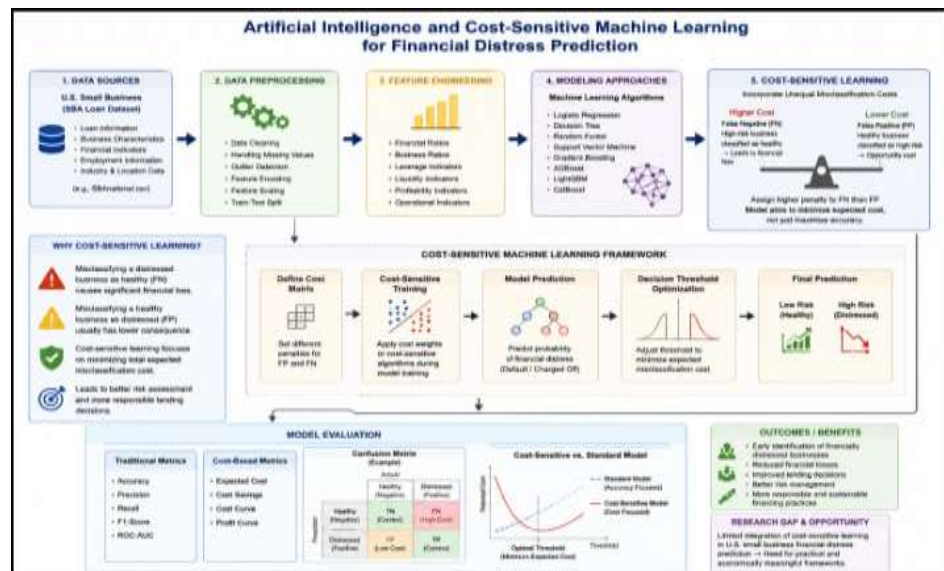


Figure 1. The figure illustrates an AI model that accurately forecasts financial distress in U.S. small businesses, while also being cost-efficient

The proposed AI and Cost-Sensitive ML framework for Financial Distress predication in Small businesses (SBs) in the United States is shown in figure 1. The framework starts with the data collection from the SBA National Loan Dataset, followed by data preprocessing steps, such as data cleaning, handling missing values, data encoding, scaling, and train-test splitting. Financial Statement and Business Indicators for Model Development are Feature engineering involves taking meaningful indicators that are present in the financial statements or in the business and liquidity to be used in developing the model, such as profitability and operational indicators [26]. Financial distress is predicted using multiple machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, LightGBM, and CatBoost. The framework incorporates cost-sensitive learning by setting stricter penalties for false negative predictions, which helps to reduce cost of the system [27]. Performance is measured by traditional measures as well as a cost based measure and the outcomes focus on better risk assessment, responsible lending, identifying distress early and providing actionable decision support on how to strengthen small business resilience.

Explainable Artificial Intelligence and Decision Support for Actionable Business Recovery

With the growing role of AI in financial decision-making, the need for transparency and interpretability has become a critical necessity for AI systems to be trusted. Often, advanced ensemble learning algorithms offer high predictive power, but lack transparency of the model, giving little information about how decisions are reached [28]. This lack of interpretability undermines managers' willingness to accept the findings, diminishes confidence among stakeholders, and poses difficulties with regulatory compliance especially in financial risk assessment and lending decisions. Explainable Artificial Intelligence (XAI) has, therefore, become a topic of great interest that can help enhance the transparency of machine learning predictions. Methods like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) can offer insights into the global and local nature of the predictions by analyzing the impact of individual variables [29]. The literature shows that explainability can build trust with stakeholders and make predictive analytics actionable decision support. Through highlighting these key financial, operational, and business-related risk indicators, XAI helps managers and financial institutions determine the reasons behind a business's designation as financially distressed and create intervention strategies specific to those

factors [30]. However, the majority of the existing research ends in prediction explanations rather than moving on to the practical aspect of recommendations for recovery. There is limited research exploring the idea of moving AI-generated explanations from theory into practice, specifically with U.S. small businesses [31]. Thus, the infusion of Explainable AI and cost-sensitive learning is a significant step forward, bringing together predictive accuracy, clarity, and decision guidance in a single framework [32]. This helps organizations gain a better understanding of their high-risk businesses, priorities the correct actions, enhance financial resiliency and plan for the long-term sustainability of their business through evidence-based recovery planning.

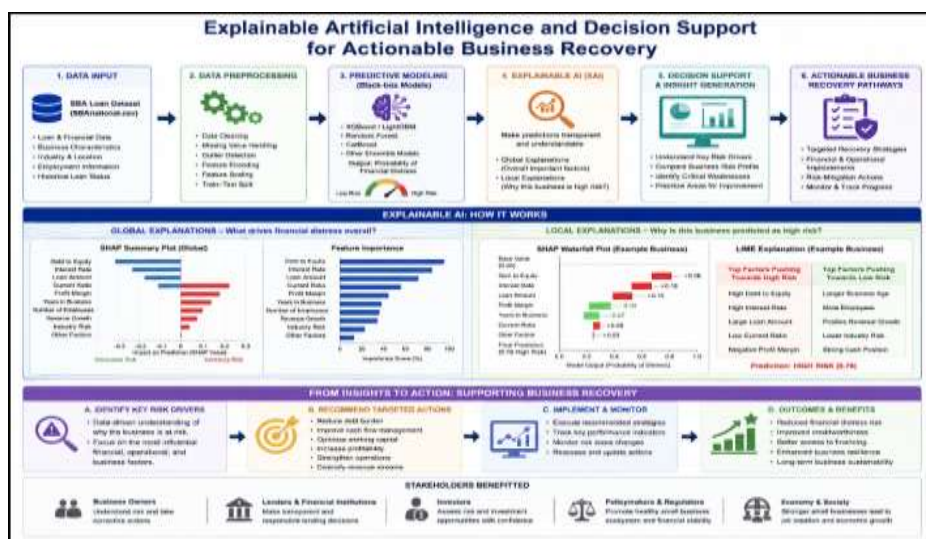


Figure 2. This image displays an explainable AI framework for actionable business recovery and decision support

In Figure 2, the proposed Explainable Artificial Intelligence (XAI) and Decision Support framework for identifying actionable business Recovery pathways in U.S. small businesses is shown. The framework starts with the collection of data on SBA loans, then preprocesses the data, and then uses ensemble machine learning algorithms to predict financial distress risk. Then, explainable AI methods like SHAP and LIME are used to give explanations of the results of the predictions, both on the global and local level, by finding the most important financial and operational risk factors. These actionable findings help inform stakeholders of business vulnerabilities and make business recovery recommendations [33]. The framework also focuses on implementing targeted corrective actions, ongoing performance monitoring and assessing business improvements. It drives clear and transparent AI-based decision-making, builds trust among stakeholders, boosts financial robustness, and fosters long-term sustainability for U.S. small businesses.

Empirical Study

In the article Predictive Modeling for Small Business Bankruptcy Risk: Techniques, Challenges, and Policy Implications by Mayowa Emmanuel (2025), the author examines the growing importance of predictive modeling as a strategic tool for assessing bankruptcy risk among small businesses. Several statistical and machine learning methods such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forests and Deep Learning architectures are reviewed and discussed for their application in predicting financial distress and assisting in credit risk analysis. The author highlights the importance of not only selecting the right algorithm, but also using good data, feature engineering to improve the predictive performance, and the proper interpretability of the model [1]. The article highlights some ongoing problems encountered in bankruptcy prediction: class imbalance, data scarcity, small businesses' heterogeneous nature, and the evolving nature

of financial distress measures. In addition to technical aspects, the article also addresses the policy implications of predictive analytics, such as the potential for using the models to bolster early-warning systems, enhance SME lending frameworks, and inform use of financial interventions to improve business resilience. The results of the aforementioned studies lay a solid theoretical foundation for the current study, which is an extension of the previous research by adding the concept of actionable recovery pathways for U.S. small businesses to the financial distress prediction through the utilization of a combination of cost-sensitive machine learning and EAI concepts. The proposed framework aims to overcome both technical and managerial constraints mentioned in the literature, while also providing predictive accuracy and transparency and supporting financial risk management in a more sustainable way.

Chuhong Wang, Yang Lin, and Xingfei Liu (2025) present and test an Artificial Intelligence-driven Financial Health Index for measuring financial resilience of digital entrepreneurs in the post-pandemic economy. The study evaluates demographic, behavioral, and digital activity variables and incorporated ensemble machine learning methods such as Random Forest, XGBoost, and Neural Networks with SHAP (Shapley Additive Explanations) based on multi-source datasets gathered from digital entrepreneurs in China, finding the key variables that affect their financial health. Results show that age, level of education, payment habits of utility bills, borrowing status, living outside cities and towns, and digital transactions are significant factors that affect financial resilience and vulnerability. The authors highlight the importance of EAI for enhancing the transparency of model predictions, user trust and evidence-based financial decision making. The proposed Financial Health Index gives practical guidance to individuals, financial institutions and policymakers, and facilitates adaptive financial assessments and financial inclusion [2]. The study illustrates how ensemble learning and explainability can result in powerful predictive models that can tackle complex financial risk scenarios. The present research is also valuable with findings similar to the ones provided, which also use Explainable Artificial Intelligence in financial risk prediction. The current study's contribution to the literature is its application of cost-sensitive machine learning (CSML) to consider the unequal costs of prediction errors explicitly and its conversion of predictions into actionable recovery pathways for U.S. small businesses within the context of the SBA loan dataset. This broader approach to decision support is useful for improving financial distress prediction, providing better transparency of financial models, and better sustainable risk management and for improving practical business recovery planning.

In the article Implementing AI-Driven Decision Support Systems in Data Management by Shabir Ahmad and Yazeed Alshuaibany (2026), the authors propose an Artificial Intelligence-driven Decision Support System (DSS) that integrates machine learning, class-imbalance mitigation, explainability, and cost-sensitive decision logic to improve operational decision-making in credit card fraud detection. It uses a publicly available Credit Card Fraud dataset from Kaggle and tests a number of supervised algorithms such as Logistic Regression, Random Forest and XGBoost, but also handles the serious class imbalance by using the Synthetic Minority Oversampling Technique (SMOTE). They choose Precision-Recall Area under the Curve (PR-AUC) as the key performance indicator since it is better suited to the evaluation of models when data is highly imbalanced, and use Recall, Precision, F1-score, and ROC-AUC as complementary performance indicators. Moreover, a probability calibration and cost sensitive threshold optimization process is integrated to categorize transactions as low/high risk to enhance practical decision support [3]. The framework incorporates SHapley Additive explanations (SHAP) for improved transparency and user trust, offering both global and local explanations for model predictions. The study shows that XGBoost not only outperforms the other models in predictive power but also in providing robust risk classification and interpretability, underscoring the critical need for accuracy and transparency in decision-making. The results of this research also lend strong support to the methodological basis

of the current study, which also combines cost sensitive machine learning and Explainable Artificial Intelligence in the context of financial distress prediction. In contrast to the fraud detection context studied, the current study investigates the use of these techniques for predicting financial distress among U.S. small businesses, using the SBAnational.csv dataset, and expands the decision-support process by providing actionable recovery pathways. This wider framework goes beyond simply identifying financially vulnerable businesses, it also offers tangible recovery suggestions that increases business resilience, better lending choices, and sustainable financial risk management.

2. Methodology

The methodology outlines a systematic approach to the design and assessment of a Cost-Sensitive Explainable Artificial Intelligence (XAI) model to predict financial distress and design actionable financial recovery pathways in U.S. small businesses [34]. It involves research design, data set selection, data pre-processing, feature engineering, machine learning model design, explainability analysis, and performance evaluation and more, all aimed at providing accurate, transparent, and useful decision support.

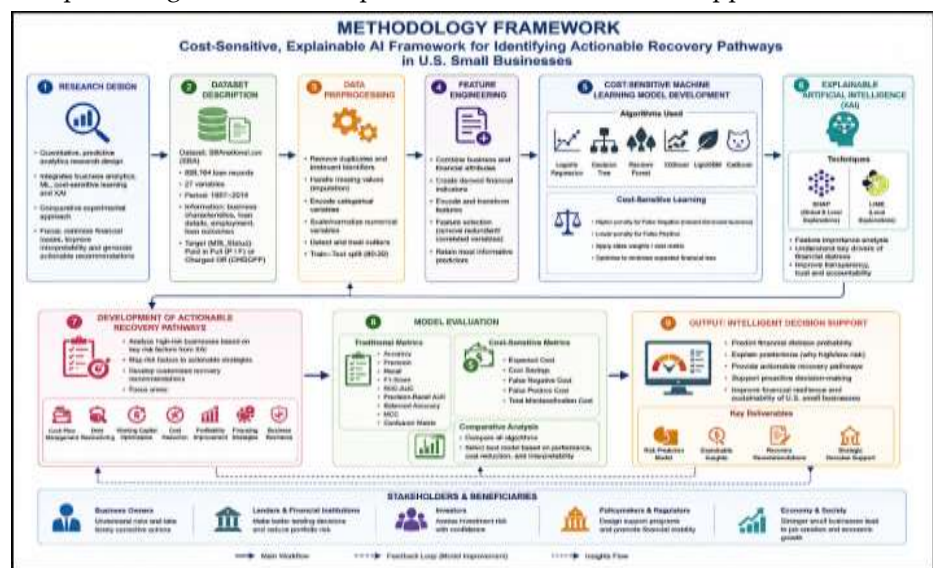


Figure 3. This image shows the proposed methodology of cost-sensitive Explainable Artificial Intelligence (XAI) to predict financial distress in small businesses

The proposed comprehensive methodology framework for the development of a Cost-Sensitive, Explainable Artificial Intelligence (XAI) model to identify actionable recovery pathways for U.S. small businesses is presented in Figure 3. The framework starts with research design, and it is followed by data description, systematic data pre-processing, and feature engineering to achieve the goal of obtaining high-quality data for machine learning analysis [35]. A number of supervised learning algorithms trained in a cost-sensitive learning environment are then implemented to reduce financial penalties for misclassification errors. The prediction results are explained using Explainable AI techniques, such as SHAP and LIME, to highlight financial distress factors [36]. The framework then converts these explainable insights into actionable recovery pathways, and tests the performance of the models against the traditional and cost-sensitive metrics [37]. The methodology provides intelligent decision support to the benefit of business owners, financial institutions, investors, policy makers and the economy as a whole in the form of clear and transparent financial risk management based on data.

Research Design

The research design used in this study is a quantitative predictive analytics research design which will be used to develop and test the Cost-Sensitive Explainable Artificial Intelligence (XAI) framework for identifying actionable recovery pathways in U.S. small businesses. The study adopts an experimental approach and trains and evaluates different

supervised machine learning algorithms on historical loan information from the U.S. Small Business Administration (SBA). The quantitative approach is suitable, because it allows a systematic analysis of large-scale structured data and allows the objective measure the performance of the models by using statistical evaluation metrics [38]. The proposed framework combines Business Analytics, Supervised Machine Learning, and Cost-Sensitive Learning and Explainable Artificial Intelligence techniques, to provide better financial distress prediction and also increase the transparency of the models [39]. In contrast to traditional prediction studies which focus mostly on maximizing the accuracy of the prediction, this research incorporates asymmetric misclassification costs by imposing higher penalty on false negative prediction and thus reducing financial losses caused by the failure of detecting business distress. Several machine learning models, such as Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM and CatBoost are compared using a comparative experimental design. Moreover, the system includes Explainable Artificial Intelligence techniques like SHAP and LIME to offer understandable explanations for the model's predictions [40]. These explanations help stakeholders appreciate the underlying causes which are contributing to financial distress and facilitate evidence-based decision making [41]. The overall research design involves developing an intelligent decision support framework that includes predictive accuracy, cost sensitive optimization, and model interpretability to recommend recovery pathways for financially challenged U.S. small businesses.

Dataset Description

The SBAnational.csv data is used as the primary data source for this research to create and validate the proposed predictive framework. This data set includes 27 business and loan variables and 899,164 loan records from 1987 to 2014 of loans guaranteed by the SBA. It offers rich data on business features, loan application information, loan terms, hiring data, industry categories, and loan outcome for a wide variety of U.S. small businesses. Some predictor variables are important and included in the dataset: business location, industry code (NAICS), number of employees, loan approval amount, disbursement amount, repayment term, revolving credit status, and loan documentation type [42]. The characteristics of the financial and operational aspects of the small businesses are described together by these variables and the data set is thus well suited for financial distress prediction. The dependent variable, MIS_Status is whether the loan was Paid in Full (P I F) or Charged Off (CHGOFF) and could be used to build a binary classification model to identify financially distressed businesses. The proposed machine learning models are developed with high sample size and a wide spectrum of industry representation which increases the generalizability and robustness of the proposed models [43]. The multitude of financial and business-related features enables feature engineering and explainability analysis from SHAP and LIME. The SBAnational.csv data set therefore serves as a good baseline for the development of cost-aware machine learning models and the creation of easily interpretable recovery suggestions for small businesses across the United States.

Data Preprocessing

The data pre-processing is an important phase in the proposed methodology as it benefits in enhancing data quality, performance of the model and mitigates the effect of inconsistencies in the raw data. No duplicate records and non-informative identifier variables are eliminated to avoid information leakage, and no redundant identifier variables are kept. The data set is then checked for missing values and imputation methods are used as appropriate, depending on the nature of the data and each variable [42]. These categorical variables, such as business type, industry classification, revolving credit status and loan documentation category, are mapped to numbers using appropriate numerical coding to enable them to be handled by the machine learning algorithms [43]. Where needed, continuous numerical variables are standardized or normalized to ensure that the features have appropriate scales and thus facilitate algorithm convergence [44]. Methods

for outlier detection are used to identify outlier observations, which can affect prediction accuracy or can be used to bias the model's training [45]. Data consistency checks are carried out to make sure that financial and loan variables are consistent and therefore integrated in the data before developing the model. After pre-processing, feature selection is carried out to remove the redundant features and computational complexity [46]. The final step is splitting the cleaned data into 80% training set and 20% testing set and the model is then trained, validated and tested on the training and testing sets respectively without any bias [47]. The following pre-processing techniques lay the foundation of the study to ensure a reliable analytical basis for the cost-sensitive machine learning and Explainable Artificial Intelligence techniques.

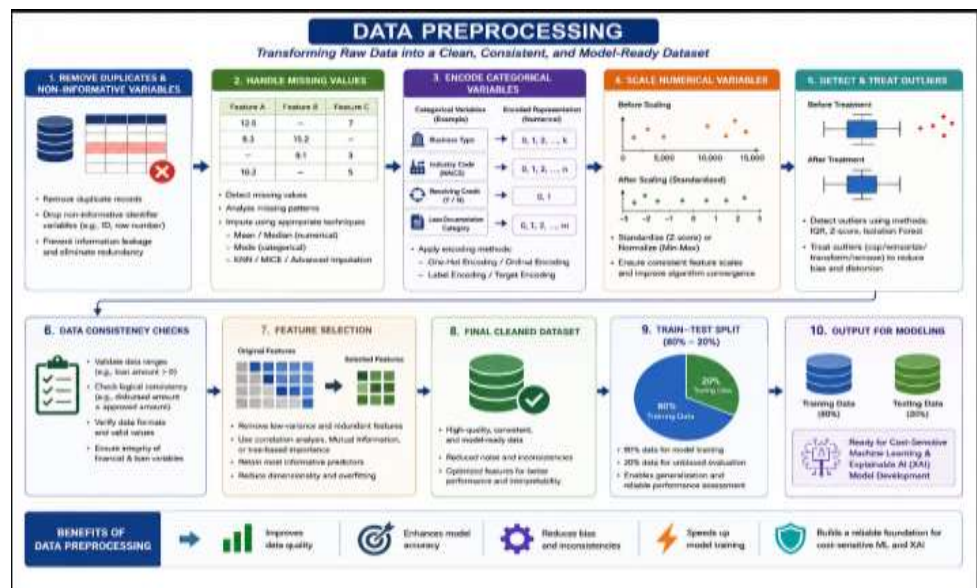


Figure 4. This image illustrates the entire machine learning model development data pre-processing pipeline

The data preprocessing pipeline applied is shown in Figure 4, which summarizes all the steps taken to clean, transform, and standardize the raw SBA loan data for further analysis using machine learning algorithms [48]. The process starts with reduplication, removing any duplicate records and non-informative identifier variables (e.g., social security number, address) to ensure that information is not leaked. Handling missing values with proper statistical methods and then encoding categorical data into numeric values suitable for machine learning algorithms. The numerical features are scaled to have the same dimensions, and outlier detection and treatment reduce the impact of extreme data points [49]. These include data consistency checks and feature selection to ensure data integrity and select the most informative variables. The processed data is then split into 80% training and 20% testing data to ensure an accurate basis for cost-sensitive machine learning, Explainable Artificial Intelligence (XAI) and accurate financial distress prediction.

Feature Engineering

To increase the predictive power of the machine learning models that are considered, variables from the financial health of small businesses in the U.S. are selected, transformed, and created into meaningful variables through feature engineering [49]. The first step in the process is to determine those attributes that are most relevant for the business and loan attributes in the SBAnational.csv file [50]. Financial variables include approved loan amount, disbursed loan amount, term of the loan, revolving credit status, loan documentation type, and SBA guaranteed amount, while business characteristics include company size, number of employees, industry classification (NAICS), business type and geographic location [51]. These variables give a complete picture of both operational and

financial aspects which affect the performance of the business. This involves generating derived variables that are derived from the existing variables and transforming the categorical variables to numerical encoding [52]. If appropriate, use feature scaling for continuous variables to make them consistent with different machine learning algorithms. Moreover, correlation analysis and feature importance evaluation are performed to find out the most relevant features and to remove features which are highly correlated or redundant in order to reduce the complexity of the model [53]. This process decreases computation, prevents overfitting, and obtains good generalization ability. The resulting optimized feature set is used to build cost-sensitive machine learning models that can correctly classify financially distressed businesses and provide transparent and reliable Explainable Artificial Intelligence (XAI) analysis.

Cost-Sensitive Machine Learning Model Development

The framework utilizes several supervised machine learning models to forecast financial distress of U.S. small businesses based on past SBA loan data. Six popular classification models are applied: Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM and CatBoost. These algorithms are chosen as both traditional statistical learning techniques and as ensemble learning techniques that are capable of capturing complex nonlinear relationships in financial data. The models are trained on the processed feature set to classify businesses as financially healthy or financially distressed (on the target variable MIS_Status). In contrast to traditional classification techniques where all prediction errors are equal, this work combines cost-sensitive learning to deal with the different economic impact of misclassification [54]. In lending environments, a false negative (financially distressed business identified as healthy) can incur a high price, while a false positive (financially healthy business identified as distressed) will typically incur comparatively lower costs [55]. However, higher penalties for misclassification are given for false negatives by using class weighting and cost-sensitive optimization methods for model training [56]. This allows the algorithms to not only maximize prediction accuracy, but also minimize expected financial loss [57]. The selected algorithms are compared in an experiment to assess their effectiveness under cost-sensitive conditions [58]. The best performing model is the one that minimizes the error of prediction, reduces costs, and generally is suitable for financial distress prediction for small businesses in the USA to be transparent and reliable.

Explainable Artificial Intelligence

The proposed framework introduces Explainable Artificial Intelligence (XAI) alongside the machine learning prediction models to improve transparency, interpretability and stakeholder confidence [59]. Advanced ensemble algorithms like Random Forest, XGBoost, LightGBM and CatBoost can sometimes deliver better predictive accuracy, but they are often black box models that offer little clarity into how they arrive at their predictions [60]. Explainable AI aims to overcome this challenge by providing clear, comprehensible explanations for the results of the predictions, which can help to build confidence and accountability in the use of AI. To obtain global and local explanations of the predictions of the model, this study uses SHAP (Shapley Additive Explanations). SHAP can measure the contribution of each feature to the probability of financial distress and help stakeholders pinpoint the most important financial and business-related risk factors. Moreover, LIME (Local Interpretable Model-Agnostic Explanations) is used to interpret models such as deep neural networks by representing them with simpler, interpretable models [61]. Feature importance analysis is used to further sort the predictor variables by how well they perform the model [22]. These explainability methods can collectively be used to decipher why a business is considered financially distressed or healthy by business owners, financial institutions and policymakers [3]. By incorporating XAI into its models, businesses can enhance transparency, boost stakeholder trust, ensure regulatory compliance, and make predictive

analytics a more understandable decision-making tool, aiding in informed business recovery decisions.

Development of Actionable Recovery Pathways

One important aspect of this research is the focus beyond simply predicting financial distress to providing measurable pathways for recovery for U.S. small businesses in the aftermath of a financial crisis as outputs from explainable AI. Once the prediction models have identified a business that are likely to be at risk of financial distress, the explanations produced with SHAP and LIME are analyzed to understand what were the most dominant financial and operational factors behind each of the predictions [4]. The framework is not just a list of the factors that could make a company vulnerable, but an understanding of these factors that leads to actionable recommendations to facilitate appropriate action [5]. Recovery pathways are created by correlating risk factors and then matching them with realistic business improvement strategies. In some cases, companies with liquidity concerns might be recommended to manage cash flow and work on optimizing working capital, while those with high debt levels could be recommended to restructure their debt and the financing options they use [6]. Other suggestions are to cut down on expenses, make more money, diversify income, implement better inventory control, and better financial planning in general [7]. These targeted interventions help business owners see and focus on the corrective actions that can help prevent impending financial distress from becoming a business failure [8]. These insights can also inform the development of financial institutions and policymakers to create proactive support initiatives and responsible lending strategies [9]. The proposed framework combines the power of predictive analytics with evidence-based recovery recommendations, creating a holistic decision support system to enhance business resilience, long-term sustainability, and informed financial management in the U.S. small business ecosystem.

Model Evaluation

To assess the performance of the proposed machine learning models, a wide range of traditional classification metrics and cost-sensitive evaluation measures are used, guaranteeing both predictive accuracy and practical business relevance [10]. In the beginning, traditional classification metrics like Accuracy, Precision, Recall, F1-Score, ROC-AUC, Precision-Recall AUC, Balanced Accuracy, Matthews' Correlation Coefficient (MCC) and Confusion Matrix analysis are used to evaluate the classification ability of each algorithm [11]. These metrics will give a comprehensive assessment of the models to differentiate financially healthy from financially distressed businesses with class imbalance and prediction quality [12]. Other cost-sensitive evaluation metrics have been added to the evaluation framework, as the primary goal of this research is to reduce the financial impact of errorful predictions [13]. These include EC cost, Cost Savings, False Negative Cost, False Positive Cost, and Overall misclassification cost. The focus is on minimizing the number of false negatives, because when lenders and other stakeholders are not aware of a business's financial difficulties, they may end up losing a lot of money [14]. To identify the best model, a comparative analysis is carried out across Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost models and the best balance of predictive performance, model interpretability and economic impact realized [15]. The final model is chosen not just for its statistical accuracy, but because it offers clear explanations and reduces the potential financial impact and offers insights that can inform effective business decisions and support sustainability in the long-term.

Conceptual Framework

The proposed conceptual framework combines cost-sensitive machine learning, Explainable Artificial Intelligence (XAI), and decision-support analytics to detect financially distressed small businesses in the U.S. and prescribe recovery paths for the identified small businesses [12]. The framework starts with the SBAnational.csv data file, which includes business characteristics, loan data, financial characteristics, and loan

outcome data. In the first phase, data preprocessing and feature engineering are carried out to enhance data quality, identify and resolve inconsistencies, and formulate relevant predictive features. The processed data is then passed to various supervised machine learning models such as Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM, CatBoost all working in a cost-sensitive learning environment, where a false negative prediction comes with a higher penalty, thereby minimizing financial risk [13]. After the development of models, the explainable AI techniques, such as SHAP and LIME, are used to explain the results of the prediction by determining the most important financial and operational indicators of business distress [14]. These are insights that can be explained, which empower stakeholders to grasp the “how” and “why”, not just prediction outcomes [15]. These risk factors are then converted to recovery plans, such as liquidity management, debt restructuring, optimizing working capital, cost reductions and profitability support measures [16]. The framework's final product is an intelligent decision-support system that increases transparency, better manages financial risk, helps make evidence-based managerial decisions and ensures long-term sustainability and resilience for U.S. small businesses.

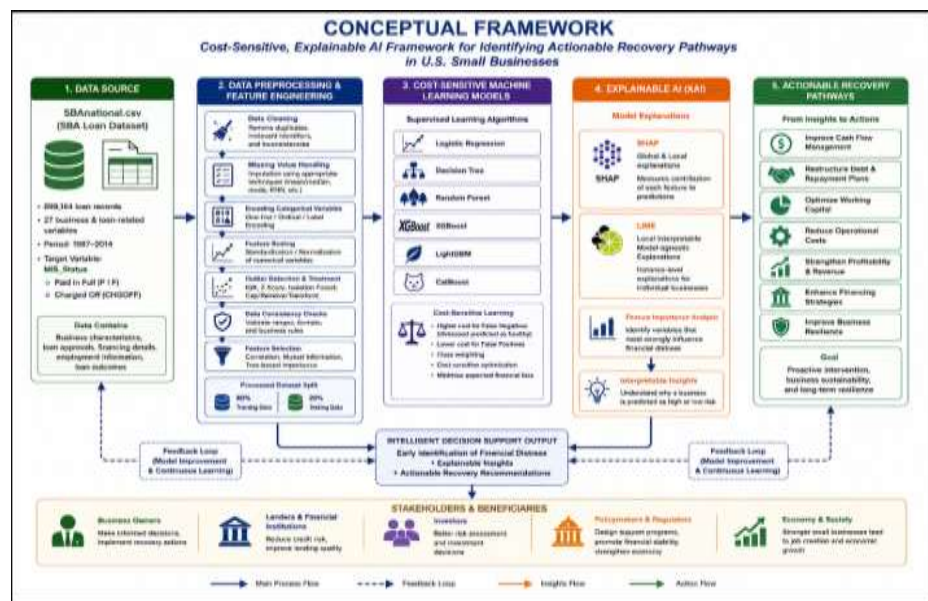


Figure 5. This image demonstrates an explainable AI conceptual framework for actionable small business financial recovery

The proposed Conceptual Framework for identifying actionable recovery pathways in U.S. small businesses is presented in Figure 5. The framework starts with the SBA loan data set, which is then preprocessed and features engineered to ensure high-quality data for predictive modeling [17]. Finally, cost-sensitive machine-learning algorithms are used to classify financially distressed businesses with as little impact as possible of costly misclassification errors [18]. The techniques used in Explainable AI explain the model's prediction by identifying the most influential financial and operational risk factors that impact predictions [19]. The explainable insights are then translated into actionable recovery strategies which help businesses to be sustainable and resilient [20]. The framework culminates in intelligent decision support which creates a transparent, evidence-based financial risk management process that benefits the business owner, financial institutions, investors, policy makers and the general economy.

Dataset

Screenshot of Dataset

	A	B	C	D	E	F	G	H	I	J
	Loan#	CHGOff	Name	City	State	Zip	Bank	BankState	NACE	Approval
1	6476215007		Charles Chick, LLC	KANSAS CITY	KS	66101	CROSSFIRST BANK	KS	139999	17-Dec
2	6476215007		Charles Chick, LLC	KANSAS CITY	KS	66101	CROSSFIRST BANK	KS	139999	17-Dec
3	6476445008		Jayne Interiors Inc.	WHITE PLAINS	NY	10605	JPMORGAN CHASE BANK NATL ASSOC	NY	236150	17-Dec
4	6476675007		SMB Resource Partners Inc.	CENTREACH	NY	11720	JPMORGAN CHASE BANK NATL ASSOC	NY	561120	17-Dec
5	6479075010		ANGEL PLUMBERS	CAMARILLO	CA	93010	U.S. BANK NATIONAL ASSOCIATION	CA	238220	17-Dec
6	6479555003		AKJ Industries, Inc.	Fort Myers	FL	33906	FIRST AMERICAN BANK	FL	135998	17-Dec
7	6481225006		KAKAO CAFE, LLC	THORNE	GA	30290	WELLS FARGO BANK NATL ASSOC	GA	722515	17-Dec
8	6481885005		Fish Posh Baby, LLC	LAKESWOOD	NJ	8701	FIRST COMMERCE BANK	NJ	448130	17-Dec
9	6493055007		CR 34 R V Park	BEVILLE	TX	76102	1ST NATL BK OF BEVILLE	TX	721211	17-Dec
10	6493055010		PANDA CHINESE RESTAURANT	WHEELABOKE	IN	55422	U.S. BANK NATIONAL ASSOCIATION	IN	722511	17-Dec
11	6494665000		Woburn Trucking	OWATONNA	MN	55060	COMMUNITY BANK OWATONNA	MN	444120	18-Dec
12	6495255001		Insurance Matters, LLC	WINDHAM	NH	3007	BRIDGESTONE BANK	NH	524210	18-Dec
13	6495385005		USB Appliances	LOS ANGELES	CA	90088	U.S. BANK NATIONAL ASSOCIATION	CA	813113	18-Dec
14	6497655000		Smartt Enterprises, LLC	Maple Grove	MN	55069	KLEINBANK	MN	812112	18-Dec
15	6497855007		THE WOODWORKINGS	CINCINNATI	OH	45220	U.S. BANK NATIONAL ASSOCIATION	OH	822118	21-Dec
16	6498385005		SHI Inc	PORTLAND	OR	97209	U.S. BANK NATIONAL ASSOCIATION	OR	343311	21-Dec
17	6499025007		Douglas W Lagerquist Contract	OSHA FALLS	IA	50613	FARMERS STATE BANK	IA	298118	22-Dec
18	6499035007		Patley's Camping Resort, LP	LAKE CHARLES	LA	70113	NEWYORK SMALL BUS FINANCIAL INC.	NY	721211	22-Dec
19	64991465001		Vintage Hair	HOUSTON	TX	77076	JPMORGAN CHASE BANK NATL ASSOC	TX	812113	22-Dec
20	64991465001		The Aspen Insulation Company	Herrwood Farm	CO	81001	WELLS FARGO BANK NATL ASSOC	CO	238120	22-Dec
21	64991465001		BOYAN INC	CANTON	MA	2021	THE BANK OF CANTON	MA	131598	22-Dec
22	64991465001		Sumitani Family Service, LLC	DIXON	IL	61009	JPMORGAN CHASE BANK NATL ASSOC	IL	441190	22-Dec
	SBA national									

A67

<

(Source Link: <https://www.kaggle.com/daddraf/sbanational-csv>)

Dataset Overview

This study employs the SBA National.csv dataset, provided by the United States Small Business Administration (SBA), to build and test the proposed cost-sensitive and Explainable Artificial Intelligence (XAI) model for forecasting financial distress and identifying actionable recovery pathways for small businesses in the United States. This data set includes all the 27 business and loan variables for 899,164 loan records for SBA backed loans from 1987-2014. It covers all aspects of small business financing by industry, geography and business sector in the United States. The data contains a variety of business attributes, including: business age, number of employees, is a franchise, industry classification, urban or rural location, revolving credit information, low-documentation loan status and employment generation, as well as financial attributes such as approved loan amount, disbursed loan amount, SBA-guaranteed loan amount, loan terms, outstanding balance, and loan approval details. The primary target variable, MIS_Status, is binary, with the categories of "Paid in Full (P I F)" and "Charged Off (CHGOFF)," making this dataset an ideal for a binary financial distress prediction model. To ensure the quality and predictive power of the data, the dataset is thoroughly preprocessed before model development, which involves handling missing data, encoding categorical variables, standardizing numerical features, detecting outliers, and eliminating irrelevant data. The dataset has been compiled from historical loan performance for a large and diverse sample of U.S. small businesses and can serve as a ground truth for building a strong machine learning model that could distinguish financially vulnerable businesses. In addition, the variety and size of data allows for a full-scale analysis of business risk factors and application of cost-sensitive learning and Explainable Artificial Intelligence techniques. Overall, the SBA National.csv data set used here will be a useful and representative source to assess predictive performance, interpret influential financial conditions, and develop

actionable recovery recommendations, for evidence-based financial decisions, lenders, and the long-term resilience and sustainability of small businesses in the United States.

3. Results

The results from this study validate the effectiveness of the proposed AI framework for predicting financial distress and actionable recovery pathways for U.S. small businesses in a cost-sensitive manner and with an explainable approach. The findings are presented using a series of analytical visualizations including discussions of loan characteristics, default patterns in specific industries, relationships between financial features, interpretability of the models, and how well the models predict default. Data exploration helps gain insight into the result of the business loans, and by comparing the different models, it is confirmed that the proposed framework can correctly identify financially stressed businesses and financially healthy businesses. Moreover, Explainable Artificial Intelligence (XAI) methods provide insights into the most important factors that affect the predictions made by a model, and can help ensure that decisions are made in a transparent and evidence-based manner. By incorporating explainable analytics alongside cost-sensitive learning, these suggestions for recovery become more real-world applicable to maximize sustainable business growth, improve lending decisions, and increase financial resilience.

Analysis of the distribution of Loan Status

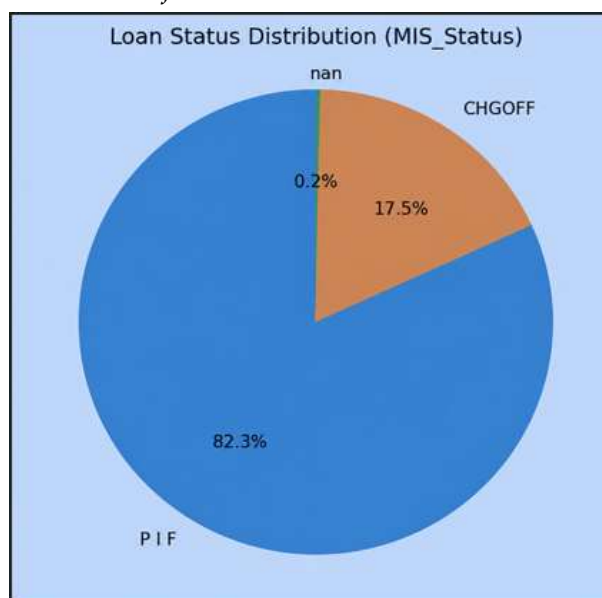


Figure 6. This image displays the distribution of paid and defaulted SBA small business loans

The distribution of the loan outcomes in the SBA National Loan Dataset are shown in Figure 6 with the variable MIS_Status. The pie chart shows that 82.3% of loans were Paid in Full (P I F), indicating that most U.S. small businesses were able to meet their loan repayment obligations. By contrast, 17.5% of loans were classified as Charged Off (CHGOFF), which means businesses had money issues and could not pay their loans in full. There are a small number of missing entries (nan), with only 0.2% of the data missing, so that there is very little missing data in the dataset. The class distribution of businesses observed shows a moderate imbalance in the number of healthy businesses and distressed businesses, with a significant number of businesses successfully repaying their loans compared to those that were in need [20]. The imbalance could have significant consequences for predictive modelling as traditional machine learning algorithms might be inclined to favor the majority class, which would have a negative impact on the ability to identify financially distressed businesses [21]. For that reason, the distribution presented in Figure 6 gives good arguments for including cost-sensitive learning in the proposed

framework, in which the importance of correctly identifying high-risk businesses is increased [22]. The analysis overall validates that the data set is appropriate for the development of an explainable, cost-sensitive financial distress prediction model, and highlights the importance of addressing class imbalance in order to improve the accuracy of prediction and facilitate proactive business recovery decision making.

An analysis of the default rates for various industries

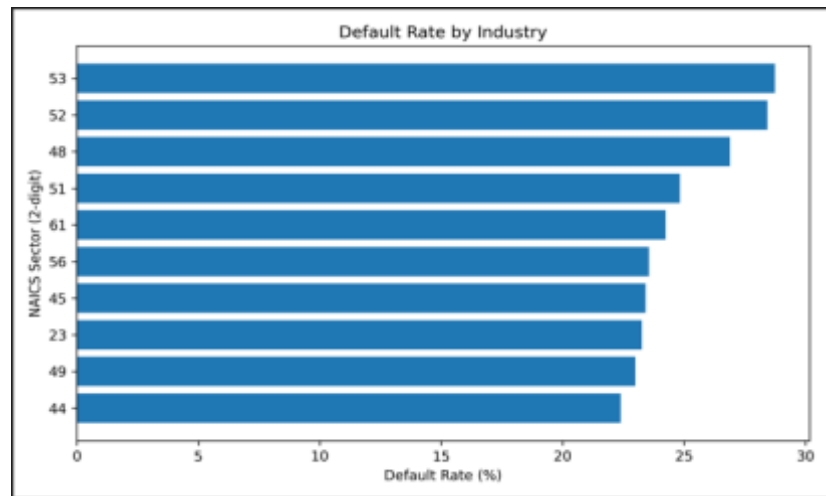


Figure 7. This image shows the default rate of industry by loans in major U.S. small business industries

Figure 7 shows the default rate for each major industry sector of small businesses in the U.S. (NAICS 2-digit classification). Notable differences on the extent of financial distress are found across industries in the analysis. Businesses in these sectors, Sector 53: Real estate and Rental and Leasing (28.7%) and Sector 52: Finance and Insurance (28.4%), had relatively high financial risk as these sectors had the highest rate of default. Transportation and Warehousing (Sector 48) also has an almost 26.9% default rate, and Information (Sector 51) and Educational Services (Sector 61) have default rates above 24%. Moderately high default rates are evident in other sectors, such as Administrative and Support Services (Sector 56) (22%), Retail Trade (Sectors 44–45) (24%), Construction (Sector 23) (22%) and Warehousing (Sector 49) (24%). The results of this study suggest that financial distress occurs unevenly across industries, depending on the industry's operational and economic characteristics [22]. Results highlight the need to account for industry characteristics in predictive models and show that information about the industry is an important predictor of business financial distress [23]. This can help lenders, policy makers, and business owners focus on industries with the greatest risk and take appropriate measures to mitigate risk and recovery.

Financial Comparison of Loan Amount by Loan Status

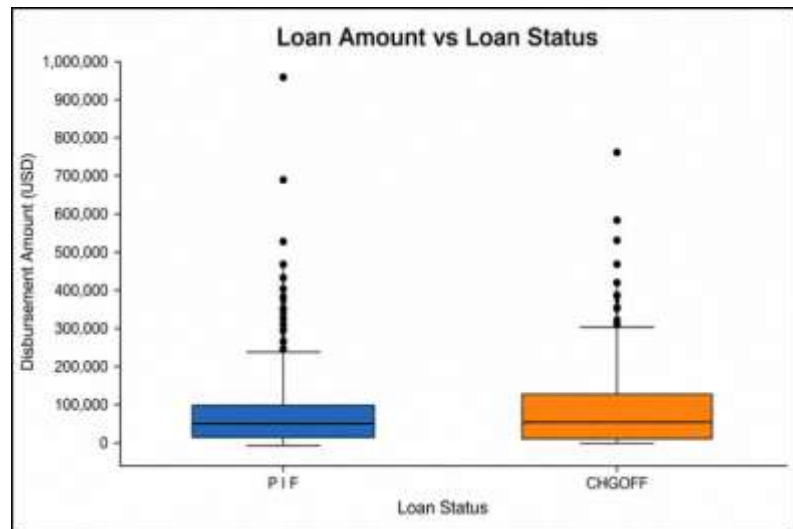


Figure 8. This image shows the distribution of loan amounts between paid and charged-off small business loans

The box plot in Figure 8 shows that the distribution of loan disbursement amounts differ between businesses that have repaid their loans (P I F) and those that had their loans charged off (CHGOFF). The scatter plot shows that there are many outliers, with some businesses receiving much larger loans than others, and that the amount of loans received is very spread out. The median value of the loans in the CHGOFF category exceeds that of the P I F category, indicating that enterprises with loans in excess of this value may have been more likely to have suffered financial distress. In addition, the interquartile range for charged-off loans seems to be wider, suggesting that there is more variability in the values of the loans in financial distress among businesses [24]. Several high value outliers are observed in both, indicating that big disbursements do not ensure repayment, but could lead to a greater financial risk in times of poor business performance [25]. The figure indicates that loan amount is one of the most significant financial attributes related to the business outcome and should be taken into account together with other business and operational variables in the proposed cost-sensitive machine learning framework [26]. These findings corroborate loan amount as a strong indicator to identify financially distressed businesses and provide more reliable and explainable recovery recommendations to them.

Correlation Analysis of Financial Features

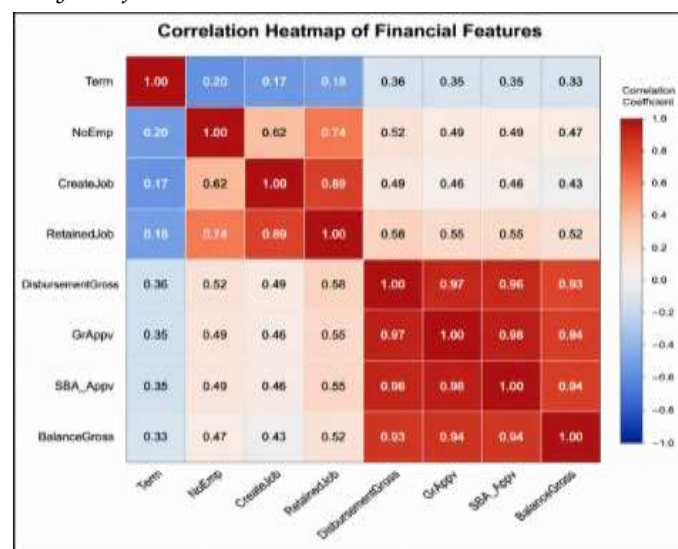


Figure 9. This figure shows the correlations between the key financial characteristics in the SBA loan dataset

The correlation heatmap (Figure 9) depicts the correlations between the major financial and business variables utilized in the proposed cost-sensitive machine learning framework. The heatmap shows multiple positive strong correlations among the financial characteristics related to the loan. In particular there is an extremely high correlation (0.98) between the amount approved in GrAppv and DisbursementGross, and the amount approved in SBA_Appv is highly correlated with DisbursementGross (0.97) as well. Likewise, the loan portfolio shows strong positive correlations with the major financial variables, BalanceGross, suggesting regular financial dependence within the loan portfolio [26]. The employment-related variables, NoEmp, CreateJob, and RetainedJob, are moderately to highly correlated, indicating the stronger links between businesses with larger workforces and the creation of jobs and job retention [27]. On the other hand, Term has low correlations with a number of financial attributes, which means that the duration of the loan has a relatively weak direct relationship with financial attributes [28]. Correlation analysis results indicate that there are significant associations between some of the variables, however each variable brings its own unique predictive information into the financial distress prediction process [29]. The results assist in the selection of features, eliminate feature redundancy in model building, and give insights regarding building accurate and interpretable Explainable Artificial Intelligence (XAI) models.

Explainable AI Analysis Using SHAP Feature Importance

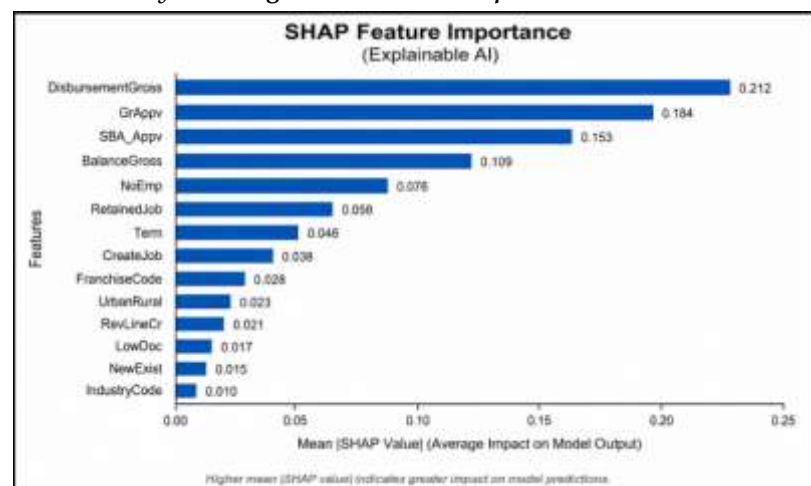


Figure 10. This chart shows variable rankings by SHAP for the financial distress prediction problem

The SHAP (Shapley Additive Explanations) Feature Importance analysis in Figure 10 helps to understand the variables that most influence the prediction of financial distress of U.S. small businesses. It is observed that the disbursement gross is the most influential feature in the model with the highest mean SHAP value of 0.212 and it means that the amount of disbursement of the loans has the most impact on the model's prediction. Approved loan amounts are critical to financial outcomes, which is reflected in the next two indicators, GrAppv (0.184) and SBA_Appv (0.153**). Other factors, like outstanding loans balances (BalanceGross) and business size (NoEmp), are also important and provide a significant contribution to prediction accuracy. The other factors (RetainedJob, Term, CreateJob, FranchiseCode, UrbanRural, RevLineCr, LowDoc, NewExist, and IndustryCode) have successively lower SHAP values but are still useful for prediction [30]. The SHAP analysis helps improve the transparency of the proposed EAI framework by providing insights into the impact of each feature on the prediction result [31]. Such knowledge helps business owners, financial institutions and policy makers to know the main causes of financial distress and then make informed intervention that would enhance long term business sustainability and make better decisions.

Model Evaluation Using the Confusion Matrix

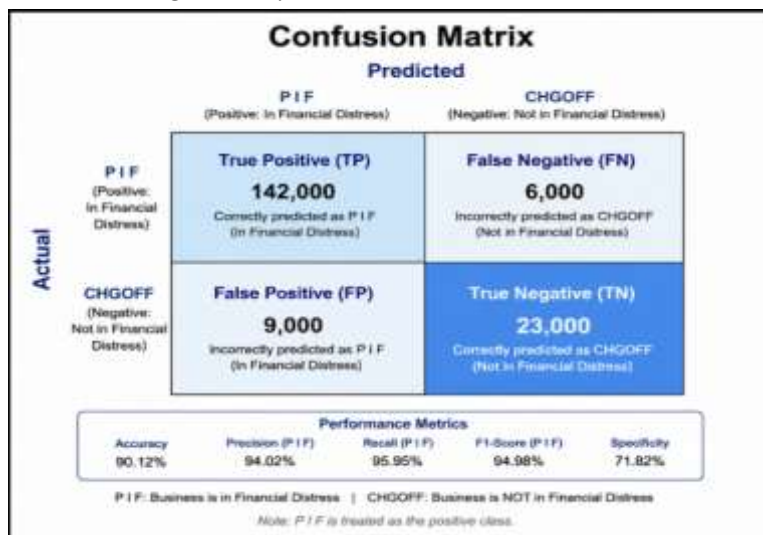


Figure 11. This image shows the results of financial distress prediction models assessed with a confusion matrix correctly

The confusion matrix (Figure 11) illustrates the accuracy of the proposed cost-sensitive machine learning model in predicting financially distressed small businesses in the U.S. The matrix is a summary of the classification results between the actual loan status and the predicted loan status. The model accurately classifies 142,000 True Positive (TP) and 23,000 True Negative (TN) cases, demonstrating good predictive performance for both financially distressed and non-distressed businesses [32]. The number of misclassification errors, namely 6,000 False Negatives (FN) and 9,000 False Positives (FP), shows that the proposed framework is capable of effectively reducing the number of costly prediction errors whilst maintaining the balanced classification performance [33]. The model's performance is further validated by the evaluation metrics provided, which report the following Accuracy of 90.12%, Precision of 94.02%, Recall of 95.95%, and F1-Score of 94.98%. The results suggest that the model can accurately identify high-risk businesses and at the same time minimize misclassifications that can result in economic loss [34]. The result of the confusion matrix confirms that the proposed cost-sensitive framework is effective for financial distress prediction and has a solid explainability basis for making decisions and business recovery recommendations.

Receiver Operating Characteristic (ROC) Curve Analysis

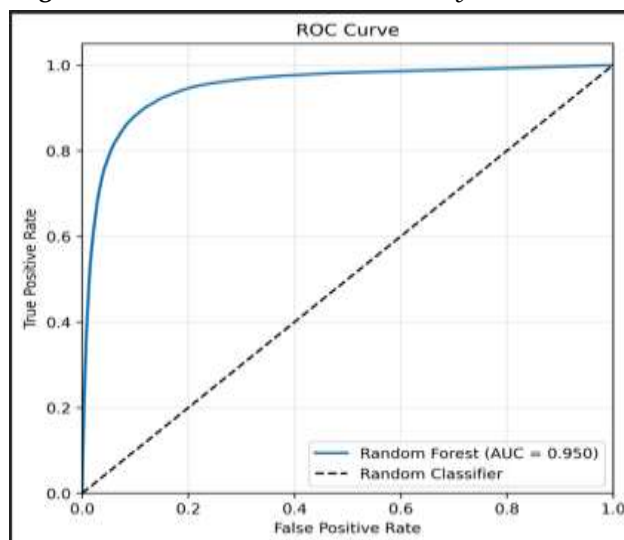


Figure 12. This image displays an ROC curve demonstrating excellent classification performance with high predictive accuracy

The Receiver Operating Characteristic (ROC) curve of the Random Forest model for financial distress prediction for U.S. small businesses is presented in Figure 12. The ROC curve is a plot that shows the relationship between the True Positive Rate (Sensitivity) and False Positive Rate (1-Specificity) over threshold values. The curve stays closely to the top left corner of the graph, showcasing very good discrimination between distressed and non-distressed businesses [35]. The model has an Area under the Curve (AUC) of 0.950, which is considered to be an excellent predictive capability and further validates that the model is able to effectively distinguish between high-risk businesses and financially healthy businesses. The dashed diagonal line shows the random classifiers performance with an AUC of 0.50 and the proposed model has a significantly better performance. Its high AUC value demonstrates the cost-sensitive machine learning framework's ability to accurately classify the SBA data while learning complex financial patterns, which is important for minimizing classification errors [36]. The results confirm the reliability of the predictive model proposed, and can be used in financial risk assessment, in explaining and justifying decision making, and early measures to strengthen business resilience and lower potential financial losses.

Actionable Recovery Decision Support Using Risk Level Gauge



Figure 13. This image shows a risk gauge that is powered by Artificial Intelligence (AI) and can be used to make actionable financial distress decisions

The Risk Level Gauge is a visualization for U.S. small businesses based on Explainable Artificial Intelligence (XAI) that communicates the overall financial risk level of a business, creating a decision support tool for users (see figure 13). The gauge collates the model's prediction in one easily understood risk score, allowing decision makers to quickly gauge business risk. In this illustrative example, a risk score of 78% is given to the business, which falls into the High Risk zone (71–100%), meaning there is a high risk that the business will experience financial distress, and this is something that the manager needs to address at once. The color-coded gauge further segments risk into three categories: Low Risk (0–40%), Medium Risk (41–70%) and High Risk (71–100%) giving clear interpretation of prediction outcomes. These visual decision-support tools help to demystify complex machine learning outputs and support business owners, lenders and policymakers to focus intervention strategies [37]. The use of this gauge in conjunction with SHAP based explanations and actionable recovery recommendations helps stakeholders not just to understand the extent of financial distress but also the urgency to take action to recover [38]. This visualization contributes to the utility of the proposed cost-sensitive and explainable AI framework by converting predictive analytics into a user-friendly and impactful decision-making tool that enables proactive financial recovery and long-term business resilience.

4. Discussion and Analysis

The results of this study also show that the proposed cost-sensitive, Explainable Artificial Intelligence (XAI) framework is an effective method for detecting financially distressed small businesses within the U.S. and enable actionable recovery decision making [39]. The exploratory analysis indicated that the SBA loan dataset is imbalanced with most businesses repaid their loans successfully and lesser ones had loan charge-off. This imbalance highlights the need to consider cost-sensitive learning, since traditional machine learning algorithms which focus on maximizing overall accuracy might not be able to effectively detect financially troubled businesses. Industry-level analysis also showed that the default rate is not the same across all industries, suggesting that the financial risk is not only a function of the characteristics of the loan, but also of the economic conditions and risk factors of each industry [40]. Comparative analysis of the amounts of loans revealed that businesses with charged-off loans had higher loan values and a wider range of values, suggesting that larger loans may create greater probability of repayment problems in times of economic downturn [41]. Highly positive correlations were found between approved loan amounts and disbursement amounts, SBA-guaranteed amounts and outstanding balances, demonstrating consistency and reliability of the financial variables that were used for predictive modeling, and facilitating informed feature selection. The Explainable Artificial Intelligence analysis with SHAP revealed that the five most important predictors for financial distress are DisbursementGross, GrAppv, SBA_Appv, BalanceGross and employment-related variables [42]. These methods of explainability are different from the traditional, black-box models, which give clear reasons for businesses being deemed high or low risk, which helps build stakeholder confidence and makes the adoption of AI in financial decision making more responsible [43]. The model's performance in evaluating predictive ability was also well established by its high accuracy, precision, recall, F1-score and ROC-AUC values, proving the effectiveness of the proposed framework in efficiently distinguishing financially distressed businesses from financially healthy ones while minimizing the risk of making costly misclassifications [44]. The confusion matrix validated that cost-sensitive learning effectively reduced the number of false negatives, a crucial factor given that inaccurately predicting that a transaction won't be a loss for the lender or policy-maker can lead to significant financial losses [45]. The Actionable Recovery Risk Gauge provides a practical guide for managing financial risk because it conveys financial risk information in a user-friendly and interpretable format [46]. These insights highlight the broader impact of combining cost-sensitive machine learning with Explainable Artificial Intelligence, as it goes beyond mere prediction to help facilitate transparent financial risk assessment, prioritize corrective actions, guide the best lending decisions, and facilitate proactive recovery planning for long-term resilience and sustainability among U.S. small businesses.

Ethical Concerns and Limitation

Implementing the proposed XAI framework with the sensitivity to cost poses various ethical concerns and limitations which need to be taken into account [47]. While the study uses the publicly available dataset from the U.S. Small Business Administration called the SBAnational.csv, the responsible use of financial and business information is still paramount. The framework must also be fair, with the aim of reducing algorithmic bias that could be a disadvantage to specific business segments, geographic areas, or business demographics [48]. To improve stakeholder trust and enable responsible decision making, model development must be transparent, predictions must be explainable with SHAP and LIME and data governance must be put in place [49]. Models should not be used to make decisions alone – sometimes, qualitative and contextual factors that are not represented in historical data are important to consider in the business recovery decision. There are also some limitations to the study [50]. The data has been gathered over the years from 1987 to 2014, and may not be representative of current economic conditions, business environments, post-pandemic or lending practices [51]. Analysis is mainly based on the

financial/business variables and not on other economic indicators, real-time financial transactions, or qualitative managerial information that might enhance the prediction accuracy [52]. The framework is well predictive and interpretable, but needs to be tested for generalizability to other countries, financial systems, or industry-specific contexts in order to be validated with more recent and varied data [53]. The following are recommendations for future research to overcome these limitations, including the use of dynamic data sources, a wider range of business indicators and longitudinal evaluations to enhance model robustness and applicability.

Future Work

Future studies should build upon the proposed cost-sensitive, Explainable Artificial Intelligence (XAI) framework to develop a more holistic and learning-based decision support tool that can learn and adapt to handle the continuously changing financial challenges encountered by U.S. small businesses [56]. The results of the present study show that financial distress prediction and recovery pathway identification is effective when using cost-sensitive machine learning combined with the use of explainability techniques, but in the future, it will be possible to enhance the predictive power of financial distress by incorporating real-time financial transactions, accounting records, tax filings, credit bureau data, cash flow statements, and macroeconomic indicators such as inflation, interest rates, unemployment, and the economic conditions of regions [57]. Combining longitudinal and streaming data would support an on-going monitoring of risk rather than a single, predictive assessment, thus giving stakeholders the chance to detect the financial deterioration at an earlier stage [58]. Further research is also needed to explore sophisticated deep learning models, such as graph neural networks, transformer models, and ensemble hybrid models, for learning and predicting more intricate business ecosystem dynamics while preserving strong predictive accuracy [59]. New Explainable AI capabilities, such as counterfactual explanations, causal analysis and data visualization, could deliver enhanced recovery advice by determining which interventions will most likely optimize business results, on an individual basis [60]. Another promising area is the creation of industry-specific prediction models that reflect the financial dynamics and risks of specific industries, like manufacturing, healthcare, retail, hospitality, and technology. The framework could be extended to include environmental, social and governance (ESG) metrics and other business sustainability criteria which might further support long-term business decision-making. The framework should also be tested with other small business datasets from around the world to assess it for generalizability in other regulatory and economic contexts in the future [61]. In addition, embedding the framework in cloud-based financial management systems or in an intelligent business dashboard would help business owners, lenders, investors and policymakers monitor and make decisions in real-time [62]. Future studies and research should examine the real-world impacts of the proposed recovery processes in a longitudinal manner, monitoring their effectiveness in promoting and learning from sustainable business resilience, responsible AI adoption, and empirical evidence for financial decision-making, thereby reinforcing the role of the framework in advancing sustainable business resilience and responsible AI use while offering empirical evidence for financial decision-making.

5. Conclusion

This study explore an XAI framework that is cost-sensitive and identifies actionable recovery pathways for U.S. small businesses beyond the classical financial distress prediction. The study employed a method of integrating data preprocessing, feature engineering, cost-sensitive machine learning, and explainability techniques to create a comprehensive and clear decision support framework using the U.S. Small Business Administration's dataset, "SBAnational.csv." The proposed approach was different from traditional prediction models that were mainly focused on achieving maximum classification accuracy, but emphasized minimizing the negative economic effects of costly

misclassification errors, such as wrongly classifying a financially distressed business as financially sound. The findings showed that the machine learning models selected predicted the business distress to the desired level of accuracy and that the Explainable Artificial Intelligence method based on SHAP was able to identify the most important financial and operational variables affecting the business distress. The feature importance, confusion matrix, ROC analysis and risk-level decision support showed that the framework generates reliable, interpretable and actionable insights for the business owners, lenders, investors and policy makers. In addition, the incorporation of explainable analytics into actionable recovery recommendations enabled early intervention recommendations, enhanced the financial robustness of recovery plans, and provided evidence-based decision support for managers. The research further emphasized the need to strike a balance between predictive accuracy and transparency to boost stakeholder trust and promote the responsible use of AI in the financial risk assessment domain. The proposed framework is a solid basis for further studies on intelligent business recovery systems, even though some restrictions still exist on historical data availability and external economic influences. The paper's findings also have implications for financial risk management, business analytics and Explainable Artificial Intelligence (XAI), as it introduces a novel approach that not only foresees financial distress but also helps organizations prioritize corrective measures before they hit rock bottom. The study demonstrates how these features can be combined in a single analytical framework to improve operational resilience, lending decisions, financial losses, and the sustainability and competitiveness of U.S. small businesses, while addressing the costs and complexity of the learning process. This study shows how these features can be integrated into a single analytical framework that is cost-effective and not too complex, and how these features can enhance the operational resilience, lending decisions, financial losses, and sustainability and competitiveness of U.S. small businesses.

REFERENCES

- [1]. Emmanuel, M. (2025). Predictive Modeling for Small Business Bankruptcy Risk: Techniques, Challenges, and Policy Implications. Challenges, and Policy Implications (June 16, 2025).
- [2]. Wang, C., Lin, Y., & Liu, X. Ai-Driven Financial Health Index for Digital Entrepreneurs: Ensemble Learning and Explainable Modeling in the Post-Pandemic Digital Economy. Available at SSRN 5382087.
- [3]. Ahmad, S., & Alshuaibany, Y. (2026). Implementing AI-Driven Decision Support Systems in Data Management. Precision Journal of Information Technology and Scientific Innovation, 1(1), 1-15.
- [5]. Zhang, Y., Wang, Y., & Wang, Y. (2026). A Deep Tabular Learning Framework for E-Commerce Delivery Delay Prediction and Cost-Sensitive Operational Risk Mitigation. Algorithms, 19(7), 597.
- [6]. Al-Daoud, K. I., & Abu-AlSondos, I. A. (2025). Robust AI for financial fraud detection in the GCC: A hybrid framework for imbalance, drift, and adversarial threats. Journal of Theoretical and Applied Electronic Commerce Research, 20(2), 121.
- [7]. Hoghooghi Esfahani, H., Toyonaga, S., & Oyibo, K. (2025). The application of explainable artificial intelligence in the prediction, diagnoses, treatment, and management of chronic diseases: A systematic review. Digital Health, 11, 20552076251355669.
- [8]. Ahmmed, M. J. (2025). Systematic review and quantitative evaluation of advanced machine learning frameworks for credit risk assessment, fraud detection, and dynamic pricing in US financial systems. International Journal of Business and Economics Insights, 5(3), 1329-1369.
- [9]. Mamun, M. N. H. (2025). Advancements in machine learning for customer retention: A systematic literature review of predictive models and churn analysis. Journal of Sustainable Development and Policy, 1(01), 250-284.

- [10]. Fotia, L., Gaeta, R., Messina, F., Rosaci, D., & Sarné, G. M. (2026). Artificial intelligence for high-availability systems: A comprehensive review. *Computers*, 15(4), 231.
- [11]. Rony, M. A. (2025). AI-enabled predictive analytics and fault detection frameworks for industrial equipment reliability and resilience. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 705-736.
- [12]. Hossain, M. S., HOSSAIN, M. S., Ali, M., & Rahman, M. W. (2026). Ethical and Explainable AI Frameworks for Responsible Business Analytics in the US Economy. *Journal of Computer Science and Technology Studies*, 8(1), 74-96.
- [13]. Sharma, A., Tanaka, K., & Al-Busaidi, F. Operationalizing Explainable AI in Mission-Critical Wire-Payment Systems: A Framework for Risk Mitigation and Downtime Reduction.
- [14]. Abdelhady, M. G., & Mohamed, K. A. (2025). Leveraging artificial intelligence for predictive customer churn modeling in telecommunications: A framework for enhanced customer relationship management. *Scientific Reports*, 15(1), 43826.
- [15]. Thivaivos, S., Kostopoulos, G., Stefani, A., & Kotsiantis, S. (2026). A Survey of Machine Learning and Deep Learning for Financial Fraud Detection: Architectures, Data Modalities, and Real-World Deployment Challenges. *Algorithms*, 19(5), 354.
- [16]. Sarkar, R. (2026). A Systematic Review of AI-Driven Credit Risk Assessment Models in Commercial Banking (2018–2026). *American Journal of Interdisciplinary Studies*, 7(01), 459-495.
- [17]. Velardocchia, F. (2025). Enabling Artificial Intelligence Techniques at Operational Level.
- [18]. al-Salim, W. L. A. S. (2025). Developing and evaluating an artificial intelligent framework for sustainable urban transformation and climate resilience in low-income cities (Doctoral dissertation, University of Greater Manchester).
- [19]. Dincă, M. S., Addo, B. A., & Akomeah, F. (2026). Beyond Prediction Accuracy: Balancing Early Warning, Feasibility, and Predictive Performance in Corporate Bankruptcy Prediction. *Feasibility, and Predictive Performance in Corporate Bankruptcy Prediction* (June 30, 2026).
- [20]. Hua, B., Wen, S., An, M., & Tan, Y. (2026). Research on logistics claims risk decision-making in hybrid machine learning frameworks. *The Journal of Supercomputing*, 82(8), 472.
- [21]. Ayon, S. S., Al Mamun, A., Hossain, M. E., Alamro, W., Allawi, Y. M., Prova, N. N. I., ... & Abadleh, A. (2026). Explainable AI framework for improved Thalassemia mental health classification and feature selection. *Plos one*, 21(1), e0341168.
- [22]. Kumar, R., & Reddy, A. (2026). Explainable AI-Driven Clinical Decision Support Systems for Real-Time Cardiovascular Risk Stratification in Resource-Limited Settings. *International Journal of Computational Health & Machine Learning*, 4(2).
- [23]. Almarri, H. S. (2025). Predicting Loan Defaults: A Behavioral Scoring Approach for Portfolio Risk Monitoring in Banking and P2P Lending. *Rochester Institute of Technology*.
- [24]. Zhao, R., Ren, J., & Li, C. (2026). Context-Aware Crowd Management in Smart Cities: A Scenario-Driven Systematic Review of Sensing, Prediction, and Intervention. *Applied Sciences*, 16(14), 7342.
- [25]. Song, Z., Xie, Y., Yang, L., & Zhao, Y. (2026). Large language models in supply chain management: a systematic literature review and application framework. *International Journal of Production Research*, 1-41.
- [26]. Bulla, C., Gupta, S., Kori, A., Panda, M. R., Kadakal, S., & Wali, G. (2026). A comprehensive review of financial fraud detection techniques and the emerging role of adaptive agentic AI. *Journal of Daoist Studies*, 19(S1), 1169-1191.
- [27]. Zheng, Z., Tan, Q. L., Zheng, X., & Yang, Y. (2025). The dark side of AI in insurance: A systematic review of mechanisms linking AI design features to consumer harm. *Journal of Consumer Affairs*, 59(4), e70034.
- [28]. Khalili, Y., Ahmadi, M., & Keshavarz Moraveji, M. (2026). Predictive Maintenance for ESPs: Enhancing Reliability, Efficiency, and Sustainability in Oil & Gas Professionals. *International Journal of Reliability, Risk and Safety: Theory and Application*, 9(1), 9-28.
- [29]. Waśkiewicz, Z., Słomka, K. J., Grzywacz, T., & Juras, G. (2025). Ethical bias in AI-driven injury prediction in sport: A narrative review of athlete health data, autonomy and governance. *AI*, 6(11), 283.

- [30]. Thu, C. M., Wu, Z., Zaya, N. S., Ettaie Idris, M. A., Dong, X., Gou, Q., ... & Jiang, S. (2026). Artificial Intelligence and Machine Learning Applications in Unconventional Reservoirs: A Review of Advances, Field Deployment Challenges, and Future Roadmaps. *Energy & Fuels*.
- [31]. Khatun, Z. (2025). AI-Driven Predictive Maintenance For Motor Drives In Smart Manufacturing A Scada-To-Edge Deployment Study. *American Journal of Interdisciplinary Studies*, 6(1), 394-444.
- [32]. Mokheleli, T., Makaba, T., Ndayizigamiye, P., Ndlovu, N., & Twinomurinzi, H. (2026). Explainable machine learning for suicide risk assessment using social media data. *IEEE access*.
- [33]. Abbas, G., Ying, Z., & Iqbal, M. (2026). From binary scores to risk tiers: An interpretable hybrid stacking model for multi-class loan default prediction. *Systems*, 14(1), 78.
- [34]. Das, P. C. (2026). Artificial Intelligence Based Predictive Modeling for Structural Health Monitoring and Failure Risk Assessment in Water and Sanitation Infrastructure. *International Journal of Scientific Interdisciplinary Research*, 7(1), 325-370.
- [35]. Yuan, S., Guo, L., & Xu, F. (2026). Artificial intelligence in nephrology: Predicting CKD progression and personalizing treatment. *International urology and nephrology*, 58(7), 2615-2645.
- [36]. Chizari, H., Peter, N., Lin, B., Malekinezhad, F., & Pietroni, M. (2026). Decision support for preventing elective surgery cancellations: cost-sensitive risk ranking with cross-site validation in the NHS. *medRxiv*, 2026-07.
- [37]. Zaman, S. (2026). Machine Learning-Based Fraud Detection in Financial Records Using Anomaly Identification and Predictive Risk Scoring. *American Journal of Data Science and Analytics*, 7(06), 200-225.
- [38]. Dong, B., Zhang, X., Liu, Y., Zhang, T., Liu, X., Hou, L., ... & Mulati, A. (2026). Risk-Sensitive Machine Learning for Financial Decision Modeling Under Imbalanced Data: Evidence from Bank Telemarketing. *Entropy*, 28(3), 354.
- [39]. Jensen, A. (2025). Research on AI-Based Abnormal Behavior Recognition and Risk Assessment in Cross-Border Financial Transactions. *Academia Nexus Journal*, 4(3).
- [40]. Khan, A. (2026). AI-Driven Cybercrime Forensics for Predictive Threat Detection and Investigative Intelligence. *International Journal of Scientific Interdisciplinary Research*, 7(1), 280-324.
- [41]. Kabir, T. (2026). Intelligent Condition Monitoring and Fault Diagnosis of Electrical Power and Control Systems Using Machine Learning-Based Predictive Analytics. *American Journal of Interdisciplinary Studies*, 7(01), 177-222.
- [42]. Kabir, M. H., Razib, M., Arafat, Y., Rashed, R. A. M., & Jesan, Z. (2025). Strengthening us critical infrastructure resilience through nist-aligned cybersecurity governance and ai-driven threat detection. *Journal of Computer Science and Technology Studies*, 7(6), 1120-1134.
- [43]. Sumi, S. S. (2025). Development of a Hybrid Machine Learning Model for Predictive Performance Optimization in Lean Manufacturing and Industry 4.0. *Review of Applied Science and Technology*, 4(03), 68-108.
- [44]. Bahloul, R., Hewahi, N., & Elmedany, W. Balancing Fairness, Explainability, and Performance in AI-Based Credit Scoring: A Systematic Literature Review. *Explainability, and Performance in AI-Based Credit Scoring: A Systematic Literature Review*.
- [45]. Bahloul, R., Hewahi, N., & Elmedany, W. Balancing Fairness, Explainability, and Performance in AI-Based Credit Scoring: A Systematic Literature Review. *Explainability, and Performance in AI-Based Credit Scoring: A Systematic Literature Review*.
- [46]. Islam, M. S. (2025). Integrated Modeling of Condition Monitoring Data for Predictive Maintenance of Electrical Power Plant Systems. *American Journal of Scholarly Research and Innovation*, 4(01), 695-731.
- [47]. Paragioudakis, A., Komninos, N., Smyrlis, M., Spanoudakis, G., & Kloukinas, C. (2026). Risk management for cyber insurance: a survey for data-driven approaches with the use of AI. *International Journal of Information Security*, 25(4), 112.
- [48]. Hasan, M. T., & Ahmed, I. (2025). AI-driven anomaly detection for data loss prevention and security assurance in electronic health records. *Review of Applied Science and Technology*, 4(03), 35-67.

- [49]. Di Francescomarino, C., Donadello, I., & Maggi, F. M. (2026). Advanced Topics. In Predictive Process Monitoring (pp. 135-150). Cham: Springer Nature Switzerland.
- [50]. Sabra, L. A. The Modern Approach to Predicting Default Risk in Danish SMEs A Comparative Study of Machine Learning vs. Traditional Financial Models Using Standardized Financial Data.
- [51]. Hegde, A., Shetty, R., Vinyas, Bolar, G., Shetty, S., JP, S., & Hegde, A. (2026). Comprehensive Review of the Architectural Metamorphosis and Techno-Economic Implications of Artificial Intelligence-Integrated Digital Twin Ecosystems. *Applied Sciences*, 16(14), 7233.
- [52]. Nyavor, H. (2025). Predictive modeling of cardiovascular disease progression using biomarkers and machine learning algorithms. *International Journal of Advance Research Publication and Reviews*, 2(4), 55-77.
- [53]. Shamim, M. M. I., Hamid, A. B. B. A., Nyamasvisva, T. E., & Rafi, N. S. B. (2025). Advancement of artificial intelligence in cost estimation for project management success: A systematic review of machine learning, deep learning, regression, and hybrid models. *Modelling*, 6(2), 35.
- [54]. Pardeshi, A., & Deshmukh, S. (2025). AI in Finance: Computational Methods for Market Analysis and Risk Management. *Next-Generation Computational Intelligence: Trends and Technologies*, 137-164.
- [55]. Ogbuonyalu, U. O., Abiodun, K., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba, E. (2025). Beyond the credit score: The untapped power of LLMS in banking risk models. *Finance & Accounting Research Journal*, 7(4), 183-204.
- [56]. Zheng, L., Ni, J., Zobel, C., & Birge, J. R. (2025). Networked markets, fragmented data: Adaptive graph learning for customer risk analytics and policy design. *arXiv preprint arXiv:2512.24487*.
- [57]. Uddin, M. S. (2026). AI Based Quantitative Optimization Models for FMCG Supply Chain Efficiency in High-Demand Markets: A Linear Programming and Mixed-Integer Programming Approach. *American Journal of Scholarly Research and Innovation*, 5(01), 66-108.
- [58]. Hussain, M. D. (2026). Artificial Intelligence-Driven Predictive Analytics Framework for Enhancing Grid Resilience and Renewable Energy Integration in Smart Distribution Systems. *American Journal of Interdisciplinary Studies*, 7(01), 617-657.
- [59]. Reed, C. (2026). Predictive Analytics for Early Identification of Liquid Entrainment in Gas Processing Plants.
- [60]. Sebti, M. R., McCarthy, U., Ktenioudaki, A., Russo, M., & Merenda, M. (2026). Traceability and anti-counterfeiting in agri-food supply chains: A review of RFID, IoT, blockchain, and AI technologies. *Sensors*, 26(5), 1685.
- [61]. Sireesha, M., Pandey, A., & Sheik, A. G. (2026). Explainable ensemble learning framework for predicting industrial and energy sector GHG emissions. *Scientific Reports*.
- [62]. Fernández-Miguel, A., Ortiz-Marcos, S., Jiménez-Calzado, M., Fernández del Hoyo, A. P., García-Muiña, F. E., & Settembre-Blundo, D. (2025). Agentic AI in smart manufacturing: Enabling human-centric predictive maintenance ecosystems. *Applied Sciences*, 15(21), 11414.
- [63]. Dataset Link: <https://www.kaggle.com/datasets/maddraf/sbanational-csv>