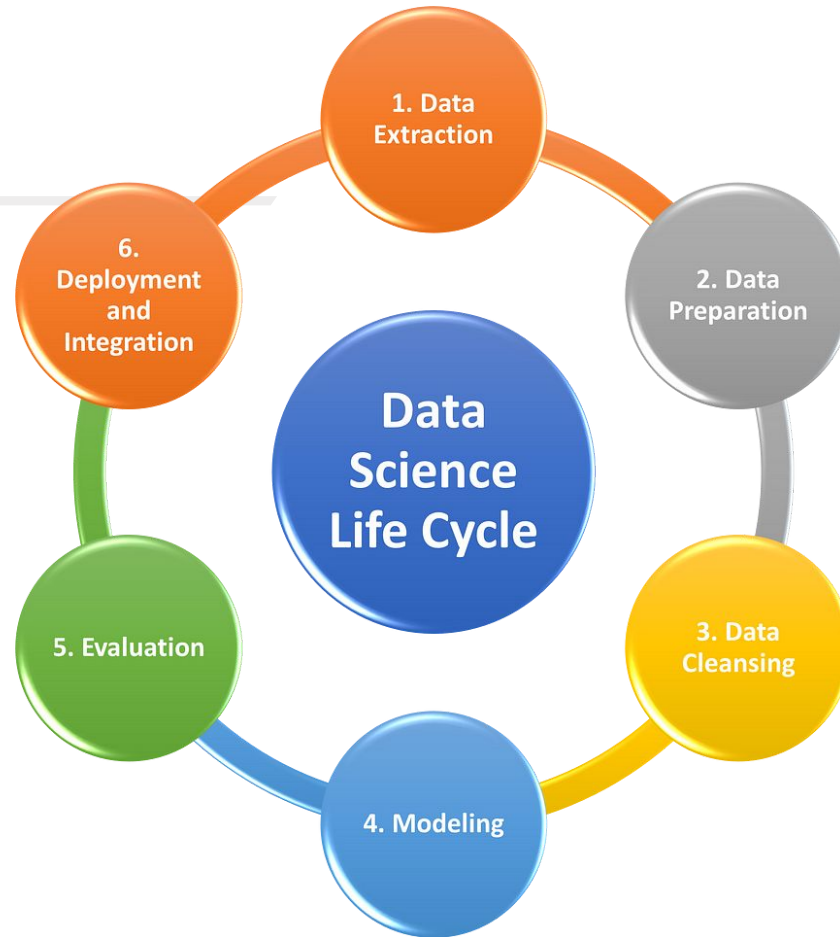# OPTIMIZING HOME LOAN APPROVAL WITH MACHINE LEARNING

27/04/2024

Agenda

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

Data Science Lifecycle

Business Problem:

Standard Bank is embracing the digital transformation wave and intends to use new and exciting technologies to give their customers a complete set of services from the convenience of their mobile devices.

As Africa's biggest lender by assets, the bank aims to improve the current process in which potential borrowers apply for a home loan. The current process involves loan officers having to manually process home loan applications. This process takes 2 to 3 days to process upon which the applicant will receive communication on whether or not they have been granted the loan for the requested amount.

Business Objective:To improve the process Standard Bank wants to make use of machine learning to assess the credit worthiness of an applicant by implementing a model that will predict if the potential borrower will default on his/her loan or not, and do this such that the applicant receives a response immediately after completing their application.

Project Overview

Hypothesis: Based on historical data we can use machine learning to predict the loan status of a potential borrower such that the time taken for them to receive their respective statuses is reduced significantly

Process Overview / Solution

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

Data

- The data was divided into two datasets:
- The train dataset is used to train the machine-learning model whereas the test dataset is used to evaluate the performance of the model.
- The train data had 13 features whereas the test data had 12 features.
- The number of records are 614.
- ▪Target/Loan Status – Y (422) vs N (192)

Analysis

-The Exploratory data analysis process involved identifying the general patterns in the data which included identifying features and outliers which might be unexpected.

Modeling

One machine learning model trained and AutoML used as well.
-Bespoke model required preprocessing -AutoML did not
-AutoML had a higher accuracy than BespokeML

Model Evaluation

-For our modelling process we use autoML with autosklearn which had an accuracy of 0.79 whereas the bespoke ML had an accuracy of 0.65.
- Where accuracy is the sum of all the correct predictions made by the model over all predictions made.

Recommendations

❖ Bespoke ML > AutoML

❖ We understand exactly what went in, how it went in and what algorithm was used to achieve the objective

❖ Less time training (works in our favour if we train and predict in real time – maybe not applicable in this use case)

❖ AutoML is best used as a baseline model.