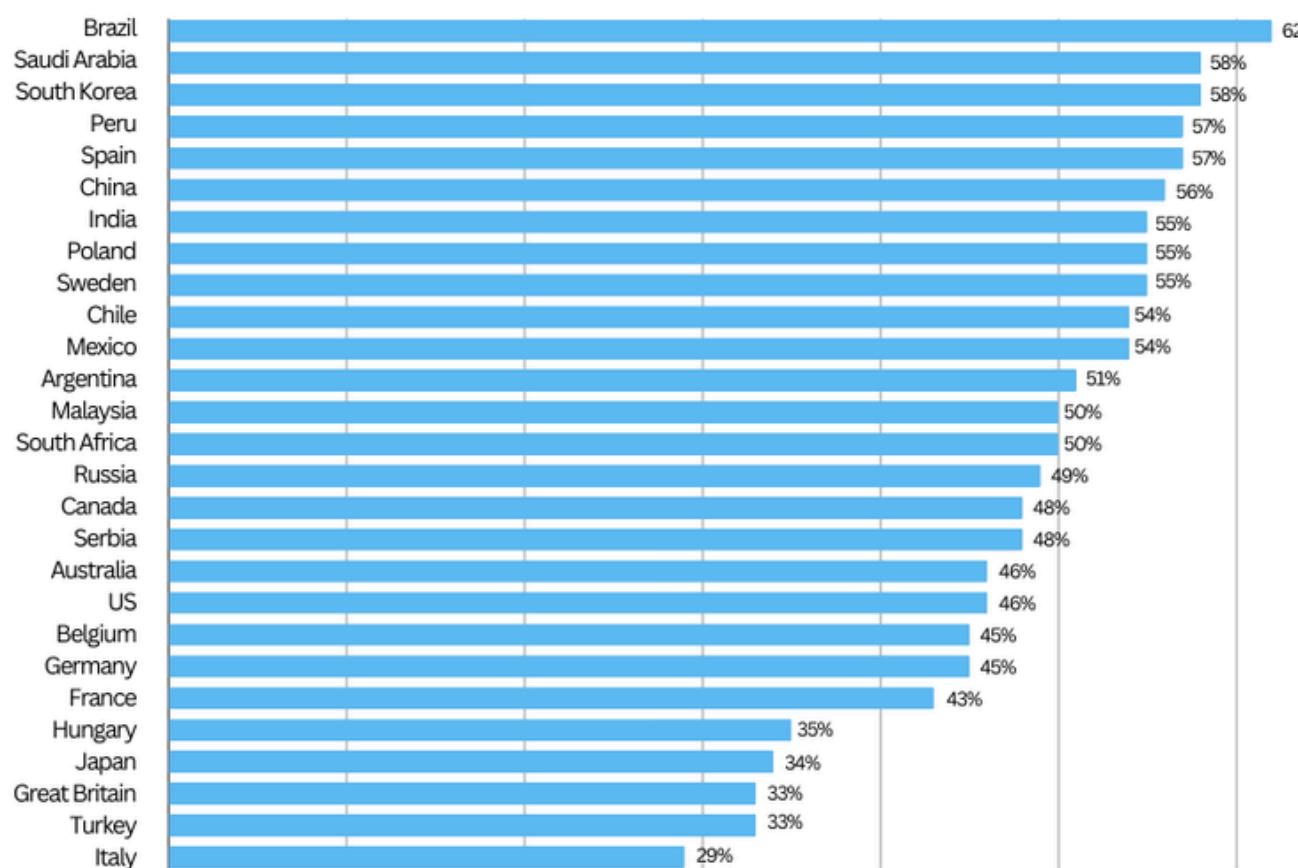


Fake News Prediction Using Machine Learning



Why is Fake News Detection Important?

Share of Respondents Who Initially Believed Fake News to Be Reliable News
From Selected Countries



Redline

Source: Ipsos MORI Social Research Institute

62% of respondents in Brazil believed fake news was reliable, showcasing the global reach of misinformation.



A screenshot of a news article from the website "ScienceInsider". The article is titled "Fake news spreads faster than true news on Twitter—thanks to people, not bots". It states that tweets containing falsehoods were 70% more likely to be retweeted than truthful tweets. The article is dated March 8, 2018, and is written by Katie Langin. The page includes a navigation bar with links to "NEWS", "CAREERS", "COMMENTARY", "JOURNALS", "Science", "DONATE", and a search bar.

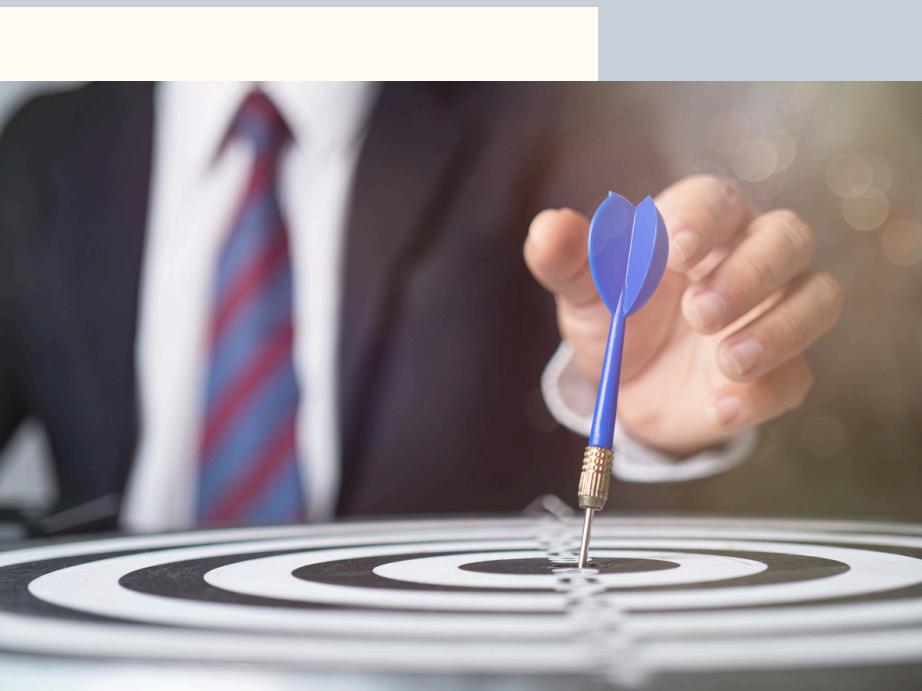
The Problem :

- Fake news spreads rapidly on social media, influencing public opinion and decision-making.
- Traditional fact-checking is too slow to counter the volume of misinformation.

Project Objective

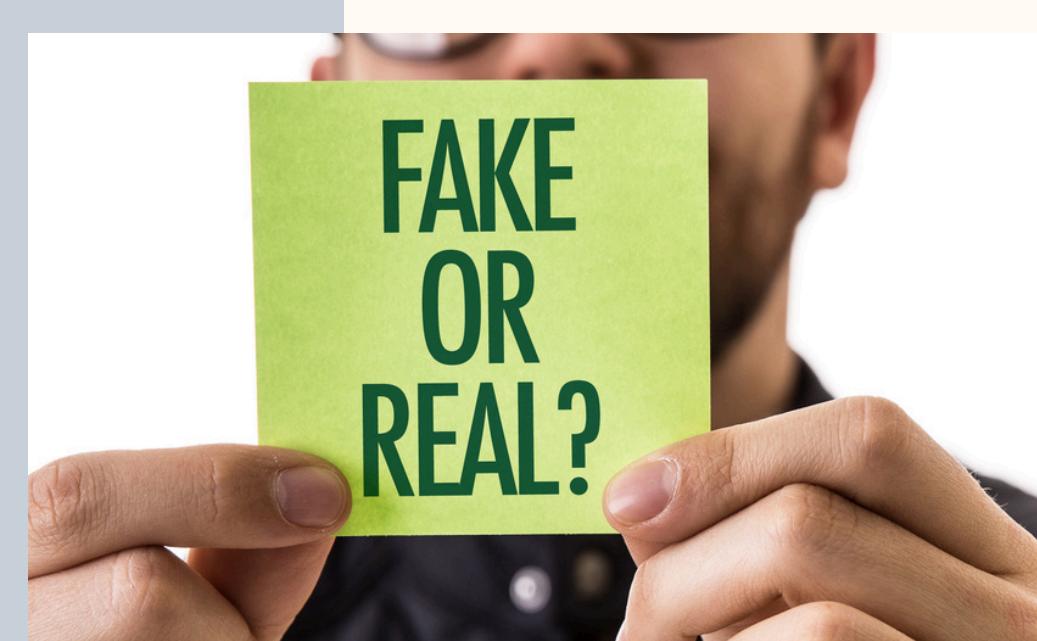
Goal

To build a machine learning model that can classify news articles as FAKE or REAL



Outcome

A robust and scalable model with high accuracy



Why This Matters:

- Helps media platforms filter out false information.
- Promotes informed decision-making among readers.

Dataset Overview

Dataset Source :

<https://www.kaggle.com/datasets/rajatkumar30/fake-news>

Dataset includes 6335 news articles, with columns:

- **Title: News headline.**
- **Text: Full article content.**
- **Label: REAL or FAKE.**

Label Distribution:

- **Approximately 50% FAKE and 50% REAL (balanced dataset).**

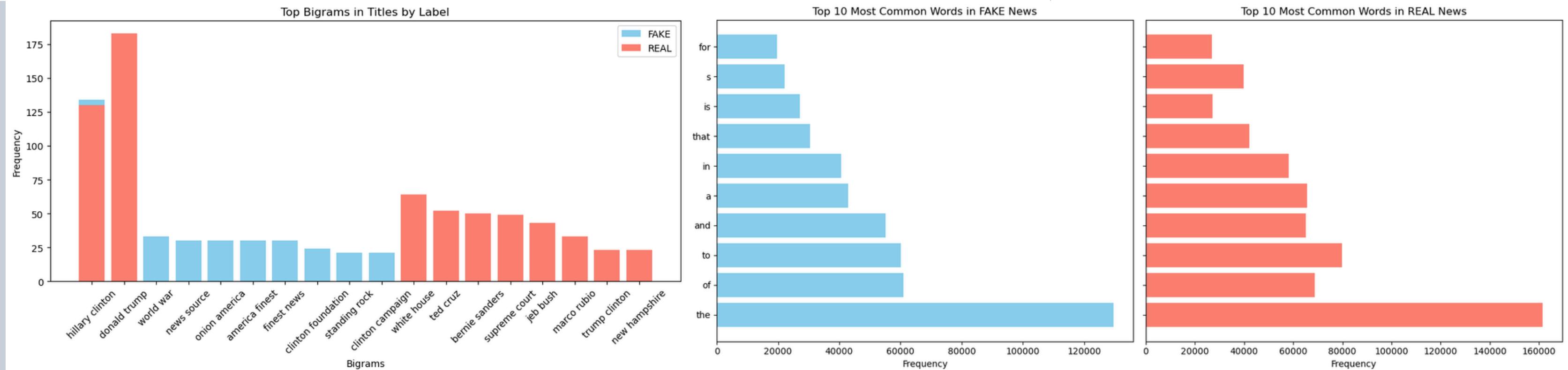
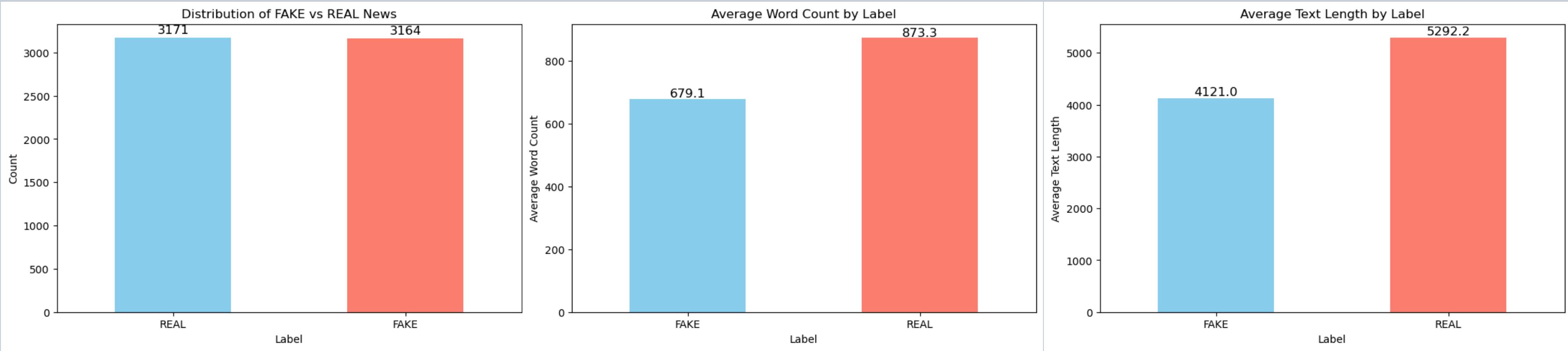
Data Example :

title	text	label
You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL



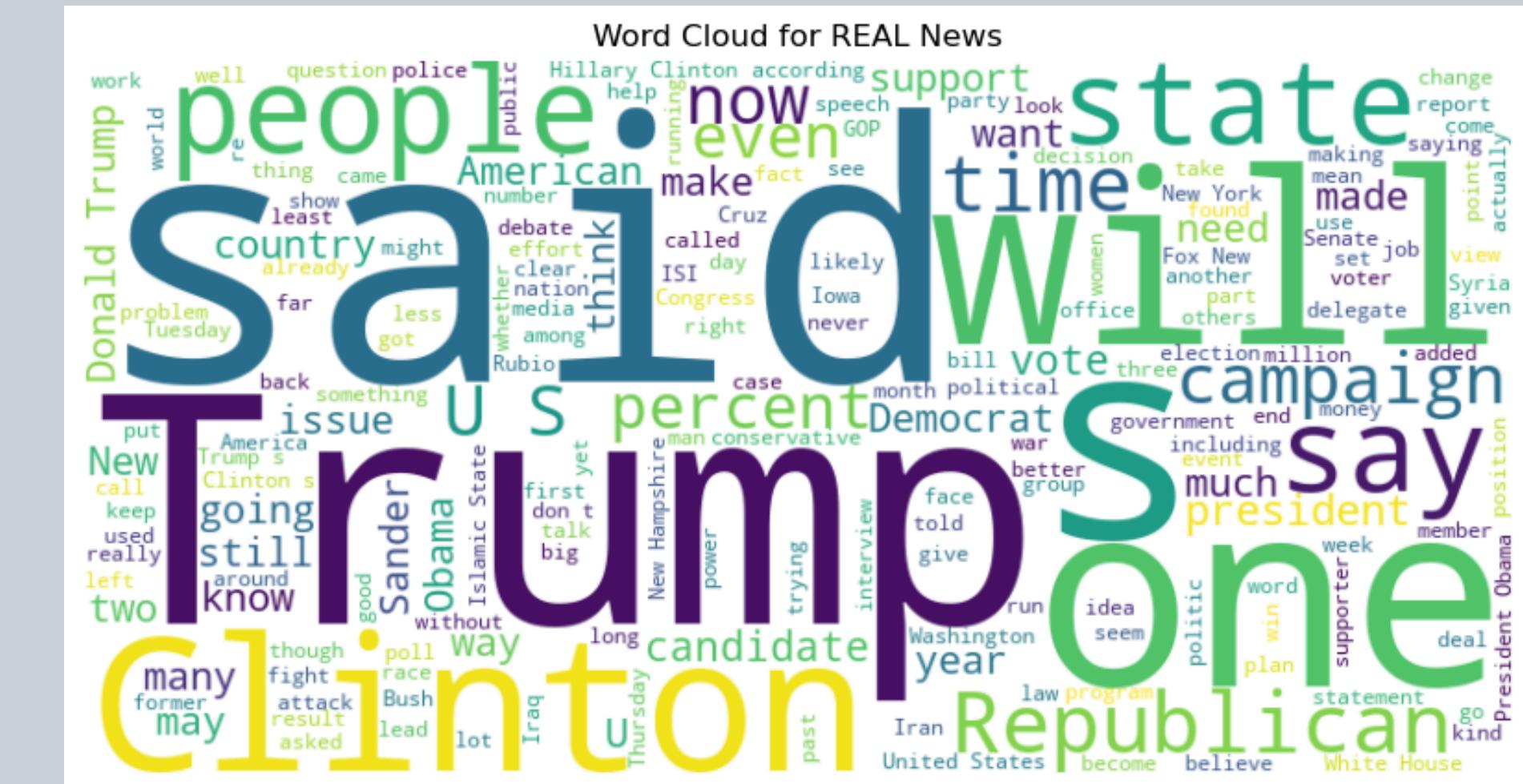
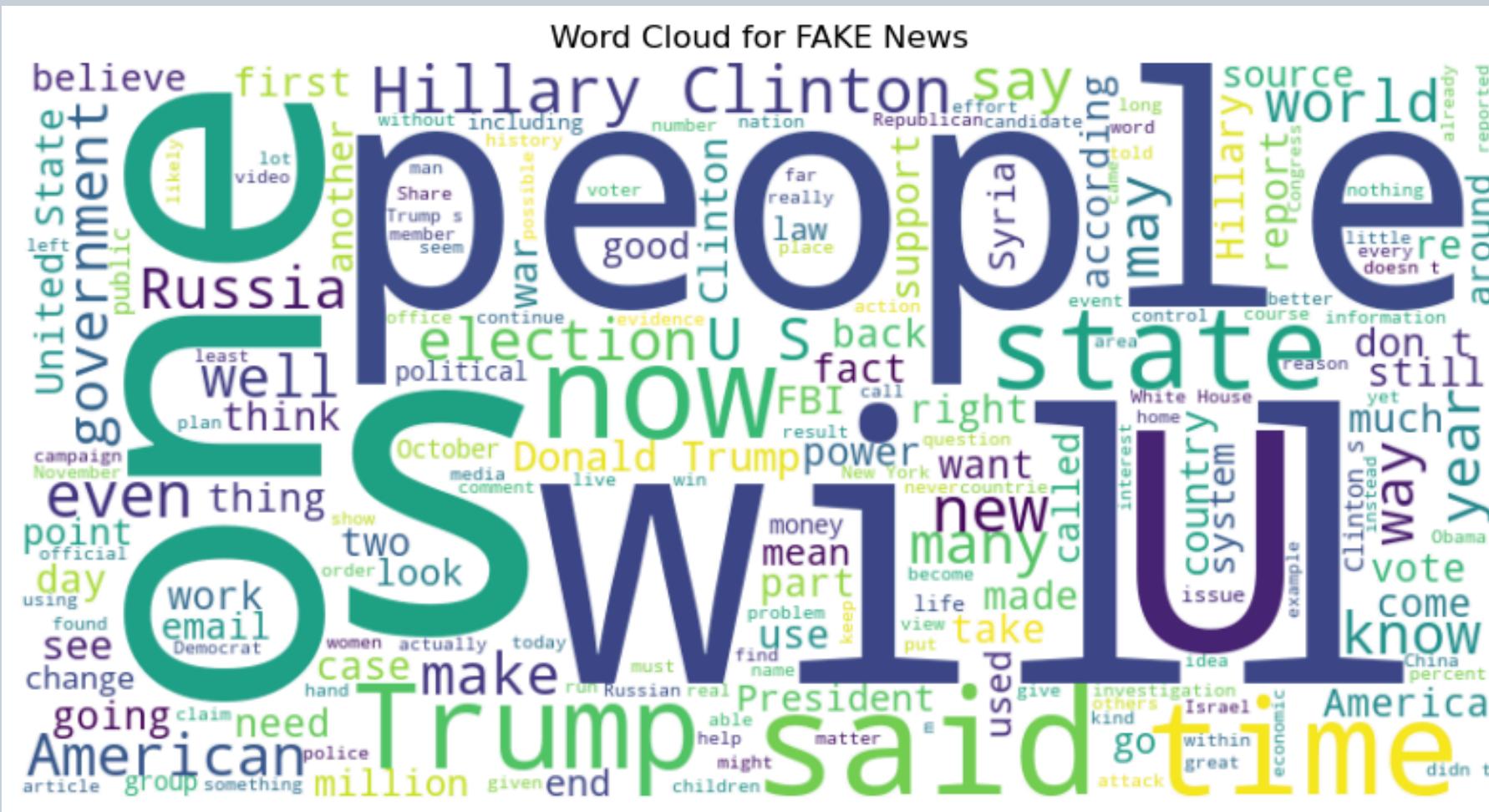
Exploratory Data Analysis

Data Distribution



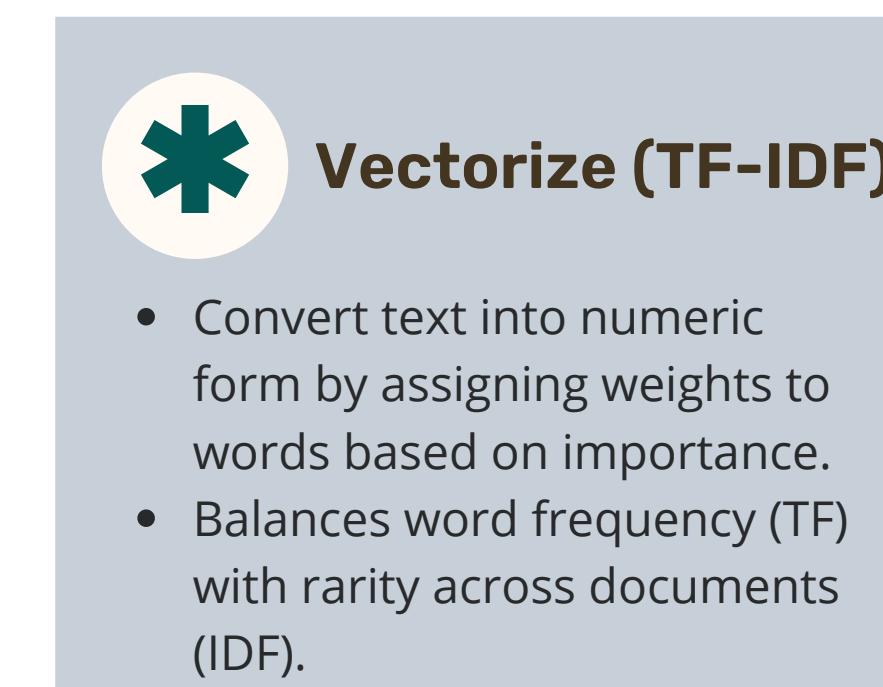
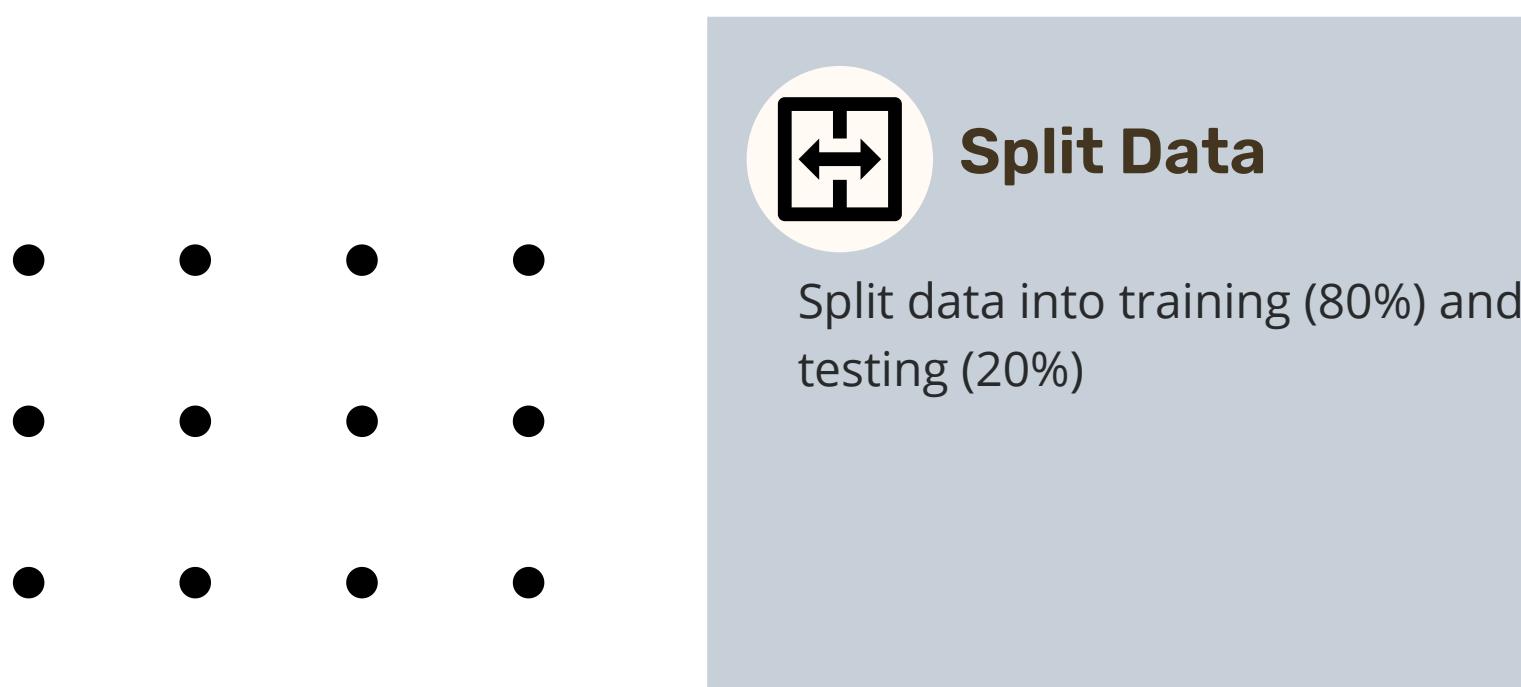
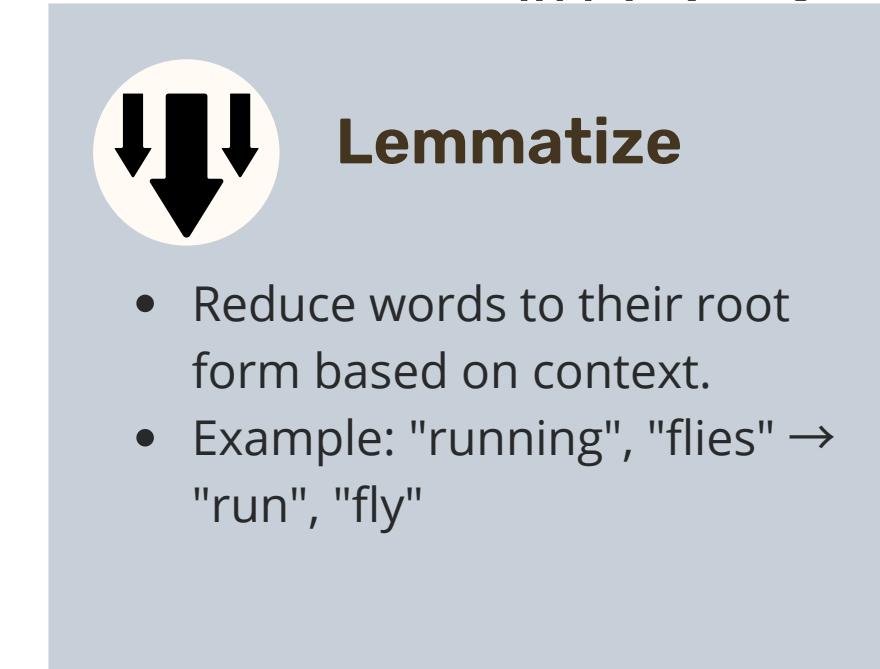
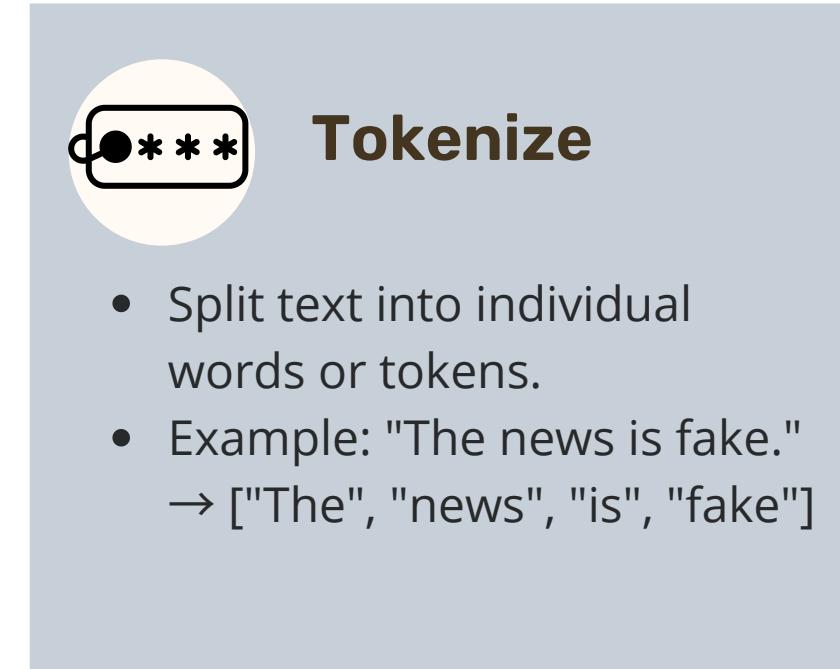
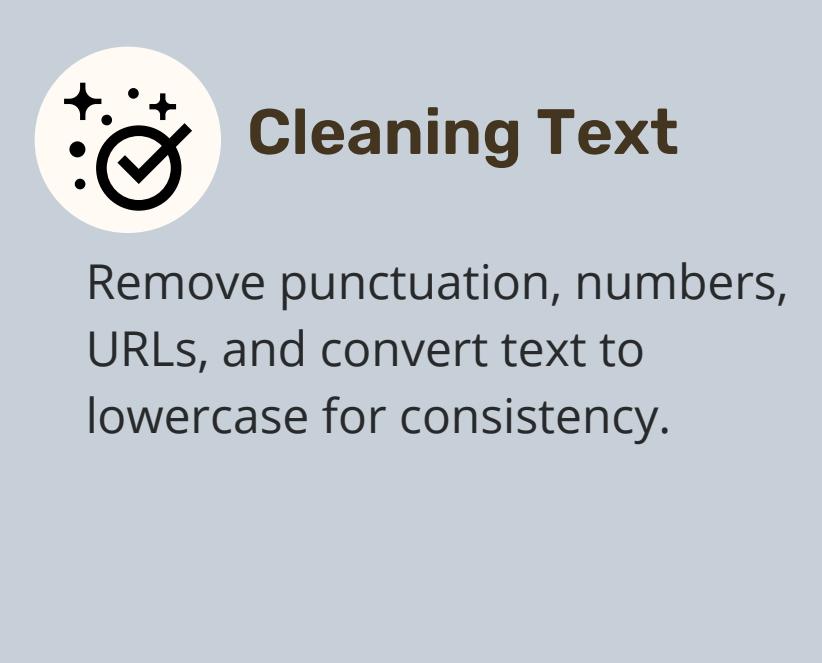
Exploratory Data Analysis

Word Cloud



Building the Prediction Model

Preprocessing :

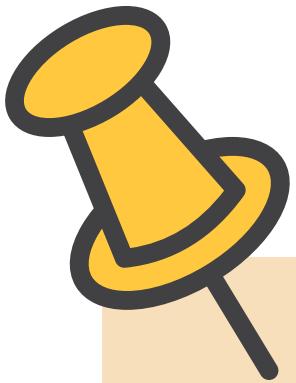




Algorithm Tested

- Logistic Regression
- SVM
- Random Forest
- Naive Bayes
- XGBoost

Model Evaluation



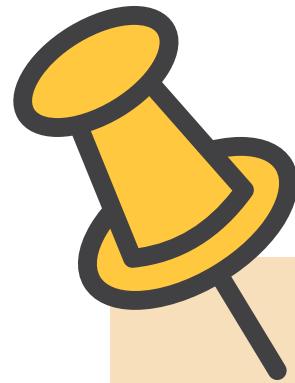
Logistic Regression

Logistic Regression Accuracy: 0.8958

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.90	0.89	0.89	628
1	0.90	0.90	0.90	639

accuracy			0.90	1267
macro avg	0.90	0.90	0.90	1267
weighted avg	0.90	0.90	0.90	1267



SVM

SVM Accuracy: 0.9013

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.90	0.90	0.90	628
1	0.91	0.90	0.90	639

accuracy			0.90	1267
macro avg	0.90	0.90	0.90	1267
weighted avg	0.90	0.90	0.90	1267

Model Evaluation



Random Forest

Random Forest Accuracy: 0.9148

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.91	0.92	0.91	628
1	0.92	0.91	0.92	639

accuracy			0.91	1267
macro avg	0.91	0.91	0.91	1267
weighted avg	0.91	0.91	0.91	1267



Naive Bayes

Naive Bayes Accuracy: 0.8548

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.84	0.88	0.86	628
1	0.87	0.83	0.85	639

accuracy				0.85
macro avg		0.86	0.85	0.85
weighted avg	0.86	0.85	0.85	1267



XGBoost

XGBoost Accuracy: 0.9132

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.91	0.92	0.91	628
1	0.92	0.91	0.91	639

accuracy				0.91
macro avg		0.91	0.91	0.91
weighted avg	0.91	0.91	0.91	1267

Conclusion

- Random Forest achieved the highest performance with 91.48% accuracy, strong precision, recall, and F1-scores.
- XGBoost followed closely, showing competitive results with 91.32% accuracy.
- Ensemble models (Random Forest, XGBoost) outperform linear models (Logistic Regression, Naive Bayes) on this dataset.
- Preprocessing techniques like TF-IDF vectorization and tokenization significantly enhanced model performance.
- Random Forest demonstrated excellent balance in identifying both fake and real news, minimizing false predictions.

Next Steps

- Optimize hyperparameters for Random Forest and XGBoost to further improve accuracy and generalizability.
- Integrate the best model into a real-world system for fake news detection on news or social media platforms.
- Gather a more diverse dataset to address biases and ensure robustness for various types of fake news.
- Combine fake news detection with sentiment analysis to better understand emotional patterns in misinformation.

Thank You