
To cite this article: Angga Ardiansyah et al 2020 J. Phys.: Conf. Ser. 1641 012048

View the article online for updates and enhancements.

Angga Ardiansyah¹, Sopian Aji², Dany Pratmanto³, Sandra Jamu Kuryanti⁴, Octa Pratama Putra⁵, Cep Adiwihardja⁶

¹,²,³Program Studi Sistem Informasi, STMIK Nusa Mandiri Jakarta, Indonesia
⁴,⁵,⁶Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia

E-mail: angga.axr@nusamandiri.ac.id, sopian.sop@nusamandiri.ac.id, dany.dto@nusamandiri.ac.id, sandra.sjk@bsi.ac.id, octa.opp@bsi.ac.id, cep.caw@bsi.ac.id

Abstract. The development of technology in the industrial revolution 4.0 is very rapid in various sectors in trade, communication and other sectors; both nationally and internationally. One of the fields, it is the telecommunications sector. Becoming one of the digital based providers of ‘by.u’ from Telkomsel which is the first digital operator in Indonesia. The algorithm testing uses Support Vector Machine (SVM) from 300 user review data applications. The ‘by.u’ review consists of 150 negative reviews which refers to dislike or bad expressions. Next, 150 positive reviews which refers the constructive suggestion or support the text in using Indonesian. The resulting model gets the results of testing accuracy by using Support Vector Machine with accuracy value shows that: 87.24

1. Introduction

The development of technology in the industrial revolution 4.0 is very rapid in various sectors. There are: in trade, communication, and other sectors; both nationally and internationally. One of the fields, it is in the telecommunications sector.

In the telecommunications sector, it has a significant development, such as services in the provider companies. It happens when cellular service providers have begun to switch to the digital world. Based on the observations made by researchers, it has not many digital providers in Indonesia yet. As being stated, there is only a few of digital providers that has provided the services in Indonesia, including ‘by.u’ from Telkomsel and Switch Mobile.

Being seen from Google Playstore, ‘by.u’ has been downloaded more than 1 million by smartphone users, with 59,474 reviews. As for the Switch Mobile, it has been downloaded by more than 100 thousand users and 6259 reviews.

‘by.u’ and the Switch Mobile, they compete with each other in providing and giving digital service providers. Both have the target each other; it is toward the youth of the generation, as known as the teenager or the adolescents.

Becoming one of the digital based providers of ‘by.u’ from Telkomsel which is the first digital operator in Indonesia.
Through its new service of ‘by.u,’ this operator is removing the offline needs and replacing it with digital processes [1]. Digital services are an innovation of technological advances and massive internet penetration [2]. With all-digital services, it is able to make it easier for consumers or provider-users to make transactions, where transactions can be done independently through the ‘by.u’ as the application. That thing is able to be downloaded from a smartphone via Google Playstore.

With the existence of internet, a person will easily express his or her feelings and emotions through ratings and reviews. Due to the enhancement of textual data, so there will be a need to analyze concepts of expressing sentiments and calculate insights to explore business [3].

In a previous study entitled "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naïve Bayesian Classification" where the data obtained was taken from tweets on cell phone service providers with an accuracy of 79.

In this study, sentiment analysis which is stabilized by researchers only from the sentence level using the Support Vector Machine (SVM) method, and the data taken is a digital provider review taken on the Google Play Store, using the method it is expected that researchers can produce more accuracy results higher than the previous study.

2. Methods
The method used by researchers to process the data is using the Vector Machine Support (SVM) Method. Data is taken from Google play store on ‘by.u’ application. The data that has been collected then, it will be selected according to the needs of researchers, by eliminating some of unnecessary data. After the data is selected, then the data is classified. At this stage, the data are grouped based on user reviews separated into positive and negative categories.

It has been many previous sentiment analysis studies by using the Support Vector Machine (SVM) method with different determination objects. One of them, it is about a study entitling the Sentiment Analysis of Fintech Users by Using Support Vector Machines and Particle Swarm Optimization Method [5].

The sentiment analysis aims to group user reviews into either positive or negative point of view with reference to feelings, emotions, opinions and attitudes. The classification of sentiment analysis is divided into three levels, they are the document level, sentence level, and aspect level. [6]

Support Vector Machine (SVM) is a supervised learning method that analyzes data and recognizes patterns used for classification [7]. The researcher uses the Support Vector Machine (SVM) method to analyze the sentiments of digital providers especially at the digital provider ‘by.u’ by measuring the accuracy of experimental data using the RapidMiner Studio 9.7.1 application.

In conducting trials or experiments, researchers used data comments or reviews of the Google Play Store on the ‘by.u’ application that has been filtered and obtained for about 300 data; consisting of 150 positive reviews and 150 negative reviews. So, from the positive review data and negative review data that has been separated by the same amount, it shows the ability of each review or being stored in the notepad application with extension ‘.txt’ for making easy processing toward the data.

Then from the data that has been saved with the notepad application, it will be processed by using the application RapidMiner Studio 9.7.1. As for the text processing that researchers use, it uses the Tokenize, Transform Cases, Stop Words (directory). The testing is done by selecting the Support Vector Machine (SVM) feature selection in order to get the accuracy value.

From the accuracy value of the equation confusion matrix proportion model, the number of predicted values is obtained, so it is seen of the ROC graph with the AUC (Area Under Curve) value.
Figure 1, it explains the framework. In this study, it is began by taking a user review of ‘by.u’ application from one of the links. Then, it processes the text by using tokenize, next the change cases and stop words (directory). At the feature selection stage, it is carried out by using Classification Algorithm with Support Vector Machine (SVM). Then the results on the accuracy model is obtained by the accuracy values on the concept of data mining and ROC Curve; or the graphic results within this research.

Table 1. The Text Processing by using Tokenize, Transform Cases, and Stop words (directory)

<table>
<thead>
<tr>
<th>Text Review</th>
<th>Tokenize</th>
<th>Transform Cases</th>
<th>Stop words (Directory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalah sama situs</td>
<td>Kalah sama situs</td>
<td>kalah sama situs</td>
<td>kalah sama situs</td>
</tr>
<tr>
<td>judi online depo</td>
<td>judi online depo</td>
<td>judi online depo</td>
<td>judi online depo</td>
</tr>
<tr>
<td>lagsung masuk</td>
<td>lagsung masuk</td>
<td>lagsung masuk</td>
<td>lagsung masuk</td>
</tr>
<tr>
<td>La ini. Gk masuk</td>
<td>La ini Gk masuk</td>
<td>la ini gk masuk</td>
<td>la ini gk masuk</td>
</tr>
<tr>
<td>Saldo udah</td>
<td>Saldo udah</td>
<td>saldo udah</td>
<td>saldo udah</td>
</tr>
<tr>
<td>kepotong. Tai</td>
<td>kepotong Tai</td>
<td>kepotong tai</td>
<td>kepotong</td>
</tr>
<tr>
<td>...!!!!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNAPA BAYAR</td>
<td>KNAPA BAYAR</td>
<td>knapa bayar</td>
<td>knapa bayar</td>
</tr>
<tr>
<td>LEWAT DANA</td>
<td>LEWAT DANA</td>
<td>lewat dana gak</td>
<td>lewat dana gak</td>
</tr>
<tr>
<td>GAK BISA....??????</td>
<td>GAK BISA</td>
<td>bisa gagal teros</td>
<td>bisa gagal teros</td>
</tr>
<tr>
<td>GAGAL TEROS.....!!</td>
<td>PDHL MURAH</td>
<td>pdhl murah gk</td>
<td>pdhl murah gk</td>
</tr>
<tr>
<td>PDHL MURAH</td>
<td>GK KENA</td>
<td>kena admin</td>
<td>kena admin</td>
</tr>
<tr>
<td>GK KENA ADMIN</td>
<td>ADMIN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the text review. Some examples of text taken from user reviews in the ‘by.u’ application who has negative properties. If it is seen after going through Text Processing using Tokenize, Transform Cases, Stop words (directory), then the text turns into positive text. Each Text Processing has a different function, if it is paid attention to the text through Tokenize, all punctuation will be omitted such as punctuation using dots, exclamation marks and so on. Then, if the text is through the Transform Cases, all text review that uses uppercase will change become lower. Well for the last it is the case letters, if the text is through Stop words (directory) negative vocabulary or words will automatically be deleted, where the text containing negative or unpolite elements; such as the example in table 1 uses the negative word ‘t*s*i’ (which means
human excrement) after going through Stop words (directory) this word is automatically deleted. In this test, the study conducted a test of using the Support Vector Machine (SVM) algorithm to obtain accuracy. It will be seen as follows:

**Table 2. The Accuracy Algorithm Support Vector Machine (SVM)**

<table>
<thead>
<tr>
<th></th>
<th>True Negative</th>
<th>True Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. Negative</td>
<td>508</td>
<td>84</td>
</tr>
<tr>
<td>Pred. Positive</td>
<td>57</td>
<td>456</td>
</tr>
</tbody>
</table>

For the table 2, ‘by.u’ application’s negative review data shows 456, which are classified into positive reviews according to predictions made by the SVM method. For 84 data, they are predicted to be positive reviews but it turns out that the results of predictions are negative reviews. For by.u’s negative review data, 508 were classified as negative reviews according to predictions made with the SVM method. While for 57 data that predicted negative reviews, it turned out to be positive review results. With accuracy by using Support Vector Machine (SVM), it has the accuracy value of: 87.24In this study, to get the value of model accuracy such as confusion matrix and ROC Curve, researchers used RapidMiner Studio by using the Support Vector Machine (SVM) algorithm, it is called by applying the Modelling Validation Testing as follows:

**Figure 2. Modelling Validation Testing Support Vector Machine**

In Figure 2, that is clearly seen where the Text Processing process uses Tokenize, Transform Cases, Stop words (Directory). Likewise, the algorithm used by researchers uses Support Vector Machine (SVM). Meanwhile, to get a graph or the results of the ROC Curve in this study, it is by applying the Modelling Validation Testing on Support Vector Machine (SVM) as follows:
In Figure 3, the results of the ROC Curve Support Vector Machine (SVM) study on ROC graph research with AUC states: 0.942 +/- 0.012 (micro average: 0.942) (positive class: Positive).

3. Conclusion
The algorithm testing uses Support Vector Machine (SVM) from 300 user review data applications, 'by.u' review consists of 150 negative reviews or unpolite expression. Then, 150 positive reviews or constructive suggestion or supporting the text in using Indonesian. The resulting model gets the results of testing accuracy using Support Vector Machine with accuracy value: 87.24

References