JOURNAL LA MULTIAPP

VOL. 05, ISSUE 05 (524-539), 2024
DOI: 10.37899/journallamultiapp.v5i5.1541

Impact of Feature Extraction on Multi-Aspect Sentiment Classification for Livin'byMandiri Using BiLSTM

Balqis Sayyidahtul Atikah[1] , Yuliant Sibaroni[1], Diyas Puspandari[1]

[1]*Faculty of Informatics, Universitas Telkom, Bandung*

**Corresponding Author: Balqis Sayyidahtul Atikah*
Email: balqissatikah@student.telkomuniversity.ac.id

Article Info

Article history:
Received 1 August 2024
Received in revised form 16 August 2024
Accepted 1 September 2024

Keywords:
Aspect-Based Sentiment Analysis
Bidirectional LSTM
TF-IDF
Word2Vec
Application Review

Abstract

Mobile applications are currently experiencing very rapid development including applications in the financial sector. Livin'byMandiri is one of the mobile applications used to transact online without the need to go to the bank. This makes it very easy for customers to transact anywhere and anytime. Application reviews are user reviews that reflect the reputation of the application among the community, these application reviews can be found anywhere, so many companies use application reviews as a reference in developing their applications in the future. However, people's opinions on apps can vary and are influenced by many aspects. Therefore, aspect-based sentiment analysis can be applied to app reviews to get better results. This research focuses on analyzing the sentiment of Livin'byMandiri app reviews on the Google Play Store. In this research, the Bidirectional LSTM (Bi-LSTM) method is combined with TF-IDF and Word2Vec feature extraction. From the results of the experiments that have been carried out, the best accuracy results for the access aspect are 81.18% and F1-Score of 81.03%, the service aspect produces an accuracy of 82.82% and F1-Score of 82.74%, and for the convenience aspect produces an accuracy of 77.28% and F1-Score of 77.19%. In this experiment, it is also found that feature extraction has an effect on sentiment analysis, this is evidenced by an increase in accuracy of more than 1% for each aspect when TF-IDF feature extraction is added and also the combination of TF-IDF and Word2vec in the initial model built using only the Neural Network embedding layer.

Introduction

Based on a survey conducted by the Indonesian Internet Service Providers Association (APJII), in 2021-2022 it was found that the number of Indonesians connected to the internet reached 210-272 million people (Arif, 2024). Among users who access the internet, it is known that 1.37% often access electronic wallets. From this data it can be concluded that the public's need for electronic wallets is quite high, one of the electronic wallet applications that is widely used in Indonesia is the Bank Mandiri mobile banking application, better known as Livin'byMandiri. Since this application was released in 2021, the mobile banking application has been downloaded more than 10 million times on the Google Play Store platform.

Livin'byMandiri on Google Play Store is a *mobile banking* application with one of the most users in Indonesia. The Google Play Store platform provides a feature to provide reviews and ratings that can be accessed by anyone, both users and potential users. Therefore, many companies utilize this feature to monitor the assessment of their application among users which will be used as a reference in future application development. Reviews given by users

ISSN: 2716-3865 (Print), 2721-1290 (Online)

can be positive, negative, or neutral reviews that are influenced by various aspects. With the many aspects discussed in the reviews, an aspect-based sentiment analysis is carried out which sees better results.

In this research, a sentiment analysis will be conducted that considers specific aspects related to the Livin'byMandiri application, such as access to the application, ease of use, and services provided. Research on sentiment analysis of banking applications has previously been conducted. Research on SVM kernel model comparison analysis in sentiment analysis of BCA Bank user opinions on Twitter resulted in an accuracy value for the RBF kernel of 73.3%, while for the Linear kernel it was 72%, and the polynominal kernel was 67.3%. Another study, entitled Sentiment Analysis of Digital Bank Reviews on the Google Play Store Using the Support Vector Machine (SVM) Method where this study analyzes sentiment on digital banks Bank Jago, Neobank, Seabank using data of 1500 reviews and classifies them using the Support Vector Machine (SVM) model resulting in an average accuracy of 91%, this study concluded that Seabank has a higher percentage of positive responses than the other three digital banks, this study also suggests using a larger dataset. Furthermore, research on user sentiment analysis of Indonesian Islamic bank applications using the Support Vector Machine (SVM) algorithm and produces a training accuracy value of 85.87%, and projected analysis results of 85.87% (Aryati & Sibaroni, 2023; Mardiyanto et al., 2023; Sujjadaa et al., 2023). The weakness in this study is that the accuracy value obtained is still relatively low, so it is recommended to use cross validation to achieve a better accuracy value. So it can be concluded that the above research still uses old machine learning methods that can still be optimized with deep learning.

Based on the above review, the authors conducted aspect-based sentiment analysis research on Livin'byMandiri application reviews using the Bi-LSTM method which has been proven to be able to perform better classification, and also compared the results if TF-IDF and Word2Vec feature extraction were added.

Similar research has been conducted, by analyzing customer satisfaction sentiment on expedition services using Bi-LSTM and Bi-GRU which produces the greatest accuracy for the Bi-LSTM algorithm of 71.9% and for the Bi-GRU algorithm produces an accuracy of 70.1% (Pranida & Kurniawardhani, 2022). The selection of BiLSTM in this study is based on previous research that analyzed the comparison of the LSTM and Bi-LSTM methods for the classification of *phonocardiogram* heart signals, which resulted in an accuracy value of 81% for the LSTM method and 89% for the Bi-LSTM method (Rizky et al., 2021).

Literature Review

Sentiment Analysis is a technique used to automatically extract text data so that information is obtained to determine the tendency of assessing a topic or object consisting of positive, negative, and neutral reviews (R. K. Dewi et al., 2023). Aspect based sentiment analysis is a technique used to determine the types of aspects contained in a text so that the sentiment analysis process will be more detailed and in-depth because the text has been grouped into several different aspects. So that multi-aspect based labeling is needed which aims to identify and determine words that have the same category so that relevant aspects can be determined. (Dewi et al., 2018).

Related research on this research has previously been conducted. Research conducted by Gracia Radiena et al. (2023) Aspect Sentiment Analysis on KAI Access Application Reviews using CRISP-DM (*Cross-Industry Starndard Process for Data Mining*) and Support Vector Machine Method to perform classification, analyzed 8,079 reviews on three aspects, namely learnability, effecienty, errors, and Satisfaction. The best classification results of accuracy, perception recall, and f1-Score values obtained from each aspect are, Learnability 94.73%, 100.00%, 89.50%, and 94.64%, Efficiency 94.38%, 72.00%, 100.00%, and 94.46%, Errors

ISSN: 2716-3865 (Print), 2721-1290 (Online)

85.13%, 97.11%, 72.41%, and 82.96%, Satisfaction 87.26%, 98.46%, 73.78%, and 84.20% (Radiena & Nugroho, 2023). Other research, related to aspect-based sentiment analysis on the JKN Mobile application review, this study modeled aspects / topics using the Latent Dirichlet Allocation (LDA) method and Naïve Bayes and Lexicon Based analysis. The resulting aspects are Services and Features, Register and Login, and User Satisfaction. The sentiment analysis results of the Naïve Bayes method are better than Lexicon Based, namely, Naïve Bayes gets the highest accuracy at 94.75% while Lexicon based with Inset Lexicon gets an accuracy of 59.99% (Roiqoh et al., 2023). The next research discusses aspect-based sentiment analysis on Gojek application reviews, to identify relevant topics researchers use Latent Dirichlet Allocation (LDA) which produces three aspects, namely user experience, service, and payment. Sentiment analysis is carried out with the BERT model which produces accuracy, precission, recall, and f1-score on each aspect above 85% with the best performance of the BERT model on gojek reviews contained in the service aspect of 98.78% accuracy, 96% precision, 94.73% recall, and F1-Score 95.36% (Simanungkalit et al., 2024).

Research conducted by Ardian Nur Romadhan et al. related to the classification of public opinion using the BiLSTM method and using two types of term weighting, namely, TF-IDF and Word2Vec which produces the best performance at a data ratio of 80: 20 with a precision value of 85%, recall of 86.85%, and F1-Score of 85.92% where this research shows that the comparison of the BiLSTM and Word2Vec models is superior to the BiLSTM model with TF-IDF. (Romadhan et al., 2023). Further research conducted by Darwis Alwan et al, raised the topic of Average Word2Vec as feature extraction on sentiment analysis of Movie reviews on IMDb using Artificial Neural Network (ANN) resulting in the best accuracy of 88.52% which proves this method can classify reviews very well with an area under the curve (AUC) of 95%. (Darwis Alwan & Ridla, 2024).

In research conducted by Setyo Adji Pratomo et al. it was found that by combining TF-IDF and Lexicon features in sentiment analysis of movie reviews using the KNN method resulted in accuracy that was not higher than using only TF-IDF extrasu features, namely 73.31% and the use of IG selection features with the right threshold can optimize performance results. The author concludes that the addition of feature extraction can affect the performance of the model built and also the *deep learning* model has better accuracy. So the author conducted research using the BiLSTM model by combining TF-IDF and Word2Vec feature extraction (Pratomo et al., 2021).

Methods

The system built in this research has several processes described in Figure 1 Sentiment analysis begins by using data scrapping techniques. After that, the collected data will be labeled as positive, neutral, or negative. Next, the dataset goes through a pre-processing stage to transform the data into something more acceptable to the system. Feature extraction is used to see its effect on the accuracy of the system to be created. The model classification process with BiLSTM is carried out on the training data and the test data is processed with the sentiment prediction model. The processed sentiment model is then evaluated to determine the performance of the system.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

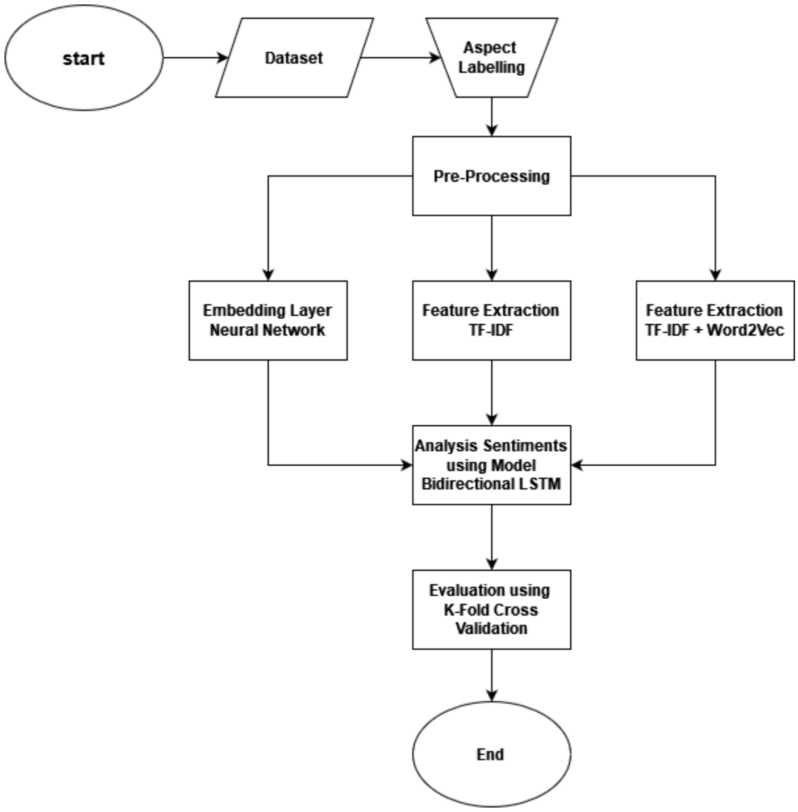

Figure 1. Flow of Aspect-based Sentiment Analysis System

Dataset

The dataset used in this research is taken from reviews found on the Google Play Store by scrapping data using the API provided by Google Play, this API helps retrieve data from the Google Play store with various features or attributes that we need. The data taken uses Indonesian language reviews. Table 1 is an explanation of the attributes of the data taken from the Google Play Store.

Table 1. Feature Description of the Dataset

Features	Description
Username	Username
Score	The value given by users to the quality of the application
At	User date wrote the review
Content	It is a message left by the user that contains a sentence

Aspect Determination

Determination of the aspects used is based on the Delone Mclean model discovered by William Delone and Ephram R. McLean in 1987. In the Delone Mclean model, there are six variables that affect the success of information systems, namely Information Quality, System quality, Service quality, Usage Intentions, User Satisfaction, and System Benefit. Of the six dimensions above, the authors use three aspects, namely, system quality called access, service quality or service, and user satisfaction or comfort (Saputra et al., 2023). Each review will be classified into aspects and labeled with the appropriate sentiment if it discusses one of the things as below.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Access (System Quality)

Aspects of Livin'byMandiri app access include app performance, app security, app proficiency, and 24-hour access.

Service Quality

The service aspect of the Livin'byMandiri application refers to the quality and completeness of the services provided and also includes the application's ability to meet user needs.

Comfort (User Satisfaction)

The comfort aspect of the Livin'byMandiri application focuses on the level of user satisfaction in using the application and includes ease of use, application appearance, and application effectiveness.

Data Labelling

After collecting data, the next step is to label the data or commonly referred to as Data Labelling. Data Labelling is a process of giving values or identifying data including Positive, Neutral, or Negative reviews which is done manually by the author. The author labeled 4926 data. The data labels used are in the form of numbers, where "-1" represents negative values, "1" represents positive values, and "0" represents neutral values. Table 2 will show an example of labeled data and Table 3 will show the distribution of data that has been labeled.

Table 2. Label Data

Aspects			**Content**
Access	**Services**	**Comfort**	
-1	-1	-1	apa apa susah dilivin ni, udahlah trasaksi malam tak bisa terus, asik maintenance aja
-1	-1	0	Tdk bisa di pakai saat melewati jam malam saat urgent, sudah di hubungi tapi tidak ada kejelasan nya atas masalah yang saya alami.
-1	0	0	Mobile banking terburuk yang dari semua bank di Indonesia !! , Bank BUMN kualitas APK kok kaya gini ...mobile banking terburuk !!!! Sistem KAYAK GINI MASIH aja DI PAKAI .
0	0	1	Sya kasi bintang 5 sekalian aplikasi ny sangat membantu

Table 3. Data Distribution by Aspect

Aspect Name	**Label**			**Total**
	Negative	**Neutral**	**Positive**	
Access	2147	1949	906	4926
Services	1598	2350	1054	
Comfort	1332	2236	1431	

Data Pre-Processing

The next step after Data Labelling is pre-processing. The purpose of this pre-processing is to transform the dataset into information that is more easily accepted, more efficient, and useful by the system without changing the meaning of the dataset. The first step is Case Folding or changing all letters in the review to lowercase letters, then cleaning the data or removing things that are not too important and can affect accuracy, namely, removing emojis, punctuation marks, repeated words, and removing sentences that have no meaning. Then Tokenizing *is* done or breaking sentences into tokens or words. After tokenization, data normalization is carried out, namely the elimination of slangwords or unofficial language varieties that are generally used in daily conversation into more official words. After that, the process of removing words that are considered unimportant by doing stopword removal. And

ISSN: 2716-3865 (Print), 2721-1290 (Online)

the last is the stemming process or changing the review to the basic word form (Munawi et al., 2023). Table 4 shows the pre-processing process carried out.

Table 4. Pre-Processing Data

Pre-Processing	Sentence	Explanation
Initial sentence	Transaksi sring gagal, APLIKASI LIVIN SANGAT MENGECEWAKAN!	-
Case Folding	transaksi sring gagal, aplikasi livin sangat mengecewakan!	Capital letters in sentences are changed to lowercase letters.
Cleaning Data	transaksi sring gagal aplikasi livin sangat mengecewakan	The punctuation marks in the sentence are omitted. Such as the comma in the kamilat and the exclamation mark at the end of the sentence.
Tokenizing	['transaksi', 'sring', 'gagal', 'aplikasi', 'livin', 'sangat', 'mengecewakan']	Sentences are converted into tokens.
Data Normalization	['transaksi', 'sering', 'gagal', 'aplikasi', 'livin', 'sangat', 'mengecewakan']	Words that are not standardized are changed into standardized words. For example, the word 'sring' was changed to 'sering'.
Stopword Removal	['transaksi', 'gagal', 'aplikasi', 'livin', 'mengecewakan']	Stopword removal. Such as the word 'sering' in the sentence
Stemming	transaksi gagal aplikasi livin kecewa	In the last stage, stemming is done, such as the word 'mengecewakan' into its base word 'kecewa'.

K-Fold Cross Validation Evaluation

The training process in this study uses K-Fold Cross Validation, the use of cross-validation is to test the performance of the model in the classification of data that has been designed (Fuadah et al., 2022). In this test, the value of k in k-fold cross validation is 10, where the dataset used will be randomly divided into 10 parts, where 90% of the data will be used as *test* data and 10% will be used as *testing* data. Figure 2 illustrates the dataset divided into K-fold Cross Validation.

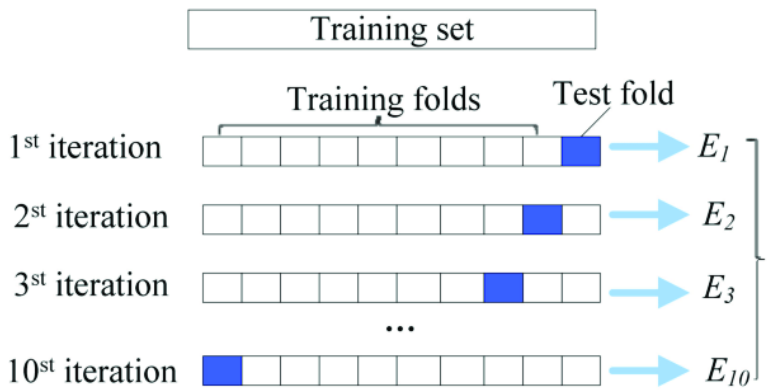

Figure 2. Illustration of K-Fold Cross Validation

TF-IDF Feature Extraction

TF-IDF (Term-Frequency-Inverse Document Frequency*)* Feature Extraction is one of the word weighting methods where Term Frequency (tf(w,d)) is considered to have a value or

ISSN: 2716-3865 (Print), 2721-1290 (Online)

proportion of importance according to how much it appears in the text or document (Gifari et al., 2022). While Inverse Document Frequency *(IDF)* is a token weighting method that has a function to monitor the appearance of tokens in the text set. So, it can be concluded that the greater or more often a word appears in a document will give a higher term frequency and the less occurrence of a word in a document will result in a higher level of importance (IDF) for keywords searched in a particular document (Safira, 2023).

In this research, TF-IDF feature extraction is used to convert text data into numeric vectors. The implementation of TF-IDF feature extraction uses the TF-IDFVectorizer python library from SkicitLearn to build TF-IDF vectors. Equation 1 is the formula of TF-IDF for the word symbolized i in the document symbolized j.

$$W_{i,j} = tf_{i,j} \times \log \frac{N}{df_j}$$

Description:

$W_{i,j}$ = weight of word i against document j

$tf_{i,j}$ = number of occurrences of word i in document j

N = number of documents

df_j = number of documents containing word i

Word2Vec Feature Extraction

Word2Vec computes word representations into vectors using a neural network that aims to group similar words and compute word representations into vectors (Djaballah et al., 2019). Word2Vec architecture consists of only 3 layers namely Input, Projection (Hidden Layer), and Output. Word2Vec input is a one-hot encode vector with a length that represents the number of unique words in the training data.

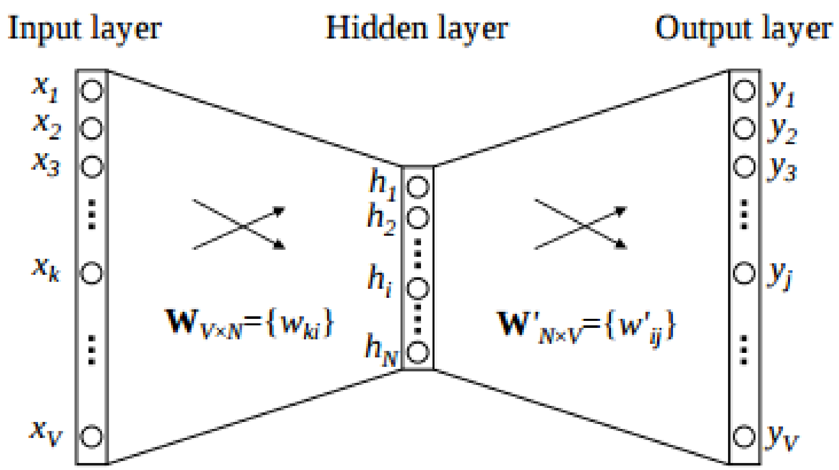

Figure 3. Word2Vec Model Architecture

There are two types of architectures in the Word2Vec neural network, namely, Skip gram and *Continous Bag of Word* (CBOW). The CBOW architecture predicts the target word based on its context or surrounding words so that the specific position of the word in a sentence has no effect, while the skip-gram uses the word to predict the target context (Rahmadanisya et al., 2022). This study uses one of the architectures of Word2Vec, namely, Continous Bag of Word (CBOW) because the CBOW architecture has the advantage that certain positions in words do not matter. The CBOW model also requires relatively faster training time and produces better accuracy for frequent words (Af'idah et al., 2021).

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Bidirectional LSTM (BiLSTM)

Bidirectional LSTM or BiLSTM is a model developed to overcome the shortcomings of RNN and LSTM. BiLSTM is one of the deep learning algorithms that consists of two LSTM layers. BiLSTM utilizes backward context and forward context by processing information from two directions with separate hidden layers and connecting the two opposing hidden layers into one output (Mutmainah et al., 2014; Mutmainah et al., 2023). Figure 4 is an overview of the BiLSTM architecture.

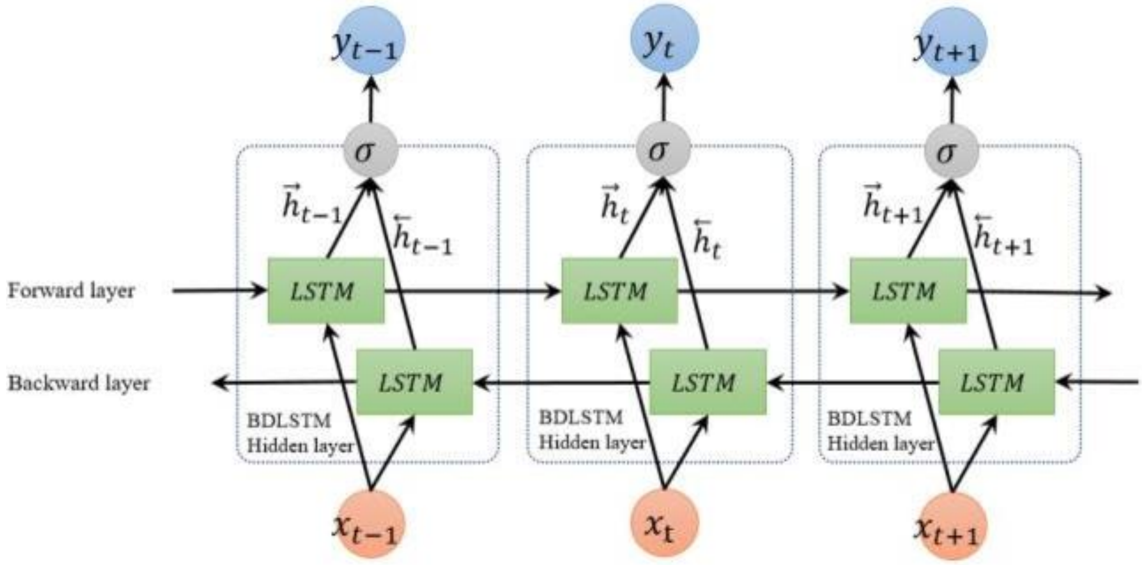

Figure 4. BiLSTM Model Architecture

Because Bidirectional LSTM is a two-way LSTM model, before calculating the value of Bidirectional LSTM, the value of LSTM is needed (Puteri, 2023). LSTM consists of equations (2), (3), (4), (5), (6), (7), (8). While Bidirectional LSTM itself has three layers, namely forward layer, backward layer, and output layer (Puteri et al., 2024). Equations (9), (10), and (11) are equations for each layer.

$f_t = \sigma(W_{fh}[h_{t-1}], W_{fx}[X_t], b_f)$

$i_t = \sigma(W_{ih}[h_{t-1}], W_{ix}[X_t], b_i)$

$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t)$

$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t)$

$O_t = \sigma(W_{oh}[h_{t-1}], W_{ox}[X_t], b_0)$

$y_t = \sigma(\overrightarrow{h_t}, \overleftarrow{h_t})$

$h_t = O_t * \tanh\,(C_t$

$i_t = \sigma(W_{ih}[h_{t-1}], W_{ix}[X_t], b_i)$

$\overrightarrow{h_t} = LSTM(x_t, h_{t-1})$

$\overleftarrow{h_t} = LSTM(x_t, h_{t+1})$

Description:

f_t = forget gate

ISSN: 2716-3865 (Print), 2721-1290 (Online)

i_t = Gate input

O_t = Output Gate

σ = sigmoid function

x_t = current input value

W = Weight

b_f = bias forget gate LSTM

b_i = LSTM gate input bias

b_0 = LSTM gate output bias

h_{t-1} = output of t-1

$\tilde{C}_t$ = current cell condition

C_t = new cell condition

C_{t-1} = memory cell state of the previous cell

h_t = final output

$\overrightarrow{h_t}$ = forward LSTM

Results and Discussion

Research Results

This research tests three scenarios to get a model with the best accuracy results. In each result, the calculation of the increase in accuracy value and F1-Score against the baseline will be carried out to compare the accuracy improvement of each scenario against the accuracy of the baseline model. The test scenario can be seen in Table 5.

Table 5. Testing Scenario

Scenario	Scenario Explanation	Scenario Objective
I	Test the model using the embedding layer provided by the neural network and the BiLSTM model.	Determining the baseline model
II	Test the baseline model by applying TF-IDF.	Getting the best accuracy of the model after applying the combination of TF-IDF type of Unigram features.
III	Test the baseline model by replacing the embedding layer with word2vec combined with TF-IDF.	Get the best accuracy of the model after applying word2vec + TF-IDF.

Scenario I

In the initial stage of experimentation, the author focused on obtaining the best accuracy of the base model by using the embedding layer provided by the neural network. The author adjusted the hyperparameters and the number of data divisions in the model to produce the best output. In this scenario, the author used the embedding layer provided by the hard library with an input_dim of 1200 and output_dim of 50, K-fold with iterations of 10 which can also mean that the data is divided into 90% training data and 10% test data, batch size of 32, epoch of 32, dropout of 0.4, and used 3 layers of BiLSTM with many layer units are 128, 256, 128, the optimizer used is Adam, and Sparse Categorical Crossentropy as a loss function to determine the accuracy value of the model that has been built.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Table 6. Scenario I Results

K-Fold	Aspects					
	Access		Services		Comfort	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
1	82.93%	81.81%	77.64%	77.08%	76.02%	75.97%
2	72.36%	71.39%	82.32%	81.76%	77.85%	77.19%
3	70.12%	68.15%	77.64%	75.51%	75.20%	74.24%
4	80.49%	78.41%	81.10%	79.77%	72.76%	71.23%
5	77.44%	75.60%	82.11%	81.73%	68.29%	67.63%
6	81.50%	80.43%	84.55%	83.73%	66.67%	67.32%
7	69.51%	67.43%	81.91%	80.69%	75.00%	74.83%
8	80.49%	78.25%	79.07%	77.53%	71.34%	72.27%
9	81.91%	80.50%	85.16%	85.18%	73.58%	72.52%
10	79.47%	78.20%	81.71%	80.72%	73.58%	72.89%
Avarage	**77.62%**	**76.02%**	**81.32%**	**80.37%**	**73.03%**	**72.61%**

In the Access aspect, the best accuracy is 82.93% and F1-Score is 81.81% and the average accuracy is 77.62% and F1-Score is 76.92%. For the Service aspect, the best accuracy is 85.16% and F1-Score is 85.18% with an average accuracy of 81.32% and F1-Score of 80.37%, while for the convenience aspect, the best accuracy is 77.85% and F1-Score is 77.19% with an average accuracy of 73.03% and f1-score of 76.61%.

Scenario II explored the impact of incorporating Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction into the baseline model, aiming to assess how this addition influences the model's performance. Specifically, this scenario utilized unigram TF-IDF features, constrained to a maximum of 1000 features, to potentially enhance the model's capacity for capturing relevant information from the data.

The results, detailed in Table 7, indicate a notable improvement in model performance across several aspects. In the Access aspect, the model achieved its highest accuracy of 82.72% and an F1-Score of 80.49% when TF-IDF was applied. The average accuracy in this aspect was 80.19%, with an average F1-Score of 77.82%. These figures reflect a substantial enhancement in accuracy and F1-Score, suggesting that TF-IDF feature extraction significantly contributes to refining the model's ability to classify and predict access-related features.

In the Services aspect, the addition of TF-IDF led to an impressive peak accuracy of 83.33% and an F1-Score of 82.54%. The average performance metrics were also strong, with an average accuracy of 81.67% and an F1-Score of 81.06%. This improvement underscores the effectiveness of TF-IDF in enhancing the model's performance in service-related predictions, likely due to TF-IDF's ability to weigh terms based on their importance across documents.

For the Comfort aspect, TF-IDF resulted in a maximum accuracy of 80.89% and an F1-Score of 81.03%. The average accuracy and F1-Score were 74.93% and 75.01%, respectively. Although the average metrics for Comfort are lower compared to Access and Services, the peak values achieved demonstrate that TF-IDF can still positively impact performance, though the extent of improvement varies across different aspects.

Table 7. Scenario II Results

K-Fold	Aspects					
	Access		Services		Comfort	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
1	82.35%	80.01%	82.35%	81.78%	74.65%	74.67%
2	78.70%	75.19%	80.53%	80.42%	75.05%	75.38%

ISSN: 2716-3865 (Print), 2721-1290 (Online)

3	79.51%	77.37%	81.34%	80.86%	76.47%	76.47%
4	81.74%	80.77%	81.14%	80.49%	74.24%	74.41%
5	78.50%	76.20%	82.15%	81.36%	71.81%	71.27%
6	78.70%	75.92%	82.56%	82.05%	73.83%	74.10%
7	82.72%	80.49%	80.49%	79.93%	80.89%	81.03%
8	81.30%	79.04%	82.32%	81.91%	73.37%	73.53%
9	79.67%	77.01%	80.49%	79.24%	74.59%	74.59%
10	78.66%	76.19%	83.33%	82.54%	74.39%	74.63%
Avarage	**80.19%**	**77.82%**	**81.67%**	**81.06%**	**74.93%**	**75.01%**

In the aspect of Access, the best accuracy is 82.72% and F1-Score is 80.49% and the average accuracy is 80.19% and F1-Score is 77.82%. For the Service aspect, the best accuracy is 83.33% and F1-Score is 82.54% with an average accuracy of 81.67% and F1-Score of 81.06%, while for the convenience aspect, the best accuracy is 80.89% and F1-Score is 81.03% with an average accuracy of 74.03% and f1-score of 75.01%.

In Scenario III of this research, the author will combine TF-IDF and Word2Vec feature extraction in the previously built model. In this scenario, a BiLSTM model is constructed, incorporating two feature extraction techniques: Word2Vec using the CBOW model with a dimension size of 100, and TF-IDF unigram with a maximum of 1000 features.

Table 8. Scenario III Results

K-Fold	**Aspects**					
	Access		**Services**		**Comfort**	
	Accuracy	**F1-Score**	**Accuracy**	**F1-Score**	**Accuracy**	**F1-Score**
1	83.57%	83.47%	82.96%	82.98%	76.47%	76.53%
2	79.72%	79.72%	82.76%	82.76%	79.31%	79.29%
3	78.70%	78.54%	83.37%	83.26%	76.27%	76.11%
4	84.18%	84.22%	82.96%	82.92%	78.70%	78.60%
5	81.74%	81.55%	83.98%	83.91%	77.89%	77.80%
6	80.12%	80.02%	83.57%	83.57%	75.66%	75.55%
7	83.33%	83.08%	79.88%	79.79%	81.50%	81.50%
8	83.13%	83.05%	81.91%	81.96%	75.61%	75.60%
9	81.91%	81.70%	82.52%	82.50%	75.81%	75.81%
10	78.05%	78.05%	82.93%	82.82%	75.41%	75.41%
Avarage	**81.45%**	**81.34%**	**82.68%**	**82.65%**	**77.26%**	**77.22%**

Table 8 presents the results for Scenario III, showing further improvements in model performance across the Access, Services, and Comfort aspects. For Access, the best accuracy was 84.18% with an F1-Score of 84.22%, while the average accuracy was 81.45% and the average F1-Score was 81.34%. In the Services aspect, the best accuracy was 83.98% with an F1-Score of 83.91%, and the average accuracy was 82.68% with an F1-Score of 82.65%. For Comfort, the best accuracy achieved was 81.50%, with an F1-Score of 81.50%, and the average accuracy was 77.26% with an F1-Score of 77.22%.

Analysis of Test Results

The analysis of the test results provides a thorough and insightful summary, effectively capturing the key findings of the study. It emphasizes the significant impact that feature extraction techniques, specifically TF-IDF and Word2Vec, have had on improving the accuracy of the BiLSTM model. The discussion carefully highlights how these techniques contributed to marked increases in accuracy and F1-Score across various aspects, including Access, Services, and Comfort. This clear emphasis on the percentage gains in these metrics

ISSN: 2716-3865 (Print), 2721-1290 (Online)

underscores the importance of feature extraction in enhancing model performance, especially when dealing with complex data.

Moreover, the figures presented in the analysis serve as a strong visual complement to the numerical data, allowing for an easier interpretation of the improvements achieved. These visual aids not only reinforce the written findings but also provide a more intuitive understanding of how the model's performance evolved with the introduction of TF-IDF and Word2Vec. The improvement trends in accuracy and F1-Score are effectively communicated, making it apparent that the integration of these feature extraction methods has had a substantial positive impact on the model.

However, while the analysis effectively conveys the overall improvements, it would benefit from a more in-depth exploration of certain nuances in the results. For instance, the Comfort aspect did not see as significant gains as Access and Services, which raises questions about the effectiveness of the applied techniques in different contexts. A deeper dive into why Comfort lagged behind could reveal important insights into the limitations or challenges of using TF-IDF and Word2Vec in this particular scenario. Understanding these nuances could help refine the model further or guide future research in addressing such discrepancies.

In addition to this, a discussion on the potential trade-offs or limitations of the applied techniques would add valuable depth to the analysis. While the improvements in accuracy and F1-Score are noteworthy, it is essential to consider whether these gains come at the expense of other factors, such as computational efficiency or the model's ability to generalize to new data. A more comprehensive exploration of these aspects would provide a balanced perspective on the model's performance, offering a clearer understanding of its strengths and weaknesses. Overall, while the analysis is strong, incorporating these additional elements would enhance its depth and provide a more holistic view of the model's capabilities and limitations.

Based on the above experiments, it has been proven that the addition of feature extraction affects the accuracy of the model created. The highest accuracy of the BiLSTM model was obtained after the addition of two feature extractions, namely TF-IDF and Word2Vec with the highest accuracy value for the access aspect of 81.45%, for the service aspect of 82.68%, and for the security aspect, an accuracy of 77.26% was obtained.

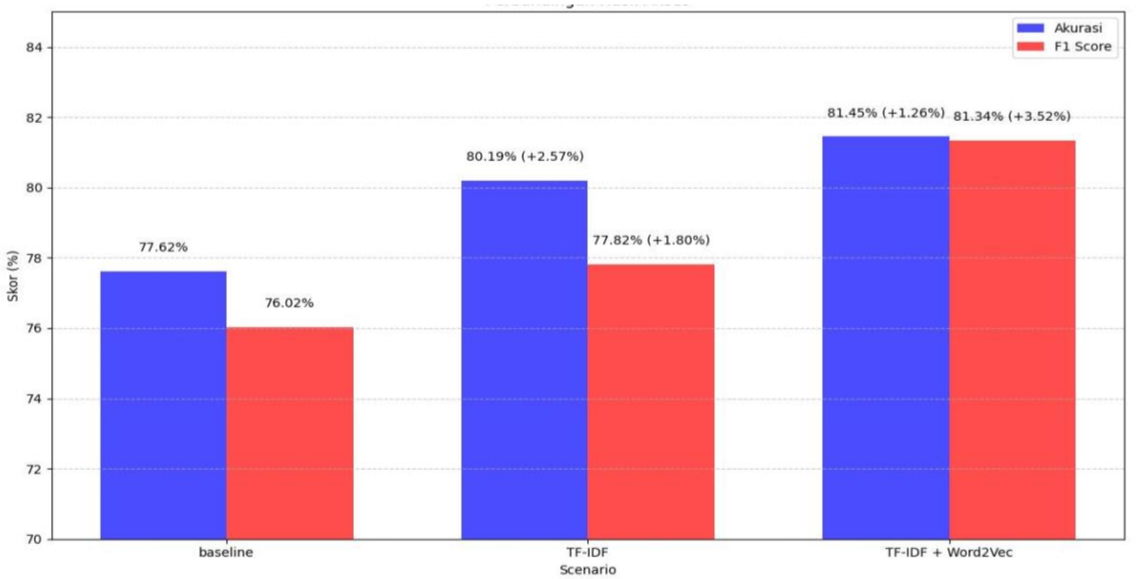

Figure 5. Accuracy and F1-Score Increase for Access Aspects

From Figure 5, it can be seen that the baseline model for access aspects after the addition of TF-IDF feature extraction experienced an increase in accuracy of 2.57% and an increase in

ISSN: 2716-3865 (Print), 2721-1290 (Online)

F1-Score of 1.80%, after the application of Word2Vec feature extraction the model experienced another increase in accuracy of 1.28% and also an increase in F1-score of 3.25%.

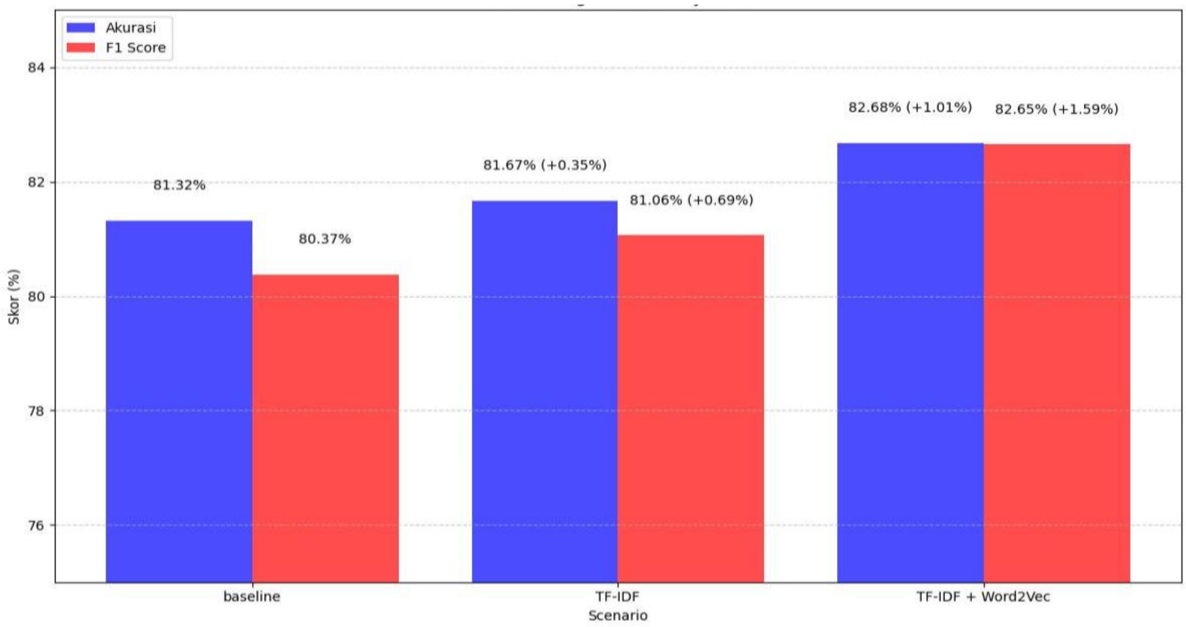

Figure 6. Increase in Accuracy and F1-Score for Service Aspects

From Figure 6, it can be seen that the baseline model for service aspects after the addition of TF-IDF feature extraction experienced an increase in accuracy of 0.35% and also an increase in F1-Score of 0.69%. And after the application of Word2Vec feature extraction the model experienced another increase in accuracy of 1.01% and also an increase in F1-score of 1.59%.

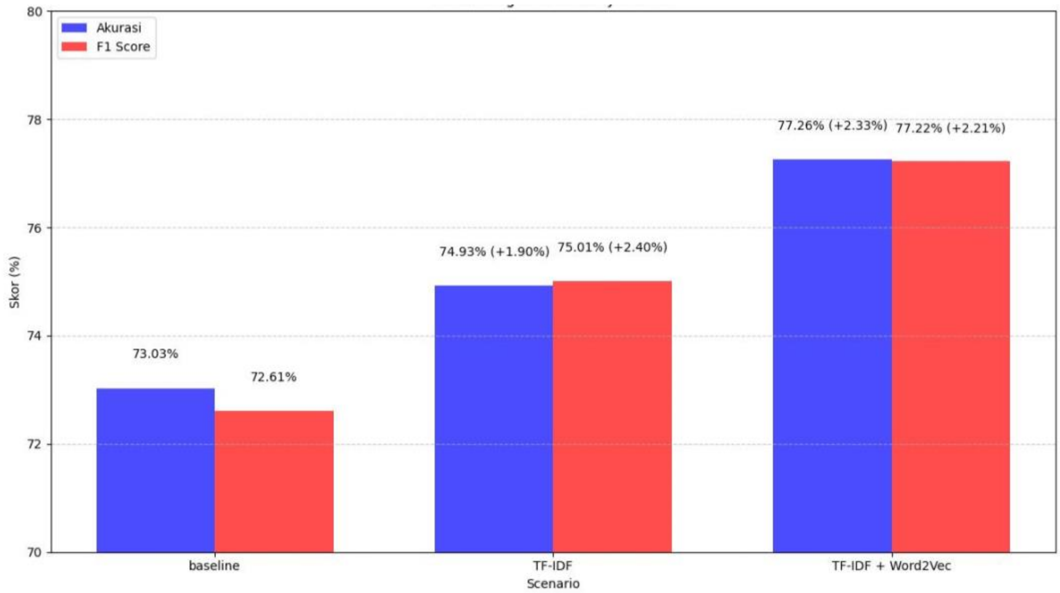

Figure 7. Accuracy and F1-Score Increase for Comfort Aspect

From Figure 7, it can be seen that the baseline model for the comfort aspect after the addition of TF-IDF feature extraction experienced an increase in accuracy of 2.40% and also an increase in f1-Score of 1.90%. And after the application of Word2Vec feature extraction the model experienced another increase in accuracy of 2.33% and also an increase in f1-score of 2.21%.

By doing all the scenarios that have been made, there are changes in the accuracy and F1-Score values of the BiLSTM model. Such as the application of TF-IDF in all scenarios

ISSN: 2716-3865 (Print), 2721-1290 (Online)

resulted in the largest accuracy change at 2.57% and the largest F1-Score change at 3.65%. As well as changes in accuracy and f1-score values in the application of word2vec which in each aspect produces accuracy and f1-score values of more than 75%. This proves that each feature extraction that has been added can affect the final results of the research.

Conclusion

Today With the importance of internet for everybody whether in private life or in practical live, we can imagine the world with out of internet specially with the smart application that presented to us, all the objectives around our environment can be connected to internet with a large network containing different sensors with standard protocol for IoT, and it provides the chance for people to control things over distance Without the need to be in a specific place to deal with a specific device, hence the IoT system that is used in this work is successfully using raspberry pi as an IoT device and HTTP post request for transferring the captured image file and server for receiving and store image. The future work will be about processing the image that stored in the server using training system and robost algorithm for processing using also hardware platform

References

Af'idah, D. I., Dairoh, D., Handayani, S. F., & Pratiwi, R. W. (2021). Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(3), 156–161. https://doi.org/10.30591/jpit.v6i3.3016

Arif, M. (2024). *Profil Internet Indonesia 2022*. Survei.Apjii.or.Id.

Aryati, S. T. R., & Sibaroni, Y. (2023). Analisis Perbandingan Model Kernel Support Vector Machine dalam Analisis Sentimen Opini Pengguna Bank BCA di Twitter. *Jurnal Tugas Akhir Fakultas Informatika*.

Darwis Alwan, & Ridla, M. A. (2024). Averaged Word2vec sebagai Ekstraksi Fitur pada Analisis Sentimen Ulasan Film di IMDb menggunakan Artificial Neural Network (ANN). *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, *9*(1), 36–45. https://doi.org/10.32528/justindo.v9i1.1204

Dewi, Herdiani, A., & Kusumo, D. S. (2018). Multi-Aspect Sentiment AnalysisKomentar Wisata TripAdvisordenganRule-Based Classifier(Studi Kasus : Bandung Raya. *E-Proceeding of Engineering*, *5*(1), 1589–1596.

Dewi, R. K., Tantular, B., Suprijadi, J., & Pravitasari, A. A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi E-Samsat Provinsi Jawa Barat Menggunakan Metode BiGRU. *Inferensi*, *1*(1), 1. https://doi.org/10.12962/j27213862.v1i1.19113

Djaballah, K. A., Boukhalfa, K., & Boussaid, O. (2019). Sentiment Analysis of Twitter Messages using Word2vec by Weighted Average. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 223–228. https://doi.org/10.1109/SNAMS.2019.8931827

Fuadah, Y. N., Ubaidullah, I. D., Ibrahim, N., Taliningsing, F. F., SY, N. K., & Pramuditho, M. A. (2022). Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *10*(3), 728. https://doi.org/10.26760/elkomika.v10i3.728

Gifari, O. I., Adha, Muh., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, *2*(1), 36–40. https://doi.org/10.46229/jifotech.v2i1.330

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Mardiyanto, R. O., Kusrini, & Wibowo, F. W. (2023). Analisis Sentimen Pengguna Aplikasi Bank Syariah Indonesia dengan Menggunakan Algoritma Support Vector Machine (SVM). *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, *4*(1), 9–15. https://doi.org/10.46764/teknimedia.v4i1.85

Munawi, H. A., Alwi, M. H., Nevita, A. P., Santoso, R., & Indrawati, E. M. (2023). Alat Pengendali Stop Kontak Berbasis Wemos D1. *G-Tech: Jurnal Teknologi Terapan*, *7*(1), 29–36. https://doi.org/10.33379/gtech.v7i1.1823

Mutmainah, S., Fudholi, D. H., & Hidayat, S. (2023). Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine Pada Google Play Menggunakan BiLSTM dan LDA. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(1), 312. https://doi.org/10.30865/mib.v7i1.5486

Pranida, S. Z., & Kurniawardhani, A. (2022). Sentiment Analysis of Expedition Customer Satisfaction using BiGRU and BiLSTM. *Indonesian Journal of Artificial Intelligence and Data Mining*, *5*(1), 44. https://doi.org/10.24014/ijaidm.v5i1.17361

Pratomo, S. A., Faraby, S. Al, & Purbolaksono, M. D. (2021). Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF Dan Lexicon Pada Ulasan Film Menggunakan Metode KNN. *EProceedings of Engineering*, *8*(5), 10116–10126.

Puteri, D. I. (2023). Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, *11*(1), 35–43. https://doi.org/10.34312/euler.v11i1.19791

Puteri, D. I., Darmawan, G., & Ruchjana, B. N. (2024). Prediksi Harga Saham Syariah menggunakan Bidirectional Long Short Term Memory (BiLSTM) dan Algoritma Grid Search. *Jambura Journal of Mathematics*, *6*(1), 39–45. https://doi.org/10.37905/jjom.v6i1.23297

Radiena, G., & Nugroho, A. (2023). Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Kai Access Menggunakan Metode Support vector Machine. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, *6*(1), 1–10. https://doi.org/10.37792/jukanti.v6i1.836

Rahmadanisya, A., Setiawan, E. B., & Adytia, D. (2022). The Influence of Sentiment on Bank Mandiri (BMRI) Stock Movements Using Feature Expansion with Word2vec and Support Vector Machine Classification. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, 287–292. https://doi.org/10.1109/ICoICT55009.2022.9914853

Rizky, M. G., Jusak, & Puspasari, I. (2021). Analisis Perbandingan Metode LSTM dan BiLSTM Untuk Klasifikasi Sinyal Jantung Phonocardiogram. *JCONES - Journal of Control and Network Systems*, *10*(2), 44–49.

Roiqoh, S., Zaman, B., & Kartono, K. (2023). Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(3), 1582. https://doi.org/10.30865/mib.v7i3.6194

Romadhan, A. N., Utami, E., & Hartanto, A. D. (2023). Analisis Sentimen Opini Publik Menggunakan Metode BiLSTM Pada Media Sosial Twitter. *SEMIOTIKA: Seminar Nasional Teknologi Informasi Dan Matematika*, *2*(1), 22–33.

Safira, A. D. (2023). *Deteksi Hoax di Media Sosial Menggunakan Metode Bidirectional Long Short-Term Memory (Bi-LSTM) dan 1 Dimensional-Convolutional Neural Network (1D-CNN)*. Universitas Telkom.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Saputra, D., Haryani, Surniandari, A., & Sidauruk, J. (2023). Implementation Of DeLone And McLean Models To Measure The Success Of Online-Based Tutoring System. *Bulletin of Computer Science and Electrical Engineering*, *4*(1), 8–23.

Simanungkalit, A., Naibaho, J. P. P., & Kweldju, A. De. (2024). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, *13*(1), 659. https://doi.org/10.35889/jutisi.v13i1.1826

Sujjadaa, A., Somantri, Novianti, J. N., & Indra Griha Tofik Isa. (2023). Analisis Sentimen Terhadap Review Bank Digital Pada Google Play Store Menggunakan Metode Support Vector Machine (SVM). *Jurnal Rekayasa Teknologi Nusa Putra*, *9*(2), 122–135. https://doi.org/10.52005/rekayasa.v9i2.345

ISSN: 2716-3865 (Print), 2721-1290 (Online)