

1. INTRODUCTION

The number of internet users increases every year. According to the Datareportal report, In January 2023, the number of internet users in Indonesia reached 212.9 million people, indicating a 5.2% increase compared to the previous year. It shows that the internet has been used by approximately 77.0% of the Indonesian population, which totals 276.4 million people [1]. Technological advances have influenced human behavior, including shopping habits. The shift from in-store shopping to online shopping via applications shows a significant change in society's behavior [2].

The online market in Indonesia has undergone significant and rapid growth. The comfort, convenience, and wide array of benefits given to the users are among the primary reasons for its growth. One of the most prominent advantages offered by the online market or e-commerce is that users can shop 24 hours a day without any limitations and without the need to visit physical stores [3]. The Google Play Store is a platform that offers applications for online shopping. One of the features of the Google Play Store is that it allows users to provide ratings and reviews about the applications and services they use. The review is in the form of text containing an assessment and comment on a work or product that plays an important role as a recommendation for application use for new users [4]. This platform has served more than 200.000 transactions every day. One of the well-known e-commerce in Indonesia is Shopee. Recently, it has been downloaded more than 100 million times and gotten up to 8 million reviews on the Google Play Store [5]. User reviews provide information about the strengths and weaknesses of the app. This can influence customers to determine which application to use. Therefore, a sentiment analysis classification process is needed to improve the quality and features of the application so that it can be used as an evaluation material for quality improvement for the application provider or owner.

Sentiment analysis is a methodology employed to obtain and understand information from user assessment of an application, product, or service. These assessments are subsequently categorized into positive or negative sentiment classifications [6]. Meanwhile, multi-aspect sentiment analysis is the classification of sentiments grouped based on different aspects that have been determined from the data used with the aim of the analysis process obtaining more optimal results [7].

To figure out the sentiment of society, it is important to apply an analysis method that can summarize the reviews. Thus, it is suggested to apply RNN with QER feature selection in this research. The RNN is commonly used in natural language processing tasks, usually known as Natural Language Processing (NLP). This method is proposed based on the previous research mentioned above as they demonstrate that RNN can achieve excellent accuracy and efficiency compared to other methods. The ability of RNNs to retain information from preceding steps in a data sequence is highly beneficial in scenarios where historical information or data sequences hold significant importance [8]. Besides, the RNN can analyze sequential text, such as customer reviews, and determine either the positive or negative sentiment of the text. Meanwhile, QER can reduce the number of less relevant features during the classification stage without reducing the quality of sentiment analysis. Therefore, this research uses the combination of RNN and QER to analyze customer reviews to achieve an optimal level of accuracy.

Deep learning is extensively used in sentiment analysis due to previous findings showing its ability to yield better classification results across various data types such as text, video, image, and sound. Recurrent Neural Network (RNN) is one type of deep learning whose work pattern is like how neural networks work in the human brain. In RNN, the input sequence will be mapped into a vector that has a fixed size [9]. In research [10], the researchers utilized RNN to analyze BPJS Health user sentiment and achieved an accuracy of 86.67%, and an F1 score of 86.63%. The RNN algorithm was chosen because it is recognized as the best algorithm inspired by biological neural networks. In another research [11] the RNN model achieved an accuracy rate of 87.42% by applying to the IMDB film review dataset, and to measure its performance, researchers achieved a recall of 87.17% and a precision of 87.53%. The classification performance results of research [12] Query Expansion Ranking (QER) got an accuracy of 77.62% by using 80% features with the highest value in Multinomial Naïve Bayes. In research [13] using QER and classification with Multinomial Naïve Bayes obtained an accuracy results of 75%, precision of 85.21%, and f1 score of 70.76%. There have been several research conducted multi-aspect sentiment analysis, such as analyzing sentiment from hotel reviews in various aspects including food, service, rooms, facilities, price, and location using the LSTM method to get notable performance [14]. Meanwhile, another aspect-based research is also implemented by the researcher to analyze aspects of mosques, halal food, and toilets at halal tourist attractions in Asian countries using the CNN method also obtained good performance [15].

There are several previous research used as related studies to create sentiment analysis systems using the Recurrent Neural Network (RNN) and QER method. The following are some related published studies.

In research [10], sentiment analysis is used on reviews from BPJS Health users using a dataset obtained from Twitter. It obtained an accuracy rate of 86.67%, a precision value of 87%, a recall rate of 86.66%, and an F1 score of 86.63%. The most optimal average value of the performance of this sentiment analysis system was obtained using word2vec weighting and RNN with Long Short-Term Memory (LSTM) type. The data partition ratio used is 90:10, which means 90% training data and 10% test data.

Another study conducted an analysis of sentiment in tweets related to the COVID-19 vaccine utilizing Recurrent Neural Network (RNN) and Naïve Bayes. The study involved data classification using the TF-IDF weighting method and compared the performance of the RNN model with that of Naïve Bayes. This research uses 5000 data grouped into several categories, which are 3800 expressing positive sentiment, 800 with negative sentiment, and 400 conveying



neutral sentiment. The study reported that the RNN model achieved an optimal accuracy of 97.77%, while the Naïve Bayes model achieved 80% accuracy [16].

Furthermore, research on sentiment analysis [17] has been conducted using Genetic Algorithm-Support Vector Machine (GA-SVM) combined with QER, resulting in high accuracy. The precision value obtained was 96.78%, the recall was 96.76%, and the f-measure was 96.75%. These results demonstrate a significant increase in performance in sentiment classification using this method.

In addition, research [18] conducted sentiment of the Merdeka Curriculum by comparing the performance of Multinomial Naïve Bayes and Bernoulli Naïve Bayes with QER. This research evaluated 106 tweets with positive sentiment and 164 tweets with negative sentiment. The results showed that the Multinomial Naïve Bayes method achieved an accuracy of 98.889%, a recall of 98.131%, a precision of 99.057%, and an f-measure of 98.591%. On the other hand, the Bernoulli Naïve Bayes method achieved an accuracy of 94.815%, a recall of 87.850%, a precision of 98.947%, and an f-measure of 93.069%.

Research related to aspect-based sentiment analysis using the Long Short-Term Memory (LSTM) method obtained impressive accuracy result of 79% [14]. Additionally, another research [15] conducted aspect-based sentiment analysis employed the CNN deep learning method for data classification. The approach yielded an accuracy of 98.299% for aspect classification, and a sentiment classification accuracy of 93.96%. Moreover, the Naïve Bayes algorithm was employed in research [19] to classify reviews from the society of Tegal City based on four aspects, which are tourism/entertainment, education, public facilities/services, and culinary aspects. The accuracy value obtained from the research is up to 75% on testing data and surpassed 90% on training data. Likewise, Likewise, the Naïve Bayes algorithm was utilized in research [20] to assess customer reviews of the Bakso President Malang restaurant across the aspects of food, service, and atmosphere. Classification in this research generates a high level of accuracy. It obtained an accuracy of 88% in the aspect of food, 76% in the aspect of services, and 84% in the aspect of atmosphere. Other research [7] on sentiment also used the Naïve Bayes algorithm to evaluate culinary tourism reviews on the TripAdvisor website. The aspects used include icons, atmosphere, service, price, facilities, and taste. This analysis yielded an impressive accuracy rate of 98.67%, surpassing the results of prior research.

Based on the background explained, the research question in this study is how to determine the optimal combination of RNN and QER for sentiment analysis and how the optimal performance combination of the RNN with QER models is in analyzing sentiment regarding aspects of delivery costs, delivery speed, and application access on the Shopee application.