

# LOANGUARD: ENHANCING LOAN DEFAULT PREDICTION AND CREDIT LIMIT OPTIMIZATION

DATA SCIENCE BOOT CAMP

MANDELA BRIGHT QUASHIE  
OMEIZA OLUMUYE  
REID HARRIS

DECEMBER 2, 2024



THE ERDŐS INSTITUTE  
Helping PhDs get jobs they love.  
Helping you hire the PhDs you need.

# PROBLEM STATEMENT

- Inaccurate loan default predictions
- Limited credit scoring indicators
- Need for precise risk assessments
- Demand for advanced ML models
- Operational inefficiencies
- Regulatory compliance challenges
- Scalability issues with large datasets

# DATA OVERVIEW

- Source: Lending Club dataset
- Key Features:
  - ▶ Financial Metrics: Loan amount, interest rate, installment, annual income
  - ▶ Credit History: FICO scores, delinquencies, revolving balance
  - ▶ Public Records: Bankruptcies, tax liens
- Data Size:
  - ▶ 1.13 million records
  - ▶ 180 features
- Preprocessing:
  - ▶ Missing value imputation
  - ▶ Feature scaling and encoding
  - ▶ Outlier removal
- Balanced dataset using undersampling for training

# MODELS AND ACCURACY

<b>Model</b>	<b>Accuracy (%)</b>	<b>ROC AUC (%)</b>
Logistic Regression	99.65	99.49
Naive Bayes	94.72	88.95
Decision Tree	99.57	99.58
Random Forest	99.85	99.70
Gradient Boosting	99.78	99.62
XGBoost	99.97	99.94

**Table:** Model Performance Comparison

# FEATURES USED IN THE MODEL FOR THE APPLICATION

## ■ **Financial Metrics:**

- ▶ Average Current Balance
- ▶ Total Collection Amount
- ▶ Total Installment High Credit

## ■ **Credit History:**

- ▶ Months Since Oldest Revolving Trade Line Open
- ▶ Number of Accounts Ever 120 Days Past Due
- ▶ Number of Trade Lines Open Past 12 Months

## ■ **Active Credit Lines:**

- ▶ Number of Active Revolving Trade Lines
- ▶ Number of Bankcard Trade Lines
- ▶ Number of Installment Trade Lines

## ■ **Public Records:**

- ▶ Number of Public Record Bankruptcies
- ▶ Number of Tax Liens

# SIMULATION RESULTS

The model achieves an impressive overall accuracy of 97.10%, indicating robust performance across the dataset.

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
<b>High</b>	0.80	0.78	0.79	1,585
<b>Low</b>	0.99	0.99	0.99	328,504
<b>Moderate</b>	0.91	0.90	0.91	10,721

**Table:** Classification report showing precision, recall, F1-score, and support for each class.

# LOANGUARD APPLICATION

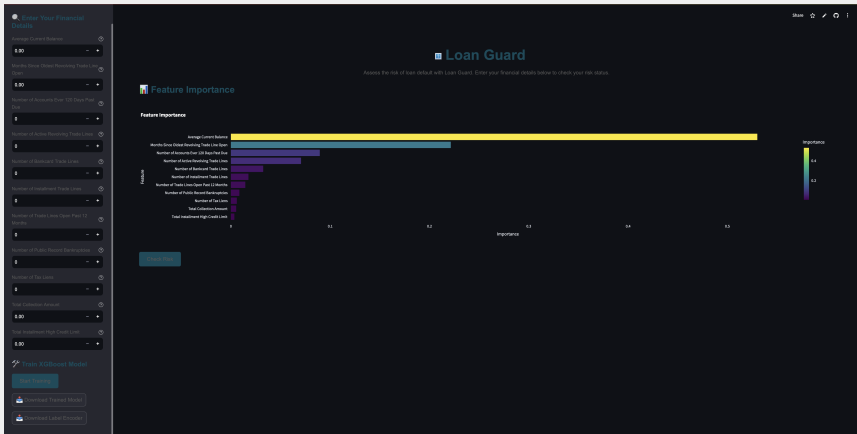


Figure: Full View

# LOANGUARD APPLICATION

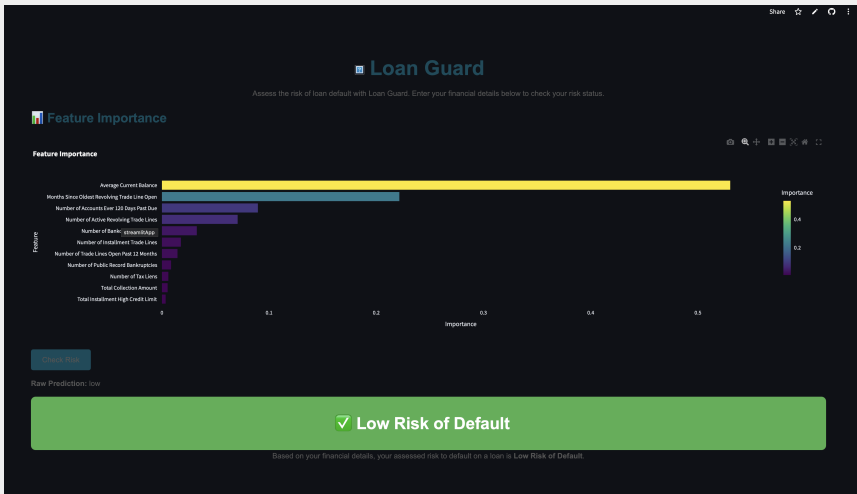


Figure: Low Risk



# LOANGUARD APPLICATION



Figure: Moderate Risk

# LOANGUARD APPLICATION



Figure: High Risk