

1. Introduction

The film industry continuously produces works, leading to the presence of numerous online platforms that provide a space for movie enthusiasts to provide evaluations and comments on the movie. Online media, particularly movie review websites such as IMDb, Rotten Tomatoes, and Metacritic, have a significant influence on the industry. These movie review websites not only serve as platforms for experts to critique movie but also allow everyone to provide their evaluations and comments. These reviews are useful for movie enthusiasts to determine whether a film is worth purchasing or not [1]. Due to the abundance of data from these reviews, many studies have been conducted, including sentiment analysis.

Opinions are an important aspect of our lives as they can influence behavior and decisions. In the movie industry, movie production companies always seek to know the opinions of consumers or the public about their movies. Movie enthusiasts also want to know other peoples opinions on movie they intend to buy or watch, which is why sentiment analysis has become a supporting field of study [2]. Sentiment analysis is a field that analyzes opinions and sentiments towards a service or product in the form of text [3]. This field analyzes peoples reviews of a movie and categorizes them as positive or negative reviews.

The classification performance of sentiment analysis depends on the features present in the text, which can sometimes pose challenges that may reduce the classification performance. Therefore, feature extraction and selection play a crucial role in addressing such issues [4]. This research is expected to contribute and assist future researchers in selecting methods for data that utilize the Indonesian language, as well as helping and facilitating the public in choosing the films they want to watch. In this study, the feature used is Word2Vec. In a research [5] conducted by Serkan Ballı 2019, Word2Vec has been shown to achieve very high accuracy of 99.64%. The method used in this research is Random Forest, which has proven to be one of the most successful methods among others.

The Random Forest approach will be employed as the chosen method for this study. This method has the advantage of producing high accuracy and efficiently handling large datasets [6]. In a study [7] conducted by Isaiah Steinke 2022, Random Forest achieved an accuracy of 85.27% using a dataset consisting of movie reviews as well.