RESEARCH FORUM 19TH UGRF

Dr .Ensaf Hussein

2025-AIS411-4

Stylometric Analysis Tool

Introduction

Stylistic analysis is a method for studying writing styles, identifying authors, and analyzing sentence construction, word choice, and stylistic elements. With the rise of digital texts and machine learning techniques, natural language processing models like LLAMA are effective tools. Enhancing Large language models with RAG enhances their performance and reduces hallucinations. We used RAG on a dataset of texts and their authors to perform stylometric analysis.

Methodology

Our analysis focuses on developing a stylometric analysis tool to detect authorship for unique writing traits using state-of-the-art natural language processing (NLP) techniques.

The study utilizes Llama models with a retrieval-augmented generation framework for effective stylometric analysis. Three models, Llama 3.2, Llama 3.1, and Llama 2, were used for stylometric analysis. The models were improved through RAG and prompt engineering, and text samples were transformed into high-dimensional embeddings using Hugging Face embeddings.

Dataset:
The dataset includes 19,579 text samples written by three writers, Edgar Allan Poe (EAP), H.P. Lovecraft (HPL), and Mary Shelley (MWS). Each instance is structured into three attributes: an Id, the body of text for analysis, and the corresponding author label.

RAG setup:
We use Langchain to construct our RAG setup.

Results

Three models, Llama 2 7B, Llama 3 8B, and Llama 3.23B, were used for authorship identification, with results including accuracy scores, confusion matrices, and predicted author distributions.

Results

Comparisons of Accuracies :
Table I summarizes the accuracy of the three models and the Weighted Voting mechanism:

• Llama 2 7B: Achieved an accuracy of 70%, The model worked well with some authors while struggling with others

.• Llama 3 8B: Delivered the lowest accuracy at 54%, suggests its limited ability to differentiate between authorial styles

.• Llama 3.2 3B: Emerged as the best-performing model, achieving an impressive accuracy of 94%,

• Weighted Voting: Attained an accuracy of 93%, closely matching Llama 3.2 3B's results.

Model	Accuracy (%)
Llama 2 7B	70
Llama 3 8B	54
Llama 3.2 3B	94
Weighted Voting	93

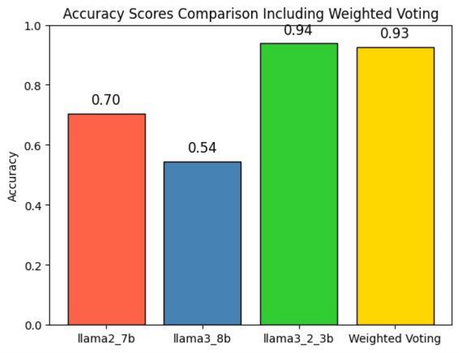

Fig. 1. Bar Chart Comparing Model Accuracies.

predicted Author Distributions:

Llama 2 7B predicted Edgar Allan Poe most frequently, followed by Mary Shelley. Llama 3 8B had a consistent bias, while Llama 3.2 3B balanced predictions across EAP, MWS, and HPL.

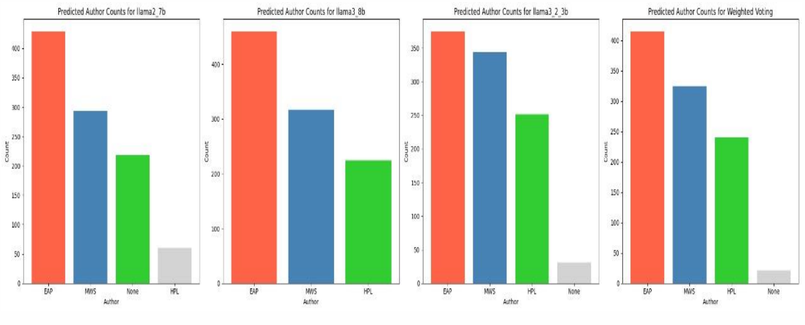

Fig. 2. Predicted Author Distributions for Each Model.

Results

Confusion Matrices:

Confusion matrices provide an overview of classification accuracy by highlighting correct and incorrect predictions:
Llama 2 7B showed moderate performance, correctly classifying EAP and MWS texts. Llama 3 8B struggled with misclassifications, but Llama 3.2 3B delivered accurate predictions. Weighted Voting improved classification reliability by leveraging strengths and reducing errors.

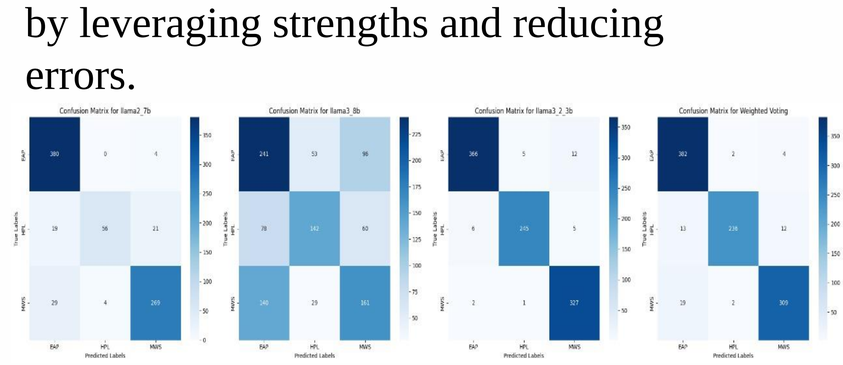

Fig. 3. Confusion Matrices for Each Model and Weighted Voting.

Conclusion

Our results highlight the powerful capabilities of advanced NLP techniques, particularly Llama models with a retrieval-augmented generation framework, in uncovering the unique stylistic fingerprints of authors. By combining sophisticated tools for text representation and efficient retrieval methods, these models excelled at capturing the nuances of individual writing styles.

Github Link

Team