

Arabic Offensive Language Detection

Mentors: Hamdy Mubarak, Kareem Darwish, Ahmed Abdelali {hmubarak, kdarwish, aabdelali}@hbku.edu.qa

Motivation: Social media sites (Twitter, YouTube..) try to **filter out** obscenities and offensive content but they **fail** in many cases. On Twitter, many offensive tweets still appear, and some normal tweets are marked as offensive!

Available Resources: 1,100 tweets (Egyptian dialect) and 32,000 comments on Aljazeera classified (obscene, insult/attack/threat, normal)
More info on our paper: <https://aclanthology.info/papers/W17-3008/w17-3008>

قناة الجزيرة @AJArabic
هذه قصة المواطن البريطاني علي عيسى الذي يتهم الأمن الإماراتي بضربه واعتقاله بسبب تشجيعه #قطر في كأس آسيا

Sakib Bishara
Replying to @AJArabic
يا قناة الجزيرة يا قناة الخزي والعار يا ماجورين يا مرتزقه..
O channel of Aljazeera, O channel of shame, you are mercenaries.

Show additional replies, including those that may contain offensive content Show

Open: normal tweets!



Comment	Type
كذاب ابن لك ** ابن ق *** Liar, son of the ***	Obscene
الارهابي انت و اجدادك You and grandparents are terrorists	Attack
كلب وجب قتله A dog who must be killed	Violence
لأن العرب عبيد وسيبقون كذلك Arabs are slaves and will remain	Racism
عيب تحكمننا واحدة ست Shameful to be ruled by a woman	Sexism

Examples of offensive comments

❖ Solution

Build a **website** to do the following tasks:

- * For any input tweet or comment, use different classification **algorithms** (fastText, DNN, etc.) to detect offensiveness level, and **spot** offensive words.
- * Identify pro-offensive **users** and crawl their history of tweets.
- * Send **notifications** to authors of offensive content (or report to Twitter)
- * **Augment** data: collect more offensive tweets, study linguistic patterns, apply to other dialects, etc.

❖ Skills

- * Experience in Java/Python
- * Experience in web development
- * Some experience in machine learning
- * Understanding Arabic is a plus!