**Individual Report**

**Member name: Sangeeth Santhosh**

**Evaluated by:** [Justin Young](mailto:jtyoun15@asu.edu)[Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Date:** 09/23/23

**Tasks Assigned:**

* Preparing Midterm Report
* Preparing Individual Midterm Review Report
* Preparing and Evaluating Individual Progress Report
* Preparing and Evaluating Individual In-depth Report
* Performing the literature review of research papers assigned.
* Preparation of an in-depth report of one research paper among the research papers found – ‘Spam Filtering of Bilingual Tweets Using Machine Learning’.

**Summary:**

* Reviewed the research paper – ‘Spam Filtering of Bilingual Tweets Using Machine Learning’.
* Analysis of different classification techniques for spam filtering of tweets, mainly in the Roman Urdu language, is done in this research paper.
* The five main steps proposed in the research paper for the spam detection process are:
  + Collection of tweets – The tweets collected are either in Roman Urdu or English language.
  + Pre-processing of tweets – Removing unimportant characters from tweets and discarding small size tweets.
  + Data Preparation for Classification – Labeling tweets as spam or legitimate is done in this step by domain experts. This step is almost 100% accurate as almost no legitimate users would be classified as spam, mainly because of the efficiency of the domain experts.
  + Classification – Done using commonly used machine learning algorithms – Naïve Bayes Multinomial, Liblinear, LibSVM, DMNBText and J48. The parameters for comparison of these algorithms are accuracy and ROC AUC.
* From the results, it is obtained that Naïve Bayes Multinomial is the best machine learning algorithm for classification, owing to its accuracy, which is highest among all the algorithms and a smaller number of false negatives generated.
* All other tasks have been successfully completed. These tasks include adding the relevant portions to the midterm report, preparation of individual midterm review reports, preparing the individual progress and individual in-depth reports, evaluation of one team member’s individual progress and individual in-depth reports.

**Outcome:**

Filtering of Bi-Lingual Tweets into spam and non-spam is achieved successfully and the best machine learning algorithm for it is obtained by comparing accuracy and ROC AUC parameters.

**References** *(with citation)*

[32] H. Afzal and K. Mehmood, "Spam filtering of bi-lingual tweets using machine learning," *2016 18th International Conference on Advanced Communication Technology (ICACT)*, PyeongChang, Korea (South), 2016, pp. 710-714, doi: 10.1109/ICACT.2016.7423530..

**Evaluation of Report  
  
Evaluation by:** [Justin Young](mailto:jtyoun15@asu.edu)

**Date: 9/24/2023**

**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, all major results are covered.

**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 are 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 version of the report is satisfactory.

**Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 9/24/2023**