Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

Multi-Aspect Sentiment Analysis Using Elman Recurrent Neural Network (ERNN) Method for TripAdvisor App User Reviews

Fahrul Raykhan Ridho*, Yuliant Sibaroni, Dyas Puspandari

Fakultas Informatika, Universitas Telkom, Bandung, Indonesia
Email: fahrulrehan@student.telkomuniversity.ac.id, yuliant@telkomuniversity.ac.id, diyaspuspandari@telkomuniversity.ac.id
Correspondence Author Email: fahrulrehan@student.telkomuniversity.ac.id
Submitted: **07/08/2024**; Accepted: **10/09/2024**; Published**: 12/09/2024**

Abstract−TripAdvisor is the world's largest travel platform, and it assists 463 million travelers each month in making their trips the best they can be. Users of TripAdvisor can provide reviews, comments, and ratings of travel destinations. However, reviews on TripAdvisor are considered insufficient in helping prospective travelers understand the strengths and weaknesses of a hotel. Therefore, a multi-aspect sentiment analysis of TripAdvisor reviews on hotels was conducted to identify commonly discussed rating aspects among visitors and to determine specific evaluations. In this study, the Elman Recurrent Neural Network (ERNN) method was employed to build a classification system for multi-aspect sentiment analysis of user reviews on the TripAdvisor application. The aspects examined in this research include Service, Cleanliness, Location, Value, Rooms, and Overall Experience, aiming to provide insights into the hotels under consideration. The results indicate that the ERNN method can deliver superior outcomes in multi-aspect sentiment analysis of TripAdvisor hotel reviews. The ERNN model's performance in multi-aspect sentiment analysis shows optimal accuracies: 81.35% for Service aspect, 98.71% for Cleanliness aspect, 74.87% for Location aspect, 93.84% for Value aspect, and 71.52% for Rooms aspect. These findings can assist travelers in better understanding the strengths and weaknesses of accommodations.

Keywords: Multiaspect sentiment, TripAdvisor, Recurrent Neural Network (RNN), and Elman Recurrent Neural Network (ERNN)

1. INTRODUCTION

TripAdvisor is the world's largest travel platform, serving as a tool to help users determine their travel destinations or accommodations. This app has assisted 463 million travelers each month in making their trips the best they can be. Travelers worldwide use TripAdvisor's website and app to browse over 859 million reviews and opinions on 8.6 million accommodations, restaurants, experiences, airlines, and cruises. Whether planning or already on a trip, TripAdvisor is used to compare prices, book tours, and find popular attractions [1]. However, ratings on TripAdvisor are considered insufficient in helping potential travelers understand a hotel's strengths and weaknesses, as reviews often contain various aspects of assessment and mixed sentiments (both positive and negative) [2]. Therefore, a multi-aspect sentiment analysis of TripAdvisor reviews is necessary to identify the aspects frequently discussed by visitors and to determine specific evaluations of the chosen hotel. The platform's pervasive usage highlights its pivotal role within the travel industry, where travelers depend on it to assess prices, arrange tours, and identify popular attractions. Despite its extensive user base and the vast quantity of user-generated content, the current rating system on TripAdvisor is often inadequate for providing potential travelers with a comprehensive understanding of a hotel's strengths and weaknesses. This limitation arises because reviews typically encompass a multitude of aspects and sentiments, both positive and negative, which makes it challenging for users to derive clear insights from the overall ratings alone.

Multi-aspect sentiment analysis and deep learning models have emerged as highly effective tools for classifying and identifying sentiments across a range of different aspects. Nowadays, public opinion has become an important source in someone's decision-making regarding a product. Multi-aspect Sentiment Analysis aims to understand opinions, emotions, and events expressed by someone in their reviews. This analysis is essential to identify aspects of a given entity and determine the sentiment expressed in each aspect [3]. The classification and identification process of each aspect in multi-aspect sentiment analysis can be done using Artificial Intelligence (AI) models such as deep learning. Deep learning, as a subfield of machine learning, uses artificial neural networks consisting of layers of artificial neurons to process and transform input data into desired output [4]. With the backpropagation algorithm, these neural networks can learn hierarchically from complex and large-scale data, such as review texts, to extract relevant features [5]. Among the many deep learning methods available, Recurrent Neural Network (RNN) is a more suitable choice for multi-aspect sentiment analysis due to its ability to handle sequential data and maintain context from previous data. RNNs can remember previous information in a data sequence, which is very useful in understanding context and nuances in review texts [6]. One type of RNN that is very effective in these tasks is the Elman Recurrent Neural Network (ERNN). ERNN has a context layer that stores information from the previous time step and uses this information to influence the current output [7].

This research is supported by a series of case studies conducted previously with the objective of exploring a variety of methods in the field of multi-faceted sentiment analysis. In a related study, Liu et al. (2020) employed Aspect-Based Sentiment Analysis (ABSA) on Amazon product reviews, achieving an accuracy rate of 87.5%. In a similar vein, Wang et al. (2019) employed the attention-based LSTM approach to analyze restaurant reviews on Yelp, achieving an accuracy rate of 89.3%. In a similar vein, Kim et al. (2018) employed a convolutional neural network

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

(CNN) for the analysis of hotel reviews, achieving an accuracy rate of 91%. In the context of multi-faceted sentiment analysis, Zhang et al. (2021) employed ERNN to analyze movie reviews on IMDB, achieving 93% accuracy. This method proved effective for understanding the context and nuances present in the review text.

Previous research in sentiment analysis has largely focused on single-aspect evaluations or has utilized methods that do not fully exploit the sequential characteristics of review data. This study seeks to address these gaps by implementing a more sophisticated approach, leveraging the ERNN's strengths in handling sequential data and maintaining contextual integrity. The research aims to contribute to the field by offering a more nuanced and accurate method for sentiment analysis, which could significantly enhance the utility of online reviews for travelers.

This study proposes the use of the Elman Recurrent Neural Network (ERNN) for multi-aspect sentiment analysis of user reviews on the TripAdvisor platform. The choice of ERNN is motivated by its capability to capture the sequential nature of review texts and to conduct supervised training through the backpropagation algorithm, which enhances the accuracy and detail of sentiment classification. By employing ERNN, this research aims to achieve a more precise multi-aspect sentiment analysis of TripAdvisor reviews, thereby providing travelers with a clearer understanding of a hotel's various aspects, such as Service, Cleanliness, Location, Value (hotel value for fee), Rooms (room quality), and Overall Experience. It is hoped that these aspects will assist visitors in understanding the assessment of a hotel based on reviews provided by other users.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The workflow in this research begins with dataset collection based on the rules that have been established and explained in the previous chapter. The work steps in this research are presented in Figure 1.

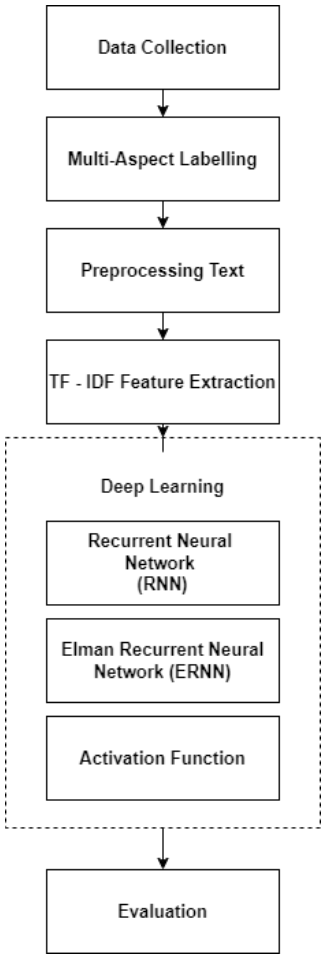

Figure 1. Research Stages

The data collection process employs the use of existing datasets, which are accessible on Kaggle.com. These datasets undergo a visualization and labeling process to represent the data, identify patterns, and recognize trends through text processing. Following this, feature extraction is conducted using the TF-IDF method. This method determines the value and frequency of words or texts within multiple reviews. The TF-IDF method is a valuable tool for assessing the relevance of text in user-generated reviews.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

Following the extraction phase, the data is subjected to deep learning processing by a computer. In this study, this is achieved through the use of RRN and ERNN, which are employed for the classification of multi-aspect sentiment analysis of TripAdvisor application user reviews. The final stage of the process is to assess the efficacy of the methodology employed through the evaluation of a randomly selected sample of reviews.

2.2 Data Collection

In this research, the data used is a public dataset obtained from the website kaggle.com with the link https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews [15]. This dataset was obtained by crawling the TripAdvisor application regarding hotel reviews. The dataset used consists of a total of 12,813 hotel review data. The visualization of the dataset used is presented in Table 1.

Table 1. Dataset Visualization

Review	Rating
recommend hotel Did reviewers actually stay hotel did, it good thing the hotel location is really close to Leidseplein; the shared facilities were filthy, the toilet floor was cleaned for a month, the facilities were not cleaned 3 days, and there disgusting staff…	1
barcelona rocks, stayed hotel jazz girlfriend 3 nights end august.the hotel excellent location carrer pelai, close placa catalunya ramblas appreciate buzz city removed respite mayhem crowds, caught airport bus barcelona costs 7 euros person…	4
ok hotel good location stayed night way beijing rawa island, hotel service room ok. location great shopping restaurants, probably stay, opinion nice 3 star hotel...	3
great service nice pool ok beach lovely grounds small rooms stayed 5 days 4 nights 1st 5th, quick cab ride hotel, no line checkin requested king bed guaranteed offered doubles b club level floor, wasnt thrilled went, grounds hotel beautiful,…	4
surprising treat spent weekend july 15/16 2006 cartwright hotel based purely recommendations read site, actually expecting like small older european hotel, cartwright amazing, small elegant pleasant staff knowlegable city, room…	5

In examining the hotel reviews provided, it becomes clear that travelers have varying experiences based on the quality of service, location, and overall cleanliness of the accommodations. The ratings, ranging from 1 to 5, reflect the satisfaction level of each reviewer.

1. Rating 1 is the lowest rating in the table reflects a deeply unsatisfactory experience.
2. Rating 4 is a positive review, reflects satisfication towards experience
3. Rating provides a more neutral perspective
4. Rating 5 is the most positive that reviewer thoroughly satisfied.

2.3 Aspect Labelling

Aspect labeling is the process of assigning labels to hotel reviews based on specific aspects to be analyzed [18]. In this study, the aspects to be analyzed are Service, Cleanliness, Location, Value (hotel value for money), and Rooms (room quality). Label determination is done based on the presence of relevant keywords in each review to determine whether the aspect is rated as good (Good), bad (Poor), or neutral (Neutral). The purpose of aspect labeling is to facilitate multi-aspect sentiment analysis of user reviews in specific research or applications. In this study, the aspect labeling process will be done manually by examining each review that has been provided beforehand. After the Aspect Labeling process, a new dataset is generated with various predetermined aspects (Service, Cleanliness, Location, Value, Rooms, and Overall Experience) along with their respective values (Good, Poor, or Neutral). For the labeling process, star ratings are not used because using stars only allows for a general assessment of whether a review is positive or negative, without providing insights into specific aspects. Therefore, the labeling process involves keyword-based searches to assess the various predetermined aspects. The visualization of the data resulting from the Aspect Labeling process is presented in Table 2.

Table 2. Aspect Labeling Result Dataset

Review	Rating	Service	Cleanliness	Location	Value	Rooms
recommend hotel di d reviewers actually stay hotel did, good …	1	-	-	1	-	-
barcelona rocks, sta yed hotel jazz girlfr iend 3 nights end au gust…	4	-	-	1	-	-

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

ok hotel good location stayed night way beijing rawa island …	3	1	-	1	-	-
great service nice pool ok beach lovely grounds small rooms…	4	1	-	-	-	1
dazed confused having read reviews tripadvisor hoping nice budget…	2	-	-	-	-	0

Table 2 shows the results of the Aspect Labeling process. As seen in Table 2, based on the given reviews, they are processed and divided into several aspects with predetermined values. Thus, from each predetermined aspect, an analysis process can be carried out using a deep learning model. It can be seen in Table 2 that the data in the review column is data that has been preprocessed previously, so the words in the reviews are simpler and more concise. This also has an impact on the built model, resulting in an optimal and efficient training process.

2.3 Research Stages

Data Preprocessing is a series of steps to prepare unstructured text data [16]. Therefore, a process is needed to transform unstructured data into structured data. This process is known as text processing. In this stage, a data cleaning process will be carried out with several steps, namely case folding, tokenizing, stopwords, normalization, and stemming, to obtain an accurate dataset before proceeding to the next stage [17]. The following are the stages in text processing, among which are the following :

a. Case Folding
Case folding is a process of changing uppercase letters to lowercase letters and removing punctuation marks. Doing case folding helps to ensure all the words are uniform regardless of their case. It is shown in Table 3 that the case folding result will improve in search of search engine algorithms by eliminating case sensitivity. Also, it will reduce the redundancy of new text in data storage.

Table 3. Words of the results of the case folding process

Original Text	Case Folding
"Hated," "inn", "terrible", "Room-service", "a" "horrible", "Staff", "un-welcoming", "décor", "recently", "updated", "lacks", "complete", "looke", "management", "staff", "horribles"	“hated”, “inn”, “terrible”, “room-service”, “a” “horrible”, “staff”, “un-welcoming”, “décor”, “recently”, “updated”, “lacks”, “complete”, “looke”, “management”, “staff”, “horribles”

b. Tokenizing
Tokenizing is a process of cutting sentences or paragraphs in a document into words called "tokens," making it easier to distinguish a specific characters.

Table 4. Tokenizing Process Results

Original Text	Tokenizing
Hated inn terrible, Room-service a horrible Staff un-welcoming, decor recently updated lacks complete looke, managment staff horribles	['Hated', 'inn', 'terrible', 'Room-service', 'horrible', 'Staff', 'un-welcoming', 'decor', 'recently', 'updated', 'lacks', 'complete', 'looke', 'managment', 'staff', 'horribles']

Table 4 presents a word-by-word breakdown of the text, with each word separated by a space, in order to facilitate text analysis and anabling frequency analysis for machine learning models that require text data to be in structured format.

c. Stopwords
Stopwords is the process of removing connecting words that do not have meaning such as "in", "its", "and", "which", and so on. In this research, the process of removing stopwords will use a library from the Corpus Natural Language Toolkit called Stopwords. This Stopwords library provides a list of common words in various languages that often do not add value in text analysis, so these words can be removed to improve the quality of the analysis results. The example of stopword could be seen in the Table 5.

Table 5. Stopwords Data Table

Stopwords

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

“a”, “an”, “and”, “are”, “as”, “at”, “be”, “by”, “for”, “from”, “has”, “in”, “is”, “it”, “its”, “of”, “on”, “that”, “the”, “to”

d. Normalization
Normalization is the process of changing words that are not used according to their standard form into standard words, such as changing 'whats' to 'what' and so on.

Table 6. Normalization Process Results

Original Text	Normalization
“looke”, “horribles”, “running”, “plays”	“look”, “horrible”, “run”, “play”

To summarize, the normalization process, as illustrated in Table 6, is employed to correct writing errors or to convert slang into standard language that will understand by machine.

e. Stemming
Stemming is the process of finding the root word by removing affixes and then applying word normalization. In this research, the stemming process will utilize the PorterStemmer library from the Natural Language Toolkit (NLTK) as illustrated in table 7..

Table 7. Stemming Process Results

Original Text	Normalization
"terrible", "Room-service", "horrible", "un-welcoming", "recently", "updated", "lacks", "complete", "looke", "management", "horribles."	“terribl”, “Room- servic”, “horribl”, “un- welcom”, “recent”, “updat”, “lack”, “complet”, “look”, “manag”, “horribl”

2.4 TF - IDF Feature Extraction

The research data to be used is textual and unstructured, but computer programming languages can only process structured data in the form of tables. Therefore, the unstructured textual data needs to be converted into numbers or vectors, so a conversion method is needed. TF-IDF is an algorithm often used in textual data processing, where this algorithm will convert textual data into a vector space model or, in other words, convert documents into vector values [19]. In this research, the researcher uses TF-IDF to calculate word weights and achieve good results using the formula:

$$w = tf\ x\ idf = tf\ x\ \frac{1}{idf} \quad (1)$$

Description:
The frequency of word T in document D is TF, used to calculate the word's ability to describe the document. IDF represents the frequency of document D containing word T in the corpus, used to calculate the word's ability to distinguish the document. If the frequency of a word is high in its own document but low in other documents, this word has a strong ability to distinguish from other documents and is assigned a high weight.

2.5 Recurrent Neural Network (RNN)

After the data cleaning stage, the next stage is RNN as a converter of independent activation into dependent activation by giving the same weights and biases to all layers and categorizing them into three layers, namely the input layer, hidden layer, and output layer [20]. After that, the weights, input types, number of hidden neurons, learning rate, and momentum factor are initialized. Then, the weights and input-output that have been applied to the activation function will be determined [21]. For the visualization of the RNN model, it is given in Figure 2.

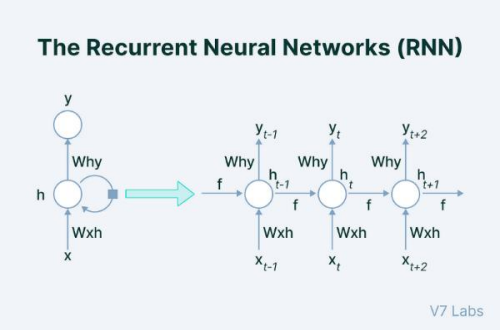

Figure 2. RNN Architecture

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

In the RNN architecture in Figure 2, a node or neuron represents data. All data are connected to each other and perform flow from input to hidden layers back to the start of the recurrence and last to the output. Each data text has weight to provide insight into learning the patterns.

2.6 Elman Recurrent Neural Network (ERNN)

Elman Recurrent Neural Network (ERNN) is used in this research for sentiment analysis of Traveloka application reviews, utilizing a contextual layer that stores information from the hidden layer at the previous time step, enabling the recognition of temporal patterns in text data. Text data is processed into vectors through TF-IDF and passed through the input layer, contextual layer, and hidden layer, which combine information to determine neuron activation. The output layer produces sentiment classification (positive, negative, or neutral) of user reviews. The ERNN training process involves adjusting weights and biases through the backpropagation through time (BPTT) algorithm, using activation functions like sigmoid or tanh to handle data non-linearity [22]. By considering temporal context, ERNN is expected to improve sentiment analysis accuracy compared to other methods that do not consider this context. The visualization of the ERNN process is shown in Figure 3.

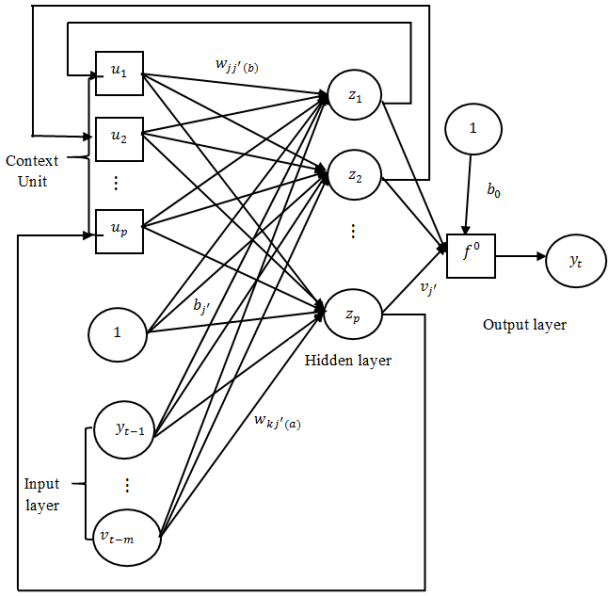

Figure 2. ERNN Process Visualization

Figure 3 illustrates the visualization of the ERNN process. As seen, the Elman Recurrent Neural Network (ERNN) model consists of several layers: the input layer, hidden layer, context layer, and output layer. In this study, two variations of the ERNN model are built. Model 1 consists of one 'SimpleRNN' layer with 128 units that receive input and produce an output of size (None, 1, 128). This output is then processed through a Dropout layer to prevent overfitting, followed by a 'Flatten' layer that transforms the output into a 1D vector of size (None, 128). After that, there is a 'Dense' layer with 64 units and ReLU activation, followed by another Dropout layer, and finally, a 'Dense' layer with two units for two-class classification. This model has a total of 2,084,546 trainable parameters. Model 2 is a more complex variation with two 'SimpleRNN' layers. The first layer has 128 units, producing an output of size (None, 1, 128), followed by the second 'SimpleRNN' layer with 256 units, producing an output of size (None, 256). Next, there are Dropout and 'Flatten' layers followed by two consecutive 'Dense' layers with 64 and 32 units, each followed by a Dropout layer. This model ends with a final 'Dense' layer with two units for two-class classification. This model has a total of 2,193,314 trainable parameters. With the addition of RNN and Dense layers, Model 2 is expected to capture more complex patterns compared to Model 1. The layers used in each ERNN model in this research are further detailed in Table 8.

Table 8. Stemming Process Results

Model	**Layer**		
Model 1 ERNN	**Layer (type)**	**Output Shape**	**Param #**
	simple_rnn (SimpleRNN)	(None, 1, 128)	2076160
	dropout (Dropout)	(None, 1, 128)	0
	flatten (Flatten)	(None, 128)	0
	dense (Dense)	(None, 64)	8256
	dropout_1 (Dropout)	(None, 64)	0
	dense_1 (Dense)	(None, 2)	130
	Layer (type)	**Output Shape**	**Param #**

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

Model 2 ERNN	simple_rnn_16 (SimpleRNN)	(None, 1, 128)	2076160
	simple_rnn_17 (SimpleRNN)	(None, 256)	98560
	dropout_19 (Dropout)	(None, 256)	0
	flatten_9 (Flatten)	(None, 256)	0
	dense_14 (Dense)	(None, 64)	16448
	dropout_20 (Dropout)	(None, 64)	0
	dense_15 (Dense)	(None, 32)	2080
	dropout_21 (Dropout)	(None, 32)	0
	dense_16 (Dense)	(None, 2)	66

Table 8 shows the layers used for the multi-aspect classification model on hotel reviews in the TripAdvisor application. The ERNN Model 1 uses an architecture consisting of several layers. First, a 'SimpleRNN' layer with 128 units is used to extract features from the sequential input of hotel review data. This layer has 2,076,160 adjustable parameters, allowing the model to learn complex representations from the text data. The 'Dropout' layer after 'SimpleRNN' is used to reduce overfitting by randomly ignoring some units during the training process. Then, the 'Flatten' layer transforms the output from 'SimpleRNN' into a 1D vector, preparing the data for input into the 'Dense' layer. The 'Dense' layer with 64 units applies an activation function to produce a more abstract representation of the previously extracted features. The second 'Dropout' is used again after the 'Dense' layer to reduce overfitting. Finally, the last 'Dense' layer with two units and the softmax activation function is used to produce the classification output, which represents the positive or negative sentiment of the hotel review.

ERNN Model 2, on the other hand, is more complex with the addition of a second 'SimpleRNN' layer with 256 units. This aims to increase the model's capacity to capture more intricate patterns from the text data. 'Dropout' layers are used after both 'SimpleRNN' layers to reduce overfitting. Afterward, the 'Flatten' layer transforms the output from both 'SimpleRNN' layers into 1D vectors. The first 'Dense' layer with 64 units and 'Dropout' are used to process the previously extracted features. The second 'Dense' layer with 32 units and another 'Dropout' is used to further refine the feature representation. Finally, the last 'Dense' layer with two units and softmax is used to produce the final prediction regarding the positive or negative sentiment of the hotel review.

2.6 Performance Measurement

After the model testing process, performance evaluation is conducted using the accuracy metric. The confusion matrix is an important evaluation tool in machine learning to measure the performance of classification models. It consists of four main components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [23]. TP indicates the number of instances correctly predicted as positive, TN indicates the number of instances correctly predicted as negative, FP is the number of instances incorrectly predicted as positive, and FN is the number of instances incorrectly predicted as negative. By using the confusion matrix, we can calculate precision, recall, and F1 scores. Precision is the proportion of correct positive predictions from the total positive predictions. Recall, or sensitivity, measures how well the model detects positive cases. F1-score is the harmonic mean of precision and recall, providing a balance when there is an imbalance between positive and negative classes.

$$\text{Presisi} = \frac{TP}{(TP + FP)} \tag{1}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{2}$$

$$\text{F1} - \text{score} = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{3}$$

3. RESULT AND DISCUSSION

In this study, a multi-aspect sentiment analysis will be conducted on hotel reviews in the TripAdvisor application using a deep learning method, namely the Elman Recurrent Neural Network. The data used are user reviews from the TripAdvisor application, consisting of a total of 20,492 review data. Based on this data, assessments will be made on various aspects, namely Service, Cleanliness, Location, Value (hotel value for money), and Rooms (room quality), with ratings of good (Positive), poor (Negative), or neutral (Neutral). Thus, the assessment of these aspects is expected to help provide detailed hotel recommendations to users. The distribution of data used in this study is presented in Table 9.

Table 9. Distribution of Aspect Data

Aspect	**Value**		
	Positive	**Negative**	**Neutral**
Service	2745	437	406
Cleanliness	115	22	15

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

Location	4004	367	413
Value	801	76	75
Rooms	3510	883	600

Table 9 shows the number of data distributions for each aspect assessed in this study. From Table 9, it can be seen that for various aspects of hotel reviews on TripAdvisor, most reviews for the aspects of Service, Cleanliness, Location, and Value tend to be neutral, indicating that many customers do not have strong opinions about these aspects. However, the Rooms and Overall Experience section has more positive reviews than negative ones, although there is also a significant number of negative reviews. This shows that although many customers have positive experiences, there are still some customers who experience negative things, especially related to the aspects of Rooms and Overall Experience. This indicates that these aspects have a significant impact on customer perception and may require further improvement to increase overall customer satisfaction. This table presents an analysis of five main aspects evaluated, namely Service, Cleanliness, Location, Value, and Rooms, with each aspect assessed in three categories: Positive, Negative, and Neutral. The service aspect has the second-highest number of positive reviews after location, with 2745 positive reviews, but it also has a significant number of negative reviews, namely 437 reviews. Cleanliness has the fewest reviews among all aspects, with 115 positive reviews, 22 negative reviews, and 15 neutral reviews, indicating that this aspect may be less of a concern for users. Location is the aspect with the most positive reviews, with 4004 reviews, but still has a considerable number of neutral and negative reviews, 413 and 367, respectively. The Value aspect shows 801 positive reviews, 76 negative reviews, and 75 neutral reviews, indicating that this aspect is generally considered positive even though there are a small number of complaints. Lastly, Rooms has 3510 positive reviews but also shows the second-highest number of negative reviews after Service, namely 883 reviews and 600 neutral reviews, indicating a significant difference of opinion among users regarding room quality. It can also be seen from the results obtained that the available data is limited because the labeling process is done manually, so each review is checked one by one, and there are some data that do not discuss these aspects. The visualization of the complete data distribution is given in Figure 5.

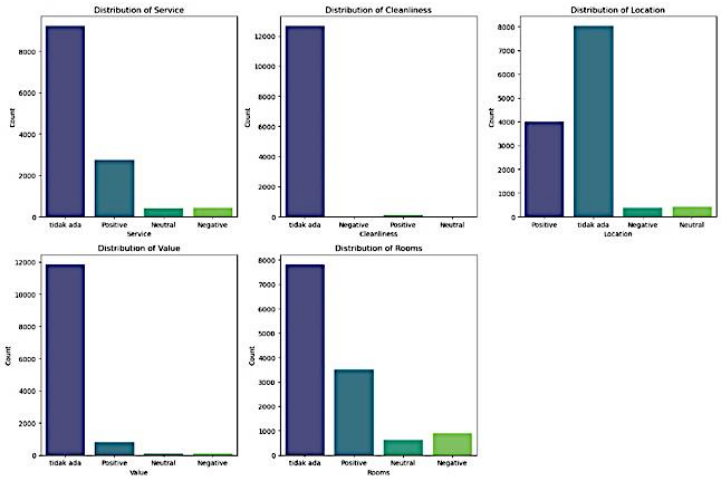

Figure 4. Data Distribution

The models built in this research were trained using an epoch or training iteration value of 20 for each aspect, utilizing 20% of the training data as validation data to assess the model's validation performance before the testing process. The models also employed a callback value called Early Stopping Callback, which halts the training process prematurely if there is no improvement in the validation loss (monitor='val_loss') over several epochs (patience=3). In the early stopping callback, the value 'restore_best_weights=True' is used, meaning that the model weights will be restored to the best weights saved during training upon completion. This helps prevent overfitting on the validation data. Therefore, with this approach, the built models are expected to learn data patterns optimally and efficiently. After the model construction, the training and testing process is conducted using the previously processed dataset and the Elman Recurrent Neural Network model. The accuracy performance results of each model are presented in Table 10.

Table 10. ERNN Model Testing Accuracy

Model	**Aspect**				
	Service	**Cleanliness**	**Location**	**Value**	**Rooms**
Model 1 ERNN	81.35	98.71	78.15	93.60	74.29
Model 2 ERNN	82.78	98.71	74.87	93.84	71.52

Table 10 shows the performance of two ERNN (Elman Recurrent Neural Network) models in performing multi-aspect labeling of hotel reviews on TripAdvisor, focusing on five aspects: Service, Cleanliness, Location, Value,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

and Rooms. In the Service aspect, Model 2 ERNN excels with an accuracy of 82.78, slightly higher than Model 1 ERNN with an accuracy of 81.35, indicating that Model 2 ERNN is more effective in capturing the nuances of service reviewed by users. For the Cleanliness aspect, both models show identical performance with a very high accuracy of 98.71, indicating that both Model 1 and Model 2 ERNN are capable of detecting cleanliness reviews very well and with almost no errors. However, in the Location aspect, Model 1 ERNN outperforms with an accuracy of 78.15 compared to Model 2 ERNN with an accuracy of 74.87, indicating that Model 1 ERNN is more accurate in identifying reviews related to hotel location. The Value aspect shows almost balanced performance between the two models, with Model 2 ERNN slightly superior with an accuracy of 93.84 compared to Model 1 ERNN with an accuracy of 93.60, indicating that both models are capable of capturing the value perceived by users well, although Model 2 is slightly more accurate. Lastly, in the Rooms aspect, Model 1 ERNN has an advantage with an accuracy of 74.29 compared to Model 2 ERNN with an accuracy of 71.52, showing that Model 1 is more effective in detecting reviews related to room quality. Overall, although both models show strong performance in various aspects, Model 1 ERNN tends to be superior in the Location and Rooms aspects, while Model 2 ERNN is better in the Service and Value aspects, with equal performance in the Cleanliness aspect. This analysis provides deep insights into the relative strengths and weaknesses of the two ERNN models in handling multi-aspect reviews, with Model 1 ERNN overall superior in the Location and Rooms aspects, while Model 2 ERNN excels in the Service and Value aspects. For a visualization of the accuracy comparison in graphical form, it is provided in Figure 5.

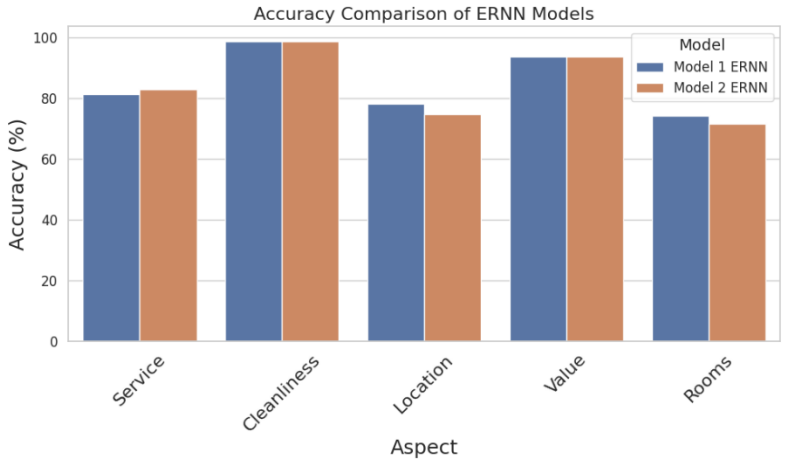

Figure 5. Visualization of Model Accuracy Results in Each Aspect

After conducting the analysis by examining the model's accuracy, the performance calculation process can be carried out by observing the values of the used evaluation metrics. In this research, the analysis process will utilize values from the confusion matrix, namely precision, recall, and F1-score. Precision is the ratio between the number of correct positive predictions and the total number of positive predictions given by the model. It measures the accuracy of the model's positive predictions. Recall is the ratio between the number of correct positive predictions and the total number of actual positive samples in the data. It measures the model's ability to find all positive samples. F1-score is the harmonic mean of precision and recall, providing a balanced view of the model's performance, especially when there is an imbalance between the number of positive and negative samples. It is useful for evaluating models in situations where precision and recall have equal importance. The precision, recall, and F1-score values after testing using ERNN model 1 are presented in Table 11.

Table 11. Precision, Recall, and F1-Score Results on ERNN Model 1

Model	Testing	Aspect				
		Service	**Cleanliness**	**Location**	**Value**	**Rooms**
Model 1 ERNN	Precision	75%	97%	73%	92%	69%
	Recall	81%	99%	78%	94%	74%
	F1-Score	78%	98%	75%	92%	70%

Table 11 shows the performance of Model 1 ERNN based on precision, recall, and F1-score values for five aspects of hotel reviews on TripAdvisor: Service, Cleanliness, Location, Value, and Rooms. In the Service aspect, the model achieves a precision of 75%, recall of 81%, and F1-score of 78%. The higher recall value indicates that the model is better at detecting all relevant reviews related to service, although slightly less accurate in terms of precision. For the Cleanliness aspect, the model shows excellent performance with a precision of 97%, recall of 99%, and F1-score of 98%. This indicates that the model is very accurate and efficient in identifying reviews related to cleanliness, with a very low error rate. In the Location aspect, the model achieves a precision of 73%, recall of 78%, and F1-score of 75%, indicating that although the model is quite good at detecting reviews related to location, there is still room for improvement in terms of accuracy. For the Value aspect, the model has a precision of 92%, recall of 94%, and F1-

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

score of 92%, demonstrating very good performance in detecting the value perceived by users, with a good balance between precision and recall. Lastly, in the Rooms aspect, the model achieves a precision of 69%, recall of 74%, and F1-score of 70%, indicating decent performance but still requiring improvement to increase accuracy and consistency in detecting reviews related to room quality. Overall, Model 1 ERNN shows good performance in the tested aspects, with relatively high precision, recall, and F1-score values, especially in the Cleanliness and Value aspects. Furthermore, to learn more about the model's performance, an analysis is also conducted on Model 2 ERNN, which is presented in Table 12.

Table 12. Precision, Recall, and F1-Score Results on ERNN Model 2

Model	Testing	Aspect				
		Service	**Cleanliness**	**Location**	**Value**	**Rooms**
Model 2 ERNN	Precision	75%	97%	70%	92%	65%
	Recall	81%	99%	75%	94%	72%
	F1-Score	78%	98%	71%	92%	66%

Table 12 shows the performance of Model 2 ERNN in multi-aspect labeling of hotel reviews on TripAdvisor, focusing on five aspects: Service, Cleanliness, Location, Value, and Rooms, based on precision, recall, and F1-score values. In the Service aspect, the model achieves a precision of 75%, recall of 81%, and F1-score of 78%, indicating a reasonably good performance in detecting relevant reviews related to service, with a balance between accuracy and the ability to detect correct reviews. For the Cleanliness aspect, the model demonstrates excellent performance with a precision of 97%, recall of 99%, and F1-score of 98%, indicating that the model is very accurate and efficient in identifying cleanliness reviews with a very low error rate. In the Location aspect, the model has a precision of 70%, recall of 75%, and F1-score of 71%, showing that although the model is quite good at detecting reviews related to location, there is room for improvement in terms of accuracy. The model is not fully capable of detecting location reviews with high precision, meaning there are still errors in classifying location reviews. For the Value aspect, the model achieves a precision of 92%, recall of 94%, and F1-score of 92%, indicating that the model has a very good performance in detecting the value perceived by users, with a good balance between precision and recall. Lastly, in the Rooms aspect, the model has a precision of 65%, recall of 72%, and F1-score of 66%, showing a decent performance but still requiring improvement to increase accuracy in detecting reviews related to room quality. Overall, Model 2 ERNN shows solid performance with high values in the Cleanliness and Value aspects, but there are some areas, such as Location and Rooms, that need improvement to achieve better results.

After the model testing process using precision, recall, and F1-score values, it is evident that Model 1 ERNN generally shows better and more consistent performance in most of the tested aspects. In the Service, Cleanliness, Location, Value, and Rooms aspects, Model 1 ERNN has good precision and recall, especially in the Cleanliness and Value aspects, with precision values of 97% and 92%, respectively, and recall of 99% and 94%, demonstrating a very good ability to identify relevant reviews. Although Model 1 shows solid performance, there is a decrease in the Location and Rooms aspects, with precision of 73% and 69% and recall of 78% and 74%, indicating room for improvement in terms of accuracy. Conversely, Model 2 ERNN shows solid performance in the Cleanliness and Value aspects with precision of 97% and 92% and recall of 99% and 94%, very similar to Model 1. However, Model 2 has more difficulty in the Location and Rooms aspects, with precision of 70% and 65% and recall of 75% and 72%, showing that this model faces challenges in detecting relevant reviews for these two aspects. Overall, Model 1 ERNN is more reliable and consistent in classification performance in most aspects, while Model 2 ERNN shows competitive performance in some aspects but needs improvement, especially in the Location and Rooms aspects.

4. CONCLUSION

A multi-aspect sentiment analysis was conducted on hotel reviews on TripAdvisor using the Elman Recurrent Neural Network (ERNN) deep learning model. The data used includes 20,492 reviews with a focus on five main aspects: Service, Cleanliness, Location, Value, and Rooms. This study examines the effectiveness of the Elman Recurrent Neural Network (ERNN) in performing multi-aspect sentiment analysis on user reviews from the TripAdvisor platform. The study focuses on various dimensions of user experience, including service, cleanliness, location, value, rooms, and overall experience. It demonstrates the efficacy of the Elman Recurrent Neural Network (ERNN) in accurately classifying sentiments across these dimensions. The model demonstrated remarkable accuracy in evaluating specific aspects, such as cleanliness (98.71%) and value (93.84%), which highlights its potential to provide detailed insights into specific areas of hotel service. These findings indicate that the ERNN approach can markedly enrich the comprehension of user feedback by providing comprehensive, aspect-specific insights, which can be beneficial for both travelers seeking well-informed decisions and hotel management striving to enhance service quality.

The analysis results show that Model 1 ERNN overall performs better than Model 2 ERNN in most aspects. Model 1 ERNN demonstrates high accuracy in the Cleanliness and Value aspects and solid performance in the Service, Location, and Rooms aspects, although there is room for improvement in the Location and Rooms aspects. Meanwhile, Model 2 ERNN excels in the Service and Value aspects but shows lower performance in the Location and Rooms

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1034−1044
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5746

aspects. Based on these results, Model 1 ERNN is the better model due to its better consistency in classification across various aspects of the reviews. For future research, it is hoped that the testing process can be conducted with more complex models such as Long Short-Term Memory (LSTM) or Convolutional Neural Network (CNN) methods.

REFERENCES

[1] J. Miguéns, R. Baggio, and C. Costa, "Social media and Tourism Destinations: TripAdvisor Case Study," in *Proceedings of the IASK International Conference on 'Advances in Tourism Research*, Aveiro, 2008.
[2] D. Sharma, A. Kulshreshtha, and P. Paygude, "Tourview : Sentiment Based Analysis on Tourist Domain," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 3, pp. 2318–2320, 2015.
[3] W. Paulina, F. A. Bachtiar, and A. N. Rusydi, "Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Terhadap Kertanegara Premium Guest House Menggunakan Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1141–1149, 2020.
[4] A. Mathew, P. Amudha, and S. Sivakumari, "Deep Learning Techniques: An Overview," in *Advanced Machine Learning Technologies and Applications*, A. Hassanien, A., Bhatnagar, R., Darwish, Ed., Springer, Singapore, 2021. doi: 10.1007/978-981-15-3383-9_54.
[5] I. Castiglioni *et al.*, "AI applications to medical images: From machine learning to deep learning," *Phys. Medica*, vol. 83, no. February, pp. 9–24, 2021, doi: 10.1016/j.ejmp.2021.02.006.
[6] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, 2020, doi: 10.1016/j.physd.2019.132306.
[7] A. Sadeghi-Niaraki, P. Mirshafiei, M. Shakeri, and S. M. Choi, "Short-Term Traffic Flow Prediction Using the Modified Elman Recurrent Neural Network Optimized through a Genetic Algorithm," *IEEE Access*, vol. 8, pp. 217526–217540, 2020, doi: 10.1109/ACCESS.2020.3039410.
[8] A. Shafaee, H. Issa, S. Agne, S. Baumann, and A. Dengel, "Aspect-Based Sentiment Analysis of Amazon Reviews for Fitness Tracking Devices," in *Conference: DMDA 2014, PAKDD Workshop on Data Mining and Decision Analytics for Public Health and Wellness*, Taiwan, 2014, pp. 50–61. doi: 978-3-319-13186-3_6.
[9] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for aspect-level sentiment classification," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, Texas: Association for Computational Linguistics, 2016, pp. 606–615. doi: 10.18653/v1/d16-1058.
[10] P. A. Aritonang, M. E. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using CNN (Case Study : Peduli Lindungi Application)," *Ultim. Infosys J. Ilmu Sist. Inf.*, vol. 13, no. 1, pp. 54–61, 2022.
[11] A. A. Zia, "SENTIMENT ANALYSIS FOR MOVIEREVIEWS USING ARTIFICIAL NEURAL NETWORKS AND RECURRENT NEURAL NETWORKS," Rajshree Institute of Management & Technology Bareilly-243001 (U.P.), 2023.
[12] M. H. Alam, W. J. Ryu, and S. K. Lee, "Joint multi-grain topic sentiment: Modeling semantic aspects for online reviews," *Inf. Sci. (Ny).*, vol. 339, pp. 206–223, 2016, doi: 10.1016/j.ins.2016.01.013.
[13] W. Wijanarto and S. P. Brilianti, "Peningkatan Performa Analisis Sentimen Dengan Resampling dan Hyperparameter pada Ulasan Aplikasi BNI Mobile," *J. Eksplora Inform.*, vol. 9, no. 2, pp. 140–153, 2020, doi: 10.30864/eksplora.v9i2.333.
[14] R. Wahyudi and G. Kusumawardhana, "Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021.
[15] S. Supriyatna and E. Fahrudin, "Pemanfaatan Algoritma Text Mining Dalam Menemukan Pola Risiko Bencana Sebagai Pengetahuan Kebencanaan Dari Dokumen Kajian Risiko Bencana (KRB)," *J. Inform. Utama*, vol. 2, no. 1, pp. 35–42, 2024, doi: 10.55903/jitu.v2i1.xx.
[16] N. Charibaldi, A. Harfiani, and O. S. Simanjuntak, "Comparison of the Effect of Word Normalization on Naïve Bayes Classifier and K-Nearest Neighbor Methods for Sentiment Analysis," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 25–31, 2024, doi: 10.25139/inform.v9i1.7111.
[17] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis Haiyun," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, vol. 34, no. 05: AAAI-20 Technical Tracks 5, pp. 8600–8607, 2020, doi: 10.1609/aaai.v34i05.6383.
[18] G. M. Raza, Z. S. Butt, S. Latif, and A. Wahid, "Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models," *2021 Int. Conf. Digit. Futur. Transform. Technol. ICoDT2 2021*, pp. 1–6, 2021, doi: 10.1109/ICoDT252288.2021.9441508.
[19] D. Durstewitz, G. Koppe, and M. I. Thurm, "Reconstructing Computational Dynamics from Neural Measurements with Recurrent Neural Networks," *Nat. Rev. Neurosci.*, vol. 24, no. 11, pp. 693–710, 2023, doi: 10.1038/s41583-023-00740-7.
[20] M. Fetanat, M. Stevens, P. Jain, C. Hayward, E. Meijering, and N. H. Lovell, "Fully Elman Neural Network: A Novel Deep Recurrent Neural Network Optimized by an Improved Harris Hawks Algorithm for Classification of Pulmonary Arterial Wedge Pressure," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1733–1744, 2022, doi: 10.1109/TBME.2021.3129459.
[21] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," *Conf. Int. Conf. Soft Comput. Artif. Intell. Mach. Learn.*, pp. 01–14, 2020, doi: 10.5121/csit.2020.100801.