

## 1. Introduction

The film industry is a type of industry that is experiencing an increase every year, both from newly produced films and films that have been produced before [1]. This is due to the rapid development of technology. The rapid development of technology affects people to make it easier to find and watch various movies from various websites. With a large selection of movies, it is often confusing for someone to choose which movie they will watch [2]. People previously needed to know the quality of various kinds of movies by looking at reviews or other people's responses to the movie. With various reviews from the audience on a film, a sentiment analysis is needed to categorize the reviews into a sentiment [3].

The process that can automate the collection of opinions, emotions, and views from speech, text, and data sources by Natural Language Processing is called sentiment analysis [4]. The results of sentiment analysis can help in understanding audience satisfaction with a movie. Sentiment analysis evaluates the polarity of text in a sentence or data to establish which opinions expressed are positive, negative, or neutral [3]. The positive nature indicates that the comments given to the movie have a good value, it can be concluded that the reviewed movie is a great movie. Meanwhile, the negative nature indicates that the comments given to the movie have a less good value, it can be concluded that the reviewed movie is a bad movie. However, there is also a neutral trait which means it indicates that the movie is not too good and not too bad. Sentiment analysis can help benchmark the success of a movie by comparing it to similar films. The success of a movie can help the public so that they are not confused in choosing a movie that they will watch. This research can also determine the effectiveness of an algorithm and sentiment analysis techniques that are well used to process movie review data from various sources.

In sentiment analysis, there are several algorithms that can be used. In this research, the feature extraction used is Word2Vec. This feature extraction aims to map information or words in a sentence into a vector. This feature extraction was chosen because research [5] explains that Word2Vec has the advantage of being able to show the contextual similarity of two words in the resulting vector. In sentiment analysis, the imbalanced distribution of class data can affect classification performance. Therefore, Modified Balanced Random Forest is used as a classification method in this research. This classification method was chosen because the data used in this research is imbalanced, it is in accordance with the purpose of this method which is to handle imbalanced data. In addition, this method can also improve accuracy and reduce time complexity. In research [6] the feature selection used is Chi-square and Modified Balanced Random Forest as a classification method was successful in producing an accuracy of 81.75% and f1-score of 71.90%.