College of Engineering
Effat University

CS4083 Final Project Presentation

Detecting Fake News Using NLP, Machine Learning, and Deep Learning Techniques

Presented by : Araa AlMarhabi
Instructor : Dr. Naila Miror

Text Mining and Natural language processing

Table of contents

Introduction

In the digital age, the spread of fake news has become a significant challenge, impacting public trust and decision-making.

This project aims to use Natural Language Processing (NLP), machine learning, and deep learning to build a system that identifies fake news articles. By analyzing a dataset of over 44,000 labeled news articles, the project extracts meaningful features, evaluates multiple models, and identifies the most effective approach for detecting fake news.

Literature Review

Author(s)	Title	Technique(s) Used	Findings	Gaps
J. Jouhar et al., 2024	Fake News Detection using Python and Machine Learning	Machine Learning (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, PAC)	XGBoost outperformed other models with high accuracy, precision, recall, and F1-score. Emphasized preprocessing and feature engineering.	Difficulty with multimedia content and platform adaptation. Suggested exploration of advanced vectorizers and real-time monitoring strategies.
R. Oshikawa et al., 2018	A Survey on Natural Language Processing for Fake News Detection	Review of NLP methods (SVM, Naive Bayes, LSTMs, CNNs, Attention Mechanisms)	Reviewed datasets, methodologies, and challenges in NLP for fake news detection. Identified strengths of various techniques.	Dataset limitations (narrow scope, bias), over-reliance on metadata, and lack of multimodal approaches.
K. Shu et al., 2017	Beyond News Contents: The Role of Social Context for Fake News Detection	TriFN framework integrating social context with content analysis	TriFN outperformed content-based methods in early-stage detection and modeled user-publisher-news relationships.	Dependence on social data, binary classification simplification, scalability, and privacy concerns.

Literature Review

Author(s)	Title	Technique(s) Used	Findings	Gaps
A. Thota et al., 2018	Fake News Detection: A Deep Learning Approach	Deep Learning (TF-IDF, Bag of Words, Word2Vec, Neural Networks)	Achieved 94.21% accuracy using TF-IDF with engineered features. Highlighted challenges with stance detection.	Limited dataset scope, difficulty generalizing to diverse formats, and poor performance with pre-trained embeddings for long texts.
N. N. Prachi et al., 2022	Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms	ML and DL (Logistic Regression, Decision Tree, Naive Bayes, SVM, LSTM, BERT)	BERT achieved the highest accuracy (98%), followed by LSTM (95%). Showcased deep learning robustness.	Dataset dependency, focus on textual data, ethical concerns, and challenges in generalization.
Z. Khanam et al., 2021	Fake News Detection Using Machine Learning Approaches	ML (Random Forest, SVM, Naive Bayes, XGBoost)	XGBoost achieved the highest accuracy (75%), emphasizing preprocessing and feature selection.	Limited dataset scope, reliance on textual data, and challenges with nuanced misinformation types.

Methodology

Dataset

kaggle

Description

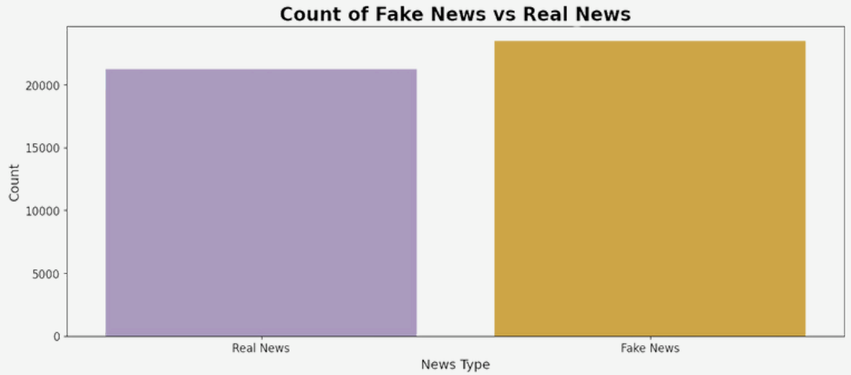
Specifically curated for fake news detection tasks, divided into Fake.csv and True.csv.

Size

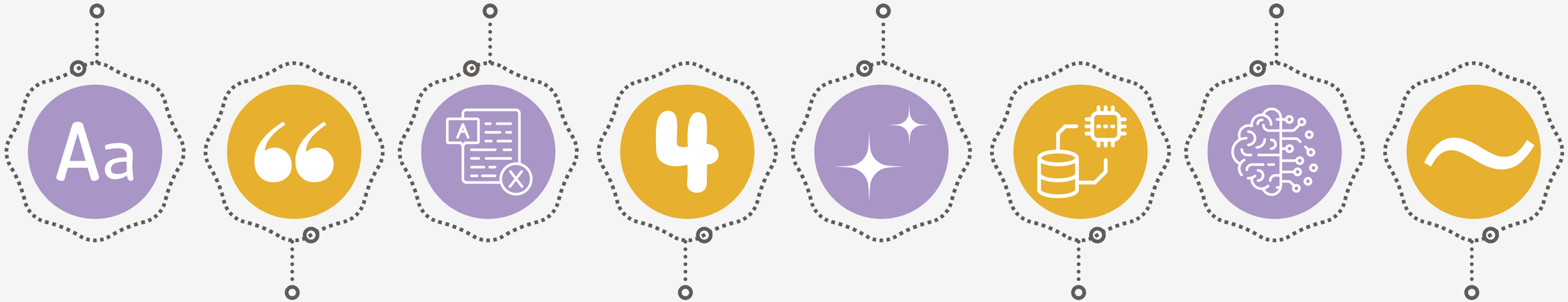
- Fake news: 23,502
- Real news: 21,417
- Overall: 44,919 articles.

Features

Includes title, text, subject, and publication date of each article.



Data Preprocessing



Data Preprocessing

Before

```
0    Indonesia reopens Bali airport as wind clears ...
1    Two Republican senators blast Trump as party f...
2    MUST WATCH VIDEO! TRUMP’S MINI-ME Steals The S...
3    Obama says it is possible Russia would try to ...
4     Angry That Benghazi Panel Couldn’t Bury Hilla...
Name: news, dtype: object
```

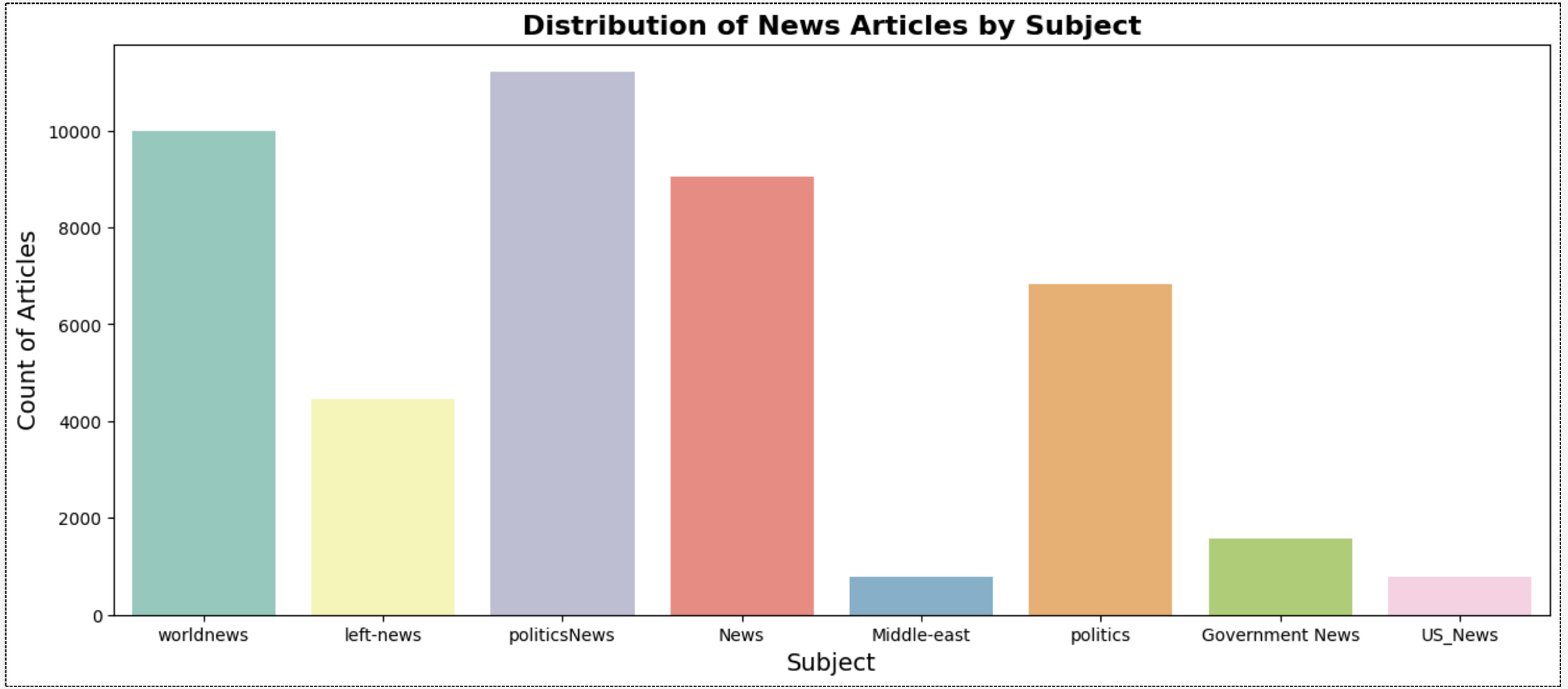
After

```
0    indonesia reopen bali airport wind clear volca...
1    two republican senat blast trump parti feud de...
2    must watch video trump ’ minim steal show “ do...
3    obama say possibl russia would tri sway us ele...
4    angri benghazi panel ’ buri hillari rightw nut...
Name: news, dtype: object
```

#Test

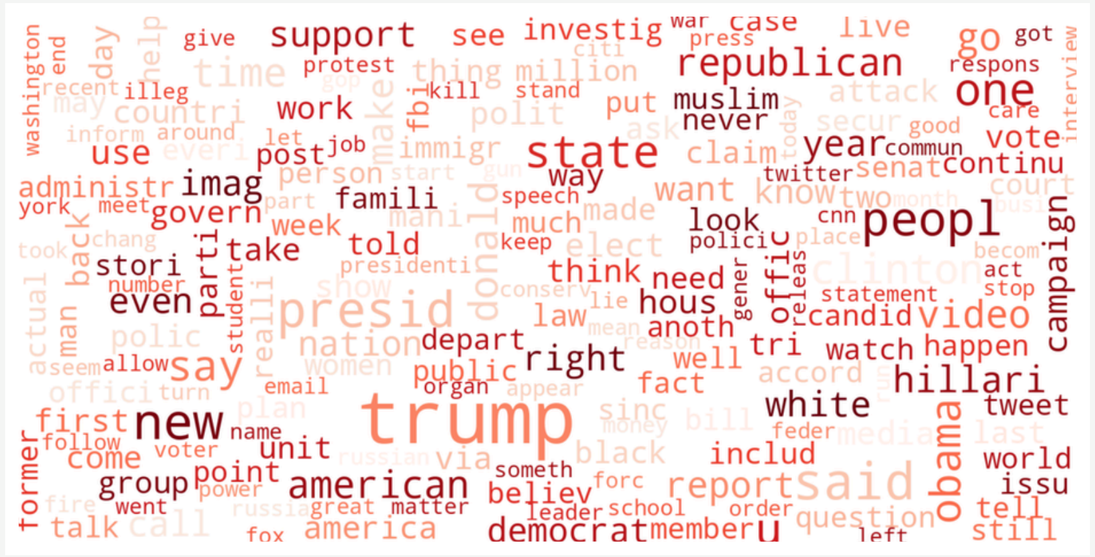
```
Please enter a news text: Breaking News! 🚀 Visit https://news.com for details. The event happened at 9 AM today.

Processed Text:
break news visit httpsnewscom detail event happen today
```

Exploratory Data Analysis (EDA)



Exploratory Data Analysis (EDA)

Word Cloud Visualization of
Real News Headlines and Text

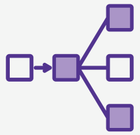


Word Cloud Visualization of
Fake News Headlines and Text

Feature Extraction

TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical method that identifies important and distinctive words in text, making it effective for fake news detection by highlighting terms that differentiate fake from real news articles.



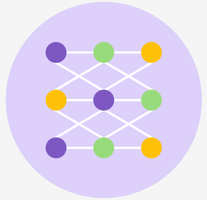
Word2Vec

Word2Vec is a deep learning-based embedding technique that converts words into dense vector representations, capturing their semantic relationships, which is useful in fake news detection for understanding context and identifying patterns in textual data.

Train & Evaluation Models

Machine Learning

Machine learning models, such as Logistic Regression and Random Forest, classify fake and real news by analyzing patterns in text features like word frequency.

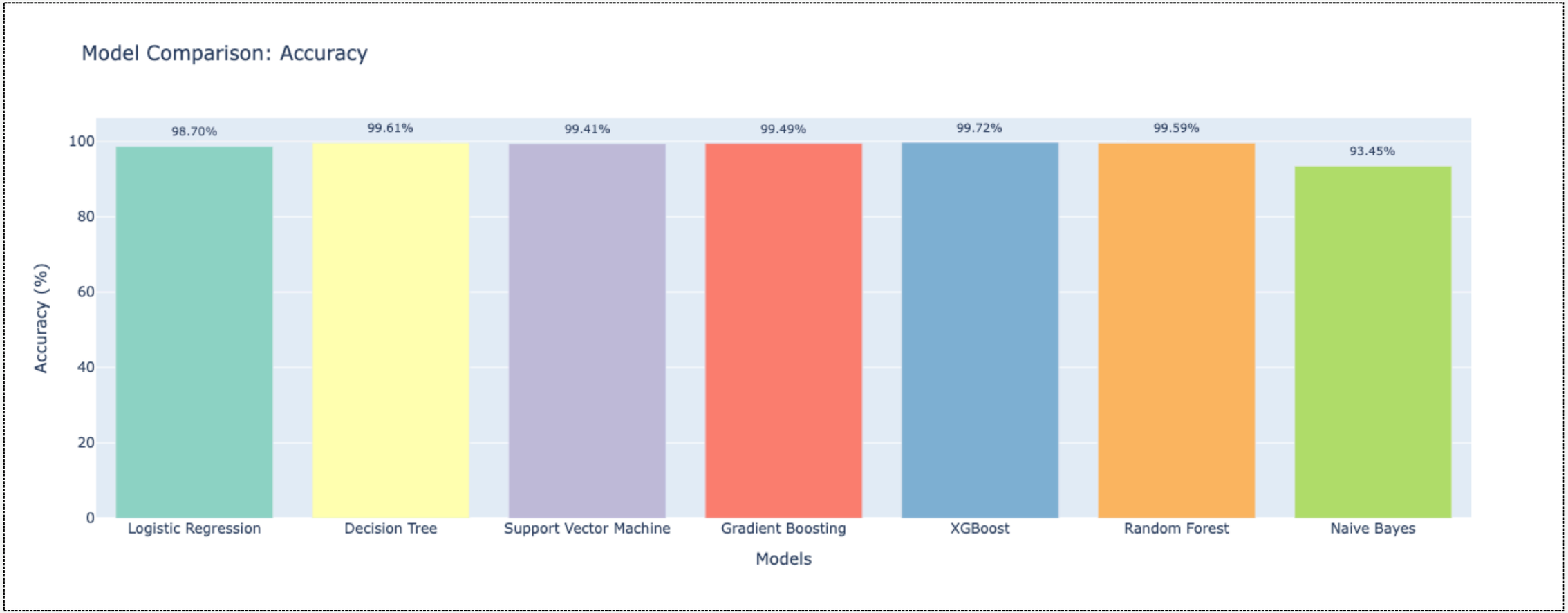


Deep Learning

Deep learning models, such as LSTM and CNN, leverage word embeddings to capture complex patterns and semantic relationships in text for fake news detection.

Train & Evaluation Models

Machine Learning



Model Testing

Machine Learning

Input

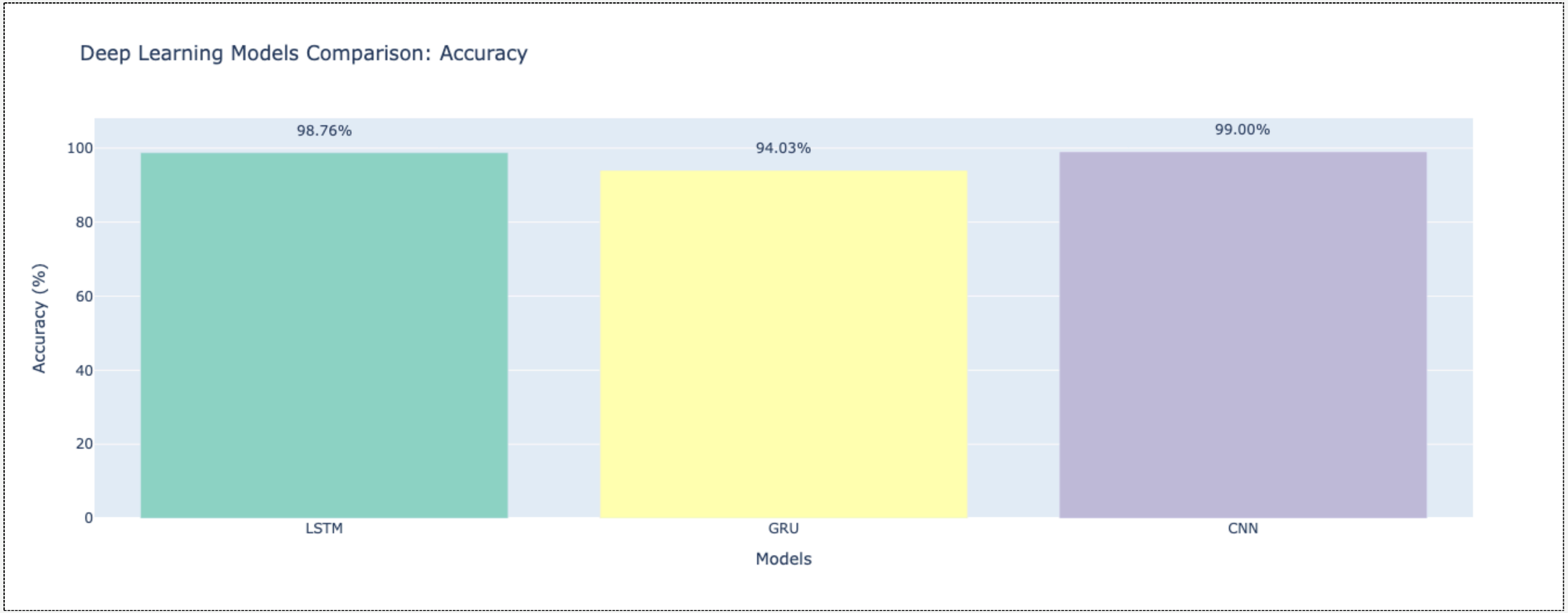
NASA discovers a new exoplanet with potential signs of life.

Output

```
Enter a news article to test if it's Fake or Not Fake:
NASA discovers a new exoplanet with potential signs of life.

Model Predictions:
Logistic Regression (LR): Fake News
Decision Tree (DT): Fake News
Support Vector Machine (SVM): Fake News
Gradient Boosting Classifier (GBC): Fake News
XGBoost (XGB): Fake News
Random Forest (RF): Fake News
Naive Bayes (NB): Fake News
```

Train & Evaluation Models

Deep Learning



Model Testing

Deep Learning

Input

```
Breaking: Famous celebrity endorses weight loss pill proven to work in 2 days.
```

Output

```
Enter a news article to test if it's Fake or Not Fake:
Breaking: Famous celebrity endorses weight loss pill proven to work in 2 days.
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 34ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 38ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 23ms/step

Deep Learning Model Predictions:
LSTM: Fake News
GRU: Fake News
CNN: Fake News
```

Topic Modeling

BERT

- Pre-trained language model considering preceding and succeeding words.
- Enables nuanced understanding of language.
- Enhances model's ability to discern subtle cues and semantic nuances.

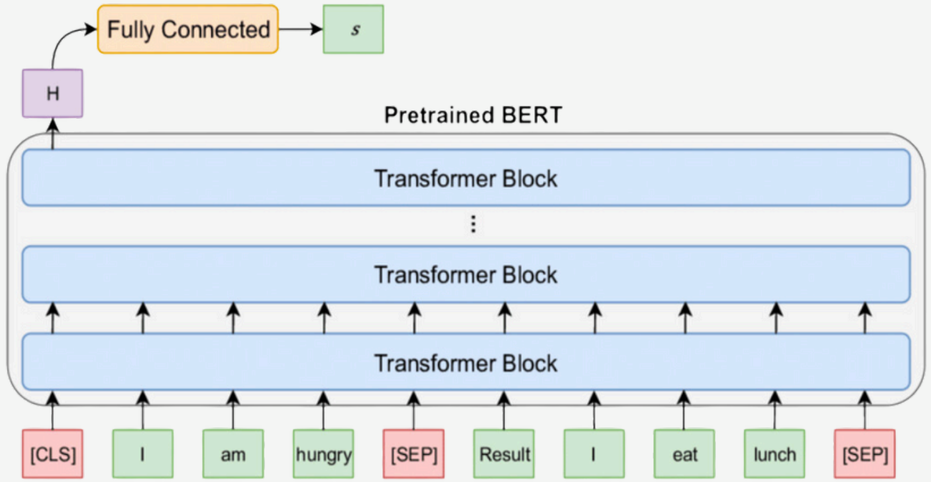
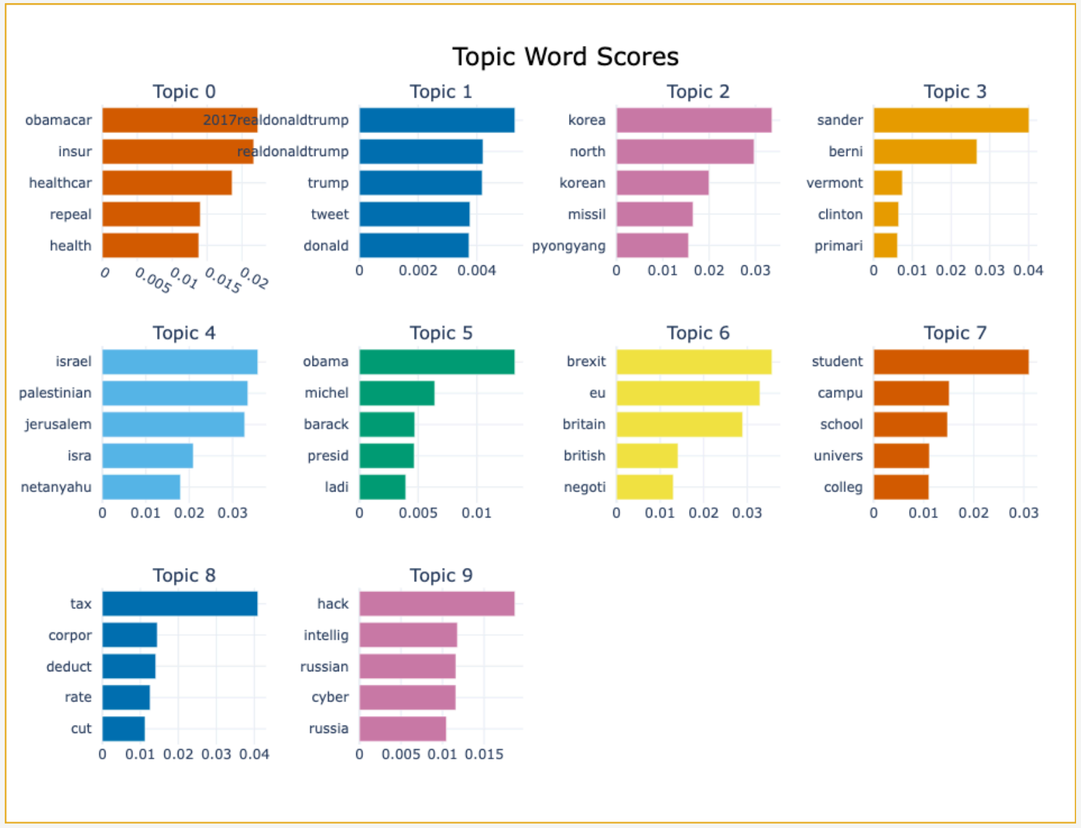
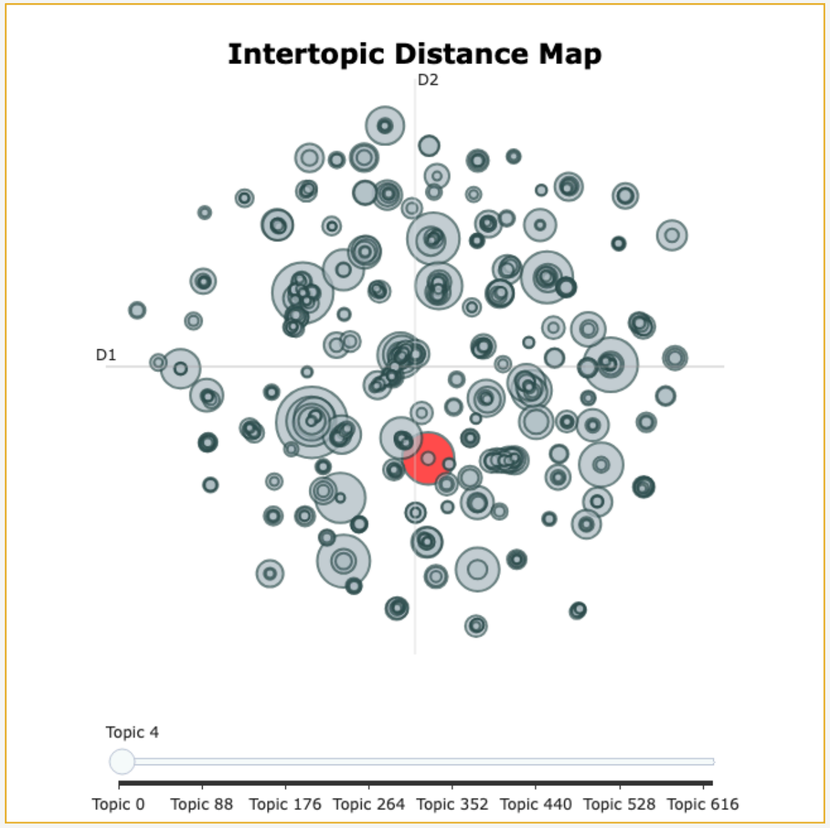


Figure 2. Architecture of BERT technique.

Topic Modeling

BERT





Conclusion

This project shows how NLP, machine learning, and deep learning can help detect fake news by finding patterns in news articles. It highlights the best models for this task and how they can be used to reduce the spread of fake news, making information more reliable.

Thanks

Do you have any questions?