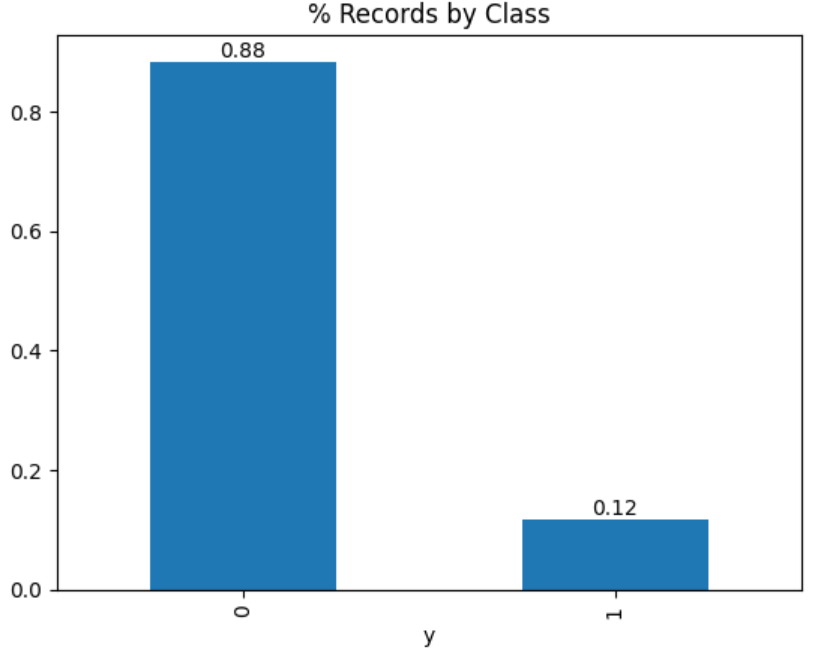
Direct Marketing Campaign



The dataset captures information from phone-based direct marketing campaigns conducted by a Portuguese bank. The objective is to classify whether a client will subscribe to a term deposit, indicated by the variable "y" as either "yes" or "no." Often, multiple contacts were made to the same client to determine their decision.

This analysis helps identify key factors influencing client subscriptions, offering valuable insights for improving campaign strategies. By leveraging this data, the bank can enhance targeting, refine resource allocation, and increase the effectiveness of future marketing efforts.

The dataset exhibited significant **class imbalance**, with only **12% of records corresponding to term deposit subscriptions**. The primary **classification objective** was to accurately predict term deposit subscriptions while prioritizing a **higher True Positive Rate (TPR)**, with slight increase in **False Positive Rate (FPR)**.

s

**Data Level approach to tackle class imbalance**

* Over-sampling may lead to over-fitting as it makes exact copies of the minority samples.
* Under-sampling may discard potentially useful majority samples

Ensemble Methods: Combine algorithmic and data approaches to incorporate different misclassification costs for each class in the learning phase.

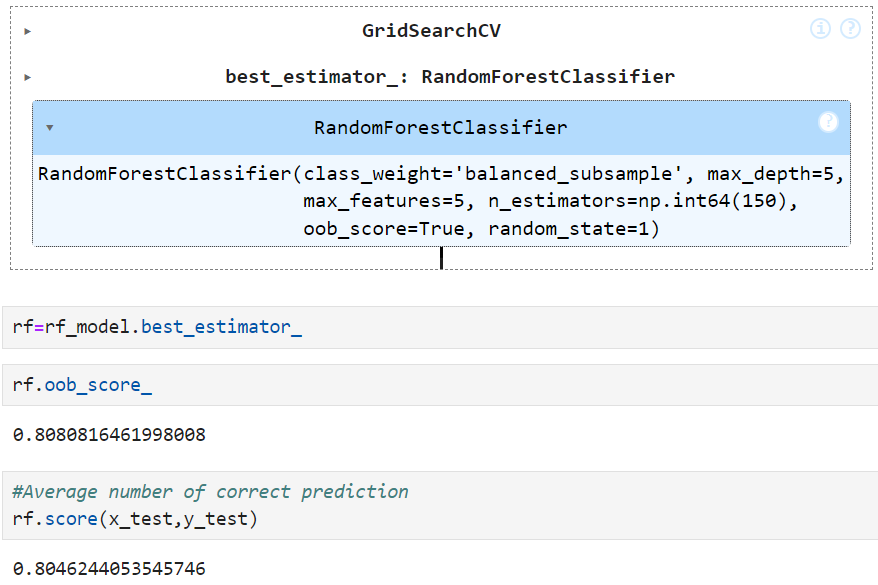
* Since the data is imbalanced will use class weights to weight classes.

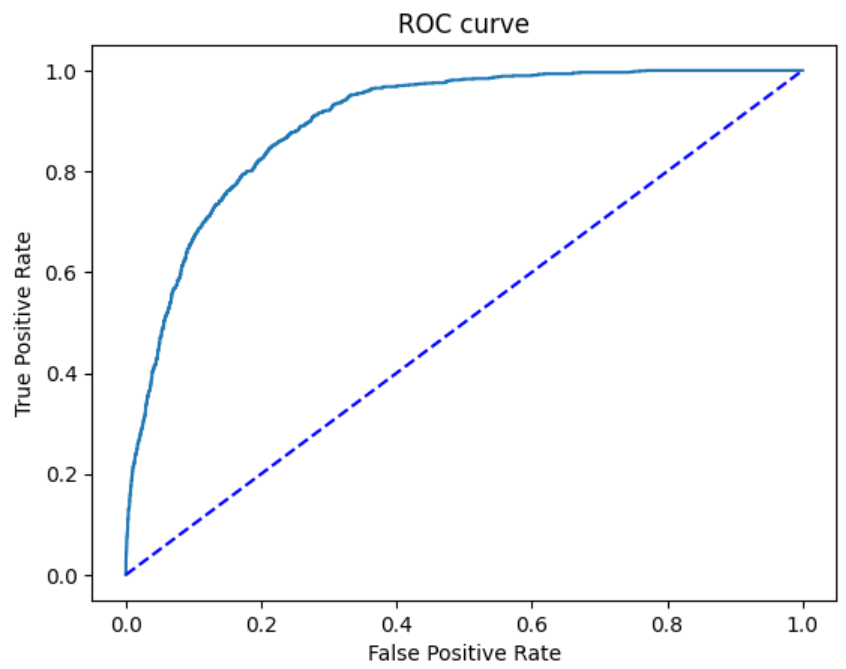
SMOTE: In SMOTE, the training set is altered by adding synthetical (artificially) generated minority class instances, instead of merely copying existing instances in the data.

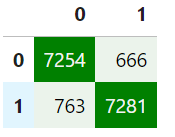
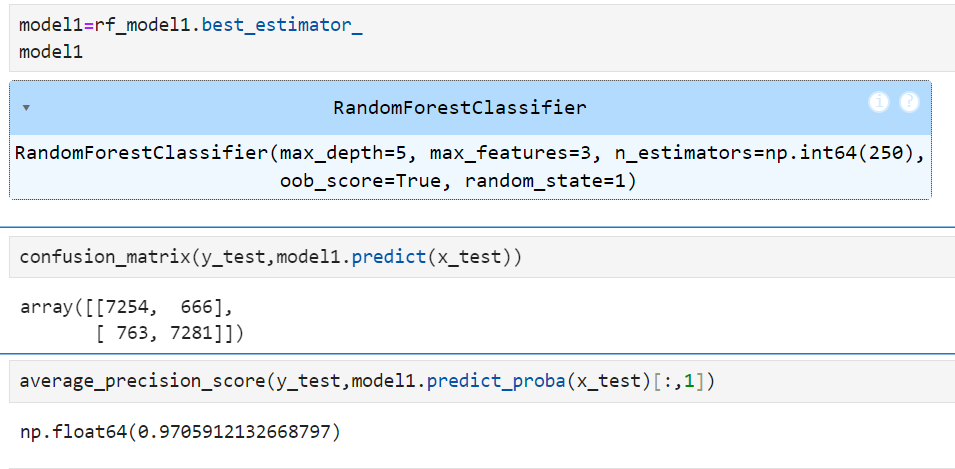
A close-up of a data processing diagram

Description automatically generated

* With **class balancing via the class\_weight parameter**, the **XGBoost model** correctly predicted **62.2% of term deposit subscriptions**.
* Using a **0.4 cut-off**, the **XGBoost model (xgb1)** achieved a **higher TPR** without applying SMOTE.
* By applying **SMOTE for class balancing** and training a **Random Forest model**, the prediction accuracy for term deposit subscriptions improved significantly to **97%**.







**For final predictions**:

* The **SMOTE + Random Forest model ('rf')** is suitable as it provides **valid probability outputs in (0,1)**.
* However, the **XGBoost model ('xgb1')** without SMOTE, using a **0.4 cut-off**, remains a viable option due to its **higher TPR** performance.