Fake News Detection Report

Date: June 22, 2025

Prepared by: Viswa Teja & Kasi Viswanadh Maddala

Business Objective

The spread of fake news has become a significant challenge in today's digital world. With the massive volume of news articles published daily, it's becoming harder to distinguish between credible and misleading information. This creates a need for systems that can automatically classify news articles as true or fake, helping to reduce misinformation and protect public trust.

Problem Statement

In the digital era, the rapid spread of information through online platforms has significantly increased the circulation of fake news. This phenomenon poses serious threats to public opinion, political stability, and societal trust. Given the overwhelming volume of content generated daily, manual verification of news authenticity is impractical and inefficient.

To address this challenge, there is an increasing demand for **automated systems** capable of accurately classifying news articles as **true or fake** based on their **semantic content** rather than superficial textual features. Traditional text classification methods often depend heavily on **keyword matching** or **syntactic patterns**, which can fail to capture the **subtle contextual meanings** embedded in language.

The goal of this project is to develop a Semantic Classification model that leverages Word2Vec embeddings to understand and represent the semantic relationships in textual data. This vector-based representation allows supervised learning models to more effectively differentiate between credible and deceptive news articles by focusing on context and meaning. By building and evaluating models such as Logistic Regression, Decision Tree, and Random Forest, this project aims to identify the most effective approach for accurate and reliable fake news detection.

🧪 Methodology

The project followed a structured pipeline to classify news articles as **fake** or **true** using semantic analysis and supervised learning techniques:

1. Data Preparation

- Merged two datasets containing true and fake news articles.
- Assigned binary labels: **1** for true news, **0** for fake news.
- Removed null values and duplicate entries to ensure clean input data.

2. Text Preprocessing

- Converted all text to lowercase for consistency.
- Removed punctuation, special characters, and stopwords to reduce noise.
- Applied **lemmatization** to normalize words to their base forms, enhancing semantic consistency.

3. Train-Validation Split

- Split the dataset into **70% training** and **30% validation** using **stratified sampling** to maintain class balance.

4. Exploratory Data Analysis (EDA)

- Analyzed word distributions, article lengths, and class balance.
- Generated **word clouds** and **frequency plots** to visualize common terms in true vs. fake news.

5. Feature Extraction

- Employed **Word2Vec embeddings** to capture the semantic meaning of words.
- Represented each article by averaging the word vectors to create fixed-length feature vectors.

6. Model Training and Evaluation

- Trained three supervised classifiers:
  - **Logistic Regression**
  - **Decision Tree**
  - **Random Forest**
- Evaluated models using:
  - **Accuracy**
  - **Precision**
  - **Recall**
  - **F1-Score**

🛠️ Techniques Used

1. Natural Language Processing (NLP)

- Applied for text cleaning, tokenization, stopwords removal, and lemmatization.

2. Word2Vec Embeddings

- Used to encode semantic relationships between words in vector space, enabling context-aware feature representation.

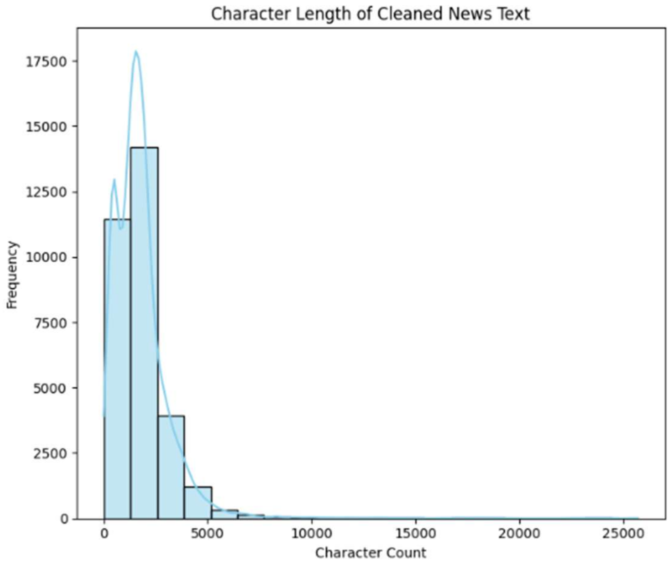
3. Supervised Learning Models

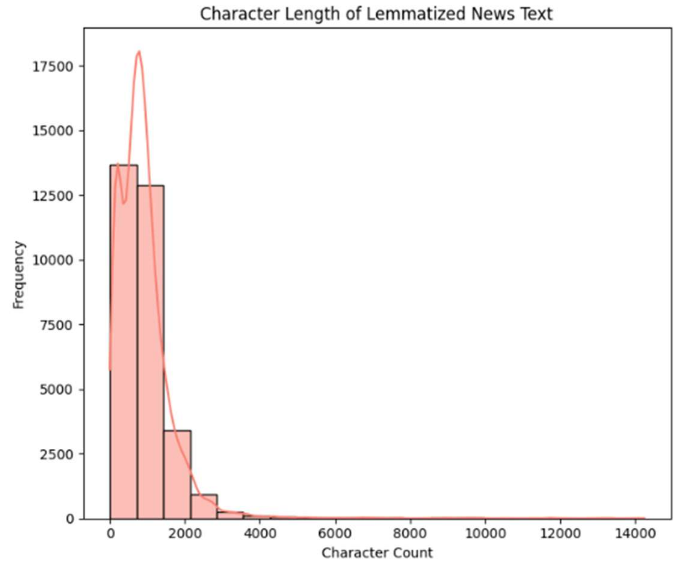
- **Logistic Regression**: A linear baseline classifier.
- **Decision Tree**: A non-linear model capable of capturing complex patterns.
- **Random Forest**: An ensemble method that enhances accuracy through multiple decision trees.

4. Evaluation Metrics

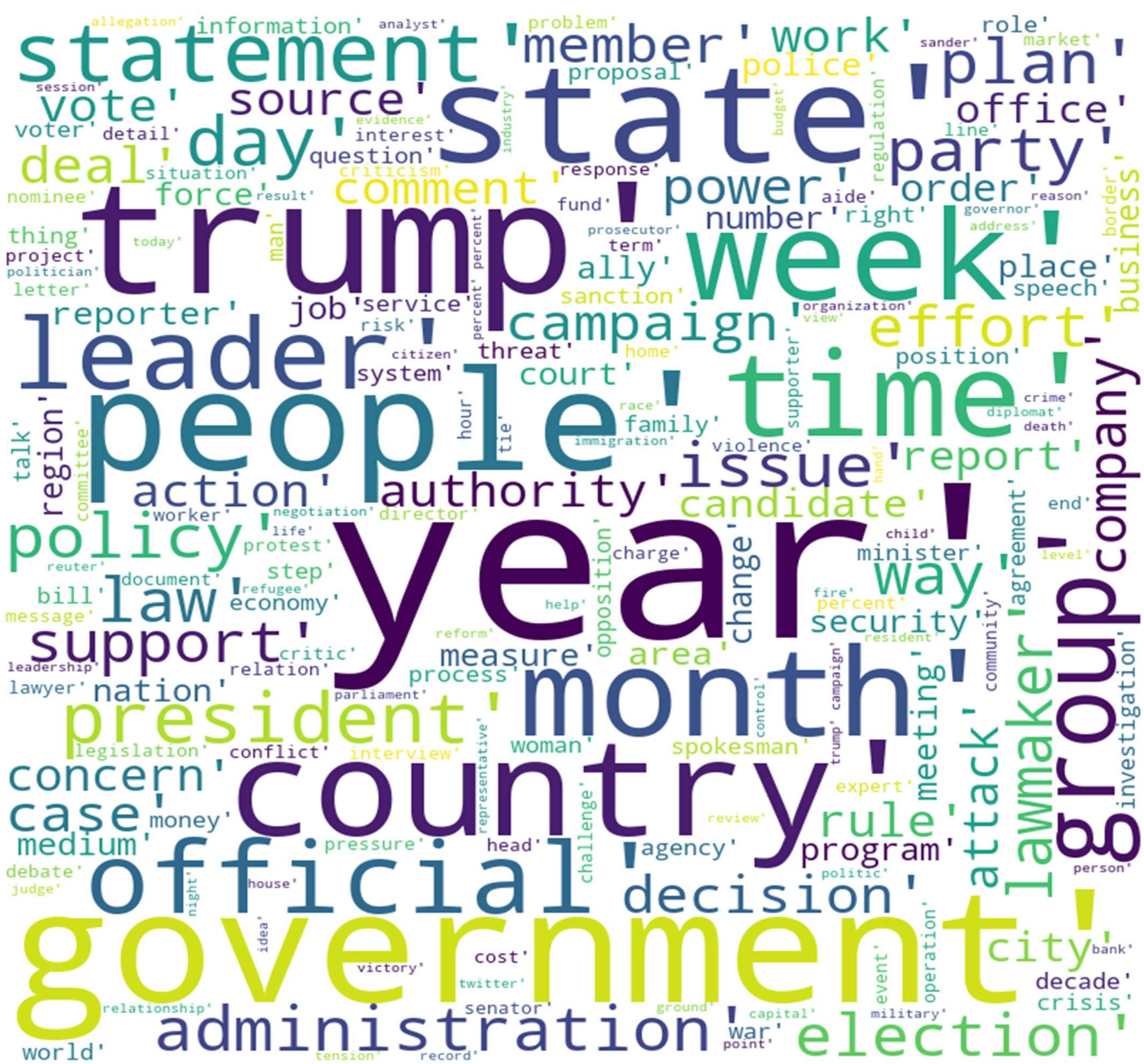
- Used **Accuracy**, **Precision**, **Recall**, and **F1-Score** to assess and compare model performance.

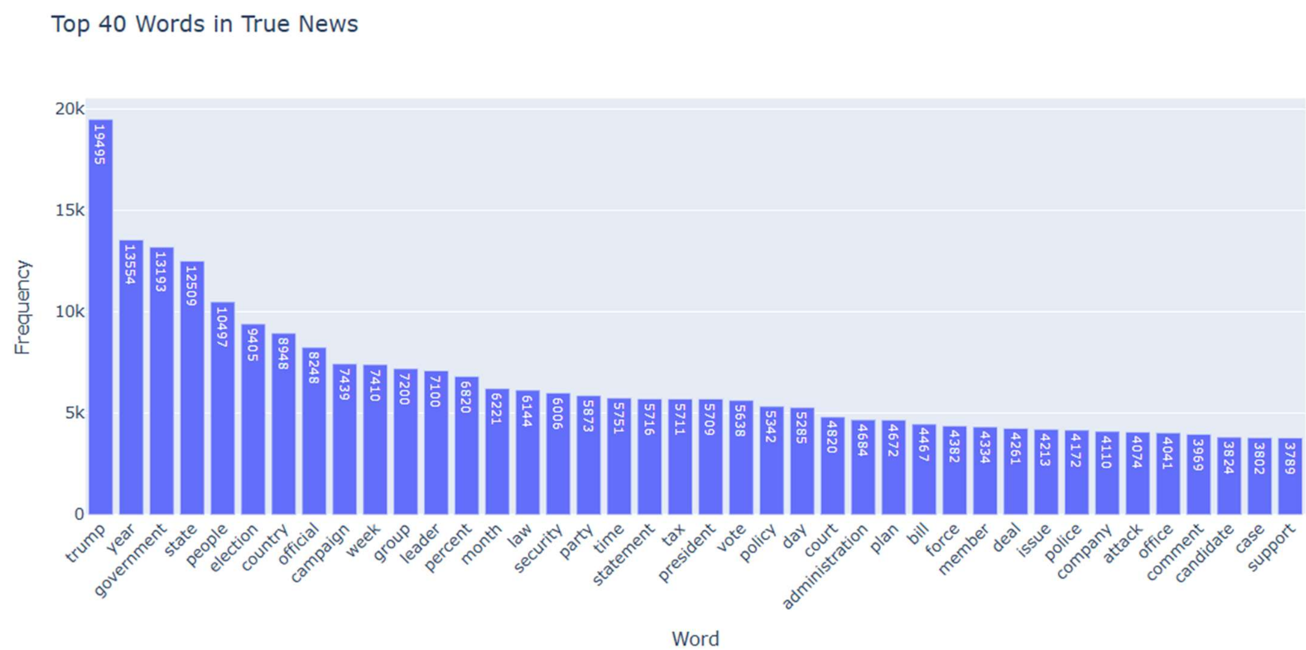
Graphs/plots



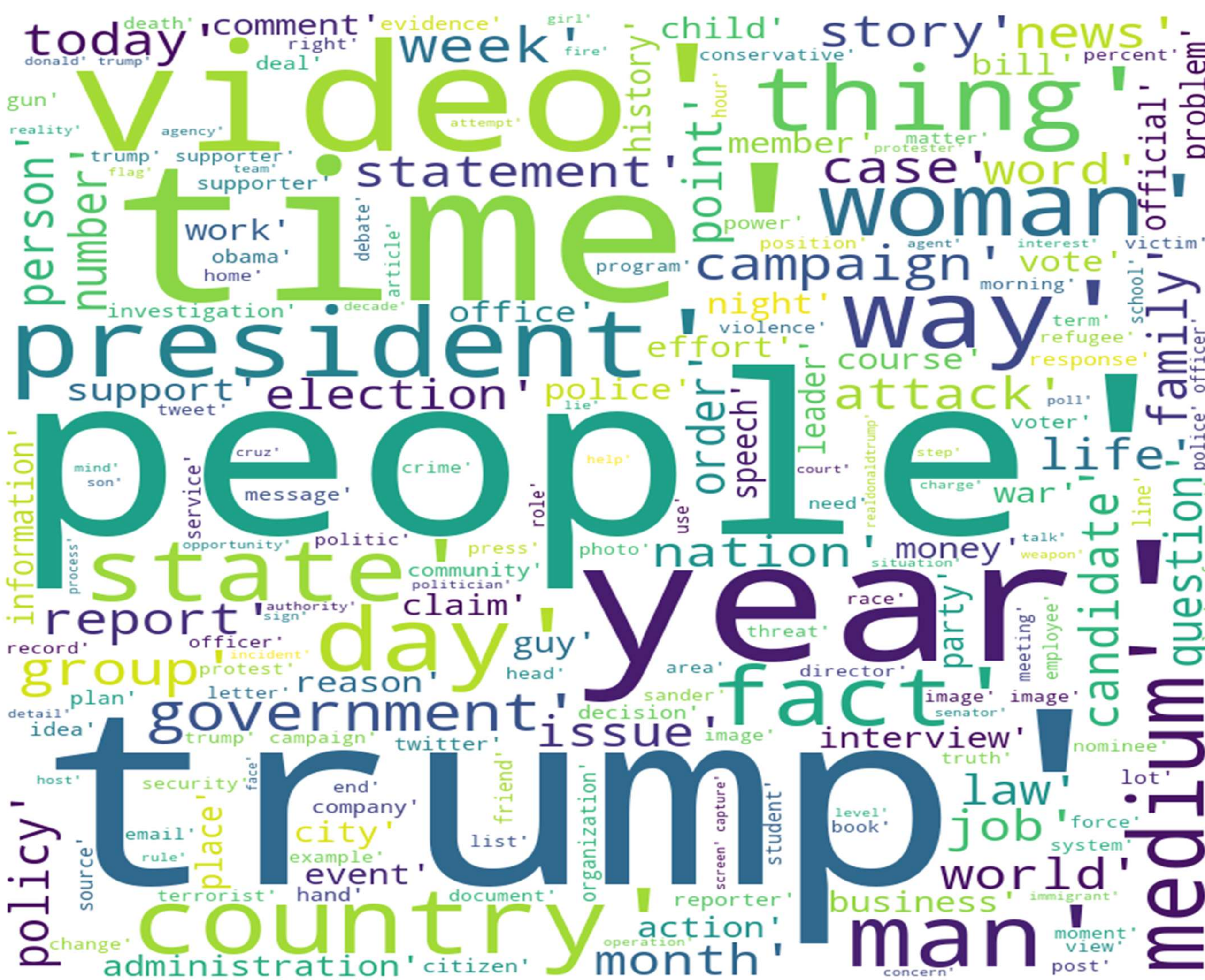


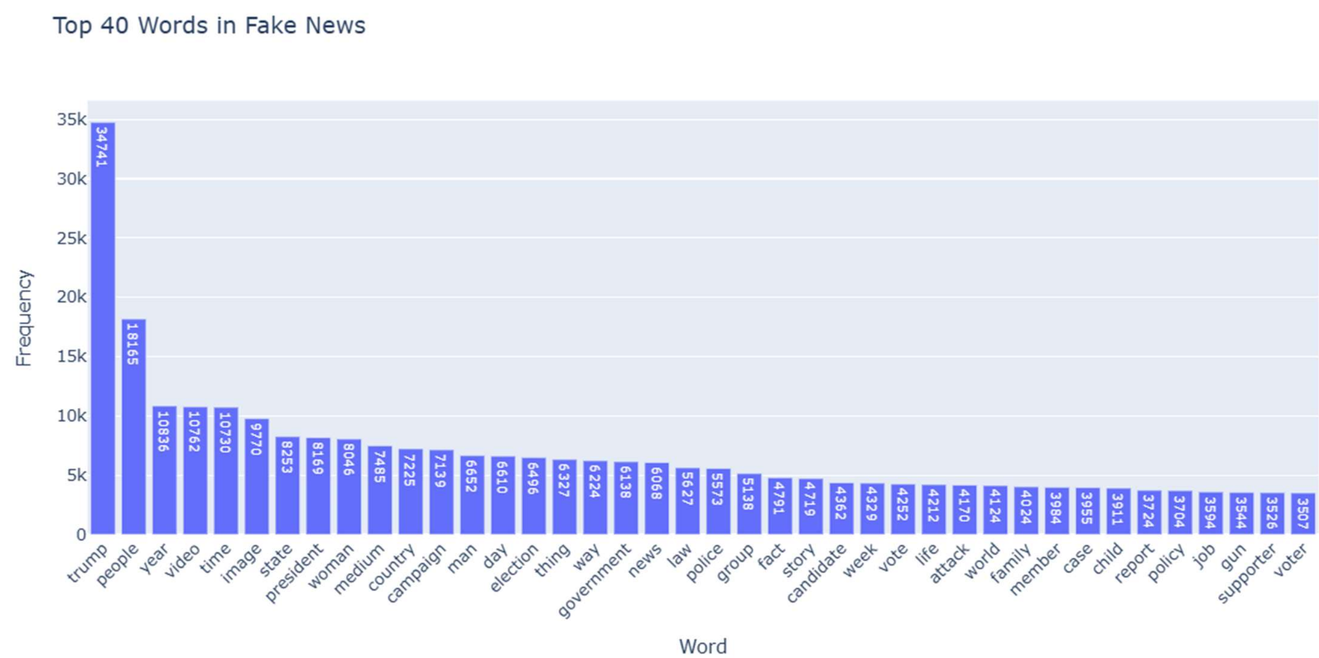
Word cloud with top 40 words by frequency among true news in the training data

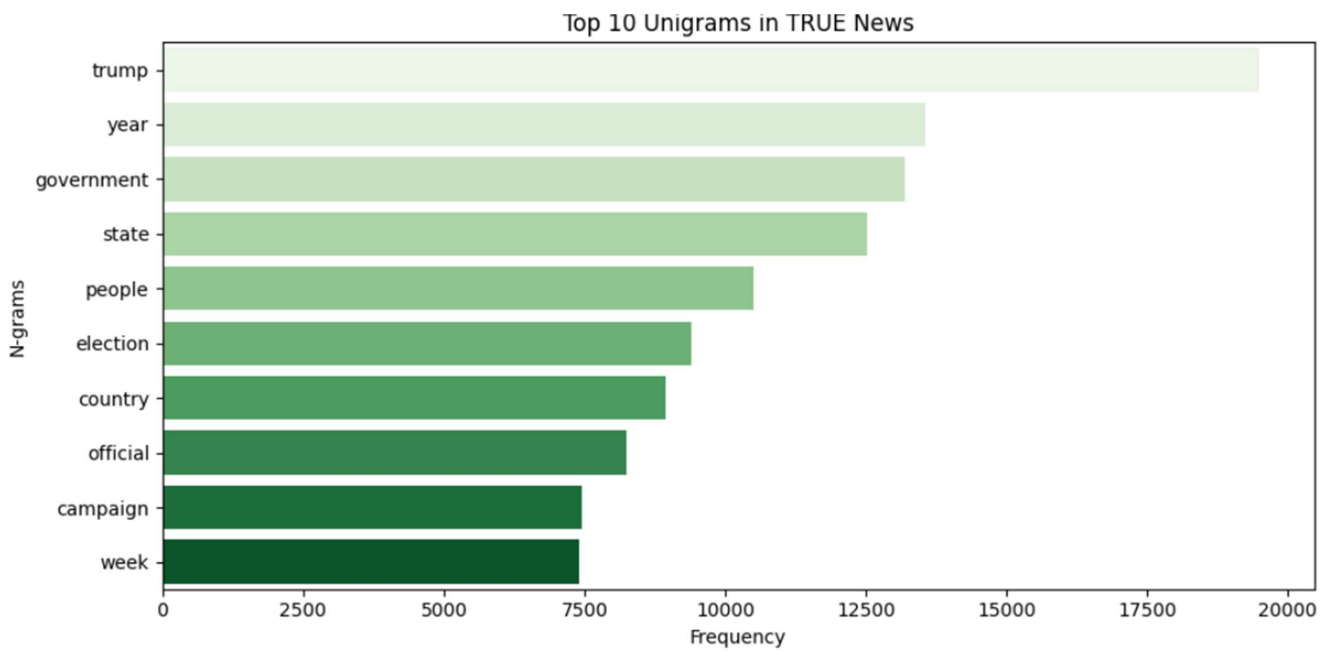


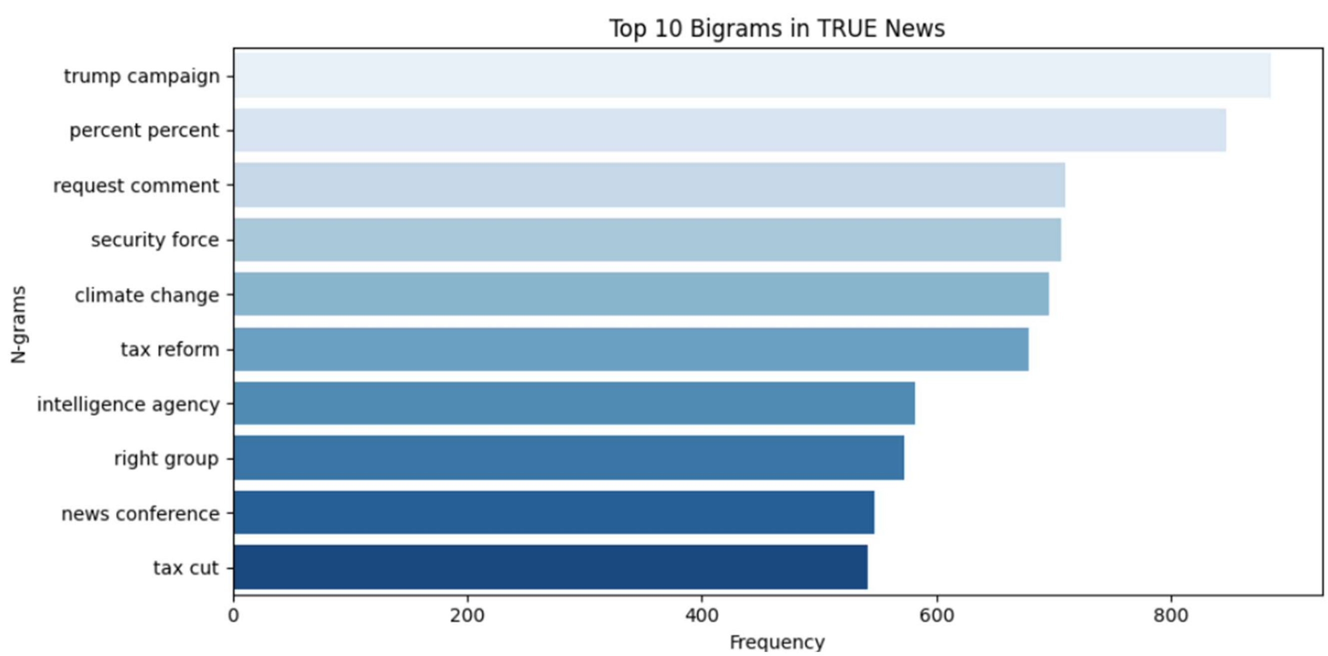


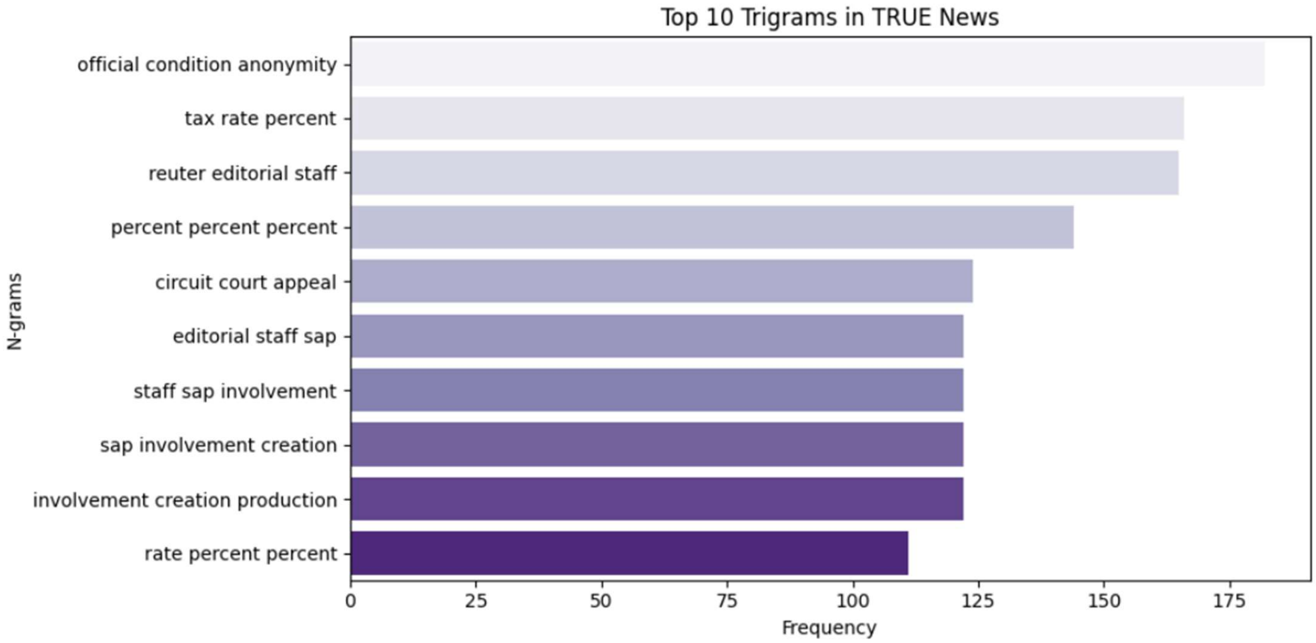
Word cloud with top 40 words by frequency among fake news in the training data

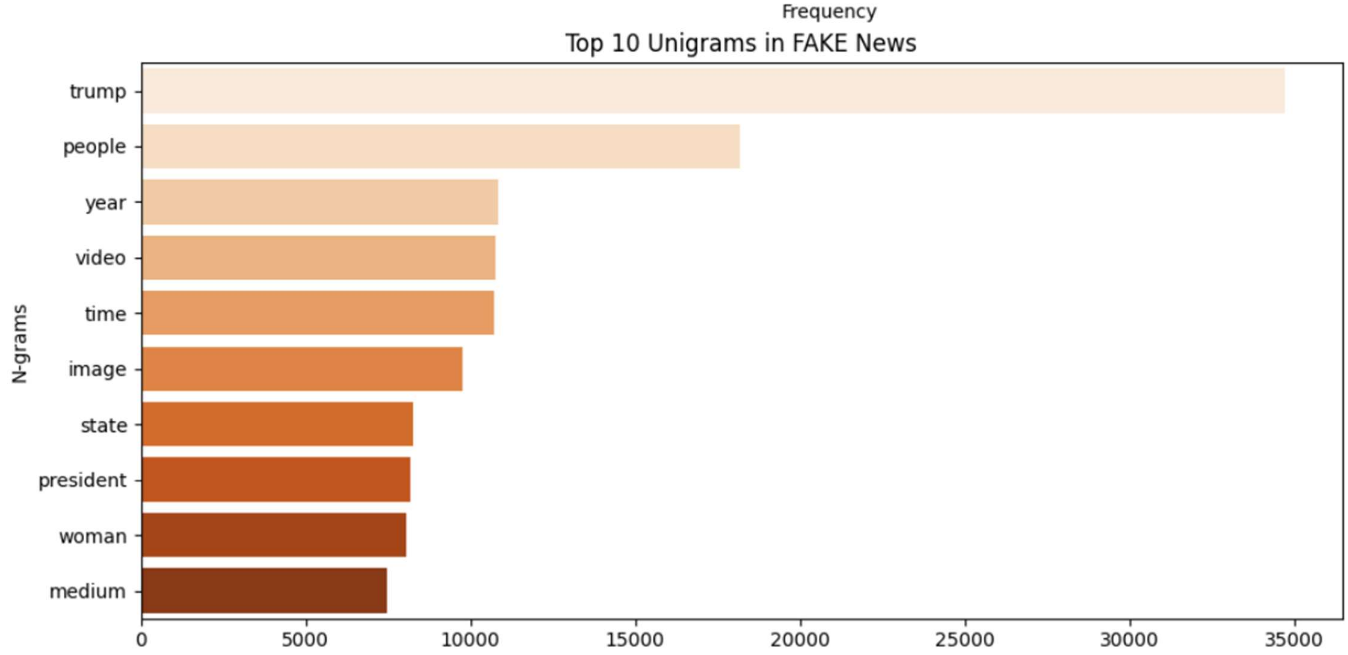


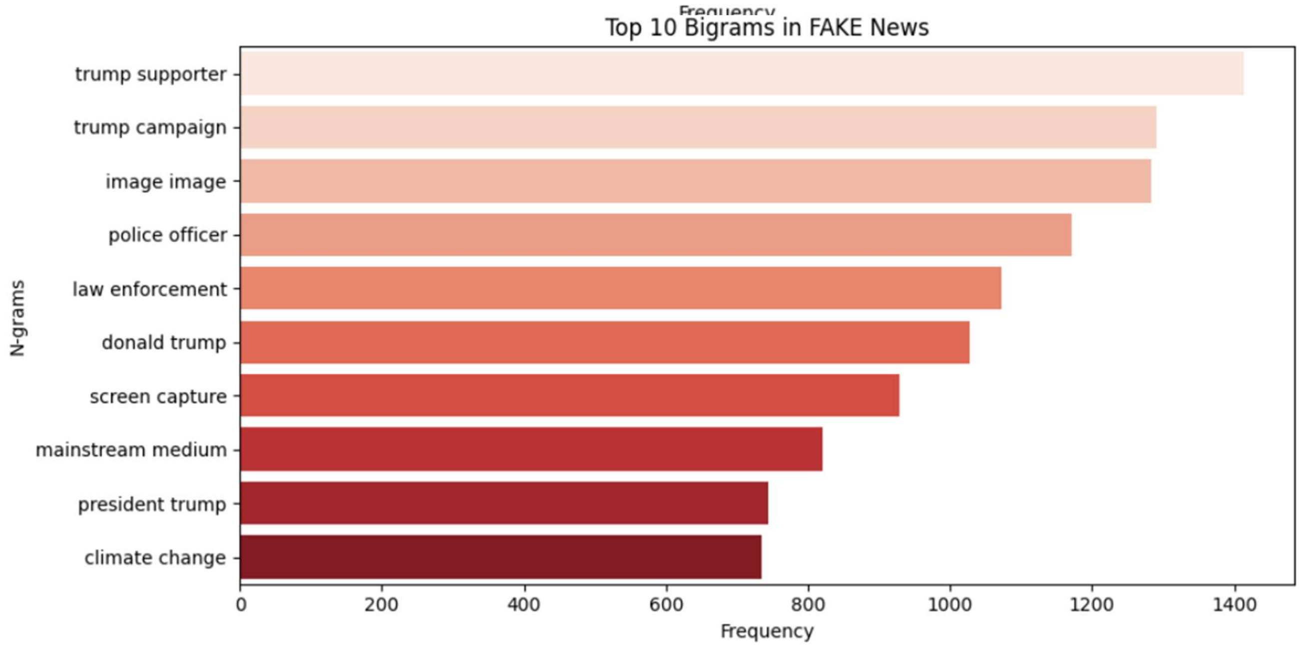

Top 40 Words in Fake News
Frequency
Word
0
5k
10k
15k
20k
25k
30k
35k
trump 34741
people 18165
year 10836
video 10762
time 10730
image 9770
state 8253
president 8169
woman 8046
medium 7485
country 7225
campaign 7139
man 6652
day 6610
election 6496
thing 6327
way 6224
government 6138
news 6068
law 5627
police 5573
group 5138
fact 4791
story 4719
candidate 4362
week 4329
vote 4252
life 4212
attack 4170
world 4124
family 4024
member 3984
case 3955
child 3911
report 3724
policy 3704
job 3594
gun 3544
supporter 3526
voter 3507

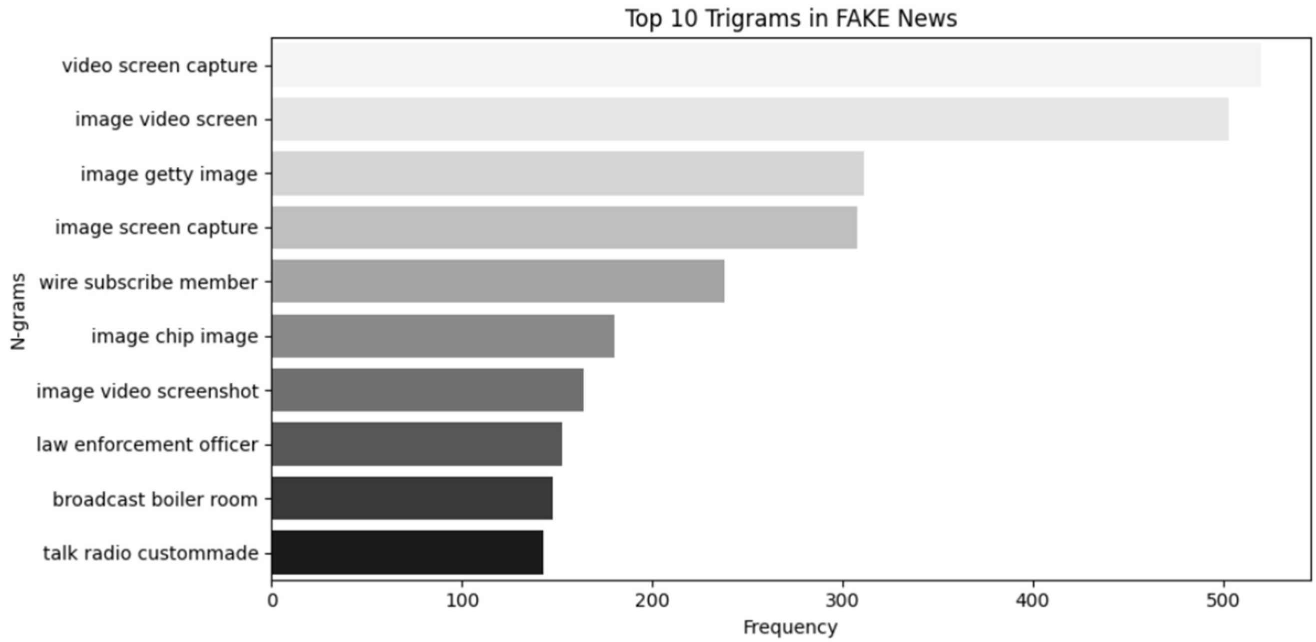

Top 10 Unigrams in TRUE News
N-grams
Frequency
trump
year
government
state
people
election
country
official
campaign
week
0
2500
5000
7500
10000
12500
15000
17500
20000


Top 10 Bigrams in TRUE News
N-grams
Frequency
trump campaign
percent percent
request comment
security force
climate change
tax reform
intelligence agency
right group
news conference
tax cut
0
200
400
600
800

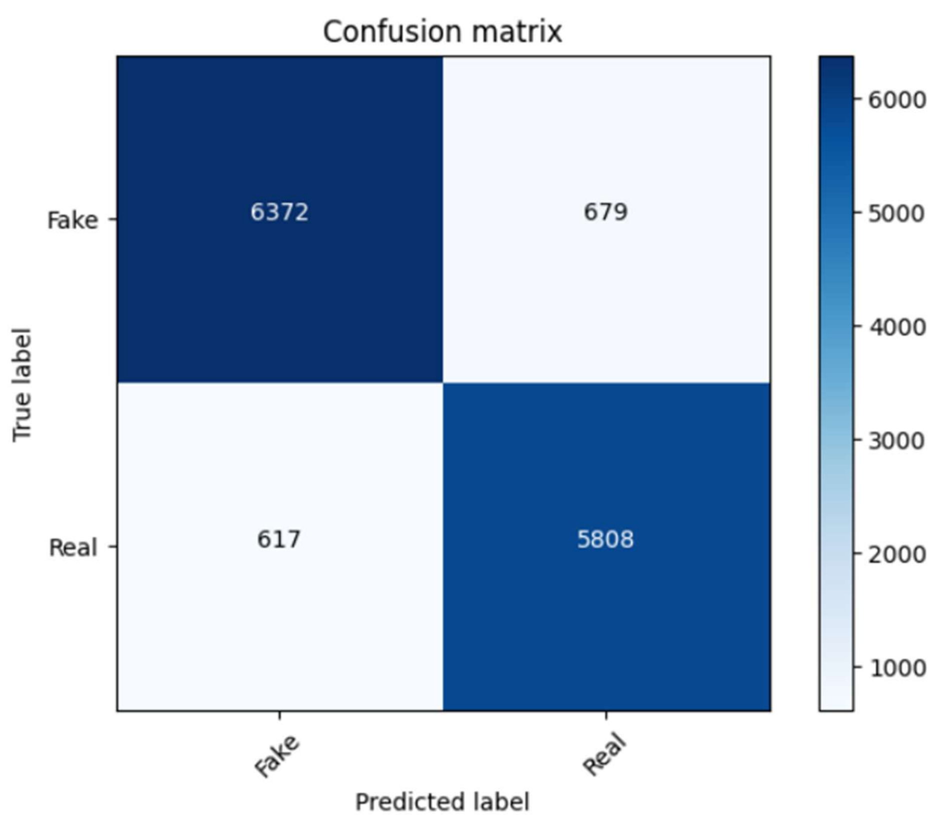

Top 10 Trigrams in TRUE News
N-grams
official condition anonymity
tax rate percent
reuter editorial staff
percent percent percent
circuit court appeal
editorial staff sap
staff sap involvement
sap involvement creation
involvement creation production
rate percent percent
0
25
50
75
100
125
150
175
Frequency


Top 10 Unigrams in FAKE News
N-grams
trump
people
year
video
time
image
state
president
woman
medium
0
5000
10000
15000
20000
25000
30000
35000

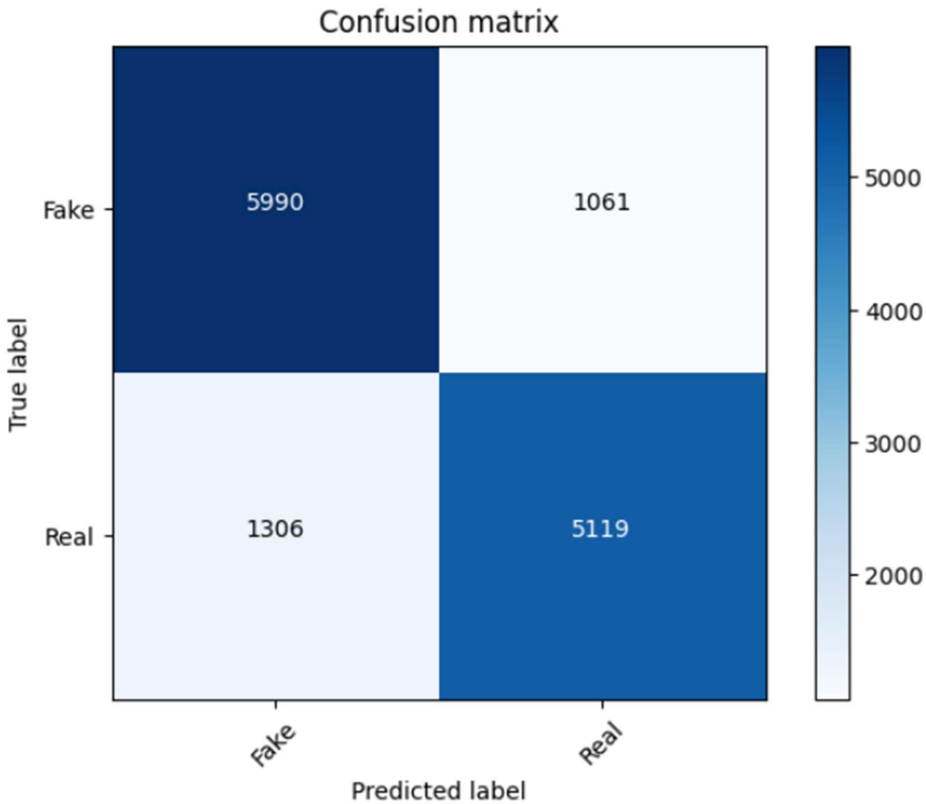

Top 10 Bigrams in FAKE News
N-grams
trump supporter
trump campaign
image image
police officer
law enforcement
donald trump
screen capture
mainstream medium
president trump
climate change
0
200
400
600
800
1000
1200
1400
Frequency



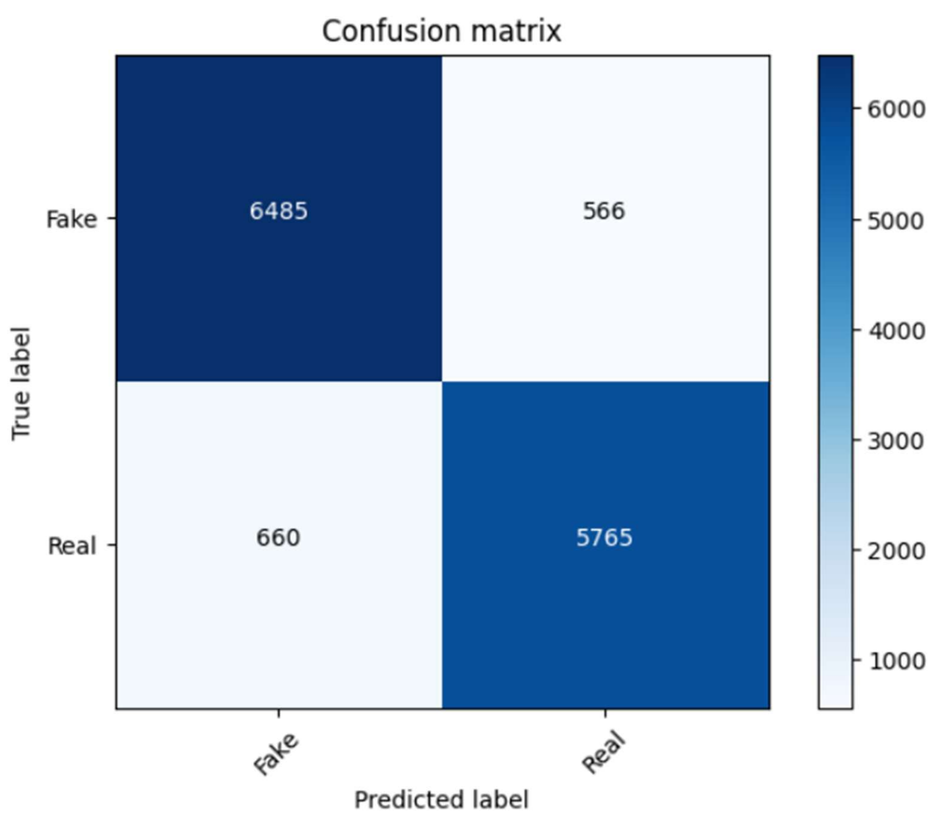
Confusion Matrix for Logistic Regression Model :



Confusion Matrix for Decision Tree Model :



Confusion Matrix for Random Forest Model :



Key Insights

The project successfully demonstrated the effectiveness of semantic classification using Word2Vec embeddings combined with supervised learning techniques for fake news detection. The core insight lies in the ability of Word2Vec to capture the underlying **semantic structure** of text, allowing the models to distinguish between true and fake news based on **contextual meaning**, rather than surface-level keywords.

- ✧ **Model Performance Insights**:
  - **Logistic Regression** exhibited strong performance, achieving **accuracy (0.90) and F1-score (0.90)**. Model has balanced precision and recall values indicated an effective ability to detect both fake and true news articles.
  - **Decision Tree** model has slightly low performance than Logistic Regression, also achieving an **accuracy of 0.82** and **F1-score of 0.81**. While decision trees offer intuitive interpretability and flexibility, they can be prone to overfitting.
  - **Random Forest** slightly outperformed the other models with an **accuracy of 0.91** and the **highest F1-score of 0.90**, along with the best **precision (0.91)**. This made it the most reliable classifier for the task, particularly effective at reducing false positives (i.e., misclassifying true news as fake).
- ✧ **Semantic Patterns Observed**:
  - **True news**: Exhibited consistent, fact-based language and stronger semantic coherence. These articles often followed structured journalistic norms, making them easier to identify using word embeddings.
  - **Fake news**: Tended to use exaggerated, emotionally charged, or contextually inconsistent language. These linguistic patterns made syntactic methods less effective, but were well captured by the semantic nature of Word2Vec.
- ✧ **Model Selection and Evaluation**:
  - The **Random Forest** model was chosen as the final model due to its superior balance across all performance metrics.
  - The **F1-score** was prioritized as the key evaluation metric, as it offers a harmonic balance between **precision** and **recall**—critical in applications where both false positives and false negatives carry significant consequences.

✧ Impact of Semantic Classification:

- **Enhanced Contextual Understanding**:
  The use of **Word2Vec embeddings** significantly improved the model's ability to grasp the **semantic meaning and contextual relationships** between words. This allowed the classifiers to go beyond surface-level keyword matching and make more informed predictions based on the overall linguistic structure of the text.

- **Beyond Keyword Detection**:
  Traditional fake news detection methods often rely on keyword frequency or syntactic patterns, which can be easily manipulated. In contrast, the semantic approach captured **deeper linguistic signals**, such as tone, coherence, and contextual relevance—making it more resilient and practical for real-world applications.

- **Robustness Through Ensemble Learning**:
  The **Random Forest model**, with its ensemble architecture, further amplified the benefits of semantic embeddings. By aggregating multiple decision trees, it produced **robust and generalizable predictions**, reducing the risk of overfitting and improving reliability across diverse news content.