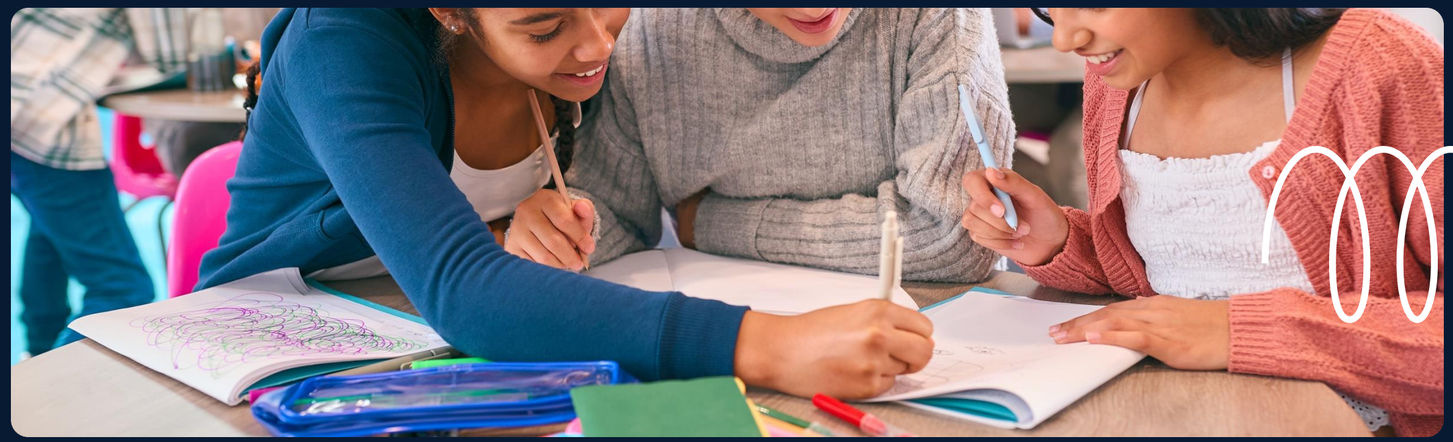
Group Assignment

Fake News Detection Case Study



Team members: Gowtham, Namratha

Pipeline Stages

- Data Preparation
- Text Preprocessing
- Train Validation Split
- EDA on Training Data
- Feature Extraction
- Model Training & Evaluation

Key Findings

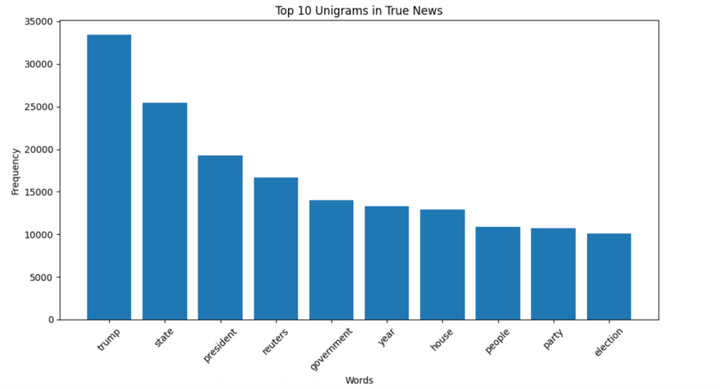
- Most articles fall between 500–1500 characters, with little difference between cleaned and lemmatized text, confirming effective and consistent text preprocessing.
- True news articles emphasize official institutions and entities, suggesting structured and policy-based reporting. Ex: President, Government, State et.,
- Fake news content tends to focus on political figures and controversies, indicating sensationalism or conspiracy bias. Ex: Trump, Obama, Hilary etc.,
- True news articles consistently use formal, factual language centered around institutions and timelines—evident from frequent n-grams like "white house", "president trump", and "new york times"—which indicates structured, credible reporting typical of journalistic standards.
- Fake news content prominently features named political figures and controversial topics—reflected in phrases like "hillary clinton email", "email scandal", and "wikileaks email dump"—revealing a pattern of emotionally charged and repetitive language aimed at driving engagement over accuracy.
- Overall – Logistic regression seems to be working better compare to other model types for this use case

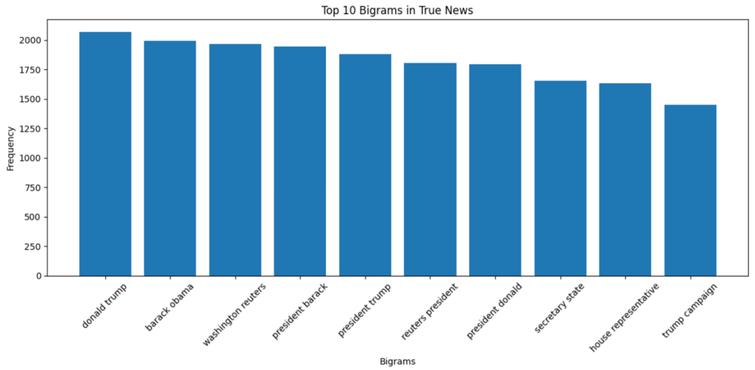
True News

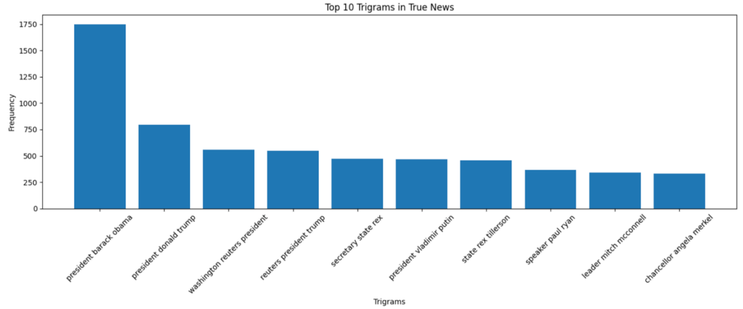
Analysis

Unigram, Bigram, Trigram

Word cloud





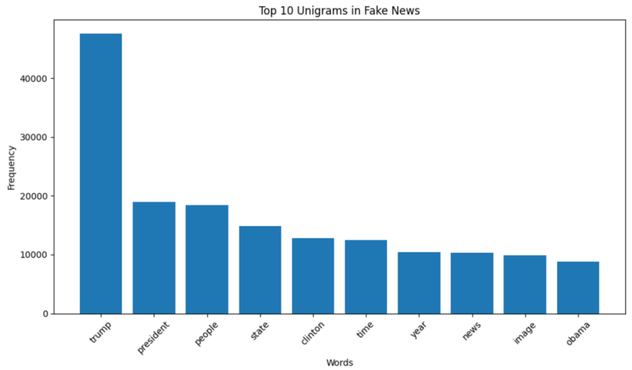


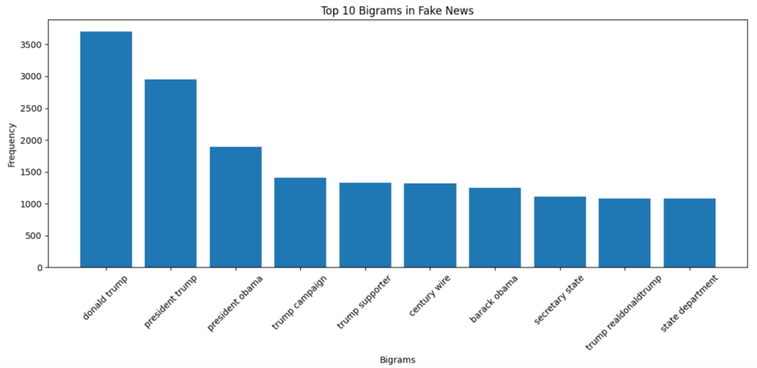
Fake News

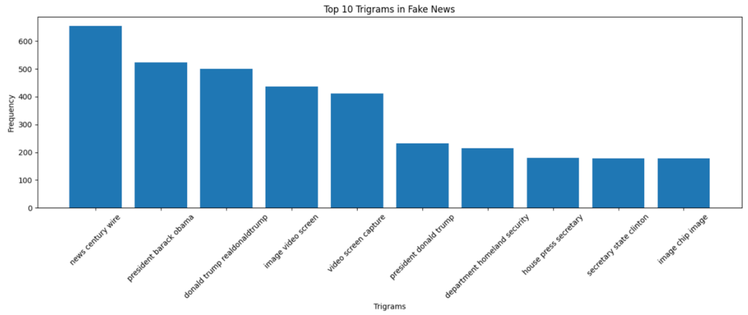
Analysis

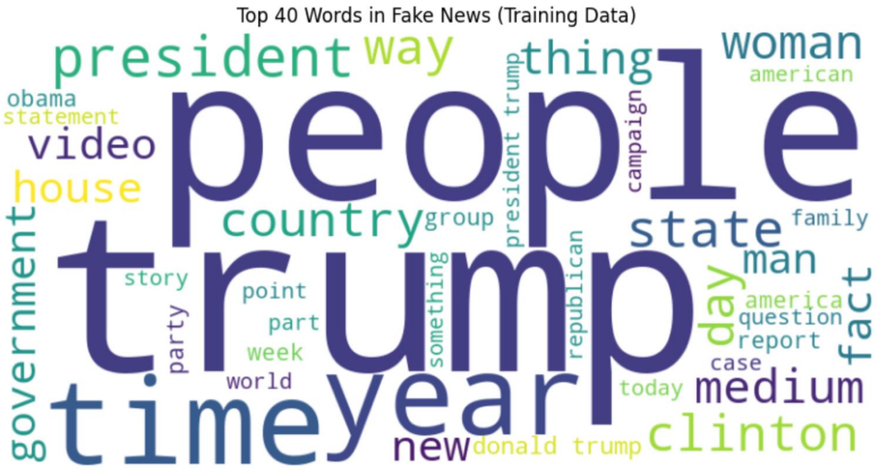
Unigram, Bigram, Trigram

Word cloud









Model Comparison

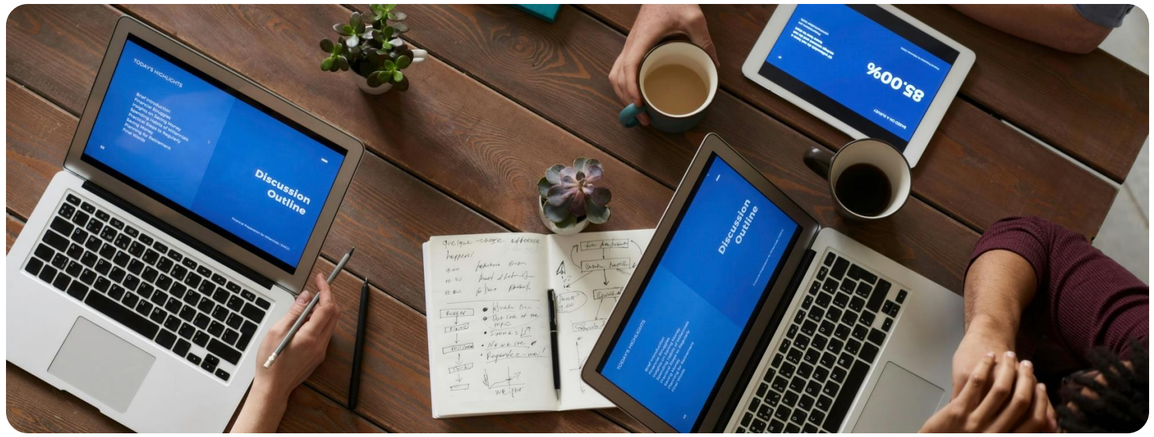
Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.932873	0.924105	0.936187	0.930107
Decision Tree	0.849484	0.854220	0.825370	0.839547
Random Forest	0.927527	0.932941	0.913774	0.923258

Best Model: Logistic Regression

Why: Achieved the highest F1-score, with reliable generalization performance and interpretability.

Impact: This model helps in automatically identifying fake news with high precision and recall, potentially aiding social media platforms, content moderation, and fact-checking tools.

Thank you for your attention and support



Email

gowtham.balram@gmail.com

grandhi.namm@gmail.com