



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Saudi Authority for Data and
Artificial Intelligence
Data Science Bootcamp

Arabic tweets sentiment analysis (MVP)



Done by: Sarah Hamad Alhussiny

The goal of this project is to understand the fetchers that more affect tweets sentiment, and how these fetchers relate to whether a tweet's category is positive or negative.

To start exploring this goal, I used a logistic regression model with one feature to product tweet category.

We can evaluate the model using by confusion matrix as paper in figure (1) this is the result when using (word density, sentence density) features.

```
array([[3035, 2559],  
       [2789, 2936]])
```

FIGURE 1: CONFUSION MATRIX 1

And figure (2) shows the confusion matrix when using TF-IDF (word embedding) as a feature.

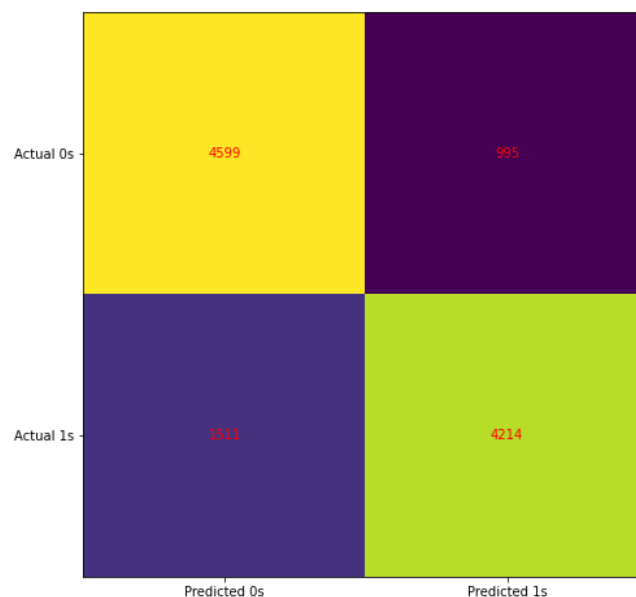


FIGURE 2: CONFUSION MATRIX 2

This result suggests that TF-IDF (word embedding) may have a significant impact on tweets' sentiment. However, the magnitude of the model's residuals makes it clear that is not the only important factor.