



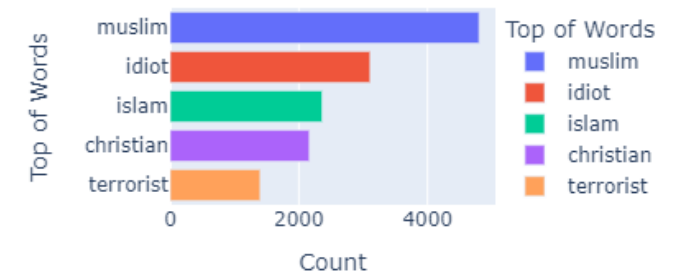
Background

What happened?

[illegible]

Most of many **users** type
in words like **bully, school,**
muslim, girl, idiot, dumb,
nigger, fuck when they
cyberbully on twitter

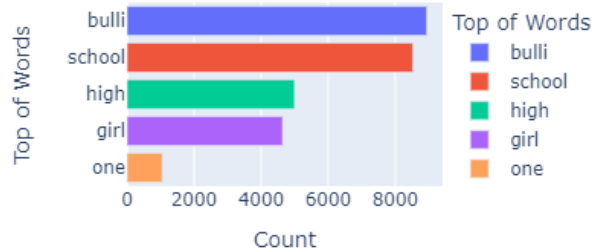
What words are often used in gender & religion cyberbullying?



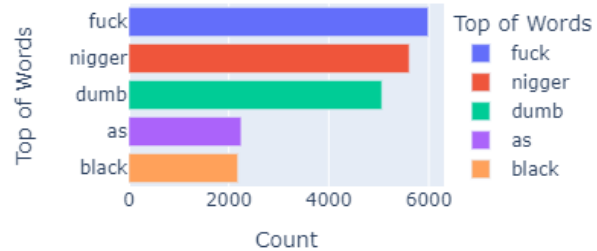
Most of many **users** type in words like **joke**, **rape**, and **gay** when they did **gender cyberbully** on twitter. Meanwhile, the word **muslim** is the word that is **most** often mentioned when Twitter users carry out **religion cyberbullying**

What words are often used in age & ethnic cyberbullying?

Top 5 Words in Age cyberbullying Tweet



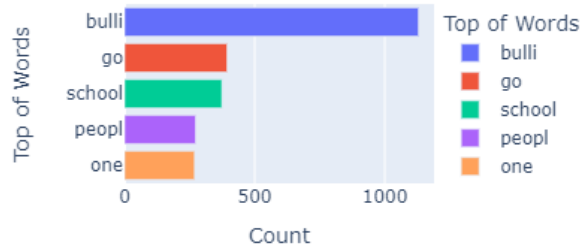
Top 5 Words in Ethnic Cyberbullying Tweet



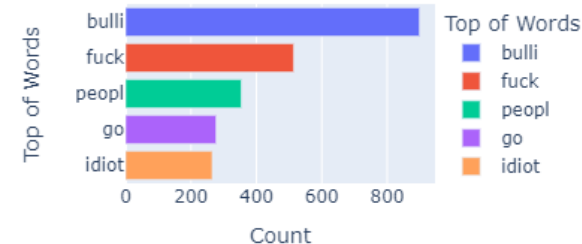
Users often tweet with ridicule in the form of **bully** and **school** when doing **age cyberbullying**. Meanwhile, when mocking certain **ethnicities**, tweeter users often use the words **fuck**, **nigger**, and **dumb**

What words are often used in other & not cyberbullying?

Top 5 Words in not cyberbullying Tweet



Top 5 Words in Other Cyberbullying Tweet



Users who do **not justify cyberbullying** behavior **tweet** by including the **word bully** in the sentences they type. **As well as for users** who do cyberbullying with **other categories**. Not only that, the word **fuck** is also typed when other users are cyberbullying in **other categories**

Pre-processing

I removed stopwords, case folding, stemming, and lemmatizer words as follows.

tweet_text	cyberbullying_type	clean_tweet
In other words #katandandre, your food was cra...	not_cyberbullying	word food crapilici

Features Extractions

I did TDIDF method on feature extraction

Model Development & Evaluation

- **Model:** I use **Logistic Regression, Support Vector Machine, and Multinomial Naïve Bayes classifier**
- **Evaluation:** I used a **cv score** with the **best parameters** and the result was that **SVC** produced the best performance with a cv score of **0.833** with a **penalty parameter** of the error term of **1**

Conclusion

The support vector machine is the best machine learning model for predicting cyberbullying tweets with **83% of accuracy**. It has a **weighted average precision** percentage of **82%**. Even so, the **sensitivity** obtained is **83%**. It means that about **17% of positive cases missed the model's predictions**

Click [here](#) to see mycode