**Individual In-depth Report**

**Member name:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)

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

**Date:** 09/23/23

**Tasks Assigned:**

* Review of Spam Filtering of Bilingual Tweets using Machine Learning

**Summary:**

* This paper focuses on filtering spam content on bilingual tweets in social media and analyzing the different classification techniques that are used for the same.
* The performance of different machine learning algorithms is used to analyze and evaluate the spam detection in Roman Urdu tweets in this paper.
* The steps proposed in this paper to detect spam tweets in the Roman Urdu language are:
  + Tweet collection – The tweets are collected from five major cities and the language of each of them is either Roman Urdu or English.
  + Pre-processing of Tweets – In this step, less informative aspects of the tweets are removed. Also, tweets of smaller size are discarded.
  + Data Preparation for Classification – Labeling tweets as spam or legitimate is done in this step. In addition to this, each word present in a tweet is converted into a vector, with the first number showcasing the numeric representation of a number and the second number showcasing its frequency. As this step is done by domain experts, the chances of incorrect classification of legitimate tweets as spam is considerably reduced. Thus, the integrity of an honest user would not be compromised in most cases.
  + Classification – Different machine learning algorithms are used for classification. In this paper, Naïve Bayes Multinomial, Liblinear, LibSVM, DMNBText and J48 are used, with two parameters – accuracy and ROC AUC used for performance measurement.
* The results of classification are analyzed. Highest accuracy is obtained by Naïve Bayes Multinomial with 95.42%. This algorithm also performs the best on less data as compared to other algorithms. Owing to this, it is concluded that Naïve Bayes Multinomial is the best algorithm for classification of tweets.
* Overall, it is concluded from the paper that Naïve Bayes Multinomial is the best technique while SVM is slow on bigger datasets and hence does not perform well. The DMNBText also performed well but it generated a greater number of false positives.

**Outcome:**

Successful classification of tweets, mainly in the Roman Urdu language was achieved and the best machine learning algorithm for classification was obtained by comparing the 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 summary with justification.**

The Naïve Bayes Multinomial model is proven to be the most accurate in filtering Bilingual spam on social media through classification (utilizing Twitter as the test platform), after undergoing a process consisting of the following steps: Tweet collection, tweet pre-processing, data preparation, classification

**The quality of the major result(s) with justification.**  
The research in this study includes a comparative analysis of various ML algorithms on this particular issue to determine the most effective model.

**The usefulness of the paper to the overall project.**   
The results of this study provides a model that can be implemented to detect and prevent a significant security issue on social media

**Other comments**

**Evaluation Approval  
  
Evaluation by:** [Justin Young](mailto:jtyoun15@asu.edu) **Date: 9/24/2023**

**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 in-depth report is complete with all the major results.
* The best machine learning algorithm for classification was obtained by comparing the accuracy and ROC AUC parameters.

**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. The report summarizes filtering spam content on bilingual tweets in social media and analyzing the different classification techniques that are used for the same.

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

Yes, the quality of the written report is satisfactory.

**Approval.  
  
Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 9/24/2023  
  
Is the quality of this written in-depth study report and Evaluation report satisfactory? If not, then why not? (limit: 40 words)**

Yes, the in-depth report and evaluation report summarized the research paper with all the guidelines followed. Both reports are satisfactory and concise.