# Company Bankruptcy Forecasting

# INTRODUCTION

- The process of predicting the likelihood of a company entering into bankruptcy within a future time period.
- It involves the use of financial data, economic indicators, and predictive modeling techniques to assess a firm's financial health and predict potential insolvency.
- For creditors and investors, forecasting is a vital tool in risk assessment, enabling more informed decision-making regarding lending and investment strategies.
- Early identification of financial distress allows companies to take preemptive measures, potentially avoiding the dire consequences of bankruptcy.

# UNDERSTANDING BANKRUPTCY

**Chapter 11 - Reorganization:** Management retains control of daily operations but must seek court approval for significant business decisions, aiming to return the company to profitability under supervised guidance.

**Chapter 7 - Liquidation:** The business ceases all activities, assets are liquidated, and the company effectively goes out of business, with the proceeds distributed to creditors as per the court's liquidation plan.

**Economic Implications:** Market Dynamics, Credit and Lending

**Social Implications:** Employment, Community and Suppliers

# DATASET OVERVIEW

Origin of the dataset: SEC filings.

Timeframe: 1999-2018 to predict future values

Scope: 8,262 companies listed on NYSE and NASDAQ.

Total observations: 78,682 firm-year instances.

The dataset is fully complete, with no missing entries, artificial data, or values that have been filled in.

# PROBLEM STATEMENT

This project aims to develop a predictive model that uses financial, liabilities, market, and asset parameters from a 10-year historical dataset to forecast the bankruptcy status of companies over the next five years, enhancing decision-making for stakeholders and improving economic stability.

The target variable, 'status_label,' classifies companies into either 'Alive' or 'Failed,' reflecting their bankruptcy status for the next 5 years.

# FEATURE OVERVIEW

The dataset has 18 features spanning from X1-X18 representing different financial factors that contribute for each company's year wise bankruptcy status.

The 18 predictor variables can be categorized into groups as:

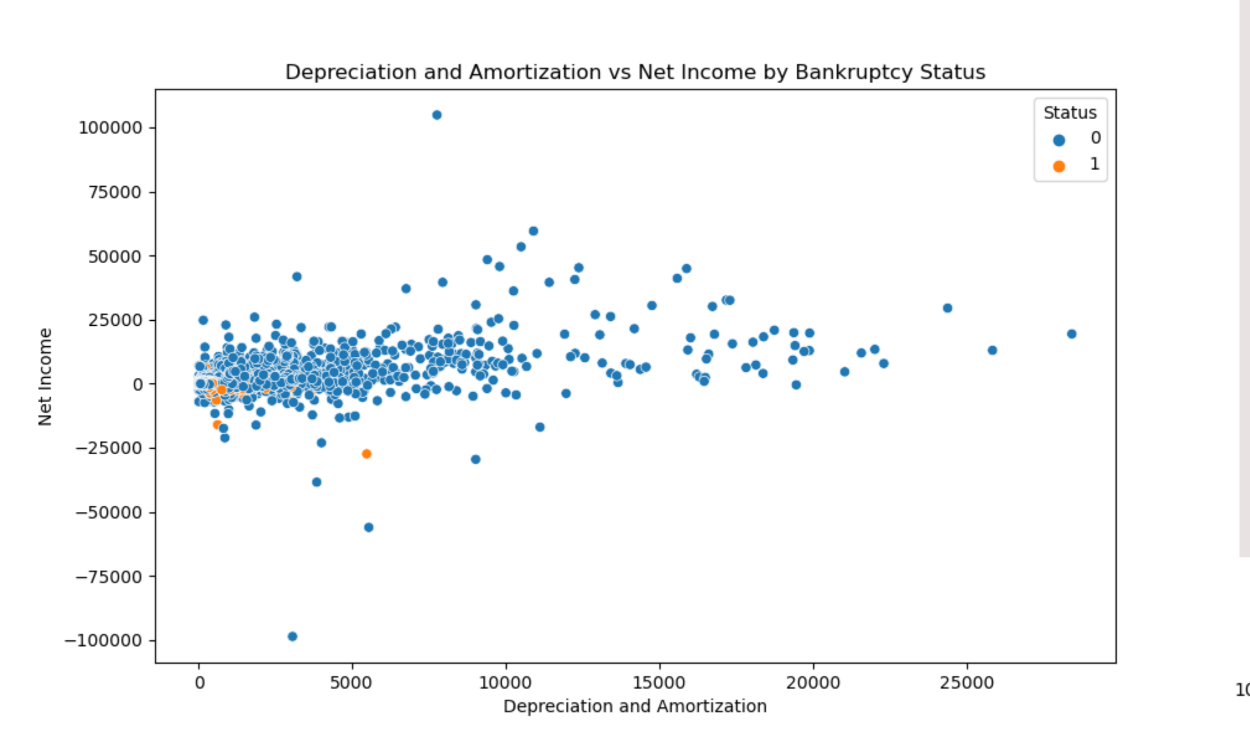
**Financial Performance Metrics:** X4 - EBITDA, X6 - Net Income, X12 - EBIT, X13 - Gross Profit, X16 - Total Revenue, X18 - Total Operating Expenses

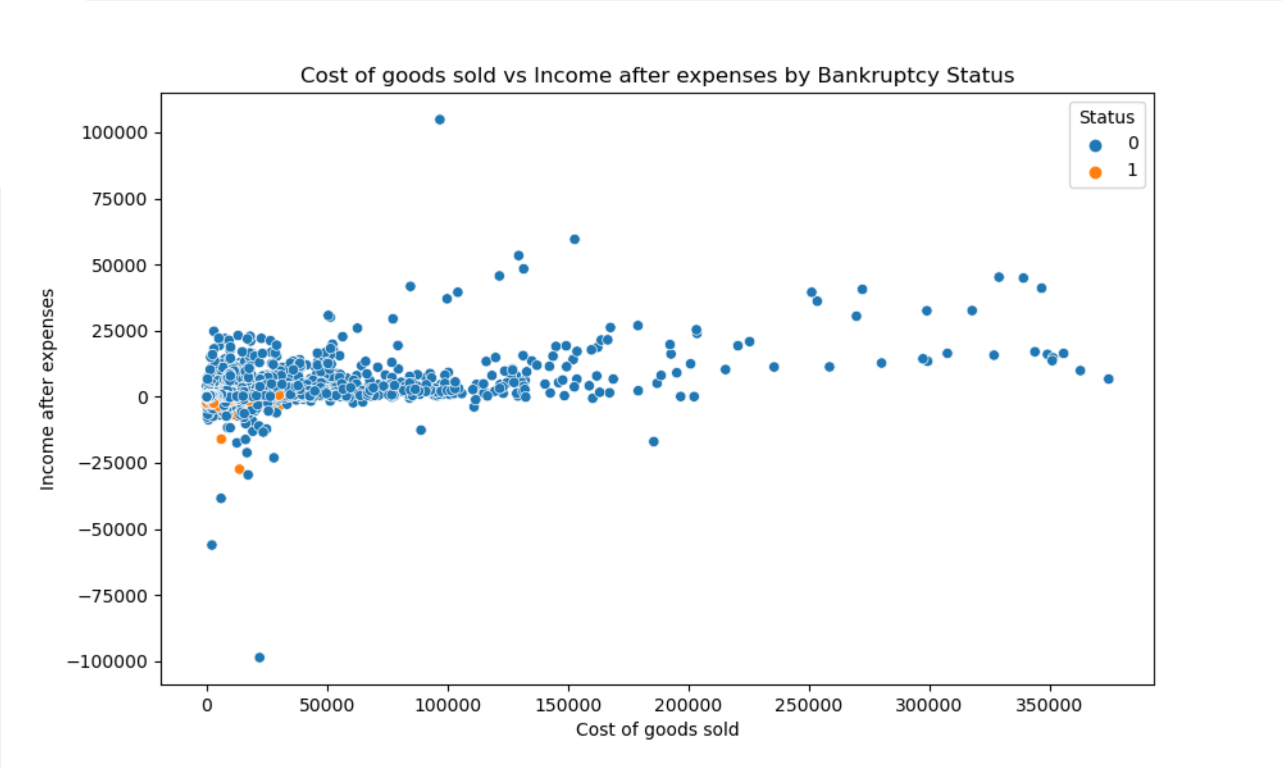
**Asset-related Metrics:** X1 - Current assets, X5 - Inventory, X7 - Total Receivables, X10 - Total assets, Liability-related Metrics: X11 - Total Long-term debt, X14 - Total Current Liabilities, X15 - Retained Earnings X17 - Total Liabilities

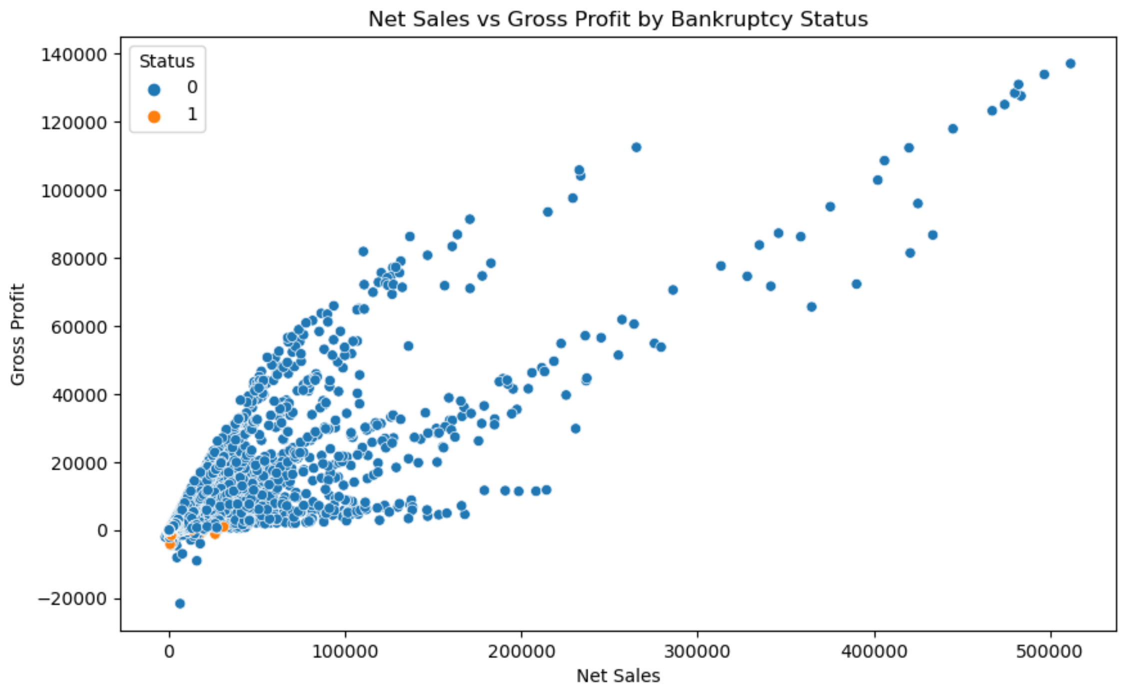
**Market-related Metrics:** X8 - Market value (Market capitalization)

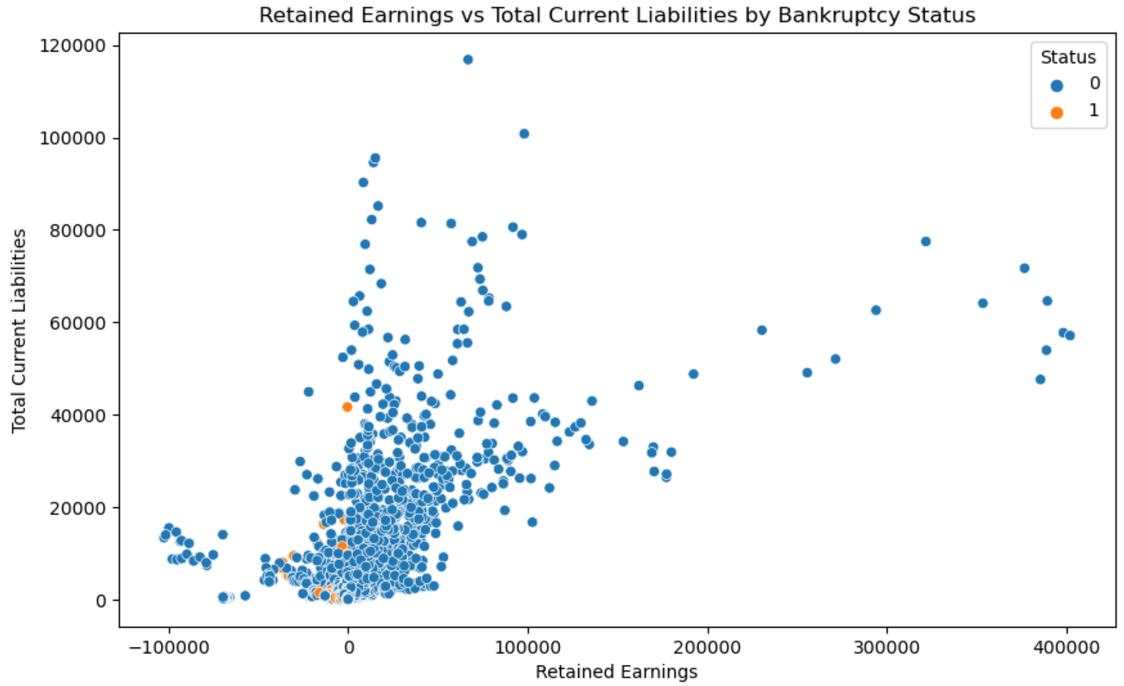
**Cost-related Metrics:** X2 - Cost of goods sold X3 Depreciation and amortization

**Revenue-related Metrics:** X9 - Net sales.


Depreciation and Amortization vs Net Income by Bankruptcy Status
Status
0
1
Net Income
Depreciation and Amortization


Cost of goods sold vs Income after expenses by Bankruptcy Status
Status
0
1
Income after expenses
Cost of goods sold


Net Sales vs Gross Profit by Bankruptcy Status
Status
0
1
Gross Profit
140000
120000
100000
80000
60000
40000
20000
0
−20000
0
100000
200000
300000
400000
500000
Net Sales


Retained Earnings vs Total Current Liabilities by Bankruptcy Status
Status
0
1
Total Current Liabilities
120000
100000
80000
60000
40000
20000
0
−100000
0
100000
200000
300000
400000
Retained Earnings

# EXPLORATORY DATA ANALYSIS(EDA)

1

**Tackling multicollinearity:** Highly correlated features like X9 with X13 and X16 and X4 with X12

2

**Feature Engineering :** Near-perfect correlation between X2 and X18, so a new variable X19 'Other costs' was added which accounts for X18 - X2, which relates to all non-goods cost, and X18 will then be dropped.

3

**Incorporating control variable:** Years 2008-2010 experienced abrupt financial conditions. A dummy variable/feature 'global_crisis' was added to differentiate between normal operation and those influenced by significant economic disruptions.

# DATA PREPROCESSING

**1. DATA STANDARDIZATION:**

Apply standard scaling to these numerical features implemented by StandardScaler() to ensure that all features contribute equally to the model's performance

**2. HANDLING CLASS IMBALANCE:**

Implemented an approach involving the resampling of the minority class to mitigate the bias towards the majority class and ensuring a more equitable representation of classes within our dataset.

The minority class was then subjected to an up-sampling process, wherein samples were randomly selected with replacement

**3. FEATURE SELECTION:**

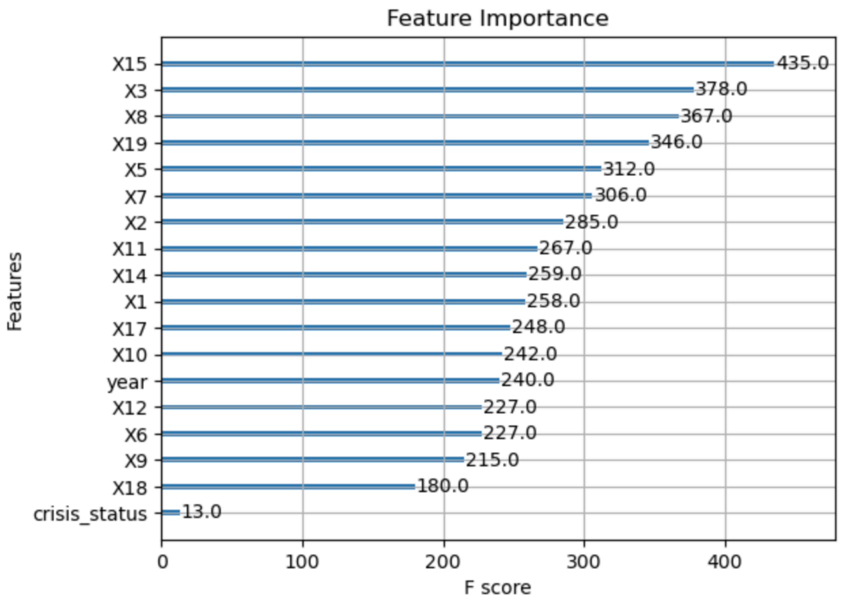
Features were selected based on their impact to the overall performance and to reduce the complexity of the model, improve its performance, and potentially avoid overfitting using XgBoost regressor.

4. **SEQUENCE GENERATION:**

A function create_sequences is used to generate sequences of 10 years of financial data for each company as features, and the subsequent 5 years' status as targets.

# FEATURE SELECTION

- **Top bankruptcy indicators:**
- Retained Earnings(X15)
- Depreciation and Amortization(X3)
- Market value(X8)

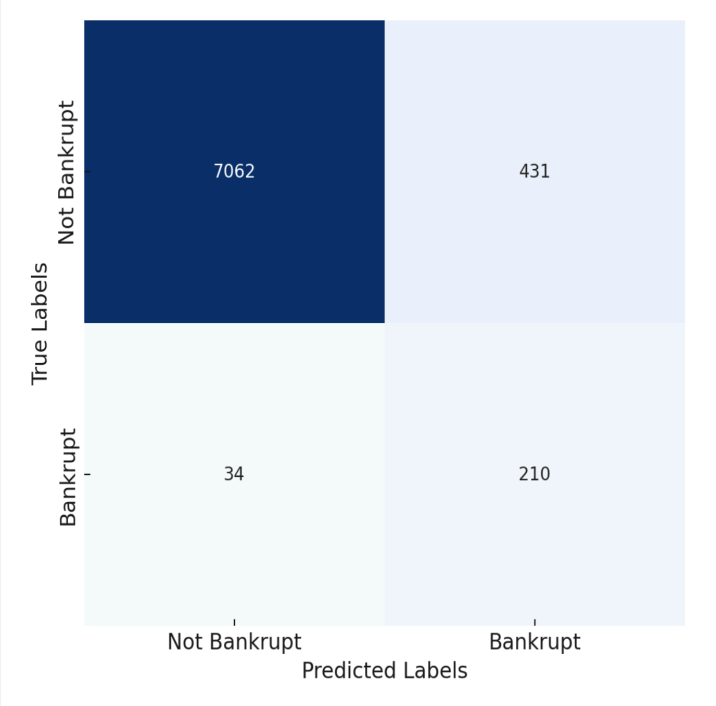


# MODELING APPROACHES

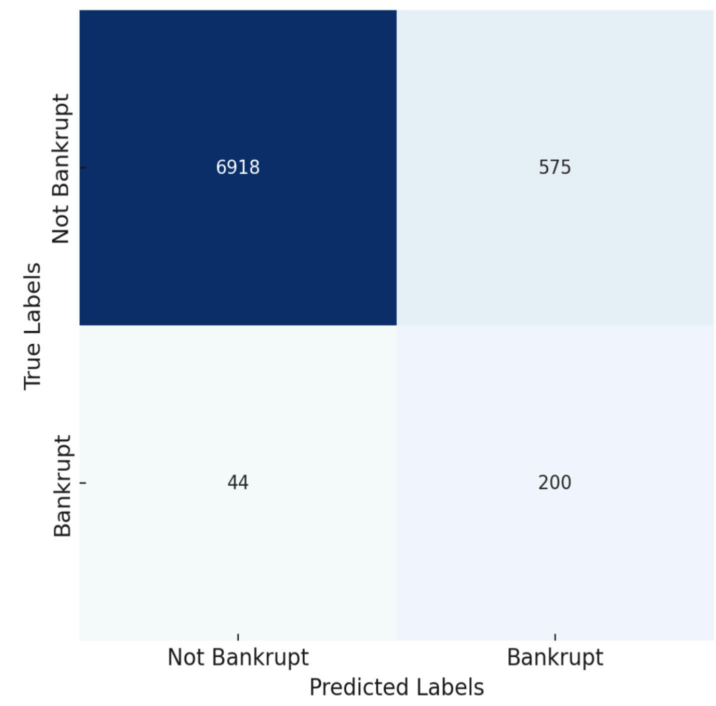
- GRU (Gated Recurrent Unit)
- LSTM (Long Short-Term Memory)
- Transformers.

# Model Evaluation and Comparison

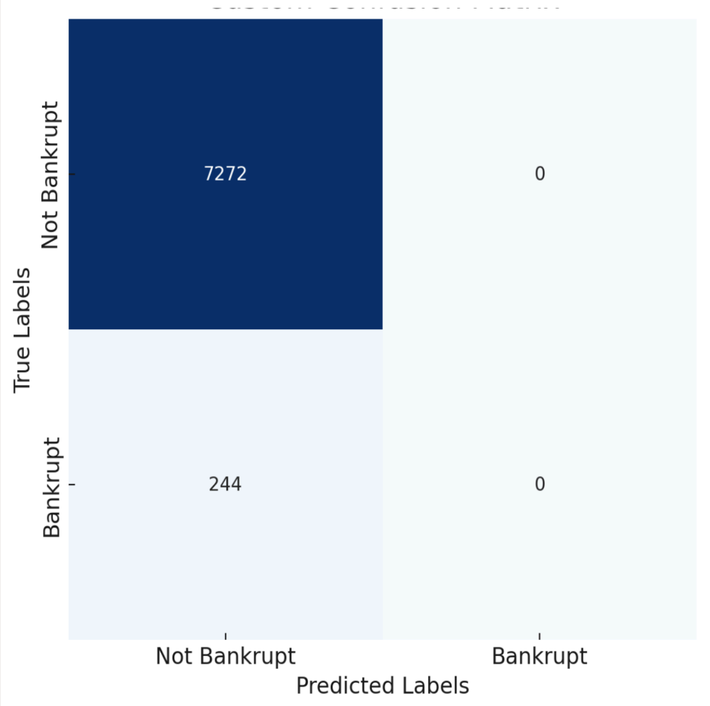
LSTM – 94%
Positive class accuracy – 86%
Negative class accuracy  - 94.2%




GRU – 92%
Positive class accuracy – 81.9%
Negative class accuracy – 92.3%




Transformers – 97%
Positive class accuracy – 0%
Negative class accuracy – 97%



# FUTURE WORK

- We aim to integrate ensembling methodologies to optimize model accuracy to achieve better results.