**Individual In-depth Report**

**Member name:** [Avani Mundra](mailto:amudra@asu.edu)

**Evaluated by:** [Rahul Nayak](mailto:rrnayak@asu.edu)

**Date: 09/23/2023**

**Tasks Assigned:**

* This report summarizes the paper ‘A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms’ in depth.

**Summary:**

* The paper tackles the escalating issue of cyberbullying within the context of social media, with a primary focus on Twitter. The rapid expansion of social media platforms has ushered in a significant concern, with cyberbullying adversely impacting a large user base.
* The study's methodology involves the collection of 1000 labeled tweets related to cyberbullying from Twitter. Data preprocessing includes text cleaning, removal of non-English words, lowercasing, and stop word removal.
* Two critical machine learning models are developed: one for detecting the sentiment of cyberbullying instances and the other for categorizing cyberbullying into different types.
* Feature extraction techniques, including Term Frequency-Inverse Document Frequency (TFIDF) and Bag of Words (BoW), are used alongside machine learning algorithms like Logistic Regression, Multinomial Naive Bayes, Random Forest, and Linear Support Vector Machines (SVM).
* The paper introduces a novel conceptual framework, comprising three integral modules: User Interaction, Analysis, and Decision Making. This framework serves as a dynamic system tailored to real-time cyberbullying detection across diverse social media platforms
* The study highlights the significance of its framework as a generalized prevention measure for all social media platforms. It acknowledges certain limitations, such as the use of a limited dataset, reliance on a small set of NLP techniques, and a focus on Twitter data

**Outcome:**

* The findings underscore the superiority of the Random Forest algorithm with TFIDF embedding. It achieved impressive results with an F1 score of 80.8% for identifying cyberbullying and 58.4% for classifying the type of cyberbullying.
* The paper calls for future research to expand the scope to other social media platforms and address different forms of cyberbullying, including memes and trolling.

**References**

[37] S. Mitra, T. Tasnim, M. A. R. Islam, N. I. Khan and M. S. Majib, "A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms," 2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2021, pp. 1-4, doi: 10.1109/IC4ME253898.2021.9768450.

**Evaluation of Report**

**Evaluation summary with justification.**

The paper presents a cyberbullying detection approach on Twitter using machine learning and a novel conceptual framework. While it demonstrates promise, its reliance on a limited dataset and a small set of NLP techniques raises concerns about generalizability. Further validation on larger datasets and broader social media platforms is needed for robustness.

**The quality of the major result(s) with justification.**

Key Result:The model achieved an impressive F1 score of 80.8% for identifying cyberbullying and 58.4% for classifying its types. This quality is significant because it demonstrates the effectiveness of the chosen algorithm-embedding combination in tackling cyberbullying on Twitter, providing valuable insights for future research and applications.

**The usefulness of the paper to the overall project.**   
  
The Machine learning techniques used in this paper can be further used to detect cyber bullying which is a type of malicious activity on social media, successfully targeting the main goal of this project.

**Other comments**

No Comments

**Evaluation Approval  
  
Evaluation by:** [Rahul Nayak](mailto:rrnayak@asu.edu) **Date: 09/24/2023**

**Is the written report of the in-depth study complete with all the major result(s) of the paper(s)? If not, provide as many examples of the major result(s) missing in the written report as possible. (in bullet form). [Normally within 100 words]**Yes the In depth study is comprehensive and summarizes all the results effectively.

**Is each section of the guidelines sufficiently completed? If not, point out what is missing. [Normally within 40 words].**

Yes, each section of the guidelines is sufficiently completed.

**Is the quality of this version of the written report satisfactory? If not, then why not? [Normally within 40 words]**

Yes the quality of this version of the report is satisfactory.

**Approval.  
  
Approved by:**[Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date:09/25/2023  
  
Is the quality of this written in-depth study report and Evaluation report satisfactory? If not, then why not? (limit: 40 words)**

The quality of this written in-depth study report and Evaluation report is satisfactory. The evaluation report signifies correct evaluation and the report itself justifies the project topic by addressing the detection process of suspicious (cyberbullying) activities in Twitter dataset.