Week 9 deliverable

Group name: LatinosDS

Name: Sebastián J. Castro, Ignacio Solórzano, Freddy Tapia, Omar Jazouli

Country: Argentina, Mexico, Spain

Specialization: Data Science

Github repo link: https://github.com/Naquiao/Bank-Marketing-Campaign/tree/main/2-%20EDA

Problem Statement: ABC Bank wants to sell its term deposit product to customers and before launching the product 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).

Problems found:

- The dataset contains several features with an “unknown” value for some instances. This value will be treated as a NaN value.
- The target variable y is imbalanced.
- The feature duration is left-skewed.
- There are many outliers in many features.

In order to overcame these problems, we will try different approaches. For NaN values we will analyze them and determine what is the best solution: imputation or deletion. We will experiment with these two possibilities and we will determine what it the best one for this case. For outliers we will try two approaches: IQR and WOE. IQR and WOE are both the best approaches for treating outliers as part of data preparation for machine learning algorithms. Besides, for overcoming the unbalanced target we will used SMOTE technique, which is one the most common solutions for this kind of problem. Best performace have been observed in machine learning algorithm after implementing SMOTE for oversampling.

Please find ipynb notebook and more details for each strategy at:

WOE: https://github.com/Naquiao/Bank-Marketing-Campaign/tree/main/2-%20EDA/WOE

IQR: https://github.com/Naquiao/Bank-Marketing-Campaign/tree/main/2-%20EDA/IQR