ISSN 2723-6471

A Lexicon-Based Long Short-Term Memory (LSTM) Model for Sentiment Analysis to Classify Halodoc Application Reviews on Google Playstore

Rina Refianti[1], Achmad Benny Mutiara[2, *], Ryan Arya Putra[3]

[1,2] *Faculty of Computer Science and Information Technology, Gunadarma University, Jl. Margonda Raya No.100, Depok 16464, Indonesia*

[3] *Department of Informatics Engineering, Gunadarma University, Jl. Margonda Raya No.100, Depok 16464, Indonesia*

(Received: November 23, 2023; Revised: December 8, 2023; Accepted: January 14, 2024; Available online: January 29, 2024)

Abstract

The development of information and communication technology is developing very quickly, has made many new breakthroughs. One of these technological advances is in the health sector, the creation of telemedicine applications. During the Covid-19 pandemic, it is difficult for people to get access to health. Therefore, telemedicine applications are needed. Halodoc is one of the telemedicine applications that has successfully become the top health application on the Google PlayStore. The application has been used by more than ten million users throughout Indonesia and received a rating of 4.6. To be able to see ratings and satisfaction from the public, user reviews are needed. The very large number of reviews often contain errors, making them difficult to decipher. Based on this, this research aims to create a web application, which can classify user reviews of the Halodoc application, using a proposed lexicon-based Long Short-Term Memory (LSTM) Model. Application is built using the Flask framework and the Python programming language. Models are created and trained using the TensorFlow library. The results of the model evaluation get an accuracy of 85.3% with an average precision value of 85.3%, a recall value of 85.6% and an f1-score of 85.3%. The proposed LSTM model can be used to classify Halodoc review sentiment classes.

Keywords: Sentiment Analysis, Classification, Long Short-Term Memory, Deep Learning, Flask

1. Introduction

Along with the times, the development of information and communication technology is developing very quickly. There have been many technological breakthroughs present in society. One such technology is internet technology. The Internet itself is a computer network throughout the world that is interconnected with each other. The internet makes it easy to find information that can be in the form of images, text, audio, and so on in the form of electronic media. The use of the internet currently almost entirely dominates human activities. Based on data from the survey results of the Indonesian Internet Service Providers Association (APJII) for the 2019-2020 (Q2) period, the total Indonesian internet users currently reach 196.7 million users with a penetration of 73.3 percent of Indonesia’s total population of around 266.9 million. One of the media that is often used to access the internet is a smartphone [9].

The advancement of digital technology in the health sector has been proven by the existence of health applications that can be accessed via smartphones. This can happen because people’s current lifestyles cannot be separated from smartphones. One of the health applications contained in smartphones is Halodoc. Halodoc was launched in 2016 by MHealth Indonesia. This application is a health application that serves and helps people to get information about health and consult with doctors without having to visit clinics or hospitals easily and quickly.

The COVID-19 virus pandemic has made the world community have to get used to a new way of life because of limited access to various fields such as health, education, economy and so on. Due to limited access, especially in health access, many parties are trying to make solutions so that people have access to health without having to face to face with doctors to suppress the spread of the COVID-19 virus [6]. One of these efforts is to create a telemedicine service. Telemedicine services are health services that function to reduce face-to-face contact with doctors. This service allows patients to consult qualified specialists online quickly and efficiently. The Halodoc application is a health service

***Corresponding author: Achmad Benny Mutiara (amutiara@staff.gunadarma.ac.id)**
DOI: https://doi.org/10.47738/jads.v5i1.160

ISSN 2723-6471

provider application found on the Google Play Store and is one of the top health applications. Google Play Store is a digital content provider service that provides various products such as applications, games, books, and music. Google Play Store provides convenience to its users because it can be accessed through the PlayStore website and application. There are many features in the Google Play Store, one of which is the review feature, this feature allows users to be able to leave reviews on the applications they use. A review is a text that contains an assessment of a work. This feature is very important, because reviews can be information to see public ratings based on positive reviews such as appreciation or suggestions, and negative reviews such as complaints to improve the quality of the product. A very large number of user reviews often contain errors in writing that are difficult to interpret.

Sentiment analysis is very useful for analyzing these reviews to later be used as more meaningful information. Sentiment analysis is one of the natural language processing (Natural Language Processing) [13, 18, 19]. Sentiment analysis aims to classify the sentimental polarity of a given text as negative, positive, or other classes [5, 10, 11]. Deep learning is one of the techniques in machine learning that has a deeper architecture than other machine learning techniques in solving prediction and classification problems [5]. Deep learning can help you solve problems at various levels from sample data and training. One of the existing methods in deep learning used for classification is the Long Short-Term Memory (LSTM) method. In this study, we propose a lexicon-based Long Short-Term Memory (LSTM) model. By using this model, sentiment analysis was carried out on reviews of the Halodoc application. Lexicon based is a commonly used method for conducting sentiment analysis because of its practicality. Lexicon based performs sentiment analysis using a dictionary as a data source. LSTM is a modified method of Recurrent Neural Network (RNN) where cells are added to store information for long periods of time to make up for the lack of RNN. LSTM was created by Hochreiter and Schmidhuber in 1997 [7]. LSTM has been shown to cover the shortcomings of the RNN model related to its inability to store memory so that it can be selected for each word according to its context.

From the explanation above, this study aims to conduct sentiment analysis of Halodoc application user reviews using a lexicon- based Long Short-Term Memory (LSTM) model and analyze the value of model accuracy by evaluating the proposed LSTM model. Then we created a website to classify sentiment on user reviews of the Halodoc application using the Flask framework [1–3].

2. Literature Review

2.1. Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a computational technique used to determine and understand the sentiment or emotion expressed in a piece of text. This advanced natural language processing (NLP) technology has gained significant prominence in recent years due to the explosion of user-generated content on social media platforms, product reviews, and other online forums. The primary goal of sentiment analysis is to classify the sentiment of a given text as positive, negative, or neutral, providing valuable insights into public opinion, customer feedback, and trends.

One key aspect of sentiment analysis involves the use of machine learning algorithms to analyze and categorize text data. These algorithms are trained on large datasets containing labeled examples of positive, negative, and neutral sentiments. By learning patterns and associations within the data, the algorithms can then predict the sentiment of new, unseen text. This enables businesses, researchers, and organizations to automate the process of evaluating public sentiment, making it a powerful tool for market research, brand management, and customer relationship management.

Sentiment analysis is particularly useful in the business world, where companies can leverage the insights gained from customer feedback to improve their products and services. By monitoring social media mentions, reviews, and comments, businesses can identify trends, address customer concerns, and capitalize on positive sentiment. Additionally, sentiment analysis can be applied to financial markets, political discourse, and healthcare, providing a comprehensive understanding of public perception in various domains.

Despite its numerous applications and advantages, sentiment analysis faces challenges, such as the nuances of language, context-dependent expressions, and cultural variations in sentiment. Ambiguous language, sarcasm, and the evolving nature of language can pose difficulties in accurately interpreting sentiments. Researchers and developers continually work to enhance sentiment analysis models to overcome these challenges and ensure more accurate and nuanced results.

In conclusion, sentiment analysis is a valuable tool in the realm of natural language processing, offering insights into the emotions expressed in textual data. Whether applied to social media monitoring, customer feedback analysis, or market research, sentiment analysis plays a crucial role in understanding and responding to public sentiment, ultimately aiding businesses and organizations in making informed decisions.

2.2. LSTM

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to overcome the limitations of traditional RNNs in capturing and remembering long-range dependencies in sequential data. Introduced by Hochreiter and Schmidhuber in 1997, LSTMs have become a fundamental building block in the field of deep learning, particularly for tasks involving sequential data such as natural language processing, speech recognition, and time-series analysis.

The key innovation of LSTM networks lies in their ability to selectively store and retrieve information over extended periods. Unlike standard RNNs, LSTMs have a more sophisticated memory cell, consisting of three gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information, allowing the network to decide what information to store, discard, and output. This mechanism enables LSTMs to effectively handle long-range dependencies, making them well-suited for tasks where context and temporal relationships are crucial.

One of the main advantages of LSTMs is their capability to mitigate the vanishing gradient problem, which is common in traditional RNNs. The vanishing gradient problem occurs when the gradients of the loss function become extremely small during backpropagation, hindering the learning process. LSTMs address this issue by maintaining a constant error flow through time, facilitating the training of deep networks on sequential data.

LSTMs have found wide applications in various fields. In natural language processing, they excel in tasks such as machine translation, sentiment analysis, and text generation. In speech recognition, LSTMs can effectively model long-term dependencies in audio sequences. Moreover, LSTMs have been successfully employed in financial forecasting, weather prediction, and other domains where understanding temporal dependencies is critical.

Despite their effectiveness, LSTMs are not without challenges. Training large LSTM networks can be computationally expensive, and selecting appropriate hyperparameters is crucial for optimal performance. In recent years, researchers have also explored more advanced architectures, such as Gated Recurrent Units (GRUs) and Transformer models, as alternatives to LSTMs in certain applications.

In summary, Long Short-Term Memory networks have revolutionized the field of deep learning, particularly in handling sequential data with long-range dependencies. Their ability to capture and retain information over extended periods makes them a powerful tool for a wide range of applications, contributing significantly to the success of deep learning in real-world scenarios.

2.3. Deep learning and Flask

Deep learning and Flask represent a powerful combination for deploying and serving machine learning models in production environments. Deep learning involves the training of complex neural networks on large datasets to enable machines to learn and make predictions. Flask, on the other hand, is a lightweight and versatile web framework for building web applications in Python. When these two technologies are integrated, developers can create web applications that leverage the capabilities of deep learning models.

One common use case for combining deep learning and Flask is deploying a trained neural network as a web service. Once a deep learning model is trained and saved, Flask can be used to create a web application that exposes an API endpoint. This endpoint can then receive input data from users, pass it to the pre-trained deep learning model for prediction, and return the results to the user.

Flask's simplicity and flexibility make it an ideal choice for serving machine learning models through web applications. Developers can design user interfaces, handle user input, and communicate with the deep learning model seamlessly. Flask's modular structure also allows for the easy integration of other components, such as databases or external services, to enhance the functionality of the application.

To deploy a deep learning model with Flask, the model is typically loaded into memory when the Flask application starts. This ensures that the model is readily available to handle incoming requests. Several Python libraries, such as TensorFlow or PyTorch, are commonly used for building and training deep learning models, and these models can be integrated into Flask applications using their respective APIs.

Security considerations are crucial when deploying deep learning models via Flask. Input data should be validated and sanitized to prevent potential security vulnerabilities. Additionally, rate limiting and other security measures should be implemented to protect against abuse or malicious attacks.

In conclusion, the combination of deep learning and Flask offers a powerful solution for deploying and serving machine learning models in web applications. This integration enables the creation of user-friendly interfaces that leverage the capabilities of trained models, making it easier to share and apply deep learning insights in real-world scenarios. As the field of deep learning and web development continues to advance, the synergy between these technologies is likely to play a key role in delivering intelligent and accessible applications.

3. Method

The steps of the process carried out in this study are depicted on a research methodology flow chart as shown in Figure 1.

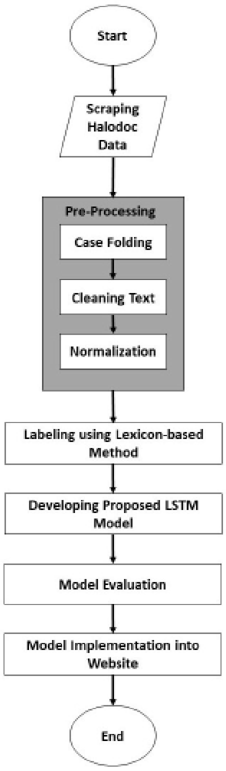

Figure 1. Research methodology

3.1. Data Collection

The process of data collection or crawling is carried out in this study uses scrapping techniques with the help of library google play scraper The use of this library requires several parameters to be entered including url, basha (”lang”), country (”country”), sort (”sort”), and data retrieval restrictions (”count”), which are then entered into the dataFrame and deleted duplicate words.

The data taken is the Indonesian halodoc application review data from the GooglePlay Store. The data collected in this process was 19,830 data, which was then stored into a dataframe.

ISSN 2723-6471

3.2. Pre-processing

The preprocessing process is necessary, because the data that has been collected is raw data and there is a lot of noise. The pre-processing stages carried out include case folding, text cleaning and normalization.

1) Case Folding. Case Folding is a process for converting all document text to lowercase. In this study, the review data collected in the dataset was converted into lowercase.
2) Cleaning Text. Cleaning Text is a method used to remove noise in data such as symbols, punctuation, numbers, spaces, emojis, URLs, and other unique characters.
3) Normalization. Normalization is the process of changing short word forms and slang words, which are not in accordance with Enhanced Spelling (EYD), into standard words according to the Big Indonesian Dictionary (KBBI) [14]

3.3. Data Labeling

The labeling process is carried out by labeling the review data, to find out the sentiment of the review data, whether the review has positive, negative and neutral sentiment. Labeling is done automatically using the lexicon-based method, with the lexicon sentiment dictionary Indonesian taken from Fajri91's github [15]. The results of this process obtained data with positive sentiment as much as 11387, negative sentiment as much as 4502, and neutral sentiment as much as 3334. The process of labeling the data is depicted in a flowchart as Figure 2.

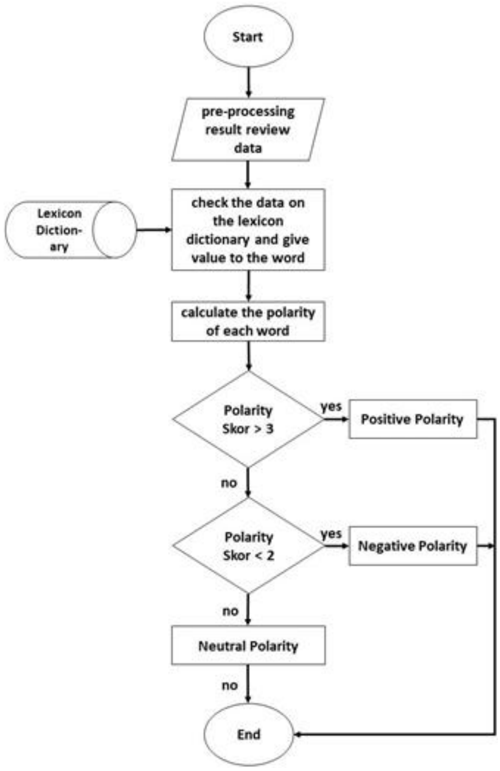

Figure 2. Data Labeling Process

3.4. Data Splitting

Data splitting is done by dividing the data into 90% as training data and 10% as test data. Before data splitting, a random undersampling process was carried out to overcome data imbalances in each class. The data is then fed into a one-hot encoding process to make the dataset categorical, which is necessary because it is used to classify three types of sentiments.

ISSN 2723-6471

3.5. Model Development

3.5.1. Model Design

LSTM Model Architecture is designed by initializing embedding layers as input, in order to find vector data representations in the Word2Vev result array [16, 20]. The results of Word2Vec are then processed into the Bi-LSTM layer which then goes into the fully connected layer. There are three hidden layers on the fully connected layer with Relu activation. After going through the dense layer or output layer with the sigmoid activation function which we can use for classification of 3 sentiments. Model is created and trained using the Tensorflow library [17].

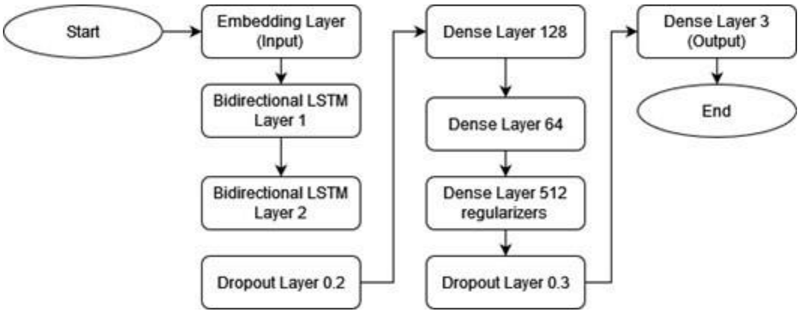

Figure 3. Model Design

3.5.2. Training

The training process is carried out to train a model that has been designed using the data that has been obtained. The parameters initialized in the Bi-LSTM model are dropout, output activation, optimizer, and learning rate. The optimizer used in this study was RMSprop [8, 16] with a learning rate of 0.0001.

3.5.3. Testing

The testing process is carried out after the training process. Testing is carried out to validate the results that have been produced in the training process. Testing is carried out by evaluating accuracy measurements against test data in all combinations. The testing process requires test data and uses all models from the training results on each combination of parameters.

3.6. Model Evaluation

The evaluation process is carried out using a confusion matrix to measure the performance of the model. Confusion matrix or so-called error matrix, provides information comparing the results of classification carried out by a system (model) with the actual classification results [4, 12].

3.7. Application Design

3.7.1. Use Case Diagram.

Use case diagram describes a series of activities performed by users on the system. The use case diagram of this application is illustrated in Figure 4.

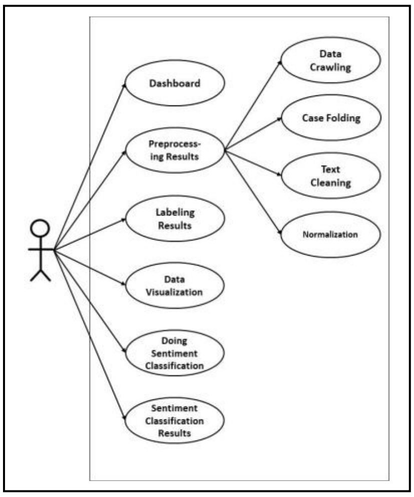

Figure 4. Use Case Diagram

3.7.2. Navigation Structure.

The navigation structure is created to know the relationships and working chains of several different pages. The navigation structure of application is shown in figure 5.

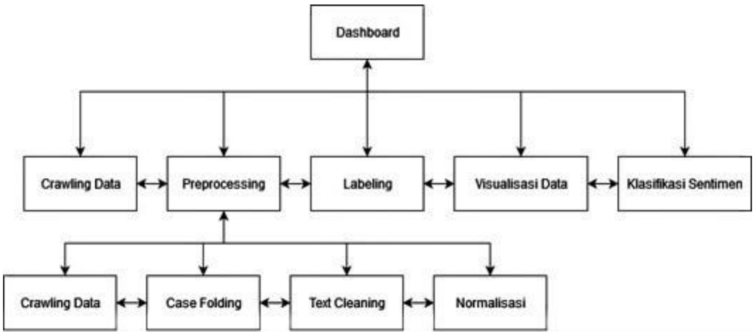

Figure 5. Navigation Structure

3.7.3. Interface Design.

This design is based on the navigation structure that has been created in the previous process. In this design there are 5 pages including dashboard, preprocessing, labeling, visualization, and sentiment classification. A sample of the dashboard page design can be seen in Figure 6.

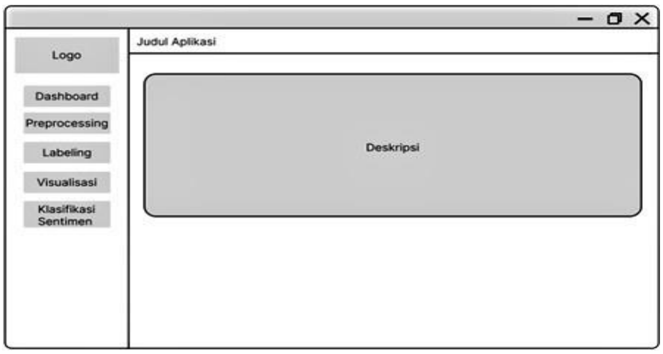

Figure 6. Dashboard Display Design

ISSN 2723-6471

4. Result and Discussion

4.1. Results

This research resulted in an application that can classify reviews. The proposed Bidirectional Long Short Term Memory model showed good enough results for sentiment analysis of Halodoc app reviews.

1) Data Crawling Results. A sample of the data collection can be seen in Table 1.

Table 1. Sampel from Data Collection

No.	Reviews
1	Sayang banget sama halodoc bintang 10 untuk halodoc ♥♥
2	Sangat membantu saya..konsul dokter, resep/ obat tinggal dikirim.. sangat praktis dan bermanfaat. ♠♠
3	Aplikasi cacat, download kartu kendali ga bisa2, bisa nya di web..mana masuknya cuma pake nomor doang.
4	Salut dgn tetap menjaga kualitas pelayanannya. Halodoc solusi untuk menjaga dan merawat kesehatan

2) Preprocessing Results. The results of each prepocessing stage are explained as follows: Table 2 shows the results of the case folding process, Table 3 shows the results of the text cleaning process, Table 4 shows the results of the normalization process.

Table 2. Sample of Case Folding Process

No.	Data before Case Folding	Data after Case Folding
1	Sayang banget sama halodoc bintang 10 untuk halodoc ♡♡	sayang banget sama halodoc bintang 10 untuk halodoc ♡♡
2	Sangat membantu saya..konsul dokter, resep/ obat tinggal dikirim.. sangat praktis dan bermanfaat ♠♠	sangat membantu saya.. konsul dokter, resep/ obat tinggal dikirim.. sangat praktis dan bermanfaat.. ♠♠
3	Aplikasi cacat, download kartu kendali ga bisa2, bisa nya di web..mana masuknya cuma pake nomor doang.	aplikasi cacat,download kartu kendali ga bisa2,bisa nya di web.mana masuknya cuma pake nomor doang
4	Salut dgn tetap menjaga kualitas pelayanannya. Halodoc solusi untuk menjaga dan merawat kesehatan	salut dgn tetap menjaga kualitas pelayanannya. halodoc solusi untuk menjaga dan merawat kesehatan

Table 3. Sample of Text Cleaning Process

No.	Data before Text Cleaning	Data after Text Cleaning
1	Sayang banget sama halodoc bintang 10 untuk halodoc ♡♡	sayang banget sama halodoc bintang untuk halodoc
2	Sangat membantu saya..konsul dokter, resep/ obat tinggal dikirim.. sangat praktis dan bermanfaat ♠♠	sangat membantu saya konsul dokter resep obat tinggal dikirim sangat praktis dan bermanfaat
3	Aplikasi cacat, download kartu kendali ga bisa2, bisa nya di web..mana masuknya cuma pake nomor doang.	aplikasi cacat,download kartu kendali ga bisa bisa nya di webmana masuknya cuma pake nomor doang
4	Salut dgn tetap menjaga kualitas pelayanannya. Halodoc solusi untuk menjaga dan merawat kesehatan	salut dgn tetap menjaga kualitas pelayanannya halodoc solusi untuk menjaga dan merawat kesehatan

Table 4. Sample of the Normalization Process

No.	Data before Normalization	Data after Normalization
1	sayang banget sama halodoc bintang untuk halodoc	sayang banget sama halodoc bintang untuk halodoc
2	sangat membantu saya konsul dokter resep obat tinggal dikirim sangat praktis dan bermanfaat	sangat membantu saya konsul dokter resep obat tingal dikirim sangat praktis dan bermanfat
3	aplikasi cacat download kartu kendali ga bisa bisa nya di web mana masuknya cuma pake nomor doang	aplikasi cacat download kartu kendali tidak bisa bisa di web mana masuknya cuma pakai nomor saja
4	salut dgn tetap menjaga kualitas pelayanannya halodoc solusi untuk menjaga dan merawat kesehatan	salut dengan tetap menjaga kualitas pelayananya halodoc solusi untuk menjaga dan merawat kesehatan

3) Lexicon-Based Labeling Results After the data is processed at the preprocessing step, labeling is then carried out using the Lexicon Based method. The results of this labeling are classified into positive, negative, and neutral sentiment. The table shows the results of the labeling process.

 ISSN 2723-6471

Table 5. Sample of Lexicon Based Labeling Process

No.	Content	Value	Sentiment
1	sayang banget sama halodoc bintang untuk halodoc	7	Positive
2	sangat membantu saya konsul dokter resep obat tingal dikirim sangat praktis dan bermanfat	13	Positive
3	aplikasi cacat download kartu kendali tidak bisa bisa di web mana masuknya cuma pakai nomor saja	-20	Negative
4	salut dengan tetap menjaga kualitas pelayananya halodoc solusi untuk menjaga dan merawat kesehatan	1	Neutral

From Figure 7, it is known that most reviews contain positive sentiment, which is around 59.2%. In other words, the halodoc application is well responded to by its users. The negative sentiment value ranked second while the neutral sentiment value ranked third.

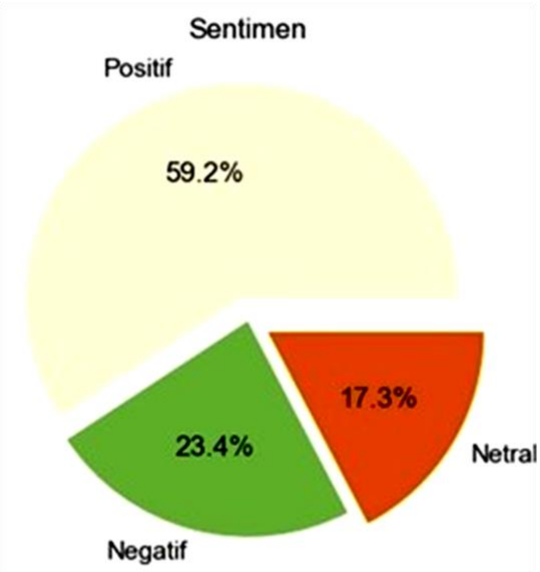

Figure 7. Results of Lexicon-based Labeling

4) Evaluation Results with Confusion Matrix. The results of the evaluation with the confusion matrix prove that the model can perform sentiment analysis quite well. It is proven from the accuracy results obtained by 85.3% with an average precision value of 85.3%, recall value of 85.6% and f1-score of 85.3%.

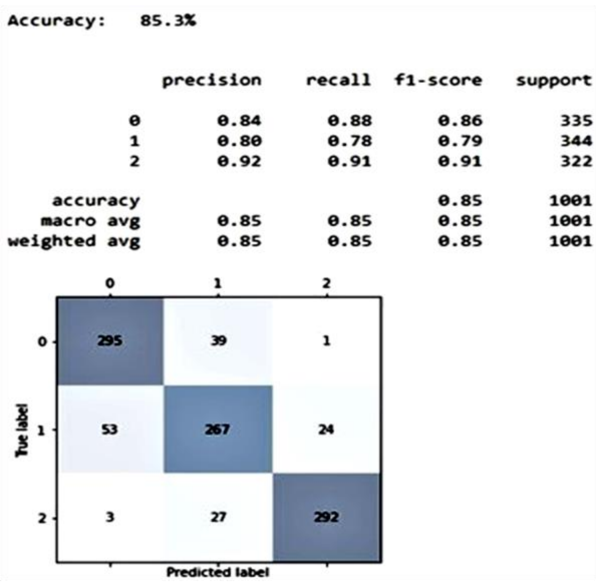

Figure 8. Confusion Matrix

4.2. Result of Developing Web-Based Application

Application development is done using the flask framework. The model used for classification is the proposed LSTM model that has been created in the previous process. The following Figure 9 is a result sample of the proposed LSTM model implementation.

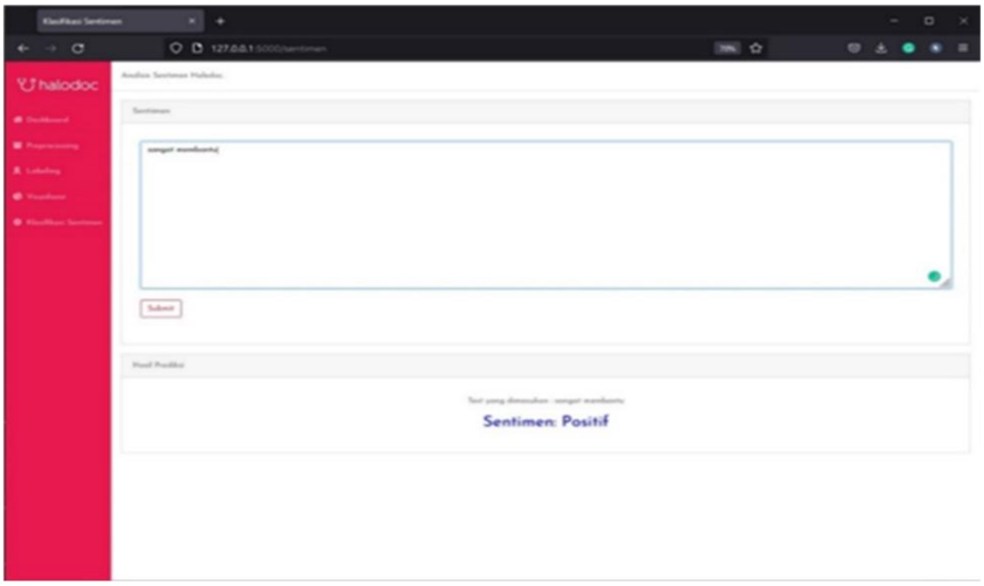

Figure 9. Dashboard Page of the Proposed LSTM Model Implementation

5. Conclusion

Developing web-based applications is able to classify sentiment analysis quite well. In this study, 19,830 review data were used. The classification results obtained at the lexicon-based step resulted in positive sentiment as much as 11387, negative sentiment as much as 4502, and neutral sentiment as much as 3334. Then the evaluation obtained from the proposed LSTM model showed quite good results with an average accuracy value of 85.3%, and an average precision value of 85.3%, a recall value of 85.6% and an f1-score of 85.3%.

6. Declarations

6.1. Author Contributions

Conceptualization: R.R., A.B.M., and R.A.P.; Methodology: A.B.M.; Software: R.R.; Validation: R.R., A.B.M., and R.A.P.; Formal Analysis: R.R., A.B.M., and R.A.P.; Investigation: R.R.; Resources: A.B.M.; Data Curation: A.B.M.; Writing Original Draft Preparation: R.R., A.B.M., and R.R.; Writing Review and Editing: R.A.P., A.B.M., and R.R.; Visualization: R.R.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank the Gunadarma Education Foundation for financial support.

References

[1] M. Lathkar, Building Web Apps with Python and Flask : Learn to Develop and Deploy Responsive RESTful Web Application Using Flask Framework. New Delhi: BPB Publications, 2021.

[2] M. Copperwaite, C. Leifer, Learning flask framework: Build dynamic, data-driven and modern applications with flask. Birmingham: Packt, 2015.

[3] J. Zhao and H. Lu, "Meniscus Effect on the In Vitro Dosimetry of the T25 Flask Under 2.45 and 5.25 GHz Exposures," *in IEEE Microwave and Wireless Components Letters,* vol. 23, no. 2, pp. 105-107, Feb. 2013, doi: 10.1109/LMWC.2013.2240675.

[4] G. Zeng, ”On the confusion matrix in credit scoring and its analytical properties”, *Communications in Statistics - Theory and Methods*, vol. 49, no. 9, pp. 2080-2093, 2020.

[5] W. Li, D. Li, H. Yin, L. Zhang, Z. Zhu, ans P. Liu, ”Lexicon-Enhanced Attention Network Based on Text Representation for Sentimen Classification,” *Applied Science*, vol. 9, no. 18, pp.3717-3729, 2019.

[6] J. Samuel et al., "Feeling Positive About Reopening? New Normal Scenarios From COVID-19 US Reopen Sentiment Analytics," *in IEEE Access,* vol. 8, pp. 142173-142190, 2020, doi: 10.1109/ACCESS.2020.3013933.

[7] J. Patterson and A. Gibson, Deep Learning: A Practitioner’s Approach. O’Reilly Media, Inc., Sebastopol, 2017.

[8] W.K. Sari, D. Palupi, R.F. , dan I.S.B. Ahar, ”Klasifikasi Teks Multilabel pada Artikel Berita Menggunakan Long Short Term Memory dengan Word2Vec,” *Jurnal RESTI (Rekayasa Sistem dan Teknologin Informasi),* vol.4 no.2, pp. 276-285, 2020.

[9] Anonymous, ”Hasil Survei Internet APJII 2019-2020-Q2,” 2020 [online]. Available: https://apjii.or.id/content/read/39/521/ Laporan-Survei-Internet-APJII-2019-2020-Q2 [Accessed: 22 Agustus 2022]

[10] Rianto, A.B. Mutiara, E.P. Wibowo, and P.I. Santosa, ”Improving Stemming Techniques For Non-Formal Indonesian Sentences Using INCORBIZ, ” *ICIC Express Letters*, vol. 15, no. 1, pp. 67-74, January 2021.

[11] Rianto, A.B. Mutiara, E.P. Wibowo, and P.I. Santosa, ”Improving The Accuracy of Text Classifcation Using Stemming Method, A Case of Non-Formal Indonesian Conversation, ” *Journal of Big Data*, vol. 8, no. 26, pp. 2-16, 2021.

[12] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe and A. Ralescu, "Confusion-Matrix-Based Kernel Logistic Regression for Imbalanced Data Classification," *in IEEE Transactions on Knowledge and Data Engineering,* vol. 29, no. 9, pp. 1806-1819, 1 Sept. 2017, doi: 10.1109/TKDE.2017.2682249.

[13] D. Anggraini, A. B. Mutiara, T. M. Kusuma and L. Wulandari, “Algorithm for Simple Sentence Identification in Bahasa Indonesia”, *The 3rd International Conference on Informatics and Computing (ICIC),* Palembang, Indonesia, vol. 1, no. 1, pp.1-6, 2018.

[14] KBBI Daring, https://kbbi.kemdikbud.go.id/ [Accessed on February 20, 2022]

[15] F.Koto, https://github.com/fajri91 [Accessed on April 20, 2022]

[16] A. Nurdin, B.A.S. Aji, A. Bustamin, Z. Abidin, ”Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks,” *Jurnal Teknokompak*, vol. 14, no.2, pp. 74-79, 2020.

[17] Y.R. Yunius, ”Implementasi Algoritma Convolutional Neural Network Dengan Framework Tensorflow Pada Aplikasi Mobile Pendeteksi Penyakit Melanoma Dengan Memanfaatkan Webservice Framework Flask,” *Jurnal Infact*, vol. 1, no. 2, pp. 1-11, Aug.2017.

[18] Y. Goldberg, ”Neural network methods for natural language processing,” *Synthesis lectures on human language technologies*,

ISSN 2723-6471

vol. 10, pp. 1-309, 2017.

[19] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf and M. Funtowicz, ”Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 2020, pp. 38-45.

[20] R.S. Yuwana, E. Suryawati, H.F. Pardede, ”On Empirical Evaluation of Deep Architectures for Indonesian POS Tagging Problem,” *2018 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, pp. 204-208, 2018