**Project Proposal**

**<Candidate Name>: <Mobile Number>**

**University of Memphis**

**Purnakumar Thota**

**Title: Smart banking customer targeting using ML Ensemble for improved business operational efficiency through reduced cost per call (CPC)**

**Introduction:** Traditionally businesses reach out to prospect customer for encashing potential opportunities (can be in terms of cross sell or upsell). Though this process of targeting customer for potential cross sell or upsell in observed in many industries, it is very wide and frequent seen in banking industry were a customer holding an with a bank will be targeted for cross selling opportunities like loans, fixed deposits and term deposits etc., Banks traditionally uses various channels to reach their customer and one of the major such channel is Telemarketing. Though telemarketing seems like a very easy way to reach a customer at the same time it is very costly. As per industry standard typical Cost per Call (CPC) is around $2.7 to $5.6 [1] (it might difference from business to business), based this statistic we can estimate the possible impact of targeting a wrong prospect and importance of accurate targeting strategy. In the current project we plan to address this problem of high operating cost due to inaccuracy customer targeting using machine learning. As part of the analysis, we will be using banking telemarketing call data for predicting the propensity of a customer opting for cross selling, which can later be used by banking businesses for making better call plan as well customer target list

**Data Source:**

Source: UCI Machine Learning Repository

Dataset: Bank Marketing Dataset

Link: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

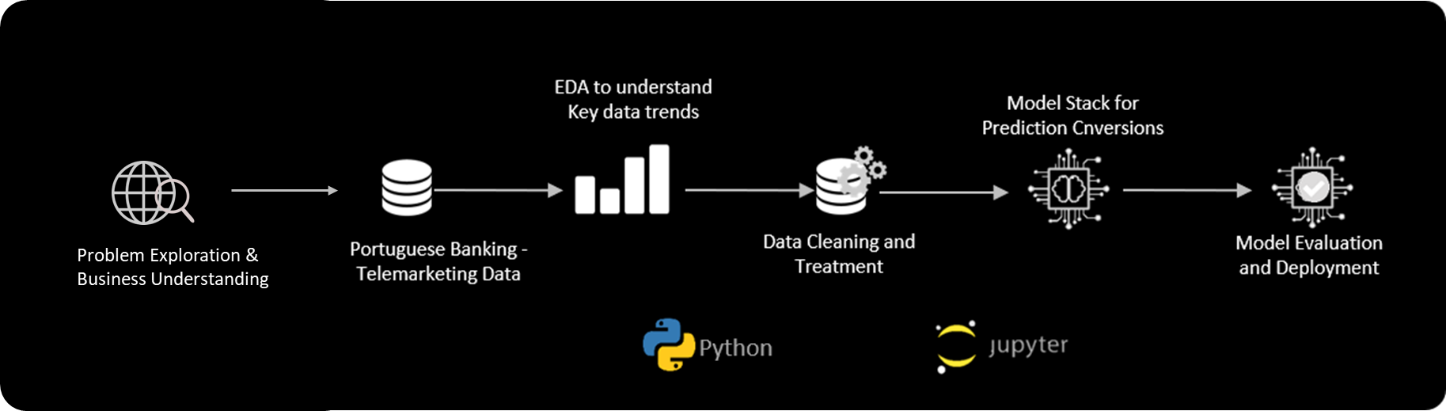
Method of Extraction: Direct download

Dataset Description (As per source): The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

**Proposed Solution Approach:** As the objective of the current solution is to predict the customer propensity, we need to build a classification model for the current problem. Also, for this kind of problem, businesses would be keen to look at the drivers of recommendation as it helps them in planning better and deep learning model due to their black box nature might be not ideal for implementation. Hence, we will be using traditional machine learning models as they offer better visibility into the ‘why?’ part of recommendations. The following stack ML model will be test for the current business problem and an ensemble of individual model recommendation will be used for final prediction

* Logistic Regression (baseline model)
* Random Forest
* XGBoost

Based on the initial look at the data we found the data suffers from class imbalance problem hence appropriate class balancing technique like over or under sampling will be using for addressing this problem. Since the problem we are dealing with is classification problem MAPE would be used as primary measure for model performance (on test, validation and train datasets).



Finally, an interactive web application generating real-time recommendation on customer’s propensity once users provide customer details.

**Tools:**

* Programming Languages:
  1. Python: For programming entire modeling pipeline (IDE: Jupyter Notebook)
  2. UI development: Streamlit, Javascript, HTML and CSS (IDE: VS Code)
* Libraries/Frameworks:
  1. Sklearn: Machine learning models
  2. Steamlit: Web Application

**References:**

1. <https://www.liveagent.com/customer-support-glossary/cost-per-call/>