Group Detail

Name: Jose Vicente Solorzano

Email: solorzano.vco@gmail.com

Country: Argentina

College: Montpellier Business School

Specialization: Data Science

Problem Description

ABC bank is about to launch its new product, a term deposit. Before the launching, they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

We were provided with a dataset of clients and features in csv format ("bank-additional-full.csv").

The dataset has the following information:

Gitub Repo Link

https://github.com/jvsolorzano96/bank_marketing_campaign/tree/main/Week%2011%20-%20EDA%20presentation%20%26%20modeling%20propose

Recommended Model

To know what model would perform better in our business problem, we implemented pycaret library.

We implemented pycaret.classification, to try different models in our three datasets:

1) KNN imputation

2) Drop all missing values

3) "Unknown" as a category itself

The better results were from dataset 3. Surely, because we are not deleting valuable information that we do if we use 1 or 2. Our results are based mainly in AUC.

As our recommendation, we are going to try with the 2 best models:

1) Light Gradient Boosting Machine

2) CatBoost Classifier

Next week we are going to implement some Grid Search and hyperparameters optimization to get the ideal model.