Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

The Comparison of RNN and Maximum Entropy on Multi-Aspect Sentiment Analysis of Gojek Application

Nida Umulhoir*, Yuliant Sibaroni, Fitriyani

Faculty of Informatics, Informatics, Telkom University, Bandung, Indonesia
Email: nidaumulhoir@student.telkomuniversity.ac.id, yuliat@telkomuniversity.ac.id, fitriyani@telkomuniversity.ac.id
Correspondence Author Email: nidaumulhoir@student.telkomuniversity.ac.id
Submitted: **10/08/2024**; Accepted: **10/09/2024**; Published**: 12/04/2024**

Abstract−The use of mobile applications is increasingly in demand by the public because it can facilitate users in their daily activities. Gojek application is one of the applications used as online transportation which ranks first with the highest download rate in Indonesia. Despite getting to the top level, gojek experienced a significant decline in downloads compared to the previous download average. This is an opportunity for researchers to use sentiment analysis material to find user reviews of gojek applications in several aspects. This research compares machine learning and deep learning methods. Maximum Entropy and Recurrent Neural Network (RNN) methods are used by combining the two methods with Chi-Square as feature selection and TF-IDF as feature extraction. The comparison was carried out on 4052 data and gave positive or negative sentiments on the aspects of availability, system, comfort, and transactions. The results show that the value contained in the RNN method gets better results than using the Maximum Entropy method with an accuracy value on the availability aspect of 90.47% and an F1-Score value of 83.85%, an accuracy value on the system aspect of 88.97% and an F1-Score value of 85.36%, an accuracy value for the comfort aspect of 81.79% and an F1-Score of 69.275% and an accuracy value on the transaction aspect of 94.05% and an F1-Score of 90.45%. Based on these results, this study can prove an increase in accuracy of more than 5% in each aspect used.

Keywords: Gojek; Setiment Analysis; Maximum Entropy; Recurrent Neural Network; Chi-Square

1. INTRODUCTION

Transportation is an essential part of society, enabling people to carry out daily activities. However, technological advances can also change transportation. One such innovation is online-based transportation, which people can access via smartphones anytime and anywhere. Gojek is an application offering online transportation that has been operating in Indonesia since 2009 [1]. Based on the results of The State of Mobile 2024 report published by data.ai and quoted from the website portal databoks.katadata.co.id [2], that Gojek is the best online transportation application with 957 thousand downloads per month from smartphone users in Indonesia. However, this Figure is down by 29% from the previous average, reaching 1.35 million downloads per month in 2022, while the Gojek app could reach 1.65 million monthly downloads in 2020. As such, 2023 is the year with the lowest number of downloads from 2020 to 2023.

Information about product reviews and ratings is an example of information that is very important for decision-making. Reviews and information about a product are stored in text form, therefore text mining is a solution for retrieving information in text form. Based on the above report, one way to find out why the number of downloads declined is to look at the reviews made by users for the Gojek application on the Google Play Store. These reviews provide their opinions or thoughts about their experience with the app and consist of positive, negative, and neutral reviews. That way, the company can use these reviews to improve its services related to online transportation. To collect data through reviews in the Google Play Store app, a data processor is needed that uses aspect-based sentiment analysis to identify reviews that are positive or negative toward certain elements. One data mining component is Aspect-Based Sentiment Analysis (ASBA), which can be used to obtain more detailed information about aspects derived from user reviews [3].

Several previous studies discussed online transportation through various approaches. Through research [4], Support Vector Machine (SVM) and Naïve Bayes models were used to examine online transportation applications using sentiment on Twitter. The GrabID application has the highest accuracy with a precision value of 66.57%, recall of 57.14%, and accuracy of 84.08%. Research [5] on improving the accuracy of Amazon food review sentiment classification with Bidirectional LSTM and BERT Embedding found that the method obtained better accuracy results of 93%. This is because deep learning focuses on the context of the sentence when doing the word vector extraction process so that it can produce a good value representation of the word.

Based on the above research, the author compared the Machine Learning and Deep Learning methods, namely Maximum Entropy with Recurrent Neural Network (RNN) on Gojek application reviews on the Google Play store. Based on previous research, in this study, researchers prove that the use of Deep Learning methods are better at classifying data than Machine Learning methods. So, the author uses the Maximum Entropy and RNN methods that have never been studied in previous studies and adds Chi-Square as feature selection and TF-IDF as feature extraction.

This research mainly focuses on sentiment analysis of online transportation service reviews on the Gojek application based on Availability, System, Comfort, and Transaction factors. To perform this sentiment analysis, the dataset used comes from user reviews of the Gojek application using the Indonesian language on Google Play store. This data is crawled from August 2023 to September 2023. This research has two objectives: first, to evaluate the level of model performance in terms of Availability, System, Comfort, and Transaction. The next goal is to evaluate

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

the comparison results of the Maximum Entropy method with the Recurrent Neural Network (RNN). This is done by using Chi-Square for feature selection and TF-IDF for feature extraction for each method.

Aspect-based Analysis (ABSA) is an analysis method that aims to identify sentiment aspects in text, whether it consists of one or several relevant aspects [6]. ABSA also aims to classify relevant sentiment elements through text, such as aspect terms or aspect categories, sentiment polarity, and opinion terms. Previous research [7] used the Random Forest Classifier algorithm to perform aspect-based sentiment analysis on Tentrem Yogyakarta hotel reviews. The results showed that the room aspect has the highest accuracy rate compared to other aspects; this aspect has a value of 90 percent accuracy value and a 90 percent F1 score. These values were obtained by using the best number of trees and tree depth parameters which greatly influenced the prediction results.

On the other hand [8], research discussed the effect of normalization, TF-IDF, and feature set selection on sentiment classification using Maximum Entropy in Grab and Gojek case studies. This study found that the use of the Maximum Entropy method resulted in an accuracy value of 90.67% and an F1-Score value of 84.3%, indicating that the use of TF-IDF does not significantly affect the classification process's results but can still improve system performance. In research [9] on comparing machine teaching algorithms for aspect-based sentiment analysis on female daily reviews. This study used three sentiment categories: Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). The feature extraction is done using Chi-Square through K-Best Selected Feature. The results show that the SVM algorithm with linear kernel gets the best accuracy value of 67.10%; the Random Forest algorithm gets an accuracy value of 65.76%, and the KNN algorithm with K 50 gets a lower accuracy value of 60%.

Furthermore, in research [10] aspect-based sentiment analysis was carried out on Mybluebird application reviews using N-Gram and logistic regression algorithms. The aspects studied include taxi user experience, function app and user interface. Compared to the bigram, trigram, and unigram models, the unigram model is the best. By using the unigram model, the curation value is 0.888, the accuracy value is 0.987, the recall value is 0.877, the F1-Score value is 0.925, and the auc score value is 0.95. In [11], the researchers used the Recurrent Neural Network (RNN) method to learn visitors' feelings about Bali beaches on the trip advisor website. In this study, the authors classified the five best beaches in Bali. They found that Kelingking Beach received the highest accuracy value of 90%, with a testing ratio of 80:20 as a reference.

2. RESEARCH METHODOLOGY

2.1 Methodology

In this study, building a system that can classify sentiments and categories regarding review data on the gojek application. The description of the system to be built is in Figure 1:

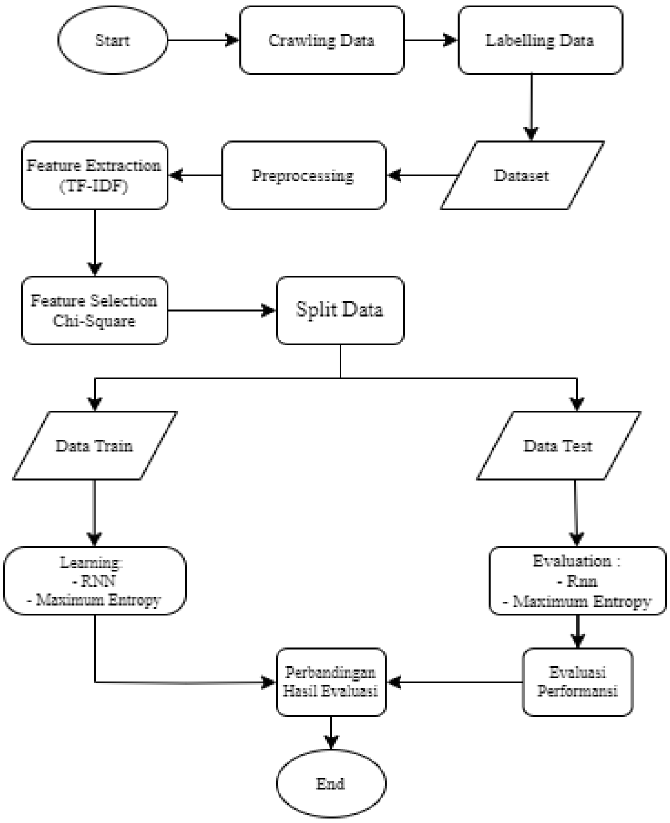


Figure 1. System Design

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

2.2 Dataset

The dataset used in this study was collected from Gojek application reviews on *Google Playstore* using the *Google Play scraper* API. The author then manually labeled the data to be grouped into two sentiment classes and four category aspects. The 4052 review datasets used are in Indonesian.

a. Aspect Grouping Based on Words in Documents

Accessibility, system, convenience, and transaction are the four elements used in this research. This process aims to categorize; a review containing a keyword will be included in one of the element categories. The table 1 below the categorized elements used:

Table 1. Description of Each Aspect

Aspect	Description	Word Example
Availability	The availability aspect of the Gojek application can include transportation for delivering or picking up users.	jemput, cancel
System	The Gojek application system includes aspects of application performance against the device's response, such as errors, access problems to the system, and no response from the system.	akses, server, eror, sistem
Comfort	The comfort aspect of the Gojek application includes the level of user satisfaction and comfort with the services provided.	aman, selamat, nyaman, percaya, kecewa, lama
Transaction	Gojek application transactions include aspects of the application, such as discounts and the number of relatively expensive or cheap tariffs.	potongan, mahal, tarif, promo

b. Labelling Data

Labeling is the process of manually labeling a data set. In this research, the labels assigned consist of two sentiment classes: positive and negative. Reviews with positive values are assigned a value of 0, while reviews with negative values are assigned a value of 1. The review data and an explanation of the sentiment towards a given element are shown in table 2 below:

Table 2. Example data and explanation of each aspect of the labeled data

No	Review	Availability	System	*Comfort*	Transaction	Description
1	Bad server takes a long time to load, even though the network is good	0	1	0	0	The server experienced a delay in responding, so it was given a negative sentiment on the system aspect.
2	The application is good and useful, but the gojek driver often has problems	0	0	1	0	The application provides good benefits for users, but users do not get comfort because Gojek drivers often make mistakes. So that negative sentiment is given to the comfort aspect.
3	The driver is far away, even though there are many close ones, why should they be far away	1	0	1	0	The availability of distant vehicles makes users uncomfortable, therefore given sentiment negative to availability dan comfort aspects.
4	The application is slow, a lot of extra costs cost. Better competitor apps	0	1	1	1	The system experienced delays, often providing additional costs to transactions so that users felt uncomfortable. So that they were given a negative sentiment on the system, comfort and transactions.

2.3 Data Preprocessing

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

Data Preprocessing is the process of removing some data, which interferes with processing due to inconsistent data structures [12]. Data preprocessing consists of several steps, including case folding, data cleaning, word tokenization, word normalization, stop words, and stemming. Table 3 below shows the preprocessing process for sentences, along with an explanation for each process:

Table 3. Preprocessing a sentence

Process	Description
Case Folding	Change to lowercase all letters in a document.
Cleaning Data	Remove things that do not significantly affect the document, such as punctuation, repetitive words, *white spaces*, and words consisting of three letters.
Tokenizing	The process of separating sentences into word forms is called tokens.
Word Normalization	Converting a non-standard language into a standard language. For example, the word 'sedikitpun' becomes 'a little'
Stopword Removal	It is removing of words that do not affect the document or text, such as the words "from", "for", "to", "in".
Stemming	Removing suffixes is either a suffix or a prefix to return the word to its basic form.

2.4 Feature Extraction (Term Weighterd TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is a method that has an essential value for each word in the document [13]. The feature extraction method can give a value or weight to each word in the dataset. Term Frequency (TF) is a word frequency (t) that appears in the document (d) and is referred to as word frequency. In contrast, inverse document frequency (IDF) measures how often the word is used. The TF value will increase and the IDF value will decrease if the word appears with a high frequency [14]. The equation used for weight calculation with TF-IDF is below:

$$TF_{t,d} = f_{t,d} \tag{1}$$

In equation (3.1) is the tf formula used to express the number of frequencies of the occurrence of terms in the document, for the calculation of IDF is found in equation (2):

$$IDF_t = log\,(\frac{D}{df_t}) \tag{3}$$

Equation (3.2) is a formula for the calculation of IDF where D expresses the sum of all documents, while *df* is used to express the number of documents containing the term *j*. After the values *of tf* and *idf* are obtained, the weight value of a term expressed by W in equation (3.3) will be calculated:

$$W_{t,d} = tf_{t,d} \times idf_t \tag{3}$$

2.5 Feature Selection Chi-Square

Feature selection is used to eliminate features that are considered less relevant in the classification process. Chi-Square feature selection uses statistical theory to test the independence of a term (t) with its category or sentiment (c). Based on statistical theory, there are two events in the selection of Chi-Square features: the emergence of features and the emergence of categories [15]. Furthermore, each term value will be sorted from the highest based on equation (4):

$$x^2(t,c) = \frac{N*(AD-BC)^2}{(A+C)*(B+D)*(A+B)*(C+D)} \tag{4}$$

From the data above, N is the total number of all documents used. A is the number of terms (t) included in the class c document. B is the number of terms (t) that are not included in the document of class c. C is the number of terms that are not terms (t) that are included in class c. D is the number of terms that are not terms (t) and are not included in class c.

2.6 Maximum Entropy

Maximum Entropy is a classification method that uses entropy values or probabilities to classify data by looking for the highest entropy value [16]. Entropy is a probabilistic value or, in other words, an uncertain value. The probability, in this case, is taken based on the prediction results of the information contained in a review or sentence. The probability of information from a review will be calculated using maximum Entropy and then grouped into negative or positive sentiments. The general function of Entropy is found in equation (3.5):

$$Entropy(S) = -\sum_{i=1}^{n} P(S_i) \times log_2 P(S_i) \tag{5}$$

So, n is the number of values contained in sentiment attributes such as negative and positive. In P(S_I)states the sample probability for class i. Suppose with an example of a review "The application is good and useful, but Gojek

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

drivers are often problematic". Based on the review, several words can be extracted, such as "good", "useful", and "problematic drivers". Of the three words, 2 positive sentiments and 1 negative sentiment will be taken to calculate the probability value as a sentiment classification based on the features contained in the review.

In this study, the author defined that the result is a member of a set (y), and the condition that affects the result is defined as the context, where the context is a member of the set (x). The set (x) is a set of features used in this classification process. Furthermore, the output obtained by using the maximum Entropy is the probability of the set (y) to (x) represented by the function f(x,y) in the following equation (3.6):

$$f(x,y) = \begin{cases} 1, & jika\ y' = y\ dan\ cp(x) = benar \\ 0, & lainnya \end{cases} \tag{6}$$

From the equation above, the feature function will be valued at 1 if y' = y; other than that, it will be given a value of 0. In the training process, training data and features are used. Furthermore, the *conditional probability* or *cp(x)* for a state (y|x) is calculated in the following equation (3.7):

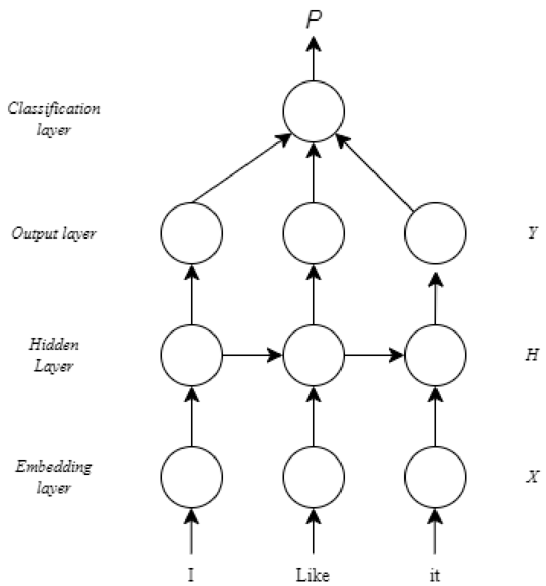
$$P(y|x) = \frac{\prod_i \alpha_i f_i(x,y)}{Z_\alpha(X)} \tag{7}$$

And $\prod_i \alpha_i f_i(x,y)$ is the value of a specific event in a document against a class, which is a parameter or weight value related to a feature and $\alpha i f i f i(x,y)$ is the probability of the term x contained in class y. While the value of Z is the normalized value of each word contained in the set (x) with the formula in the equation below (3.8):

$$Z_\alpha(x) = \sum_y \prod_i \alpha_i f_i(x,y) \tag{8}$$

2.7 Recurrent Neural Network

Recurrent Neural Network or RNN is one of the neural network architectures that can be used for sentiment analysis. RNN also process input sequences based on state, so they can be used to perform NLP tasks [17]. Not only that, RNN can also store a previous input that has sequence data using memory such as text, genome or time series. Because basically, the RNN algorithm will perform a process on each input of a given word with a certain time step, which then RNN will convert the input sequence into a vector of a fixed size [18]



Gambar 2. Recurrent Neural Network

From Figure 2, it shows that there are four layers in the RNN algorithm, namely the embedding layer, hidden layer, output layer, and classification layer. On layer X = {x1, x2, ... xn} is used as the embedding word of every word in a sentence. Layer H = {h1, h2, ... hn} is the iterative layer where each node is in H is a repeating unit and the nodes hj to jth are determined by the input vector xj and the last iteration is hj-1. Output layer or layer Y = {y1, y2.... yn} is the output of the RNN classification. The last layer, the Player, is the logistic classification layer, where the results of this layer can show positive sentiment on probability values. The value of each layer can be calculated using equations (3.9), (3.10), (3.11):

$$\text{h1} = \text{f}(Wx_j + Uh_1 + b_h) \tag{9}$$

$$y_1 = f(Vh_j + b_y) \tag{10}$$

$$p = \frac{1}{1+e^{-\theta Y}} \tag{11}$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

2.8 K-Fold Cross Validation Evaluation

K-Fold Cross Validation is an evaluation model used to test results and evaluate algorithms to get better results [19]. This research uses a k value of 10. In a way, it is divided into ten equal-sized parts, with ninety percent of the training data and ten percent of the test data divided. Figure 3 below shows the division of data in each fold:

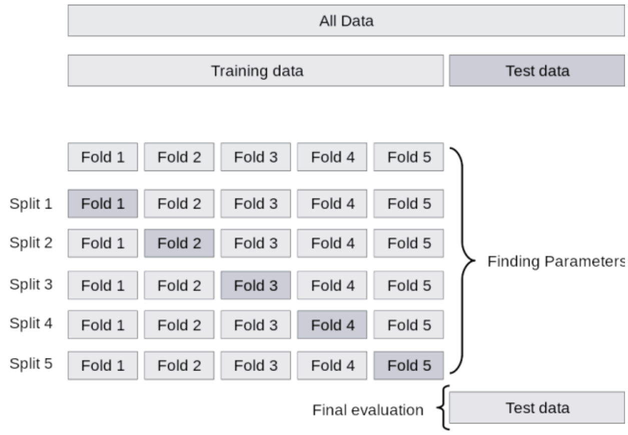


Figure 3. The illustration K-Fold Cross Validation

2.9 Performance Evaluation

To see the level of performance of the system that has been built, we need to go to the next stage: evaluating through measuring accuracy, precision, recall, and F1-Score values. Then, the results obtained from the prediction can be represented using the confusion matrix. The following formula can be used to calculate the value of performance evaluation

a. Accuracy

The accuracy value is the percentage of data that is predicted correctly compared to the entire amount of data.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{12}$$

b. Precision

Precision is the value obtained by comparing the positive values predicted correctly with all the predicted data.

$$precision = \frac{TP}{TP+FP} \tag{13}$$

c. Recall

Recall is the ratio of data that shows accurate positive predictions to all available positive data.

$$recall = \frac{TP}{TP+FN} \tag{14}$$

d. F1-Score

F1-Score is a performance metric that includes a portion of the recall and precision values.

$$F1 - Score = 2x\frac{precision\ x\ recall}{precision+recall} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{15}$$

Table 4. Performance Evaluation

	Actual Values	
Predicted Values	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

A confusion matrix is a method for representing predictions and actual conditions from data [20]. The matrix also has four values that are used as a reference in calculations, as stated in Table 4. Based on the confusion matrix data above, four values will be calculated including: First, when the prediction result and the actual class value are correct, it will be called True Positive (TP). Second, when the prediction result is wrong and the actual class value is wrong, it will be called True Negative (TN). Third, when the prediction result is correct and the actual class value is wrong, it will be called False Positive (FP). And finally, when the prediction result is bad and the actual class value is correct, it is called False Negative (FN).

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

3. RESULT AND DISCUSSION

At the evaluation stage, researchers conducted tests on the system built to determine the system's performance in each method used. Furthermore, in the classification stage, researchers used two methods, namely Maximum Entropy and Recurrent Neural Network (RNN), on reviews on the Gojek application using 4052 review data with data distributed based on aspects in Table 5.

Table 5. Data Distribution Based on Aspects

Aspects	Label		Total Amount
	Positive	Negative	
Availability	2802	1250	4052
System	2483	1569	
Comfort	2720	1332	
Transaction	2756	1295	

The dataset used has undergone a feature extraction process using TF-IDF, feature selection using Chi-Square, and model evaluation using K-Fold Cross Validation with 10 folds, which can be interpreted as 90% for training data and 10% for testing data. The results in this test take the accuracy and f1-score values in each classification, and a comparison is made between the two models used. Result and Analysis Result on Maximum Entropy Testing

In this research, classification is done through Maximum Entropy by testing the number of folds on four aspects. The results show that the first fold can provide the most accurate value for each aspect studied. For this classification, the author performed ten iterations on each aspect with a min_lldelta of 0.001 and used weighted to get the best value on F1-Score. Thus the research obtained the best value by the Availability Aspect value of 72.91% and F1-Score of 61.73% with an average accuracy value of 69.22% and F1-Score 56.73%.

Table 6. The Result of Maximum Entropy Test

K-Fold	Aspect							
	Availability		System		Comfort		Transaction	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
1	**72.91%**	**61.73%**	62.56%	48.15%	**70.20%**	**57.90%**	62.81%	48.46%
2	72.91%	61.48%	**65.27%**	**51.81%**	69.21%	56.62%	66.01%	52.75%
3	66.91%	53.65%	62.22%	47.73%	67.41%	54.28%	68.40%	56.01%
4	70.12%	58.06%	56.05%	40.26%	68.40%	55.56%	70.12%	57.81%
5	67.41%	54.28%	61.73%	47.63%	65.19%	51.45%	**71.85%**	**60.08%**
6	68.15%	55.24%	63.21%	48.96%	65.19%	51.45%	67.16%	53.97%
7	68.64%	56.13%	59.75%	44.70%	67.41%	54.28%	65.43%	51.76%
8	71.11%	59.11%	59.51%	44.40%	67.65%	54.60%	67.41%	55.16%
9	69.14%	56.52%	59.51%	44.40%	63.70%	49.58%	70.62%	58.46%
10	64.94%	51.13%	63.70%	49.58%	66.91%	53.65%	70.86%	59.03%
Average	**69.22%**	**56.73%**	**61.35%**	**46.76%**	**67.13%**	**53.94%**	**68.07%**	**55.35%**

3.1 Result and Analysis Result on Recurrent Neural Network

The best accuracy value in each aspect was found at different folds using the RNN classification model. The authors of this study used a simple RNN with three layers, 512, 256, and 128. They also used two layers with a dropout value of 0.5 to overcome overfitting on the dataset with a batch size of 64, epoch 50, used Adam as an optimizer, and iterated on kfold evaluation of 10. That way, the best accuracy results on transactions can be found with a value of 95.56% and F1-Score 92.50%. The average accuracy value is 94.0% and F1-Score is 90.45%.

Table 7. The Result of Recurrent Neural Network

K-Fold	Aspek							
	Availability		Sistem		Comfort		Transaksi	
	Akurasi	F1-Score	Akurasi	F1-Score	Akurasi	F1-Score	Akurasi	F1-Score
1	**93.10%**	**87.27%**	89.90%	86.20%	79.80%	64.96%	94.09%	91.89%
2	91.63%	83.96%	91.87%	88.58%	84.73%	72.07%	92.86%	89.30%
3	89.38%	83.00%	90.12%	86.21%	80.25%	64.60%	93.58%	89.43%
4	88.15%	78.38%	84.44%	81.42%	80.99%	69.80%	92.10%	86.44%

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

5	89.38%	82.73%	88.40%	84.49%	82.22%	70.97%	93.83%	89.27%
6	90.37%	83.82%	**91.85%**	**88.96%**	79.26%	64.41%	94.81%	91.95%
7	91.11%	85.94%	86.67%	82.35%	82.47%	68.72%	94.32%	91.39%
8	89.88%	81.78%	89.88%	86.98%	82.72%	71.31%	94.57%	91.13%
9	91.60%	86.18%	88.15%	84.81%	81.48%	70.59%	**95.56%**	**92.50%**
10	90.12%	85.40%	88.40%	83.62%	**83.95%**	**75.29%**	94.81%	91.21%
Average	**90.47%**	**83.85%**	**88.97%**	**85.36%**	**81.79%**	**69.27%**	**94.05%**	**90.45%**

3.2 The Comparison Result between Maximum Entropy and RNN

Based on the results of the data experiments above, it can be proven that RNN classification has the highest accuracy value and F1-Score. The highest average accuracy results are obtained in the transaction aspect with an accuracy value of 94.05% and F1-Score of 90.45%.

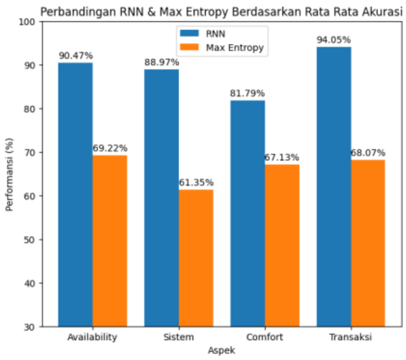


Figure 4. The Comparison Result Based on the Accuracy

Based on the illustration in Figure 5 below, we can see the evaluation results of the RNN & Maximum Entropy classification comparison based on the f1-score value with Chi Square weighting. In this evaluation, the f1-score value is obtained better when using the RNN classification with a value of 90.45%. The transaction aspect with a significant difference in F1-Score results against RNN classification with the best value of 56.73% is in the Availability aspect.

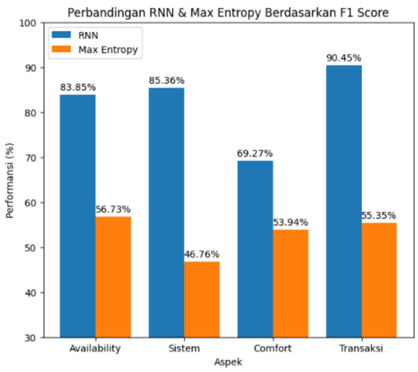


Figure 5. The Comparison Result Based on F1-Score

In this research, Recurrent Neural Network is known to be able to handle a constraint well even with a limited amount of data. However, this result is inversely proportional to the Maximum Entropy model which is required to have a sufficient amount of data for each class in the training dataset because this model is highly dependent on proper feature representation. Thus, when the RNN model can provide better results, this can also be due to its ability to capture the context in a text. This can be proven by the performance evaluation, which gets quite balanced results compared to the maximum entropy classification.

4. CONCLUSION

Based on the results of the above research, it can conclude that the R and Maximum Entropy methods for sentiment analysis of Gojek applications, which include aspects of availability, system, comfort, and transactions, obtained the results that in the use of the TF-IDF feature extraction process and Chi-Square feature selection, the performance for each method is 90: 10 in each division of training and training data. In the Recurrent Neural Network (RNN) classification, the average transaction accuracy value is 94.05 percent, and the F1-Score value is 90.45 percent. This research also shows that the deep learning RNN method using chi-square for feature weighting and TF-IDF for feature

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1083−1091
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5767

extraction can produce higher accuracy values on transaction aspects. The two methods can provide a significant difference in value. However, when using classifications maximum Entropy or Recurrent Neural Network classification, both classifications have advantages and disadvantages, depending on the specific characteristics of the data and the problem at hand. Researchers suggest that future research should consider using a larger number of datasets when conducting research, either on Maximum Entropy classification or Recurrent Neural Network classification. Further research is also expected to use various feature selections or other feature extractions to provide improvements in research.

REFERENCES

[1] Santika Erlina F, "Aplikasi Transportasi Online Terbanyak Diunduh di RI 2023, Gojek Juaranya," 2023, DATABOKS.

[2] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujm

[3] H. G. Sulistio and A. Handojo, "Aspect-Based Sentiment Analysis pada Ulasan ECommerce dengan Metode Support Vector Machine untuk Mendapatkan Informasi Sentimen dari Beberapa Aspek," *J. Infra*, vol. 10, no. 2, pp. 1–5, 2022.

[4] E. Dwianto and M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 1, p. 94, 2021, doi: 10.22441/format.2021.v10.i1.009.

[5] D. J. M. Pasaribu, K. Kusrini, and S. Sudarmawan, "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan Bert Embedding," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, 2020, doi: 10.35585/inspir.v10i1.2568.

[6] S. Fan, J. Yao, Y. Sun, and Y. Zhan, "A summary of aspect-based sentiment analysis," *J. Phys. Conf. Ser.*, vol. 1624, no. 2, 2020, doi: 10.1088/1742-6596/1624/2/022051.

[7] H. Chyntia Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[8] M. F. Putra, A. Herdiani, and D. Puspandari, "Analisis Pengaruh Normalisasi, TF-IDF, Pemilihan Feature-set Terhadap Klasifikasi Sentimen Menggunakan Maximum Entropy (Studi Kasus : Grab dan Gojek)," *e-Proceeding Eng.*, pp. 8520–8529, 2019.

[9] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.

[10] Novirianto, "Analisis Sentimen Berbasis Aspek Pada UlasanAplikasi Mybluebird Dengan Implementasi N-GramDan Algoritma Logistic Regression," 2023, [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/76656/1/IQBAL FARIZ NOVIRIANTO-FST.pdf

[11] N. K. Shadrina, E. Sutoyo, and V. P. Widartha, "Sentiment Analysis in Reviews About Beaches in Bali on Tripadvisor Using Recurrent Neural Network (RNN)," *Proc. - 2021 IEEE 7th Inf. Technol. Int. Semin. ITIS 2021*, 2021, doi: 10.1109/ITIS53497.2021.9791501.

[12] D. Atika, A. Ari Aldino, S. Informasi, J. Pagar Alam No, L. Ratu, and K. Kedaton, "Term Frequency-Inverse Document Frequency Support Vector Machine Untuk Analisis Sentimen Opini Masyarakat Terhadap Tekanan Mental Pada Media Sosial Twitter," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 4, p. page-page, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[13] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.

[14] F. M. Rizky, J. Jondri, and K. M. Lhaksmana, "Twitter Sentiment Analysis of Kanjuruhan Disaster using Word2Vec and Support Vector Machine," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 219–227, 2023, doi: 10.47065/bits.v5i1.3612.

[15] S. Anisah, A. S. Honggowibowo, and A. Pujiastuti, "Klasifikasi Teks Menggunakan Chi Square Feature Selection Untuk Menentukan Komik Berdasarkan Periode, Materi Dan Fisikdengan Algoritma Naivebayes," *Compiler*, vol. 5, no. 2, pp. 59–66, 2016, doi: 10.28989/compiler.v5i2.171.

[16] F. A. Tamardina, H. Yasin, and D. Ispriyanti, "Analisis Sentimen Review Aplikasi Cryptocurrency Menggunakan Algoritma Maximum Entropy Dengan Metode Pembobotan Tf, Tf-Idf Dan Binary," *J. Gaussian*, vol. 11, no. 1, pp. 1–10, 2022, doi: 10.14710/j.gauss.v11i1.34004.

[17] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2098–2117, 2022, doi: 10.1016/j.jksuci.2022.02.025.

[18] M. K. Hasan and E. B. Setiawan, "Sentiment Analysis of Twitter Data on Bank Central Asia Stocks (BBCA) Using RNN and CNN Model with GloVe Feature Expansion," *Proceeding - COMNETSAT 2023 IEEE Int. Conf. Commun. Networks Satell.*, no. November 2023, pp. 195–200, 2023, doi: 10.1109/COMNETSAT59769.2023.10420731.

[19] K. S. Putri, I. R. Setiawan, and A. Pambudi, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier," *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.

[20] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.