Bankruptcy

How to pick a winning investment

By the Three Wolf Moon of Wall Street
Binita, Juan, and Faye

Table of contents

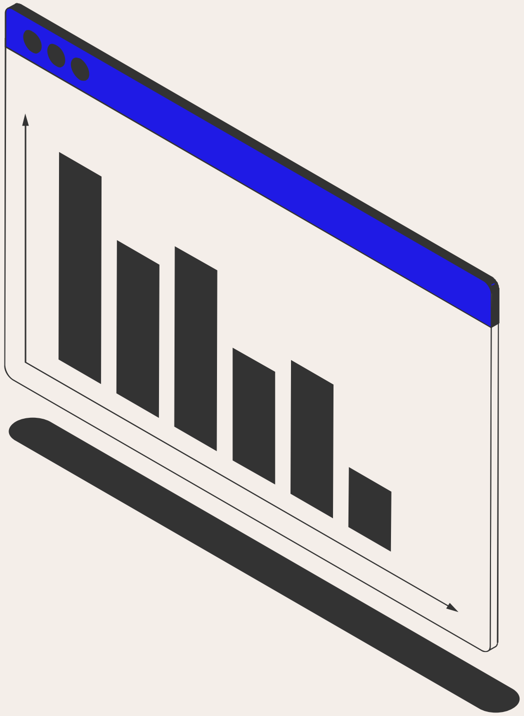
01 Objective

We have been contracted to advise a hedge fund looking to add prudent investments to their portfolio. The client is risk-averse.

We will build a classification model that predicts whether a company will succeed or go bankrupt in order to better advise our client which companies to invest with and which to avoid.

02

Exploratory data analysis of bankruptcy



The dataset

US Company Bankruptcy Prediction Dataset (1999 - 2018)

- 8,971 distinct companies:

  8,362 are in business "alive"

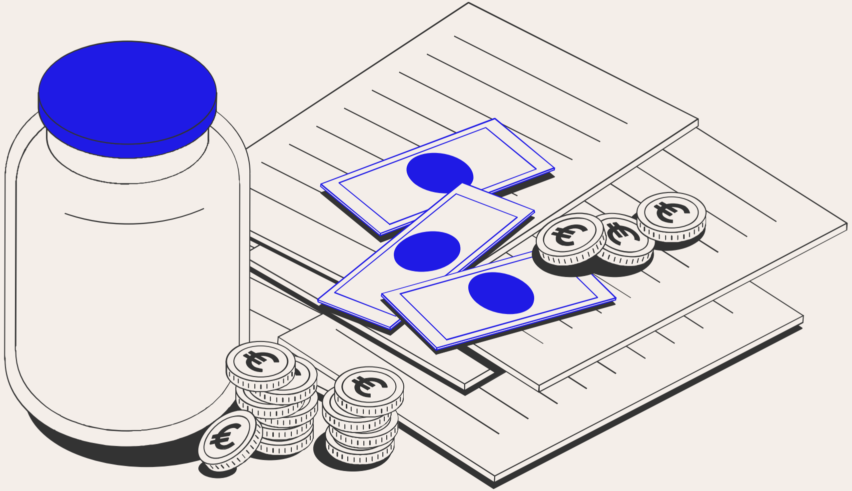
  609 are bankrupt

- 18 financial health features such as:

  Total assets

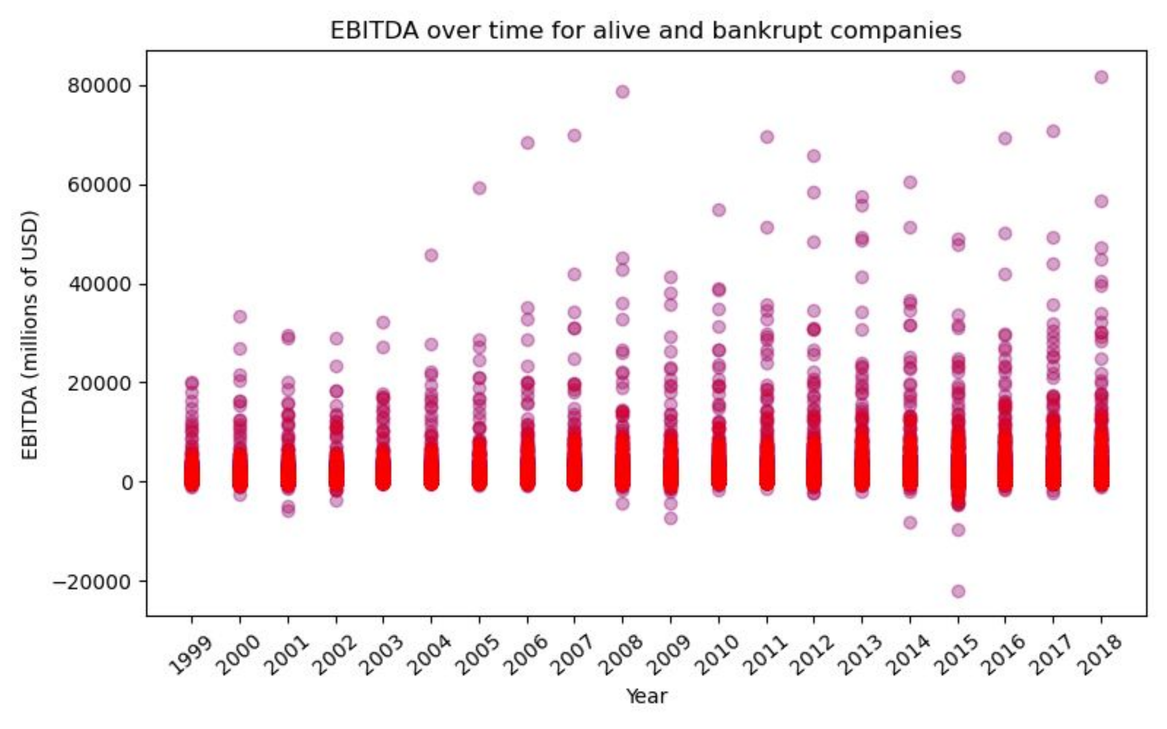
  Earnings before interest and taxes

  Total long-term debt

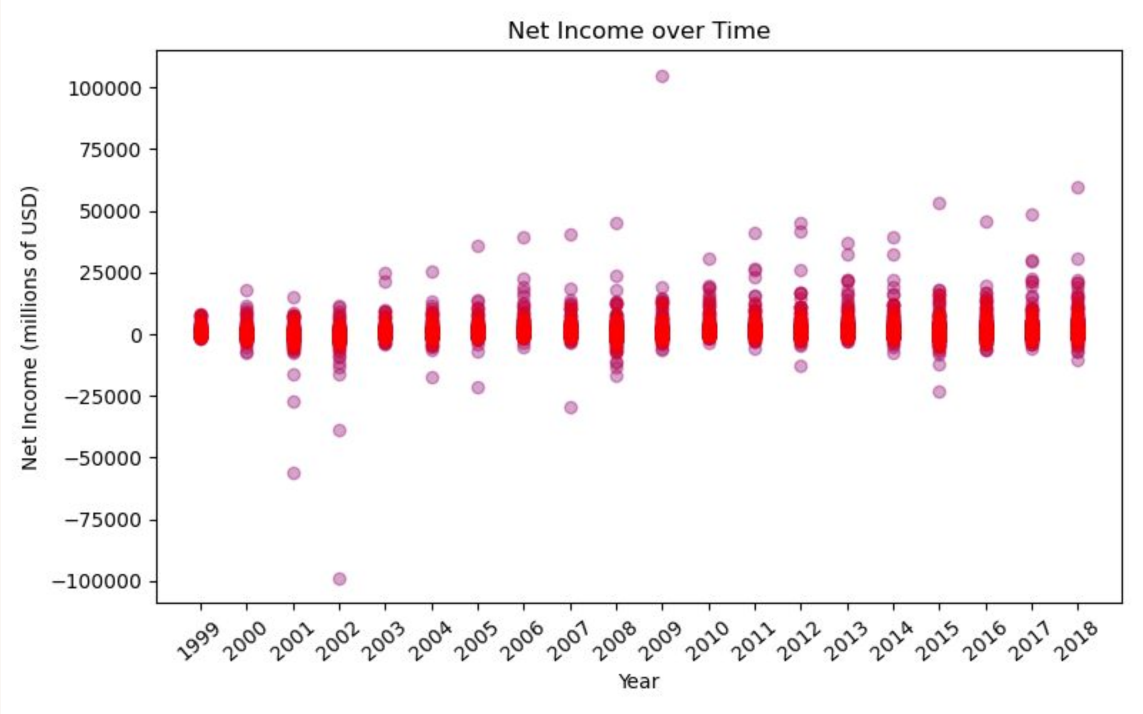


Link to dataset:
https://www.kaggle.com/datasets/utkarshx27/american-companies-bankruptcy-prediction-dataset

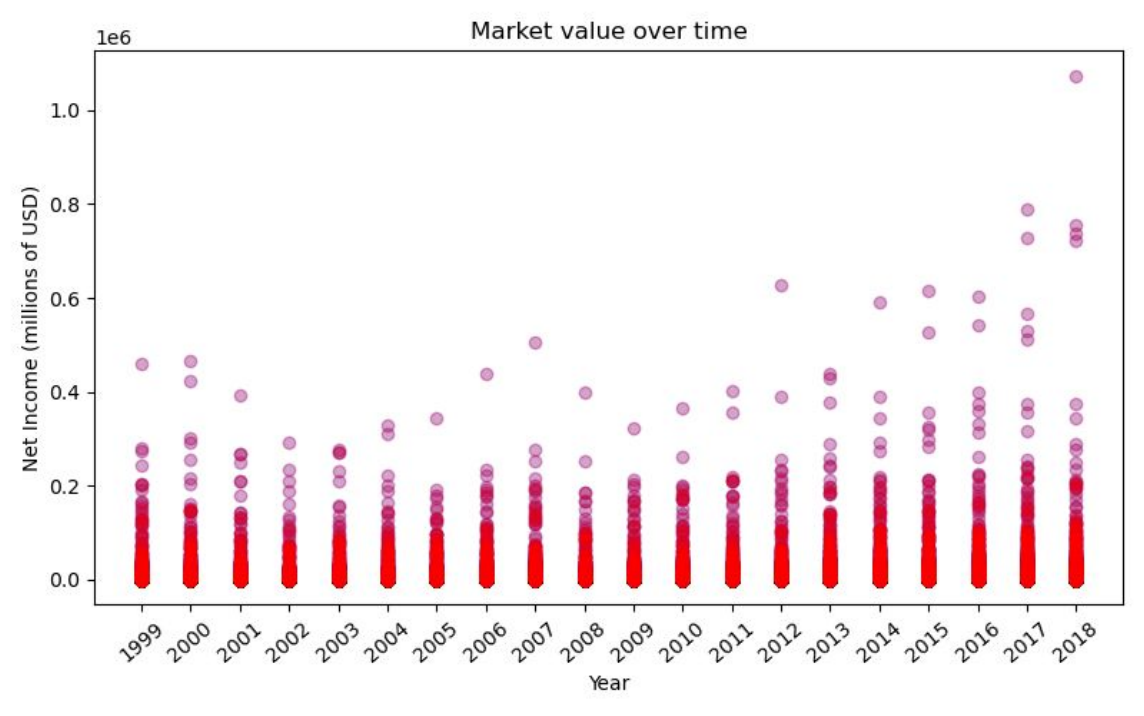
The relationship between EBITDA and bankruptcy



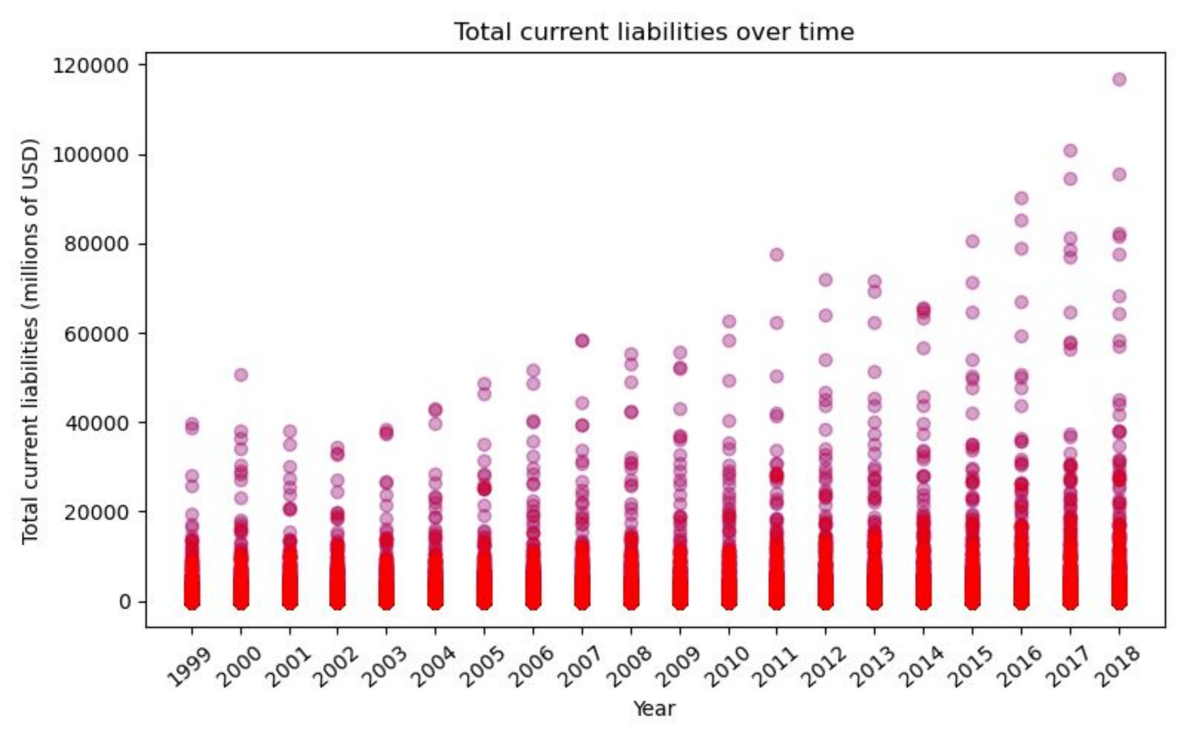
The relationship between Net income and bankruptcy



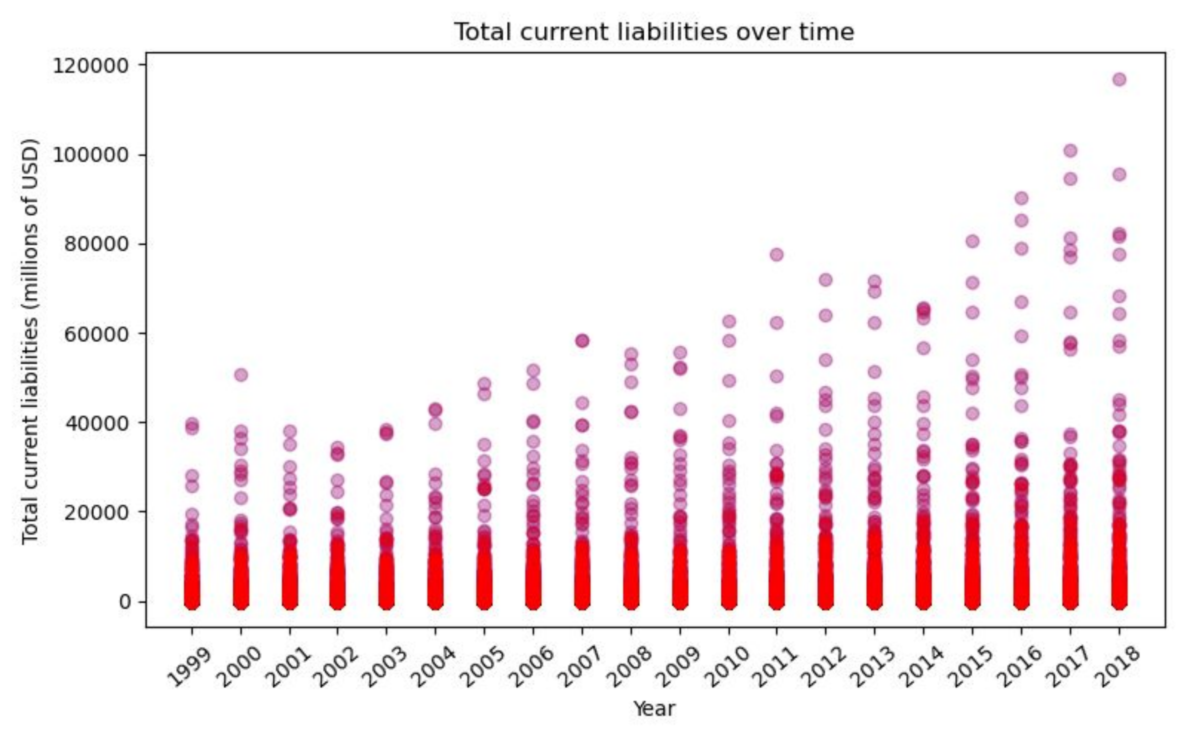
The relationship between market value and bankruptcy



The relationship between total liabilities and bankruptcy

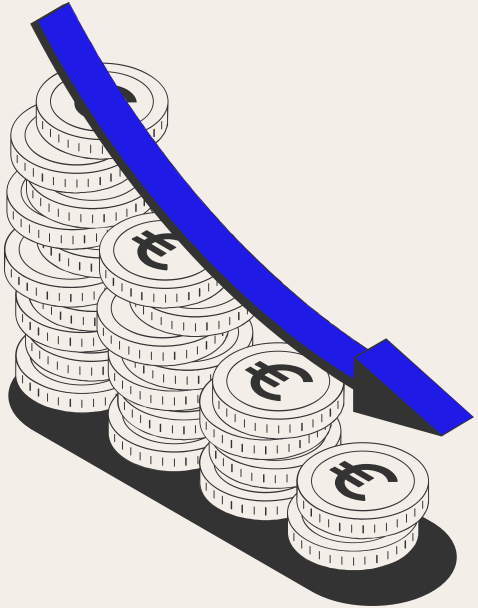


The relationship between operating costs and bankruptcy



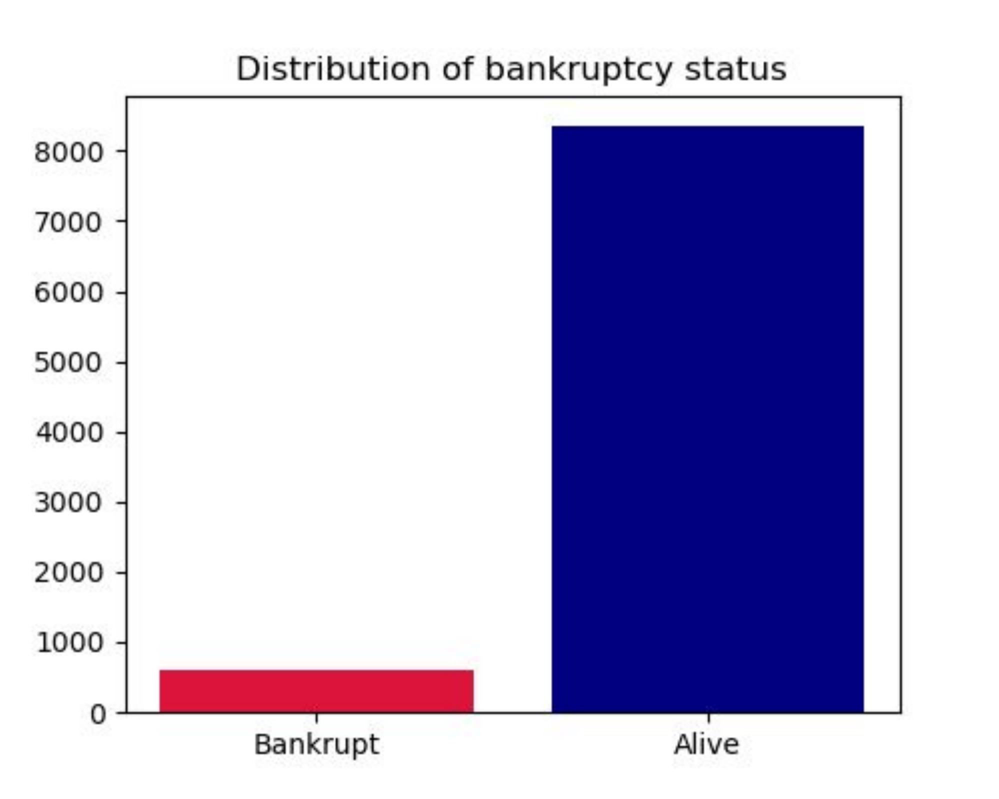
03

Tree models for classification



Handling imbalanced classes

Correcting class imbalances before training a model is important to reduce bias, improve generalization, ensure accurate performance metrics, and facilitate better decision-making.



Random Forest + RandomSearchCV

Test set accuracy: 0.9407

Precision (TP / (TP + FP)): 23%

Recall (TP / (TP + FN)): 64%

True Positive	False Positive
310	995
172	18194
False Negative	True Negative

Random Forest

Feature Importances:

X8 Market value	0.092
X15 Retained Earnings	0.073
X6 Net Income	0.071
X3 Depreciation and amortization	0.067
X7 Total Receivables	0.063

XGBoost

Test set accuracy XGBoost: 0.9395

Precision (TP / (TP + FP)): 35%

Recall (TP / (TP + FN)): 57%

True Positive	False Positive
459	846
345	18021
False Negative	True Negative

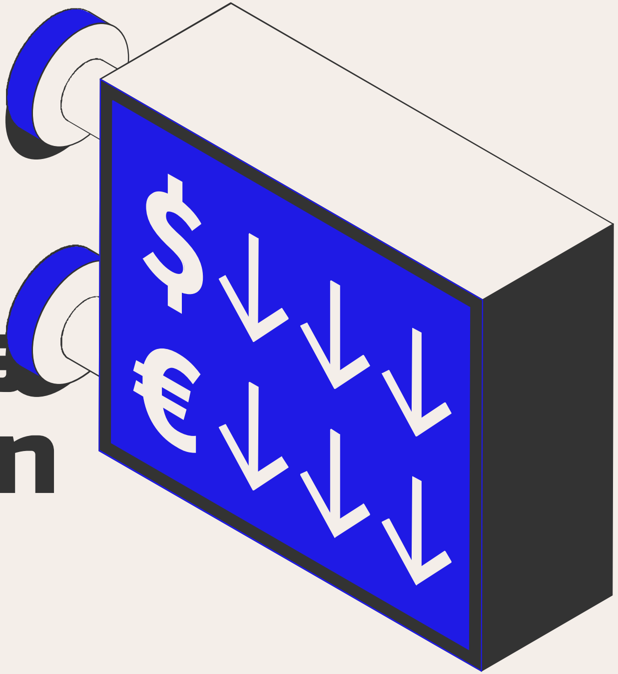
XGBoost

Feature Importances:

X15 Retained Earnings	2771
X8 Market value	2459
X3 Depreciation and amortization	2397

04

Neural networks for classification



Neural Network model

Test set accuracy: 0.8343

Precision (TP / (TP + FP)): 45%

Recall (TP / (TP + FN)): 19%

True Positive	False Positive
594	711
2549	15817
False Negative	True Negative

Recurrent Neural Network model

Explored LSTM, Dense, Dropout, BatchNormalization, and EarlyStopping to optimize

Only achieved Test set accuracy: 0.9025

05

Key Findings

Accuracy over baseline