1. INTRODUCTION

The internet is commonly used by people to access social media. Social media is a platform that builds networks between users by sharing information such as text, videos, photos, audio, or sentiment information [1][2]. One of the benefits of social media can be used in improving the quality of products or movies through reviews [2]. Social media is used as a means to express their perspective on various topics discussed such as on the Twitter application. The Twitter application allows users to show their interest by sharing opinions about the knowledge they have, such as giving reviews of movies [3]. Movie reviews are someone's opinion in giving an assessment of the movie that has been watched. With movie reviews, filmmakers can use these reviews as evaluation material in making better movies [4]. However, the large number of reviews resulted in discrepancies in assessments of the movie [5]. In addition, a movie review can include several aspects of a movie such as plot, acting, and director aspects [5][6]. This creates difficulties in determining whether a review tends to be a positive or negative opinion. Therefore, a study is needed to help classify opinions on this problem. The classification results in this study can be used as analytical material in making films so as to increase audience satisfaction.

Sentiment analysis is a technique for classifying and evaluating individual opinions based on the emotions contained in a person's subjective information [7][8]. Sentiment analysis detects opinion polarity by classifying positive, negative, or neutral opinions depending on the words used. Sentiment analysis can be applied at several levels, one of which is the aspect level by identifying sentiment on certain aspects [7]. In this research, it focuses on aspect-based sentiment analysis by identifying sentiments on aspects of plot, acting, and director in movies. Sentiment analysis can be done using deep learning techniques, which are part of machine learning. One of the algorithms used in deep learning is Convolutional Neural Network (CNN) which has independent computational analysis capabilities [9].

Some previous research has been done on sentiment analysis, especially using CNN classification and word embedding methods. Research [6] used Naïve Bayes on a movie review dataset to analyze sentiment based on aspects. The aspects used in this research are screenplay, music, acting, plot, movie, and direction. The Naïve Bayes method produces the highest accuracy value of 79.372%. Another study in 2020 [10], compares various methods such as CNN, Naïve Bayes, SVM, and ANN in analyzing sentiment based on aspects. The dataset used is movie reviews. The results stated that the CNN method obtained the best accuracy value of 72.17%. Research in 2020 [11] also uses the CNN method with Word2Vec feature extraction in analyzing sentiment based on aspects. The dataset used is a review of bukalapak.com online stores with the classification of aspects, that is accuracy, quality, service, price, packaging, and delivery. This research resulted in an accuracy of 85.54%. While in research [12] comparing the performance of word embedding, that is Word2Vec, Glove, and Fasttext in classifying text using the CNN method, it was concluded that FastText produced better performance. In research [13] using the CNN method in sentiment analysis. This research uses word embedding, i.e., Word2Vec, FastText and GloVe. The dataset used comes from Twitter social media in a product review. It was concluded that the CNN method using FastText has a higher F-measure value, i.e., in the skip-gram scheme of 91.6% and the CBOW scheme of 86.3%.

Another study conducted in 2022 [14], Sukma and Erwin discussed sentiment based on aspects by applying TF-IDF as feature extraction and FastText as feature expansion to improve performance on the NBSVM classification model. The research resulted in an F1-score of 91.24% on the signal aspect and 88.75% on the service aspect by applying SMOTE in overcoming unbalanced data labeling. And in other aspect-based sentiment research [15] Alhakiem and Erwin in 2022, applied TF-IDF feature extraction and FastText feature expansion to logistic regression, then overcame unbalanced data by using SMOTE. It was concluded that by using SMOTE, F1-Score increased by 3.33% in the signal aspect and 12.91% in the service aspect. In research [16] using the SVM method in aspect-based sentiment. The dataset used is tweet data related to products from Telkomsel with signal and service aspects. This study uses TF-IDF feature extraction, as well as FastText as feature expansion using a corpus of tweets, Indonews, and a combination of tweet corpus and Indonews. In this study, the problem of unbalanced datasets was handled using SMOTE. The results obtained are this research produces an F1-score of 95.93% in the signal aspect and 94.53% in the service aspect. Meanwhile research in 2022 [17], Rimdani and Erwin implemented the Gradient Boosting Decision Tree classification technique with SMOTE and Random Undersampling on signal and service aspects. The results of this study show that the use of SMOTE can produce the best performance value compared to Random Undersampling. The final performance value obtained is the F1 score for the signal aspect of 96.035%, an increase of 25.260%, and the F1 score for the service aspect of 90.256%, an increase of 27.918%. Also in research [18], Aditya and Erwin carried out an aspect-based sentiment analysis on Twitter by applying the Glove feature expansion to the Random Forest classification. The accuracy value obtained on the signal aspect is 80.37% and on the service aspect is 80.12%. The usage of the GloVe feature expansion is also able to increase the accuracy value by 13.15% in the signal aspect and 5.37% in the service aspect.

Based on previous research, these studies have not implemented a sentiment analysis based on movie aspects using CNN with added the FastText feature expansion. Therefore, in this research conducted a sentiment analysis based on the movie aspect using the CNN classification model with TF-IDF feature extraction and FastText feature expansion. In addition, unbalanced dataset labeling is handled using SMOTE to maximize the performance of the model.

This research aims to analyze sentiment based on aspects such as plot, acting, and director using the CNN method and FastText feature expansion by overcoming data imbalance using SMOTE. The results of this study provide accuracy and F1-Score values for each aspect as well as an analysis of the effect of FastText feature expansion on the resulting accuracy. The problem limitation in this research is that the dataset used is Indonesian language movie review tweet data, and the method used is FastText feature expansion and CNN classification algorithm. The aspects used in this study are based on the movie review tweet data that has been collected i.e., plot, acting, and director.