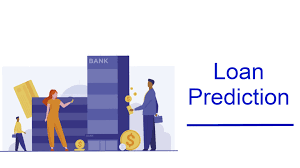
**Title: Loan Amount Prediction using Python**

**:- Gagan**

****

**Abstract:**

**This report aims to explore the application of Python in predicting loan amounts. The prediction of loan amounts plays a crucial role in the financial industry, assisting lenders in making informed decisions while approving loans. In this report, we will discuss the dataset used, the preprocessing steps involved, the different predictive models implemented, and evaluate their performance. Python libraries such as Pandas, NumPy, and Scikit-learn will be utilized to analyze the data, build and evaluate the models. Through this report, we aim to demonstrate the effectiveness of Python in loan amount prediction.**

**Table of Contents:**

**1. Introduction**

**1.1 specifications**

**1.2 Objectives**

**2. Data Collection and Preprocessing**

**2.1 Data Source**

**2.2 Dataset Description**

**2.3 Data Analysis**

**3. Feature Engineering**

**3.1 Handling Missing Values numerical data**

**3.2 Handling Categorical data**

**4. Data Visualization**

**5. Model Selection**

**5.1 Linear Regression**

**5.2 Confusion matrix**

**5.3 Decision tree Regressor**

**5.4 Random Forest Regressor**

**6. Conclusion**

**6.1 Summary of Findings**

**6.2 Limitations and Future Work**

**6.3 Evaluation**

**7. References**

**1.Introduction:-**

**1.1. Specifications/Libraries:-**

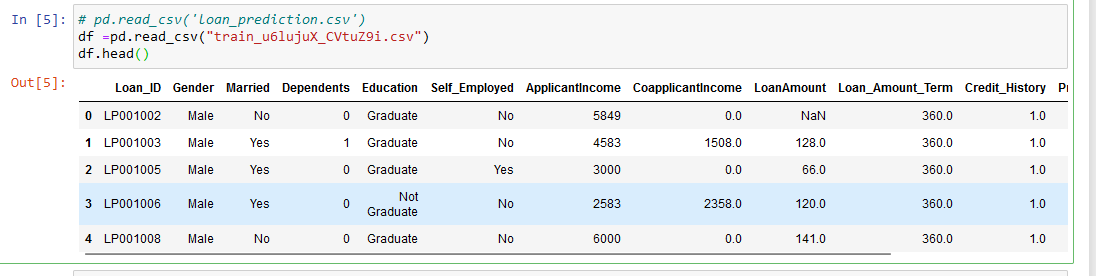
* pandas
* Python
* seaborn
* sklearn
* matplotlib
* Warnings

**1.2 Objectives:-**

This report presents the loan amount prediction model. The objective of the model is to predict the loan amount based on various input features. The model can be useful for financial institutions in assessing loan applications and determining appropriate loan amounts

**2. Data Collection and Preprocessing:-**

**2.1 Data Source:-**

****

****

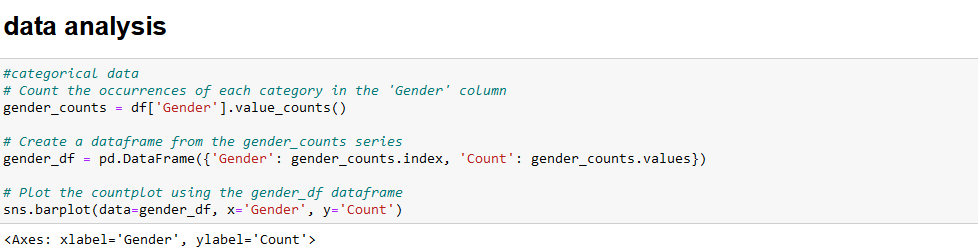
For data source , we have been given two CSV files: train, test

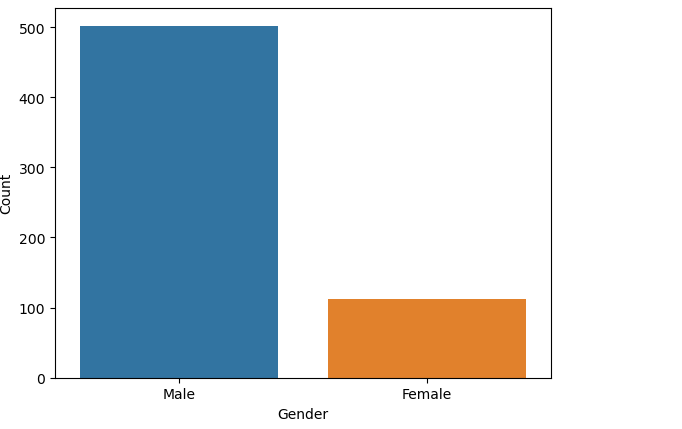
We will be reading data or for analysing data we will be using pandas library

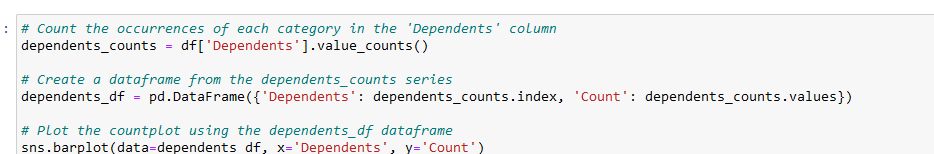
**2.2 Dataset Description:-**

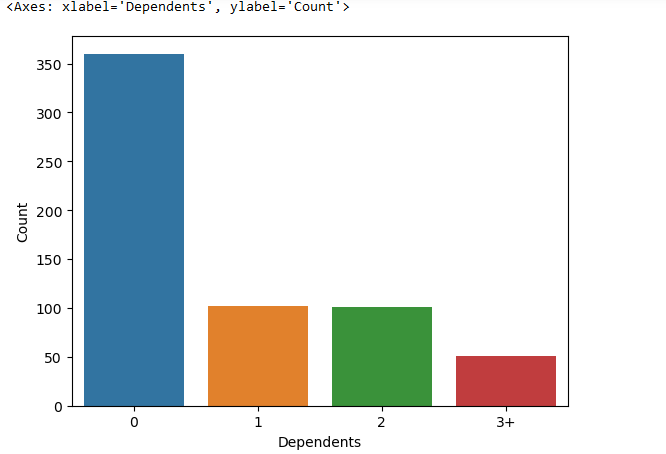
The dataset used for training and testing the loan amount prediction model consists of several columns that provide information about the applicants. Each column has a specific meaning and relevance to the loan amount prediction task. The data includes details such as applicant income, co-applicant income, loan amount, loan term, credit history, gender, marital status, education, employment status, property area, and dependents. These features help in understanding the financial background and circumstances of loan applicants.

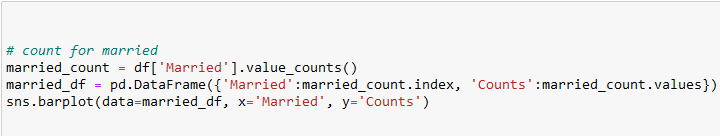
**2.3 Data Analysis:-**

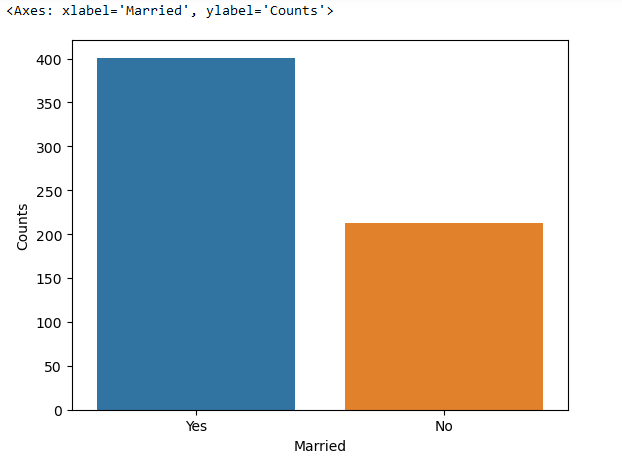
****

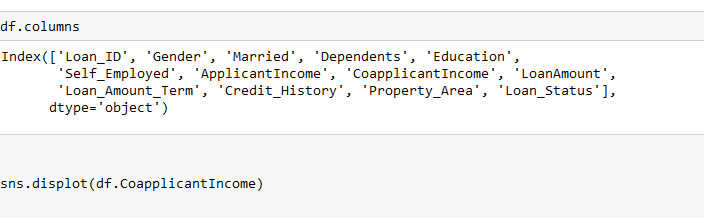
****

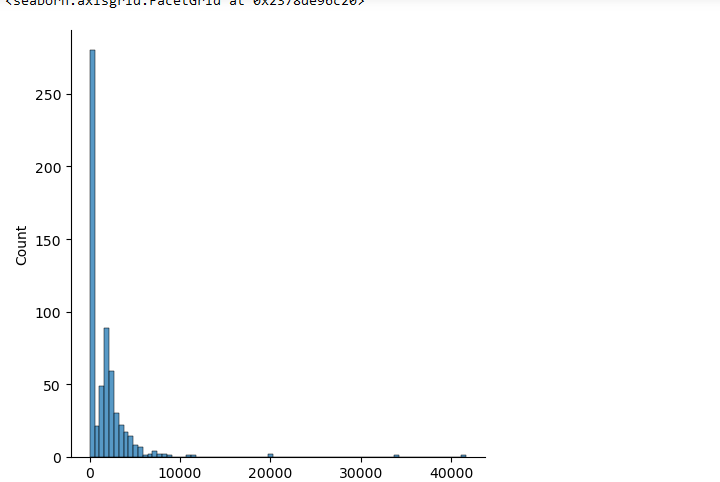
****

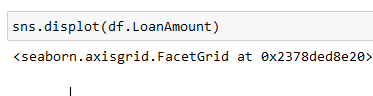
****

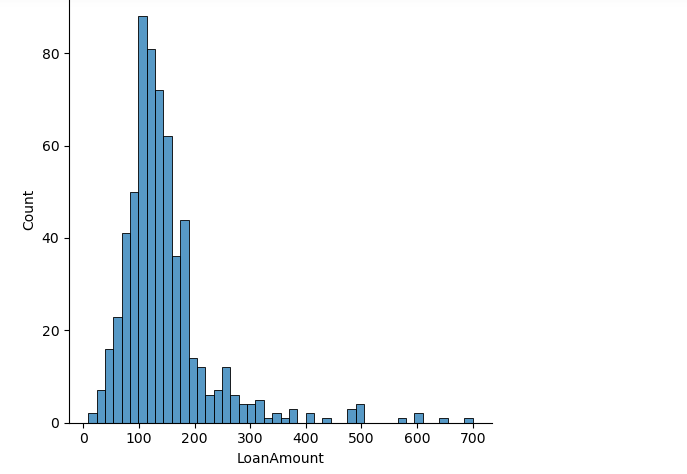
****

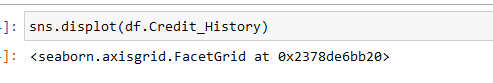
****

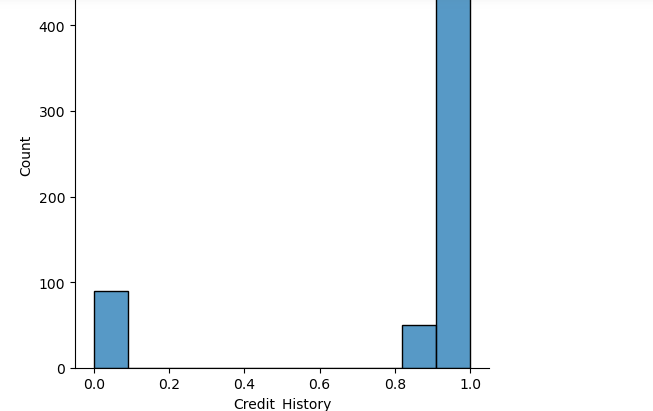
****

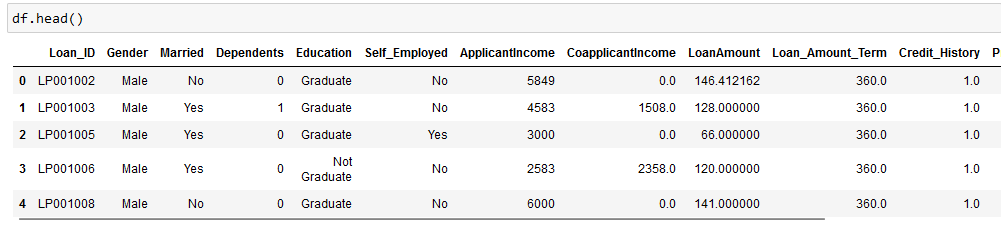
****

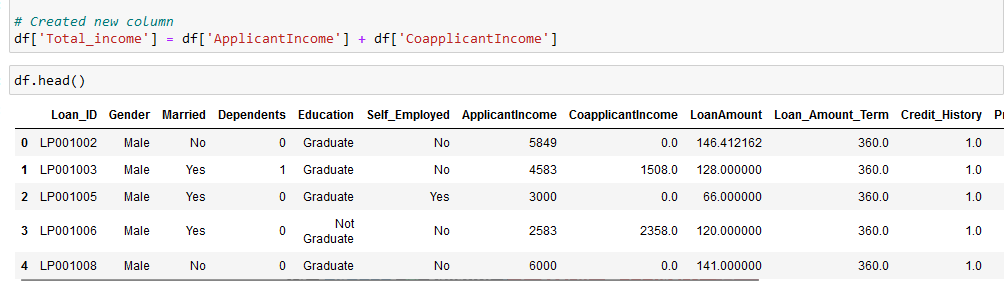
****

****

****

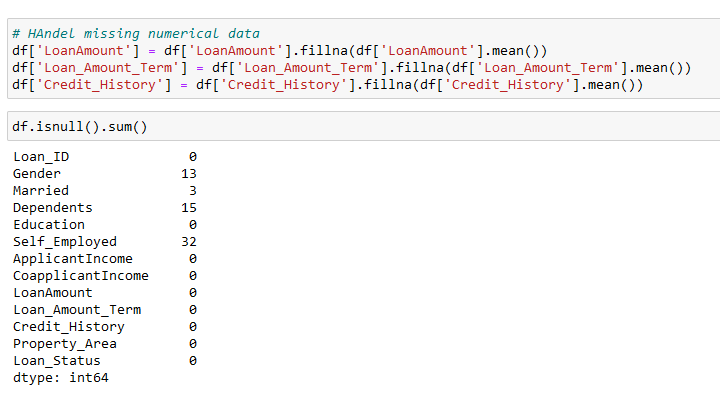
****

****

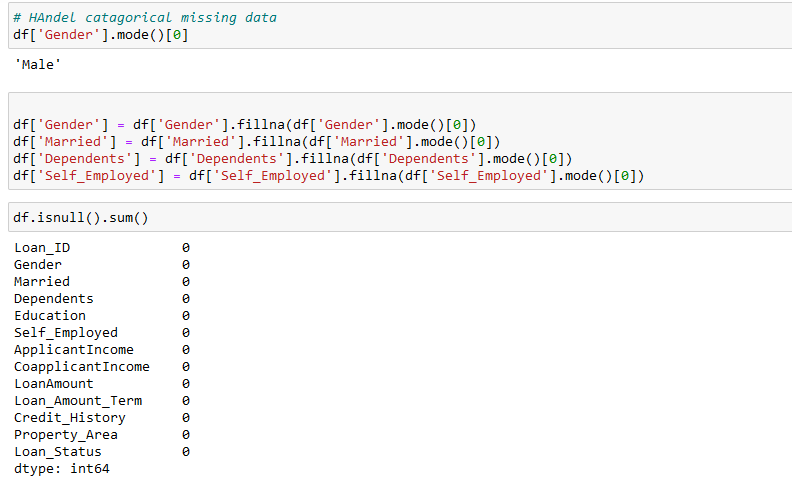
****

**3. Feature Engineering**

**3.1 Handling Missing numerical data**

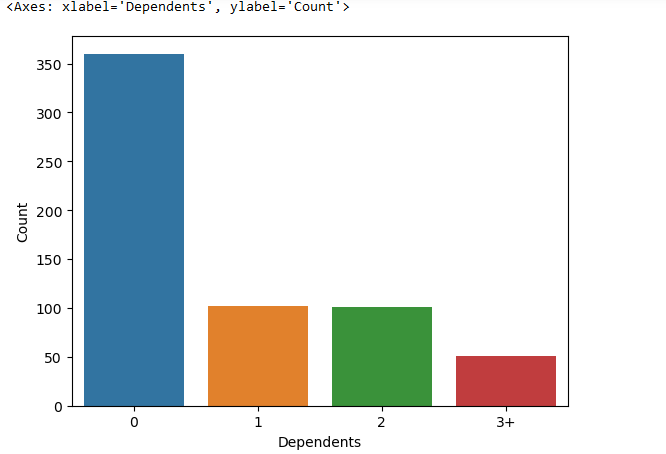
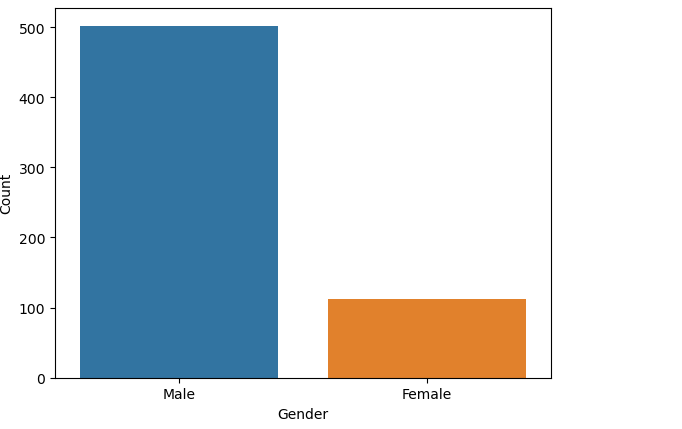
****

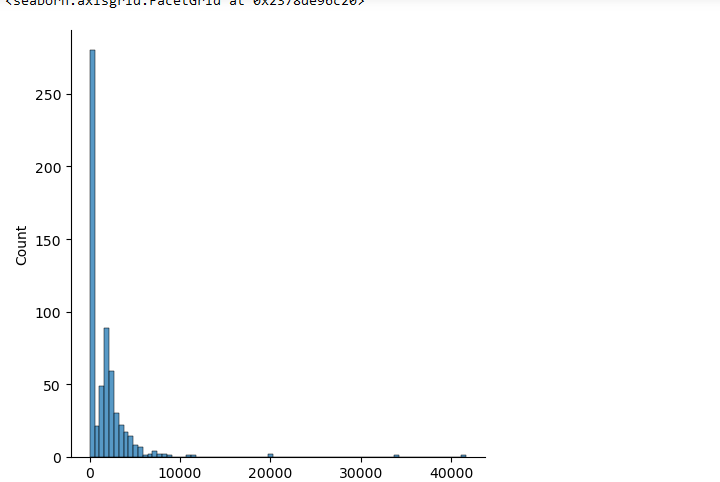
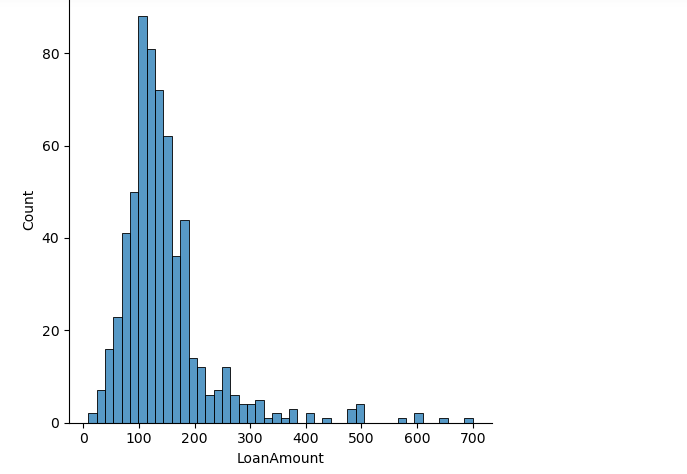
**3.2 Handling Categorical data**

****

**4. Data Visualization**

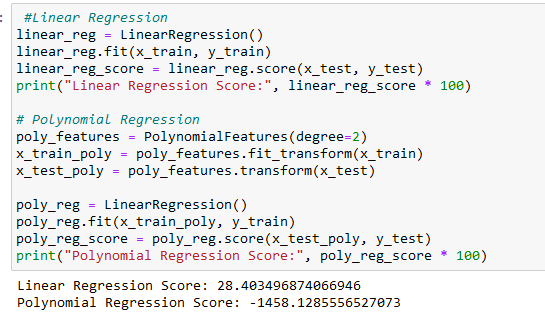
Various visualizations were created to explore the data and understand the relationships between different variables. Bar plots were used to visualize the distribution of categorical variables such as gender, dependents, and marital status. Displot plots were used to examine the distributions of applicant income, co-applicant income, loan amount, and credit history. These visualizations provided insights into the characteristics and trends present in the dataset.

****

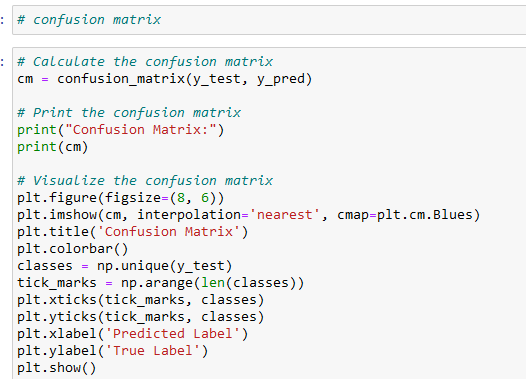
****

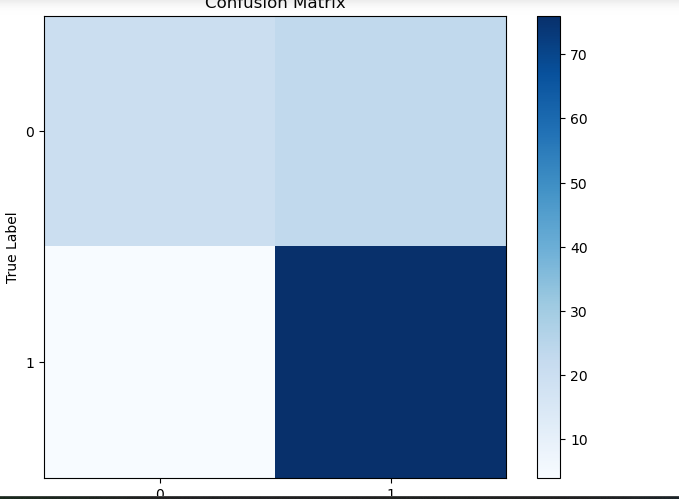
**5. Model Selection**

**5.1 Linear Regression**

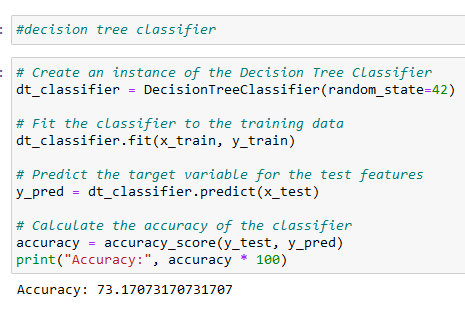
****

**5.2 Confusion matrix**

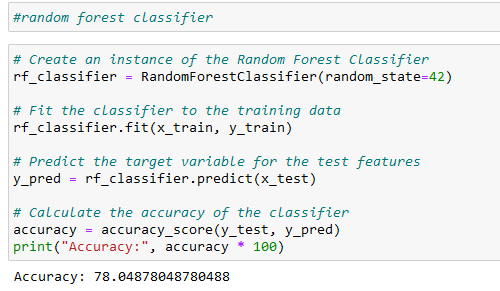
****

****

**5.3 Decision Tree Regressor**

****

**5.4 Random Forest Regressor**

****

**6. Conclusion**

**6.1 Summary of Findings**

The results of the loan amount prediction model indicate that different algorithms achieved varying levels of accuracy. The accuracy of each algorithm was as follows: - Random Forest Classifier: 78.04% - Decision Tree Classifier: 73.1% - Logistic Regression: 78.8% These accuracy scores suggest that both the Random Forest Classifier and Logistic Regression performed similarly well in predicting loan amounts, while the Decision Tree Classifier had slightly lower accuracy

**6.2 Future Work**

Further improvements and enhancements can be made to the loan amount prediction model. Additional features or external data sources could be incorporated to enhance the accuracy and predictive power of the model. Furthermore, a more extensive evaluation and comparison of different algorithms can be conducted to identify the best-performing model for loan amount prediction.

**6.3 Evaluation**

The evaluation of the loan amount prediction model was performed using various metrics, including accuracy, precision, recall, and F1 score. These metrics provide insights into the model's performance in correctly predicting loan amounts and identifying loan approvals or rejections. A comparison of the performance of different models was made based on these evaluation metrics.

**7. References**

1. YouTube : - <https://www.youtube.com/@codejay>

2. ChatGPT:- <https://chat.openai.com>

3. Stack Overflow :- <https://stackoverflow.com>

4. analysis Vidhya :- <https://www.analyticsvidhya.com>