Loan Data Analysis

By Nur Inna Alfianinda

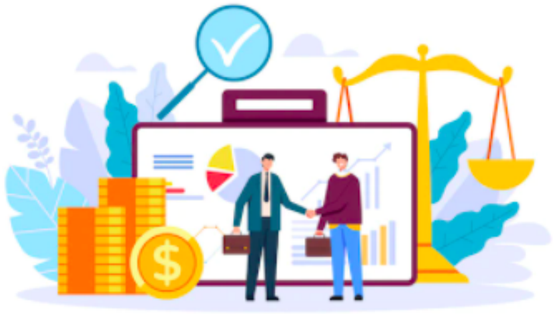
IM Finance

IM Finance

An financial technology company that is designed to protect cash flow and help entrepreneurs grow their businesses over the long-term.

We also lend money to purchase commercial real estate, buy new or used equipment, and a variety of other business needs.
We work in partnership with other financial institutions to find the best solution for each business.

Business Process



Customer Management

Credit Analysis

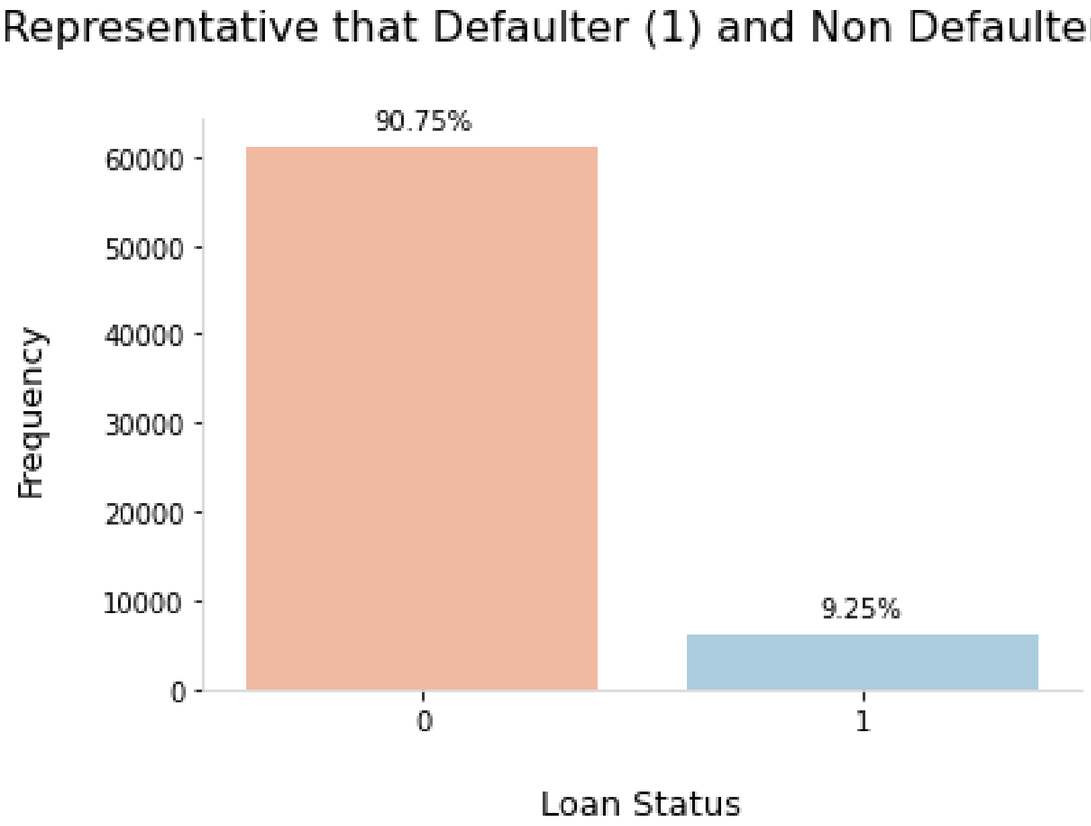
Credit Presentation

Decisioning and Approvals

Covenants Monitoring

Portfolio Risk Management

Problem Identification



Objective

Create a system to help predict and calculate if a person will be a loan defaulter or not automatically

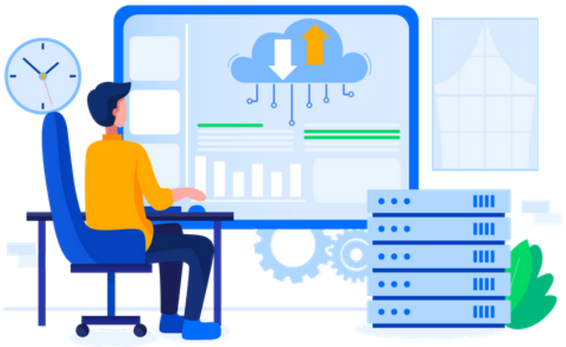
Business Metrics

- Minimise losses of the lenders
- Reducing the risk by predicting customers who are likely to default
- Increasing the profits by improving loan accessbility the cutomers

Dataset Overview

Dataset

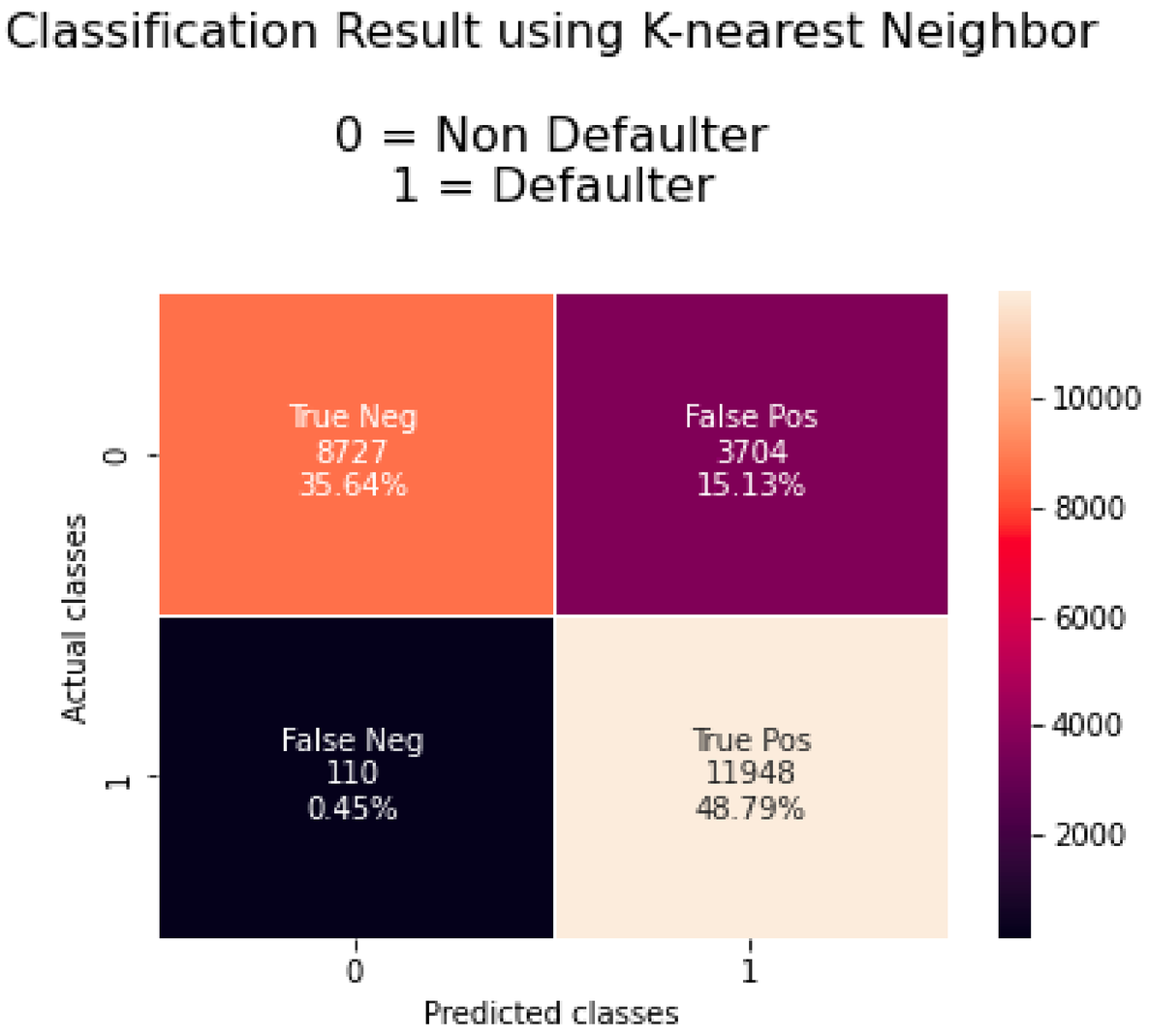
Historical loan data of customers and their loan status



Features

- ID
- Loan Amount
- Funded Amount
- Funded Amount Investor
- Term
- Batch Enrolled
- Interest Rate
- Grade
- Sub Grade
- Employment Duration
- Home Ownership
- Verification Status
- Payment Plan
- Loan Title
- Debit to Income
- Delinquency - two years
- Inquires - six months
- Open Account
- Public Record
- Revolving Balance
- Revolving Utilities
- Total Accounts
- Initial List Status
- Total Received Interest
- Total Received Late Fee
- Recoveries
- Collection Recovery Fee
- Collection 12 months Medical
- Application Type
- Last week Pay
- Accounts Delinquent
- Total Collection Amount
- Total Current Balance
- Total Revolving Credit Limit
- Loan Status

Model Evaluation Result



- We will use Accuracy metrics to evaluate the model performance because each label has the same importance. The goal is to produce a more accurate way to predict loan defaulters
- The best prediction performance score using accuracy metric is **84%**
- Using feature importance graph, the feature that influences the most is the '*Home Ownership*'

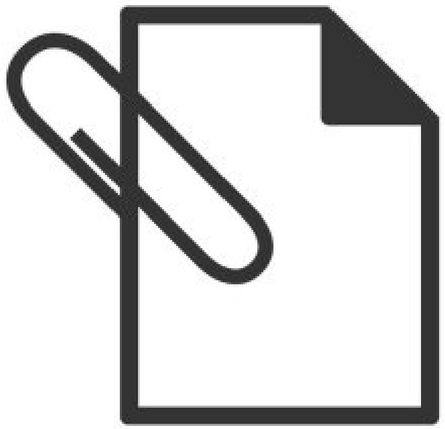
Conclusions

- The algorithm that has the best performance (84%) in evaluating the results of this prediction is K-nearest Neighbor
- Applicants who have been predicted as non defaulter should have higher chances of loan approval
- From of all the algorithms carried out, the feature that influences the most is the 'Home Ownership'

- Home ownership could signal that a person is responsible and capable of handling loans
- Another view is that home owners could use their house as collateral when asking for loan so that they are charged with the lower interest rate.

Business Recommendation

Create customer segments to facilitate the evaluation of borrowers and loan amounts so that it can specifically target these customers



APPENDIX

Model Evaluation

	Accuracy	Precision	Recall	F1 Score	AUC Score
Logistic Regression	0.54	0.53	0.54	0.53	0.54
Logistic Regression GridSearch	0.54	0.53	0.56	0.54	0.54
K-nearest Neighbor	0.84	0.76	0.99	0.86	0.85
Decision Tree	0.94	0.89	1.00	0.94	0.94
Bagging: Random Forest	1.00	1.00	1.00	1.00	1.00
Boosting: AdaBoost	0.55	0.54	0.55	0.54	0.55
Boosting: XGBoost	0.60	0.58	0.63	0.61	0.60
VotingClassifier	0.95	0.91	1.00	0.95	0.95
Neural Network	0.81	0.80	0.81	0.81	0.81


thank
you