REPORT ON

SPAM NEWS DETECTION

With the rise of digital media and social networking, the distribution of fake or misleading news articles, often referred to as "spam news," has become a significant issue. Such misinformation can have adverse effects on public opinion, political discourse, and much more. This project focuses on developing a system for the automated detection of spam news articles using machine learning techniques.

Objective:

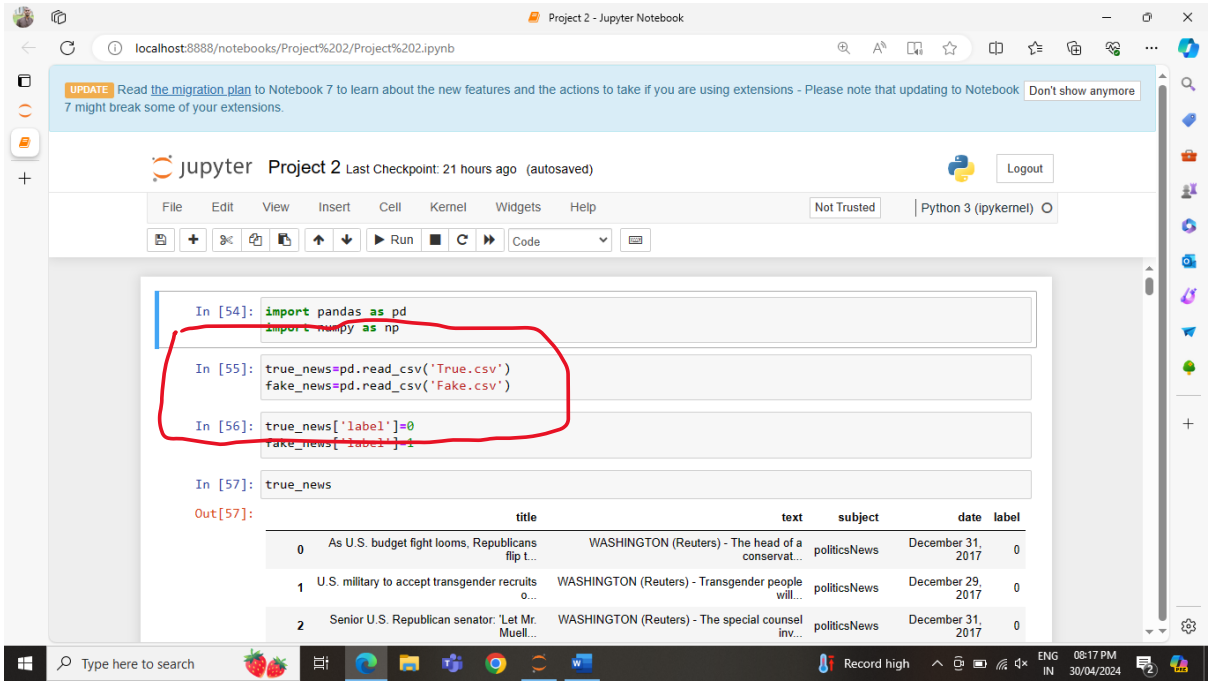
The main objective of this project is to design and implement a spam news detection system capable of accurately identifying and filtering out fake or misleading news articles from legitimate sources.

This Project is coded in python programming where the program is programmed in 5 sectors they are:-

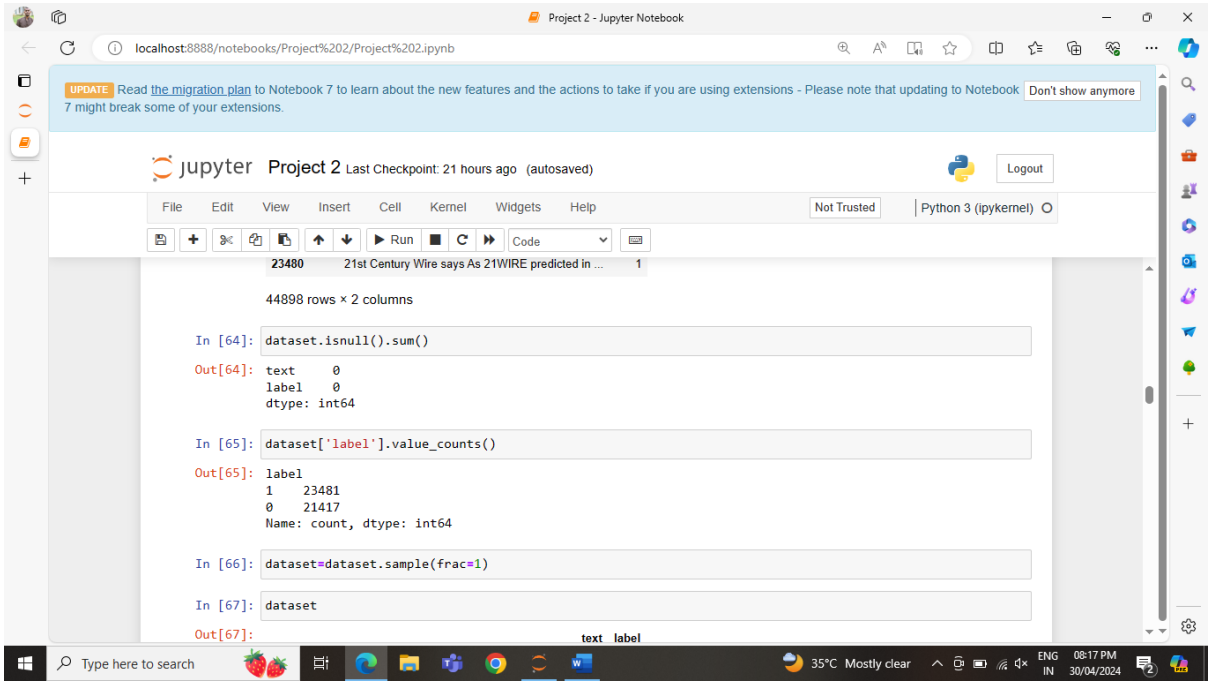
1. Loading the Dataset:-
   The given set of data with features are loaded in the form of either Mammogram Images or CSV File and here two datasets named True news and Fake news have been loaded.
2. Cleaning the loaded data:-
   The collected data undergo preprocessing techniques such as resizing, normalization, and noise removal to enhance their quality and suitability for analysis.
   Features such as word frequency, sentiment analysis, and readability scores are extracted from the pre-processed text using natural language processing (NLP) techniques.
3. Conversion of Categorical Data:-
   The conversion of categorical data involves transforming textual features extracted from news articles into numerical representations that can be used by machine learning algorithms. This conversion process facilitates the analysis and classification of news articles by converting categorical variables such as word presence or absence, sentiment scores, or readability metrics into numerical values, enabling the models to effectively learn and distinguish between spam and legitimate news content.
4. Fitting the training Data through the relavant Training Model:-
   The trained Model is Fitted into the relavant training Models like KNN,NAÏVE BAYES,LOGISTIC REGRESSION and much more
5. Prediction:-
   The performance of the developed models is evaluated using metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) through cross-validation techniques and Once the model achieves satisfactory performance, it is deployed as a user-friendly application or integrated into existing healthcare systems for real-time breast cancer detection.

Some of the screenshots of the Model is displayed below:-

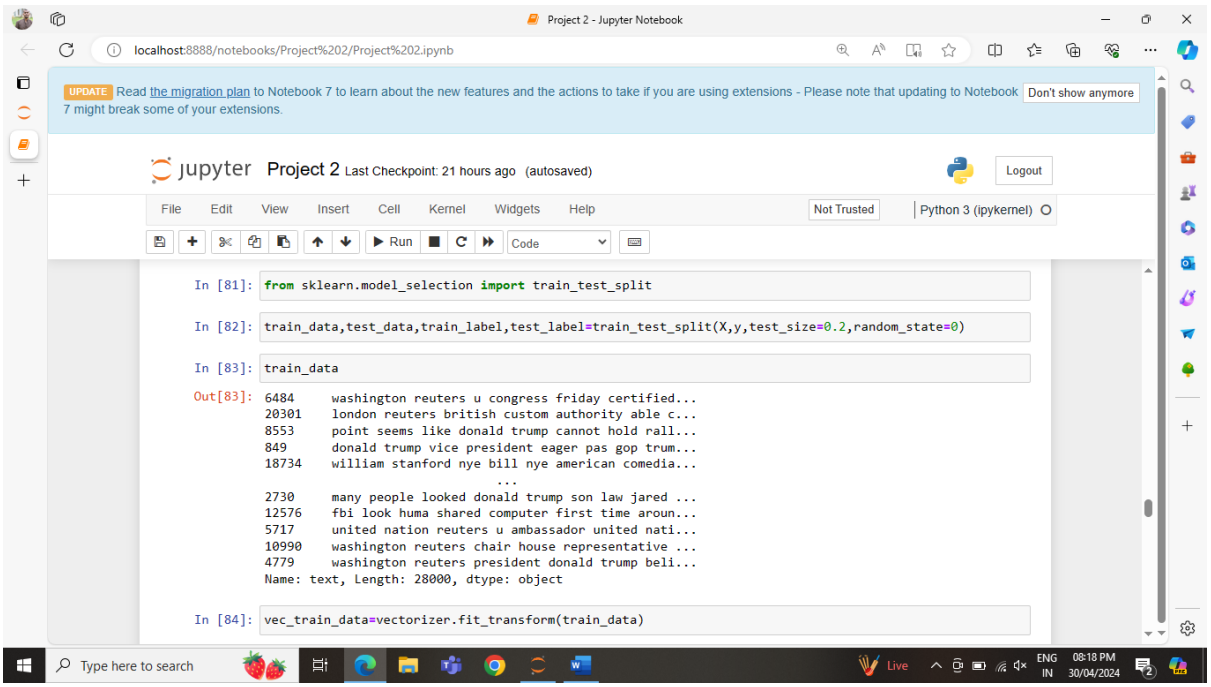
1. Loading the Dataset:-



2. Cleaning the loaded data:-

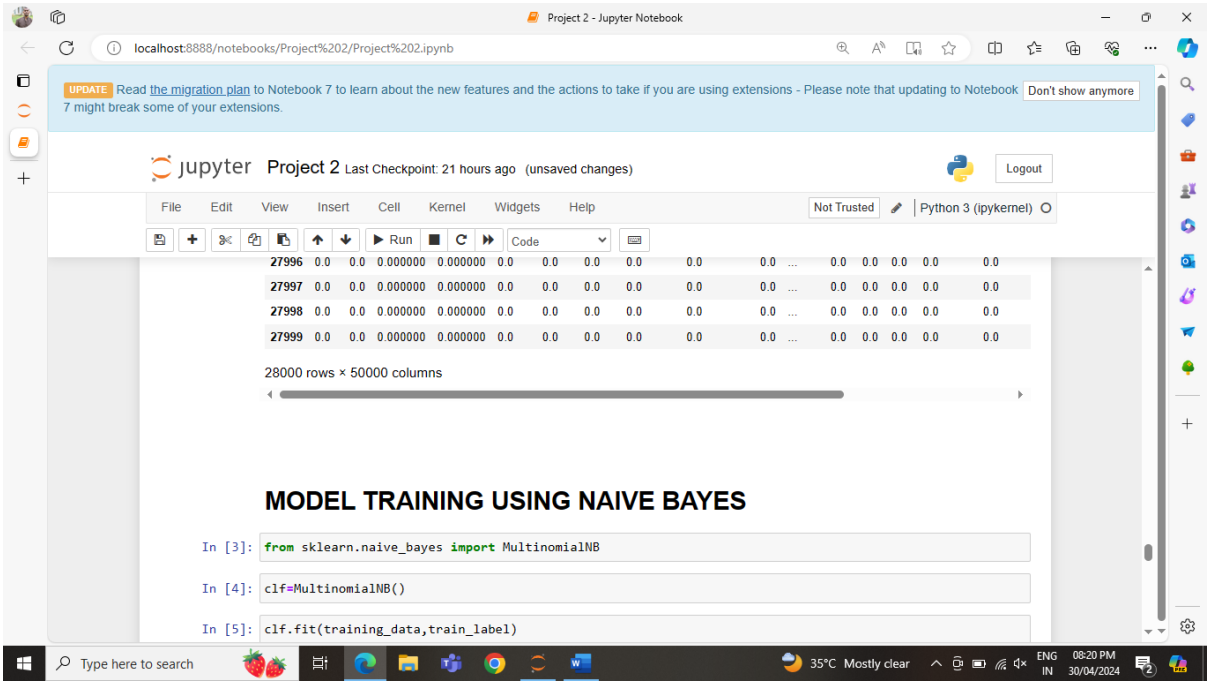


3. Conversion of Categorical Data:-

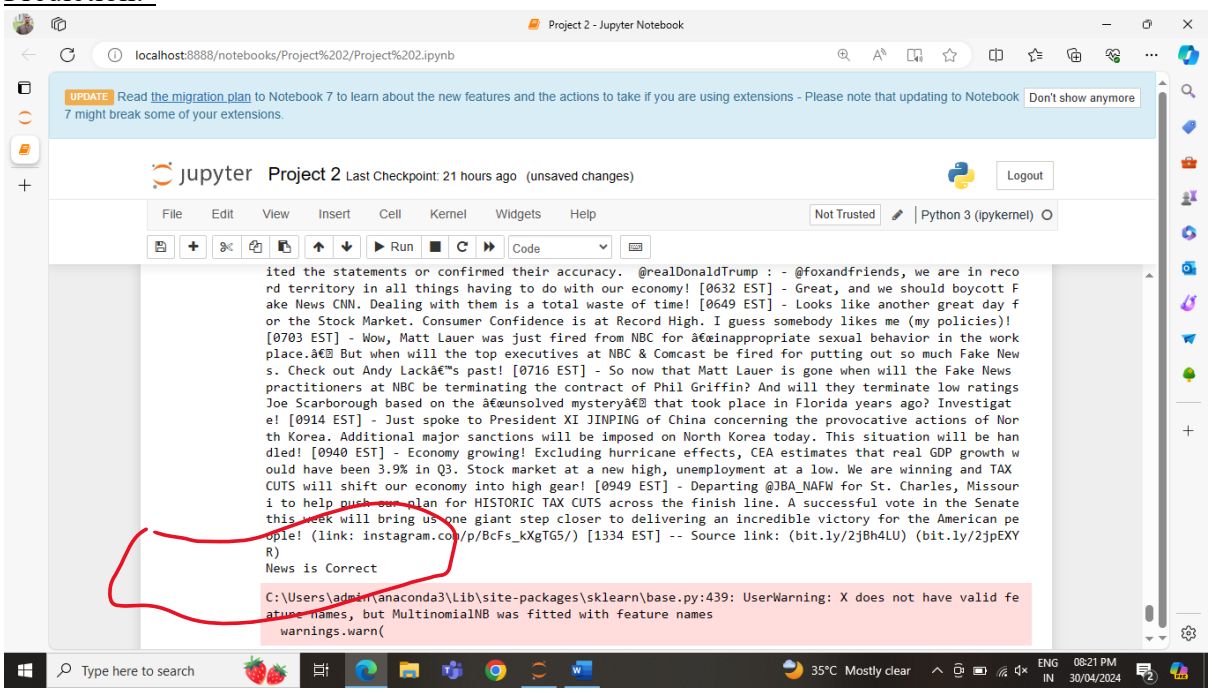


4. Fitting the training Data through the relavant Training Model:-

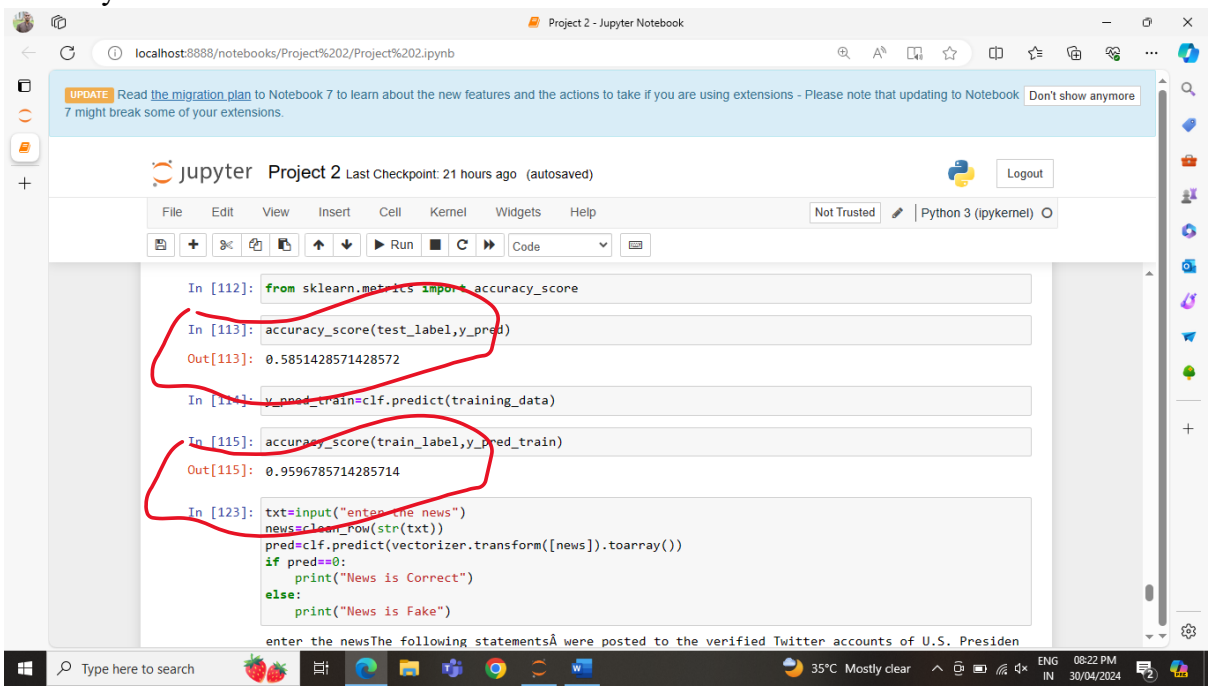
Modelling is done through Naïve Bayes Model as it is efficient in giving the accuracy for the trained and tested data



5. Prediction:-



As per the modelling Naïve bayes Modelling the accuracy of trained data is 95% and accuracy of tested data is 58%



.

Conclusion:-

The project successfully demonstrates the feasibility and effectiveness of using machine learning techniques for automated spam news detection. The developed system can play a crucial role in combating the spread of misinformation online and promoting media literacy among internet users.