Introduction

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.

Data Description

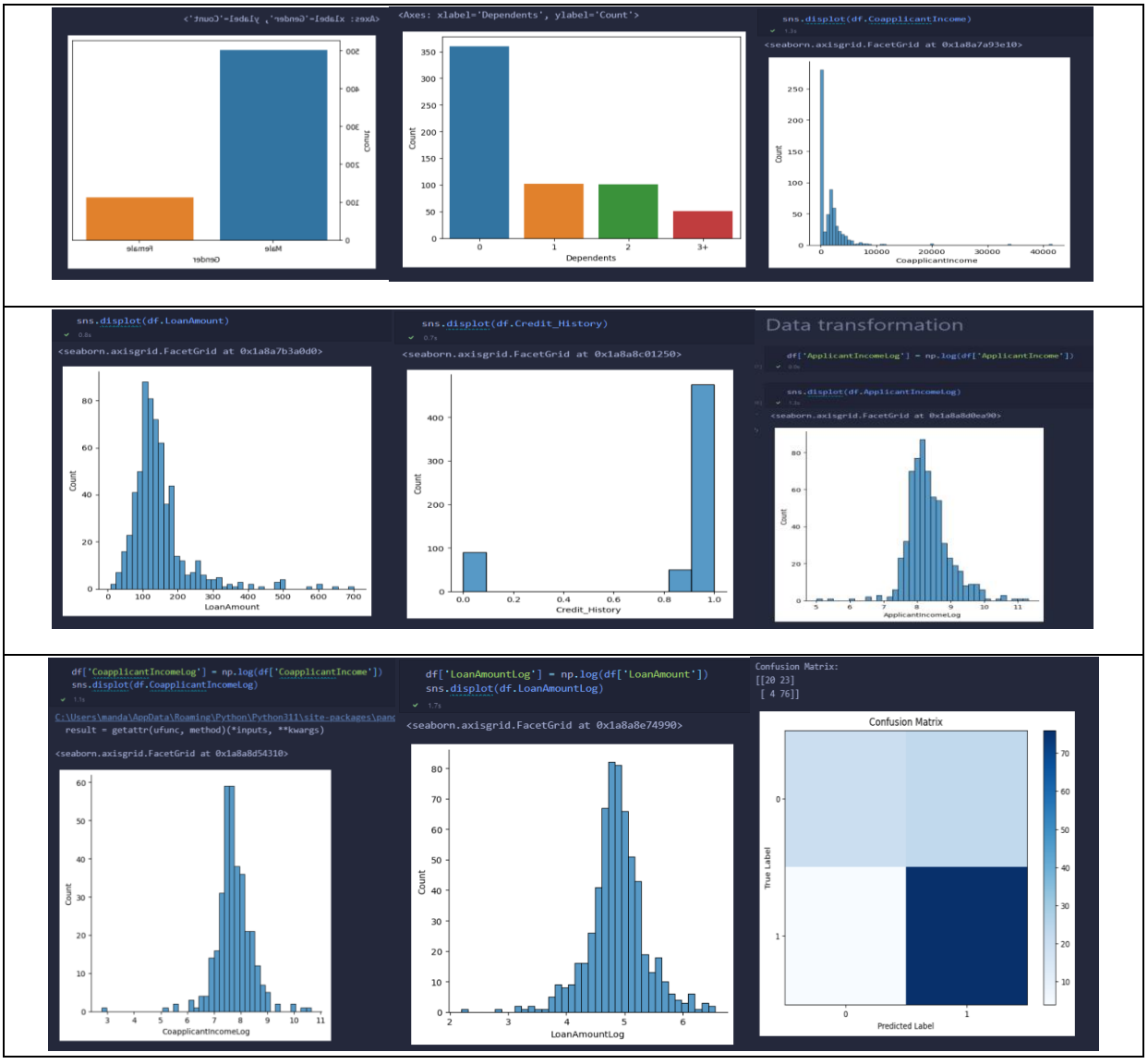
The dataset used for training and testing the loan amount prediction model consists of several columns that provide information about the applicants. Each column has a specific meaning and relevance to the loan amount prediction task. The data includes details such as applicant income, co-applicant income, 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.

Approach

The approach taken to develop the loan amount prediction model involves several steps. Firstly, the missing values in the dataset were handled by imputing the mean or mode values. Then, the data was preprocessed by performing log transformations on some attributes to improve their distribution. Categorical variables were encoded using the LabelEncoder. The dataset was split into training and testing sets. Different classification algorithms were applied, including RandomForestClassifier, DecisionTreeClassifier, and LogisticRegression. The performance of each model was evaluated using accuracy scores and confusion matrices.

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.



Algorithms

Multiple machine learning algorithms were utilized to build the loan amount prediction model. The following algorithms were applied and compared based on their accuracy and performance measures:

1. Random Forest Classifier: This ensemble learning algorithm combines multiple decision trees to make predictions.

RandomForestClassifier

```
# Create an instance of the Random Forest Classifier
rf_classifier = RandomForestClassifier(random_state=42)

# Fit the classifier to the training data
rf_classifier.fit(x_train, y_train)

# Predict the target variable for the test features
y_pred = rf_classifier.predict(x_test)

# Calculate the accuracy of the classifier
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy * 100)
```

1.3s

```
Accuracy: 78.04878048780488
```

2. Decision Tree Classifier: This algorithm constructs a tree-like model to make decisions based on feature values.

Decision Tree classification

```
# Create an instance of the Decision Tree Classifier
dt_classifier = DecisionTreeClassifier(random_state=42)

# Fit the classifier to the training data
dt_classifier.fit(x_train, y_train)

# Predict the target variable for the test features
y_pred = dt_classifier.predict(x_test)

# Calculate the accuracy of the classifier
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy * 100)
```

3. Logistic Regression: This algorithm models the relationship between the input features and the loan amount using logistic functions.

Logistic regression

```
# Create an instance of the Logistic Regression model
logreg = LogisticRegression(random_state=42)

# Fit the model to the training data
logreg.fit(x_train, y_train)

# Predict the target variable for the test features
y_pred = logreg.predict(x_test)

# Calculate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy * 100)
```

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.

Result and Discussion

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.

Conclusion

In conclusion, the loan amount prediction model developed in this project shows promise in accurately predicting loan amounts based on applicant information. The Random Forest Classifier exhibited the best performance among the evaluated algorithms. This model can assist financial institutions in making informed decisions regarding loan approvals and determining appropriate loan amounts for applicants.

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.

References

1. YouTube : - https://www.youtube.com/@codejay
2. ChatGPT:- https://chat.openai.com
3. Stack Overflow :- https://stackoverflow.com