Loan Decision Modeling

By Marcos Dominguez

Project Idea

- Predicting credit default risk of loan applicant
  - Features: 771 columns of anonamized, numerical data
  - Target: 0-100, default amount
    - 0 → no to very low loan loss: approve loan
    - >0 → some to very high loan loss: decline loan

Workflow of data collection and modeling

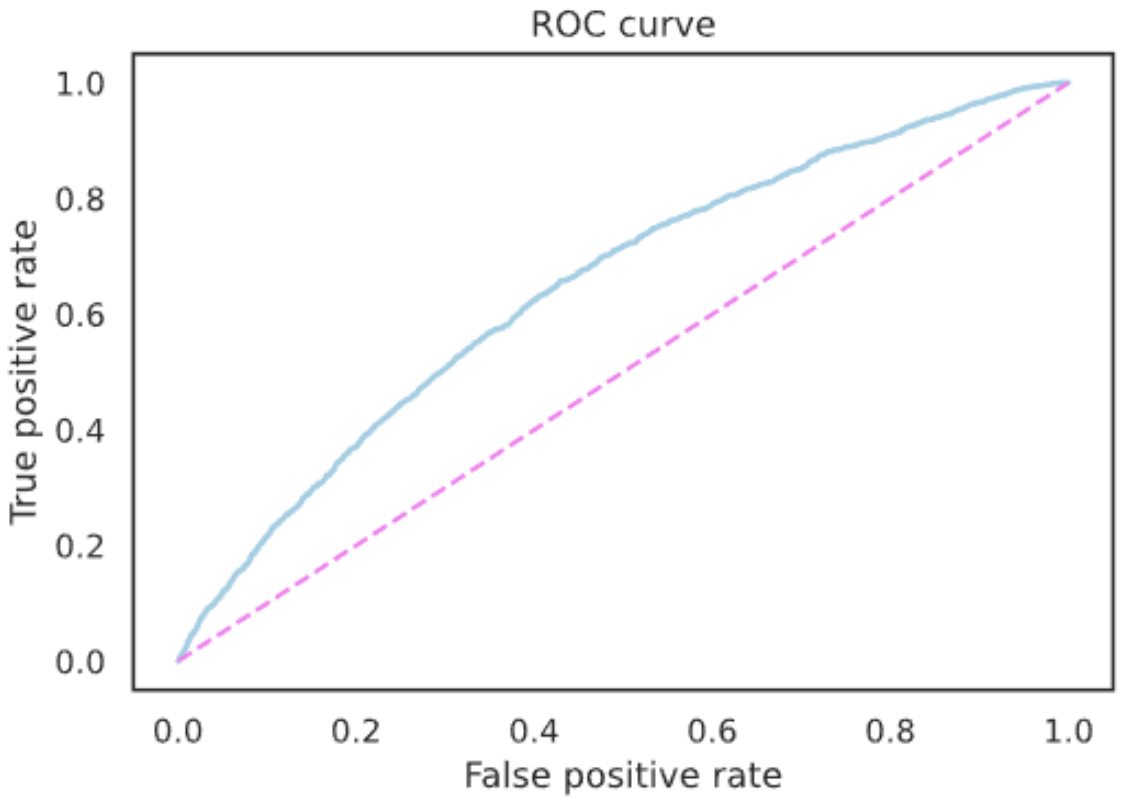
- Data size: 100k x 771 dataframe
- Kaggle competition: Loan Default Prediction
  - Imperial College of London
  - Kaggle Dataset
- Classification
  - Predict binary outcome to approve or decline
- Linear Regression
  - Predict loan loss amount to optimize portfolio
  - Minimize RMSE

Initial Results and Takeaways

- Best classification model:
  - XGBoost
  - Currently overfit
- Best linear Model:
  - R-squared of 0
  - Horrible at explaining variance
  - Working on reducing features through p-value and coefficient

Visualization

```
ROC AUC score =  0.6479420373869114
```



Next Steps

- Improve R-squared of multiple regression model
- Create streamlit app containing process and results