**Problem Statement:**

The Portuguese Bank had run a telemarketing campaign in the past, making sales calls for a term-deposit product. Whether a prospect had bought the product or not is mentioned in the column named 'response'. The marketing team wants to launch another campaign, and they want to learn from the past one. You, as an analyst, decide to build a supervised model in Python and achieve the following goals: Reduce the marketing cost by X% and acquire Y% of the prospects (compared to random calling), where X and Y are to be maximized

**Solution:**

We are having a dataset of 20 features which can be used to predict whether a customer will start a term deposit or not. So instead of making sales calls randomly to various customers and trying to convince them for a term deposit, we can filter the customers seeing their features and past trends. A model can be trained using different classifiers which will be learning from previous trends given in the dataset and can classify whether the user may be interested or not. This will reduce the cost of labour considerably as we can segregate the users.

Classifiers which I used for the task are:

1.   Logistic Regression Classifier
2.   Decision Tree Classifier
3.   K-Nearest Neighbours Classifier
4.   Random Forest Classifier
5.   Support Vector Machine Classifier

All the models take different run time and different parameters depending upon the dataset. The Accuracy achieved from the models are:

| | |
|---|---|
| Logistic Regression Classifier | 91.54% |
| Decision Tree Classifier | 88.90% |
| K-Nearest Neighbours Classifier | 91.75% |
| **Random Forest Classifier** | **92.25%** |
| Support Vector Machine Classifier | 91.40% |

Random Forest Classifier can give accuracy up to 92.5% which means that we can predict with an accuracy of 92% even before calling by just passing the features of the person to the trained model.

Hence, we can see that instead of calling randomly to different customers we can call to selective customers which will reduce the labour cost (i.e. Marketing Cost) considerably as well as will give more term deposits making a smaller number of calls(Y). So, using a Classifier is beneficial for the given task.

Reference:

- [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014
- https://archive.ics.uci.edu/ml/datasets/Bank+Marketing