1. INTRODUCTION

Social media is used by almost everyone today. Whether in the field of education, work, or entertainment, almost everyone from various parts of the world chooses social media as the medium they use [1]. With the development of time, the use of social media has become increasingly prevalent in everyday life. One of the popularly used social media is X, formerly known as Twitter. Based on data from Statista, X users in Indonesia reached 24 million users and is the country with the most users in fourth place [2]. On X social media, users can make posts commonly called posts about their thoughts or comments in the form of text, images, videos, GIFs, and others [3]. X has many benefits but some people misuse X to attack, intimidate, and other things related to cyberbullying. Cyberbullying refers to bullying, frightening or mistreating someone indirectly or through digital platforms [4]. Cyberbullying actions can be in the form of sending texts containing abusive messages, spreading personal information, bullying on online platforms or sending threatening messages. This problem must be addressed immediately, one form of problem solving that can be done is to create a cyberbullying detection system.

Cyberbullying detection research has been developed based on previous studies related to similar methods and research objects. Previous research on cyberbullying detection has used hybrid deep learning methods as has been done by Nur Wakhidah Fitri Amalia who uses the CNN-GRU method, as well as TF-IDF feature extraction and GloVe feature expansion in its detection [5]. In this study, a comparison of several methods such as CNN, GRU, CNN-GRU combination, and GRU-CNN combination was carried out. The four methods produce accuracy values that are not much different, but the highest accuracy is obtained by the GRU method with an accuracy value of 80.58%.

In addition, research related to cyberbullying detection was also conducted by Yudi Setiawan and his colleagues using SVM and KNN machine learning algorithms [6]. The main focus of this research is to utilize the combination of n-grams on TF-IDF to improve accuracy. The results showed that the application of a combination of machine learning algorithms with TF-IDF succeeded in increasing accuracy to 95.5%, which showed its effectiveness in detecting cyberbullying.

While cyberbullying detection focuses more on identifying unfavorable actions, sentiment analysis and other fields use similar approaches in improving model performance and relevance of extracted features. One of the studies related to sentiment analysis on movie reviews used Word2Vec combined with LDA. The main focus in this research is to use the skip-gram model in Word2Vec to improve the feature dictionary previously created by LDA analysis [7]. The results obtained show that the application of Word2Vec successfully increases the relevance of the extracted words, and can produce better and more targeted sentiment analysis.

In addition to feature extraction techniques, attention mechanisms have also proven effective in improving classification accuracy in various fields, including hate content detection and stock price prediction. Research on the recognition of hateful content in Arabic text was conducted by Abeer Aljohani and his colleagues using Convolutional Neural Network (CNN) and attention mechanism [8]. The main focus in this research is to utilize CNN in extracting features in the text and applying attention mechanisms to increase the accuracy value. The accuracy result obtained was 97.83%, indicating that the application of the combined model proved effective in increasing the accuracy of the model.

Another research using attention mechanism was conducted by Qingyang Liu and his colleagues to predict stock prices [9]. Attention mechanism in this study was combined with the LSTM model to create a more optimal model. The results obtained show that the ATT-LSTM model is able to achieve lower Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error values compared to the

LSTM model without attention, which indicates that the combination of the model is effective in improving model accuracy.

The combination of attention-mechanism with GRU and ResNet was also done by Gaurav and Pratistha Mathur for automatic image captioning [10]. In that study, attention-mechanism was used to create a proposed model that achieved a higher BLEU score compared to other models that used LSTM as a decoder. This finding shows that the combination of these models is effective in improving accuracy in generating image descriptions.

Although the application of techniques such as Word2Vec and attention mechanism is proven to be effective in improving model accuracy in other fields, the application of their combination in cyberbullying detection remains unexplored. While the CNN-GRU hybrid model has been applied in this field, its performance has not been able to surpass other approaches, which suggests there is still potential for improvement.

The main contribution of this study is to enhance the CNN-GRU hybrid model by applying attention mechanism and Word2Vec for feature expansion, which has been proven to produce good results in previous studies in other fields. This approach aims to improve the sensitivity and overall performance of the model in detecting patterns related to cyberbullying behavior, while also contributing to the development of more reliable detection systems for online platforms and paving the way for integrating multimodal data to further improve performance in the future.