**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| * **ANKUSH** * **EMAIL** - 90ankushsaini@gmail.com   # solo project |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/Ankushsaini90/-Bank-Marketing-Effectiveness-Prediction- |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  One of the industries that is being transformed the most by the recent Machine learning advances is the finance industry. Be it predicting the stock prices, or in our case predicting, a customer willingness to subscribe to a term deposit. Therefore, in our project we have come up with a solution that increases the efficiency by making fewer calls but improves the success rate.  In this report, we train a model to predict if the client will subscribe to a term deposit or not.  The data set was obtained from thedata of direct marketing campaigns (phone calls) of a Portuguese banking institution.  First we do Exploratory Data Analysis on the data set. We look for missing data values (none were found) and outliers and appropriately modify them. We also perform correlation analysis to extract out the important and relevant feature set and later perform feature engineering to modify few existing columns and drop out irrelevant ones.  We then look at several popular individual models from simple ones like Logistic Regression to more complicated ensemble ones like Random Forest . Additionally, few options for model formulation were tried - first we differentiate numerical and categorical columns, then we analyze both of them separately, we do label encoding on categorical columns too.  Finally, we also tried stacking algorithms where the predictions from the level 1 individual models were used as meta-features into a second level model (Logistic Regression, Random Forest, KNN, SVC, LGBM andDecision tree) to further enhance the predicting capabilities.  The labeled data set provided consisted of the first 19 days of each month. Hyperparameters were tuned using GridSearchCV cross validation using 5 folds on part of the provided training data set. The remaining data were used as a hold out set to test our model performance. Of all the methods and models, we found the Random Forest Ensemble method using a Single Model and Categorical Feature set an ideal choice with training and test scores and we found the best model among them. |
|  |