RFN: **Kenneth Tochukwu NWALA**

RD: **Computer Science**

RS: **Computing and Engineering Sciences**

RE: **nwalakennethtochukwu@gmail.com**

RAE: **nwala1126@pg.babcock.edu.ng**

RP: **07067017411**

RT: **Text-Based Cyberbullying Detector Using Linear Support Vector Machine Framework**

MS: **Prof. Olumide LONGE**

ME: **longeolumide@gmail.com**

SP: **080160900893**

CS: **Nil**

CE: **Nil**

CP: **Nil**

AB: Cyberbullying has become a pervasive threat to internet users, as it involves the transmission of intimidating messages through popular social networking sites like twitter, Facebook, LinkedIn and Reddit. Despite the benefits of social networking, users have been humiliated, insulted, and harassed by anonymous users, strangers, or peers online. Previous studies showed that the effects of cyberbullying on Twitter are significant on the victims, sometimes leading to depression and suicide. Currently, the efforts by the social networking sites in curbing cyberbullying is largely user centered, reactionary and ineffective because of the manual and human intervention involved in the process which makes it inefficient. Therefore, this research aimed at developing a text-based cyberbullying detector model using Linear Support Vector Machine (LSVM) framework to address the aforementioned problems.

LSVM, a machine learning technique was used to build a sentiment analysis model to detect instances of cyberbullying as the tweets are being composed. Twitter was purposively chosen for this study because it provides a rich source of public data that is easily accessible for research purposes and it is the most susceptible social networking site to cyberbullying. A total of 1.06 million tweets and 17 million combined text-based elements were gathered as data and studied for 5,000 streaming sessions from twitter platform using Twitter streaming RESTful application interface. The datasets were further trained, tested and validated in order to determine the positive and negative connotation of words. The trained data were

subjected to Natural Language Processing Tool Kit (NLTK), and Statistical Natural Language Processing (SNLP) to discern between the positive and negative connotation of words. A Natural Language Processing pipeline was used to design the model, where the output is fed into another system. Precision, F1-score, Recall was used as the metrics to evaluate the model. Support Vector Machine (SVM) and Python using Visual studio code Integrated Development Environment were further used for the classification of words that identifies only two values, 0's and 1's for negative and positive words respectively.

The results obtained from the model showed that both SVM and LSVM had a recall of 81% and 80%, F1-score had 69% and 74%, Precision had 79% and 78% respectively. The NLTK had 73% percentage accuracy while the SNLP had 79%. This sets SVM and LSVM apart in contrast to 79%, which is higher in computation and ineffective. The SVM classifier coupled with the LSVM emerged as the best classifier while using sentiment to classify text documents, with an accuracy of 94.95% when subjected to SNLP.

In conclusion, the model indicated that the framework provided a feasible solution to detect cyberbullying behavior and its severity on twitter platform. The model should be adopted by twitter and other social networking sites to reduce cyberbullying.

Keywords: Cyberbully, Machine learning, Naïve Baye algorithm, Sentiment analysis, Support Vector Machine, Text classification, Twitter

Word Count: 450