Model cards - Tweeter Gender Classification

Vijaya lakshmi Kuruba - Houda Saadaoui

March 11, 2021

Model Card - Tweeter Gender Classification

Model Details

- Developed by Houda Saadaoui and Vijaya Lakshmi Kuruba as an academic requirement of the course IFT 6390 at the University of Montreal in Winter 2021.
- This is an exploratory SVM linear classifier which predicts the gender of the user based on its Tweets and profile description.
- SKlearn API is utilised for model creation. We have used SGD classifier with Hinge loss with L2 penality which is equivalent to SVM linear classifier.
- Released under MIT License.

Intended Use

- Intended to be used for anyone interested in the relationship between type of writing and gender.
- Particularly intended for people who are interested in analysing social media use.

Factors

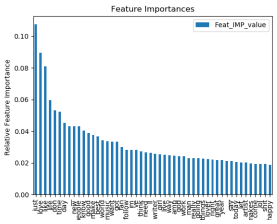
- The prediction of the model is within the scope of the social media users and particularly, Twitter user. Therefore, relevant factors are age and level of education as most of the Twitter users are young and have higher education.
- Language is another relevant factor as the model was trained with English Tweets and English profile description

Metrics

- Model is evaluated with reporting the accuracy score of the classifer.
- A ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR). ROC response of different datasets, created from K-fold cross-validation.
- Model hyperparameter is tuned using GridSearchCV by choosing the best alpha. Alpha is the regularization term and learning rate.

Training Data

- This model uses the dataset for the CrowdFlower AI team. The dataset contains 20,000 random tweets and account profile with the label corresponding to the gender of the user.
- Dataset is first filtred using only the gender ('male','female') as it has other irrelevent gender values
- Dataset is preprocessed using the tweet-preprocessor library.
- Dataset is split to 80-20 ratio to account for train and test data.
- Text features where vectorized using all the dataset vocabulary .



Evaluation Data

- Data is split to 80-20 ratio to account for train and test data
- The model has an accuracy of 66.23 %

Ethical Considerations

- Because of the population in which the study was conducted, the results should be carefully generalized to other populations.
- In no case can this model be used to make decisions involving gender discrimination.

Caveats and Recommendations

- We have considered only tweets and profile description for classification
- Model performence can be improved by including more features based on exploratory analysis.
- Model performence can be improved by using a larger dataset.

Quantitative Analyses