Cyber Bullying Results

Predictions / Models	Training Accuracy		Validation Loss		Test Loss	
	Count Vectorizer	*TFIDF Vectorizer*	*Count Vectorizer*	*TFIDF Vectorizer*	*Count Vectorizer*	*TFIDF Vectorizer*
Logistic Regression	0.826	0.822	0.419	0.448	0.416	0.451
Naive Bayes	0.770	0.767	0.686	0.632	0.707	0.635
XGBoost	0.834	0.831	0.398	0.403	0.390	0.399
SVM	0.795	0.808	0.514	0.481	0.503	0.468
Sent2Vec	0.781		0.554		0.560	
Vanilla ANN	0.883		0.551		0.531	
Bidirectional LSTM	0.883		0.419		0.414	

Machine Learning Models:

We have the following models as our Machine Learning models for the project:

1) Logistic Regression
2) Naive Bayes
3) XGBoost
4) SVM

For the Logistic regression model,
From the above predictions, we can see that Count Vectorizer would be a better vectorizer as compared to TF-IDF vectorizer as, even though the accuracies from both the vectorizers are almost similar, it gives a better training accuracy as compared to TFIDF and as we can observe, the validation loss and test loss are less. Thus, we can conclude that the Count vectorizer performs better in the case of Logistic regression model.

For the Naive Bayes model,
From the above predictions, we can see that TF-IDF Vectorizer would be a better vectorizer as compared to Count vectorizer. Even though the accuracy of the Count Vectorizer is slightly better than that of TF-IDF, the validation loss and test loss of TF-IDF are less than that of the

count vectorizer. Thus, we can conclude that the TF-IDF vectorizer performs better in the case of Naive Bayes model.

For the XGBoost model,
From the above predictions, we can see that Count Vectorizer would be a better vectorizer as compared to TF-IDF vectorizer. Even though the accuracies from both the vectorizers are almost similar, it gives a better training accuracy as compared to TFIDF and as we can observe, the validation loss and test loss are less in the case of the count vectorizer. Thus, we can conclude that the Count vectorizer performs better in the case of the XGBoost model.

For the SVM model,
From the above predictions, we can see that TF-IDF Vectorizer would be a better vectorizer as compared to Count vectorizer. Along with better accuracy of the model using TF-IDF, the validation loss and test loss of TF-IDF are less than that of the count vectorizer. Thus, we can conclude that the TF-IDF vectorizer performs better in the case of SVM model.

Among these Machine Learning models, we can conclude that XGBoost model which uses Count vectorizer gives the most accurate results along with the least amount of validation and test loss.

Word Vectors:

Using the GloVe XGBoost algorithm, we can see that it gives a good training accuracy but it is not better than the accuracy we received from the XGBoost model. The validation loss and test loss are more than that of the XGBoost model. Thus, we can say that the machine learning model, XGBoost is better than the GloVe XGBoost model.

Deep Neural Network Models:

We have the following models as our Deep Neural Network models for the project:

1) Vanilla Neural Network
2) Bidirectional LSTM

When we compare the two models listed above, we can observe that both the models have the same training accuracy. But in the case of loss, the Bidirectional LSTM has lower loss values for both validation loss and Test loss. Thus, we can conclude that the Bidirectional LSTM performs better than Vanilla Neural Network model.

Conclusion:

Among these three models, we can see that Bidirectional LSTM has better accuracy than that of XGBoost, the validation loss and test loss for XGBoost is less than that of Bidirectional LSTM.

Thus by our analysis, we can say that XGBoost is a good model for our project.