Loan Approval Prediction

By Theo Jeremiah

Lending Business Model

Revenue = (No of Loans - No of NPL) X Loan amount X Interest

How to increase revenue?

↑ Number of Loans

↓ Number of Non Performing Loans

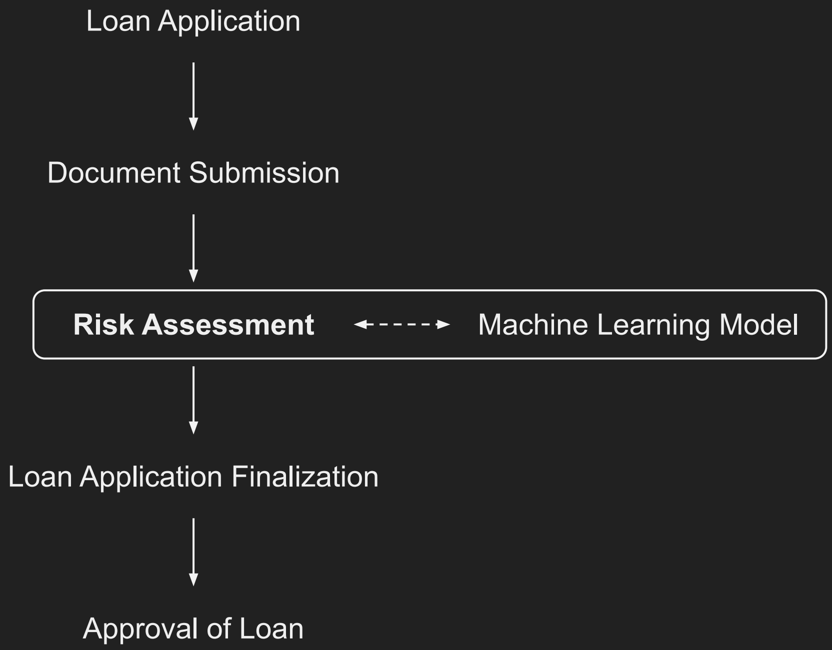
↑ Loan Amount

↑ Interest

>

High Quality Loaners!

What is a Loan Approval Prediction?

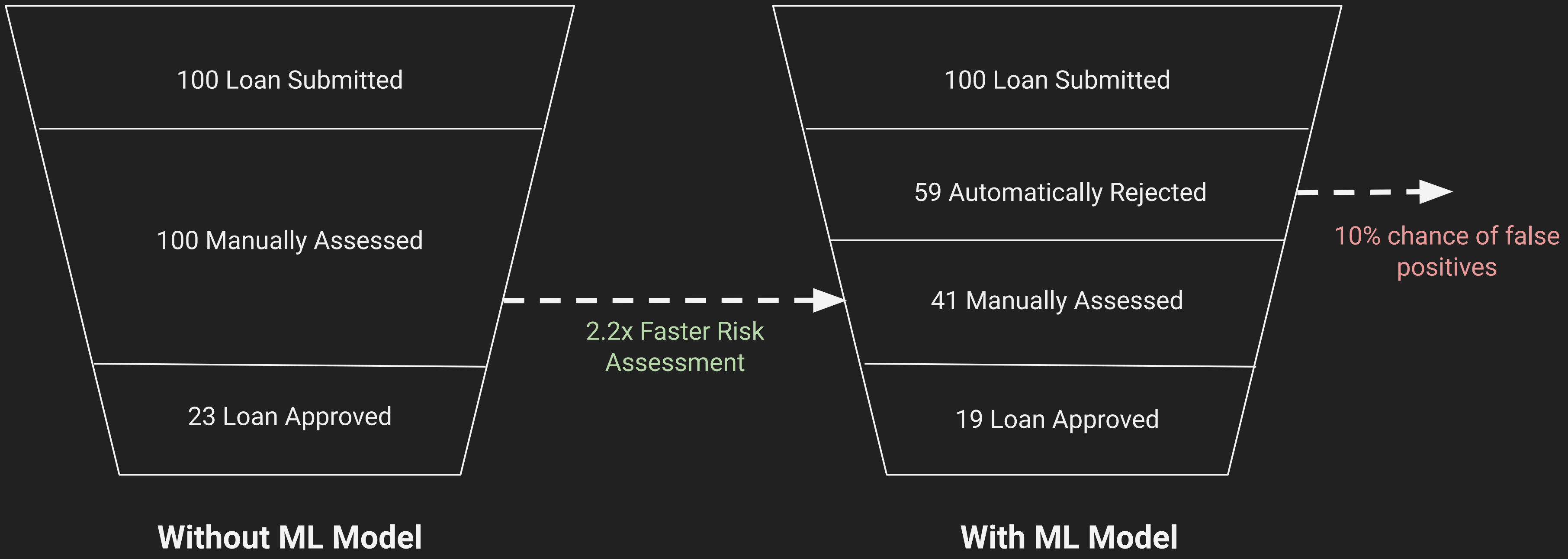


A loan approval prediction is a system to help the **Credit Analyst** on doing risk assessment by implementing a score to each application based on their multiple characteristic to boost the **speed of assessment** and **minimize human error** when doing so.

Based on this assessment, an implementation of Machine Learning Model for loan approval prediction can multiply the revenue roughly by 2x.

*assumptions and calculation on appendix xx

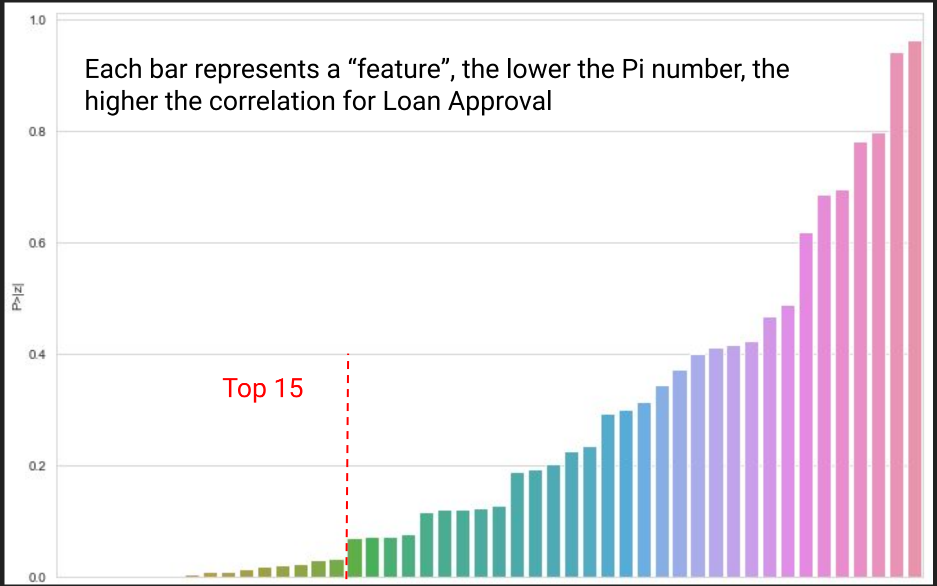
59% of applications are automatically declined w/o human intervention



Checking Account Status & Credit History are the top two characteristic for the risk assessment

	Top 5 Features
1	No Checking Account
2	Installment as Income
3	Purpose for Used Car
4	Foreign Worker
5	Credit Amount

* *Red = Negative Correlation*
**Green = Positive Correlation*



Heavy feature engineering needs to be done because of the raw data quality

Raw Train Data

↓

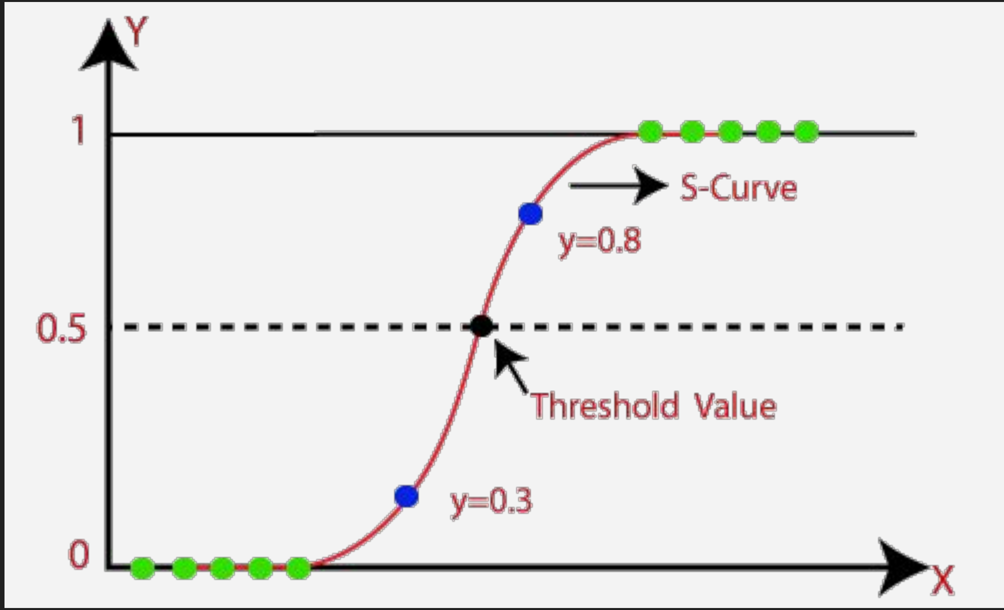
Standard Scaler for Numerical Variables

↓

Encode categorical Variables using pandas.get_dummies

↓

Data oversampling using SMOTE



Logistic Regression Model

Summary from the exercise

Suggestions:

- Machine Learning Model are best applied when the risk assessment stage becomes a bottleneck for the company revenue growth
- Credit Analyst can spend more time assessing to the most significant characteristic/ features based on the model recommended.

Things to improve:

- The amount of training data is not enough to produce a high precision model
- The quality of the raw data needs to be improved to have a more efficient model creation
- There are possibility to try different models if there are more time to assess the data

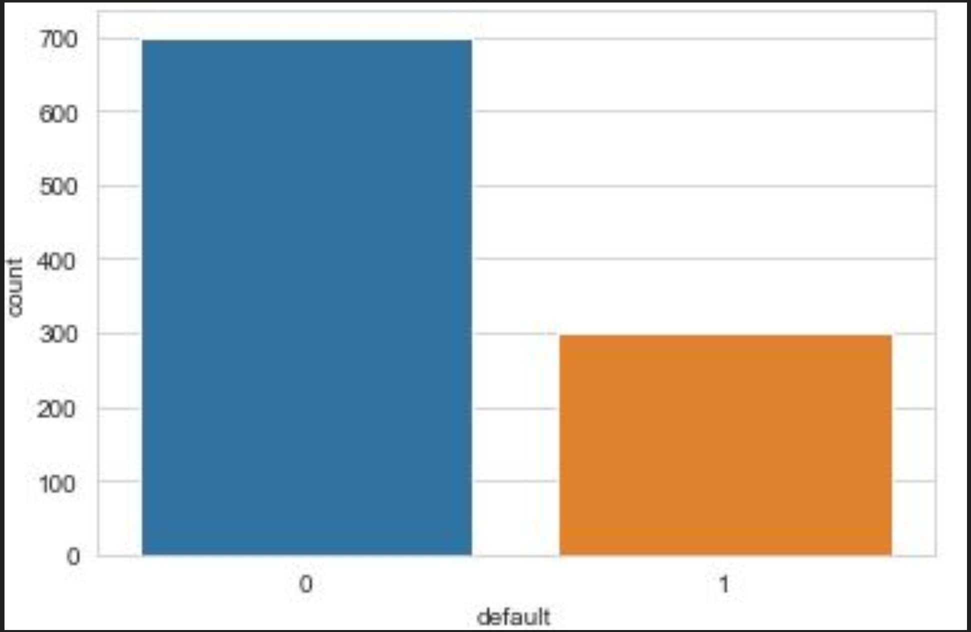
APPENDIX

Appendix 1 - Data Understanding

- The Data contains 20 independent variables
- 7 Numerical Variables and 13 Categorical Variables

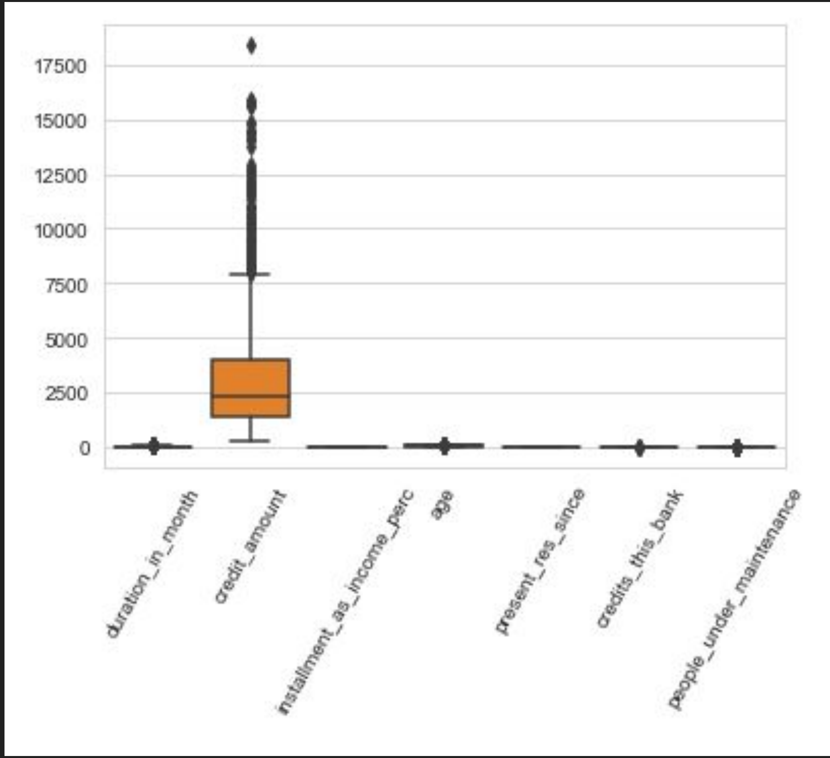
```
Data columns (total 21 columns):
default                       1000 non-null int64
account_check_status          1000 non-null object
duration_in_month             1000 non-null int64
credit_history                1000 non-null object
purpose                       1000 non-null object
credit_amount                 1000 non-null int64
savings                       1000 non-null object
present_emp_since             1000 non-null object
installment_as_income_perc    1000 non-null int64
personal_status_sex           1000 non-null object
other_debtors                 1000 non-null object
present_res_since             1000 non-null int64
property                      1000 non-null object
age                           1000 non-null int64
other_installment_plans       1000 non-null object
housing                       1000 non-null object
credits_this_bank             1000 non-null int64
job                           1000 non-null object
people_under_maintenance      1000 non-null int64
telephone                     1000 non-null object
foreign_worker                1000 non-null object
```

Appendix 2 - Target Data



Target data is not balanced, Oversampling technique need to be applied to balanced the data and get more accurate predictions

Appendix 3 - Numerical Features



Need to apply standard scaler to standardize the scale of the numerical features

Appendix 4 - Confusion Matrix

Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

	P No	P Yes
A No	55	22
A Yes	4	19

Appendix 5 - Classification Report

Classification Report

	precision	recall	f1-score	support
0	0.93	0.71	0.81	77
1	0.46	0.83	0.59	23
accuracy			0.74	100
macro avg	0.70	0.77	0.70	100
weighted avg	0.82	0.74	0.76	100

Appendix 6 - Significant Features

	P>|z|	coef
account_check_status_no checking account	2.301828e-14	-1.586928
installment_as_income_perc	4.355345e-09	0.523008
personal_status_sex_male : single	9.442988e-06	-0.842057
purpose_car (used)	3.606069e-05	-1.723244
foreign_worker_yes	6.559689e-05	2.106572
credit_amount	1.660816e-04	0.425709
credit_history_critical account/ other credits existing (not at this bank)	4.514576e-04	-1.355127
property_real estate	3.910701e-03	-0.656115
present_emp_since_4 <= ... < 7 years	7.778062e-03	-0.716524
age	9.664570e-03	-0.242301