



Facultad de Ciencias de la Administración

**Escuela de Ingeniería de Sistemas y
Telemática**

**Búsqueda y comparación contextual de llamadas
transcritas del ECU 911 utilizando modelos de lenguaje
(GPTs) para encontrar emergencias similares**

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

Autor: Ronnie Eduardo Urdiales Quinde

Director: Ing. Marcos Patricio Orellana Cordero

Cuenca – Ecuador

**Año
2025**

DEDICATORIAS

A mami, por su amor y apoyo incondicional durante toda mi vida, a papi, por ser el ejemplo de que si se hace algo se hace bien, a los dos, gracias por ser quienes son e inculcarme que, si se empieza algo, se tiene que terminar.

A mi ñaño, que me ha acompañado en todas las etapas de mi vida, este pequeño triunfo también es tuyo.

AGRADECIMIENTOS

A mi familia, por ser el ejemplo más grande de amor incondicional.

A Pablo M. por ser el mejor amigo y apoyo, no solo en la carrera, sino en toda clase de aventuras que nos ha puesto la vida por delante.

A Cami C., que además de ser mi mejor amiga y una de las mejores personas que he conocido en mi vida, me ha enseñado el valor de la amistad, y que los amigos son en realidad la familia que nosotros escogemos.

A Alejandra O., por llegar a mi vida en el momento justo, por su fuerza y su capacidad de estar presente aún en la distancia. Gracias por ser ese abrazo que siempre reconforta.

A Carolina C., por convertirse en una de esas personas que sin esperarlo se vuelven imprescindibles. Tu apoyo constante, tu alegría y tu luz han sido clave para no rendirme.

A la representación estudiantil, que me ha dado la oportunidad de ayudar a mucha gente y vivir en serio la experiencia universitaria.

A David S., que es el mayor ejemplo de disciplina que conozco, y me enseñó a ver siempre el lado feliz de las cosas.

A todos mis amigos, amigas y gente que ha repercutido en mi vida para bien.

A mi carrera, que me hace sentir cada día que todo esfuerzo tiene su recompensa.

A todas las personas que mencioné, e incluso muchas que no, gracias por ayudar, gracias por enseñarme y gracias por ser los pilares para superar cada una de mis fallas, esto es por y para ustedes.

Índice de Contenidos

DEDICATORIAS	i
AGRADECIMIENTOS	ii
Índice de Contenidos	iii
Índice de Figuras	iv
Índice de Tablas	v
1. RESUMEN	vi
2. ABSTRACT	vi
3. Introducción	1
4. Objetivos	2
4.1 Objetivo General.....	2
4.2 Objetivos Específicos.....	2
5. Marco Teórico	3
Métodos basados en contenido	3
Métodos basados en contexto	4
6. Revisión de Literatura	4
7. Métodos	6
7.1 Comprender el negocio	8
7.2 Comprender los datos	11
7.3 Procesar la información	13
7.4 Clusterizar	14
7.5 Desarrollar el modelo	16
8. Resultados	17
9. Evaluación	20
10. Conclusión	24
11. Referencias	25

Índice de Figuras

Figura 1. <i>Diagrama CRISP-DM de la metodología</i>	6
Figura 2. <i>Diagrama SPEM de la metodología de la investigación</i>	8
Figura 3. <i>Distribución de la cantidad de mensajes por llamada</i>	10
Figura 4. <i>Distribución de la cantidad de mensajes por llamada</i>	10
Figura 5. <i>Palabras que más se repiten en el dataset</i>	11
Figura 6. <i>Nube de palabras de la matriz TF-IDF</i>	13
Figura 7. <i>Método de “Codo” y “Silhouette Analysis” para determinar el número óptimo de clústeres</i>	15
Figura 8. <i>Gráfico de dispersión de los clusters y embeddings</i>	18
Figura 9. <i>Histograma de similitudes semánticas por cuartiles</i>	19
Figura 10. <i>Estructura del formulario para la aplicación de la evaluación TAM-BDA</i>	22
Figura 11. <i>Resultados cuantitativos de la evaluación TAM-BDA</i>	23

Indice de Tablas

Tabla 1. <i>Explicación de cada columna del dataset.</i>	9
Tabla 2. <i>Fragmento del dataset original.</i>	9
Tabla 3. <i>Ejemplo de llamada con preprocesamiento de datos.</i>	12
Tabla 4. <i>Ejemplo de llamada tokenizada.</i>	12
Tabla 5. <i>Visualización de palabras clave del Clúster 0 y cálculo de similitud semántica.</i>	15
Tabla 6. <i>Comparación de llamada con clúster.</i>	16
Tabla 7. <i>Similitud semántica entre llamadas</i>	17
Tabla 8. <i>Resultados de la comparación de llamadas con clústeres del tercer cuartil.</i>	19
Tabla 9. <i>Preguntas evaluación TAM-BDA</i>	20

1. RESUMEN

Este estudio presenta un enfoque innovador para la búsqueda y comparación contextual de llamadas transcritas del servicio de emergencia ECU 911, utilizando modelos de lenguaje GPT. La metodología se basa en CRISP-DM y un diagrama SPEM, e incluye preprocesamiento de datos, tokenización con modelos BERT, cálculo de similitud semántica, clusterización y comparación. El modelo utiliza K-Means para la clusterización de llamadas, mientras que KeyBERT se emplea para la extracción de palabras clave relevantes. El modelo está diseñado para ser intuitivo y accesible para los operadores de emergencias, y ha demostrado precisión en la identificación de emergencias similares. La validación del modelo se realizó a través de la evaluación por expertos y un análisis de aceptación tecnológica de Big Data. Los resultados indican que el modelo es útil y preciso en la identificación de patrones de emergencia. Se sugiere que futuros trabajos exploren la implementación del modelo en un plan piloto y su aplicación en llamadas en tiempo real para optimizar la capacidad de respuesta. Este enfoque promete mejorar la eficiencia operativa y la toma de decisiones en situaciones de emergencia.

Palabras clave: ECU 911, modelos de lenguaje, GPT, inteligencia artificial, procesamiento de datos, tokenización, similitud semántica, clusterización, KeyBERT, K-Means, Big Data

2. ABSTRACT

This study presents an innovative approach for contextual searching and comparing transcribed calls from the ECU 911 emergency service using GPT language models. The methodology uses CRISP-DM and a SPEM diagram, including data preprocessing, tokenization with BERT models, semantic similarity calculation, clustering, and comparison. The model employs K-Means for clustering calls, while KeyBERT extracts relevant keywords. The model is designed to be intuitive and accessible for emergency operators and has demonstrated accuracy in identifying similar emergencies. Model validation was conducted through expert evaluation and a Big Data technology acceptance analysis. Results indicate that the model is valid and accurate in identifying emergency patterns. Future work is suggested to explore the implementation of the model in a pilot plan and its application to real-time calls to optimize response capabilities. This approach promises to enhance operational efficiency and decision-making in emergencies.

Keywords: ECU 911, language models, GPT, artificial intelligence, data processing, tokenization, semantic similarity, clustering, Big Data



Marcos Orellana

0102666209