Dots

Dilmurodov Javohir
Ergashev Suxrob
Bekturdieva Indira
Abduhalilova Oysha

Task introduction & process

Task: Data preprocessing and building a scoring model to identify the default of clients

Problem:

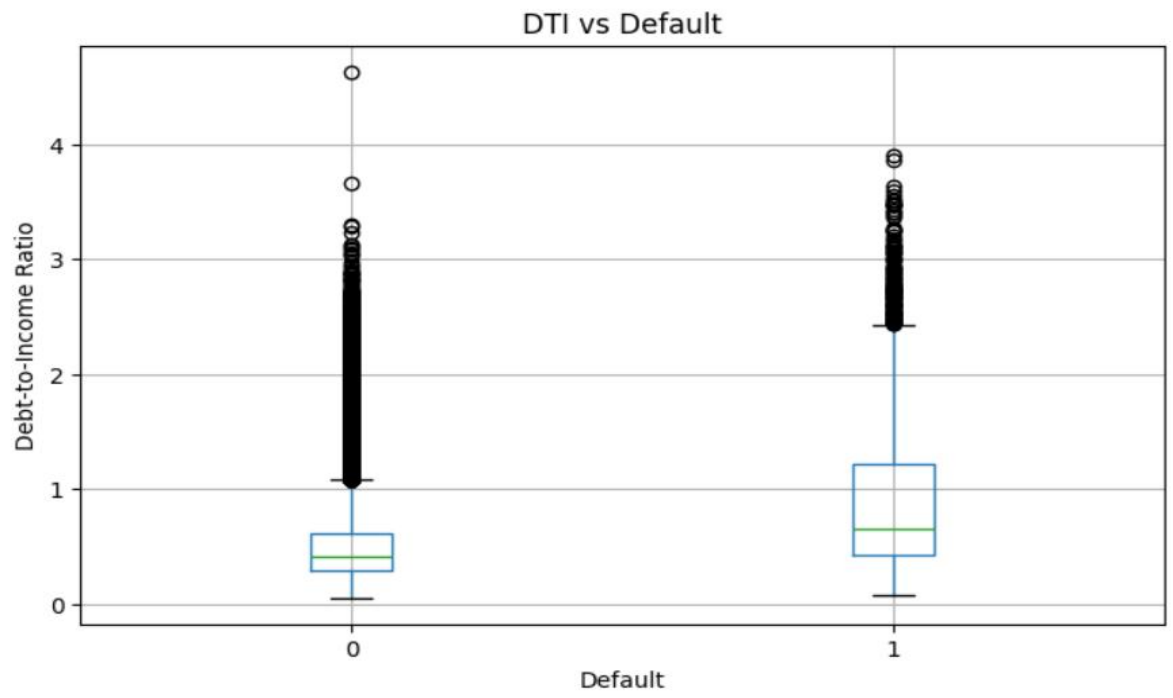
- Description of dataset
- Imbalance distribution of the target column(**default**)
- Selection of the best model

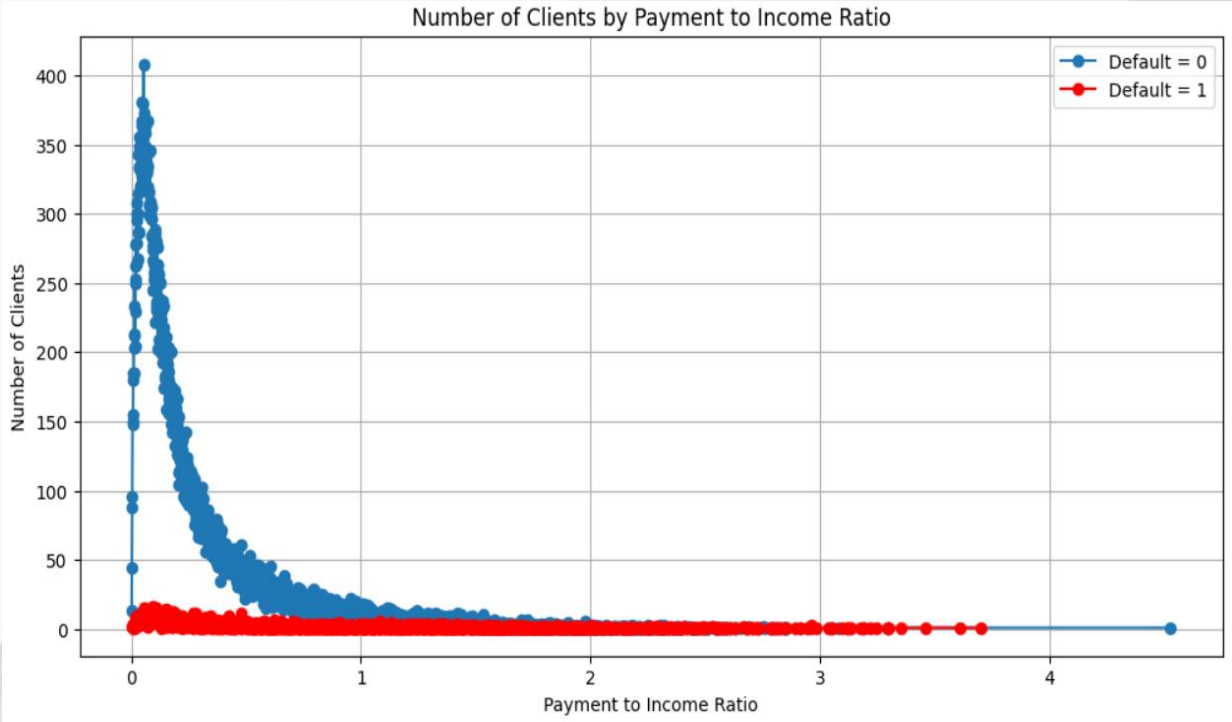
Process & Execution:

- Gathering data from different sources(JSON, CSV, XML, XLSX, PARQUET)
- Performing data cleaning and Exploratory Data Analysis
- Testing and model with tree-based(Xgboost, LGBM) and Linear models(Logistic regression)

Analysis and Visualizations

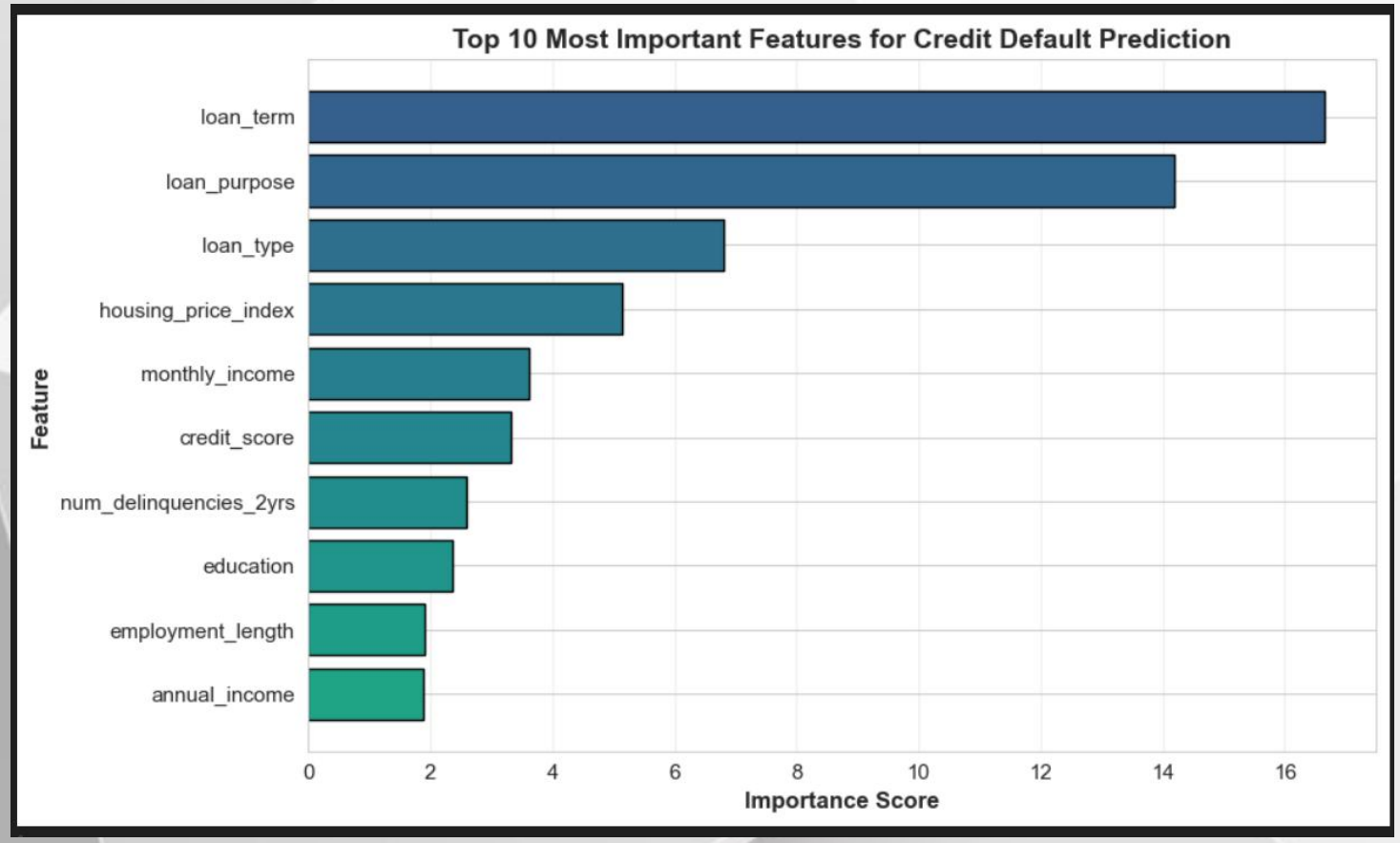
- As DTI Default Visual shows that The higher DTI rate is, the higher risk of getting Default

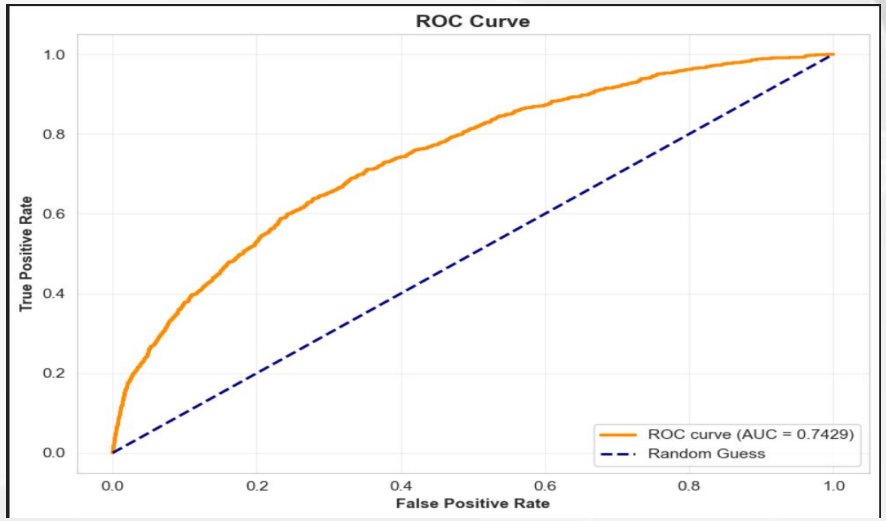




- This graph shows that as PTI rate increases, the probability of getting default is also increases, meaning Income and Payment depends on each other

Model Evaluation






```
FINAL RESULTS
==========================================================================
AUC:       0.7429
F1 Score:  0.2360
Recall:    0.2612
Precision: 0.2152
Accuracy:  0.9137
Threshold: 0.4500

Classification Report:
              precision    recall  f1-score   support

  No Default       0.96      0.95      0.95     17081
     Default       0.22      0.26      0.24       919

    accuracy                           0.91     18000
   macro avg       0.59      0.60      0.60     18000
weighted avg       0.92      0.91      0.92     18000
```

- Finally, we chose the model with best performance.
- AUC: 74,3%
- Accuracy: 91,3%

THANKS