Midterm examination



At Santander our mission is to help people and businesses prosper. We are always looking for ways to help our customers understand their financial health and identify which products and services might help them achieve their monetary goals.

In this challenge, you are invited to help us identify which customers will make a specific transaction in the future, irrespective of the amount of money transacted. The data provided for this competition has the same structure as the real data we have available to solve this problem.

**Data Description:**

You are provided with an anonymized dataset containing numeric feature variables, the binary *target* column, and a string *ID\_code* column.

The task is to predict the value of *target* column in the test set.

1. EDA --- Exploratory Data Analysis

2. PCA --- Principal Component Analysis

3. LR --- Logistic Regression

4. SVM --- Support Vector Machine

5. NN --- Neural Network

6. AUC --- Area under Curve

**Hint:**

**sklearn.preprocessing.StandardScaler**

Standardize features by removing the mean and scaling to unit variance.

**sklearn.model\_selection.train\_test\_split**

Split arrays or matrices into random train and test subsets.

**sklearn.metrics.roc\_auc\_score**

Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC) from prediction scores.