



Módulo 3 Retos - Análisis de sentimientos en Pueblos Mágicos Mexicanos

Presentado por: Dr. Miguel Ángel Álvarez
Carmona

Junio 2025

Contenido

1. Introducción
2. Descripción del reto
3. Objetivos del reto
4. Datos
5. Requisitos para los participantes
6. Metodología de evaluación
7. Impacto
8. Calendario tentativo
9. Comité organizador del reto

Introducción

Este reto tiene como objetivo evaluar el desempeño de modelos de **Procesamiento de Lenguaje Natural (PLN)** en la tarea de **análisis de sentimientos (en español)** aplicado a opiniones de turistas sobre **Pueblos Mágicos Mexicanos**. Se utilizarán datos del foro de evaluación internacional **Rest-Mex**, el cual contiene más de **300,000 opiniones en español**, categorizadas por **polaridad del sentimiento (1 a 5, de negativo a positivo)** y **tipo de destino (atracción, hotel o restaurante)**. Para este reto solo se usaran unas 10 mil opiniones para entrenamiento y unas 5 mil para prueba.

Los participantes deberán construir modelos que sean capaces de predecir correctamente ambas etiquetas. Se proporcionará un **conjunto de entrenamiento con etiquetas** para que los participantes desarrollen sus modelos y, posteriormente, un **conjunto de prueba sin etiquetas**, cuyos resultados serán evaluados comparándolos con el **ground truth** mantenido por los organizadores.

Descripción del reto

Este reto plantea un problema real y desafiante en la aplicación del **análisis de sentimientos en español** dentro del sector turístico. La información que se analizará tiene un alto impacto social y económico, ya que puede ayudar a mejorar la experiencia de los turistas y la calidad de los servicios ofrecidos en los **Pueblos Mágicos Mexicanos**.

Los desafíos específicos del reto incluyen:

- Manejo de texto en español con expresiones coloquiales y variantes lingüísticas.
- Modelado de sentimientos en una escala ordinal (1-5).
- Clasificación del tipo de destino (atracción, hotel, restaurante).
- Evaluación de modelos en un entorno competitivo y con datos no vistos.

Objetivos del reto

- Diseñar modelos de PLN que clasifiquen correctamente la polaridad del sentimiento y la categoría del destino.
- Aplicar técnicas avanzadas como **modelos de lenguaje preentrenados (Transformers)**.
- Evaluar el rendimiento de los modelos utilizando métricas estándar en PLN.
- Fomentar el desarrollo de soluciones de IA para la industria turística.

Datos

Los datos se obtendrán del conjunto Rest-Mex, que contiene opiniones etiquetadas sobre destinos turísticos en México. Se utilizarán un total de 10,000 opiniones para el conjunto de entrenamiento y 5,000 opiniones para el conjunto de prueba.

Ejemplo de opiniones con sus respectivas etiquetas:

- Opinión negativa (Polaridad: 1, Categoría: Hotel)
 - *"El hotel estaba sucio y el personal fue muy grosero. No lo recomiendo para nada."*
 - Etiquetas: Polaridad = 1, Categoría = Hotel
- Opinión positiva (Polaridad: 5, Categoría: Atracción)
 - *"La visita al parque ecológico fue increíble. La atención y los paisajes son espectaculares."*
 - Etiquetas: Polaridad = 5, Categoría = Atracción

El flujo de datos será:

1. Liberación del conjunto de entrenamiento: Datos etiquetados para el desarrollo de modelos.
2. Desarrollo de modelos: Los participantes entrenan y ajustan sus modelos.
3. Liberación del conjunto de prueba: Datos sin etiquetas para la evaluación final.
4. Envío de predicciones: Los participantes envían sus predicciones.
5. Evaluación y publicación de resultados: Comparación con el ground truth y ranking de desempeño.

Requisitos para los participantes

- Conocimientos en **Procesamiento de Lenguaje Natural y aprendizaje automático**.
- Experiencia en programación en **Python** y uso de **Google Colab**.
- Familiaridad con herramientas de PLN como **spaCy**, **NLTK**, **scikit-learn** y **Transformers**.

Para la participación en el reto, se recomienda el uso de:

- **Google Colab** para el desarrollo y ejecución de modelos.
- **Bibliotecas de PLN y aprendizaje profundo**, tales como:
 - scikit-learn
 - TensorFlow / PyTorch
 - Transformers (Hugging Face)
 - spaCy y NLTK
- Computadora con acceso a internet para ejecutar los modelos en Google Colab.

Metodología de enseñanza y evaluación

Se brindará **seguimiento continuo a los participantes** a lo largo del reto mediante **coaching especializado**, asesorando en la optimización de modelos, ajustes de hiperparámetros y buenas prácticas en PLN. Se ofrecerán sesiones de retroalimentación y se facilitará un espacio de consulta para resolver dudas y mejorar la experiencia de aprendizaje práctico.

Los modelos serán evaluados en función de su precisión en la predicción de ambas etiquetas mediante las siguientes métricas:

- **F1-score macro**: Se calculará para cada tarea de clasificación.
- **Exactitud global**: Porcentaje de instancias clasificadas correctamente en ambas etiquetas.
- **Ranking de desempeño**: Los modelos serán ordenados según sus puntuaciones en F1-score y exactitud global.

Los resultados enviados por los participantes se compararán con el **ground truth**, y los modelos con mejor rendimiento serán destacados.

Impacto

Este reto tiene un impacto significativo en aplicar la IA en un problema real en la industria turística, ya que:

- Permite mejorar la experiencia de los turistas al analizar comentarios sobre destinos y servicios.
- Proporciona herramientas de IA para empresas del sector turístico en México.
- Promueve el uso de modelos de PLN en **español**, un área con menos recursos en comparación con el inglés.

Calendario tentativo para la impartición del módulo

Fecha	Actividad
16 de junio	Inicio del reto y liberación del conjunto de entrenamiento
17 de junio	Desarrollo y ajuste de modelos por los participantes
18 de junio	Liberación del conjunto de prueba y envío de predicciones
19 de junio	Evaluación de los modelos
20 de junio	Presentación de los resultados

Modalidades de participación

Por equipo

Listado de integrantes

Es importante mencionar que los 3 integrantes de esta propuesta tienen experiencia en organizar retos similares a este en foros internacionales como en el IberLef.

- Miguel Ángel Álvarez Carmona (Representante)
- Centro de Investigación en Matemáticas
- Doctor en Ciencias Computacionales.
- IA y Procesamiento de Lenguaje Natural.
- Ángel Ramón Aranda Campos
- Centro de Investigación en Matemáticas

- Doctor en Ciencias con orientación en Ciencias de la Computación
- Inteligencia Artificial

- Ángel Díaz Pacheco.
- Universidad de Guanajuato
- Doctor en Ciencias Computacionales.
- Inteligencia Artificial.