I. INTRODUCTION

According to the Aqueduct Global Flood Analyzer analysis, Indonesia is a country with the 6th largest population affected by floods in the world, which is around 640,000 people every year [1]. Every community group, whether affected by the flood disaster or not, obtaining information, not a few have used social media to find various information about floods, one of which is Twitter.

Flood disasters that always occur occasionally give rise to opinions as if there is no effort from various parties in overcoming flood disasters. This opinion is then analyzed to determine its polarity with sentiment analysis. Sentiment analysis is part of text mining that can classify the polarity of the text that is carried out to see how the polarity of a tweet is whether the opinion given is positive or negative. The method chosen for this sentiment analysis research is the Naïve Bayes Classifier method with N-gram features because it is considered a potentially good method for classifying data than other classification methods in terms of accuracy and computation.

Text classification is the supporting technology of some information processing tasks, including controlled vocabulary index, the filtering on the content (spam, pornographers, etc.), information security, and others. Instead of manually classifying the documents, many machine learning algorithms have been trained to classify the documents automatically based on annotated training documents. The Naive Bayes (NB) classifier is often used as a basic text classification. [2]. Naïve Bayes will create probability from the fitted data, assuming all the variables are unique. This classifier is a good model for analyzing unique data like this tweet data. N-Gram tokenization could support the Naïve Bayes model by creating a weight of the data text. This weight contains numbers of the data text indexes, and each data will have a different weight combination. Research related to the journal entitled Comparison of Final Project Classification of Informatics Engineering Students Using the Naïve Bayes Classifier And K-Nearest Neighbour Method by Yusra et al. Get the results that the Naïve Bayes Classifier method produces a better accuracy value, which is 87% compared to the test on the K-Nearest Neighbour method which produces an accuracy value of 84% [3]. Based on this explanation, research on sentiment analysis on Twitter to see the response of Twitter users to the flood disaster uses the Naïve Bayes Classifier algorithm to classify positive and negative tweets by applying the use of the N-gram feature. Data should be prepared before it goes to fit the model. The result of Bayesian inference depends strongly on prior probabilities, which must be available to apply the method directly[4]. Related scientific work also discusses the use of N-grams in sentiment analysis of the Jakarta Regional Head Election using the Naïve Bayes algorithm by Wahyu Candra Indhiarta, the results of the average value of the greatest accuracy are found in the use of bigram, which is 0.823, this shows that by using bigram accuracy the accuracy of the system is better than using a unigram or trigram. The highest precision value is also found in the use of bigram with 0.76[5]. Research by [6] conducts a sentiment analysis of the presidential candidates of the Republic of Indonesia in 2019. In this research, the class classification or the level of public sentiment towards the presidential candidates of the Republic of Indonesia uses Naïve Bayes. This research also compares several methods, namely Naive Bayes, SVM, and KNN where Naive Bayes has the highest accuracy reaching 75.58% compared to other methods. Research by [7] analyzes sentiment on economic stabilization and overcoming the spread of the virus during the COVID-19 pandemic through Twitter. The research will analyze the Indonesian people's sentiments towards the New Normal policy. This study uses the naive Bayes method using trigrams to get an accuracy value of 84%, precision 84%, and recall 86%. Research by [8]analyzes social media sentiment on the Covid-19 vaccine. The method used to classify in this research is Nave Bayes. The result of this research is to classify positive and negative tweets in a period of one week. In this research, we do not do a model search first and make predictions. In this study, the sentiment results obtained were relatively balanced between positive and negative sentiments, namely 39% positive and 56 negative sentiments. Naïve Bayes has fairly good effectiveness and accuracy from previous research even though the data used differs. Every study conducted also uses the N-gram function to increase the amount of accuracy. From previous research, researchers want to try to analyze flood tweets using Naïve Bayes in the hope of separating positive and negative tweets so that they can classify floods that occur, including natural disasters or not.