ANALYSIS SENTIMENT ON AIRLINE CUSTOMER SAISFACTION USING RECCURENT NEURAL NETWORK

Astriyer J. Nahumury, Danny Manongga, Ade Iriani
Universitas Kristen Satya Wacana, Indonesia
Email: astriyer08@gmail.com, dmanongga@gmail, adeiriani@gmail.com

ARTICLE INFO	ABSTRACT
	When talking about customer satisfaction, Twitter as a large and great media could be used to get sentiment or opinion on a product and service of a business. The sentiment will be in a form of tweet that was posted on Twitter that referred to hot debated issues subjectively. The tweet data then will be processed using machine learning to analyze the sentiment of a certain topic. This study aimed to analyze the sentiment of Indonesian public on one of the Indonesian airlines using Deep Learning, Recurrent Neural Network (RNN) method based on the training for Long Short-Term Memory (LSTM), validation and prediction. The tweet will be selected in the span of three years (2017-2020) through the triangulation sentence sentiment process. The LSTM model gives a result of 98.5% accuracy and 92.2% validation accuracy in the data training. Whereas, the LSTM model's data testing gives a result of 56.5% negative sentiment higher than the positive and neutral sentiment. It could be assumed that the factors which affect the negative sentiment could be used as an input to improve any business process.
KEYWORDS:	
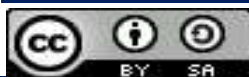
INTRODUCTION

Customer satisfaction is a measure how well a company's product, service or experience has met customers' expectations (Kasiri, Cheng, Sambasivan, & Sidin, 2017). Customer satisfaction ties with the general to specific psychological from the customers' experience on the product and service from the company, meaning customer's feedbacks will be affected by customers' sentiment, emotions.

Needed to be known that customer satisfaction is the key for business to stand in a long-range along with establishing the quality and service of the company (Athiyah, 2016). Business that has a high level of customer satisfaction would also have a high quality service (Kasiri et al., 2017). For companies such as airlines, questionnaires that was whether done traditionally by handing out flyers or done by online is an often-used method in order to see the customer feedback. The traditionally done customer satisfaction or sentiment analysis data collection was seen to be easy but is lacking as lots of respondents gave false and irrelevant answers which makes the data invalid and was wasted during the data cleansing. On the other hand, Twitter was used by airlines as a mine to collect the customer's sentiment regarding the products and services as it was instant and reliable.

According to Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) survey in 2018, Indonesia's internet infiltration reached 171,17 million from 264,16 million people that is equivalent to 64,8% of population in Indonesia. Which is the major reason on the use of social media that reach the approximate 18.19%, namely Facebook, Instagram, twitter, and many other (Sundari, 2019). Correspond to Country Industry Twitter in Indonesia that claimed Indonesia has one of the most active users and is one of the countries with the largest growth of Twitter users (Clinten & Nistanto, 2019). The increase growth of Twitter users makes social media data analyzing evolve and was merged with the field of study such as Social Network Analysis, multimedia management, social media analytic, and sentiment or opinion mining (Cambria, Olsher, & Rajagopal, 2014). Twitter contains tweets or messages that are personal or messages that was influenced by public statements or recent talks events (Boerman & Kruikemeier, 2016). Data that was retrieved from those tweets then would be used in opinion mining or sentiment analysis. Related to the airline companies that was interested in customer feedback to know what are their rate quality of the products and services of the said airlines? was the responses positive or negative towards their products and services? would the customer recommend their products and services to other possible customer?

Data that will be used for this journal is the data from the tweets that will be used in the process of analysis sentiment is tweets relevant to the service of one of the most popular airlines in Indonesia (xy airline). The airline has a positive sentiment or was favored because of the affordable price with route that almost cover all of the region in Indonesia and even got the Operational Safety Audit (IOSA) certification. However, there were also some negative sentiment from the customers as to the delay flights, free baggage removal, and airplane accidents (Arenggoasih & Wijayanti, 2020). In the tweets that was voiced by the airline customers textual data (knowledge) could be found and extracted (text mining) and analyzed using tools (He et al., 2013; Javed & Muralidhara, 2018) so that customers' ideas and sentiment in correlation to customer satisfaction could be seen. Sentiment analysis was done to analyze and describe problems faced by the customers as it was focused on analyzing and understanding the emotions from text pattern review (Gajakosh & Jayaraj, 2015).The sentiment research on airlines have been one with Naïve Bayes's method and Information Gain feature selection using Support Vector Machine (SVM) method through Twitter or flight ticket buying

websites (Prasetiarini, 2020). The result of the two research displays the accuracy of the used learning machine.

Neural Network Deep Learning will be used as a tool to build Machine learning in the sentiment analysis process approach for this study. Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) is ideal to be applied for text classification aspect or xy airline sentiment (Miller et al., 2017). Furthermore, LSTM is more advance in analyzing emotion in a long sentence and for that multi-classification for text emotional attributes will be used as a LSTM language model (Yang et al., 2018). Explicitly, RNN-LSTM has a higher accuracy value compared to other machine learning like Vector Machine K – Nearest Neighbor, Naïve Bayes dan Decision Tree (Wazery, Mohammed, & Houssein, 2018).

With that, this sentiment analysis case study on one of the airlines in Indonesia aimed to know how the products and service consumer in Indonesia's sentiment (customer feedback) is. For the analysis on customer satisfaction to be done successfully and to get what factors are affecting the sentiment of the customer, Deep learning, RNN based on LSTM will be used as an approach to analyzed the sentiment classification. This study supports the two main findings to be aimed, (i) the process of RNN, LSTM that was used in the twitter data text classification that choose a specific sample in Indonesia to generate the accuracy sentiment value. (ii) Machine learning that has been built could be generalized with inputting new data so that complement and complain (customer feedback), the variety of sentiment in a specific period of time could be identified.
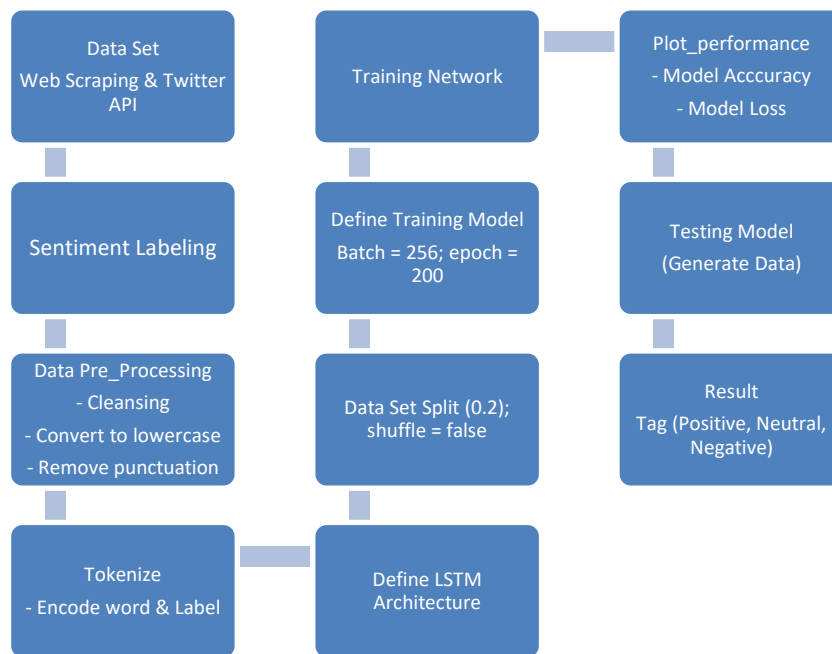
RESEARCH METHOD

The process of the research will be depicted in Picture 1, process in the RNN stage in the data set sentiment that is implemented in the LSTM architecture.

Data Set & Sentiment Labeling

Preparing the data set and sentiment labeling was a long and tough process in this research. The data was collected by selecting the tweet through web scarping using the searched keyword on Twitter. Tweet related to xy airline will be taken from January 1st 2017 to January 1st 2020. The three years' time range was taken with the purpose to avoid any biased to only one incident. If there were by chance happened a big incident, the tweet (sentiment) would also increase (Tsolmon, Kwon, & Lee, 2012). The tweet that has been selected will then be processed in the pre-cleansing to get the tweet data set sentiment by the amount of 27.462 tweet. The process intent to clean the irrelevant data set, like tweet from an account that sells beauty products and uses the #xyairline hashtag to promote their products, which in other word oust the tweet that do not have any sentiment, opinions, and information related to xy airline. However, news accounts will not be ousted as it was part of the sentiment actor. News and personal account are both have the same status in twitter which are the actors of people in the online community (Tomasoa, Iriani, & Sembiring, 2019).

The collected data set will the be labelled with sentiment positive, neutral, and negative manually. The result of the labeling data next will go through the

triangulation process to acquire the valid labeling data (Lemon & Hayes, 2020). The triangulation method is done according to the triangulation theory that referred to the book Sentiment Analysis Mining (Liu, 2019) and Triangulation observer (expert judgement) by a language expert (humanities). To specify the positive, neutral, and negative sentiment category (expert judgement) the data set will be used as an semantic analysis approach. Theoretically, if the Natural Semantic Metalanguage theory was used, there would be 65 primary lexicons (natural). In relation to this study, the sentiment aspect including the 'feel' of the natural semantic that akin to emotional expression, for instance angry, dissatisfied, completely dissatisfied, very angry, and many other; whereas the satisfied expressions being flattered, happy, feeling good and others; where neutral expressions gave an unbiased and indifferent responses.

Data Set Web Scraping & Twitter API	Training Network	Plot_performance - Model Acccuracy - Model Loss
Sentiment Labeling	Define Training Model Batch = 256; epoch = 200	Testing Model (Generate Data)
Data Pre_Processing - Cleansing - Convert to lowercase - Remove punctuation	Data Set Split (0.2); shuffle = false	Result Tag (Positive, Neutral, Negative)
Tokenize - Encode word & Label	Define LSTM Architecture	

Picture 1
Stages of Sentimen Analisis in the LSTM architecture

Data Pre-processing

The pre-processing is a process to help the algorithm learning in data training. This process was done to alter the unstructured data to a more structure data in order to simplify the data processing. Tweet in a regional or abbreviation language will be translated so that the data will be processed and aligned with Indonesian language. There are a few steps in the Text pre-processing in this study, which are:

a) Cleansing: Cleansing is where the non-alphabetic characters are removed to reduce the noise. Punctuations, symbols like '@' to mention accounts, hashtags (#), emoticons, and link from websites were the characters that was removed in this process.

b) Case Folding: In this process, tweet that has been through the cleansing process will then convert all the characters to lower case.
c) Tokenizing: Is the process of words separation from the composing sentence called token or term. In this step, data training and sentiment labels will make vocabulary into the dictionary mapping index (replacing the words in tweet to integrator).

LSTM Architecture

This study uses RNN/LSTM (Long Short-Term Memory) that work as an artificial architect from Recurrent Neural Network in deep learning *Learning* to form the model (Sherstinsky, 2020). LSTM is appropriate to learn the experience (deep learning) in classification, process, and predicting time series with an unpredictable intermission (Azzouni & Pujolle, 2017). The main advantage of LSTM is retrieving the data output order in the previous process to later be used for deciding the sentiment of the words. A model-driven Deep learning will be using tensorflow backend. Picture 2 shows the LSTM structure that will later be used in the sentiment analysis. Embedding layer, functions to transform word token (integrator) into certain embedding measure, while LSTM layer is determined by the hidden state dims and the number of layers. Full connected layer will map the LSTM output layer into the decided measure. Softmax Activation layer, will transform all of the output value into scores between 0 and 1. *Output Softmax*, final *ouput* dari *network*.
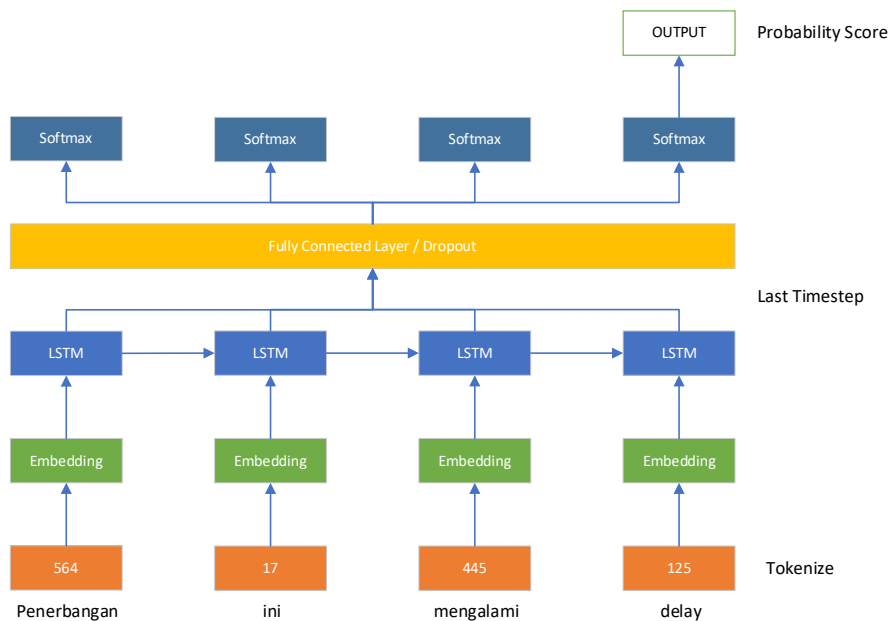
Picture 2
LSTM Architecture for Sentimen Analisis

Based on the built model archutecture in Picture 3, layer 1: embedding layer with the vector 256 sixe with the set of 50 max length per sentence. Layer 2, droput network posed as a regulation to prevent overfitting in the neural network train model. Layer 3, which has 2 layers LSTM that stackked on top of each other. The

first layer of LSTM takes singular parameter input with 256 parameter output, whilst the second layer of LSTM has 256 parameter input and return the same number to the parameter output so that in the last layer there will be 256 parameter length. Layer LSTM will be applied using tensorflow NVIDIA® CUDA® Deep Neural Network (cuDNN) kernel for maximum performance model along wih supporting the DNN implementation.

```
Model: "sequential_2"
_____
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 50, 256)           1280000
_____
dropout_2 (Dropout)          (None, 50, 256)           0
_____
lstm_4 (LSTM)                (None, 50, 256)           525312
_____
lstm_5 (LSTM)                (None, 256)               525312
_____
dense_2 (Dense)              (None, 3)                 771
=================================================================
Total params: 2,331,395
Trainable params: 2,331,395
Non-trainable params: 0
_____
```

Picture 3
LSTM Model's Parameter

Training Model

In this stage, data set will be divided into data tarin and data test (Picture 4). With the ratio 0,2 or 80:20, 80% data training and 20% data testing. Which means, 27.426 tweet was divided into 21.968 data training and 5.492 data testing. Training model has *batch_size* = 256 that implies the model sample before updated, and *epoch* = 200 meaning the number of training sample in 1 batch. The *batch_size* and epoch selection were done several times to get the great *accuracy* dan *val_accuracy* percentage.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=True)

#Training Model
batch_size = 256
epochs = 200

history = model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size, verbose=2, validation_data=(X_test,y_test))
```

Picture 4
**Split and Training Model**

RESULT AND DISCUSSION

## LSTM TRAINING AND TESTING RESULT

Picture 5
Time Stamp Model Fit

Correspond to the statement, here are the explanations of the training model:

- Loss decrease, Accuracy increase. Meaning that Neural Network adjust with the weight and bias of the model with decreasing the Loss, making the model run with no problem in every optimization. Whilst the increase in Accuracy metric, gives percentage to the algorithm performance that run smoothly. The lessen the Loss, the better the model. This percentage shows how the model is more accurate than the actual data.

- Validation Loss decrease, Validation Accuracy increase. This means the built model learning run smoothly. It could be assumed that the model could be generalized for the new data that have yet to be seen by the model.
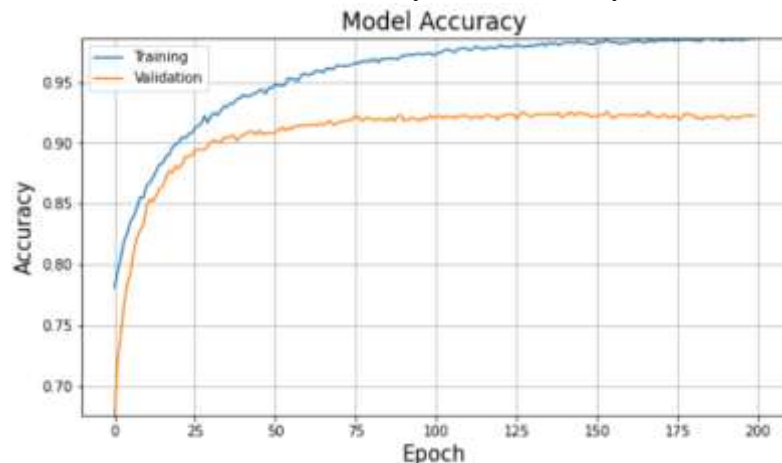
Picture 6
**Model *Accuracy* from *Training* and *Validation* using LSTM Network**
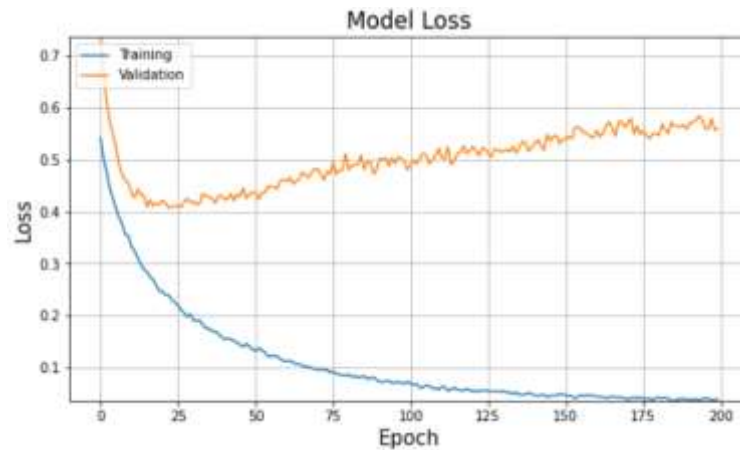
Picture 7

Model *Loss* from *Training* and *Validation* using LSTM Network
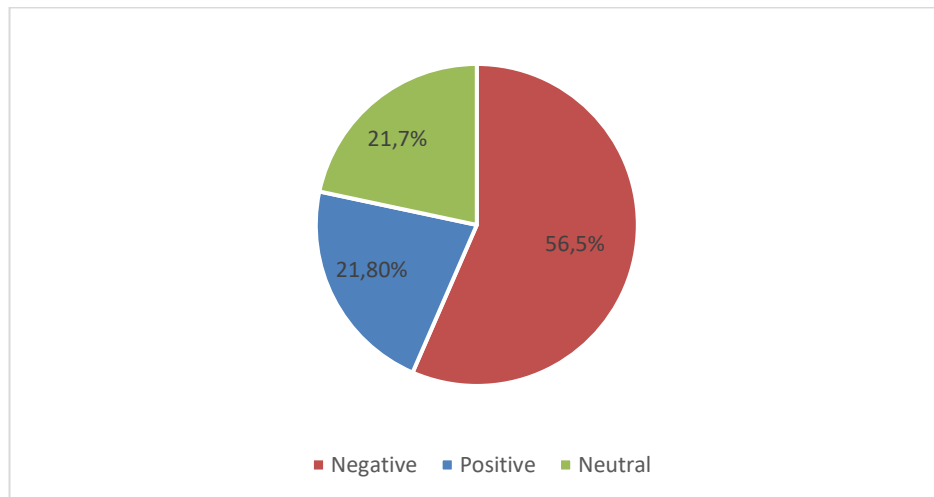SENTIMENT ANALYSIS TO ANALYZE CUSTOMER FEEDBACK

After testing the model on 5.493 data testing (tweet), the achieved data could be seen in Picture 8. The graphic present the outcome of text classification sentiment that has a real tagging and prediction tagging value. According to the graphic, real negative from the airline sentiment after model testing (LSTM Network) display the Prediction Negative with 174 tweets difference. Sentence level sentiment analysis was concluded into a document level that shows xy airline sentiment in the past 3 years is 56.5% Negative sentiment, followed by 21.8% Positive sentiment, and 21.7% Neutral sentiment (Picture 9).
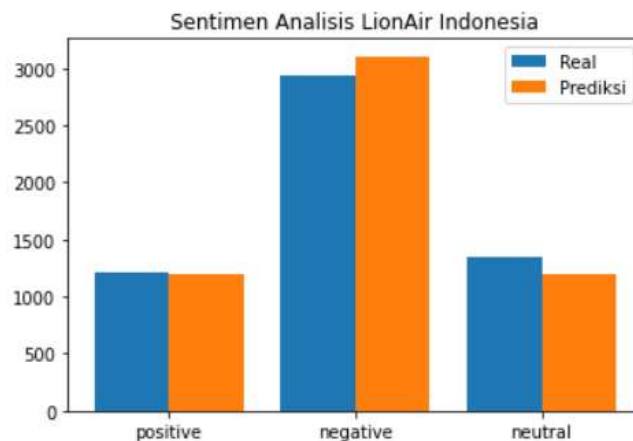
Picture 8

**Real and Prediction comparison Sentimen Analisis**

Picture 9
Result of airline's Sentimen *Document* Level

Simply put, the result of the sentiment analysis that was done, was used to analyze customer feedback from products and services of xy airline. Deep learning model has labeled tweet from customer or from observer as customer feedback to positive, neutral, and negative. The majority of the sentiment against xy airline is the negative sentiment. From the negative sentiment label, word cloud will be made to show the text data visually to portray the words frequency used in the negative tweet sentiment. Negative word cloud sentiment could be seen in Picture 10.

From the word cloud, the word "Xy airline", "delay", "victim", "baggage", "officer" was dominate. It could be assumed that those are the factors that caused the negative sentiment against xy airline. The factors would be explained more clearly in the graphic on Picture 10.
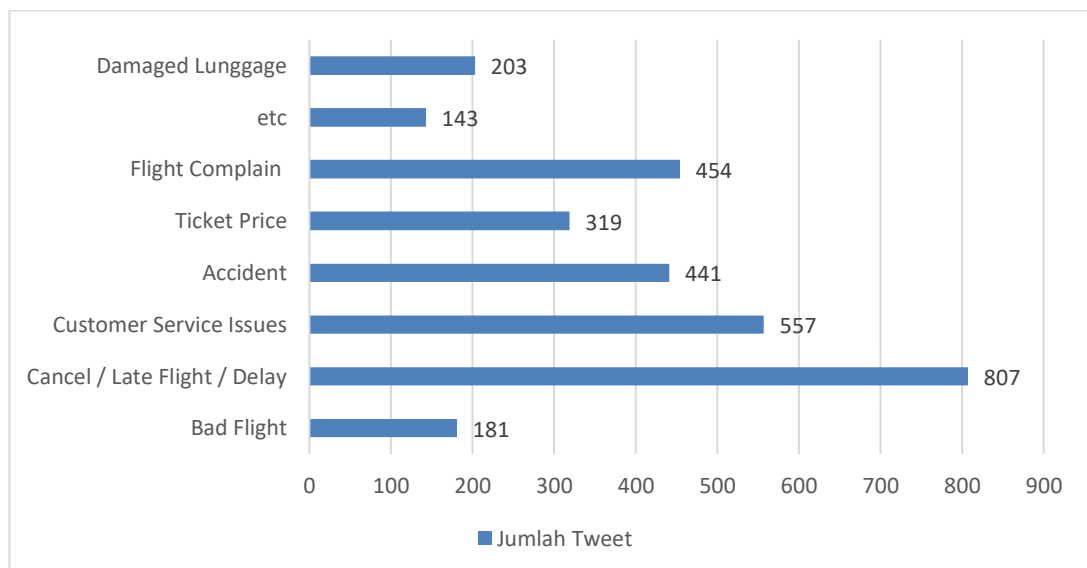
Picture 10
Reason Graph *Sentiment Negative*

From Picture 10, delay flight, followed by the lack of customer service, flight complains, recent airplane accidents, and plane tickets are the main reason why the negative sentiment has the highest value. Biased case that happened to the pilot and stewardess, the protests of the victim accident were accumulated to be the other reason of the negative sentiment tweet

CONCLUSION

Through this study, the application of sentiment analysis was discussed in xy airline in Indonesia using Deep learning, RNNLSTM as a model/ architecture. Needed to know that NLP in Deep learning could use Indonesian language as an input, with the requirement that the provided data set has to be validated (triangulation) before it was put in the model.

LSTM model training implemented in the data set gain 98.5% accuracy and 92.2% validation accuracy, which could be interpret that the model run well and could learn the data input that have not been seen (generalization). The outcome of the data training gave output that the sentiment in xy airline has more than 56.5% negative sentiment than positive and neutral sentiment.

From the presented percentage, the factors that underlie the negative sentiment was drawn from Word cloud tweet negative sentiment and resulted in some reasons how the negative sentiment of the xy airline's product and services arise. The rationale learnt from this study is expected to be a source to help the Long-term business continuity in the future.

REFERENCES

Arenggoasih, Wuri, & Wijayanti, Corona Raisa. (2020). Pesan kementerian agama dalam moderasi melalui media sosial instagram. *Jurnal Jurnalisa*, *6*(1).

Athiyah, Ummu. (2016). Pengembangan Alat Permainan Maze" Papan Laju Warna" Untuk Menstimulasi Kesiapan Membaca Anak Kelompok B Di Tk Aba As-Salam. *E-Jurnal Skripsi Program Studi Teknologi Pendidikan*, *5*(8), 417–424.

Azzouni, Abdelhadi, & Pujolle, Guy. (2017). *A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction*.

Boerman, Sophie Carolien, & Kruikemeier, Sanne. (2016). Consumer responses to promoted tweets sent by brands and political parties. *Computers in Human Behavior*, *65*, 285–294.

Cambria, Erik, Olsher, Daniel, & Rajagopal, Dheeraj. (2014). SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Clinten, B., & Nistanto, R. K. (2019). *Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak, kompas. com*.

Gajakosh, A. M., & Jayaraj, M. (2015). an important Hepatoprotective medicinal plant. *Research Article Chemical And Pharmacognostic Investigation On*.

Kasiri, Leila Agha, Cheng, Kenny Teoh Guan, Sambasivan, Murali, & Sidin, Samsinar Md. (2017). Integration of standardization and customization: Impact on service quality, customer satisfaction, and loyalty. *Journal of*

Retailing and Consumer Services, *35*, 91–97.

Lemon, Laura L., & Hayes, Jameson. (2020). Enhancing trustworthiness of qualitative findings: Using leximancer for qualitative data analysis triangulation. *Qualitative Report*, *25*(3), 604–614.

Liu, Bing. (2019). Sentiment Analysis Mining Opinions, Sentiments, and Emotions. In *Journal of Chemical Information and Modeling* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Miller, Victoria, Mente, Andrew, Dehghan, Mahshid, Rangarajan, Sumathy, Zhang, Xiaohe, Swaminathan, Sumathi, Dagenais, Gilles, Gupta, Rajeev, Mohan, Viswanathan, & Lear, Scott. (2017). Fruit, vegetable, and legume intake, and cardiovascular disease and deaths in 18 countries (PURE): a prospective cohort study. *The Lancet*, *390*(10107), 2037–2049.

Prasetiarini, Tantri Ayu. (2020). *Analisis Sentimen Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM)*. Universitas Pembangunan Nasional Veteran Jakarta.

Sherstinsky, Alex. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*(March), 1–43. https://doi.org/10.1016/j.physd.2019.132306

Sundari, Cisilia. (2019). Revolusi Industri 4.0 Merupakan Peluang Dan Tantangan Bisnis Bagi Generasi Milenial Di Indonesia. *Prosiding Seminar Nasional Fakultas Ekonomi Untidar 2019*.

Tomasoa, Lyonly, Iriani, Ade, & Sembiring, Irwan. (2019). Ekstraksi Knowledge tentang Penyebaran #Ratnamiliksiapa pada Jejaring Sosial (Twitter) menggunakan Social Network Analysis (SNA). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *6*(6), 677. https://doi.org/10.25126/jtiik.2019661710

Tsolmon, Bayar, Kwon, A. Rong, & Lee, Kyung Soon. (2012). Extracting social events based on timeline and sentiment analysis in twitter corpus. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7337 LNCS*, 265–270. https://doi.org/10.1007/978-3-642-31178-9_32

Wazery, Yaser Maher, Mohammed, Hager Saleh, & Houssein, Essam Halim. (2018). Twitter sentiment analysis using deep neural network. *2018 14th International Computer Engineering Conference (ICENCO)*, 177–182. IEEE.

Yang, Bo Yi, Qian, Zhengmin Min, Li, Shanshan, Fan, Shujun, Chen, Gongbo, Syberg, Kevin M., Xian, Hong, Wang, Si Quan, Ma, Huimin, & Chen, Duo Hong. (2018). Long-term exposure to ambient air pollution (including PM1) and metabolic syndrome: The 33 Communities Chinese Health Study (33CCHS). *Environmental Research*, *164*, 204–211.