Fact Finder

Saurabh Chaudhary, Kavita Kamtekar, Suraj Salunke, Sai Tata, Manthan Kale,

Maheswari Vidyadharani , Gayatri Kulkarni, Rushabh Shingala

Seidenberg School of Computer Science and Information Systems

Pace University, New York, NY, USA

Abstract— **Fact Finder is a web app designed to combat online misinformation by differentiating between fake and real news. Additionally, it can create a summary of the news. For anyone concerned about online misinformation and facing time constraints for reading news articles. Who uploaded the news article the fact finder web app is a web app that uses machine learning and deep learning algorithms that can detect misinformation unlike believing any information on the internet. Our application verifies the authenticity of information and saves time by creating summaries of articles."Fact Finder" can provide users with reliable, accurate, and up-to-date information on various topics. It aims to combat misinformation and fake news by presenting well-researched facts, data, and evidence. The website can cover a wide range of subjects, including current events, science, history, health, and more, catering to users who seek trustworthy information for personal or professional purposes. Additionally, it offers tools or resources for users to verify the accuracy of claims they encounter online.**

***Keywords*— Machine Learning, TensorFlow, News Articles, Logistic Regression Model Firebase, Machine Learning, Fake News, Real News, Firebase, scikit-learn.**

INTRODUCTION

Misinformation and the Proliferation of Fake News: Combating the Scourge through Technological Innovation and Media Literacy.The pervasive problem of fake news and misinformation has significant global implications, impacting public opinion, politics, and societal trust. The rise of social media platforms has facilitated the rapid dissemination of false and misleading information, which can lead to harmful consequences such as public panic or misinformed decision-making. Addressing this challenge requires a multifaceted approach that combines technological solutions, media literacy education, and cross-sector collaborations.This paper presents Fact Finder, a web application that aims to combat the spread of fake news by providing users with reliable information and rigorous analysis. Through continuous updates and advancements, Fact Finder strives to stay ahead of emerging trends in misinformation tactics. By empowering users with tools to discern truth from falsehood, this platform contributes to fostering a more informed society .The proposed system leverages cutting-edge technologies in data mining, natural language processing, and information retrieval to automatically identify and debunk misinformation. Furthermore, it incorporates robust fact-checking methodologies that involve cross-verification from multiple trusted sources, ensuring transparency and credibility in its processes.Recognizing the pivotal role of education and media literacy, Fact Finder incorporates interactive modules and resources designed to enhance users' critical thinking abilities and awareness of common misinformation tactics. By promoting these essential skills, the platform aims to cultivate a more discerning and resilient public capable of navigating the complexities of the modern information landscape.

Underpinning Fact Finder's approach is a strong emphasis on collaborations with experts from diverse fields, including journalism, data science, and technology [6]. This interdisciplinary cooperation ensures that the platform benefits from the collective wisdom and expertise necessary to tackle the multifaceted challenges posed by fake news effectively.The following sections of this paper delve into the technological architecture, data processing pipelines, and user interface design of Fact Finder, while also exploring the platform's educational components and strategies for disseminating reliable information. Additionally, the paper discusses the system's performance evaluation, future development roadmap, and potential societal impact in fostering a more trustworthy and informed digital ecosystem

I. Literature review

The proliferation of misinformation and fake news has emerged as a major challenge in the digital age, with far-reaching implications for public discourse, decision-making, and societal trust. The rise of social media and the ease with which information can be created and disseminated online have exacerbated this problem.. Consequently, there has been a growing body of research dedicated to developing technological solutions, promoting media literacy, and fostering cross-sector collaborations to combat the spread of misinformation.Our Fact Finder web app aims to combat the spread of fake news by providing users with reliable information and analysis.Through continuous updates and improvements, we strive to stay ahead of emerging trends in misinformation.By empowering users with tools to discern truth from falsehood, we contribute to fostering a more informed society.Education and media literacy are essential components in the fight against fake news, and our platform seeks to promote these values. Collaborations with experts in journalism, data science, and

technology are integral to our mission of combating fake news effectively.

Detecting fake news automatically has become a crucial research area, with various approaches being proposed in recent years . This literature review focuses on the effectiveness of using logistic regression and TF-IDF vectorization for fake news detection, comparing it with other less effective methods.

Traditional and Machine Learning Approaches Early attempts at fake news detection often relied on manual fact-checking by domain experts, which is time-consuming and not scalable. Rule-based approaches using predefined linguistic patterns or writing styles have also been explored but are limited by the need for manual feature engineering. Machine learning techniques such as Naive Bayes classifiers and decision tree-based methods have shown promising results but may not effectively capture the complex patterns in text data.

Logistic Regression with TF-IDF Vectorization Logistic regression combined with TF-IDF vectorization has emerged as a powerful approach for fake news detection. Logistic regression is a simple yet effective linear classification algorithm that can handle high-dimensional sparse data, making it well-suited for text classification tasks. TF-IDF vectorization is a widely used technique for representing text data as numerical features, capturing the importance of words based on their frequency in a document and rarity across the corpus.

Methodology

The methodology employed in the development of the FactFinder web application primarily involves the utilization of machine learning (ML) techniques and natural language processing (NLP) algorithms to distinguish between actual and fake news. The core components of the methodology include data preprocessing, feature extraction, model training, and summary generation.

Data Collection and Preprocessing

Two datasets were collected to train and evaluate the model: one containing actual news articles and the other containing fake news articles. These datasets were merged randomly, resulting in a combined dataset. A binary classification label, denoted as "classification," was assigned to each article, with a value of 1 indicating fake news and 0 indicating actual news.

To prepare the textual data for analysis, preprocessing steps were performed using Python libraries such as re and porterStemer. These steps involved the removal of special characters, punctuation, and stop words, as well as the stemming of words to their root forms, to ensure consistency and reduce noise in the dataset.

Feature Extraction and Model Training

The preprocessed data was then transformed into feature vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique. This process converts each article into a numerical representation, capturing the importance of each word in the document relative to the entire corpus.

The dataset was split into training and testing sets using the train_test_split function, with a certain percentage allocated for training and the remainder for testing. A logistic regression model was chosen as the classifier due to its simplicity and effectiveness in binary classification tasks.

The logistic regression model was trained on the training data, where it learned to map the feature vectors to the corresponding classification labels. The trained model was then evaluated on the test data to assess its performance in classifying news articles as either real or fake.

Summary Generation Using Recurrent Neural Networks (RNN)

In addition to classifying news articles, FactFinder employs a Recurrent Neural Network (RNN) to generate summaries of the input news. RNNs are well-suited for sequential data processing tasks and are capable of capturing dependencies between words in a text.

The RNN takes the preprocessed input news as input and generates a concise summary by sequentially processing the words and learning their contextual representations. The summary provides users with a brief overview of the main points of the news article, aiding in quick comprehension and decision-making.

Evaluation Metrics

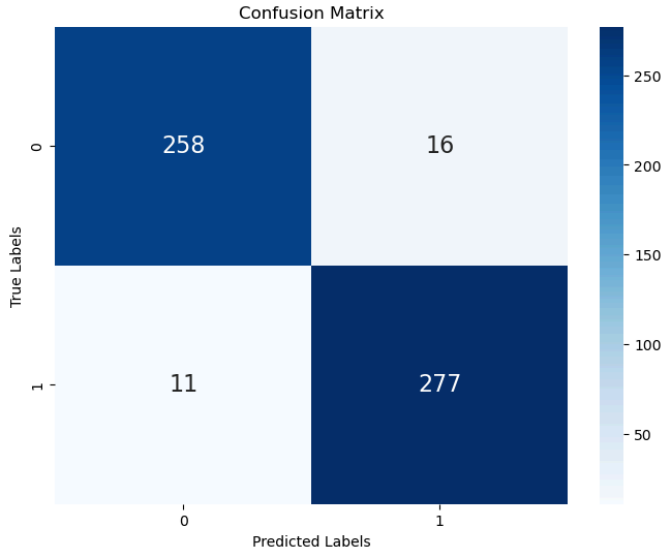
The performance of the logistic regression model was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to correctly classify news articles as real or fake, as well as its overall effectiveness in discriminating between the two classes.

Results

The results of the FactFinder model showcase its efficacy in discerning between actual and fake news articles.

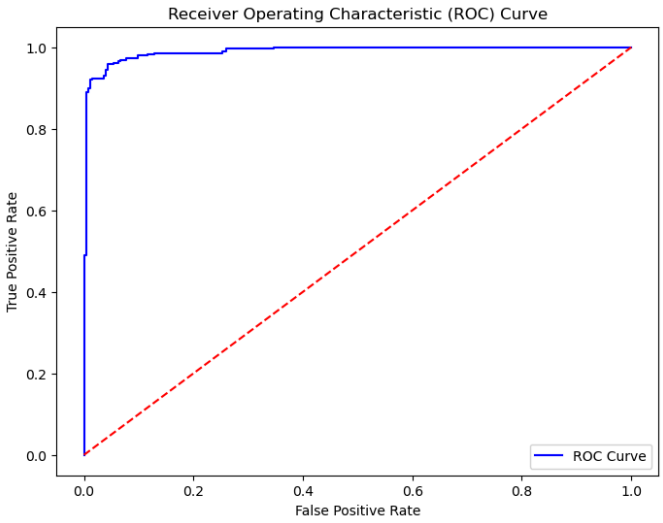
Confusion Matrix:

The breakdown of predicted labels reveals accurate identification of true negatives and true positives, with minimal false classifications. True Negatives (TN): 258 predicted labels of 0, correctly identifying actual news articles, False Positives (FP): 16 predicted labels of 1, incorrectly classifying actual news articles as fake, False Negatives (FN): 11 predicted labels of 0, failing to identify actual fake news articles, True Positives (TP): 277 predicted labels of 1, accurately identifying fake news articles.



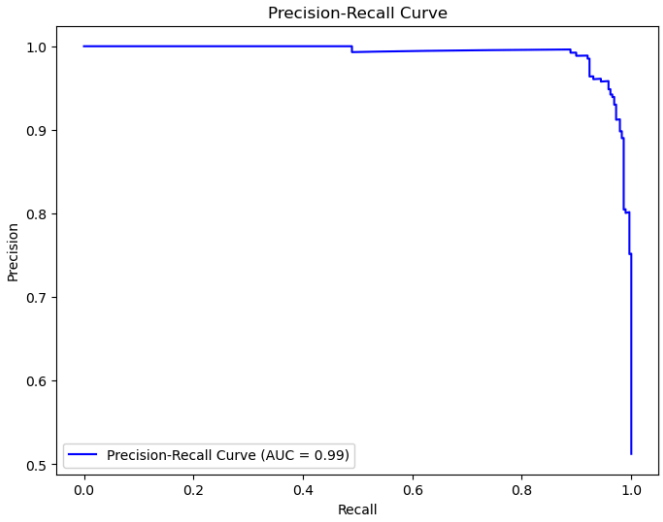
ROC Curve:

The ROC curve illustrates the trade-off between the model's true positive rate (TPR) and false positive rate (FPR). A higher TPR and lower FPR indicate better model performance in distinguishing between actual and fake news articles.



Precision-Recall Curve:

With a precision of 0.94 and recall of 0.96, the model strikes a balance between accurately identifying actual news articles and minimizing false positives.



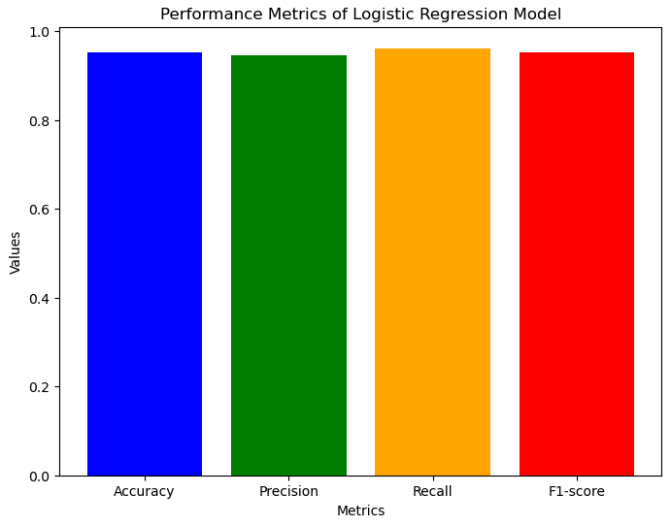
Performance Metrics:

Accuracy Score: 0.95 reflects the high proportion of correctly classified predictions.

Precision Score: 0.94 signifies the accuracy of identifying actual news articles among positive predictions.

Recall Score: 0.96 indicates the model's effectiveness in correctly identifying actual news articles.

F1 Score: 0.95 harmonizes precision and recall, indicating robust overall performance.



Media Literacy and Education

Recognizing the importance of equipping individuals with the skills to critically evaluate information sources, several studies have focused on media literacy interventions. Curricula have been developed to teach strategies for identifying misinformation tactics, assessing source credibility, and fact-checking claims [12,13]. The FactFinder aims to contribute to these efforts by incorporating educational resources and promoting media literacy as an essential component in combating fake news [9].

Moreover, researchers have explored the role of digital literacy [14] and the effectiveness of game-based learning [15] in promoting media literacy and critical thinking, which could inform FactFinder's educational approach.

Cross-Sector Collaborations

Tackling the complex challenge of misinformation requires interdisciplinary collaboration among various stakeholders, including technology companies, news organizations, academia, and policymakers. The FactFinder acknowledges the importance of such collaborations, particularly with experts in journalism, data science, and technology, to enhance the effectiveness of its mission [9]. Initiatives such as the Trust Study[16] and the Credibility Coalition have brought together experts from diverse fields to develop standards, share best practices, and foster greater transparency and accountability in the dissemination of information, which could serve as models for FactFinder's collaborative efforts.

As misinformation tactics evolve, researchers must continuously adapt and refine their approaches. Promising avenues for future work include leveraging advancements in deep learning and multimodal analysis to detect misinformation across various media formats, exploring the role of human-AI collaboration in fact-checking processes, and investigating the psychological and societal factors that contribute to the spread and acceptance of misinformation.

III. CURRENT SOLUTIONS

Our objective is to create a seamless web application enabling users to verify the authenticity of news articles effortlessly. Through the implementation of advanced Machine Learning algorithms, our platform discerns between genuine and fabricated content, scrutinizing data accuracy with precision. Users can conveniently upload articles and receive immediate results. Moreover, registration and login functionalities enable users to engage in discussions and connect with individuals sharing similar interests, fostering a vibrant community of informed users. This study necessitates the incorporation of the following pivotal features:

1)OS (platform-independent): - Users should be able to access the interface of the application from any operating system to view the results of their news articles conducted by the Logistic Regression model.

2)Text Format: - The application should have the func-tionality to identify and categorize the given text or data according to their findings.

3)Logic behind the Features: - The application should be able to process the user's news data and retrieve the results generated by the Logistic Regression model.

VI. PRODUCT RESULTS

So far, the accuracy of our model is around 84% which is good enough considering that the model is learning and improving every day. The model is very good at detecting fake news and the real news that are clear and perfectly giving the output result whether it's actual news or fake news.. We have doubled our test data and every day try to get in as much data as possible so that the model performance improves and accuracy. Although we have come a both on the front-end part of the FactFinder website as well as the backend part. The UI/UX of the FactFinder website looks great and has incorporated features like User Profile Creation, Verify News,

Login Page, and About FactFinder details and About us section.

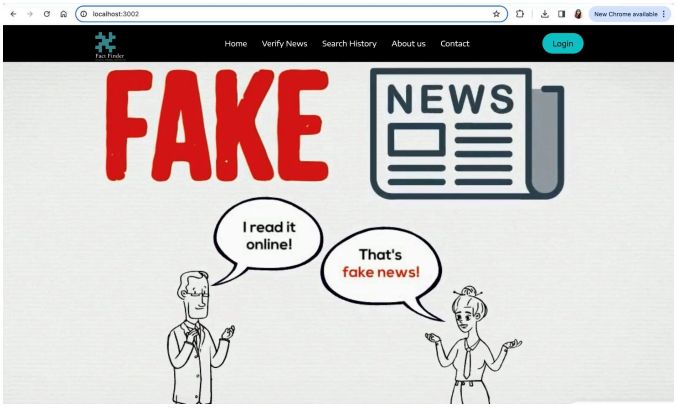


Fig: FactFinder Home Page

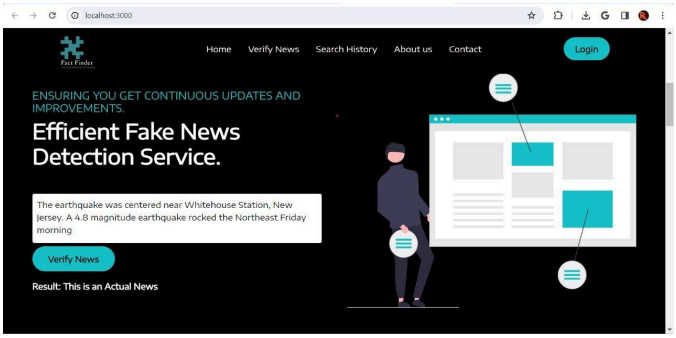


Fig: Verify Actual News

.

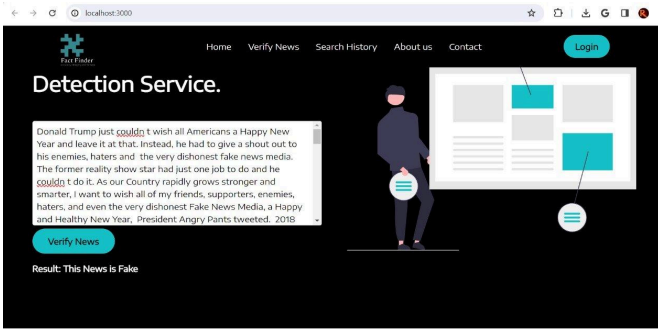


Fig: Verify Fake News

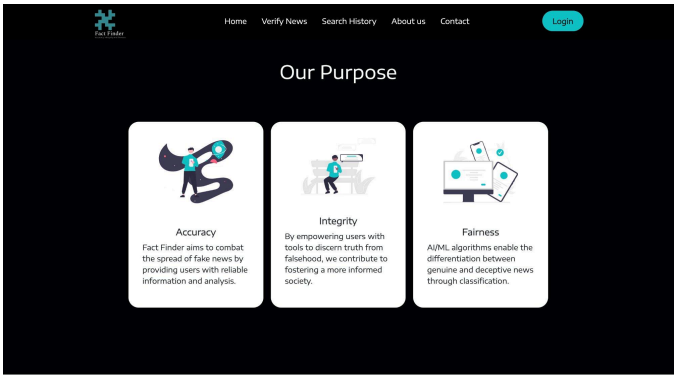


Fig: About us Tab

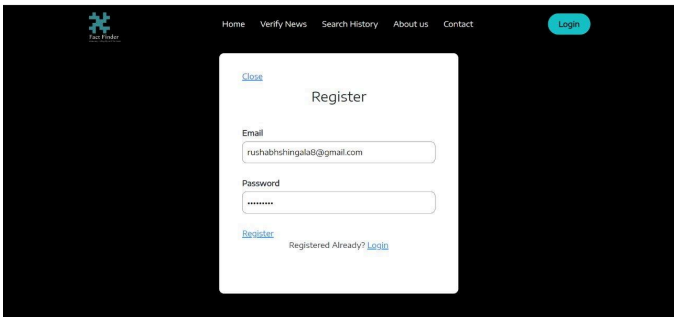


Fig: Registration Tab

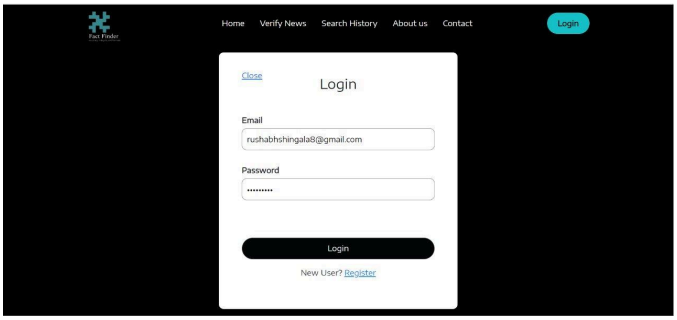


Fig: Login Tab

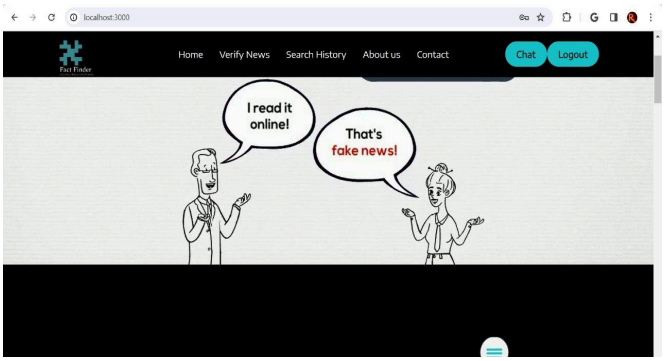


Fig: Chat Feature

II. Conclusions

The version of this template is V2. Most of the formatting instructions in this document have been compiled by Causal Productions from the IEEE LaTeX style files. Causal Productions offers both A4 templates and US Letter templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEEtran.cls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained. Causal Productions has used its best efforts to ensure that the templates have the same appearance.

Causal Productions permits the distribution and revision of these templates on the condition that Causal Productions is credited in the revised template as follows: "original version of this template was provided by courtesy of Causal Productions (www.causalproductions.com)".

Acknowledgment

References

[1] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[2] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[3] FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

[4] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[5] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[6] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[7] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

[8] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.

[9] Khanam Z., Alkhaldi S. "An Intelligent Recommendation Engine for Selecting the University for Graduate Courses in KSA: SARS Student Admission Recommender System." In: Smys S., Bestak R., Rocha Á. (eds) Inventive Computation Technologies. ICICIT 2019. Lecture Notes in Networks and Systems, vol 98. Springer, Cham. 2019.

[10] Zhang, Jiawei, Bowen Dong, and S. Yu Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network." 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, 2020.

[11] Cardoso Durier da Silva, F., Vieira, R., & Garcia, A. C. (2019, January). Can machines learn to detect fake news? a survey focused on social media. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

[12] Bovet, Alexandre, and Hernán A. Makse. "Influence of fake news in Twitter during the 2016 US presidential election." Nature communications 10.1 (2019): 1-14. The science of fake news.

[13] Veronica Perez-Rosas et al. Available at: https://www.researchgate.net/publication/319255985_Automatic_Detection_of_Fake_News August, 2017.

[14] Prabhjot Kaur et al. " Hybrid Text Classification Method for Fake News Detection." International Journal of Engineering and Advanced Technology (IJEAT) , 2388-2392. 2019.

[15] I. Traore et al. "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques." International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments (pp. 127–138). Springer International Publishing . 2017.