Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

Multi-aspect Sentiment Analysis of Shopee Application Reviews using RNN Method with Query Expansion Ranking

Ariqoh Novitasari, Yuliant Sibaroni*, Diyas Puspandari

Informatics, Faculty of Informatics, Telkom University, Bandung, Indonesia
Email: ariqohnovitasari@student.telkomuniversity.ac.id, yuliant@telkomuniversity.ac.id, diyaspuspandari@telkomuniversity.ac.id
Correspondence Author Email: yuliant@telkomuniversity.ac.id
Submitted: **17/07/2024**; Accepted: **08/09/2024**; Published**: 09/09/2024**

Abstract−Online shopping using e-commerce is a common activity society does in this digital era. Shopee is one of the well-known e-commerce in Indonesia. There are a lot of e-commerce platforms that can easily be accessed through mobile applications like Google Play Store. Users are allowed to review and rate the application they have downloaded. The review from users can influence potential customers who read online reviews to choose a product or service. Reviews from users also become an opportunity for e-commerce companies to advance their performances and services, thus increasing customer satisfaction. Therefore, to enhance the understandability of user reviews, a system that can efficiently analyze the sentiment is needed. This study aims to design and establish a system that can perform sentiment analysis on the selected aspects. The amount of data used in this research is 4000 reviews. Sentiment classification is implemented by using the Recurrent Neural Network (RNN) algorithm with Query Expansion Ranking (QER) feature selection to classify Shopee application reviews into two classes, which are positive and negative. Feature selection is used to reduce less useful features so that the classification model conducts the classification process optimally and more efficiently. In conclusion, the evaluation results based on an 80:20 data split ratio indicate that the combination of RNN with QER achieves the highest accuracy of 95% in the delivery cost aspect, 93% in the delivery speed aspect, and 86% in the application access aspect. The combination of RNN with QER feature selection in this study achieved the best performance in analyzing sentiment for each aspect studied.

Keyword: Google Play Store; Recurrent Neural Network; Reviews; Shopee; Query Expansion Ranking

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 thats play an important role as a recommendations 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

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

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 dan 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.

2. RESEARCH METHODOLOGY

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

2.1 System Design

Figure 1 below is a flowchart of the research design system used for this study. The flowchart illustrates the sequence of stages carried out in this research.

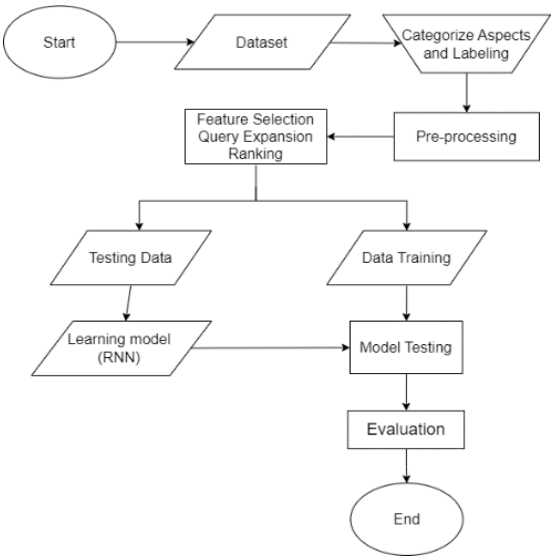

Figure 1. Research Design System

Figure 1 shows the initial research process, which starts from acquiring the dataset from user reviews of the Shopee application on the Google Play Store. Subsequently, the aspects are identified and manually labeled. These aspects include delivery cost, delivery speed, and application access. The next step is dataset pre-processing, followed by feature selection utilizing QER and applying the RNN model. After that, the data is segregated into training and testing sets, followed by the modeling phase. After the completion of the modeling process, testing and evaluation ensue.

2.2 Dataset

The dataset used in this research is obtained from an analysis of the Shopee application on the Google Play Store. The data was collected through web scraping using the Google Play Scraper library in the Python programming language. This library facilitated the extraction of review data from Shopee users who submitted their reviews on the Google Play Store. Specifically, the dataset comprises reviews in Indonesian, sorted by relevance, with a total of 4000 data.

2.3 Categorize Aspect and Labelling

Categorize aspect is the process of determining the group aspect of a text. This process involves text analysis to identify entities or specific concepts in the form of categories from user reviews within a document. For example, in a Shopee application review, these aspects can encompass details about delivery costs, accessibility of the application, and the speed of delivery. This process helps in analyzing and organizing the user reviews into categories that enhance the system's ability to extract specific and relevant information more efficiently.

Labeling is a stage conducted after categorizing aspects by assigning a specific label or category to each aspect that has been determined. For instance, the aspect of delivery cost is categorized as either positive or negative based on user reviews, including assessments of delivery speed and application access. This procedure helps the segmentation of reviews or data into more readily interpretable categories, enabling further in-depth analysis. The process of labeling helps to identify sentiment or overall performance within previously delineated parameters. The results of the comparison between positive and negative data are available in Table 1 below.

Table 1. Data labeling result

Aspect/label	Positive	Negative
Delivery Cost	3294	706
Delivery Speed	3280	720
Application Access	3067	933

2.4 Pre-processing

Before proceeding to the next stage, it is required to conduct pre-processing of the dataset. This process involves a series of actions designed to cleanse, organize, and prepare data to make it more suitable for analysis or further processing. The steps involved in this process are as follows:

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

a. Lowering is the process of changing all of the text in the dataset to lowercase. This process aims to ensure consistency in the representation of words so that words that have the same meaning but are written in upper or lower case will be considered the same or identical.
b. Removing symbols is the process of removing unnecessary or irrelevant characters or symbols in the text. These symbols can be punctuation marks, numbers, or special characters that do not provide essensial information.
c. Tokenization, this stage involves segmenting the text into smaller units known as tokens. This process helps the preparation of the text for further analysis by considering each token as an individual entity.
d. Stopword removal, stopwords are common words that appear frequently but do not provide high information value. This step involves eliminating stopwords that do not significantly add to the understanding of the content or meaning of the document. This research uses the spaCy library, which is a Python-based tool used in the field of natural language processing (NLP). It is specifically designed for processing large amounts of words or text.
e. Stemming is the process of finding the base form of a word by removing its endings. The purpose is to reduce the variation of words which have the same root. For example, "berlari", "berlarian", and "berlari-lari" can all be grouped as "lari." In this study, researchers used a literary stemmer, which is widely used for stemming Indonesian sentences.

2.5 TF-IDF Feature Weighting

TF-IDF is a method for weighting words to perform classification. This method gives a weight or score to each word in a document based on how often the word appears (Term Frequency) and how unique the word is in the entire text corpus (Inverse Document Frequency). The following equation represents the TF-IDF feature weighting:

$$TF_{t,d} = \sum_{x \in d} f_t(x) \quad (1)$$

$$IDF_t = \log(\frac{D}{df_t}) \quad (2)$$

$$W_{(t,d)} = TF_{(t,d)} * IDF_t \quad (3)$$

The following is a table of the number of terms extracted in the delivery cost aspect, delivery speed aspect, and application access aspect:

Table 2. Number of extracted terms

Aspect Studied	Number of extracted terms
Delivery cost	5615
Delivery speed	5453
Application Access	5466

The following is a table of the top 10 terms with the highest TF-IDF values:

Table 3. Term with the highest TF-IDF values

No.	Name of term	TF-IDF value
1	Shopee	182.41461855466554
2	Belanja	156.30506805438537
3	Nya	125.86643557044226
4	Kirim	125.02513507301633
5	Aplikasi	121.44337978851772
6	Bagus	108.6573314464487
7	Barang	107.3107911911249
8	Mudah	105.22004798352381
9	Banget	79.4817024622823
10	Bantu	75.23652744497221

2.6 Feature Selection Query Expansion Ranking

QER method is one of the selection features based on the Query Expansion Technique, which is useful for assessing whether a word is influential or not in a study [21]. In this research, the features or attributes selected are words used to represent documents or text. Selecting features helps to reduce the size of the data and improve efficiency and performance. The working principle of QER is based on the term weighting method [9]. Query Expansion enriches a user's search query by adding additional relevant words or phrases. The term weighting method provides scores and rankings based on these scores. Below is the equation for QER:

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

$$pf = \frac{df_{+}^{f}+0.5}{n^{+}+1} \quad (4)$$

$$qf = \frac{df_{-}^{f}+0.5}{n^{-}+0.5} \quad (5)$$

$$score\ f = \frac{pf-qf}{|pf+qf|} \quad (6)$$

The following table shows several examples of results obtained through the QER process. The following are 4 data samples used in this research:
Data 1 : aplikasi bantu pengiriman cepat (positive review)
Data 2 : aplikasi berat ongkos kirim mahal lambat sampai (negative review)
Data 3 : barang bagus harga murah pengiriman cepat (positive review)
Data 4 : aplikasi jelek ongkos kirim mahal (negative review)
The explanation regarding the QER feature selection process will be demonstrated using manual calculations as follows

Table 4. Example data of manual QER calculation

No.	Positive words	Number of words	Negative words	Number of words
1	Pengiriman	2	Aplikasi	2
2	Cepat	2	Ongkos	2
3	Aplikasi	1	Kirim	2
4	Bantu	1	Mahal	2
5	Barang	1	Berat	1
6	Bagus	1	Lambat	1
7	harga	1	Sampai	1
8	Murah	1	Jelek	1
	Total Dokumen	10		12

2.7 Data Splitting

Data splitting is a step that must be completed before doing classification modeling process. The main purpose of data splitting is to ensure that the developed data can be tested on data that was not involved in the training process. This helps to provide an objective evaluation of the model's performance on new data. In this process, the data is devided into training data and testing data. Training data is used to train the classification model, while testing data is utilized to conduct trials on the trained models. This research uses a comparison of training and testing data, which is 80;20, 70;30, and 60;40. The selection of this ratio was based on previous research [22]. In this study, the optimal dataset splitting ratio for achieving the highest accuracy is determined to be 80% for training data and 20% for testing data.

After the data splitting process, a class balancing process is implemented to address the disparity between the quantities of positive and negative data. Therefore, class balancing is implemented in data training. It makes the classification model impartial to a class. Undersampling was chosen for class balancing because the training data already contained a substantial number of samples. If oversampling had been used, it would have introduced random data, potentially biasing the classification model.

2.8 Modeling RNN

The RNN structure used in this research involves an input layer that receives a sequence of data, followed by a recurrent layer that uses a cycle mechanism to process and understand the data sequentially. Each time step in the data sequence produces a hidden state that passes to the next step. Each input or hidden layer in the RNN utilizes the same parameters and performs the same function to reduce the complexity of the obtained output results, distinguishing it from other neural networks [23]. The architecture used for sequential data in RNN is depicted in Figure 2 below.

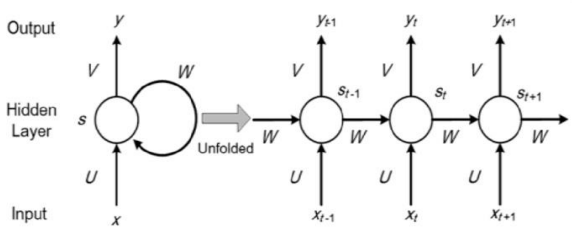

Figure 2. The architecture of RNN

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

RNN has other layers such as a dropout layer to reduce overfitting and an output layer to cc predictions based on understanding the sequence of data obtained. The RNN model was built by using a library from the Python programming language, known as TensorFlow, with details in Table 5 below.

Table 5. RNN Architecture

Layer	Parameter	Total Number of parameter
Input	Shape = (row_len, 1, col_len)	0
Reccurent		
SimpleRNN1	Units = 256, return_sequences = True	1780224
Dropout	Rate = 0.7	0
SimpleRNN2	Units = 128, return_sequences = True	49280
Dropout	Rate = 0.7	0
SimpleRNN3	Units = 64	12352
Dropout	Rate = 0.7	0
Output		
Dense	Units = 1, Activation = Sigmoid	65

The above parameter is the best result of the experiment conducted by the author to create a more optimal classification process. This research belongs to the Biner class. Therefore, the loss parameter used is binary_crossentropy, and the optimizer used is the Adam optimizer. This research also uses an early stopping parameter to avoid overfitting in the classification process.

2.9 Model Evaluation

The data in this research is divided into 2 parts, which are training data and testing data. The author divides the data into 3 times with the comparison of 80;20, 70;30, and 60;40. This ratio is based on the previous research [19]. The best accuracy is obtained from the ratio of 80;20 where 80% is training data and 20% is testing data. RNN evaluation that has been created is conducted, and training of the RNN model was carried out. To get performance value, a Confusion Matrix is needed so that it can determine the model that has been built.

Table 6. Confusion Matrix

True Label	Predicted Label	
	Negative	Positive
Negative	True Negative (TN)	False Positive (FP)
Positive	False Negative (FN)	True Positive (TP)

Table 6 above consists of several values, in which True Negative is the value obtained when the prediction and the true label are negative. Meanwhile, the False Positive value is obtained when the prediction result is positive and the true label is negative. Moreover, the False Negative is the value obtained when the prediction result is negative and the actual label is positive. Next, the True Positive is the value obtained when the prediction result and the true label are positive.

a. Accuracy is a metric that can describe how well a classification model provides correct prediction results. The accuracy equation can be seen as follows

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP} * 100\% \quad (7)$$

b. F1 Score is a metric that can describe the weighting of recall metrics and precision metrics. The F1-Score equation can be seen as follows.

$$F1\ Score = \frac{2TP}{2TP+FP+FN} * 100\% \quad (8)$$

3. RESULT AND DISCUSSION

3.1 Test Result

In the feature selection process utilizing Query Expansion Ranking (QER), the feature values from the trained data are gathered and then sorted from the smallest to the largest value. The QER values calculated in this research are detailed in Table 7 below. Accuracy is a metric that can describe how well a classification model provides correct prediction results. The accuracy equation can be seen as follows.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

Table 7. QER Value Calculation Result

Ranking	Feature	pf	qf	$score\ f$
1	Internet	0.000264	0.000013	0.050500
2	Cache	0.000218	0.000013	0.061679
3	Pinjam	0.000616	0.000038	0.066145
4	Verifikasi	0.000203	0.000013	0.066611
5	Manfaat	0.000188	0.000013	0.072408
6	Koin	0.000524	0.000038	0.078427
7	Hadiah	0.009157	0.000013	0.087703
8	Rasa	0.000157	0.000013	0.087703
9	Daftar	0.000149	0.000013	0.092600
10	Keranjang	0.000142	0.000013	0.098079
⋮	⋮	⋮	⋮	⋮
6697	Baik	0.001811	0.001807	455.149914

This study used three ratios of training data and testing data. Here are the accuracy values obtained from three data split trials with ratios of 80:20, 70:30, and 60:40.

Table 8. Accuracy comparison with data splitting

Trial	Accuracy of Delivery Cost Aspects	Accuracy of Delivery Speed Aspects	Accuracy of Application Access Aspects
1	92%	93%	86%
2	90%	92%	83%
3	90%	92%	82%
Average	90.66%	92.33%	83.66%

Based on the data in Table 8, it can be concluded that the highest accuracy was achieved when using a data split ratio of 80:20 for training and testing. Thus, the accuracy results from the 80:20 data split ratio were used for this research. The study examined the selection of Query Expansion Ranking (QER) features, including variations in the percentage of QER features (100%, 80%, 60%, 40%, 20%) relative to the total features across three aspects of delivery cost, delivery speed, and application access. The findings from these tests will be presented in the following sections.

a. Aspect of Delivery Cost
The evaluation of the RNN model's performance on the delivery cost aspect, constructed using the QER feature selection, yielded confusion matrix results. These results are detailed in Table 9 below.

Table 9. Confusion matrix on delivery cost aspect

True Label	**Predicted Label**	
	Negative	**Positive**
Negative	107	24
Positive	43	626

The accuracy value and f1 score obtained from the classification performance results on the delivery cost aspect can be seen in Table 10 below.

Table 10. Classification performance results in delivery cost aspect

Number of Feature	**Accuracy**	**F1 score**
100%	92%	92%
80%	92%	92%
60%	92%	92%
40%	93%	93%
20%	95%	95%

b. Aspect of Delivery Speed
The evaluation of the RNN model's performance on the delivery speed aspect, constructed using the QER feature selection, yielded confusion matrix results. These results are detailed in Table 11 below.

Table 11. Confusion matrix on delivery speed aspect

True Label	**Predicted Label**	
	Negative	**Positive**

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

Negative	134	16
Positive	37	613

The accuracy value and f1 score obtained from the classification performance results on the delivery speed aspect can be seen in Table 12 below.

Table 12. Classification performance results in delivery speed aspect

Number of Feature	Accuracy	F1 score
100%	93%	93%
80%	94%	94%
60%	93%	93%
40%	94%	94%
20%	89%	89%

c. Aspect of Aplication Access
The evaluation of the RNN model's performance on the aspect of application access, constructed using the QER feature selection, yielded confusion matrix results. These results are detailed in Table 13 below.

Table 13. Confusion matrix on the aspect of application access

True Label	Predicted Label	
	Negative	**Positive**
Negative	152	38
Positive	76	534

The accuracy value and f1 score obtained from the classification performance results in the aspect of application access can be seen in Table 14 below.

Table 14. Classification performance results in the aspect of application access

Number of Feature	Accuracy	F1 score
100%	86%	86%
80%	85%	85%
60%	85%	85%
40%	86%	86%
20%	84%	82%

3.2 Analysis of Test Result

After doing the test and obtaining the result of the RNN model, in this stage, comparison of the performance of each aspect againts the RNN algorithm is conducted. The following is a graph of the comparison results for each accuracy aspect of delivery costs based on the percentage of the number of QER features used.

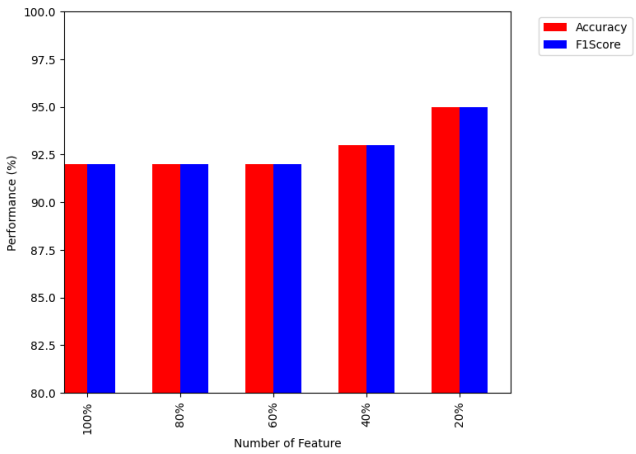

Figure 3. Comparison of accuracy and f1 score in delivery cost aspect

Based on the above Figure 3, the accuracy comparison results of the combination of RNN Method with QER percentage of features used. Notably, at a QER feature percentage of 20%, there is a significant performance

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

improvement compared to the QER percentage of other features, with an accuracy and f1-score of 95%, making it the feature percentage with the best accuracy compared to the total percentage of other features.

The following graph illustrates the comparison results for each accuracy aspect of delivery speed based on the percentage of the number of QER features used.

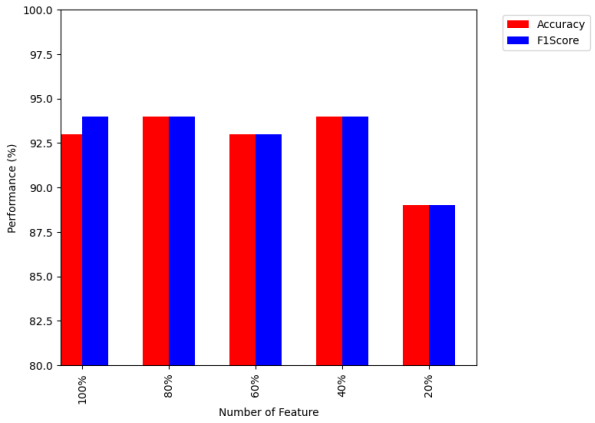

Figure 4. Comparison of accuracy and f1 score in delivery speed aspect

Based on the above Figure 4, the accuracy comparison results of the combination of RNN Method with QER percentage of features used. The optimum accuracy in this aspect is 40% of the QER feature percentage with an accuracy and f1-score of 94%, which has the best performance and is more efficient than others. The highest accuracy and the fewest number of features make this model faster in classifying.

The following graph shows the comparison results for each accuracy aspect in the aspect of application access based on the percentage of the number of QER features used.

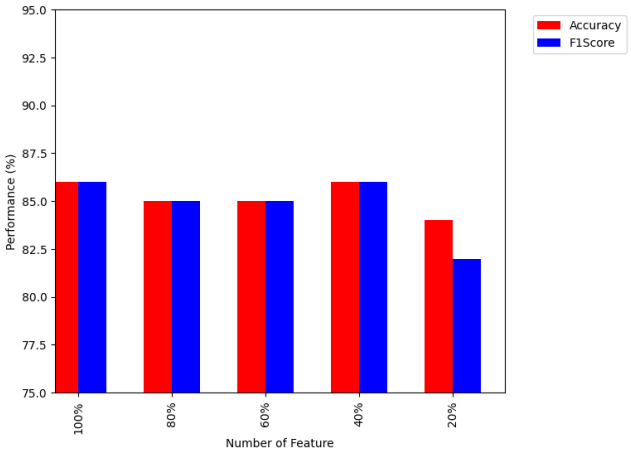

Figure 5. Comparison of accuracy and f1 score in aspect of application access

Based on the above Figure 5, the accuracy comparison results of the combination of RNN Method with QER percentage of features used. The optimal accuracy is achieved with the QER feature at a 40% percentage with an accuracy and f1-score of 86%, outperforming other QER percentages.

4. CONCLUSION

In this research, a system was developed for the sentiment analysis of 4000 Indonesian data taken from Google Play Store reviews about the Shopee application. This research analyzes sentiment from 3 aspects: delivery costs, delivery speed, and application access on the Shopee application. This research uses the Recurrent Neural Network (RNN) method and Query Expansion Ranking (QER) feature selection with binary_crossentropy parameters and early stopping parameters to provide optimal accuracy values. Based on Figure 3, Figure 4, and Figure 5 above, it is found that the combination of the RNN model with QER feature selection performs the best classification using 40% or 20% of the best features. Therefore, using QER can improve performance compared to not using QER. Accuracy and f1-score in the delivery cost aspect obtained results of 95%. Accuracy and f1-score in the delivery speed aspect obtained results of 94%. Accuracy and f1-score in the application access aspect obtained results of 86%. The research concludes

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 825−834
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5605

that the collaborative performance of the RNN method and QER feature selection achieves optimal accurasy with overall accuracy exceeding 80% in sentiment classification across multiple aspects.

REFERENCES

[1] K. Simon, "Digital 2023 : Indonesia," Datareportal. Accessed: Jan. 05, 2024. [Online]. Available: https://datareportal.com/reports/digital-2023-indonesia
[2] A. N. Ardianti and Widiartanto, "Pengaruh Online Customer Review dan Online Customer Rating terhadap Keputusan Pembelian melalui Marketplace Shopee (Studi pada Mahasiswa Aktif FISIP Undip)," 2019. [Online]. Available: http://teknonisme.com
[3] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012043.
[4] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and Samudi, "ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE," 2020.
[5] M. Afdal and L. Waroka, "Shopee Application Review Classification Using Probabilistic Neural Network Algorithm And K-Nearest Neighbor," *IJIRSE Indones. J. Inform. Res. Softw. Eng.*, vol. 2, pp. 49–58, 2022.
[6] F. V. Sari and A. Wibowo, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," *J. SIMETRIS*, vol. 10, no. 2, 2019.
[7] S. A. Azzahra and A. Wibowo, "ANALISIS SENTIMEN MULTI-ASPEK BERBASIS KONVERSI IKON EMOSI DENGAN ALGORITME NAÏVE BAYES UNTUK ULASAN WISATA KULINER PADA WEB TRIPADVISOR," . *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 4, 2020, doi: 10.25126/jtiik.202071907.
[8] D. T. Adherda, M. Hikmatyar, and Ruuhwan, "GENDER CLASSIFICATION BASED ON VOICE USING RECURRENT NEURAL NETWORK (RNN)," *Antivirus J. Ilm. Tek. Inform.*, vol. 17, no. 1, pp. 111–122, Oct. 2023, doi: 10.35457/antivirus.v17i1.3049.
[9] L. Kurniasari and A. Setyanto, "Sentiment Analysis using Recurrent Neural Network," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1742-6596/1471/1/012018.
[10] F. Faturohman, B. Irawan, and C. Setianingsih, "ANALISIS SENTIMEN PADA BPJS KESEHATAN MENGGUNAKAN RECURRENT NEURAL NETWORK," 2020.
[11] A. Patel and A. K. Tiwari, "Sentiment Analysis by using Recurrent Neural Network," 2019. [Online]. Available: https://ssrn.com/abstract=3349572
[12] L. M. Siniwi, A. Prahutama, and A. R. Hakim, "Query Expansion Ranking Pada Analisis Sentimen Menggunakan Klasifikasi Multinomial Naïve Bayes," *J. Gaussian*, vol. 10, pp. 377–387, 2021.
[13] Y. A. V. Gunawan, N. A. S. ER, I. B. M. Mahendra, I. M. Widiartha, I. G. N. A. C. Putra, and I. G. A. G. A. Kadyanan, "Analisis Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Multinomial Naïve Bayes dan Query Expansion Ranking," *J. Elektron. Ilmu Komput. Udayana*, vol. 11, 2022, [Online]. Available: http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/
[14] W. Astriningsih and D. H. Fudholi, "Multi Aspek Sentimen Analisis pada Review Hotel Menggunakan Deep learning," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 3, p. 433, 2023, [Online]. Available: https://jurnal.mdp.ac.id/index.php/jatisi/article/view/5321
[15] R. Naquitasia, D. H. Fudholi, and L. Iswari, "Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning," *J. Teknoinfo*, vol. 16, no. 2, p. 156, 2022, doi: 10.33365/jti.v16i2.1516.
[16] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, Aug. 2021, doi: 10.29207/resti.v5i4.3308.
[17] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, "A Combination of Query Expansion Ranking and GA-SVM," *Procedia Comput. Sci.*, 2021.
[18] M. Yusran, S. Siswanto, and A. Islamiyati, "Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking," 2024. [Online]. Available: http://sistemasi.ftik.unisi.ac.id
[19] N. F. Putri, M. F. Hidayattullah, and D. I. Af'idah, "Sentimen Analisis Kota Tegal Berbasis Aspek Menggunakan Algoritma Naïve Bayes," *Infomatek*, vol. 26, no. 1, pp. 45–54, 2024, doi: 10.23969/infomatek.v26i1.11209.
[20] P. Whita, F. A. Bachtiar, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naive Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1090–1099, 2020, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7134
[21] H. Atsqalani, N. Hayatin, and C. S. K. Aditya, "Sentiment Analysis from Indonesian Twitter Data Using Support Vector Machine And Query Expansion Ranking," *J. Online Inform.*, vol. 7, no. 1, p. 116, Jun. 2022, doi: 10.15575/join.v7i1.669.
[22] A. D. A. Putra and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, 2021, [Online]. Available: http://jurnal.mdp.ac.id
[23] D. Yolanda, K. Gunadi, and E. Setyati, "Pengenalan Alfabet Bahasa Isyarat Tangan Secara Real-Time dengan Menggunakan Metode Convolutional Neural Network dan Recurrent Neural Network," 2020.