Logistic Regression Algorithm

Logistic regression is a supervised learning classification algorithm that can be used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there would only be two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (success/yes) or 0 (failure/no). As to achieve our main objective here, we would be using Logistic Regression algorithm as one of the models to predict the cyberbullying in social media posts which we had chosen Formspring.me as the platform to mine the desired data. As to proceed, we decided to train our dataset with binary logistic regression model which would be the simplest form of logistic regression. The result of this prediction would turn out as either 1 (true/bully post) or 0 (false/normal post). To conduct this analysis, we used python language as to gauge the quality of this model. Here are the results after running the model:

Classification Report: precision recall f1-score support

0 0.97 0.99 0.98 2177

1 0.88 0.61 0.72 145

accuracy 0.97 2322

macro avg 0.93 0.80 0.85 2322

weighted avg 0.97 0.97 0.97 2322

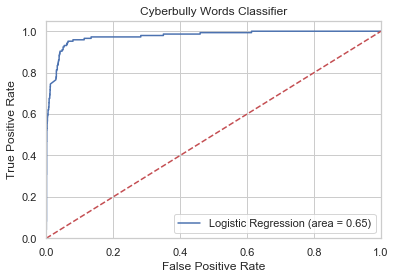
Confusion Matrix:

[2165 12]

[ 56 89]

As for interpretations of these results, we could conclude that this model did churn out an average result even with good accuracy since the recall is barely above 0.5 which is 0.61 as a good model should have a recall which is near to 1. Looking at the confusion matrix, we managed to classify 2165 words correctly which is true positive and 89 words incorrectly which is false negative.

ROC Curve:



As to assess the model more, we managed to carve out AUC score for this model which turned out to be 0.65. Having it close to 0.5 making this classifier as slightly not suitable in achieving our target as the closer the score to 1, the better the classifier. In conclusion, we would decide to move on with another classifier that possesses better results and quality than logistic regression model does.