

ABSTRACT

One of the most used and famous social media by the public in general is Twitter and this social media have big amounts of users all around the world and they use it as communication as well as information sharing service on internet. The aim of this research is classifying the data that have been gathered before from Twitter and aside from to classify those datas, this research also has a purpose to compare 2 of most used data mining algorithm in classification method. Both of the algorithms are Naive Bayes Algorithm and K-Nearest Algorithm in data mining algorithms that are used in classification of data. There are some distant differences of classification result in both of the algorithm in performance and accuracy, which is the main aim of this research. To prove and compare the performance and accuracy of the result of classification, the research need to be done systematically. The steps from gathering data itself, cleaning the data to become a proper data use in process, processing the data using data mining algorithm, visualize the result of the process, and lastly presenting the result and compare them. With the data mining tools R Studio as the tool to gather data, creating a dataset to be processed in this research, processing the dataset with classification algorithm, and analyze performance with visualizing the result using confusion matrix method. In the final step, the confusion matrix value contents will be used as the values needed to calculate F1 score and compare the overall result from both classification algorithm that which algorithm have the bigger score and suitable in processing dataset in this research. From this research, it is resulting that Naive Bayes algorithm is better in classifying dataset in this research rather than K-Nearest Neighbor algorithm in both of accuracy and performance terms. Eventhough the accuracy of both algorithms have slight gap differences, the performance is has big difference in gap between Naive Bayes and K-Nearest Neighbor. With the research results, it is also helping the research on developing the data mining classification algorithm in the next research for the future reference.

Keywords—Covid-19, Classification, Algorithm, Naïve Bayes, K-Nearest Neighbor, Confusion Matrix, F1 Score.