**Individual In-depth Report**

**Member name:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)

**Evaluated by:** [Justin Young](mailto:jtyoun15@asu.edu)

**Date:** 09/29/23

**Tasks Assigned:**

* Review of Ecosystem of Spamming on Twitter: Analysis of Spam Reporters and Spam Reportees

**Summary:**

* This research paper focuses on the analysis of spam activity on Twitter and the corresponding reporters and reportees of such activities.
* The process of classification of these classes of users is done using the techniques of Decision Tree, K-Nearest Neighbors and Random Forest Classifier.
* The algorithm proposed in the paper takes a unique approach by focusing on the attributes of the reporter and reportee. This would help in accurately identifying spam users, while also preventing honest users from incorrectly getting classified as spam and getting their Twitter accounts suspended.
* The main steps in the algorithm are:
  + Data Collection – This step involves collection of data from Twitter using Python(...). From the data of each of the tweets collected, each tweet data is categorized into reporters and reportees.
  + Categorization and Feature Set Description – In this crucial step, different classes of reporters and reportees are identified.

The classes of reportees are: Active Spammers, Promoters, Genuine and Spammers.

The classes of reporters are: Automated Account, Genuine Users and Popular/Promotional/Celebrities.

The classification of reporters and reportees can be done only using metrics or attributes. These attributes are described as feature sets. These sets are created separately for classification of reporters and reportees using different features.

* + Application of data mining algorithms – The final step involves the application of Decision Tree, K-Nearest Neighbors and Random Forest Classifier algorithms for data mining. The implementation is done using R language. The accuracy of these algorithms is measured and verified with manual classification. (About the algorithms)
* From the results of the algorithm, it is found that the decision tree is the best classifier for reportees and reporters. (Quantitative Measures - Accuracy etc..)

**Outcome:**

The result obtained from the paper is that the Decision Tree algorithm is the best algorithm for reporter and reportee classification owing to its high accuracy value.

**References** *(with citation)*  
  
[33] P. Sinha, O. Maini, G. Malik and R. Kaushal, "Ecosystem of spamming on Twitter: Analysis of spam reporters and spam reportees," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp. 1705-1710, doi: 10.1109/ICACCI.2016.7732293.

**Evaluation of Report**

**Evaluation summary with justification.**

This research paper covers a new approach to analyzing spam activity on Twitter that focuses on the attributes of the reporters and reportees. A comparative analysis of models from the study also concluded that the Decision Tree model performed the best in classifying spam.

**The quality of the major result(s) with justification.**  
This paper provides an algorithm that can be implemented to detect and classify spam on social media, along with data regarding the performance of models we might choose for a system like this.

**The usefulness of the paper to the overall project.**   
The results from this study are relevant to our group’s topic, as spam is a significant issue when ensuring security on social media.

**Other comments**

**Evaluation Approval  
  
Evaluation by:** [Justin Young](mailto:jtyoun15@asu.edu) **Date:** 10/1/23

**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, all major results of the paper are covered.
* The in-depth study report is complete.

**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 have been 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 written report is satisfactory.

**Approval.  
  
Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 10/1/2023  
  
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.