**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| Name - V Bhavya Reddy  Email - [bhavya.reddy0711@gmail.com](mailto:bhavya.reddy0711@gmail.com)  Contribution - whole project. |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/coolphotography/Bank-Marketing-Effectiveness-Prediction.git> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Problem statement:**  We are given the data of direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit (target variable y). This project will determine the success of Bank Telemarketing.  **Approach:**  We begin with the data set overview, in which we briefly analyze the observations and note several characteristics of numeric and categorical features. We take note of all the unique values present in every feature. In addition, we find the number of NULL values - none were found.  The next step is Exploratory Data Analysis (EDA) on the dataset, where we analyze numeric and categorical features through bar charts, box plots, countplots, etc. We also use heatmaps to analyze correlation, and VIF calculations are carried out to check for multicollinearity.  In feature Engineering, we drop unnecessary columns, modify the dataset, and remove unwanted observations. Later, Data encoding followed by train-test split is performed on the data set.  With the train data we train our models using Logistic regression, Decision Tree, Random Forest Classifier, K Nearest Neighbours, Naive Bayes algorithms. Several hyperparameters are tuned to optimize the performance. We try to find the most suitable model by comparing the evaluation metrics of various models. We found that the Naive Bayes Model had the best performance. Worst performance is seen by Random Forest Classifier. In order to gain a better understanding of the data, several plots are drawn (Accuracy, Precision, Recall, ROC-AUC, etc.).  **Conclusion:**  1. The key features or attributes that helped in the prediction of the term deposit were - poutcome, age, balance, previous, loan.  2. KNN's prediction is heavily influenced by the majority class, so it seems to be the poorest model for our imbalanced data.  3. As for this project, I have considered the AUC parameter to be significant over other metrics since it considers TPR and FPR. Except for KNN model, AUC scores are similar for other models. AUC for Naive Bayes is slightly lower than that of logistic regression, decision trees, and random forests.  4. Naive Bayes remains the right fit for term deposits due to the recall score (59.4%) being quite high compared to other models.  5. In this project, I considered recall to be more significant than precision. This assumption was made based on intuition. |