Developing an NLP Model for Efficient Summarization of Legal and Financial Documents

In this project we are employing Natural Language Processing (NLP) algorithm to generated summary from given text data. Summary contains main topic from large text data. NLP will remove special symbols and stop words from large text and then calculate frequency of each sentences and then select all those sentences which has highest similarity score.

To train NLP model we have utilize legal document dataset which can be downloaded from below URL

<https://github.com/lauramanor/legal_summarization/blob/master/tldrlegal_v1.json>

Above dataset is in JSON format which contains TEXT data and its summary and NLP will get trained on above dataset and then evaluate performance by using various metrics such as Recall, FMEASURE and precision.

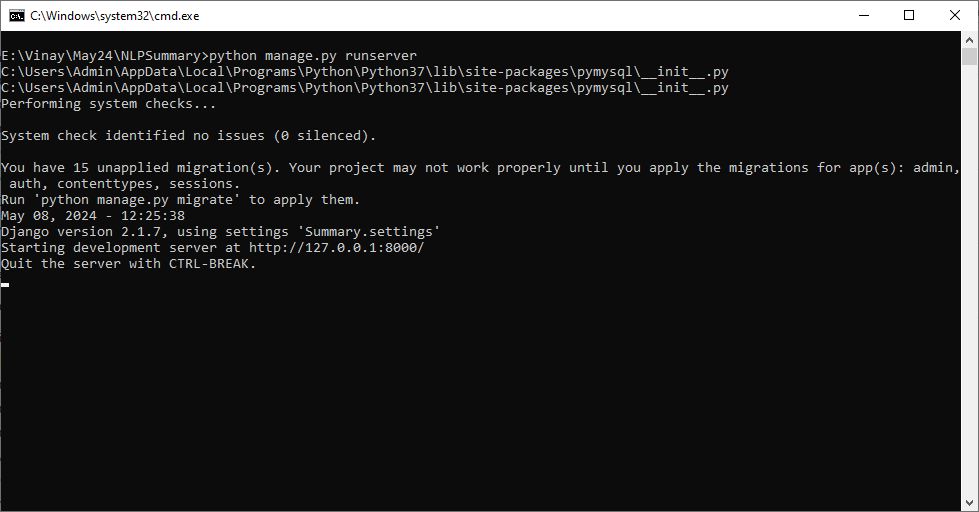
To implement this project we have designed following modules

1. New User Sign up: using this module user can sign up with the application
2. User Login: after sign up user can login to system
3. Train NLP Summary Model: after login user can utilize this model to trained NLP algorithms on Legal dataset and then perform prediction on test data to calculate precision and other metrics.
4. Generate Summary: using this module user can input sample text and then NLP will generate from given text data.

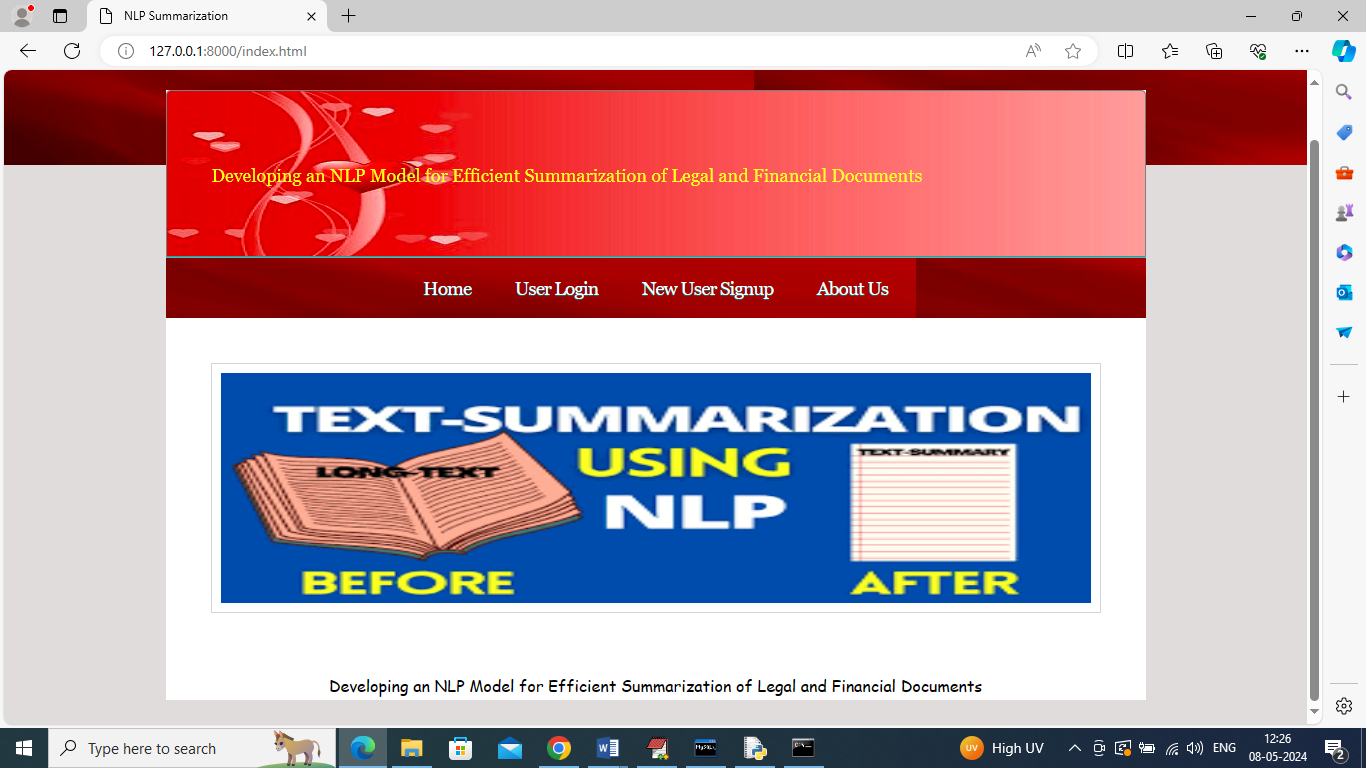
SCREEN SHOTS

To run project install MYSQL database and then copy content from DB.txt and paste in MYSQL console to create database.

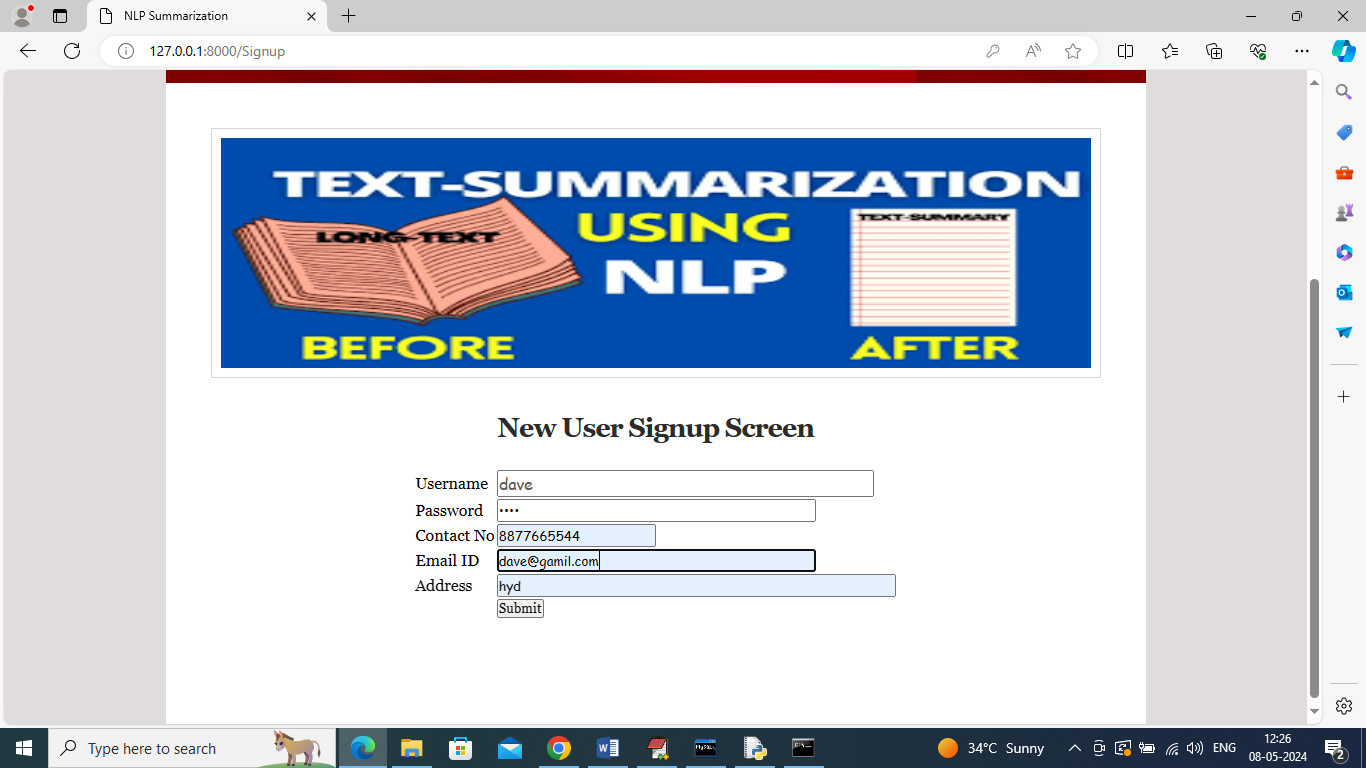
Install python 3.7 and then install all packages given in requirement.txt file and then double click on run.bat file to get below page



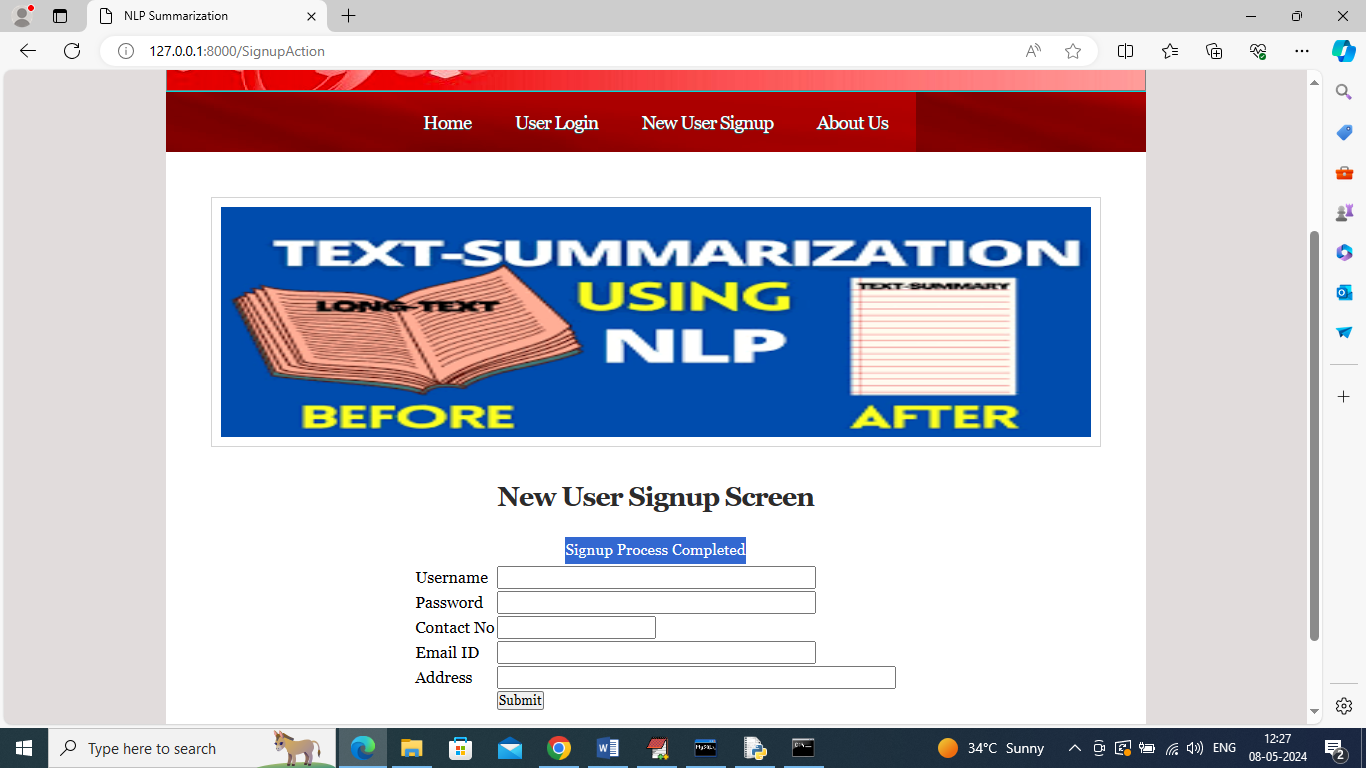
In above screen python server started and now open browser and enter URL as <http://127.0.0.1:8000/index.html> and press enter key to get below page



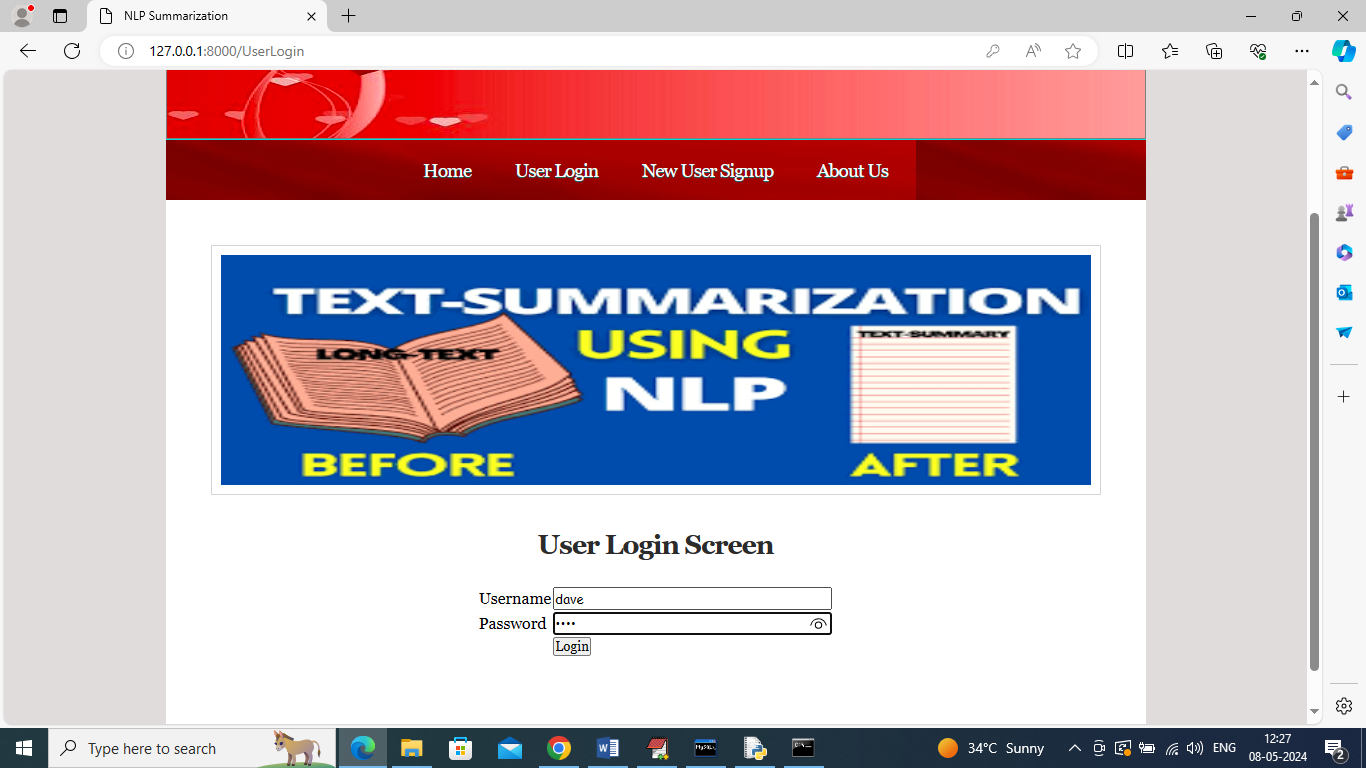
In above screen click on ‘New User Sign up’ link to get below page



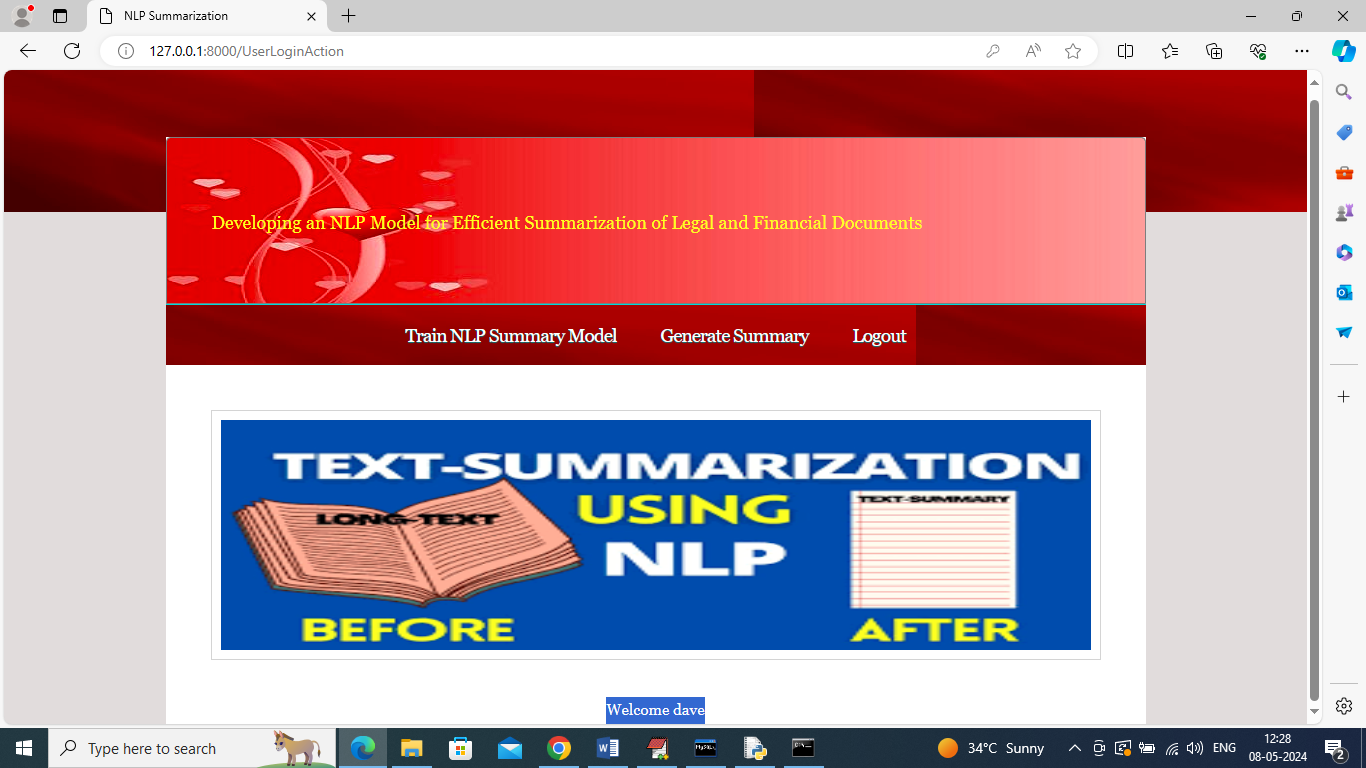
In above screen user is entering sign up details and then press button to get below page



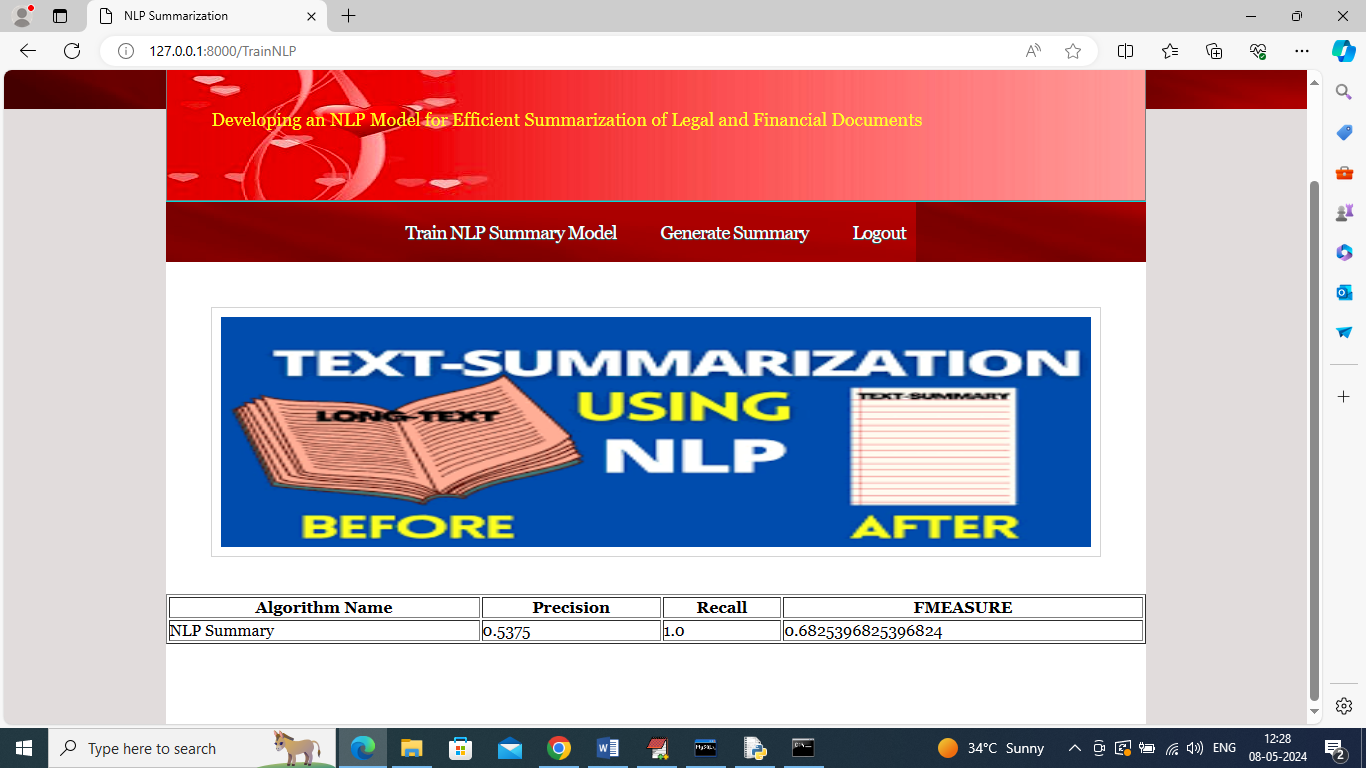
In above screen sign up task completed and now click on ‘User Login’ link to get below page



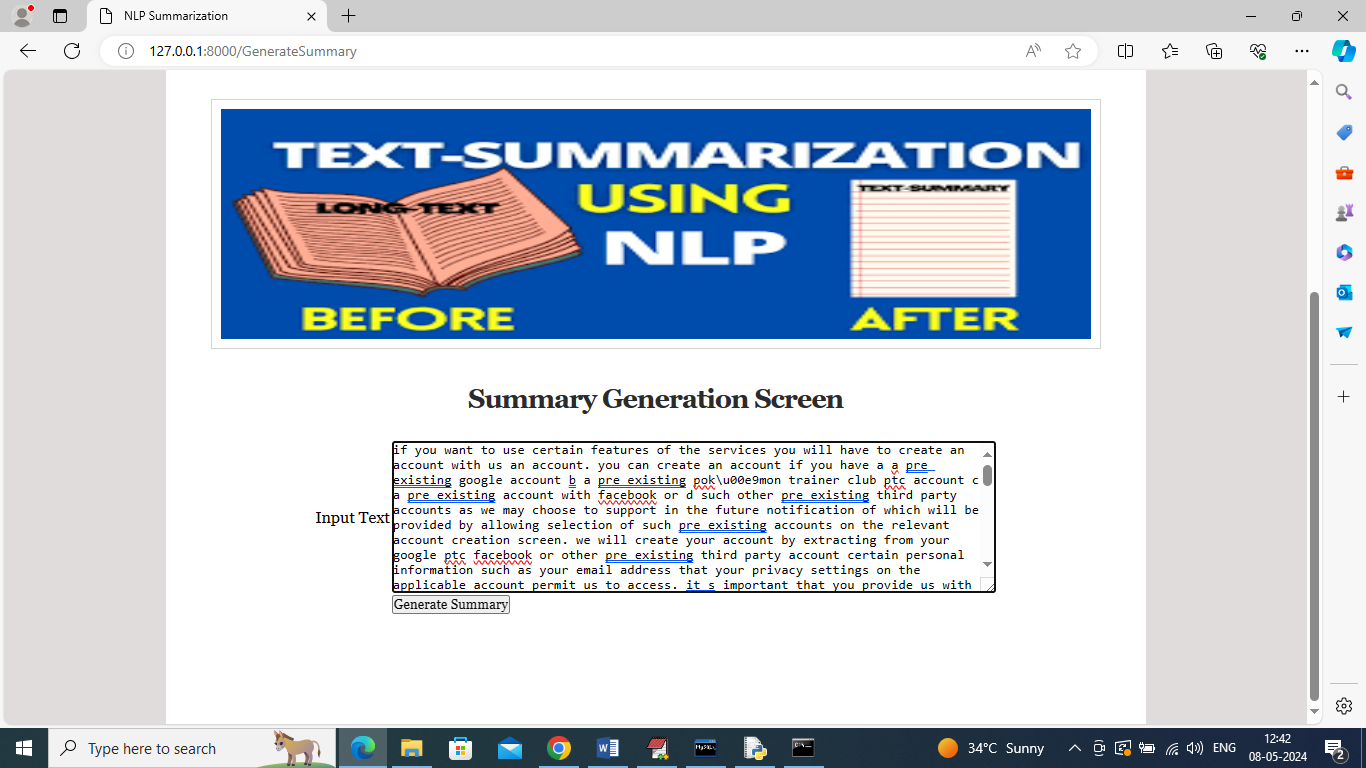
In above screen user is login and after login will get below page



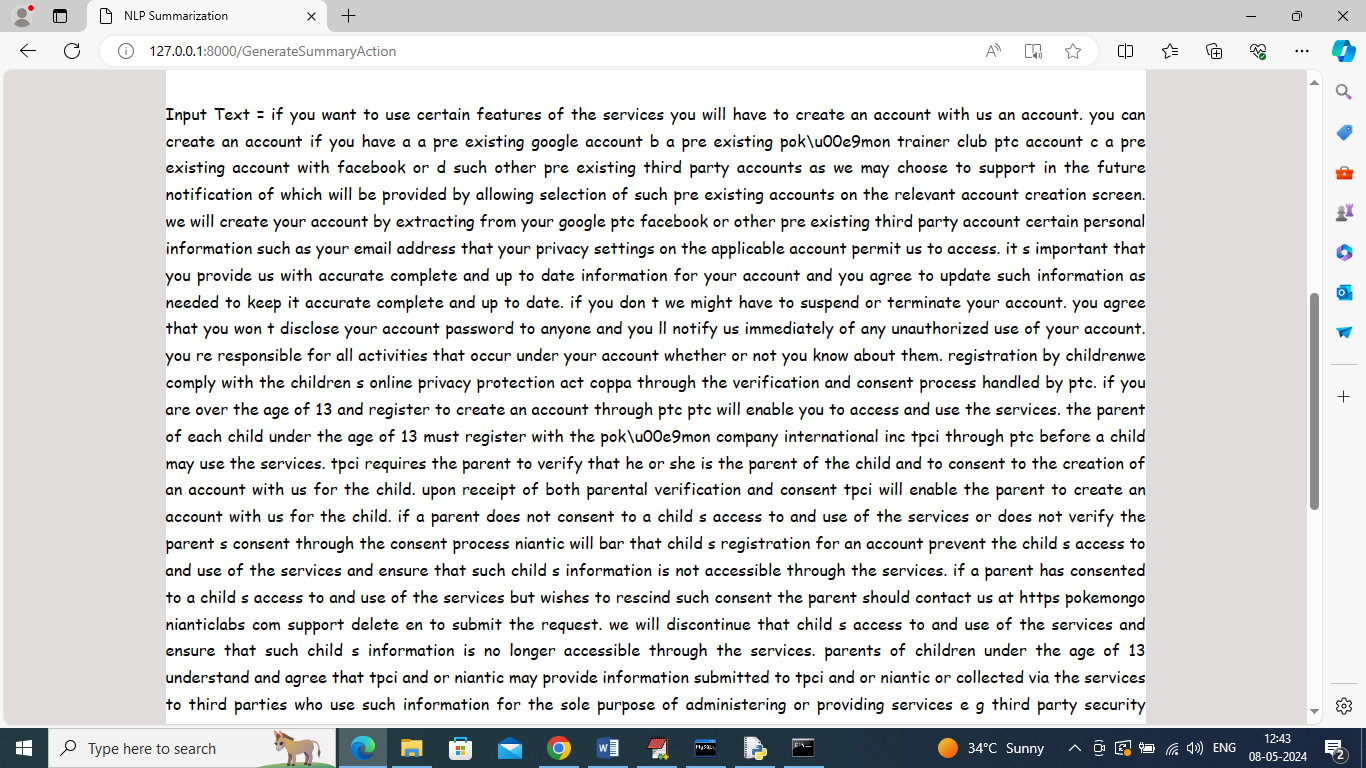
In above screen user can click on ‘Train NLP Summary Model’ link to train NLP algorithm on LEGAL dataset and get below output



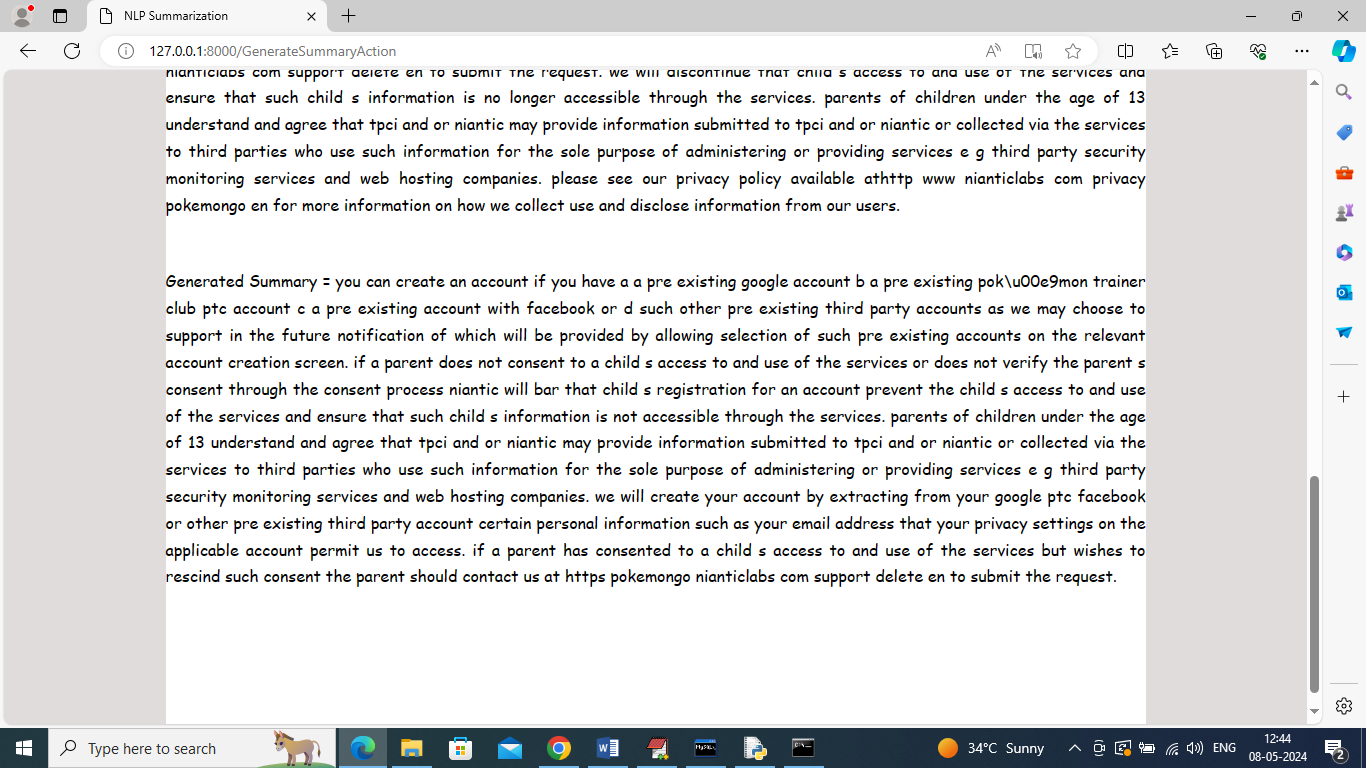
In above screen can see performance metrics of NLP on summary generation and got recall as 100%. Now click on “Generate Summary” link to get below page



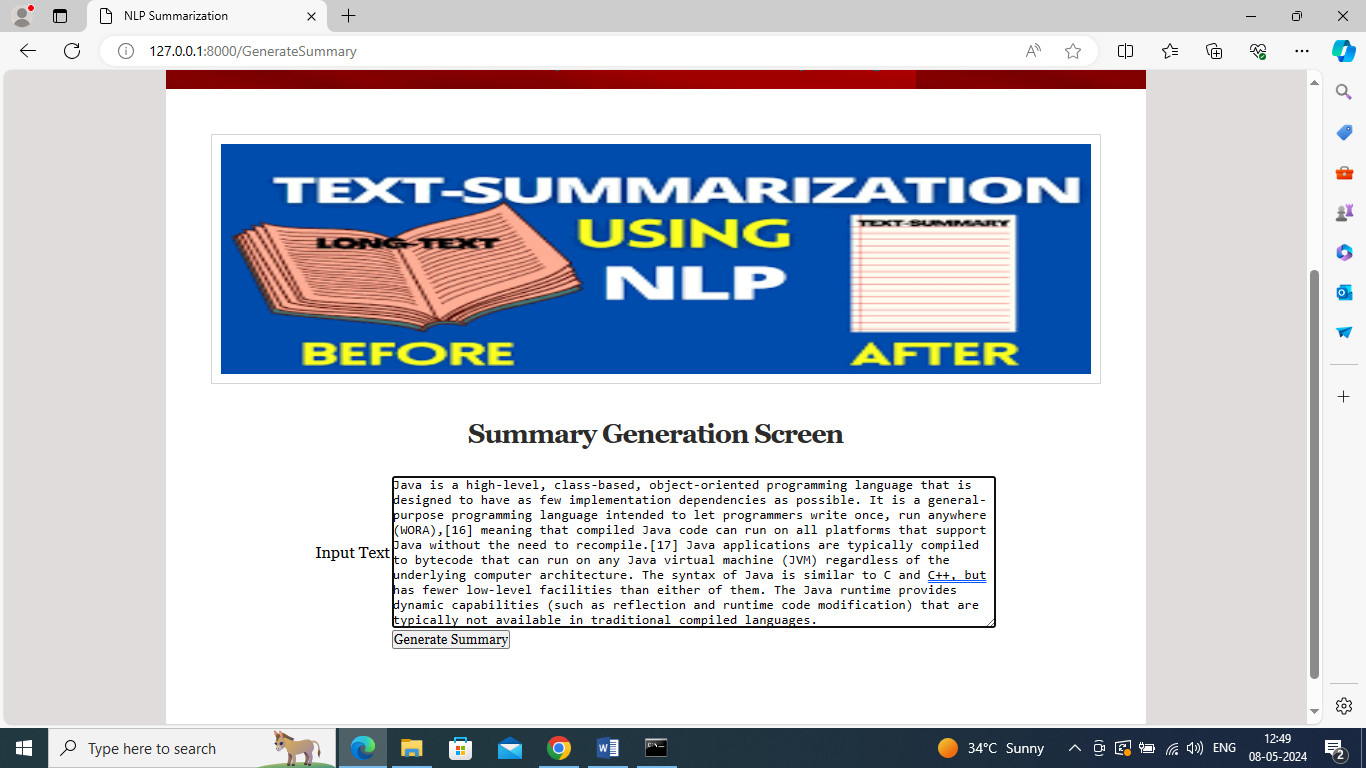
In above screen you can enter some text and then click on ‘Generate Summary’ button to get below summary output



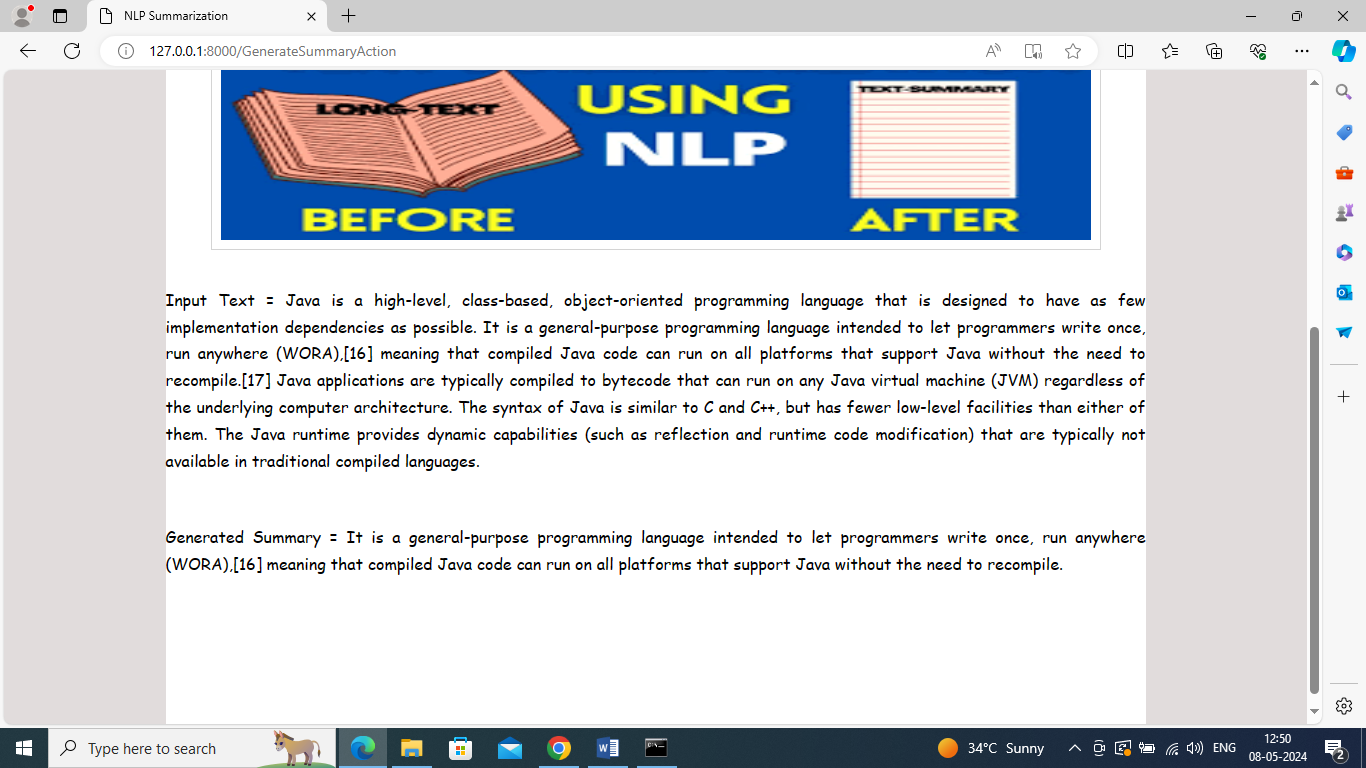
In above screen can see INPUT TEXT Data and below is the summary generated from above TEXT data



In above screen in second para can see generated summary from given long text and similarly you can give any text to get summary



In above screen entered some other text and then click button to get below output



In above screen first para is the original INPUT TEXT and second para contains generated summary