AI-Powered Multilingual Fake News Detection with Evidence Retrieval

SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE COURSE OF
DIPLOMA IN BIG DATA ANALYSIS

SUNBEAM INSTITUTE OF INFORMATION TECHNOLOGY, PUNE

Submitted By:

Kavin Sreeram AG (92623)

Saurabh Rathore (92984)

Mr.Nitin Kudale
Centre Coordinator

Mrs.Manisha Hingne
Course Coordinator

SUNBEAM

CERTIFICATE

This is to certify that the project work under the title ‘Walmart Stores Sales Prediction’ is done by Dattatray Hake & Udit Deshmukh in partial fulfillment of the requirement for award of Diploma in Big Data Analysis Course.

Mr. Aniket P	Mrs. Manisha Hingne
Project Guide	**Course Coordinator**

Date:
04/02/2026

ACKNOWLEDGEMENT

A project usually falls short of its expectation unless aided and guided by the right persons at the right time. We avail this opportunity to express our deep sense of gratitude towards Mr. Nitin Kudale (Center Coordinator, SIIT, Pune) and Mrs. Manisha Hingne (Course Coordinator, SIIT ,Pune) and Project Guide Mr. Aniket P.

We are deeply indebted and grateful to them for their guidance, encouragement and deep concern for our project. Without their critical evaluation and suggestions at every stage of the project, this project could never have reached its present form.

Last but not the least we thank the entire faculty and the staff members of Sunbeam Institute of Information Technology, Pune for their support.

Kavin Sreeram AG
DBDA August 2025 Batch,
SIIT Pune

Saurabh Rathore
DBDA August 2025
Batch,SIIT Pune

.

TABLE OF CONTENTS

1. Introduction

1.1 Introduction And Objectives:

In recent years, the rapid growth of digital media and social networking platforms has led to the widespread circulation of news and information. While this has improved access to information, it has also resulted in the increasing spread of fake and misleading news. Such false information can influence public opinion, create social unrest, and negatively impact decision-making processes.

To address this issue, this project focuses on developing an AI-Based Fake News Detection System that automatically verifies the authenticity of news claims. The proposed system uses advanced Natural Language Processing (NLP) techniques, Machine Learning models, and Retrieval-Augmented Generation (RAG) to analyze user-provided claims and compare them with reliable online sources.

The system retrieves relevant evidence from trusted fact-checking websites and online resources, generates semantic embeddings, and evaluates the credibility of the claim using a BERT-based classification model. Based on this analysis, the system provides a final verdict such as True, False, or No Direct Evidence, along with supporting reasons and evidence.

Objectives of the Project

The main objectives of this project are:

- To design and implement an automated fake news detection system using AI and NLP techniques.
- To retrieve relevant evidence from online sources using a RAG-based pipeline.
- To classify news claims using a BERT-based deep learning model.
- To generate clear and understandable explanations for the verification results.
- To provide a user-friendly interface using Streamlit for easy interaction.
- To reduce the manual effort required in fact-checking and improve verification accuracy.

1.2 Why this problem needs To be Solved?

The spread of fake news has become a major challenge in the digital era. False information can quickly go viral and mislead large sections of society, leading to misunderstandings, panic, and even social conflicts. Manual fact-checking is time-consuming and requires significant human effort, making it difficult to handle the large volume of online content. An automated fake news detection system is essential to ensure that users receive reliable and verified information. By providing quick and accurate verification, such systems can help individuals make informed decisions and avoid being influenced by misleading content.

This problem also affects journalists, researchers, policymakers, and media organizations, who rely on accurate data for reporting and analysis. An efficient fact-checking system can support these professionals by reducing verification time and improving content credibility. Furthermore, automated verification tools can enhance public trust in digital platforms by filtering unreliable content. By integrating machine learning and real-time evidence retrieval, this project helps in building a more transparent and trustworthy information ecosystem.

In addition, the proposed system can be extended for use in social media monitoring, content moderation, and misinformation control, making it highly valuable in today's information-driven society..

1.3

Dataset Description

The dataset used in this project was collected from reliable fact-checking platforms such as Bharat Fake News Kosh and other verified online sources. It contains structured information related to news claims, verification results, and supporting evidence. The dataset is used for training and evaluating the fake news detection system.

The dataset consists of the following main attributes:

1. News and Claim Information

- **ID:** Unique identification number for each record
- **Statement:** Original news claim or statement
- **Eng_Trans_Statement: English** translated version of the statement

- **News Body**: Detailed news content
- **Eng_Trans_News_Body:** English translated news content

These fields represent the core textual data used for classification and analysis.

2. Source and Author Details

- **Author_Name**: Name of the author or publisher
- **Fact_Check_Source**: Organization that verified the claim
- **Source_Type**: Type of source (website, social media, news portal, etc.)
- **Fact_Check_Link**: URL of the verification article

These attributes help in identifying content credibility and source reliability.

3. Publication and Category Information

- **Publish_Date**: Date of publication
- **News_Category**: Category of news (politics, health, social, etc.)
- **Language**: Language of the original content
- **Region**: Geographic region of origin
- **Platform**: Platform where the content was published

These fields support temporal and regional analysis.

4. Media Type Information

- **Text**: Indicates presence of textual content
- **Video**: Indicates presence of video content
- **Image**: Indicates presence of image content
- **Media_Link**: Link to multimedia content

These attributes describe the format of news content.

5. Verification Label

- **Label**: Final verification result of the claim
    - 0 – False
    - 1 – True
    - 2 – No Direct Evidence

This column is used as the target variable for model training.

2. Problem Definition and Algorithm:

2.1 Problem Definition

The main problem addressed in this project is the automatic detection and verification of fake news and misleading claims available on digital platforms. Due to the rapid spread of information through social media and online news portals, false and unverified content can easily reach a large audience. Manual fact-checking is time-consuming and cannot handle the massive volume of data generated daily. Therefore, there is a strong need for an intelligent system that can analyze a given news claim and determine whether it is true, false, or lacks sufficient evidence.

The proposed system takes a textual claim as input and processes it using Natural Language Processing techniques. It retrieves relevant information from trusted online sources, generates semantic embeddings to measure similarity, and applies a BERT-based classification model to evaluate credibility. By combining Retrieval-Augmented Generation (RAG) with deep learning, the system provides accurate verification results along with supporting evidence. The objective is to minimize incorrect classifications and improve the reliability of automated fact-checking.

2.2 Algorithm Definition

In this project, advanced Natural Language Processing and deep learning techniques are used for detecting fake news and verifying online claims. The core components of the proposed system include BERT-based text classification, Retrieval-Augmented Generation (RAG), semantic embedding generation, and similarity-based evidence matching.

BERT (Bidirectional Encoder Representations from Transformers)

BERT is a transformer-based deep learning model developed for understanding natural language context. Unlike traditional models, BERT processes text bidirectionally, enabling it to capture both previous and next word relationships. In this project, BERT is fine-tuned on the fake news dataset to classify news claims into True, False, or No Direct Evidence categories.

Retrieval-Augmented Generation (RAG)

RAG is used to retrieve relevant evidence from trusted online sources before generating the final decision. The system performs web-based search and collects related articles and fact-check reports. These documents are converted into embeddings and compared with the input claim to identify the most relevant information.

Sentence Embeddings and Similarity Matching

The retrieved documents and input claims are transformed into numerical vectors using embedding models. Cosine similarity is applied to measure the semantic closeness between the claim and retrieved documents. Higher similarity scores indicate stronger evidence support.

LLM-Based Summary Generation

After identifying relevant evidence, a language model is used to generate concise summaries and explanations. This helps users understand why a particular claim was classified as true or false.

Final Decision Module

The outputs from the BERT classifier and RAG pipeline are combined to generate the final verdict. If strong evidence supports the claim, it is labeled as true. If evidence contradicts it, it is labeled as false. If sufficient evidence is not available, the system returns "No Direct Evidence".

3. Experimental Evaluation:

3.1 Methodology:

The objective of this project is to automatically verify news claims using artificial intelligence and external evidence sources. The dataset used in this project was collected from trusted fact-checking platforms such as Bharat Fake News Kosh and other verified repositories. It contains labeled news statements, translated content, publication details, and verification results.

Initially, the dataset was loaded and preprocessed to remove noise, handle missing values, and normalize text data. Tokenization, stopword removal, and text normalization were performed as part of preprocessing. Publication dates and categorical attributes were standardized.

The cleaned dataset was then divided into training and testing sets. The BERT model was fine-tuned using the training data, while the RAG pipeline was used to retrieve supporting documents from the web. During testing, unseen claims were evaluated by both modules.The performance of the system was measured using Accuracy, Precision, Recall, and F1-score. Experimental results demonstrated that the combined BERT and RAG approach improves verification reliability compared to standalone classifiers.

Loading in raw data

```
master_df.head()
df = pd.read_csv("fake_news_dataset.csv")
print(df.shape)
df.head()
```

Preprocessing:

After loading the dataset, several preprocessing steps were applied to prepare the data for analysis and model training.First, missing values were identified using isna() functions. Columns such as translated statements, news body, media information, and publication dates contained missing values. Instead of removing these records, missing textual values were replaced with empty strings, and categorical missing values were handled appropriately to avoid data loss.

Publication dates were normalized using regular expressions and converted into standard datetime format. From the parsed dates, additional features such as year, month, and day of the week were extracted for temporal analysis.

```
df["publish_date_parsed"] =
df["Publish_Date"].apply(clean_publish_date)
df["pub_year"] = df["publish_date_parsed"].dt.year
df["pub_month"] = df["publish_date_parsed"].dt.month
```

Handling Missing Values

The dataset contained several missing values in important columns such as translated statements, news body, publication date, author name, media information, and source details. These missing values needed to be cleaned to ensure accurate analysis and reliable model performance.

Initially, the number of missing values in each column was identified using the isna() function. Since a large portion of the missing values occurred in textual fields, removing such records would result in significant data loss. Therefore, instead of deleting these rows, missing text fields were replaced with empty strings, and missing categorical values were handled appropriately.

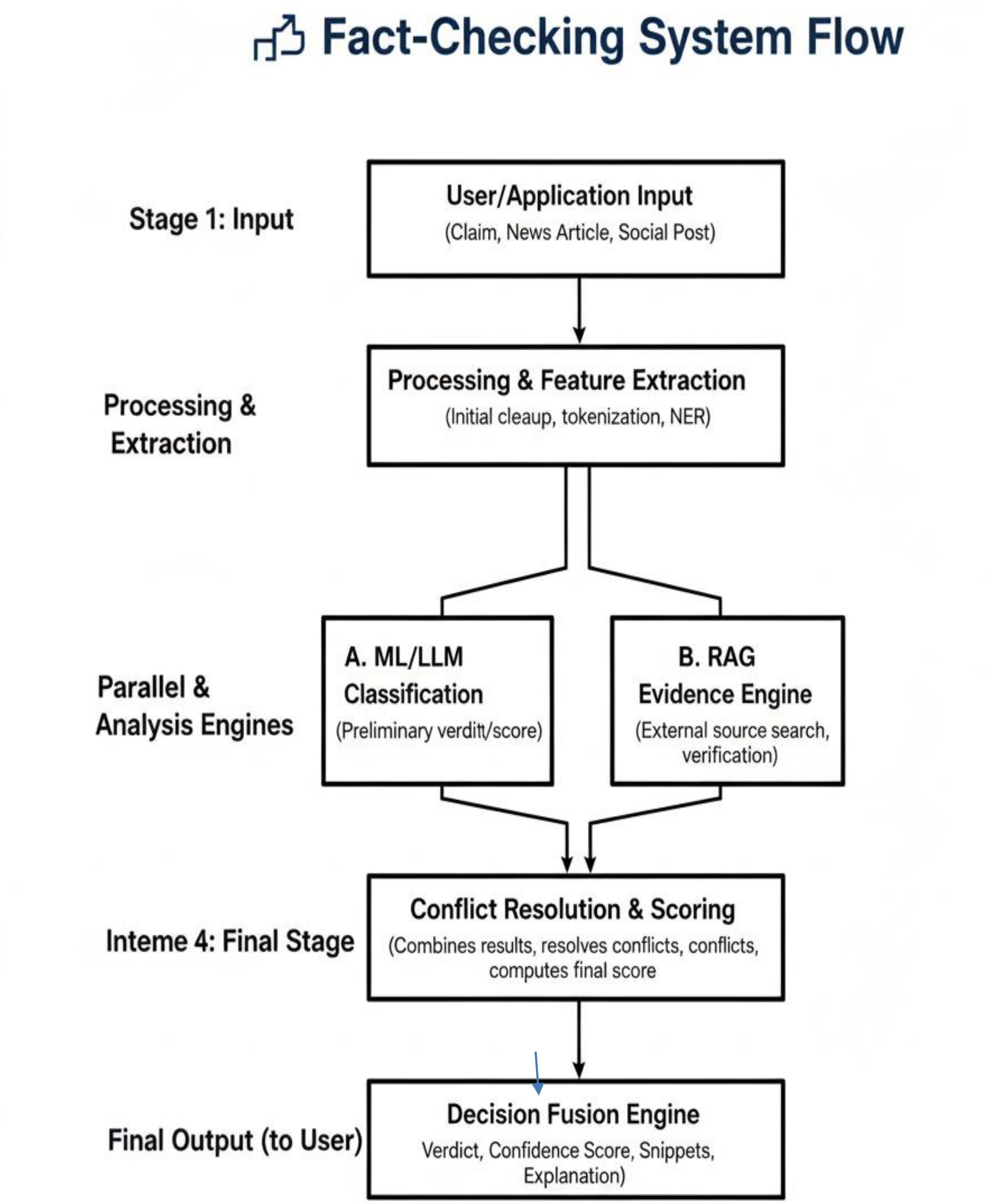
```
print(df.isna().sum())

df["Statement"] = df["Statement"].fillna("")
df["Eng_Trans_Statement"] = df["Eng_Trans_Statement"].fillna("")
df["News Body"] = df["News Body"].fillna("")
df["Eng_Trans_News_Body"] = df["Eng_Trans_News_Body"].fillna("")
df["Author_Name"] = df["Author_Name"].fillna("Unknown")
df["Fact_Check_Source"] = df["Fact_Check_Source"].fillna("Unknown")
```

Missing publication dates were converted to NaT (Not a Time) and excluded from time-based analysis when required. Boolean-like media fields such as Text, Video, and Image were normalized and missing values were treated as “No”.

```
df["Text"] = df["Text"].fillna("No")
df["Video"] = df["Video"].fillna("No")
df["Image"] = df["Image"].fillna("No")
```

After handling missing values, the dataset was re-evaluated to ensure that no critical null values remained. This preprocessing step helped in maintaining data integrity and improving the effectiveness of the classification model.

Flow Diagram :


Fact-Checking System Flow
Stage 1: Input
User/Application Input
(Claim, News Article, Social Post)
Processing &
Extraction
Processing & Feature Extraction
(Initial cleaup, tokenization, NER)
Parallel &
Analysis Engines
A. ML/LLM
Classification
(Preliminary verditt/score)
B. RAG
Evidence Engine
(External source search,
verification)
Inteme 4: Final Stage
Conflict Resolution & Scoring
(Combines results, resolves conflicts, conflicts,
computes final score
Final Output (to User)
Decision Fusion Engine
Verdict, Confidence Score, Snippets,
Explanation)

3.2 Exploratory Data Analysis

The sales for each week is plotted for 3 years. This shows only a slight relationship as the weekly sales increased towards the end of the year.

Exploratory Data Analysis (EDA) was performed to understand the structure, distribution, and patterns present in the fake news dataset. Various statistical and visual techniques were applied to analyze textual, categorical, and numerical features.

Author-wise Distribution

The frequency of news articles published by different authors was analyzed using a bar chart (Fig. 2). From the visualization, it can be observed that a small number of authors contributed a large portion of the total content. Authors such as admin, Chendur Pandian, and Yogesh Karia appeared frequently, indicating their major role in content creation and fact-checking activities.

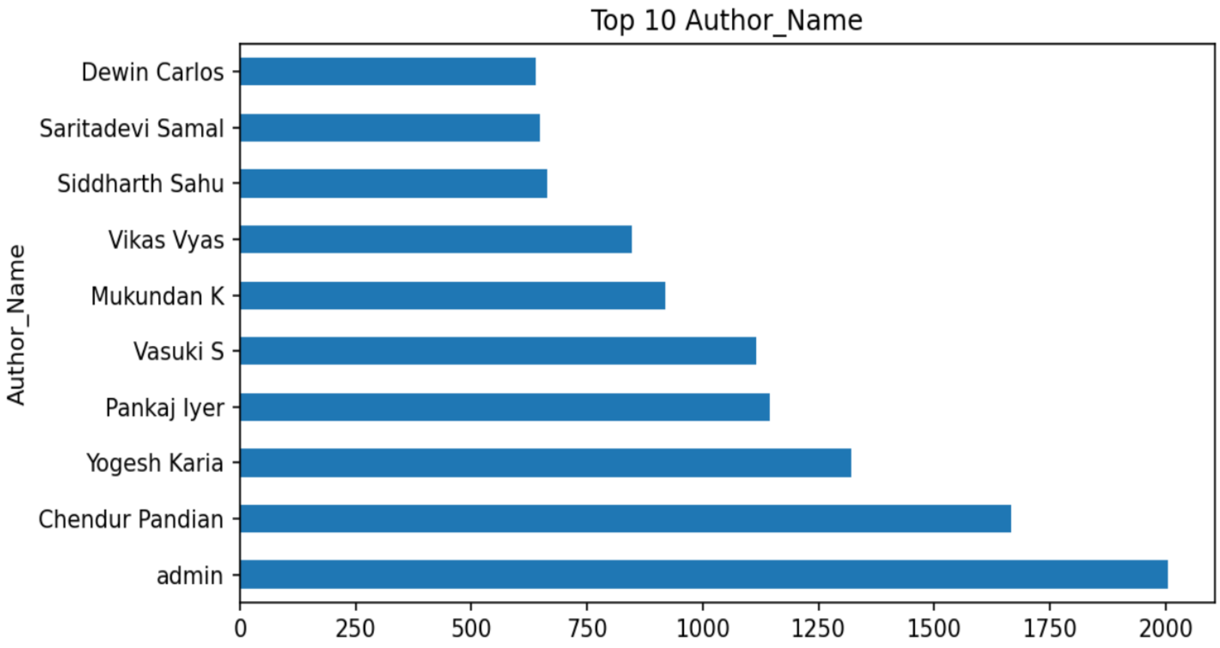


Fig 2: Top 10 Authors by Number of Articles

Language Distribution

The distribution of languages used in the dataset was analyzed using a bar plot (Fig. 3). The results show that English is the most dominant language, followed by regional languages. This indicates that English-translated content plays a significant role in automated analysis and model training.

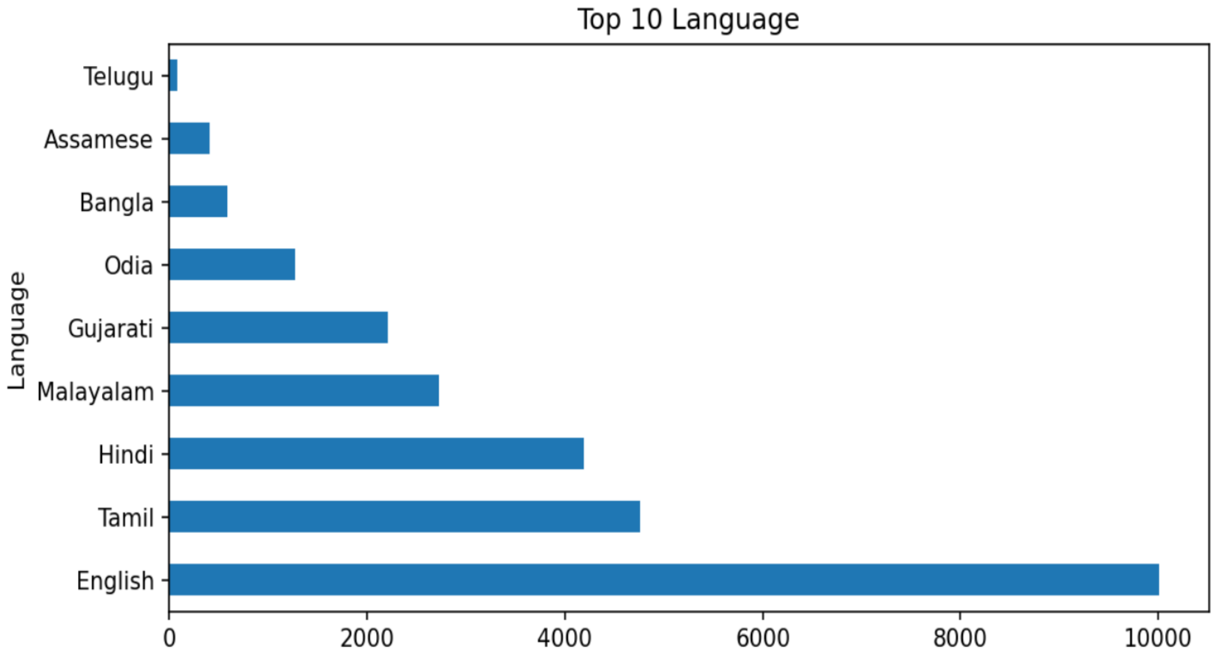


Fig 3: Language-wise Distribution of News Articles

News Category Analysis

The dataset was examined based on different news categories such as politics, health, social media, and public affairs (Fig. 4). Political and social-related news constituted a major portion of the dataset, reflecting the high spread of misinformation in these domains.

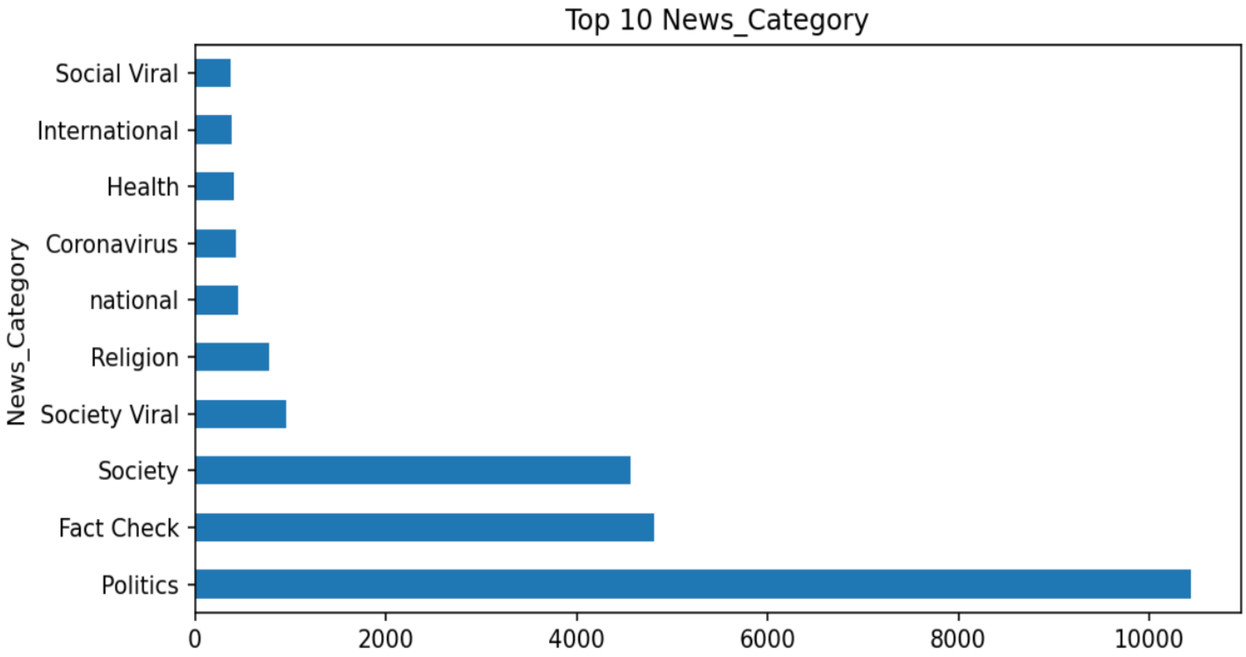


Fig 4: Distribution of News Categories

Label Distribution

The distribution of verification labels was visualized using a bar chart (Fig. 5). The analysis shows the proportion of true, false, and unverifiable news in the dataset. This distribution is important for understanding class imbalance and ensuring proper model training.

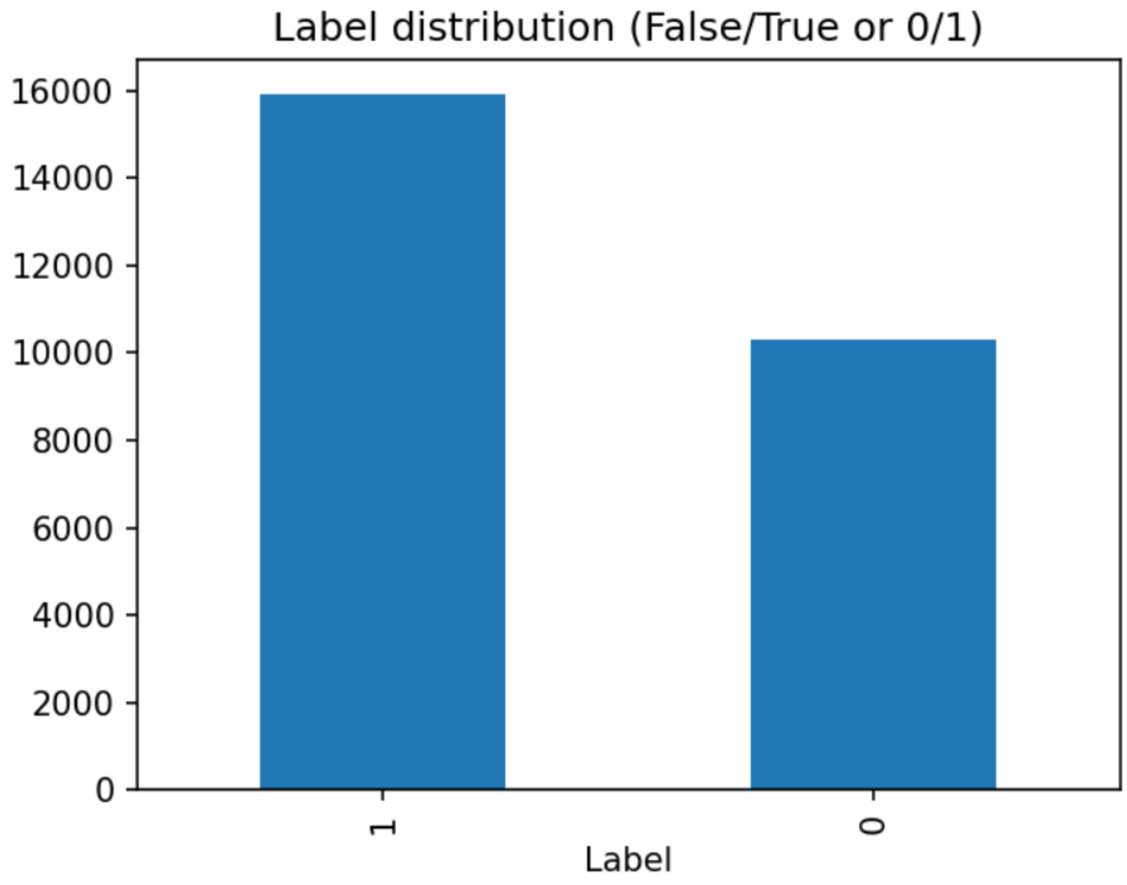


Fig 5: Distribution of Verification Labels

Media Type Analysis

The presence of different media types such as text, video, and images was analyzed (Fig. 6). Most of the news records contained textual content, while a smaller portion included videos and images. This indicates that text-based analysis is the primary focus of the system.

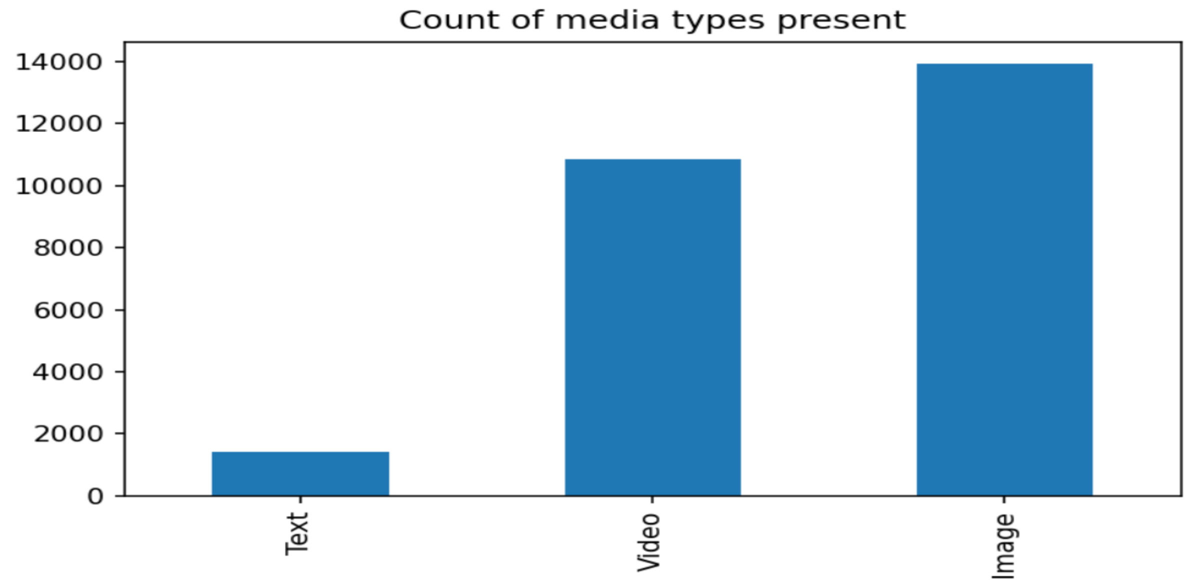


Fig 6: Media Type Distribution

Missing Data Analysis

A heatmap was used to visualize missing values across different columns (Fig. 7). Some columns such as translated text and publication dates contained missing values. These were handled during preprocessing by appropriate imputation techniques to maintain data quality.

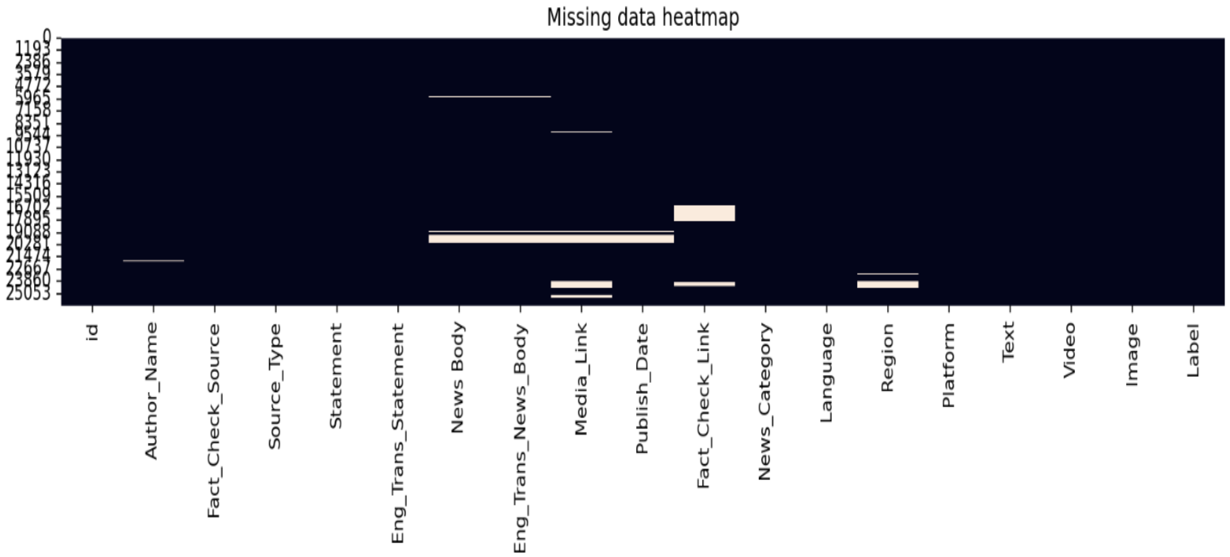


Fig 7: Missing Data Heatmap

Word Cloud and Keyword Analysis

Word clouds were generated from translated news content to identify frequently occurring words (Fig. 8). Common terms related to politics, government, public figures, and social issues were observed, highlighting dominant misinformation topics.

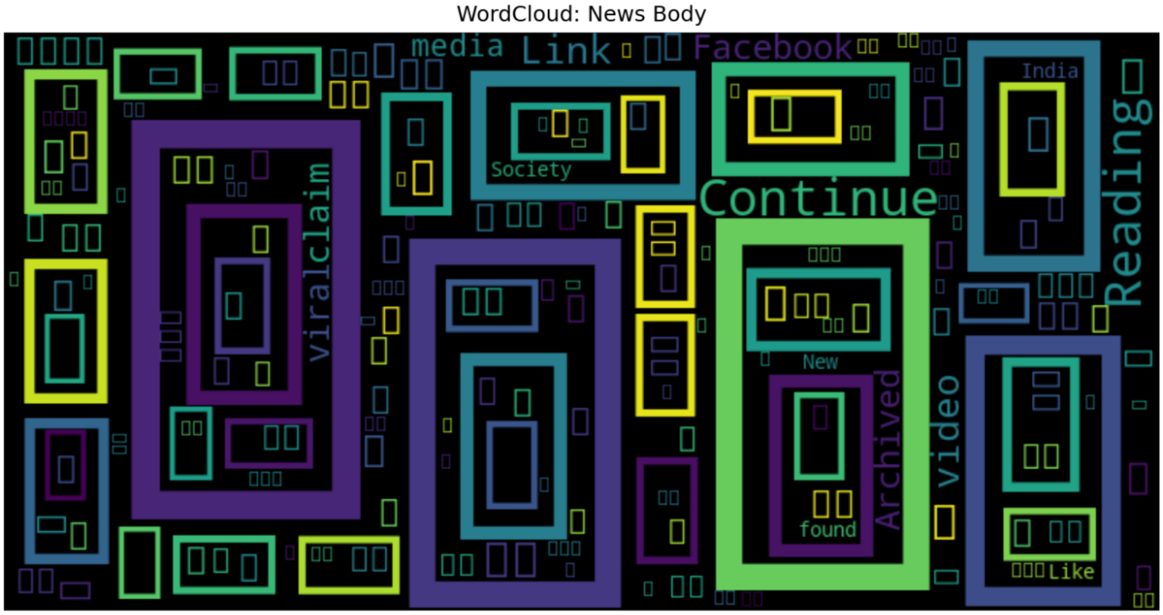
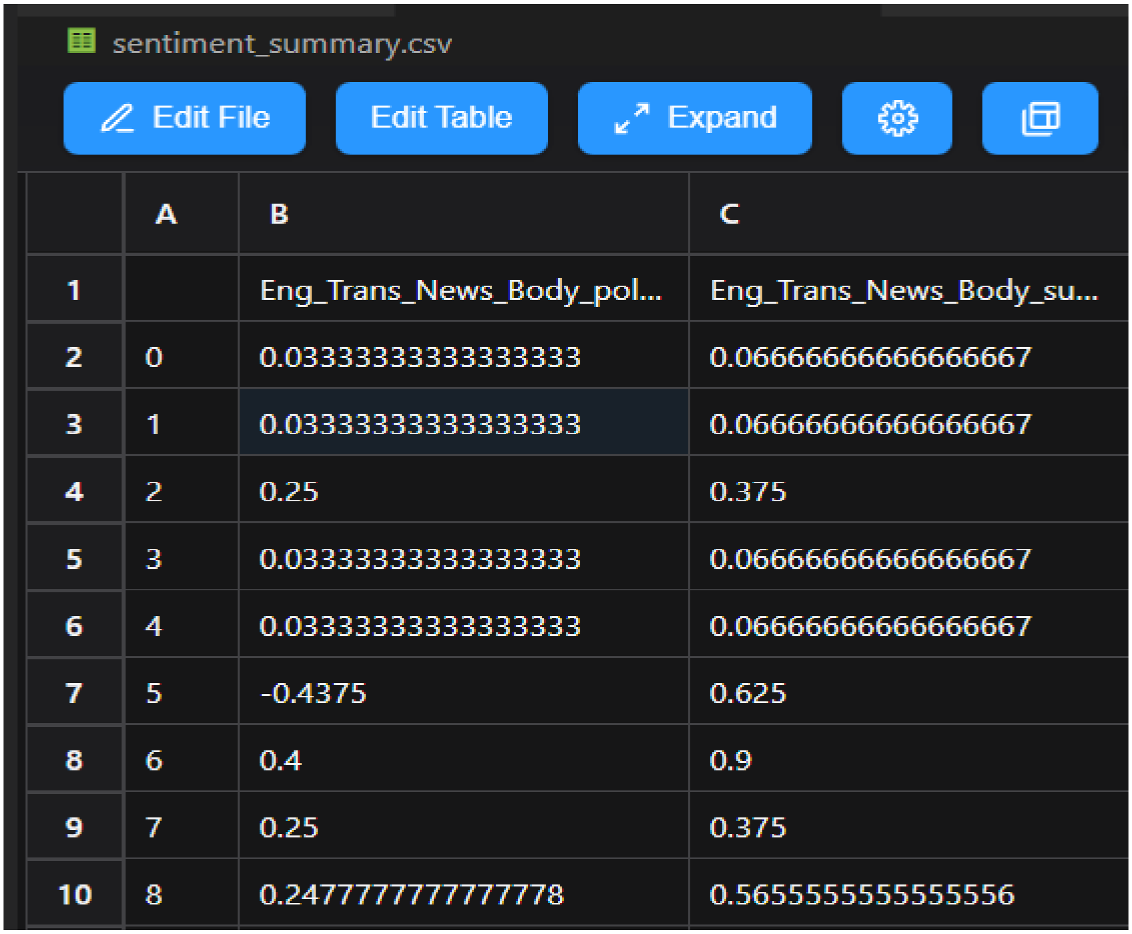


Fig 8: Word Cloud of News Content

Sentiment Analysis

Sentiment analysis was performed on English-translated news bodies using TextBlob (Fig. 9). The polarity and subjectivity scores were analyzed across different labels. It was observed that fake news articles tend to have higher emotional polarity and subjectivity compared to verified news.

Fig 9: Sentiment Distribution by Label



	A	B	C
1		Eng_Trans_News_Body_pol...	Eng_Trans_News_Body_su...
2	0	0.03333333333333333	0.06666666666666667
3	1	0.03333333333333333	0.06666666666666667
4	2	0.25	0.375
5	3	0.03333333333333333	0.06666666666666667
6	4	0.03333333333333333	0.06666666666666667
7	5	-0.4375	0.625
8	6	0.4	0.9
9	7	0.25	0.375
10	8	0.2477777777777778	0.5655555555555556

Correlation Analysis

A correlation heatmap was generated to study relationships between numerical features such as word count, sentiment scores, media indicators, and labels (Fig. 10). The analysis showed moderate correlation between text length, sentiment polarity, and misinformation labels, indicating their usefulness in classification.

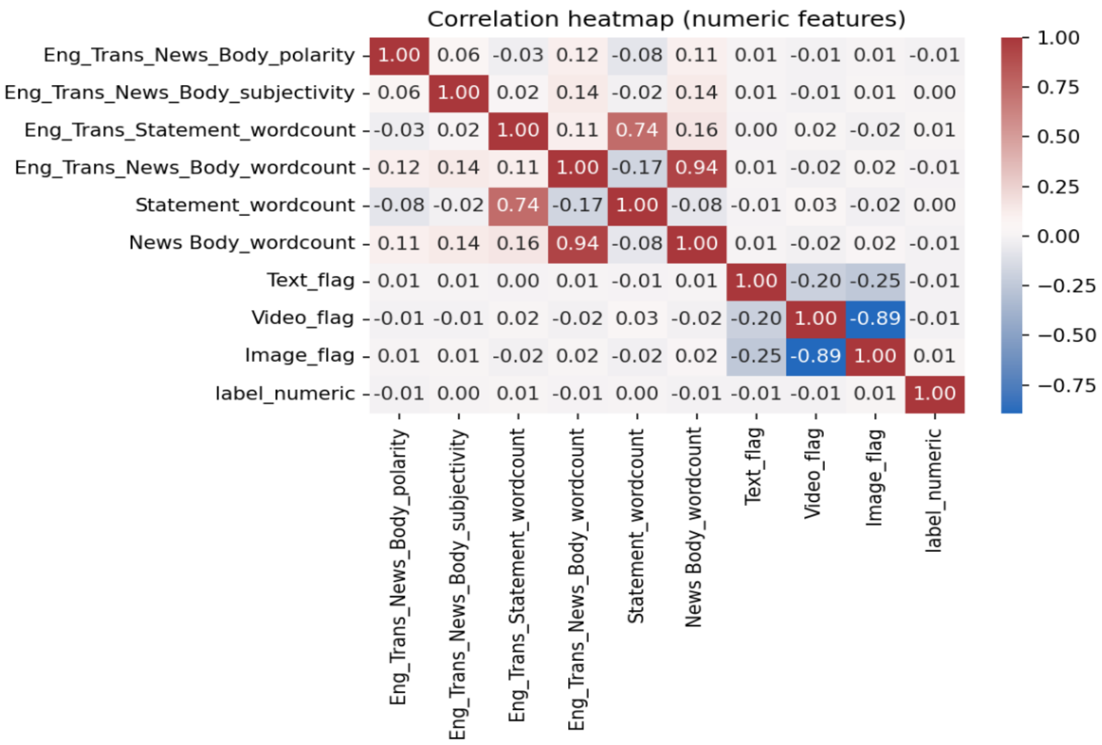
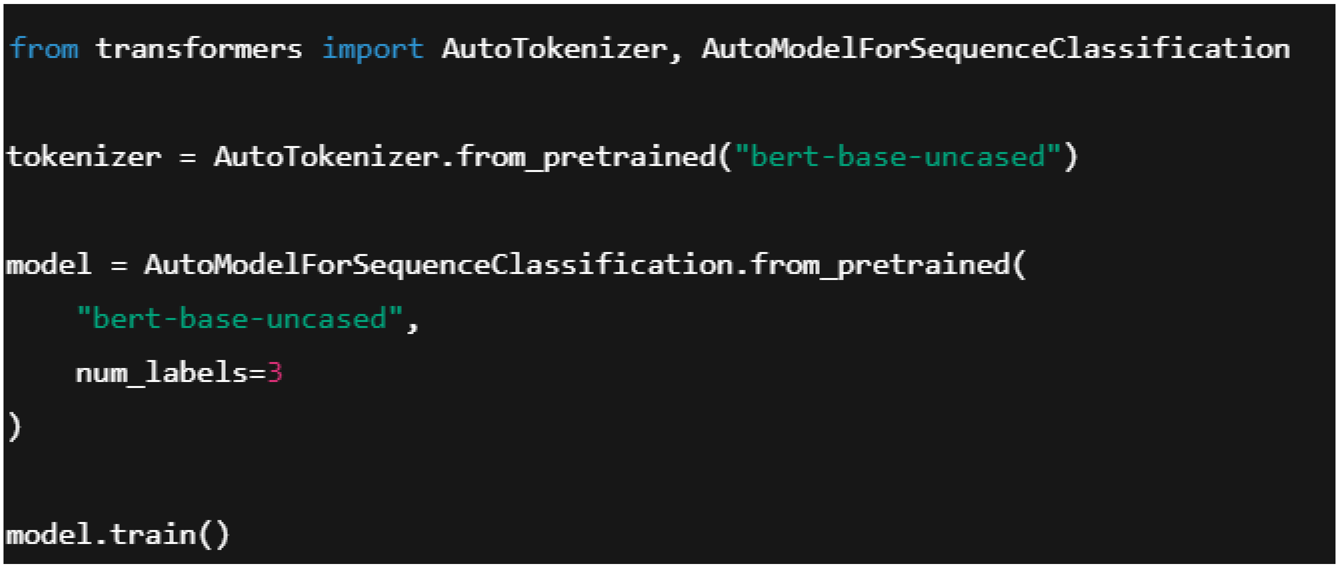


Fig 10: Correlation Heatmap of Numerical Features

4. Results and discussion:

Model Training and Evaluation

In this project, a BERT-based deep learning model was used for classifying news claims as True, False, or No Direct Evidence. The dataset was first divided into training and testing sets. The input text was tokenized using the BERT tokenizer, and necessary padding and truncation were applied to ensure uniform input size. The pre-trained bert-base-uncased model was fine-tuned on the fake news dataset to learn domain-specific patterns and contextual relationships. BERT was selected because of its strong ability to understand natural language context, which is essential for detecting misinformation.



```python
from transformers import AutoTokenizer, AutoModelForSequenceClassification

tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")

model = AutoModelForSequenceClassification.from_pretrained(
    "bert-base-uncased",
    num_labels=3
)

model.train()
```

Performance Evaluation

After training, the model was evaluated on the test dataset using standard classification metrics such as Accuracy, Precision, Recall, and F1-score. These metrics help in measuring how effectively the model distinguishes between fake, true, and unverifiable news. The trained model was used to generate predictions, and its performance was analyzed using the classification report. The results showed that the BERT-based model achieved high accuracy and reliable generalization, making it suitable for real-world fact-checking applications.

```
      Accuracy: 0.5914085914085914
      precision   recall  f1-score   support


       0      0.66      0.66      0.66      1209
       1      0.48      0.49      0.49       793


   accuracy                          0.59      2002
  macro avg       0.57      0.57      0.57      2002
weighted avg       0.59      0.59      0.59      2002

Confusion Matrix:
 [[796 413]
 [405 388]]
```

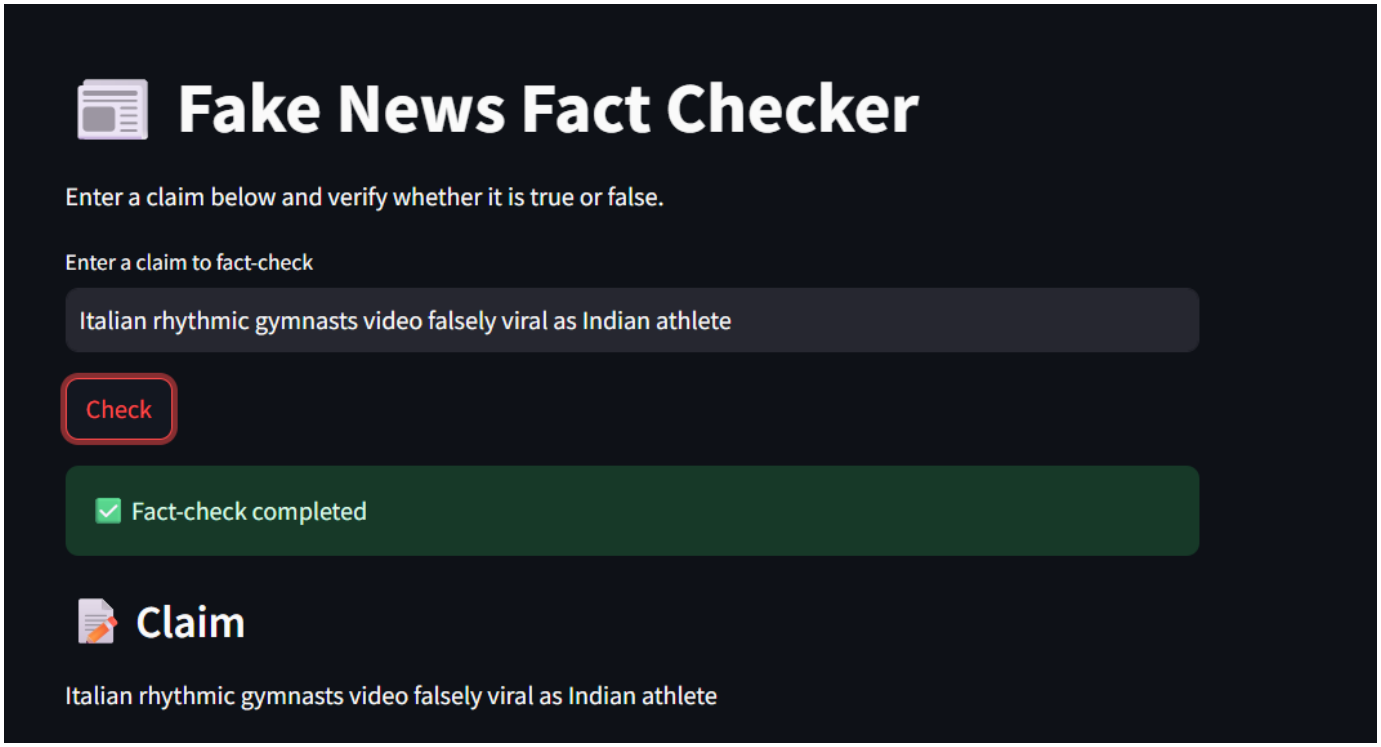
5. GUI:

Graphical User Interface Using Streamlit

The graphical user interface (GUI) of the proposed Fake News Detection System was developed using the Streamlit framework. Streamlit is an open-source Python library designed for building interactive and data-driven web applications with minimal effort. It enables rapid development of user interfaces directly from Python scripts without requiring extensive knowledge of front-end technologies.

In this project, Streamlit is used to provide a simple and user-friendly interface for interacting with the fake news detection system. Users can enter a news claim through a text input field and submit it for verification. The application then processes the input using the integrated BERT model and Retrieval-Augmented Generation (RAG) pipeline and displays the final verdict along with supporting evidence and explanations.

Streamlit supports automatic updating of application components, interactive widgets, and easy deployment. It also integrates seamlessly with machine learning and data analysis libraries such as PyTorch, Transformers, and Scikit-learn. Due to its simplicity, fast development cycle, and strong compatibility with AI models, Streamlit was selected as the frontend framework for implementing the user interface of this system.



Fake News Fact Checker

Enter a claim below and verify whether it is true or false.

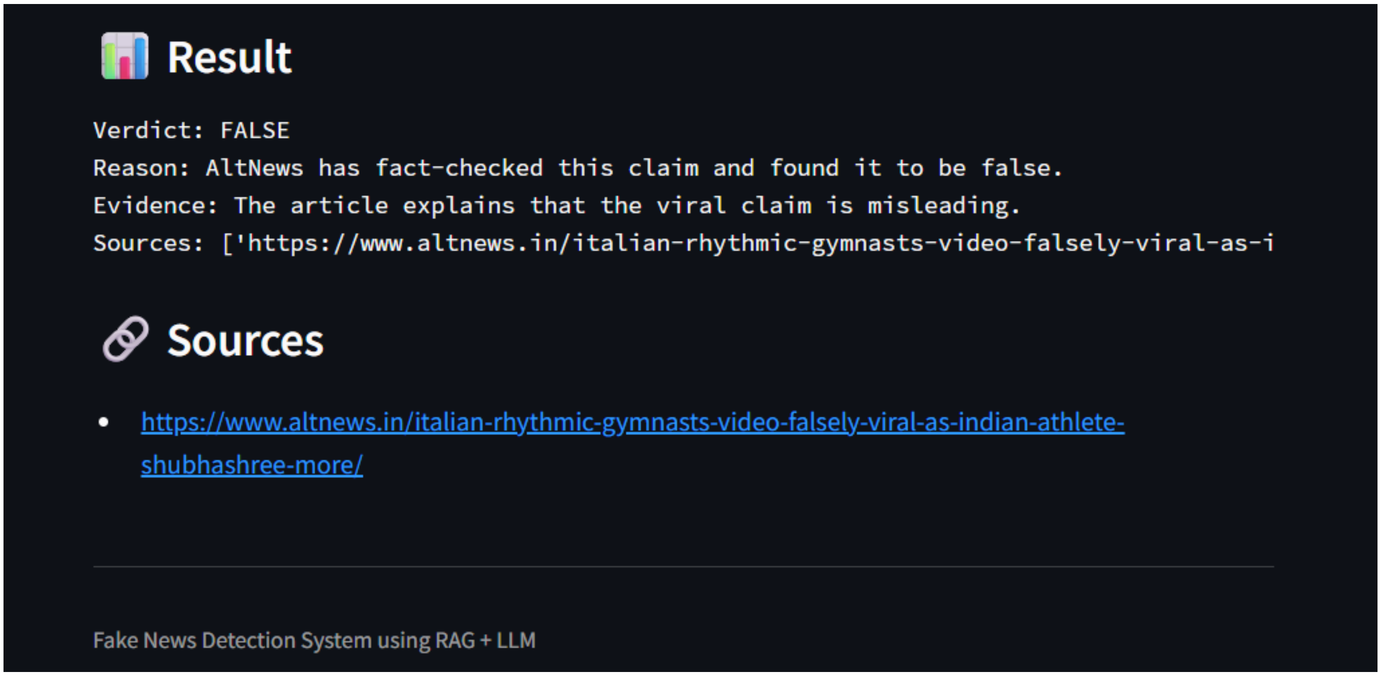
Enter a claim to fact-check

Italian rhythmic gymnasts video falsely viral as Indian athlete

✅ Fact-check completed

Claim

Italian rhythmic gymnasts video falsely viral as Indian athlete



```
Verdict: FALSE
Reason: AltNews has fact-checked this claim and found it to be false.
Evidence: The article explains that the viral claim is misleading.
Sources: ['https://www.altnews.in/italian-rhythmic-gymnasts-video-falsely-viral-as-i
```

- https://www.altnews.in/italian-rhythmic-gymnasts-video-falsely-viral-as-indian-athlete-shubhashree-more/

6.GitHubLink:

https://github.com/Saurabh09821/FAKE-NEWS-CLASSIFIER

7.Future work And Conclusion

7.1Future Work:

The proposed Fake News Detection System can be further enhanced in several ways to improve its performance and applicability. In future work, the system can be extended to include real-time monitoring of social media platforms such as Twitter, Facebook, and WhatsApp to detect misinformation at an early stage. Integration with more trusted fact-checking organizations and multilingual sources can also improve the coverage and reliability of evidence retrieval.

Advanced deep learning models such as RoBERTa, DeBERTa, and GPT-based architectures can be explored to improve classification accuracy. The system can also be optimized for faster processing by using model compression and caching techniques. In addition, mobile application support and browser extensions can be developed to make the system easily accessible to a wider audience. Incorporating user feedback mechanisms can further help in continuously improving the system's accuracy and trustworthiness.

7.2 Conclusion:

This project successfully developed an AI-based Fake News Detection System using Natural Language Processing, BERT-based classification, and Retrieval-Augmented Generation techniques. The system is capable of analyzing user-provided news claims, retrieving relevant evidence from reliable online sources, and generating accurate verification results.

The exploratory data analysis revealed that political and social issues dominate misinformation content and that most fake news is text-based. The BERT model demonstrated strong performance in understanding contextual meaning and classifying news effectively. The integration of the RAG pipeline enhanced the reliability of predictions by providing supporting evidence.

The Streamlit-based user interface enabled easy interaction and real-time verification for users. Overall, the proposed system reduces manual fact-checking effort and contributes to building a trustworthy information ecosystem. With further improvements and scalability enhancements, the system has strong potential for real-world deployment in combating misinformation.