JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FAKE NEWS DETECTION DESKTOP APPLICATION

SREEGIRI S L[1], RISHI KRISHNA SUGESH[1], VYSHNAN V CHANDRAN[1], Prof. SABIRA A S[2]

[1]UG Fellow , Department of Computer Science and Engineering ,
[2]Asst.Professor, Department of Computer Science and Engineering,
UKF College of Engineering and Technology, Parippally , Kerala

Abstract : The proliferation of misinformation has significant social and political consequences. In response, the development of automated tools for accurately classifying and distinguishing fake news from real news has become crucial. This paper presents a machine learning-based approach to identify fake news articles using a range of classification algorithms, including Logistic Regression, Decision Tree, Gradient Boosting, and Random Forest classifiers. The models were trained and evaluated on a dataset comprising labeled news articles, utilizing Natural Language Processing (NLP) techniques for feature extraction. Preliminary results demonstrate the efficacy of the proposed methods in automating the detection of fake news, providing a scalable solution to mitigate the spread of false information.

I. INTRODUCTION

The advent of digital media has facilitated the rapid dissemination of information, which, while beneficial in many respects, also poses challenges due to the spread of "fake news." This term refers to misinformation and disinformation purposely crafted to mislead readers. The consequences of fake news are wide-reaching, affecting political, social, and economic sectors. Thus, developing automated methods to detect and flag such content is imperative.

Traditional approaches to content verification have relied heavily on manual fact-checking by human experts, a method that struggles to keep pace with the volume of content generated daily. To address this gap, this study explores the application of various machine learning techniques to develop an automated fake news detection system. The system processes textual content to predict the veracity of news articles, thus supporting efforts to maintain the integrity of information on social platforms and news outlets.

II. METHODOLOGY

Data Collection and Preprocessing

The dataset used in this project was sourced from Kaggle, containing diverse news articles labeled as "fake" or "real." Preprocessing involved several steps to convert raw text into a machine-readable format:

- **Text Cleaning:** Removing special characters, URLs, and HTML tags to standardize the text.
- **Tokenization:** Breaking down text into individual words or tokens.
- **Vectorization:** Transforming text into numerical format using TF-IDF vectorization, which helps in evaluating the importance of a word in a document relative to the dataset.

Model Implementation and Training

Four machine learning models were implemented to classify the news articles:

1. **Logistic Regression:** Chosen for its efficiency in binary classification problems.
2. **Decision Tree Classifier:** Provides a model of decisions and their possible consequences.

3. **Gradient Boosting Classifier:** An ensemble technique that combines several weak learning models to create a strong predictive model.
4. **Random Forest Classifier:** A robust method that uses multiple decision trees to improve classification accuracy.

Each model was trained using the transformed feature set obtained from the TF-IDF vectorization of the preprocessed text.

Evaluation Metrics

The models were evaluated based on accuracy, precision, recall, and F1-score, calculated from the test data. These metrics help in understanding the effectiveness of each model in classifying the news accurately. Additional experiments involved manual testing where new articles were classified to observe the models' real-world applicability.

Analysis of Model Performance

The evaluation of four machine learning models—Logistic Regression (LR), Decision Tree Classifier (DT), Random Forest Classifier (RFC), and Gradient Boosting Classifier (GBC)—demonstrates each model's effectiveness in accurately classifying fake and real news articles. The performance metrics, which include precision, recall, f1-score, and support for each class, provide insights into the strengths and limitations of each model. The dataset used in the evaluation consists of 11,220 articles, divided into 5,896 real news articles (class 0) and 5,324 fake news articles (class 1).

Logistic Regression (LR)

- **Precision:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 99%
- **Recall:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 98%
- **F1-Score:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 99%

Logistic Regression shows a nearly perfect precision and recall, indicating a high level of accuracy in identifying both real and fake news. The slight difference in recall for fake news suggests a minimal number of fake articles might be misclassified as real.

Decision Tree Classifier (DT)

- **Precision:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 100%
- **Recall:**
  - Class 0 (Real): 100%
  - Class 1 (Fake): 99%
- **F1-Score:**
  - Class 0 (Real): 100%
  - Class 1 (Fake): 99%
  -

The Decision Tree Classifier performs excellently, especially in identifying real news with a 100% recall. This model is particularly effective in minimizing false negatives for real news, ensuring almost no real article is mislabeled as fake.

Random Forest Classifier (RFC)

- **Precision:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 99%
- **Recall:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 98%
- **F1-Score:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 99%

Random Forest exhibits a consistent performance across all metrics, closely matching that of Logistic Regression. Its robust nature and ensemble method contribute to its high precision and recall, making it a reliable choice for this application.

Gradient Boosting Classifier (GBC)

- **Precision:**
  - Class 0 (Real): 100%
  - Class 1 (Fake): 99%
- **Recall:**
  - Class 0 (Real): 99%
  - Class 1 (Fake): 100%
- **F1-Score:**
  - Class 0 (Real): 100%
  - Class 1 (Fake): 100%

Gradient Boosting Classifier stands out with the highest metrics, showcasing its capability to adapt and improve. It is particularly effective in identifying fake news with a perfect recall of 100%, indicating no fake news article is misidentified as real.

III.COMPARATIVE ANALYSIS

- **Performance Consistency:** All models demonstrate high consistency in their performance with a general accuracy around 99% to 100%, which reflects advanced capabilities in handling binary classification tasks like fake news detection.
- **Strengths and Weaknesses:** While all models perform well, the GBC shows a slight advantage in balancing recall and precision across classes, making it potentially the most robust model in scenarios where both types of classification errors are equally undesirable.
- **Model Choice:** The choice of model may depend on specific needs:
  - For avoiding false negatives (not missing any fake news), GBC or DT might be preferable.
  - For a balanced approach, GBC would be ideal given its symmetrical precision and recall.

IV. CONCLUSION

This section will summarize the performance comparisons among the different models and discuss the implications of deploying such a system in a real-world scenario. Future work could explore deeper linguistic and semantic analysis and the integration of multi-modal data (e.g., text and images) to further enhance the accuracy of fake news detection.We can demonstrates technological adaptability for effective implementation. Integration of machine learning and deep learning will help us for robust fake news detection. Preliminary results demonstrate the efficacy of the proposed methods in automating the detection of fake news, providing a scalable solution to mitigate the spread of false information.

V. REFERENCE

[1] C. Wu, F. Wu, Y. Huang, and X. Xie, ‘‘Personalized news recommendation: Methods and challenges,’’ ACM Trans. Inf. Syst., vol. 41, no. 1, pp. 1–50, Jan. 2023.

[2] S. Gaillard, Z. A. Oláh, S. Venmans, and M. Burke, ‘‘Countering the cognitive, linguistic, and psychological underpinnings behind suscepti_x0002_bility to fake news: A review of current literature with special focus on the role of age and digital literacy,’’ Front. Commun., vol. 6, Jul. 2021, Art. no. 661801.

[3] M. L. D. Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, and L. de Alfaro, ‘‘Automatic online fake news detection combining content and social signals,’’ in Proc. 22nd Conf. Open Innov. Assoc. (FRUCT), May 2018, pp. 272–279.

[4] E. C. Tandoc, Z. W. Lim, and R. Ling, ‘‘Defining ‘‘fake news’’: A typology of scholarly definitions,’’ Digit. Journalism, vol. 6, no. 2, pp. 137–153, Feb. 2018.

[5] A. Galli, E. Masciari, V. Moscato, and G. Sperlí, ‘‘A comprehensive benchmark for fake news detection,’’ J. Intell. Inf. Syst., vol. 59, no. 1, pp. 237–261, Aug. 2022.