Fake News Detection System Using Featured-Based Optimized MSVM Classification

Fake news spread false information in the community which can create havoc and leads person to make incorrect decision. Often in election fake news can help correct candidates in gaining popularity and may win election also. In the past manual detection by human are employed to capture fake news which can be error prone and required lots of time to analyse millions of data.

To overcome from above issue author of this paper employing machine learning algorithms such as Multi-Layered Principal Component Analysis to extract relevant features from NEWS data and then apply MSVM (Multi support vector machine) along with Firefly to select important features from dataset and then ignore least important features which can help algorithm in accurately detecting or classifying giving data as True or False.

To evaluate Propose algorithm performance author has used 10 different fake news dataset but processing all those dataset may take large amount of time so we have experimented with Politifact dataset.

Politifact dataset will be cleaned and processed using NLP algorithms such as Stop word removal, special symbols removal, stemming and lemmatization. All cleaned news data will be input to PCA algorithm to extract relevant features and then input to Firefly algorithm to select important features. Selected features will get trained with MSVM algorithm to classify news as False or True.

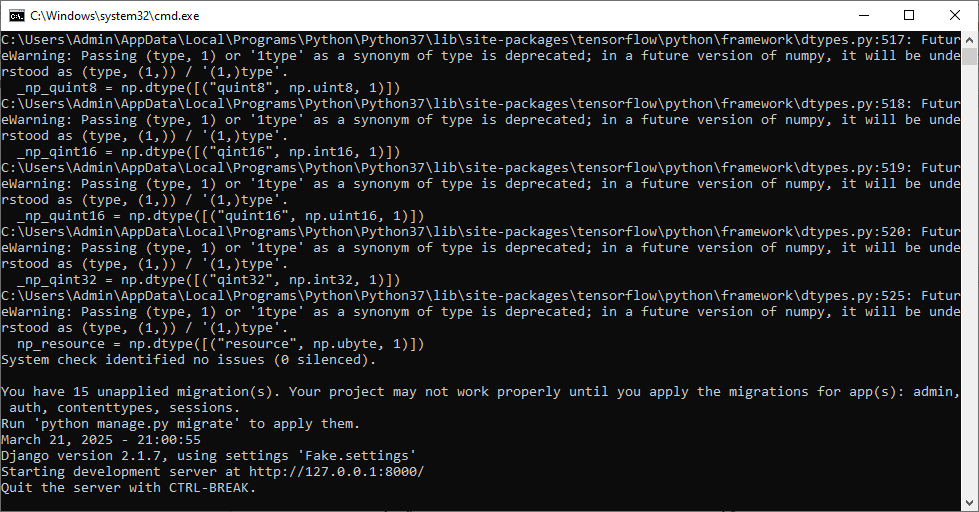
Propose MSVM algorithm is compared with existing LSTM algorithm and each algorithm performance is evaluated in terms of accuracy, precision, recall and FSCORE.

To implement this project we have designed following modules

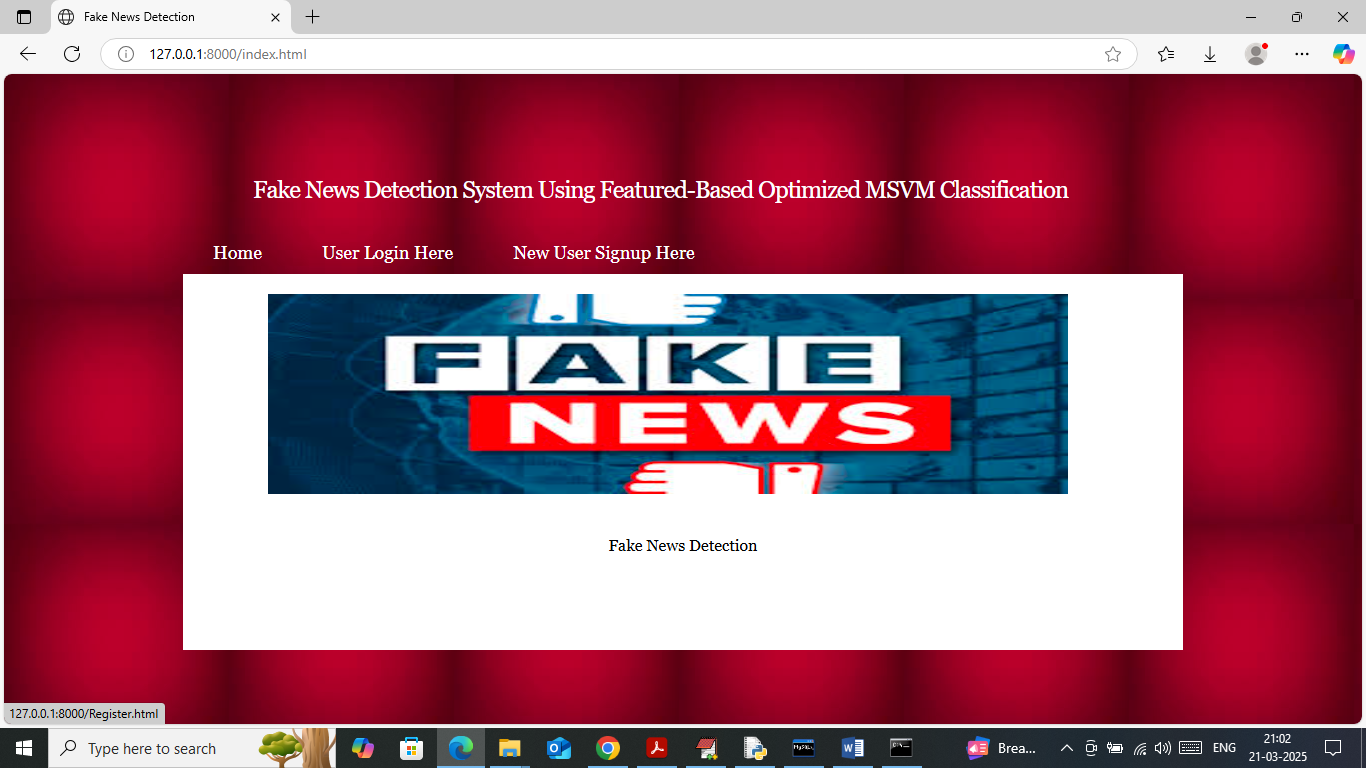
1. New User Sign up: using this module user can sign up with the application
2. User Login: using this module user can login to system
3. Load Fake News: using this module user can upload fake news dataset to application and then apply NLP algorithms on loaded news text data to remove stop words and special symbols
4. Run MPCA & Firefly Features Selection: cleaned news data will be input to MPCA algorithm to extract features and then apply Firefly algorithm to select relevant features. Selected features will be split into train and test where application using 80% data for training and 20% for testing
5. Run MSVM Algorithm: 80% training data will be input to MSVM and LSTM algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
6. Predict News: using this module either user can enter NEWS data or upload news file and then system will apply all algorithms to classify that news or False or True.

SCREEN SHOTS

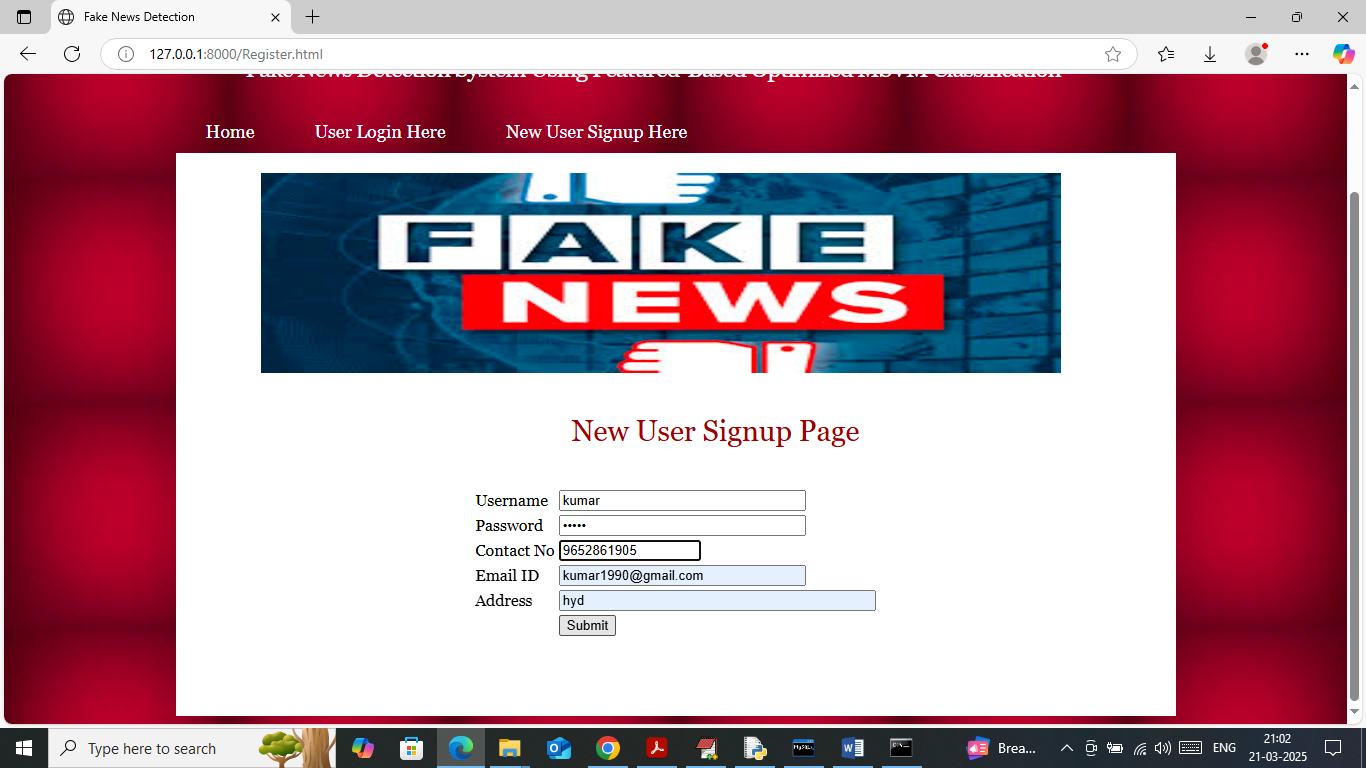
To run project double click on ‘run.bat’ file to start python server and then will get below page



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



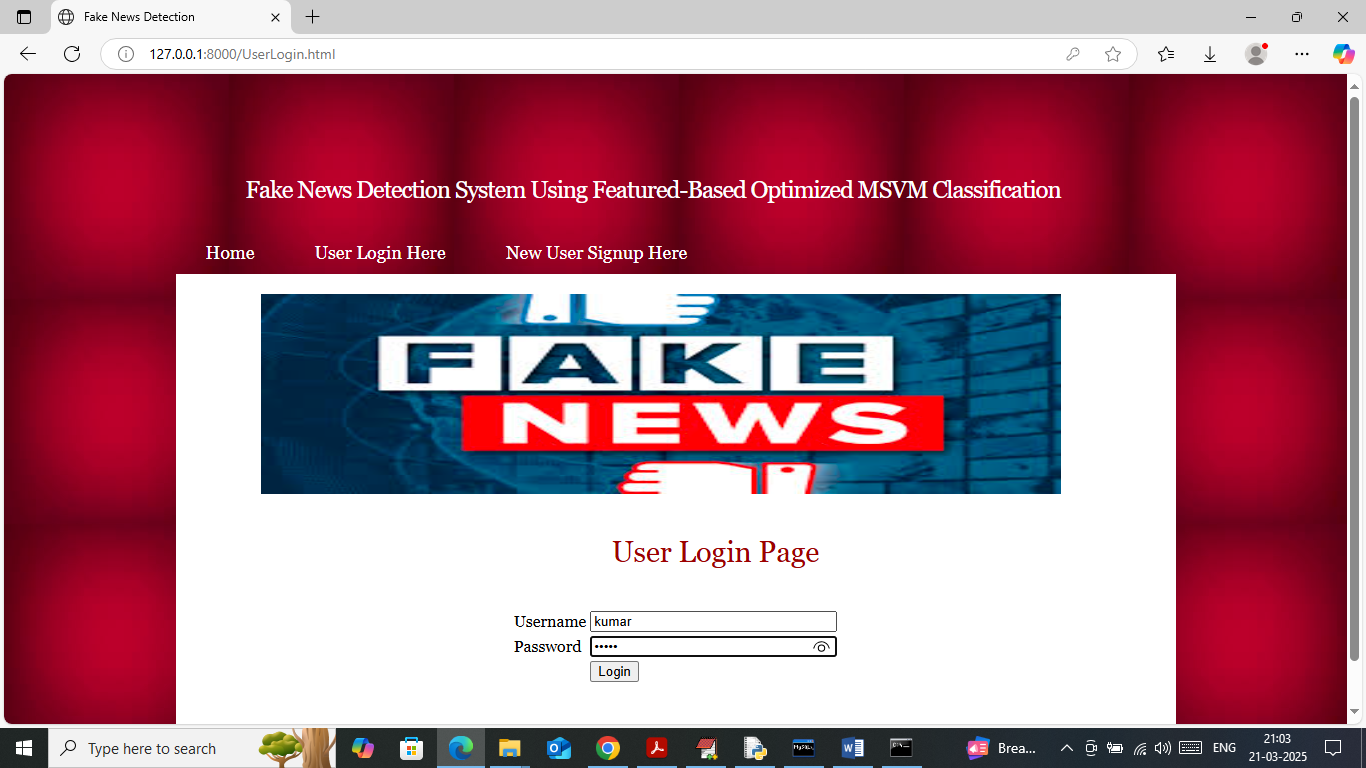
In above screen click on ‘New User Sign up’ link to get below page



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



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



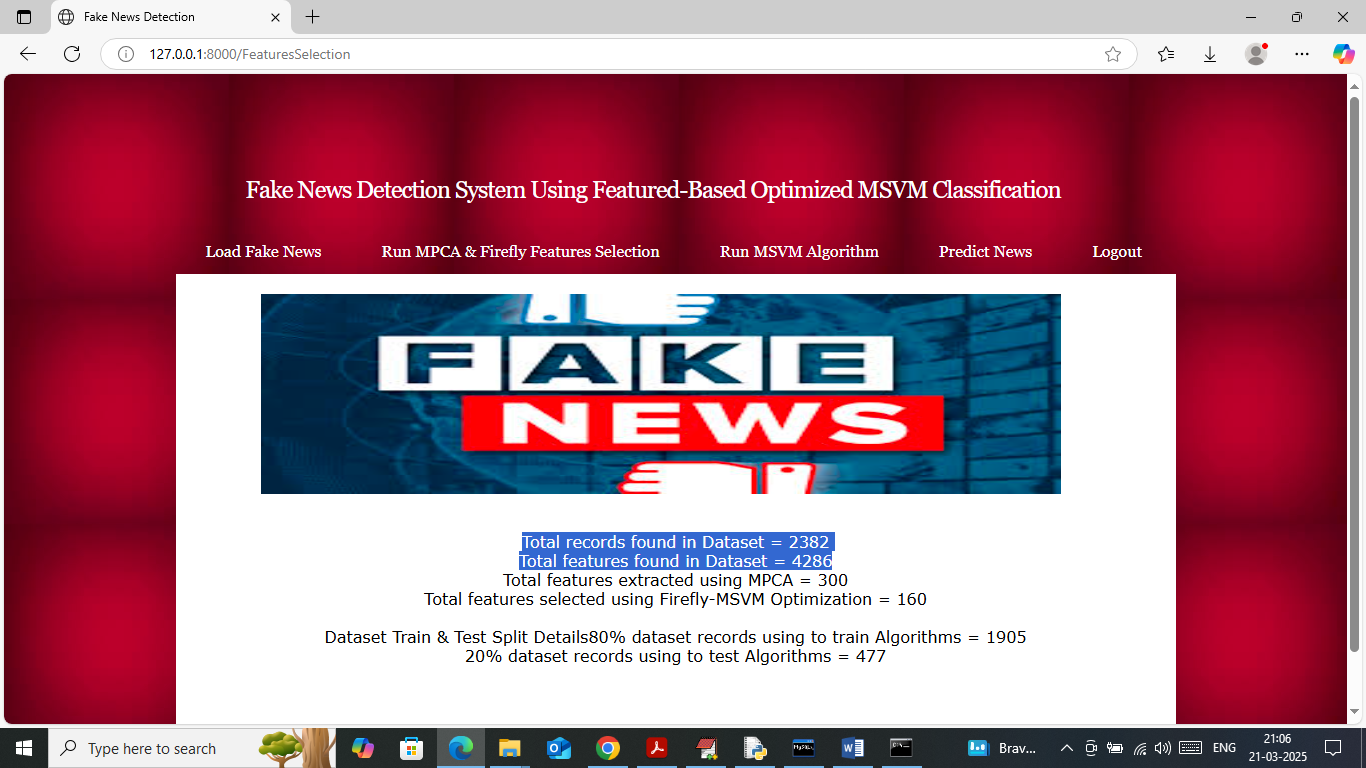
In above screen user is login and after login will get below page



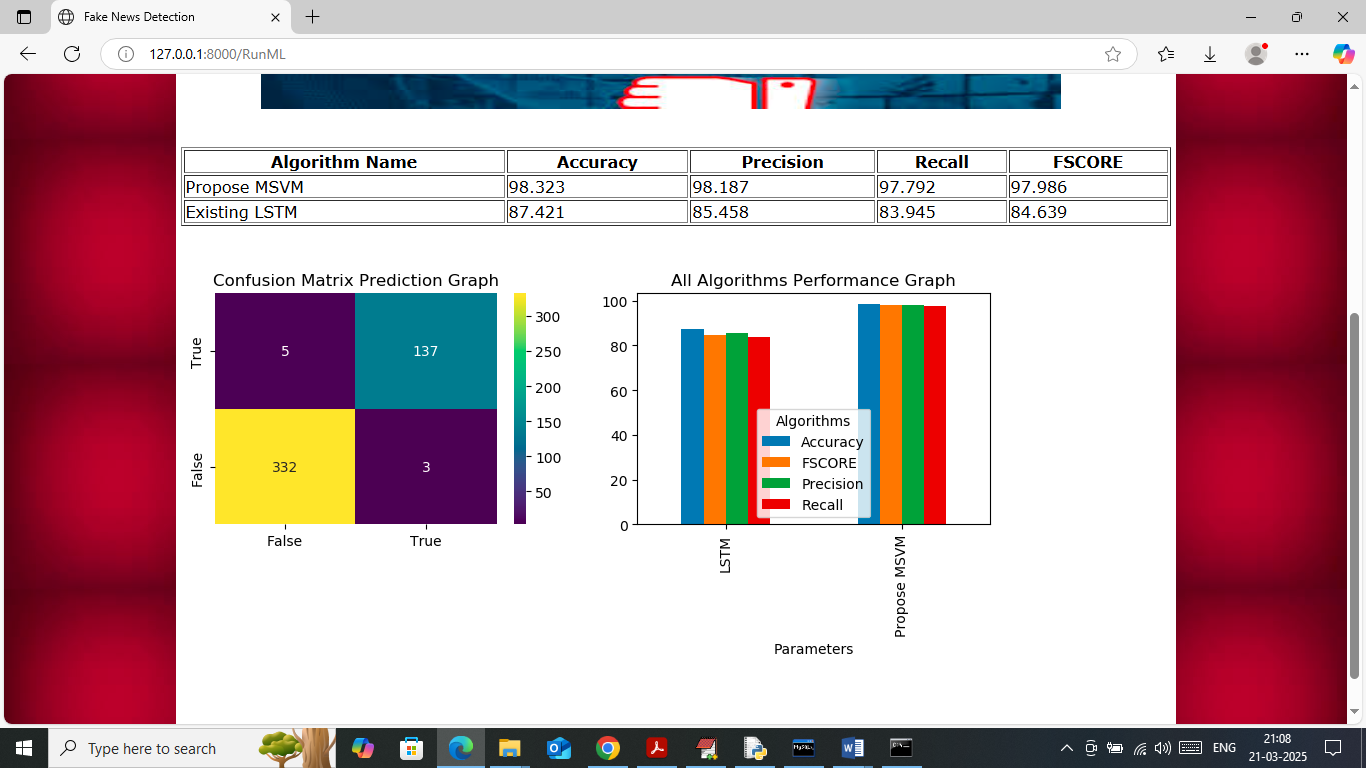
In above screen click on ‘Load Fake News’ link to load dataset and then will get below page



In above screen dataset loaded where first column contains news data and second column contains Class label and FALSE or TRUE. Now click on ‘Run MPCA and Firefly’ link to extract and select features from dataset and then will get below page



In above screen in first two lines can see number of records and features available in dataset and then MPCA extracted 300 features and then Firefly selected 160 important features and now click on ‘Run MSVM Algorithm’ link to train algorithms and then will get below page



In above screen in table format can see accuracy, precision, recall and FCSORE of Propose MSVM and existing LSTM algorithm. In above table can see MSVM got 98% accuracy and Existing LSTM got 87% accuracy. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then yellow and light green boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count. In bar graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. Now click on ‘Predict News’ link to get below page



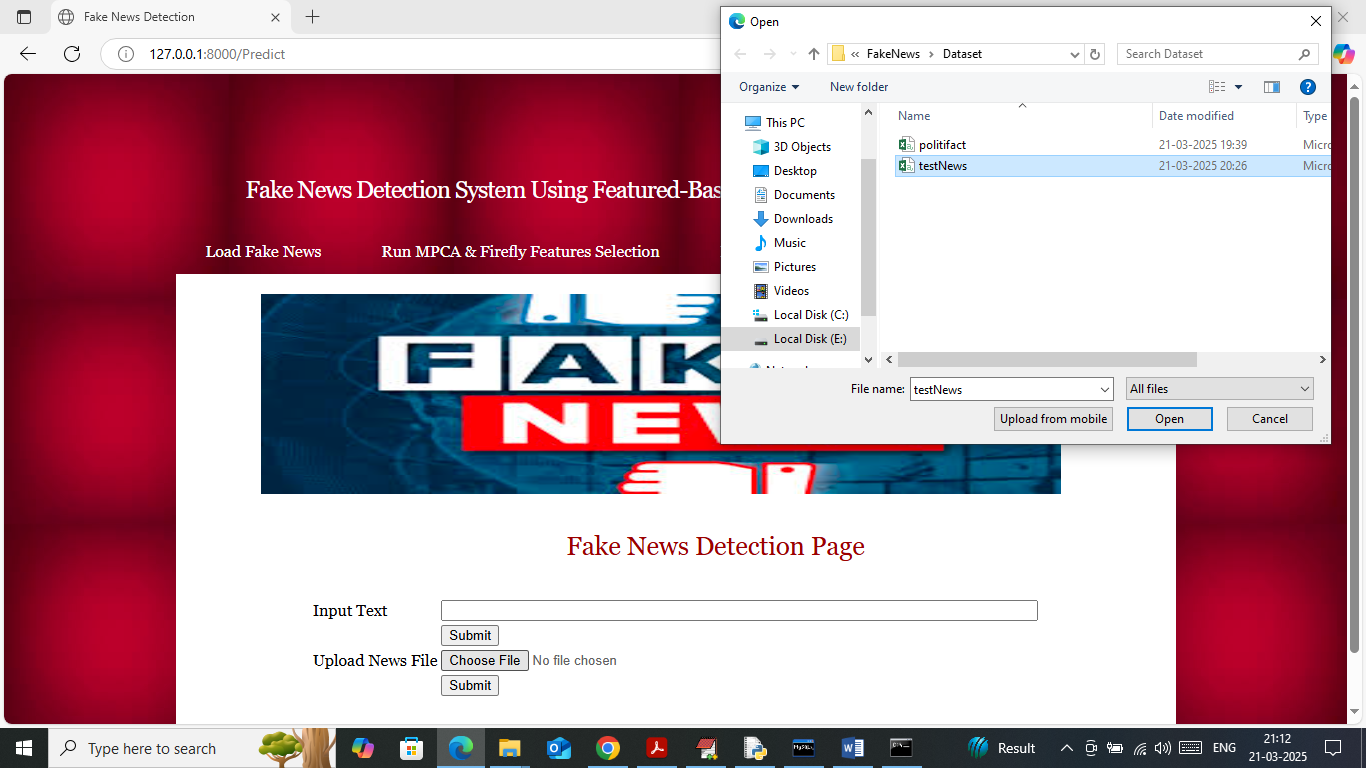
In above screen in text field you can enter some news data and then press button to classify that news text data as False or True or you can upload news data file



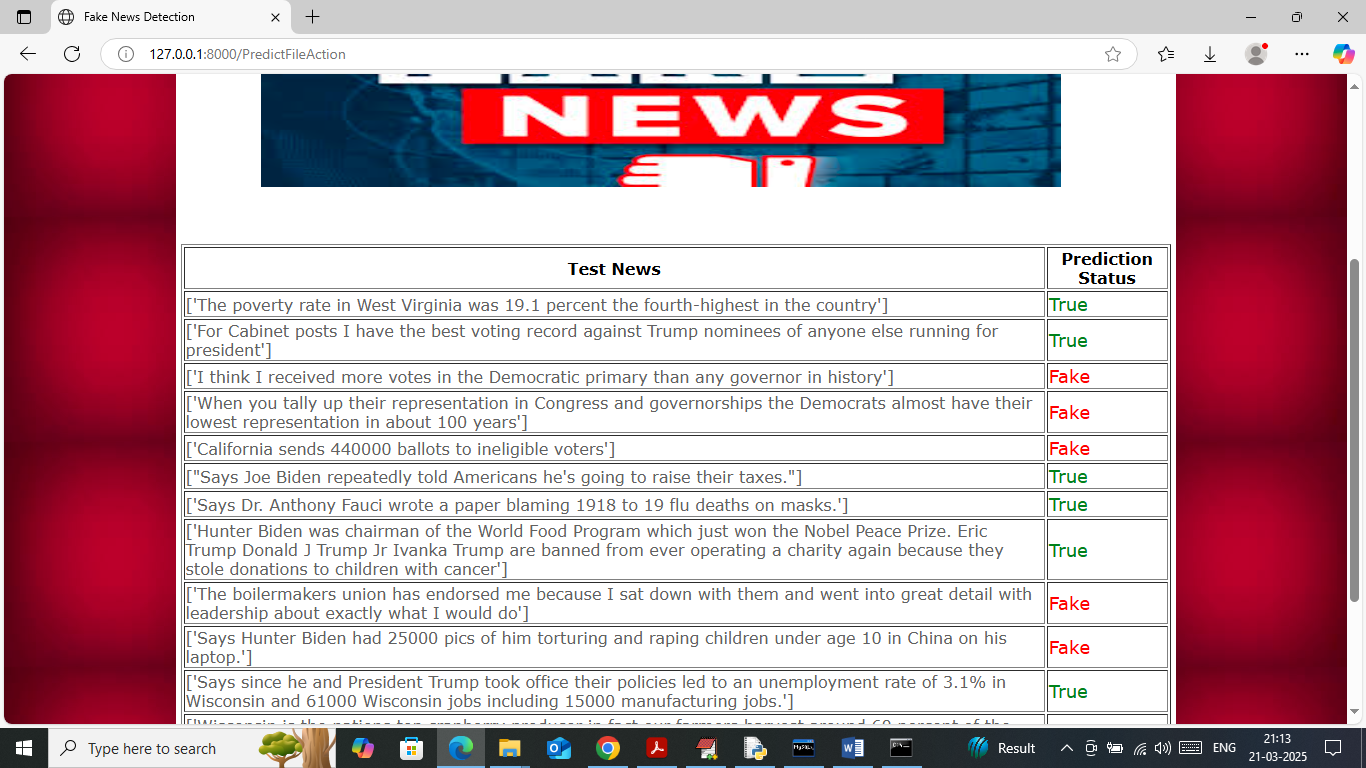
In above screen in text field I entered some news data and then press button to get below page



In above screen given news data predicted as ‘Fake’ and similarly you can test any other news. Now upload news data file



In above screen uploading ‘testNews.csv’ file and then click on ‘Open’ and second submit button to get below page



In above screen in first column displaying news text data and in second column displaying classification output as ‘False or True’.

Note: in requirement you mention to click picture and then predict fake or real but propose algorithm can only apply on text data not on images.

You sent some template to include but that template was not working as you include everything in single file so we cannot use that template.