FAKE NEWS DETECTION USING MACHINE LEARNING

TEAM-22

GIREESHA DHULIPUDI(21B21A44E9)

JAMI NARASIMHA SATYA SAIRAM(21B21A44C5)

ALLADA MAHESHBABU(21B21A44D9)

BANDI MOHAN SURYA(21B21A44F2)

SAKA SRI RAMA KRISHNA(22B25A4403)

ABSTRACT

Fake News has become one of the major problems in the existing society. Fake News has high potential to change opinions, facts and can be the most dangerous weapon in influencing society.

The proposed project uses NLP techniques for detecting the 'fake news', that is, misleading news stories which come from non-reputable sources. By building a model based on a K-Means clustering algorithm, the fake news can be detected. The data science community has responded by taking actions against the problem. It is impossible to determine whether the news was real or fake accurately. So, the proposed project uses the datasets that are trained using the count vectorizer method for the detection of fake news and its accuracy will be tested using machine learning algorithms.

In this research, we concentrate on how to spot fake news in internet news sources. We are dedicated in two ways. In order to determine the percentage of correct news that is phony, we will use multiple datasets of actual and fake news. We provide a thorough description of the selection, justification, and approval process as well as a few exploratory analyses on the observable evidence of etymological differences in false and legitimate news material. In order to create precise false news identifiers, we focus a lot of learning studies. Additionally, we provide close examinations of the automatic and manual evidence of bogus news. Python can be used to spot fake news posted on social media.

With the rising use of social media platforms, false news has become a severe problem in recent years. Finding fake news is a difficult problem that necessitates the use of several computer techniques, such as data mining, machine learning, and natural language processing. In this abstract, the current state of false news detection will be discussed, along with its challenges and potential solutions. Finally, it will consider how cutting-edge technology like blockchain and artificial intelligence may be used in the future to improve the efficiency and precision of fake news detection.

As a result, there is a larger than ever need for accurate and reliable techniques to distinguish fake news. The field of fake news detection has rapidly evolved as a result of researchers and engineers developing a number of techniques and tactics to identify and combat misleading information. These methods include human factchecking by educated professionals as well as sophisticated computers that use machine learning to examine and classify news content. Automated processes are also a part of them.

It is important to research and create fake news detection, but it is also a challenging and complex problem. The ability to recognise fake news requires knowledge of linguistic nuance, social and cultural contexts, and the complex network dynamics of online communication. Despite these challenges, work has been done to establish effective methods for spotting false news, and the area is still developing as new tools and technology are created.

REQUIREMENTS

Functional Requirements

- **User Input:** Users can enter a news headline or article for verification.
- **Preprocessing:** The system should clean and preprocess the input text (removing stop words, stemming, etc.).
- **Feature Extraction:** Extract meaningful features such as word frequency, sentiment analysis, and source credibility.
- **Classification Model:** A machine learning model (e.g., Naïve Bayes, Random Forest, or Deep Learning) should classify the news as real or fake.
- **Database Integration:** The system should store analyzed articles and their classification results.
- **Confidence Score:** Display a probability score indicating the likelihood of the news being fake.
- **User Feedback Mechanism:** Users should be able to provide feedback if they find a misclassification.
- **Web Interface:** A simple web-based interface to allow users to interact with the system.

Use Case Scenarios

Checking the Authenticity of a News Article

Actors: User, System

Precondition: User has a news article to verify.

Steps:

1. The user inputs a news article into the system.
2. The system processes the text and extracts relevant features.
3. The machine learning model classifies the article as real or fake.
4. The system displays the classification result with a confidence score.
5. The user can provide feedback if the classification seems incorrect.

Postcondition: The user receives a classification result and can take action accordingly.

Updating the Machine Learning Model

Actors: System Administrator

Precondition: New labeled data is available for training.

Steps:

1. The system administrator uploads new datasets.
2. The system preprocesses the new data.
3. The machine learning model is retrained with updated data.
4. The updated model is deployed to improve classification accuracy.

Postcondition: The system improves its accuracy in detecting fake news.

User Feedback Mechanism

Actors: User, System

Precondition: The user is dissatisfied with the system's classification.

Steps:

1. The user provides feedback by marking an article as incorrectly classified.
2. The system records the feedback.
3. If multiple users report the same issue, the model is flagged for retraining.

Postcondition: The system adapts based on user feedback to improve classification accuracy.

FUNCTIONS

The Fake News Detection System comprises several key functions:

1. Text Preprocessing:
    - Tokenization
    - Stop-word removal
    - Lemmatization
    - Part-of-speech tagging
2. Feature Extraction:
    - Sentiment analysis (positive, negative, neutral)
    - Source credibility scoring
    - N-gram frequency analysis
    - Readability score calculation
3. Machine Learning Classification:

- Training models using labeled datasets
- Deploying classifiers such as Naïve Bayes, Logistic Regression, or Neural Networks
- Outputting classification results with confidence scores

4. User Interface:

- Web-based interface for inputting news articles
- Displaying results in an easy-to-understand format
- Allowing user feedback submissions

5. Database Management:

- Storing analyzed news articles and their classification results
- Maintaining a record of user feedback

- Updating datasets for future training

6. System Feedback and Improvement:
   - Collecting user feedback for misclassified articles
   - Periodically retraining models with new data
   - Enhancing accuracy based on real-world use