BANQUE MISR

CREDIT RISK PREDICTION

JOSEPH OLADOKUN

BANQUE MISR

CREDIT OVERVIEW

47%

Overall Default rate

~46k

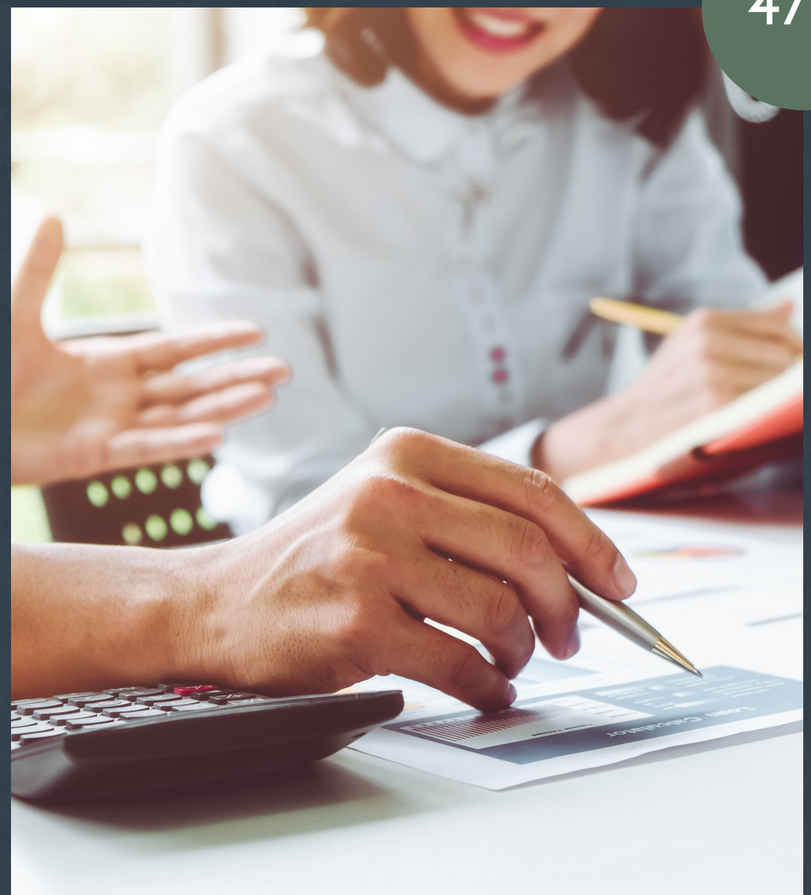
Total Clients

187k

Avg Income of loan applicant

BANQUE MISR

PROBLEM



47%

High Default Rate

The high default rate is a problem that needs to be reduced

Determine who to give credit to

Data-driven system to determine who to give loan to

BANK MISR

ANALYSIS APPROACH

Weight of Evidence and Information value

I calculated the information value and weight of evidence on each of the parameters

Machine Learning Appraoch

I applied three (3) machine learning algorithms to the data and calculated the feature importance on the algorithm with the highest accuracy to get the most important parameters to look out for when disbursing loans

BANK MISR

WEIGHT OF EVIDENCE AND INFORMATION VALUE

IV	Ability to predict
<0.02	Almost no predictive power
0.02~0.1	weak predictive power
0.1~0.3	Moderate predictive power
0.3~0.5	Strong predictive power
>0.5	Predictive power is too strong, need to check variable

WoE & IV

IV is based on an analysis of each independent variable without considering other predictor variables. WOE - Closely related to the IV value, WOE measures each grouped attribute's strength in predicting the Dependent Variable's desired value.

This table shows the value and the predictive power

BANQUE MISR

WEIGHT OF EVIDENCE AND INFORMATION VALUE

	variable	IV
6	income_category	0.052583
8	marital_status	0.051942
16	occupation	0.046588
7	education	0.040091

0,052

Highest IV

The income category has the highest IV, based on the table on the previous page, this table shows that these parameters have low predictive value.

This result shows that WoE and IV alone is not enough to arrive at a conclusion, i will use machine learning to get the accurate result

BANQUE MISR

MACHINE LEARNING APPROACH

LOGISTIC REGRESSION

60.5%

Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.

DECSION TREE

92.8%

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

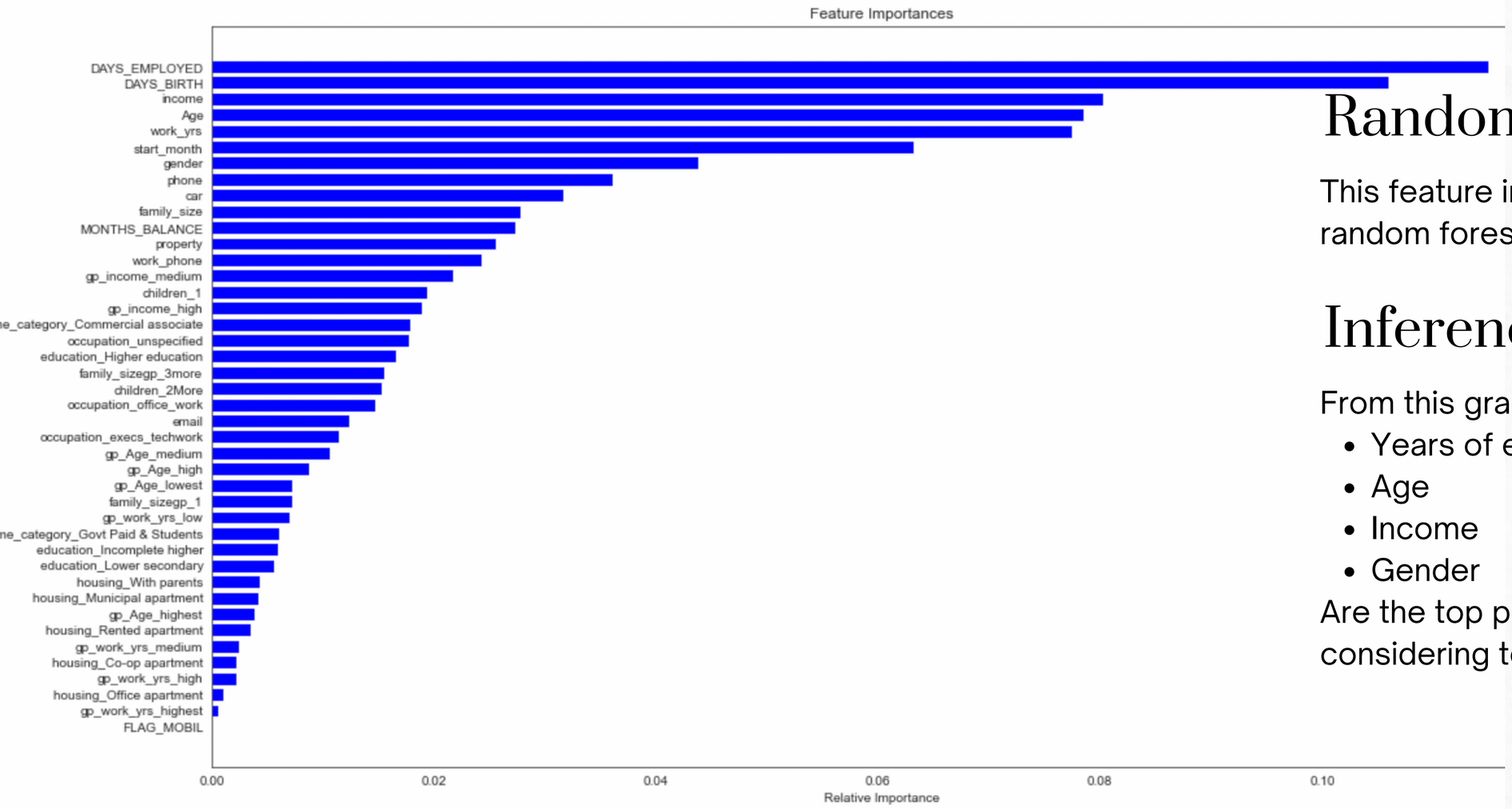
RANDOM FOREST

97.8%

The random forest has the highest accuracy. Random forests are an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

BANQUE MISR

FEATURE IMPORTANCE



Random Forest

This feature importance was created on the random forest model

Inference

From this graph, we can observe that

- Years of employment,
- Age
- Income
- Gender

Are the top parameters to consider when considering to give loan

BANK MISR

CONCLUSION & PROJECTION

30%

reduction in default rate

this model is projected to reduce default rate by 30% in the next 6 months, thereby saving the company million of dollars of potential bad debt

75%

Reduction in time required make loan decision

This model will reduce the time spent in making loan decision drastically

BANQUE MISR

RECOMMENDATION

USE WITH HUMAN

The best way to use this model is to combine it with human intelligence at the early stages while we work on fixing the other edge cases

RETRAIN AND SET THRESHOLDS

The model is still at the infant stage, retraining is required and thresholds for model accuracy should be set, once the model accuracy drops to a certain threshold, it should alert the machine learning engineer

JOSEPH SAYS

THANK YOU

FOR YOUR ATTENTION