

Sentiment Analysis of the Use of Telecommunication Providers on Twitter Social Media using Convolutional Neural Network

Zafiratul Amalia
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
zafiratulamalia@gmail.com

Mohamad Irfan
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
irfan.bahaf@uinsgd.ac.id
<https://orcid.org/0000-0003-4963-287X>

Dian Sa'adillah Maylawati
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
diansm@uinsgd.ac.id
<https://orcid.org/0000-0002-1193-3370>

Agung Wahana
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
wahana.agung@uinsgd.ac.id
<https://orcid.org/0000-0002-6468-0741>

Wildan Budiawan Zulfikar
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
wildan.b@uinsgd.ac.id
<https://orcid.org/0000-0002-7387-7300>

Muhammad Ali Ramdhani
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
m_ali_ramdhani@uinsgd.ac.id
<https://orcid.org/0000-0002-6492-067X>

Abstract— Telecommunication technology continues to develop starting from 1G, 2G, 3G, 4G, and currently entering the 5G era. The Global System for Mobile Communications (GSM) based telecommunication industry in Indonesia consists of three big names: Telkomsel, XL, and Indosat. During the Covid-19 pandemic, activities carried out outside the home should be done online. People hope that the internet network can work properly. However, the reality is not as expected, because many networks are experiencing slow internet problems and many complaints are delivered through social media. Therefore, this research aims to find the insight opinions that have been conveyed to the telecommunications operator in social media. This research used the Convolutional Neural Network (CNN) algorithm to classify text sentiment (negative or positive) about telecommunication providers. The experiment with text data from Twitter is conducted after preprocessing and weighting of the Word2Vec process. The confusion matrix experiment shows that the CNN algorithm's performance reaches an average accuracy value of around 86.22%. The experiment was carried out by dividing the training data and testing the data 5 times in 10 times. The study results indicated that disruption of cellular telecommunications operators provided many sentiments, especially negative sentiment at the beginning of the COVID-19 pandemic.

Keywords— convolutional neural network, deep learning, sentiment analysis, telecommunication provider

I. INTRODUCTION

The development of communication technology in Indonesia has entered a new phase with the development of the information technology industry. Cellular phone coverage has reached all provinces in Indonesia with some regencies/cities in Indonesia. Telecommunication technology is also providing more types of telecommunications services, ranging from fixed telephones, wireless telephones, mobile telephones, etc. The role of the cellular telecommunications industry in people's lives and the national economy is crucial. The telecommunications sector tends to increase from 2017 to 2020 [1]. Mobile telecommunications technology on the internet network will continue to develop, starting from 1G, 2G, 3G, and until now entering the 4G era. However, several

telecommunications companies are already preparing to enter the latest era, namely the 5G era. Along with the times, until now the GSM-based telecommunications industry occupies three big names in the telecommunications operator market in Indonesia, namely: Telkomsel, XL Axiata, and Indosat [2]–[4].

In 2017, three telecommunications operator companies provided internet services, which ranked in the top 3 based on a survey of mobile-based internet users. 43% of users use Telkomsel, 18.1% use Indosat, and 18% use XL. The main reason Indonesian people use mobile operators for mobile-based internet service providers is that the strongest signal at home gets 52.3% survey results, affordable package prices get 18.3% survey results, and the number of promos and bonuses with 8.5% of survey results [5]. Moreover, in 2021, Telkomsel, Indosat, and XL still be the biggest telecommunication provider in Indonesia [6], [7]. Then, Regulation of the Minister of Communication and Informatics No: 48/PER/M.KOMINFO/11/2009 concerning internet access service providers in the sub-district internet universal telecommunications service area to support internet access services for the community as well as to encourage the accelerated improvement of services and for the use of information and communication technology which has the aim of increasing the intelligence of citizens and the welfare of the community, a mobile internet network service is needed [8].

The Covid-19 pandemic has forced all Indonesians to carry out daily activities only at home. School children or students who are currently studying are affected by this pandemic. They are required to take online learning. Some office workers were also affected by requiring them to work online. The Indonesian people hope that the telecommunication operator signals they use will run smoothly without any problems from the following online activities. However, many Indonesians may complain about the telecommunication operator's network that is used frequently experiencing disturbances. Based on research released by Hootsuite, in January 2020, Indonesia's Internet speed had an average of only 20.1 Mbps or far below the

world's average internet speed of 73.6 Mbps. In his research, Singapore occupies a country with a record internet speed of up to 200.1 Mbps [9].

This condition gave rise to many opinions including on social media Twitter. Twitter is the most data source that used in text analytics research. This is because of the easy access and permission to pull data from Twitter compared to other social media. However, apart from more uncomplicated data collection techniques, Twitter is a place that is more open, open-minded, egalitarian, and a place for various trending news [10]. Compared to Facebook, which has many fake accounts and is spreading hoax news, and Instagram, which is mostly used for self-promotion and self-exposure, Twitter has far more educated users. The lack of mental judgment shows this, and if there is news that is a hoax, has a toxic smell, and is fighting with each other, Twitter users are wiser in dealing with it. In essence, social media analysis technology is currently one of the research strengths in the digital era [11]. Therefore, Twitter is considered to be more in demand by the public because it is easy to express their opinions. With so many Twitter users expressing this opinion, it can be used to find information. However, its use requires good analysis so that the information obtained can help many parties to support a decision.

Today, many sentiment analysis researchers use the Deep Learning method for Natural Language Processing (NLP), including sentiment analysis. The Deep Learning method is considered to have higher accuracy and is consistent when faced with big data. Deep Learning can produce improved accuracy as a result of these advantages [7]. There are Indonesia's largest NLP communities, such as IndoLEM, are developing Indonesian NLP benchmarks (Indonesian Language Evaluation Montage) [12], IndoNLU (Indonesian Natural Language Understanding) [13], and IndoNLG (Indonesian Language for Natural Language Generation) [14]. These studies also perform NLP research, which includes sentiment analysis utilizing Deep Learning. Deep learning, which is the creation of artificial neural networks, is widely used in sentiment analysis: (1) sentiment analysis research which proves sentiment analysis using CNN has a better accuracy value than using other neural network models, such as: Recursive Neural Network (RNN) and Matrix-Vector Neural Network (MV RNN). From the dataset that has been obtained, as many as 10,662 film reviews, half of which are positive data and half are negative data. The accuracy value obtained using the CNN algorithm reaches 80% [15]; (2) short text sentiment analysis to prove the effectiveness of the convolutional neural network algorithm compared to traditional methods in sentiment analysis of comments from users' attitudes based on various texts collected from Weibo [16]; (3) sentiment analysis of the Russian community for reviews of Russian-language e-commerce products located in Russia. In the test by including emoji, the f-measure result was 75.45% [17]; (4) compare the advantages of each algorithm, namely the CNN algorithm and the Support Vector Machine (SVM) in sentiment analysis. In this journal, a sentiment analysis model that uses CNN and SVM algorithms can work effectively and improve the performance of text classification [18], [19]; (5) Sentiment analysis for ranking the popularity of tourist destinations which aims to find out popular tourist attractions according to social media users based on the results of likes,

statuses, and accounts that discuss these tourist attractions [20]; and so on.

From the ongoing problems regarding the speed of the internet network in Indonesia, which is experiencing disruption or slowness, Indonesia is still stuck on 3G and 4G networks even though several countries have started accessing 5G networks. With the analysis of data from Twitter regarding the use of telecommunication operators. It is hoped that it can provide information to operator users to be more selective in choosing and help internet service providers improve their internet network and as a benchmark whether it is feasible to enter the 5G internet network. Therefore, this study aims to analyze the public sentiment about the use of telecommunication providers through Twitter using Convolutional Neural Networks.

II. RESEARCH METHODS

A. Research Activities

This research was conducted with a data science approach using the data science methodology Cross-Industry Standard Process for Data Mining (CRISP-DM) [21], [22]. Activities in the CRISP-DM methodology consist of business understanding, data understanding, data preparation, modeling, model evaluation, deployment, and model management. This research conducts the business understanding until the evaluation of the model. The deployment and model management are not conducted.

B. Convolutional Neural Network

Convolutional Neural Networks (CNN) is a type of neural network that is widely utilized in image and text processing. Convolution, or simply convolution, is a filtering and classification matrix for images and text [23]. Filtering is done in each phase using the layers of a Convolutional Neural Network. The training procedure is the name of the technique. The training procedure is divided into three layers: Convolutional, Pooling, and Fully-connected layers [24][25]. The CNN layer architecture is shown in Figure 1. The Convolutional layer is the main layer for generating new features from input data. Like the Convolutional Layer, the Pooling layer is responsible for shrinking the Convolved Feature's spatial size. Because of the dimensionality reduction, the amount of computing power required to process the data is reduced. It also helps with training by extracting rotational and positional invariant dominant characteristics. Pooling can be divided into two types: maximal pooling and average pooling. The maximum value from the image's Kernel-covered area is returned by Max Pooling. Average Pooling returns the average of all values from the Kernel's data segment. After that, the Fully-Connected layer starts learning a non-linear function. Learning non-linear combinations of the high-level qualities represented by the convolutional layer's output using a Fully-Connected layer is a cheap strategy for learning non-linear combinations of the high-level qualities represented by the convolutional layer's output.

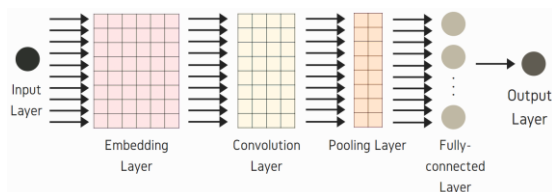


Fig. 1. CNN Architecture Illustration

III. RESULT AND DISCUSSION

A. Business Understanding

Based on research that has been done, internet speed in Indonesia has a meager average compared to several countries in Asia. Indonesia is even still trapped in 3G and 4G networks even though several countries have started accessing networks to 5G, and some have even entered the 6G network. If only the existing networks were more sophisticated, the costs would be even lower. The uneven internet service infrastructure is one of the causes of the current slow internet network. Development is only concentrated in big cities, while it is challenging to get access in remote areas and has not been served.

So far, telecommunications operators are competing for positions in areas that are crowded with users. This business pattern encourages telecommunications operators to develop their internet network infrastructure in regions with many uses, such as offices and shopping centers. Meanwhile, the development of the internet network in residential areas has received less attention because, in terms of business, the economy of scale in residential areas is low. This resulted in residential areas experiencing slow access to the internet network, which is widely felt today. With the existence of Large-Scale Social Restrictions (PSBB) and the large number of people doing Work from Home (WFH) in various regions, this has resulted in high community activities from home. To work from home, people also need internet access which is as fast as internet access when they work in the office.

In the use of comment data in the form of text tweets from social media, Twitter is often used to process each comment's sentiment in the form of tweets. Sentiment analysis system from users of telecommunication operators for internet network services becomes a means to express user satisfaction and complaints from telecommunication operators providing internet services. Therefore, in overcoming and knowing the problems of complaints that often occur to telecommunications operators, a system that can analyze social media users' opinions, especially on Twitter social media, is proposed.

B. Data Understanding

With the popular use of social media for people in Indonesia, they are indirectly connected to each other and exchange information that they know. Freedom in its use also gives rise to many opinions or opinions that make them active in responding to social issues that are happening both at home and abroad. The data taken for this sentiment analysis research is taken by utilizing the Twitter API provided by Twitter for free so that data retrieval can be done more efficiently than other types of social media.

On social media, Twitter is the official account owned by an internet service provider that is used to interact with customers or their users. The data taken is in the form of

tweets that are directly mentioned to the official accounts @Telkomsel, @IndosatCare, and @myXLCare. Tweet data is taken starting from May 15, 2020 to August 10, 2020. The number of users taken as the example above for the number of mentions on official accounts for the period May 16, 2020 - May 31, 2020, namely: a total of 491 mentions @Telkomsel, 407 mentions @IndosatCare, and 318 mentions of @myXLCare. The data will be grouped or classified into two sentiments, there are positive sentiments and negative sentiments. A lot of tweet data is in the form of criticism or complaints as well as the satisfaction felt with this sentiment analysis system. That's why only positive and negative sentiments are taken, and no neutral sentiments are included.

From the dataset that has been collected, it will be used as training data (training data) and testing data (test data). Experiments on the system will be carried out five times with variations in training and testing data used. The experiment was carried out successively using 90% training data and 10% testing data, 80% training data and 20% testing data, 70% training data and 30% testing data, 60% training data and 40% testing data, and 50% training data and 50% of testing data.

C. Data Preparation

The data taken from Twitter is in raw or unstructured data, which causes the data to be processed first. Unstructured data is not well used in data mining because the data still contains a lot of noise, possibly incomplete data. Noise is a very sensitive thing in data mining. Data needs to be processed or improved so that the resulting output becomes higher quality data, and of course, the data will be cleaner, namely, text preprocessing.

Text preprocessing is an essential step in the data preparation stage of the mining process. Text preprocessing guarantees that input data is of high quality, guaranteeing that the outcomes are as expected [26], [27]. Case-folding and tokenizing, deleting superfluous letters, removing stop-words, and stemming are among the text preprocessing techniques used in this study. The Sastrawi library is used for Indonesian language stemming, with the Nazief-Adriani stemming method Word Embedding using Word2Vec

Word embedding is a type of word representation that allows for representing words that have similar meanings [28]. There are many word embedding representations, such as Word2Vec [29], GloVe [30], and FastText [31]. The embedding layer in this study is Word2Vec. Word2vec is a method for expressing words in N-dimensional vector form [32]. Word2Vec uses a neural network to determine contextual and semantic similarity (contextual and semantic similarity) in the form of one-hot encoded vectors when presenting a word. Contextual and semantic similarity can be used to illustrate a word's relationship to other words. Word embedding representation varies by language.

Word weighting (Word2Vec) is done after going through the data preprocessing process. In Word2Vec words will be represented in the form of vectors and create a vocabulary with models from training data. These vocabulary words will form vectors that place words, so words with the same meaning will be close together in the vector space. Vector values that are owned by almost the same word will have almost the same value. Word2vec has various dimensions, there are 300, 250, 200, 150, 100, and 50. It would be easier

if the analyzed data had the same size. To adjust the length and short of a sentence will be added a space to fill the empty matrix. After generating the matrix of the same length, the data is ready to be used in the following process, namely as input to the model process of the convolutional neural network algorithm that will be used.

D. Modeling Process with Convolutional Neural Network

By using the linear regression layer as the output layer, we can create a simple CNN model for our purpose. The task's CNN architecture is given here. The embedding layer, which is the initial layer of the model, is found within the CNN algorithm. Words are represented as real-valued vectors in a high-dimensional space using layer embedding. This layer allows for the pre-trained word vector matrix to be used to initialize vocabulary word vectors. The input tweets are converted into numeric word token sequences, such as t_1, t_2, \dots, t_N , where t_N represents the original word, and N is the token vector length. To keep the result size identical for tweets of varying length, the authors limit the maximum value of N to the maximum tweet length of all tweets. Any tweet was shorter than N will be populated to N using zeros. In the convolutional layer, the m filter is used to extract the local n -gram features from the matrix of the previous layer.

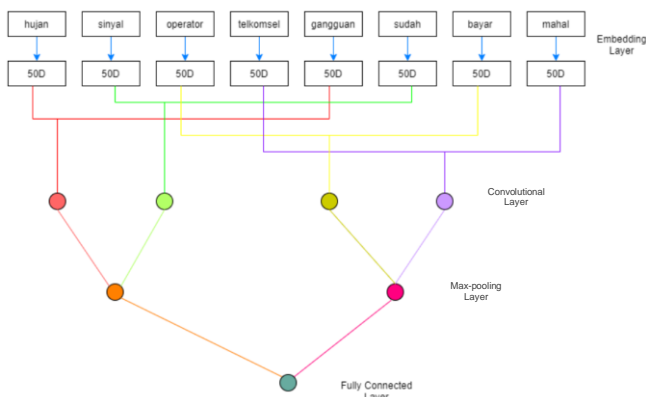


Fig.2. Illustration of CNN Modeling Process

There are several stages in the classification process, namely Convolutional Layer, Pooling Layer, and Fully Connected Layer. After the preprocessing stage, the module's word (tweet) input will be converted into an array of tokens with specific dimensions using word embedding and then mapped in the feature matrix or sentence matrix by the embedding layer. Furthermore, in the data convolution layer, the weight is calculated layer by layer with a filter layer to perform calculations on each embedding layer. This layer is used to form the weights that the input layer has with the hidden layer. Then, the largest value weight is taken from each filter layer using a pooling layer. The fully connected layer finally composes these features to output the final regression results with a linear decoder. After doing the pre-training then it is done to train the backpropagation algorithm to train the error rate. A loss function is an error in the training data set. Loss validation is an error after running a data validation set.

Fig 2 illustrates the CNN modeling process. At the pooling layer, the unification operation will be carried out which is used to combine vectors generated from different convolution layers into one-dimensional vectors. This

process is done by taking the maximum value observed in the vector generated from the convolution layer. This vector will later capture the features that most relate to the text. From the illustration above, one of the examples of tweets used in this research is used. Each word has a weighting in the form of a 50×50 matrix, but in the illustration above, it only takes 3 matrices.

Furthermore, in the convolution layer, the number of weighting results is calculated and then divided by five because the coding stores the kernel which is worth 5. After getting the results from the calculations in the convolution layer, the largest value for each row is sought from convolution one and convolution two columns and the convolution 3 and 4 columns. The comparison results will enter the max-pooling layer, find the largest value, and end up with the final connected delayer. The value in the final connected layer is the largest value after comparing it to the max-pooling layer.

E. Evaluation

Evaluation is carried out to adjust whether the results of the modeling with the desired business understanding are appropriate or not. Before the evaluation is carried out, it will be tested first. By testing, we can find out whether the system built can achieve the desired business understanding. Twenty tests using epoch 50 and distribution between each official account @Telkomsel, @IndosatCare, and @myXLCare, with 90% sharing of data as training data with 10% testing data, 80% training data with testing data as much as 20%, training data as much as 70% with 30% testing data, 60% training data with 40% testing data, 50% training data with 50% testing data.

This research used the confusion matrix to evaluate the sentiment classification result model [33]. There are values of accuracy, precision, and Recall [34]–[36]. Precision is calculated by dividing the total sentiment classifications that are correctly predicted to be positive (True Positive) by the number of sentiment classifications that are correctly predicted to be positive and the correct sentiment classifications are predicted to be negative in the document (True Negative), so that the prediction results correctly determine whether the sentiment classification is correct, both predicted by humans and predicted by the system is correct. In a document, Recall is equal to the total number of predicted true sentiment classifications divided by the number of false-positive predicted and false-positive sentiment classifications. Recall evaluates the proportion of sentiment classifications predicted by humans and those generated by the system. At the same time, accuracy is defined as the proportion of the total sentiment classification that is predicted correctly and incorrectly divided by the total sentiment classification in a document.

F. Experiment Scenario and Result

There are 20 times experiment scenarios with 1,480 tweets data. Experiment I was conducted for the entire dataset consisting of mentions of official accounts @Telkomsel, @IndosatCare, and @myXLCare was carried out with 90% training data with 1,332 data and 10% testing data with 148 data. Experiment II for the entire dataset consisting of mentions of official accounts @Telkomsel, @IndosatCare, and @myXLCare was carried out with 80% training data with 1,184 data and 20% testing data with 296 data. Experiment

III for the overall dataset consisting of mentions of official accounts @Telkomsel, @IndosatCare, and @myXLCare was carried out with 70% training data with 1,036 data and 30% testing data with 444 data. Experiment IV for the entire dataset consisting of mentions of official accounts @Telkomsel, @IndosatCare, and @myXLCare was carried out with 60% training data with 888 data and 40% testing data with 592 data. Experiment V test for the entire dataset consisting of mentions of official accounts @Telkomsel, @IndosatCare, and @myXLCare was carried out with 50% training data with 740 data and 50% testing data with 740 data. Experiment VI, to mention to the official account @Telkomsel was carried out with 90% training data with 452 data and 10% testing data with 50 data. Experiment VII to mention to the official account @Telkomsel was carried out with 80% training data with 402 data and 20% testing data with 100 data. Experiment VIII test to mention to the official account @Telkomsel was carried out with 70% training data with 351 data and 30% testing data with 151 data. Experiment IX to mention to the official account @Telkomsel was carried out with 60% training data with 301 data and 40% testing data with 201 data. Experiment X to mention to the official account @Telkomsel is carried out with 50% training data with 251 data and 50% testing data with 251 data.

Experiment XI to mention to the official account @IndosatCare was carried out with 90% training data with 438 data and 10% testing data with 49 data. Experiment XII to mention to the official account @IndosatCare was carried out with 80% training data with 390 data and 20% testing data with 97 data. Experiment XIII to mention to the official account @IndosatCare was carried out with 70% training data with 341 data and 30% testing data with 146 data. Experiment XIV test to mention to the official account @IndosatCare was carried out with 60% training data with 292 data and 40% testing data with 195 data. Experiment XV to mention to the official account @IndosatCare was carried out with 50% training data with 244 data and 50% testing data with 243 data. Experiment XVI for mentioning the official account @myXLCare was carried out with 90% training data with 442 data and 10% testing data with 49 data. Experiment XVII to mention the official account @myXLCare was carried out with 80% training data with 393 data and 20% testing data with 98 data. Experiment XVIII to mention the official account @myXLCare was carried out with 70% training data with 344 data and 30% testing data with 147 data. Experiment XIX test to mention to the official account @myXLCare was carried out with 60% training data with 295 data and 40% testing data with 196 data. Last, experiment XX to mention to the official account @myXLCare was carried out with 50% training data with 246 data and 50% testing data with 245 data. Table I shows the accuracy result of those experiment.

TABLE I. ACCURACY RESULT OF EXPERIMENT

Splitting Percentage	Accuracy Percentage of Experiment Result of-			
	Whole	@Telkomsel	@IndosatCare	@myXL
90%-10%	93.59	90.19	93.87	83.99
80%-20%	93.24	89.1	95.91	85.85
70%-30%	91.44	88.74	92.66	91.21
60%-40%	92.39	88.11	92.83	92.89
50%-50%	91.48	83.73	94.67	91.05
Average Value	92.428	87.974	93.988	88.998

G. Discussion

In Table IV, the predictions generated from the convolutional neural network algorithm depend on the amount of data being trained. This is because the accuracy value of each test will be different with the use of different testing data. After testing 5 times on all datasets, mentioning the official accounts @Telkomsel, @IndosatCare, and @myXLCare, the average test results were 92.42% for testing all datasets; 87.97 for testing the @Telkomsel mention; 93.98% for testing the @IndosatCare mention; and 88.99% for testing the @myXLCare mention using epoch 50.

This experiment result is influenced by several factors. This is due to various factors. The first factor is that less data is used for this deep learning method. The data used is only 1480 tweets, this is due to the limited data collection time. This is because when crawling data through the tweet archive feature on the google spreadsheet, it can only retrieve 100 tweets per crawl. The next factor is the data used between negative data and positive data is not balanced. In the tests carried out, it was shown that the system was negative when classifying test tweets because 92% of the total tweet data that mentions the official accounts @IndosatCare and @myXLCare was labeled negative the system was learning it could not represent data that was labeled positive. Unlike the mention of @Telkomsel which has a balanced amount of positive and negative data.

Factors that affect the positive precision in the test are how many true negative and false positive values, while the negative precision is how many true negative values and false negative values. Furthermore, the factors that affect positive Recall are how many true positive and false negative values, while negative Recall is how many true negative values and false positive values. From the results of the labeling that has been done, it is found that negative sentiments have more numbers than positive sentiments. Prediction of negative sentiment is more accurate than positive sentiment because more negative data is tested than positive data.

CONCLUSION

The implementation of the sentiment analysis system using the Convolutional Neural Network algorithm has been successfully carried out by categorizing positive sentiment and negative sentiment. When the test was carried out five times for each data collection on mentions with official accounts @Telkomsel, @IndosatCare, and @myXLCare, the average test results were 87.97 for tests on mentions @Telkomsel; 93.98% for testing the @IndosatCare mention; and 88.99% for testing the @myXLCare mention using epoch 50. Several factors affect the level of accuracy: the first factor is that less data is used for this deep learning method. The data used is only 1480 tweets. This is due to the limited data collection time. The next factor is the data used between negative data and positive data is not balanced. In the tests carried out, the system when classifying test tweets became negative, because 92% of the tweet data that mentions the official accounts @IndosatCare and @myXLCare was labeled negative, so that when the system did the learning it could not represent data that was labeled positive. So it can be concluded that the telecommunications operator internet network services have negative sentiments from Twitter users. For future works, it needs more data collection is

