NAME: JAIVARDHAN TAMMINANA

REGISTRATION NUMBER: 21BCE0512

DIGITAL ASSIGNMENT

PROJECT TITLE: DETECTING BANKRUPTCY

About Project:

Bankruptcy prediction models have long been a cornerstone of economic research, with their roots tracing back nearly a century. These models serve a critical purpose in assessing the financial health of companies, offering insights into their viability and long-term prospects within the market. By integrating various econometric indicators, these models aim to anticipate the likelihood of a company facing financial distress or declaring bankruptcy.

In the context of this study, the focus is on developing a predictive model tailored to the **dynamics of the Dutch market**. The dataset under examination spans from 2002 to 2014, encompassing both bankrupt firms and those still operational as of 2010. This timeframe allows researchers to analyze a comprehensive spectrum of companies, capturing both successful and struggling enterprises over a significant period.

The predictive model incorporates a range of financial ratios and indicators from the initial year of observation. These metrics serve as fundamental inputs, providing valuable insights into the financial structure and performance of the companies under scrutiny. By analyzing these indicators alongside other relevant factors, such as industry trends and macroeconomic conditions, researchers can construct a robust framework for predicting bankruptcy risk.

The classification cases within this framework are delineated based on the forecasting period, with each case representing a distinct time horizon for predicting bankruptcy. These cases allow for a nuanced analysis of bankruptcy risk across different timeframes, enabling stakeholders to tailor their strategies and interventions accordingly.

Accompanying these classification cases are corresponding class labels indicating the bankruptcy status of each company. These labels serve as the ground truth for model training and evaluation, facilitating the assessment of predictive performance and the refinement of the predictive model over time.

Overall, the development of a bankruptcy prediction model for businesses operating in Poland represents a significant contribution to the field of economic research. By leveraging econometric techniques and empirical data, researchers can enhance our understanding of the factors driving financial distress and bankruptcy, ultimately empowering stakeholders to make informed decisions in managing risk and fostering economic resilience.

About the Functionalities:

1. **Data Collection and Preparation:**
    a. The dataset includes financial rates from the second year of the forecasting period and predicts bankruptcy status after four years. There are 10,173 instances, with 400 representing bankrupted companies and 9,773 representing firms that remained solvent.
2. **Feature Selection:**
    a. Feature selection is a crucial step in building an effective bankruptcy prediction model. It involves identifying the most relevant and informative features from the available data that contribute significantly to predicting the likelihood of bankruptcy.
    b. We tested out for 3 separate classifiers, namely **Decision Tree Classifier, Random Forest Classifier and XGBoost Classifier**
    c. Each generated its own 10 prominent attributes
3. **Model Development and Training:**
    a. Model development and training are central stages in building a bankruptcy prediction model. These steps involve selecting an appropriate machine learning algorithm, preparing the data, training the model, and evaluating its performance.
    b. We prepared data for all three classifiers chosen before with a 70-30 train-test split.
    c. All the classifiers were trained with the same random dataset and their accuracies along with other evaluation matrices are compared to pick the right classifier to use to continue developing the model
4. **Evaluation Matrix:**
    a. Evaluation metrics are essential for assessing the performance of a bankruptcy prediction model. These metrics quantify how well the model predicts the bankruptcy status of companies and help stakeholders understand its effectiveness.

Below is the evaluation matrix of all three classifiers.

MODEL	TEST ACCURACY	TRAIN ACCURACY	SELECTED FEATURES
DECISION TREE CLASSIFIER	88.70	100	['Attr5', 'Attr6', 'Attr9', 'Attr27', 'Attr34', 'Attr44', 'Attr46', 'Attr54', 'Attr58', 'Attr60']
RANDOM FOREST CLASSIFIER	95.68	100	['Attr5', 'Attr6', 'Attr27', 'Attr34', 'Attr36', 'Attr44', 'Attr46', 'Attr54', 'Attr58', 'Attr60']

XGBoost	96.02	100	['Attr5', 'Attr6', 'Attr27', 'Attr33', 'Attr34', 'Attr44', 'Attr46', 'Attr54', 'Attr58', 'Attr60']

About the Attributes in Dataset:

Attr1: net profit / total assets

Attr2: total liabilities / total assets

Attr3: working capital / total assets

Attr4: current assets / short-term liabilities

Attr5: [(cash + short-term securities + receivables short-term liabilities) / (operating expenses depreciation)] * 365

Attr6: retained earnings / total assets

Attr7: EBIT / total assets • Attr8: book value of equity / total liabilities

Attr9: sales / total assets

Attr10: equity / total assets

Attr11: (gross profit + extraordinary items + financial expenses) / total assets

Attr12: gross profit / short-term liabilities

Attr13: (gross profit + depreciation) / sales

Attr14: (gross profit + interest) / total assets

Attr15: (total liabilities * 365) / (gross profit + depreciation)

Attr16: (gross profit + depreciation) / total liabilities

Attr17: total assets / total liabilities

Attr18: gross profit / total assets

Attr19: gross profit / sales

Attr20: (inventory * 365) / sales

Attr21: sales (n) / sales (n-1)

Attr22: profit on operating activities / total assets

Attr23: net profit / sales • Attr24: gross profit (in 3 years) / total assets

Attr25: (equity share capital) / total assets

Attr26: (net profit + depreciation) / total liabilities

Attr27: profit on operating activities / financial expenses

Attr28: working capital / fixed assets

Attr29: logarithm of total assets

Attr30: (total liabilities cash) / sales

Attr31: (gross profit + interest) / sales

Attr32: (current liabilities * 365) / cost of products sold

Attr33: operating expenses / short-term liabilities

Attr34: operating expenses / total liabilities

Attr35: profit on sales / total assets

Attr36: total sales / total assets

Attr37: (current assets inventories) / long-term liabilities

Attr38: constant capital / total assets

Attr39: profit on sales / sales

Attr40: (current assets inventory receivables) / short-term liabilities

Attr41: total liabilities / ((profit on operating activities + depreciation) * (12/365))

Attr42: profit on operating activities / sales

Attr43: rotation receivables + inventory turnover in days

Attr44: (receivables * 365) / sales

Attr45: net profit / inventory

Attr46: (current assets inventory) / short-term liabilities

Attr47: (inventory * 365) / cost of products sold

Attr48: EBITDA (profit on operating activities depreciation) / total assets

Attr49: EBITDA (profit on operating activities depreciation) / sales

Attr50: current assets / total liabilities

Attr51: short-term liabilities / total assets

Attr52: (short-term liabilities * 365) / cost of products sold)

Attr53: equity / fixed assets • Attr54: constant capital / fixed assets

Attr55: working capital

Attr56: (sales cost of products sold) / sales

Attr57: (current assets inventory short-term liabilities) / (sales gross profit depreciation)

Attr58: total costs /total sales

Attr59: long-term liabilities / equity

Attr60: sales / inventory

Attr61: sales / receivables

Attr62: (short-term liabilities *365) / sales

Attr63: sales / short-term liabilities

Attr64: sales / fixed assets

Class: did not get bankrupt (0) or got bankrupt (1)

Observations and Additional Functionality:

Observations:

1. Financial Ratios: Various financial ratios calculated based on companies' financial statements, including profitability ratios (e.g., return on equity, net profit margin), liquidity ratios (e.g., current ratio, quick ratio), leverage ratios (e.g., debt-to-equity ratio, interest coverage ratio), and efficiency ratios (e.g., asset turnover ratio, inventory turnover ratio).
2. Accounting Metrics: Accounting metrics such as revenue, expenses, assets, liabilities, equity, earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA), net income, cash flow from operations, and free cash flow.
3. Market Data: Market-related metrics reflecting companies' performance and valuation, including stock prices, market capitalization, trading volumes, price-to-earnings (P/E) ratio, price-to-book (P/B) ratio, and dividend yield.
4. Bankruptcy Status: A binary or categorical label indicating whether a company has declared bankruptcy or faced financial distress within a specified timeframe. This serves as the target variable for training the bankruptcy prediction model.

Additional Features:

1. **Deployment & Continuous Monitoring:** Deploy the trained model for making predictions on new data and monitor its performance over time. Update the model as necessary to maintain effectiveness.
2. **New Features such as Recommendations:** Trained model can identify trends on how a company is growing and can suggest company owners on how to either avoid losing their cashflow or pinpoint the issue with their current model.

TableAU Figures:

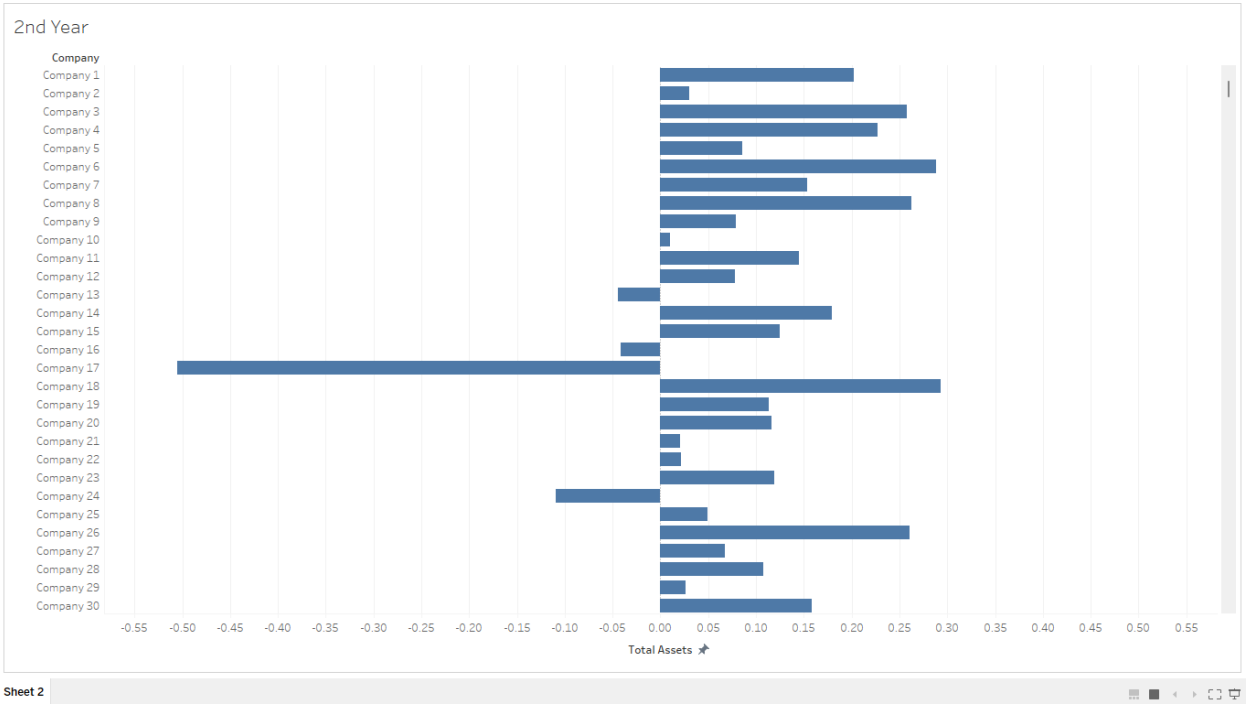


Fig: Company vs Total Assets (Attr1)

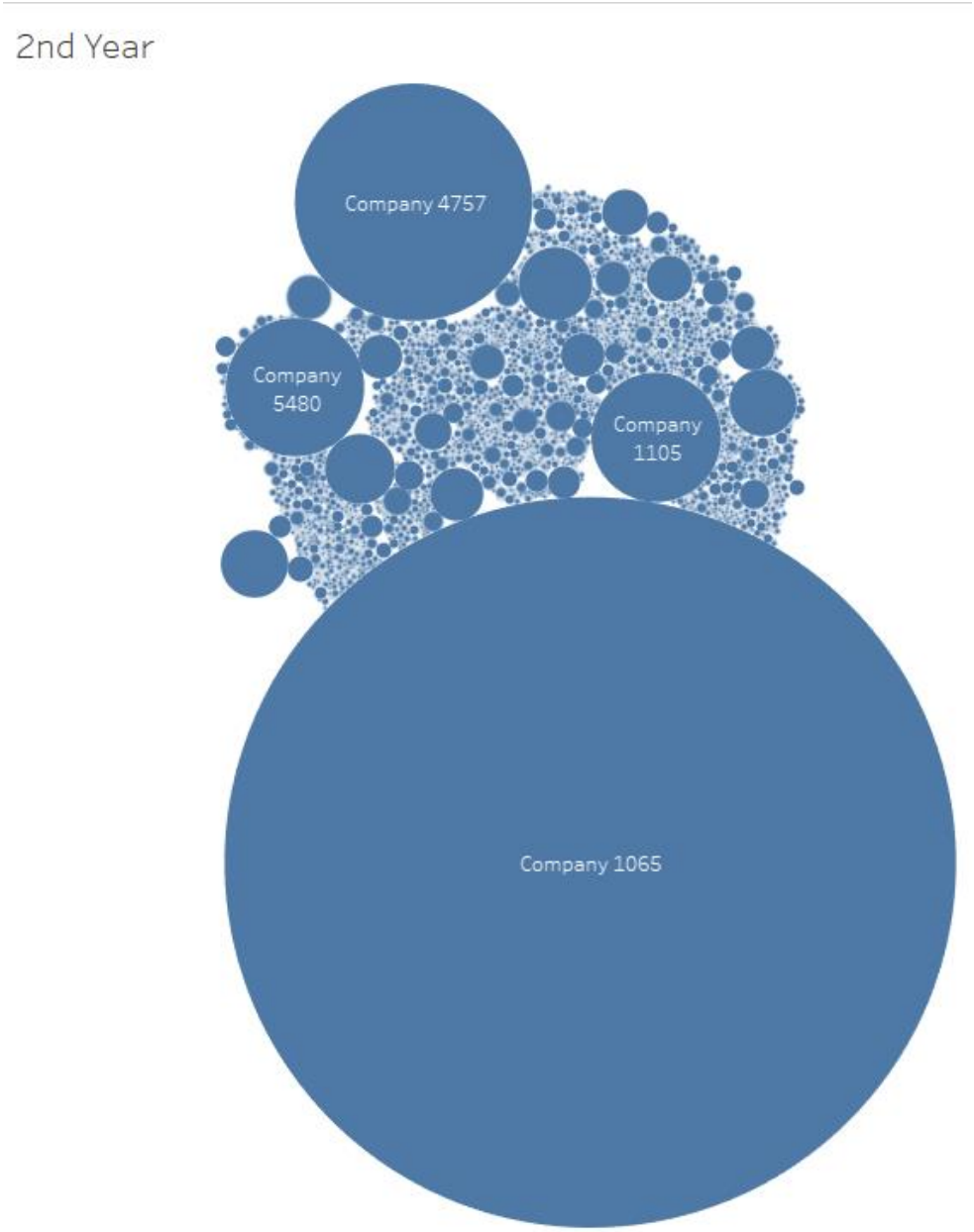


Fig: Company vs Long-term Liability (Attr59)

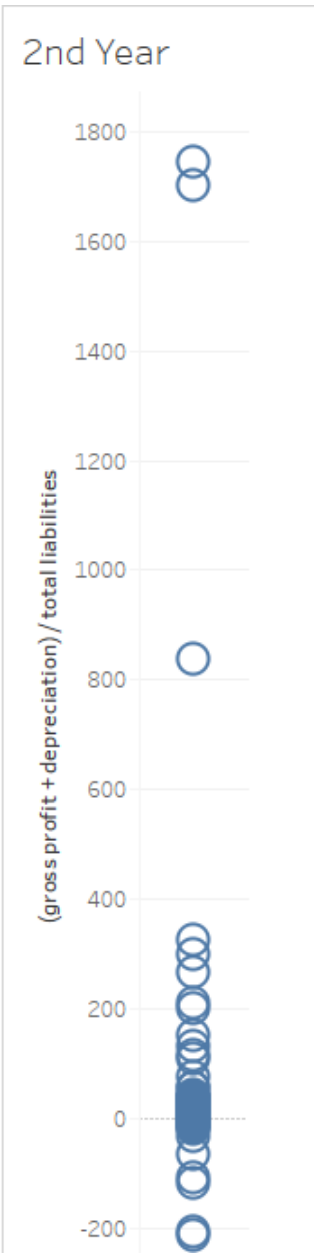
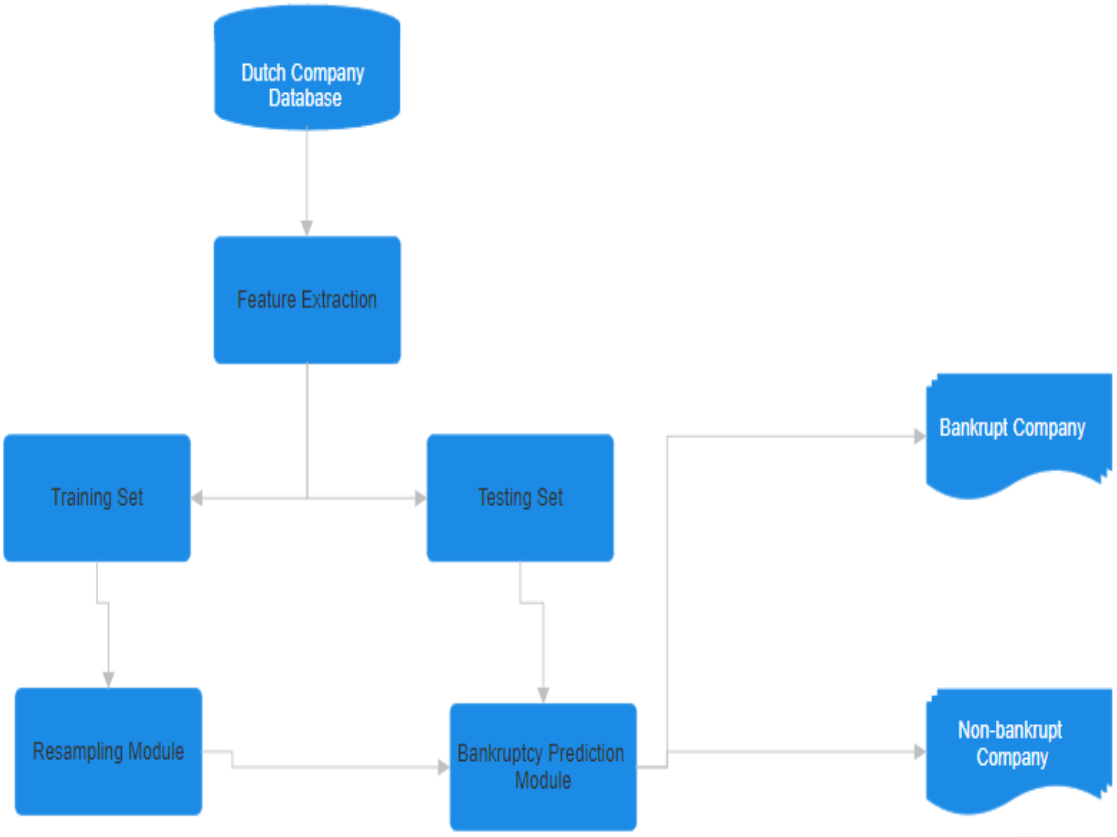


Fig: Company vs (Gross Profit + Depreciation) / Total Liabilities

Workflow Diagram of the Project:



Source Code:

```
# Create a feature selection object
rfe_xgb = RFE(
    estimator = DecisionTreeClassifier(),
    n_features_to_select = 10
)

# Fit the feature selection object to the data
rfe_xgb.fit(x, y)

# Create a list of all selected features
xgb_selected_features = []
for i, col in zip(range(x.shape[1]), x.columns):
    if rfe_xgb.support_[i]:
        xgb_selected_features.append(col)

# Print the selected features
print(f"XGB Classifier : \n{xgb_selected_features}")
# Test-train Split
xgb_x_train, xgb_x_test, xgb_y_train, xgb_y_test = train_test_splitting(
    x = x.drop(columns = [col for col in x if col not in xgb_selected_features], axis = 1),
    y = y
)
# Initialize the model
xgb_classifier = XGBClassifier(max_depth = 8)

# Fitting the data on the model
xgb_classifier.fit(xgb_x_train_smote, xgb_y_train_smote)

# Get the predictions
xgb_prediction = xgb_classifier.predict(xgb_x_test)
```

```
xgb_train_prediction = xgb_classifier.predict(xgb_x_train)
```

Prediction:

```
[ ] bankrupt = xgb_classifier.predict([[
        38.803,46.705,2.872,0.59628,0.2375,
        0.053954,1.2819,0.22599,0.77401,3.1534
    ]])[0]

[ ] if bankrupt:
        print("The company is likely to go BANKRUPT.")
    else:
        print("The company is SAFE.")

    The company is SAFE.
```

Github Repository link: https://github.com/jaiVIT123/DSAssignment