Table of Contents

Data Extraction, Cleaning and Transformation

Facts

Columns

Rows

Cleaning – Removed usernames,

(Before)

Data Extraction and Transformation

Added cleaned tweets column

Deleted original tweets column

Cleaned data exported as a CSV file for machine learning

(After)

Data Model implementation and Optimization

Cleaned Data, Normalized Data, Standardized data prior to modeling

Initialized, trained, and evaluated model or loads a pretrained model

Documented evidence of model optimization using tuning logic

Model demonstrates predictive power, region, prediction classification accuracy

(Bar Graph)

(Add Confusion Matrix)

Conclusion:

Best classification model is:

Accuracy:

Without stop words better than with stop words

Ngrams better learning and more accurately classified hate speech.

Future Considerations

User twitter API to obtain updated dataset

Pull annual dataset from twitter to analyze hate speech over time

Run through trained model

Pull new dataset from twitter

Multi-label classifications for hate/offensive/neutral

Problems and Limitations

Deciding on the best classification model

Time constraints prevented us from further optimization: multi-label classification