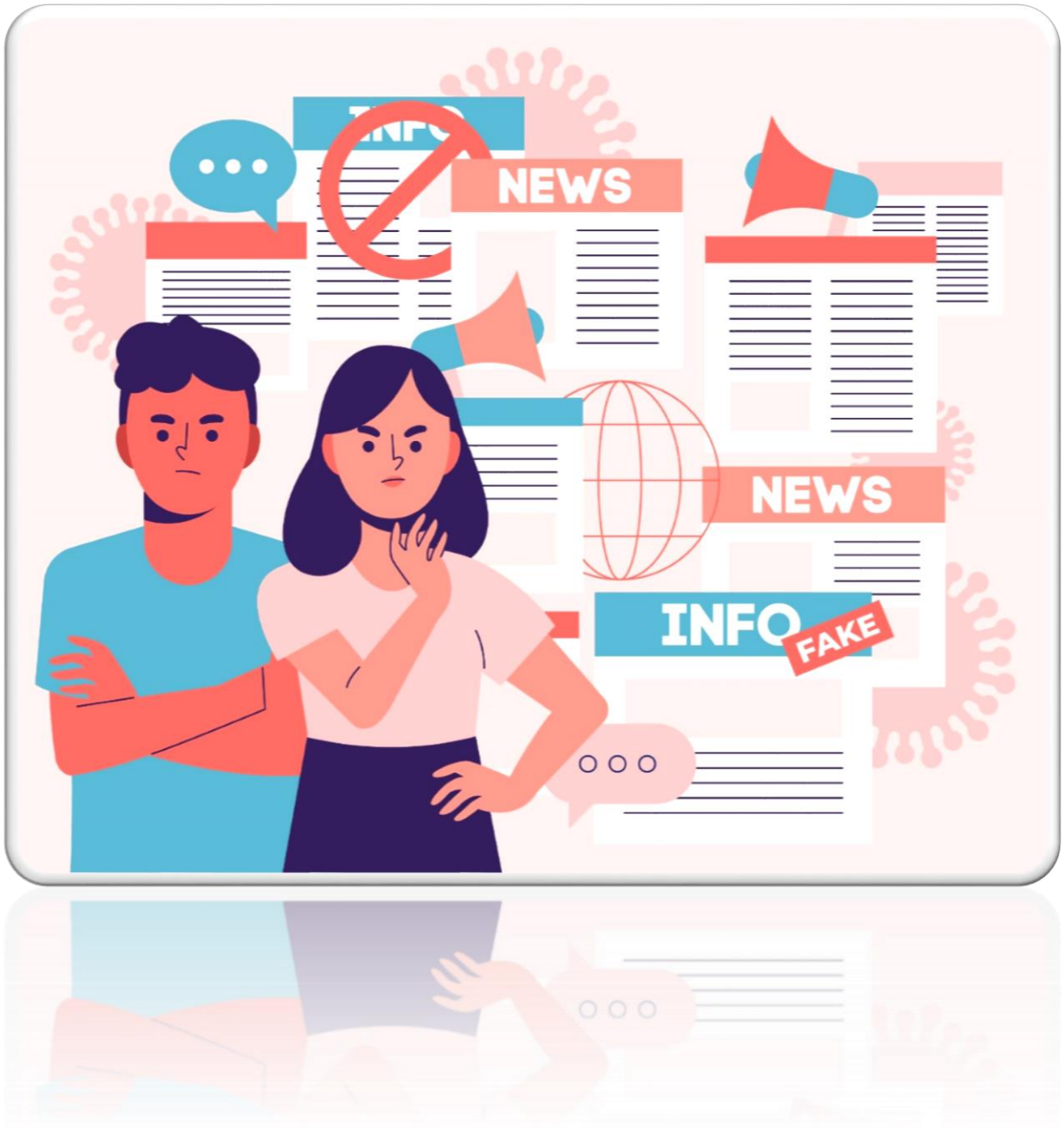
# Fake News Detection Using NLP

## Phase - 5



Team Member:

| Name | Reg No |
|---|---|
| ABDUR RAHIM S | 821021104003 |
| SiVA PRAKASH S | 821021104043 |
| VARUN RAJ M | 821021104055 |
| VIJAY RAJ B J | 821021104057 |
| SIVA G | 821021104304 |

# Table of Content

# Problem Statement

Design and develop an AI-based fake news detection system to automatically identify and categorize false or misleading information in textual content from online sources. The system should be capable of analyzing news articles, social media posts, and other forms of digital text, and provide a confidence score or classification indicating the likelihood of the content being fake or genuine. The goal is to assist users in making informed decisions about the credibility of the information they encounter on the internet, thereby combating the spread of misinformation and disinformation.

# Design Thinking

## DATA COLLECTION

***Collect and curate a diverse and extensive dataset of text-based content, including both fake and genuine news articles, from various sources and domains.***

## FEATURE ENGINEERING

***Develop techniques to extract relevant features from the text data, such as linguistic patterns, sentiment analysis, source credibility, and contextual information.***

## MACHINE LEARNING MODELS

***Implement and train machine learning models (e.g., natural language processing, deep learning, or ensemble models) on the labeled dataset to classify text as fake or genuine. Fine-tune the models for optimal performance.***

## REAL-TIME ANALYSIS

***Create a system that can process and analyze text content in real-time, allowing users to submit URLs, articles, or text snippets for verification.***

## CONFIDENCE SCORING

***Provide a confidence score or probability estimation for each classification to indicate the level of certainty in the model's prediction.***

## EXPLAINABILITY

***Incorporate interpretability techniques to explain why the AI system made a particular classification, enhancing user trust and transparency.***

## USER INTERFACE

***Design a user-friendly interface (e.g., web application or browser extension) that allows users to interact with the system easily and receive real-time feedback on the credibility of the content they encounter.***

## SCALABILITY

***Ensure the system can handle a high volume of requests and continuously update its dataset and models to adapt to evolving fake news tactics.***

## ETHICAL CONSIDERATIONS

***Address ethical concerns related to privacy, bias, and fairness in the data and model development process.***

## EVALUATION

***Evaluate the system's performance through rigorous testing, including metrics such as accuracy, precision, recall, and F1-score, using both historical and real-time data.***

## DEPLOYMENT

***Deploy the AI fake news detection system in a production environment, making it accessible to a wide range of users.***

## MONITORING AND MAINTENANCE

***Establish a monitoring system to detect and address model drift, data quality issues, and emerging fake news trends, ensuring the system's ongoing effectiveness.***

# The Phase Of The Development

## Phase -1

### **Module 1 :** Introduction to Artificial intelligence

Artificial intelligence (AI) is **the intelligence of machines or software, as opposed to the intelligence of humans or animals**. It is also the field of study in computer science that develops and studies intelligent machines.

### **Module 2 :** Python For Data Science

Python is a programming language widely used by **Data Scientists**. Python has **in-built mathematical libraries** and functions, making it easier to calculate **mathematical problems** and to perform data analysis. We will provide practical examples using Python.

# Phase -2

## **Module 3 :** Data Wrangling techniques

Data wrangling—also called data cleaning, data remediation, or data munging—refers to a variety of processes designed to transform raw data into more readily used formats. The exact methods differ from project to project depending on the data you're leveraging and the goal you're trying to achieve.

## **Module 4 :** Introduction to Neural Networks

A neural network is a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. It is a type of machine learning process, called deep learning, that uses interconnected nodes or neurons in a layered structure that resembles the human brain.

# Phase -3

## **Module 5 :** Tensorflow & keras ANN

Tensorflow is a powerful machine learning library to create models and neural networksTensorFlow is an open-sourced end-to-end platform, a library for multiple machine learning tasks, while Keras is a high-level neural network library that runs on top of TensorFlow. Both provide high-level APIs used for easily building and training models, but Keras is more user-friendly because it's built-in Python.

## **Module 6 :** Convolutional Neural Network

A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

# Phase - 4

## **Module 7 :** Object Detection using Yolo

YOLO is a real-time object detection algorithm that is faster and more accurate than many other CNN-based object detection algorithms. YOLO is also able to detect multiple objects in a single image, while many other CNN-based algorithms can only detect one object at a time.

## **Module 8 :** Recurrent Neural Network

A recurrent neural network is a type of artificial neural network commonly used in speech recognition and natural language processing. Recurrent neural networks recognize data's sequential characteristics and use patterns to predict the next likely scenario.

# **Module 10 :** Natural Language Processing

Natural language processing (NLP) is **a branch of artificial intelligence (AI) that allows machines to understand human language**. NLP has been around for over 50 years and has roots in linguistics. NLP's main purpose is to build systems that can make sense of text and automatically perform tasks like: Spell-check, Text translation, Topic classification

# Phase - 5

## **Module 11 :** IBM Cloud Watson AI Services

This section will explore the step-by-step process of various approaches to deploying machine learning models using popular frameworks like Python, Flask, Django, and Streamlit. Whether you want to build RESTful APIs with Flask, create scalable web applications using Django, or develop interactive interfaces using Streamlit, this section will help you understand how to successfully deploy your machine learning models and make them accessible to users. Let's dive in and learn how to bring your models to life in production environments.

## **Module 12 :** Build & Deploy ML Application

Model deployment in machine learning means integrating a trained machine-learning model into a real-world system or application to automatically generate predictions or perform specific tasks. For example, imagine a healthcare company developing a model to predict the chances of readmission for patients with chronic diseases. Model deployment would involve taking the trained model and implementing it within the company's existing electronic health record system. Once deployed, the model can analyze patient data in real-time, offering insights to healthcare professionals to help them identify high-risk patients and take proactive measures to avoid patient readmissions.

# ML Model Deployment in Python

***1. Data Preprocessing***

***2. Model Optimization and Training***

***3. Model Serialization***

***4. Prepare the Deployment Environment***

***5. Build The Deployment API***

***6. Test And Validate The Deployment***

***7. Deploy The ML Model***

***8. Monitor And Maintain The Deployment***

# Machine Learning Model Deployment Using Flask

1. Data Preprocessing And Preparation
2. Model Training And Optimization
3. Model Serialization
4. Set Up a Flask Application
5. Define an API Endpoint
6. Model Loading
7. Incoming Data Preprocessing
8. Generate Predictions
9. Run the Flask Application
10. Deploy the Flask Application
11. Test and Monitor the Deployment

# Dataset :

**Dataset Name : FAKE AND REAL DATASET**

**Dataset Link :**

**https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset**

**This Dataset File have Two Files :**

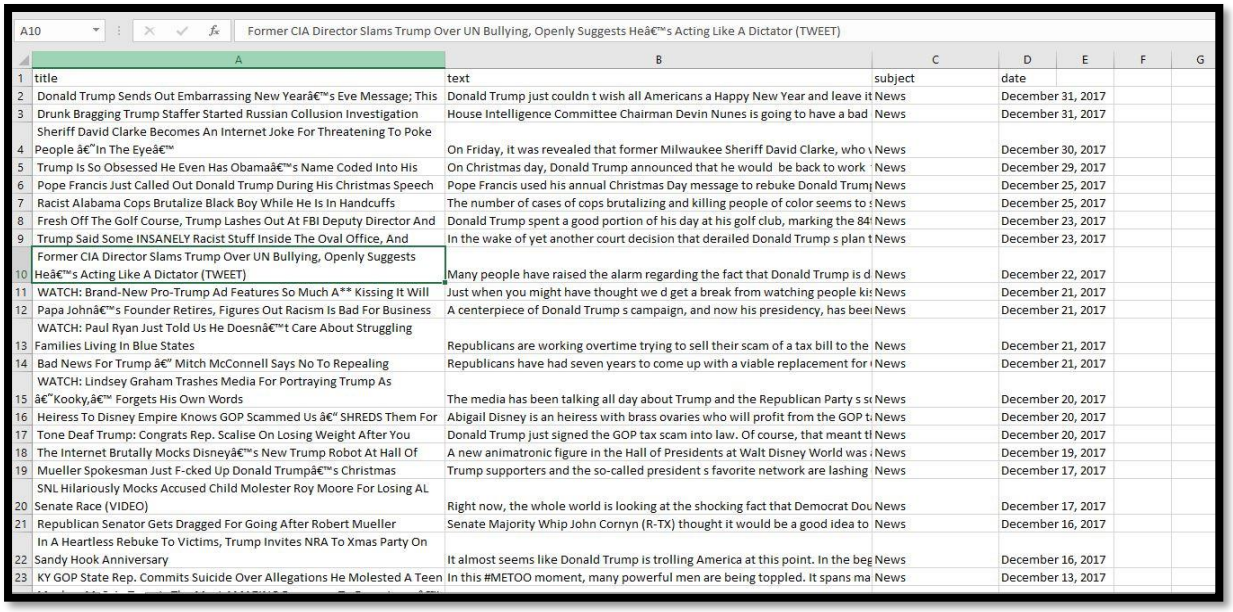
- **Fake.csv**
- **True.csv**

**About Dataset :**


1. *Ahmed H, Traore I, Saad S. "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018.*
2. *Ahmed H, Traore I, Saad S. (2017) "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127-138).*

## Fake.csv :

This dataset contains a list of articles considered as "fake" news.

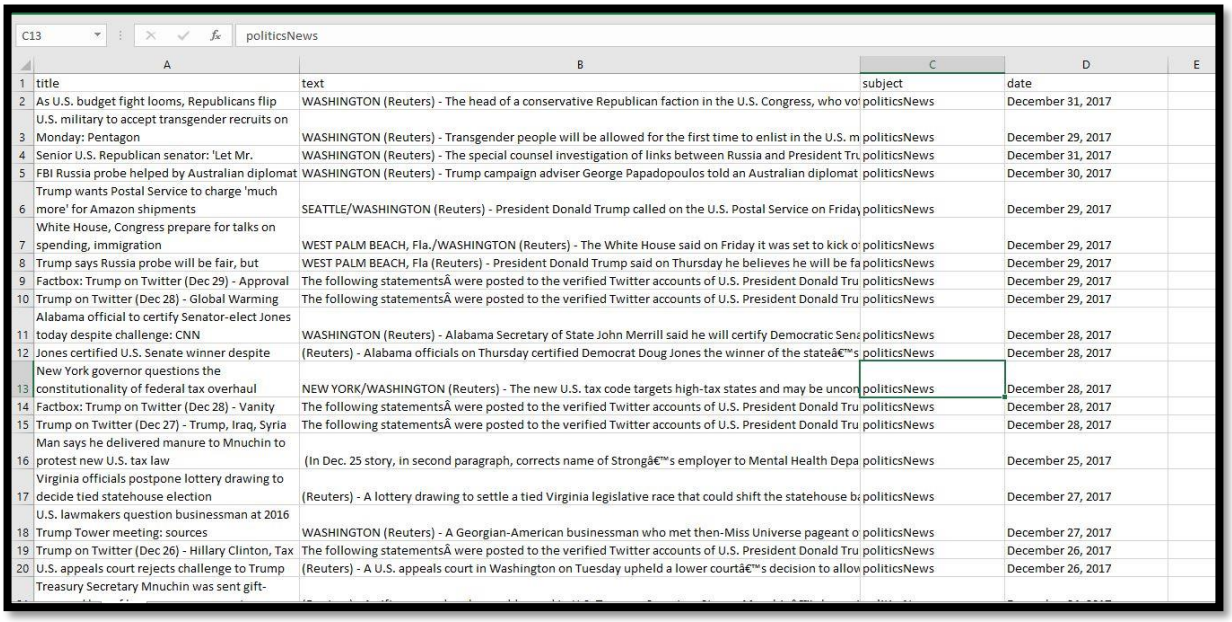


A10 — Former CIA Director Slams Trump Over UN Bullying, Openly Suggests Heâ€™s Acting Like A Dictator (TWEET)

| | A | B | C | D |
|---|---|---|---|---|
| 1 | title | text | subject | date |
| 2 | Donald Trump Sends Out Embarrassing New Yearâ€™s Eve Message; This | Donald Trump just couldn t wish all Americans a Happy New Year and leave it | News | December 31, 2017 |
| 3 | Drunk Bragging Trump Staffer Started Russian Collusion Investigation | House Intelligence Committee Chairman Devin Nunes is going to have a bad | News | December 31, 2017 |
| 4 | Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People â€˜In The Eyeâ€™ | On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who | News | December 30, 2017 |
| 5 | Trump Is So Obsessed He Even Has Obamaâ€™s Name Coded Into His | On Christmas day, Donald Trump announced that he would be back to work | News | December 29, 2017 |
| 6 | Pope Francis Just Called Out Donald Trump During His Christmas Speech | Pope Francis used his annual Christmas Day message to rebuke Donald Trump | News | December 25, 2017 |
| 7 | Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs | The number of cases of cops brutalizing and killing people of color seems to | News | December 25, 2017 |
| 8 | Fresh Off The Golf Course, Trump Lashes Out At FBI Deputy Director And | Donald Trump spent a good portion of his day at his golf club, marking the 84 | News | December 23, 2017 |
| 9 | Trump Said Some INSANELY Racist Stuff Inside The Oval Office, And | In the wake of yet another court decision that derailed Donald Trump s plan t | News | December 23, 2017 |
| 10 | Former CIA Director Slams Trump Over UN Bullying, Openly Suggests Heâ€™s Acting Like A Dictator (TWEET) | Many people have raised the alarm regarding the fact that Donald Trump is d | News | December 22, 2017 |
| 11 | WATCH: Brand-New Pro-Trump Ad Features So Much A** Kissing It Will | Just when you might have thought we d get a break from watching people kis | News | December 21, 2017 |
| 12 | Papa Johnâ€™s Founder Retires, Figures Out Racism Is Bad For Business | A centerpiece of Donald Trump s campaign, and now his presidency, has bee | News | December 21, 2017 |
| 13 | WATCH: Paul Ryan Just Told Us He Doesnâ€™t Care About Struggling Families Living In Blue States | Republicans are working overtime trying to sell their scam of a tax bill to the | News | December 21, 2017 |
| 14 | Bad News For Trump â€“ Mitch McConnell Says No To Repealing | Republicans have had seven years to come up with a viable replacement for | News | December 21, 2017 |
| 15 | WATCH: Lindsey Graham Trashes Media For Portraying Trump As â€˜Kooky,â€™ Forgets His Own Words | The media has been talking all day about Trump and the Republican Party s s | News | December 20, 2017 |
| 16 | Heiress To Disney Empire Knows GOP Scammed Us â€“ SHREDS Them For | Abigail Disney is an heiress with brass ovaries who will profit from the GOP t | News | December 20, 2017 |
| 17 | Tone Deaf Trump: Congrats Rep. Scalise On Losing Weight After You | Donald Trump just signed the GOP tax scam into law. Of course, that meant t | News | December 20, 2017 |
| 18 | The Internet Brutally Mocks Disneyâ€™s New Trump Robot At Hall Of | A new animatronic figure in the Hall of Presidents at Walt Disney World was | News | December 19, 2017 |
| 19 | Mueller Spokesman Just F-cked Up Donald Trumpâ€™s Christmas | Trump supporters and the so-called president s favorite network are lashing | News | December 17, 2017 |
| 20 | SNL Hilariously Mocks Accused Child Molester Roy Moore For Losing AL Senate Race (VIDEO) | Right now, the whole world is looking at the shocking fact that Democrat Dou | News | December 17, 2017 |
| 21 | Republican Senator Gets Dragged For Going After Robert Mueller | Senate Majority Whip John Cornyn (R-TX) thought it would be a good idea to | News | December 16, 2017 |
| 22 | In A Heartless Rebuke To Victims, Trump Invites NRA To Xmas Party On Sandy Hook Anniversary | It almost seems like Donald Trump is trolling America at this point. In the beg | News | December 16, 2017 |
| 23 | KY GOP State Rep. Commits Suicide Over Allegations He Molested A Teen | In this #METOO moment, many powerful men are being toppled. It spans ma | News | December 13, 2017 |

## True.csv :

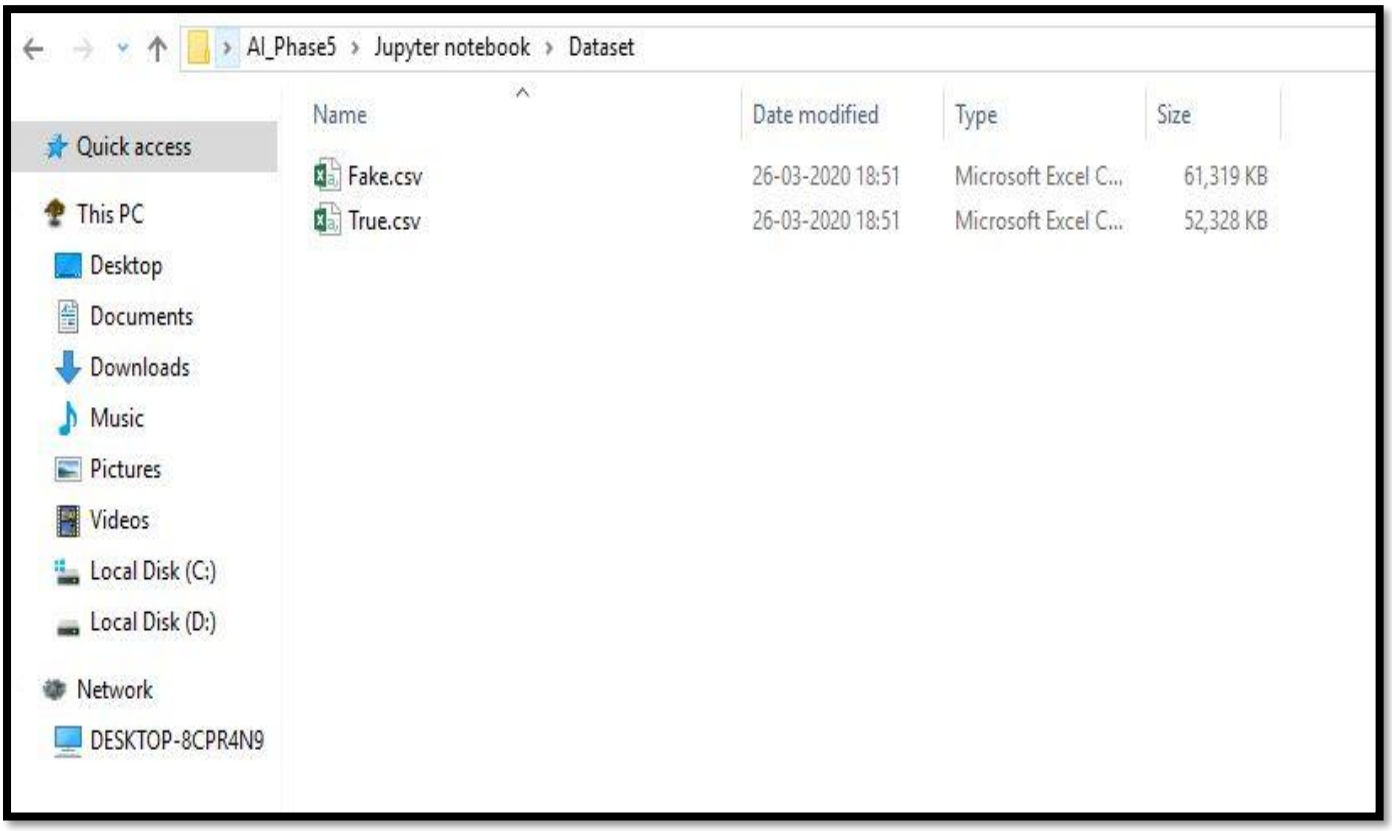
This dataset contains a list of articles considered as "real" news.



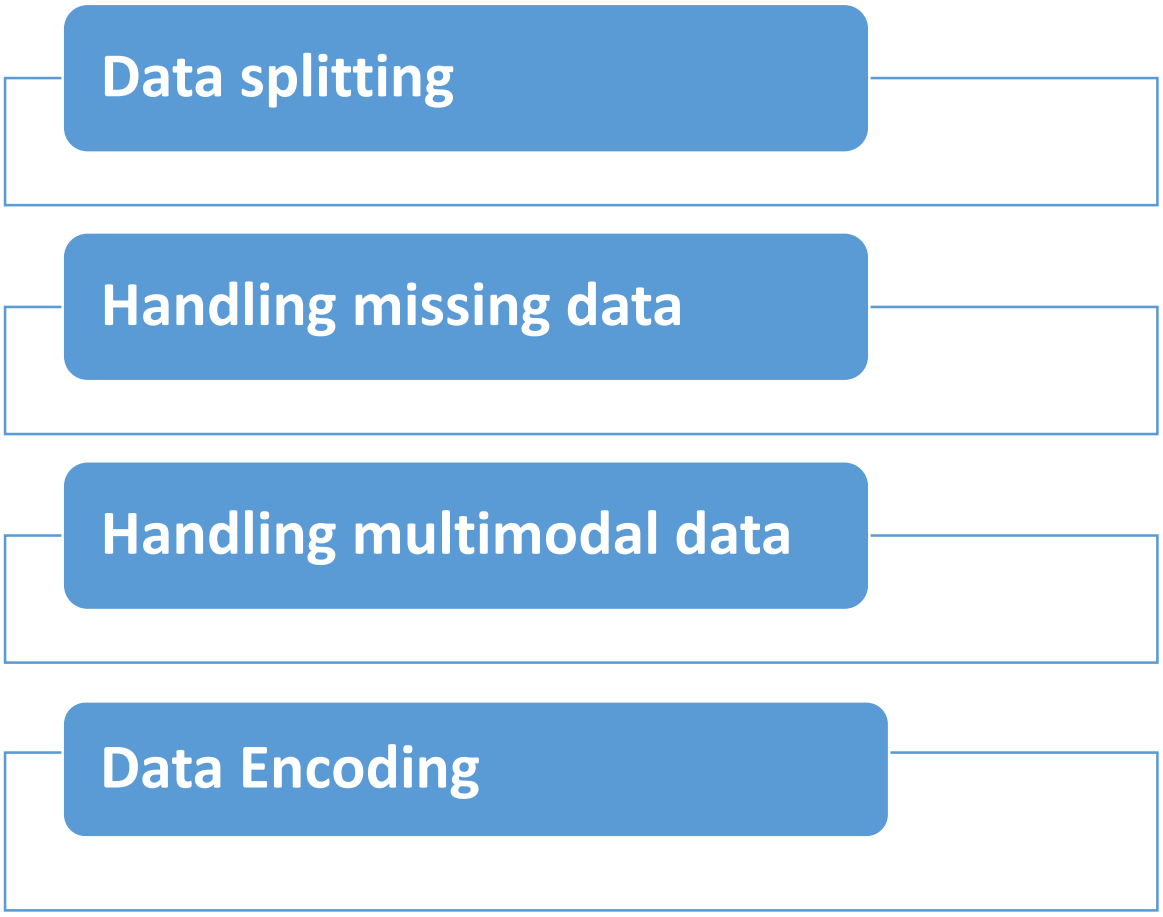
C13 — politicsNews

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | title | text | subject | date | |
| 2 | As U.S. budget fight looms, Republicans flip | WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who vo | politicsNews | December 31, 2017 | |
| 3 | U.S. military to accept transgender recruits on Monday: Pentagon | WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. m | politicsNews | December 29, 2017 | |
| 4 | Senior U.S. Republican senator: 'Let Mr. | WASHINGTON (Reuters) - The special counsel investigation of links between Russia and President Tru | politicsNews | December 31, 2017 | |
| 5 | FBI Russia probe helped by Australian diplomat | WASHINGTON (Reuters) - Trump campaign adviser George Papadopoulos told an Australian diplomat | politicsNews | December 30, 2017 | |
| 6 | Trump wants Postal Service to charge 'much more' for Amazon shipments | SEATTLE/WASHINGTON (Reuters) - President Donald Trump called on the U.S. Postal Service on Friday | politicsNews | December 29, 2017 | |
| 7 | White House, Congress prepare for talks on spending, immigration | WEST PALM BEACH, Fla./WASHINGTON (Reuters) - The White House said on Friday it was set to kick o | politicsNews | December 29, 2017 | |
| 8 | Trump says Russia probe will be fair, but | WEST PALM BEACH, Fla (Reuters) - President Donald Trump said on Thursday he believes he will be fa | politicsNews | December 29, 2017 | |
| 9 | Factbox: Trump on Twitter (Dec 29) - Approval | The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Tru | politicsNews | December 29, 2017 | |
| 10 | Trump on Twitter (Dec 28) - Global Warming | The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Tru | politicsNews | December 29, 2017 | |
| 11 | Alabama official to certify Senator-elect Jones today despite challenge: CNN | WASHINGTON (Reuters) - Alabama Secretary of State John Merrill said he will certify Democratic Sena | politicsNews | December 28, 2017 | |
| 12 | Jones certified U.S. Senate winner despite | (Reuters) - Alabama officials on Thursday certified Democrat Doug Jones the winner of the stateâ€™s | politicsNews | December 28, 2017 | |
| 13 | New York governor questions the constitutionality of federal tax overhaul | NEW YORK/WASHINGTON (Reuters) - The new U.S. tax code targets high-tax states and may be uncon | politicsNews | December 28, 2017 | |
| 14 | Factbox: Trump on Twitter (Dec 28) - Vanity | The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Tru | politicsNews | December 28, 2017 | |
| 15 | Trump on Twitter (Dec 27) - Trump, Iraq, Syria | The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Tru | politicsNews | December 28, 2017 | |
| 16 | Man says he delivered manure to Mnuchin to protest new U.S. tax law | (In Dec. 25 story, in second paragraph, corrects name of Strongâ€™s employer to Mental Health Depa | politicsNews | December 25, 2017 | |
| 17 | Virginia officials postpone lottery drawing to decide tied statehouse election | (Reuters) - A lottery drawing to settle a tied Virginia legislative race that could shift the statehouse ba | politicsNews | December 27, 2017 | |
| 18 | U.S. lawmakers question businessman at 2016 Trump Tower meeting: sources | WASHINGTON (Reuters) - A Georgian-American businessman who met then-Miss Universe pageant o | politicsNews | December 27, 2017 | |
| 19 | Trump on Twitter (Dec 26) - Hillary Clinton, Tax | The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Tru | politicsNews | December 26, 2017 | |
| 20 | U.S. appeals court rejects challenge to Trump | (Reuters) - A U.S. appeals court in Washington on Tuesday upheld a lower courtâ€™s decision to allow | politicsNews | December 26, 2017 | |
| | Treasury Secretary Mnuchin was sent gift- | | | | |

**Project Path :** ***..\..\AI_Phase5\Jupyter notebook\Dataset***



# Data Preprocessing :


Data splitting
Handling missing data
Handling multimodal data
Data Encoding

## Classification Algorithm and Model Training Process.

### Data Collection:

Gather a dataset of news articles labeled as either "real" or "fake." This dataset is used for training and evaluating the model.

### Data Preprocessing:

- Text Cleaning: Remove HTML tags, special characters, and other noise from the text.
- Tokenization: Split the text into words or subword tokens.
- Stopword Removal: Eliminate common words that don't carry much information.
- Lemmatization or Stemming: Reduce words to their base form.
- Feature Extraction: Convert text into numerical features (e.g., TF-IDF, Word Embeddings like Word2Vec or GloVe).

### Data Splitting:

Divide the dataset into training, validation, and test sets. Typically, an 80-20 or 70-30 split is used.

**Model Selection:**

- Choose a machine learning or deep learning model for classification. Common models include:
  - Logistic Regression
  - Naive Bayes
  - Random Forest
  - Support Vector Machine (SVM)
  - Recurrent Neural Networks (RNN)
  - Convolutional Neural Networks (CNN)
  - Transformers like BERT or GPT

**Model Training:**

- Train the selected model on the training data.
- Model parameters are adjusted to minimize the loss function, which measures the difference between predicted and actual labels.
- Hyperparameter tuning may be performed to optimize model performance.

**Model Evaluation:**

- Assess the model's performance using the validation dataset. Common evaluation metrics include accuracy, precision, recall, F1-score, and confusion matrix.
- Fine-tune the model based on validation results.

**Model Testing:**

- Test the final model on the held-out test dataset to evaluate its generalization to unseen data.

**Post-processing:**

- Apply post-processing techniques to refine model predictions, such as threshold adjustment or rule-based filtering.

**Deployment:**

- Integrate the model into a real-time system or application for fake news detection.

**Monitoring and Maintenance:**

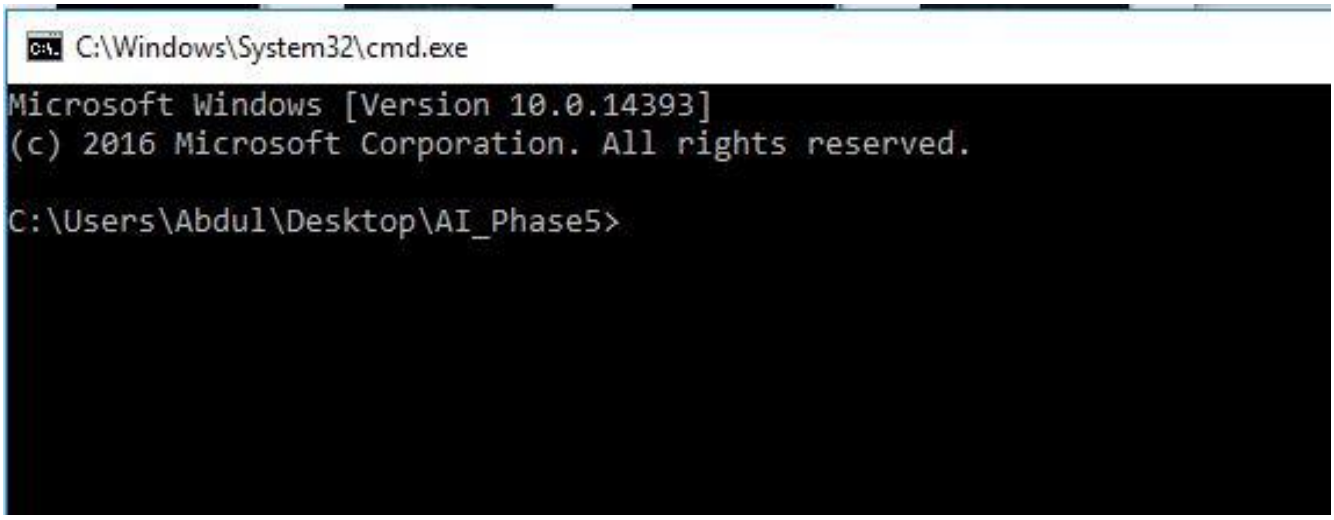
- Continuously monitor the model's performance and retrain it periodically to adapt to evolving fake news patterns.

## Project Structure :

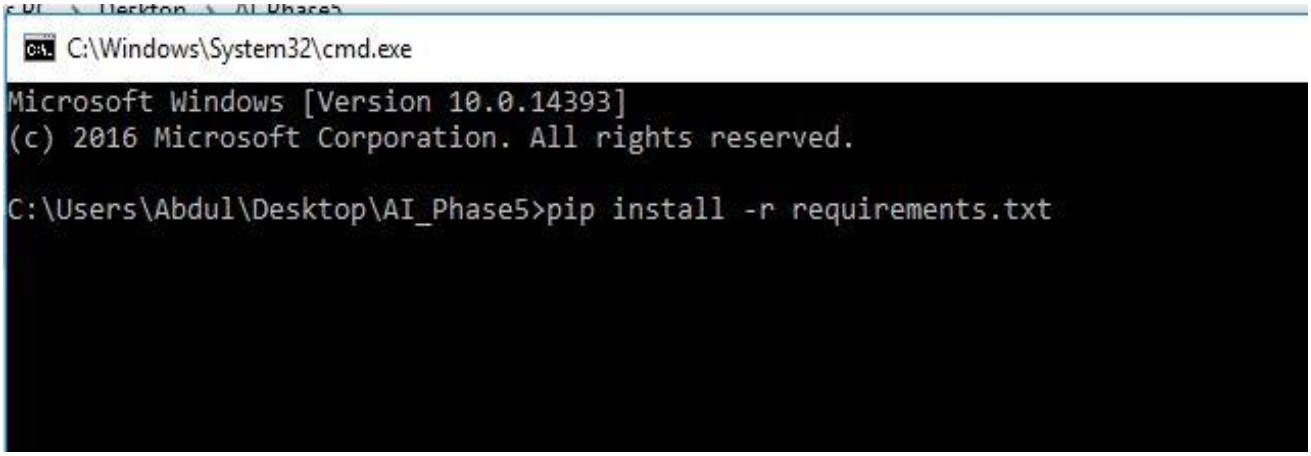
- **Jupyter notebook**
  - **Dataset**
    - *Fake.csv*
    - *True.csv*
  - **Static**
    - *style.css*
  - **templates**
    - *index.html*
    - *scripts.js*
- *Main.py*
- *model.pkl*
- *news.csv*
- *requirements.txt*

# Execution Steps :

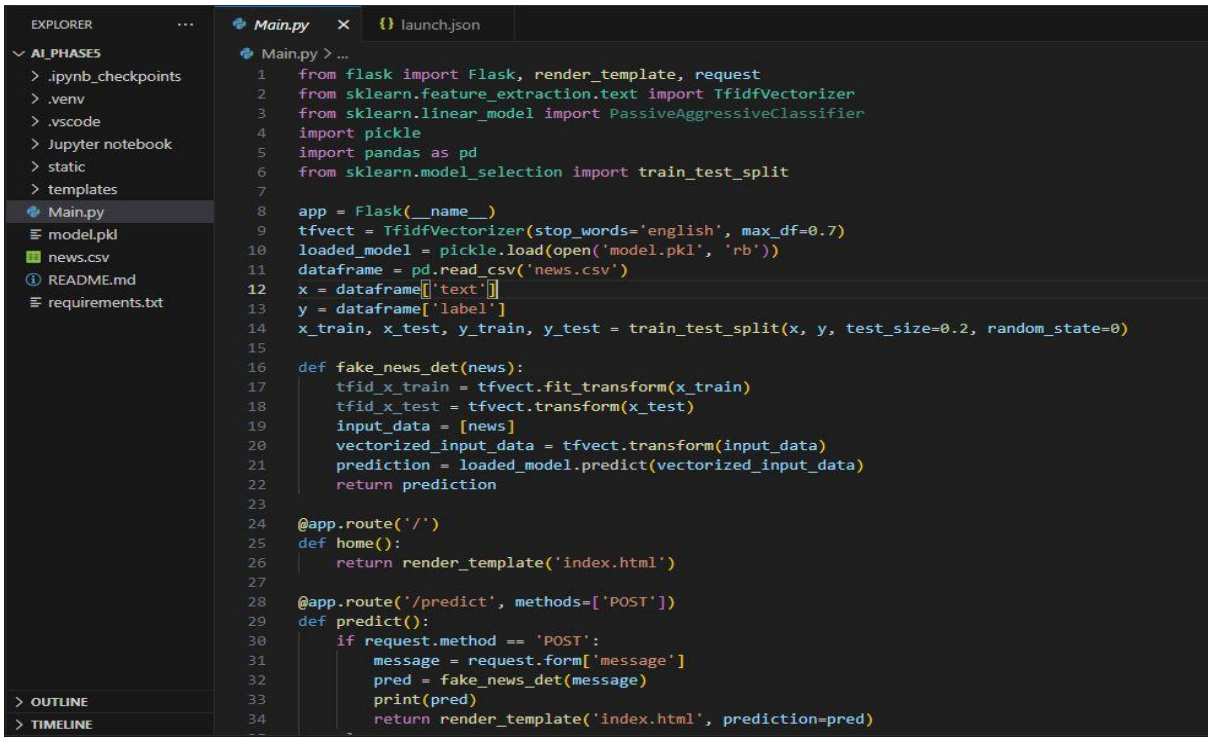
- Open **CMD** in working directory.



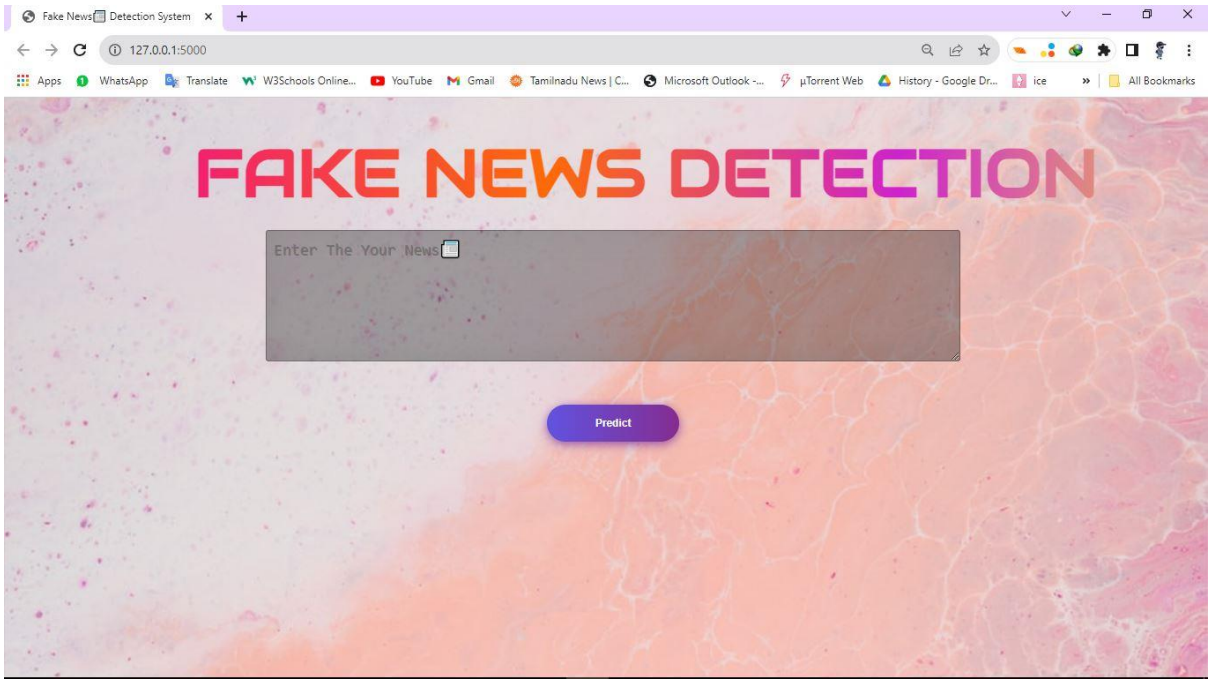

- Run `**pip install -r requirements.txt**`



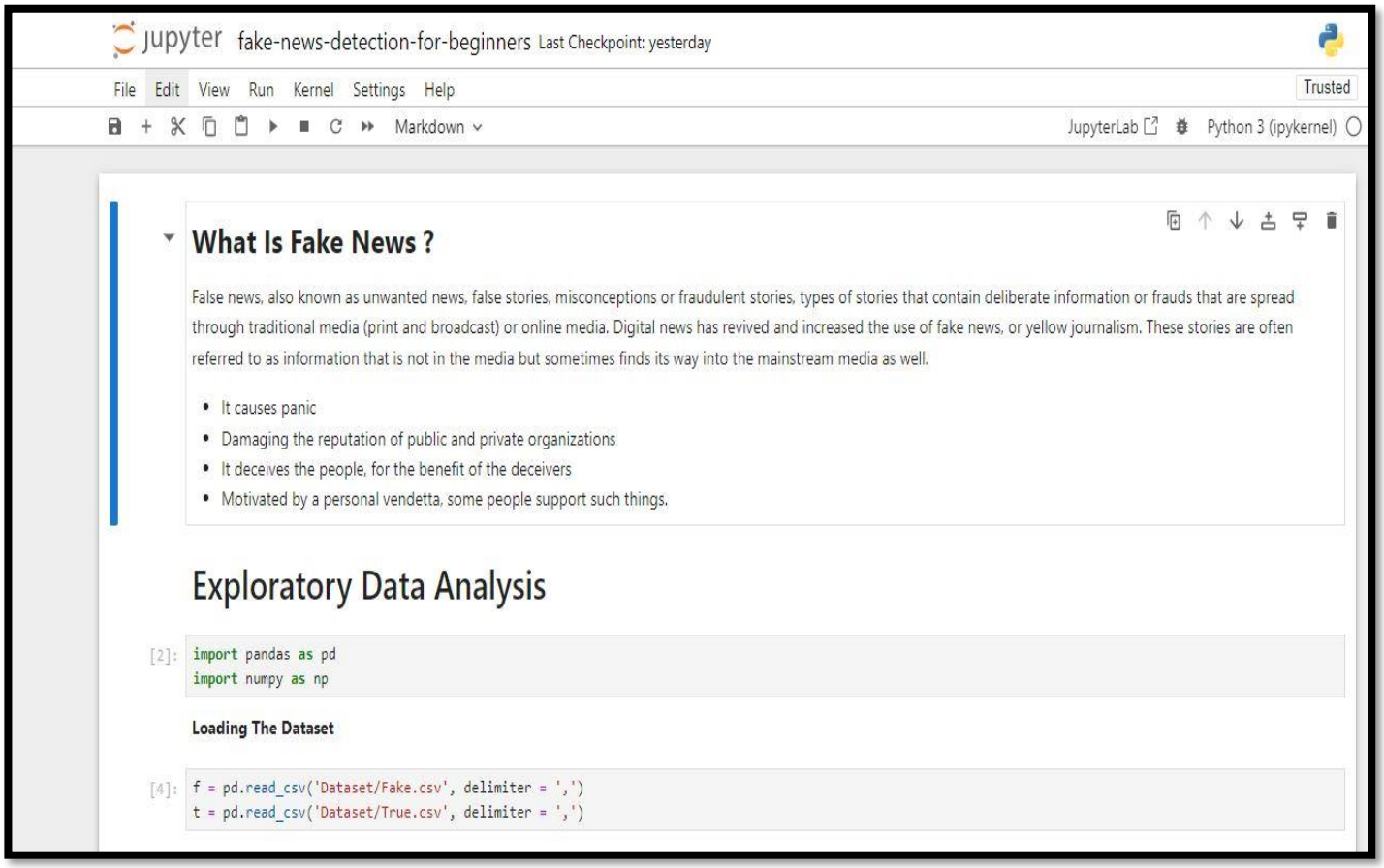
➢Open project in any IDE(Pycharm or **VSCode**)




➢Run `**Main.py**`, go to the `***http://127.0.0.1:5000/***`

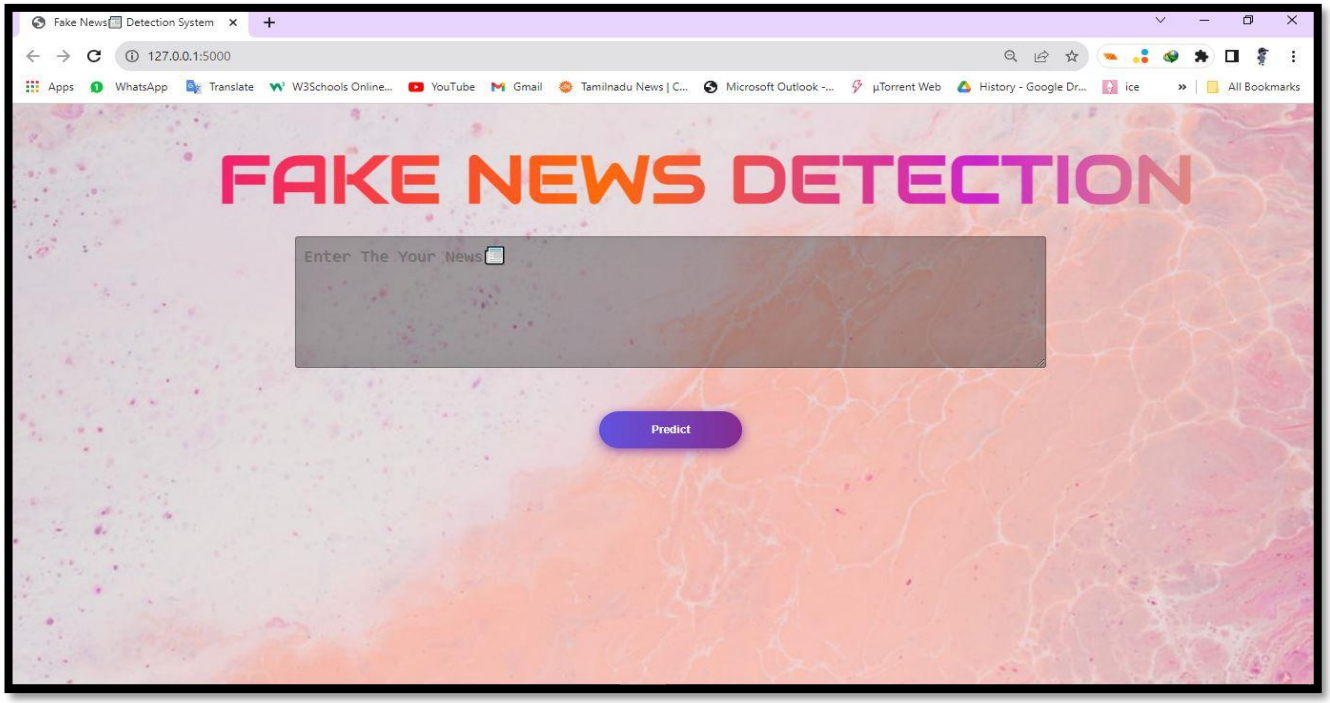


➢If you want to build your model with the some changes, you can check the `**Jupyter notebook/Fake_News_Detection.ipynb**`.



# Deploy :

➢Open The Chrome Output:



➢Enter The news in Text area Box :

Input : U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sundayâ€™s unity march against terrorism.


Then Click Predict button.

Input : 18 mins ago 3 Views 0 Comments 0 Likes It's fair to say, Europe's been shocked by Trump's Europe's Trump's RT LIVE http://rt.com/on-air

## Conclusion :

**In this part of your project, you've loaded and preprocessed the fake news dataset. Preprocessing is a critical step in building a reliable fake news detection model. After preprocessing, you can proceed to feature engineering, model selection, training, and evaluation. Your choice of machine learning or deep learning algorithms will depend on the dataset size and complexity. Make sure to continuously monitor and fine-tune your model to improve its performance. Lastly, be prepared to handle real-world scenarios where fake news can evolve and adapt, so regular updates and retraining might be necessary for a robust solution.**