ABSTRACT

SENTIMENT ANALYSIS OF TWITTER ON COOKING OIL EXPENSIVE USING NAIVE BAYES METHOD AND SUPPORT VECTOR MACHINE

Oleh

Moh.Aminullah Al Fachri

18102166

The nine basic commodities or what we know as groceries are necessities that support everyday human life. The need for basic commodities never decreases so that the market must be able to meet consumer demand. In addition to sufficient supply, the market must also be able to provide prices that buyers can afford. At the end of 2021, the public was shocked by the drastically reduced supply of cooking oil and the sudden increase in price. The scarcity and the high price make people talk a lot about it, both directly and through their respective social media, one of which is Twitter. With a total of 19.9 million Twitter users, the Indonesian people have responded to the issue of the high price of cooking oil, both with positive and negative responses. The large number of responses expressed by the community made a comparison between the two responses difficult to observe. This study aims to determine the comparison of positive responses and negative responses. Machine learning with various methods including Naïve Bayes and Support Vector Machines can overcome this problem. The research conducted will examine how the comparison between positive responses and negative responses and which method has a higher level of accuracy. The data used in this study were taken from netizens' tweets on Twitter, with a total of 9,194 lines of data ready to be processed. The data used in the analysis phase is in the form of Indonesian language tweets collected by crawling. The tweets taken are tweets containing the search word "expensive cooking oil". To produce a better level of accuracy, the preprocessing process is carried out. The preprocessing stage will change the raw text data into data that is mature enough to be analyzed. The analysis phase with the Naïve Bayes method and the Support Vector Machine has different results. the Naïve Bayes method with 1839 test data obtained an accuracy of 74.06% by predicting two positive tweets and 1837 negative tweets. The SVM method was tested on the four kernels, namely linear, polynomial, RBF and sigmoid kernels. The four kernels produce better overall accuracy than the Naïve Bayes method. The kernel with the highest accuracy value is the sigmoid kernel with an accuracy value of 81.8% with predictions of 266 positive tweets and 1573 negative tweets.

Keyword : *Machine Learning, Naïve Bayes, Text Processing,* **SVM**, *Twitter*