

Practical 5

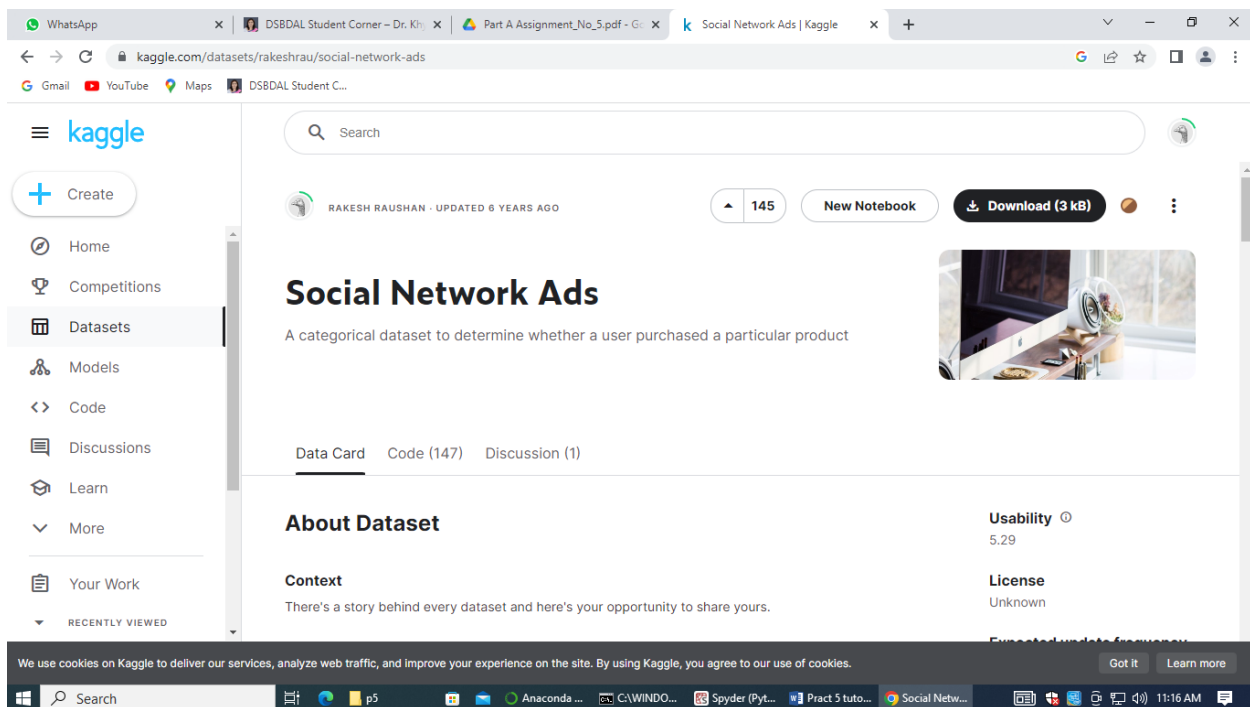
Tutorial

In the 5th practical we will be studying logistic regression and sigmoid function. The theory can be found in the lab manual which is provided.

Here we will be using an external data set from kaggle.

We will be using the the Social_Media_Adv Dataset.

Link on kaggle: <https://www.kaggle.com/datasets/rakeshrau/social-network-ads>



If the data set is in zipped format unzip it.

After downloading the database head over to anaconda and launch spyder.

To create a logistic regression model follow these Steps :

Step 1: Import libraries and create alias for Pandas, Numpy and Matplotlib

```
import pandas as pd
```

Step 2: Import the Social_Media_Adv Dataset

```
df = pd.read_csv(r"C:\Users\comp\Documents\sushant\Social_Network_Ads.csv")
```

Step 3: Initialize the data frame

```
df['Gender']
```

Step 4: Perform Data Preprocessing

- **Convert Categorical to Numerical Values if applicable**

- **Check for Null Value**

```
df.isnull()
```

- **Covariance Matrix to select the most promising features**

```
df.dtypes
```

```
df['Gender'] = df['Gender'].map({'Male': 1, 'Female': 0})
```

```
df['Gender']
```

- **Divide the dataset into Independent(X) and Dependent(Y) variables.**

```
x = df.drop(['Age'],axis=1)
```

```
y= df['Age']
```

- **Split the dataset into training and testing datasets**

```
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.25, random_state=0)
from sklearn.preprocessing import StandardScaler
st_x= StandardScaler()
xtrain= st_x.fit_transform(xtrain)
xtest= st_x.transform(xtest)
```

- **Scale the Features if necessary.**

Step 5: Use Logistic regression (Train the Machine) to Create Model

```
from sklearn.linear_model import LogisticRegression
classifier= LogisticRegression(random_state=0)
classifier.fit(xtrain, ytrain)
```

Step 6: Predict the y_pred for all values of train_x and test_x

```
y_pred= classifier.predict(xtest)
y_pred
```

Step 7: Evaluate the performance of Model for train_y and test_y

Step 8: Calculate the required evaluation parameters

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(ytest, y_pred)
cm
```

In this way we have done data analysis using logistic regression for Social Media Adv. And evaluate the performance of model.

Lastly save the file and create a text doc containing the final code and output.