

# Procesamiento del Habla e Introducción a LLMs

## Portfolio de Trabajos | Ciclo 2025

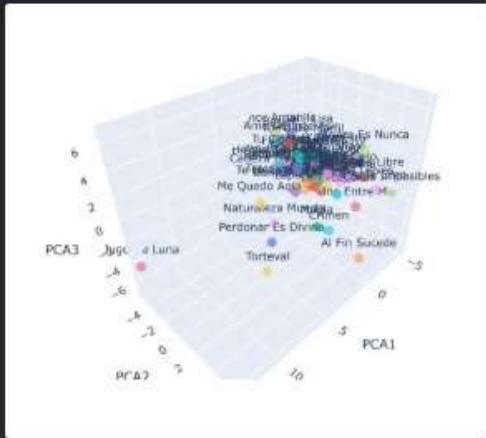
Presentado por:

Araceli Valeria Cornes

IFTS 24 - Tecnicatura Superior en Ciencias de Datos e IA

Prof. Matías Barreto

# Análisis de Canciones de Gustavo Cerati con Web Scraping



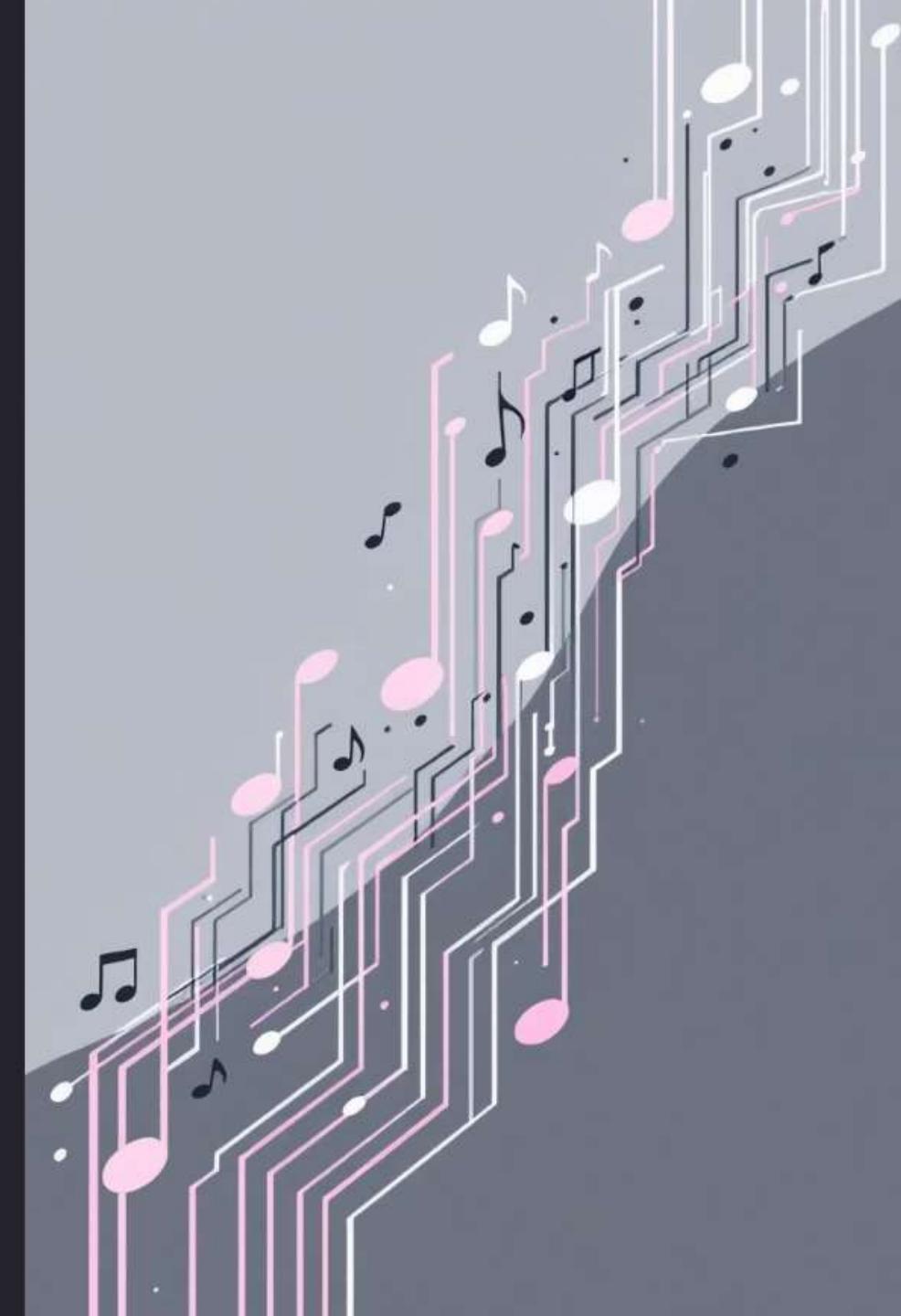
Exploramos las letras del icónico artista Gustavo Cerati para descubrir patrones ocultos y la evolución de sus temas a lo largo del tiempo.

## Análisis Semántico de Canciones Evolución Temporal de Palabras

**Herramientas:** Web Scrapping, Word2Vec, Gradio

Este proyecto combina técnicas de **web scraping** para la recolección de datos y **modelos de lenguaje** como Word2Vec para entender las relaciones semánticas de las palabras en las letras. Finalmente, se visualizan los resultados a través de una interfaz interactiva desarrollada con Gradio.

[Repositorio del proyecto en GitHub](#)



# Trabajos Prácticos de Cursada

Un recorrido por los hitos y aprendizajes clave del ciclo, mostrando la aplicación práctica de conceptos fundamentales en Ciencias de Datos e IA.

## Análisis Semántico

- 1 Exploración de técnicas de **Web Scrapping** y creación de **Embeddings** para comprender la estructura y significado del lenguaje. Se abordó la recolección de datos masivos y su transformación en representaciones vectoriales para análisis.

## Redes Neuronales

- 2 Implementación de **Redes Neuronales Multicapa** utilizando **PyTorch**, enfocándose en la arquitectura y entrenamiento de modelos para tareas de clasificación y regresión.

## MVP Deploy

- 3 Desarrollo y despliegue de un Producto Mínimo Viable (MVP). Utilización de **GitHub** para control de versiones y **Hugging Face Spaces** para la publicación de modelos de IA.

## MVP y LLM RAG

- 4 Proyecto final enfocado en la construcción de un MVP con un **Large Language Model (LLM)** y la integración de **Retrieval Augmented Generation (RAG)** a través de una API para mejorar la contextualización y relevancia de las respuestas.

# Proyecto Final: MIA Chatbot

MIA es una **IA interactiva de estudio** diseñada para facilitar el aprendizaje. Se basa en contenidos de **Minería de Datos** para ofrecer asistencia personalizada y respuestas precisas.

Consulta información del curso usando un modelo local con RAG.

LLM cargado Llamamiento: gemini\_2.0\_fish

Vectordate cargado correctamente.

Preguntas:

K-MEANS

Enviar

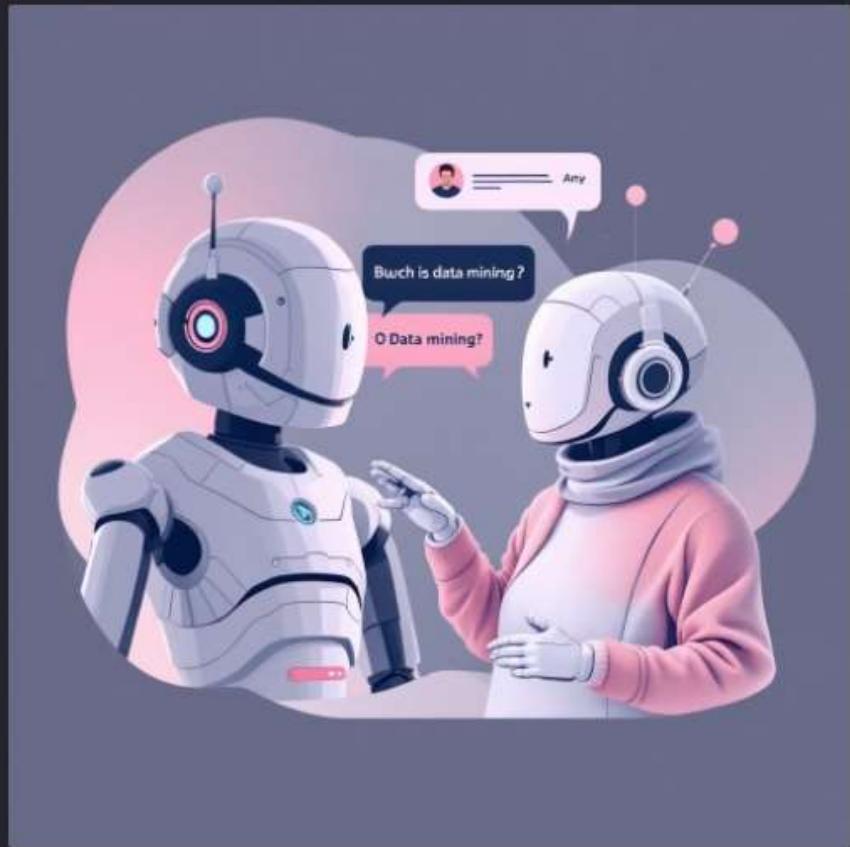
Responde:

K-means:

- La agrupación de K-means es sensible a la cantidad de agrupaciones definidas.
- Si la agrupación en Clusters no funciona bien con valores atípicos y conjuntos de datos con ruido.
- Las dimensiones variables de los puntos de datos no afectan al algoritmo K-means.
- La agrupación en clusters de K-means es más eficiente para grandes conjuntos de datos.
- A diferencia de DBSCAN, K-means requiere que se especifique el número de clústeres de antemano.

3 documentos utilizados

El objetivo principal de MIA es crear una experiencia de aprendizaje más dinámica y accesible, transformando la forma en que los estudiantes interactúan con la información compleja de minería de datos.



MIA integra un **LLM** robusto con una base de conocimientos específica, permitiendo un diálogo fluido y educativo. Es una herramienta poderosa para el repaso de conceptos, resolución de dudas y profundización en temas complejos.

Explora el código y la implementación de MIA en GitHub

Spaces Cumbiera-Intelectual-07 mia\_chatbot private • Running Logs

App Files Community Settings

## Sistema RAG - Minería de Datos

Consulta información del curso usando un modelo local con RAG.

LLM cargado correctamente: gemini-2.5-flash

Vectorstore cargado correctamente.

Pregunta:

Escribe tu pregunta aquí...

Enviar