Stock Market Prediction via Multi-Source Multiple Instance Learning

Peoples are heavily dependent on stock market forecasting results to invest money in suitable stock to gain revenue. Incorrect forecasting may leads to huge loss and many existing algorithms fails to give accurate forecasting as they get trained on single data source called quantitative data.

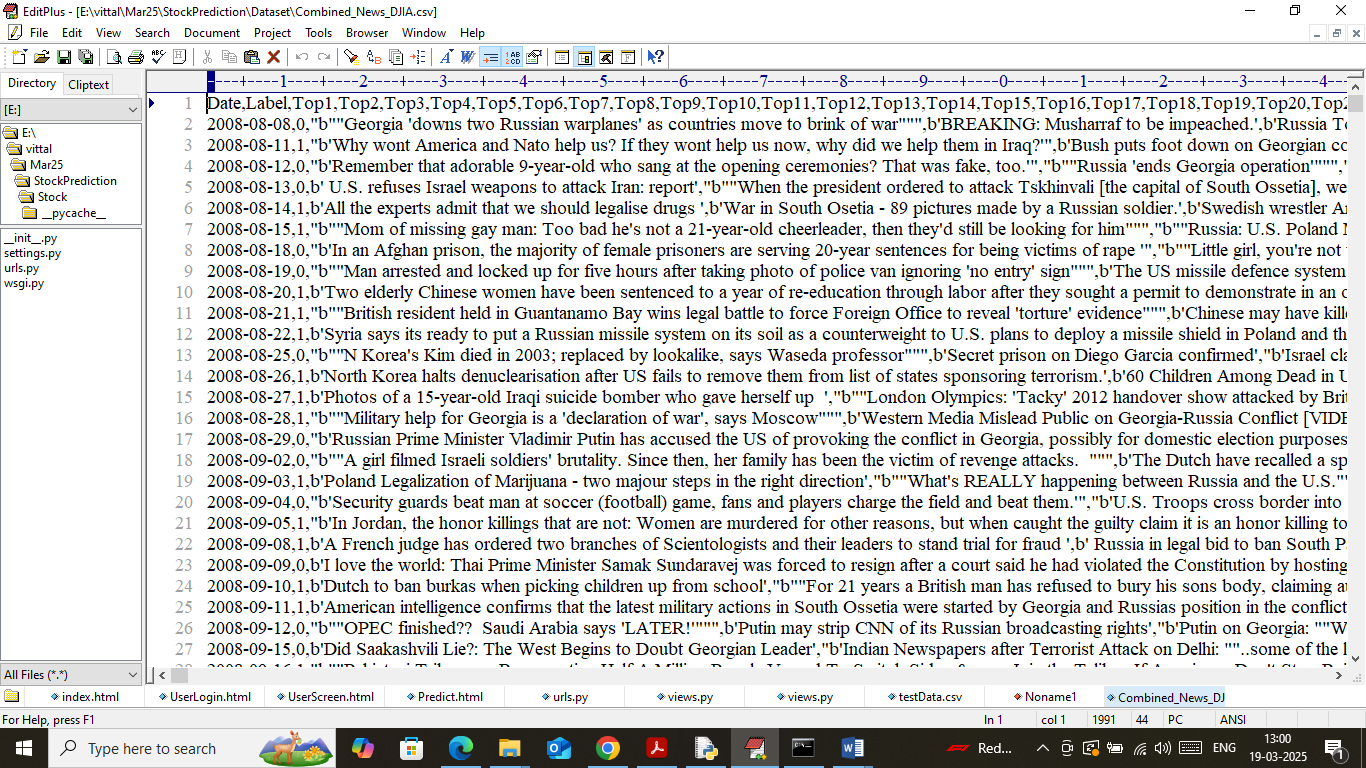
To overcome from above issues author of this paper employing multi-source data to extract useful stock events and then merged with quantitative data (stock prices) to make accurate forecasting. In propose work author using news, social and stock market data. All this data will be processed to extract EVENTS, Sentiments and sentence2vector generation. Generate vector will be trained with multi-instance SVM algorithm for future stock prediction such as ‘Decline or Rise’.

To extract features from multi-source data author has used following algorithms

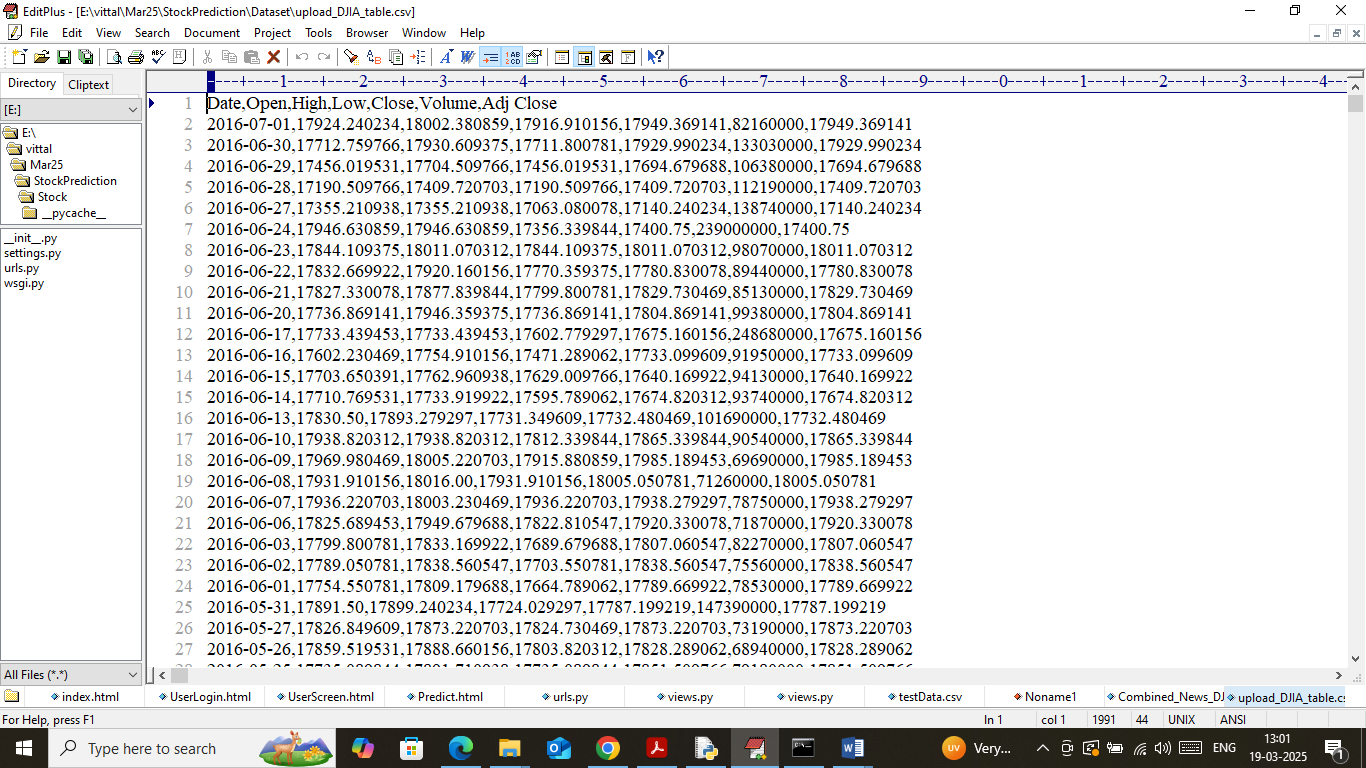
1. Structured event extraction: HanLP algorithm employ to capture the syntactic structure of a sentence. The root node denotes the core verb, and the nodes of the second layer are the subject of the verb and the object of the verb respectively. The child of the subject is the modifier who is the nearest to the subject in the sentence, and so is the child of the object. This algorithm will extract events from NEWS data
2. Training with RBM: this algorithm is used to capture dimension for the training features based on available events data
3. Training with sentence2vec: all extracted events will be input to generate vector from sentences based on dimension calculated by RBM algorithm.
4. LDA Algorithm: LDA algorithm will be applied to extract sentiments from processed news data.

All the above extracted features will be merge with stock market quantitative data and then trained with Multi-instance SVM algorithm to predict stock prices. Propose multi-source multi-instance algorithm accuracy will be compared with existing SVM algorithm which will get trained on single dataset.

To train and test above algorithm performance we have used below dataset which contains News and stock and social data



In above dataset we have stock news data with labels as 0 (decline) and 1 (rise) and below is the quantitative dataset



In above screen we have stock data and both datasets will be merged to extract features and then train with algorithms.

Extension Concept

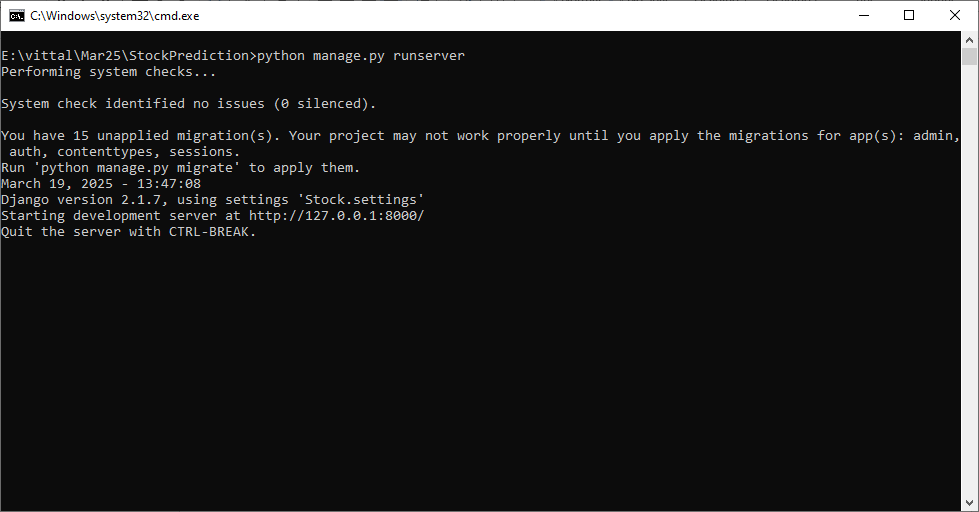
In propose paper author has used traditional SVM algorithm with extracted features so as extension we have employed advance Machine Learning algorithm called XGBOOST which will utilize 100’s of decision tree to optimize training features which can help algorithm in gaining maximum accuracy.

To implement this project we have designed following modules

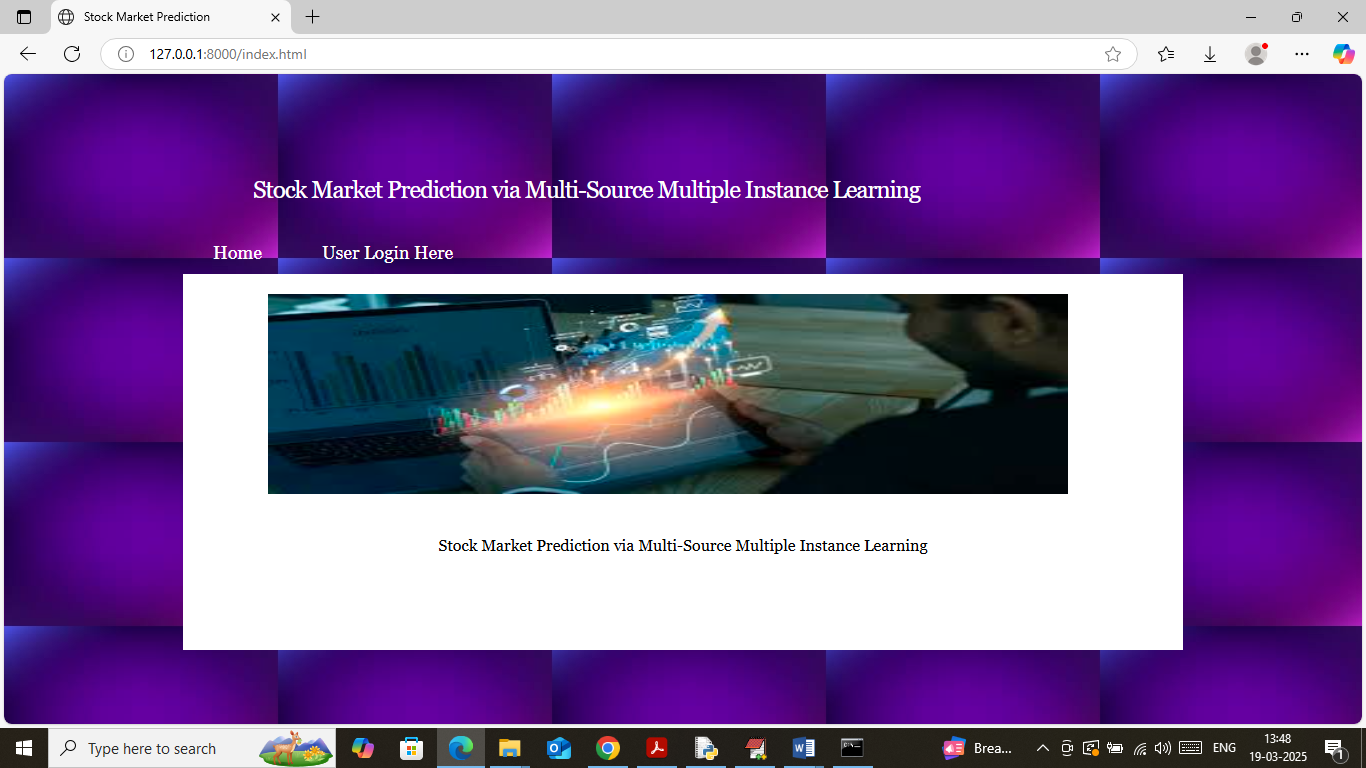
1. User Login: user can login to system using username and password as ‘admin and admin’.
2. Load Dataset: after login user can execute this module to load dataset with NEWS and quantitative data and then clean all text data by converting to lower case and then remove all special symbols
3. Event, Sentiments & Vector Generate: all processed news text will be further scanned to extract HANLP events, sentiments and vector generation. Generated vector will be split into train and test where application using 80% data for training and 20% for testing
4. Multi Instance Training: 80% training data will be input to propose and other algorithms to train a model and this model will be applied on 20% test data to calculate prediction accuracy and other metrics
5. Predict Market: using this module will upload test data and then extract features and input to extension algorithm to predict market status as Decline or Rise.

SCREEN SHOTS

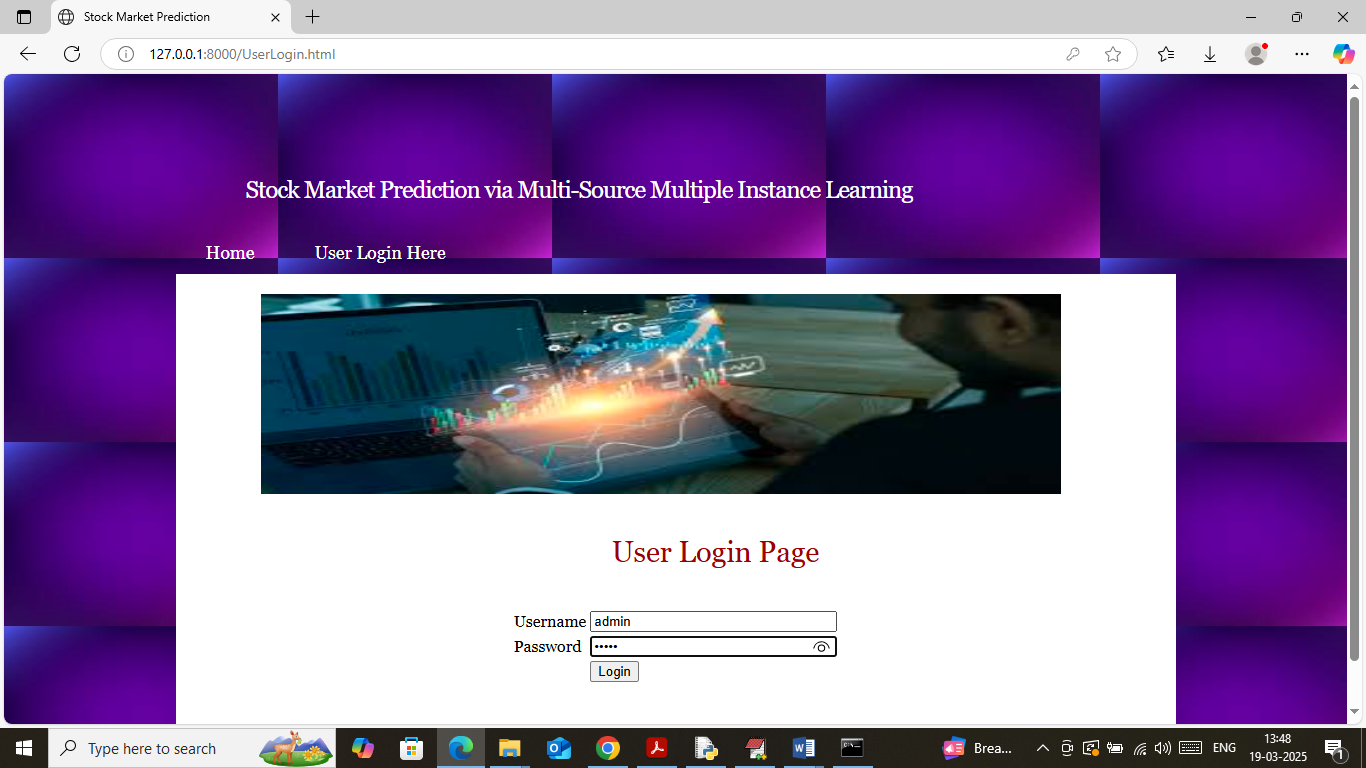
To run project install Python 3.7.2 and then install all packages given in requirements.txt file and then double click on ‘run.bat’ file to start python web server and then will get below page



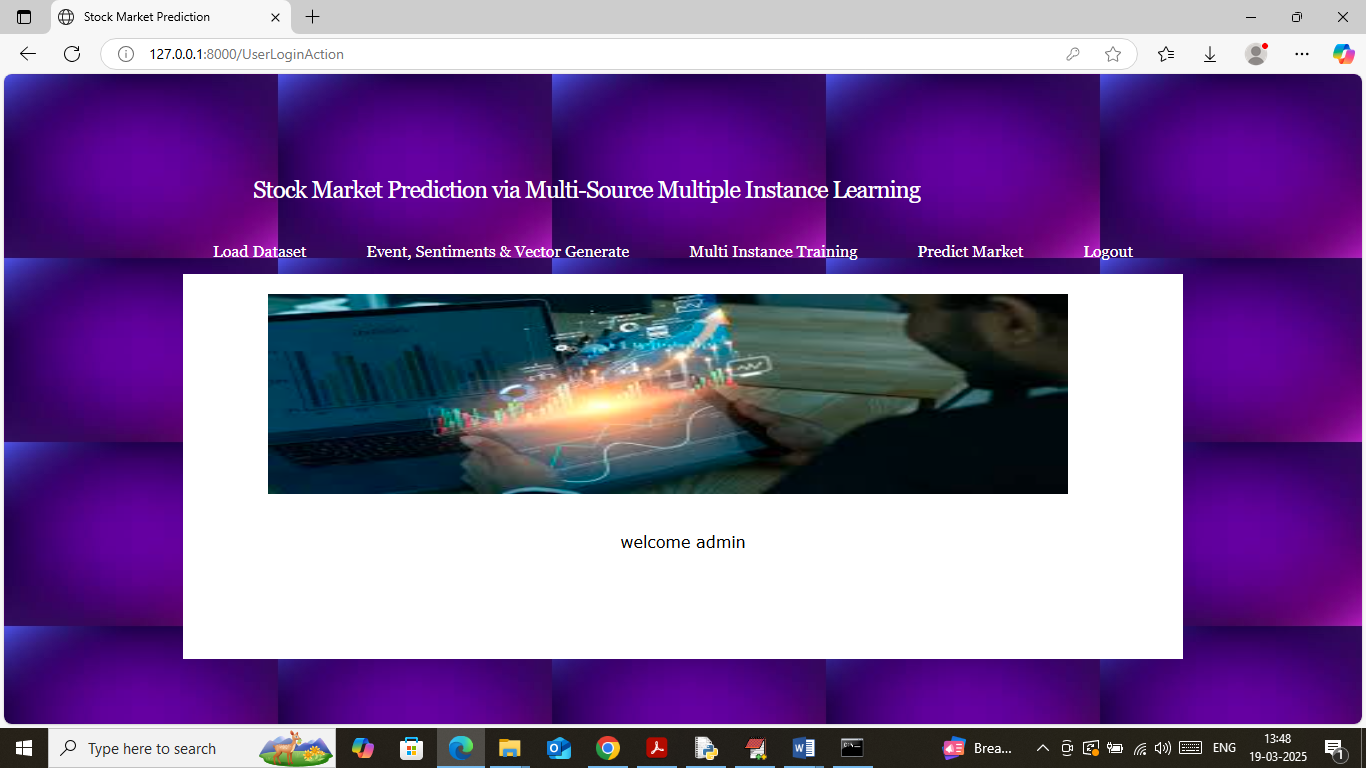
In above screen python web server started and now open browser and enter URL as <http://127.0.0.1:8000/index.html> and then press enter key to get below page



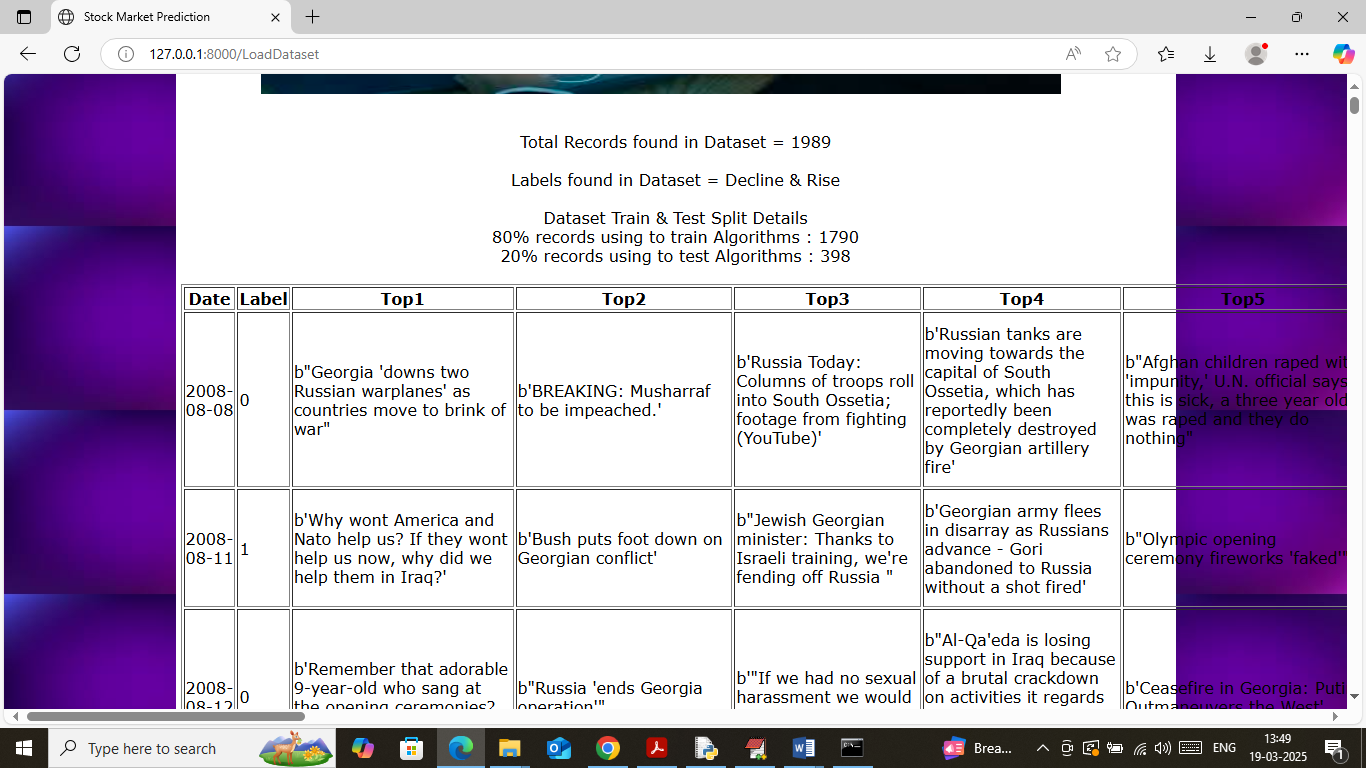
In above screen click on ‘User Login’ link to get below page



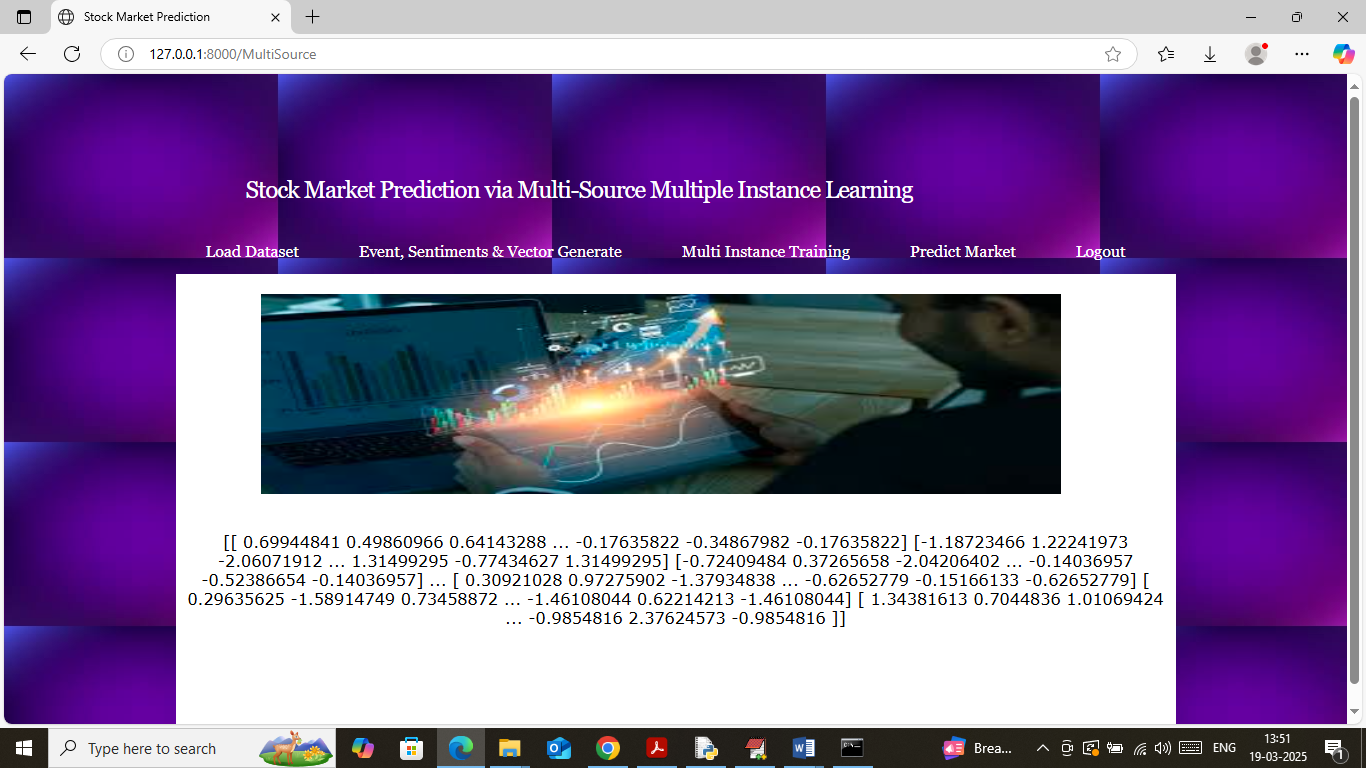
In above screen user is login and after login will get below page



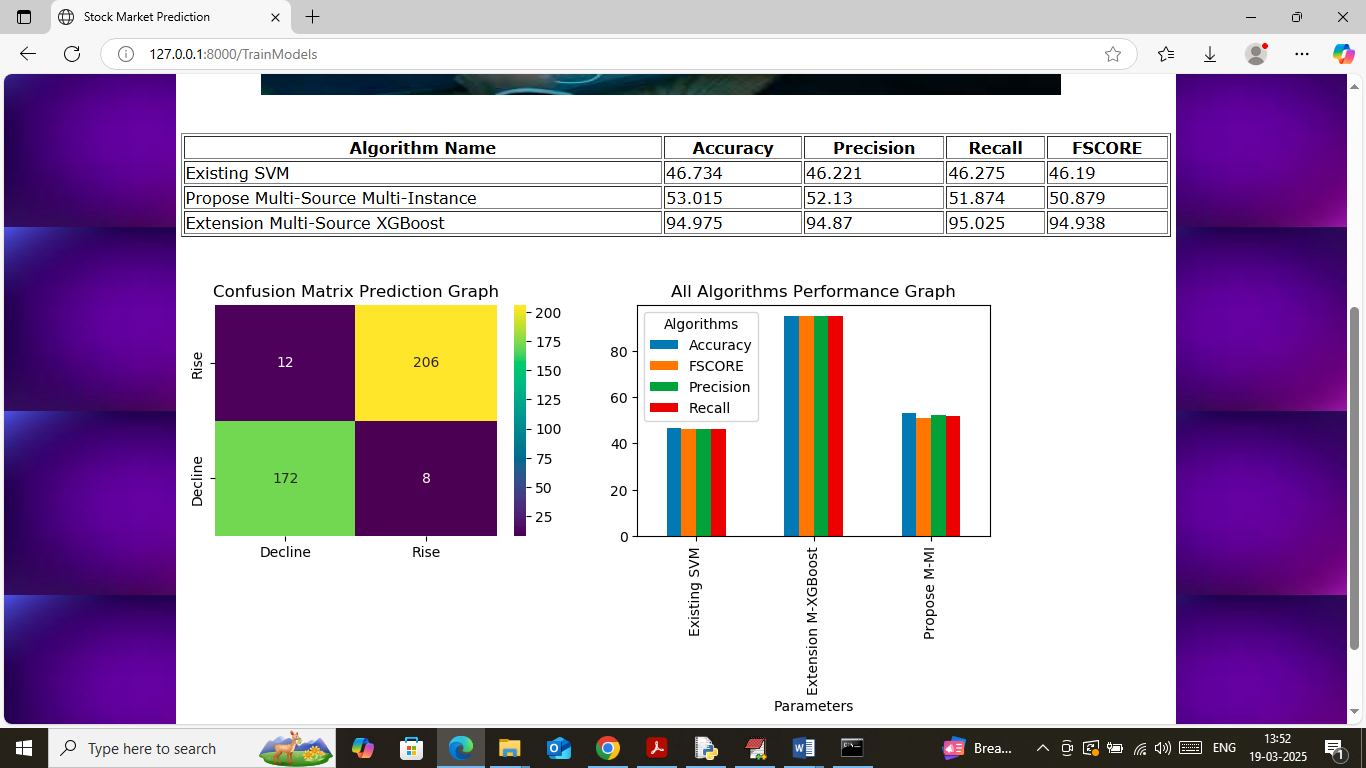
In above screen click on ‘Load Dataset’ link to load dataset and then will get below page



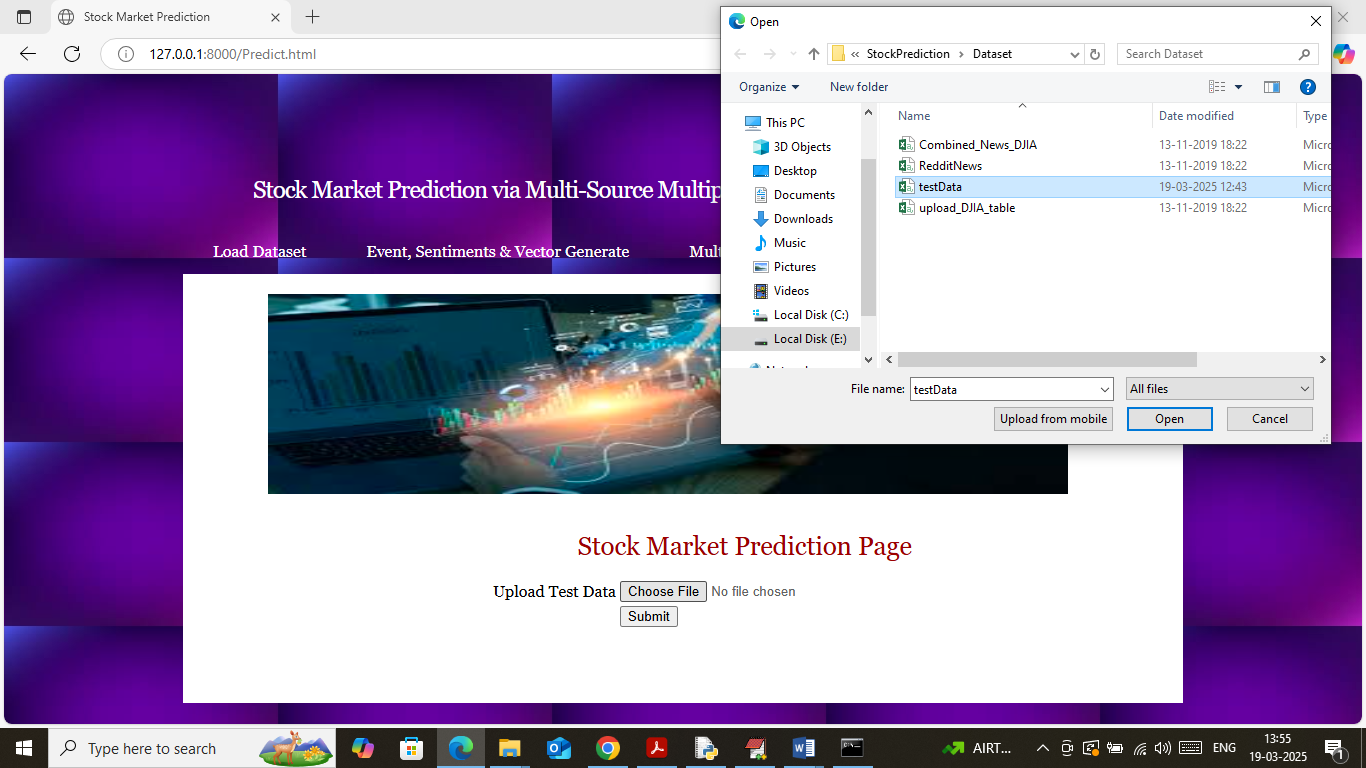
In above screen in first line can see total number of records found in dataset and then can see train and test size. In table format we can see all news data along with stock data. Now click on ‘Event, Sentiments & Vector Generate’ link to extract events, sentiments and vector from train and test data and then will get below page



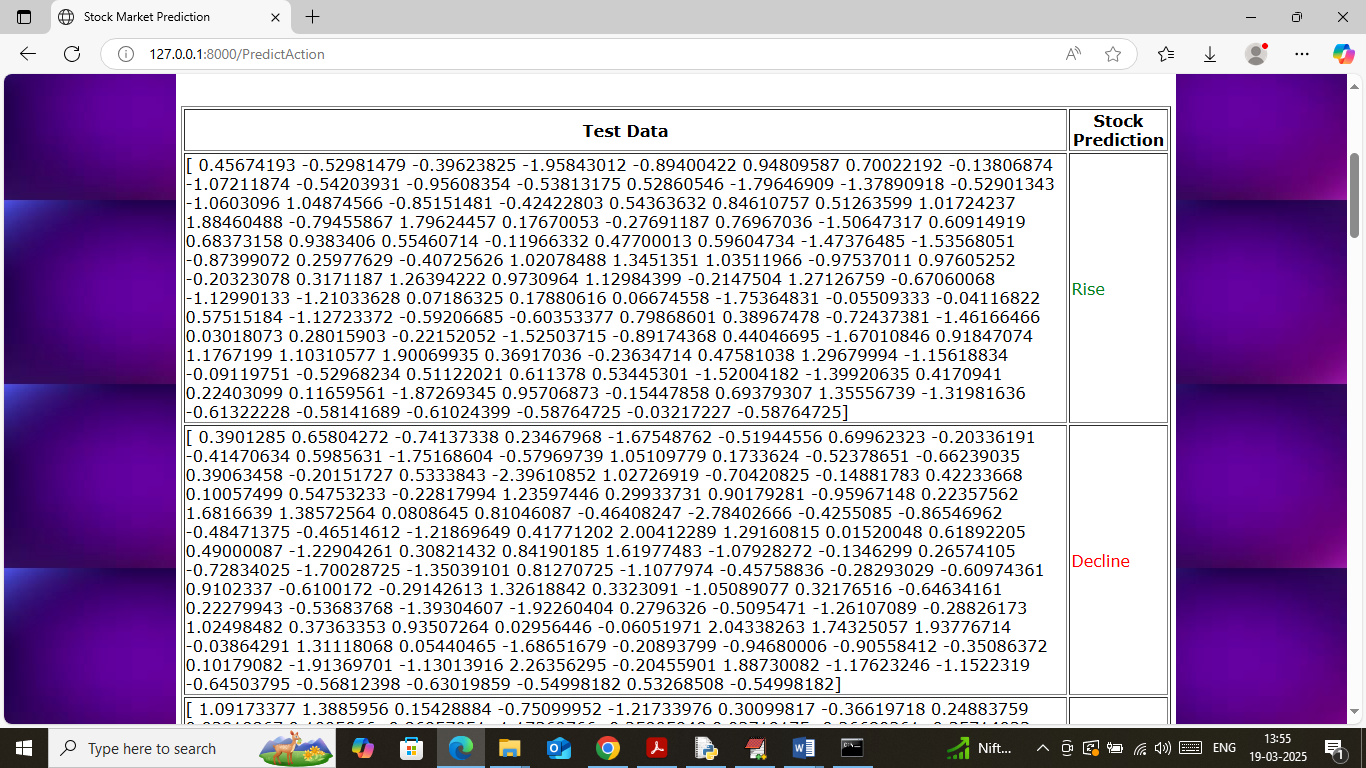
In above screen displaying some vector values generated from news and stock data and now click on ‘Multi Instance Training’ link to train all algorithms and then will get below page



In above screen in table format can see accuracy, precision, recall and FCSORE of Existing SVM, Propose Multi-Instance and extension XGBOOST algorithm and then can see XGBOOST got high accuracy compare to all algorithms. In above table can see Propose got 53% accuracy and extension XGBOOOST got 94% accuracy. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then yellow and green boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count which are very few. In bar graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. Now click on ‘Predict Market’ link to get below page



In above screen selecting and uploading test dataset and then click on buttons to load dataset and then will get below page



In above screen in first column can see generated vector from News and stock test data and then in second column displaying stock prediction status as “Rise or decline’ and based on prediction user may invest amount.

Similarly by following above screens you can run entire project