**Guvi Datathon 1.0**

**Customer Conversion Prediction**

**Problem Statement**

You are working for a new-age insurance company and employ multiple outreach plans to sell term insurance to your customers. Telephonic marketing campaigns remain one of the most effective ways to reach out to people however they incur a lot of cost. Hence, it is important to identify the customers that are most likely to convert beforehand so that they can be specifically targeted via call. We are given the historical marketing data of the insurance company and are required to build a ML model that will predict if a client will subscribe to the insurance.

**Data**

The historical sales data is available as a compressed file [here](https://drive.google.com/file/d/1BJ_Q8Q-kDRisAQyLltBQggeb0QmdWGZy/view?usp=sharing).

**Features**:

* age (numeric)
* job : type of job
* marital : marital status
* educational\_qual : education status
* call\_type : contact communication type
* day: last contact day of the month (numeric)
* mon: last contact month of year
* dur: last contact duration, in seconds (numeric)
* num\_calls: number of contacts performed during this campaign and for this client.
* prev\_outcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

**Output variable (desired target):**

* y - has the client subscribed to the insurance?

**Minimum Requirements**

It is **not** sufficient to just fit a model - the model must be analysed to find the important factors that contribute towards the price. AUROC must be used as a metric to evaluate the performance of the models.

**Procedure:**

Data cleaning, Exploratory Data Analysis, Modelling and Feature Importance.

**Technologies used:**

Python (Pandas, Numpy, Seaborn, Matplotlib)

Machine Learning algorithms

**ML algorithms:**

Since it is a classification problem, I have used Logistic Regression, Decision Tree Classifier, KNN classifier, XG Boost Classifier, Random Forest Classifier.

The ML algorithms were used to fit the respective models and each model was evaluated to calculate the performance of the model: Accuracy score and AUROC score.

And based on the evaluation metric scores of each model the best model is concluded.