

Rest-Mex 2022

Sentiment Analysis And Recommendation System for Mexican Tourist Texts

Erika Rivadeneira.
Cipriano C. Hdz.

Mentors:

Adrian Pastor
Fernando S. Vega

Sentiment Analysis

*Given an **opinion of Tripadvisor** about a Mexican tourist place, the goal is to **determine the polarity**, between 1 and 5, of the text. **and the type of opinion** (hotel, restaurant or attraction)*

Bert-based Approach

— — —

Data (Example):

"Un callejón donde tienes que besar a tu amante por años de felicidad, en el amor es parte de un mito en esta ciudad especial. El callejón estrecho con escalones no es muy especial en sí mismo. Lo que lo hace especial es toda la historia a su alrededor."

- **Polarity:** 5 (Very positive)
- **Type:** Attractive

Results:

(BERT) BETO: 0.8858

(BERT) roBERTuito: 0.8789

Recommender System

*This subtask is a **classification task** where the participating system can **predict the degree of satisfaction** that a tourist will have when recommending a destination*

The data:

- Tourist
 - Gender: Male
 - Place of residence: Jalisco
 - Self-description: 'Me gusta Tomar fotografías por los lugares a donde viajo'
 - All user reviews that the tourist has previously issued*
- Place
 - Name: El muelle de San Blas
 - Description: Se encuentra justo al final de la autopista Tepic. Sobre a las calmadas aguas de la Bahía de Matanchén, se encuentra este muelle de más de 200 metros de longitud. Cuenta con distintas bancas semi-techadas para disfrutar de la brisa del mar e iluminación nocturna. En su extremo marítimo, se encuentra la estatua de “La Loca de San Blas”, un tributo a Rebeca Méndez Jiménez cuya historia se hizo famosa por la canción del grupo Maná “En el Muelle de San Blas”, la cual menciona al muelle en la cabecera municipal y no éste nuevo.
 - Some characteristics: Public, beach, familiar.
 - Satisfaction degree: 5 (Most positive)

Approach:



Only 922 of 1400 instances had complete information in train dataset.
587 of 680 instances had complete information in test dataset

BOW([Opinion, Place's Name, Place's Description]) + TFIDF

Classifiers: XGBoost and Gaussian Process (Equiparable to Ordinal Regression)

Results:

BOW of text then split -> **MAE 0.0** (1st run)

Split text then BOW-> MAE 0.6 (2nd run)

Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$