



Predicting Loan Default with Logistic Regression: Empowering Lenders with Risk Assessment

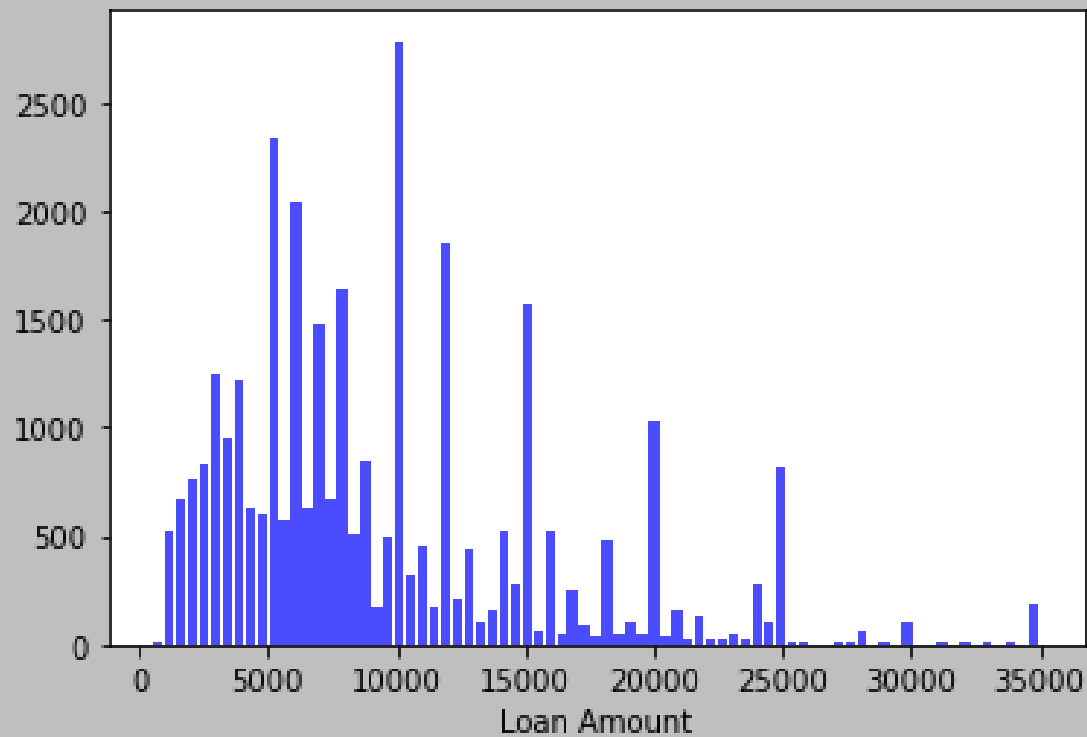
CHRISTOPHER A. N. EREFOROKUMA

INTRODUCTION

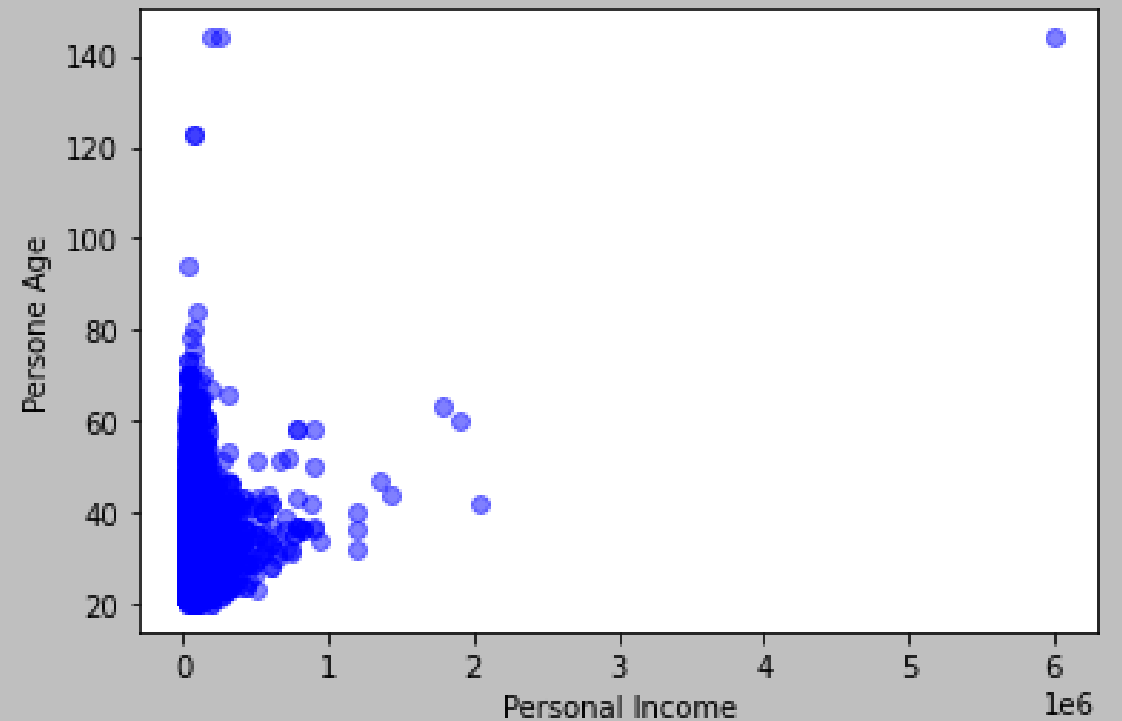
This is a logistic regression project for predicting loan default. It includes data exploration, data preparation, logistic regression modeling, model evaluation, and a prototype application for loan default prediction.

EXPLORATORY DATA ANALYSIS

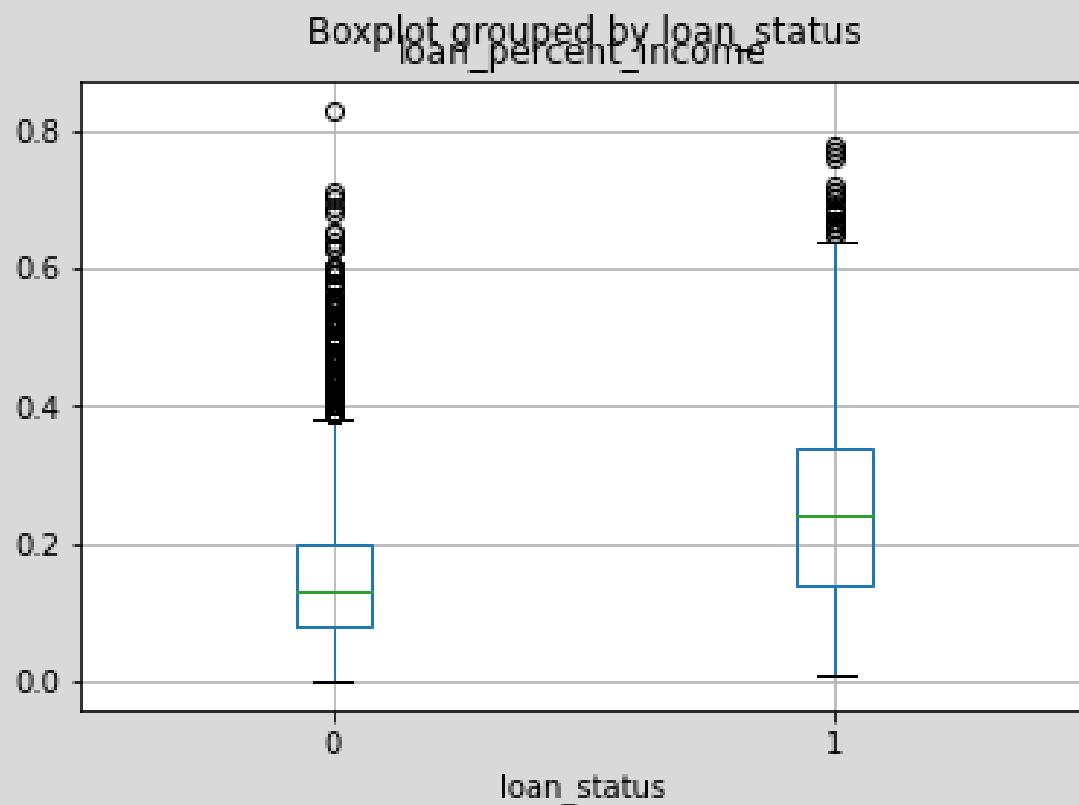
The project starts with exploring the dataset and performing exploratory data analysis (EDA). It includes histograms, scatter plots, cross tables, and box plots to understand the loan amounts, income-age relationship, loan intent, home ownership, and more.



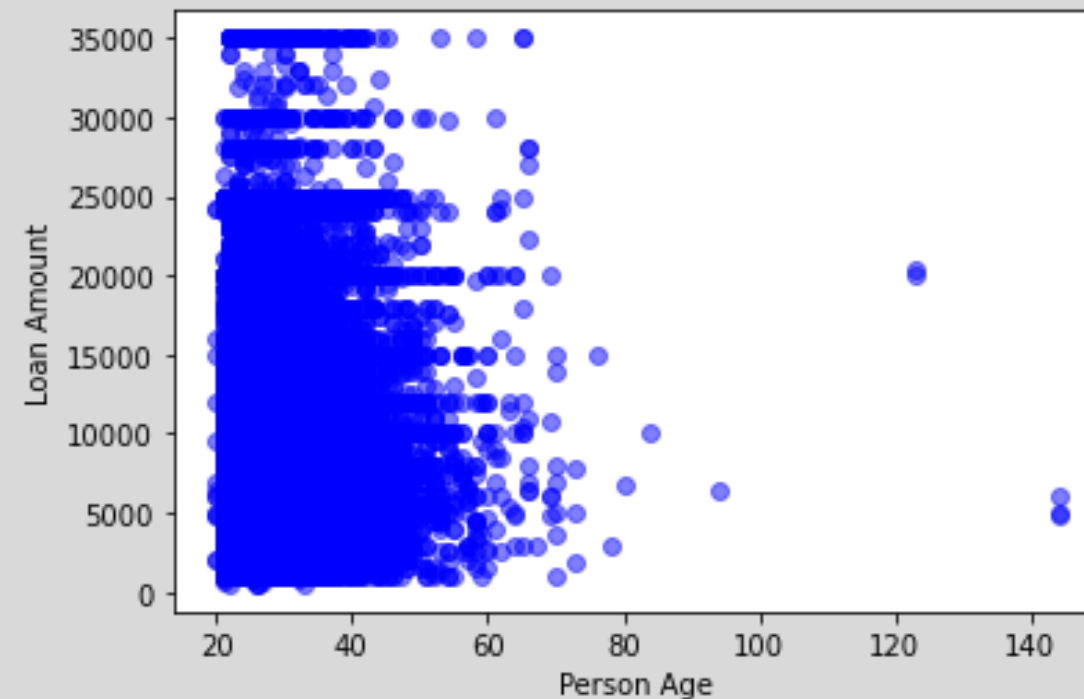
Plot of the distribution of loan amounts



Plot of Income vs. Age



Plot of Percentage Income by Loan Status



Plot for Age vs Loan Amount



DATA PREPARATION

Data preparation involves handling missing values and encoding categorical variables. Missing values are filled using appropriate strategies, and categorical variables are encoded using one-hot encoding.

Handling Missing Values

```
cr_loan['loan_int_rate'].fillna(cr_loan['loan_int_rate'].mean(), inplace=True)
cr_loan.isnull().sum()
```

```
indices = cr_loan[cr_loan['person_emp_length'].isnull()].index
cr_loan.drop(indices, inplace=True)
cr_loan.isna().sum()
```

Encoding the ordinal categorical variables person_home_ownership'

```
cut_mapping = {'RENT': 0, 'MORTGAGE': 1, 'OWN': 2, 'OTHER': 3}
cr_loan.person_home_ownership = cr_loan.person_home_ownership.map(cut_mapping)
```

Encoding the ordinal categorical variable 'loan_intent'

```
cut_mapping = {'EDUCATION': 0, 'MEDICAL': 1, 'VENTURE': 2, 'PERSONAL': 3,
'DEBTCONSOLIDATION': 4, 'HOMEIMPROVEMENT': 5}
cr_loan.loan_intent = cr_loan.loan_intent.map(cut_mapping)
```


TRAIN-TEST SPLIT

To evaluate the model, the dataset is split into training and testing sets using the `train_test_split` function. Around 70% of the data is used for training, and 30% is used for testing.

```
# Train-Test Split
X = cr_loan.drop('loan_status', axis=1)
y = cr_loan[['loan_status']]
```

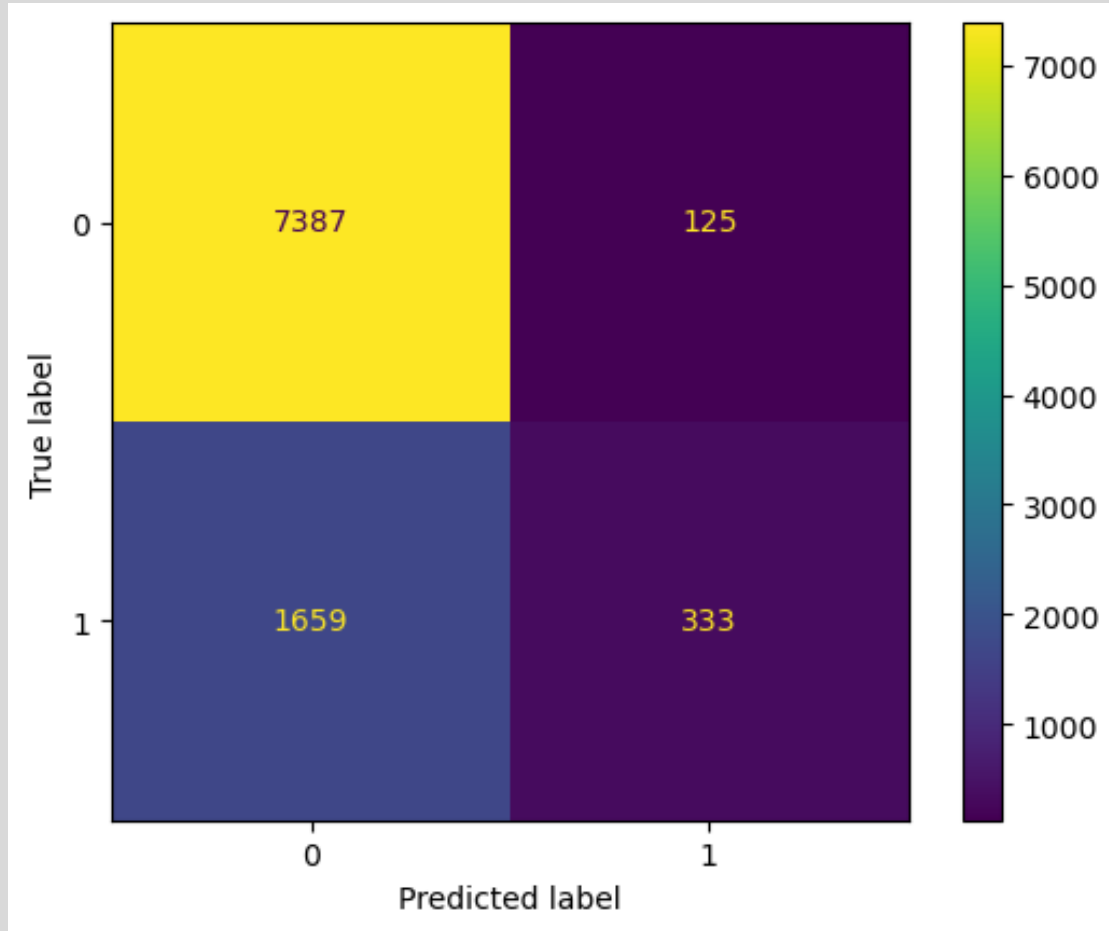

MODEL BUILDING

The logistic regression model is trained using the Logistic Regression class from scikit-learn. It learns the patterns and relationships between the input features and the loan default status.

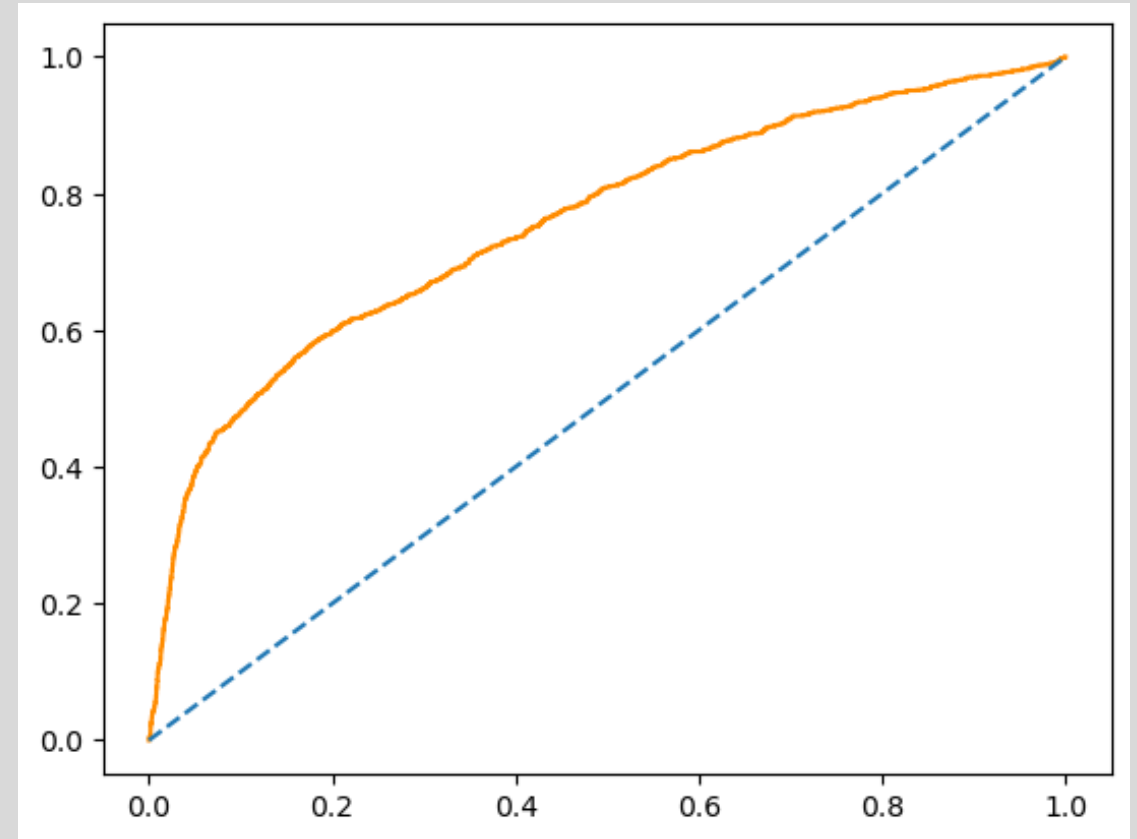
```
# Logistic Regression Modeling
clf_logit = LogisticRegression(solver='lbfgs')
clf_logit.fit(X_train, np.ravel(y_train))
```

MODEL EVALUATION

The model is evaluated using various metrics and visualizations. These include classification reports, ROC curves, AUC scores, and confusion matrices to assess the model's performance in predicting loan default.



Confusion Matrix



ROC Curve

STREAMLIT PROTOTYPE

LOAN DEFAULT PREDICTOR

Enter the characteristics of the applicant:

| | |
|--|--|
| Age of Applicant: | Loan Grade: |
| <input type="text" value="25.00"/> - + | <input type="text" value="A"/> ▾ |
| Income of Applicant: | Loan Amount: |
| <input type="text" value="9500.00"/> - + | <input type="text" value="5000.00"/> - + |
| Type of Residence: | Interest Rate: |
| <input type="text" value="OWN"/> ▾ | <input type="text" value="14.00"/> - + |
| Years of Employment: | Percentage of Loan to Income |
| <input type="text" value="3.00"/> - + | <input type="text" value="0.51"/> - + |
| Loan Intent: | Previous History of Default: |
| <input type="text" value="MEDICAL"/> ▾ | <input type="text" value="N"/> ▾ |
| | Length of Credit History (In Years) |
| | <input type="text" value="2.00"/> - + |

Application Result: YES

A prototype application is developed to showcase the loan default prediction. It allows users to enter applicant characteristics and provides a prediction of whether the loan is likely to default or not.

The code for this streamlit Prototype web application can be accessed on the GitHub Repo ([link](#))

CONCLUSION

At the heart of this project, we aimed to understand and predict loan default risk. Through extensive data exploration, meticulous data preparation, and model evaluation, we uncovered valuable insights into the factors that contribute to loan default. Our user-friendly Streamlit app empowers lenders to make informed decisions and proactively manage risks, ensuring a fair and secure lending process for both applicants and financial institutions. Together, we can foster a stronger financial ecosystem by leveraging data-driven approaches to mitigate loan default risks and promote responsible lending practices.

THANK YOU

Christopher A. N. Ereforokuma

cereforokuma@gmail.com

[GitHub Repo \(link\)](#)

[LinkedIn Page \(link\)](#)

[Article on Medium \(link\)](#)