Project Title: Recession prediction

Milestone:
Final Project Report

Group 6

Sai Venkat Madamanchi(Student-1)
Akshay Pagare(Student-2)

(781)-579-1295(Student-1)
(857)-675-0570(Student-2)

madamanchi.s@northeastern.edu
pagare.a@northeastern.edu

Percentage of effort contributed by student-1 : 50%
Percentage of effort contributed by student-2 : 50%

Signature of student 1: Sai Venkat Madamanchi
Signature of student 2: Akshay Pagare

Submission Date: 20/04/2023

Table of contents

Page No

Problem Setting:

A recession is a significant period of economic downturn, characterized by a decline in gross domestic product (GDP) for at least two consecutive quarters. This decline is often accompanied by a rise in unemployment, a decrease in consumer spending and investment, and a reduction in industrial production. Businesses may struggle and fail, and people may lose their jobs, their homes and their savings. It can be a difficult time for individuals, families and the economy as a whole. It is important to take necessary steps to mitigate the effects of a recession and to help the economy recover as quickly as possible.

Problem definition:

Our machine learning project aims to utilize advanced predictive models to accurately forecast the likelihood of a recession by analyzing a wide range of economic indicators such as employment, GDP growth, and consumer spending. We will utilize a machine learning techniques to identify patterns and trends in the data that may indicate a potential recession. Additionally, we will utilize historical data to train and validate our models, providing a more accurate prediction. The ultimate goal of this project is to provide actionable insights and early warning signals, allowing them to better prepare for and potentially mitigate the effects of an economic downturn.

Data sources:

Data extraction involves collecting information from various sources and converting it into a format that can be utilized. In our scenario, the data will be obtained from multiple datasets located on different websites, which may necessitate utilizing by direct downloading. Once the data is extracted, it can be processed, cleaned, and transformed into a analyzable format. The websites are as follows,

U.S. BUREAU OF LABOR STATISTICS website:

https://beta.bls.gov/dataViewer/view/timeseries/LNS14000000,

https://beta.bls.gov/dataViewer/view.

Federal Reserve Bank of st Louis FRED website :-

https://fred.stlouisfed.org/series/CPALTT01USM659N, https://fred.stlouisfed.org/series/FPCPITOTLZGUSA,

https://fred.stlouisfed.org/series/INDPRO,https://fred.stlouisfed.org/series/PCEC96.

Data description:

We are creating final data set from various data sets. Currently, we have 10 attributes and 1 target attribute - "Recession" , which indicates if the instance is a legitimate(0) or a recession(1). The attributes are as follows Gross Domestic Product(GDP), Unemployment, Consumer price index, Inflation, Consumer expenditure, House prices, interest rate, Country/State, Time period, and Earnings by industries. Our data set consists of 300 instances.

Gross Domestic Product (GDP): A decrease in the overall economic output for two consecutive quarters is considered a recession.

Employment: A decrease in the number of employed individuals and an increase in thenumber of unemployed individuals can indicate a recession.

Inflation: A decrease in consumer prices, also known as deflation, can indicate a recession asit is a sign of weak demand.

Housing Market: A decrease in housing prices, construction, and sales can indicate a recession.

Stock Market: A decrease in stock prices, market volatility, and investor confidence canindicate a recession.

Trade Balance: A decrease in exports and an increase in imports can indicate a recession.

Interest Rates: An increase in interest rates can slow down economic growth and may indicate a recession.

Industrial Production: A decrease in industrial production can indicate a slowdown in economic activity and may indicate a recession.

Credit Markets: Tightening credit standards and a decrease in lending can indicate a recession as businesses and consumers may struggle to access capital.

Retail Sales: A decrease in retail sales can indicate a decrease in consumer spending, whichcan be a sign of a recession.

Consumer Debt: An increase in consumer debt levels can indicate that consumers are struggling financially and may decrease their spending, leading to a potential recession.

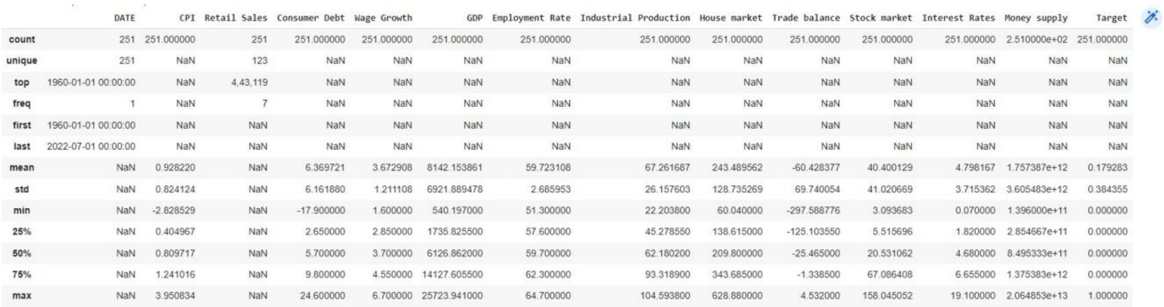
Consumer Spending: A decrease in consumer spending can indicate a decrease in consumer confidence and may indicate a potential recession.

Data Collection:

To create a comprehensive and reliable data set, we searched and retrieved information from several sources on the internet. We then organized and merged the data into a single dataset, ensuring that there was no duplication of information. This approach enabled us to obtain a more extensive and accurate representation of the data, making it easier to analyze and draw valid conclusions. By aggregating the data from various websites, we were able to obtain a more comprehensive picture of the topic under investigation.

Data Exploration:

Figure 1. Data Description



	DATE	CPI	Retail Sales	Consumer Debt	Wage Growth	GDP	Employment Rate	Industrial Production	House market	Trade balance	Stock market	Interest Rates	Money supply	Target
count	251	251.000000	251	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	2.510000e+02	251.000000
unique	251	NaN	123	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	1960-01-01 00:00:00	NaN	4,43,119	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
first	1960-01-01 00:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
last	2022-07-01 00:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	0.928220	NaN	6.369721	3.672908	8142.153861	59.723108	67.261687	243.489562	-60.428377	40.400129	4.798167	1.757387e+12	0.179283
std	NaN	0.824124	NaN	6.161880	1.211108	6921.889478	2.685953	26.157603	128.735269	69.740054	41.020669	3.715362	3.605483e+12	0.384355
min	NaN	-2.828529	NaN	-17.900000	1.600000	540.197000	51.300000	22.203800	60.040000	-297.588776	3.093683	0.070000	1.396000e+11	0.000000
25%	NaN	0.404967	NaN	2.650000	2.850000	1735.825500	57.600000	45.278550	138.615000	-125.103550	5.515696	1.820000	2.854667e+11	0.000000
50%	NaN	0.809717	NaN	5.700000	3.700000	6126.862000	59.700000	62.180200	209.800000	-25.465000	20.531062	4.680000	8.495333e+11	0.000000
75%	NaN	1.241016	NaN	9.800000	4.550000	14127.605500	62.300000	93.318900	343.685000	-1.338500	67.086408	6.655000	1.375383e+12	0.000000
max	NaN	3.950834	NaN	24.600000	6.700000	25723.941000	64.700000	104.593800	628.880000	4.532000	158.045052	19.100000	2.064853e+13	1.000000

The image above provides a detailed representation of the dataset.

As the data in the set was obtained from multiple sources, each source may have used different units or scales to represent the variables. This can result in variations in the measurement units, making it difficult to compare and analyze the data. Therefore, when dealing with a data set that was collected from multiple sources, it is essential to ensure that the variables are appropriately scaled and that they can be compared with each other. It is necessary to standardize the variables to make sure that they are on the same scale and can be accurately analyzed to produce meaningful insights.

Data Processing:

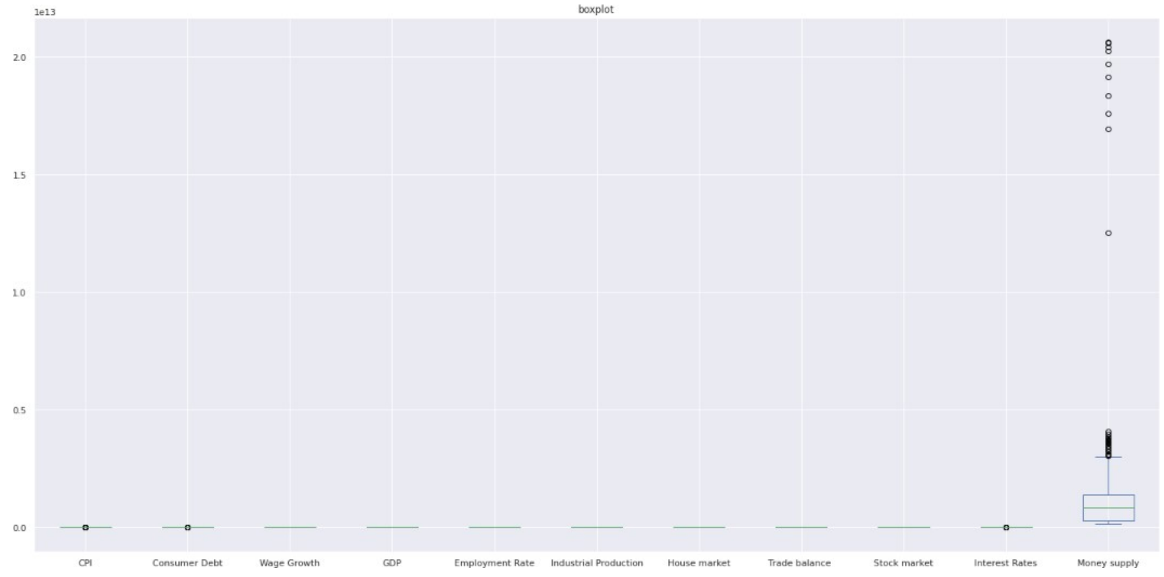
In order to maintain consistency and accuracy, we needed the data for all variables to be within the designated time frame (1960-2022). We utilized the left join function extensively to merge the data into a single dataset within this time frame. We also transformed the data into a usable format that could be analyzed to provide meaningful insights.

However, we encountered some missing values in the data set, which could impact our analysis. To address this, we utilized imputation, which involves replacing missing values with either a fixed value or a value calculated from other observed data. By doing this, we could effectively eliminate null values from the dataset, ensuring that the data was complete and usable for analysis.

Data Visualization:

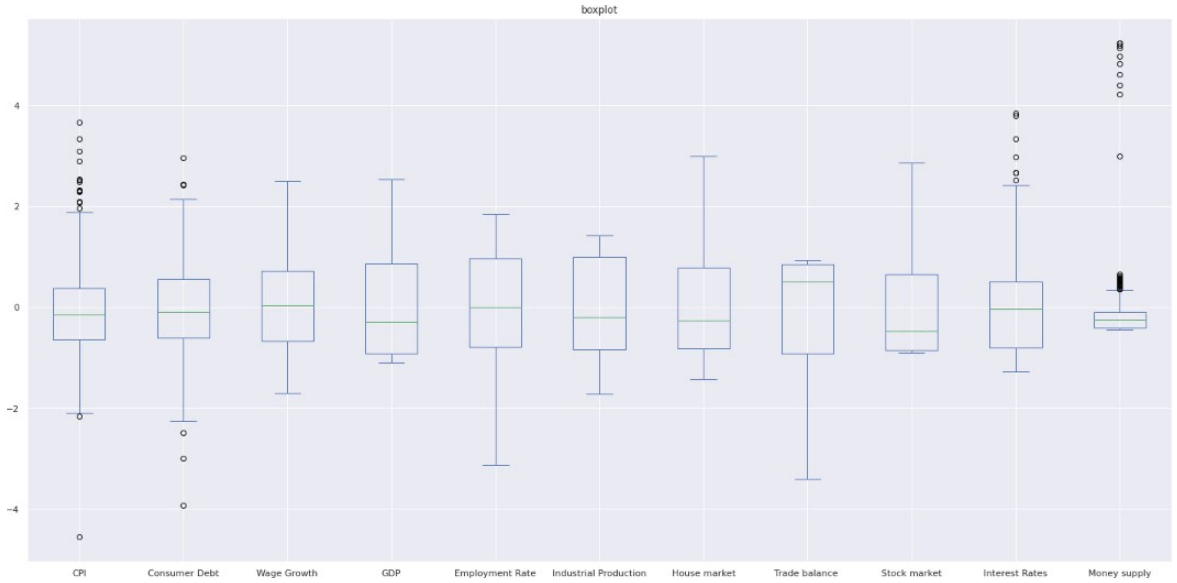
We recognized the importance of visual representation in understanding complex data. Therefore, to gain insights into the details of the data set, we used a few meaningful graphs. By doing this, we were able to visualize the relationship between different variables in the data set, identify any patterns or trends, and make valid conclusions. This approach also allowed us to communicate our findings more effectively by presenting our analysis in a clear and concise way that is easily understandable. Additionally, using meaningful graphs, we were able to highlight any important insights that could be overlooked when analyzing data in a tabular form.

Figure 2. Data distribution before scaling the data set:



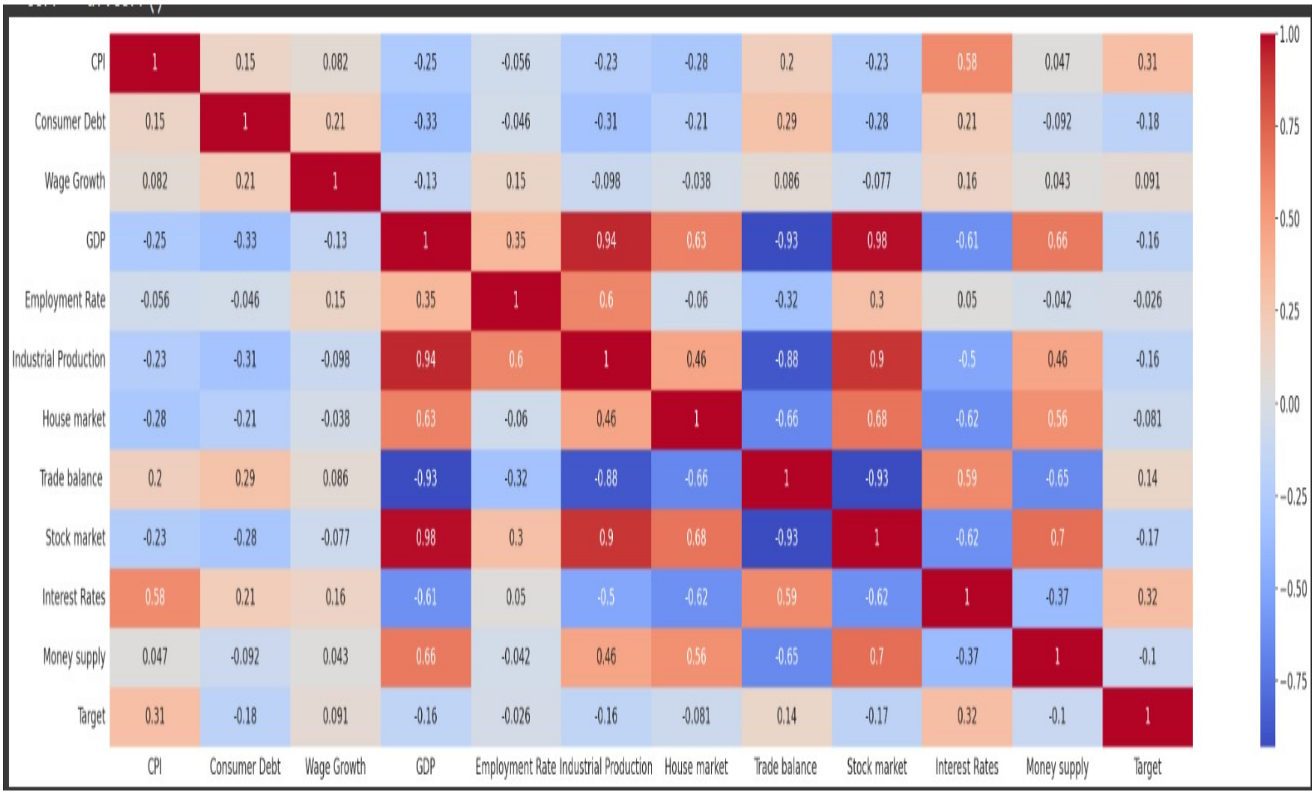
Insight:- From the series of box plots, we were able to observe a significant variation in the scales of each variable.

Figure 3. Data distribution after the scaling :-



Insight:-The data became much more clear after scaling, as evidenced by the series of box plots above.

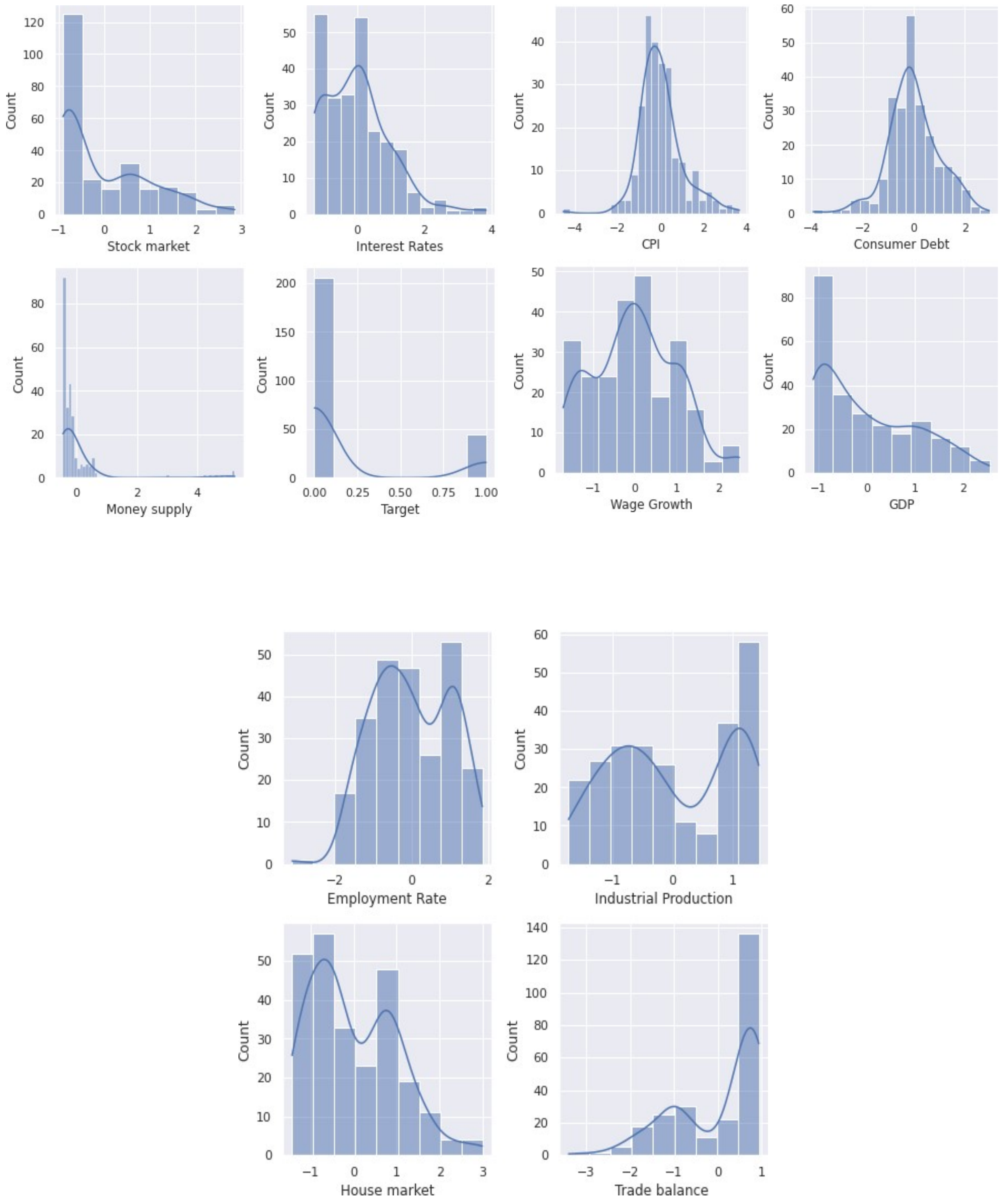
Figure 4. Correlation Heat map for the scaled data:



	CPI	Consumer Debt	Wage Growth	GDP	Employment Rate	Industrial Production	House market	Trade balance	Stock market	Interest Rates	Money supply	Target
CPI	1	0.15	0.082	-0.25	-0.056	-0.23	-0.28	0.2	-0.23	0.58	0.047	0.31
Consumer Debt	0.15	1	0.21	-0.33	-0.046	-0.31	-0.21	0.29	-0.28	0.21	-0.092	-0.18
Wage Growth	0.082	0.21	1	-0.13	0.15	-0.098	-0.038	0.086	-0.077	0.16	0.043	0.091
GDP	-0.25	-0.33	-0.13	1	0.35	0.94	0.63	-0.93	0.98	-0.61	0.66	-0.16
Employment Rate	-0.056	-0.046	0.15	0.35	1	0.6	-0.06	-0.32	0.3	0.05	-0.042	-0.026
Industrial Production	-0.23	-0.31	-0.098	0.94	0.6	1	0.46	-0.88	0.9	-0.5	0.46	-0.16
House market	-0.28	-0.21	-0.038	0.63	-0.06	0.46	1	-0.66	0.68	-0.62	0.56	-0.081
Trade balance	0.2	0.29	0.086	-0.93	-0.32	-0.88	-0.66	1	-0.93	0.59	-0.65	0.14
Stock market	-0.23	-0.28	-0.077	0.98	0.3	0.9	0.68	-0.93	1	-0.62	0.7	-0.17
Interest Rates	0.58	0.21	0.16	-0.61	0.05	-0.5	-0.62	0.59	-0.62	1	-0.37	0.32
Money supply	0.047	-0.092	0.043	0.66	-0.042	0.46	0.56	-0.65	0.7	-0.37	1	-0.1
Target	0.31	-0.18	0.091	-0.16	-0.026	-0.16	-0.081	0.14	-0.17	0.32	-0.1	1

Insight : Since we have both positive and negative target variables with respect to the recession, we can observe the correlation between the variables from the graph above.

Figure 5. Distribution of all variables



Insight:- By analyzing the internal distribution of each variable, we were able to gain insights into the behavior of the elements in the data set. We discovered a variety of distributions, including left-skewed, perfectly symmetrical bell curves, unimodal, bimodal, and trimodal distributions throughout the elements in the data set.

Data Mining tasks:

Dimension Reduction:

During data pre-processing, exploratory data analysis, and visualization in our scenario, we reduced a few variables. During the process of data dimension reduction. The 99% of the variation in the data set is extracted using principal component analysis. In light of the findings, we decreased the number of predictors from 11 to 9. However, the two predictors still only account for 1% of the overall variation.

The Final data set consists of the following variables :- 'Trade balance ', 'GDP', 'Stock market', 'Industrial Production', 'House market', 'Employment Rate', 'Money supply', 'CPI', 'Consumer Debt'.

Data Models:

In our project we have numeric predictors and categorical response. So, we choose

1) Logistic Regression

2) Decision Tree Classifier

3) K neighbours Classifier

4) Naive Bayes

5) Neural network

6) Discriminant Analysis

Logistic Regression:

Logistic Regression is a classification algorithm that is used to predict the probability of a binary or categorical target variable based on one or more independent variables. The goal of logistic regression is to find the best fitting S-shaped curve that can estimate the probability of the target variable. The equation for logistic regression is typically of the form p = 1 / (1 + e^-(mx + b)), where p is the probability of the target variable, x is the independent variable, m is the slope of the curve, and b is the intercept.

Advantages of Logistic Regression:

- It is a simple and effective algorithm for binary classification problems.
- It can handle both linear and nonlinear relationships between variables.
- It can be easily interpreted and can provide insights into the factors that influence the target variable.

Disadvantages of Logistic Regression:

- It is not suitable for multi-class classification problems.
- It assumes a linear relationship between the independent variables and the log-odds of the target variable.
- It is sensitive to overfitting when there are too many independent variables.

Decision Tree Classifier:

Decision Tree Classifier is a tree-based algorithm that is used for classification and regression problems. The goal of decision tree classifier is to divide the dataset into smaller subsets based on the values of independent variables, in order to predict the target variable. The algorithm creates a tree structure where each node represents a test on an independent variable, and each branch represents the outcome of the test.

Advantages of Decision Tree Classifier:

- It can handle both categorical and numerical variables.
- It can capture non-linear relationships between variables.
- It can be easily interpreted and can provide insights into the factors that influence the target variable.

Disadvantages of Decision Tree Classifier:

- It is prone to overfitting when the tree is too deep.
- It is sensitive to the order of the variables and the splitting criteria.
- It can be unstable and produce different results for different subsets of the data.

K Nearest Neighbors Classifier:

K Nearest Neighbors (KNN) Classifier is a non-parametric algorithm that is used for classification and regression problems. The goal of KNN classifier is to find the K nearest data points to a new data point based on the distance metric, and predict the target variable based on the majority class of the K neighbors.

Advantages of KNN Classifier:

- It is a simple and effective algorithm for classification problems.
- It can handle both linear and nonlinear relationships between variables.
- It can be easily adapted to handle multi-class classification problems.

Disadvantages of KNN Classifier:

- It is computationally expensive and slow for large datasets.
- It is sensitive to the choice of distance metric and the value of K.
- It cannot handle missing data and requires all features to be standardized.

Naive Bayes Classifier:

Naive Bayes Classifier is a probabilistic algorithm that is used for classification problems. The goal of Naive Bayes classifier is to predict the target variable based on the probability of the independent variables given the target variable. The algorithm assumes that the independent variables are conditionally independent given the target variable.

Advantages of Naive Bayes Classifier:

- It is a simple and fast algorithm for classification problems.
- It can handle high-dimensional data and is robust to irrelevant features.
- It can be easily adapted to handle multi-class classification problems.

Disadvantages of Naive Bayes Classifier:

- It assumes that the independent variables are conditionally independent given the target variable, which may not be true in practice.
- It cannot capture complex relationships between variables.
- It may suffer from the problem of zero frequency.

Neural Network:

Neural Network is a powerful and flexible algorithm that is used for classification, regression, and other machine learning tasks. The goal of Neural Network is to learn a complex mapping between the input and output variables by using multiple layers of interconnected neurons. Each neuron applies a mathematical function to the weighted sum of its inputs, and the output of one layer becomes the input to the next layer.

Advantages of Neural Network:

- It can capture complex and nonlinear relationships between variables.
- It can be trained to handle a wide range of machine learning tasks.
- It can handle large datasets and high-dimensional data.

Disadvantages of Neural Network:

- It is computationally expensive and requires a large amount of data to train.
- It can be difficult to interpret and understand the inner workings of the model.
- It is prone to over fitting when the model is too complex or the data is insufficient.

Discriminant Analysis:

Discriminant Analysis is a classification algorithm that is used to predict the class of a target variable based on one or more independent variables. The goal of Discriminant Analysis is to find a linear combination of the independent variables that can best separate the classes of the target variable. The algorithm can be either linear or quadratic, depending on the distribution of the independent variables.

Advantages of Discriminant Analysis:

- It can handle both binary and multi-class classification problems.
- It can capture the correlations between the independent variables.
- It can be easily interpreted and can provide insights into the factors that influence the target variable.

Disadvantages of Discriminant Analysis:

- It assumes a linear or quadratic relationship between the independent variables and the target variable.
- It is sensitive to overfitting when there are too many independent variables.
- It may require assumptions about the distribution of the data, which may not always be accurate.

Model Performance Evaluation and Interpretation:

Performance evaluation methods are essential in assessing the effectiveness of a model and determining if it can produce accurate and reliable predictions. These methods are used to compare the predicted values of a model with the actual values of a data-set. Some of the commonly used evaluation metrics and tools include:

Mean Absolute Error (MAE): This metric measures the average absolute difference between the predicted values and the actual values. It is computed by taking the average of the absolute differences between the predicted and actual values. The lower the MAE, the better the model's performance.

Root Mean Squared Error (RMSE): This metric measures the square root of the average squared difference between the predicted and actual values. The lower the RMSE, the better the model's performance.

F1 score: F1 score is a harmonic mean of precision and recall. It is often used to evaluate classification models when the number of classes is not balanced.

For Logistic Regression

Logistic Regression, has an accuracy of 0.79, which means that the model has correctly predicted the target variable in 79% of the cases. The precision score of 0.75 indicates that out of all the positive predictions, only 75% of them are actually correct.

The F1 score of 0.75 suggests that the model has a reasonable balance between precision and recall. The MSE and RMSE values are 0.21 and 0.45, respectively, which indicates that the model has a moderate level of error.

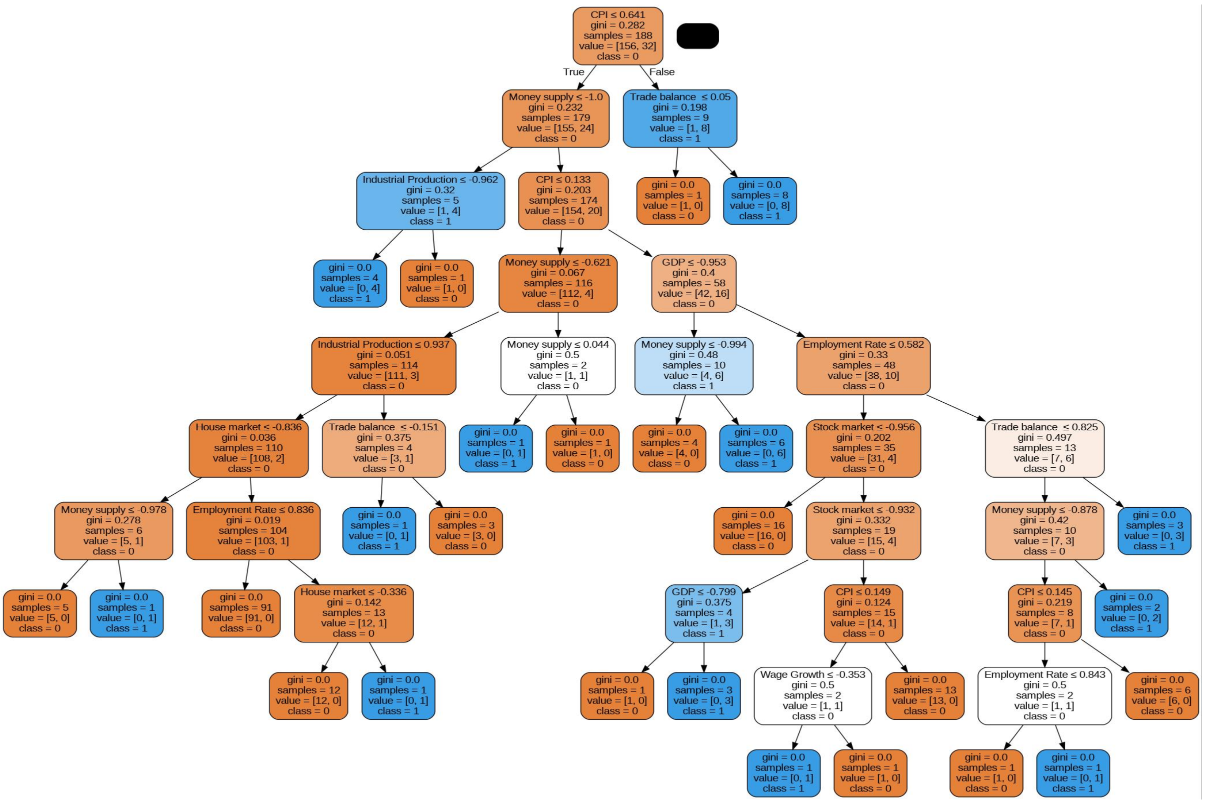
Overall, Logistic Regression is a reasonable model, with good accuracy and a reasonable balance between precision and recall.

For Classification Tree

The second model, Classification Tree, has an accuracy of 0.78, which is slightly lower than the Logistic Regression model. The precision score of 0.74 is similar to that of Logistic Regression, indicating that out of all the positive predictions, only 74% of them are actually correct.

The F1 score of 0.75 is slightly higher than that of Logistic Regression. The MSE and RMSE values are 0.22 and 0.47, respectively, which indicates that the model has a moderate level of error.

Figure 6. Classifiction tree


[redacted]

CPI ≤ 0.641 gini = 0.282 samples = 188 value = [156, 32] class = 0
True
False
Money supply ≤ -1.0 gini = 0.232 samples = 179 value = [155, 24] class = 0
Trade balance ≤ 0.05 gini = 0.198 samples = 9 value = [1, 8] class = 1
Industrial Production ≤ -0.962 gini = 0.32 samples = 5 value = [1, 4] class = 1
CPI ≤ 0.133 gini = 0.203 samples = 174 value = [154, 20] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 8 value = [0, 8] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
Money supply ≤ -0.621 gini = 0.067 samples = 116 value = [112, 4] class = 0
GDP ≤ -0.953 gini = 0.4 samples = 58 value = [42, 16] class = 0
Industrial Production ≤ 0.937 gini = 0.051 samples = 114 value = [111, 3] class = 0
Money supply ≤ 0.044 gini = 0.5 samples = 2 value = [1, 1] class = 0
Money supply ≤ -0.994 gini = 0.48 samples = 10 value = [4, 6] class = 1
Employment Rate ≤ 0.582 gini = 0.33 samples = 48 value = [38, 10] class = 0
House market ≤ -0.836 gini = 0.036 samples = 110 value = [108, 2] class = 0
Trade balance ≤ -0.151 gini = 0.375 samples = 4 value = [3, 1] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 4 value = [4, 0] class = 0
gini = 0.0 samples = 6 value = [0, 6] class = 1
Stock market ≤ -0.956 gini = 0.202 samples = 35 value = [31, 4] class = 0
Trade balance ≤ 0.825 gini = 0.497 samples = 13 value = [7, 6] class = 0
Money supply ≤ -0.978 gini = 0.278 samples = 6 value = [5, 1] class = 0
Employment Rate ≤ 0.836 gini = 0.019 samples = 104 value = [103, 1] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 3 value = [3, 0] class = 0
gini = 0.0 samples = 16 value = [16, 0] class = 0
Stock market ≤ -0.932 gini = 0.332 samples = 19 value = [15, 4] class = 0
Money supply ≤ -0.878 gini = 0.42 samples = 10 value = [7, 3] class = 0
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.0 samples = 5 value = [5, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 91 value = [91, 0] class = 0
House market ≤ -0.336 gini = 0.142 samples = 13 value = [12, 1] class = 0
GDP ≤ -0.799 gini = 0.375 samples = 4 value = [1, 3] class = 1
CPI ≤ 0.149 gini = 0.124 samples = 15 value = [14, 1] class = 0
CPI ≤ 0.145 gini = 0.219 samples = 8 value = [7, 1] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.0 samples = 12 value = [12, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 3 value = [0, 3] class = 1
Wage Growth ≤ -0.353 gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.0 samples = 13 value = [13, 0] class = 0
Employment Rate ≤ 0.843 gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.0 samples = 6 value = [6, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1

Overall, Classification Tree is a reasonable model, but it performs slightly worse than Logistic Regression.

For Naive Bayes

Naive Bayes, has the lowest accuracy score of 0.54, which means that the model has correctly predicted the target variable in only 54% of the cases. The precision score of 0.73 is similar to the other models, but the F1 score of 0.58 is significantly lower than that of the other models.

The MSE and RMSE values are 0.46 and 0.68, respectively, indicating that the model has a high level of error.

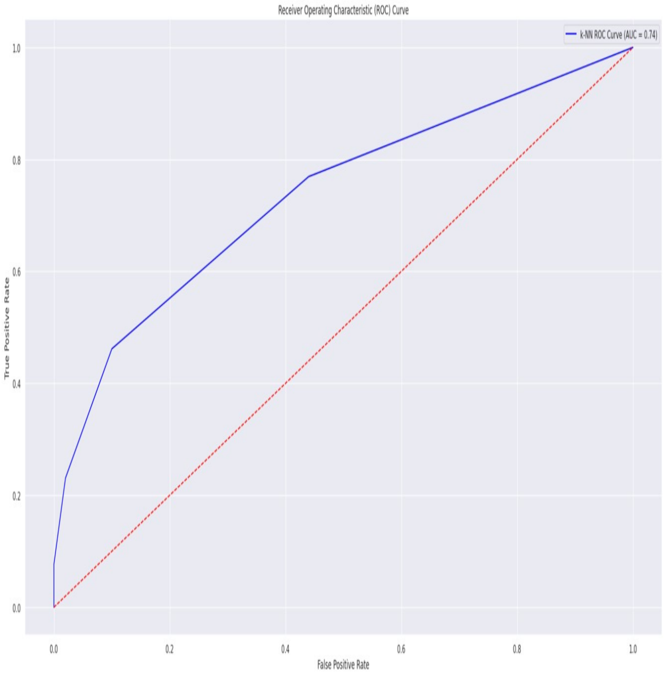
Overall, Naive Bayes is the worst-performing model in terms of accuracy and F1 score, and it has a high level of error.

For K neighbours Classifier

The fourth model, K-NN, has the highest accuracy score of 0.83, which means that the model has correctly predicted the target variable in 83% of the cases. The precision score of 0.81 is the highest among all the models, indicating that out of all the positive predictions, 81% of them are actually correct.

The F1 score of 0.79 is also the highest among all the models, indicating that the model has a good balance between precision and recall. The MSE and RMSE values are 0.17 and 0.42, respectively, indicating that the model has a low level of error.

Overall, K-NN is the best-performing model in terms of accuracy, precision, F1 score, and error rate.



For Neural Network

Neural Network had an accuracy of 77.78%, precision of 75.31%, and an F1 score of 76.18%. The model had a MSE of 22.22% and RMSE of 0.47. The model's performance was similar to that of the Classification Tree model.

For Discriminant Analysis

Discriminant Analysis had an accuracy of 79.37%, precision of 77.70%, and an F1 score of 78.32%. The model had a MSE of 20.64% and RMSE of 0.45. The model performed similarly to the Logistic Regression model, but with slightly better precision and F1 score.

Project Results:

Figure 7. RMSE of Models

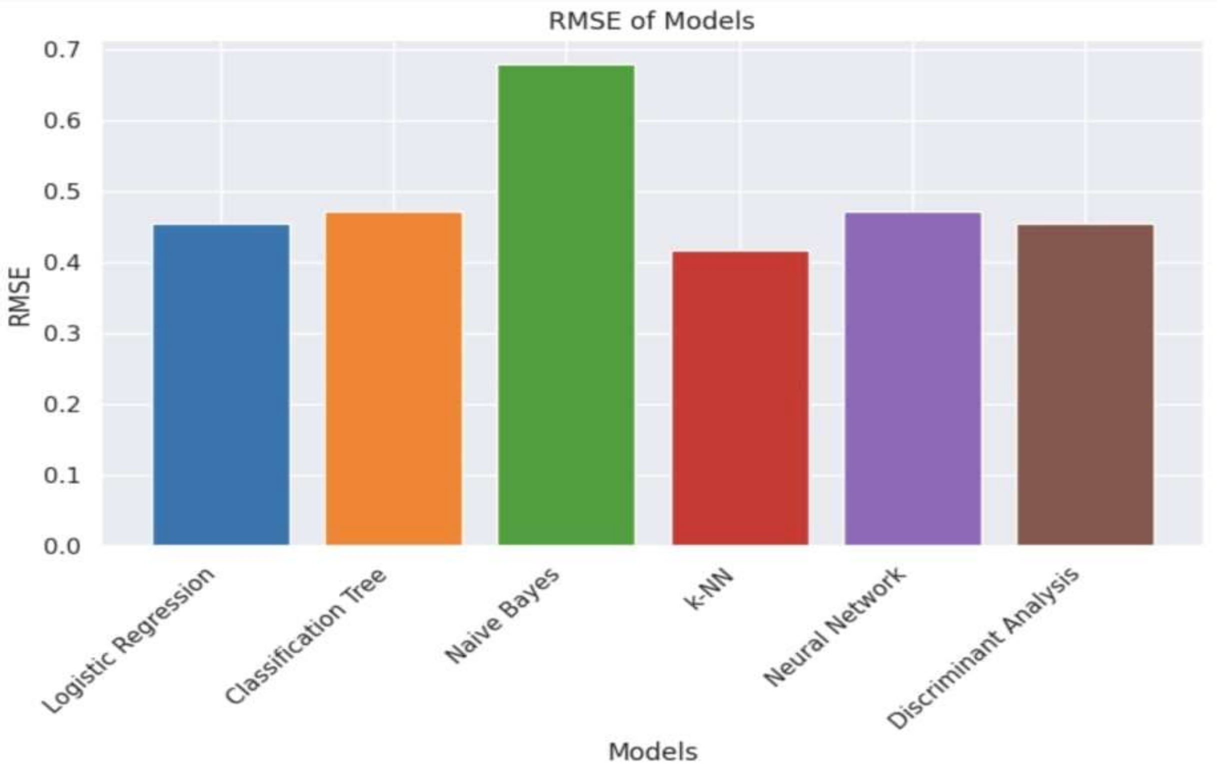
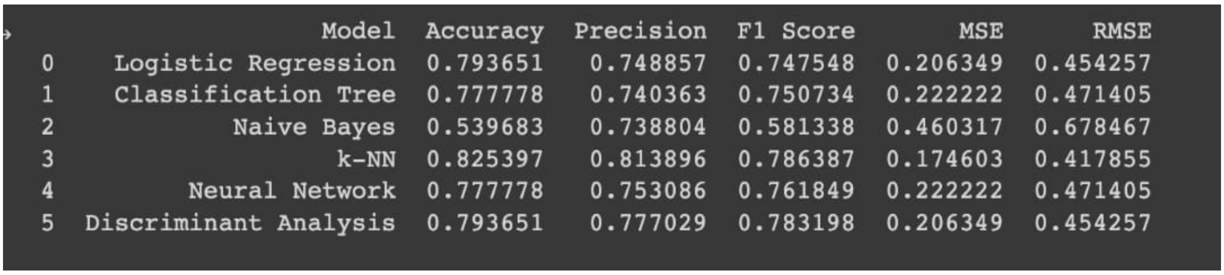


Figure 8. Overall, Metrics of Models.



```
                    Model  Accuracy  Precision  F1 Score       MSE      RMSE
0     Logistic Regression  0.793651   0.748857  0.747548  0.206349  0.454257
1     Classification Tree  0.777778   0.740363  0.750734  0.222222  0.471405
2             Naive Bayes  0.539683   0.738804  0.581338  0.460317  0.678467
3                    k-NN  0.825397   0.813896  0.786387  0.174603  0.417855
4          Neural Network  0.777778   0.753086  0.761849  0.222222  0.471405
5  Discriminant Analysis  0.793651   0.777029  0.783198  0.206349  0.454257
```

Overall, the k-NN model performed the best among the six models, with high accuracy, precision, and F1 score, and low MSE and RMSE. The Logistic Regression and Discriminant Analysis models also performed relatively well. However, the Classification Tree and Neural Network models had lower accuracy and higher MSE and RMSE, while the Naive Bayes model had the lowest accuracy and highest MSE and RMSE.

The choice of model depends on the specific requirements of the problem being addressed, but k-NN can be considered as a promising candidate for classification problems.

Impact of Project Outcomes:

The impact of the project outcomes can be significant, as accurate predictions of an economic recession can provide valuable insights for policymakers, businesses, and individuals. By identifying early warning signals of a potential recession, proactive measures can be taken to mitigate the effects of an economic downturn, potentially minimizing the negative impact on the economy and society.

For policymakers, accurate forecasting of a recession can enable them to implement policies and measures to stabilize the economy and protect vulnerable groups. For example, they could increase government spending or decrease interest rates to stimulate economic growth and reduce unemployment. Businesses can also benefit from the predictive power of machine learning models, as they can adjust their operations, investment, and hiring plans accordingly. Individuals may also be able to take measures to protect their financial well-being, such as reducing debt levels or increasing savings.

However, it is important to note that predictive machine learning models are not infallible and should not be relied upon exclusively. Economic recessions can be complex and multifaceted, and there may be factors that the models do not consider. Additionally, false positives and false negatives can occur, leading to incorrect predictions. Therefore, the models should be used in conjunction with other economic indicators and analyses to provide a more comprehensive understanding of the economy.