RESEARCH FORUM 19TH UGRF

AIS411-4

Stylometric Analysis Tool

Mohamed Saied, Nada Mokhtar, Michael Adel, Eslam Atwan, Philopater Boles
School of Information Technology and Computer Science,
Nile University, Egypt

***Abstract*—This paper demonstrates the effectiveness of Llama models with a retrieval-augmented generation (RAG) framework in identifying unique authorial styles. By combining advanced text representation and retrieval techniques, these models excel in stylistic analysis, with applications in forensic analysis, literature studies, and digital communications.**

***Index Terms*—NLP, Llama Models, Retrieval-Augmented Generation, Stylistic Analysis**

I. Introduction

Stylometric analysis studies writing styles to determine authorship by examining patterns like sentence structure, word choice, and stylistic elements. Applications include forensic science, internet communication, and literature. Advanced NLP models like LLAMA analyze these nuances with high precision. Retrieval-Augmented Generation (RAG) further improves performance by reducing hallucinations. This study employs RAG for stylometric analysis on texts authored by Edgar Allan Poe, H.P. Lovecraft, and Mary Shelley.

II. Methodology

A. Dataset

We utilized a dataset of 19,579 text samples written by Edgar Allan Poe, H.P. Lovecraft, and Mary Shelley. Each instance included an ID, the body of text, and the corresponding author label. HuggingFace embeddings were used for text representation, and ChromaDB indexed embeddings for efficient retrieval.

B. Model Setup

We employed three Llama models: Llama 2 (7B): 7 billion parameters, Llama 3.1 (8B): 8 billion parameters, Llama 3.2 (3B): 3 billion parameters

All models were quantized and integrated with RAG using Langchain and FAISS for similarity search. Outputs were aggregated using a Weighted Voting mechanism to improve prediction accuracy.

III. Results

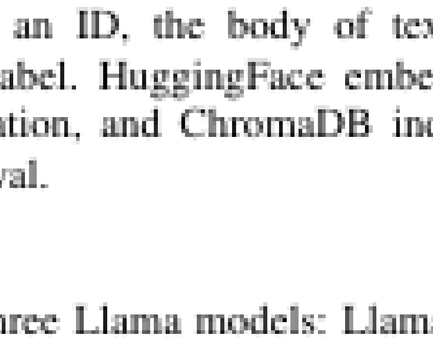

Fig. 1. Bar Chart Comparing Model Accuracies.

Observations: Llama 3.2 (3B) achieved the highest accuracy, demonstrating robust understanding of stylistic nuances. The Weighted Voting mechanism closely matched its performance.

A. Predicted Author Distributions

Llama 2 predicted Edgar Allan Poe most frequently, with moderate misclassifications. Llama 3.1 displayed bias toward Poe, struggling with other authors. Llama 3.2 balanced predictions across all authors with minimal errors. Weighted Voting combined strengths of all models, yielding reliable predictions.

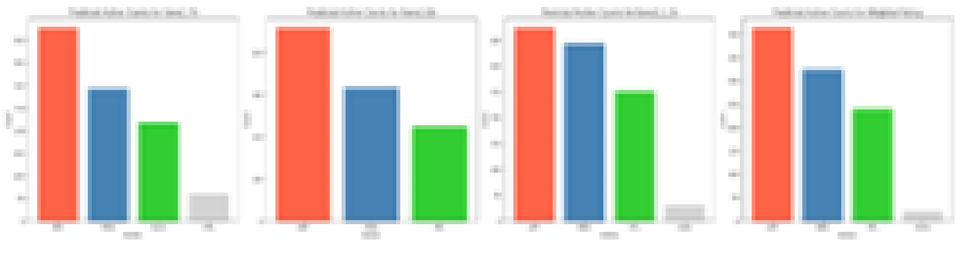

Fig. 2. Predicted Author Distributions for Each Model.

B. Confusion Matrices

Llama 2 had moderate accuracy with occasional misclassifications. Llama 3.1 had frequent misclassifications with inconsistent patterns. Llama 3.2 was most accurate with clear distinctions between authors. Voting outperformed individual models by mitigating errors and improving reliability.

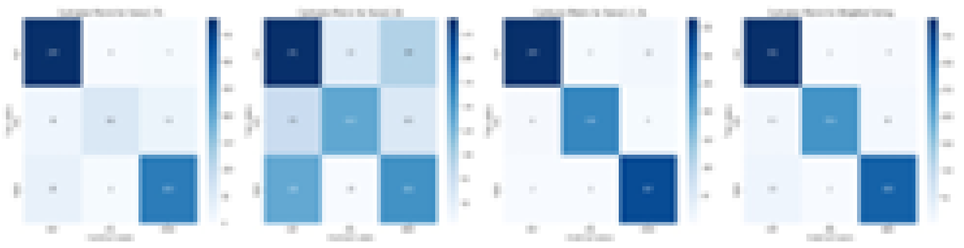

Fig. 3. Confusion Matrices for Each Model and Weighted Voting.

IV. Conclusion

This study highlights the efficacy of Llama models integrated with RAG for stylometric analysis. By leveraging advanced text representation and retrieval techniques, our approach demonstrated significant accuracy in identifying unique writing styles. The findings have broad implications for forensic analysis, literature studies, and digital communications.