US Company Bankruptcy Prediction

Outline

- Introduction
- Feature Selection
- Challenge
- Model Workflow
- Training Result
- Insights
- Conclusion

Introduction

About our dataset

1. Related to American Public Companies, listed on the NewYork Stock Exchange and NASDAQ
2. Provide an understanding of a company's financial health

company_name	year	is_bankruptcy	Current assets	Cost of goods sold	Depreciation and amortization	EBITDA	Inventory	Net Income	Total Receivables	Market Value
C_1	1999	0	511.267	833.107	18.373	89.031	336.018	35.163	128.348	372.7519
	2000	0	485.856	713.811	18.577	64.367	320.590	18.531	115.187	377.1180
	2001	0	436.656	526.477	22.496	27.207	286.588	-58.939	77.528	364.5928
	2002	0	396.412	496.747	27.172	30.745	259.954	-12.410	66.322	143.3295
	2003	0	432.204	523.302	26.680	47.491	247.245	3.504	104.661	308.9071

Net Sales	Total Assets	Total Long-term Debt	EBIT	Gross Profit	Total Current Liabilities	Retained Earnings	Total Revenue	Total Liabilities	Total Operating Expenses
1024.333	740.998	180.447	70.658	191.226	163.816	201.026	1024.333	401.483	935.302
874.255	701.854	179.987	45.790	160.444	125.392	204.065	874.255	361.642	809.888
638.721	710.199	217.699	4.711	112.244	150.464	139.603	638.721	399.964	611.514
606.337	686.621	164.658	3.573	109.590	203.575	124.106	606.337	391.633	575.592
651.958	709.292	248.666	20.811	128.656	131.261	131.884	651.958	407.608	604.467

Income Statement

Total Revenue
- sales return, allowance, and discounts
= Net Sales
- Cost of goods sold
= Gross Profit
- Total Operating Expenses (R&D, Salaries and Wages, Rent and Utilities, etc,.)
= EBITDA
- Depreciation and amortization
= EBIT
- Interest and Tax
= Net Income

Balance Sheet

Assets	Liabilities and Equity
Total Assets	Total Liabilities
Current assets	Total Long-term Debt
Inventory	Total Current Liabilities
Total Receivables	Retained Earnings

Goal Prediction

A company's possibility of facing bankruptcy

Who cares about ?

1. Company

2. Investor

Feature Selection

SelectKBest

SelectFromModel

- **score_func = f_classif**

- **Lasso() = L1 regularization**

	param_select__k	mean_test_score
17	18	0.503538
16	17	0.502628
15	16	0.502628

	param_select__estimator_ _alpha	param_select__threshold	mean_test_score
0	10	mean	0.503539
1	10	median	0.503528
2	100	mean	0.503515

Optimal number of features: 18

Optimal number of features: 18

Challenge

Imbalanced Data

Proportion of target variable

Failed
6.6%
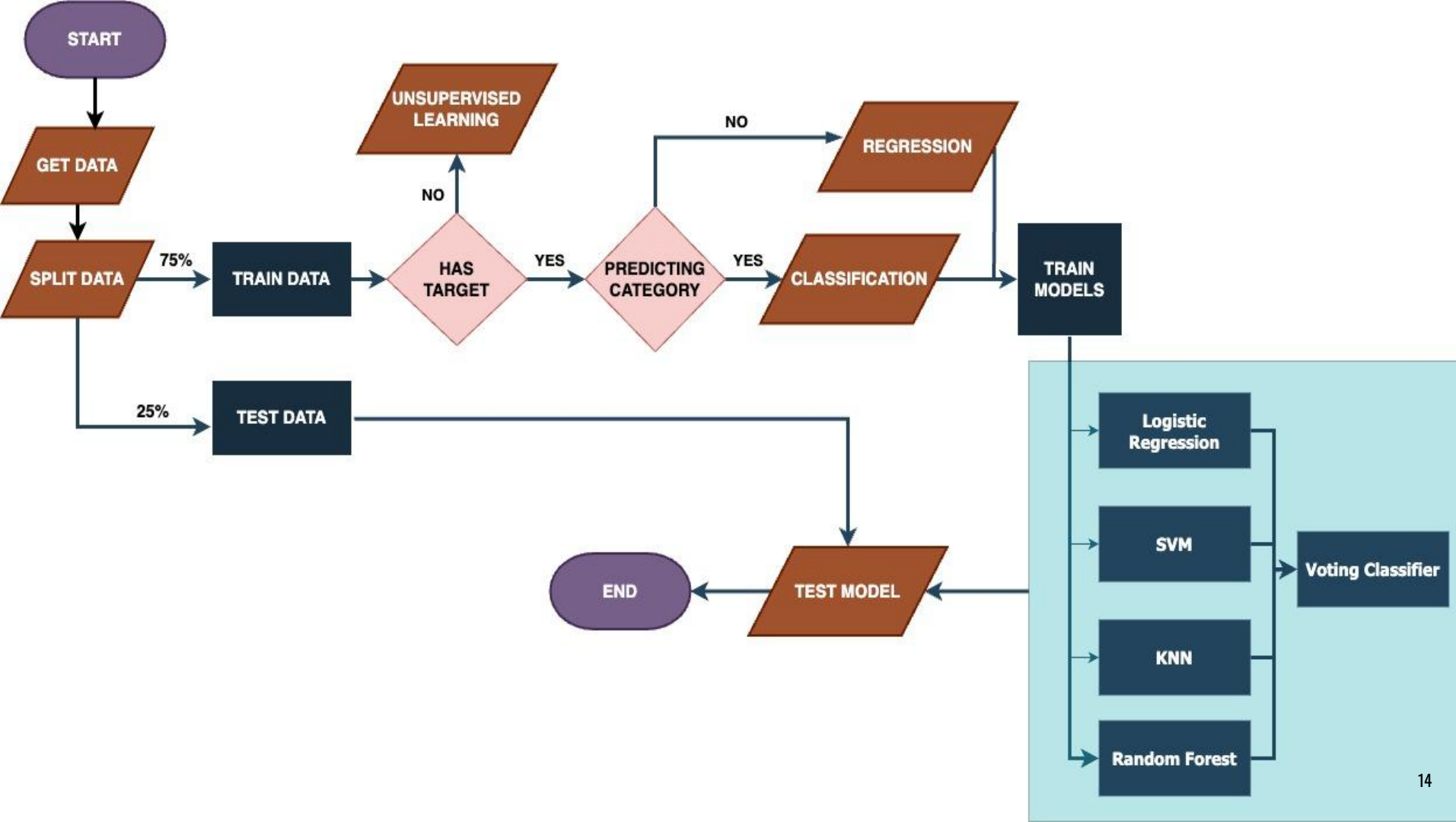
Alive
93.4%
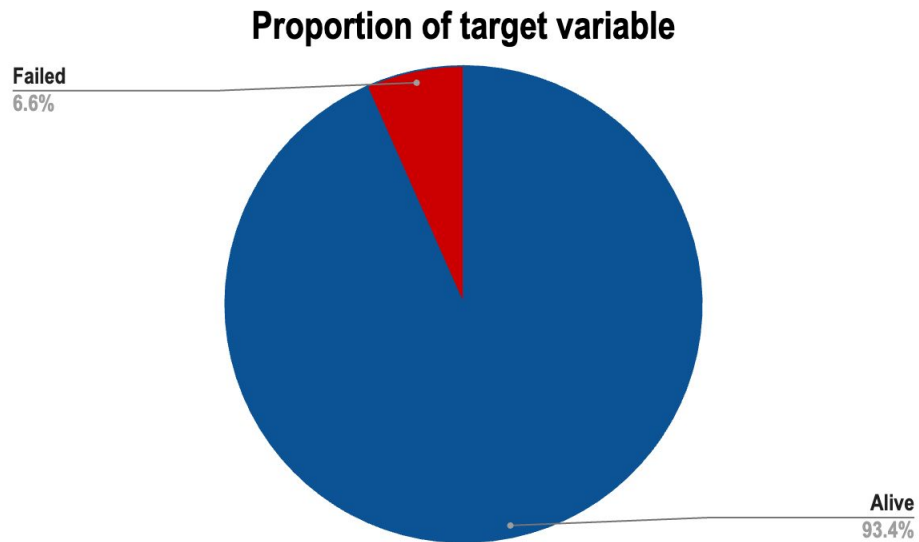
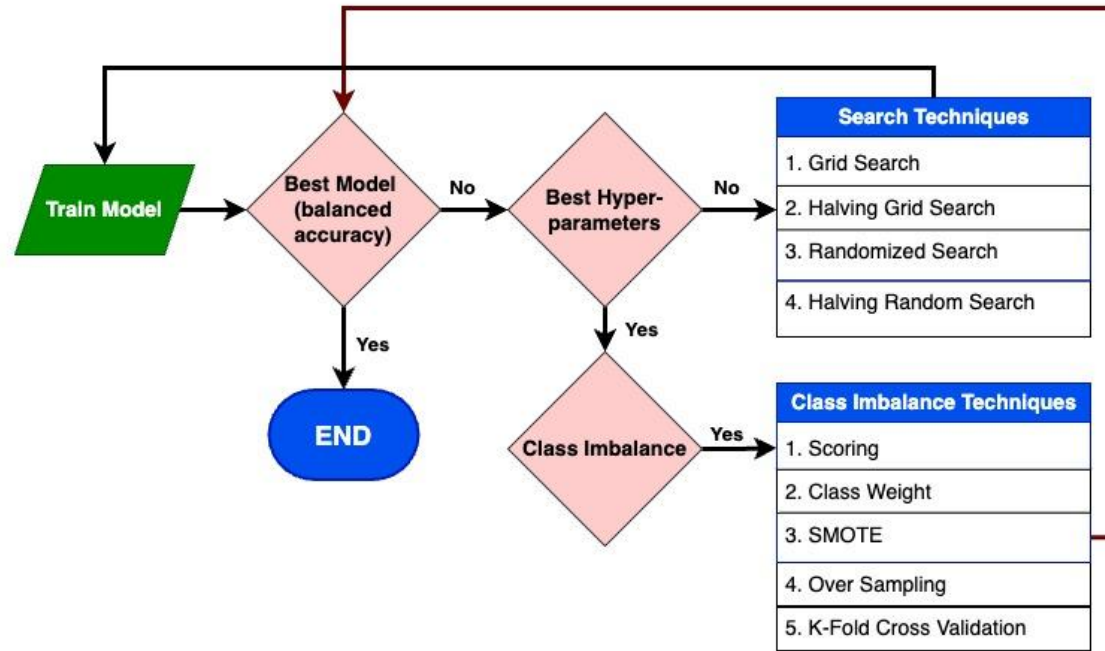1. scoring='balanced_accuracy'
2. class_weight='balanced'
3. Bootstrap
4. SMOTE
5. K-fold cross-validation

Model WorkFlow

START

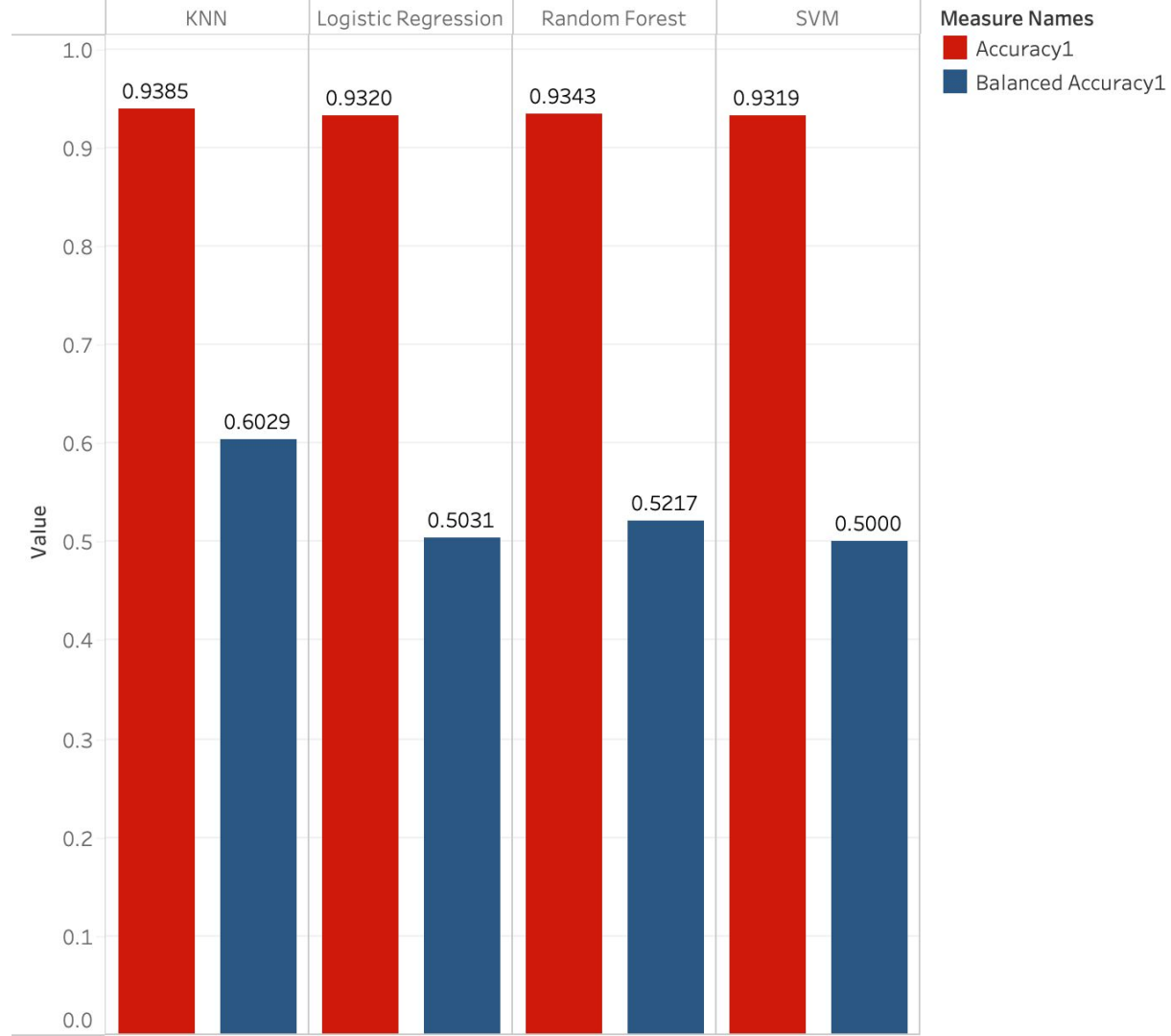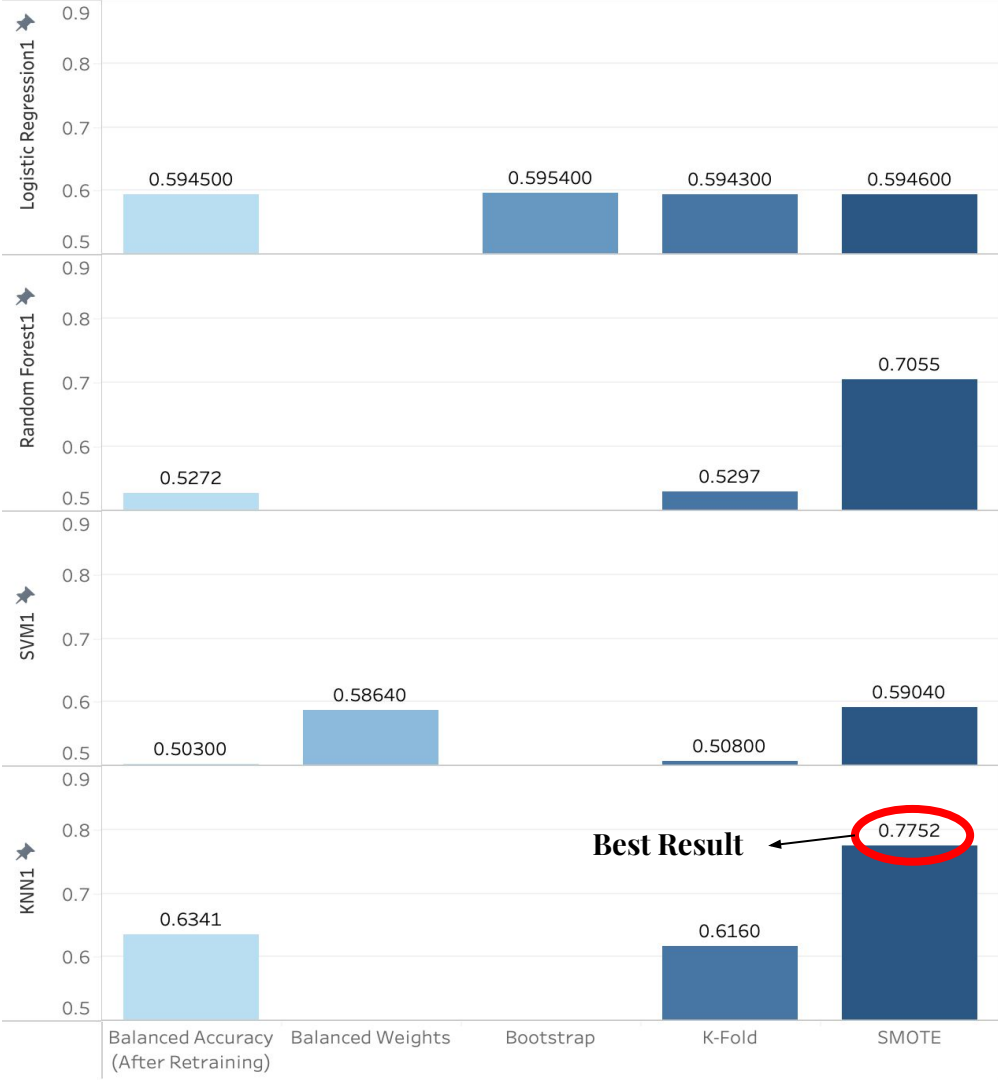GET DATA

SPLIT DATA

75% → TRAIN DATA

25% → TEST DATA

HAS TARGET

NO → UNSUPERVISED LEARNING

YES → PREDICTING CATEGORY

NO → REGRESSION

YES → CLASSIFICATION

TRAIN MODELS

Logistic Regression

SVM

KNN

Random Forest

Voting Classifier

TEST MODEL

END

14

Inside Training Models

Training Result

Balanced Accuracy

Voting
0.7643

KNN SMOTE

Confusion matrix:

	Predicted 0	Predicted 1
True 0	17237	1095
True 1	522	817
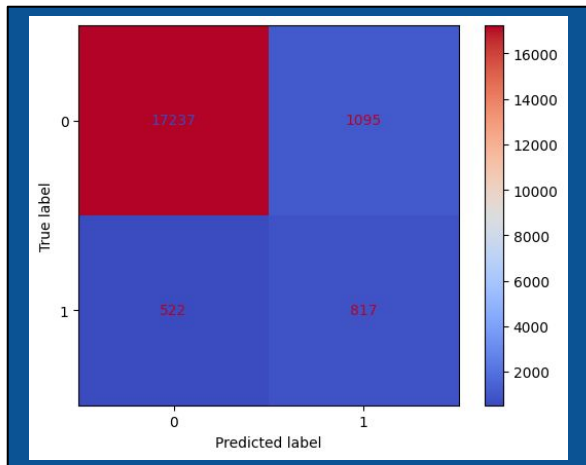
classification report=

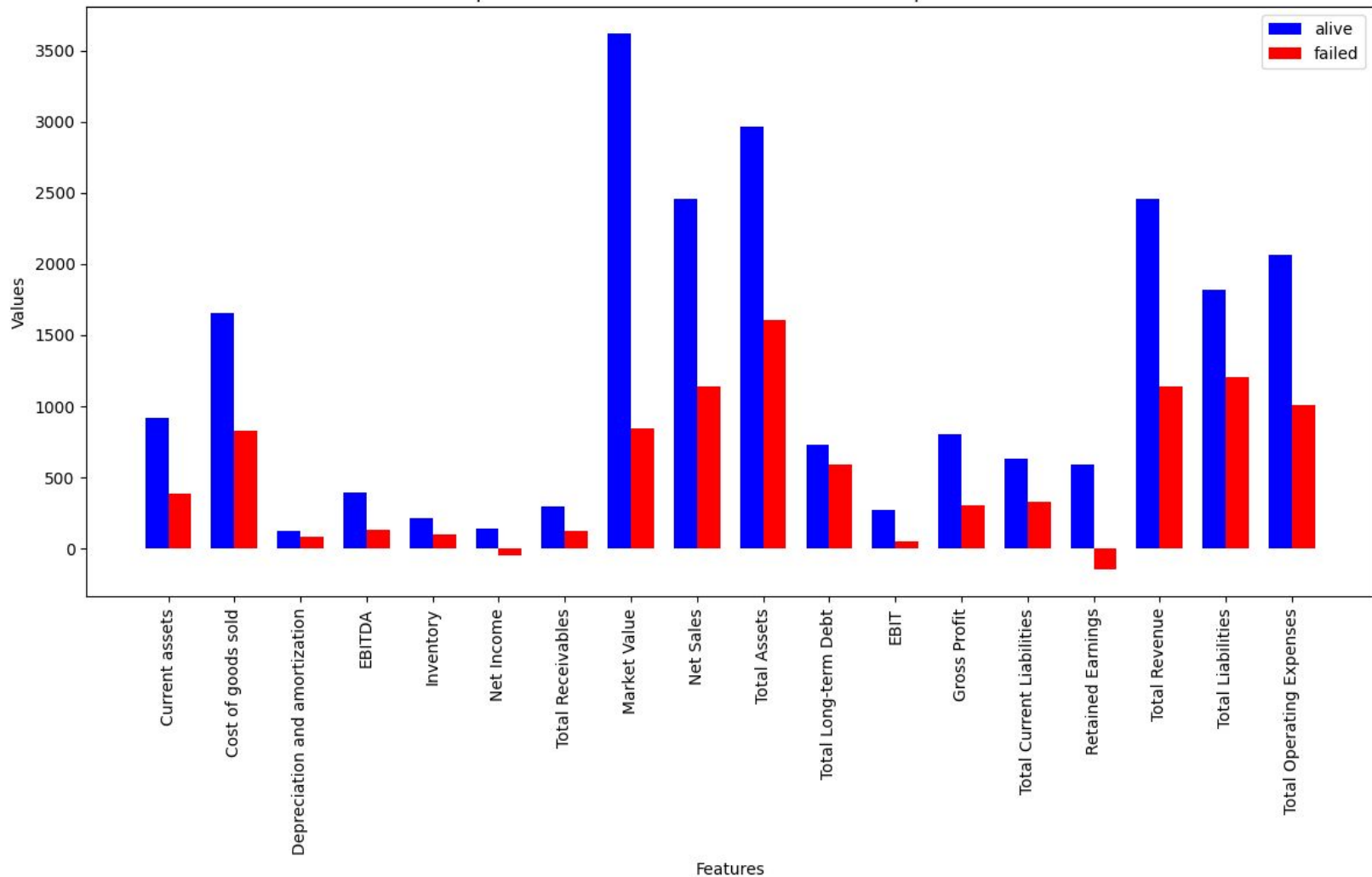|	precision	recall	f1-score	support
0	0.97	0.94	0.96	18332
1	0.43	0.61	0.50	1339
accuracy			0.92	19671
macro avg	0.70	0.78	0.73	19671
weighted avg	0.93	0.92	0.92	19671

Accuracy: 0.9178
Balanced Accuracy: 0.7752

Pipeline (AUC = 0.82)
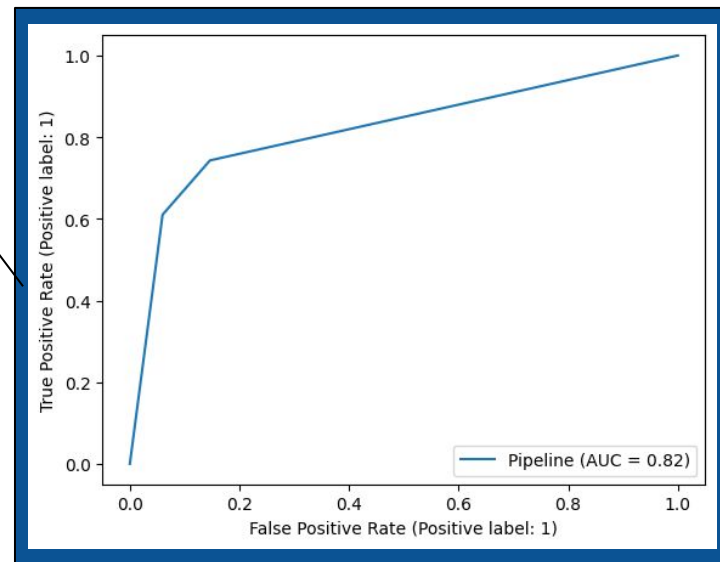
Insights

1. Financial metrics
2. Liability
3. Market Value

Comparison of Features between different bankrupt condition

Conclusion

1. Model Summary

2. Investors Perspective

3. Companies Perspective

THANK YOU

Colab Link:
https://colab.research.google.com/drive/14wfN6q9cSq-w6vTqYDpSOqfuQAYxn3pD?usp=sharing