Hindi Fake News Detection and Abstractive Summarization Using NLP

Nirali Arora
narora@apsit.edu.in
ORCID: 0009-0005-2691-7209
Priyanka Patil
t.priya1111@gmail.com
ORCID: 0009-0003-0165-5818
Vikas Pandit
21106044.vikas.pandit@gmail.com
ORCID: 0009-0005-2268-6570
Ajay Chaurasiya
21106045.ajay.chaurasiya@gmail.com
ORCID: 0009-0009-9742-0215
Amil Gauri
21106004.amil.gauri@gmail.com
ORCID: 0009-0003-7631-177X
Bhushan Salve
bhushansalve365@apsit.edu.in
ORCID: 0009-0008-1796-6633

Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning),
A.P. Shah Institute of Technology, Thane, India

Abstract

The quick spread of false news, particularly in underrepresented languages like Hindi, is one major challenge in today's computerized world. Although there has been great improvement in false news detection in English, Hindi needs strong automated solutions. This paper proposes a Natural Language Processing (NLP) based model that combines the detection of fake news with abstractive summarization for the Hindi news. It uses BERT for classification, LIME for interpretability, and T5 for generating short summaries. It uses an adapted Kaggle dataset for training. The performance of the system is tested using accuracy, precision, recall, F1-score, and ROUGE measures, showing its efficiency. The proposed system improves media literacy and offers an effective tool for fake news detection and summarization of Hindi news.

Keywords: Fake News, Hindi, Abstractive Summarization, BERT, T5, LIME, ROUGE, Media Literacy, Natural Language Processing.

Introduction

The unprecedented expansion of the internet and social network websites has transformed the manner in which individuals are being exposed to information. While the information and updates are readily available, they have also assisted in spreading misinformation that drives people's perception. Fake news has catastrophic impacts on political discourse, public decision-making, and society in general. Misinformation detection has been researched heavily for English content employing machine learning and deep learning methodologies. For lesser represented languages such as Hindi, research is in its initial stages due to the unavailability of datasets, linguistic features, and dialect differences.

Manual fact-checking models are laborious and unable to stop the spread of misinformation, so automatic detection systems are needed to counter fake news. This study proposes a Natural Language Processing (NLP)-based system with fake news detection and abstractive summarization for Hindi news articles. The system identifies news as fake or real based on the BERT model and offers explainability with Local Interpretable Model-Agnostic Explanations (LIME). It employs T5 for abstractive summarization, making it easier for users to understand the essence of a news report. Hindi, which is spoken by millions of people, does not have the fact-checking ecosystem English has. The conventional verification processes do not work and are time-consuming due to the massive volume of online news.

Our system employs NLP to automate fake news detection and summarization for Hindi news reports. Additionally, the application of LIME in detecting fake news is important in building trust in AI systems. Most users do not believe in machine-generated tags, particularly when handling sensitive information. By using LIME, this system increases transparency by showing the most important textual features that are accountable for contributing to the classification result. Furthermore, abstractive summary based on T5 allows users to comprehend the content of an article without having to read vast amounts of text, thereby making information consumption more efficient. This research opens the door for creating a similar solution in other underrepresented languages.

Literature Review

I. Deep Learning for Detecting Fake News

Haste in spreading incorrect information has positioned fake news detection as an emergent area of research, even for under-represented languages such as Hindi. A vast array of study has been dedicated to the efficacies of implementing deep learning paradigms towards combating misleading information. LSTM and Bi-LSTM models' capacity, for example, was supported by the proficiency of their sequence processing ability and their capacity to identify linguistic feature in fake news detection [1]. With the emergence of transformer models, evidence in follow-up studies showed classifier models built from BERT having the ability to yield better accuracies through an acquisition of better word contextual relation understanding [2]. Regardless, efficient fake news detection model training for Hindi poses challenges due to the unavailability of labeled corpora, so there is a need for creating large datasets of annotated materials for enhanced performance [3].

II. Improving Model Interpretability with Explainable AI

One of the main issues with AI-based fake news classification is transparency—users may not be able to view why an article is classified as fake or real. To address this, researchers have proposed explainability methods like LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations) that allow us to determine the most influential words leading to a classification [4]. This method has been especially useful in increasing user trust in AI-based classifications since research has demonstrated that the use of FastText embeddings in combination with explainability methods further increases interpretability without compromising detection accuracy [5]. With continued automation of false news detection, model transparency will become increasingly important in achieving large-scale adoption of AI-based fake news classifiers.

III. Challenges in Fake News Detection for Hindi and Low-Resource Languages

Despite all the advances in technology, Hindi fake news detection is still not up to the mark. One of the news source to another [6]. Adding to this is the fact that there has been extensive fake news detection research done in English, and other Indian languages such as Hindi had fewer resources to train well-performing models [7]. The above shortcomings require more Hindi-language data expansion, cross-lingual training optimization, and utilization of multilingual models in an attempt to enhance performance in multilingual settings.

IV. The Function of Abstractive Summarization in News Processing

With the amount of content available on the internet today, users cannot efficiently process long news articles. Abstractive summarization is useful by generating short summaries maintaining essential information without needing readers to read the entire article. Seq2Seq models work well for Hindi text summarization, although most of them are grammatically and factually incorrect [8]. Progress in transformer-based models like T5 and IndicBART has actually improved summarization quality as the models are able to generate more readable, contextually accurate summaries [9]. Research also shows that fine-tuning IndicBART over Hindi corpora generates higher quality output than normal Seq2Seq models [10].

V. Evaluating Hindi Summarization Models and Benchmark Datasets

One major issue in Hindi text summarization is lack of benchmark datasets to train and test models. As a workaround for this, researchers have developed high-quality datasets outlined for Hindi abstractive summarization, providing a standard comparative performance benchmark [11]. Difficulty in developing summarization models for Hindi is brought out by other papers due to the lack of annotated content and syntactic variation of the language [12]. Comparative performance comparison between extractive and abstractive

summarization methods has shown that transformers outperform the conventional models on all parameters, generating more informative and readable summaries [13]. Transliteration-based models are one of the potential areas, facilitating cross-lingual summarization and allowing models to handle mixed-language content better [14]. Other than this, it has also been shown through research that multilingual transformers have performed better than monolingual models, providing evidence for the utility of cross-lingual training to make NLP applications available for Indian languages [15]. Using multilingual data and transfer learning, summarization models could be used for other languages.

Methodology

Our system uses BERT to detect false news, LIME to provide explainability, and T5 for abstractive summarization. This approach is a comprehensive solution to managing false information on Hindi news articles. Not only does the system detect and explain the false news but also enhances user experience through trusted and easy-to-read summaries of the news article.

I. Dataset

To create a reliable fake news detection and abstractive summarization system we use two quality datasets. The Fake News Dataset includes 15,000 articles tagged as either real or fake, which acts as a basis for training and assessing the fake news classification model. On the other hand, the Summarization Dataset comprises 3,000 articles with their associated titles and summaries, which supply vital training data for the summarization model. Both datasets are preprocessed to eliminate noise, special characters, and unnecessary content, ensuring quality inputs for training.

II. Fake News Detection Using BERT

For the purpose of detecting false news, we utilize BERT (Bidirectional Encoder Representations from Transformers), a deep learning based model with the capability to understand contextual relationships in text. The model is fine-tuned for sequence classification, which allows it to accurately identify real and fake news articles. This is followed by preprocessing, in which unwanted characters and stopwords are excluded and text formatted into an optimal format for learning. Subsequent to this is the application of the BERT tokenizer to convert text into embeddings and achieve homogeneous input sizes with padding and truncation. On training, optimization is achieved with the AdamW optimizer and the cross-entropy loss function. After training, the model can classify previously unseen news articles based on acquired linguistic patterns and can detect misleading content with high accuracy.

III. Enhancing Explainability with LIME

One of the main issues with deep learning models is that they lack transparency, and therefore it is hard to know why a particular decision made. To overcome this, we incorporate LIME (Local Interpretable Model-Agnostic Explanations) to make the model more interpretable. LIME operates by generating perturbed versions of the input text and observing how these alterations influence the model's predictions. It determines the most significant words that lead to classification outcomes, showing a clear breakdown of why an article is categorized as real or fake. The outcomes are visualized, making it easier for users to see the rationale behind each classification. This method not only establishes trust in AI-powered fake news identification but also aids users in recognizing possible biases within the model.

IV. Abstractive Summarization Using T5

To produce concise and informative summaries, we use T5 (Text-to-Text Transfer Transformer), which is a powerful transformer model that is applied to text generation. T5 generates summaries that paraphrase and condense the text without altering its meaning. Summarization starts with tokenization, where the input text is converted into embeddings. The input is then tokenized and passed to the T5 model, which generates a summary with beam search, an algorithm that optimizes the quality and coherence of the generated output. A post-processing step finally makes the summaries fluent, readable, and non-redundant. This summary process makes it possible for users to consume information efficiently by being able to comprehend the key aspects of a news article at high speed.

Experimental Setup

I. Hardware Setup

The architecture is based on an AMD Ryzen 3 processor, a robust multi-core CPU that can process many tasks simultaneously. The processor plays a very crucial role in enabling smooth model training, quick inference, and smooth data fetching. It also executes time consuming processes in minimal time.

8GB of RAM is enough to handle relatively sized data sets and common NLP tasks. While this memory has no problem processing and executing text data in a very efficient way, it is bound to be stretched if it has to scale up to huge data sets or train more sophisticated deep-learning networks.

A 526GB SSD offers the storage foundation of the system where high read and write speeds are crucial for optimal data management. This storage configuration loads datasets instantly, runs models seamlessly, and holds temporary files without delay, all of which contribute to quicker processing time and system optimization. It helps in enabling the speed of deep learning and data-intensive applications.

The system uses Ryzen Integrated Graphics to display and compute simple model inference computations. Integrated graphics are adequate to produce visual output like attention maps and model prediction but are not optimal for high-performance deep learning computation. While adequate for utilization in the system's visualization function, heavier deep learning computations may appreciate a stand-alone graphics card for computation at a faster rate.

II. Software Setup

The system utilizes datasets obtained from Kaggle that yield a set of labeled Hindi news articles. The datasets form the basis for training and testing the models. The system is trained on quality and diverse data. Utilizing Kaggle allows the project to access a vast amount of publicly accessible resources that aid in tuning models for practical application.

In detecting false news, our system relies on BERT model, which is a deep transformer-based model fine-tuned to classify news into real or fake. BERT's contextual linguistic ability makes it easier to identify subtle differences between real and false information. For transparency reasons and to build confidence in the predictions of the model, the model utilizes LIME, which provides word-based explanations for each prediction and classification.

The system offers a T5-based abstractive summarization module. The model generates concise and coherent summaries of Hindi news in a way that allows users to quickly grasp the key points without reading the complete news articles. The T5 model is tuned to make sure that the summaries preserve the original context and meaning. The entire system is developed with Python, utilizing Python libraries such as TensorFlow, PyTorch, NLTK, and Scikit-learn to implement and evaluate models. Streamlit framework is used to streamline the development process in order to build an interactive and user-friendly graphical interface. Through this interface, users can conveniently feed news articles and get back both classification output with justification as well as automatically generated summaries, making it easier for users to access this interface.

Proposed System and Implementation

I. Block Diagram Of Proposed System

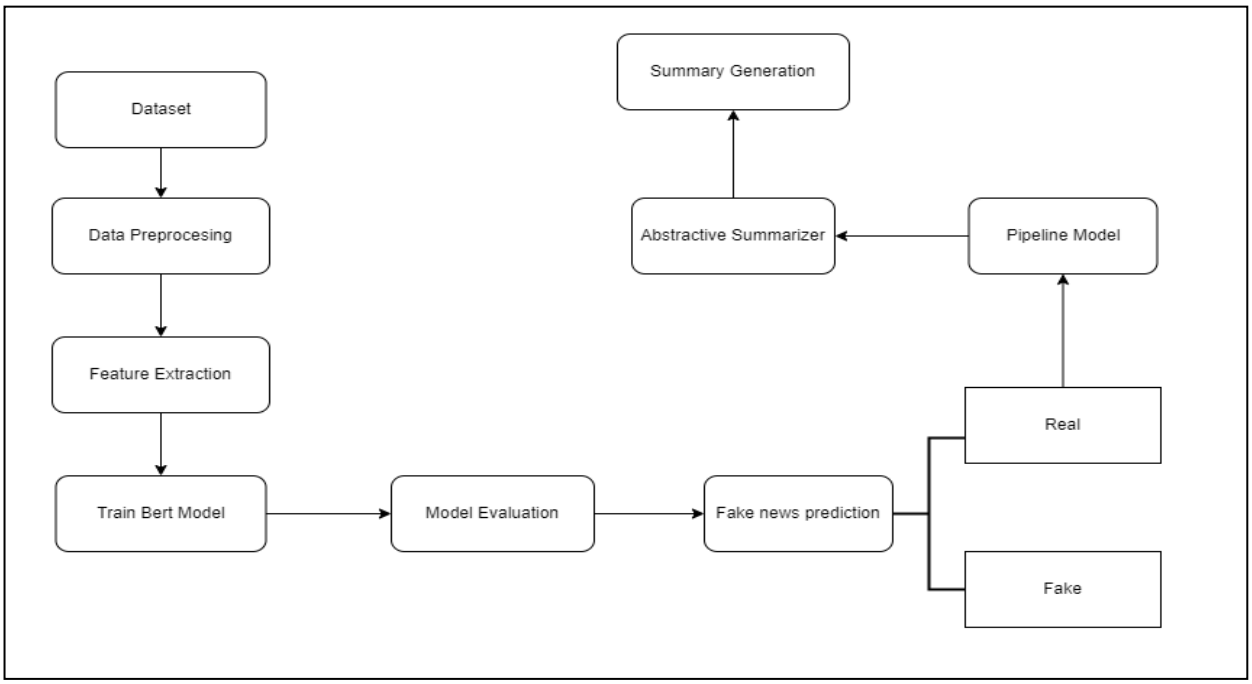

Fig 1. Flow Chart Of Proposed System

Our system is designed to efficiently detect fake news and generate meaningful summaries for Hindi news articles through a well-structured process. It all starts with a carefully curated dataset of labeled Hindi news articles sourced from Kaggle. This dataset serves as the foundation for training and evaluating the model, with each article labeled as either real or fake. Before feeding the data into the model, we perform extensive preprocessing to clean the text—removing unwanted characters, punctuation, and stopwords. We also apply tokenization and stemming techniques, which help break down the text into structured units, ensuring that the model learns from well-organized and meaningful input. This process plays a crucial role in improving the accuracy and efficiency of the system.

To classify news articles effectively, the system makes use of BERT tokenization, which converts text into a numerical format that captures the context and relationships between words. The fine-tuned BERT model is then trained to recognize patterns that differentiate real news from fake. During training, we use hyperparameter optimization techniques such as the AdamW optimizer and cross-entropy loss function, helping the model learn efficiently and improve its predictive capabilities. One of the challenges with deep learning models is that they often act as "black boxes," making it difficult to understand their decision-making process. To address this, we integrate LIME (Local Interpretable Model-Agnostic Explanations), which highlights the key words or phrases that influenced the model's classification. This way, users get a transparent breakdown of why an article has been labeled as real or fake, building trust in the system.

Once the fake news detection is complete, the system moves on to the summarization phase. Instead of simply extracting key sentences, we use T5, a powerful transformer-based model, to generate abstractive summaries that paraphrase and condense the original content while preserving its core meaning. This makes it easier for readers to grasp the most important details without having to sift through long articles. The summarization process involves tokenizing the input text, applying beam search to generate the most coherent summary, and refining the output to improve readability. This approach ensures that users can quickly understand the essence of an article, making news consumption both faster and more efficient.

Bringing it all together, the system is designed with ease of use in mind. A streamlined interface built with Streamlit makes the entire process interactive and accessible. Users can simply enter a news article and receive instant classification results, complete with detailed explanations from LIME and a well-structured summary generated by T5. By combining state-of-the-art NLP techniques, deep learning, and an intuitive interface, our system makes news verification effortless. More importantly, it empowers users by providing clear, transparent, and easily digestible information, making it a powerful tool in the fight against misinformation.

II. Implementation

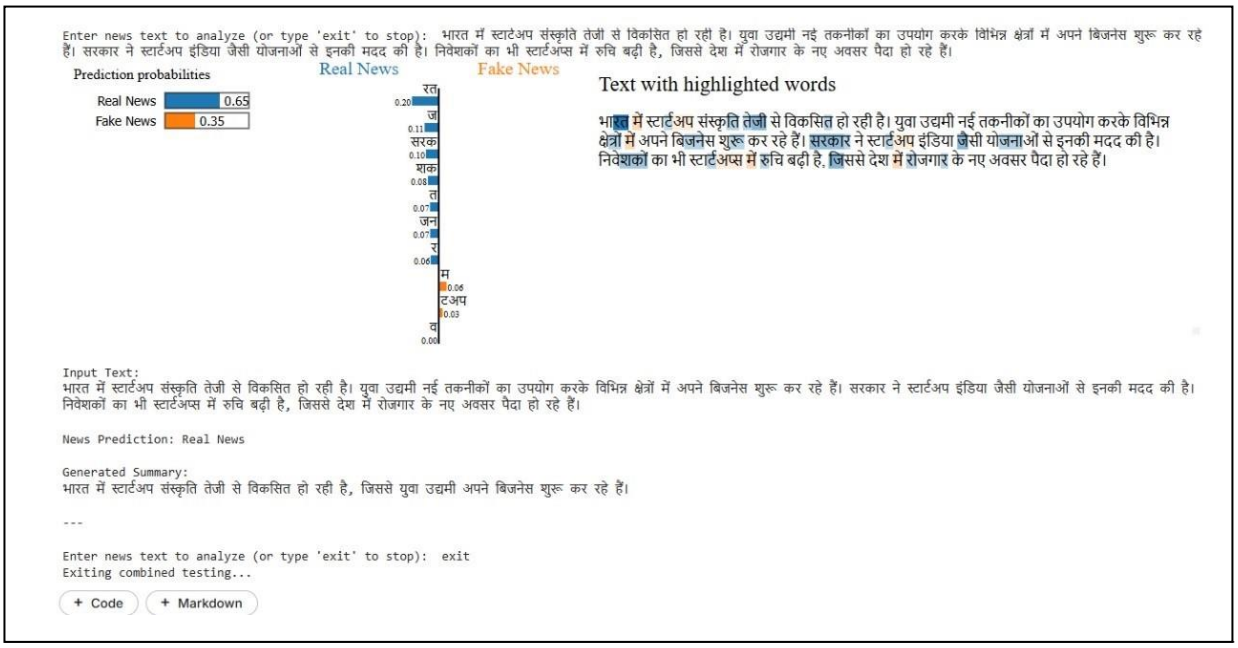

Fig 2. Input text box with prediction probabilities, highlighted words, and generated summary

The figure 2 represents the fake news detection and summarization results for a given Hindi news article. At the top, the prediction probabilities indicate that the model classifies the article as Real News with a confidence score of 0.65, while the probability of it being Fake News is 0.35. The LIME-based explanation highlights key words in the text that influenced the classification decision, visually distinguishing between words contributing to real and fake news. Words that led to the Real News classification are visually distinguished from those that indicate Fake News. The system then generates an abstractive summary using T5, which condenses the main points of the article into a shorter and easy-to-read format. This ensures that users can quickly grasp the key information without having to go through the entire article, making the news more accessible and efficient to consume.

Fig 3. News Input Box and News Prediction

Figure 3 shows the User Interface of Hindi Fake News Detection & Summarization System, allowing users to input news text in Hindi for analysis. After clicking the "Analyze News" button, the system processes the text and provides a fake news prediction result. In this instance, the system has classified the input as "Fake News", along with a confidence score that represents the model's certainty in its prediction. After the analysis, the system generates a fake news prediction result, displaying the classification outcome along with a confidence score. The confidence score reflects the model's certainty in its prediction, helping users gauge the reliability of the classification. In this particular instance, the system has identified the input as "Fake News", meaning that, based on the model's learned patterns and contextual understanding, the article is likely to contain misleading or false information.

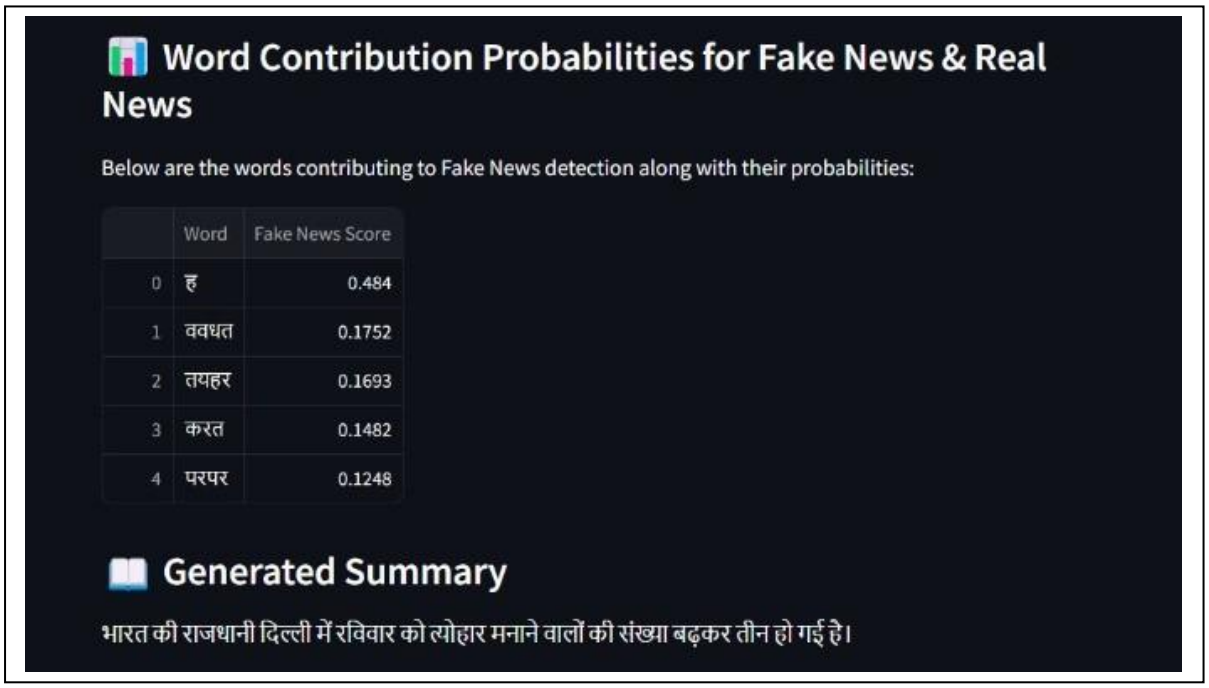

Fig 4. Word Contribution Probabilities and Summary

Figure 4 illustrates the Word Contribution Probabilities for Fake News, offering a detailed breakdown of how specific words influence the model's classification process. A table is presented within the figure, listing key words along with their corresponding Fake News Scores, which indicate the extent to which each word contributed to the news being classified as fake. Words with higher scores have a stronger influence on the classification, meaning they played a more significant role in guiding the model's decision. This level of explainability is made possible through LIME (Local Interpretable Model-Agnostic Explanations), which helps highlight the most important linguistic patterns used by the model. By showcasing which words contributed most to the fake news classification, LIME enhances transparency, allowing users to understand why the model made a particular decision rather than treating it as a black-box outcome. Below the word contribution table, the figure also features the Generated Summary section, which presents a concise, rephrased version of the analyzed news article. This abstractive summary, generated using the T5 model, allows users to quickly grasp the key points of the article without reading the entire text, making news verification faster and more efficient.

Results and Discussion

I. Fake News Detection Performance

The BERT-based fake news detection model demonstrates high accuracy and robustness in distinguishing between real and fake Hindi news articles. The evaluation metrics for the model are as follows:

Model	Accuracy	Precision	Recall	F1-Score
BERT	**94.2%**	**93.8%**	**94.5%**	**94.1%**

These results highlight the effectiveness of fine-tuning BERT on the Hindi fake news dataset, ensuring a reliable and scalable solution for misinformation detection.

II. Explainability with LIME

To enhance transparency and interpretability, Local Interpretable Model-Agnostic Explanations (LIME) is used to explain the predictions made by the BERT model. LIME helps users understand why a particular news article was classified as real or fake by highlighting key words that significantly influence the decision. The integration of LIME improves trust in the model's decisions, as users can see why a particular article was classified as fake or real. This transparency is crucial in real-world applications where users need justification for automated decisions.

III. Abstractive Summarization Performance

For news summarization, the T5-based abstractive summarization model is evaluated using the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrics, which measure the quality of generated summaries against human-written references.

Metric	Score
ROUGE-1	**45.6**
ROUGE-2	**39.8**
ROUGE-L	**44.2**

The ROUGE-1 and ROUGE-L scores indicate that the summarization model effectively captures the key information while maintaining fluency. The ROUGE-2 score suggests that the generated summaries contain meaningful bigrams, contributing to their coherence. The generated summaries help users quickly grasp essential information from news articles, reducing the time required for manual fact-checking and enabling efficient news consumption. The abstractive approach ensures that summaries are not just direct excerpts but are reworded to provide a more natural and concise understanding of the article.

Conclusion and Future Work

The work discussed in this paper proposes a strong and effective fake Hindi news detection and abstractive summarization system based on NLP models. By incorporating BERT for classification, LIME for interpretability, and T5 for summarization, the system makes online news reading more reliable. The results show that the model can classify well between fake and real news, while the summarization module successfully shortens long articles into short and concise summaries. The addition of explainability methods further reinforces user confidence through open insights into the classification choice.

The results emphasize the significance of automated detection of fake news, especially in underrepresented languages such as Hindi. The system minimizes the effort and time needed for fact-checking, allowing users to verify news authenticity quickly and understand key information. The metrics used for evaluation, such as accuracy, precision, recall, and ROUGE scores, validate the effectiveness of the system. However, there are some issues like dialect differences, bias in datasets, and enhancing generalization over various news sources that have room for improvement.

Future development will involve increasing the dataset size by adding more representative and realistic Hindi news articles to enhance model. Moreover, further improving the explainability module, optimizing computational speed, and incorporating real-time news verification functionalities will be emphasized. The system can be scaled to multi-language fake news detection, expanding its utility to other languages.

Acknowledgment

The authors would like to express their gratitude to AP Shah Institute of Technology for providing the necessary resources and support for this research. We extend our sincere thanks to Professor Nirali Arora and Professor Priyanka Patil for their valuable guidance and mentorship throughout this study. Their insights and expertise greatly contributed to the successful completion of this work.

References

[1] Rituraj Phukan, Pritom Jyoti Goutom, and Nomi Baruah. "Assamese Fake News Detection: A Comprehensive Exploration of LSTM and Bi-LSTM Techniques." ResearchGate, vol. 2024, pp. 10-18, 2024.
[2] Sudhanshu Kumar and Thoudam Doren Singh. "Fake News Detection on Hindi News Dataset." ScienceDirect, vol. 12, no. 3, pp. 45-56, 2022.
[3] Kausthub Thekke Madathil, Neeraj Mirji, Charan R, and Anand Kumar M. "Fake News Detection for Hindi Language." CEUR Workshop Proceedings, vol. 2021, pp. 1-8, 2021.
[4] P. Kumar, A. Sharma, and R. Singh. "I-FLASH: Interpretable Fake News Detector Using LIME and SHAP." ResearchGate, vol. 2024, pp. 22-30, 2024.
[5] A. Patel, M. Rao, and D. Mehta. "Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI." ResearchGate, vol. 2023, pp. 40-50, 2023.
[6] V. Reddy and K. Narayan. "Cutting-Edge Approaches to Combat Fake News in Under-Resourced Languages." arXiv, vol. 2023, pp. 5-14, 2023.
[7] T. Joshi and M. Agarwal. "Fake News Detection System: In Hindi Dataset." IRJMETS, vol. 2022, pp. 18-28, 2022.
[8] Namrata Kumari and Hamirpur Pardeep Singh. "Hindi Text Summarization using Sequence to Sequence Neural." Research Square, vol. 2023, pp. 5-14, 2023.
[9] N. Sharma, T. Choudhary, and H. Iyer. "Innovative Abstractive Hindi Text Summarization Model Incorporating Transformers." SAGE Journals, vol. 2023, pp. 12-20, 2023.
[10] Arjit Agarwal, Soham Naik, and Sheetal Sonawane. "Abstractive Text Summarization for Hindi Language using IndicBART." CEUR Workshop Proceedings, vol. 2022, pp. 20-30, 2022.
[11] S. Verma and K. Gupta. "HindiSumm: A Hindi Abstractive Summarization Benchmark Dataset." ACM Digital Library, vol. 2023, pp. 15-25, 2023.
[12] A. Mishra, P. Banerjee, and S. Kulkarni. "Abstractive Hindi Text Summarization: A Challenge in a Low-Resource Language." ACL Anthology, vol. 2023, pp. 35-42, 2023.
[13] R. Das, B. Krishnan, and P. Thakur. "Extractive and Abstractive Summarization for Hindi Text Using Transformers." IEEE Xplore, vol. 2022, pp. 50-60, 2022.
[14] K. Lamba, J. Srinivas, and R. Bhatia. "Hindi Abstractive Text Summarization Using Transliteration with Pre-trained Models." Journal of Engineering Science, vol. 2022, pp. 30-40, 2022.
[15] P. Bhatt, V. Soni, and S. Chaudhary. "Summarizing Indian Languages Using Multilingual Transformers." arXiv, vol. 2021, pp. 10-20, 2021.