Bankruptcy Prevention

STORIES OF
BANKRUPTCY

Mr. Rohit Gusain
Mrs Megha Patil
Mr.Aayesh Azim Abdul
Mr. Muthamilarasu S

Bankruptcy Overview

Bankruptcy is a legal process that provides relief to individuals and businesses that are unable to pay their debts. It allows them to either eliminate or restructure their debts under the protection of the bankruptcy court.

- Bankruptcy Also known as liquidation bankruptcy, this type of bankruptcy involves selling off assets to pay off debts.
- This type of bankruptcy is designed for businesses that want to restructure their debts and continue operating.
- This type of bankruptcy is designed for individuals with regular income who want to restructure their debts and pay them off over time.

Bankruptcy Prediction Model Evaluation

Model Overview
The bankruptcy prediction model was developed to identify companies at risk of bankruptcy based on financial and non-financial factors. The model uses a combination of machine learning algorithms and financial analysis techniques to generate a bankruptcy risk score for each company.

Effectiveness
The model has been effective in identifying companies that have gone bankrupt in the past. However, there have been instances where the model failed to predict bankruptcy for companies that eventually went bankrupt. This suggests that there may be additional factors that the model is not capturing.

Areas for Improvement

- Incorporating additional non-financial factors, such as industry trends and market conditions, could improve the model's accuracy.
- The model could be updated more frequently to account for changes in a company's financial position.
- The model's performance could be evaluated on a more regular basis to ensure it is still effective in predicting bankruptcy.

Bankruptcy Prediction Model Overview :

Key Factors for Bankruptcy Prediction

- Financial ratios, such as debt-to-equity ratio and current ratio.
- Industry-specific factors, such as market trends and competition.
- Company-specific factors, such as management experience and customer base.

Machine Learning Model Implementation

The machine learning model will incorporate the identified key factors for bankruptcy prediction. The model will be trained on historical financial data and industry trends to accurately predict the likelihood of bankruptcy for a given company.

Catching the Doom, Before it Happens

1. **Problem Statement**
2. **Exploratory Data Analysis**
3. **Applying the Models**
4. **Model Selection and validation**
5. **Model deployment in streamlit**

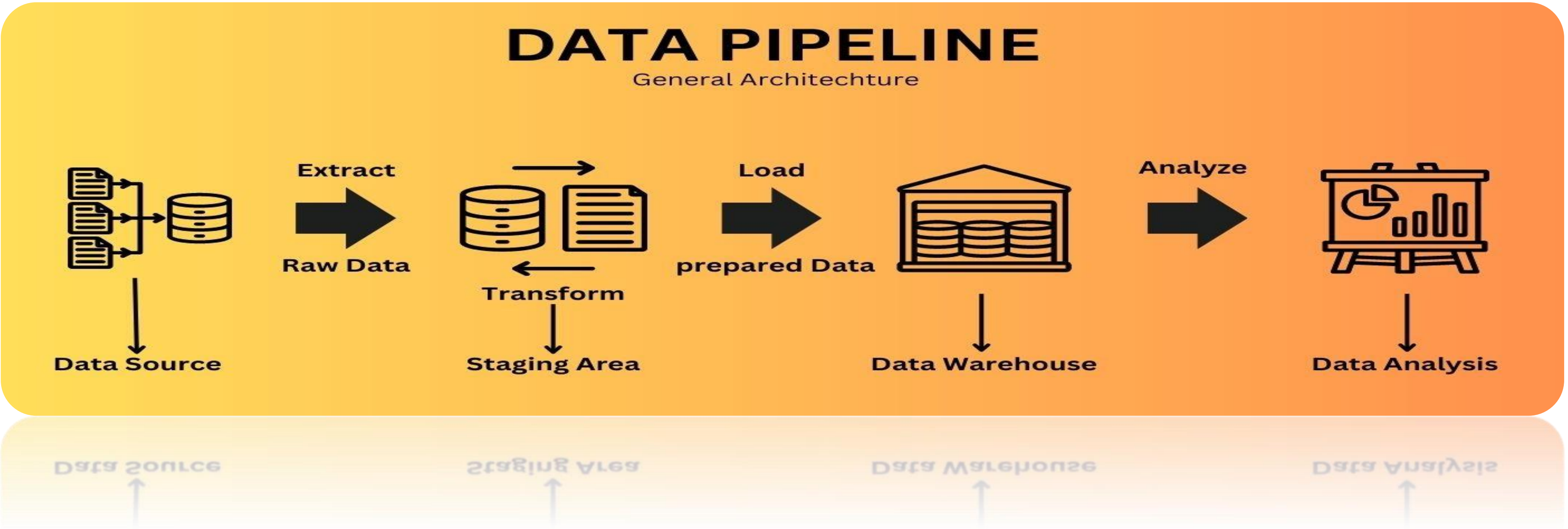
Problem Statement

- Prediction of bankruptcy is a phenomenon of increasing interest to firms who stand to lose money because of unpaid debts. Since computers can store huge dataset pertaining to bankruptcy, making accurate predictions from them beforehand is becoming important.

- For this project, we have aimed to curate a model that captures the bankruptcy patterns among companies in the industry. This model will work to provide early signs of financial downturn in the corporations

Data Pipeline

- Cleaning the Data: The data was checked for null values, categorical values and primary inspection was performed.
- Feature Selection: Techniques such as VIF, p-value, L-1 Regularization and Information Gain were performed to select important features.
- EDA: Exploratory analysis was performed to review the skewness in the data, outliers and correlation patterns.
- Model Testing: Combination of different models and feature selection techniques were used to determine optimal results.
- The dataset consisted of 7 columns with mainly of continuous features in 250 rows and total 1750 values.



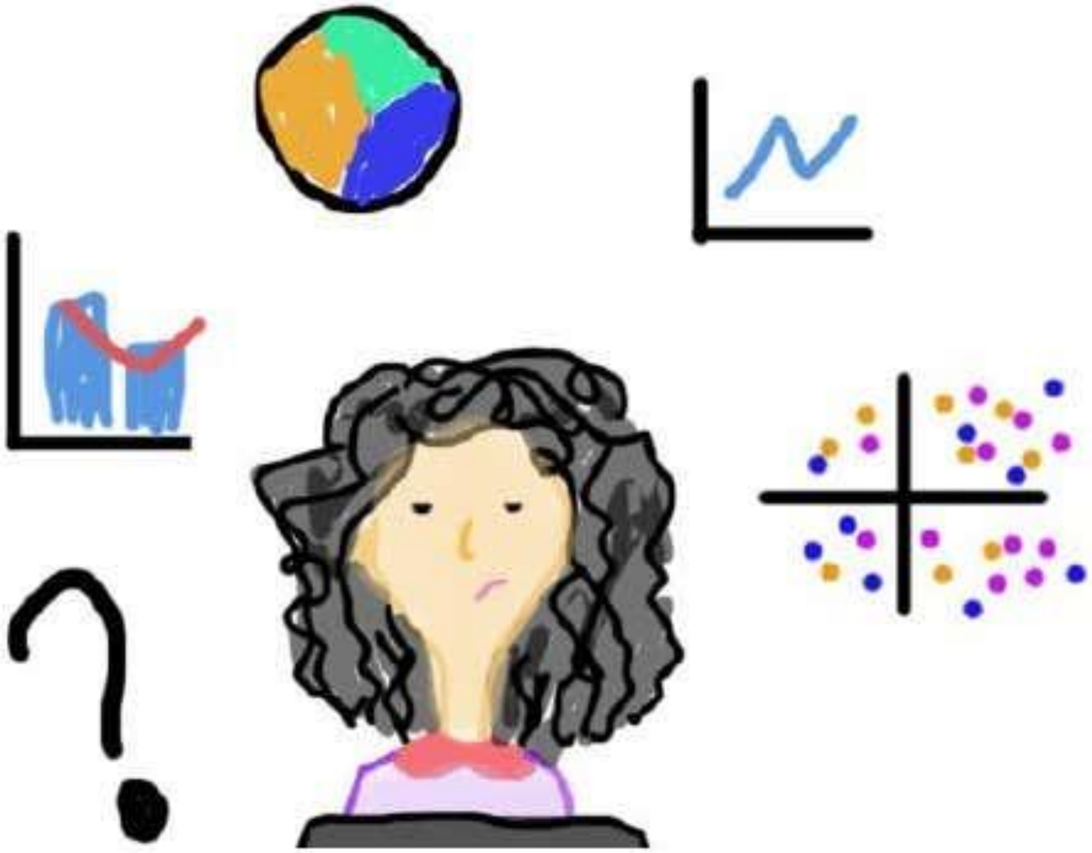
Data Summary

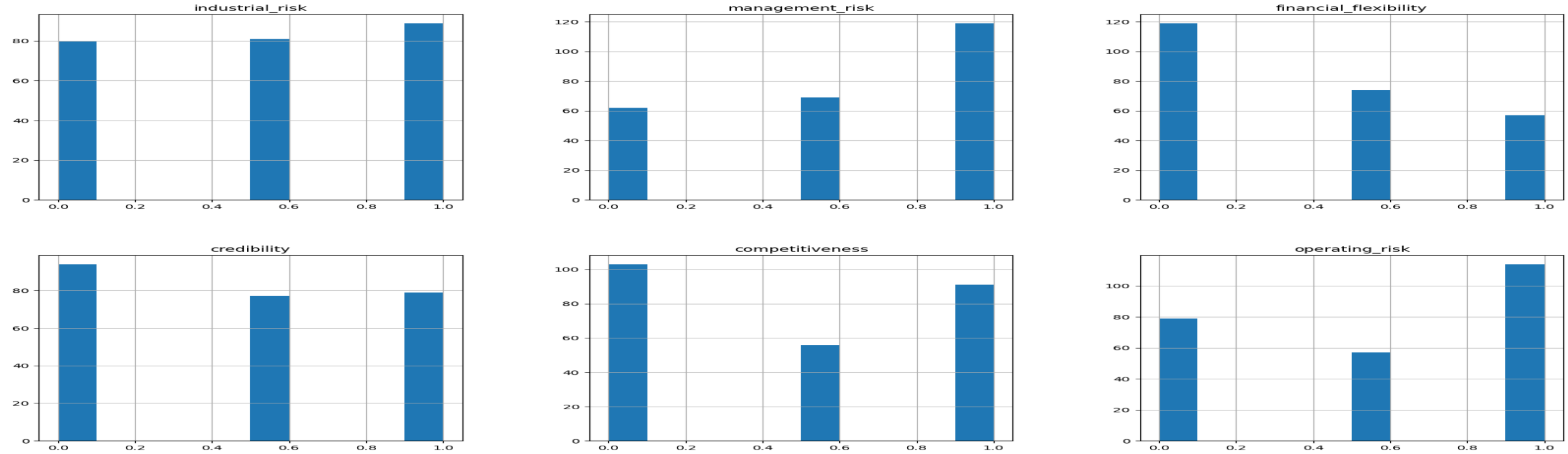
The data file contains 7 features about 250 companies

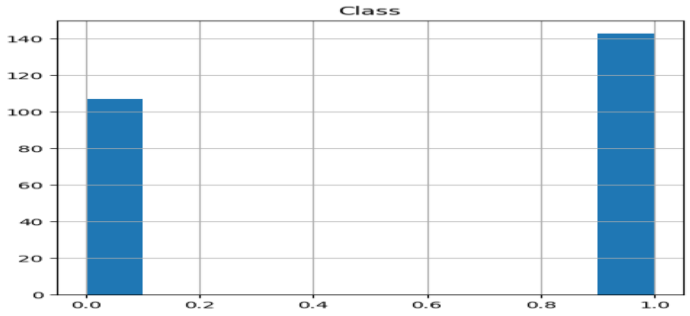
The data set includes the following variables:

1. industrial_risk: 0=low risk, 0.5=medium risk, 1=high risk.
2. management_risk: 0=low risk, 0.5=medium risk, 1=high risk.
3. financial flexibility: 0=low flexibility, 0.5=medium flexibility, 1=high flexibility.
4. credibility: 0=low credibility, 0.5=medium credibility, 1=high credibility.
5. competitiveness: 0=low competitiveness, 0.5=medium competitiveness, 1=high competitiveness.
6. operating_risk: 0=low risk, 0.5=medium risk, 1=high risk.
7. class: bankruptcy, non-bankruptcy (target variable).

Exploratory Data Analysis





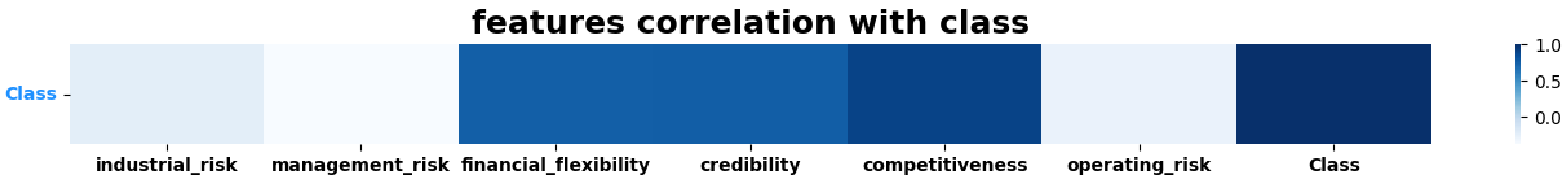



```
non_bankruptcy = df[df['Class'] == 1]            # Couting bankruptcy and non-bankruptcy cases.
bankruptcy = df[df['Class'] == 0]
print("non_bankruptcy : ", non_bankruptcy.shape[0],"\n","bankruptcy : ",bankruptcy.shape[0])
```

```
non_bankruptcy :  143
 bankruptcy :  107
```

Insights

- In above bar graph from 1-6 fig. we are counting the number of low risk, medium risk and high risk by plotting bar graph.
- In 7th fig. we are counting the bankruptcy and non-bankruptcy rate by plotting bar graph with 143 non-bankruptcy and 107 bankruptcy case.




```
realtion = corr['Class'].head(31).to_frame()     #here also same we are checking the relation but is in table form
cm = sns.light_palette("red", as_cmap=True)
realtion.style.background_gradient(cmap=cm)
```

	Class
industrial_risk	-0.227823
management_risk	-0.370838
financial_flexibility	0.751020
credibility	0.755909
competitiveness	0.899452
operating_risk	-0.279786
Class	1.000000

Insights :

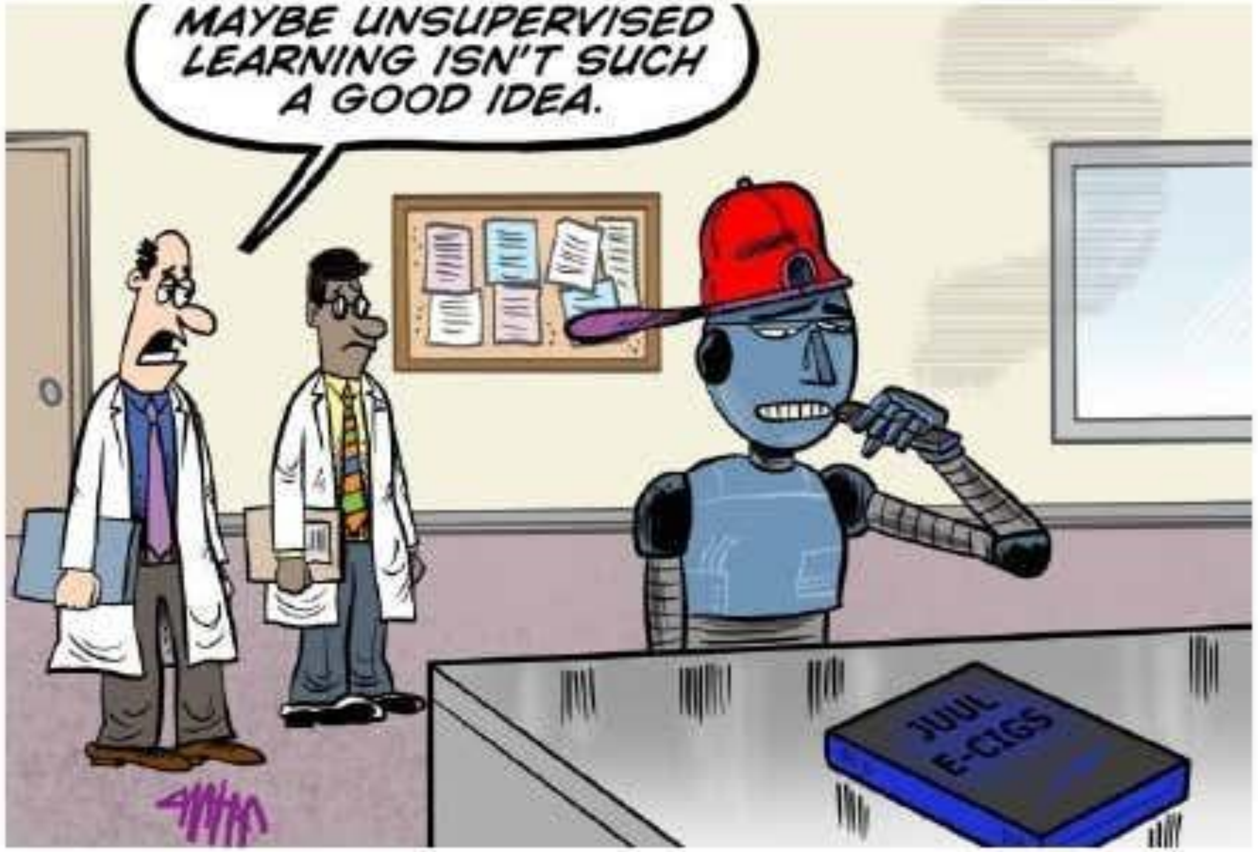
In above Fig. we are checking the correlation of feature classes w.r.t target variable with heat map diagram and also with numerical values.

Conclusion on EDA

- The Problem at hand is not an anomaly detection problem. It is a classification problem with not highly imbalanced dataset.

- We will use different combinations of feature selection techniques, classification models to reach a solution

Applying Models



Applying Model: Logistic Regression

Logistic Regression is a statistical method used to analyze a dataset in which there are one or more independent variables that determine an outcome. It is commonly used for binary classification problems, where the goal is to predict whether an observation belongs to one of two classes.

Applications in Bankruptcy Prediction Models

Logistic Regression is commonly used in bankruptcy prediction models to determine the likelihood of a company going bankrupt. The model takes into account various financial ratios and other factors to predict the probability of bankruptcy within a certain time period.

```
from sklearn.linear_model import LogisticRegression

logisticclassifier = LogisticRegression()
```

```
logisticclassifier.fit(x_train, y_train)

logisticclassifier.coef_ # coefficients of features
```

```
array([[-0.4736723 , -0.84998822,  2.20149964,  2.33553912,  3.72655283,
        -0.45035203]])
```

er traing the model then we prediction on test data

```
y_pred = logisticclassifier.predict(x_test)
y_pred
```

```
array([1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0,
       1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1,
       1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1])
```

```
y_pred_df=pd.DataFrame({"Actual":y_test,"Predicted":y_pred})
y_pred_df
```

	Actual	Predicted
142	1	1
6	0	0
97	0	0
60	0	0
112	1	1
...	...	...
168	1	1
229	1	1
233	1	1
156	1	1
117	1	1

63 rows × 2 columns

```
# Train Accuracy

train_acc_logist = np.mean(logisticclassifier.predict(x_train)== y_train)
train_acc_logist
```

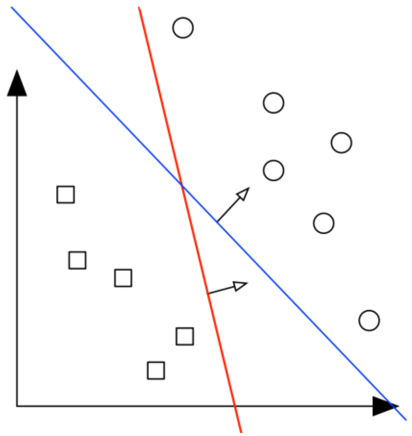
```
0.9946524064171123
```

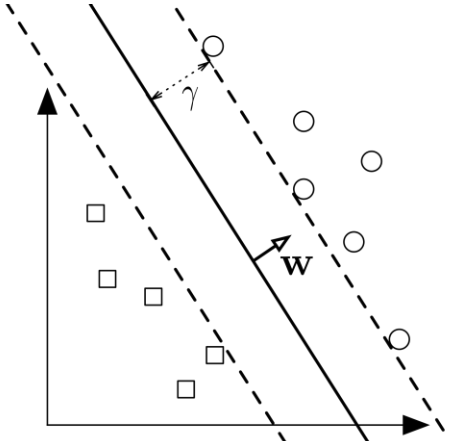
```
# Test Accuracy

test_acc_logist = np.mean(logisticclassifier.predict(x_test)== y_test)
test_acc_logist
```

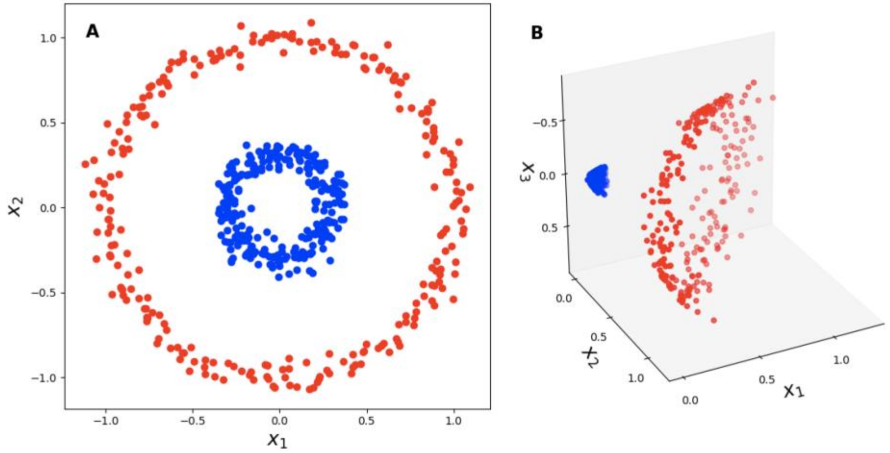
```
1.0
```

Applying Model: Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a popular machine learning algorithm used for classification and regression analysis. They work by finding the hyperplane that best separates the data points into different classes, with the largest margin possible.







Applications
SVMs have been successfully applied in various fields, such as image classification, text classification, bioinformatics, and finance.

Kernel Trick
The kernel trick is a technique used to transform the input data into a higher-dimensional space, where it may be more easily separated by a hyperplane. This allows SVMs to solve non-linear problems, where the data cannot be separated by a linear hyperplane in the original feature space.

```
[ ] from sklearn.svm import SVC
```

Kernel = Linear model

```
model_linear = SVC(kernel = 'linear')

model_linear.fit(x_train, y_train)

pred_test_linear = model_linear.predict(x_test)

np.mean(pred_test_linear==y_test)
```

```
0.9841269841269841
```

```
[ ] y_pred_df4=pd.DataFrame({"Actual":y_test,"Predicted":pred_test_linear})
    y_pred_df4
```

	Actual	Predicted
142	1	1
6	0	0
97	0	0
60	0	0
112	1	1
...	...	...
168	1	1
229	1	1
233	1	1
156	1	1
117	1	1

63 rows × 2 columns

Kernel = ploy model

```
[ ] model_poly = SVC(kernel = "poly")
    model_poly.fit(x_train,y_train)
    pred_test_poly = model_poly.predict(x_test)

    np.mean(pred_test_poly==y_test) # Accuracy
```

```
0.9841269841269841
```


```
y_pred_df4=pd.DataFrame({"Actual":y_test,"Predicted":pred_test_linear})
y_pred_df4
```

	Actual	Predicted
142	1	1
6	0	0
97	0	0
60	0	0
112	1	1
...	...	...
168	1	1
229	1	1
233	1	1
156	1	1
117	1	1

63 rows × 2 columns

Other Algorithms Test Accuracy Chart

	Model_Name	Test_Accuracy
0	Logistic_Regression	100
1	KNN	98
2	Naive_Bayes	100
3	SVM	98
4	Decision_Tree	93
5	Random_Forest	100
6	Bagging Classifier	98
7	Gradient Boosting Classifier	98
8	AdaBOOST Classifier	98

From Above chart we can check the test accuracy for different algorithms which is performed in project.

Streamlit

Streamlit is a popular Python library for creating web applications with minimal effort.

1. **Easy-to-Use**: Streamlit is designed with simplicity in mind. You can create interactive web apps with just a few lines of Python code. This ease of use is a significant advantage for both developers and non-developers.

2. **Rapid Prototyping**: Streamlit is excellent for quickly prototyping data-driven applications. You can iterate on your app's design and functionality without a steep learning curve.

3. **Data Integration**: Streamlit easily integrates with popular data science libraries like Pandas, NumPy, and Matplotlib. This makes it a powerful tool for data analysis and visualization.

4. **Customizable UI**: Although Streamlit is beginner-friendly, it's highly customizable. You can create a tailored user interface with widgets, plots, and text, allowing for a wide range of applications.

5. **Sharing Insights**: Streamlit apps are easy to share with colleagues, clients, or the public. You can deploy them on cloud platforms, making it simple to showcase your data and insights.

6. **Community Support**: Streamlit has a growing and active community. You can find a wealth of resources, from example apps to documentation, making it easier to get started and solve problems.

Streamlit

7. **Interactive Widgets**: Streamlit provides a variety of widgets like sliders, buttons, and text inputs, making your apps interactive and engaging.

8. **Machine Learning Integration**: Streamlit works well with machine learning libraries, allowing you to create data-driven apps that leverage predictive models.

9. **Deployment Options**: Streamlit apps can be deployed locally or in the cloud, and you can use various hosting options such as Streamlit Sharing, Heroku, AWS, and more.

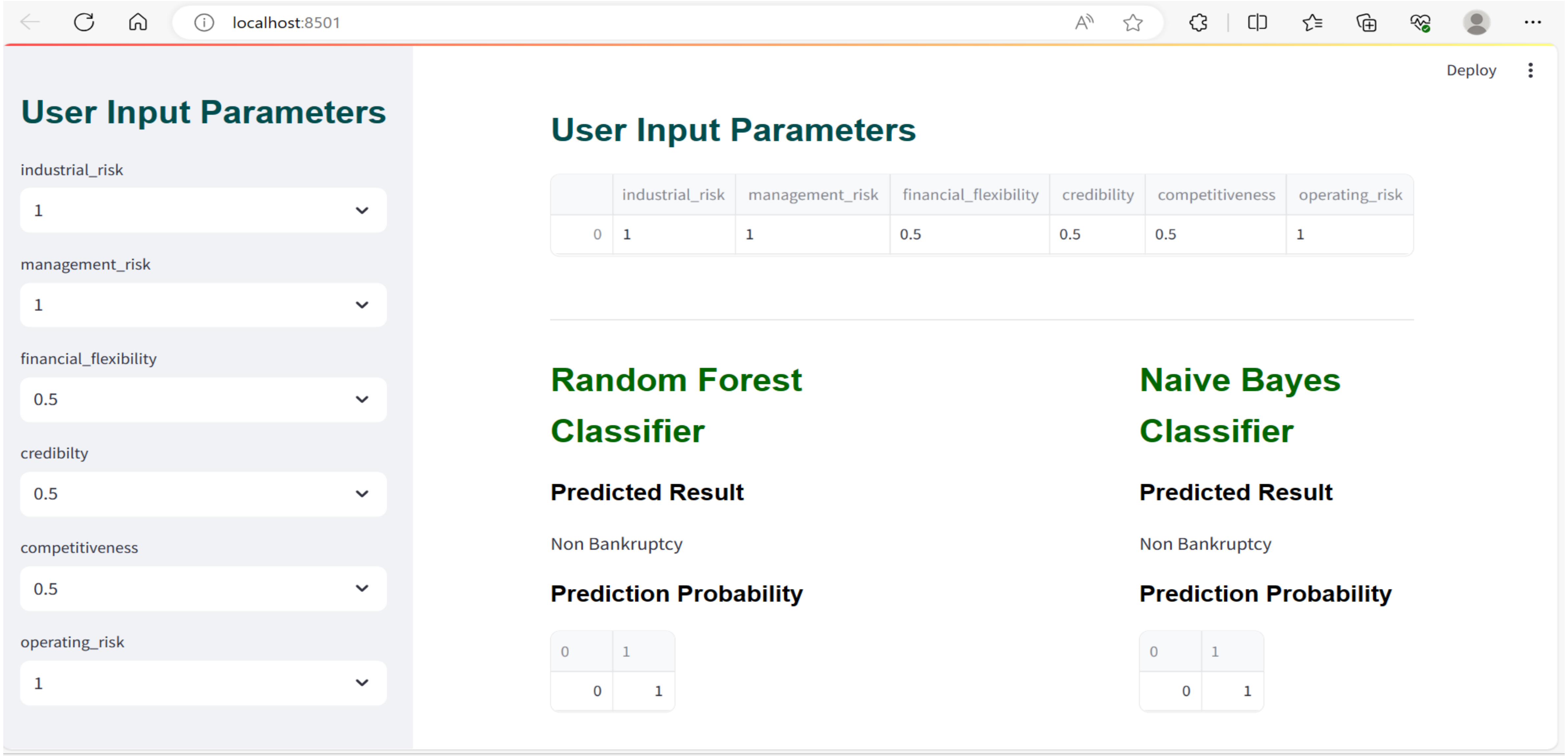
10. **Open Source**: Streamlit is open-source, which means it's free to use and has an active development community.

11. **Documentation and Tutorials**: Streamlit offers extensive documentation and tutorials to help users get started and make the most of the library.

12. **Version Control**: Streamlit apps are written in Python, which allows for version control using tools like Git. This is essential for collaboration and maintaining codebase integrity.

13. **Integration with Other Web Technologies**: You can integrate Streamlit apps with other web technologies like Flask, Dash, and HTML/CSS, allowing for more complex and customized web applications.

Output of Model using Streamlit



Conclusion

1. Improved Accuracy: Machine learning models can significantly improve the accuracy of bankruptcy prediction compared to traditional statistical methods. These models can analyze a wide range of financial and non-financial data, including historical financial statements, market trends, customer behavior, and more, to make more precise predictions.
2. Early Warning: Machine learning models can provide early warning signals of financial distress by identifying patterns and anomalies in data. This allows businesses and financial institutions to take proactive measures to mitigate risk or restructure their operations before bankruptcy becomes imminent.
3. Feature Importance: Machine learning models can identify the most important factors contributing to bankruptcy risk. This information can help businesses and financial institutions focus their attention on key drivers of financial distress and make informed decisions.
4. Model Interpretability: Understanding the reasons behind a bankruptcy prediction is essential for decision-makers. Some machine learning models offer interpretable features that can explain why a particular prediction was made, which can be crucial for taking appropriate actions.
5. Data Challenges: One of the biggest challenges in using machine learning for bankruptcy prediction is the availability and quality of data. Obtaining comprehensive and reliable data can be difficult, especially for smaller businesses or startups. Additionally, historical data may not always capture rapidly changing market conditions.

6. Model Selection: The choice of machine learning algorithms and techniques can greatly impact the performance of bankruptcy prediction models. It's important to select the most appropriate algorithms based on the nature of the data and the specific objectives of the analysis.
7. Regulatory Compliance: In some cases, regulatory requirements may dictate the use of specific models or methodologies for bankruptcy prediction. Machine learning models must be developed and deployed in a manner that complies with relevant regulations.
8. Ethical Considerations: As with any application of machine learning, ethical considerations are important. Bias and fairness issues should be carefully addressed to ensure that predictions do not discriminate against certain groups or individuals.
9. Continuous Monitoring: Machine learning models should be regularly updated and monitored to adapt to changing market conditions and evolving business dynamics. Failure to do so may result in outdated or inaccurate predictions.

In conclusion, machine learning has the potential to enhance bankruptcy prediction and analysis by providing more accurate and timely insights. However, it should be used in conjunction with domain expertise and a thoughtful consideration of data quality, ethics, and regulatory requirements to make informed decisions in the context of bankruptcy assessment.

Thank you