Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator

In this paper author is describing concept to detect fake news from social media or document corpus using Natural Language Processing and attribution supervised learning estimator. News documents or articles will be uploaded to application and then by using Natural Language Processing to extract quotes, verbs and name entity recognition (extracting organizations or person names) from documents to compute score, verbs, quotes and name entity also called as attribution. Using supervised learning estimator we will calculate score between sum of verbs, sum of name entity and sum of quotes divided by total sentence length. If score greater than 0 then news will be consider as REAL and if less than 0 then new will be consider as FAKE.

3 techniques will be used to calculate score

1. Source: any person who is writing news will give his name or a person name on which he writing articles
2. CUE: using this we will extract VERBS or VERBS phrases, if news is real then it will have verb types of words
3. Quotes: all articles will be on some topics and person will describe that topic name under quotes. So we will look for quotes in articles to determine fake or real news.

Document examples

"When “Mitt Romney” was governor of Massachusetts, we didnt just slow the rate of growth of our government, we actually cut it."

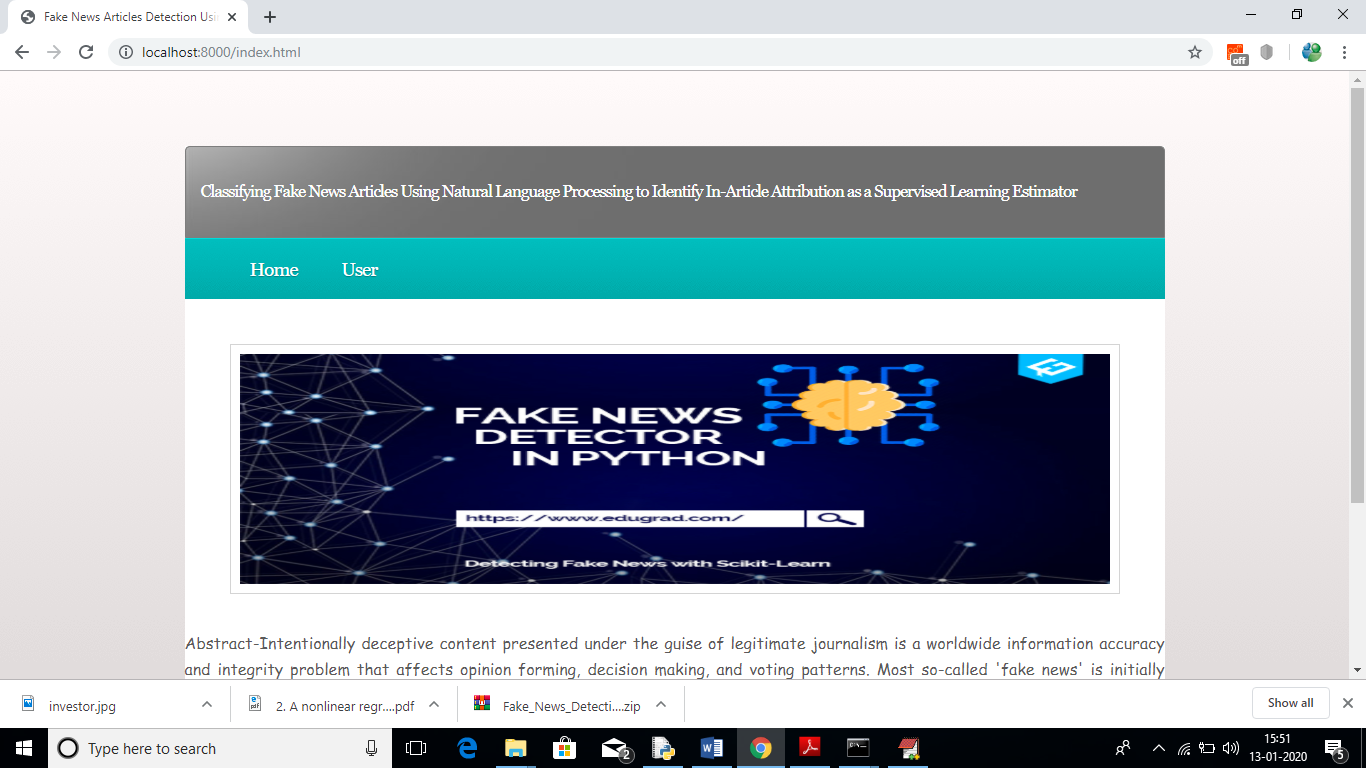
In above sentence quotes are there and it’s talking about ‘Mitt Romney’ and it’s contains some verbs such as ‘was, didn’t, slow, cut'. By analysing above 3 features from articles we can come to the conclusion whether news is FAKE or REAL.

All FAKE peoples will not write such statements in their articles so we can detect by applying this techniques.

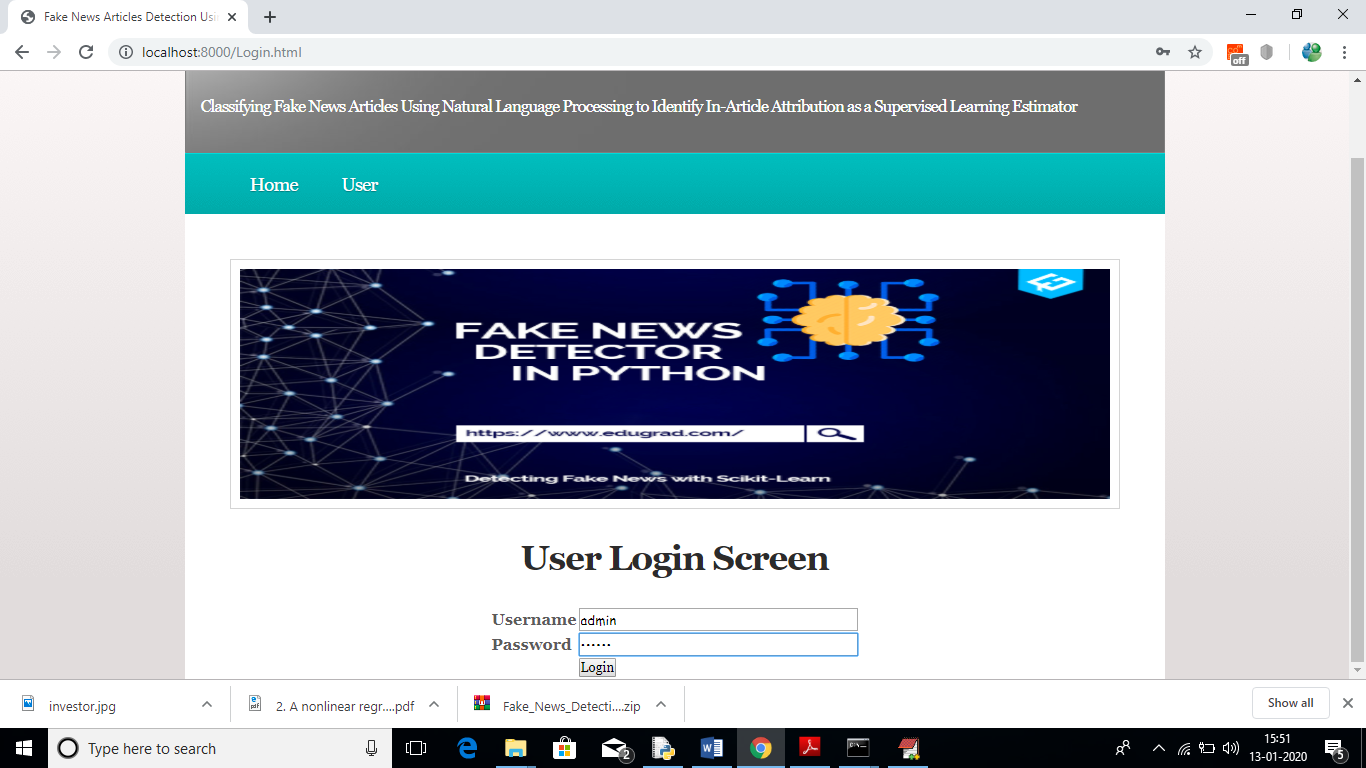
To implement this project we are using ‘News’ dataset and then by applying above technique we can detect whether this news are fake or real. This dataset I kept inside dataset folder. Upload this dataset when you are running application.

To run this project deploy ‘FakeNews’ folder on ‘django’ python web server and then start server and run in any web browser. After running code in web browser will get below page.

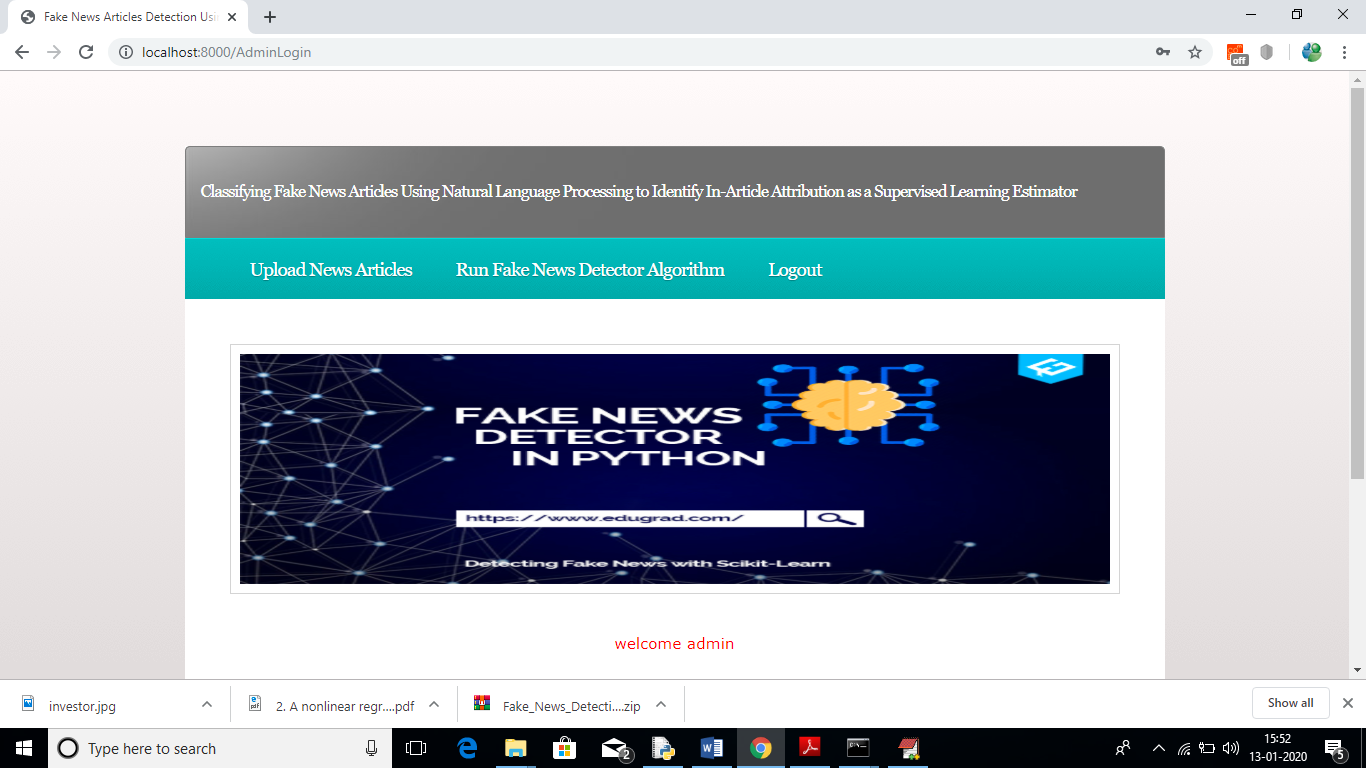
Screen shots



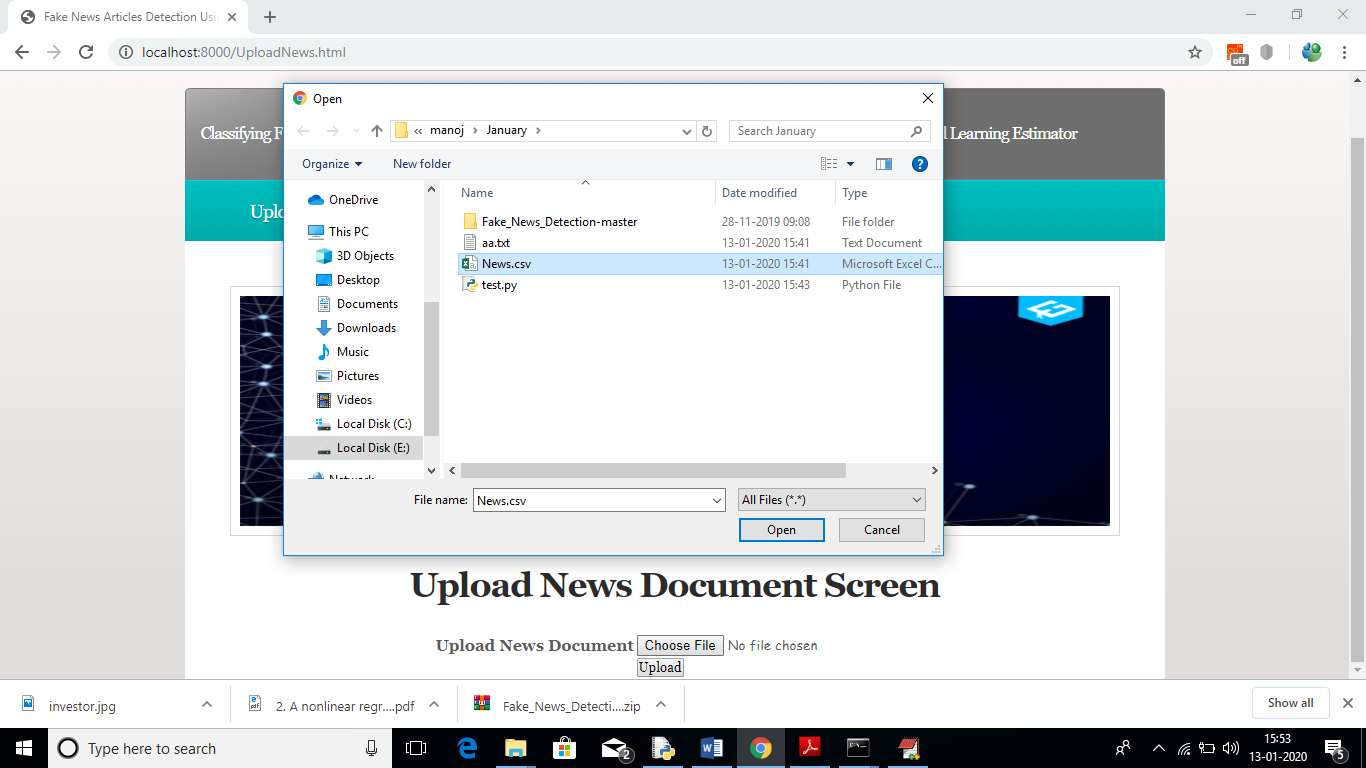
In above screen click on ‘User’ link to get below screen



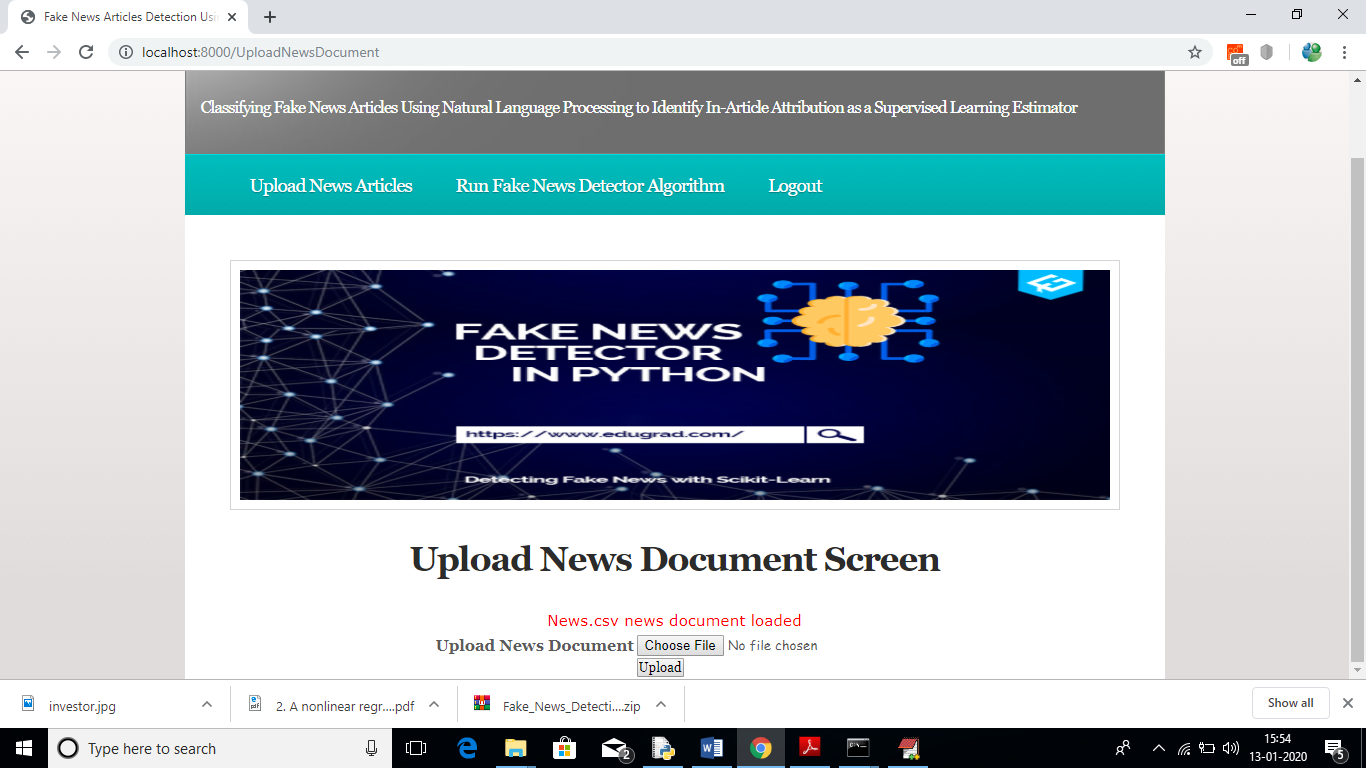
In above screen enter username and password as ‘admin’ and then click on ‘Login’ button to get below screen



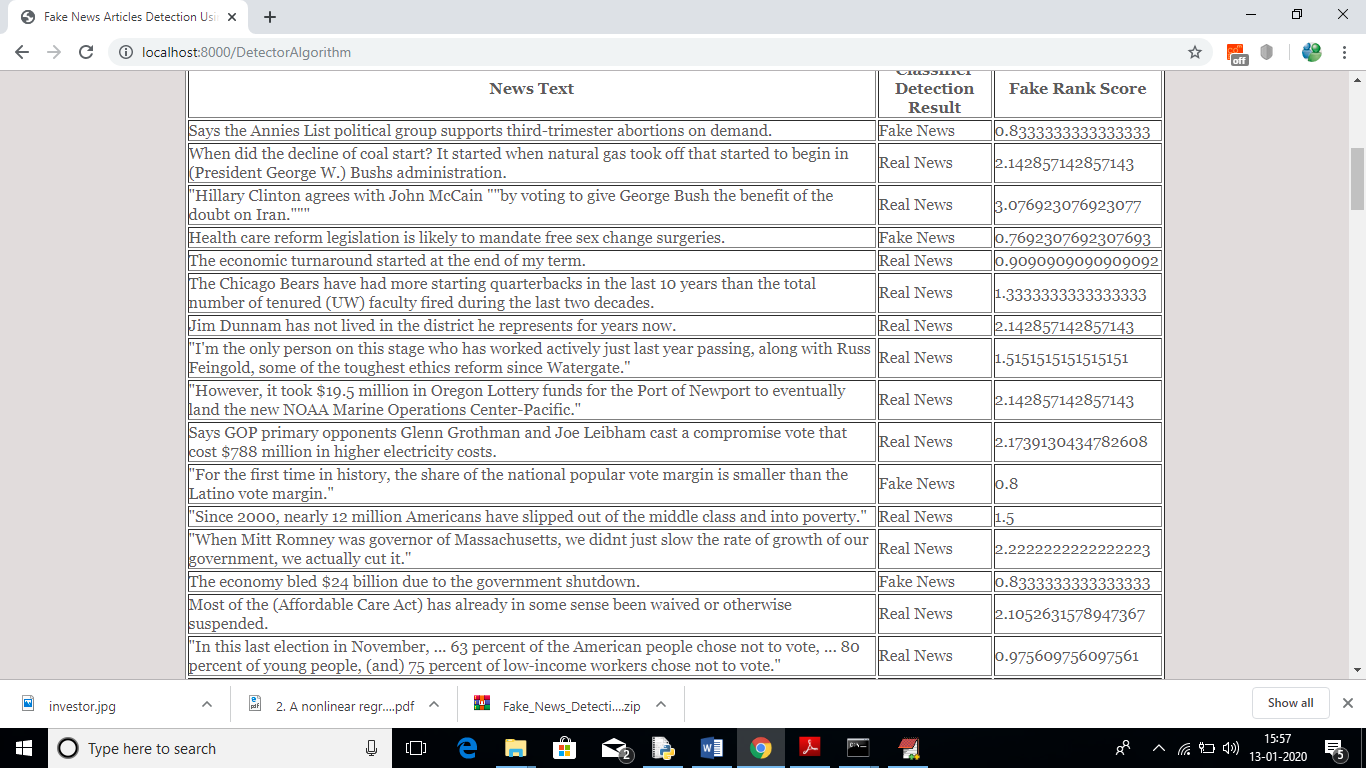
In above screen click on ‘Upload News Articles’ link to upload news document



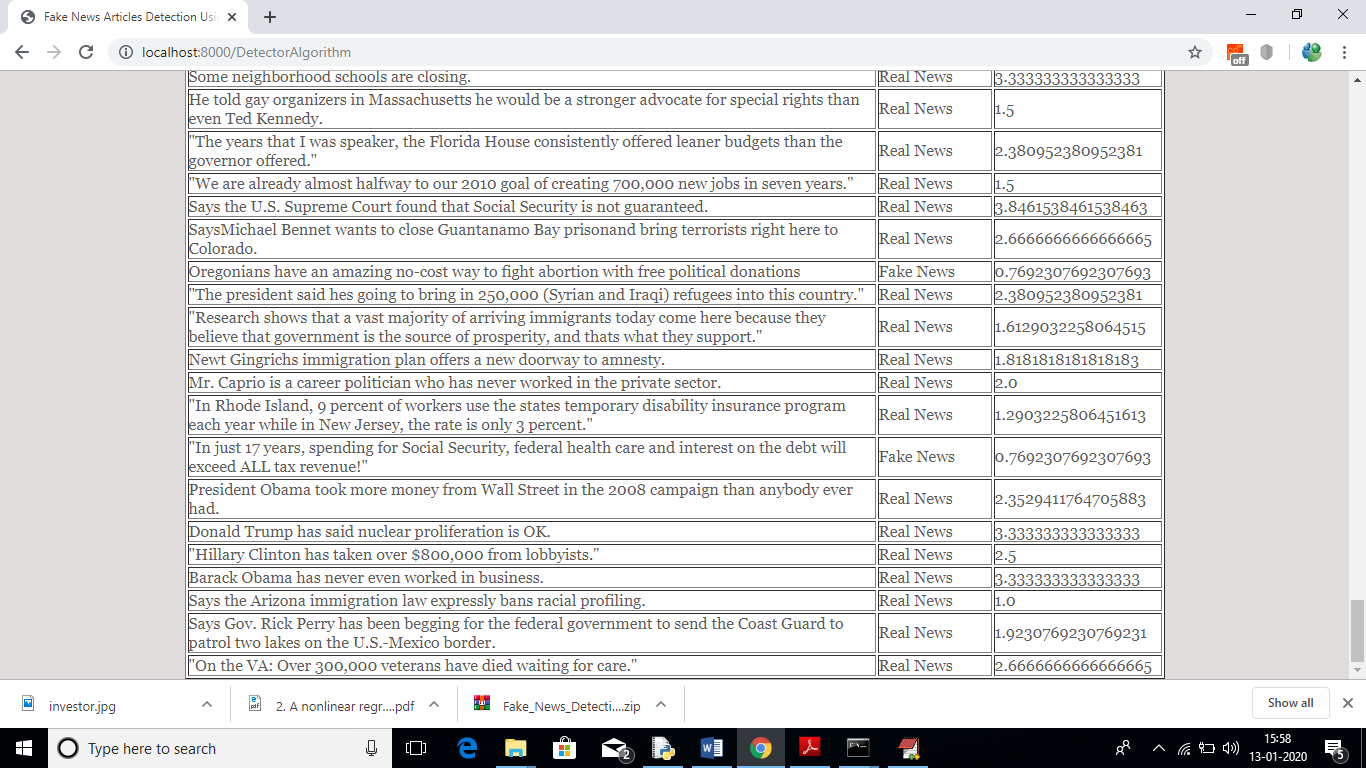
In above screen I am uploading ‘News.csv’ file which contains 150 news paragraphs. After uploading news will get below screen



In above screen news file uploaded successfully, now click on ‘Run Fake News Detector Algorithm’ link to calculate Fake News Detection algorithm score and based on score and naïve bayes algorithm we will get result.

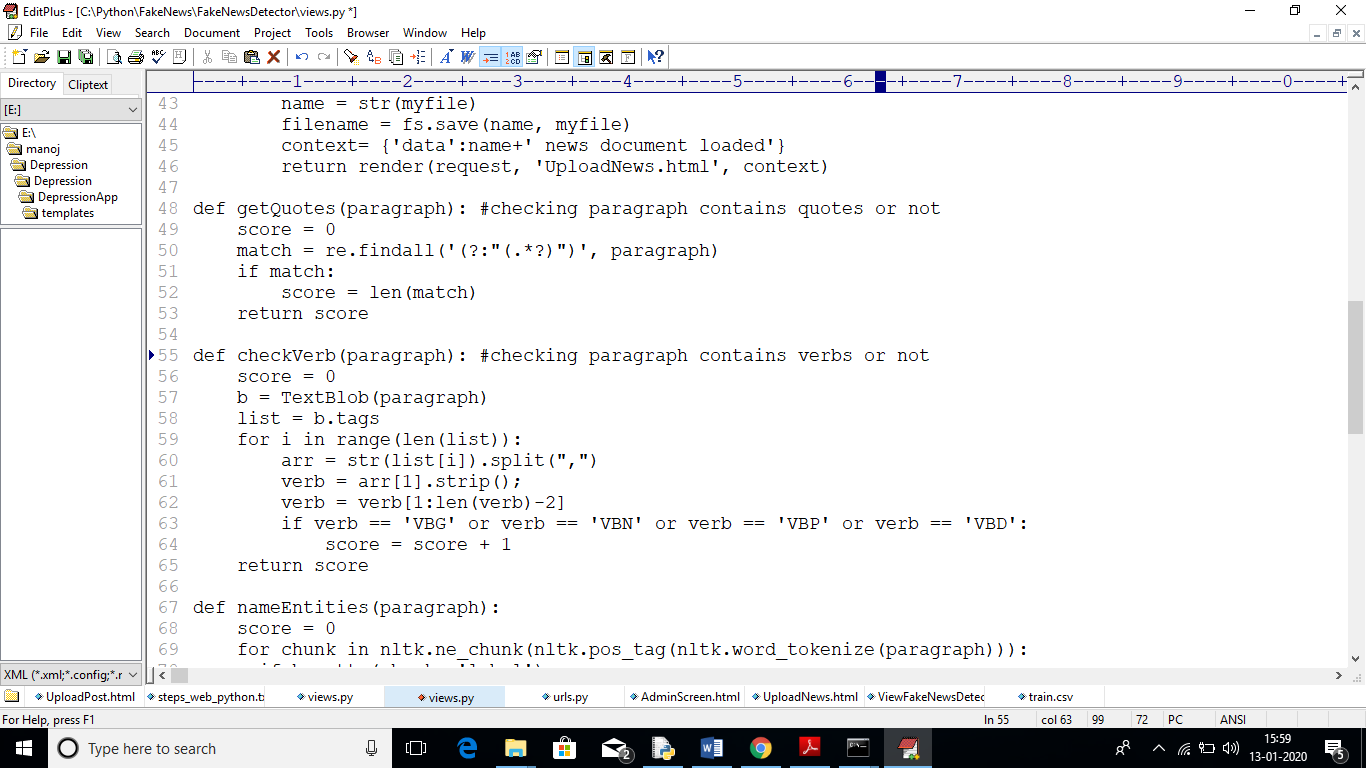


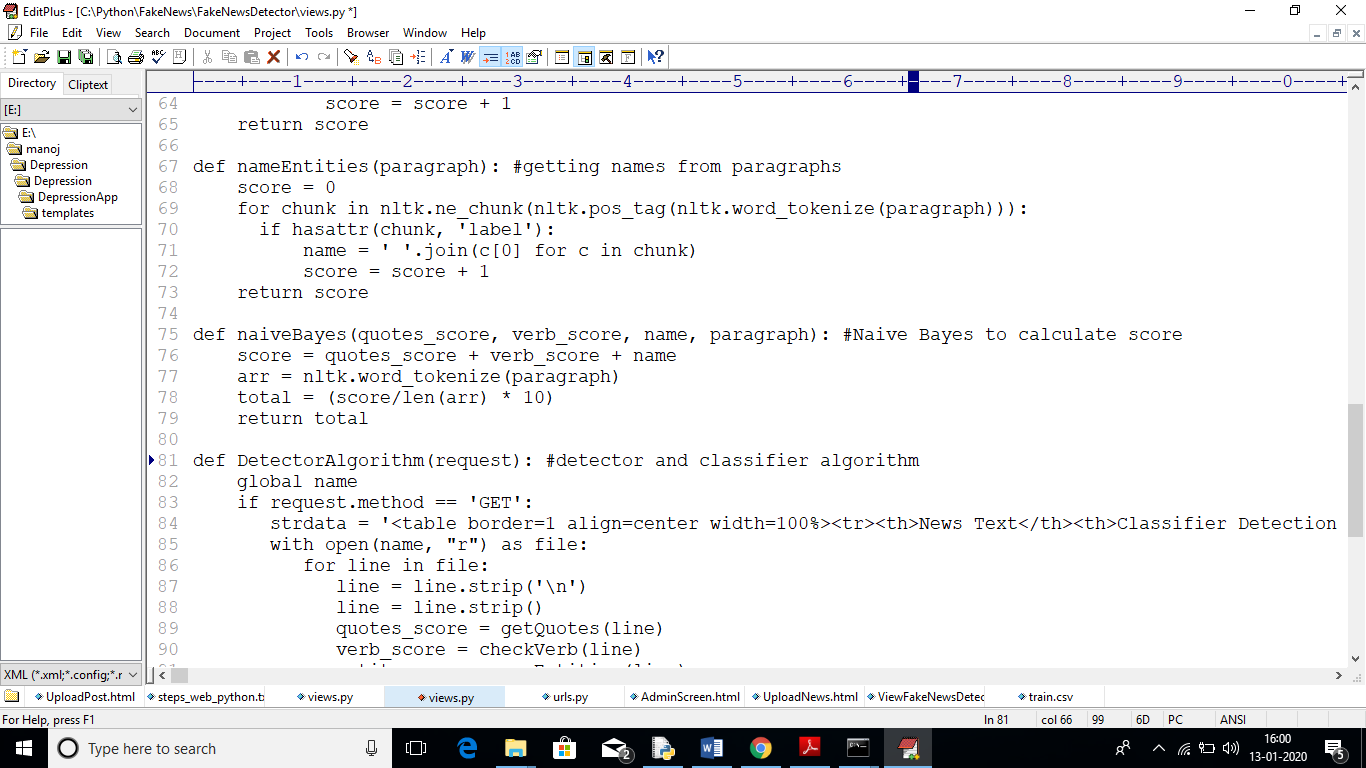
In above screen first column contains news text and second column is the result value as ‘fake or real’ and third column contains score. If score greater > 0.90 then I am considering news as REAL otherwise fake.



For all 150 news text articles we got result as fake or real.

See below screen shots of code calculating quotes, name entity and verbs from news paragraphs





Above code you can see inside ‘Views.py’ program