# **Abstract**

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 which offers online merchant financing to vendors on e-retail platforms.

The data used for the model is existing loan data from loans lent using an expert based excel model , which consists of loans currently labeled as fully paid , partially paid, or defaulted. The share of defaulted loans is 73% hence the need to build a data-based credit scoring model to enable Bora Credit to make credit decisions based on data-driven insights rather than subjective judgments which are more prone to errors.

The target variables are :

* Loan status – business classification of the loan.
* Amount Paid – Total amount repaid by the customer against the loan.

To achieve this goal, some preliminary steps include :

* Exploratory data analysis (EDA) to gain insights into the data.
* Feature engineering to create new features from the existing data.

We then run two different machine learning algorithms, for each output variable to compare performance and choose the model that performs better. The best-performing models are selected for making credit decisions.

# **Technical Approach**

The data used is readily available first loan cycle data. We first conducted an exploratory data analysis to understand the distribution and relationships of the features in the dataset and selected the features. In addition to the raw variables from customer historical loan data, we generated additional variables.

The applicable machine learning methodology for the project are:

Table

Description automatically generated

We split the dataset into training and testing sets, where the training set was used to train the model, and the testing set was used to evaluate the model's performance. Using a combination of abstract model performance metrics (precision, recall and overall accuracy) and the company’s gain (or cost reduction) by using the model compared to not using it the best performing model is selected.

Finally, we deployed the model to create a web application that can predict customer status and loan amount in real-time. The application takes input from the user, preprocesses the data, and returns the predicted loan status and loan amount.

To ensure the model remains accurate and serves its intended purpose , we will regularly monitor its performance through regular data analysis, performance metrics tracking and tracking of key performance indicators. In addition, customer feedback on improvements will be incorporated.