



Fake/Real News Classification.

Efforts by –
Tanmay Pathak



contents.

INTRODUCTION

NLP AND TECHNIQUES USED

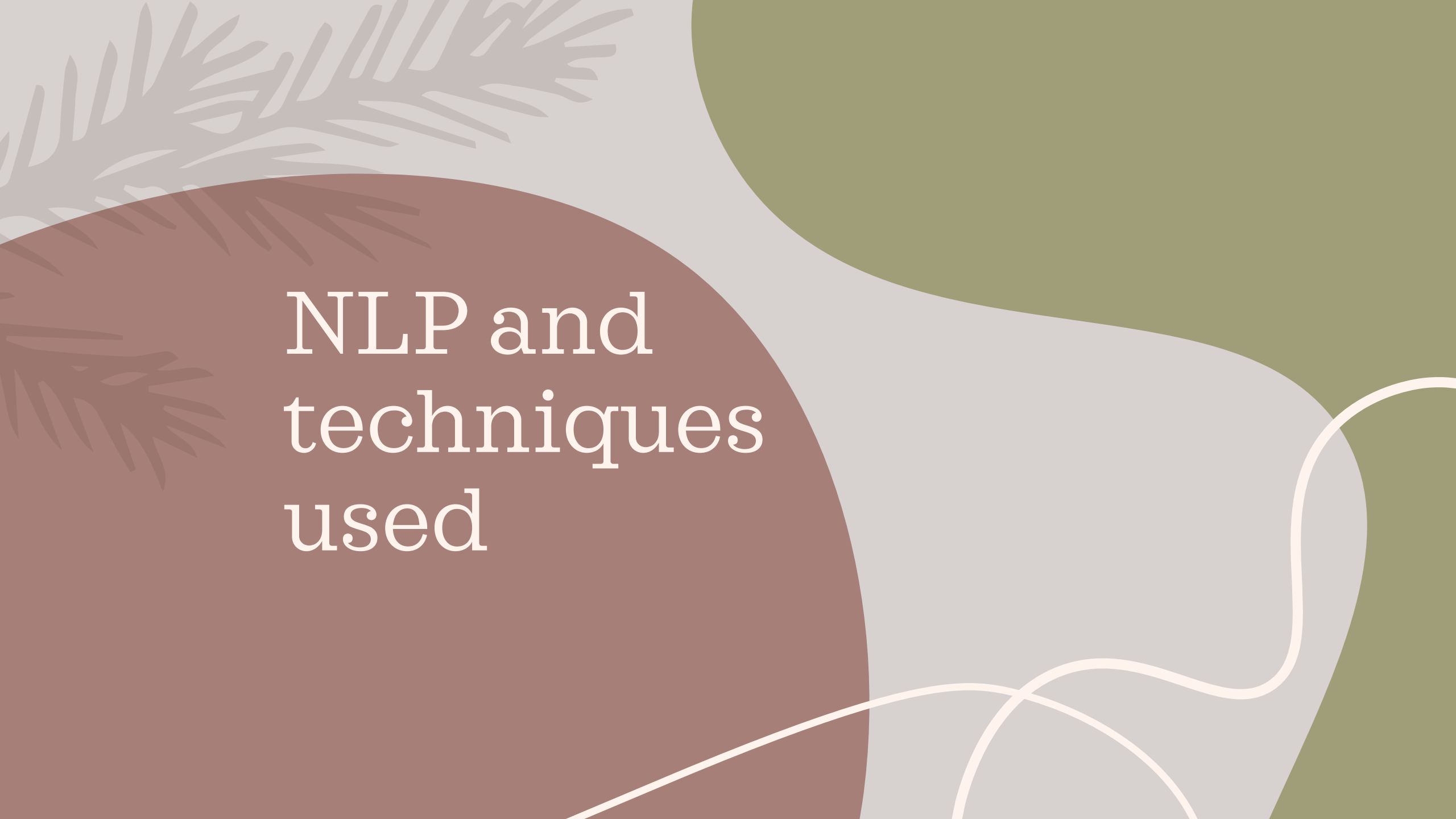
CHALLENGES IN FAKE NEWS
DETECTION

CODE

CONCLUSION

introduction

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.

The background features a light gray base with large, overlapping organic shapes in muted olive green and dusty rose. In the top left, there are stylized, layered illustrations of foliage or branches in shades of gray and rose. A thin, white, wavy line curves across the bottom right portion of the image.

NLP and techniques used

NLP

NLP is a branch of artificial intelligence that focuses on the interaction between computers and humans using natural language. It involves the development of algorithms and computational models to analyze, understand, and generate human language. NLP can be used to detect fake news by analyzing various aspects of the language used in the news articles. Some of the techniques used for fake news detection include:

Text Classification: This involves training an algorithm to identify the characteristics of fake news articles, such as sensational headlines, biased language, and poor grammar. The algorithm can then classify new articles as either fake or real based on these characteristics.

Sentiment Analysis: This involves analyzing the emotions and opinions expressed in a news article to determine whether it is real or fake. For example, fake news articles may use emotionally charged language to manipulate readers' opinions.

Natural Language Generation: This involves generating news articles that are similar to real news articles but contain fake information. NLP algorithms can be used to detect these fake articles by analyzing the language and comparing it to a database of known real news articles.

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.”

challenges in fake news detection

Data Collection

The quality and quantity of data available for fake news detection are crucial. The lack of large datasets with labeled fake news examples can make it difficult to train machine learning algorithms to detect them accurately.

Diversity in Fake News

Fake news can take many different forms, including clickbait, satire, propaganda, and conspiracy theories. Each type requires a different approach to detection, and it can be challenging to develop a single algorithm that can detect them all.

Language Ambiguity

Natural language is often ambiguous, making it difficult to determine the intent and meaning behind a piece of text. Fake news creators often use language that is deliberately vague or misleading, making it harder to identify and distinguish from real news.

Speed of Dissemination

Fake news can spread rapidly through social media and other online platforms. Detecting and removing fake news quickly is critical to preventing its widespread dissemination and potential harm.

Adversarial Attacks

Fake news creators may attempt to circumvent detection algorithms by introducing subtle changes to the language or format of their articles. These adversarial attacks can make it challenging for algorithms to identify fake news accurately.

code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

pd.set_option("display.max_columns",50)
pd.set_option("display.max_rows",50)

import nltk
from nltk.stem.porter import PorterStemmer
from wordcloud import WordCloud
from wordcloud import STOPWORDS

from PIL import Image

from sklearn.model_selection import train_test_split

from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
import warnings
warnings.filterwarnings('ignore')

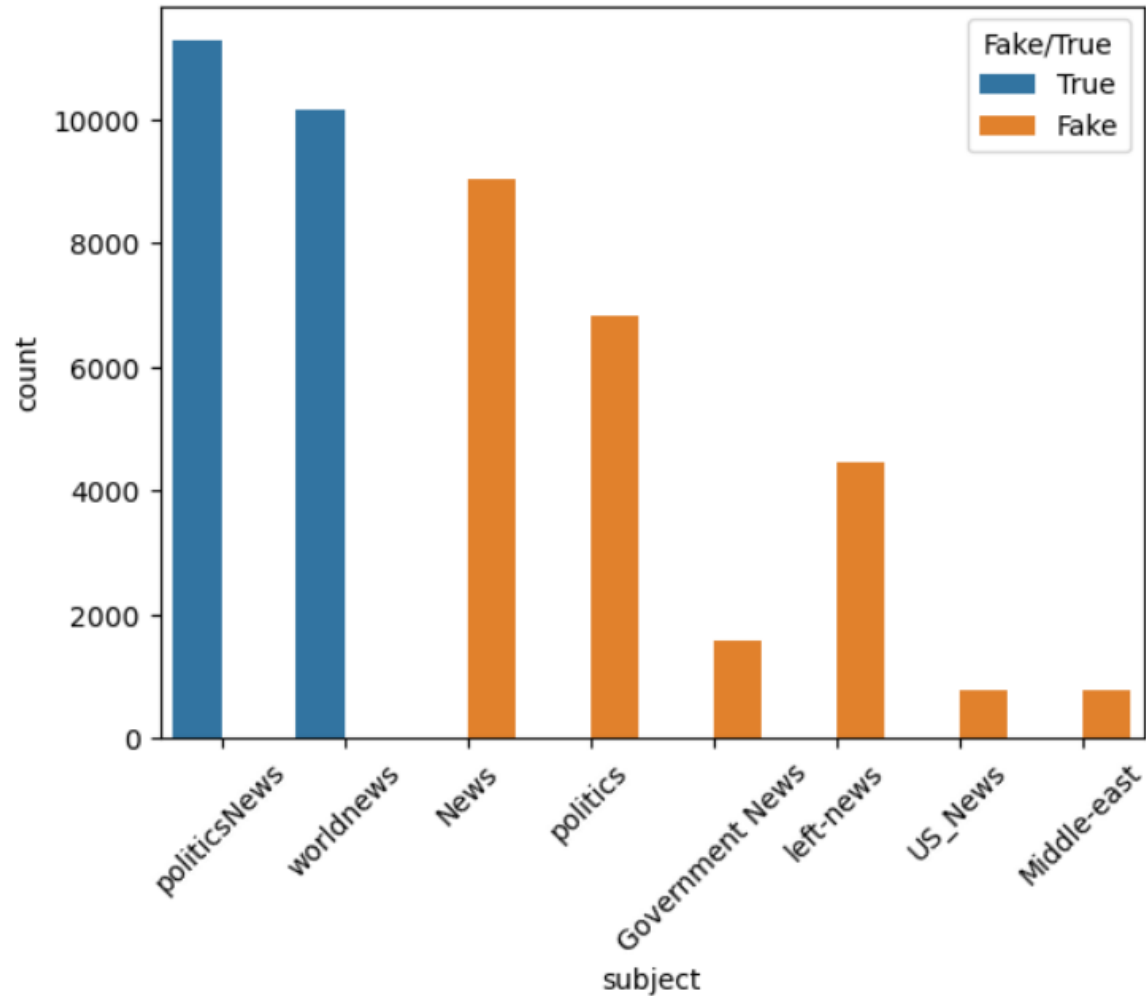
from nltk.corpus import stopwords

stopwords = stopwords.words('english') + ['of', 'the', 'to', 'be', 'that', 'with', 'on', 'for', 'by', 'at', 'one']

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

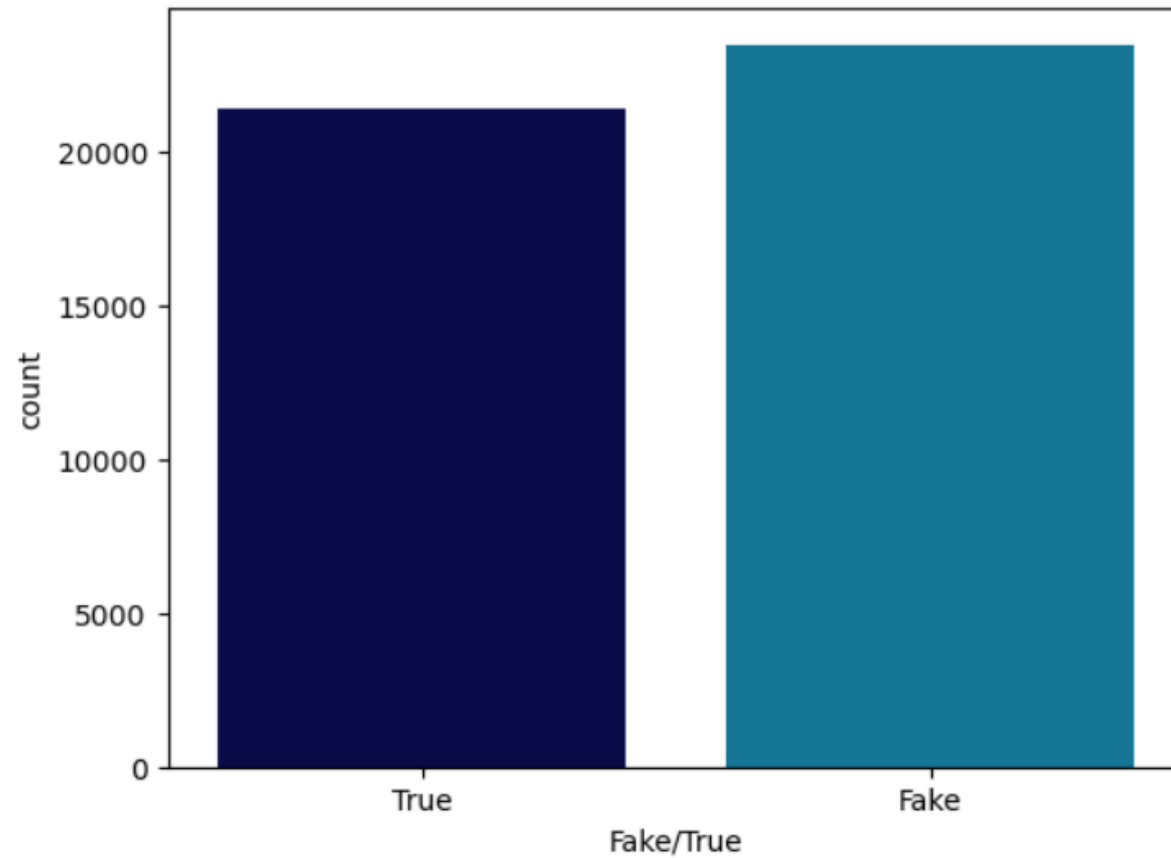
DATA VISUALIZATION

```
In [22]: sns.countplot(x='subject',hue='Fake/True', data=df)  
plt.xticks(rotation=45);
```



TRUE TO FALSE RATIO VISUALIZATION

```
In [24]: sns.countplot(x='Fake/True',data=df,palette='ocean');
```



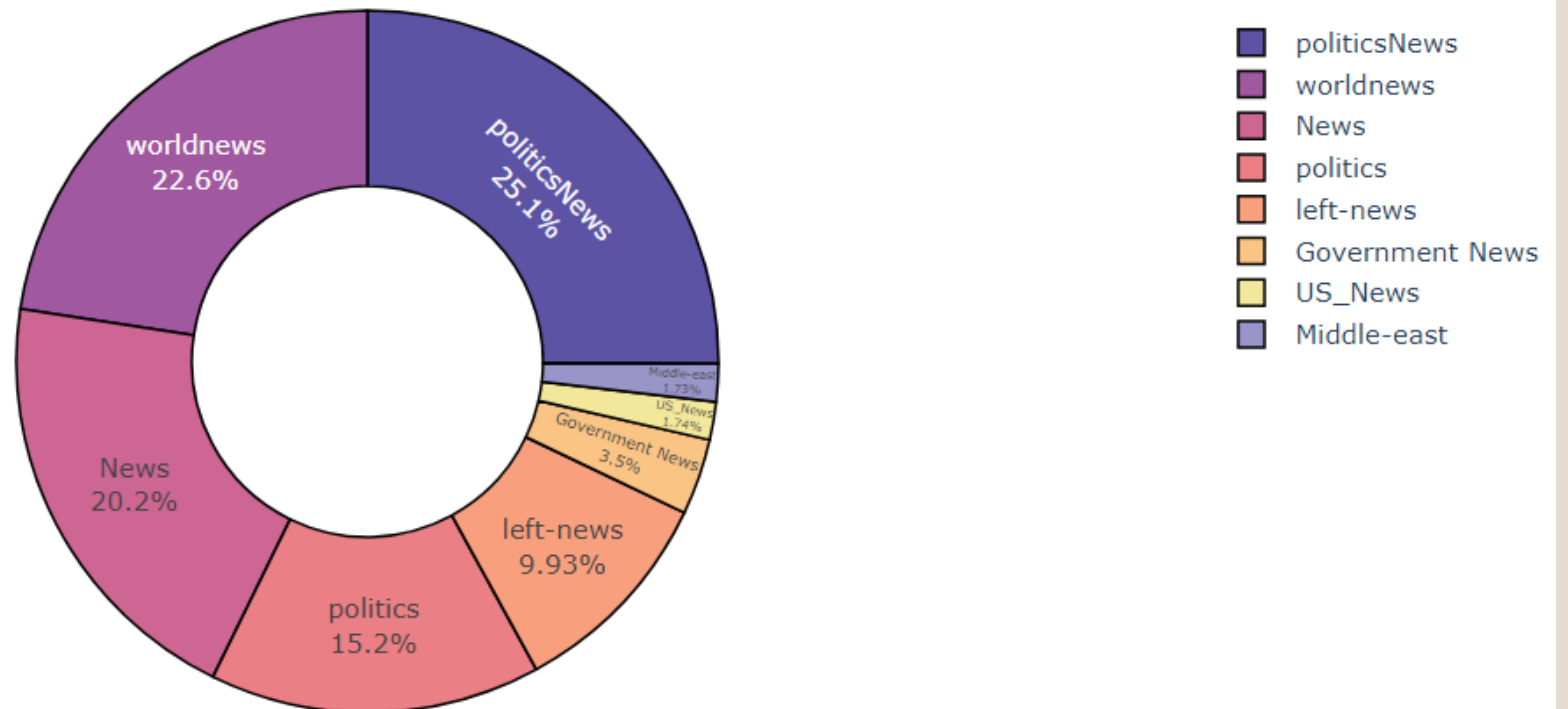
PIE CHART OF DFFERENT KINDS OF NEWS

```
In [25]: import plotly.express as px

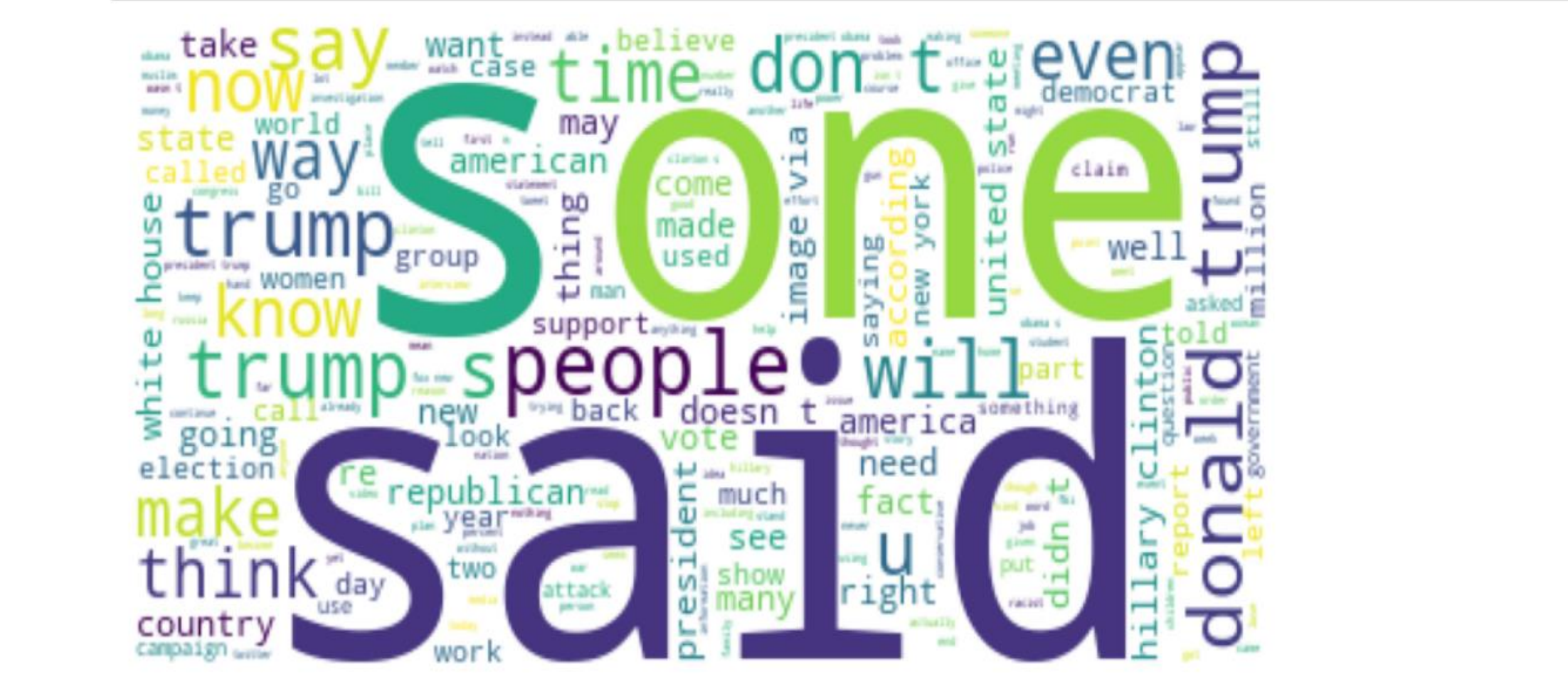
fig = px.pie(df, names = "subject", title = "News Subject", hole = 0.5,
            width = 1000, height = 500, color_discrete_sequence = px.colors.sequential.Sunset_r)

fig.update_traces(textposition = "inside", textinfo = "percent+label",
                  marker = dict(line = dict(width = 1.2, color = "#000000")))
```

News Subject

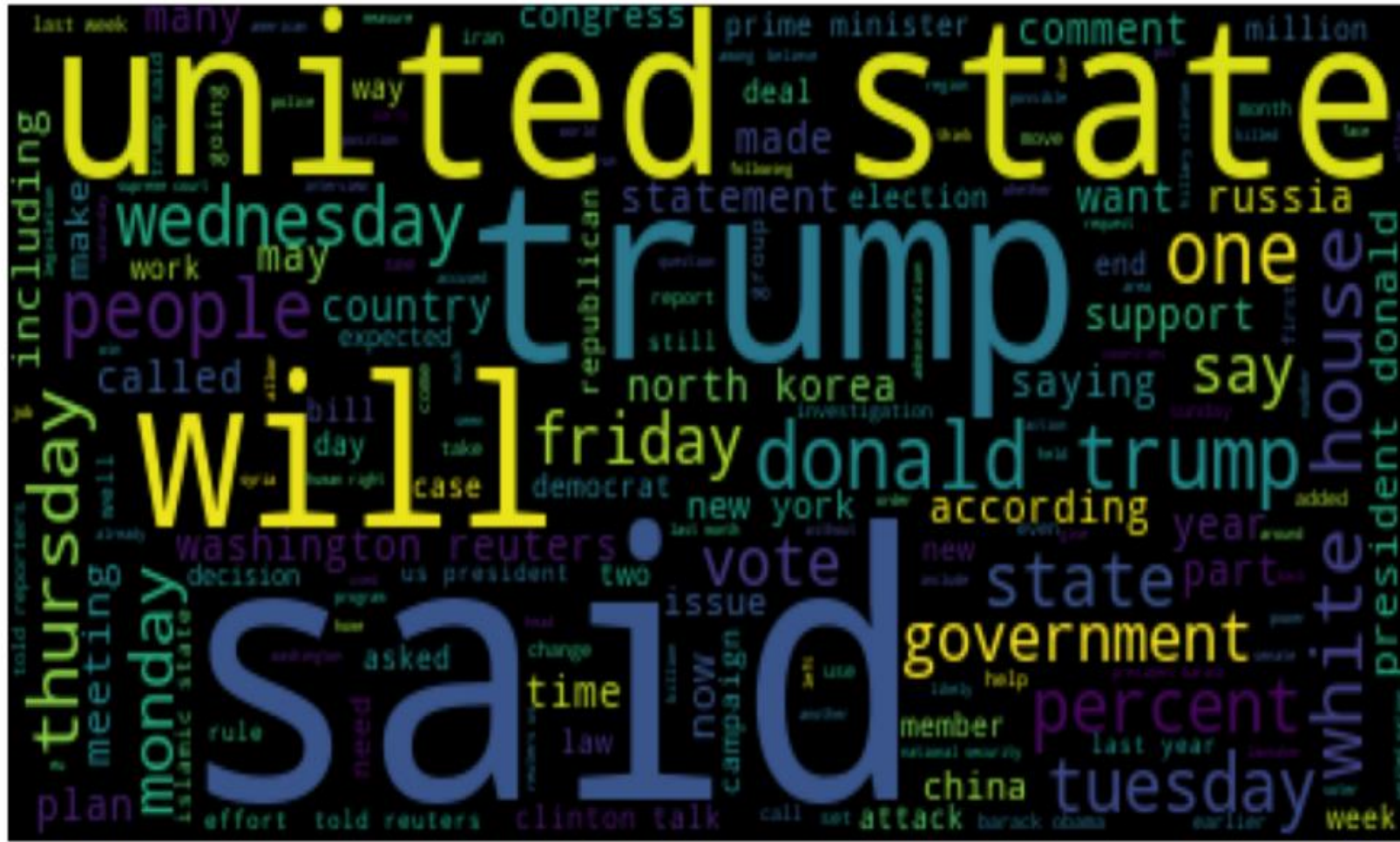


```
In [30]: wc(Fake, 'white')
```



TRUE NEWS (USING WORDCLOUD)

```
In [31]: wc(true, 'black')
```



MIDDLE EAST NEWS (USING WORDCLOUD)

```
In [32]: wc(middle_east,'white')
```



Prediction -1

```
In [30]: prediction=mb.predict(x_test)
```

```
In [31]: accuracy_score(y_test,prediction)
```

```
Out[31]: 0.9600593912397921
```

Prediction-2

```
In [34]: prediction2=d.fit(x_train,y_train).predict(x_test)
```

```
In [35]: accuracy_score(y_test,prediction2)
```

```
Out[35]: 0.9962880475129918
```


conclusion

NLP can be a powerful tool for detecting and combatting fake news. By analyzing various aspects of language used in news articles, NLP algorithms can identify patterns and characteristics that distinguish real news from fake news. As fake news continues to be a problem, NLP can play a critical role in ensuring that readers have access to accurate and trustworthy information.