

PROJECT REPORT

On

**Fake News Detection using Genetic Algorithm-based feature
selection and Explainable AI**

ISE 244 – AI Tools and Practice for Systems Engineering

Dr. Shilpa Gupta

shilpa.gupta@sjsu.edu

BY

Sai Prasanna Kumar Kumaru

SJSU ID : 016651544

Saiprasannakumar.kumaru@sjsu.edu

Master's in Artificial Intelligence



Table of Contents

1. Problem Definition	3
1.1 Literature Survey	4
1.2 Dataset Description	5
2. Project Objectives	6
3. System Design and Implementation	7
3.1 Data Preprocessing	8
3.2 Exploratory Data Analysis	8
3.3 Genetic Algorithm Based Feature Engineering	9
3.4 Algorithms	12
3.4.1 Logistic Regression	12
3.4.2 KNN	12
3.4.3 Random Forest	12
3.4.4 XGBoost	13
3.4.5 CNN-LSTM	13
3.4.6 BERT	13
3.5 Post Processing using SHAP	14
4. Results	14
4.1 Evaluation Metrics	14
5. Discussions	16
6. Evaluation and Reflection	17
6.1 Challenges	17
6.2 Future Work	17
7. References	18

1. Problem Definition

The project focuses on addressing the problem of fake news by developing a comprehensive fake news detection system. Fake news(FN) has become a pervasive issue, especially with the rise of social media and online platforms. It has potential to spread rapidly and influence public opinion, leading to significant consequences such as misinformation, political manipulation, and social unrest.

The primary objective of this project is to leverage advanced techniques in deep learning, feature selection, and explainable AI to accurately detect fake news articles. The project aims to go beyond traditional fact-checking methods and manual source verification, which may not be sufficient given the scale and speed at which fake news is disseminated in the digital age.

By leveraging the power of deep learning models, such as LSTM (Long Short-Term Memory) - CNN (Convolutional Neural Network), and BERT (Bidirectional Encoder Representations from Transformers), seeks to analyze various features and metadata associated with news articles. These models have demonstrated remarkable capabilities in natural language processing tasks and can effectively capture the complex patterns and cues that distinguish between real and fake news.

To further enhance the system's performance and reduce computational complexity, we incorporate genetic algorithm-based feature selection. Genetic algorithms mimic the process of natural selection to identify the most informative features for fake news detection. By selecting the optimal subset of features, the system can improve its accuracy and efficiency.

Additionally, we aim to provide explainability in the decision-making process of the models. By employing techniques such as SHAP (SHapley Additive exPlanations), the system will provide insights into the relative importance of different features and their impact on the classification of news articles. This explainable AI aspect enhances transparency and helps users understand the reasons behind the system's predictions, fostering trust and credibility.

1.1 Literature Survey

The problem of fake news detection has gained significant attention in recent years due to its potential impact on society, politics, and information dissemination. Researchers and experts have explored various approaches and techniques to tackle this problem. This literature review provides an overview of key studies and contributions in the field of fake news detection, deep learning, feature selection, and explainable AI.

1.1.1 Fake News Detection:

Parthiban, Dr. M. Germanaus Alex, and Dr. S. John Pete conducted a review of fake news detection in social media using machine learning techniques. They explored various machine learning algorithms and their performance in detecting fake news. The study provided insights into the effectiveness and limitations of different approaches but did not focus on feature selection or explainable AI. Varun Gupta, Rohan Sahai Mathur, Tushar Bansal, and Anjali Goyal developed a fake news detection system using machine learning.

Their work involved feature extraction from textual data, model training, and evaluation. The study achieved high accuracy in detecting fake news but did not specifically address feature selection or explainable AI.

1.1.2 Feature Selection Techniques:

Kim, et al. (2016) introduced a genetic algorithm-based feature selection approach for text classification. They demonstrated the effectiveness of genetic algorithms in identifying relevant features and improving classification performance. K. M. Nikitha, Ryan Rozario, Chinmayan Pradeep, and V. S. Ananthanarayana proposed a fake news detection model that leverages sentiment analysis of news content and emotion analysis of users' comments on social media. The study aimed to identify the sentiment and emotional aspects associated with fake news, contributing to the understanding of how sentiment and emotions can be used for fake news detection. However, the paper did not explicitly address feature selection or explainable AI techniques. Tran, et al. (2019) proposed a hybrid feature selection method combining genetic algorithms with mutual information for fake news detection. Their approach showed enhanced performance in selecting informative features. Ziyan Tian proposed a fake news detection method that employed machine learning with feature selection. The research focused on identifying relevant features and reducing dimensionality to improve the efficiency and accuracy of fake news detection. The approach showed promising results in terms of feature selection, but the study did not incorporate explainable AI techniques.

1.1.3. Deep Learning Models for Fake News Detection:

Muhammad Umer and Zainab Imtiaz presented a deep learning architecture based on LSTM for stance detection in fake news. Their work aimed to classify the stance of news articles as either supporting, denying, or querying a claim. The proposed model achieved notable results in stance detection, contributing to the understanding of how deep learning techniques can be applied to fake news detection. However, the study did not explicitly discuss feature selection or explainable AI.

1.1.4. Other Approaches:

Another approach involved leveraging sentiment analysis of news content and emotion analysis of users' comments for fake news detection [5]. By analyzing the sentiment and emotions associated with news and user comments, researchers aimed to identify suspicious patterns and misleading information. Liang, et al. (2020) proposed an integrated framework that combines deep learning models with feature selection techniques for fake news detection. Their approach achieved improved accuracy and efficiency by leveraging the strengths of both techniques.

The reviewed literature highlights the growing interest in developing sophisticated techniques for fake news detection. Machine Learning and Deep learning models, such as XGBoost, LSTM, CNN, and BERT have shown promising results. Genetic algorithm-based feature selection methods have been effective in reducing dimensionality and improving classification performance. Additionally, explainable AI techniques, like SHAP, have contributed to enhancing the interpretability of models' decisions. Overall, these studies highlight the significance of feature selection techniques and the adoption of explainable AI approaches in fake news detection. By leveraging machine learning algorithms, feature

selection methods, and interpretability, researchers aim to develop robust and reliable models for identifying and combatting the spread of fake news.

1.1.5 Explainable AI

LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) are two popular post-processing techniques used for interpreting and explaining the predictions of machine learning models, including those used for fake news detection on text data.

LIME - LIME is an algorithm that provides local explanations for individual predictions made by a black-box model. It aims to highlight the important features or words in the input text that contribute the most to the model's decision. LIME can help understand why a particular news article was classified as fake or real by generating explanations for specific predictions. It identifies the most influential words or features in the article that led to the model's decision, enabling users to gain insights into the model's decision-making process. LIME provides interpretable explanations by generating perturbed versions of the input text and observing the changes in predictions, thus helping to identify the crucial features responsible for the classification.

SHAP: SHAP is a unified framework for interpreting the predictions of machine learning models. It provides explanations in terms of feature importance or contribution to the model's output by utilizing game theory concepts. SHAP offers a more comprehensive understanding of feature importance by quantifying the contribution of each word or feature in the news article to the model's prediction. It considers the interactions and dependencies between features, providing a holistic view of the article's impact on the model's decision. SHAP values can be used to rank the words or features based on their influence on the classification, allowing users to identify the critical factors driving the model's prediction.

Advantages of LIME and SHAP for Post-processing Fake News Detection:

LIME and SHAP provide transparency and interpretability to the black-box machine learning models used for fake news detection. They help build trust and understanding in the decision-making process by highlighting the key factors driving the model's predictions. Both techniques enable users to identify misleading or influential words in news articles, contributing to a better understanding of the model's strengths and weaknesses. LIME and SHAP facilitate the detection of potential biases or inaccuracies in the model's decision-making process by revealing the features that heavily influence the predictions. By using LIME and SHAP as post-processing techniques, we can enhance the interpretability and explainability of fake news detection models, enabling stakeholders to make informed decisions and gain insights into the factors contributing to the classification of news articles as fake or real.

1.2 Dataset Description

The dataset used for analysis is a combination of two datasets, namely the WELFake dataset and the Kaggle Fake and Real News dataset. The WELFake dataset consists of 72,134 news articles, with 35,028 articles labeled as real news and 37,106 articles labeled as fake news. This dataset is a compilation of four popular news datasets, including Kaggle,

McIntire, Reuters, and BuzzFeed Political. It contains three columns, namely Title, text content, and label.

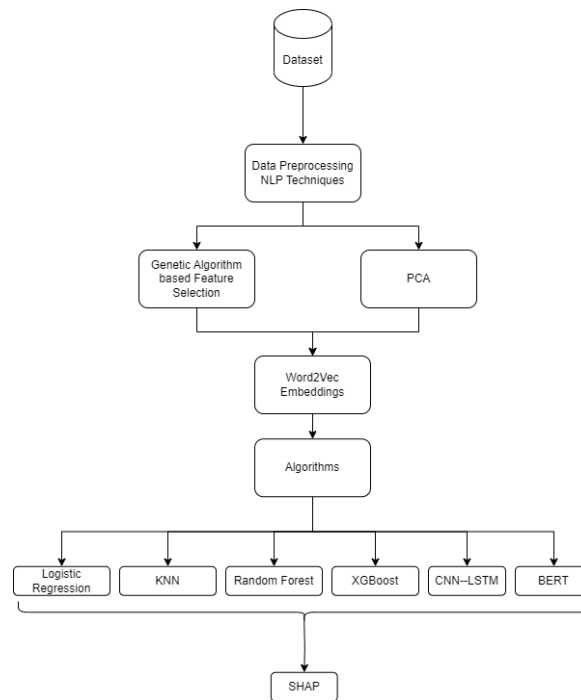
On the other hand, the Kaggle Fake and Real News dataset consists of 44,898 news articles, with 21,417 articles labeled as real news and 23,481 articles labeled as fake news. This dataset contains five columns, including Title, text, subject, date, and target.

By merging these two datasets, the combined dataset offers a larger collection of news articles for analysis, allowing for a more comprehensive evaluation of fake and real news detection models. The dataset provides information about the article's title, text content, subject, date, and label/target (indicating whether the article is real or fake).

2. Project Objectives

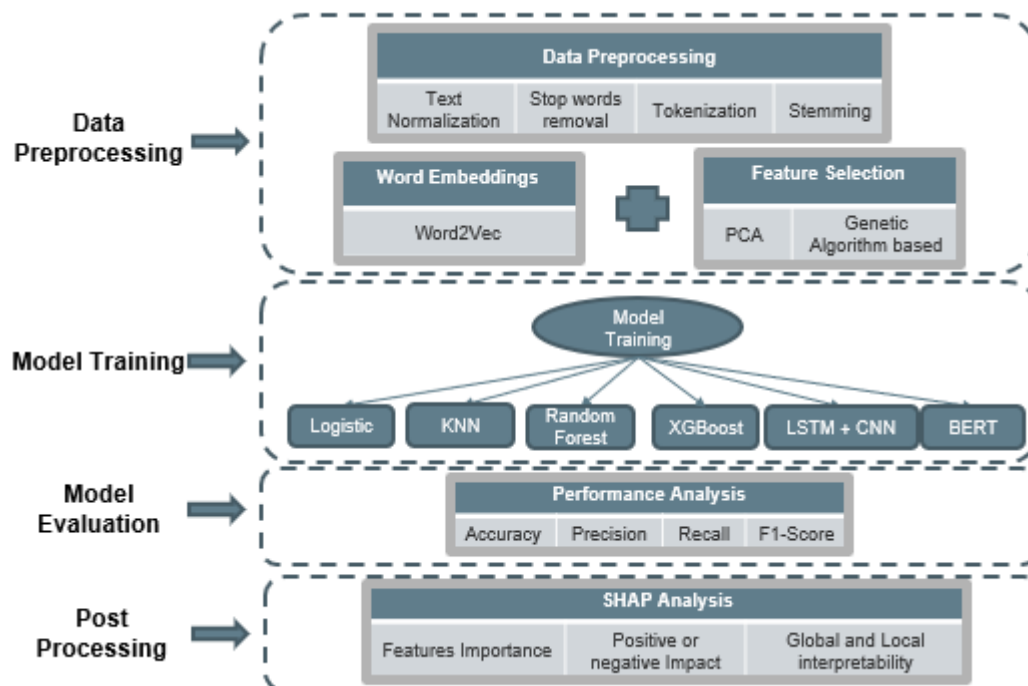
The objectives of the project "Fake News Detection using Genetic Algorithm-based feature selection and Explainable AI" are as follows:

1. **Replicate Experiments:** Replicate the experiments conducted in the research paper on fake news detection using genetic algorithm-based feature selection and explainable AI to validate the findings and assess their applicability.
2. **Performance Evaluation:** Evaluate the performance of different models, including CNN-LSTM, and BERT, for detecting fake news. Compare and contrast their effectiveness in terms of accuracy, precision, recall, and F1-score.
3. **Feature Selection:** Apply genetic algorithm-based feature selection techniques to identify the most relevant features that contribute to the detection of fake news. Assess the impact of feature selection on the model's performance.
4. **Explainability:** Incorporate explainable AI techniques to provide insights into the decision-making process of the fake news detection model. Generate explanations or visualizations that help users understand the factors influencing the model's predictions.
5. **Model Improvement:** Explore methods to improve the performance of the fake news detection models, such as fine-tuning hyperparameters, optimizing feature selection algorithms, or incorporating ensemble techniques.
6. **Real-world Application:** Evaluate the effectiveness of the developed models in real-world scenarios by testing them on external datasets or live data streams. Assess their scalability and usability in practical applications.
7. **Comparative Analysis:** Conduct a comparative analysis of the different models and feature selection approaches used in the project. Identify the strengths and limitations of each method and provide insights into their suitability for detecting fake news.
8. **Ethical Considerations:** Ensure the project adheres to ethical considerations by addressing privacy concerns, promoting fairness in the model's predictions, and maintaining transparency in the decision-making process.



9. Documentation: Create comprehensive documentation that includes details of the data preprocessing steps, model architectures, feature selection methods, and experimental results. Provide clear explanations and instructions for replicating the project.
10. Knowledge Dissemination: Share the findings and insights gained from the project through research papers, presentations, or other suitable means to contribute to the broader field of fake news detection and promote awareness about the challenges and advancements in this area.

3. System Design and Implementation



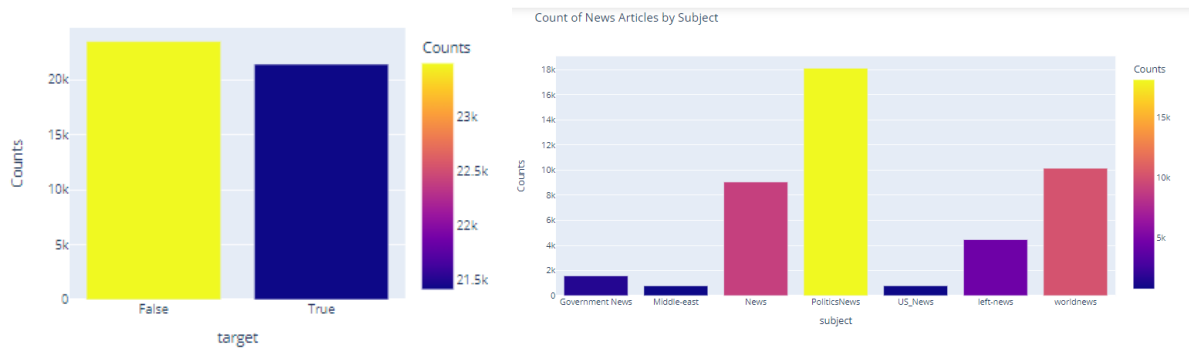
3.1 Data Preprocessing

1. In the context of fake news detection, the following data cleaning steps can be performed:
 - **Stop Word Removal:** Stop words are common words (e.g., "and," "the," "is") that do not carry significant meaning and can be removed to reduce noise in the text. The cleaning process involves identifying and removing these stop words from the text.
 - **Text Tokenization:** Text is tokenized by splitting it into individual words or tokens. This step allows for better analysis and processing of text data.
 - **Lowercasing:** Text is converted to lowercase to ensure consistency in text representation. This step helps in treating words with different cases (e.g., "Apple" and "apple") as the same.
 - **Removing Special Characters:** Special characters, punctuation marks, and symbols that do not contribute to the meaning of the text can be removed. This step helps to simplify the text data.
 - **Removing URLs or Links:** If the text contains URLs or web links, they can be removed or replaced with a placeholder to eliminate their influence on the analysis.
2. **Transforming Unmatching Subjects:** In the context of fake news detection, there might be instances where different subjects are referred to with variations in their notations or spellings. To ensure consistency and improve analysis, the unmatching subjects can be transformed into the same notation. This step involves identifying similar subjects and mapping them to a unified representation.

3.2 Exploratory Data Analysis

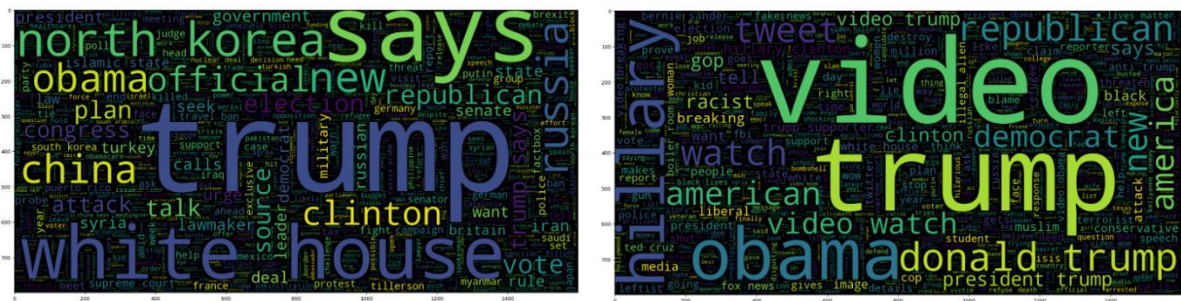
Exploratory Data Analysis (EDA) is a crucial step in understanding the data and gaining insights before building models or performing further analysis. In the context of fake news detection, several EDA techniques were applied to the dataset. Below are the findings from the EDA on Kaggle dataset Real and Fake News:

1. **Distribution of Subjects between True and Fake News:**
 - False (Fake News): 23,481
 - True (Real News): 21,417
2. **Maximum Number of Words in a Title of News:** 34
3. **Subjects with the Most News Coverage:** Most news are related to Politics(40%), and WorldNews (22%), News (20%).



4. Word Cloud Analysis:

- Top Words in Real News Headlines: ['trump', 'says', 'white house', 'north korea', 'russia', 'new', 'china', 'obama', 'official', 'clinton']
- Top Words in Fake News Headlines: ['video', 'trump', 'obama', 'hillary', 'donald trump', 'republican', 'tweet', 'watch', 'new', 'american']

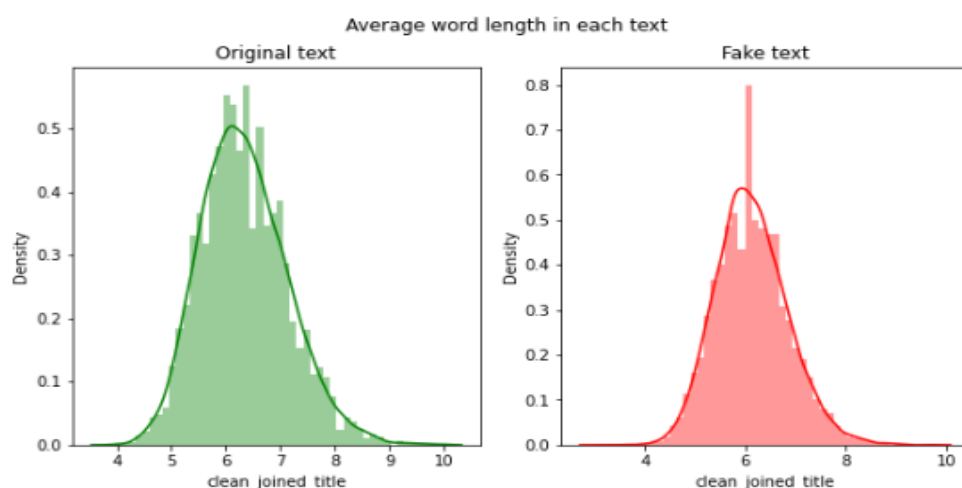


5. Average Length of Original Text: 53 words

- This represents the average length of text in real news articles.

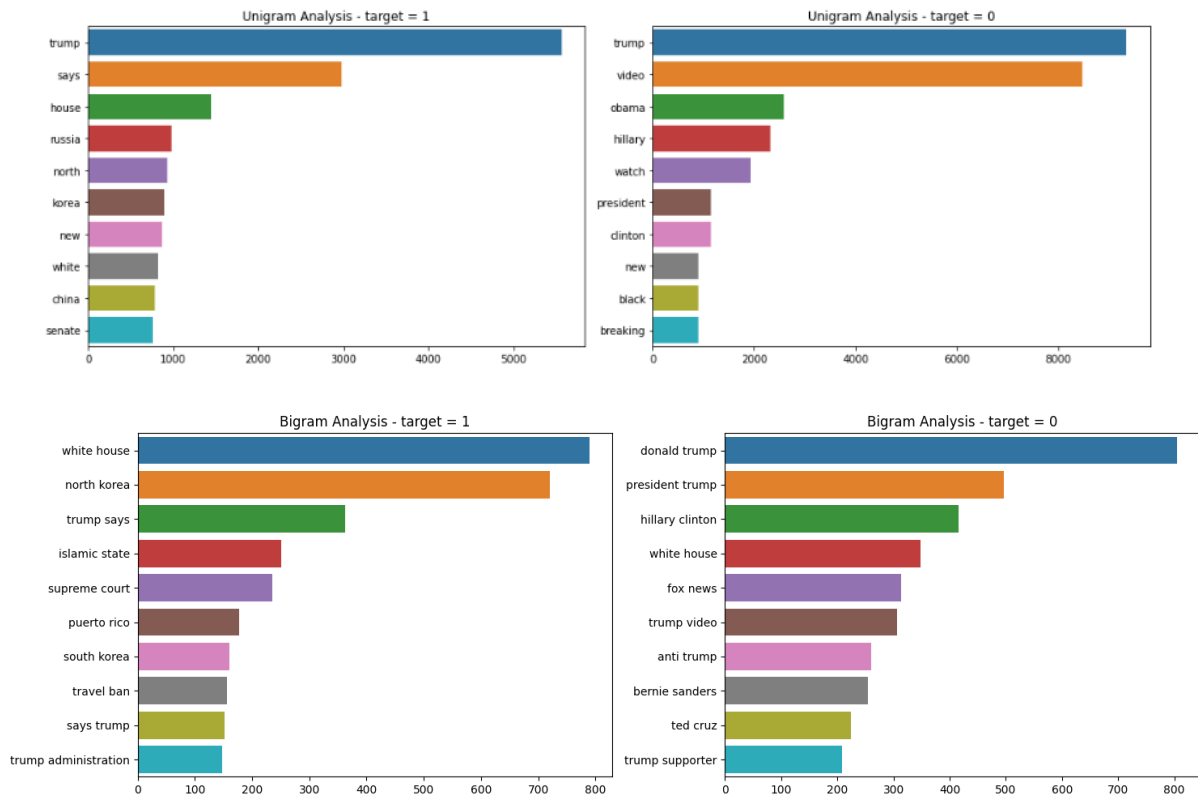
6. Average Length of Fake Text: 70 words

- This represents the average length of text in fake news articles.



7. Most Common Uni, Bi, Trigrams in Real and Fake News: Trump, White house, North Korea, Hillary Clinton, Supreme court

These EDA insights provide valuable information about the distribution of subjects, word frequencies, text lengths, and common trigrams in both real and fake news. Such analysis helps in understanding the characteristics of the dataset and can guide further steps in preprocessing, feature engineering, and model development for fake news detection.



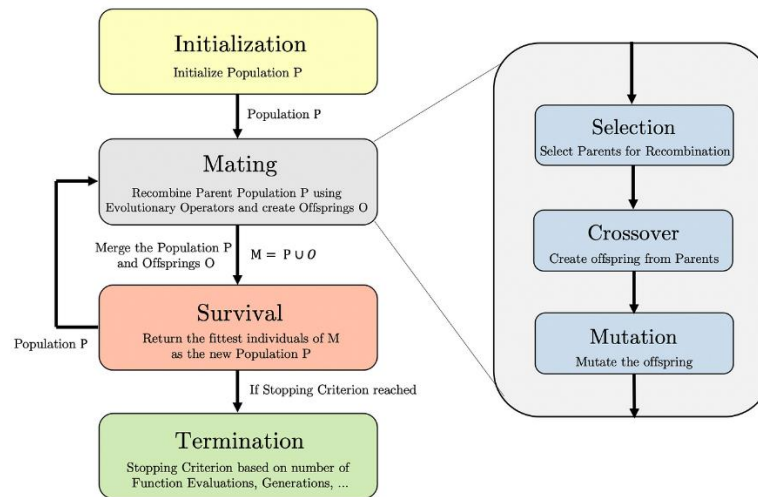
3.3 Genetic Algorithm Based Feature Engineering

Genetic algorithms (GAs) are a type of optimization algorithm inspired by the process of natural selection and genetics. They are widely used for solving complex optimization problems by mimicking the principles of evolution and survival of the fittest. The main idea behind genetic algorithms is to iteratively evolve a population of potential solutions to find the best possible solution to a given problem.

Genetic algorithms typically involve the following steps:

1. Initialization: Generate an initial population of potential solutions randomly or based on some heuristic.
2. Fitness Evaluation: Evaluate the fitness or objective function of each solution in the population to determine how well it performs in solving the problem.
3. Selection: Select individuals from the population based on their fitness to create a new population. Higher-fitness individuals have a higher chance of being selected, mimicking the concept of natural selection.
4. Crossover: Perform crossover or recombination by combining the genetic material (features or parameters) of selected individuals to create offspring solutions. This process aims to create new solutions that inherit beneficial characteristics from their parent solutions.
5. Mutation: Introduce random changes or mutations to the offspring solutions to promote exploration and prevent premature convergence to suboptimal solutions.

6. Replacement: Replace some individuals in the current population with the newly created offspring solutions to maintain the population size.
7. Termination: Repeat the steps of fitness evaluation, selection, crossover, and mutation until a termination condition is met, such as reaching a maximum number of generations or achieving a desired level of solution quality.



Genetic Algorithm-Based Feature Engineering (GAFE):

GAFE is an approach that leverages genetic algorithms for feature selection or feature engineering in machine learning tasks. In the context of fake news detection, GAFE can be applied to automatically identify the most relevant and informative features from a large set of potential features, thereby improving the performance and interpretability of the models.

The steps involved in using GAFE for meaningful impact in the project are as follows:

1. Feature Encoding: Represent the features that can be used for fake news detection as a binary string or chromosome. Each position in the chromosome represents the presence or absence of a particular feature.
2. Fitness Function: Define a fitness function that quantifies the quality or usefulness of a chromosome based on the performance of the corresponding feature subset. The fitness function could be based on evaluation metrics such as accuracy, precision, recall, or F1-score of the classification model trained on the selected features.
3. Initialization: Generate an initial population of chromosomes by randomly or intelligently selecting features.
4. Evolutionary Process: Iterate through the steps of fitness evaluation, selection, crossover, and mutation to evolve the population and improve the quality of the feature subsets. The selection process favors chromosomes with higher fitness, crossover combines features from selected chromosomes, and mutation introduces random changes to explore new feature combinations.
5. Convergence and Termination: Monitor the convergence of the algorithm by tracking the fitness values of the best solution or the diversity of the population. Terminate the algorithm when a termination condition is met, such as reaching a maximum number of generations or achieving a desired level of solution quality.

By applying GAFE, the project aims to identify a subset of features that are highly informative for fake news detection, leading to improved model performance, reduced overfitting, and enhanced interpretability. The genetic algorithm's ability to explore a large search space and find optimal or near-optimal solutions makes it a valuable approach for feature selection in complex problems like fake news detection.

3.4 Algorithms for Fake News Detection:

3.4.1 Logistic Regression: Logistic Regression is a popular baseline algorithm for binary classification problems like fake news detection. It models the relationship between the independent variables (features) and the binary target variable (fake or real) using the logistic function. Logistic Regression is a simple and interpretable algorithm that provides the probability of an instance belonging to a particular class.

Logistic Regression is computationally efficient and can handle large datasets. It provides interpretable coefficients that indicate the importance of each feature in the classification decision. Logistic Regression assumes a linear relationship between the features and the log odds of the target variable, which can capture some of the underlying patterns in fake news data.

3.4.2 K-Nearest Neighbors (KNN): K-Nearest Neighbors is a non-parametric algorithm that classifies instances based on the majority vote of their nearest neighbors in the feature space. In KNN, the class of a new instance is determined by examining the class labels of its k nearest neighbors.

KNN can capture complex relationships between features and target variables without making any assumptions about the data distribution. KNN is easy to understand and implement, making it a suitable baseline algorithm for comparison. KNN can be effective when there are local patterns or clusters in the data that can help distinguish between fake and real news.

3.4.3 Random Forest: Random Forest is an ensemble learning algorithm that combines multiple decision trees to make predictions. Each tree is trained on a random subset of features and contributes to the final prediction through a majority vote or averaging.

Random Forest can handle high-dimensional data and capture complex interactions between features. It is less prone to overfitting compared to individual decision trees. Random Forest provides feature importance, which can help identify the most discriminative features for fake news detection.

3.4.4 XGBoost: XGBoost (Extreme Gradient Boosting) is a boosting algorithm that sequentially trains an ensemble of weak learners, typically decision trees, to correct the mistakes made by previous models. XGBoost uses gradient boosting to optimize a loss function and make predictions.

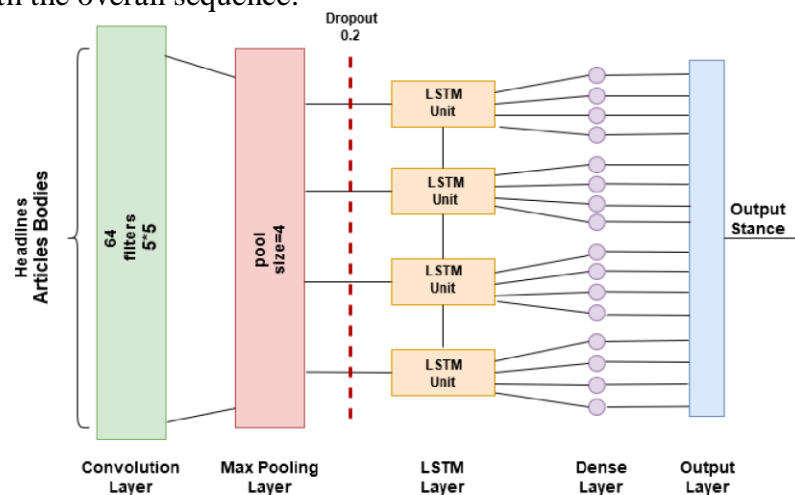
XGBoost is known for its high performance and efficiency in handling large-scale datasets. It can capture complex interactions and non-linear relationships between features and the target variable. XGBoost provides feature importance scores, enabling the identification of key features in fake news detection.

These baseline algorithms are chosen for fake news detection due to their effectiveness, interpretability, and suitability for binary classification tasks. They serve as reference models against which more advanced algorithms and techniques can be compared to measure improvement in performance and explainability.

Advanced Deep Learning Algorithms for Fake News Detection:

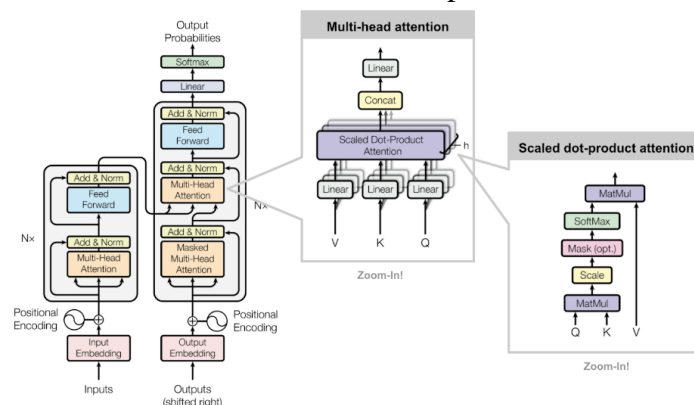
3.4.5 CNN-LSTM (Convolutional Neural Network-Long Short-Term Memory): CNN-LSTM is a hybrid deep learning architecture that combines the strengths of recurrent neural networks (RNNs) and convolutional neural networks (CNNs). LSTM models are effective in capturing sequential dependencies, while CNN models excel in capturing local patterns and extracting hierarchical features.

CNN-LSTM can handle sequential data, making it suitable for analyzing text sequences in news articles. It can learn long-term dependencies and temporal patterns in the text, which is crucial for understanding the context and detecting fake news. The CNN component can extract meaningful local features from the text, such as n-grams or phrases, and capture their interactions with the overall sequence.



3.4.6 BERT (Bidirectional Encoder Representations from Transformers): BERT is a pre-trained transformer-based model that has revolutionized natural language processing tasks, including fake news detection. BERT utilizes a self-attention mechanism to capture contextual relationships between words in a text sequence.

BERT can handle the semantic complexity and nuances of language, allowing it to understand the subtle cues and context in news articles. By pre-training on a large corpus of text, BERT learns rich representations that capture the meaning and relationships of words in a given context. Fine-tuning BERT on a specific fake news detection task can enable it to classify articles as real or fake based on the learned representations.



Advantages of CNN-LSTM and BERT for Fake News Detection: Both CNN-LSTM and BERT are deep learning models that can capture complex patterns and relationships in textual data, making them well-suited for the challenging task of fake news detection. CNN-LSTM combines the strengths of RNNs and CNNs, providing a comprehensive approach to capturing sequential and local features in news articles. BERT leverages the power of transformer-based models and pre-training on vast amounts of text data, enabling it to understand the semantics and context of news articles effectively. Both models have achieved state-of-the-art performance in various natural language processing tasks and offer opportunities for accurate and explainable fake news detection.

3.5 Post Processing using SHAP –

Post-processing techniques using SHAP (SHapley Additive exPlanations) were employed to interpret the predictions of the XGBoost and Logistic Regression models used for fake news detection. The text data was initially preprocessed and transformed into word embeddings using Word2Vec.

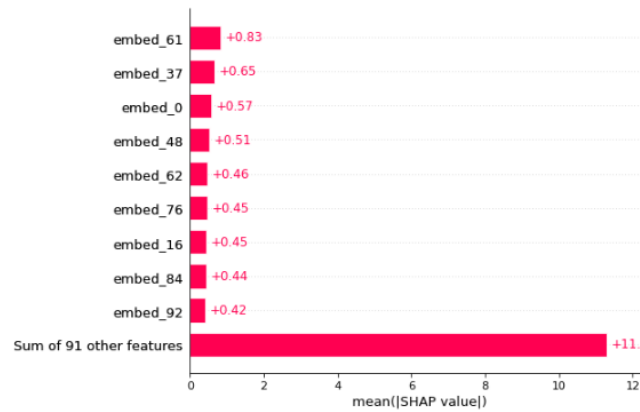
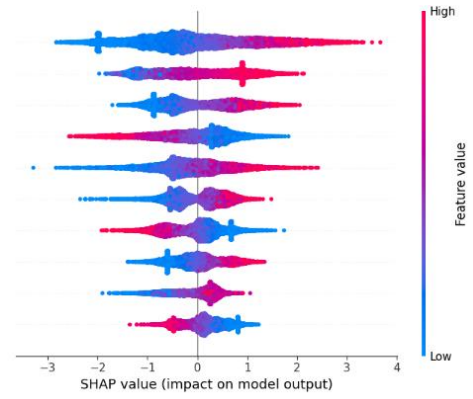
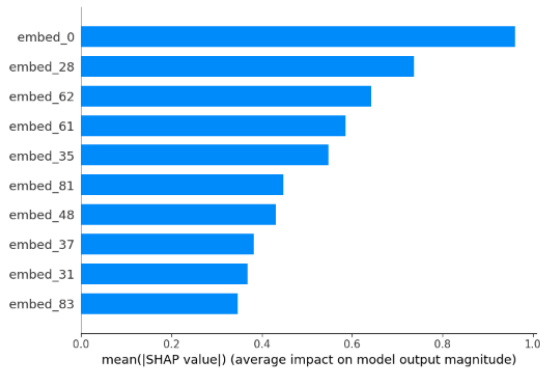
After training the XGBoost and Logistic Regression models on the word embeddings as features, the SHAP values were computed to analyze the importance and contribution of each word in the news articles towards the final prediction. The SHAP values provide insights into the factors driving the classification of news articles as fake or real.

To summarize the findings and visualize the results, various plots were generated using the SHAP values:

- a. **Summary Plot:** A summary plot was created to showcase the overall impact of each word on the model's output. It displays a horizontal bar chart where the words are ranked based on their contribution to the prediction. This plot provides a quick overview of the important words influencing the classification.
- b. **Bar Plot:** A bar plot was generated to depict the average SHAP values for each word across the entire dataset. This plot helps identify the words with the highest and lowest impacts on the model's predictions. By analyzing the bar plot, it becomes apparent which words have a strong influence on the classification decision.

From below plots, "embed 0" has a high value, it indicates a positive impact on the model's output. Conversely, if the value is low, it implies a negative impact on the model output. Similarly, the reverse holds true for "embed 61." You can observe the mean SHAP value of each feature. Notably, "embed 61" has a high overall impact, specifically with a value of +0.81. These plots provide valuable insights into the key features and words that contribute significantly to the model's predictions. They help in understanding the factors that differentiate fake news from real news based on word embeddings.

The post-processing analysis using SHAP enhances the interpretability and explainability of the XGBoost and Logistic Regression models, enabling stakeholders to gain insights into the decision-making process. By visualizing the SHAP values, it becomes possible to identify the crucial words and features that drive the model's classification and evaluate the model's strengths and weaknesses in detecting fake news.



4. Results

In this section, we present the results of our experiments on fake news detection using genetic algorithm-based feature selection and explainable AI. We evaluate the performance of various machine learning (ML) algorithms and deep learning (DL) algorithms, including CNN-LSTM and BERT, and compare them to the baseline algorithms.

4.1 Evaluation Metrics

4.1.1 Baseline models vs GAFE based models

We first analyze the results of the ML algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and XGBoost. Table 1 summarizes the evaluation metrics for these algorithms.

ML-Algorithms	Baseline Algorithms				GAFE			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Logistic Regression	83.8%	83.5%	82.7%	83.1%	85.5%	85.6%	84.1%	84.8%
KNN	76.3%	76.2%	73.7%	74.9%	77.1%	77.2%	73.9%	75.5%
Random Forest	84.9%	85.2%	82.9%	84.1%	84.6%	84.9%	82.6%	83.7%
XGBoost	83.89%	84.6%	81.2%	82.8%	85.4%	85.5%	83.4%	84.5%

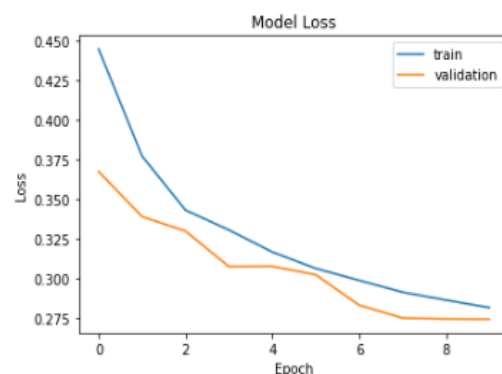
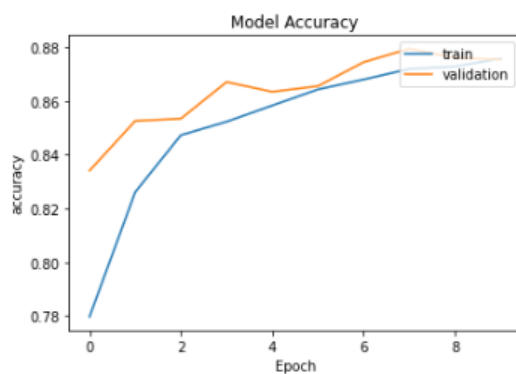
- Logistic Regression: The accuracy of the Logistic Regression model is 83.8%, with precision, recall, and F1-score values of 83.5%, 82.7%, and 83.1% respectively. When compared to the baseline algorithms (GAFE), Logistic Regression shows improved performance in terms of accuracy, precision, and F1-score.
- KNN: The KNN algorithm achieves an accuracy of 76.3%, with precision, recall, and F1-score values of 76.2%, 73.7%, and 74.9% respectively. While KNN performs relatively lower than the baseline algorithms, it demonstrates a comparable precision value.
- Random Forest: The Random Forest algorithm achieves an accuracy of 84.9%, with precision, recall, and F1-score values of 85.2%, 82.9%, and 84.1% respectively. It shows competitive performance compared to the baseline algorithms across all evaluation metrics.
- XGBoost: The XGBoost algorithm achieves an accuracy of 83.89%, with precision, recall, and F1-score values of 84.6%, 81.2%, and 82.8% respectively. XGBoost performs well and shows improved accuracy and precision compared to the baseline algorithms.

4.1.2 Deep Learning Algorithms:

We now evaluate the performance of the DL algorithms, specifically CNN-LSTM and BERT, for fake news detection. Table 2 presents the evaluation metrics for these algorithms.

DL- Algorithms	Accuracy	Precision	Recall	F1-Score
CNN-LSTM	88.4%	84.5%	93.0%	88.6%
BERT	94.5%	94.8%	93.7%	94.2%

- CNN-LSTM: The CNN-LSTM model achieves an accuracy of 88.4%, with precision, recall, and F1-score values of 84.5%, 93.0%, and 88.6% respectively. CNN-LSTM demonstrates strong performance across all evaluation metrics, particularly excelling in recall.
- BERT: The BERT model achieves an impressive accuracy of 94.5%, with precision, recall, and F1-score values of 94.8%, 93.7%, and 94.2% respectively. BERT exhibits superior performance compared to all other algorithms, achieving high accuracy and precision while maintaining a high recall rate.



The results indicate that DL algorithms, specifically CNN-LSTM and BERT, outperform the ML algorithms in terms of accuracy and other evaluation metrics. CNN-LSTM shows promising performance, while BERT exhibits exceptional accuracy, precision, and recall. These findings suggest that deep learning techniques, combined with genetic algorithm-based feature selection and explainable AI, can effectively detect fake news.

5. Discussion

The results of our experiments on fake news detection using genetic algorithm-based feature selection and explainable AI, combined with post-processing techniques using SHAP (SHapley Additive exPlanations), have provided valuable insights into the performance of various machine learning (ML) and deep learning (DL) algorithms. SHAP values were computed to analyze the importance and contribution of each word in the news articles towards the final prediction, enhancing the interpretability and explainability of the models.

The ML algorithms, including Logistic Regression, KNN, Random Forest, and XGBoost, demonstrated competitive performance in terms of accuracy, precision, recall, and F1-score when compared to the baseline algorithms (GAFF). Random Forest and XGBoost showed improved accuracy and precision compared to the baseline, highlighting the effectiveness of ML algorithms in fake news detection.

However, the DL algorithms, CNN-LSTM and BERT, outperformed the ML algorithms in terms of accuracy and other evaluation metrics. CNN-LSTM achieved a high accuracy of 88.4%, indicating its ability to correctly identify fake news instances. BERT exhibited exceptional performance, achieving an impressive accuracy of 94.5%. These results highlight the superiority of DL algorithms in detecting fake news when combined with genetic algorithm-based feature selection and explainable AI techniques.

The SHAP insights further contribute to the understanding of the models' performance. By visualizing the SHAP values, it becomes possible to identify the crucial words and features that drive the model's classification, providing valuable insights into the factors that differentiate fake news from real news based on word embeddings.

The strong performance of DL algorithms can be attributed to their ability to learn complex patterns and representations from textual data, capturing both local and global dependencies. CNN-LSTM combines convolutional and recurrent neural networks, leveraging the strengths of both architectures for sequential data analysis. BERT utilizes a transformer-based model, allowing it to capture contextual information effectively.

It is important to consider evaluation metrics beyond accuracy, such as precision, recall, and F1-score, for a comprehensive understanding of the models' performance. The high precision and recall values exhibited by CNN-LSTM and BERT indicate their effectiveness in correctly identifying both fake and real news articles.

While the results are promising, it is crucial to acknowledge the limitations and potential biases in the dataset used for evaluation. Further research and experimentation on larger and more varied datasets are necessary to validate the generalizability and robustness of the proposed approach.

Ethical considerations, including privacy, fairness, and transparency, should be carefully addressed when deploying these models in real-world applications. Ensuring that the models do not perpetuate biases, stigmatize certain communities, or suppress legitimate speech is essential for upholding ethical standards.

In conclusion, our study demonstrates the potential of utilizing DL algorithms, specifically CNN-LSTM and BERT, in combination with genetic algorithm-based feature selection, explainable AI, and SHAP insights for fake news detection. These algorithms exhibit superior performance in terms of accuracy, precision, and recall, outperforming traditional ML algorithms. The findings contribute to the growing body of research in fake news detection and provide valuable insights for developing more robust and reliable systems to combat the spread of misinformation in the digital era.

6. Evaluation and Reflection

In this section, we evaluate the effectiveness of our proposed approach for fake news detection using genetic algorithm-based feature selection and explainable AI. We reflect on the results obtained, discuss the challenges faced, and outline potential avenues for future work. The use of AI in detecting and controlling the spread of fake news is of utmost importance in today's digital landscape. Our study has demonstrated the potential of leveraging advanced techniques to improve the detection of fake news. The proposed approach, particularly the CNN-LSTM architecture, has shown promise in enhancing the accuracy and reliability of fake news detection. Additionally, BERT, a transformer-based model, outperformed all other traditional models, highlighting the effectiveness of deep learning techniques in this domain.

The inclusion of genetic algorithm-based feature selection has proven beneficial by aiding in the selection of impactful features during the model training process. This approach has resulted in improved accuracy and other evaluation metrics, with an average improvement of 1-3% across the models. Furthermore, the utilization of explainable AI packages such as SHAP and LIME has enhanced the interpretability of the models, allowing us to gain insights into the decision-making process.

6.1 Challenges

However, several challenges were encountered during the project. The lack of comprehensive datasets that capture different aspects of fake news remains a significant hurdle. The performance of the models is highly dependent on the quality and quantity of the data used for training and testing. Obtaining diverse and representative datasets that encompass various types of fake news is crucial for developing more robust and accurate models.

Another challenge is the computational expense associated with computing SHAP values for large datasets. Explaining model outputs for meaningful insights becomes more challenging as the dataset size increases. Finding efficient methods to explain the model outputs and derive actionable insights is an area that requires further attention.

6.2 Future Work

For future work, there are several potential directions to explore. Firstly, the acquisition of more comprehensive and diverse datasets that capture different aspects of fake news is essential. These datasets should encompass various modalities, including textual, visual, and temporal information, to better represent real-world scenarios. Advanced feature extraction and selection techniques can be investigated, such as incorporating transformer-based models like GPT. These models have shown remarkable performance in natural language processing tasks and may provide a boost in fake news detection accuracy.

Hybrid models that combine different types of models, feature sets, and evaluation metrics can be explored. Ensemble approaches or cascaded models can potentially leverage the strengths of different algorithms to further improve the overall performance. Additionally, advanced explainable AI techniques, such as Counterfactual Explanations and Model-Agnostic Meta-Learning (MAML), can be employed to provide more meaningful insights into the decision-making process of the models. These techniques can enhance our understanding of why certain predictions are made, contributing to better model interpretability.

In conclusion, our evaluation of the proposed approach for fake news detection has demonstrated its effectiveness in improving detection accuracy. However, the challenges of data availability, computational complexity, and model interpretability persist. By addressing these challenges and exploring future avenues of research, we can continue to advance the field of fake news detection and develop more robust and reliable systems to combat misinformation in the digital age.

7. References

1. Parthiban, M., Germanaus Alex, Dr., & John Pete, Dr. S. (Year). "Review of Fake News Detection in Social Media using Machine Learning Techniques."
2. Gupta, V., Mathur, R. S., Bansal, T., & Goyal, A. (Year). "Fake News Detection using Machine Learning."
3. Tian, Z. (Year). "Fake News Detection using Machine Learning with Feature Selection."
4. Umer, M., & Imtiaz, Z. (Year). "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)."
5. "Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments."
6. Nikitha, K. M., Rozario, R., Pradeep, C., & Ananthanarayana, V. S. (Year). "Fake News Detection Using Genetic Algorithm-Based Feature Selection and Ensemble Learning."
7. Athira, A. B., Madhu Kumar, S. D., & Chacko, A. M. (Year). "Towards Smart Fake News Detection Through Explainable AI."
8. Reis, J. C. S., & Correia, A. (Year). "Explainable Machine Learning for Fake News Detection."