International Journal Science and Technology

IJST Vol 2 No. 3 | Nov 2023 | ISSN: 2828-7223 (print), ISSN: 2828-7045 (online), Page 36-48

SENTIMENT ANALYSIS USING LONG TERM MEMORY (LSTM) BOOK CASE STUDY: UNSOLICITED ADVICE FOR MURDERERS BY VERA WONG'S

Tri Sulistyorini[1*], Erma Sova[2], Nelly Sofi[3], Revida Iriana Napitupulu[4]
[1,3]Faculty Industrial Technology, Gunadarma University, Indonesia
[2,4]Faculty of Computer Science and Technology, Gunadarma University, Indonesia

Article History
Received : Nov 2023
Revised : Nov 2023
Accepted : Nov 2023
Published : Nov 2023

Corresponding author*:
tri_s@staff.gunadarma.ac.id

Cite This Article:
Tri Sulistyorini, Erma Sova, Nelly Sofi, and Revida Iriana Napitupulu, "SENTIMENT ANALYSIS USING LONG TERM MEMORY (LSTM) BOOK CASE STUDY: UNSOLICITED ADVICE FOR MURDERERS BY VERA WONG'S ", IJST, vol. 2, no. 3, pp. 36–48, Nov. 2023.

DOI:
https://doi.org/10.56127/ijst.v2i3.995

Abstract: Reading books is one of the most effective ways to reduce stress. In today's digital era, access to finding and buying books is getting easier, so reader reviews are important in choosing books that match interests. However, with a large number of reviews, Natural Language Processing (NLP) with the Long Short-Term Memory (LSTM) method is used to help analyze positive and negative sentiments from many book reviews. This sentiment analysis is useful for readers to evaluate the quality of books, as well as for authors and sellers to find out the opinions of readers and improve the quality of their work. In this study, the book review dataset "Vera Wong's Unsolicited Advice for Murderers" from the Goodreads website is used, which is then divided into training data and validation data with a ratio of 75%: 25%. The Long Short-Term Memory (LSTM) method is used to analyze the sentiment of the reviews. The model architecture built consists of Embedding Layer, LSTM Layer with 128 neuron units, 3 Dense Layer with ReLU activation function, 3 Dropout Layer, and Fully Connected Layer with and Sigmoid activation function, Binary Cross Entropy loss function, and RMSprop optimizer. The model training process was conducted with 30 epochs. The evaluation results show that the model achieved an accuracy of 90%, indicating the model performs relatively well in correctly classifying positive sentiments.

Keyword: Sentiment Analysis, Reviews, Books, Goodreads, LSTM

INTRODUCTION

Reading books can be one way to reduce stress from everyday life. A 2009 study at the University of Sussex found that reading can reduce stress by 68%. Reading is more effective and faster in relieving stress compared to other relaxation methods such as listening to music, going for a walk or drinking a glass of warm tea. In today's digital era, people are facilitated in accessing many things. One of them is the use of online platforms to search and buy products, including books. Not only books in physical form, there are also digital books in electronic or softcopy form.

One important aspect of making a decision before buying a book is reviews or reviews from previous customers or readers. If you don't check the reviews of the book, of course, readers will find it difficult to find a book that suits their tastes. On Goodreads' website, the book "Vera Wong's Unsolicited Advice for Murderers" has received a considerable number of reviews. However, with such a large number of reviews, it can be difficult for readers to read and analyze each review manually. Therefore, technology that method can help to process and analyze reviews into positive or negative sentiment categories.

By analyzing sentiment using the LSTM method can help readers to evaluate whether the book is worth reading or not. This sentiment analysis can also provide valuable benefits for book authors, such as being able to find out the opinions of readers, so that they can evaluate the strengths and weaknesses of their work. Not only that, for booksellers, sentiment analysis can help sellers and book publishers to improve the quality of the books offered. Research on the analysis of positive and negative sentiment in book reviews using NLP is very relevant and important in helping buyers and sellers of books in making informed decisions. This study aims to build a model that can process and analyze positive or negative sentiments.

RESEARCH METHODOLOGY

Reviews are evaluations or ratings given by individuals or groups to a product, service, or work such as books, movies, or music. In the Big Indonesian Dictionary (KBBI), a review is a peel or interpretation or commentary. Meanwhile, according to Waluyo (2016), reviews or reviews or also called review texts are texts that contain considerations or reviews about a book or work. Often, reviews contain opinions, impressions or subjective experiences of users of the product. Reviews can be written in various media, such as online media or print media. However, with the all-digital era like today, reviews are more often found in online media.

Artificial Intelligence (AI)

Artificial Intelligence (AI) is a part of computer science that studies how to create computers and systems have the ability to do works with a level of quality and ability that is equal to or even better than that of humans. According to John McCarthy (1956) AI is to identify and model human thought processes and design machines to be able to mimic human behavior. Using AI can help human work in various fields such as science, business, health, transportation, and etc. Examples of AI applications are machine translators such as paraphrase, speech recognition applications such as Google Assistant or Siri, and others [1].

Natural Language Processing

Natural Language Processing (NLP) is a branch of Artificial Intelligence that aims to make computers able to understand and understand sentences or words written in human language. Natural Language Processing was created to assist users in interacting with computers using everyday language. The develop of NLP has undergone significant improvements including Language modeling [2].

Sentiment Analysis

Sentiment Analysis also called opinion mining is one example a application Natural Language Processing. Sentiment analysis as a provider of information understanding for the general public to analyze different sayings and reviews [3]. This method involves the process of extracting features, such as words or phrases that indicate a particular sentiment, and using statistical models or neural networks to classify that sentiment. Sentiment analysis is very useful in a variety of contexts, especially in the business and social media industries. In addition, sentiment analysis also provides valuable insights into user desires and preferences. With a better understanding of customer sentiment, decisions can be made more precisely and data-driven, and improve the overall user experience.

Recurrent Neural Network

Recurrent Neural Network is a type of deep learning architecture used to process sequential data such as text, audio, and time series. The main difference between Recurrent Neural Networks and other types of architectures is the ability to remember and utilize information from previous data while processing the current data. Recurrent neural networks use memory units called "cells" to store previous information and stream the information to the next step in the data sequence. Recurrent Neural Network, a more flexible model of encoding temporal context in feedback connections system [4]. Thus, when analyzing text, can account for previous words in a sentence to understand context and influence subsequent predictions.

Long Short-Term Memory

A variant of Recurrent Neural Network (RNN), LSTM has more complex structure with memory units called "cells" and "gate" mechanisms that regulate the flow of information within them [5]. Memory cells can store long-term information from sequences of previously processed data. The forget gate regulates when unneeded information is deleted by a cell, allowing the cell to remember new information.

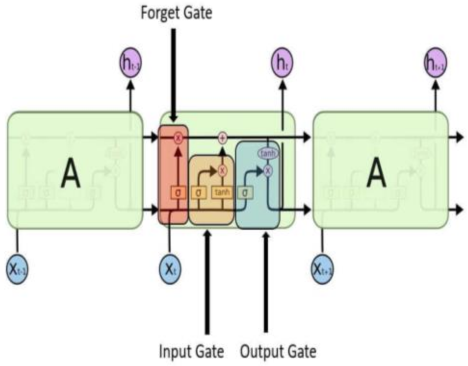

Figure 1. Long Short-Term Memory Architecture

Confusion Matrix

Confusion Matrix is a table use to analyze the performance of classification models. Confusion Matrix will record the number of occurrences true or actual classification values to prediction classifications [6]. This table will give you an idea of how well the model classify data into the correct class. The confusion matrix consists of four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Tabel 1. Tabel Confusion Matrix

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

RESULT AND DISCUSSION

Work Flow

To analyze sentiment towards the review of the book "Vera Wong's Unsolicited Advice for Murderers" from the Goodreads website used the Long Short-Term Memory model which is a type of Recurrent Neural Network [7].

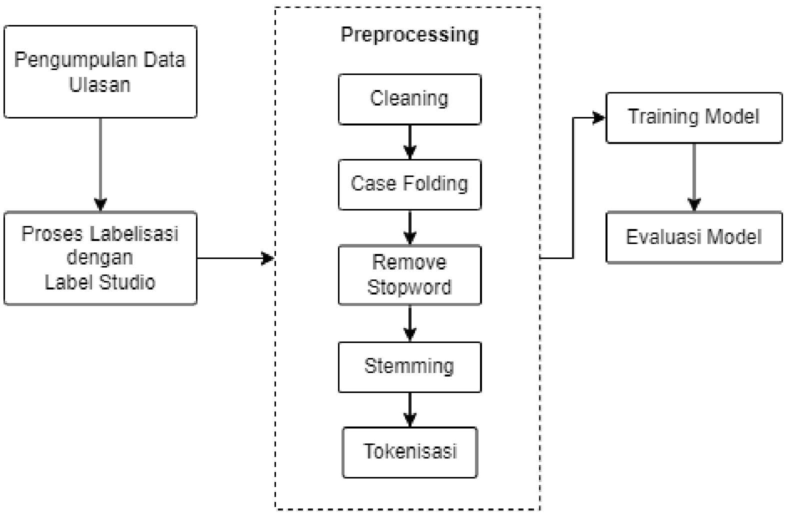

Figure 2. Work Flow

Data Collection

The first step is to collect review data from the Goodreads site, the data that has been collected is then used as a dataset. Then the dataset is labeled with positive and negative sentiments. This labeling process is to determine sentiment of each review. After that, the data preprocess stage is carried out to prepare unstructured data. The preprocessing process involves several stages, specially data cleaning, case folding, remove stopword, stemming and tokenization. Next The dataset will be divided into training data and validation data [8]. Finally, the classification results will be evaluated by looking at the accuracy value to assess the performance of the model. In data from Goodreads.com sites were used in the form of reader reviews from the book "Vera Wong's Unsolicited Advice for Murderers". Data is collected manually by copying reviews. Then, the data is saved in a csv extension file.

Data Labeling

Review data will be labeled based on the sentiment contained in the text. Labeling is categorized into two categories: positive and negative. This is done by taking into account the connotation of the word and the context of the review [9]. Connotation refers to the emotional meaning or values associated with a word, beyond its denotative meaning. Connotations can vary depending on the specific context and culture or even depend on the collective understanding and interpretation of the people who use the language. Therefore, labeling is done by following the following rules:

1) Positive
   a. There are words that have positive meanings, such as fun, love, good, enjoy, cute.
   b. If the review has positive context.
2) Negative
   a. There are words that mean negative, such as bad, boring, hate.
   b. If the review has negative context.

The results of this labeling are used as a reference for the model to study patterns and perform sentiment analysis on unlabeled data, with the aim of producing accurate predictions. Labeling data is done using the Label Studio program. The program will run a web server with a website display that can be accessed through a browser using a URL http://localhost:8080.

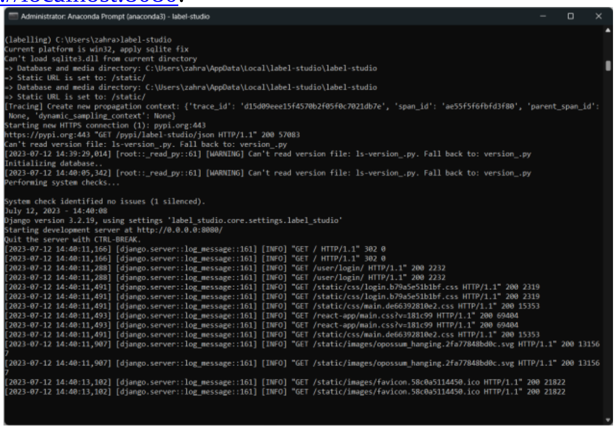

Figure 3. Running Label Studio

After logging into the site, create a new project to label the dataset as shown in Figure 4

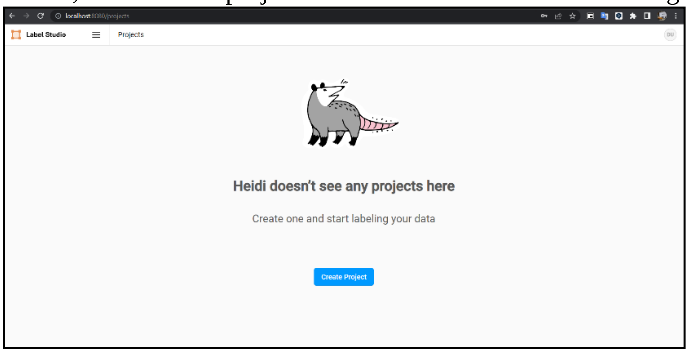

Figure 4. Create a New Project

Give the project a name to be created as shown in Figure 5.

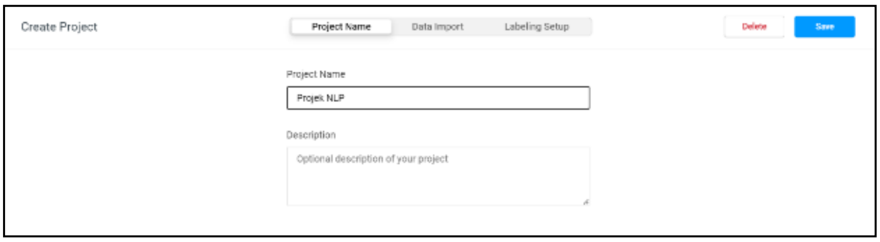

Figure 5. Name the Project

The collected dataset is uploaded into the project to be labeled as in Figure 6

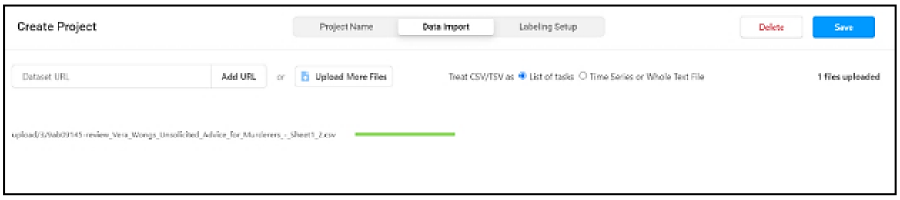

Figure 6. Uploading the Dataset into Project

The collected dataset is uploaded into the project to be labeled as in Figure 7.

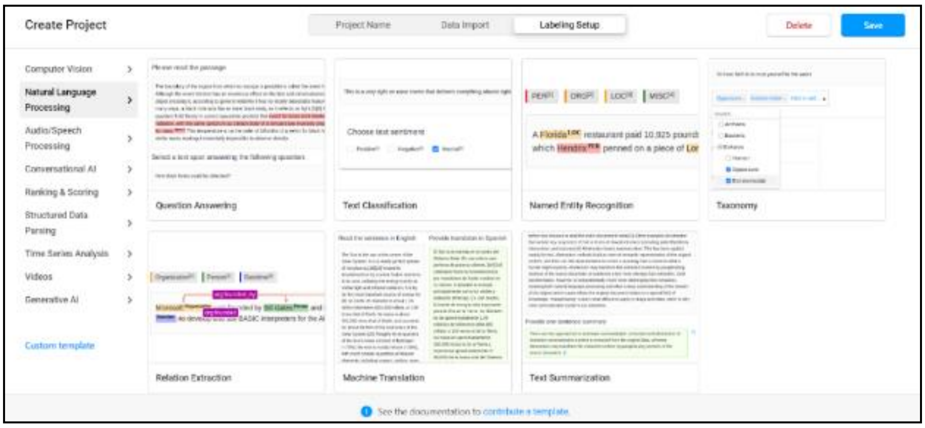

Figure 7. Select the labeling type

Next will be redirected to the Configure data page, select the column containing the text you want to label, then save the project.

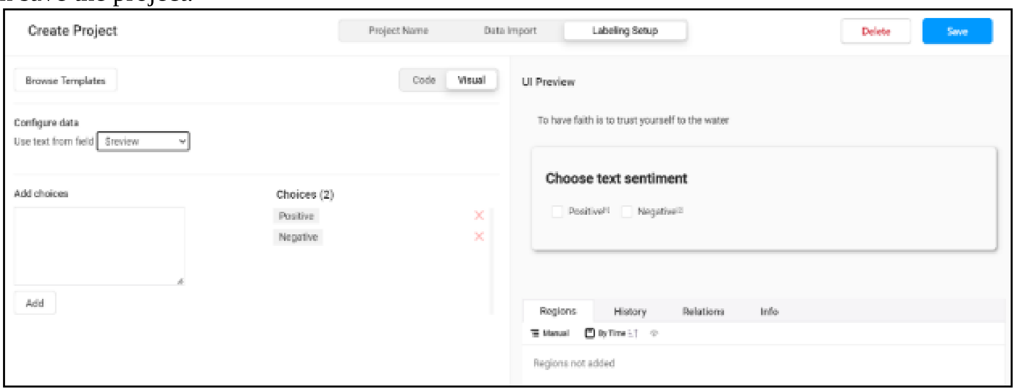

Figure 8. Select the Columns You Want to Label

After the project is saved, the uploaded dataset will appear, then select Label All Tasks and start labeling the data with positive or negative sentiment. If all data is already labeled, then select Submit to save it.

Figure 9. Labeling on Data

Then Export dataset to download the labeled dataset. An example of labeling results is shown in Table 2.

Table 2. Example of Dataset Labeling Results

REVIEW	SENTIMENT
"SO MUCH LOVE FOR VERA!! This was a delight. A funny and heartwarming story of murder, found family, friendships, and FOOD. I hope this isn't the last we see of Vera and her tea shop! ARC Provided by NetGalley"	Positive
Not what I expected but a very heartwarming read. In some ways it reminded me of a man called Ove. Something about found family trope just always works for me	Positive
Sweet, funny story. But not my cup of tea.	Negative

From the labeling results, data were obtained as many as 204 positive value reviews and 131 negative value reviews in Figure 10

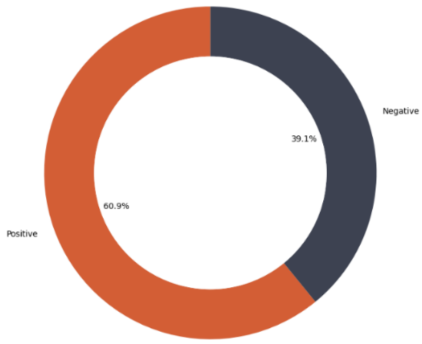

Figure 10. Distribution dataset

Labeling process for this dataset is not carried out by linguists in their fields, so consequently there are some comment sentiments that do not match the existing reviews.

Preprocessing Data

Data preprocessing aims to clean, transform, and organize data for analysis, so data preprocessing is an important stage data analysis process. Stages of data preprocessing such as cleaning data of irrelevant information, changing capital letters to lowercase, removing meaningless common words, breaking data into smaller units such as words or sentences, and converting words into basic forms or root words [10].

Case Folding

The first stage is case folding, which is the process of converting all letters in data into lowercase. The goal is to equalize the representation of letters in the text, so that the difference in uppercase and lowercase letters does not affect the further analysis process. With case folding can avoid ambiguity in words that actually have the same meaning but are written in different letters. To perform case folding, the lower function of Python is used [11]. The results of the case folding process on Figure 3.10

	review	sentiment
0	sixty-year old, vera wong, is a widow, mother ...	Positive
1	jesse q. sutanto's fascinating new cozy myster...	Positive
2	i was a little suspect to begin this novel as ...	Positive
3	sixty-year-old vera wong zhuzhu owns and runs ...	Positive
4	wanted to love this but it was pretty unsubsta...	Negative

Figure 11. Dataset after case folding

Data Cleaning

Data cleaning is the process of cleaning data from punctuation, special characters, numbers, symbols, excessive spaces, and characters that repeat two or more. The purpose of this process is to ensure the quality of the data used in the analysis is well structured, so that the analysis process can be carried out better and efficiently. The regex (Regular Expression) library from Python is used to remove punctuation, numbers, excessive spaces and other string manipulations. In addition, to replace emoji into appropriate words, you can use the emoji library by using the demojize method [12]. A comparison of the original data with the cleaned data in Table 3.

Table 3. Comparison of Data Cleaning Results

BEFORE	AFTER
I'll read anything Jesse Sutanto writes 😍	i ll read anything jesse sutanto writes smiling_face_with_heart eyes
this was so wholesome 😭 i love all the characters so much 😭 i laughed so hard, and cried too. a mystery making you feel so many emotions!! ahhh i love it so much jesse sutanto can do no wrong. all hail vera!	this was so wholesome loudly_crying_face i love all the characters so much loudly_crying_face i laughed so hard and cried too a mystery making you feel so many emotions ahh i love it so much jesse sutanto can do no wrong all hail vera

Stopword

Stopwords are common words that are less meaningful. These common words are usually like "the", "is", "and", "in", and so on. The removal of stopwords leaves words that have significance and relevance, which can help reduce the dimensions of the data and improve the quality of text analysis. To remove stopwords, you can use the stopwords dictionary provided by the NLTK (Natural Language Toolkit) library [13]. A comparison of the original data with the data that has been removed stopwords in Table 4.

Table 4. Comparison of Stopword Removal Results

BEFORE	AFTER
Cute & cosy mystery story that is perfect for for a lazy afternoon	cute cosy mystery story perfect lazy afternoon

Stemming

Stemming is the process of turning words into their basic form. In English, words often have a variety of forms, such as verbs with the suffix "-ing" or "-ed", nouns with the suffix "-s" or "-es", and so on. This process removes the suffix or prefix of the word to get the basic form of the word. For example, the words "running" and "runs" will be converted into their basic form "run". To do stemming, you can use the SnowballStemmer module from the NLTK library which implements the snowball stemming technique [14]. Comparison of original data with data that has been stemmed in Table 5.

Table 5. Comparison of Stemming Results

BEFORE	AFTER
I loved this book! It was cute and smart and made me laugh out loud a few times. I listened on audible and highly recommend. It's a 4.5 for me!	love book cute smart made laugh loud time listen audibl high recommend

Tokenization

Tokenization is the process of breaking text into smaller units called tokens. A token a specific word, phrase, or symbol that has meaning in the context of the language. To perform this process, the Tokenizer function of the TensorFlow library is used [15]. The Tokenizer function converts text into sequences and builds a word index. A comparison of the original data with tokenized data in Table 6.

Table 6. Comparison of tokenization results

BEFORE	AFTER
thoroughly enjoyable read vera is a force of nature and i really loved the descriptions of all the teas she mixes and brews there is palpable joy in reading about the small family she inadvertently assembles around her	['thoroughly', 'enjoyable', 'read', 'vera', 'is', 'a', 'force', 'of', 'nature', 'and', 'i', 'really', 'loved', 'the', 'descriptions', 'of', 'all', 'the', 'teas', 'she', 'mixes', 'and', 'brews', 'there', 'is', 'palpable', 'joy', 'in', 'reading', 'about', 'the', 'small', 'family', 'she', 'inadvertently', 'assembles', 'around', 'her']

Data Preprocessing Results

After going through the preprocessing process, wordcloud is used to visualize the most words in the data. In the wordcloud results, it can be seen that the word "vera" is the word that appears the most in both positive and negative sentiment categories, followed by words related to book reviews. Figure12 shows wordcloud results from a positive review. Worldcloud's results from the negative review are shown in figure 13.

Figure 12. Word Cloud Positive Review

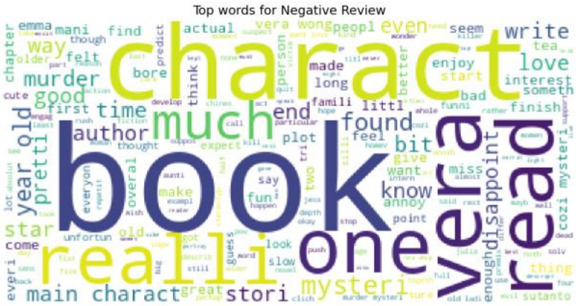

Figure 13. Word Cloud Negative Review

Model Training

Before entering the model training stage, target encoding, word embedding with GloVe, and data sharing will be carried out. Next is to make a model, for this research will be used LSTM. The model will be trained with a previously prepared dataset[16]

Target Encoding

Target encoding is the process of converting labels from word form to numerical form. This process is done because the model can only process numerical data, so it is necessary to encode the label. To perform encoding, the LabelEncoder class from the scikit-learn library is used [17]. The encoding results will be stored in a new column, specially sentiment_encoded, where positive sentiment is represented with the number 1 and negative sentiment is represented by the number 0. The results in Figure 14.

	review	sentiment	sentiment_encoded
0	sixti year old vera wong widow mother owner sm...	Positive	1
1	jess q sutanto fascin new cozi mysteri vera wo...	Positive	1
2	littl suspect begin novel featur yet anoth old...	Positive	1
3	sixti year old vera wong zhuzhu own run vera w...	Positive	1
4	want love pretti unsubstanti throughtout didnt...	Negative	0

Figure 14. Target Encoding Results

Pad Sequence and Glove Methods.

The pad sequence method is used to convert text into a sequence of words of equal length, so that the input data has fixed dimensions. In a sequence pad, text shorter than the specified length will be filled with the specified value, usually using zeros, while the longer text will be truncated to fit the specified length. The maximum length value is determined from the longest sentence in the dataset and uses the padding type post, so padding will be added at the end of the sentence to equalize the length [18]. The results of the pad sequence in Figure 15.

```
array([[ 409,   42,   19, ...,    0,    0,    0],
       [  64,  125,   39, ...,    0,    0,    0],
       [  47,   23,  122, ...,    0,    0,    0],
       ...,
       [  19,  387, 2252, ...,    0,    0,    0],
       [ 119,  139,  573, ...,    0,    0,    0],
       [   2, 1161,  112, ...,    0,    0,    0]], dtype=int32)
```

Figure 15. Sequence Pad Results

GloVe (Global Vectors for Word Representation) Method

Used after the pad sequence method to provide a vector representation each word that describes the relationship between words based on their occurrence in the text. The GloVe Embedding data used to obtain the embedding vector was obtained from Kaggle. The result of this process is matrix embedding [19]. If a word can be found in the GloVe data, the embedding vector for that word will be inserted into the corresponding row. However, if the word is not found in the GloVe data, it will be represented with zero. The Embedding Matrix in Figure 16.

```
array([[ 0.        ,  0.        ,  0.        , ...,  0.        ,
         0.        ,  0.        ],
       [ 0.28591001, -0.18667001, -0.11585   , ..., -0.29989001,
        -0.94976002, -0.15494999],
       [-0.19744   ,  0.44830999,  0.13688999, ..., -0.56967998,
         0.0015374 ,  0.66600001],
       ...,
       [ 0.20458999,  0.40351   ,  0.052177  , ...,  0.43007001,
         0.66615999,  0.32233   ],
       [ 0.        ,  0.        ,  0.        , ...,  0.        ,
         0.        ,  0.        ],
       [ 0.8319    ,  0.17497   ,  0.075066  , ...,  0.11735   ,
        -0.16799   ,  0.40340999]])
```

Figure 16. Word Embedding Results with GloVe

Data Sharing

This process is carried out to divide the dataset into training data and validation data. The process of sharing data is necessary to test model performance and prevent overfitting [20]. Sharing data can help evaluate the model objectively. The dataset is divided by a ratio on 75% for training data and 25% for testing data. Data training is used to train models and data testing to test model performance on never-before-seen data. Data sharing is done using the train_test_split function of the scikit-learn library.

Data Training Phase

Long Short-Term Memory (LSTM) is used to build optimal models. LSTM network architecture consists of several main layers[21].

Embedding Layer

This layer is used to convert the input text into a more concise and dense vector representation. The parameters used are input_dim is the dictionary size (number of words in the corpus), output_dim is the dimension of the embedding vector, weights is the pre-trained embedding matrix, in this study GloVe is used, and input_length is the specified input length.

LSTM Layer

The LSTM Layer consists of 128 LSTM units with 0.2 recurrent dropouts. LSTM Layer is used to process sequence data such as text and understand the temporal context in that text. Using a "Bidirectional" approach, LSTM can process text sequences in two directions which can improve its performance in some situations.

Dense Layer

The dense layer is a neural network layer that has neurons fully connected to the previous layer. In making this model used three Dense Layers. In the first two layers, the ReLU activation function is worn. Meanwhile, the last layer, the sigmoid activation function (activation='sigmoid') is worn.

Droput Layer

Droput Layer is useful to reduce overfitting in the model and speed up the learning process. The model is built using three Dropout Layers.

Fully Connected Layer

The Fully Connected Layer is the last layer in the model. This layer has several parameters, namely the number of units, the activation function, and the loss function. This layer serves to connect each unit of the previous layer and perform the final calculation to produce the output of the model.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 349, 100)          225900

 bidirectional_1 (Bidirecti  (None, 349, 256)          234496
 onal)

 global_max_pooling1d_1 (Gl  (None, 256)               0
 obalMaxPooling1D)

 dropout_3 (Dropout)         (None, 256)               0

 dense_3 (Dense)             (None, 64)                16448

 dropout_4 (Dropout)         (None, 64)                0

 dense_4 (Dense)             (None, 32)                2080

 dropout_5 (Dropout)         (None, 32)                0

 dense_5 (Dense)             (None, 1)                 33

=================================================================
Total params: 478957 (1.83 MB)
Trainable params: 478957 (1.83 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 17. Architectural Information Model

Model training used a number of neurons of 128 units at the LSTM layer with an epoch count of 30. From the results of model performance during training on training data and validation data which in Figure 3.26, an average time of 13 to 18 seconds was obtained to complete one epoch. The accuracy value in the training data reached 99.60%, while in the validation data it reached 90.48%. The loss value in the training data at 0.0677 and the loss value in the validation data at 0.5685.

```
Epoch 20/30
8/8 [==============================] - 15s 2s/step - loss: 0.1843 - accuracy: 0.9363 - val_loss: 0.4815 - val_accuracy: 0.8214
Epoch 21/30
8/8 [==============================] - 15s 2s/step - loss: 0.1780 - accuracy: 0.9442 - val_loss: 1.2820 - val_accuracy: 0.8095
Epoch 22/30
8/8 [==============================] - 15s 2s/step - loss: 0.3090 - accuracy: 0.9243 - val_loss: 0.4499 - val_accuracy: 0.8214
Epoch 23/30
8/8 [==============================] - 17s 2s/step - loss: 0.1756 - accuracy: 0.9522 - val_loss: 0.3770 - val_accuracy: 0.8571
Epoch 24/30
8/8 [==============================] - 13s 2s/step - loss: 0.1431 - accuracy: 0.9602 - val_loss: 0.3763 - val_accuracy: 0.8929
Epoch 25/30
8/8 [==============================] - 15s 2s/step - loss: 0.2040 - accuracy: 0.9283 - val_loss: 0.3990 - val_accuracy: 0.8810
Epoch 26/30
8/8 [==============================] - 15s 2s/step - loss: 0.1012 - accuracy: 0.9801 - val_loss: 0.5857 - val_accuracy: 0.8810
Epoch 27/30
8/8 [==============================] - 15s 2s/step - loss: 0.0998 - accuracy: 0.9801 - val_loss: 0.4013 - val_accuracy: 0.8810
Epoch 28/30
8/8 [==============================] - 15s 2s/step - loss: 0.1059 - accuracy: 0.9641 - val_loss: 0.6353 - val_accuracy: 0.8929
Epoch 29/30
8/8 [==============================] - 13s 2s/step - loss: 0.1272 - accuracy: 0.9641 - val_loss: 0.4422 - val_accuracy: 0.9167
Epoch 30/30
8/8 [==============================] - 13s 2s/step - loss: 0.0677 - accuracy: 0.9960 - val_loss: 0.5685 - val_accuracy: 0.9048
```

Figure 18. Data Training Results

The results of model training are also shown in the accuracy graph and loss chart which in Figure 3.18

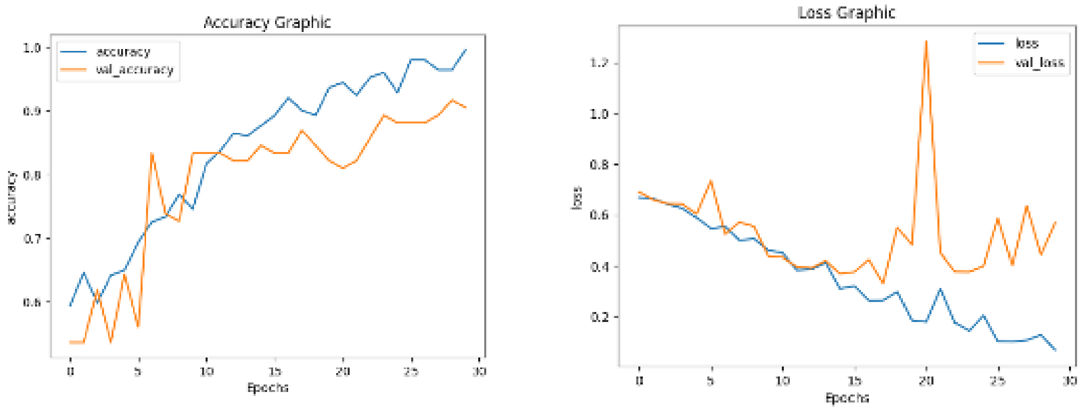

Figure 19. Accuracy Charts and Loss Charts

Based on the graph, there is still overfitting around the 20th epoch on training data accuracy and data validation accuracy.

Model Evaluation

Model evaluation is carried out on test data using confusion matrix. A heatmap plot from the Seaborn library is used to show the results of the confusion matrix [22].

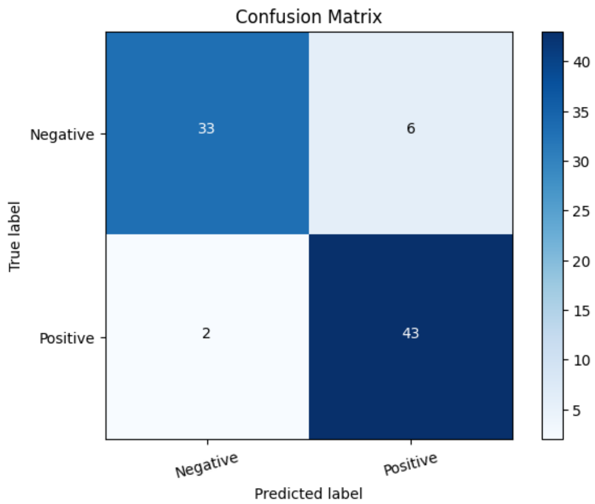

Figure 20. Confusion Matrix

From the results of the confusion matrix in Figure 3.19 it can be concluded that there are:

True Positive as many as 43 data

1.1.1.1.1.1.1 True Positive as many as 33 data

1.1.1.1.1.1.2 False Positive as many as 6 data

1.1.1.1.1.1.3 False Negative as many as 2 data

Because there are more positive sentiment data compared to negative sentiment data, the model tends to incorrectly predict negative sentiment data as positive sentiment data as seen from the higher number of False Positives which is 6 data compared to False Negative which is 2 data. Obtaining the results of the confusion matrix, the classification_report function from Scikit-learn is used calculate accuracy, precision, recall, and f1-score. The calculation results in Figure 3.20 which shows a fairly high accuracy value.

```
3/3 [==============================] - 1s 217ms/step
              precision    recall  f1-score   support

    Negative       0.94      0.85      0.89        39
    Positive       0.88      0.96      0.91        45

    accuracy                           0.90        84
   macro avg       0.91      0.90      0.90        84
weighted avg       0.91      0.90      0.90        84
```

Figure 21. Model Evaluation Classification Report from Model Evaluation Results

CONCLUSION AND SUGGESTION

An experiment was conducted on the use the LSTM method to analyze sentiment, it was concluded that the LSTM method was applied to the model to analyze sentiment using data from the Goodreads site. The model built consists of embedding Layer, LSTM Layer, 3 Dense Layer with ReLU activation function, 3 Dropout Layer, and Fully Connected Layer with Sigmoid activation function, Binary Cross Entropy loss function and RMSprop optimizer. The model performed well using 128 LSTM neurons. Accuracy in training data was obtained by 99.60% and in validation data by 90.48% with 30 epochs. From the results of the model evaluation, a model accuracy value of 90% was obtained. Precision for positive is 88% and negative is 94%. Recall for positive is 96% and negative is 85%. F1-Score for positive is 91% and negative is 89%. Based on these results, that the model has a relatively acceptable performance to correctly classifying positive sentiments. In addition, the results also showed that Vera Wong's Unsolicited Advice for Murderers had a positive review with a rating of 4.08 on the Goodreads website.

The suggestion that can be done for future research is to use more datasets so that the classes in the sample are balanced and can produce better accuracy values. In addition, it can be modified to the model architecture, using other word embedding such as Word2Vec or FastText, or using other sentiment analysis methods.

REFERENCES

[1] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[2] Y. Belinkov and J. Glass, "Analysis Methods in Neural Language Processing: A Survey", doi: 10.1162/tacl_a_00254/1923061/tacl_a_00254.pdf.

[3] Q. T. Ain *et al.*, "Sentiment Analysis Using Deep Learning Techniques: A Review," 2017. [Online]. Available: www.ijacsa.thesai.org

[4] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, "An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting," May 2017, doi: 10.1007/978-3-319-70338-1.

[5] D. Norris, "Short-term memory and long-term memory are still different," *Psychol Bull*, vol. 143, no. 9, pp. 992–1009, Sep. 2017, doi: 10.1037/bul0000108.

[6] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[7] D. Makowski, "The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science," *The Journal of Open Source Software*, vol. 3, no. 22, p. 470, Feb. 2018, doi: 10.21105/joss.00470.

[8] G. Schuh, P. Jussen, and T. Harland, "The Digital Shadow of Services: A Reference Model for Comprehensive Data Collection in MRO Services of Machine Manufacturers," in *Procedia CIRP*, Elsevier B.V., 2018, pp. 271–277. doi: 10.1016/j.procir.2018.03.318.

[9] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14749

[10] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A survey on security threats and defensive techniques of machine learning: A data driven view," *IEEE Access*, vol. 6. Institute of Electrical and Electronics Engineers Inc., pp. 12103–12117, Feb. 12, 2018. doi: 10.1109/ACCESS.2018.2805680.

[11] R. Annisa, I. Surjandari, and Zulkarnain, "Opinion mining on mandalika hotel reviews using latent dirichlet allocation," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 739–746. doi: 10.1016/j.procs.2019.11.178.

[12] V. N. Gudivada, A. Apon, and J. Ding, “Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations.” [Online]. Available: www.iaria.org

[13] I. A. El-Khair, “Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study.” [Online]. Available: http://www.unine.ch/info/clef

[14] A. Sithole, E. T. Chiyaka, P. McCarthy, D. M. Mupinga, B. K. Bucklein, and J. Kibirige, “Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges,” *Higher Education Studies*, vol. 7, no. 1, p. 46, Jan. 2017, doi: 10.5539/hes.v7n1p46.

[15] R. Heines and R. Jung, “The Tokenization of Everything: Towards a Framework for Understanding the Potentials of Tokenized Assets,” 2021. [Online]. Available: https://aisel.aisnet.org/pacis2021/40

[16] J. L. Whittaker *et al.*, “Imaging with ultrasound in physical therapy: What is the PT’s scope of practice? A competency-based educational model and training recommendations,” *Br J Sports Med*, vol. 53, no. 23, pp. 1447–1453, Dec. 2019, doi: 10.1136/bjsports-2018-100193.

[17] Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, and J. Glass, “What do Neural Machine Translation Models Learn about Morphology?,” Apr. 2017, doi: 10.18653/v1/P17-1080.

[18] E. Wallace, Y. Wang, S. Li, S. Singh, and M. Gardner, “Do NLP Models Know Numbers? Probing Numeracy in Embeddings,” Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.07940

[19] E. Hindocha, V. Yazhiny, A. Arunkumar, and P. Boobalan, “Short-text Semantic Similarity using GloVe word embedding,” *International Research Journal of Engineering and Technology*, 2019, [Online]. Available: www.irjet.net

[20] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, “MeDShare: Trust-Less Medical Data Sharing among Cloud Service Providers via Blockchain,” *IEEE Access*, vol. 5, pp. 14757–14767, Jul. 2017, doi: 10.1109/ACCESS.2017.2730843.

[21] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, “A survey on security threats and defensive techniques of machine learning: A data driven view,” *IEEE Access*, vol. 6. Institute of Electrical and Electronics Engineers Inc., pp. 12103–12117, Feb. 12, 2018. doi: 10.1109/ACCESS.2018.2805680.

[22] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.12808