**Project Proposal for Capstone 1**

**Background/Context:**

Commercial banks spend considerable time/energy in marketing various products/offerings to their clients. It makes sense for them financially to only focus their marketing efforts on those clients who they believe will be most responsive to a particular product offering(s).

End clients that I have in mind for capstone 1 project are marketing managers at commercial banks who want to be efficient in their use of time/energy at marketing various products/offerings to their clients. As result of analysis that I provide to them via this project, they would be able to classify clients expected to be most responsive to a particular product/offering and hence they would then focus their team’s effort/energy primarily or only on those clients and as a result increase ROI on their marketing efforts.

**Problem to be solved and beneficiary of solution:**

Specific problem I will be addressing for commercial banks is that of increasing subscription to term/fixed deposit accounts among bank’s clients who have current/checking accounts but don’t have term/fixed deposit accounts. Increasing subscription to term deposit accounts among clients is important for commercial banks because it ensures availability of funds for a defined term/period that the commercial banks can then lend out to corporations/individuals for profit.

As a result of this project, marketing managers at commercial banks would be able to identify those clients (that already have current accounts but don’t have term deposit accounts) that are expected to be most responsive to an offer of term deposit account. Marketing managers would then focus their efforts only on those clients vs all the clients to get high ROI on their efforts focused on increasing subscription to term deposit accounts.

**Data gathering:**

I will be using data provided at Kaggle:

<https://www.kaggle.com/prakharrathi25/banking-dataset-marketing-targets>

The data has 18 columns and about 31,600 rows.

**Method(s) to be used:**

First, data obtained from Kaggle would be explored to get better understanding of the data. Data would be cleaned/processed as necessary to prepare it for next steps i.e. training the models. Data would then be divided into training set and test set. A set of models to be explored would be chosen. Currently the set of models to be explored are the following classification models:

i) Logistic regression

ii) Decision trees

iii) Random forest

iv) Gradient boost

The training dataset would be used to train each of the models. Trained models would then be used to identify clients most likely to subscribe to an offer of term deposit from the test dataset. The outputs from the trained models would then be evaluated using combination of following metrics to decide which trained model is the most optimal one for the purpose:

i) Accuracy

ii) Confusion matrix

iii) Classification report

iv) Area under Precision Recall curve

**Deliverables:**

The deliverables would be a slide deck (powerpoint format), a written report ( word format) and python code (jupyter notebook format).