**Individual Report**

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

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

**Date: 09/17/2023**

**Tasks Assigned:**

* Preparing an individual in-depth report.
* Preparing an individual progress report.
* Approving and evaluating the Gantt Chart.
* Evaluating the individual progress report report prepared by [Krupaben Kothadia](mailto:kkothadi@asu.edu)
* Preparing and approving the Weekly Report.
* Approving 4 team members in-depth reports if there are any.
* Approving 4 team members' research domain and their research papers.
* Formatting the Gantt Chart.
* Preparing [References\_List](https://docs.google.com/document/d/15T-F8te1yxn_htvAxAdAmRKHNw10Ci9SCHduxrQxA4c/edit) doc
* Finding reference papers related to the project
* Organizing Google Drive.
* Taking meeting notes.

**Summary:**

* All the tasks have been completed successfully. These include preparing the in-depth report, individual progress report, preparing and approving the weekly report, preparing the reference list, formatting and evaluating the Gantt chart, approving 4 team members’ in-depth reports and progress reports along with finding some references related to the project.
* An In-Depth study of [**Sentiments Analysis Of Twitter Data Using Data Mining**](https://drive.google.com/file/d/1SGNHuQyCJ--drDbYgOgs12MNIApcoHFT/view?usp=drive_link) has been conducted.
* The main focus area of this paper is to find an approach to analyze the sentiments of users using data mining classifiers.
* Existing systems propose a hybrid approach of extracting opinion using direct and indirect features of Twitter data based on Support Vector Machines (SVM), Naive Bayes, Maximum Entropy and Artificial Neural Networks based supervised classifiers.
* This paper presents a mechanism to predict the overall sentiments inclination of Indian people towards political situations and issues. After collecting raw tweets various preprocessing methods get applied to clean the data.
* The sentiment analysis is performed is using the following steps:
  + **Data Collection:** Training and testing tweets collected from twitter by using twitter searched API v 1.1 for various political leaders and parties in india.
  + **Preprocessing:** All userID,twitterId, userinfo from the tweets is removed. Duplicate tweets and retweets also get removed from the training dataset. After applying all cleaning methods text that is only with the tweeted text remains.
  + **Training DataSet and Testing Dataset:** For classifying a dataset, three classes, Positive, Negative, Neutral are used. To classify tweets into these categories, “SentiWordNet 3.0.0.” dictionary is used.

**Outcome:**

2,102,52 tweets were collected about various political leaders and parties. k-nearest neighbor gives the highest accuracy of all classifiers. Prediction accuracy of k-nearest neighbor 99.6456%. While the previous 2 references focussed on the security concerns of the data mining techniques, this paper focuses on the efficiency in the data mining process of extracting Tweets from Twitter.

**References** *(with citation)*  
  
[9] Anurag P. Jain, Vijay D. Katkar, “Sentiments analysis of Twitter data using data mining,” in *“International Conference on Information Processing (ICIP)”*, Pune, India, 2015, pp.807-810

**Evaluation of Report  
  
Evaluation by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 09/18/2023**

**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]**

This report explores sentiment analysis of Indian political tweets using data mining classifiers. It achieves high predictive accuracy with the k-nearest neighbor classifier. Given Twitter's extensive user base and tweet volume, this research is crucial. It introduces a hybrid approach with supervised classifiers, offering insights into Indian users' political sentiments. Hence, the weekly member report is complete with all the major results discussed in the paper referred. Along with successfully completing all the assigned tasks.

**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 version of the written report is satisfactory.

**Approved by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date:** 09/18/2023