IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TRUTHGUARD – NLP-POWERED SENTINEL AGAINST FAKE NEWS IN SOCIAL MEDIA

[1]Vijay Sree Dhulipalla, [2]K Saketh reddy, [3]KANTHETI RAJU MITHRA, [3]Patchala Sainath, [3]Gunda Sanjay

[1]Assistant Professor, Department of CSE – Artificial Intelligence, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

[2]Co-Founder, Codegnan, Vijayawada, Andhra Pradesh, India

[3]Student, Department of CSE – Artificial Intelligence, KKR & KSR Institute of Technology and Sciences, Guntur,
Andhra Pradesh, India

Abstract: Detection of fake news on social media is a significant concern due to potential societal and national consequences. This paper introduces TruthGuard, a groundbreaking solution utilizing Natural Language Processing (NLP) to combat the rampant spread of fake news on social media platforms. The main goal is to identify the most effective machine learning model for accurately classifying news articles and social media posts as true or false. The methodology involves Python's scikit-learn, focusing on feature extraction and vectorization with tools like Count Vectorizer and Tfidf Vectorizer. The proposed system creates a special webpage and a Chrome extension for browsing with smart tools, integrating APIs for real-time fake news detection.

***Index Terms* -** Fake news, Social media, Natural Language Processing (NLP), Machine learning, Count Vectorizer, Tfidf Vectorizer, TruthGuard, Chrome extension, API integration, Real-time detection.

I. INTRODUCTION

The introduction of TruthGuard serves as a trailblazing countermeasure against the spread of false information on social media. The internet's propensity for distributing misleading information has resulted in widespread disinformation, the swaying of public opinion, and possibly even national repercussions. The study emphasizes the difficulty presented by false information and stresses the significance of fact-checking in order to solve this problem.

Social media's facilitation of information accessibility has enabled users to participate in the flow of information as well as absorb it. But this ease of access has also led to a lack of confidence in the accuracy of the information. The emergence of fake news—described as purposefully and independently verified incorrect information—poses a serious risk to free speech, media, and democracy.

It has been demonstrated that being exposed to fake news has negative consequences, such as making people feel incompetent, alienated, and cynical about communities and political candidates. Consistent misinformation efforts have even been linked to encouraging acts of violence, as the genocide that occurred in Myanmar between 2016 and 2017. Furthermore, erroneous information about COVID-19 and 5G networks has caused violent attacks and changes in the financialmarkets.

Conventional techniques for classifying fake news depend on journalists and subject matter experts to manually cross-reference statements with known facts and reliable sources. However, manual procedures are unfeasible because to the amount and pace of information on social media networks. In order to automatically classify and detect fake news, current attempts have concentrated on utilizing machine learning and natural language processing techniques. By utilizing breakthroughs Technologies NLP, Machine Learning

II. LITERATURE REVIEW

[1] The literature review explores various research efforts in fake news detection. Notable work by Noshin Nirvana Prachi et al. from North South University in Bangladesh focuses on machine learning and NLP algorithms, achieving high accuracy using techniques like logistic regression, decision trees, naive Bayes, LSTM, and BERT.

[2] Kasra Majbouri et al. propose a K-Means clustering approach for fake news detection, achieving approximately 87% accuracy.

[3] The literature also discusses the expanding role of IT departments in combating fake news, addressing risks, and proposing solutions that incorporate syntactic analysis, word count, punctuation, and bounce rates.

III. MODEL

The main goal of the proposed model is to identify the most effective machine learning model for classifying news articles and social media posts as true or false. The methodology involves using Python's scikit-learn, focusing on feature extraction and vectorization with tools like Count Vectorizer and Tfidf Vectorizer. The model experiments with various feature selection methods to enhance precision in classification. such as SVM, LSTM, Random Forest….

SUPPORT VECTOR MACHINE (SVM)

One well-liked machine learning approach for identifying fake news is called Support Vector Machine (SVM). The way SVM operates is by identifying the ideal hyperplane for classifying data points into distinct groups. When it comes to the detection of fake news, SVM analyzes a variety of textual properties, including word frequencies, sentence structures, and semantic clues, to learn how to differentiate between real and fake news pieces. In order to improve its generalization to new data, support vector machines (SVMs) strive to maximize the margin that separates the hyperplane from the nearest data points. SVM is effective at classifying news articles according to their legitimacy, with a 92% accuracy rate.

LSTM

Recurrent neural networks (RNNs) of the Long Short-Term Memory (LSTM) type are frequently used for natural language processing applications, such as the identification of false news. Text data that is sequential in nature is ideally processed using LSTM since it is well-suited for evaluating sequential data. Because LSTM can maintain long-term dependencies in data, unlike standard feedforward neural networks, it can recognize subtle patterns and correlations in text. LSTM models achieve exceptional performance in classification tasks by utilizing contextual information included in the text to identify bogus news with an astonishing 93% accuracy.

Fig a : LSTM Classifier

Random Forest Classifier

As part of its ensemble learning process, Random Forest builds a large number of decision trees during training and outputs the mean prediction (regression) or the mode of the classes (classification) for each tree. To encourage variation among the trees, each decision tree is built using a subset of the training data and a subset of the features. When it comes to detecting false news, Random Forest is particularly good at identifying intricate relationships between various textual elements that are extracted, which helps it distinguish between real and fraudulent news pieces. Random Forest is an effective machine learning model for tasks involving the identification of fake news, as seen by its strong performance in classifying news articles based on their authenticity, with an accuracy of 93.8%.

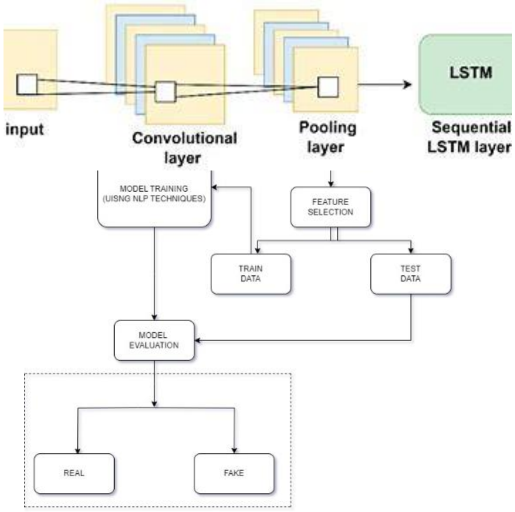


Fig b : Random Forest Classifier

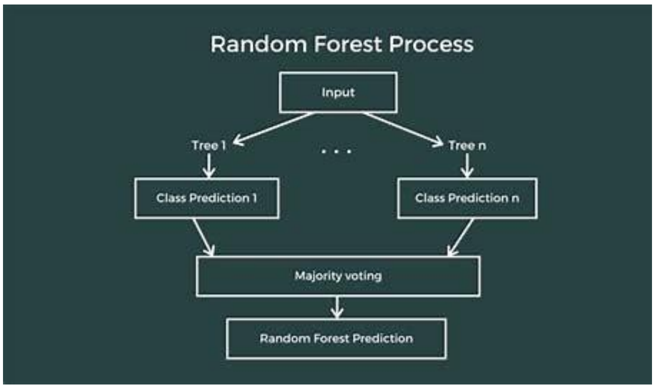


Fig c : Process Flow Diagram

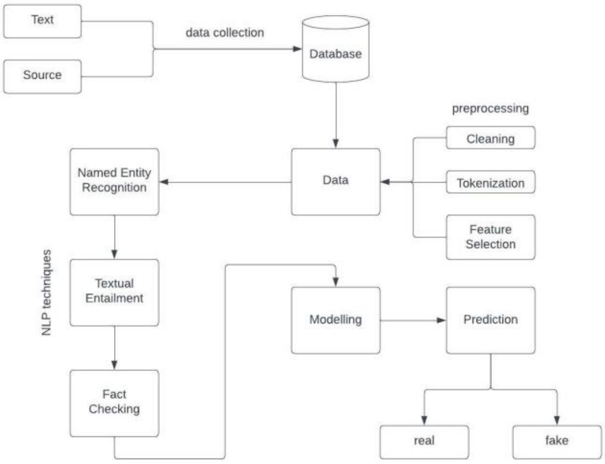


Fig d : Data Flow Diagram

IV. IMPLEMENTATION

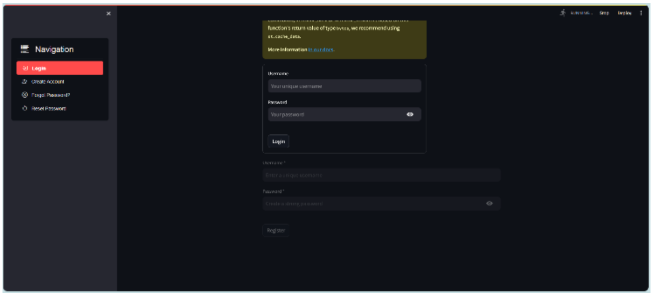
Data Gathering and Cleaning:

Collect a dataset containing labeled news articles, indicating whether each article is genuine or fake.
Clean the data by removing unnecessary elements like HTML tags, punctuation, and stopwords.
Break down the text into smaller components for analysis.

Feature Extraction:

Extract important features from the text, such as word frequencies or patterns.
Convert the text data into numerical form using techniques like TF-IDF or word embeddings.
These numerical representations will serve as input for the machine learning models.

Model Selection and Training:



function's return value of type bytes, we recommend using st.cache_data.

More information in our docs.

Username

Your unique username

Password

Your password

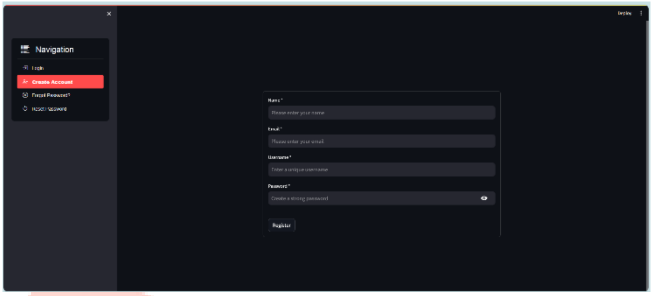
Username *

Enter a unique username

Password *

Create a strong password

Choose suitable models like Support Vector Machines (SVM), Random Forest, or LSTM neural networks for fake news detection.
Split your dataset into training and testing sets to evaluate model performance.
Train your chosen models using the training data, adjusting parameters to improve accuracy.



Name *

Please enter your name

Email *

Please enter your email

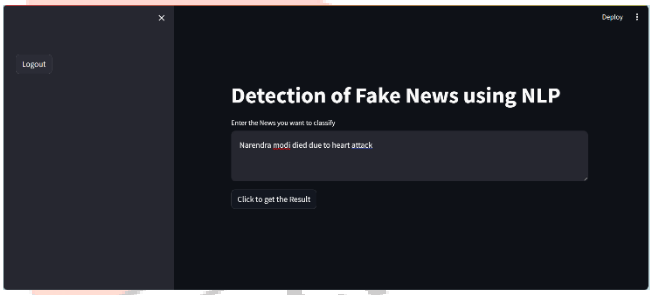
Username *

Enter a unique username

Password *

Create a strong password

V. RESULT



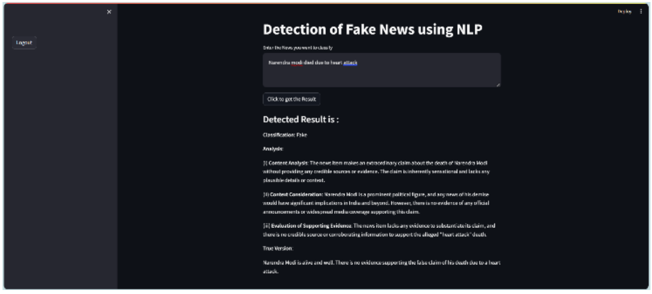
Detection of Fake News using NLP

Enter the News you want to classify

Narendra modi died due to heart attack

Fig e : Create an account
Fig f : Logged into the registered account
Fig g : Main webapp , Enter the text to detect

Fig h : Final Output screen



Detection of Fake News using NLP

Enter the News you want to classify

Narendra modi died due to heart attack

Detected Result is :

Classification: Fake

Analysis:

(i) Content Analysis: The news item makes an extraordinary claim about the death of Narendra Modi without providing any credible sources or evidence. The claim is inherently sensational and lacks any plausible details or context.

(ii) Context Consideration: Narendra Modi is a prominent political figure, and any news of his demise would have significant implications in India and beyond. However, there is no evidence of any official announcements or widespread media coverage supporting this claim.

(iii) Evaluation of Supporting Evidence: The news item lacks any evidence to substantiate its claim, and there is no credible source or corroborating information to support the alleged "heart attack" death.

True Version:

Narendra Modi is alive and well. There is no evidence supporting the false claim of his death due to a heart attack.

VI. PERFORMANCE EVALUATION

The performance evaluation of the model involves assessing accuracy, precision, recall, F-1 score, and ROC curve. The paper discusses the results obtained from experiments with different machine learning and NLP algorithms, highlighting the strengths and weaknesses of each approach.

VII. CONCLUSION & FUTURE SCOPE

In conclusion, TruthGuard presents a promising solution to combat fake news on social media platforms. The integration of advanced NLP techniques and real-time detection through API integration demonstrates the system's potential. Future scope includes further refinement of the model, expanding the dataset, and exploring additional features for improved accuracy. The proposed system also aims to integrate with social media platforms for widespread implementation.

VIII. REFERENCES

Liu, C., Wu, X., Yu, M., Li, G., Jiang, J., Huang, W., Lu, X.: A two-stage model based on BERT for short fake news detection. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 11776 LNAI, pp. 172–183 (2019). https://doi.org/10.1007/978-3-030-29563-9_17

Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. (2020). https://doi.org/10.1145/3395046

Horne, B.D., NØrregaard, J., Adali, S.: Robust fake news detection over time and attack. ACM Trans. Intell. Syst. Technol. (2019). https://doi.org/10.1145/3363818

Shu, K., Wang, S., Liu, H.: Beyond news contents: The role of social context for fake news detection. Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., Choi, Y.: Defending against neural fake news. Neurips (2020)

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., Liu, H.: Unsupervised fake news detection on social media: a generative approach. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 5644–5651 (2019)

Mohammadrezaei, M., Shiri, M.E., Rahmani, A.M.: Identifying fake accounts on social networks based on graph analysis and classification algorithms. Secur. Commun. Netw. (2018). https://doi.org/10.1155/2018/5923156 In: WSDM 2019—Proceedings of 12th ACM International Conference on Web Search Data Mining, vol. 9, pp. 312–320 (2019). https://doi.org/10.1145/3289600.3290994

Liu, Y., Wu, Y.F.B.: FNED: a deep network for fake news early detection on social media. ACM Trans. Inf. Syst. (2020). https://doi.org/10.1145/3386253

Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science 359, 1146–1151 (2018)

Jwa, H., Oh, D., Park, K., Kang, J.M., Lim, H.: exBAKE: automatic fake news detection model based on Bidirectional Encoder Representations from Transformers (BERT). Appl. Sci. 9, 4062 (2019). https://doi.org/10.3390/app9194062

Popat, K., Mukherjee, S., Yates, A., Weikum, G.: Declare: debunking fake news and false claims using evidence-aware deep learning. arXiv Preprint. http://arxiv.org/abs/1809.06416. (2018)

Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., Gao, J.: Weak supervision for fake news detection via reinforcement learning. In: AAAI 2020—34th AAAI Conference on Artificial Intelligence, pp. 516–523 (2020)

Hoens, T.R., Polikar, R., Chawla, N.: V: Learning from streaming data with concept drift and imbalance: an overview. Prog. Artif. Intell. 1, 89–101 (2012)

Kaliyar, R.K., Goswami, A., Narang, P., Sinha, S.: FNDNet—a deep convolutional neural network for fake news detection. Cogn. Syst. Res. 61, 32–44 (2020). https://doi.org/10.1016/j.cogsys.2019.12.005

Nguyen, V.H., Sugiyama, K., Nakov, P., Kan, M.Y.: FANG: leveraging social context for fake news detection using graph representation. Int. Conf. Inf. Knowl. Manag. Proc. (2020). https://doi.org/10.1145/3340531.3412046

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv (2019). https://doi.org/10.18653/v1/2020.acl-main.703

Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint. http://arxiv.org/abs/1810.04805. (2018)

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog. 1, 9 (2019)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)

Xian, Y., Akata, Z., Sharma, G., Nguyen, Q., Hein, M., Schiele, B.: Latent embeddings for zero-shot classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 69–77 (2016)