SYNOPSIS

Report on

Fake News Detection using AI/ML

by

Sujal Gupta 2426mca189
Shubhranshu 2426mca708
Shivam Chaudhary 2426mca339
Tanya Singh 2426mca2503

Session:2024-2026 (III Semester)

Under the supervision of

Miss. Neelam Rawat
Miss. Shruti Aggrawal

KIET Group of Institutions, Delhi-NCR, Ghaziabad

DEPARTMENT OF COMPUTER APPLICATIONS
KIET GROUP OF INSTITUTIONS, DELHI-NCR,GHAZIABAD-201206

ABSTRACT

In the current digital era, the rapid spread of misinformation and fake news has become a critical challenge affecting individuals, organizations, and society at large. Fake news has the potential to manipulate opinions, influence elections, and spread panic during sensitive situations. Hence, automatic detection of fake news is an urgent requirement.

This project proposes a Fake News Detection System using Artificial Intelligence (AI) and Machine Learning (ML), leveraging state-of-the-art Natural Language Processing (NLP) models such as DistilBERT (a transformer-based model). The system is trained on publicly available datasets of true and fake news articles, with preprocessing steps like text cleaning, tokenization, and vectorization.

By employing deep learning and transformer-based architectures, the model learns semantic and contextual patterns in news articles to classify them as REAL or FAKE. The system further provides evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix for performance analysis.

This project aims to provide a reliable and automated solution for combating misinformation and improving the trustworthiness of digital information.

Keywords: Fake News, Machine Learning, NLP, Transformers, DistilBERT, Text Classification.

To
The Project Mentor
KIET Group of Institution

Subject - Request for Project Approval and Permission to Form Group with Students from Another Mentor.

Respected Mam

I hope you are doing well.
I am writing to request your approval for my project titled "Fake News detection" - and also to kindly grant permission to form my project group with students who are under the guidance of another mentor (Shruti Ma'am)

This arrangement will help us work collaboratively, combine our skills, and complete the project more effectively. We assure you that we will follow all departmental guidelines and keep you updated on our progress regularly.

Kindly consider my request and grant the necessary approval.

Thank you for your time and consideration.

Yours sincerely,
Tanya Singh
Sujal Gupta
[Mentor - Neelam Ma'am]

Group Members
Sujal Gupta
Shivam Chaudhary
Subhranshu
Tanya

Shruti

Allowed

(Neelam)

12/8/25

TABLE OF CONTENTS

Introduction

The exponential growth of digital platforms and online news distribution has transformed the way people consume information, but it has also given rise to one of the most pressing issues of the modern world—misinformation and fake news. Fake news spreads rapidly across social media and online portals, influencing public opinion, creating social unrest, impacting elections, damaging reputations, and severely reducing trust in reliable media outlets. Manual verification of news articles, while accurate, is slow, resource-intensive, and unsuitable given the massive scale of information generated every second, which highlights the urgent need for automated fake news detection systems. To address this global challenge, the proposed project focuses on developing a Deep Learning-based Fake News Detection System that leverages the power of Natural Language Processing (NLP) and state-of-the-art transformer architectures. The system is implemented using PyTorch as the backend deep learning framework and employs DistilBERT, a lightweight yet highly effective transformer model derived from BERT, which provides a balance between accuracy and computational efficiency.

The overall workflow of the system begins with data preprocessing where irrelevant tokens, special characters, and noise are removed to ensure clean input data, followed by tokenization using DistilBERT’s pretrained tokenizer that encodes textual content into numerical embeddings representing semantic meaning. These embeddings are then passed through a fine-tuned DistilBERT classification model, trained to predict whether a piece of news is real or fake. To evaluate the system’s effectiveness, multiple metrics such as accuracy, precision, recall, F1-score, and confusion matrix visualizations are used to analyze both the strengths and weaknesses of predictions in terms of error distribution. Unlike traditional machine learning methods, this approach benefits from the deep contextual understanding of language offered by transformer models, allowing it to recognize subtle cues in text that often distinguish misinformation from authentic reporting. The system is designed not only for research significance but also for real-world impact and scalability; it can be deployed as a standalone web application, mobile app, browser extension, or API service, making it accessible to readers, journalists, governments, and social media platforms. Such technology holds the potential to strengthen the fight against misinformation by providing instant, scalable, and user-friendly verification tools that help safeguard democracy, restore faith in media, and promote an informed society.

Furthermore, the project lays the groundwork for future advancements by extending into multimodal detection where text and images can be analyzed together, adapting the system for multilingual support, and incorporating credibility analysis of information sources. In summary, this project aims to create a robust and efficient intelligent platform that not only detects fake news with high accuracy but also contributes meaningfully to the broader mission of combating digital misinformation and its adverse consequences on global society.

Literature Review

Evolution of Fake News Detection Approaches

1. Early Approaches: Rule-Based and Keyword Matching

The earliest attempts at fake news detection relied heavily on manual keyword matching and rule-based systems. These systems identified deceptive or misleading content based on the presence of certain words, phrases, or heuristic rules predefined by experts. While straightforward, this approach suffered from major limitations. Specifically, it was not scalable given the rapidly growing volume of digital news, and it lacked contextual understanding. For instance, the same keyword could appear in both genuine and false contexts, leading to high false positives or negatives. These limitations highlighted the urgent need for more robust, automated techniques.

2. The Machine Learning Era

With the advancement of machine learning (ML) and its adoption in natural language processing (NLP), researchers began applying supervised algorithms to fake news detection. Classical models such as Naïve Bayes, Logistic Regression, and Support Vector Machines (SVMs) were extensively used. These models relied on textual features like bag-of-words (BoW), term frequency–inverse document frequency (TF-IDF), and n-grams, which quantified patterns of word usage. Compared to keyword-based methods, ML significantly improved detection performance by capturing statistical correlations between words and labels.

However, the major limitation was that these models processed language as independent tokens without understanding deeper semantic meaning or context, thus failing to generalize when fake news involved more nuanced writing styles.

3. Deep Learning Advances

The onset of the deep learning era transformed fake news detection. Researchers began employing algorithms such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs) to capture sequential patterns and hierarchical textual features. These architectures made it possible to model contextual dependencies in news articles — an essential requirement, since fake news often uses subtle cues spread across sentences.

The most significant leap came with transformer architectures. Models like BERT (Bidirectional Encoder Representations from Transformers) introduced deep bidirectional context learning, where every word is encoded with awareness of its surrounding words. BERT-based architectures achieved state-of-the-art performance in text classification, pushing accuracy and robustness to new heights compared to classical ML and earlier deep learning models.

4. Current Trends and Innovations

The field has now moved to transfer learning and fine-tuning of pre-trained, large-scale language models. Models such as BERT, RoBERTa, DistilBERT, and XLNet are widely applied to fake

news detection, as they leverage vast amounts of prior linguistic knowledge and require relatively little task-specific labeled data to deliver strong performance.
Additionally, researchers are increasingly exploring multimodal fake news detection, which combines textual signals with visual and contextual cues (e.g., images, videos, or metadata from social platforms). This trend addresses the fact that much of today's disinformation spreads through multimedia-rich platforms, making text-only detection insufficient.

Identified Gaps and Challenges
Despite significant progress, several gaps remain:

Real-time, scalable detection:
Current transformer-based solutions, while accurate, are computationally heavy. Real-time detection of fake news across large-scale dynamic streams (e.g., social media feeds) remains a significant challenge.

Detection of nuanced disinformation:
Sophisticated fake news articles exploit stylistic subtleties, satire, and implicit bias that even advanced models struggle with. Capturing intent, tone, and cross-document context continues to be difficult.

Explainability and trust:
Although deep learning and transformer models achieve high accuracy, they often function as black boxes. The growing field of Explainable AI (XAI) in fake news detection is still underdeveloped. Without transparency, it is difficult to convince policymakers, journalists, and the general public to trust automated fake news detectors.

Project Objective

1. To develop an AI/ML-powered fake news detection system that automatically classifies news articles as REAL or FAKE.

- The main goal of the project is to build a system that can read a news article and decide whether it is genuine or fake.
- This removes the need for manual fact-checking and speeds up misinformation detection.

2. To preprocess and clean textual data from fake and real news datasets.

- Raw news data often contains noise (punctuation, stopwords, HTML tags, irrelevant text).
- Preprocessing ensures the text is cleaned, tokenized, and normalized so that models can learn meaningful patterns.

3. To implement transformer-based NLP models (DistilBERT, BERT) for high accuracy.

- Traditional ML models (like Logistic Regression, Naïve Bayes) have limitations in understanding context.
- Transformers like **BERT and DistilBERT** capture deep contextual meaning of words.
- Using them helps achieve much higher classification accuracy.

4. To evaluate system performance using standard metrics (accuracy, precision, recall, F1-score).

- Accuracy alone is not enough — for fake news detection, we must also check **precision** (how many predicted fake are actually fake) and **recall** (how many fake news were caught).
- The F1-score balances both.
- This ensures a fair and reliable evaluation of the system.

5. To visualize results with confusion matrices and classification reports.

- Visual tools like **confusion matrices** help understand where the model makes mistakes (e.g., misclassifying real as fake).
- Reports provide detailed statistics for comparison of models.
- This makes evaluation more transparent.

6. To provide a scalable framework adaptable to large datasets and real-world applications.

- The final goal is not just to test on small datasets but to design a framework that can handle **large-scale real-world news streams** (like Twitter, online media).
- Scalability ensures the system can be extended for real-time fake news monitoring in practical scenarios.

Software Requirements

Frontend (User Interface)

Framework:

Next.js (React-based framework for SSR and CSR, providing SEO optimization and performance)

Styling:

Tailwind CSS / Material UI (for modern responsive design)

Features:

User-friendly interface to input news text or upload article links

Real-time results display (genuine vs fake with probability score)

Visualization components (charts/graphs for model confidence, explanation highlights)

Integration:

REST API/GraphQL to connect with backend model

Optional integration with Flask/Streamlit prototype (for demo or admin analysis)

Backend (Core Service Layer)

Core Language: Python 3.x

Frameworks & Libraries:

Data Handling & Analysis: Pandas, NumPy

ML/NLP: Scikit-learn, PyTorch, Hugging Face Transformers (BERT, RoBERTa, etc.)

Visualization: Matplotlib, Seaborn (for result analysis)

API Creation: FastAPI / Flask (for scalable REST endpoints to serve model predictions)

Key Functionalities:

Preprocessing news articles (tokenization, lemmatization, stop-word removal)

Model training & fine-tuning (traditional + transformer models)

Fake news detection service (prediction endpoint for frontend integration)

Confidence scoring & interpretability (e.g., using LIME/SHAP for explainability)
Database (Optional for Large-Scale System)
Options:
MongoDB (NoSQL) for storing user queries, articles, model logs
MySQL/PostgreSQL (Relational DB) for structured datasets and research experiments

Usage:
Store datasets for training/validation
Keep track of user queries, detection results, and system logs
Provide history and analytics for monitoring trends
Deployment & Platform

Development Environments:

Google Colab / Jupyter Notebook (for research, prototyping, initial model training)
Local GPU-enabled workstation (if available, for fine-tuning and heavy training)
Deployment:

Frontend:
Deployed on Vercel (seamless hosting for Next.js frontend)

Backend API:
Option 1: Deploy on Vercel serverless functions (small-scale use case)
Option 2: Deploy as Dockerized API (Flask/FastAPI) on AWS/GCP/Heroku for scalable serving

CI/CD Integration:
GitHub Actions / Vercel's CI for automatic deployment on push

PROJECT FLOW

1. **Data Collection & Preprocessing**
- Load datasets (Fake.csv, True.csv).
- Clean and merge data, remove stopwords, tokenize, and label (0 = Fake, 1 = Real).

2. **Text Representation**
- Use DistilBERT Tokenizer for converting text into embeddings.

3. **Model Training**
- Train a **DistilBERT sequence classification model**.
- Fine-tune hyperparameters (epochs, batch size).

4. **Evaluation**
- Evaluate with classification report, accuracy, precision, recall, F1-score.
- Visualize confusion matrix.

5. **Prediction System**
- Implement function to test user-input news articles.
- Return label as **FAKE or REAL.**

PROJECT OUTCOMES

The outcome of this project is the development of a fully functional AI-powered fake news detection system that can automatically classify news articles as either real or fake with a high degree of accuracy. Unlike traditional keyword-based or rule-driven approaches, the system leverages the power of advanced natural language processing (NLP) through DistilBERT, a state-of-the-art transformer-based model known for its efficiency and contextual understanding. By analyzing textual content at a deeper semantic level, rather than merely focusing on surface-level word frequencies, the system achieves robust and reliable detection performance, making it more resilient against subtle and sophisticated disinformation attempts. To validate the effectiveness of the model, the outcomes are thoroughly evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score, which provide a quantitative measure of success across different dimensions of classification.

Furthermore, insights into model behavior are supported with graphical visualizations like confusion matrices and classification reports, enabling a clearer understanding of how well the model distinguishes between real and fake news as well as highlighting potential weaknesses where misclassifications occur. Beyond its performance, a major achievement of this work lies in the scalability and adaptability of the designed framework. The system is implemented to support real-time processing, making it not only academically significant for research purposes but also practically viable for deployment within online platforms. This adaptability positions it as a meaningful tool for media organizations seeking fact-checking automation, researchers analyzing information ecosystems, or end-users aiming to verify the authenticity of online content. In essence, the project bridges the gap between research and real-world applicability by delivering a solution that is accurate, scalable, user-friendly, and capable of addressing the pressing global challenge of misinformation in the digital age.

PROPOSED TIME DURATION

- **Phase 1 (Weeks 1–2):** Literature survey, problem formulation, and dataset collection.
- **Phase 2 (Weeks 3–4):** Data preprocessing, cleaning, and feature engineering.
- **Phase 3 (Weeks 5–8):** Model training, hyperparameter tuning, and fine-tuning transformer models.
- **Phase 4 (Weeks 9–10):** Model evaluation using metrics (Accuracy, Precision, Recall, F1-score) and visualization (Confusion Matrix, heatmaps).
- **Phase 5 (Weeks 11–12):** System integration, deployment preparation, testing, and final documentation.

Total Duration: ~12 weeks (approx. 3 months).

REFERENCES

- Devlin, J. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL.

- Sanh, V. et al. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. Hugging Face.

- Shu, K. et al. (2020). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD.

- Zhang, X. & Ghorbani, A. (2020). An Overview of Online Fake News: Characterization, Detection, and Discussion. Information Processing & Managem