**Individual In-depth Report**

**Member name:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Evaluated by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu)

**Date: 10/2/2023**

**Tasks Assigned:**

* Focus on my question to work upon as my area of research for this project.
* The question I chose was to do my research and find a system that maintains the balance between detecting suspicious activity in social media and safeguarding the user's privacy and maintaining data integrity.
* Review of [**Data Mining Approach for Anomaly Detection in Social Network Analysis**](https://drive.google.com/file/d/1BSRlKQV-C77yN9aP7ccNfmVOTVLWKtaU/view?usp=drive_link)

**Summary:**

* The main focus area of this paper is to discuss an approach for anomaly detection in Online Social Networks (OSNs) using data mining techniques.
* This analysis is carried out using Facebook dataset. It mentions profiling user social behavior by analyzing user updates, messages, clickstreams, photo uploads, posts, and comments.
* The discovery process consists of an iterative sequence of the following steps:
  + **Data cleaning** – This handles noisy, erroneous, missing or irrelevant data from the collection.
  + **Data integration** – In this stage multiple, heterogeneous are combined in a common source.
  + **Data selection** – The data relevant to the analysis task are retrieved from the dataset.
  + **Data transformation** – Where the data are transformed are consolidated into forms appropriate for mining by performing aggregations operations.
  + **Data mining** – It is a crucial step, where intelligent methods are applied to extract data patterns.
  + **Pattern evaluation** – In this step, interesting patterns representing knowledge are identified based on given measures.
  + **Knowledge representation** – It is a final phase, where visualization and knowledge representation techniques are used to present the mined knowledge to the user.
* DATA MINING APPROACHES TO ANOMALY DETECTION :
  + **Supervised Method** – treats the problem as a classification task with pre-labeled data, distinguishing between normal and anomalous observations.
  + **Semi-Supervised Method** – builds a model based on normal data to create a profile of normal activity. These methods work with both labeled and unlabeled data. They are useful when only a few instances of labeled normal data are available.
  + **Unsupervised Method** – trains an anomaly detection model using unlabeled data that includes both normal and abnormal instances. They are employed when labeled data with predefined labels like "anomalies" or "normal" are unavailable.
* However, cyber criminals or attackers have started to exploit these Social Networking sites by propagating malwares and carrying out scams. Some of these attacks are the Sybil attack, Identity clone attack and Socware attacks.

**Outcome:**

* The paper introduces a method for detecting anomalous users in social networks based on their distinct behavioral patterns using outlier detection along with a visual explanation.
* The analysis is conducted on a Facebook dataset to assign risk scores to users, with the notion that greater behavioral divergence implies higher risk. This evaluation aims to assess the method's effectiveness.

**References** *(with citation)*  
  
[8] M. S. Sudha, K. A. Priya, A. K. Lakshmi, A. Kruthika, D. L. Priya and K. Valarmathi, "Data Mining Approach for Anomaly Detection in Social Network Analysis," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, 2018, pp. 1862-1866, doi: 10.1109/ICICCT.2018.8472985.

**Evaluation of Report**

**Evaluation summary with justification.**

* This report discusses anomaly detection in Online Social Networks (OSNs) using data mining techniques, focusing on Facebook data. It covers various steps, from data cleaning to knowledge representation. Three data mining approaches are outlined: supervised, semi-supervised, and unsupervised methods. The report highlights the increasing threat of cyberattacks on social networking sites, such as Sybil attacks and malware propagation.

**The quality of the major result(s) with justification.**  
This report presents effective data mining methods for anomaly detection in Online Social Networks (OSNs) using Facebook data, addressing emerging cyber threats.

**The usefulness of the paper to the overall project.**   
This paper aids the project by discussing data mining methods for anomaly detection, crucial for identifying suspicious activities in social media.

**Other comments**

The generated summary is on the point and well justified as per the project definition.

**Evaluation Approval  
  
Evaluation by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 10/2/2023**

**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 written report of the in-depth study is complete with all the major results of the paper. The paper discusses an approach for anomaly detection in Online Social Networks (OSNs) using data mining techniques, utilizing a Facebook dataset. It covers the entire process, including data cleaning, integration, selection, transformation, data mining, pattern evaluation, and knowledge representation. The paper also outlines different data mining approaches for anomaly detection, such as supervised, semi-supervised, and unsupervised methods. Furthermore, it highlights the emerging threat of cyberattacks on social networking sites.

**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 adequately covered. The paper discusses anomaly detection in Online Social Networks, outlining data mining stages and methods, and acknowledges emerging cyber threats, such as Sybil and identity clone attacks.

**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: 10/2/2023  
  
Is the quality of this written in-depth study report and Evaluation report satisfactory? If not, then why not? (limit: 40 words)**

Yes, the report is evaluated by me and the report has been studied carefully with all the remarks added in the evaluation report. The quality of this written in-depth study report and evaluation report is satisfactory.