DETECTION OF FAKE NEWS ARTICLES USING NLP

JAGRITI SHARMA(11021210010)

INTRODUCTION TO WHY WE NEED TO CLASSIFY REAL & FAKE NEWS

- Fake news refers to false or misleading information presented as news. This can take many forms, including fabricated stories, manipulated images or videos, and distorted or misrepresented facts. Fake news can have serious consequences, including misinforming the public, damaging reputations, and inciting violence or social unrest. It can also undermine trust in the media and democratic institutions.

PROBLEM STATEMENT

The objective of our problem statement is to do Natural Language Processing on news articles that are fake using AI and ML.

We have used various Python Libraries and Modules to achieve the goal, some of them include:

NLTK

TENSORFLOW

KERAS

INTRODUCTION TO NLP

- Natural Language Processing (NLP) is an exciting field that combines artificial intelligence and linguistic analysis to enable computers to understand, interpret, and generate human language.

- With the power of AI&ML (Artificial Intelligence Markup Language), NLP opens up a world of possibilities for automated language processing, chatbots, and intelligent virtual assistants

- One of the most important applications of NLP is to recognize fake news articles and detection of false information

Techniques used

Logistic Regression: Logistic Regression is a "Supervised machine learning" algorithm that can be used to model the probability of a certain class or event. It is used when the data is linearly separable and the outcome is binary or dichotomous in nature.

MultinomialNB: Multinomial Naive Bayes algorithm is a probabilistic learning method that is mostly used in Natural Language Processing (NLP). The algorithm is based on the Bayes theorem and predicts the tag of a text such as a piece of email or newspaper article. It calculates the probability of each tag for a given sample and then gives the tag with the highest probability as output

"We have used MultinomialNB as it has a better accuracy score."

OUR OBJECTIVE

OUR OBJECTIVE IS TO DEVELOP A CODE THAT CAN VISUALIZE REAL AND FAKE NEWS ARTICLES USING VARIOUS PYTHON LIBRARIES. WITH THIS CODE, USERS CAN EXPLORE THE DIFFERENCES BETWEEN REAL AND FAKE NEWS ARTICLES.

WE USE THE NAIVE BAYES PRINCIPLE TO PREDICT THE ACCURACY OF THE TF-IDF MODEL, WHICH IS COMMONLY USED IN NATURAL LANGUAGE PROCESSING TASKS.

BY COMBINING THESE TECHNIQUES, OUR CODE PROVIDES USERS WITH A POWERFUL TOOL FOR ANALYZING AND DETECTING FAKE NEWS. WHETHER YOU'RE A JOURNALIST, RESEARCHER, OR CONCERNED CITIZEN, OUR CODE CAN HELP YOU BETTER UNDERSTAND THE LANGUAGE OF FAKE NEWS AND IMPROVE YOUR ABILITY TO IDENTIFY IT.

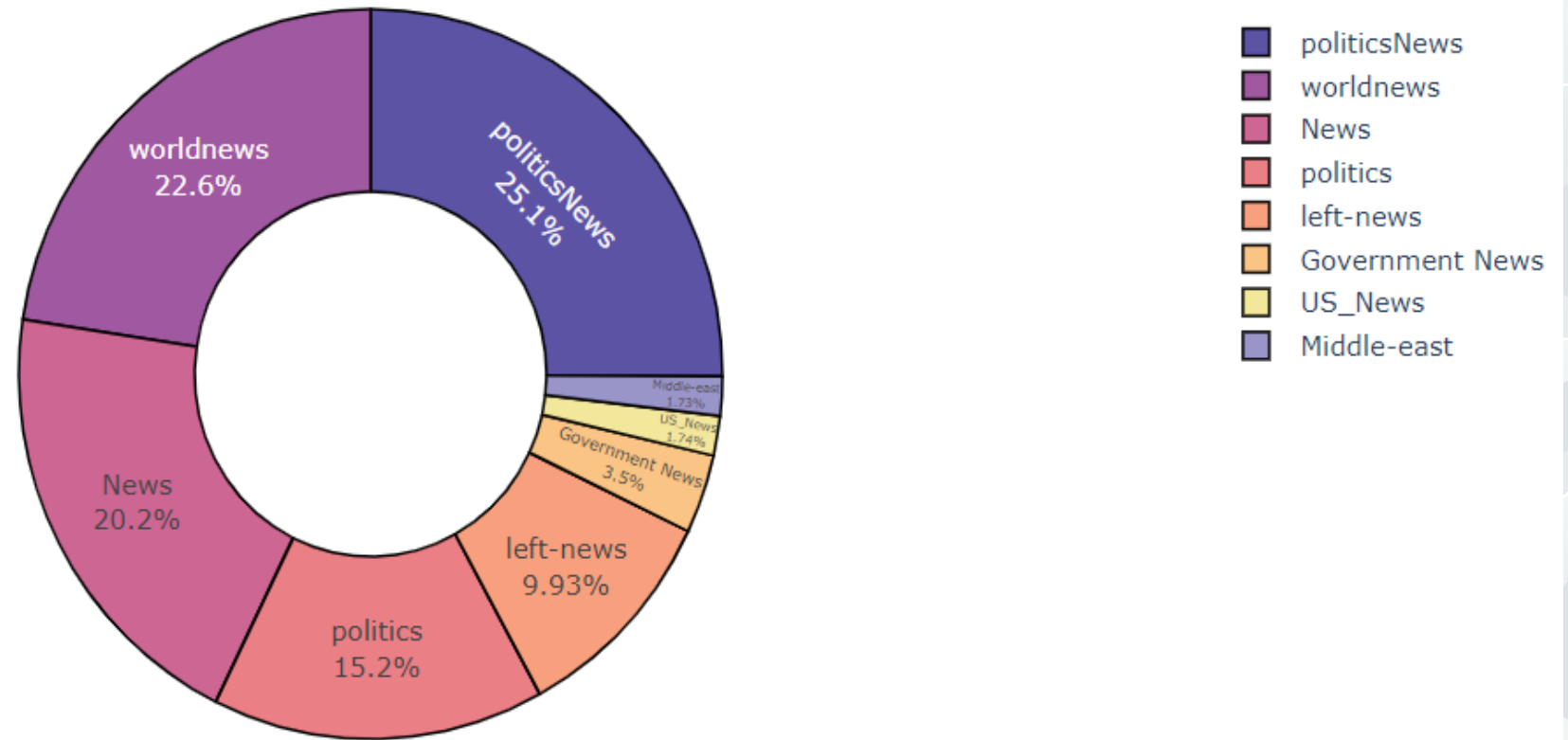# LET'S SEE SOME VISUALIZATIONS OF OUR DATA

This is a very simple yet descriptive Bar Plot which tells us about the quantity of Fake and Real news articles present in the dataset.
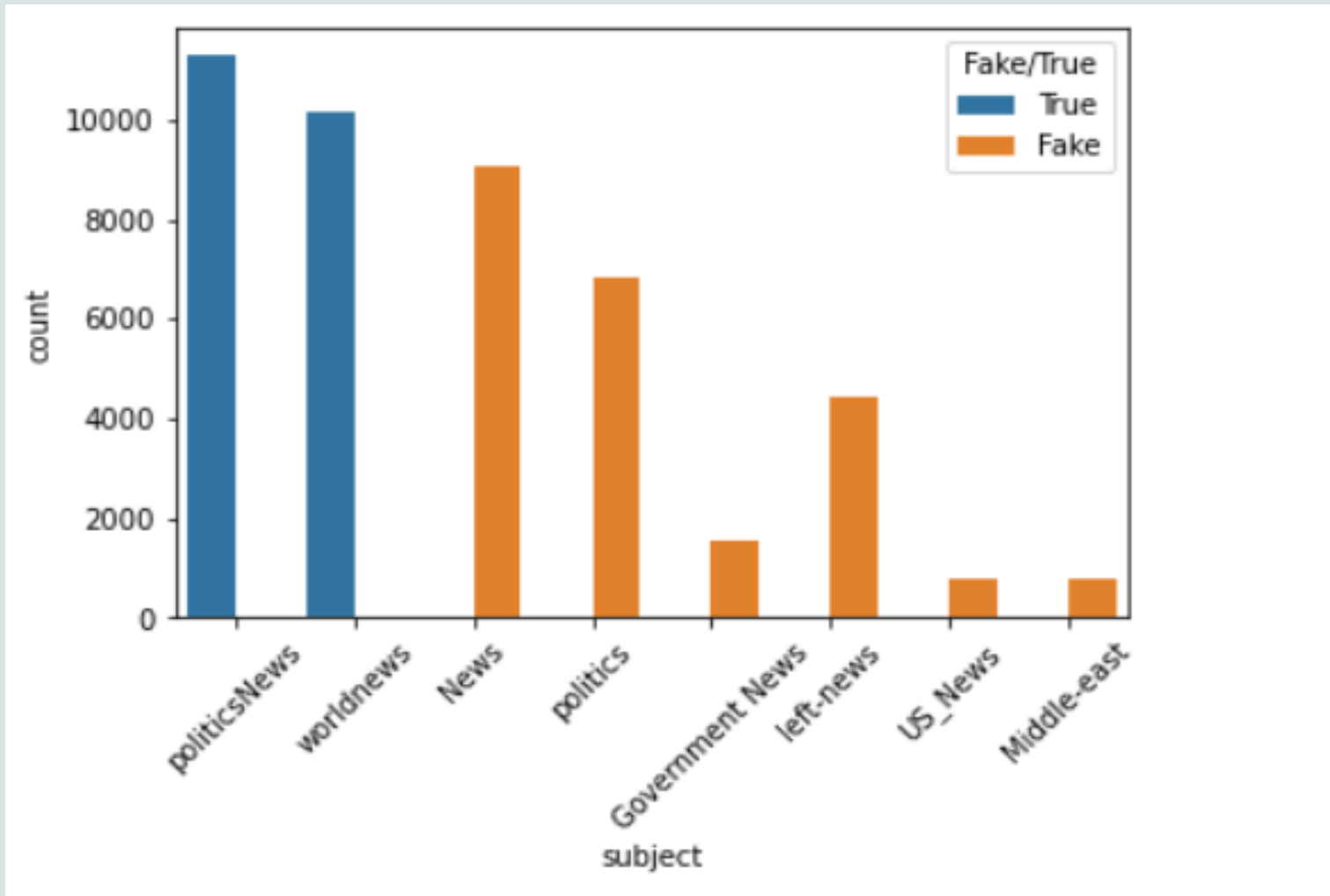

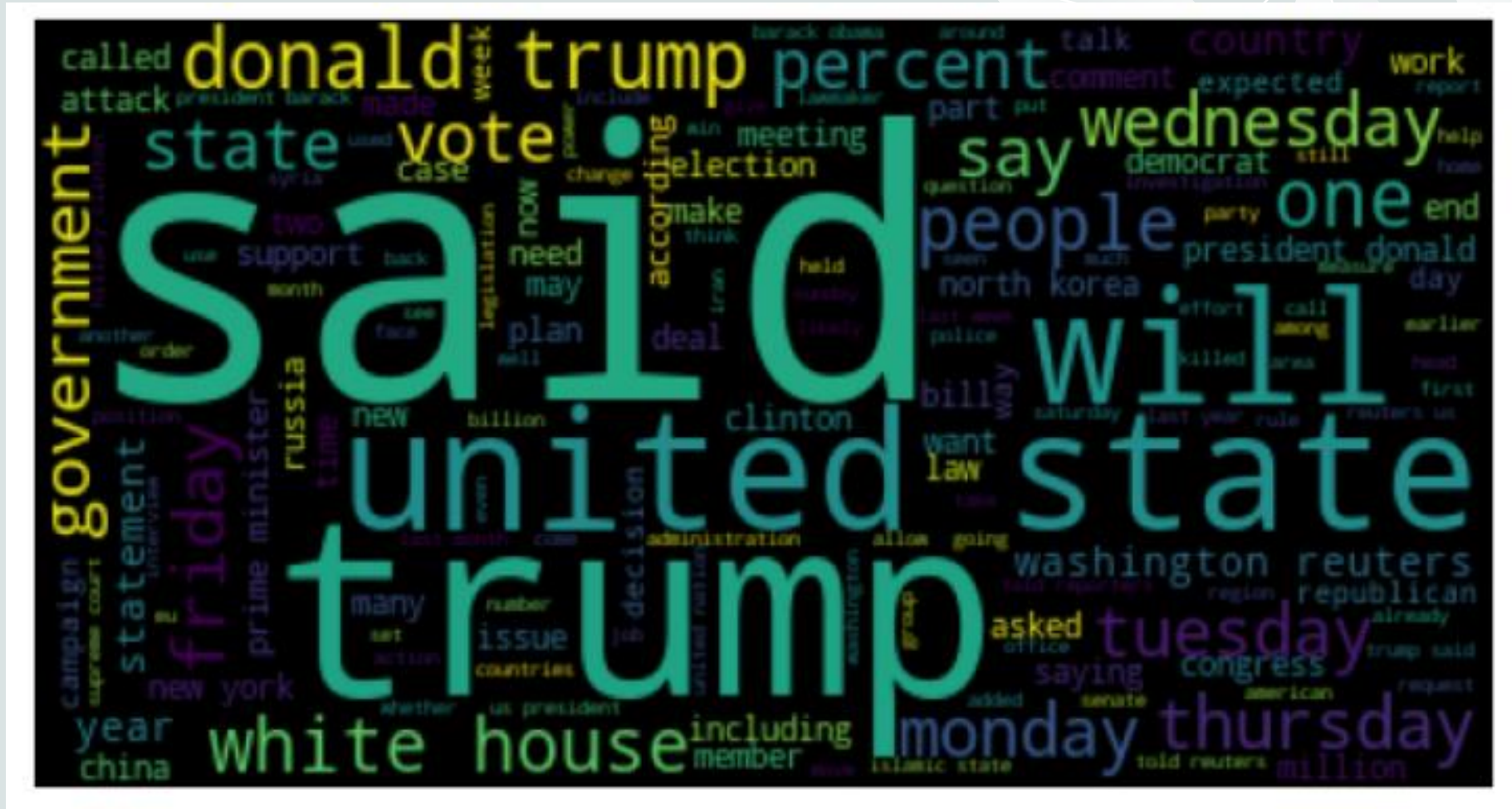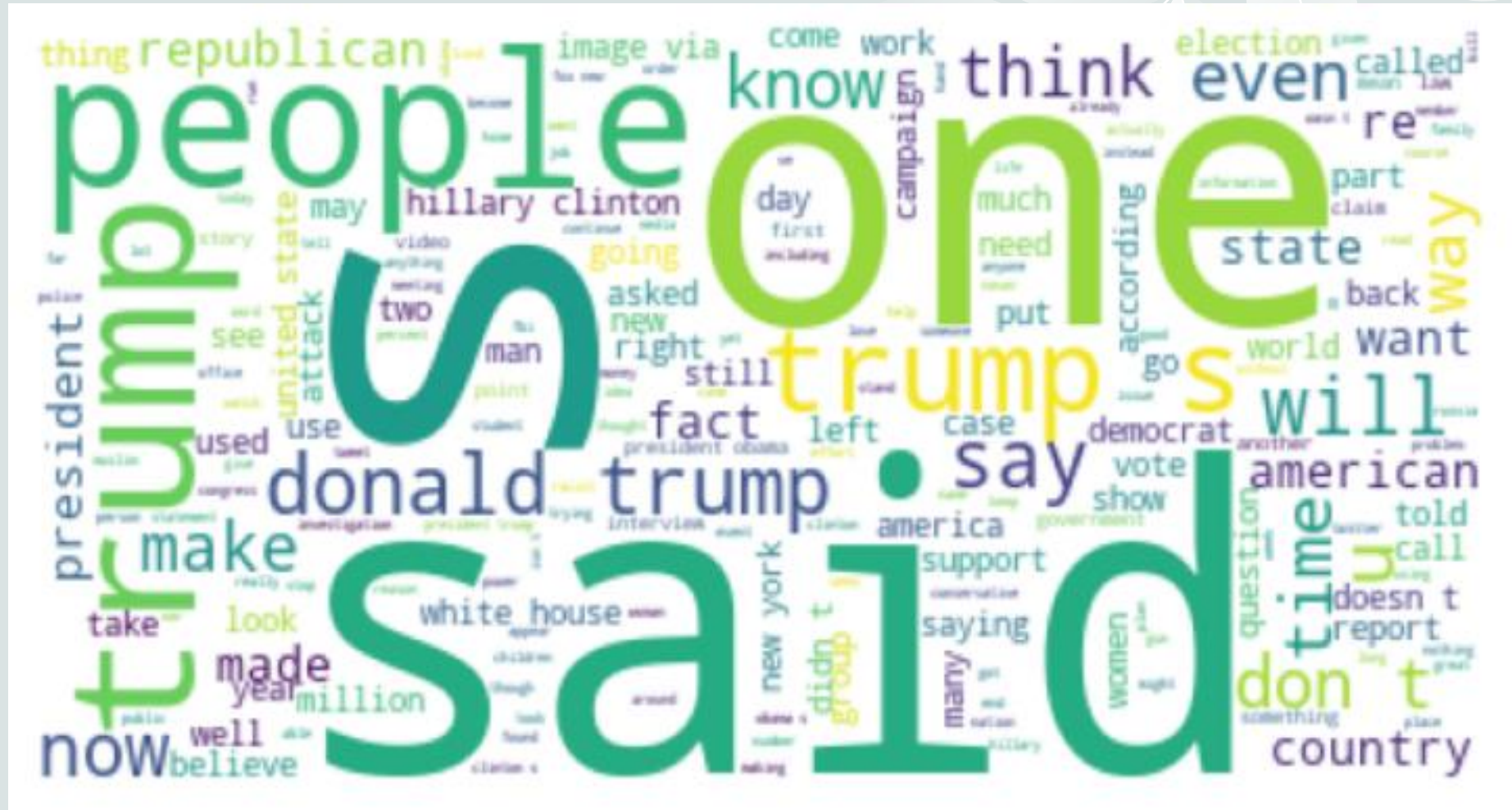QUANTITY OF FAKE AND REAL NEWS ARTICLES

DISTRIBUTION OF ARTICLES ON VARIOUS SUBJECTS

Here we can see a clear split between political news , world news and the other subjects. Fake news is most frequently found in the other subjects.

VISUALISATION OF REAL(TRUE) NEWS ARTICLES USING WORDCLOUD

VISUALISATION OF FAKE(FALSE) NEWS ARTICLES USING WORDCLOUD

Conclusion and Future Directions

- **Continued Research**

- Exploring advanced NLP techniques and deep learning models to improve fake news detection.

- **Cross-Disciplinary Collaboration**

- Combining expertise from fields like journalism, social science, and technology to tackle the fake news problem.

- **Public Education**

- Empowering users to critically evaluate online information and identify misinformation.

- **Policy Frameworks**

- Developing regulatory guidelines and best practices for the responsible use of news verification technology.

THANK YOU