INTRODUCTION

In this fast-paced technological development, many things can be accessed easily through the internet. One example that benefits from this is the tourism sector. Tourists can easily find information not only about the tourist attractions they want to visit, but tourists can also find information about hotels that tourists will use to rest after traveling.

One of the websites that provide hotel information is TripAdvisor. TripAdvisor is a website that can help tourists to book hotels quickly. Tourists can also add reviews, pictures and ratings of the hotels they visit as a reference for other tourists [1]. These reviews can be used as a consideration by other tourists to determine whether tourists want to visit a place. Research [2] shows that 89% of tourists and 64% of hotel owners believe that reviews of hotels can influence tourists' decisions to rent hotels. In research [2] also mentioned that 95% of tourists booked a hotel by first reading reviews of the hotel. It is one of the important factors that influence tourist decisions to book hotels. However, from the reviews available on the TripAdvisor website, many reviews do not contain certain aspects. This makes it difficult for other tourists to make decisions. Therefore, a multi-aspect sentiment analysis is needed for hotel reviews on the TripAdvisor website.

Previous studies have discussed sentiment analysis on hotel review data. Research [3] obtained an accuracy value of 85.96%, but multi-aspect sentiment analysis was not yet used. In research [4] multi-aspect sentiment analysis was used with seven aspect categories: room, food, location, service, cleanliness, price, and entertainment. While in [1] using the Hybrid Classifier method with five aspect categories: room, location, cleanliness, check-in, and service. The study managed to get an average accuracy score of 84%. The Random Forest method was used by research [5] in analyzing the beauty product domain. Research [5] managed to get an accuracy score of 90.48%, precision 87.27%, recall 70.13%, and F1- score 71.77%. The Word Embeddings spaCy.vector method was used in research [6] along with the Support Vector Machine (SVM) method. The combination of these methods resulted in an accuracy of 64%. Furthermore, [7] used a combination of SVM and Word Embeddings Word2Vec method for various datasets. The results of research [7] get the best accuracy score of 86.90% for the IMDB-EN dataset. Research [8] used Word2Vec for sentiment analysis in Arabic and managed to get an average accuracy score of 82.60% for the combination of Logistic Regression and Word2Vec methods. A comparison of the best Word Embeddings methods is discussed in research [9]. In the study, Word2Vec and GloVe methods were compared. Research [9] managed to get an accuracy score of 87.2% for Word2Vec and 85.7% for GloVe by using the Convolutional Neural-Network (CNN) classification method.

Seeing the potential of the Random Forest method in research [5] which produces a fairly high accuracy value (90.48%) and the Word Embeddings method in research [9] which produces the best accuracy of 87.2%. In this research, we will compare the effectiveness of Word2Vec and GloVe methods using Random Forest classifier for hotel reviews in English.