**Individual In-depth Report**

**Member name:**[Krupaben Kothadia](mailto:kkothadi@asu.edu)

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

**Date:** 09/03/2023

**Tasks Assigned:**

* I have identified 2 pivotal questions to work upon as my research domain for this project.
* These questions address the project’s aim and it also touches base with addressing an issue related to information assurance and security.
* Finalizing the research domain as Data Preprocessing and NLP techniques used to produce a robust dataset for detecting suspicious activity on social media.
* This report summarizes the NLP techniques discussed in the paper “[An Integrated approach for Malicious Tweets detection using NLP](https://drive.google.com/drive/u/3/folders/1T7fs2iwRVNtzAw-9zUeapSM3Ha7uB5_H)” which produces a dataset that can accurately detect suspicious tweets on twitter, while making sure that the user privacy has been maintained.

**Summary:**

* This paper introduces innovative data preprocessing and Natural Language Processing (NLP) techniques tailored to analyze individual tweets rather than entire user accounts, preserving user privacy effectively.
* The authors have used Twitter’s APIs to collect the data from the twitter. This dataset was publicly available and it contained pre-classified tweets as spam or not spam, and for misclassified ones, the authors reevaluated tweets linked to the same URLs. They assessed tweet content and associated URLs to distinguish between these categories. Collected tweets were organized into one specific file format.
* Furthermore, the data has been splitted into tweets with and without URLs. Data without the URLs have been considered as not spam and data with URLs have been considered as spam. Then the tweets with the URLs were partitioned into the text and the URLs, these URLs were used as one of the features.
* Now, utilizing advanced NLP techniques, the author processed the texts by removing stop words and used these separated words to identify the semantic meanings between each other.. The language model has generated many features through this text and one of their features was used to calculate the divergence ratio between the semantics of text and the page found by exploring the associated URL. This divergence ratio is considered to be one another feature used to train the classifier. It’s mentioned in the paper that this feature is helpful in calculating the divergence point (difference) amongst the posts, which eventually detects if something is different from the actual post. Hence, the authors have developed a new language model using this “divergence point” feature. This feature is helpful in classifying the tweets as spam or not spam.

**Outcome:**

The paper introduces advanced data preprocessing and NLP techniques to analyze tweets while preserving privacy. Leveraging Twitter's APIs, publicly available pre-classified data was collected, labeled, and organized. NLP technique was used to parse text. The outcome is a language model with text-URL divergence classifying tweets into spam and non-spam.

**References** *(with citation)*

[10] S. Gharge and M. K. Chavan, “An integrated approach for malicious tweets detection using NLP,” *International Conference on Inventive Communication and Computational Technologies,* Mar. 2017, doi: 10.1109/icicct.2017.7975235.

**Evaluation of Report**

**Evaluation summary with justification.**

This paper outlines the methods by analyzing tweets on Twitter and preserving privacy at the same time. The above is done using classification algorithms like the SVM classification method. Spam accounts are analyzed for detecting spam tweets from Twitter. The NLP technique is used to parse texts and classify tweets into spam and non-spam tweets.

**The quality of the major result(s) with justification.**  
Results from this paper include a discrete study of spam detection in twitter with the use of classification algorithms and NLP techniques and securing privacy at the same time. This research paper was helpful for one of the major goals of the project.

**The usefulness of the paper to the overall project.**   
The classification of spam tweets and non spam tweets is useful for detecting suspicious activities prevailing in the social media and in twitter to be precise.

**Other comments**

No Comments

**Evaluation Approval  
  
Evaluation by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date:** 09/03/2023

**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 report complete all the major results of the project
* Detection of spam accounts and spam tweets in twitter can detect some of the suspicious activities in twitter
* The use of SVM and NLP techniques provides insights to detect threats in other social networking applications too.

**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 the report is satisfactory.

**Approval.  
  
Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date:** 09/03/2023 **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 written report and evaluation report is satisfactory and concise.