**INTRODUCTION**

**This project entitled "BANK MARKETING" has been brought to fruition jointly by NEHA SINGH and SHUBHNEET TIWARI of BABU BANARSI DAS NORTHERN INDIA INSTITUTE OF TECHNOLOGY ,LUCKNOW under the mentor-ship of our course instructor Mr. Abhijeet and Mr. Amit Singh as a part of our Machine learning through python summer internship program at ETL Labs Pvt. Ltd. The project aims to find the best strategies to improve for the next marketing campaign. how can the financial institution have a greater effectiveness for future marketing campaigns? in order to answer this, we have to analyze the last marketing campaign the bank performed and identify the patterns that will help us find conclusions in order to develop future strategies. A vivid description via graphical analysis,prediction of the rates in future and regression model of the same by making use of linear regression of the available variables.**

**BANK MARKETING**

This is the classic marketing bank dataset uploaded originally in the UCI Machine Learning Repository. The dataset gives you information about a marketing campaign of a financial institution in which you will have to analyze in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

Data will be fetched by web scraping from “**https://www.kaggle.com/janiobachmann/bank-marketing-dataset** “.

**Packages used-**

**Pandas :** A Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

**Numpy :** It is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

**Matplotlib** : It is a plotting library for the Python programming language and its numerical mathematics extension NumPy.

**Sklearn:** It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

**Keywords:**

Machine learning, pattern recognition, classification, supervised learning

**DATA COLLECTION :**

This is the first stage in data analysis. In this process data is gathered from source in an established systematic way.

**HEADING COLUMNS:**

* Age loan
* Job day
* Marital month
* Education duration
* Default campaign
* Balance pdays
* Housing previous
* Contact poutcome
* Deposit

**Splitting the Data set**

As we have seen already, In Machine learning we have two kinds of datasets

* Training dataset - used to train our model
* Testing dataset - used to test if our model is making accurate predictions

Our dataset has 248 records. We are going to use 80% of it for training the model and 20% of the records to evaluate our model. copy paste the below commands to prepare our data sets

**Evaluating the model and training the Model**

We are going to apply the below four algorithms to this problem and evaluate its effectiveness. And finally choose the best algorithm and train it.

**Logistic Regression** : Logistic Regression is a classification algorithm. It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. To represent binary / categorical outcome, we use dummy variables.

**Decision tree** : Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables

**Random forest** : Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees

**CONCLUSION**

This project gives the 98% accuracy report. we conclude that this analysis is vaery useful for marketing and we come to know the applications of machine learning broadly.