

- **Abstract:**

Nowadays, people from all around the world use social media sites to share information. Twitter for example is a platform in which users send, read posts known as 'tweets' and interact with different communities. Users share their daily lives, post their opinions on everything such as brands and places. Companies can benefit from this massive platform by collecting data related to opinions on them.

- **Design:**

The clustering model was analyzed. In this project, the SVM algorithm was used to classify Arabic tweets to 4 clusters (Sports ,Religion,Politics and Economy) .

- **Data:**

Columns: ▪ User name (text). ▪ Tweet or text (text) ▪ Hashtags (text). ▪ Source (text) ▪ Label (Text).

Data size (no. of rows and no of columns): ▪ 30k rows and 4 columns.

- **Algorithms:**

Topic modeling : Latent Semantic Analysis(LSA), Non-negative Matrix Factorization(NMF).

Model selection : Logistic Regression , support vector classifier and Naiive Bayes.

Tools: jupyter, Colab, Pandas, NumPy and NLTK.

- **Communication:**

- This chart is represent the 100 common in politics words.



- This chart is represent the 100 common in economy words.



- This chart is represent the 100 common in sports words.



- This chart is represent the 100 common in religion words.

