Practice Assignment 1.2 : Classification

This Assignment deals with the following topics:

- Classification Models
- Logistic Regression

The Assignment

In this assignment, you will explore the Online Shoppers Purchasing Intention dataset to build a logistic regression model that predicts whether a visitor will make a purchase. You will analyze data in the online_shoppers_intention.csv file and use functions from the pandas module for loading, inspecting, and querying the data. You are expected to preprocess the data and use the data to train machine learning models from the 'sklearn' library to solve problems.

To begin this assignment, we have provided you with an .ipynb file in Codio, which contains the instructions as inline comments.Details about the dataset are described in the following sections.

For each question, there are clear instructions in each cell. Follow those instructions and write the code after each block of:

```
# YOUR CODE HERE
raise NotImplementedError()
```

Make sure to delete the line raising an error! Please use the exact variable name if it is specified in the comment.

We'll run a Python test script against your program to test whether the code in each cell is correct.

Dataset Overview:

We will use the Online Shoppers Purchasing Intention dataset for practice assignments 2 and 3. The link to this dataset is shown here:

https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset

The dataset consists of 10 numerical features and 8 categorical features totaling 18 input variables. Some of the key features are listed below:

- **Administrative, Administrative_Duration:** access statistics related to administrative pages
- **Informational, Informational_Duration:** access statistics related to informational pages
- **ProductRelated, ProductRelated_Duration:** access statistics related to product-related pages
- **BounceRates, ExitRates and PageValues:** metrics related to page values
- **SpecialDay:** proximity of visits to a specific special day
- **Month:** the month of the visit
- **OperatingSystems, Browser, Region, TrafficType:** technical information about the user's operating system, browser type, region and traffic type
- **VisitorType:** new visitor, returning visitor or other type
- **Weekend:** whether the visit occurred on a weekend
- **Revenue:** target variable indicating whether the visitor purchased a product

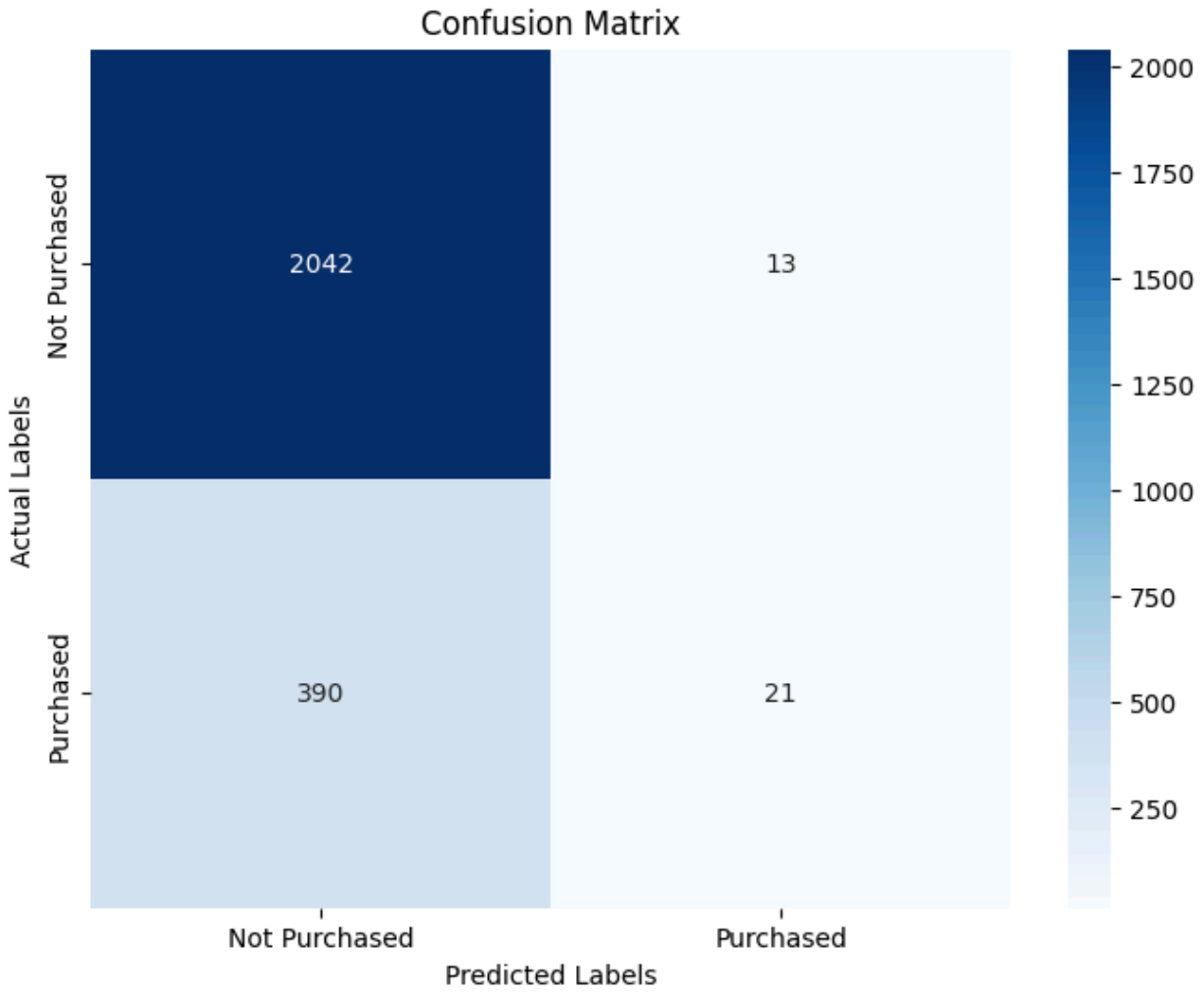
Administrative	Administrative_Durat	Informational	Informational_Durati	ProductRelated	ProductRelated_Dur	BounceRates	ExitRates	PageValues	SpecialDay	Month	OperatingSystems
0	0	0	0	1	0	0.2	0.2	0	0	Feb	1
0	0	0	0	2	64	0	0.1	0	0	Feb	2
0	0	0	0	1	0	0.2	0.2	0	0	Feb	4
0	0	0	0	2	2.666666667	0.05	0.14	0	0	Feb	3
0	0	0	0	10	627.5	0.02	0.05	0	0	Feb	3
0	0	0	0	19	154.2166667	0.015789474	0.024561404	0	0	Feb	2

What to submit

Before submitting, click “Validate” to check if your answers are right. When you are satisfied with your answers, in Codio go to the “Education” menu and select “Mark as Completed.”

Sample Plots:

Here is the sample plot for Part 06:


Confusion Matrix
Not Purchased
Purchased
Actual Labels
2042
13
390
21
Not Purchased
Purchased
Predicted Labels
2000
1750
1500
1250
1000
750
500
250