 Done By : Rawan Al-Qahtani

Instructor : Mejdal Al-Qahtani

T5 Bootcamp Project Proposal

Data scientists Arabic hate & offensive speech classifying

# Question/Problem statement:

Recognize and classify Arabic words to hate and offensive.

# Data Description:

Main data source ALT *(Arabic language technologies)***.**

**(**[**http://alt.qcri.org/~hmubarak/offensive/AJCommentsClassification-CF.xlsx**](http://alt.qcri.org/~hmubarak/offensive/AJCommentsClassification-CF.xlsx))

It’s collected from official websites

* + Columns: contains 15 features.

words , serial, link , languagecomment\_gold, insdt, bodylen, body, articletitle, languagecomment:confidence, languagecomment, \_last\_judgment\_at, \_trusted\_judgments, \_unit\_state, \_golden, \_unit\_id

* + Data Type: Integer, Float, Bool, Object.
  + Rows: contains 31693 rows
  + Data size: contains 31693 rows × 15 columns.

# Tools:

## Programs:

We will use Python ,anaconda jupyter.

## Libraries:

nltk,, pandas,matplotlib, sklearn, keras , numpy .

## Functions:

We will use several functions that we define in the project.

## Plots:

We will use one of them (lineplot, Scatter, bar, hits, pie, circle).

# MVP Goal:

The goal of this project is to classify the Arabic hate and offensive words.