Abstract

The proposed project, FinSentinal, presents a Hybrid Data Fusion Model designed to predict the likelihood of corporate financial distress or bankruptcy before it occurs. The system integrates multiple heterogeneous data sources—financial ratios, market indicators, sentiment data, and macroeconomic metrics—to provide a comprehensive, data-driven assessment of corporate stability. By leveraging machine learning algorithms and explainable AI techniques, FinSentinal computes a Financial Distress Index (FDI) to identify potential risk patterns and issue early alerts for proactive decision-making.

Traditional financial risk detection methods, such as the Altman Z-score, are static and rely on limited accounting data, often failing to detect distress until it is too late. They do not incorporate alternative indicators such as market sentiment or dynamic liquidity trends, which are critical to understanding modern financial health. As a result, these models react after distress events rather than preventing them.

FinSentinal addresses these gaps by fusing structured financial and unstructured sentiment data to develop a multidimensional early warning framework. Using models like Random Forest, XGBoost, and LightGBM, the system classifies companies as 'Healthy' or 'Distressed' and generates explainable insights using SHAP and LIME. Additionally, a visualization dashboard helps users track distress probability, analyze feature importance, and take preventive action based on clear, interpretable intelligence.

Introduction / Background

Corporate financial distress, including bankruptcies and stock market collapses, poses significant risks to economies, investors, and institutions. High-profile fail1 ures, such as Lehman Brothers in 2008, highlight the need for early detection to mitigate losses. Recent advancements in artificial intelligence and access to diverse data sources, including social media sentiment and market indicators, enable more accurate predictions. FinSentinal leverages these trends to create a hybrid AI-driven framework that offers timely, transparent, and interpretable risk assessments

Problem Statement

Conventional methods for predicting financial distress, such as ratio-based models like the Altman Z-score, are limited by their reliance on historical financial data. They fail to incorporate dynamic signals like market volatility, investor sentiment, or macroeconomic shifts, resulting in delayed or inaccurate predictions. This project aims to develop a hybrid data fusion model that integrates multi-source data to provide early, accurate, and interpretable predictions of corporate financial distress, addressing the gaps in existing reactive approaches

Objectives

- Design a hybrid data fusion model integrating financial, market, sentiment, and macroeconomic data for early distress detection.
- Develop a machine learning framework to predict high-risk companies with over 80% precision.
- Compute a Financial Distress Index (FDI) to quantify company-level risk in real time.
- Implement SHAP and LIME for explainable AI insights into distress predictions.
- Create an interactive dashboard to visualize risk scores and historical distress trends.

Scope of the Project

The project focuses on developing a prototype for FinSentinal, covering data collection, model training, and visualization. It includes:

Phase 1 (Core Implementation):

• Collect and preprocess financial and market data via APIs (e.g., Yahoo Finance).

• Perform feature engineering and train baseline machine learning models.

• Compute the Financial Distress Index (FDI) and generate basic risk alerts.

Phase 2 (Extended Implementation):

• Integrate sentiment analysis and macroeconomic indicators.

• Implement explainable AI (SHAP/LIME) and develop a visualization dashboard.

• Optionally deploy the prototype on Google Cloud Platform for scalability. The project excludes real-time market feeds and multi-market analysis, which are planned for future enhancements.

Methodology / Approach

• **Data Collection**: Gather financial ratios, stock market data (Yahoo Finance), sentiment data (NewsAPI), and macroeconomic indicators.

• **Data Preprocessing**: Clean, normalize, and merge structured and unstructured data for consistency.

• **Feature Engineering**: Derive features like volatility metrics, liquidity ratios, and sentiment indices.

• **Model Training**: Train supervised machine learning models (Random Forest, XGBoost, LightGBM) for binary classification (Healthy vs. Distressed).

• **Model Evaluation**: Assess performance using precision, recall, F1 score, and ROC-AUC, targeting over 80% accuracy.

• **Explainability**: Apply SHAP and LIME to interpret feature contributions to predictions. • Visualization: Develop an interactive dashboard using Streamlit or Flask to display FDI, risk trends, and at-risk company rankings. The system follows a modular architecture, ensuring scalability and ease of integration for future enhancements.

Tools and Technologies

• Programming Language: Python 3.10

• Data Handling: Pandas, NumPy

• Machine Learning: Scikit-learn, XGBoost, LightGBM

• Explainable AI: SHAP, LIME • Visualization: Matplotlib, Seaborn, Plotly

• Web Dashboard: Streamlit or Flask

• Cloud Platform (Optional): Google Cloud Platform (Vertex AI, App Engine)

Expected Outcomes

• A hybrid machine learning model predicting corporate financial distress with over 80% accuracy.

• A Financial Distress Index (FDI) quantifying company-level risk scores.

• Transparent AI insights via SHAP/LIME, highlighting key risk drivers.

• An interactive dashboard for real-time monitoring of financial health and risk trends.

• A scalable framework for future integration with live financial systems.

Work Plan / Timeline

Phase	**Activities**
Week 1	Data collection and preprocessing
Week 1-2	Feature engineering and exploratory data analysis
Week 2-3	Model training, hyperparameter tuning, and evaluation
Week 4	Explainability integration and dashboard development
Week 4-5	Documentation and presentation preparation

Table 1: Project Timeline

References / Bibliography

• Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance.

• Chen, M. Y., & Du, Y. (2009). Using Neural Networks to Predict Corporate Financial Distress. Expert Systems with Applications.

• Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions (SHAP). NIPS. • Kaggle Datasets: Corporate Bankruptcy Prediction Dataset.

• Yahoo Finance API – Market Data Retrieval and Analysis.

APPENDIX A: REFERENCES

• Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance.

• Chen, M. Y., & Du, Y. (2009). Using Neural Networks to Predict Corporate Financial Distress. Expert Systems with Applications.

• Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions (SHAP). NIPS. • Kaggle Datasets: Corporate Bankruptcy Prediction Dataset.

• Yahoo Finance API – Market Data Retrieval and Analysis.