UNIVERSIDAD DEL AZUAY

Facultad de Ciencias de la Administración

Carrera de Ingeniería de Sistemas y Telemática

CLASIFICACIÓN JERÁRQUICA DE LOS TEXTOS DE LAS CONVERSACIONES DEL ECU 911 UTILIZANDO MODELOS DE LENGUAJE LARGOS

Trabajo de titulación previo a la obtención del grado de Ingeniero de Sistemas y Telemática

Autor:
Juan Gabriel Flores Sánchez.

Director:
Marcos Patricio Orellana Cordero

Cuenca – Ecuador

Año
2025

DEDICATORIA

Dedico este trabajo a todas las personas que estuvieron presentes y me acompañaron durante mi formación universitaria, especialmente a mi familia por todo su apoyo. A mi esposa e hijo, quienes son mi mayor inspiración y motivación, y quienes me brindaron el tiempo y espacio necesarios para alcanzar esta meta tan importante.

En especial, dedico este trabajo a mi madrecita, quien, aunque hoy ya no está físicamente conmigo, su espíritu y amor continúan guiándome. Siempre me brindó su amor incondicional, sus oraciones y su sabiduría, ayudándome a superar cualquier desafío. Gracias, mamá, por tu infinito amor. Que Dios Jehová te tenga en su santa gloria por siempre.

AGRADECIMIENTO

En primer lugar, agradezco a Dios Todopoderoso por guiarme y bendecirme a lo largo de mi trayectoria universitaria.

Extiendo mi más sincero agradecimiento al director de mi trabajo de titulación, Ing. Marcos Orellana, quien me orientó con sus valiosos consejos durante todo el proceso de elaboración de este proyecto.

De igual manera, expreso mi gratitud al equipo de investigación y desarrollo en informática, LIDI.

Agradezco a mis amigos y compañeros de aula, quienes me acompañaron y brindaron su apoyo a lo largo de esta etapa académica.

Un agradecimiento muy especial a mi esposa e hijo, quienes me ofrecieron su apoyo y comprensión durante esta importante etapa de mi vida.

Por último, quiero expresar mi más profundo agradecimiento a mi querida madre, quien, aunque ya no está conmigo, su espíritu me acompaña y me da el aliento para seguir adelante. Mamá, gracias por todo tu amor y esfuerzo; me siento muy orgulloso de ser tu hijo y haber tenido la oportunidad de aprender tanto de ti. Te amo, mamá.

Índice de contenidos

Índice de figuras

Índice de tablas

CLASIFICACIÓN JERÁRQUICA DE LOS TEXTOS DE LAS CONVERSACIONES DEL ECU 911 UTILIZANDO MODELOS DE LENGUAJE LARGOS

RESUMEN

El estudio desarrolló un método de clasificación jerárquica para analizar las conversaciones del ECU 911, integrando modelos de lenguaje con arquitectura *LLaMA* (ej. *LLaMa 3.1 8B Instruct*), utilizando la librería *BERTopic* y técnicas de *clustering* aglomerativo. El método organizó los textos en categorías y subcategorías, capturando relaciones semánticas complejas y ofreciendo un análisis temático claro en el dominio de seguridad ciudadana. El uso de métricas de coherencia y diversidad temática aseguró asignaciones confiables, mientras que las etiquetas generadas automáticamente facilitaron la interpretación de los tópicos. El principal aporte del estudio se basa en la adaptación de estas técnicas al idioma español, proponiendo una solución innovadora para gestionar grandes volúmenes de datos textuales en los textos de las llamadas de emergencias. Aunque se evidencian limitaciones, como el solapamiento temático y la necesidad de representaciones vectoriales específicas, los resultados demuestran la viabilidad y robustez del modelo. Este trabajo establece una base sólida para futuras investigaciones y aplicaciones en clasificación jerárquica en otros contextos.

Palabras clave: *BERTopic*, clasificación jerárquica, ECU 911, *embeddings*, inteligencia artificial, lenguaje natural, LLMs, seguridad ciudadana.

HIERARCHICAL CLASSIFICATION OF TEXTS FROM ECU 911 CONVERSATIONS USING LARGE LANGUAGE MODELS

ABSTRACT

The study developed a hierarchical classification method to analyze ECU 911 conversations, integrating language models with LLaMA architecture (e.g., LLaMA 3.1 8B Instruct), using the BERTopic library and agglomerative clustering techniques. The method organized texts into categories and subcategories, capturing complex semantic relationships and providing a clear thematic analysis in the domain of public safety. The use of coherence metrics and topic diversity ensured reliable assignments, while automatically generated labels eased the interpretation of topics. The main contribution of the study lies in the adaptation of these techniques to the Spanish language, proposing an innovative solution for managing large volumes of textual data in emergency call transcripts. Although limitations such as thematic overlap and the need for specific vector representations are clear, the results prove the model's feasibility and robustness. This work sets up a solid foundation for future research and applications in hierarchical classification across other contexts.

Keywords: BERTtopic, citizen security, ECU 911, embeddings, hierarchical classification, LLMs, natural language.

Marcos Otellano
0102668209