**Bankruptcy Prediction using Machine Learning**

# 1. Introduction

This report aims to predict bankruptcy using various machine learning models. We will discuss data preprocessing, model training, testing, validation, hyperparameter tuning, and model evaluation. We will compare model performance and provide recommendations.

# 2. Data Preprocessing

## 2.1 Handling Missing Values

The dataset had missing values which were handled by imputing the median values for operational margin, assets growth, sales growth, and employee growth columns. Whereas, null values were removed from EPS, liquidity, leverage ratio, Return on Equity, market book ratio.

A screenshot of a computer

Description automatically generated

## 2.2 Splitting the data

The dataset was split into 80% training and 20% test sets to ensure that the models were evaluated on unseen data and to fine-tune hyperparameters.

## 2.3 Dealing with class imbalance

An over sampling method SMOTE was used to deal with the data imbalance problem. It helped to increase the number of data points for the minority class to improve model performance.

## 2.4 Feature Scaling

Standard scaler was used to standardize the features to have zero mean and unit variance. This was essential due to the variance in the magnitude of different features.

# 3. Model Training and Evaluation

## 3.1 Logistic Regression

### 3.1.1 Model Description

Logistic Regression is a linear model suitable for binary classification problems like bankruptcy prediction.3.1.3 Evaluation

### 3.1.2 Evaluation

The Logistic Regression model achieved an accuracy of 68%, a precision of 50%, recall of 63%, and an F1-score of 42%.

## 3.2 Random Forest

### 3.2.1 Model Description

Random Forest is an ensemble method that combines multiple decision trees to improve accuracy.

### 3.2.2 Evaluation

The Random Forest model achieved an accuracy of 99%, a precision of 55%, recall of 72%, and an F1-score of 58%.

A screenshot of a computer

Description automatically generated

# Recommendation

**1. Target High-Risk Companies**: Focus on companies flagged with high bankruptcy probability (> 70%) for immediate intervention, such as offering restructuring plans or tightening credit terms.

**2. Adjust Credit Policies**: Implement dynamic credit policies. Companies predicted as high-risk should face stricter credit terms, higher interest rates, or additional collateral requirements.

**3. Prioritize Early Warning System**: Use model predictions to create an early warning system. Proactively monitor companies with moderate risk for signs of financial distress to prevent bankruptcy.

**4. Optimize Investment Strategy**: Utilize model results to prioritize investment in low-risk companies and reassess investments in high-risk companies to mitigate potential losses.

**5. Sector-Specific Risk Management**: Apply industry-specific insights from the model to adjust loan and credit strategies for sectors with higher bankruptcy risk, tailoring risk management strategies accordingly.

**Overall Conclusion:**

The Random Forest model substantially surpasses the SVM model in all evaluation metrics, exhibiting higher precision, recall, and F1-scores for both classes. With an overall accuracy of 99%, the Random Forest model delivers a more robust and dependable classification performance, making it the optimal choice for predicting bankruptcy in this dataset.