

Probability of Loan Default

Alicja Ligas



Lending Club



- Lending Club is an online P2P lending platform
- Mainly for Credit Borrowers and Debt Consolidation
- Data available online for free

Data Description

01

Loan Status

Current, Fully Paid, In Grace Period, Late, Charged Off, Default

Loan Properties

Amount, Length, Interest Rate, Month Issued, Description provided by the borrowers

02

03

Borrower Data

Income, Home Ownership Status, Occupation, Employment Length, etc.

Credit Risk Metrics

Metrics used to measure the risk of the loan, such as, fico score, bank card utilization, current balances on accounts, etc.

04

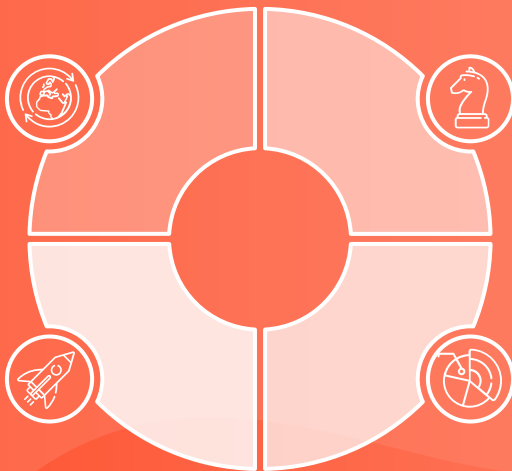
Cleaning

Size

~ 100,000 rows and
150 columns of data
per quarter

Columns

- 38 object columns
- Dummy, binarize, or drop

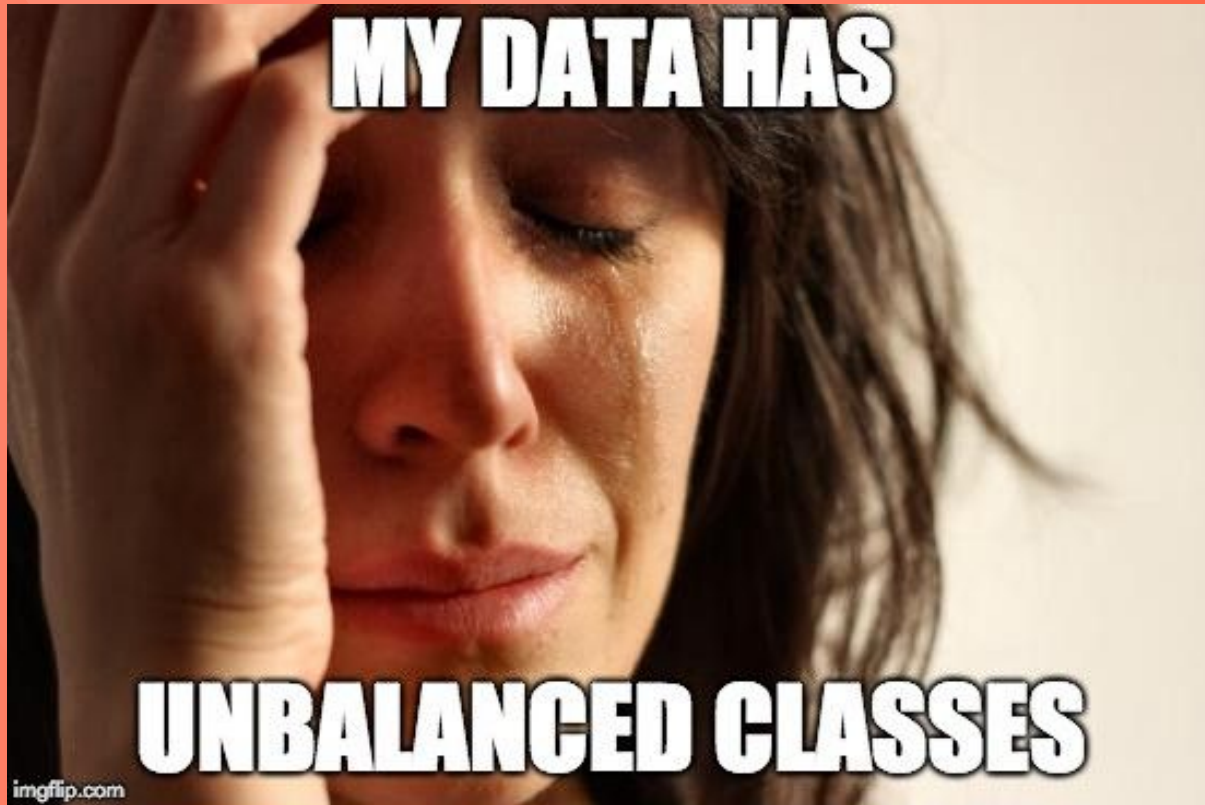


Null Values

- Strategic Imputation
- Some quarters entire columns were null values

Concatting

- Categorical columns had inconsistent values quarter to quarter



Less than 1% of loans defaulted each quarter...

Strategies

Under Sampling
of Majority Class

01

02

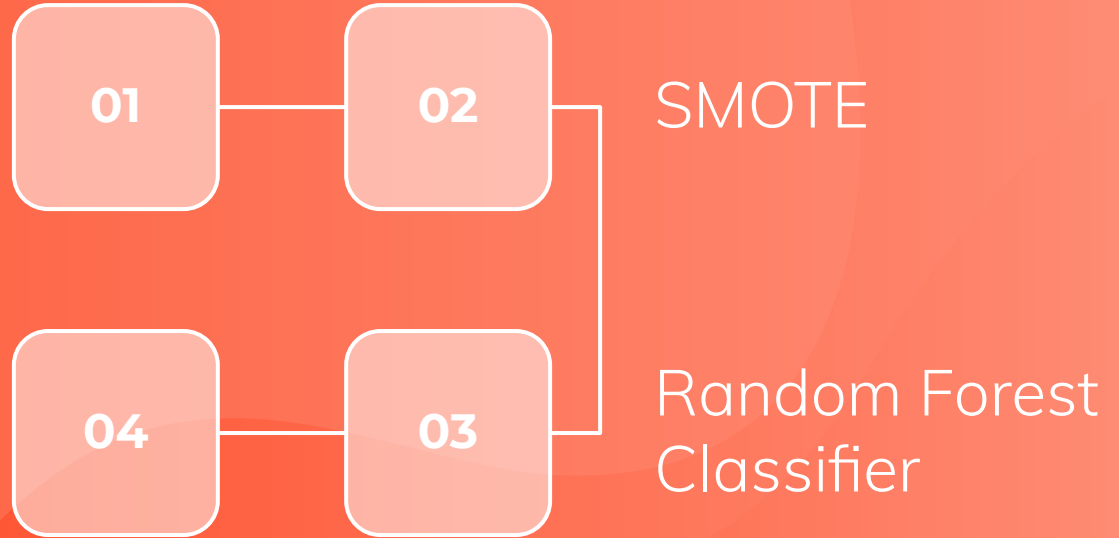
SMOTE

AdaBoost
Classifier

04

03

Random Forest
Classifier



Going Forward

- Classification with Neural Networks
- Growing database
- Flask web application





Thanks!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.