

Data 606 Capstone - Fall 2023

Machine Learning-Driven Authenticity Verification in Digital Journalism

RUTUJA KARAD

November | 2023



The Relevance of Fake News Detection in the Digital Age

Reliable Information Imperative:

- Addressing the digital age's challenge of ensuring accuracy amidst a sea of content.

Restoring Digital Journalism's Credibility:

- Countering misinformation with automated detection tools to regain public trust.

Empowering Decisions:

- Providing a filter against fake news for informed individual and organizational choices.

Diverse Research Goals:

- Evaluating algorithms, uncovering misinformation patterns, and studying its societal spread.



Dataset Overview

PRIMARY OBJECTIVES:

1. Devise a high accuracy model for predicting Fake news .
2. Develop a user-friendly platform tailored to everyday users for easy identification and flagging of fake news.

1. File Name: Fake.csv

- Size: 57.73MB
- Rows: 18,471
- Columns: 4

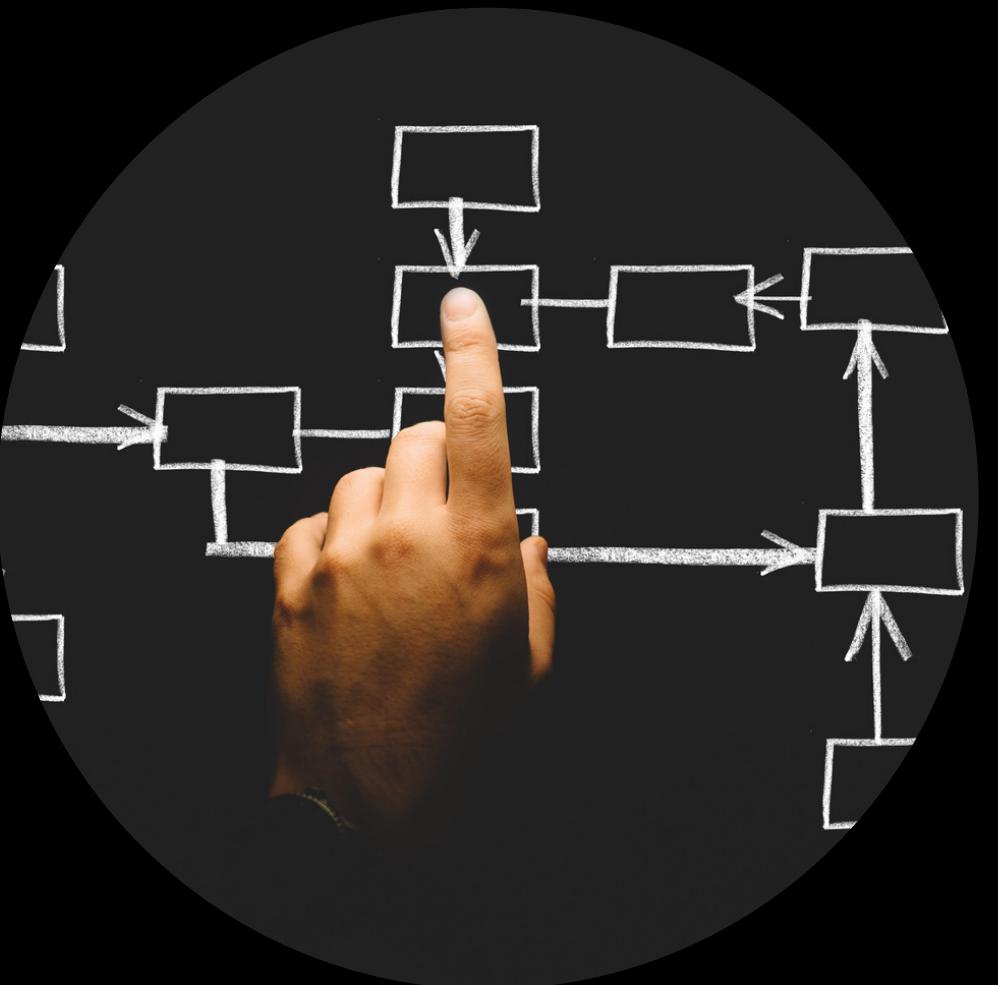
2. File Name: True.csv

- Size: 46.62MB
- Rows: 14,915
- Columns: 4



Project Methodology

1. Preprocessing and EDA
2. Model Building
3. Hyper Parameter Tuning
4. Web-app Development and Deployment



Preprocessing and EDA

1. Preprocessing Performed:

- a. Checked and removed null values
- b. Removed special characters

2. Performed EDA

- a. Analyzed the data and generated graphs
- b. Derived insights

3. NLP Preprocessing

- a. Removed stop words
- b. Stemming
- c. Lemmatization
- d. N-Gram analysis

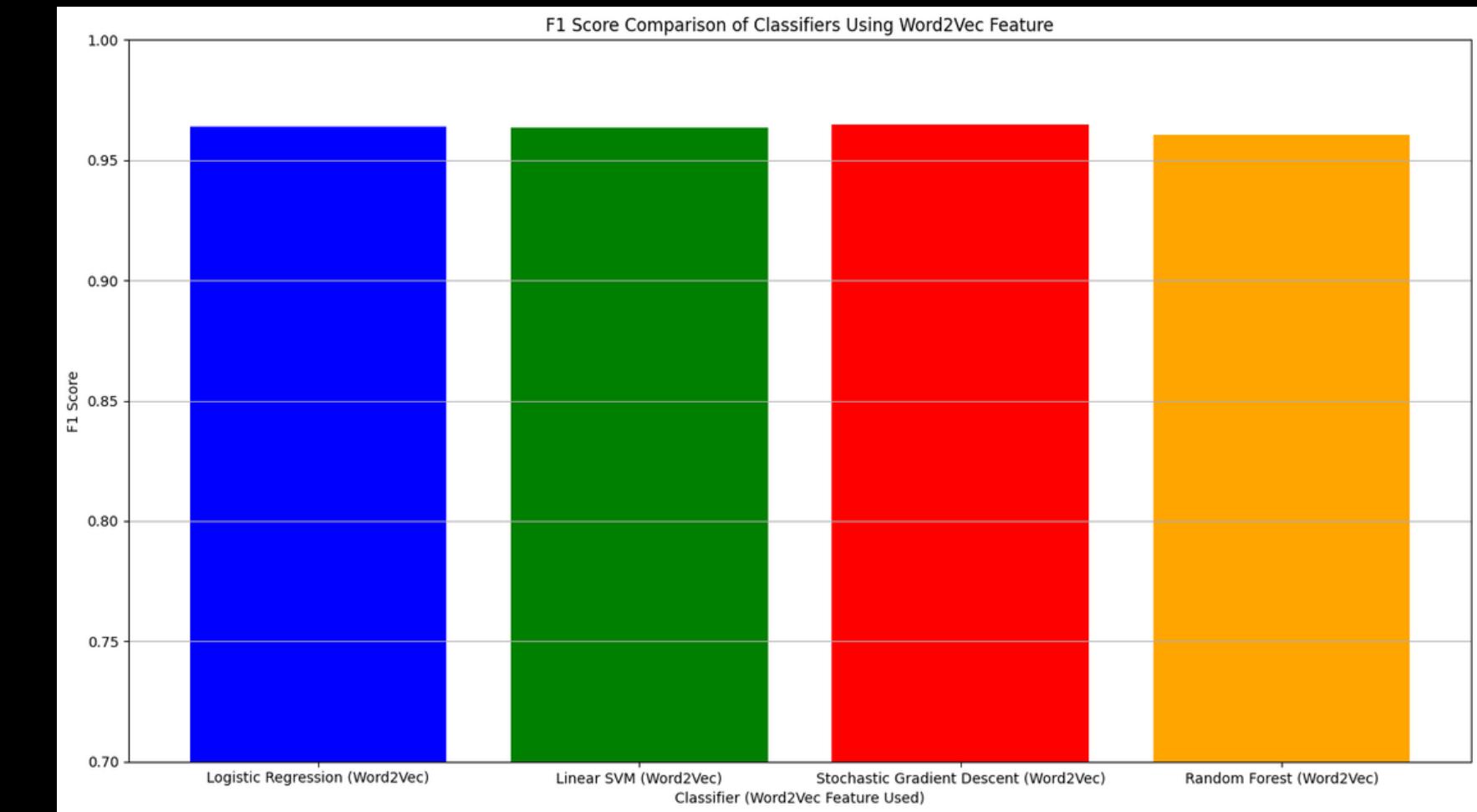
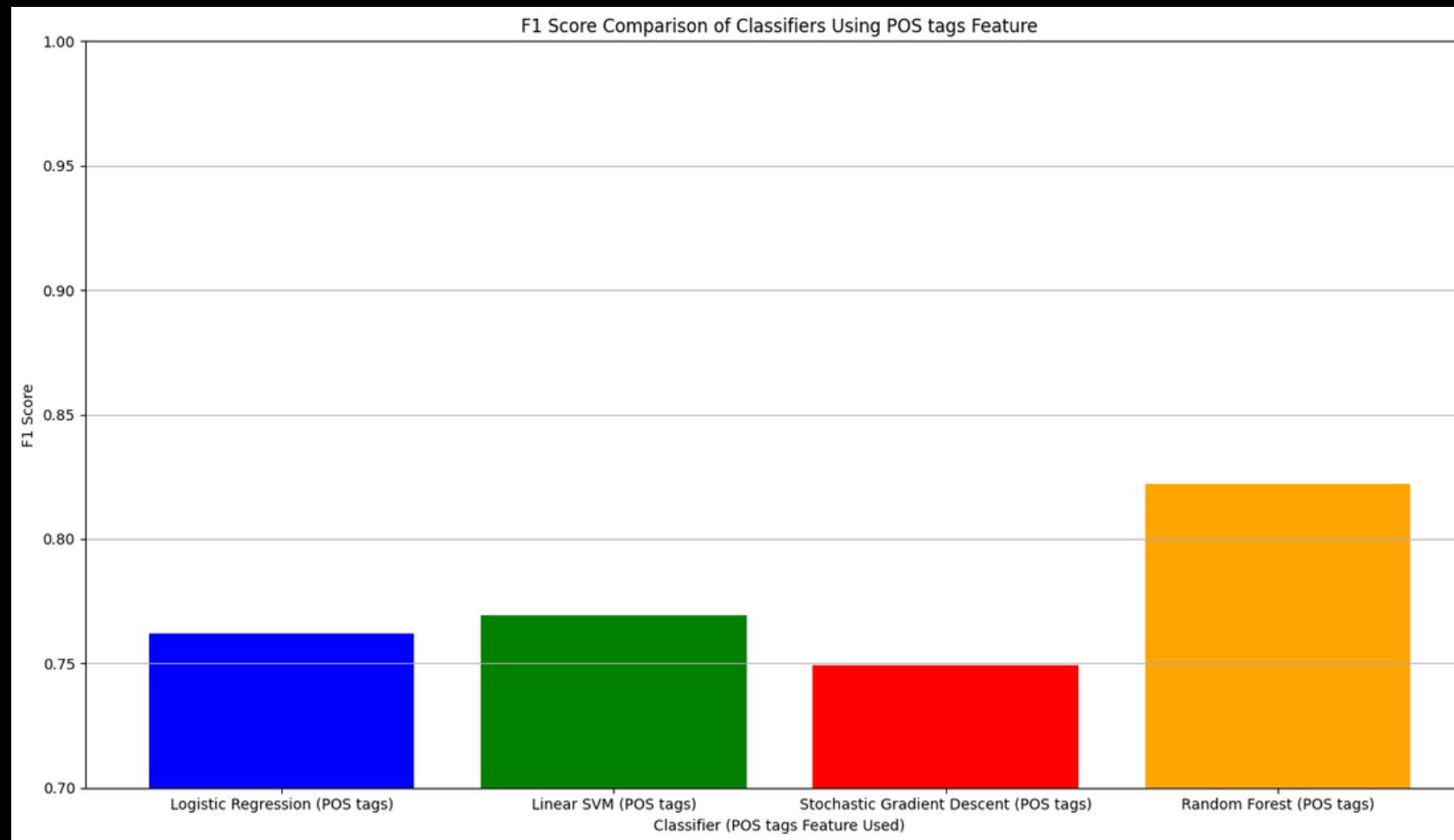


Model Building

- Used Two Methods Part of Speech Tagging and Word2Vec Embeddings
- Trained models on both separately and choose best performing model - models on Word2Vec outperformed POS.
- Classification Algorithms used:
 - Linear SVM
 - Random Forest
 - Logistic Regression
 - Stochastic Gradient Descent



Model Comparison (F1 Scores)



Hyperparameter Tuning

- Used Grid search to perform hyperparameter tuning on best performing models - Logistic Regression using w2v
- Best Parameters
 - $C \geq 10$
 - Regularization = L2

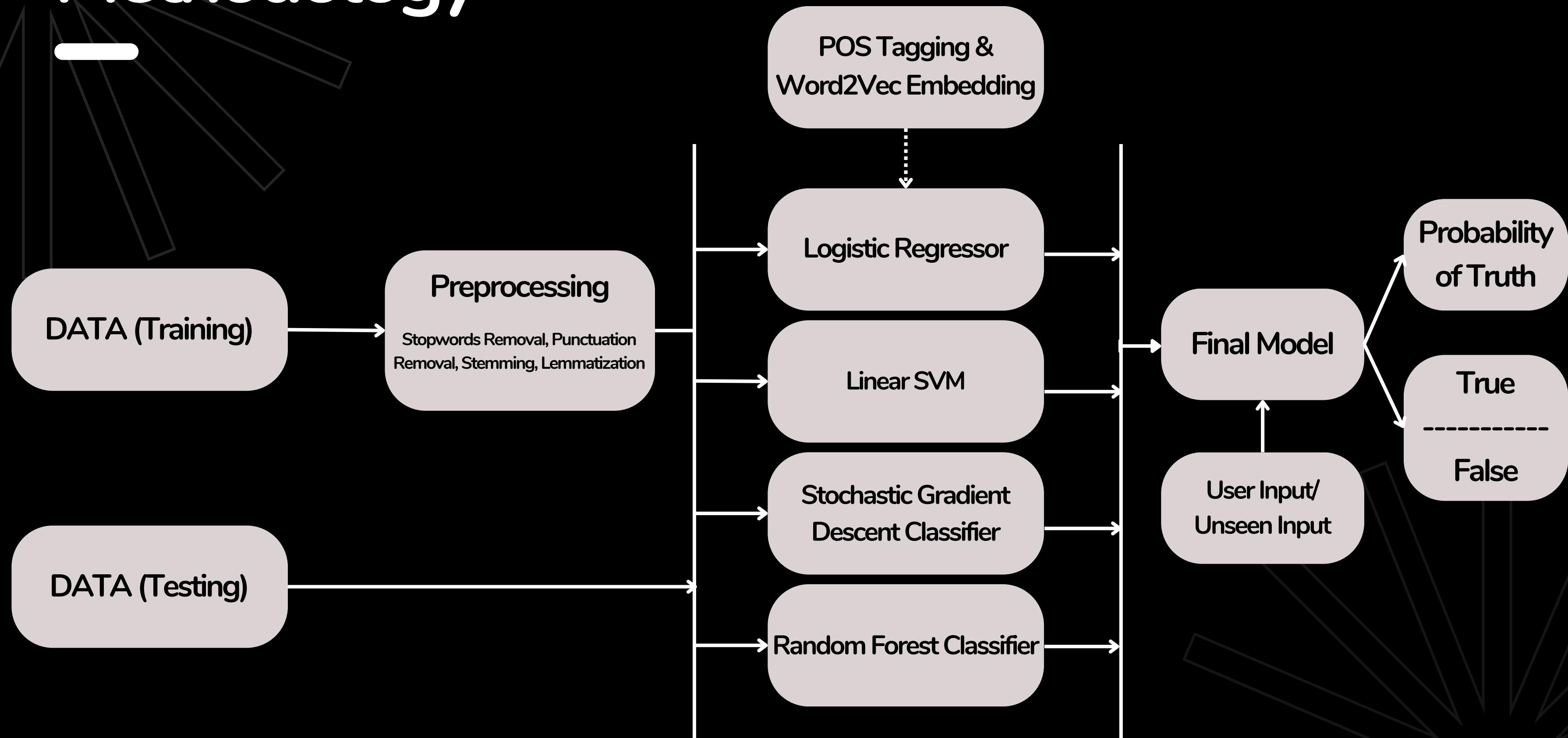


Webapp Development

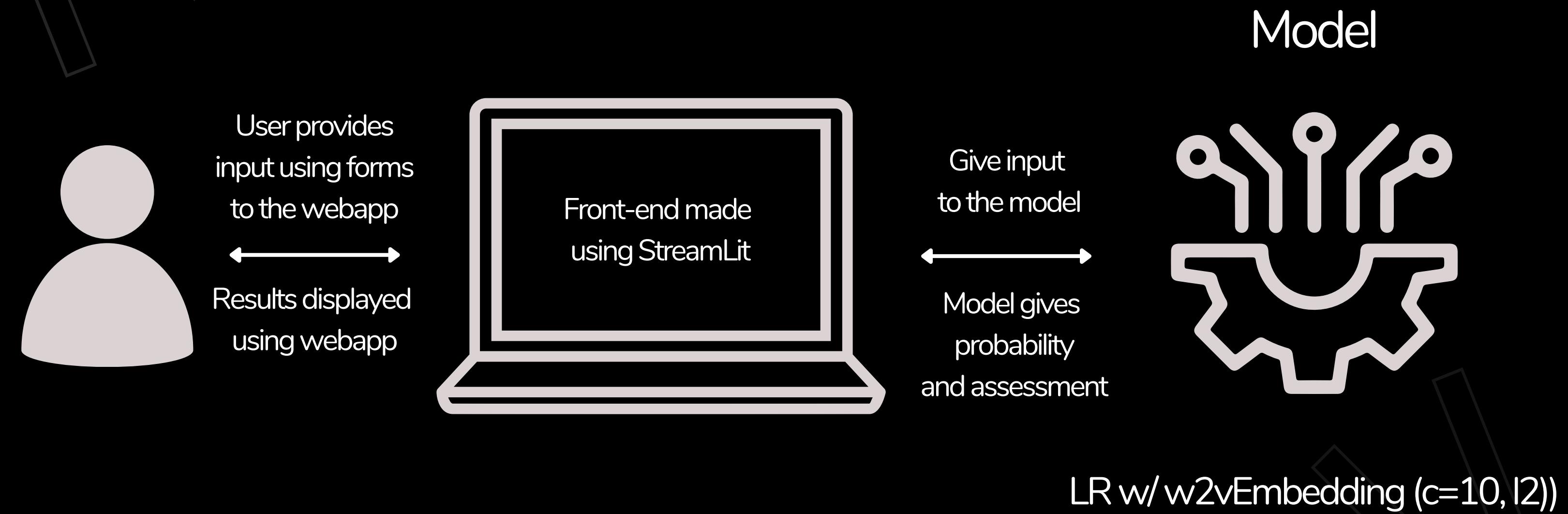
- Used Streamlit library for integration of HTML and Python
- Used a Text Input field and submit button to get input and run model on it.
- Displayed probability of input text being real or fake



Methodology



System Architecture



Deploy :

Fake News Detection

Enter Text

Donald Trump just couldn't wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his enemies, haters and the very dishonest fake news media. The former reality show star had just one job to do and he couldn't do it. As our Country rapidly grows stronger and smarter. I want to wish all of my friends. supporters. enemies. haters. and even the very

Predict

The probability of the news being fake is: 99.98%

The probability of the news being real is: 0.02%

Deploy



🔗 Fake News Detection

Enter Text

JAKARTA (Reuters) - Indonesia will buy 11 Sukhoi fighter jets worth \$1.14 billion from Russia in exchange for cash and Indonesian commodities, two cabinet ministers said on Tuesday. The Southeast Asian country has pledged to ship up to \$570 million worth of commodities in addition to cash to pay for the Suhkoi SU-35 fighter jets, which are expected to be delivered in stages starting in .

Predict

The probability of the news being fake is: 0.12%

The probability of the news being real is: 99.88%



Closing Remarks

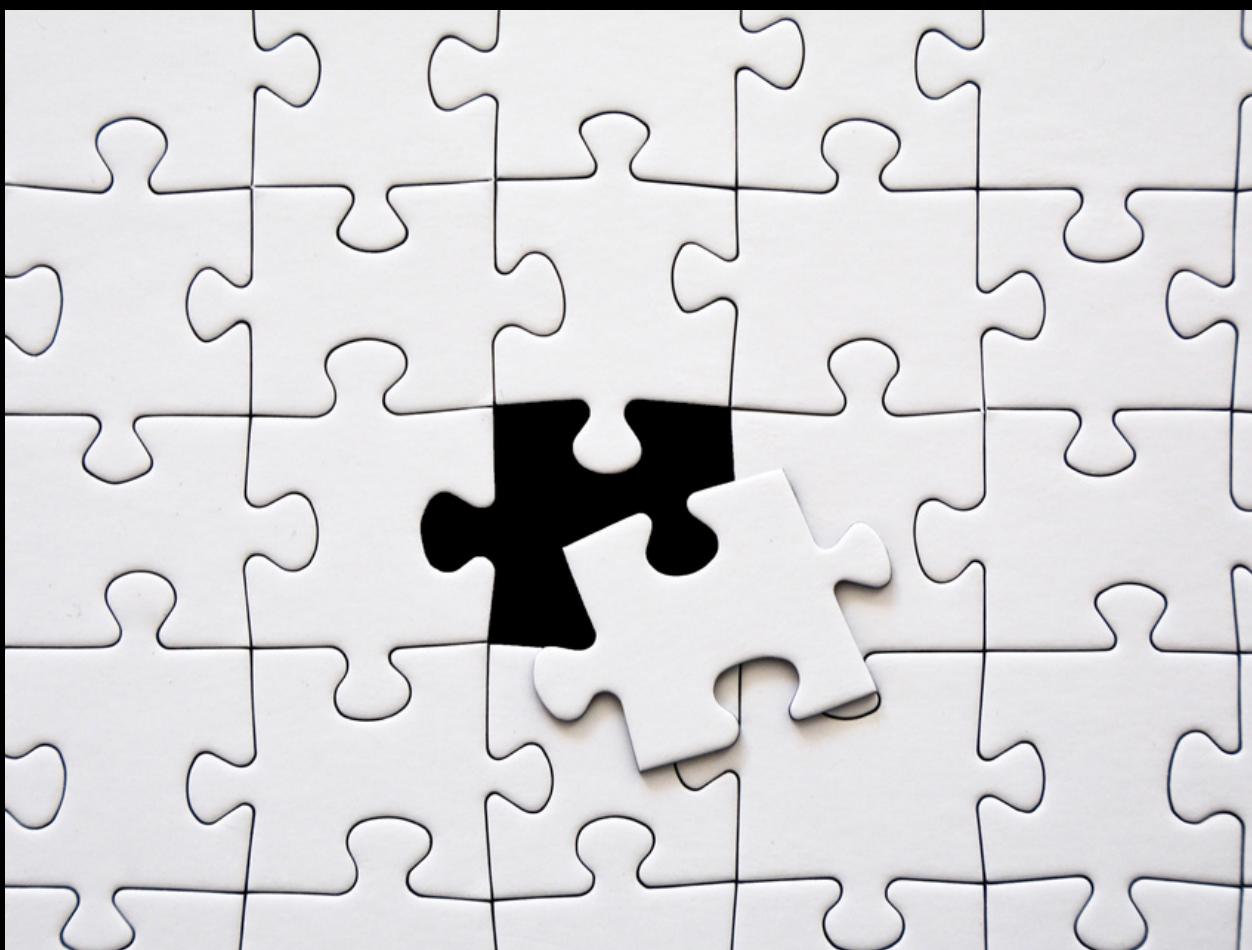
Successfully Implemented Fake news Detection using NLP and classification Algorithms.

Successfully developed and deployed a web application for inter-activeness.

Applications

Social Media Fact Checking

Utilizing NLP for fake news detection to identify and flag false information, enhancing integrity and safeguarding users from misinformation.



Election Period Turmoil

In the political arena, NLP-based fake news detection can be crucial in maintaining fair and unbiased election processes by identifying and mitigating the spread of false information that could influence voters.

Consumer Protection

Detecting fake reviews and fraudulent product descriptions online, thereby protecting consumers from deceptive practices.



Thank You



rutujak1@umbc.edu