

Análisis de Similitud Semántica entre Carreras Universitarias y la Oferta Académica de la UTPL

Paul Eduardo Espinosa Silva

Fecha de entrega: 13/06/2025

PRACTICUM 2.2-DATA & IA

Contenido

[1. Descripción del reto planteado 2](#_Toc200719243)

[2. Solución propuesta 2](#_Toc200719244)

[3. Tecnologías utilizadas 2](#_Toc200719245)

[4. Proceso de limpieza de datos 3](#_Toc200719246)

[5. Capturas de pantalla del panel 3](#_Toc200719247)

[6. Enlaces relevantes 3](#_Toc200719248)

[7. Conclusiones 3](#_Toc200719249)

[8. Recomendaciones 3](#_Toc200719250)

[9. Bibliografía 3](#_Toc200719251)

# 1. Descripción del reto planteado

El reto consistió en analizar y comparar la oferta académica de universidades ecuatorianas con las carreras ofrecidas por la UTPL, identificando coincidencias semánticas mediante modelos de lenguaje y visualizando los resultados en Power BI.

# 2. Solución propuesta

Se desarrolló un sistema que calcula la similitud semántica entre las carreras nacionales y las carreras de la UTPL usando el modelo MiniLM. El resultado fue exportado a un archivo Excel que sirvió como base para construir un dashboard dinámico en Power BI.

Resumen del Proceso de la Solución:

|  |  |  |
| --- | --- | --- |
| Fase | Descripción | Resultado |
| Limpieza | Eliminación de filas vacías y encabezados incorrectos | Dataset estructurado |
| Preprocesamiento | Asignación de IDs, nombres estandarizados | Dataset enriquecido |
| IA Semántica | Cálculo de similitud con SentenceTransformer | Archivo Excel con resultados |
| Visualización | Dashboard interactivo en Power BI | Panel con filtros, KPIs y gráficos |

# 3. Tecnologías utilizadas

- Python 3.13

- Librerías: pandas, openpyxl, sentence-transformers

- Modelo de IA: paraphrase-MiniLM-L6-v2

- Power BI Desktop

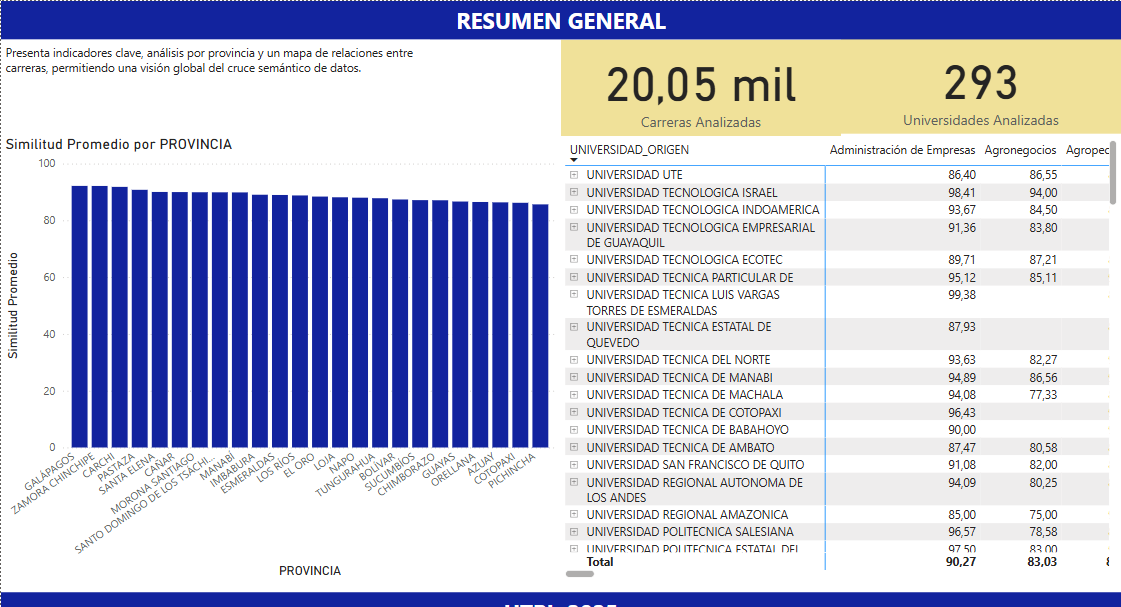
- Dataset: Oferta académica SENESCYT y carreras UTPL

# 4. Proceso de limpieza de datos

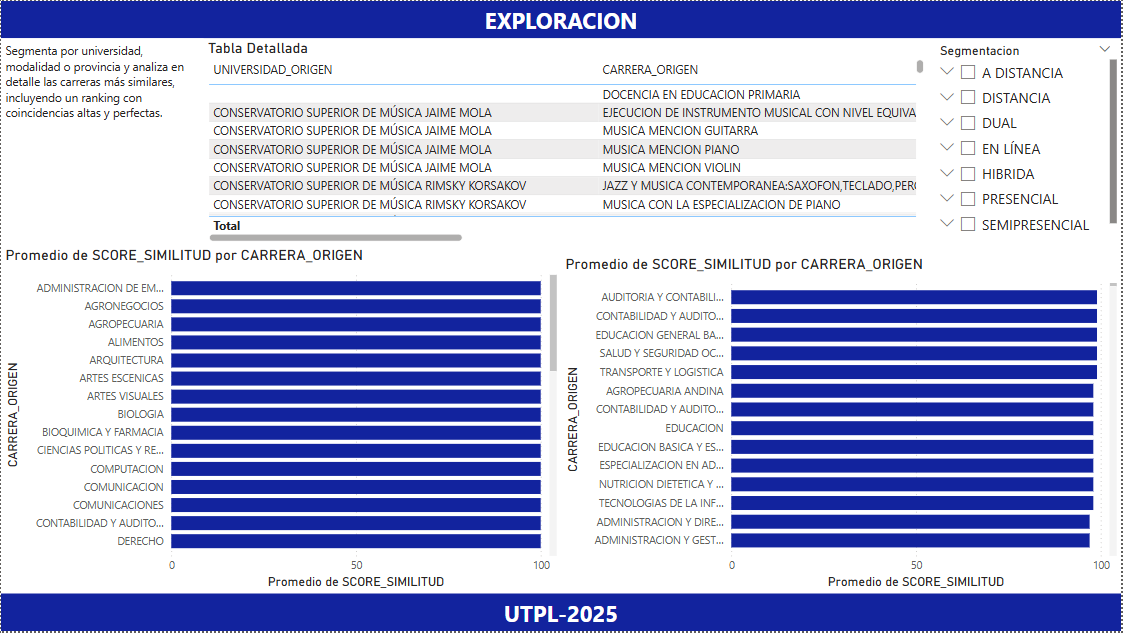
Se establecieron encabezados correctos, se eliminaron registros nulos, se estandarizaron los nombres de columnas y se generaron identificadores únicos por carrera.

# 5. Capturas de pantalla del panel

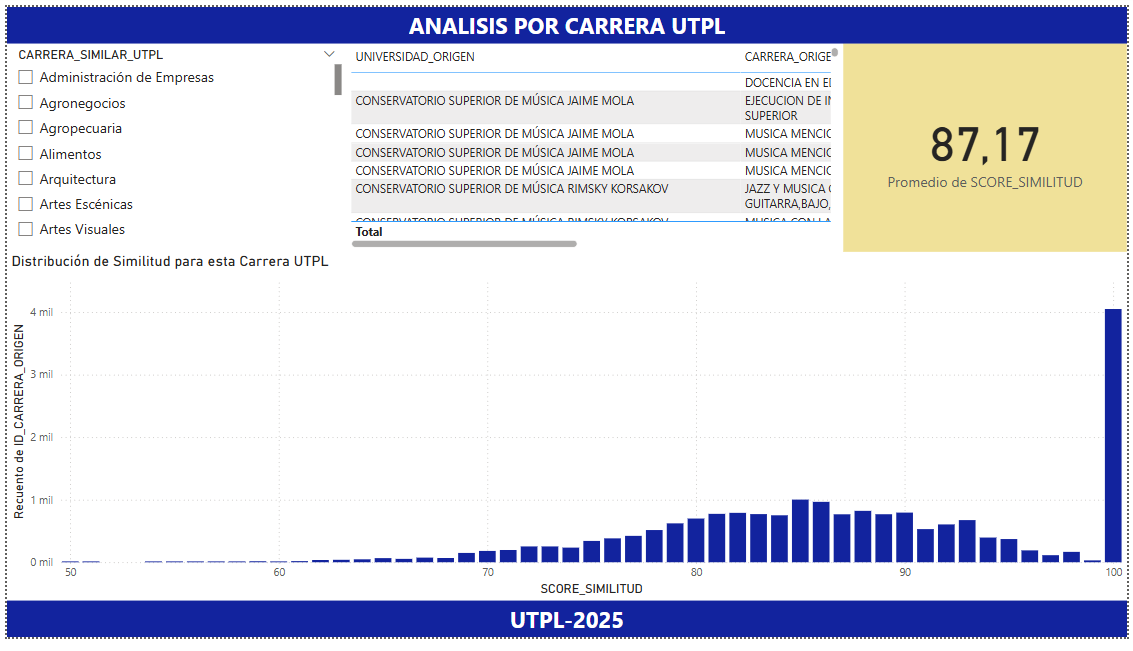
- Página 1: Resumen General → KPIs, gráfico por provincia y matriz de relación.



- Página 2: Vista Interactiva → Segmentadores y gráficos por score.



- Página 3: Análisis por Carrera UTPL → Filtro individual y análisis de distribución.



# 6. Enlaces relevantes

• Link al panel en Power BI Service: [URL]

• Link al repositorio GitHub: [URL]

• Archivo Excel generado: resultado\_semantico\_mejorado.xlsx

# 7. Conclusiones

El uso de modelos de lenguaje permite encontrar relaciones útiles entre carreras, incluso cuando los nombres varían. Power BI permitió construir una herramienta visual poderosa para explorar estos vínculos semánticos.

# 8. Recomendaciones

Para mejorar la precisión, se sugiere incluir descripciones detalladas de cada carrera. Además, se puede clasificar por campos disciplinarios para mejorar la agrupación semántica y reducir errores.

# 9. Bibliografía

- SentenceTransformers (https://www.sbert.net/)  
- Power BI Documentation (https://learn.microsoft.com/en-us/power-bi/)