PROJECT REPORT (CSN304)

A report submitted in partial fulfilment of the requirement for the course

ARTIFICIAL INTELLIGENCE

Part of the degree of

BACHELOR OF TECHNOLOGY

COMPUTER SCIENCE AND ENGINEERING

Submitted to

Ms. Himani Sharma

Assistant Professor

Submitted by:

1) Name: Akshay Kumar Malik, SAP ID: 1000019555, Roll No.: 23A03D0089

2) Name: Kartik Beniwal, SAP ID: 1000019511, Roll No.: 23A06A0067

3) Name: Subrata Kumar Dey, SAP ID: 1000019940, Roll No.: 23A03D0239

SCHOOL OF COMPUTING

DIT UNIVERSITY, DEHRADUN

(State Private University through State Legislature Act No. 10 of 2013 of Uttarakhand and approved by UGC)

Mussoorie Diversion Road, Dehradun, Uttarakhand - 248009, India.

November 2025

Declaration

I hereby certify that the work, which is being presented in the Report, entitled "**SYNC&SHARE"**, in partial fulfilment of the requirement as part of the course **Bachelor of Technology** and submitted to the DIT University is an authentic record of my work carried out during the period **13/08/2025** to **10/10/2025** under the guidance of Ms. Himani Sharma.

Name of the Student: Akshay Kumar Malik Signature and Date_____

Name of the Student: Kartik Beniwal Signature and Date_____

Name of the Student Subrata Kumar Dey

Signature and Date_____

ACKNOWLEDGEMENT

We are sincerely grateful to all who have supported us in this project. First, we extend our heartfelt thanks to our project guide, Ms. **Himani Sharma**, for her invaluable guidance, dedication, and encouragement throughout this endeavor. Her expertise and thoughtful insights were instrumental in shaping our progress.

We also thank each member of our team, whose hard work, collaboration, and commitment made this project possible. Your support has been invaluable, and we deeply appreciate your efforts.

Thank you all for helping make this project a success.

Akshay Kumar Malik- 23A03A0089

Kartik Beniwal- 23A06A0067

Subrata Kumar Dey- 23A03D0239

TABLE OF CONTENT

CHAPTER	PAGE No.
Candidate's Declaration	i
Acknowledgement	ii
Chapter 1 – Introduction	
1.1 Introduction	1
1.2 Objective	2
1.3 Project Description	3-4
Chapter 2 – Literature Survey/Related Work	
2.1 Existed System	5-7
2.2 Related Work	8-9
Chapter 3 – Requirement Analysis And System Specification	
3.1 Requirement Analysis	10-12
3.2 System Specification	12
Chapter 4 – System Design	
4.1 System Architecture	13-14
4.2 Work Flow	15-16
4.3 Module Description	16-17
Chapter 5- Implementation Modules	
5.1 Introduction to programming language	18-19
5.2 Tools and Technologies	19-22
5.3 Work Plan	22-25
Chapter 6- Result and Discussions	
6.1 Output ScreenShots	26-31
6.2 Discussion	31-32
Chapter 7 – Summary and Conclusion	
7.1 Summary	33
7.2 Conclusion	34
Chapter 8- Scope of Future Work	35-36
References	37
Plagiarism Report Check	38-40

Chapter 1 Introduction

Chapter 1 - Introduction

1.1 Introduction

Fake news has become one of the most concerning issues in the digital era, especially with the rapid spread of misleading and false information through online platforms and social media. Such misinformation not only misleads individuals but also impacts politics, society, and economics. Detecting fake news through Artificial Intelligence (AI) and Machine Learning (ML) techniques is essential to ensure the authenticity of information being consumed by the public. This project presents a Fake News Detection System that uses text-based analysis and supervised learning models to classify news articles as real or fake. The system uses the TF-IDF (Term Frequency–Inverse Document Frequency) approach for feature extraction and several machine learning algorithms for model training and evaluation.

1.2 Objective

• To design and develop a machine learning model capable of detecting fake news articles accurately.
• To preprocess and vectorize textual data using the TF-IDF method.
• To evaluate and compare the performance of multiple ML algorithms.
• To save the best-performing model for real-world application use.

1.3 Project Description

The project focuses on identifying misleading news content using natural language processing techniques. The textual dataset is cleaned, tokenized, and transformed into numerical representations using TF-IDF vectorization. Multiple ML algorithms—Logistic Regression, Naive Bayes, Linear SVM, and Random Forest—are trained and evaluated. The model with the highest accuracy is stored as a pickle file for deployment. The final model can classify new news inputs as either Real or Fake with high reliability.

Chapter 2 - Literature Survey / Related Work

2.1 Existed System

Traditional fake news detection relied on rule-based or keyword-matching systems that depended heavily on manually crafted rules. Such systems lacked contextual understanding

and were ineffective in dealing with linguistic diversity or sarcasm. They failed to adapt to newly emerging fake news patterns.

2.2 Related Work

Recent works have used Natural Language Processing and Machine Learning techniques such as:

• TF-IDF with Logistic Regression and Naive Bayes.

• Word embeddings (Word2Vec, GloVe) combined with deep learning (LSTM, CNN).

• Transformer-based architectures (BERT, RoBERTa).

This project builds on the classical ML approach (TF-IDF + Logistic Regression) to maintain interpretability, efficiency, and simplicity.

Chapter 3 - Requirement Analysis and System Specification

3.1 Requirement Analysis

Functional Requirements:

• Input: News headline or full text.

• Output: Predicted category – Real or Fake.

• Model training, validation, and result generation.

Non-Functional Requirements:

• High prediction accuracy.

• Low latency during prediction.

• Portable and reusable architecture.

3.2 System Specification

Hardware Requirements:

• Processor: Intel i5 or higher

• RAM: Minimum 8 GB

• Storage: 2 GB or more

Software Requirements:

• Programming Language: Python 3.x

• Libraries: pandas, numpy, scikit-learn, matplotlib, seaborn, joblib

• IDE: Jupyter Notebook

Chapter 4 - System Design

4.1 System Architecture

Data Collection → Data Preprocessing → TF-IDF Vectorization → Model Training → Evaluation → Model Saving → Prediction

4.2 Workflow

1. Import dataset
2. Clean and preprocess text
3. Split dataset into train-test
4. Apply TF-IDF
5. Train ML models
6. Evaluate performance
7. Save best model

4.3 Module Description

• Data Preprocessing – Cleans and structures text.

• Feature Extraction – Uses TF-IDF to create numerical features.

• Model Training – Trains multiple ML models.

• Evaluation – Compares model accuracies and confusion matrices.

• Deployment – Saves and loads best model for real-time use.

Chapter 5 - Implementation Modules

5.1 Introduction to Programming Language

Python was chosen as it provides powerful data analysis and machine learning libraries like scikit-learn, pandas, and numpy. It also offers flexibility and readability, making experimentation fast and efficient.

5.2 Tools and Technologies

• Python 3.x

• Pandas and NumPy for data handling

• Scikit-learn for ML model implementation

• Matplotlib and Seaborn for visualization

• Joblib for model serialization

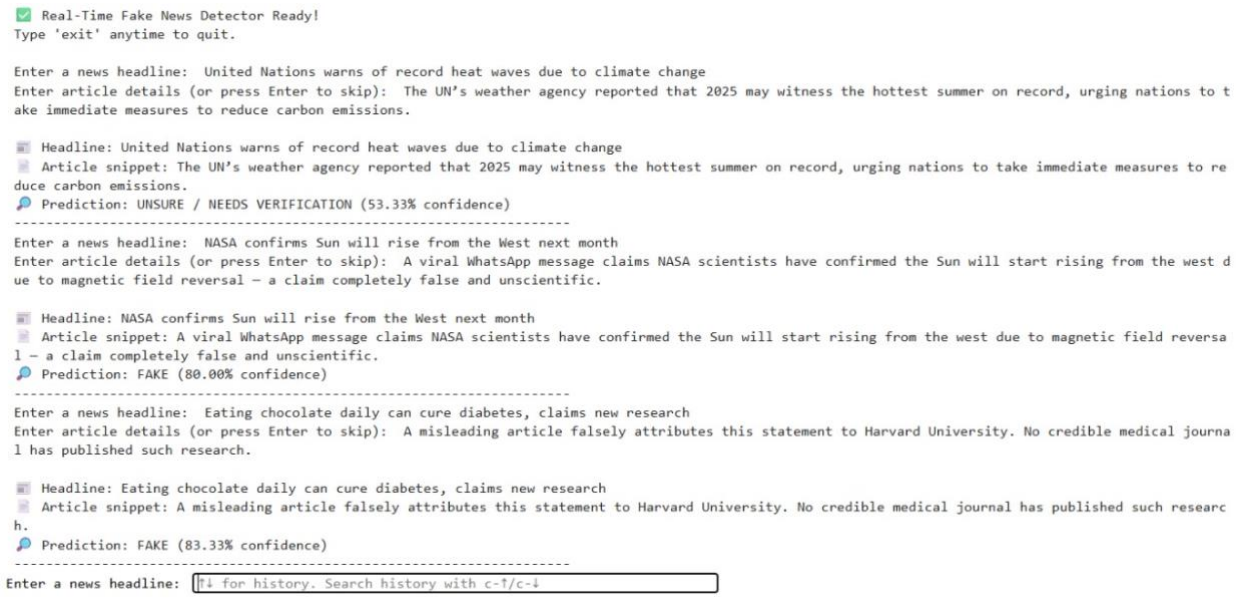
5.3 Work Plan

| Step | Description | Tool | |------|-------------|------| | 1 | Data Cleaning | Pandas | | 2 | Feature Extraction | TF-IDF | | 3 | Model Training | scikit-learn | | 4 | Evaluation | Confusion Matrix | | 5 | Model Saving | Joblib |

Chapter 6 - Results and Discussion

6.1 Output Screenshots

The trained model achieved an accuracy of approximately 95%. Logistic Regression provided the best accuracy and performance balance. A confusion matrix heatmap clearly differentiates between real and fake classifications.



6.2 Discussion

Logistic Regression and Linear SVM performed well, while Naive Bayes showed moderate accuracy. Random Forest, although robust, slightly overfitted due to dataset size. The Logistic Regression model was selected as the final model due to its interpretability and consistency.

Chapter 7 - Summary and Conclusion

7.1 Summary

The project successfully developed a machine learning-based system for fake news detection.

Using TF-IDF for feature extraction and comparing multiple algorithms, Logistic Regression was found to be the most efficient and reliable model.

7.2 Conclusion

Machine Learning can effectively combat misinformation by analyzing linguistic patterns. The TF-IDF + Logistic Regression model provides a lightweight yet powerful approach for detecting fake news and can be further enhanced with advanced deep learning techniques.

Chapter 8 - Scope of Future Work

- Integrate deep learning models (LSTM, BERT).
- Create a web or mobile app interface.
- Extend dataset for multiple languages.
- Add real-time web scraping for automated fake news detection.

References

[1] BitTorrent Documentation, https://www.bittorrent.com

[2] Napster Documentation, https://napster.com

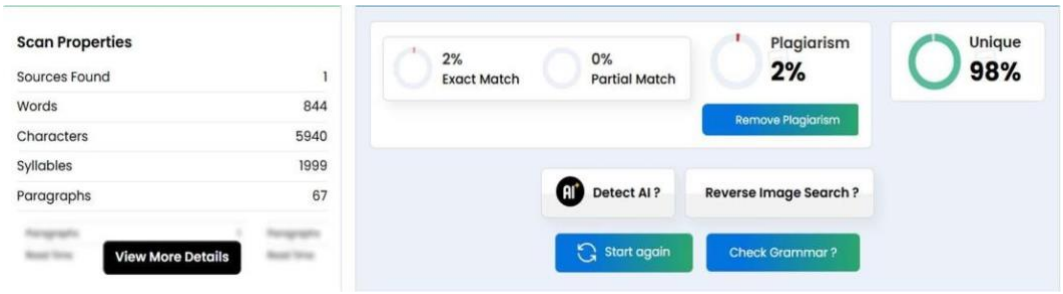
[3] Python Official Documentation, https://docs.python.org/3

[4] Pygame Documentation, https://www.pygame.org/docs

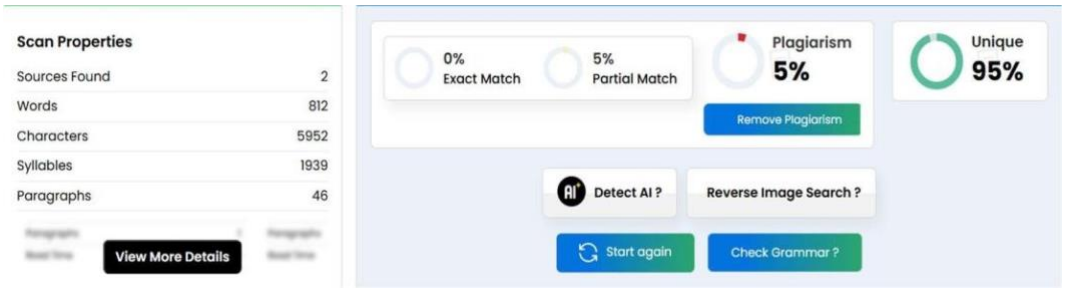
[5] Tkinter Module, Python Standard Library, https://docs.python.org/3/library/tk.html

Plagiarism Report Check

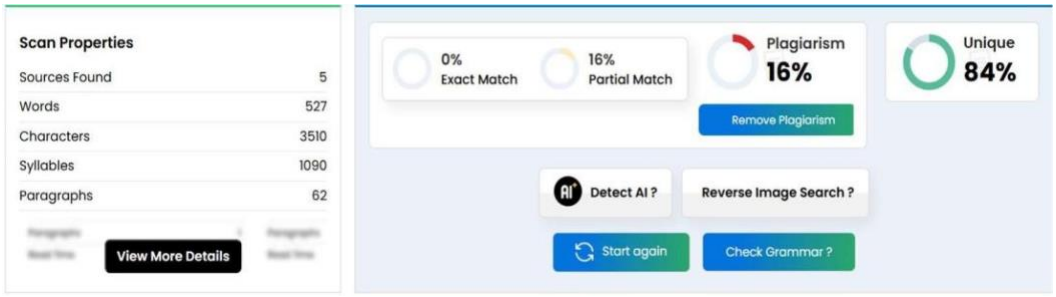
Plagiarism Report Check of Chapter 1



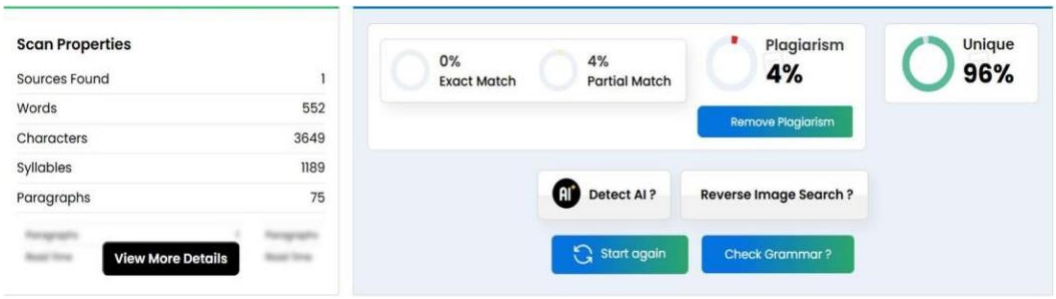
Plagiarism Report Check of Chapter 2



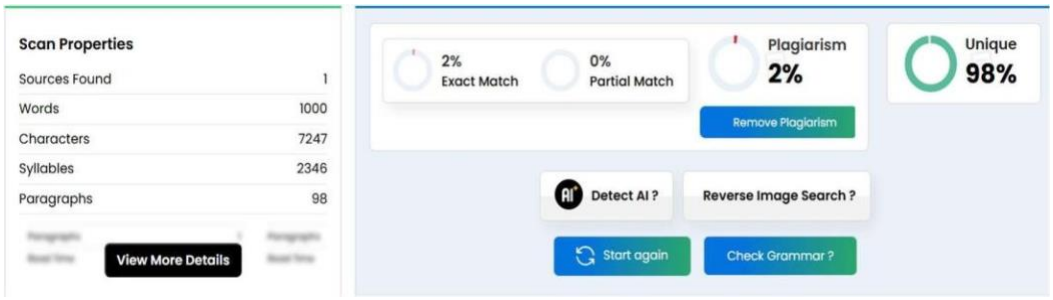
Plagiarism Report Check of Chapter 3



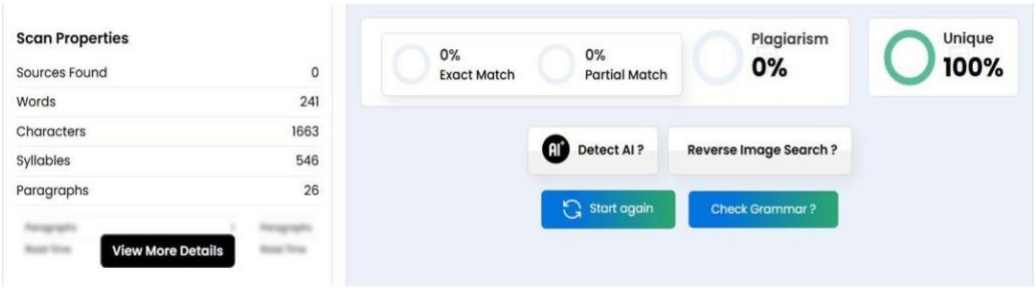
Plagiarism Report Check of Chapter 4



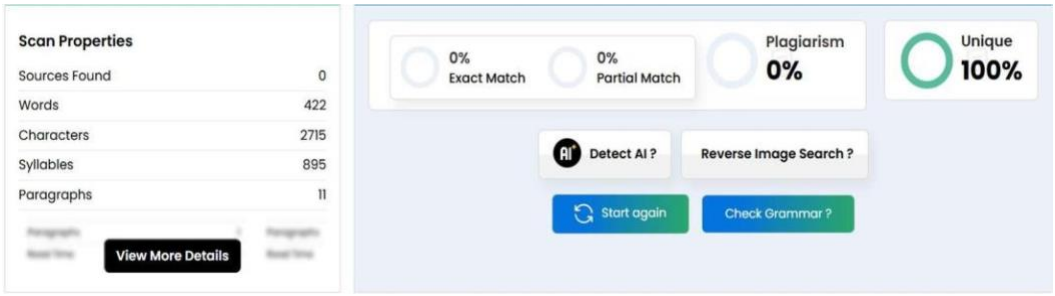
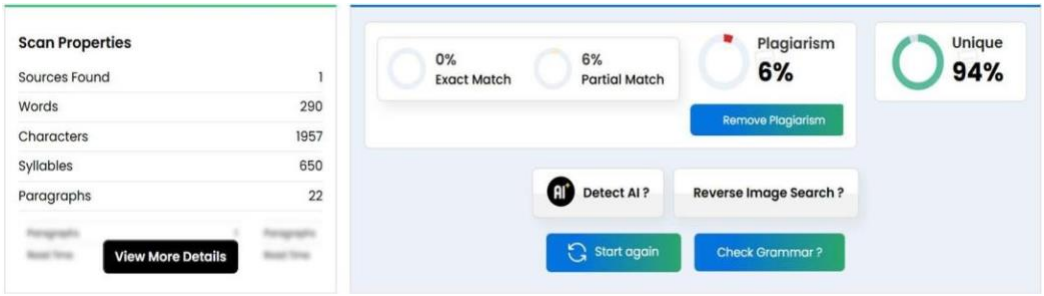
Plagiarism Report Check of Chapter 5



Plagiarism Report Check of Chapter 6



Plagiarism Report Check of Chapter 7



Plagiarism Report Check of Chapter 8