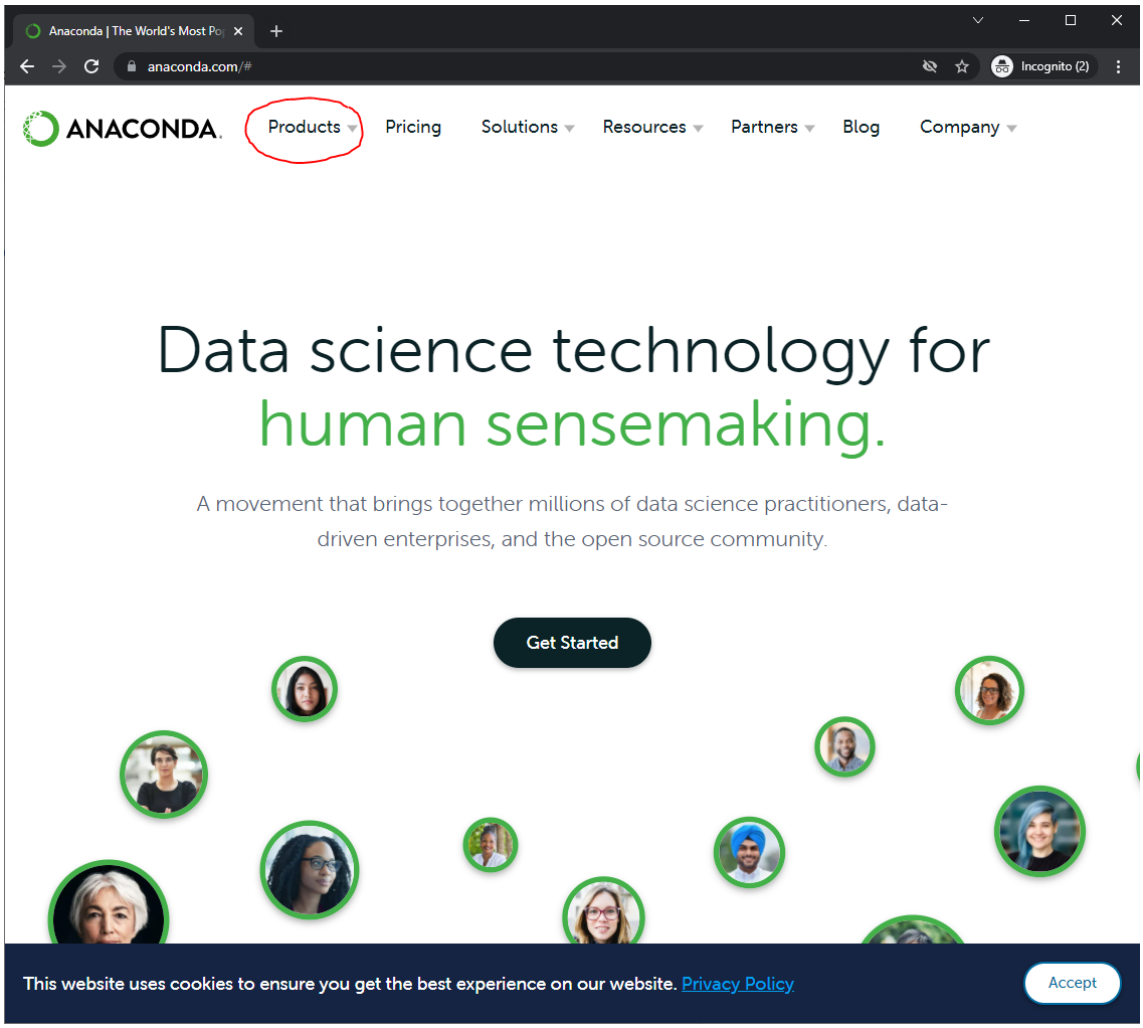
Text Classification

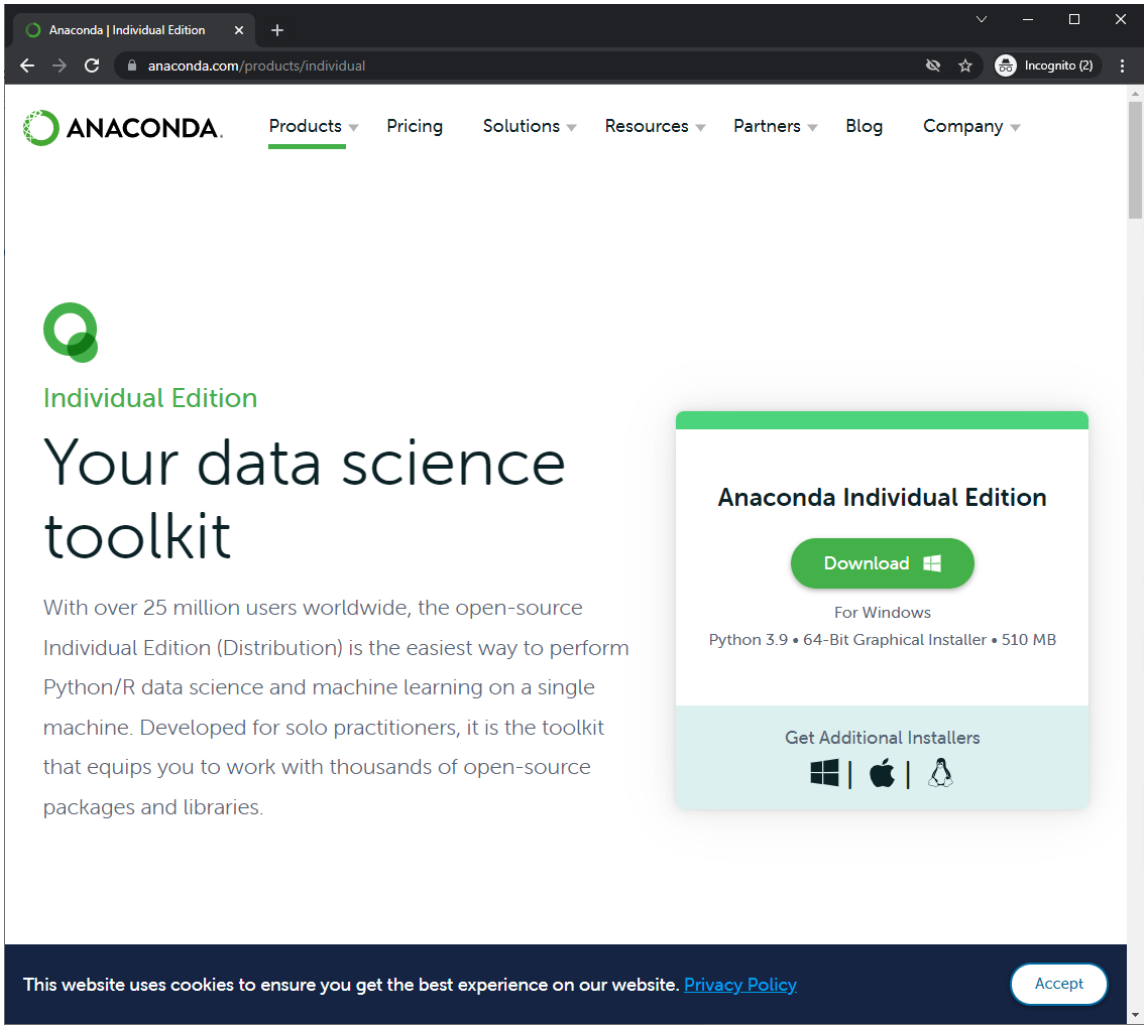
Download the latest version of Anaconda Individual edition.

https://www.anaconda.com/



Click on Products dropdown and select Individual edition.

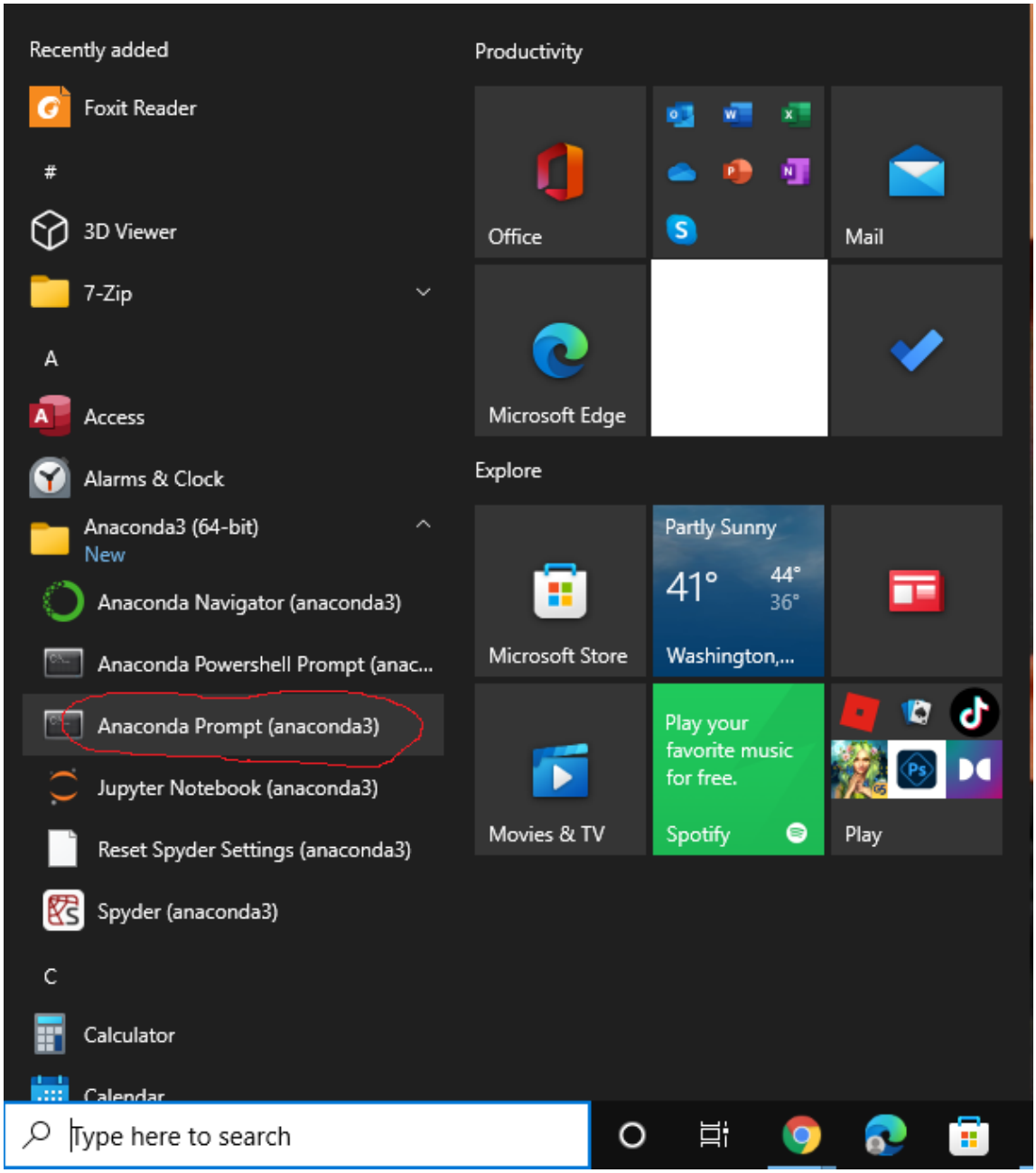
If anaconda3 is already installed then go to the section (page #3) for installing steamlit 0.84.2 version.



Click on Download.

After the software is downloaded, click on the .exe file to install.

Once the software is installed, click on the start button and look for “Anaconda Prompt (anaconda3)” within Anaconda3 (64-bit)



Click on the Anaconda Prompt (anaconda3). It opens up a command line window.

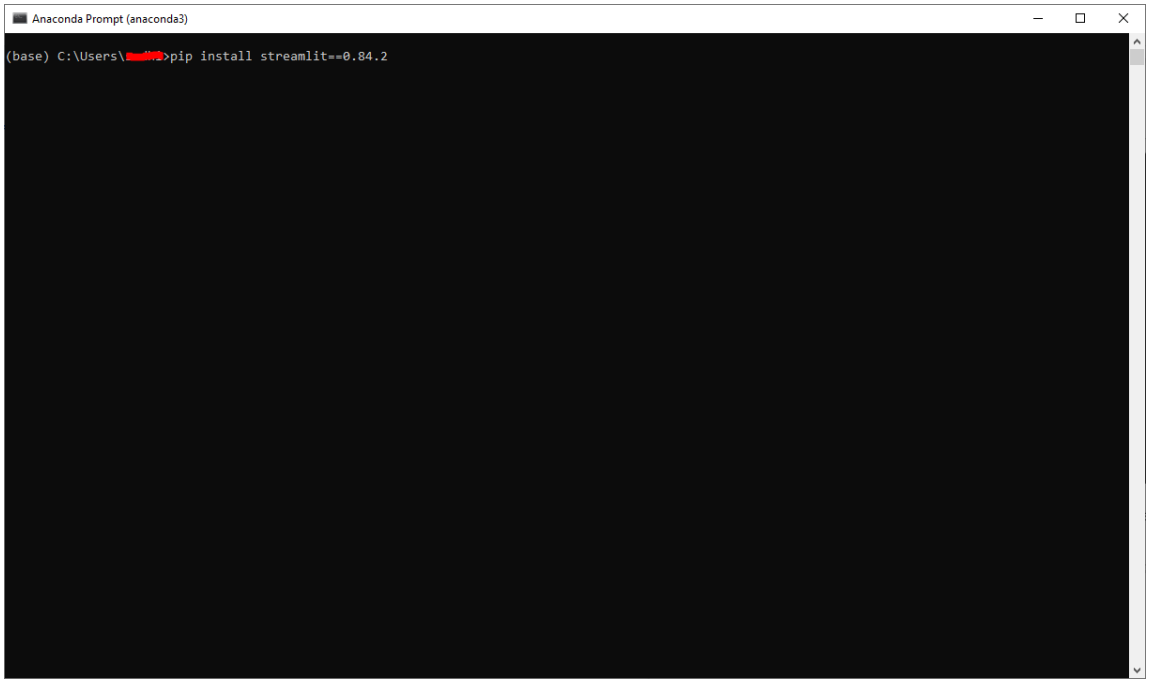
Now do a pip install to install steamlit version 0.84.2

pip install streamlit==0.84.2

If streamlit is already installed ensure it is version 0.84.2 else install 0.84.2 as code requires this version. Installation removes any prior version and then installs 0.84.2

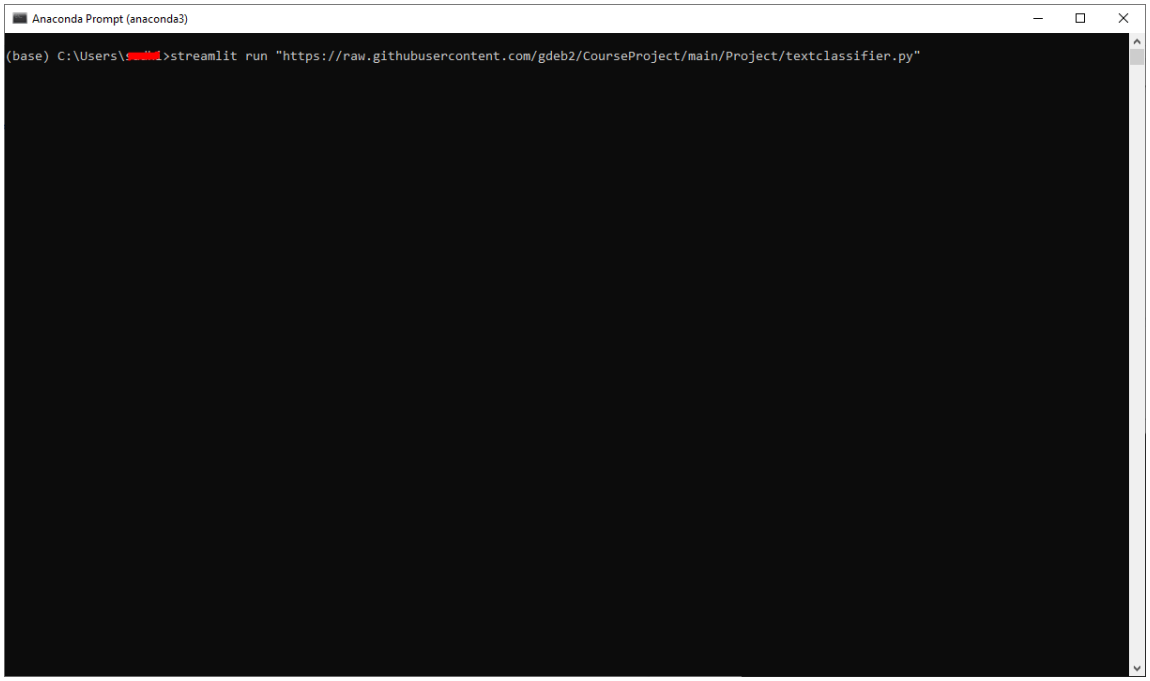
Please ensure the below packages are installed. (The latest version of Anaconda comes with these packages installed)

sklearn
matplotlib
pandas

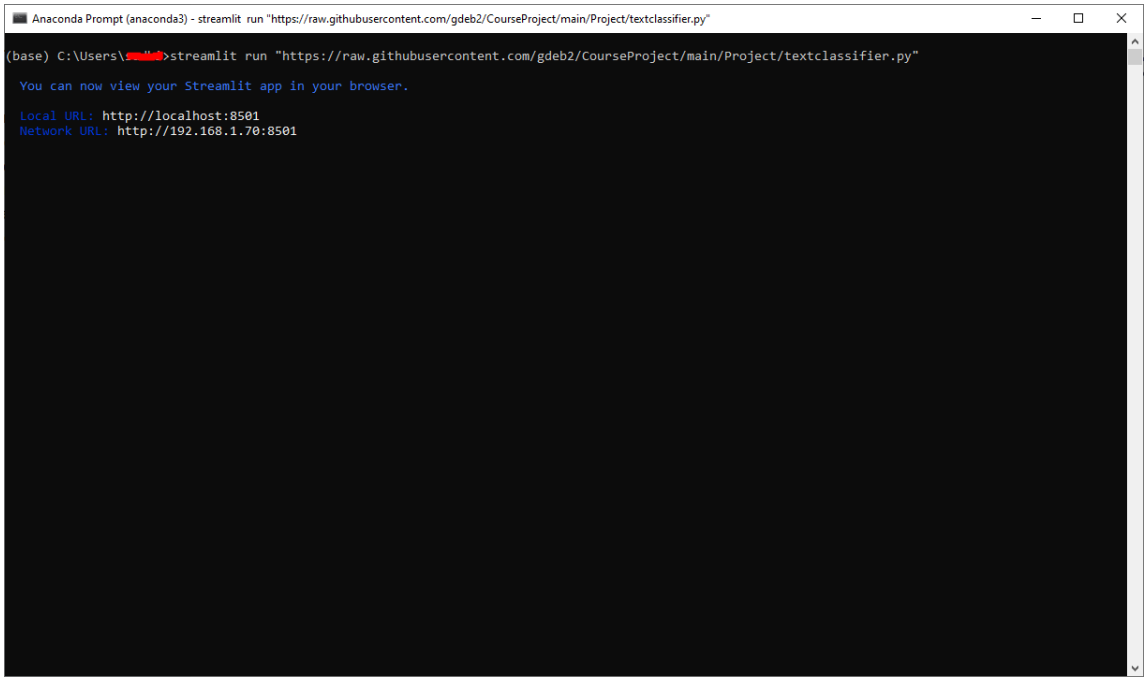


After installing streamlit, execute the below command to run the “Text Classifier” project.

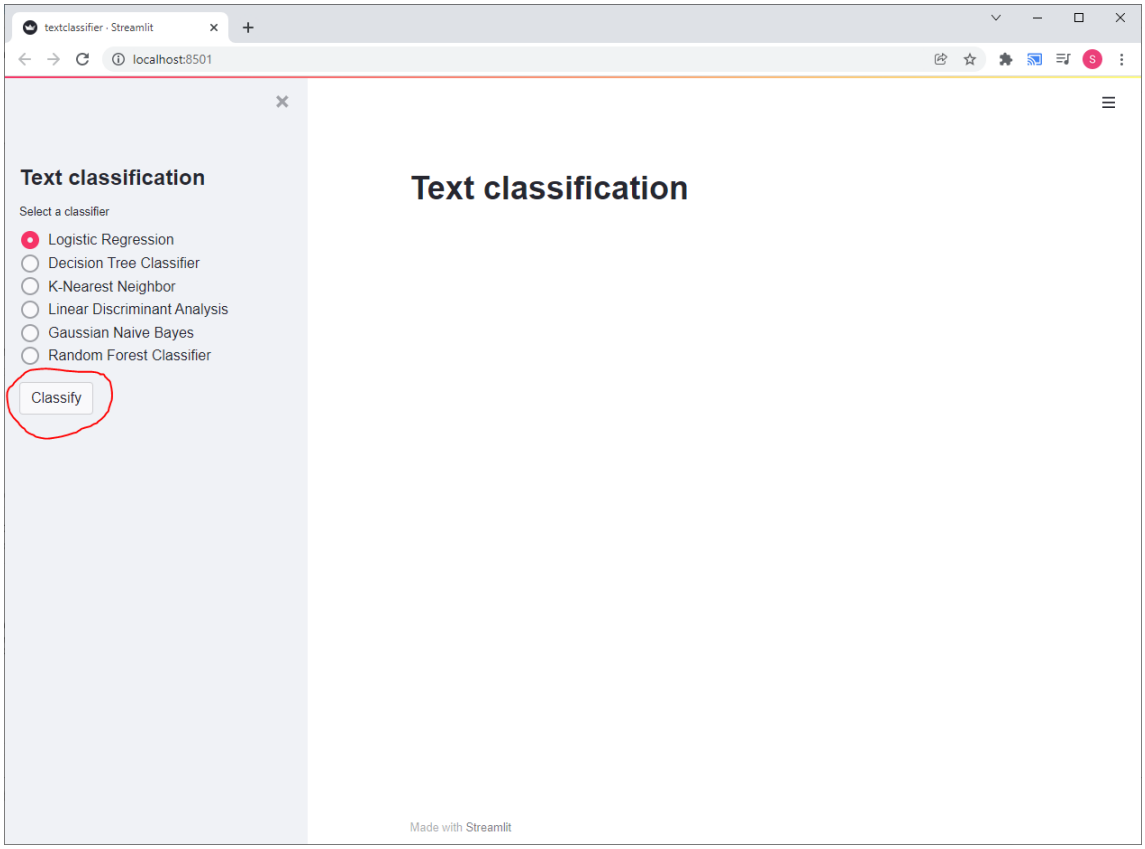
streamlit run “https://raw.githubusercontent.com/gdeb2/CourseProject/main/Project/textclassifier.py”



Once the streamlit run command is executed, Local and Network URL will be displayed as shown below.

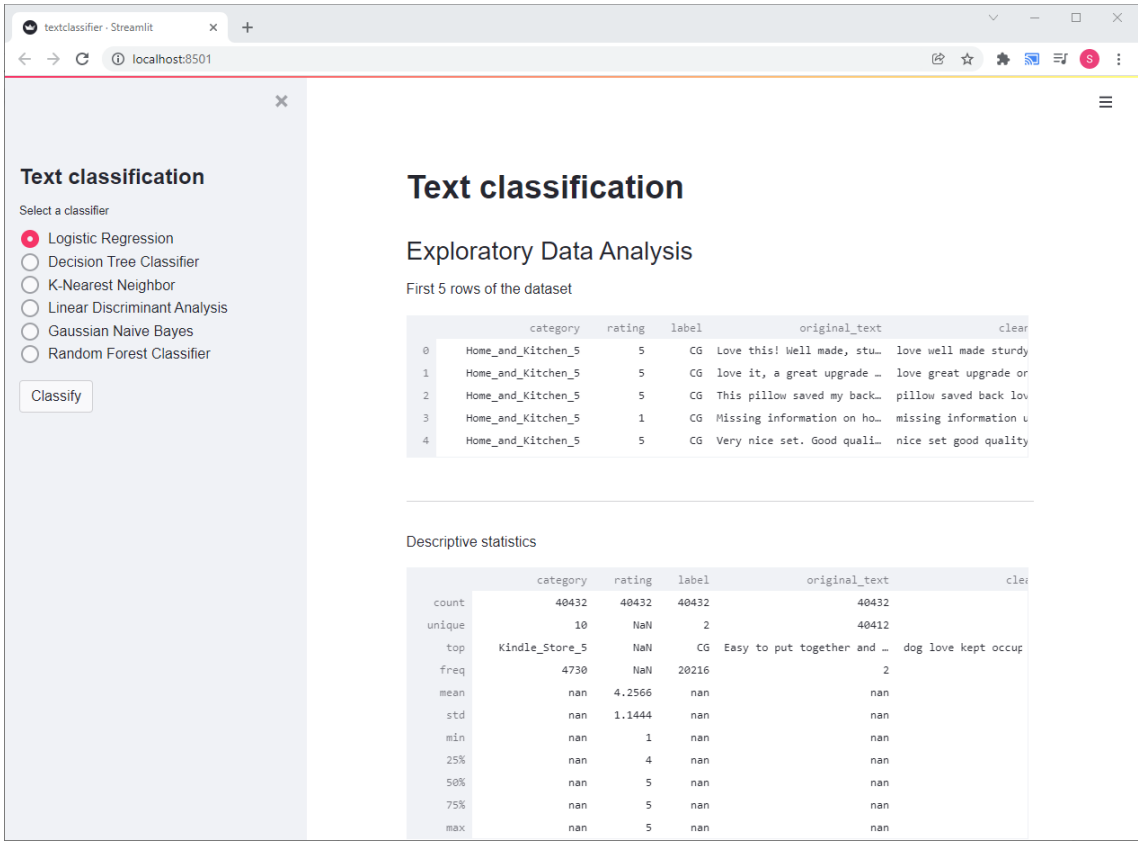


A new window opens up Text Classification.



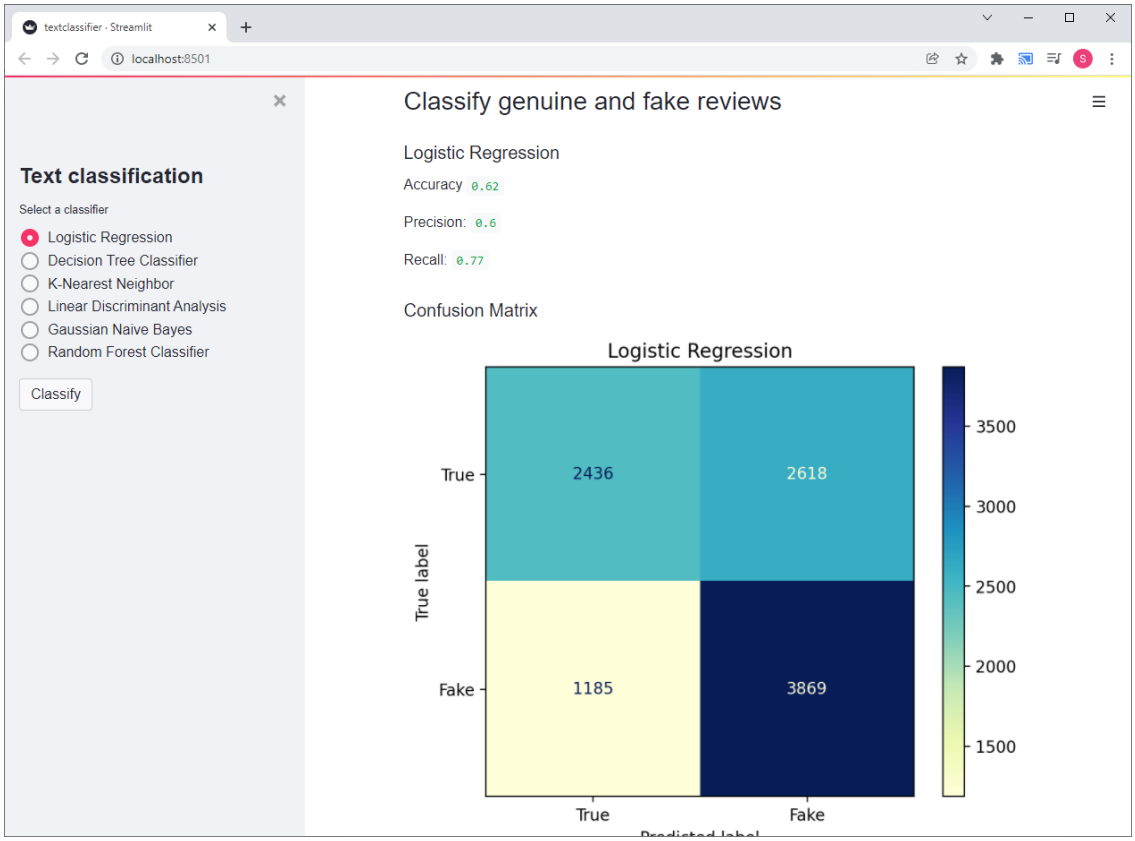
This is how the main screen looks like. The left panel lists all the classification algorithms implemented.

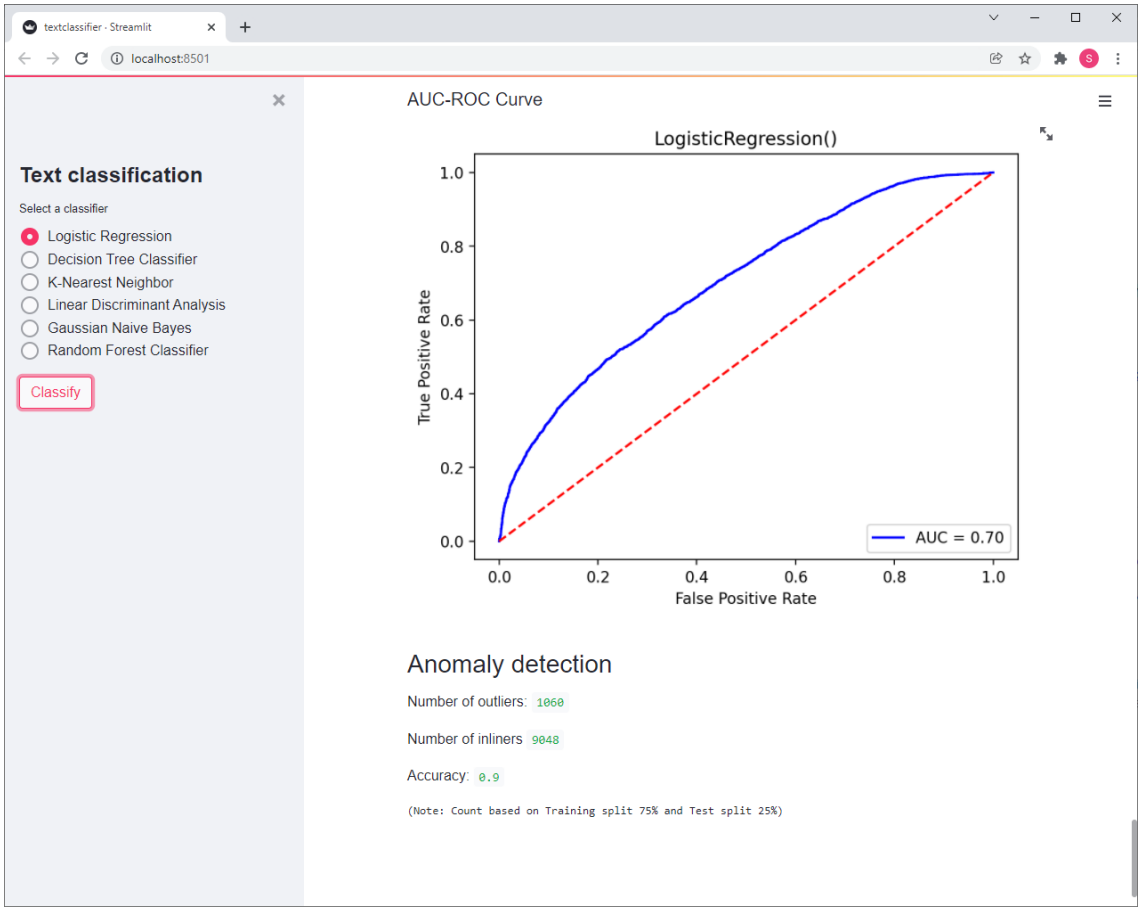
Check the radio button to select an algorithm and click the "***Classify***" button.



Exploratory Data Analysis (EDA), Model metrics (Accuracy, Precision and Recall), Confusion Matrix and ROC curve for the selected algorithm will be displayed on the right side as shown in the above screenshot. The selected algorithm will train the model prior to displaying the results.

Scroll down the screen to view all the information.


textclassifier · Streamlit
localhost:8501
Text classification
Select a classifier
Logistic Regression
Decision Tree Classifier
K-Nearest Neighbor
Linear Discriminant Analysis
Gaussian Naive Bayes
Random Forest Classifier
Classify
Classify genuine and fake reviews
Logistic Regression
Accuracy 0.62
Precision: 0.6
Recall: 0.77
Confusion Matrix
Logistic Regression
True
Fake
True label
2436
2618
1185
3869
3500
3000
2500
2000
1500
True
Fake



After reviewing the EDA, model metrics, confusion matrix and ROC curve, repeat the above step for all other algorithms.

To exit, click the “x” button on the top right corner.