

I. INTRODUCTION

In this digital era, information and communication technology is developing rapidly, including social media. Social media is one of the means that a person uses to interact with each other by giving, sharing, and exchanging information or ideas in a virtual network [1]. Based on a survey conducted by Hootsuite, in 2020, internet usage in Indonesia increased by 17% compared to the previous year, reaching 175.4 million people of which 160.0 million were active users on social media such as YouTube at 88%, WhatsApp by 84%, Facebook at 82%, Instagram at 79%, and Twitter at 56% [2].

By the existence of social media, everyone has the right to express their opinions freely with minimum restrictions. It could lead to social media misused by irresponsible peoples, either a person or a group of people, to spread hatred, racist comments, radicalism or extreme ideology, pornography, defamation, and so on, even though there is now a legal that limits social media usage. This phenomenon leads some of the researchers to distinguish some different types of toxicity in a comment to avoid undesirable things.

There were several studies on the classification of toxic comments that have conducted before. Previous research classified tweet data using the Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction and using the Support Vector Machine (SVM) method and compared the model with the Convolutional Neural Network model [3]. The best $F1 - Score$ obtained using the TF-IDF and SVM models equal to 74.88% due to SVM can classify data with two labels with only a few appearances. The preprocessing stages that yield the best result are without stopwords removal, stemming, and translation. Furthermore, detecting hateful comments on YouTube and Facebook social media using TF-IDF as feature extraction and the linear-SVM method obtained 79% of $F1 - Score$ [4]. Classification of text in Arabic using Chi Square, Information Gain, and Mutual Information compared feature selection using two different methods, namely SVM and Decision Tree. The best $F1 - Score$ value is obtain using Chi Square and SVM method, which is 79.59%. Using Chi Square, the calculation process in classification becomes faster [5].

In this study, we used the SVM method. Since, this method is proven to deliver good performance in the previous research such as in [3], [4], and [5]. Besides, we also use three different kernels for SVM, including Linear, RBF, and Sigmoid to find out which kernel obtain the best performance. Furthermore, we use TF-IDF as feature extraction and Chi Square as feature selection. The test scenario in this study is with and without using Chi Square for each SVM kernel referred to [5], while the preprocessing stage is to compare the use of stemming and stopwords and without using stemming and stopwords referred to [3]. The evaluation is using the $F1 - Score$.

The rest of the paper is organized as follows. Section II is the literature review. Followed by section III describes the methodology we used in this research. Furthermore, Section IV presents experimental results on the dataset using several scenarios. Lastly, Section V describes our conclusions and future works based on our experiment results.