Maestría En Ciencia de los Datos

Programación II

Analysis of comments on Glassdoor

Andrés Felipe Vázquez Valdés 224808017

andres.vazquez0801@alumnos.udg.mx

Description Proyect:

The project analyzed a Database from Glassdoor. This DB shows a list of employees with some opinions regarding the company they are working for. (Link on last page)

The objective of this project is to be able to categorize their comments into XOV (Bad, Neutral, Good). To do that, I took the information on Pros, Cons, and Headlines compared with the general opinion of those comments (column Recommend).

I had to merge the comments of the three columns, remove link words, and symbols.

Then pull the unigrams and bigrams (I made two models, one of each)

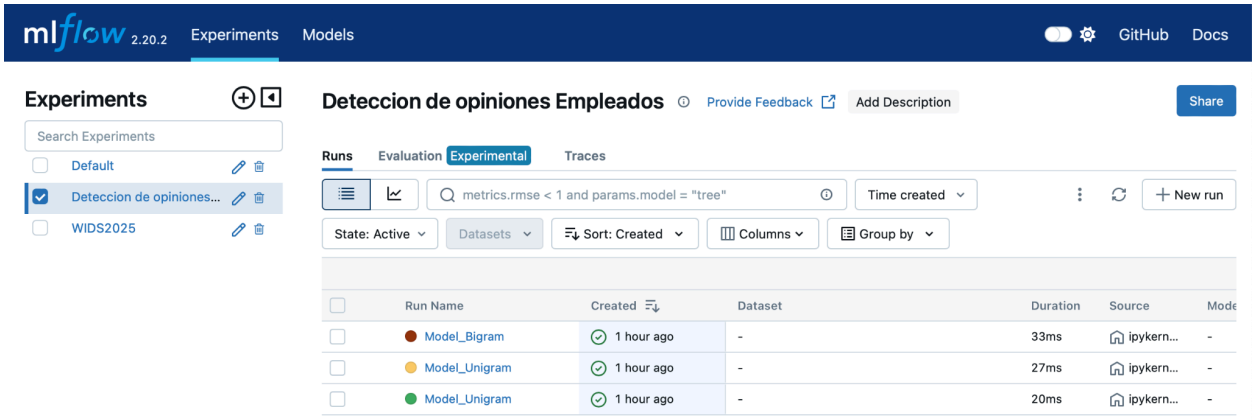
The Bigrams model gave an accuracy of 0.605 vs the accuracy of Unigrams 0.505.

Conclusion:

The model, with Bigrams worked much better than Unigrams. The thing here is that I had to drop a lot of the information to work with fewer employees because my computer wasn't able to process all the information.

It could be interesting to create another model using trigrams or more.

MLflow:



Link to Kaggle to get the information: Glassdoor Job Reviews

https://www.kaggle.com/datasets/davidgauthier/glassdoor-job-reviews