

ABSTRACT

Every year, the number of vehicles in Jakarta is increasing. However, the increase in the number of motor vehicles in Jakarta is not proportional to the addition of road segments. This condition causes the disruption of the traffic smoothness and cause the congestion points. To anticipate being caught in traffic, traffic users search and exchange information about congestion on social media. One of the social media that is often used to spread information is Twitter.

In this research, a system that implements one of data mining technique is used to classify Jakarta's traffic condition. Using the *k-Nearest Neighbor* classification method, tweets from multiple accounts with a high degree of validity will be used as a set of information that becomes the traffic condition reference for the traffic condition's classification process. Performance testing for five times with different amount of training data yields an average accuracy above 70%. In addition to the results of accuracy, this test releases the optimal k value used in this system which is eight.

Keyword : Twitter, congestion, *data mining*, *k-Nearest Neighbor*