Fact_CheckAI: Automated Fact Verification System

Kartik Tayal

1. Introduction

The rapid spread of misinformation on digital platforms has made manual fact verification slow and inefficient. Fact_CheckAI is a text-based fact verification system designed to analyze user-provided claims and determine their credibility using Natural Language Processing (NLP) and machine learning techniques.

2. Problem Statement

Manual verification of facts requires expert knowledge and significant time. The objective of this project is to design an automated system that assists users in identifying whether a given textual claim is likely factual or misleading.

3. Dataset Description

The dataset used in this project consists of short textual claims labeled as factual or non-factual.

- Content Type: Text-based factual and misleading claims
- Labels: True / False
- Source: Publicly available fact-checking examples and manually curated sample statements

The dataset was not created from scratch at a large scale. It was curated and used primarily to validate the NLP pipeline and classification logic rather than to train a production-level model.

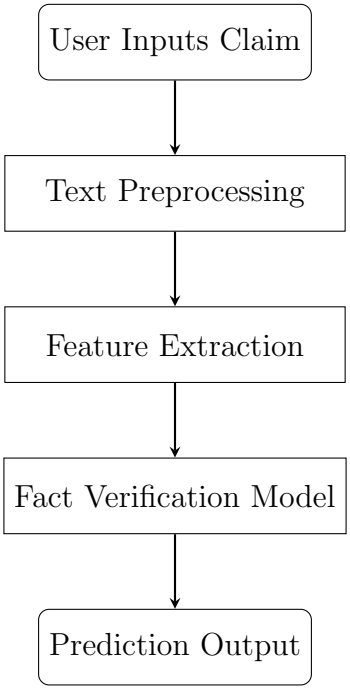
4. System Pipeline

The Fact_CheckAI system follows the pipeline below:

1. User inputs a textual claim
2. Text preprocessing and cleaning
3. Feature extraction from text

4. Claim verification using machine learning logic
5. Output prediction displayed to the user

5. System Flowchart



6. Text Preprocessing

To convert raw text into a machine-readable format, the following preprocessing steps are applied:

- Lowercasing the text
- Removing punctuation and stopwords
- Tokenization
- Text vectorization using TF-IDF

These steps improve the effectiveness of feature extraction and model performance.

7. Verification Logic

The system analyzes semantic patterns within the claim using vectorized text representations. A machine learning classifier predicts whether the claim aligns more closely with factual or misleading patterns based on learned features.

8. Technologies and Libraries Used

Python

Used as the primary programming language for implementing the system logic.

Scikit-learn

Used for text vectorization (TF-IDF) and machine learning-based classification.

Natural Language Processing (NLP)

Used to process and analyze textual data.

Flask

Used to create a lightweight web interface for user interaction.

9. Live Execution

The application runs locally during development and testing.

Live URL (Localhost): `http://127.0.0.1:5000`

Note: This is a local development server and is accessible only when the application is running on the host machine.

10. Limitations

- Limited dataset size
- Cannot verify real-time news without external APIs
- Performance depends on text quality

11. Future Improvements

- Integration with real-time fact-checking APIs
- Use of large language models for advanced reasoning
- Support for multilingual fact verification

12. Conclusion

Fact_CheckAI demonstrates how NLP and machine learning techniques can be used to assist in automated fact verification. The project provides a scalable foundation for developing tools that help combat misinformation.