**Technology Justification Document**

**Credit Card Default**

Inter-track project to build an app "**Predictz"** that can predict if a customer will default in payment using demographic data and history of past payment. Estimation of this kind of risk has been done by statistical methods through decades and with respect to recent development in the field of machine learning, there has been an interest in investigating if machine learning techniques can perform better quantification of the risk. The aim of this project is to examine which method from a chosen set of machine learning techniques exhibits the best performance in default prediction with regards to chosen model evaluation parameters. The investigated techniques were Logistic Regression, Random Forest, Decision Tree, XGBOOST, Artificial Neural Network and Support Vector Machine. The results showed that XGBOOST without any threshold adjustment provided the best results. Upon adjustment of classification threshold the XGBOOST model provided the best recall, precision value.

**Problem Description**

Financial institutions especially banks offer amounts of loans and credit facilities mostly to customers who cannot afford collaterals to get these credits from them. Despite the recent growth in this sector, the sector is faced with challenges of loan and credit repayment defaults by clients. This has caused a certain decrease in amount of these credit facilities are available to customers. In Nigeria, the credit and loan repayment percentage is very low as most customers will most likely default in loan repayment.

**Solution**

To solve the above stated problem, an app **"Predictz"** was built. Predictz would help predict if a customer will default in payment using demographic data and history of past payment. The process is fully automated and helps eliminate any form of human bias.

**Role Played**

As the data scientist in the group, the prediction model was built by me. Several machine learning methods were examined. They include; Logistic Regression, Random Forest, Decision Tree, XGBOOST, Artificial Neural Network and Support Vector Machine. From the initial probe into model accuracy, a XGBOOST gave the highest accuracy of 0.824, with the second highest being the Random Forest method. Upon optimization by adjusting the classification threshold we observe a better precision to accuracy ratio, better recall value and thus F1 score. The model was created using pickle, loaded and an API created so the model could interact with the front/user's end.

**Payoff**

A system can help predict credit default, eliminating human bias will help give loans to deserving individuals, creating an efficient system and happier customers. It will help eliminate defaulters and help lenders save money by reducing cost accrued from credit default.