**Individual In-depth Report**

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

**Evaluated by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Date: 09/17/2023**

**Tasks Assigned:**

* Reviewing the paper ‘An approach to detect abusive content incorporating Word2Vec and Multilayer Perceptron’ in depth.

**Summary:**

* The aim of the paper is to detect suspicious activities and hateful content in social media using deep learning techniques of Word2Vec and Multilayer Perceptron Neural Network.
* The paper starts with an introduction to the definition of abusive content, hate speech and how it is used by influences to manipulate the general public with incorrect and sometimes offensive information.
* The paper highlights the challenges in abusive text detection, such as grammatically correct abusive sentences, implicit abusive text, and overlapping content identification, and introduces a novel research approach using compositional vector models and word embeddings with MLP classifiers for abusive content detection on Twitter.
* The proposed abusive content detection approach utilizes word embedding and a compositional vector model to extract tweet vectors for detection using a deep learning classifier, the Multilayer Perceptron (MLP).
* The process involves four main steps: preprocessing, word embedding, compositional vector model, and MLP classification.
* Preprocessing includes several steps like removing user names, emojis, punctuation, symbols, URLs, email IDs, English stopwords, and converting line breaks into single spaces. Lemmatization is applied to reduce word inflections.
* Word embedding is based on the Word2Vec model, specifically the Continuous Bag of Words (CBOW) variant, which represents word vectors for each tweet in a semantic space with 200 vector dimensions.
* The compositional vector model computes tweet vectors from word vectors using the Bi-function, capturing bi-gram information and word interactions with hyperbolic tangent functions.
* The Multilayer Perceptron (MLP) classifier is employed to classify tweet vectors obtained from the compositional vector model. The results indicate that MLP outperforms other classifiers in abusive content detection.

**Outcome:**

* The research compared word embedding models and classifiers for abusive content detection. Word2Vec with bi-function outperformed others. Logistic Regression and MLP had the highest weighted F1 scores.
* The model achieved 86% accuracy and 0.84 weighted F1-score, suggesting potential improvements with emotion, sentiment, domain-specific lexicons, and low-resource language analysis in the future.

**References**

[15] S. Ghosal, A. Jain and D. K. Tayal, "An approach to detect abusive content incorporating Word2Vec and Multilayer Perceptron," *2022 IEEE Bombay Section Signature Conference (IBSSC)*, Mumbai, India, 2022, pp. 1-5, doi: 10.1109/IBSSC56953.2022.10037274.

**Evaluation of Report**

**Evaluation summary with justification.**

The paper summarized on detecting suspicious activities and negative content in social media using deep learning techniques of Word2Vec and Multilayer Perceptron Neural Network. The in-depth report also contains the challenges in abusive text detection which is approached using compositional vector models and word embeddings with MLP classifiers for abusive content detection on Twitter.

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

Key Result: The model achieved 86% accuracy and a 0.84 weighted F1-score, validating the success of the abusive content detection process.

**The usefulness of the paper to the overall project.**   
  
The MLP classification technique can be further used to detect abusive content on social media which are considered suspicious, successfully targeting the main goal of this project.

**Other comments**

No comments

**Evaluation Approval  
  
Evaluation by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 09/17/2023**

**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 complete with all the major results of the paper. The paper focuses on detecting abusive content on social media, which concludes that the model proposed is 86% accurate.

**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 are completed sufficiently.

**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 written report is satisfactory.

**Approval.  
  
Approved by:**[Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 09/18/2023  
  
Is the quality of this written in-depth study report and Evaluation report satisfactory? If not, then why not? (limit: 40 words)**

Yes, the quality of the in-depth study report is satisfactory, as the approach is discussed clearly. Also, the evaluation report is satisfactory as the quality of content mentioned, denotes that the study report has been thoroughly understood by the evaluator.