## *Disclaimer: The following document is intended to provide general guidance on the use of this template. Please refer to the template for the specific format and content requirements. You may also add additional information to the template as needed for your specific project.*

**Individual In-depth Report**

**Member name:** [Rahul Nayak](mailto:rrnayak@asu.edu)

**Evaluated by:** [Anuranjan Dubey](mailto:adubey37@asu.edu)

**Date: 09/30/2023**

**Tasks Assigned:**

* Work on individual in-depth report
* Preparing individual progress report
* Evaluating Avani’s individual progress report and in-depth report

**Summary:**

* Last week I worked on studying a paper related to detecting spam accounts on Sina Weibo social network.As this paper still had more content and study on Dataset collection and analysis, This week the focus will be on this topic.
* In the paper the authors manually selected 100 normal users (celebrities, companies, government accounts) and 50 spammer users for data collection.
* They developed data crawlers for both normal and spammer users, extracting information from 30,116 Weibo users.
* The authors also collected information from 500 recent messages for each user, including basic user information and message attributes.
* 11,488 spammers and 17,646 non-spammers were labeled. 80% of each group was used as training data, and the rest as testing data.
* They analyzed user-based features like the number of followers, created days, and posting frequency.It was Identified that spammers often have a large number of followers compared to followers and create new accounts frequently.
* They observed that spammers tend to post messages automatically and include URLs in their messages.
* The authors Introduced a supervised machine learning model, specifically a Support Vector Machine (SVM) with an RBF kernel, for detecting spammers.
* They achieved high accuracy in classifying both spammers and non-spammers.
* They also experimented with different ratios of spammers to non-spammers in the training dataset and found that a 1:2 ratio worked well.
* Finally they developed a prototype software for real-time spammer detection on Weibo and tested the software on a trending topic and achieved a high accuracy rate.

**Outcome:**Overall, the paper presents a comprehensive approach to detecting spammers on Weibo using machine learning(SVM and other classifiers) and provides insights into the key features that differentiate spammers from regular users**.**

**References** *(with citation)*

[1] Xianghan Zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu, Chunming Rong,

“Detecting spammers on social networks”,Neurocomputing,Volume 159,2015,

Pages 27-34,ISSN 0925-2312

**Evaluation of Report**

**Evaluation summary with justification.**

The report provides a comprehensive overview of the paper on detecting spam accounts on Sina Weibo. It covers key aspects of the paper, such as data collection, analysis, machine learning models, and real-time detection.

**The quality of the major result(s) with justification.**  
The authors collected a substantial dataset, employed a Support Vector Machine with an RBF kernel for detection, and achieved high accuracy in classifying spammers and non-spammers. Their real-time detection prototype also yielded a high accuracy rate. These achievements indicate the robustness of their approach.

**The usefulness of the paper to the overall project.**   
The machine learning techniques serve as valuable guidance for further implementation of anti-spam measures on the platform.

**Other comments**

**Evaluation Approval  
  
Evaluation by:** [Anuranjan Dubey](mailto:adubey37@asu.edu) **Date:** 9/30/2023

**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 effectively summarizes the major results of the paper on detecting spam accounts on Sina Weibo.
* 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: 9/30/2023  
  
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 in-depth report is satisfactory and meets the requirement.