**Individual In-depth Report**

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

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

**Date: 10/02/2023**

**Tasks Assigned:**

* This report summarizes the paper “[Detection of User Cluster with Suspicious Activity in Online Social Networking Sites](https://drive.google.com/file/d/1-93xUJ5x4MJQEww1yoLOcA0i-wtr9g5z/view?usp=drive_link)”, which discusses the data preprocessing and NLP techniques used to process the Social Networking Sites (SNS) data. Suspicious message identification is done using the NLP/Keyword system and Latent semantic analysis (LSA) system is used to determine the similarity between the messages.

**Summary:**

* The authors' objective is to present a comprehensive system for scrutinizing messages within Social Networking Sites (SNS) to identify clusters of suspicious users engaged in questionable activities. Given the challenging nature of accessing SNS data, the authors established a private experimental network known as 'Manipal Net.' Their system revolves around collecting data from SNS through online monitoring, capturing vital message details like sender, receiver, content, and timestamps.
* To detect suspicious messages within this vast dataset, the authors employ Natural Language Processing (NLP) and keyword-based systems, with a focus on specific topics including 'Hate Messages,' 'Terrorist Activity,' 'Delhi Gang Rape,' 'Harm to Society,' 'Narendra Modi,' and 'Confidential Keyword-Based Suspicious Message Exchange from Organization.' Here's an overview of the distinct NLP systems devised by the authors:
  + Sentiment Score Identification System: This subsystem assesses message sentiment by assigning scores to words based on their weight in the message, determining whether it conveys a positive or negative sentiment.
  + Sentiment Count Identification System: Calculating the likelihood of positive and negative sentiment word occurrence in a message, this system relies on a set of 2230 positive and 3905 negative words for assessment.
  + Training Set Based Sentiment Identifying System: This pivotal component applies sentiment analysis to messages using topic-specific training sets. It selects the appropriate training set, based on results from the Sentiment Score Identification System, to classify sentiment as positive or negative.
  + Dictionary-Based Topic Identifying System: This system assesses messages by assigning matching scores using a topic dictionary built from topic-specific information. It leverages NLP techniques to flag messages as suspicious or normal, preparing them for further processing.
* Furthermore, the authors introduce Latent Semantic Analysis (LSA) to identify groups of individuals involved in abnormal activities via SNS. LSA, grounded in Singular Value Decomposition (SVD), uncovers latent relationships between seemingly unrelated words. Combining findings from the 'Suspicious Message Identification using NLP' and 'Latent Semantic Analysis' modules facilitates the identification of clusters of suspicious SNS users. This intricate system reflects the authors' commitment to devising an effective solution for detecting suspicious activities on social networking platforms.

**Outcome:**

The authors have developed an intricate system utilizing Natural Language Processing (NLP) and keyword-based methods to identify suspicious activities within Social Networking Sites (SNS). By focusing on specific topics and employing sentiment analysis, this system successfully detects clusters of users engaged in suspicious activities, enhancing SNS security.

**References** *(with citation)*

[1] A. Kumar and S. Singh, “Detection of User Cluster with Suspicious Activity in Online Social Networking Sites,” 2013 Second International Conference on Advanced Computing, Networking and Security, Dec. 2013, doi: 10.1109/adcons.2013.17.

**Evaluation of Report**

**Evaluation summary with justification.**

The research paper explores a sophisticated system using Natural Language Processing and keyword techniques to spot suspicious behavior on Social Networking Sites (SNS). Through topic analysis and sentiment assessment, this system effectively identifies groups of users involved in dubious activities, bolstering SNS security.

**The quality of the major result(s) with justification.**  
  
The major results and outcomes of both the papers have been discussed in detail in this report.

**The usefulness of the paper to the overall project.**   
  
The paper is highly relevant and contributes significantly to the overall project.

**Other comments**

No comments

**Evaluation Approval  
  
Evaluation by:**[Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 10/02/2023**

**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, 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 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 written report is satisfactory.

**Approval.  
  
Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 10/02/2023  
  
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.