



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>)

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 \times 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.