# Twitter User Gender Classification

Given an English sentence of sufficient length, I would like to figure out the gender of the author with sufficiently high accuracy. One of the things that is missing from Twitter is an indication of a person’s gender. It is reasonable to think that males and females write words different, such as “husband” or “wife”. My goal is to determine the gender of user from a random one tweet.

User gender analysis may have useful applications, such as machine translation. Some languages employ different grammatical structures for depending on the gender of the user, whereas in English no such thing exists. The result can also work for other online forum, not only twitter.

The data are gathered from Kaggle. I decide to choose Bayesian classification to train the data, with bigrams or trigrams as feature. I also consider to use deep learning to train the sentence, since the output may only be male ore female.

Data source: https://www.kaggle.com/crowdflower/twitter-user-gender-classification