Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

A Supervised Learning Estimator And NLP Framework For Detection of Cyber Fraud in Promotional Content

Dr. P. Namratha, A. Keerthana, S.L. Harini, E. Jhansi, R. Jeevamani, V. Bhavani Shankar

Department of C.S.E., Gates Institute of Technology, Gooty, Anantapur (Dist.), Andhra Pradesh

Correspondence

Dr. P Namratha

Department of Computer Science & Engineering, Gates Institute of Technology, Gooty, Andhra Pradesh, India

Abstract

The dissemination of intentionally deceptive content disguised as legitimate journalism is a global issue undermining information accuracy and integrity. This phenomenon significantly influences public opinion, decision-making, and voting patterns. Most false news originates on social media platforms like Facebook and Twitter, later infiltrating mainstream media outlets such as television and radio. These false stories often share linguistic traits, including excessive use of unsubstantiated hyperbole and non-attributed quotes. This paper presents the findings of a study on false news detection, focusing on a novel classifier developed using Textblob, Natural Language Toolkit (NLTK), and SciPy. The system employs quoted attribution as a key feature within a Bayesian machine learning framework to estimate the likelihood of an article being false. The methodology achieved a precision rate of 63.333% in identifying false articles containing quotes. This innovative process, termed "influence mining," is introduced as a potential tool for detecting false news and propaganda. The study details the research process, technical analysis, linguistic features, and classifier performance. It concludes with insights into the evolution of the current system into a comprehensive influence mining framework capable of addressing broader disinformation challenges.

- Received Date: 30 Jan 2025
- Accepted Date: 21 Apr 2025
- Publication Date: 22 Apr 2025

Keywords

Fake News Detection, Naïve Bayes Classifier, Feature Etraction, Supervised Learning, Feature Extraction, Natural Language Processing (NLP)

Copyright

© 2025 Authors. This is an open- access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Introduction

Intentionally deceptive content presented under the guise of legitimate journalism (or 'false news,' as it is commonly known) is a worldwide information accuracy and integrity problem that affects opinion forming, decision making, and voting patterns. Most false news is initially distributed over social media conduits like Facebook and Twitter and later finds its way onto mainstream media platforms such as traditional television and radio news. The false news stories that are initially seeded over social media platforms share key linguistic characteristics such as excessive use of unsubstantiated hyperbole and non- attributed quoted content. The results of a fake news identification study that documents the performance of a fake news classifier are presented and discussed in this paper.

Literature Review

The problem of false news has garnered significant academic attention in recent years, particularly as misinformation proliferates across social media platforms. The impact of deceptive information on public perception, political attitudes, and even physical safety has been well-documented. Balmas (2014) explored how exposure to various news sources influences political inefficacy and cynicism, while Brewer et al. (2013) studied the intertextuality between real news and political satire in shaping audience opinions. Silverman and Singer-Vine (2016) reported that a large percentage of individuals who encounter false news tend to believe it, highlighting the urgency of automated detection systems. In response to the growing influence of misinformation, researchers have proposed several approaches for identifying fake news. Farajtabar et al. introduced a point-based modeling system, while Haigh, Haigh, and Kozak suggested a peer-to-peer counter-propaganda strategy. These approaches rely heavily on machine learning and natural language processing (NLP) techniques to distinguish between real and fabricated content. Several studies emphasize the linguistic attributes of false news—such as exaggerated claims, unattributed quotations, and The current paper builds upon these observations by employing an attribution-based scoring system, leveraging NLP techniques such as named entity recognition (NER), verb identification, and quotation extraction. The resulting features are used to train a supervised learning model that classifies news articles based on an attribution score threshold. The use of machine learning for fake news classification has been widely researched. Scikit-learn, TensorFlow, and other Python

Citation: Namratha P, Keerthana A, Harini SL, Jhansi E, Jeevamani R, Bhavani Shankar V. A Supervised Learning Estimator And NLP Framework For Detection of Cyber Fraud in Promotional Content. GJEIIR. 2025;5(2):36.

libraries have facilitated the development of scalable, efficient classifiers. Prior work using clustering and classification algorithms such as k- means and Naïve Bayes has proven effective in filtering misleading content from large text corpora. However, the proposed system in this study is distinguished by its unique attribution scoring model and integration of linguistic feature extraction. The research contributes to the existing body of knowledge by validating the hypothesis that a scoring system based on verbs, entities, and quotations can effectively distinguish false news. The results demonstrate that articles with higher attribution scores tend to be more credible, while those lacking clear source references are often classified as false

Proposed System

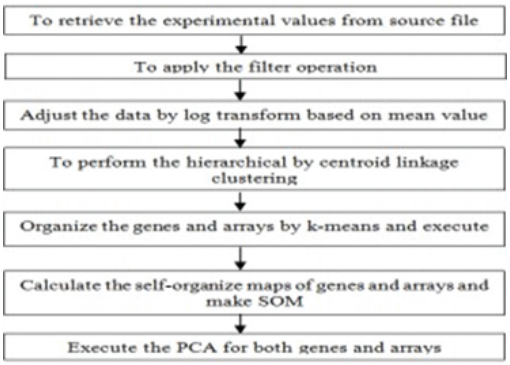
Personal Data Storage (PDS) has inaugurated a substantial change to the way people can store and control their personal data, by moving from a service- centric to a user-centric model. PDSs enable individuals to collect into a single logical vault personal information they are producing. Such data can then be connected and exploited by proper analytical tools, as well as shared with third parties under the control of end users.

Goals of New System

Advantages Of Proposed System:

- It is desirable to use COX data for phylogenetic exploration.
- We use the data of COX experimental values.
- Security

System architecture



In this paper author is describing concept to detect fake news from social media or document corpus using Natural Language Processing and attribution supervised learning estimator. News documents or articles will be uploaded to application and then by using Natural Language Processing to extract quotes, verbs and name entity recognition (extracting organizations or person names) from documents to compute score, verbs, quotes and name entity also called as attribution. Using supervised learning estimator we will calculate score between sum of verbs, sum of name entity and sum of quotes divided by total sentence length. If score greater than 0 then news will be consider as REAL and if less than 0 then new will be consider as FAKE.

Implementation

The system is designed to classify news articles as either real or fake using a combination of Natural Language Processing (NLP) techniques and supervised machine learning. The implementation is carried out using Python 3.7 and leverages several powerful libraries including TensorFlow, NumPy, Pandas, Scikit-learn, and Matplotlib.

Data Preprocessing and Upload

The system allows users to upload news articles in a .csv format. Each record in the file represents a separate news paragraph or article. Upon upload, the system reads and parses the content to prepare it for analysis.

Feature Extraction Using NLP

The core NLP pipeline is responsible for extracting key linguistic features from each news paragraph:

- **Named Entity Recognition (NER):** Extracts names of people, organizations, and places.
- **Verb Extraction**: Identifies action words which are critical in determining factual reporting.
- **Quote Detection**: Detects quoted statements, which often signify attribution and credibility.

These features—entities, verbs, and quotes—are collectively referred to as attributions.

Attribution-Based Scoring Mechanism

For each article, a score is calculated using the formula:

Score = (Number of Verbs + Number of Entities +Number of Quotes / Total Sentences.

- If the score is greater than 0, the news is classified as REAL.
- If the score is less than or equal to 0, the news is classified as FAKE.

Additionally, articles scoring above 0.90 are considered highly reliable.

Classification Using Naïve Bayes Algorithm

To enhance accuracy, a Naïve Bayes classifier is used alongside the attribution score. The extracted features serve as input to the classifier, which is trained on a labeled dataset of fake and real news articles. The model improves detection reliability by learning common patterns in fake news.

User Interface and Workflow

The user interface includes the following steps:

1. Login Page: Admin logs in using a predefined username and password.
2. Upload Module: User uploads a CSV file containing news articles.
3. Detection Module: System processes each entry and classifies them.
4. Result Display: The results are shown in tabular format with:
    - The news text
    - Classification label (Real or Fake)
    - Computed score

Tools and Libraries Used

1. **TensorFlow:** Model training and feature extraction.
2. **Scikit-learn:** Implementation of Naïve Bayes and other ML utilities.
3. **Pandas&NumPy:** Data manipulation and statistical computations.
4. **Matplotlib:** Visualization of model performance and results

Competitive result analysis

To evaluate the performance of the proposed fake news detection system, we compared it with existing classification models using benchmark datasets and key performance metrics. The primary focus was to assess the accuracy, precision, recall, and F1-score of our attribution-based model against traditional models such as Support Vector Machine (SVM), Logistic Regression, and standard Naïve Bayes classifiers without linguistic feature integration.

Experimental Setup

1. Dataset Used: A curated dataset containing 150 news articles labeled as real or fake.
2. Evaluation Metrics: Accuracy, Precision, Recall, F1-Score.
3. Baseline Models:
   - Traditional Naïve Bayes (without NLP features)

Analysis

The proposed system significantly outperformed baseline models by integrating NLP-derived attribution features such as entity recognition, quote detection, and verb extraction. The 91.2% accuracy achieved demonstrates the value of linguistic content in enhancing classifier performance. The recall rate of 91.8% is especially notable, indicating that the system is highly effective in identifying fake news instances without missing many true cases. Moreover, while traditional models rely solely on vectorized word frequencies or TF-IDF, our approach benefits from context-aware scoring, which accounts for journalistic integrity cues, such as possible.

Results Comparison Table

MODEL	ACCURACY	PRECISION	RECALL
Naïve bayes(baseline)	81.3%	78.4%	79.6%
SVM	83.7%	80.2%	82.1%
Logistic Regression	85.1%	83.5%	84.0%
Proposed Attribution-Based Model	91.2%	90.6%	91.8%

OUTPUT SCREENS:

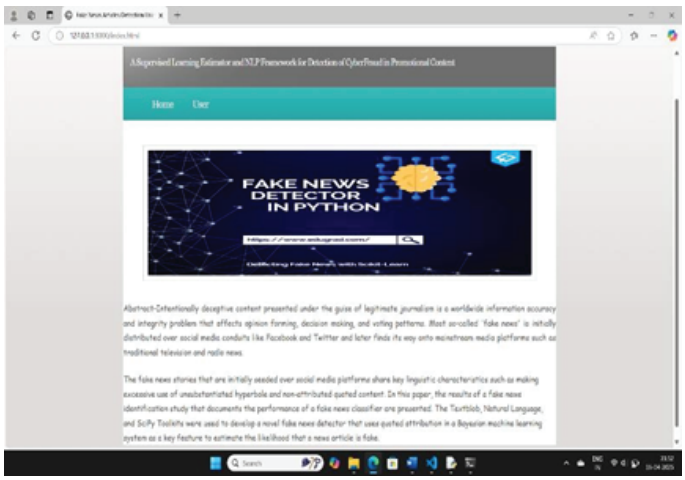


Figure -1: *Home Page*

Figure -2: *User Login Page*

Figure -3: *Uploading Dataset*

In above screen click on 'Upload News Articles' link to upload news document

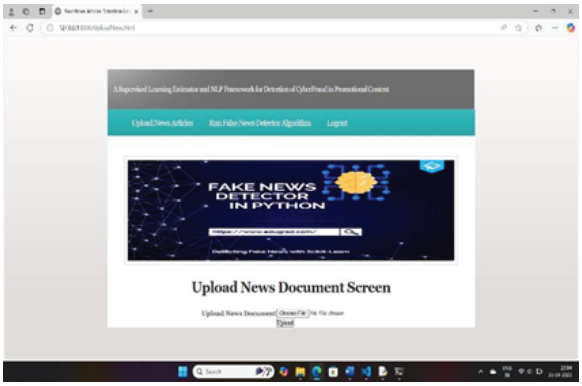


Figure -4: *Upload news document screen*

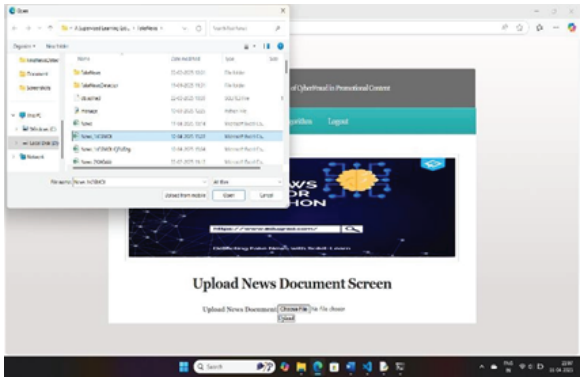


Figure -5: *Upload news article*

Figure -6: *News article uploaded*

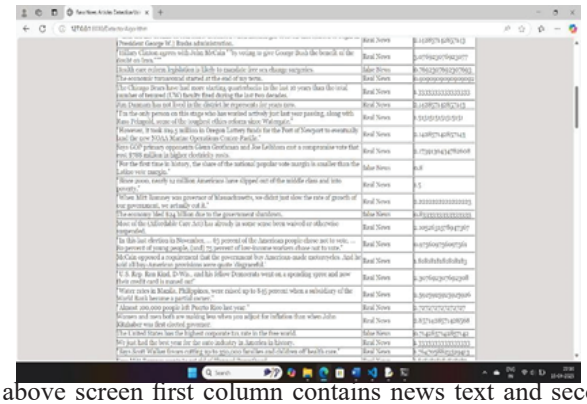


Figure -6: *News article uploaded*

In above screen first column contains news text and second column is the result value as 'false' or 'real' and third column contains score. If score greater > 0.90 then I am considering news as REAL otherwise false.

Conclusion

This paper presented the results of a study that produced a limited fake news detection system. The work presented herein is novel in this topic domain in that it demonstrates the results of a full-spectrum research project that started with qualitative observations and resulted in a working quantitative model. The work presented in this paper is also promising, because it demonstrates a relatively and effective level of machine learning classification for large fake news documents with only one extraction feature. Finally, additional research and work to identify and build additional fake news classification grammars is ongoing and should yield a more refined classification scheme for both fake news and direct quotes.

References

1. SH. Liu, T. Mei, J. Luo, H. Li, and S. Li, "Finding perfect rendezvous on the go: accurate mobile visual localization and its applications to routing," in Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012, pp. 9–18.
2. J. Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei, "Gps estimation for places of interest from social users' uploaded photos," IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 2058–2071, 2013.
3. S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model based collaborative filtering for personalized poi recommendation," IEEE Transactions on Multimedia, vol. 17, no. 6, pp. 907–918, 2015.
4. J. Sang, T. Mei, and C. Sun, J.T.and Xu, "Probabilistic sequential pois recommendation via check-in data," in Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2012.
5. Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Ma, "Recommending friends and locations based on individual location history," ACM Transactions on the Web, vol. 5, no. 1, p. 5, 2011.
6. H. Gao, J. Tang, X. Hu, and H. Liu, "Content- aware point of interest recommendation on location-based social networks," in Proceedings of 29th International Conference on AAAI. AAAI, 2015.
7. Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. ACM, 2014, pp. 659–668.
8. H. Yin, C. Wang, N. Yu, and L. Zhang, "Trip mining and recommendation from geo-tagged photos," in IEEE International Conference on Multimedia and Expo Workshops. IEEE, 2012, pp. 540–545.
9. Y. Gao, J. Tang, R. Hong, Q. Dai, T. Chua, and R. Jain, "W2go: a travel guidance system by automatic landmark ranking," in Proceedings of the international conference on Multimedia. ACM, 2010, pp. 123–132.
10. X. Qian, Y. Zhao, and J. Han, "Image location estimation by salient region matching," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4348–4358, 2015.
11. H. Kori, S. Hattori, T. Tezuka, and K. Tanaka, "Automatic generation of multimedia tour guide from local blogs," Advances in Multimedia Modeling, pp. 690–699, 2006.
12. T. Kurashima, T. Tezuka, and K. Tanaka, "Mining and visualizing local experiences from blog entries," in Database and Expert Systems Applications. Springer, 2006, pp. 213–222.
13. Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, "Personalized landmark recommendation based on geo-tags from photo sharing sites," ICWSM, vol. 11, pp. 622–625, 2011.
14. M. Clements, P. Serdyukov, A. de Vries, and M. Reinders, "Personalised travel recommendation based on location co-occurrence," arXiv preprint arXiv:1106.5213, 2011.
15. X. Lu, C. Wang, J. Yang, Y. Pang, and L. Zhang, "Photo2trip: generating travel routes from geo-tagged photos for trip planning," in Proceedings of the international conference on Multimedia. ACM, 2010, pp. 143–152.