Banco Santander Customer Transaction Prediction.

Problem Statement:

Customer preservation is essential in a variety of businesses as acquiring new customers is often costlier than keeping the current ones. Many companies are therefore always trying to answer the question “How can we predict the value of a customer over the course of his or her interactions with the business”. Santander therefore wants to know future customer transaction predictions.

Company overview

Banco Santander, S.A. is a Spanish multinational commercial bank and financial services company founded and based in Santander, Spain. Their mission is to help promote businesses and people by helping them understand their financial status and how best to achieve their monetary goals. Santander is presently working on how to accelerate its digital transformation and platform strategy to boost growth and increase profitability.

One of these transformations is to identify which customers will make at least one single transaction in the future irrespective of the amount.

This question can best be answered by using predictive analysis of the customer past interactive data with the company and then a machine learning model is used to predict the possibility of these customers being retained or lost.

From these predictions, the company will know how much more efforts they need to keep the old customers which could be through bonuses and discounts, special coupons, better customer services and many more. While also thinking of how to attract more customers into the business through marketing campaigns, offer more discounts and deals, online adverts etc.

Source of Data and description

The data for this project is taken from Kaggle, an online community of Data Scientists and machine learners owned by Google LLC.

The data set contains two comma separated value files for training and testing on the model with anonymous feature names for security purposes.

The train data set contains:

- Unique ID_code
- 200 numerical features labeled from var_0 to var_199
- Target either a 0 for no future transaction and a 1 for future transaction
- Total of 20000 unique records

Methodology

This project will undergo a series of processes outlined as:

- Exploratory Data Analysis where the different relationships between the 200 features and the target
- Inferential Statistics applied on the different features and the target.
- After these analyses, stories will then be brought out from these results depending on the visualizations gotten.
- Since this project is a classification problem with two classes (0 and 1), different machine learning models will be tested on the dataset to see which best predicts with highest model accuracy score.
- Applicable extensions of project and it's impact to the business and/or customers too.