Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets

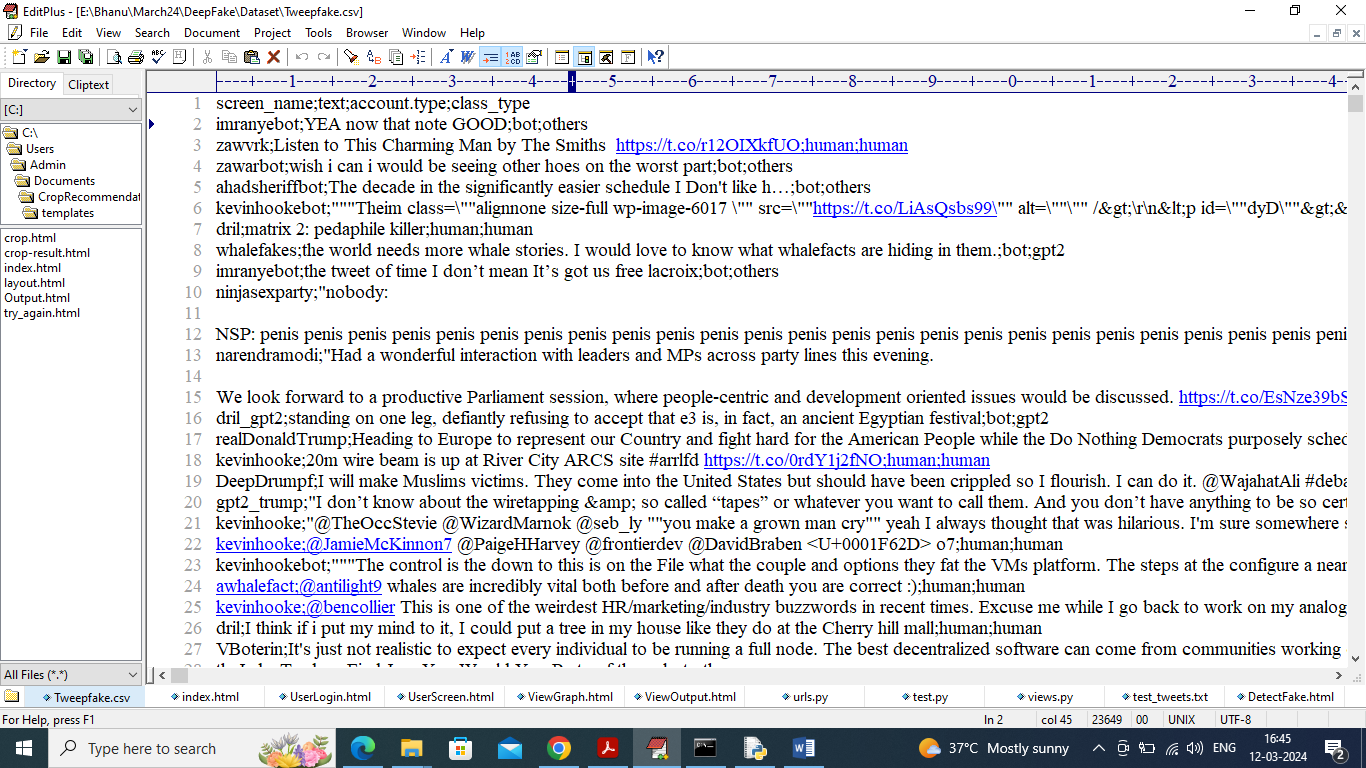
Online platform become a massive tool to spread users opinion on any topic and one get latest news across world in minutes through social media but some malicious users may utilize this platform to spread fake opinion or news. Some malicious user may develop BOT application which will spread fake news or opinions on social media. This BOT will be programmed to spread or publish tweets after some interval. Peoples will view such BOT tweets and consider as genuine tweets and spread to further user.

In propose paper author investigating performance of various TEXT Embedding tools like FASTTEXT, TFIDF (term frequency inverse frequency document) and TF (term frequency) with different machine learning algorithms like Naïve Bayes, Logistic Regression, CNN , LSTM, Decision Tree, Random Forest and Many more algorithms. Among all algorithms CNN is giving best accuracy.

Each algorithm performance is evaluated in terms of accuracy, precision, recall and FSCORE. To train above algorithms author has utilized dataset called ‘Tweepfake’ which can be downloaded from KAGGLE repository.

FASTTEXT, TFIDF or TF is an algorithm used to convert text to numeric vector which will replace each word occurrence with its average frequency. In propose paper FASTTEXT with CNN giving high accuracy so we are using this algorithms to implement project.

In below screen showing dataset details



In above dataset screen first column contains dataset column names and remaining rows contains dataset values which contains tweet text along with class labels as Human or bot where human refers to tweet written by human and bot refers to tweet publish by bot and its fake.

Extension Concept

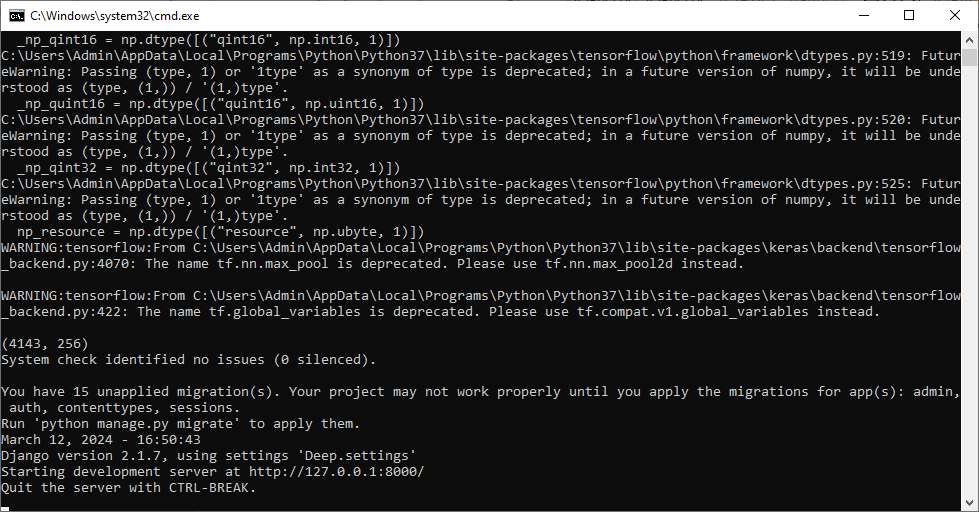
In propose paper CNN is giving best accuracy and will be consider as one of the best algorithm for optimized feature extraction. Extracted features from CNN can be retrain with any other algorithms such as Random Forest to get high accuracy. Random Forest will get trained on optimized features from CNN so its accuracy will get high. So by combing CNN features with Random Forest we are forming extension algorithm called Hybrid CNN.

To implement this project we have designed following modules

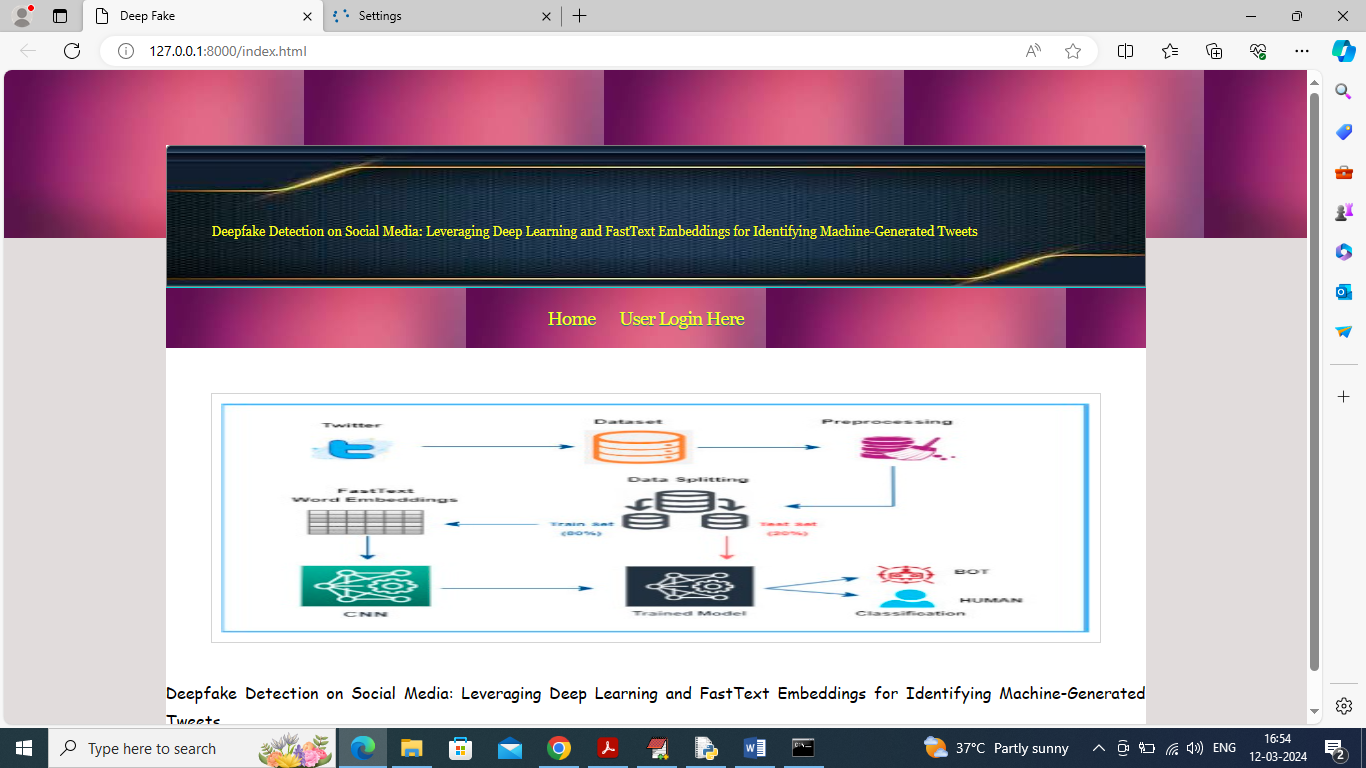
1. User Login: user can login to system using username and password as ‘admin and admin’
2. Load Dataset: after login user can click this link to load dataset to application
3. Fast Text Embedding: loaded dataset will be clean by removing stop words, special symbols and other text processing techniques and then input to FASTTEXT algorithm to generate numeric vector
4. Run All Algorithms: numeric vector will be normalized and then split into train and test and then training data will be input to all algorithms to train a model and this models will be applied on test data to calculate prediction accuracy
5. Predict Deep Fake: in this module will enter some tweets text and then CNN algorithm will predict weather tweet is written by Human or BOT

SCREEN SHOTS

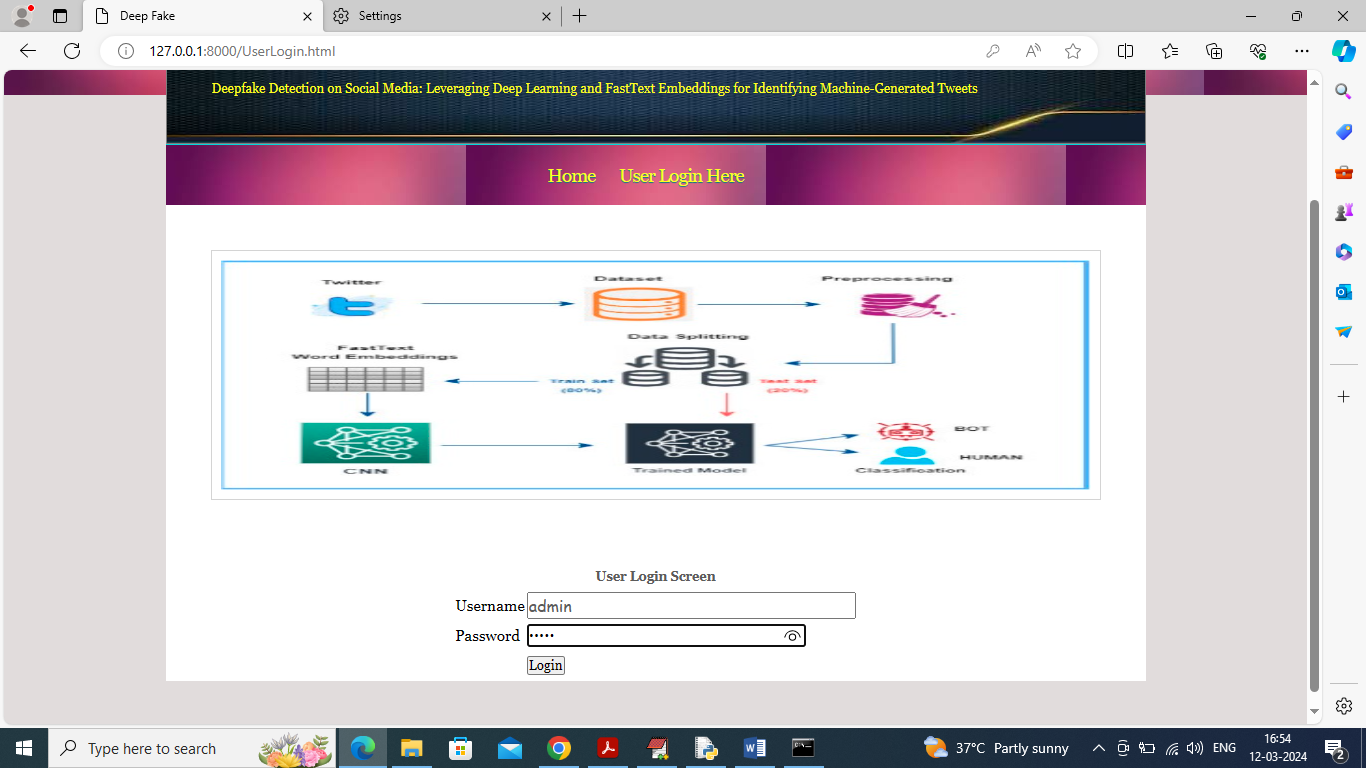
To run code double click on ‘run.bat’ file to start python server and get below page



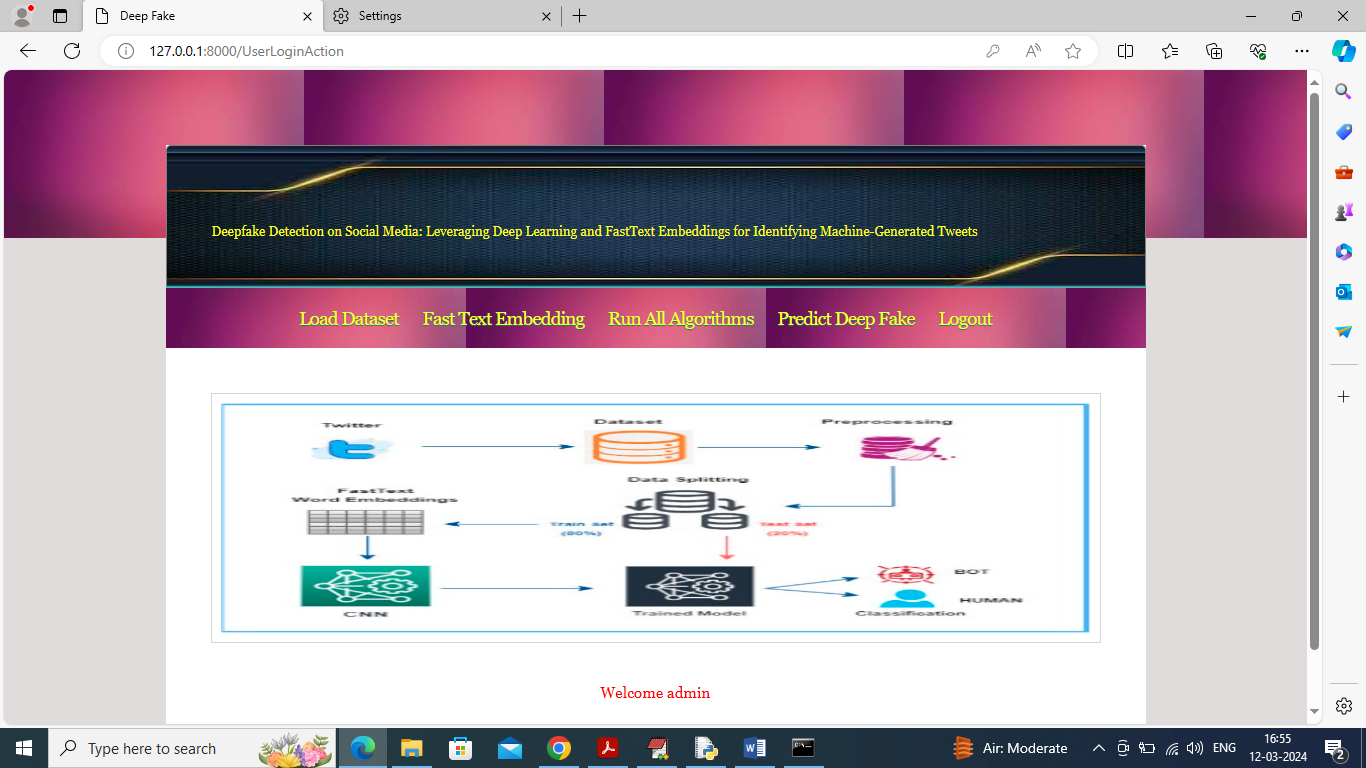
In above screen python server started and now open browser and enter URL as <http://127.0.0.1:8000/index.html> and press enter key to get below page



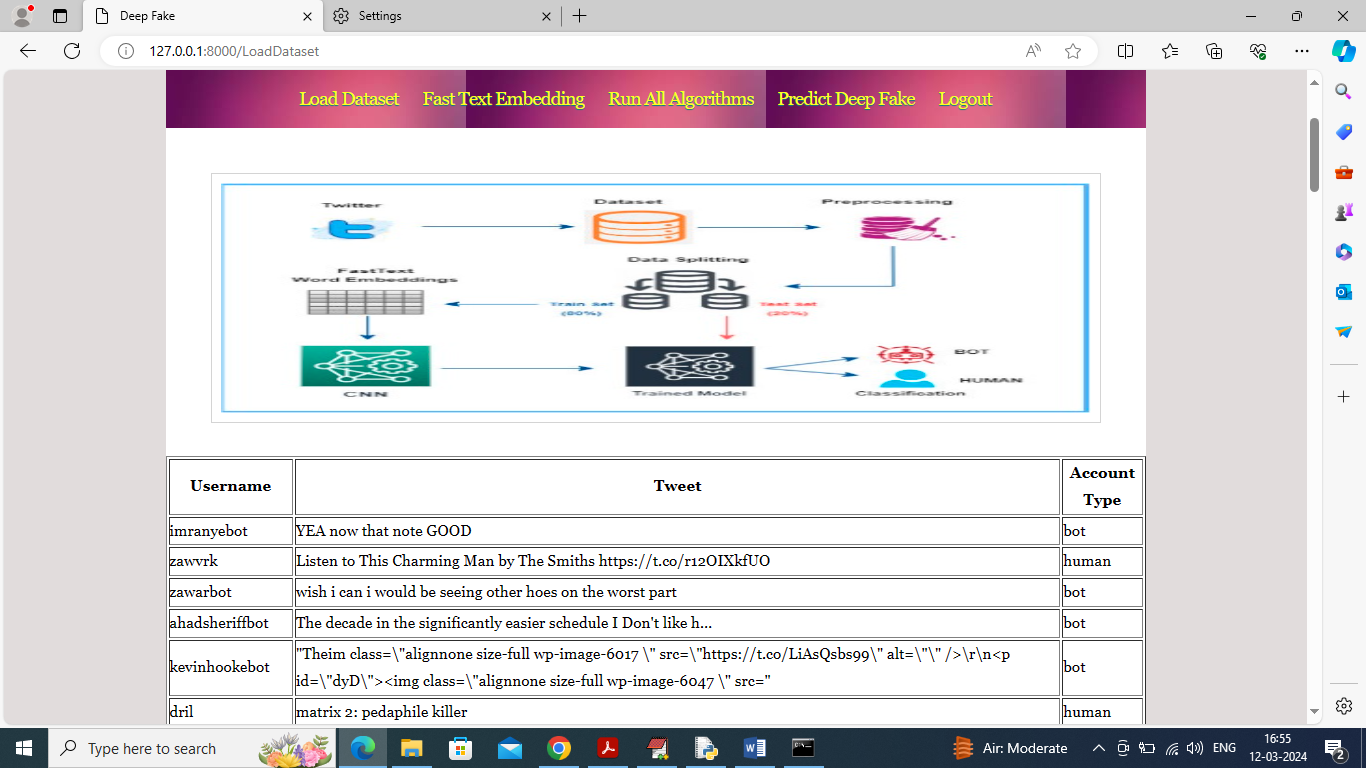
In above screen click on ‘User Login Here’ link to get below page



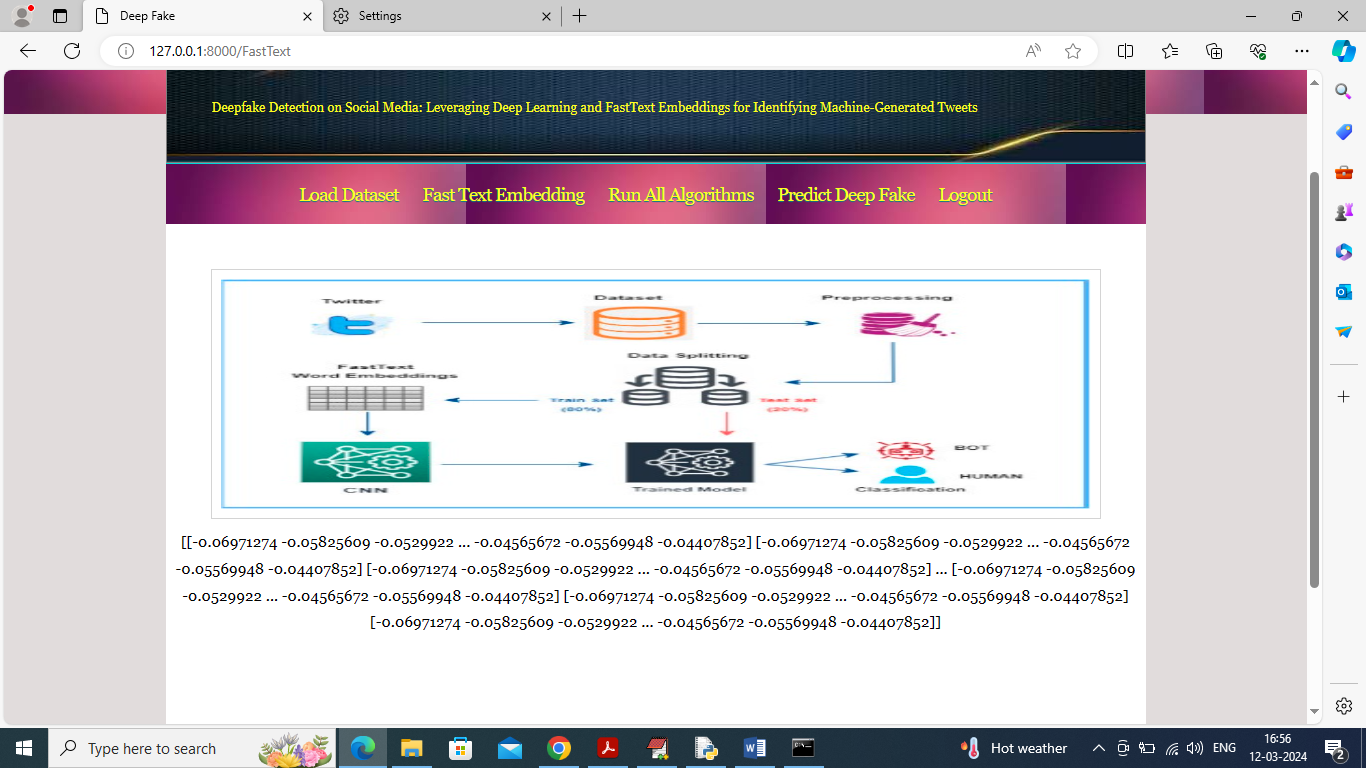
In above screen user is login and after login will get below page



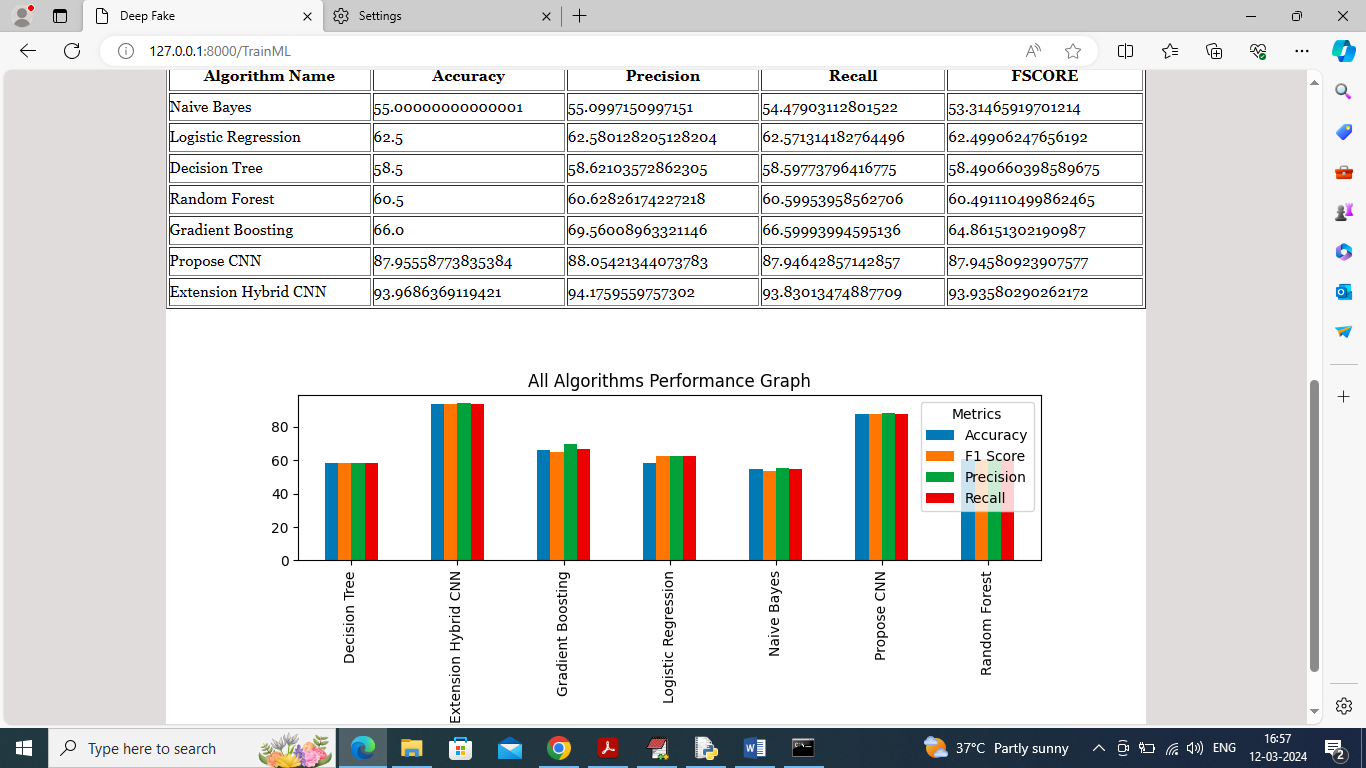
In above screen click on ‘Load Dataset’ link to load dataset and get below page



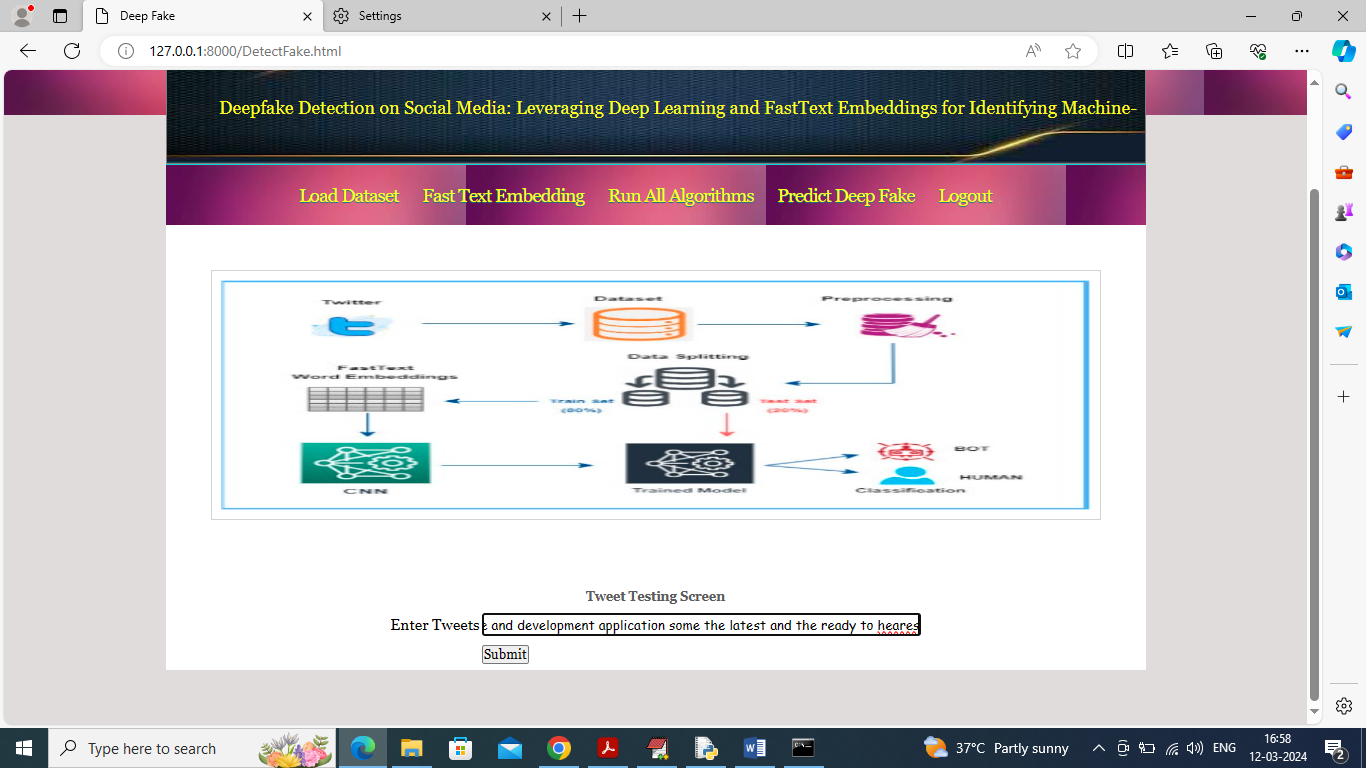
In above screen dataset loaded and now click on ‘Fast Text Embedding’ link to convert all text to numeric vector and get below page



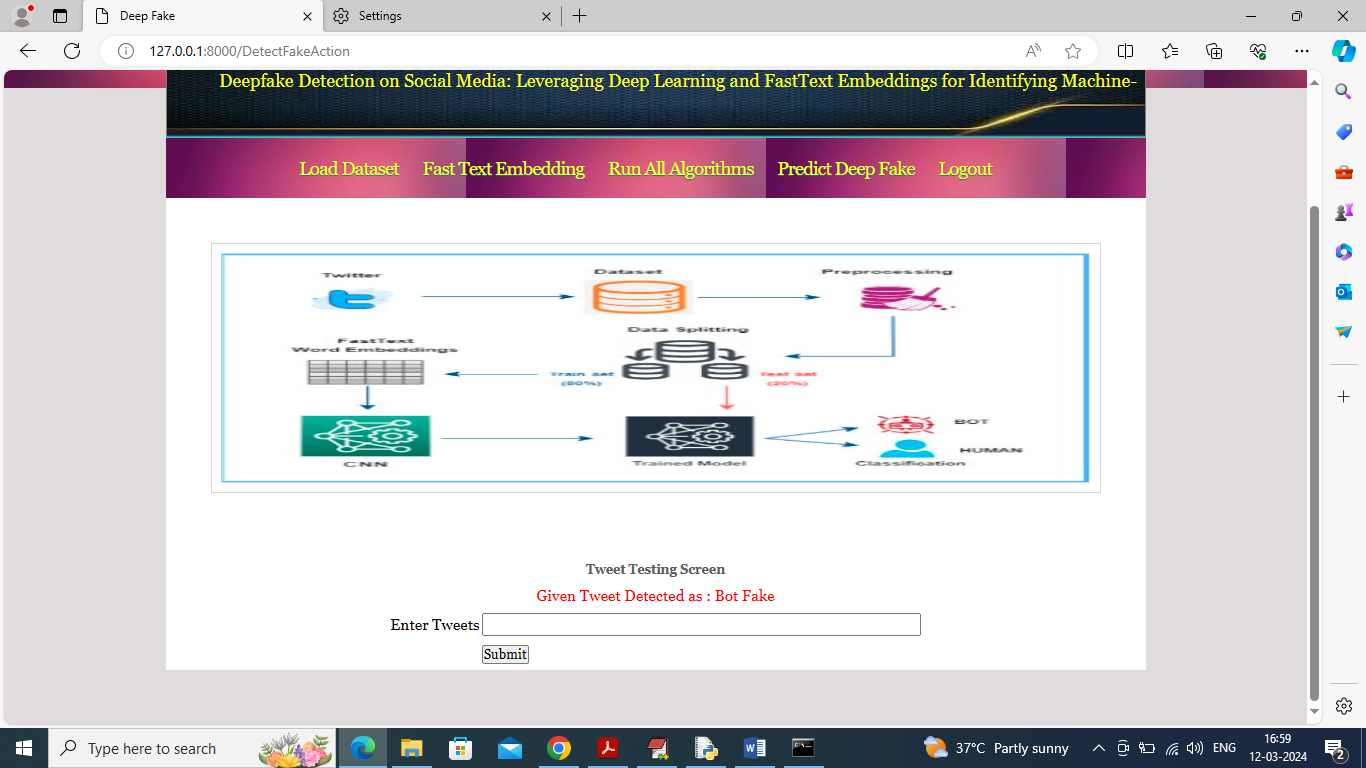
In above screen all tweets converted to numeric vector and then displaying some values from vector and now click on ‘Run All ML Algorithms’ link to train all algorithms and get below page



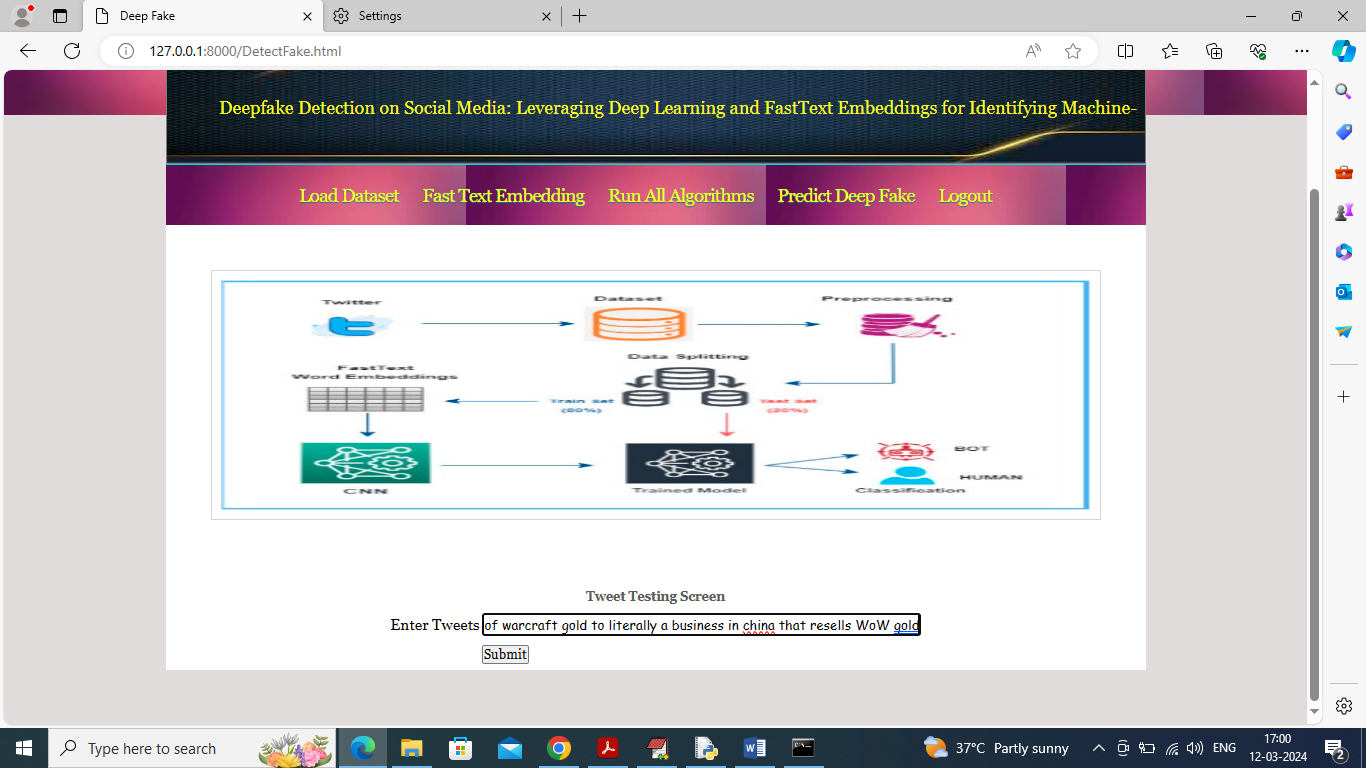
In above screen can see all algorithms result in tabular and graph format and in above screen can see propose CNN and extension hybrid CNN got high accuracy. Now click on ‘Predict Deep Fake’ link to get below page



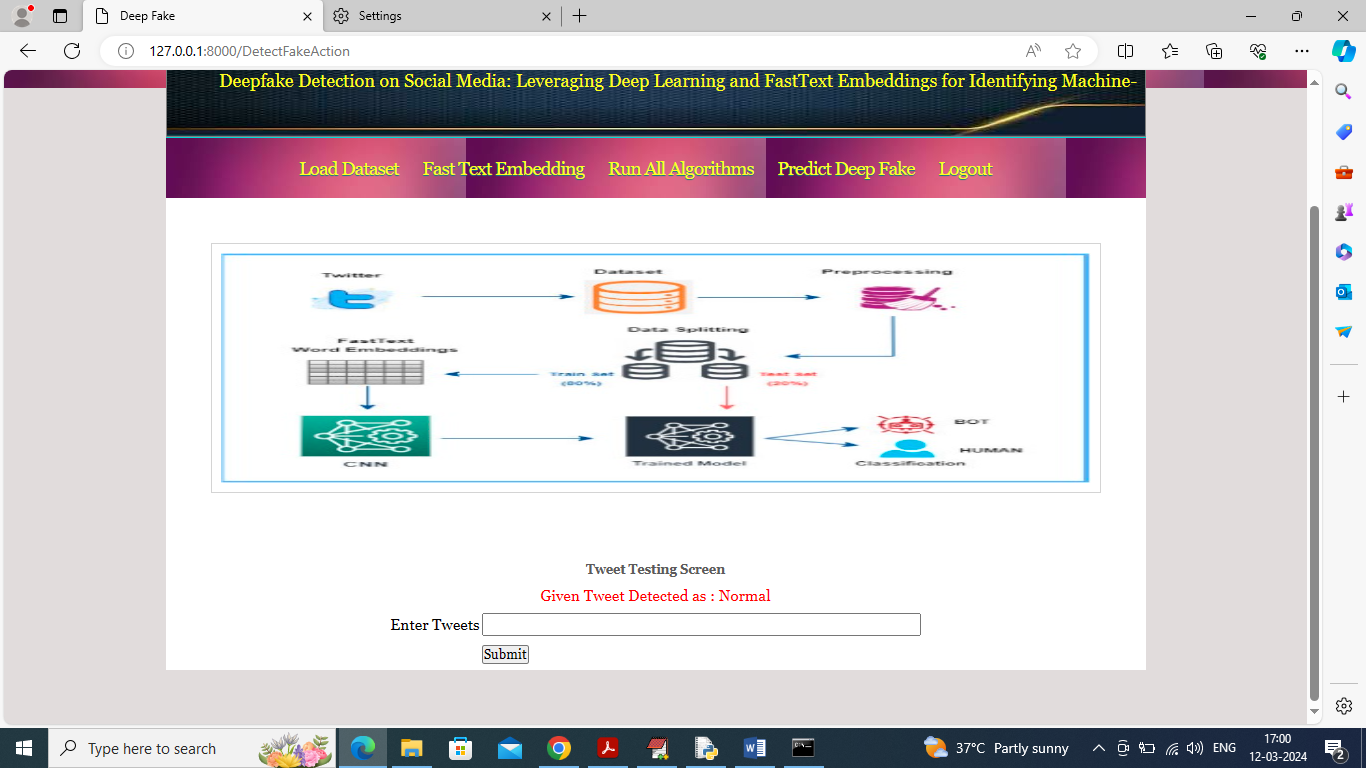
In above screen in text field enter some tweet text and then press button to get below values and if you want you can use sample tweets given in ‘test\_tweets.txt’ file



In above screen given tweet predicted as ‘Deep Bot’ means its fake tweet spread by BOT and now in below screen can see another example



In above screen entered some other tweet text and below is the output



In above screen tweet detected as normal which means tweet written by human. Similarly you can enter some tweets and get output