Team Omega

Or

Gender classification through text analytic

**The goal**

Our main goal is to be able to say if a text is written by a woman or a man.

In term of business, with this idea we aim to enhence marketing targeting. Third party advertisement companies can have better knowledge on Twitter’s gender distribution. Thus, they can advise companies that want to make advertisement on social network like Twitter.

**The dataset**

to lead our research, we use a huge dataset of twitter. The dataset records over **25’000 tweets**. Every row gives the following information:

* **User’s Gender**
* User’s name
* Pofile picture
* **Retweet’s number**
* **Sidebar Color**
* **Tweet content**

dataset’s link : <https://www.kaggle.com/crowdflower/twitter-user-gender-classification>

**The method**

1. Define objective

Help targeting in social network advertisements.

1. Collect data

Une image contenant texte

Description générée automatiquementWe collect the dataset on Kaggle about the tweet and the gender.

1. Prepare data
2. Remove the useless column
3. Remove the rows who have no gender or the rows that contain no gender but a brand
4. Split 20%-80% between train and test set.
5. Merge all the tweet by gender
6. Tokenization the text
7. Remove the stop-word and parasite
8. Stemming & lemmatization
9. Select algorithm
10. Train model
11. Test model
12. Predict
13. Deploy