Data Co-Lab Engineering Lab Test

Given various Information related to customer's behavior in online shopping websites, we want to create a model to predict whether this customer will generate revenue or not.

The steps of this project are:

- Data Analysis (with an interpretation of the results shown in the figures)
- Data cleaning and preprocessing
- Training and testing some machine learning models (stating the different possibilities, choosing two or three of them and explaining your choice)
- Training and testing a deep a learning model

You must send a folder containing a jupyter notebook file* and a requirements file** to the lab email address (data.colab.tn@gmail.com) before Monday, 26th August 2019 at 10 pm. The folder's name will be in this format: yourName_yourLastName_engineering_test

*: The jupyter notebook file must contain the commented code: when choosing a certain methodology or a certain algorithm. Name the different possibilities and explain your choice.

**: The requirements file is a .txt file and must contain the different requirements along with their versions.

The python version must be equal to or higher than 3.6

Information about the dataset:

The dataset consists of 12,330 entries, each one representing a session (belonging to a user) in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period. The dataset and the information about it are taken from Kaggle. However, changes may have been applied.

Information about the attributes:

The dataset consists of 10 numerical and 8 categorical attributes.

- **Revenue:** the class label.
- **Administrative**: the number of the administrative pages visited by the user in that session.

- **Administrative Duration**: the total time spent by the user in administrative pages
- **Informational**: the number of the informational pages visited by the user in that session.
- **Informational Duration**: the total time spent by the user in informational pages.
- **Product Related**: the number of the product related pages visited by the user in that session.
- **Product Related Duration**: the total time spent by the user in product related pages.
- **Bounce Rate** (of a web page): the percentage of users that entered and left (bounced) the website through this same webpage
- **Exit Rate** (of a web page): the percentage of pageviews for this web page that were the last in the session
- **Page Value** (of a web page): the average value for a web page that a user visited before completing an e-commerce transaction
- **Special Day:** the closeness of the visiting date to a special day (like Mother's Day, Valentine's Day). The values are between 0 and 1, 0 being far from any special day and 1 being closest to a special day.
- **operating system:**
- **browser:** the operating system used
- **region:** the region of the user
- **traffic type:** the traffic type
- **visitor type:** the visitor type (new visitor or returning visitor
- **weekend:** whether the date of the visit is weekend or not
- **month of the year:** the month of the year of the visit