1. INTRODUCTION

The use of social media is closely related to people's lives today. Social media allows each individual to communicate and interact with others through cyberspace, one of which is Twitter [1]. Based on Statista [2], the number of Twitter users in Indonesia is the fifth largest in the world with a total of 18.45 million users. There are various topics of discussion found on Twitter social media, one of which is movie reviews. Based on Statista [3], the theater or film entertainment market continued to grow from 2015 to generate 80.8 billion US dollars in 2020. This indicates that the film industry is growing from year to year. But with the increasing output, the quality of a movie is also evaluated by most netizens such as making reviews on social media. Twitter users can express their opinions in the form of tweets, such as giving a review of a movie. The reviews that Twitter users express in their tweets about a movie can be negative or positive [4]. However, due to the various opinions available, the assessment of reviews about a movie can be different [5]. The various opinions contain sentiment information that is in lines with this research such as plot, acting, and director aspects [6]. These opinions can be used as evaluations by filmmakers to increase audience satisfaction [7]. Therefore, research is needed to analyze the opinion whether it is a positive, negative, or neutral opinion.

Sentiment analysis is a means of determining whether an opinion is negative or positive or neutral from someone who can be realized in the written or oral form [8][9]. Sentiment analysis is comprised of three levels: sentence level, document level, and aspect level [10],[11]. Sentiment analysis at the aspect level is utilized to uncover sentiments expressed in reviews of a product. Considering multiple aspects of a product is crucial as a product typically has multiple facets that require evaluation. This cannot be done using the document level or sentence level because both levels cannot determine opinions specifically as in the aspect level [12],[13]. Since movie reviews have various aspects such as storyline, cast, visual effects, and so on, the implementation of aspect level can help in conducting this research.

This research was conducted based on previous research that has been done. In research [14], Rimdani conducted an aspect-based sentiment analysis on Twitter with applied the Gradient Boosting Decision Tree classification technique to signal and service aspects, as well as SMOTE and Random Undersampling. The results of this study indicate that SMOTE can produce better performance values than Random Undersampling. As a result, the final performance value obtained is 96.035% for the signal aspect, an increase of 25.260%, and 90.256% for the service aspect, an increase of 27.918%. In research [15], M. S. Mubarok et al. conducted experiments using the Naïve Bayes method in analyzing aspect-based sentiment and the dataset used is restaurant reviews. The aspects used in this study are food, service, price ambiance, and miscellaneous. With this method, the performance value of the f1-score is 78.12%. However, this research has not applied deep learning to the sentiment analysis process. Another study [16], K. Kim et al. used several deep-learning methods namely CNN, LSTM, and CNN-LSTM. This research compares the performance value of each deep learning method used. The dataset used in this research is a movie review dataset. It is concluded that the CNN-LSTM method is better than the others with an accuracy value of 89.20%. However, the research is not based on the product review aspect. In research [17], J. S. Lee et al. uses the GRU method in analyzing sentiment. The dataset used is a dataset taken from product reviews on Chinese e-commerce. This research aims to compare the GRU method to the LSTM method in potential performance improvement. The accuracy result obtained from this method on the Facebook dataset is 87%. However, this research has not implemented aspect-level sentiment analysis.

In research [18], R. Ahuja et al. conducted sentiment analysis by comparing two feature extraction methods namely TF-IDF and N-Gram on the SS-Tweet dataset. The study used six classification models such as the Decision Tree, SVM, KNN, Logistic Regression, Random Forest, and Naïve Bayes. The results of the study found that the use of TF-IDF has a better performance than N-Gram. In research [19], Aditya and Erwin applied Glove feature expansion on an aspect-based sentiment analysis on Twitter using the Random Forest classification. The accuracy value obtained for the service aspect is 80.12%, while the accuracy value for the signal aspect is 80.37%. In addition, GloVe feature expansion is able to increase signal accuracy by 13.15 percent and service accuracy by 5.37%. M. A. Raihan and E. B. Setiawan in their research [20], conducted aspect-based sentiment analysis on Twitter social media using FastText and SVM. The aspects used in the study are signal and service. The research also uses SMOTE data sampling techniques to overcome unbalanced data. The results of this study obtained the highest performance value after the application of SMOTE with an F1-Score value for the signal aspect of 95.93% and the service aspect with an F1-Score value of 94.53%.

Based on previous research, most of these studies have not conducted aspect-level sentiment analysis using GRU with feature extraction, GloVe, and SMOTE methods. The use of feature extraction, feature expansion and SMOTE can improve the performance value of the model built to be better. Therefore, In this study, an aspect-level sentiment analysis was carried out using the GRU classification.

The purpose of this research is to apply the GRU classification method to aspect-level sentiment analysis on movie review data and calculate the accuracy value generated by the model. The results of this study provide the performance value of the GRU classification model built. In this study, experiments were also conducted by

adding feature extraction with TF-IDF, feature expansion using Global Vectors (GloVe) trained based on corpus datasets, and adding SMOTE to improve model performance. The problem limitation of this research is that the data used is movie review data taken from Indonesian Twitter. The aspects used in this research are plot, acting, and director.