AI-TEXT DETECTION

Gopi Nagabhiru

*School of Computing Sciences & Computer Engineering*

*Univ. of Southern Mississippi* Hattiesburg, MS, USA [Gopi.Nagabhiru@usm.edu](mailto:Gopi.Nagabhiru@usm.edu)

Srinivasa Reddy Annapareddy

*School of Computing Sciences & Computer Engineering*

*Univ. of Southern Mississippi* Hattiesburg, MS, USA [W951458@usm.edu](mailto:W951458@usm.edu)

Nick Rahimi

*School of Computing Sciences & Computer Engineering*

*Univ. of Southern Mississippi* Hattiesburg, MS, USA [nick.rahimi@usm.edu](mailto:nick.rahimi@usm.edu)

*Abstract*—This is the basic model that can predict the input text is ai generated text or not by using basic natural language processing for data processing and logistic regression for output prediction.

Introduction and problem statement

The main problem with the texts during research is whether it is human written or not. So in order to overcome this problem we try to create a machine learning model for ai text detection.This project is mainly implemented in python by using machine learning technique for predicting whether the text is AI generated or not. The text is preprocessed by using the basic natural language processing techniques and is cleaned and sent to the machine learning model.

**Literature review and Dataset:**

Various research papers are examined for this project and the basic one is selected and in the future more models will be implemented for this project

The dataset is collected from kaggle data repository which consists of almost 5 lakh samples of data which are the combination of both ai generated and human generated text.

The text that consists of samples of data as one attribute and labelled as 0 and 1 which is human generated and ai generated text.

The text is of different lengths and the ai generated texts are a bit longer than human generate text.

**Natural language processing:**

Natural Language Processing (NLP) is an area of artificial intelligence dedicated to the communication between machines and humans using natural language. The primary aim of NLP is to allow computers to comprehend, analyze, and reply to human language in a manner that is both significant and beneficial. This encompasses various activities, such as analyzing text, creating language, and translating languages.

Concepts:

Tokenization involves dividing text into smaller components known as tokens, which can be individual words, groups of words, or single characters. This step is the initial phase of working with text and is crucial for subsequent examination. Additionally, it is necessary to eliminate tags and punctuation marks as a preliminary step.

**Data pre-processing and implementation**

1. **Cleaning**

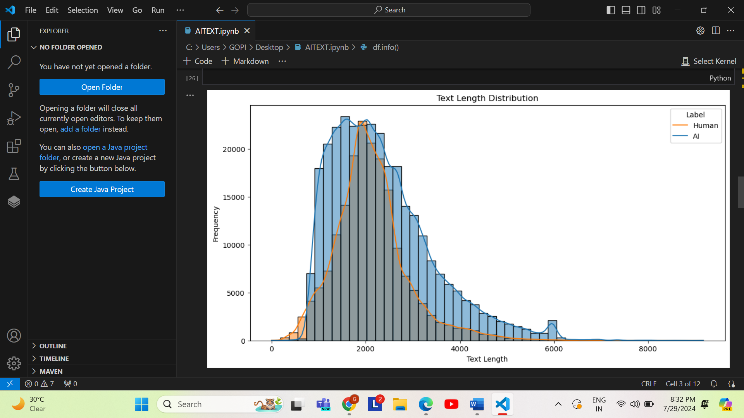
The data was preprocessed by eliminating metadata, unusual symbols, and standardizing the text to a uniform style. This action guarantees that the model is educated on the genuine content instead of irrelevant information.

1. **Tokenization**

The text was tokenized into individual words or sub words. Tokenization is a crucial step as it converts the text into a format that can be processed by machine learning models.

1. **Labeling**

Each text sample was labeled as either human-written (0) or AI-generated (1).



**Model Training**

1. **Data Splitting**

The dataset was split into training, validation, and test sets to ensure unbiased evaluation of the model. Typically, an 80-20 split was used.

1. **Classifier Selection**

A Logistic Regression model was chosen for its simplicity and effectiveness in binary classification tasks.

1. **Training**

The model was trained on the training set using the extracted features. Hyperparameters were optimized using the validation set.

**Evaluation Metrics**

The model's performance was evaluated using the following metrics:

1. **Accuracy**: The overall correctness of the model.
2. **Precision**: The proportion of true positives among the predicted positives.
3. **Recall**: The proportion of true positives among the actual positives.
4. **F1-Score**: The harmonic mean of precision and recall.

#### **Future scope:**

In the future we would like to implement more machine learning models and try to implement the classification models like lstm or bert.

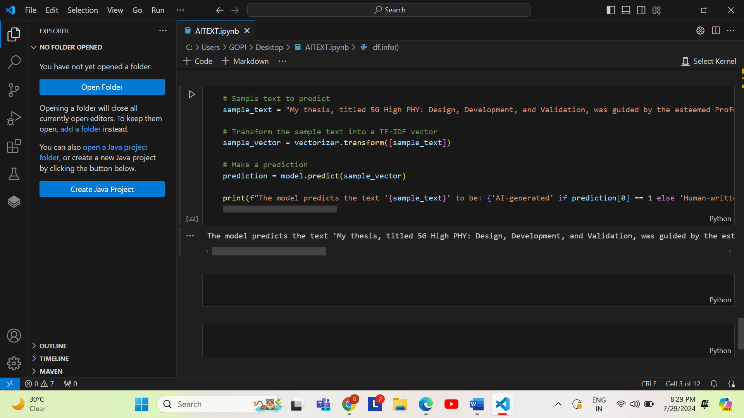
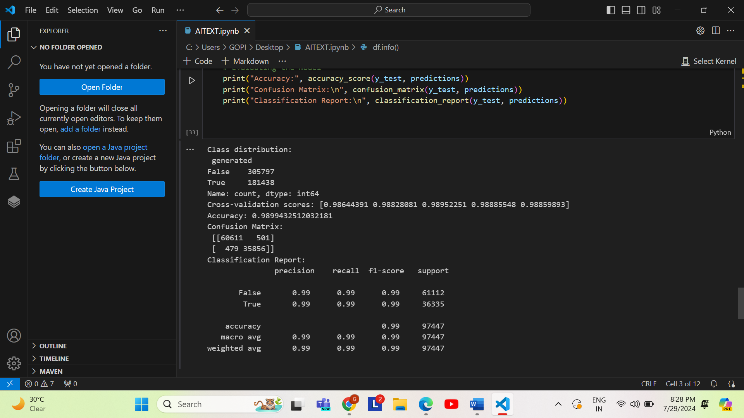
In the coming future we would try to detect which part of the text is ai generated.

##### **conclusion**

##### The developed system shows a high degree of accuracy in differentiating text generated by AI from material written by humans. Even though the logistic regression model offers a reliable starting point, other advancements in models and features can be explored to achieve even better results. Consistent upgrades and oversight guarantee the system's resilience against novel AI-generated text patterns.

Sample results:

The sample results for the accuracy of the model and cross validation scores are provided in the screenshots below



##### References

The following are some of the references used for this model

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. Available at: <https://arxiv.org/abs/2005.14165>
2. Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing* (Vol. 3). London: Pearson.

**Dataset:**

[AI Vs Human Text (kaggle.com)](https://www.kaggle.com/datasets/shanegerami/ai-vs-human-text/data)