Gender Classifier

abstract

The data was collected from Twitter, which consists of selecting random, tweet and determining the gender of the user (female, male, brand). We can take advantage of this dataset by analyzing the text to find out the most used word for each gender.

Design

The data show us the most frequent words based on gender

Data

The dataset contains 26 feature and more than 20,000+ observation. We can use it to predict the user gender.

Tools

-Pandas and numpy

-Nltk for NLP and stopwords

-CountVectoizer

-Seaborn and matplot

-For classification algorithms (Naïve bayes, Logistic regression, Random forest)