CREDIT DEFAULT PREDICTION MODEL

A Machine Learning Approach to Risk Assessment

Presenter :Brian Githinji

Date : 8/29/2024

BUSINESS UNDERSTANDING:

- **Business Context** : Credit default prediction is crucial for managing risk in lending.Understanding which customers are likely to default allows the business to take proactive measures.

- **Objective**: To build a model that accurately predicts loan default risks.

- **Importance:** Helps financial institutions reduce potential losses by identifying high-risk borrowers.

DATA OVERVIEW:

The dataset used in this project is **sourced from KaggleDatasets** and contains various features related to the applicant's demographics, financial status, and loan details.

Key Features:

- Age, Income, Employment Length, Loan Amount, Interest Rate, Home Ownership, Loan Intent.

The **target variable** is a binary indicator representing whether the applicant defaulted on the loan.

Model Selection:

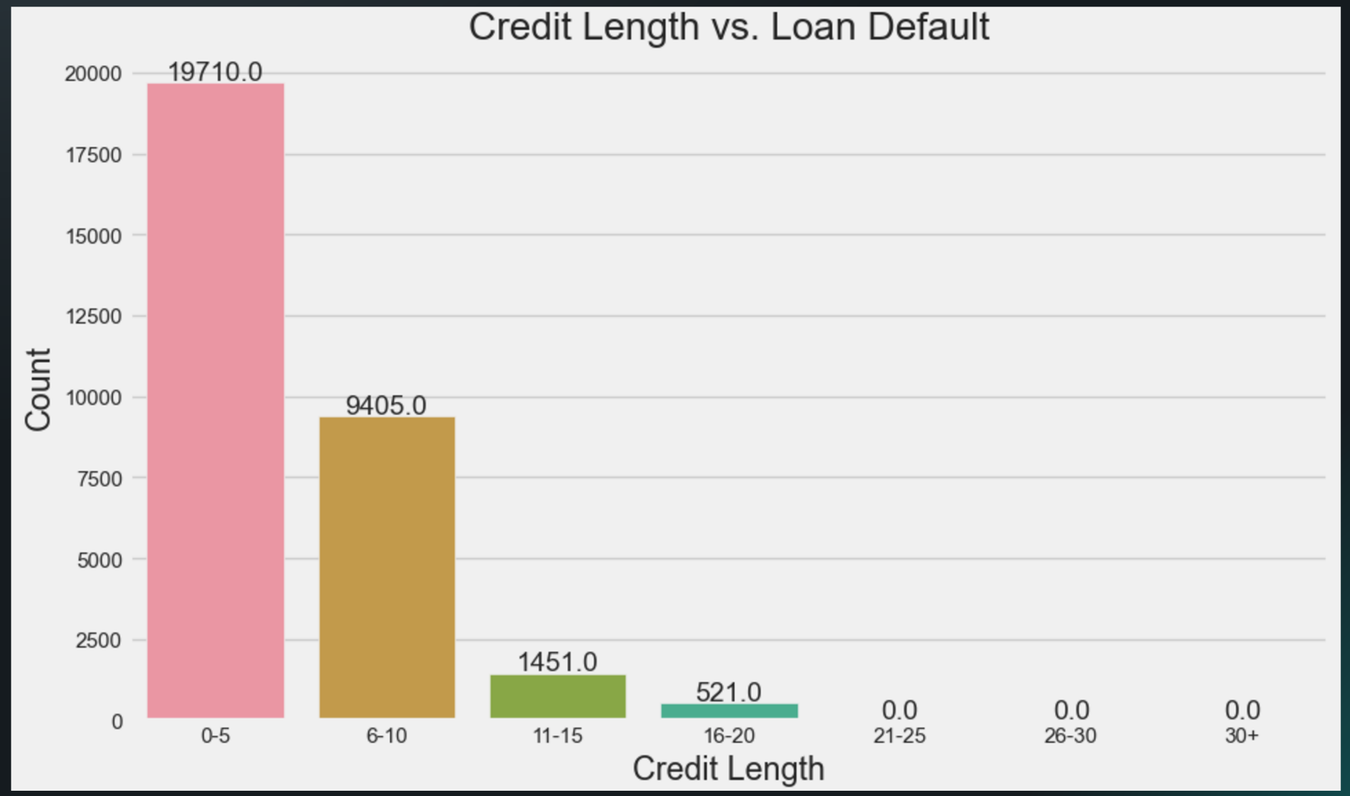
Models considered :

+1 Logistic Regression

+2 Decision Tree

+3 Random Forest

KEY FINDINGS :

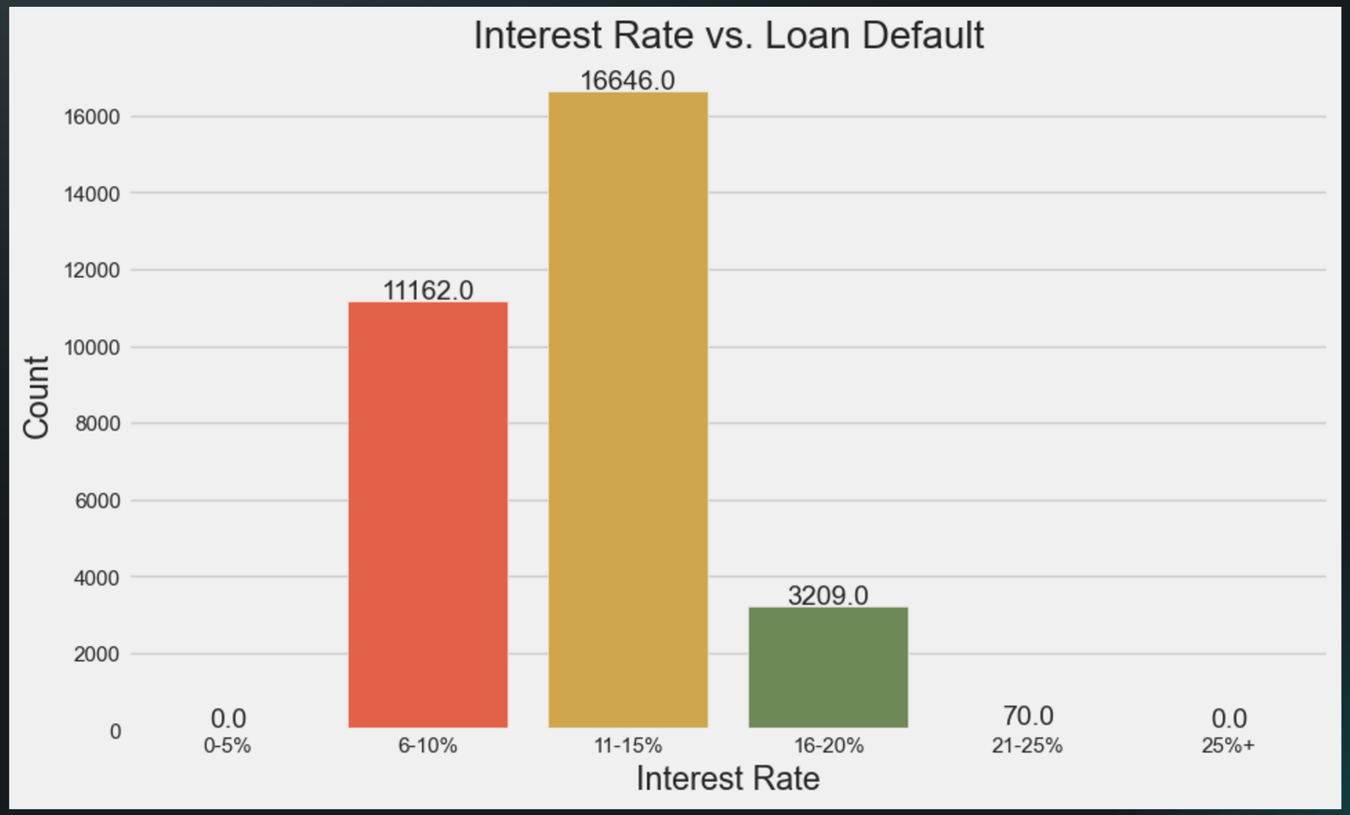


1. 63% of defaults are on short-term loans (0-5 years).

2. 52% of defaulters have 0-5 years of work experience.

3. 49% of defaulters are aged 18-25.

KEY FINDINGS :



4. 53% of defaults occur on loans with 11-15% interest rates.

5. 37% of defaults are on loans of 5k-10k

6. 20% of education loans default.

Model Perfomance & Evaluation :

+ 1 Logistic Regression

- Accuracy: ~55%
- Struggles with distinguishing defaulters, particularly in reducing false negatives.

+ 2 Decision Tree

- Accuracy: ~80%
- Better at capturing data interactions but prone to overfitting.

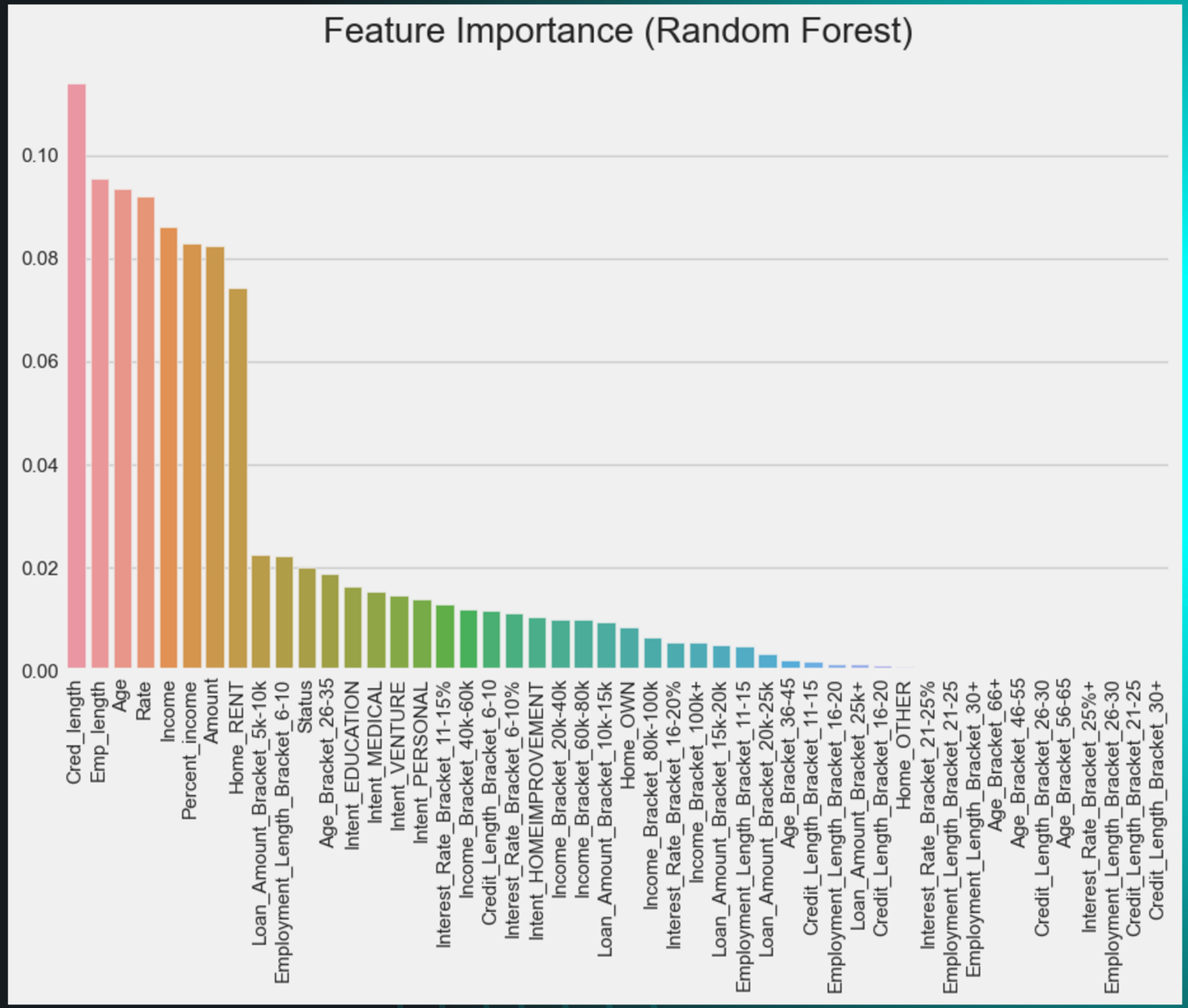
+ 3 Random Forest

- Accuracy: ~89%
- Best performance with a good balance between precision and recall.

Feature Importance :

Top 7 Features:

- Credit Length
- Employment Length
- Age
- Interest Rate
- Income
- Percent Income
- Loan Amount



MODEL SELECTION

After comparing the models, the Random Forest model was selected as the final model due to its superior performance.

Reason for Selection:

- High accuracy and effectiveness in reducing false negatives.
- Robust and interpretable, suitable for real-world application.

Conclusion

Summary:

- Successful development of a loan default prediction model.

- Random Forest chosen for its superior performance.

Impact:

- Helps financial institutions manage risk effectively.

Future Work:

- Potential enhancements with more data and advanced algorithms.

Thank You For Watching

LinkedIn profile .
LinkedIn

(https://github.com/BrianGithinji-BMG)
Github