A PRELIMENERY REPORT ON

AI CONTENT DETECTION

SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF TECHNOLOGY
(ARTIFICIAL INTELLIGENCE)

SUBMITTED BY

DEVANG PAGARE	Exam No:2020AAIT1108093
ESHAN KASLIWAL	Exam No: 2020AAIT1108091
PRATIK JADE	Exam No: 2020AAIT1111019
DEVASHRI BHOSALE	Exam No: 2020AAIT1101007

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

G H RAISONI COLLEGE OF ENGINEERING AND MANAGEMENT
WAGHOLI, PUNE 412207

AN AUTONOMOUS INSTITUTE, AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY

2023-24

RAISONI GROUP
— a vision beyond —

CERTIFICATE

This is to certify that the project report entitles

"AI CONTENT DETECTION"

Submitted by

DEVANG PAGARE	Exam No: 2020AAIT1108093
ESHAN KASLIWAL	Exam No: 2020AAIT1108091
PRATIK JADE	Exam No: 2020AAIT1111019
DEVASHRI BHOSALE	Exam No: 2020AAIT1101007

are bonafide students of this institute and the work has been carried out by them under the supervision of **Prof. Rachna Sable** and it is approved for the partial fulfillment of the requirement of An Empowered Autonomous Institute, Affiliated to Savitribai Phule Pune University, for the award of the degree of **Bachelor of Technology** (Artificial Intelligence).

Prof. R.Y. Sable
Guide & H.O.D.
Department of Artificial Intelligence

External Examiner

Dr. R. D. Kharadkar
Director
GHRCEM, Pune

Place: Pune

Date:

ACKNOWLEDGEMENT

It is our privilege to acknowledge with deep sense of gratitude towards our project guide, **Prof. Rachna Sable** for her valuable suggestions and guidance throughout course of study and timely help given in the progress of my dissertation on "**AI CONTENT DETECTION**".

It is needed a great moment of immense satisfaction to express out profound gratitude towards our H.O.D **Prof. Rachana Sable**, whose real enthusiasm was a source of inspiration for our work. Our special thanks to **Dr. R. D. Kharadkar**, Director, G H Raisoni college of Engineering and Management, Pune for his valuable support.

We would also like to thank all other faculty members of Artificial Intelligence Department who directly or indirectly kept the enthusiasm and momentum required to keep the work towards an effective dissertation alive in us and guided in their own capacities in all possible ways.

Devang Pagare (18)
Eshan Kasliwal (22)
Pratik Jade (70)
Devashri Bhosale (19)

ABSTRACT

The rise of AI-generated data, mainly from models like ChatGPT, LLAMA2 poses serious difficulties to academic integrity and raises worries about plagiarism. The current research looks at the competencies of various AI content recognition algorithms to distinguish between human and AI-authored material.

This research looks at numerous research papers, publication years, datasets, machine learning approaches, and the benefits and drawbacks of detection methods in AI text detection. Various datasets and machine learning techniques are employed, with various types of classifiers emerging as top performers.

This work creates an Extra tree classifier that can distinguish ChatGPT-produced text from human-authored content. The "ChatGPT Paraphrase" dataset was used for model training and testing. The result shows that the proposed model resulted in 80.1% accuracy and outperformed the existing models namely Linear Regression (LR), Support Vector Machine (SVM), Decision Tree, (DT), K-Nearest Neighbour (KNN), Ada Boost Classifier (ABC), Random Forest Classifier (RFC), Bagging Classifier (BG), Gradient Boosting Classifier (GBC).
.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	ILLUSTRATION
LLM	Large Language Model
TF-IDF	Term Frequency - Inverse Document Frequency
ROC	Receiver Operating Characteristic Curve

TABLE OF CONTENTS

1. INTRODUCTION

1.1 OVERVIEW

Advanced language models such as ChatGPT and LLAMA2 have the capability to generate text that is not only coherent but also mirrors the realism of human writing. These models are versatile and can be utilized for a range of applications including but not limited to paraphrasing, summarizing, and question-answering. Despite their benefits, these AI models present certain challenges. One significant issue is their potential to disrupt academic integrity and content authenticity and complicate plagiarism detection.

Recognizing these challenges, this project aims to devise effective strategies to differentiate between text written by humans and that generated by ChatGPT. The cornerstone of this project is the development of a machine-learning model that leverages the Extra Trees Classifier. This model has been designed to accurately predict the origin of a piece of text, determining whether it's human-written or produced by ChatGPT.

The implications of this project are far-reaching. It will play a crucial role in preserving academic integrity and ensuring content authenticity by providing a reliable tool for content moderators and educators. This tool will enable them to verify the originality of the text, thereby fostering responsible and ethical practices in AI-generated content creation.

1.2 MOTIVATION OF THE PROJECT

In an era marked by the pervasiveness of digital text, the capacity to differentiate between individuals or human-generated data and AI-generated data has become an increasingly important issue. The evolution in Artificial Intelligence (AI) technology, represented by sophisticated language models like ChatGPT, has heralded a new era of information distribution in which text created by algorithms like GPT- 3.5 may closely replicate human language.

This situation raises critical queries about the authenticity, credibility, and ethical implications of the text that permeates our digital landscape. The fast growth of artificial intelligence has led to a growing need for tools to identify the source of content. By identifying the distinctive characteristics of AI and human-generated content, this project aims to provide a tool for identifying misinformation, authenticity, intellectual property protection, responsible AI use, user empowerment, bridging the gap between AI and humans, ethical AI development, and meeting societal needs.

1.3 PROBLEM DEFINITION

The evolution in Artificial Intelligence (AI) technology, represented by sophisticated language models like ChatGPT, has heralded a new era of information distribution in which text created by algorithms like GPT- 3.5 may closely replicate human language. This situation raises critical queries about the authenticity, credibility, and ethical implications of the text that permeates our digital landscape. The fast growth of artificial intelligence has led to a growing need for tools to identify the source of content.

1.4 PROJECT SCOPE AND LIMITATIONS

1.4.1 PROJECT SCOPE

This project paved the way for improved AI content detection tools and has significant implications for applications in fields such as plagiarism detection and content moderation. We test 10 machine learning classifiers with different hyperparameters. Out of all 10 classifiers, Extra Tree Classifier outperformed all the classifiers.

1.4.2 LIMITATIONS

- The Current classifier only works on text data.
- Well-known classifiers such as KNN are performing poorly on available datasets.

1.5 METHODOLOGIES OF PROBLEM-SOLVING

The project's main goal is to analyze Large Language Models (LLMs) with a focus on text datasets, features, evaluation metrics, and classifiers used to differentiate between human-generated and generative model data. A novel approach, the Extra Tree classifier, is implemented to distinguish data generated by ChatGPT from human-authored content. The

project involves acquiring a dataset that comprises 51.68% human-generated text and 48.31% ChatGPT-generated text. Extensive data preprocessing and model development are carried out, with a primary emphasis on accuracy as the key evaluation metric, supplemented by the F1 score. The project relies on a valuable paraphrase dataset generated by ChatGPT, obtained from Kaggle, facilitating an in-depth exploration of paraphrasing and its applications.

- Module 1 - Data Collection: In this module, the dataset used is called "ChatGPT Paraphrases," obtained from Kaggle. The dataset comprises 419,197 rows and contains four columns. Among its text samples, approximately 51.68% are generated by humans, with the remaining 48.31% generated by ChatGPT.

- Module 2 - Data Pre-processing: This module is dedicated to data pre-processing. There are two important steps within it: TF-IDF word vectorization and transforming text data into numerical representations using the TF-IDF technique. Dataset balancing: Addressing class imbalance within the dataset, ensuring equitable representation of human-generated and ChatGPT-generated text.

- Module 3 - Data Split module: In this module, data is partitioned into training and testing sets using Scikit-Learn's train_test_split function. 80 % of the entire data is utilized for training and the rest 20 % for testing. This split is a critical step in preparing the data for model training and evaluation.

- Module 4 - Model Selection: In this module, experiments are conducted with ten different machine-learning models on the dataset. These models include LR, DT, KNN, SVM, RFC, ABC, BG, GBC, and Extra Tree Classifier models.

- Module 5 - Evaluation: In this module, the selected model's performance is evaluated. Assessment is based on the F1 score, accuracy, and confusion matrix, providing critical insights into the model's capability to differentiate between people or human-generated and ChatGPT-generated data.

- Module 6 - Testing: Module 6 represents the final testing phase. Here, the selected models are applied to new, unseen data to gauge their real-world performance.

- Module 7 – Final Prediction User input is taken and the final prediction is done classifying whether the data is Human or AI-generated.

2. LITERATURE SURVEY

TABLE 2.1: LITERATURE SURVEY

Sr. No.	Name of Paper	Publication	Year of Publications	Datasets Used
1	ChatGPT or Human? Detect and Explain	arXiv	30-01-2023	Public generated restaurant review and ChatGPT generated restaurant review
2	Distinguishing Human-Written and ChatGPT-Generated Text Using Machine Learning	IEEE	27-04-2023	Not used
3	Automatic Detection of Machine Generated Texts: Need More Tokens	IEEE	23-09-2022	Open dataset for Russian language with long texts
4	Discrimination of human-written and human and machine written sentences using text consistency	IEEE	19-02-2021	Not used
5	Distinguishing Human Generated Text From ChatGPT Generated Text Using Machine Learning	arXiv	26-05-2023	Kaggle dataset of 10000 texts (5204 - human generated text, 4796- Al generated text)
6	DetectGPT-Zero-Shot Machine-Generated Text Detection using Probability Curvature	arXiv	23-01-2023	For real data, 500 news articles from the Xsun dataset.
7	Testing of Detection	arXiv	21-07-2023	Human generated

	Tools for AI-generated Text			pilot study data and Al generated data.
8	Will ChatGPT get you caught- Rethinking of Plagiarism Detection	Springer	15-07-2023	ChatGPT written 500 words essay.
9	AI-generated research paper fabrication and plagiarism in the scientific community	CellPress	2023	Not used
10	Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text	International Journal for Educational Integrity	2023	100 words Al generated responses for given prompts
11	One-Class Learning for AI-Generated Essay detection	Applied Sciences (Switzerland)	2023	Al Generated Essays
12	ChatGPT and Academic Integrity Concerns: Detecting Artificial Intelligence Generated Content	LET Journal	2023	Not used
13	Differentiating Chat Generative Pretrained Transformer from Humans: Detecting ChatGPT-Generated Text and Human Text Using Machine Learning	MDPI	2023	Not used
14	AI vs. Human? Differentiation Analysis of Scientific Content Generation	arXiv	2023	Human-generated abstracts and Al-generated abstracts.

3. SOFTWARE REQUIREMENTS SPECIFICATION

3.1 SYSTEM REQUIREMENTS SPECIFICATION:

3.1.1 SOFTWARE REQUIREMENTS (PLATFORM CHOICE)

- Operating system - Windows 10 or higher, Mac, Linux

3.1.2 HARDWARE REQUIREMENTS

- Processor: Ryzen 5 5600G
- Ram 8 GB
- Active internet connection is needed

4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

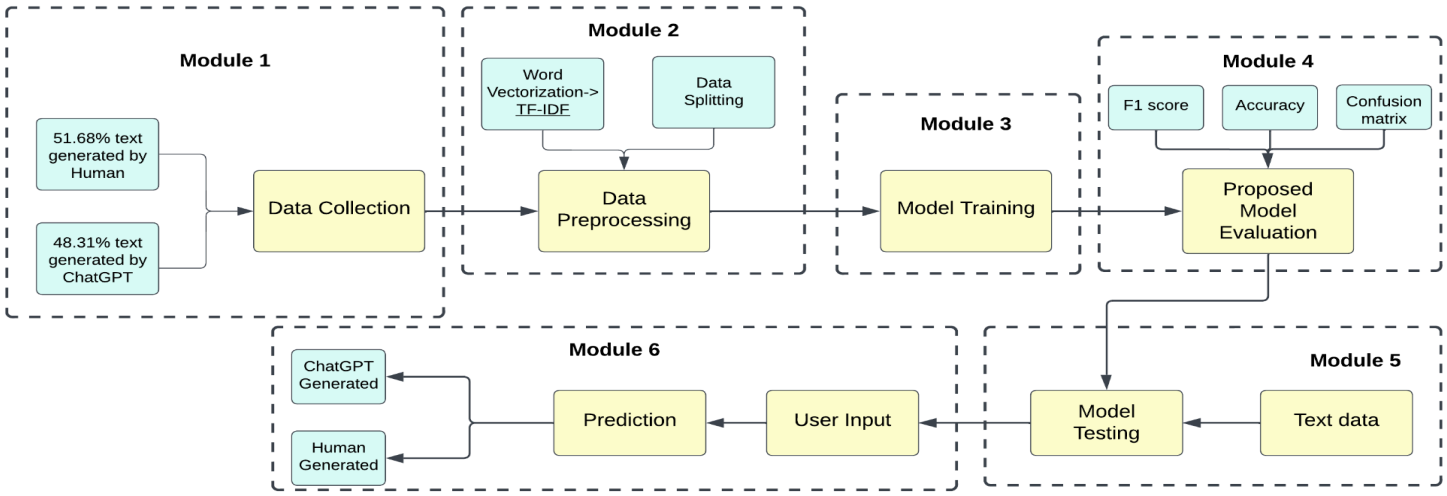

Fig. 4.1 System Architecture

4.2 DATA FLOW DIAGRAMS

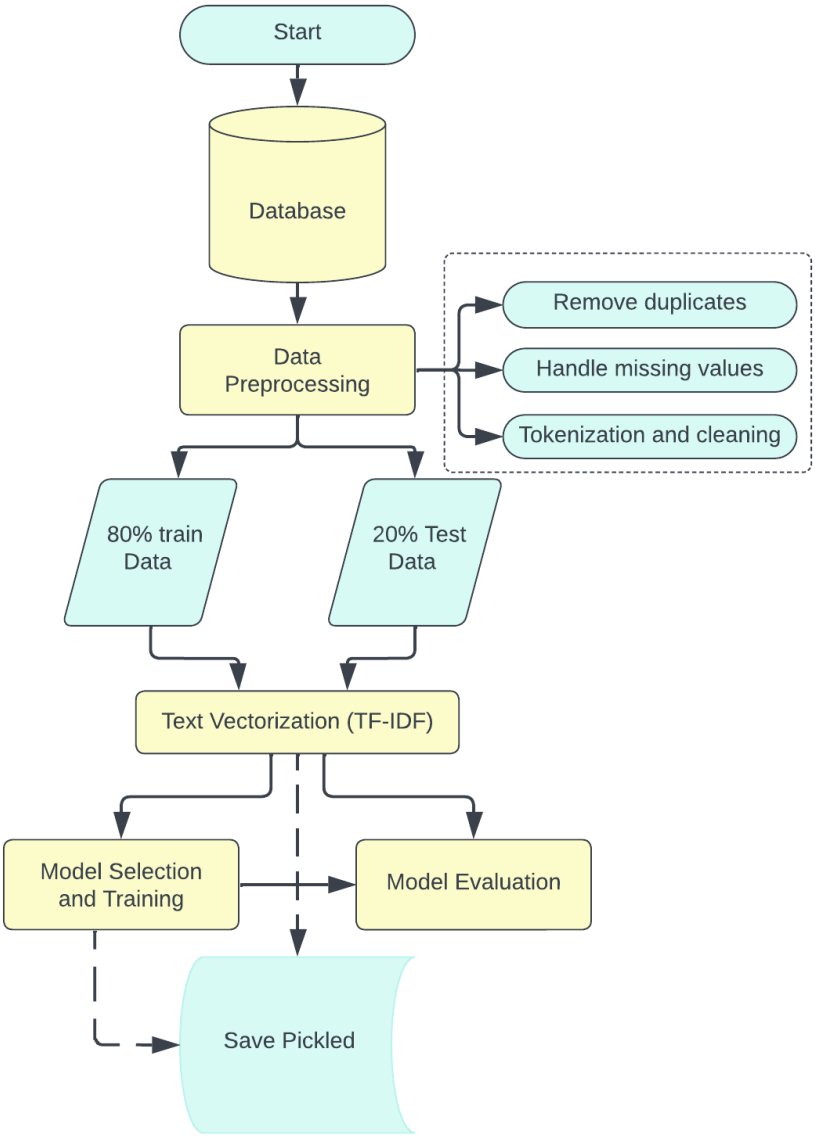

Fig.4.2 Data Flow Diagram

4.3 UML DIAGRAMS

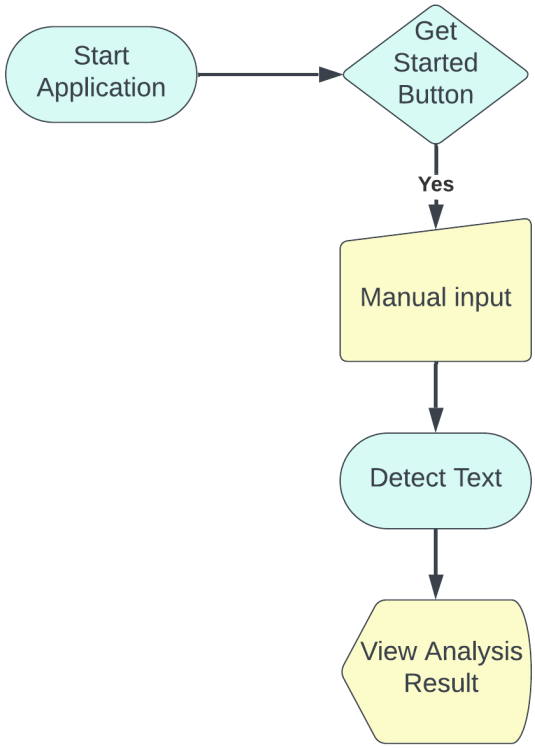

Fig.4.3.1 UML Use Case Diagram

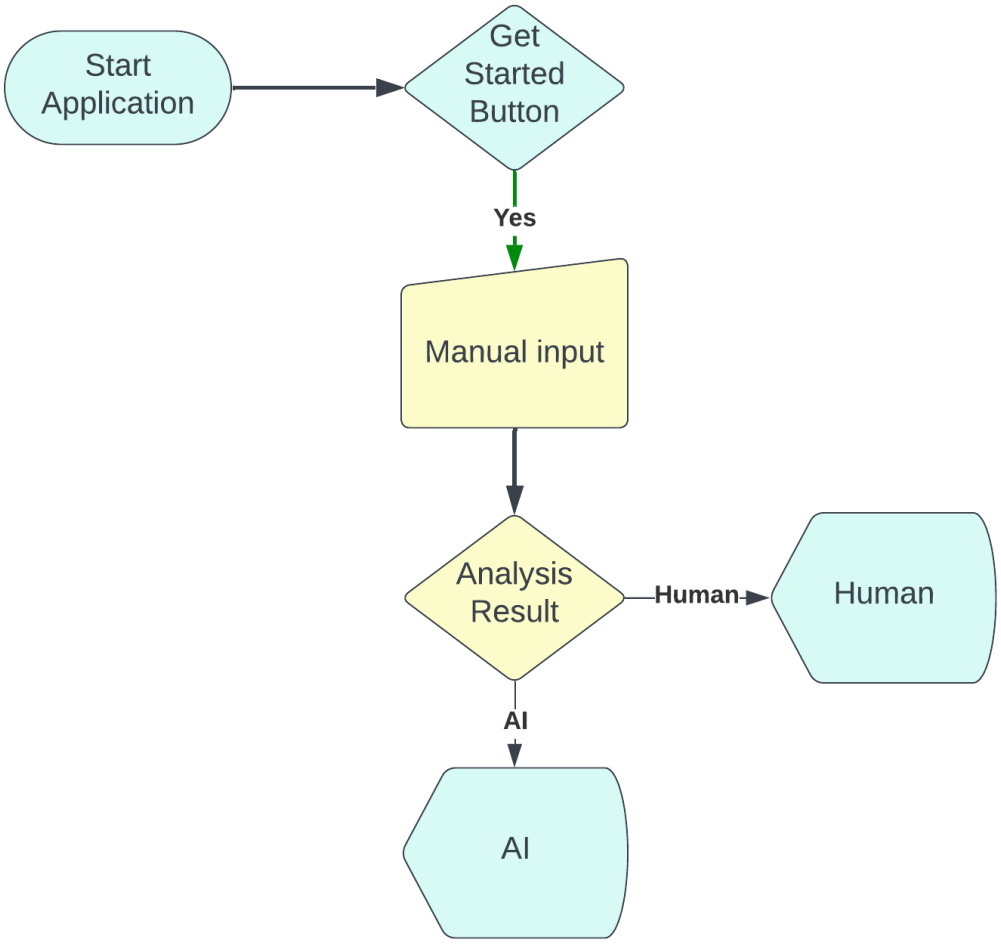

Fig.4.3.2 UMLSequence Diagram

5. PROJECT PLAN

5.1 PROJECT SCHEDULE

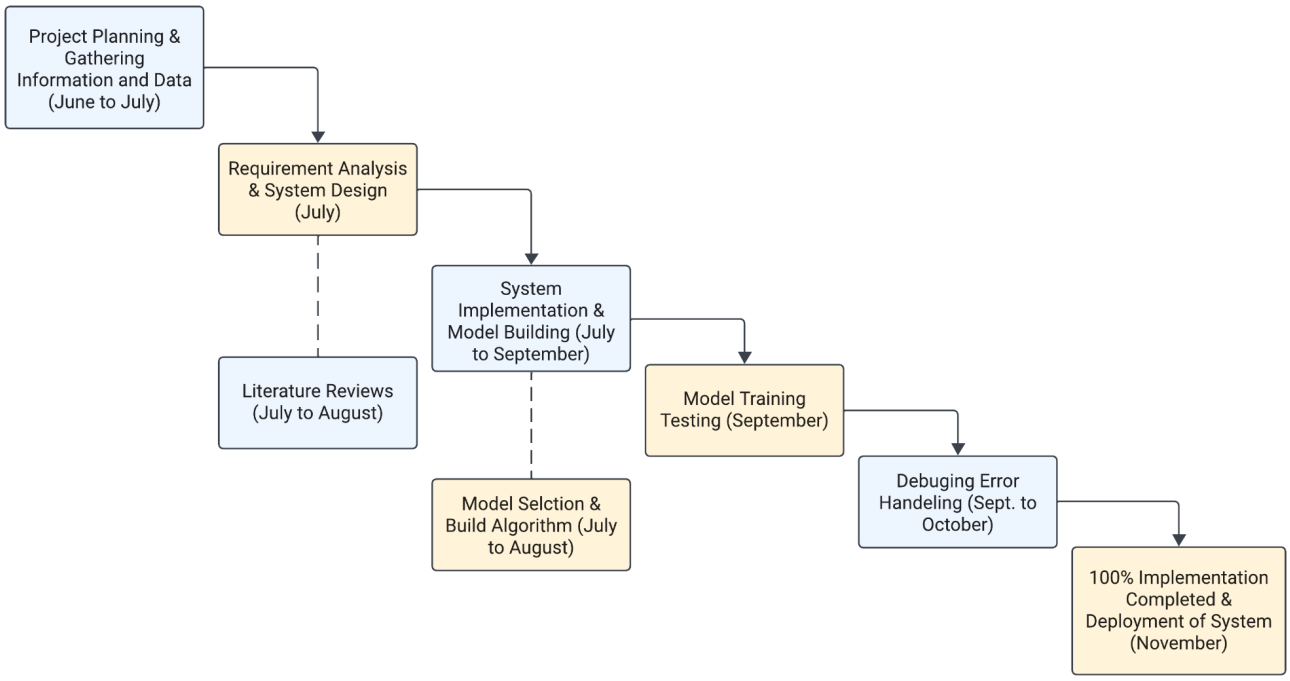

Fig 5.1 Project Schedule

5.1.1 PROJECT TASK SET

The project task set for the AI content detection project may include the following tasks:

1. Project planning and scoping
2. System design and development
3. Procurement and installation of equipment
4. Testing and quality assurance
5. Training and certification of personnel
6. Implementation and deployment
7. Monitoring and maintenance
8. Risk management and mitigation
9. Performance evaluation and reporting

Each task should be broken down into smaller subtasks and assigned a timeframe for completion. This will help ensure that the project stays on schedule and that all tasks are completed in a timely manner. A Gantt chart or similar project management tool can be used to visualize the project schedule and track progress.

6. PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

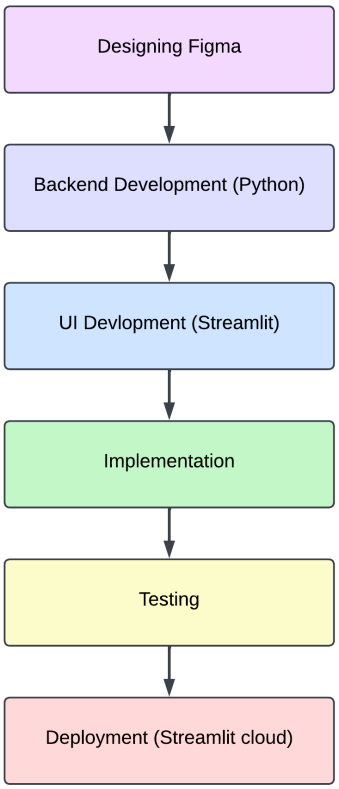

Fig 6.1 System Implementation Plan

AI content detection module:

- Word Vectorization: We start the project by using TF-IDF (Term Frequency-Inverse Document Frequency) for word vectorization. This technique transforms text into meaningful numerical data, which is a crucial step in any Natural Language Processing (NLP) task.

- Machine Learning Model: After transforming the text data, we build a machine learning model using the Extra Tree Classifier. This model is known for its high accuracy and efficiency. It's trained to distinguish whether a given text was written by a human or generated by a machine.

- Result Display: Finally, the result of the analysis is displayed to the user. The software accurately determines and shows whether the given input text was generated by a machine or was written by a human.

Other project modules include

- Data collection Module: In this module, the dataset used is called "ChatGPT Paraphrases," obtained from Kaggle. The dataset comprises 419,197 rows and contains four columns. Among its text samples, approximately 51.68% are generated by humans, with the remaining 48.31% generated by ChatGPT.

- Data Pre-processing: This module is dedicated to data pre-processing, we use TF-IDF word vectorization for transforming text data into numerical representations using the TF-IDF technique.

- Data Split module: In this module, data is partitioned into training and testing sets using Scikit-Learn's train_test_split function. 80 % of the entire data is utilized for training and the remaining 20 % for testing. This split is a critical step in preparing the data for model training and evaluation

- Model Selection: In this module, experiments are conducted with ten different machine-learning models on the dataset. These models include LR, DT, KNN, SVM, RFC, ABC, BG, GBC, and Extra Tree Classifier models. We finally used the Extra Tree classifier for our classification as it gave the best accuracy among all.

- Evaluation: In this module, the selected model's performance is evaluated. Assessment is based on the F1 score, accuracy, and confusion matrix, providing critical insights into the model's capability to differentiate between people or human-generated and ChatGPT-generated data.

- Final Prediction: In this module, we develop a web application that takes user input, gives a final prediction, and classifies whether the data is Human written or ChatGPT generated.

6.2 TOOLS AND TECHNOLOGIES USED

- Programming Languages: Python
- Frameworks: Streamlit
- Cloud Platform: Streamlit cloud, GitHub
- Development Tools: Visual Studio Code, Git

6.3 ALGORITHM DETAILS

The Extra Trees (Extreme Randomized Trees) classifier is an ensemble learning method that belongs to the family of decision tree algorithms. It is an extension of the Random Forest algorithm and shares similarities with it. Like Random Forest, the Extra Trees classifier builds multiple decision trees during training and combines their predictions to improve overall accuracy and robustness.

However, what sets the Extra Trees classifier apart is the way it constructs individual decision trees. While Random Forest randomly selects a subset of features at each split, Extra Trees goes a step further by using random thresholds for feature splits, making the algorithm even more randomized. This excessive randomization during the tree-building process often results in a diverse set of trees.

Key characteristics of the Extra Trees classifier include

Random Feature Selection: At each node of every decision tree, a random subset of features is considered for splitting, enhancing the diversity of individual trees.

Random Thresholds: Instead of finding the optimal threshold for each feature, Extra Trees uses random thresholds for splitting, which adds an extra layer of randomness.

Bootstrap Aggregating (Bagging): Like Random Forest, Extra Trees builds multiple trees using bootstrapped samples of the training data, and the final prediction is an aggregation of predictions from all trees.

Reduced Overfitting: The ensemble nature of Extra Trees and the randomization in feature selection and threshold determination help reduce overfitting, making it less sensitive to noise in the training data.

Due to its high level of randomness, the Extra Trees classifier is particularly useful in situations where the dataset is noisy, and it often performs well across various types of data. It can be an effective choice for classification tasks in machine learning, providing a balance between accuracy and resilience to overfitting.

6.3.1 ALGORITHM

Extra Tree Classifier:

An Extra Trees Classifier is an ensemble machine learning model for classification. It's a variant of decision tree-based models, similar to Random Forests, but with added randomness. During training, Extra Trees randomly selects subsets of features and threshold values for feature splits in each tree. This extra randomness reduces overfitting and speeds up training, making it a faster option than Random Forests. When making predictions, each tree in the ensemble votes for the predicted class, with the majority determining the final prediction. Extra Trees' advantages include reduced overfitting, faster training, and effectiveness with high-dimensional data. While individual trees may not be as accurate as in Random Forests, the ensemble often delivers strong predictive performance. The choice between Extra Trees, Random Forests, or other ensembles depends on the specific dataset and problem, typically requiring experimentation to determine the best fit.

- The extra tree classifier is an Ensembling algorithm.
- It divides the dataset into n equal subsets (n can be specified by the developer)
- It creates a decision tree for each subset.
- It uses bootstrap aggregation to aggregate each output from the decision tree
- Using the bagging technique, it conducts all the computations parallelly.

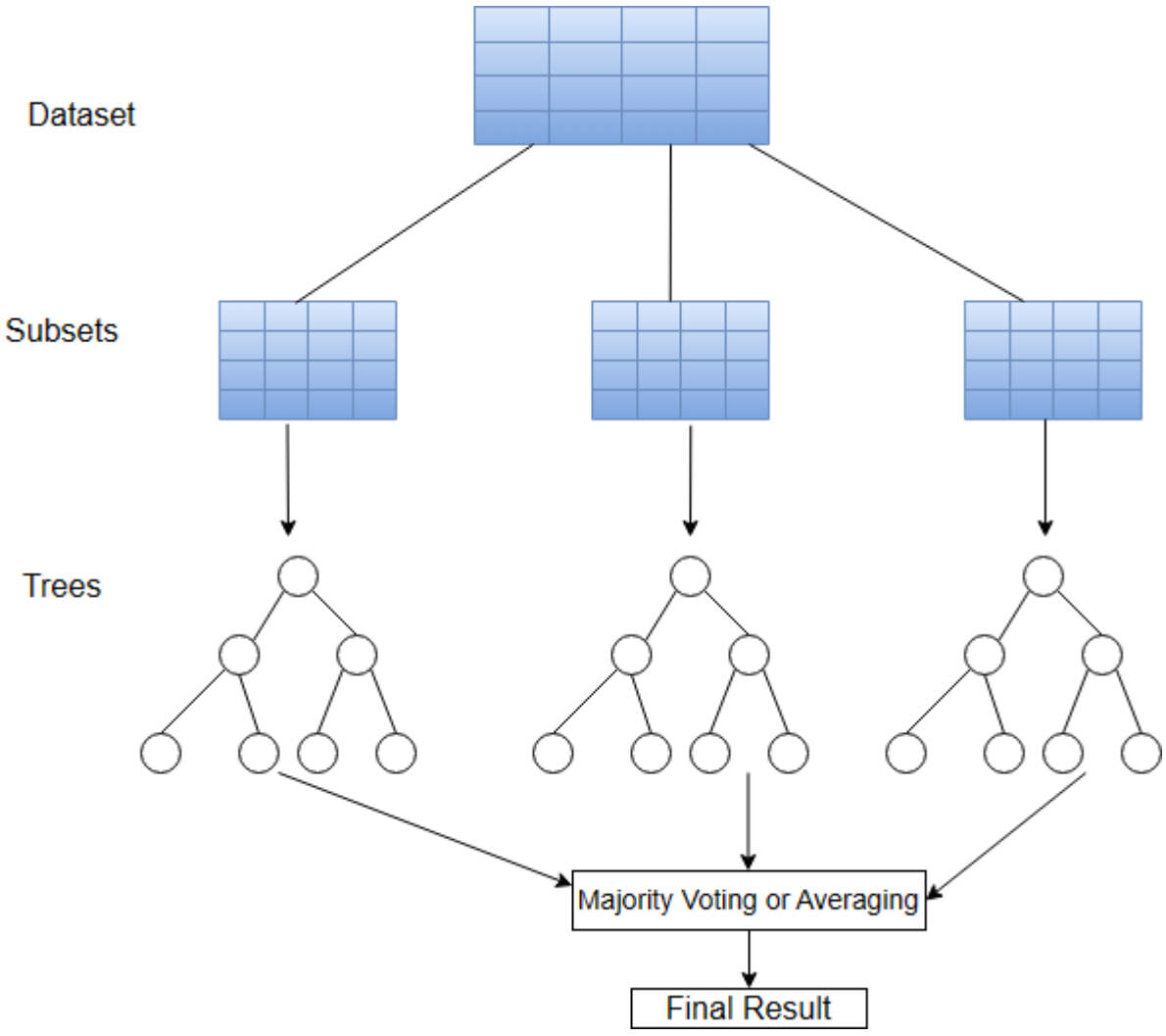

Fig 6.3.1 Extre Tree Classifier

7. TESTING

7.1 TYPES OF TESTING

Manual Testing

7.2 TESTING TECHNIQUES

TABLE 7.2: TESTING RESULTS

Sr.No.	Test Case	Model Output	Actual Output
1	Education is the cornerstone of personal and societal development. It empowers individuals with knowledge, skills, and critical thinking abilities. Beyond imparting academic content, education fosters socialization, promoting values like tolerance and collaboration. A well-rounded education system prepares students for the challenges of an ever-changing world, encouraging curiosity and adaptability. Education is not just confined to classrooms; it extends to experiential learning and lifelong pursuits. In essence, it equips individuals to contribute meaningfully to society, fostering innovation, empathy, and global understanding. As a catalyst for progress, education is an invaluable asset, shaping the future and creating a more informed and enlightened world.	Ai 76.0%	Ai
2	Nature is a harmonious tapestry of life, a serene symphony that unfolds in every corner of the Earth. From the majestic mountains to the tranquil rivers, nature captivates with its breathtaking beauty. The vibrant colors of blooming flowers and the soothing whispers of rustling leaves remind us of the delicate balance that sustains life. In nature, we find solace and inspiration, a refuge from the chaos of modern life. The intricate ecosystems teach us the importance of coexistence and adaptation. As stewards of the planet, it is our duty to preserve and cherish the wonders of nature for future generations, fostering a world where humanity and the environment thrive together.	AI 76%	AI
3	The sound of your scream can't be heard in space. The vacuum in	Human 100%	Human

	space is caused by the lack of air. Vacuums do not permit the propagation of sound waves. A 100 km radius around our planet marks the beginning of outer space. Space appears as a black blanket dotted with stars due to the absence of air to scatter sunlight.		
4	Space, the final frontier, has captivated human imagination for centuries. Its vastness, filled with celestial wonders, evokes a sense of awe and curiosity. From the mysterious depths of black holes to the dazzling beauty of distant galaxies, space is a canvas of cosmic wonders. Advancements in technology have allowed us to explore this frontier, sending probes and telescopes to unravel its secrets. The International Space Station serves as a testament to human ingenuity, fostering international collaboration in the pursuit of knowledge. As we gaze at the stars, we ponder our place in the universe, sparking dreams of interstellar travel and the possibility of extraterrestrial life.	AI 80%	AI
5	Nature is a gift of God to all living being on the earth. There is no one who is not daily helped by the goods of nature. Nature is significant to the development of life. As humans beings, we realize how important a single plant is for our survivial. The beauty of nature includes plants, animals, insects and other aspects like the mountains, hills, rivers and oceans.	Human 100%	Human
6	'Wild animals' is a term that refers to the animals that are not normally domesticated. Wild animals generally live in forests. The major wild animals of Indian are elephant, tiger, lion, rhino, bear, etc. Wild animals are a living resources. We might not see any direct connection between agriculture and animals. But most of the animals have a huge contribution to agriculture and food growing sources. There are large number of animals facing huge aggression and their life is in danger	Human 100%	Human
7	Water, a fundamental molecule for life, flows through the tapestry of our existence. Its pristine clarity mirrors the essence of purity, sustaining ecosystems and quenching the thirst of all living beings. Fluidity encapsulates its nature, seamlessly adapting to varied forms, from tranquil lakes to roaring oceans. A universal solvent, water dissolves the boundaries between compounds, fostering chemical harmony. Its cyclical dance, from vaporous clouds to liquid embrace, epitomizes the circle of life. Yet, water, a silent witness to the ages, faces the threats of pollution and scarcity, urging humanity to become stewards of this liquid	AI 60%	AI

	treasure for the well-being of our planet and future generations.		
8	Today I was a very eventful day filled with love of hurdles which tested my patience. I started by practicing meditation, then I had to go college for my submissions. After a long day at college, I came back home in the evening feeling tired. A power nap recharged me. In the evening, I went out on dinner with my friends, we talked a lot about sports and the latest movie released by our favorite superstar. Then we had dessert to wrap up the party. I was very tired, so I hit the sack as soon as I got back home.	Human 100%	Human
9	Food, a symphony of sustenance, orchestrates the dance of nourishment in the grand theatre of life. A diverse palette of flavors, textures, and aromas, it transcends cultural boundaries, weaving a tapestry of culinary heritage. From the crisp crunch of vegetables to the savory embrace of proteins, each bite is a sensorial journey. Food not only fuels our bodies but also binds communities, fostering shared experiences and traditions. I love eating, I usually prefer home cooked food, but I am very bad at cooking. Though I am slowly slowly learning how to cook according to my taste and style, I still prefer eating the cooked by my mom. She makes the best dishes and I love it	AI 72.00%	Mix
10	Today was a whirlwind of experiences. The morning greeted me with radiant sunshine, but unforeseen obstacles tested my resilience. Amidst challenges, I discovered pockets of happiness, whether in a warm cup of coffee or a heartfelt conversation. Gratitude transformed an ordinary day into one with remarkable depth and fulfillment. In the evening, I went out on dinner with my friends, we talked a lot about sports and the latest movie released by our favorite superstar. Then we had dessert to arap up the party. I was very tired, so I hit the sack as soon as I got back home.	AI 72%	Mix
11	Cats, enigmatic companions in the tapestry of domesticity, embody elegance and independence. Graceful felines, with eyes that gleam like orbs of mystery, navigate our homes with a silent prowess. From playful antics to languid stretches, they enchant with their fluid movements. Their purrs resonate as a soothing melody, a testament to contentment. The orange cats are very difficult to handle, but are equally cute. They are known for their mischievous behaviour. I personally don't like cats. I am not scared of them, but I just don't like them around me, they seem	AI 57%	Mix

	very evil and selfish. They only come to you when they are hungry. I prefer dogs. Dogs are the best. They are so jolly and friendly		
12	Today was a whirlwind of experiences. The morning greeted me with radiant sunshine, but unforeseen obstacles tested my resilience. Amidst challenges, I discovered pockets of happiness, whether in a warm cup of coffee or a heartfelt conversation. Gratitude transformed an ordinary day into one with remarkable depth and fulfilment.	AI 100%	AI

8. RESULTS

8.1 COMPARATIVE ANALYSIS BASED ON INPUT DATASET SIZE

As we experimented with diverse data sizes, we observed fluctuations in accuracy across classifiers, wherein larger datasets consistently yielded higher accuracy. Noteworthy is the consistent superiority of the Extra Tree classifier, which consistently outperformed other algorithms, demonstrating the highest accuracy in all data size batches. This robust performance underscores the effectiveness of the Extra Tree classifier across varied dataset magnitudes, suggesting its reliability and potential for superior predictive outcomes in diverse scenarios. The findings emphasize the importance of considering both algorithm selection and dataset size when developing machine learning models, with the Extra Tree classifier standing out as a reliable choice for achieving consistently high accuracy across different data scales.

TABLE 8.1: COMPARATIVE ANALYSIS BASED ON INPUT DATASET SIZE

Algorithm	15K	30K	45K	75K	90K	100K
LR	0.749	0.764	0.847	0.784	0.851	0.794
SVM	0.763	0.775	0.842	0.790	0.841	0.797
MNB	0.718	0.698	0.801	0.761	0.805	0.768
KNN	0.658	0.677	0.617	0.547	0.688	0.681
DTC	0.652	0.654	0.709	0.650	0.711	0.650
RFC	0.777	0.783	0.837	0.781	0.847	0.803
ETC	0.788	0.793	0.853	0.800	0.863	0.813
ABC	0.720	0.719	0.767	0.722	0.765	0.718
BG	0.736	0.753	0.812	0.753	0.826	0.775
GBC	0.714	0.720	0.753	0.710	0.753	0.711

8.2 OUTCOMES

- Optimum Model Selection: After experimenting with various machine learning models, we successfully identified the Extra Tree Classifier as the optimum model for our project due to its highest accuracy.
- Effective Prediction: Our model effectively takes the input data and predicts whether the given input text was written by a human or generated by ChatGPT. This showcases the model's ability to distinguish between human-written text and AI-generated text.
- High Accuracy: The model demonstrated high accuracy in its predictions, validating our choice of using the Extra Tree Classifier and the TF-IDF technique for data preprocessing.
- Successful Implementation: The successful implementation of our project signifies a step forward in the field of text classification, particularly in distinguishing between human and AI-generated text.

8.3 SCREEN SHOTS

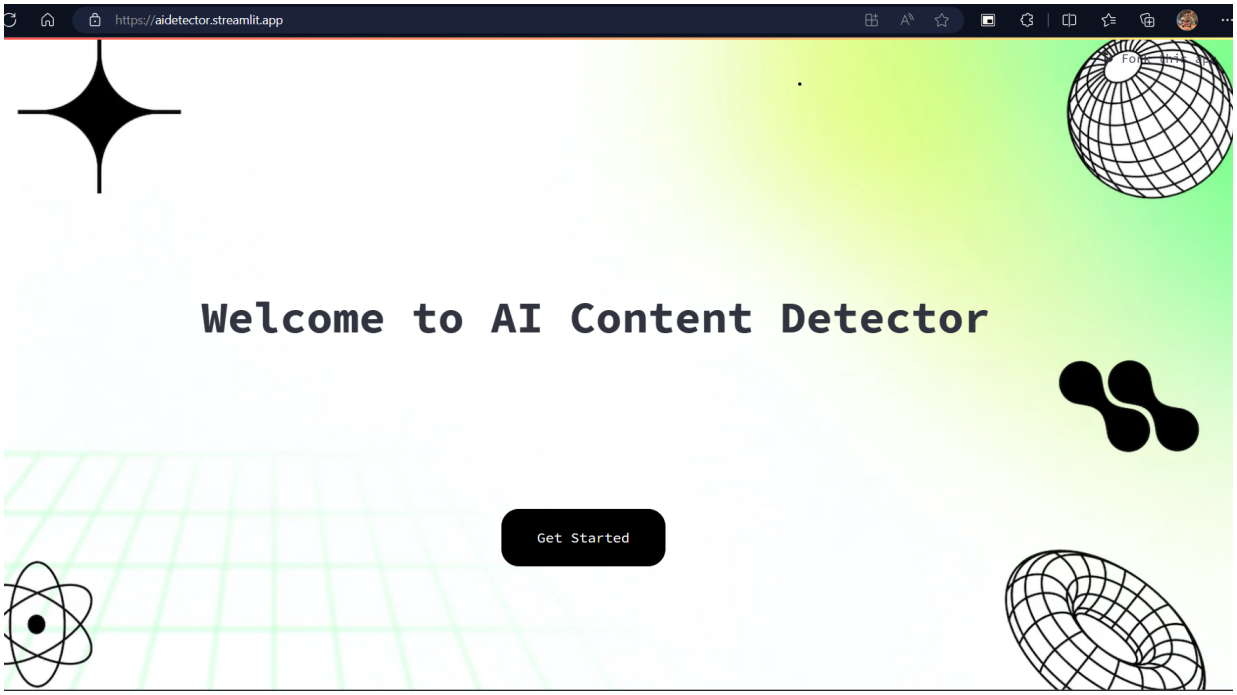

Fig.8.3.1 User Interface

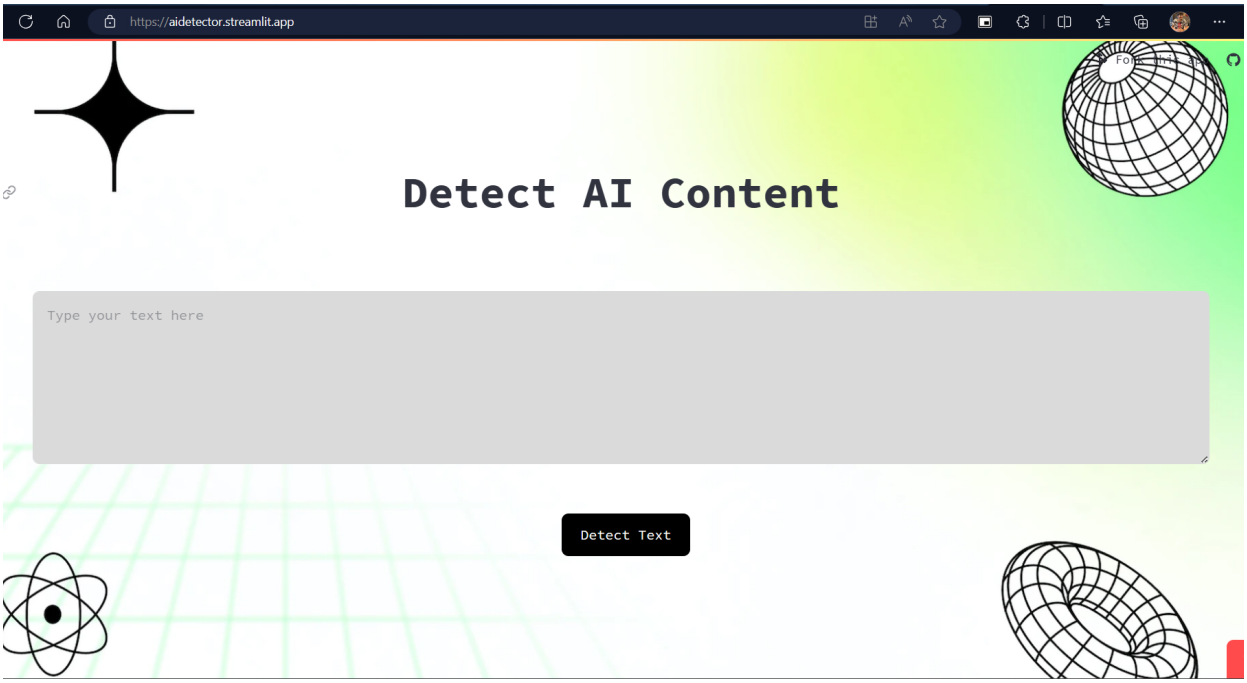

Fig.8.3.2 User Input

Fig.8.3.3 AI Content Detected

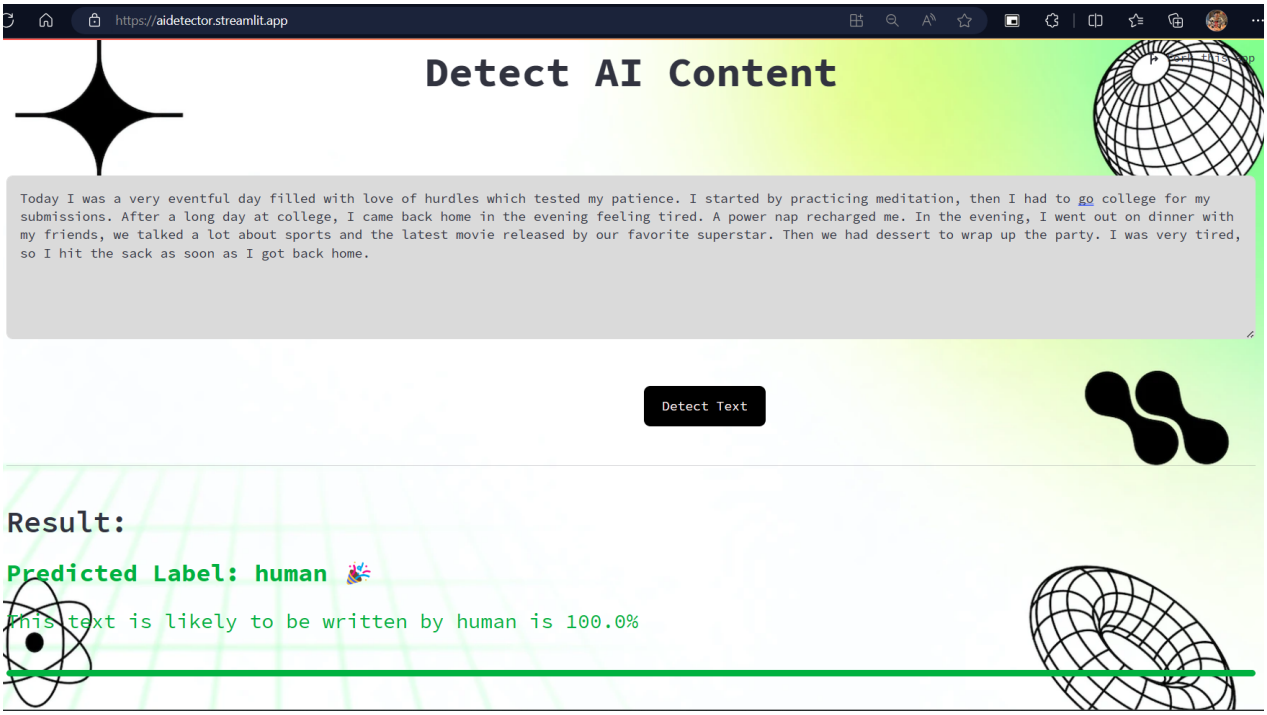

Fig.8.3.4 Human Content Detected

Fig.8.3.5 Hybrid Content Detected

9. CONCLUSIONS

9.1 CONCLUSIONS

This project aimed to distinguish between human-written and ChatGPT-generated text. We trained a model on a dataset, used TF-IDF for feature extraction, and applied an Extra Trees Classifier, which yielded the highest accuracy of 80.1%. The system takes any text as input and identifies its source. This work contributes to digital forensics and content moderation by providing a tool that can effectively discern the origin of a text. It highlights the potential of machine learning in understanding and utilizing AI-generated content.

This project has demonstrated the practical value and applicability of our model in various scenarios, thereby underscoring the potential of machine learning in enhancing our understanding and utilization of AI-generated content. This project has not only achieved its objectives but also opened up new avenues for future research and applications. It has demonstrated the potential of machine learning and natural language processing in tackling the challenges posed by the proliferation of AI-generated content.

9.2 FUTURE WORK

- Future work includes exploring multi-modal data analysis for more accurate detection of text, images, audio, and videos.
- Integration with other websites, browser extensions, or a mobile app that can allow users to easily access this tool.
- Incorporating machine learning techniques to better predict demand and optimize delivery routes in real time.
- Analyzing the impact of the system on society, ethics, and the law.
- Exploring the potential benefits and risks of the system, such as how it can help detect and prevent misinformation, plagiarism, or fraud, or how it can infringe on the privacy, rights, or creativity of the authors.
- Proposing some guidelines or policies to regulate the use of the system and ensure its ethical and responsible use.

9.3 APPLICATIONS

- Content Moderation: Your model can be used to filter out AI-generated spam, misinformation, or inappropriate content on websites and social media platforms.

- Academic Integrity: It can help educational institutions and online learning platforms identify AI-generated essays or assignments submitted by students, ensuring academic honesty.

- News and Journalism: Media organizations can use it to check for AI-generated news articles or reports, maintaining the integrity of journalism.

- Chatbot Verification: It can validate whether customer support chatbots are providing genuine human-like responses or relying on AI-generated scripts.

- AI Ethics: Researchers and organizations can use this to assess and disclose the extent of AI involvement in text content, contributing to AI ethics and transparency.

- Creative Writing: Authors and creative writers can confirm the authenticity of their work, ensuring originality and creativity in their content.

- Legal Documents: Legal professionals can verify the origin of documents, contracts, or legal content to prevent fraudulent submissions.

- AI-Human Collaboration: It can be employed to measure the balance between AI-generated and human-generated contributions in collaborative projects, ensuring fairness and accountability.

REFERENCES

1. Islam, N., Sutradhar, D., Noor, H., Raya, J. T., Maisha, M. T., & Farid, D. M., "Distinguishing Human Generated Text From ChatGPT Generated Text Using Machine Learning." arXiv, arXiv:2306.01761 [cs.CL]

2. H. Alamleh, A. A. S. AlQahtani and A. ElSaid, "Distinguishing Human-Written and ChatGPT-Generated Text Using Machine Learning," 2023 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2023, pp. 154-158, doi: 10.1109/SIEDS58326.2023.10137767.

3. G. Gritsay, A. Grabovoy and Y. Chekhovich, "Automatic Detection of Machine Generated Texts: Need More Tokens," 2022 Ivannikov Memorial Workshop (IVMEM), Moscow, Russian Federation, 2022, pp. 20-26, doi: 10.1109/IVMEM57067.2022.9983964

4. Mitrovic´, S., Andreoletti, D., & Ayoub, O., "ChatGPT or Human? Detect and Explain. Explaining Decisions of Machine Learning Model for Detecting Short ChatGPT-Generated Text." arXiv. arXiv:2301.13852 [cs.CL]

5. Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C., " DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature.", arXiv, arXiv:2301.11305 [cs.CL]

6. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L., "Testing of Detection Tools for AI-Generated Text.", arXiv., arXiv:2306.15666 [cs.CL]

7. Khalil, M., & Er, E., "Will ChatGPT get you caught? Rethinking of Plagiarism Detection.", . In: Zaphiris, P., Ioannou, A. (eds) Learning and Collaboration Technologies. HCII 2023. Lecture Notes in Computer Science, vol 14040. Springer, Cham., doi: https://doi.org/10.1007/978-3-031-34411-4_32

8. Faisal R. Elali, Leena N. Rachid, "AI-generated research paper fabrication and plagiarism in the scientific community.", CellPress , https://doi.org/10.1016/j.patter.2023.100706.

9. Elkhatat, A.M., Elsaid, K. & Almeer, S., "Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text." Int J Educ Integr,

DOI:https://doi.org/10.1007/s40979-023-00140-5

10. Yongqiang Ma, Jiawei Liu, Fan Yi, Qikai Cheng, Yong Huang, Wei Lu, Xiaozhong Liu, “AI vs. Human -- Differentiation Analysis of Scientific Content Generation”, arXiv, arXiv:2301.10416 [cs.CL]

11 .Corizzo, R., & Leal-Arenas, S.,). "One-Class Learning for AI-Generated Essay Detection.", Applied Sciences (Switzerland), 13(13). DOI: https://doi.org/10.3390/app13137901

12. A. Harada, D. Bollegala and N. P. Chandrasiri, "Discrimination of human-written and human and machine written sentences using text consistency," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 41-47, doi: 10.1109/ICCCIS51004.2021.9397237.

13. Uzun, L. "ChatGPT and Academic Integrity Concerns: Detecting Artificial Intelligence Generated Content.", Technology (LET Journal), 3, 45–54 (2023)

14. Katib, I., Assiri, F. Y., Abdushkour, H. A., Hamed, D., & Ragab, M,). "Differentiating Chat Generative Pretrained Transformer from Humans: Detecting ChatGPT-Generated Text and Human Text Using Machine Learning.", MDPI, Mathematics 2023, DOI: https://doi.org/10.3390/math11153400

APPENDIX I

List of Publications