Project Proposal
Arwa Abdulrahman AlBassam | Nouf I. Alsaeed
30/9/2021

أكاديمية سدايا
SDAIA Academy

Comparison of classifier Algorithms on bank marketing Dataset

Overview:

A bank institution located in Portuguese did a marketing campaign, where the marketing campaigns is based on phone calls, and sometimes more than one contact to the same client is required, in order to access if the product (bank term deposit) would be (or not) subscribed, using the data provided by the bank we are going to build a classification model to predict if the client will subscribe a term deposit with a comprehensive analysis with all data cleaning, exploration, visualization, feature selection, model building, and evaluation.

Dataset Description:

Using the data provided by the bank with the below features:

FEATURE	DESCRIPTION
age	Age of clints
job	Type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur","student", "blue-collar", "self-employed", "retired", "technician", "services")
marital	Marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
education	(Categorical: "unknown", "secondary", "primary", "tertiary")
default	Has credit in default? (Binary: "yes", "no")
balance	Average yearly balance, in euros (numeric)
housing	Has housing loan? (Binary: "yes", "no")
loan	Has personal loan? (Binary: "yes", "no")
contact	Contact communication type (categorical: "unknown", "Telephone", "cellular")
day	Last contact day of the month (numeric)
month	Last contact month of year (categorical: "Jan", "Feb", "Mar", …, "Nov", "Dec")
duration	Last contact duration, in seconds (numeric)
campaign	Number of contacts performed during this campaign and for this client (numeric, includes last contact)

pdays	Number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
previous	Number of contacts performed before this campaign and for this client (numeric)
poutcome	Outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")
y	Has the client subscribed to a term deposit? (Binary: "yes","no")

Tools:

- Pandas and NumPy packages to manipulate data.
- Matplotlib library for visualizing data.
- LogisticRegression model from sklearn.linear_model class to build a classification algorithm that is used to predict if the client will subscribe.
- train_test_split function in Sklearn model selection for splitting data.
- KNeighborsClassifier model from sklearn.neighbors to build a classification algorithm that is used to predict if the client will subscribe.
- DecisionTreeClassifier model from sklearn.tree to build a classification algorithm that is used to predict if the client will subscribe.
- RandomForestClassifier model from sklearn.ensemble to build a classification algorithm that is used to predict if the client will subscribe.
- Measure performance of each algorithm using precision_score, recall_score, accuracy_score, roc_auc_score and confusion_matrix from sklearn.metrics module.
- Jupyter notebook to execute the code.

Conclusion:

In the banking field, a huge amount of data is generated continuously, and this data can be used to extract meaningful information. In this project, we want to predict whether a customer will subscribe or not using multiple machine learning algorithms and compare the performance of each algorithm.