

**Conception Phase**

**M.Sc. Data Science**

**DLMDSPDSUC01 – Project : Data Science Use Case**

**20th February 2023**

**Caroline Karimi Gitari**

**92125360**

**Tutor: Prof. Sahar Qaadan**

# **Introduction**

The aim of this project is to develop a credit worthiness prediction model for Bora Capital, a digital lending startup company. Bora Capital is a non-bank financial institution operating in Kenya. It offers online merchant financing to vendors on e-retail platforms.

Bora Capital has 1300 registered customer accounts across e-retail platforms. Out of the opened accounts , Bora Capital has disbursed 1,005 first loans. These loans have been disbursed using an excel based expert-credit scoring model. The total amount disbursed over the period was KES 24,176,000 and the loan tenor for each loan was 1 month.

## **Business Challenge**

Using the expert-based credit scoring model , Figure 1 shows the loan performance of existing loans.

Diagram

Description automatically generated

Figure :Bora credit existing loans performance

While the share of defaulted loans is significantly higher than both Partially and fully paid loans , the company had initially included this possibility in its initial budget as a trade-off for “buying data” to be used in building subsequent data-based credit assessment models.

# **Objectives**

The objectives of this project are as follows:

* Build a data-based credit scoring model – Using currently available data to build a model that predicts whether a new customer is good(Fully Paid Customers) or Bad (Partially Paid or Defaulted Customers).
* Identify the most important variables in determining the credit worthiness of customers – This will enable the company to focus on collecting valuable data.
* Identify the positives and negatives of the developed model – After building the new model , this step will involve measuring its business value and accuracy.

# **Methodology and Tools**

The methodology for this project involves the following steps:

* Data -Existing Customer data will be used for model building.
* Data Preprocessing- Cleaning and formatting data to ensure consistency and accuracy. This includes dealing with missing data and handling outliers.
* Feature selection -The most important features will be selected for model building.
* Model Selection - Choosing the appropriate algorithm to build the credit worthiness prediction model. In this case, we will be using a supervised learning algorithm since we have labelled data. We will run two models and select one with the highest accuracy.
* Model Building – This will involve splitting the data into training and testing sets and training the model on the training set.
* Model Evaluation – The built models will be evaluated using their accuracy and the model with higher accuracy will be chosen as ideal.
* Model Comparison – The chosen model will be evaluated against the current expert-based model.

The tools used in this project include:

* Python programming language
* Jupyter Notebook: This will be used to write and execute Python code.
* Scikit-learn library: This will be used for machine learning algorithms.
* Matplotlib and Seaborn libraries: These will be used for data visualization.

# **Conclusion**

This project aims to develop a data-based credit scoring model credit for Bora Credit, a digital lending startup company. The model will be developed using available data on categorical variables such as gender, business registration type, savings frequency, and numerical data such as monthly amount sold, total and amount of previous loans. Originally, two different models will be built and evaluated using various accuracy metrics and the one with the highest accuracy will be selected as the ideal model for use in future credit assessments.