20XD88 - Data Mining Lab

Package Report

AI Writer Detection

Varun S, Anrutha J K

20PD36, 20PD04

09th April, 2024

Introduction

The Genuine Verify application addresses concerns stemming from the rise of large language models (LLMs), particularly regarding potential plagiarism and educational integrity in assessing writing proficiency among middle and high school students.

Using advanced machine learning techniques of classification, Genuine Verify distinguishes between essays authored by students and those generated by LLMs. By analyzing subtle artifacts within texts, it identifies distinctions indicative of machine-generated content.

This tool serves as a valuable asset for educators, helping to preserve academic integrity and nurture students' writing skills in an environment increasingly shaped by LLM technology.

Dataset

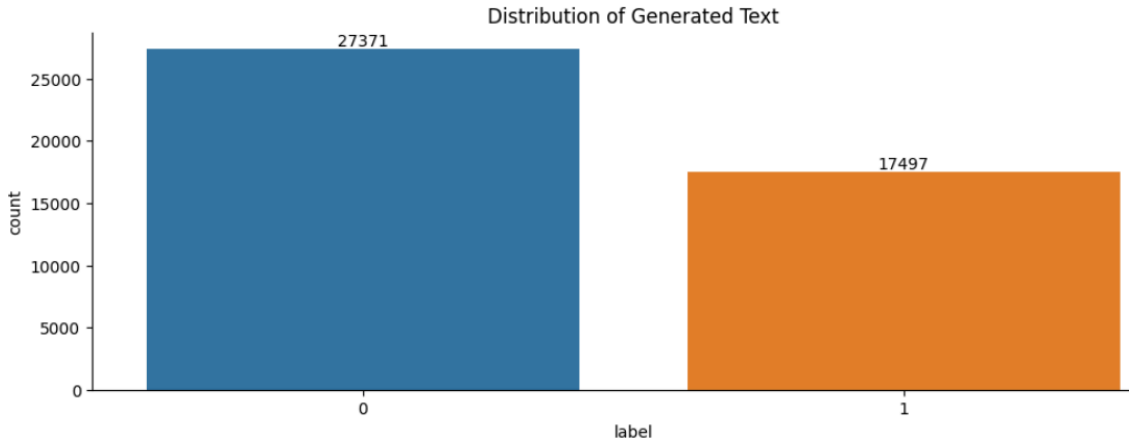
Text: The essay text itself

Generated: Whether the essay was written by a student (0) or generated by an LLM (1).

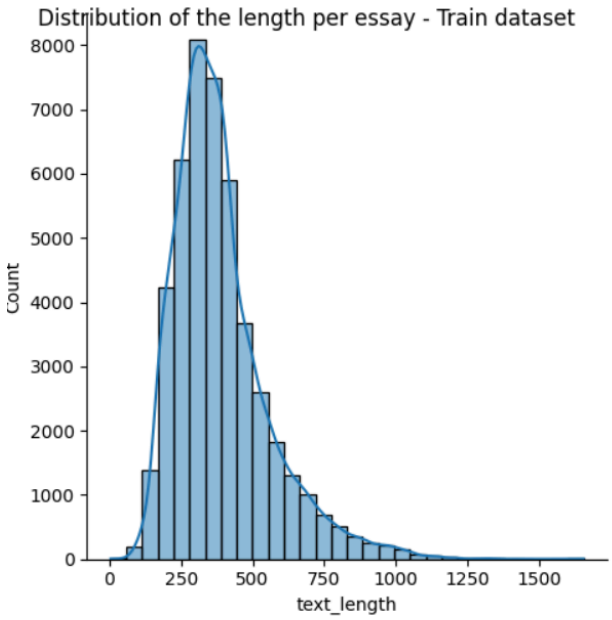
text	generated
Cars. Cars have been around since they became ...	0
Transportation is a large necessity in most co...	0
"America's love affair with it's vehicles seem...	0
How often do you ride in a car? Do you drive a...	0
Cars are a wonderful thing. They are perhaps o...	0

Exploratory Data Analysis

Label Count Distribution

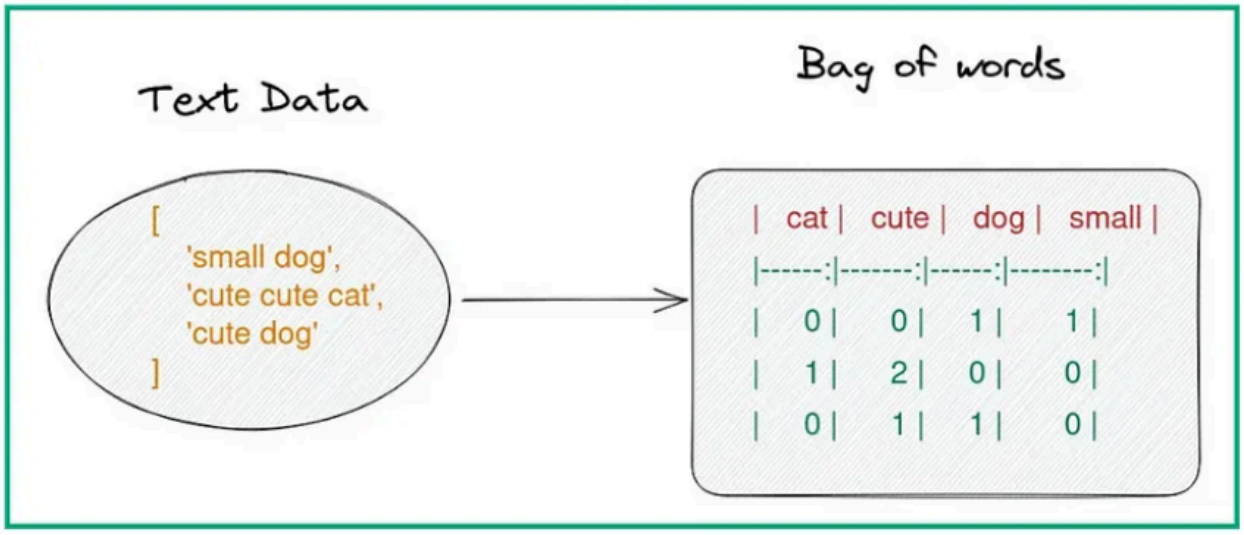


Distribution of length per essay



FEATURE ENGINEERING

The text feature was converted into vector representation using Bag of Words technique. The bag of words model is a simple way to convert words to numerical representation in natural language processing. This model is a simple document embedding technique based on word frequency.



Model Building

This is a classification task, So Logistic Regression, Naive Bayes, Random Forest, and XGBoost were employed to train models for distinguishing between AI-generated and human-generated text. Among these, logistic regression demonstrated the most effective performance.

	Model	Accuracy
0	Logistic Regression	0.960000
1	Naive Bayes	0.910000
2	Random Forest	0.957000
3	XGBoost	0.901000

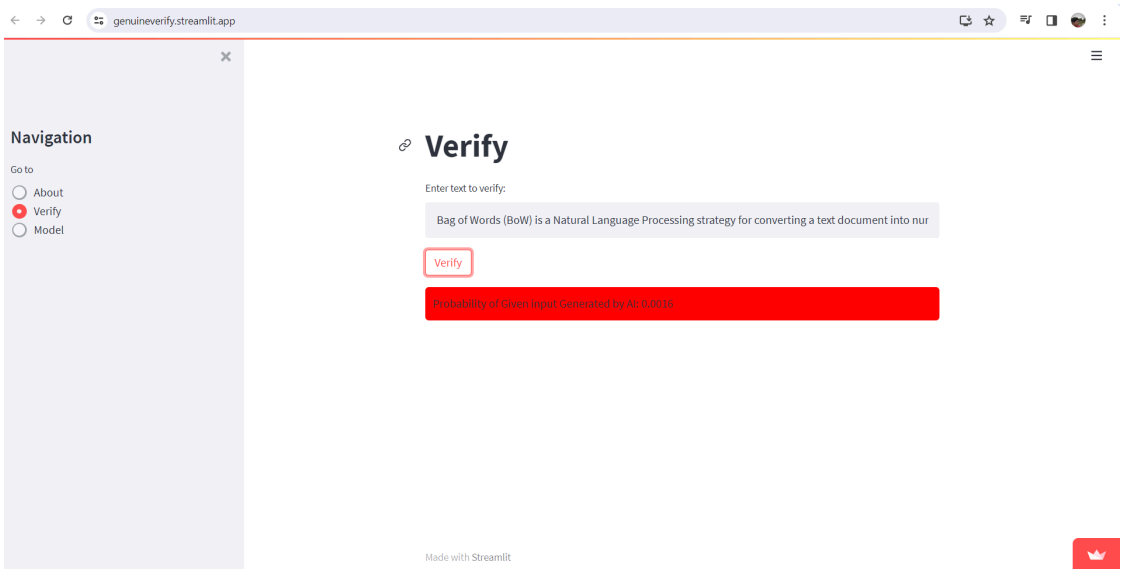
Deployment

This project has been deployed in Streamlit as a web application for a user-friendly experience. Firstly, the trained classification model was pickled for persistence of the model. The pickled model version along with the feature engineering steps was written in the backend script.

This web application can be assessed via this website.

https://genuineverify.streamlit.app/

Codebase can be assess via this website:
https://github.com/AnruthaKamal/Detect-AI-generated-Text



Conclusion:

In conclusion, the Genuine Verify application effectively addresses concerns around large language models (LLMs) by utilizing logistic regression to distinguish between student-authored essays and those generated by LLMs. It provides educators with a valuable tool to uphold academic integrity and nurture students' writing skills in an LLM-dominated environment. Deployed as a user-friendly web application via Streamlit, Genuine Verify offers seamless access to its capabilities, empowering educators to maintain authenticity in student work and foster critical thinking. It serves as a crucial safeguard against emerging technological challenges, ensuring fair academic evaluation and genuine learning outcomes.