

INTRODUCTION

Based on Katadata [1], Indonesia is one of the countries with the largest population of internet users in the world, according to reports, there are 204.7 million internet users in the country as of January 2022. This number slightly increased 1.03% from the previous year, In early 2022, Indonesia's internet penetration rate reached 73,7 percent of the total population. According to the data, nearly all internet users use social media. [2]. According to the National Indonesian Police Criminal Investigation Agency, there were 143 cyber crimes in the form of hate speech in Indonesia in 2015. In 2016, this figure increased to 199. This data, however, only includes hate speech that has been criminalized and reported to the authorities. Of course, there are numerous other hate speeches on social media. [3]. Twitter is a popular social media platform in Indonesia. Twitter and other social media and web microblogging services enable users to read and analyze their tweets in real time. This analysis can aid in the early detection of hate speech, allowing it to be contained before it spreads widely.

Research on the classification of hate speech on Twitter using the Naïve Bayes method has been widely carried out and has a high accuracy of 80-85%.[4][5], this method can be used using many features, namely the unigram feature and feature selection, these features managed to get better results than other Naïve Bayes features, such as the bi gram feature

The k-Nearest Neighbor (kNN) method is also widely used in the analysis of hate speech on Twitter, for example personality classification based on text. [6] and identification of hate speech against political figures [7] get a fairly high accuracy in the range of 70-85%.

The Support Vector Machine (SVM) method is the most widely used method in classifying hate speech on Twitter because of its high effectiveness and accuracy, SVM is proven to be able to produce an independent system that can classify tweets containing hate speech. [8][9] including hate speech in English and Spanish [10], SVM accuracy ranges from 70% to as high as 95%.

Multilayer Perceptron (MLP) method is a feedforward artificial neural network with one or more hidden layers. In MLP, there are two important parameters, namely the function activation and optimization functions [11]. Studies shows that MLP had been used for hate speech detection [12] and cancer breast classification with 97,7% accuracy using logistic activation and Adam optimization function [13].

In previous studies, the classification of hate speech was only done by one separate method. Method innovation is needed to get a higher accuracy of hate speech classification. This is the rationale in this study to implement the Hybrid Classifier method in Hate Speech Hashtag Classification, where the MLP classification method will be combined with the kNN and NB methods, with these three methods combined using Voting Classifier, it will produces higher accuracy than using just a single method to classify Hate Speech on Twitter.