**Individual Report**

**Member name:** [Anuranjan Dubey](mailto:adubey37@asu.edu)

**Evaluated by:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)

**Date:** 10/06/2023

**Tasks Assigned:**

* Literature review for Machine learning techniques applied in: “Machine learning approach for threat detection on social media posts containing Arabic text”.
* Identifying previously read papers where methods employed in these papers like feature selection, extraction with machine learning approaches were similar.
* Review and Evaluate Rahul’s Report.
* Visited the writing center appointment with Parvathy.
* Read not-so important papers.

**Summary:**

* Tasks accomplished this week: visited writing center, identified the similar methodologies, wrote an individual report and reviewed Rahul’s Indepth report and individual report.
* Writing Center Appointment with Parvathy Screenshot: [Anuranjan\_Dubey\_Writing\_Center\_Confirmation.png](https://drive.google.com/file/d/19e2a6tombKuSg_wdIrY5604jQ5tSWKFv/view?usp=drive_link)
* The research aims to develop a Convolutional Neural Network (CNN) model to classify Instagram content, including images and Arabic comments, for threat detection, filling a gap in current research on Arabic social media threat tracking.
* The Inception v3 model, a deep CNN pretrained on ImageNet, is fine-tuned using transfer learning.
* A CNN model is employed for comment classification.
* There are 1792 training samples, 384 validation samples, and 384 testing samples that are used for comment classification.
* 700 training, 150 validation, and 150 testing samples are used for image classification.
* Instagram data, comprising images and Arabic comments, is collected via the Instagram API and search engine. The dataset is manually labeled for threat categorization.
* Arabic comments are cleaned by removing non-Arabic comments, ensuring adequate word count, eliminating @mentions, punctuation, and tokenization. Images containing only text or self-images are removed.
* While the research focuses on Instagram, which is a different social media platform from most research papers seen yet, it's unclear if the method can be directly applied to other social media platforms, necessitating further investigation.
* Feature extraction techniques include Histogram Oriented Gradient (HOG) and Local Binary Pattern (LBP) for image analysis, with Haar-like features used for Viola-Jones face detection.
* The methodology involves three phases: data collection, data preprocessing, and classification, with a CNN model employed for both image and comment classification.

**Outcome:**

The model achieved a remarkable 96% accuracy for images and an outstanding 99% for comments. A hands-on experiment using a CNN model shows real promise in tracking threats on Instagram.

Future applications could be significant for social media monitoring.

**References** *(with citation)*

[1] AlAjlan, Shatha AbdulAziz and A. Khader, “Machine learning approach for threat detection on social media posts containing Arabic text,” Evolutionary Intelligence, vol. 14, no. 2, pp. 811–822, 2021, doi: https://doi.org/10.1007/s1206502000458w. Available: https://doi.org/10.1007/s1206502000458w

**Evaluation of Report  
  
Evaluation by: Sangeeth Santhosh  
Date: 07th October 2023**

**Is the weekly member report 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). [within 100 words]**

Yes, the weekly individual progress report is complete with all major tasks of the member this week covered in detail. The major tasks completed by the member include literature review of a research paper, visiting writing center and reading research papers common to every student. The research paper reviewed by the member focuses on development of a CNN model to track threats on Instagram.

**Is each section of the guidelines sufficiently completed? If not, point out what is missing. [Normally within 40 words].**

Yes, each section of the guideline has been sufficiently covered this week, with no major points missing in the report.

**Is the quality of this version of the written report satisfactory? If not, then why not? [Normally within 40 words]**

Yes, this version of the written report is satisfactory and covers all the major points that the member has done this week.

**Approved by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 10/09/2023**