1. INTRODUCTION

COVID-19 is an infectious disease caused by a newly discovered type of coronavirus. The epicentre of the spread of this virus began in Wuhan, China, in December 2019 [1]. President Joko Widodo announced that there were 2 Indonesian citizens who were confirmed positive for COVID-19 on March 2, 2020. As of February 16, 2022, the Indonesian government reported 4,966,046 confirmed cases of COVID-19 from 510 districts in 34 provinces [2]. The spread of COVID-19 is very fast. In addition to implementing health protocols, other solutions are needed to overcome the high level of spread, namely by making vaccines. The vaccine will stimulate antibodies, thus, the body will recognize the virus and reduce the risk of exposure [3]. Over time, several variants of the coronavirus that cause the disease COVID-19 emerged. On November 26, 2021, based on recommendations from the Technical Advisory Group on Virus Evolution, WHO established a new variant called Omicron [4]. Due to the outbreak of positive cases of the Omicron variant corona, the provision of booster vaccines has intensified. The booster vaccine is the third dose of vaccine given in an effort to break the chain of transmission of COVID-19 with the aim of increasing the body's immunity against the coronavirus [5]. A booster vaccination is considered controversial by the public, so many people give their opinions.

Nowadays, developments in all aspects of life are becoming completely digital. All information is very easy and fast to obtain and disseminate using digital technology. Many people express their opinions through social media. One of the social media that is quite popular for expressing opinions is Twitter. Twitter's active users will reach 206 million worldwide in 2021. While in Indonesia, there are 17.5 million users on the platform [6].

Twitter is a social networking service that allows users to send and read text-based messages, commonly called 'tweets'. Twitter was founded in 2006 by Jack Dorsey. Social media Twitter is widely used by several users to express their opinions, one of them being related to booster vaccines. In this study, data that has been tweeted by Twitter users related to the COVID-19 Booster Vaccine was collected using the Twitter API (Application Programming Interface). Twitter API is useful for easy access to information on the Twitter web.

Referring to the booster vaccine which is considered controversial, in this study, a system will be built to find out whether public opinion on the COVID-19 booster vaccine contains positive or negative sentiments. The result of the sentiment distribution will later be used to determine the performance of the Ensemble Bagging methodData on the distribution of positive and negative sentiments that are too far apart can affect the quality of the data so that it can also affect the performance of the method. In 2003 Nasukawa.T.& Yi conducted a study on sentiment analysis. According to him, sentiment analysis in Indonesian is a technique or method used to identify how sentiment is expressed using text and how that sentiment can be categorized as positive sentiment or negative sentiment. The results of the prototype system achieve high precision (75-95% depending on the data) in finding sentiments on web pages and news articles [7].

Research [8] shows that tweets data classification uses the Naïve Bayes classifier algorithm and support vector machine with the highest accuracy results obtained when using the support vector machine method with an accuracy value of 90.47%. In research [9], sentiment analysis was carried out using the Naive Bayes classifier with an accuracy of 93%. Similar research [10] used the Support Vector Machine and K-Nearest Neighbor methods, although the results were not satisfactory, the use of the SVM method had a better accuracy of 75%. Research [11] conducted research on vaccine actions using the keywords "Sinovac Vaccines" and "Red and white vaccines" using the Naïve Bayes method and Support Vector Machine, the positive sentiment results became the sentiment that dominates the two keywods, with the positive precentage results using the Naïve Bayes method for the keyword "vaksinsinovac" is 66% and negative is 34%, while for the keyword "vaksinmerahputih" the percentage of positive sentiment is 89% and 11% for negative sentiment. This study also shows that the average accuracy using the Naïve Bayes method is higher at 85.59%, while the Support Vector Machine method is 84.41%. There is also research [12] related to booster vaccines from the perspective of Indonesian citizens on Twitter using the Naive Bayes method with the keyword "booster vaccine" the results of positive sentiment is 23%, neutral sentiment is 15%, and negative sentiment is 76%, with the average results average accuracy of 89%.

In this study, an analysis of public sentiment towards the COVID-19 Booster Vaccine will be carried out using the ensemble bagging method. Ensemble Bagging is used because it is a combination of a set of machine learning algorithms. The final decision is based on voting from the classification results of the basic models. The basic model used is Naive Bayes, K-Nearest Neighbor, and Decision Tree. The results achieved in this study are related to the results of sentiment analysis and the performance of the Ensemble Bagging method.