



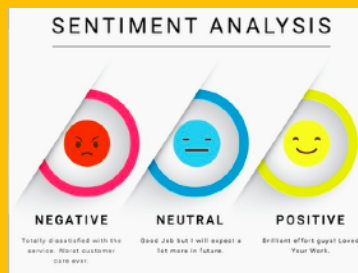
Capstone Project

Postgraduate Course on DataScience and BigData

2017-2018

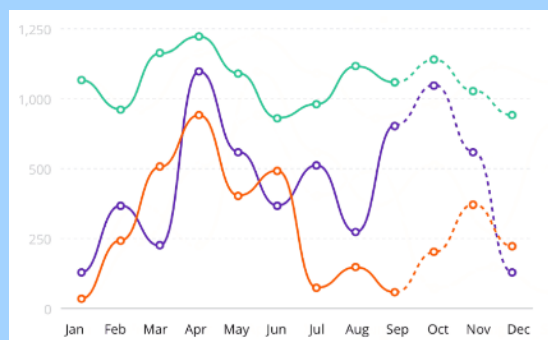
The Capstone Project

Sentiment Analysis



Data Understanding and Story Telling

What is beyond data?



Presentation

Sentiment Analysis

1st Part

Classify sentiment from the American Airlines tweets

2nd Part

Create a new dataset in Spanish

3rd Part

Classify sentiment from the new database

DEADLINE : June 15th

Twitter Airlines Sentiment Analysis

Determine the sentiments towards airlines using the data provided by Twitter.

20 teams · 10 hours to go

Overview Data Kernels Discussion Leaderboard Rules

[Join Competition](#)












Public Leaderboard

Private Leaderboard

This leaderboard is calculated with approximately 30% of the test data.

The final results will be based on the other 70%, so the final standings may be different.

[Raw Data](#) [Refresh](#)

| # | Δ1w | Team Name | Kernel | Team Members | Score ? | Entries | Last |
|----|-----|----------------------|--------|---|---------|---------|------|
| 1 | — | derichard | |  | 0.83904 | 18 | 16h |
| 2 | ▲1 | Hansel | |  | 0.82305 | 28 | 34m |
| 3 | ▼1 | acastroginard | |  | 0.82191 | 50 | 2mo |
| 4 | — | Carolina Migliorelli | |  | 0.81164 | 4 | 2mo |
| 5 | — | Tatiana | |  | 0.81050 | 23 | 14d |
| 6 | — | Pere Munar | |  | 0.80821 | 5 | 2mo |
| 7 | ▲1 | JordiBR | |  | 0.80136 | 31 | 1d |
| 8 | ▼1 | Miquel Vives | |  | 0.78995 | 6 | 2mo |
| 9 | — | Marc Roig | |  | 0.77853 | 2 | 3mo |
| 10 | — | Alberto Cruz | |  | 0.77625 | 20 | 15d |
| 11 | — | happybirds | |  | 0.76027 | 3 | 1mo |

Spanish Airlines Tweets DB

- Test your old model with the new dataset, the columns provided are the following:
 - *tweet_id*
 - *airline_sentiment*
 - *retweet_count*
 - *reply_count*
 - *text*
 - *tweet_coord*
 - *tweet_created*
 - *tweet_location*
 - *is_reply*
 - *user_timezone*

Spanish Airlines Tweets DB

| airline_sentiment | reply_count | retweet_count | text | tweet_coord | tweet_created | tweet_id | tweet_location | user_timezone |
|-------------------|-------------|---------------|--|-------------|--------------------------------|----------|----------------|---------------|
| negative | 0 | 0 | @davidrsalcedo En el mismo sitio donde regalan las α7R III. Y no es en vuelo de Iberia...sino en el mundo de los sueños :-P ^Jose | None | Wed Oct 25 10:00:48 +0000 2017 | 84222976 | None | Madrid |
| positive | 0 | 0 | Iberia suma y sigue: Destinia establece conexión NDC Nexotur https://t.co/Q9xAHYY1HC | None | Wed Oct 25 10:04:14 +0000 2017 | 46814720 | None | Ljubljana |
| negative | 0 | 0 | @aguilar_adolfo @Iberia Y que lo digas, @aguilar_adolfo hace poco en un vuelo transoceánico de más de 10 horas no p... https://t.co/DKCZS599Wn | None | Wed Oct 25 10:04:42 +0000 2017 | 20412673 | None | Hong Kong |
| neutral | 0 | 0 | @SamsungEspana @Iberia Y nadie lo devolvio? | None | Wed Oct 25 10:06:07 +0000 2017 | 29387265 | None | Madrid |
| neutral | 0 | 0 | @Iberia Cuanto tiempo pasa entre la realización de un vuelo y que se añadan los avios/PE que genera? | None | Wed Oct 25 10:06:56 +0000 2017 | 73759744 | None | Amsterdam |
| negative | 0 | 0 | Gracias @Iberia por todos los ataques de ansiedad que tengo. Antes de despegar revisar bien los aviones | None | Wed Oct 25 10:08:19 +0000 2017 | 27649536 | None | Amsterdam |

Spanish Airlines Tweets DB

Some Statistics

| | #Tweets |
|---------------|---------|
| Positive | 1.841 |
| Neutral | 3.202 |
| Negative | 4.696 |
| Total labeled | >30.000 |

Presentation/Blog

- **When:** July 1st
- **What:**
 - Online article (blog, github, website...)
 - Oral Presentation (10 minutes) explaining
- **Including:**
 - Sentiment Analysis Pipeline (data, models, evaluation, results, conclusions)
 - StoryTelling (questions, answers and visualizations)

Next Session: May 24th