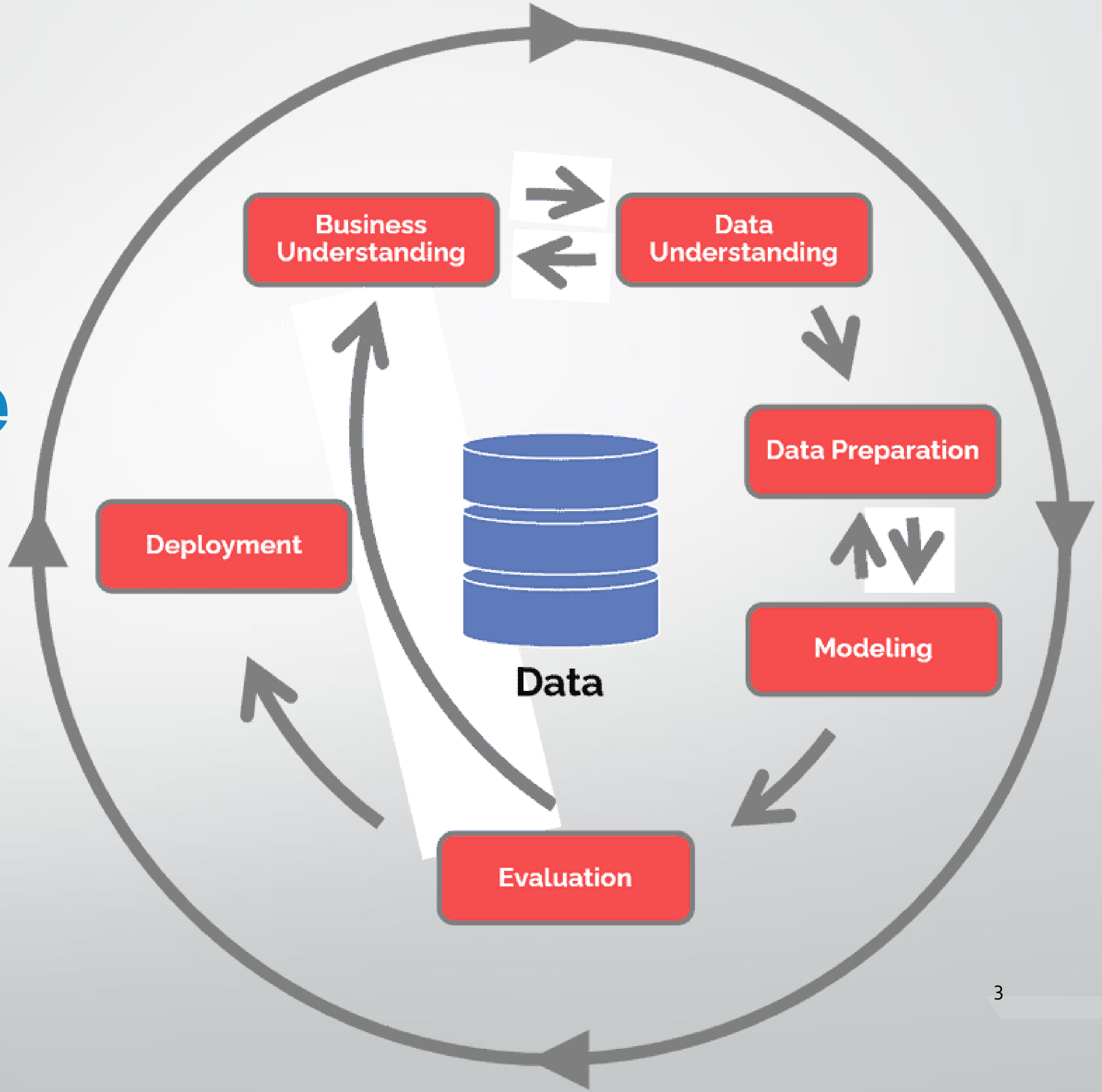
Data Analysis Home Loan Eligibility Prediction

Grace Eze

Agenda

- Data Science Lifecycle
- Project Overview
- Data
- Analysis
- Modeling
- Model Evaluation
- Recommendations

Data Science Lifecycle



Project Overview

Business Problem

The process application for a home loan involves loan officers having to manually process them which takes 2 to 3 days, thus applicant's don't learn the results of their eligibility on time.

Business Objective

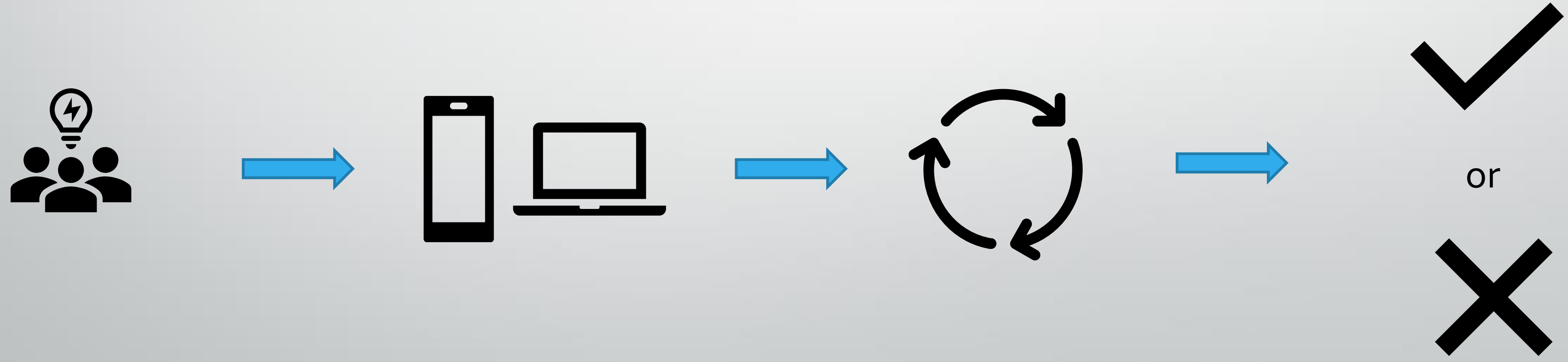
Embrace digital transformation and give the customers a complete set of services such as assessing the credit worthiness of an applicant, checking the status of their loans from the convenience of their mobile devices and much more using machine learning thus speeding up the notification process.

Hypothesis

Use of Machine Learning methods, both automated and classical algorithms to predict an applicants loan status

Project Overview

An applicant can apply on any device by filling his/her information (Gender, Marital Status, Income etc.). Upon completion the ML model makes a predict (based on historical data that it has been trained on). The prediction will appear on the device as Accept or Decline on the same device in a matter of
seconds



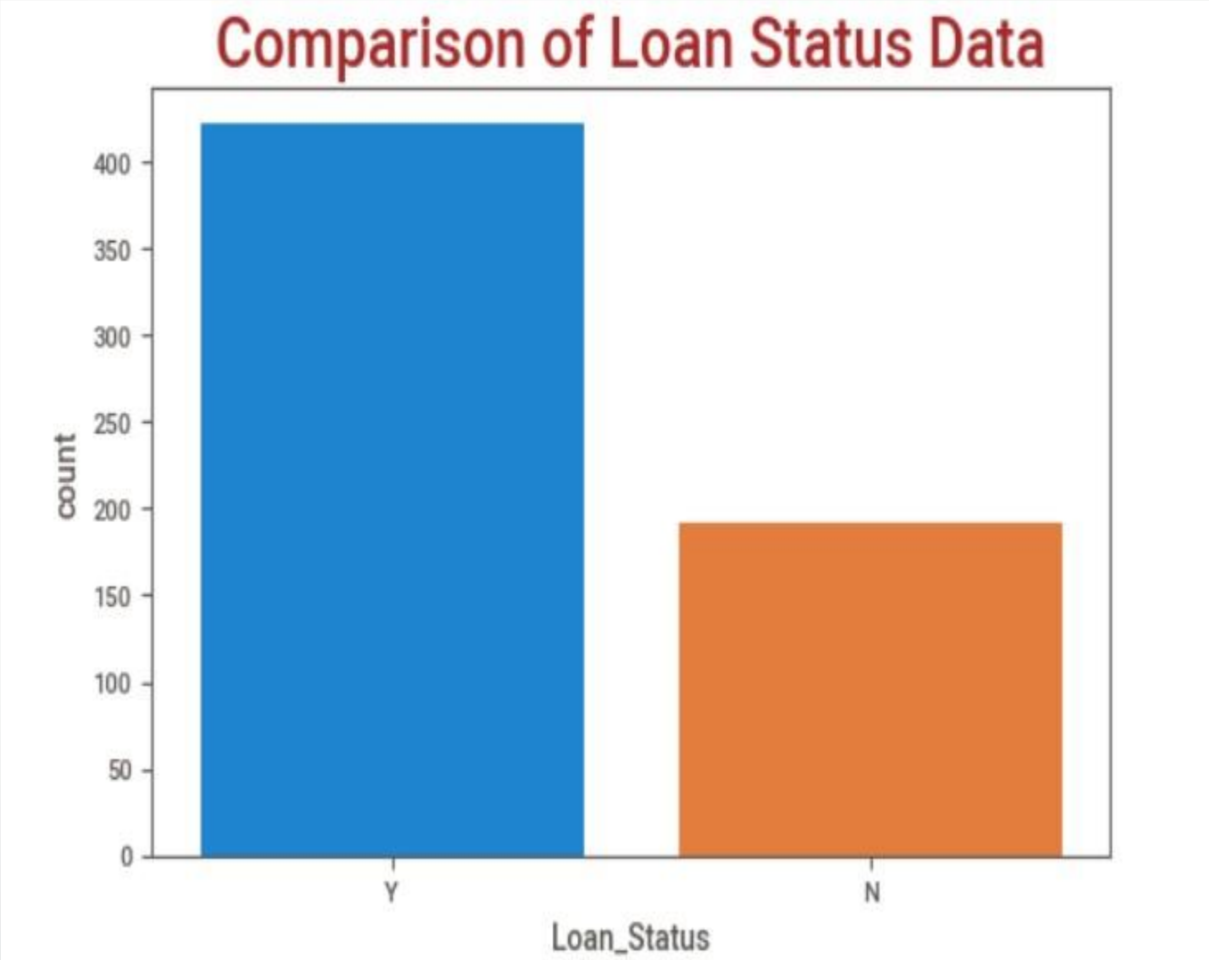
Overview of Data

- **The data focused on was the historical data i.e train data.**
- **The data has 614 unique customer entries and 13 fields containing information with respect to each customer entry, it has both numerical and categorical data.**
- **Our target variable is the Loan Status**

Analysis

The Loan Status gives an indication of customers who applied for a loan versus who didn‘t. it indicates 422 persons have an active loan application while 192 don‘t.

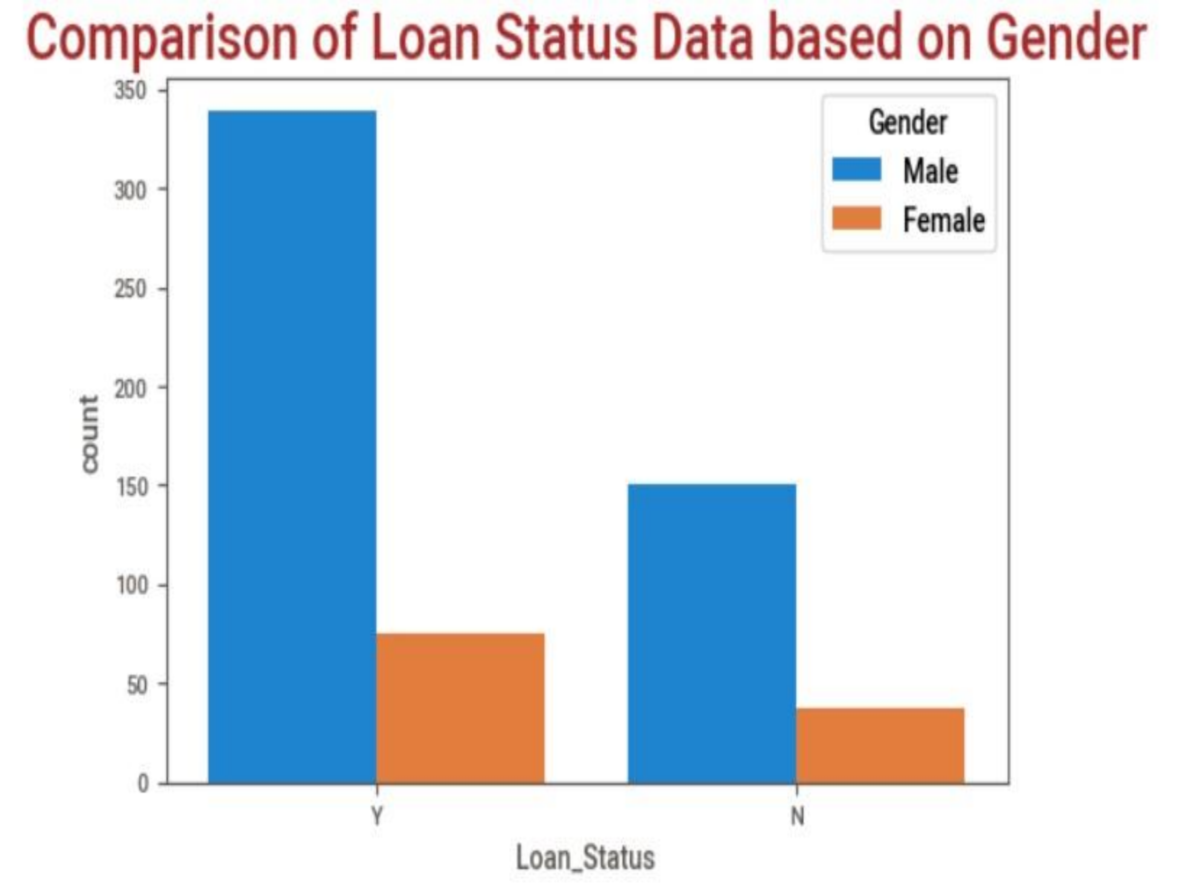
- **Yes(Y) and NO(N)**



Analysis

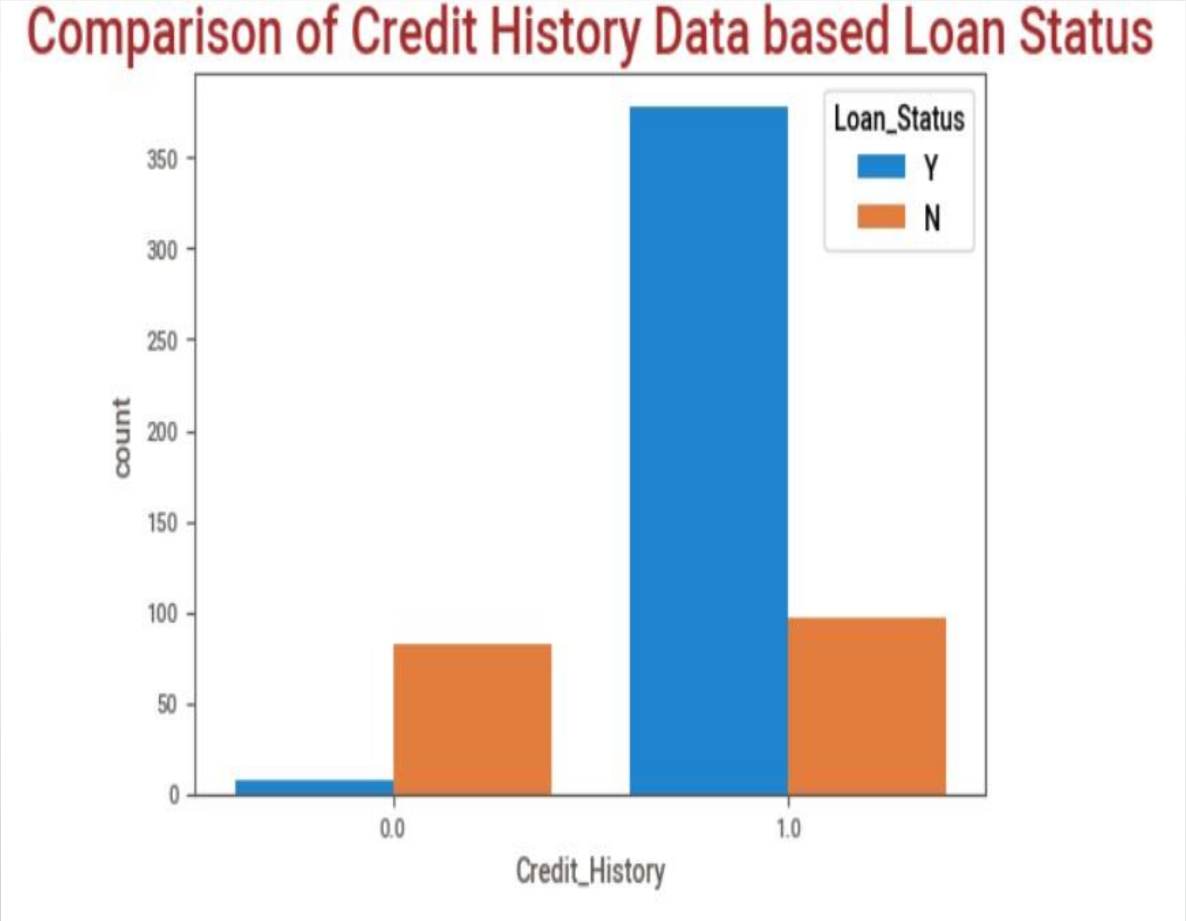
There are 489 Males and 112 Females. The Comparison shows that based on the historical dataset, it shows that men have a higher loan status than women do.

- **Yes(Y) and NO(N)**
- **Male(M) and Female(F)**



Analysis

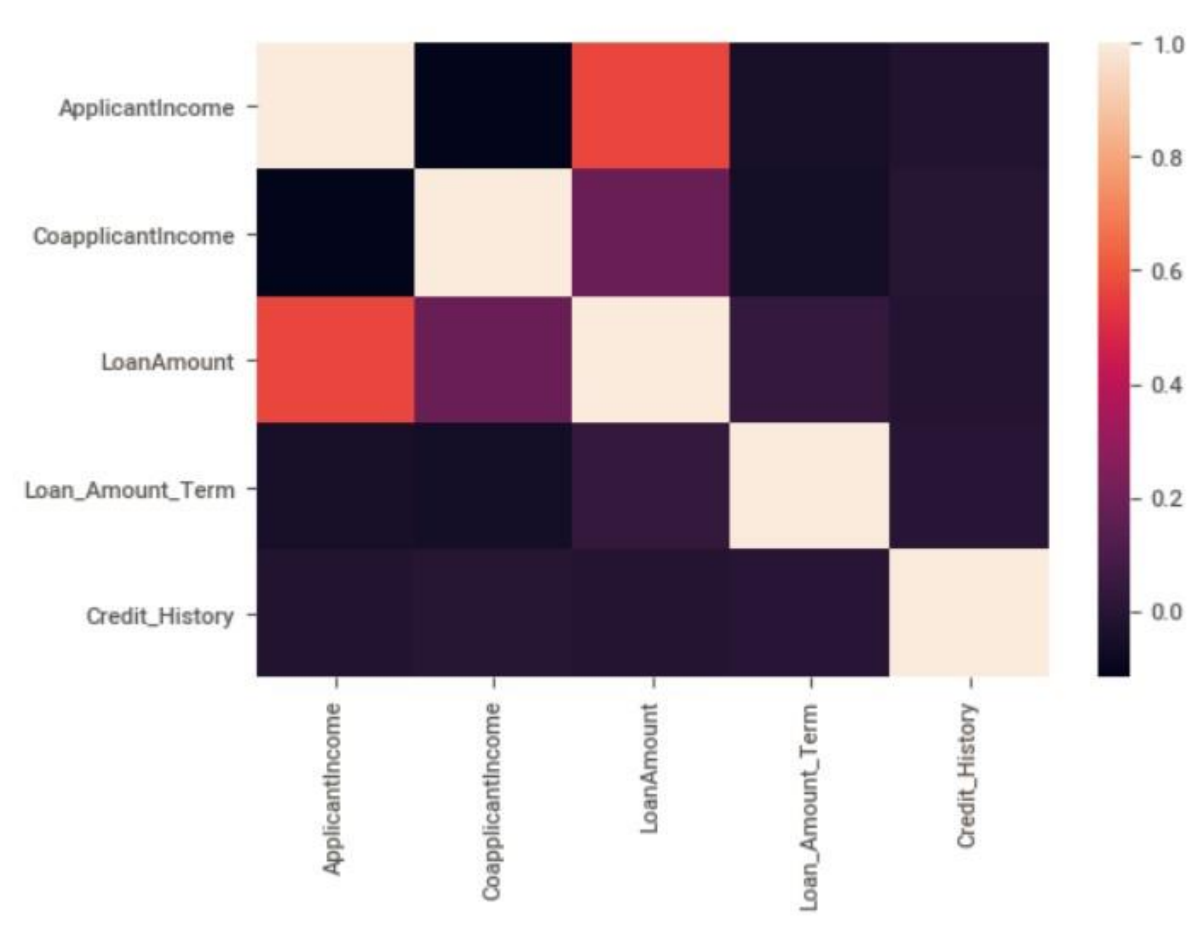
There are 475 customers with credit histories and 112 without. The comparison shows that based on the historical dataset, it shows that applicants with credit histories are more likely to default than those who don't have them.



- **Yes(Y) and NO(N)**

Analysis

The confusion matrix shows there is a correlation between Applicant's Income and the loan amount requested.



INTERNAL

Modeling

- **The AutoSklearn AutoML was used in training and making predictions on our data.**
- **The Logistic Regression model was used as the Bespoke ML model.**

Model Evaluation

The metric used to evaluate the models in this task is the **Accuracy** which helps determine the fraction of predictions our model got correctly.

Evaluation Metric	Bespoke ML	AutoML
Model Accuracy	65%	79%

Recommendations

- **The evaluation shows that the AutoML performs better than the Bespoke ML**
- **We get quicker results AutoML which helps us save on time and resources.**
- **AutoML gives us less insight on our data and less customizable compared to Bespoke ML's**
- **Using the Bespoke ML helps us fully understand all the details of our data, details of the algorithm and how it can be further customized also improved to meet the business needs for the company.**