**PAPER** • **OPEN ACCESS**

# Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method

Journal of Physics: Conference Series **1943** (2021) 012147 doi:10.1088/1742-6596/1943/1/012147

# Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method

**V Oktaviani, B Warsito, H Yasin, R Santoso, Suparti**

Department of Statistics, Diponegoro University, Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia

Corresponding author: viraoktavi10@gmail.com

**Abstract**. E-commerce is a business operation model which rapidly growing today. Many business actors and the customer take advantage of E-commerce itself. Thus, it influences people's socially and economically. Traveloka is one of the best e-commerce applications that is often visited in Indonesia. Each application allows users to post an application review. The review aims to evaluate and improve the quality of the future product. For that purpose, analysis sentiment can be used to classify the review into positive or negative sentiment. Sentiment analysis can provide information that can be extracted. From the observed data, it can provide useful information for those who need it. Some sentiment analysis stages contain sentiment data collection, data preprocessing, term weighting using TF-IDF, sentiment labeling using sentiment scoring, review data classification using the Naïve Bayes Classifier method, and text association. The model was evaluated using 10 Fold Cross-Validation. Measurements were made with the Confusion Matrix. The results obtained from the reviews given by Traveloka users on Google Play using the Multinomial Naïve Bayes was obtained overall accuracy in 91.20% and kappa accuracy in 59,56 %. The higher overall accuracy value and kappa accuracy obtained, the better performance of the classification model.

## 1. Introduction

Internet growth in Indonesia includes in the very fast category. In 2018, the number of internet users was 171.11 million, from 254.12 million Indonesian populations [1]. The increasing of internet users greatly influences various fields; one is the economic and business field. E-commerce is a business in which the process is using communication networks and computers [2]. In Indonesia, the E-commerce industry increases by up to 17% in the last 10 years and reaches 26.2 million e-commerce business units. It is predicted will continue to grow up along the increases of the business actor in Indonesia [3]. Traveloka is one of the biggest e-commerce in Indonesia. Traveloka's concept is a services activity for hotel reservations and transportation tickets sale. Traveloka application on Google Play site successfully achieves the fourth position in the e-ticketing application category, which offers 24 hours of services.

An interesting thing obtained in the Google Play site is a review feature from a particular application user. Reviews received can be an effective and efficient benchmark to find real information from a product. Monitoring and sorting all reviews obtained is not an easy thing, because there are so many reviews loaded in the Google Play site. Therefore, it is needed an extraordinary method to classify the reviews to find real information massively. This study uses Naïve Bayes Classifier and association method to discover user perception against service quality.

## 2. Literature Review

### *2.1. Sentiment Analysis*
Sentiment analysis in a review is a process to investigate reviews of a product on the internet to determine opinion or feeling against the overall products [4]. Information obtained in the form of text is currently widely available on the internet in several formats for forums, social media, and sites containing reviews. With the sentiment analysis, we can process this data into structured data.

### *2.2. Text Mining*
Text mining is a process to extract useful information of corpus documents from time to time by identifying an interesting pattern [5]. Text mining has textual data sources that are not formatted and structured. While data sources in data mining are already formatted and structured, and it is not only text. Text mining has several main stages, namely preprocessing tasks, processed document collection, and core mining operations [5].

### *2.3. Preprocessing*
The text mining data uses formatted data but not structured. Thus, it cannot be used in the next step. Those data must have a preprocessing process to clean up the data from the noise to become smaller and structured [5]. Preprocessing will be done as follows: case folding, removing number, removing punctuation, normalization, and sentence translating. Case folding is a step to generalize the size of letters, removing numbers is the stage for removing numbers in the text, while removing punctuation is removing symbols in text. After going through these stages, the text will be normalized and the text will be translated using a language other than Indonesian.

### *2.4. Sentiment scoring*
After carrying out the preprocessing process, the process of labeling the sentiment classes in the review document. The labeling section is a process to get the expected corpus representation results. This research review results will be divided into positive and negative classes. Scoring process and labeling use sentiment scoring. In sentiment scoring, it occurs the input of sentiment dictionary, negation, and boosterwords. Sentiment dictionary contains corpus words given weight with sentiment power in 1 to 5 (have positive sentiment), and -1 to -5 (have negative sentiment). Boosterwords is a word that can increase and decrease the intensity of word sentiment next to it [6]. Negation word is a word contained in a sentence that can change orientation from an opinion. [6].

### *2.5. Feature Selection*
The purpose of selection is to reduce the number of features involved in determining a target class value, and data resulting in minsunderstandings regarding the target class. Feature selection will be done as follows: stopwords removing, stemming, and tokenizing. Stopwords are used to remove words that have no effect or do not reduce the information contained in the document, but their existence often appears. Stemming is the process of changing the words in a text document into basic words. Tokenizing is the process of cutting text in a document into non-influential or independent pieces of words called tokens.

### *2.6. Term weighting*
Term weighting is a process of giving weight in a Term. Term weighting is performed by calculating the occurrence of term frequency in a document. Term Inverse Document (TF-IDF) is a combination of two weighting schemes that is Term Frequency (TF) and Inverse Document Frequency (IDF*).* If a term occurs in a document is high in this weighting, and the term's occurrence in other documents is

low, then the weight obtained is getting higher. Otherwise, if a term occurs in other text is high, then the weight is getting lower. Term Frequency-Inverse Document Frequency (TF-IDF) can be calculated using models as follows [7]:

$$W_{j,i} = \frac{n_{j,i}}{\sum_k n_{k,i}} . \log_2 \frac{D}{d_j} \tag{1}$$

With: $n_{j,i}$: the number of occurrence for a term to $j$ in the document to $i$, $\sum_k n_{k,i}$: the number of all terms occurrence in $i$ document, $D$: the number of a document generated, and $d_j$ : the number of the document contains terms to $j$.

### 2.7. *Naïve Bayes Classifier*

Naïve Bayes Classifier is one of classification algorithm discovered by Thomas Bayes. Classification is a process to g rules or models, which can classify new data never been studied yet by learning old corpus data [8]. General principle of Naïve Bayes Classifier assumes that value from an attribute does not depend and influence other attributes. Naïve Bayes uses basic concepts from Bayes theorem with calculate probability value as follows [8]:

$$P(C = c_i \mid D = d_j) = \frac{P(C=c_i \cap D=d_j)}{P(D=d_j)} \tag{2}$$

Where $P(C = c_i \mid D = d_j)$ is the category of probability if documents is known: From model (2) can make model (3):

$$P(C = c_i \mid D = d_j) = \frac{P(C = c_i \cap D = d_j)}{P(D = d_j)}$$
$$= \frac{P(D = d_j \mid C = c_i) \times P(C = c_i)}{P(D = d_j)} \tag{3}$$

Where $P(D = d_j \mid C = c_i)$ are the probability value of $d_j$ document occurrence if the documents is known have a category; $c_i$, $P(C = c_i)$ are the probability value of $c_i$, document occurrence; $P(D = d_j)$ are the value of $d_j$ document occurence. Text classification is done first to determine $c \in C$ category from a $d \in D$ document where $C = \{c_1, c_2, c_3, \ldots, c_i\}$ , $D = \{d_1, d_2, d_3, \ldots, d_i\}$, and $P(C = c_i \mid D = d_j)$ have maximum value from $P = \{P(C = c_i \mid D = d_j \mid c \in \mathrm{C} \text{ and } \mathrm{d} \in D\}$.

Naïve Bayes assumes a document as corpus words arrange the document itself, and do not notice the order of word occurrence in a document. Thus, probability calculation can calculate in the model (4).

$$P(C = c_i \mid D = d_j) = \frac{\prod_k P(W_k \mid C=c_i) \; x \; P(C=c_i)}{p(W_1, W_2, W_3, \ldots, W_k, \ldots, W_n)} \tag{4}$$

With $\prod_k P(W_k \mid C = c_i)$ is a calculation and multiplication of occurrence probability in all the $d_j$ document words.

The classification process has done by making a probabilistic model from a training document with calculating $p(w_k|c)$ value. Thus, it can be found the probability for all values using (5) and (6) model [8].

$$P(w_k = w_{kj}|c) = \frac{D_b(w_k = w_{kj}|c) + 1)}{D_b(c) + |V|} \tag{5}$$

The function that restores the amount of b document in c category, which has the word value of $w_k = w_{kj}$ can be expressed with $D_b(w_k = w_{kj,}c)$. $D_b(c)$ is a function that restores the amount of b document, which has c category, and $|V|$ is the amount of $w_{kj}$ value possibility.

Laplace Smoothing is often combined with the $D_{b\,=}(w_k = w_{kj,}c)$ equation to prevent the equation, which gets 0 value and disturb all of the classification results.

$$P\left(w_k = w_{kj} \middle| c\right) = \frac{D_b(c)}{|D|} \tag{6}$$

$D_b(c)$ is a function that restores the amount of b document, which has c category, and the amount of all training which can be expressed with |D|. Giving category in a text document can be done by choosing the maximum of c value in $P(C = c_i \mid D = d_j)$ value, such as in (7) model.

$$c^* = \underset{c \in C}{arg\,max} \prod_k P(w_k|c) \times P(c) \tag{7}$$

C category* is a category that has the maximum of $p(C = c_i \mid D = d_j)$ value.

### 2.8. *K-Fold Cross-Validation*

Cross-validation is one of the methodologies for predicting a new model built based on the corpus dataset and predicting how accurate a model is when running in their practice. The data is divided into k segments which have the same or nearly the same ratio. Performed repetitions of k times for training and testing. Training on data and validation k times with each experiment taking one different segment as test or validation data and the other k-1 segments as training data to take the average of the results of each iteration [9].

### 2.9. *Evaluation*

A classification system must rate the performance to measure the level accuracy from classification prediction, which already produced. Some of the methodologies can be used to rate classification performance, such as the confusion matrix. The number of sizes is used to evaluate the classification model obtained, such as overall accuracy and kappa accuracy. Kappa accuracy is highly recommended because its accuracy calculation uses all confusion matrix [10].

**Table 1.** Confusion matrix

| Class | Positive | Negative |
|---|---|---|
| Positive | True Positive (TP) | False Positive (FP) |
| Negative | False Negative (FN) | True Negative (TN) |

**Table 2.** The Evaluation Size of Classification Model

| Size | Formula |
|---|---|
| Overall Accuracy | $\frac{TP + TN}{P + N}$ |
| Kappa Accuracy | $\frac{([P+N][TP+TN] - ([TP+FN][TP+FP] + [FP+TN][FN+TN]))}{(P+N)^2 - ([TP+FN][TP+FP] + [FP+TN][FN+TN])}$ |

### 2.10. *Association*

Correlation is often used to state two or more variable relations quantitatively. In contrast, the association term is often interpreted as the closeness relation between two or more variables qualitatively [11]. Correlation analysis aims to find the relation between independent and dependent variables.

This study uses an association approach to get the relation between user reviews in obtaining helpful information for a particular party.

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\{n\sum x_i^2 - (\sum x_i)^2\}\{n\sum y_i^2 - (\sum yi)^2\}}} \quad (8)$$

## 3. Data and Methodology

This study used to review data that occurred in the Traveloka application on Google Play Store from September, 23 2019 – November, 23 2019. Data were obtained with a Scrapper application. Those data were analyzed using Naïve Bayes Classifier and Association method.

## 4. Result and Discussion

### *4.1. Data Preprocessing*

In this step has done the data transformation from unstructured data becomes structured data to simplify the next analysis step. The preprocessing step conducted is case folding, removing numbers, removing punctuation, translating, and normalization. Preprocessing is very required; thus, during the classification step, the result obtained is more optimal.

### *4.2. Data Labeling*

Data labeling has done using sentiment scoring. Every review in the text document, the score has calculated according to the following conditions:

1. Every word in the review which is available in sentiment dictionary will get score corresponding to sentiment dictionary, if those word is not available in sentiment dictionary will get 0 scores.
2. Every word in the review which contains negation word in the previous word will get the opposite score with a sentiment dictionary.
3. If a word in the review is in sentiment dictionary valued in >0 followed with boosterwords in the previous or next word, then the score of sentiment word is added with the boosterwords score.
4. If a word in the review is in sentiment dictionary valued in <0 followed with boosterwords in the previous or next word, then the score of sentiment word is reduced with boosterwords score.

### *4.3. Feature Selection*

The number of reviews in this study caused the amount of feature formed and meaningful word choice. Stopwords used in this study are 504 words. After a document through the stopwords removing process then proceed to the stemming. The stemming step is a step to change words in the text document to become root words. Stemming has been done to make words with the same meaning and the different forms become the same. In the tokenizing process has done document cutting becomes parts of words called token. Space is used to separating between those words.

### *4.4. Term Weighting*

Term weighting in this study used Frequency-Inverse Document Frequency (TF-IDF) calculation. In this process, all terms could be processed so that every term in the document has its weight. After that, the corpus document was immediately ready for the training process to the classification process. The value of term weighting with Term Frequency-Inverse Document Frequency (TF-IDF) could be seen in Table 3.

**Table 3.** Term weighting using TF-IDF

| review | baik | cepat | maaf | jelas | mudah | … | suka |
|---|---|---|---|---|---|---|---|
| **12** | 2.6807 | 0 | 0 | 0 | 0 | … | 0 |
| **148** | 2.6807 | 0 | 0 | 0 | 0 | … | 0 |
| **709** | 0.3351 | 0 | 0.6685 | 0 | 0 | … | 0.5605 |
| **1763** | 0.3351 | 1.2668 | 0.6685 | 0 | 0.9483 | … | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
| **4090** | 0 | 0 | 0 | 8.676 | 0 | … | 0 |

### *4.5. Classification*

The algorithm of Naïve Bayes used training data to establish a classification model. The model that was established would be used as a class prediction of new data like never before. Training data and test data used were data that already have labels with the comparison, respectively 80%: 20%.

**Table 4.** The Sentiment Class Proportion in Labeling Result of Training and Testing Data

| Classification | Positive | Negative | Amount |
|---|---|---|---|
| Training Data | 1025 | 326 | 1351 |
| Test Data | 117 | 32 | 149 |
| Amount | 1142 | 358 | 1500 |

In the classification process, it used the Naïve Bayes Classifier method done by making machine learning using training data and test data, which was taken randomly called K-Fold Cross-Validation. K-Fold Cross-Validation is used to discover machine learning in classification form. K value that is used in this K-Fold Cross-Validation is 10. The total of data reviews was 4090 if using K-Fold Cross-Validation would produce 409 reviews data for every fold. Looping has been done until 10 times, where 3272 reviews data was used as training data and 818 reviews data as testing data.

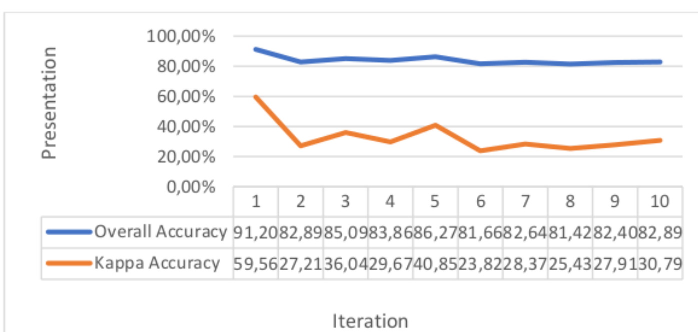


**Figure 1**. The result of Naïve Bayes classification with 10-fold cross-validation

Journal of Physics: Conference Series **1943** (2021) 012147 doi:10.1088/1742-6596/1943/1/012147

### *4.6. Evaluation*

According to Figure 18, the result accuracy using Naïve Bayes Classifier has a little different in every fold. It is obtained 91.20% of the average value accuracy. With the average value accuracy of 91.20%, the Naïve Bayes Classifier method is suitable for reviews data classification with *Bahasa Indonesia* text. The classification process has done 10 times using the Naïve Bayes Classifier model. The model has taken by the highest overall accuracy and kappa accuracy value in the first iteration with 91.20% and 59.56%, respectively.

*4.6.1. Positive Reviews Data.* Positive Reviews data is a labeling result categorized in a positive class. Labeling in the review had done using the calculation of sentiment scoring. The positive result was identified based on the term frequency in the review. The following was a visualization result of positive reviews from information extraction results obtained according to user-written reviews.

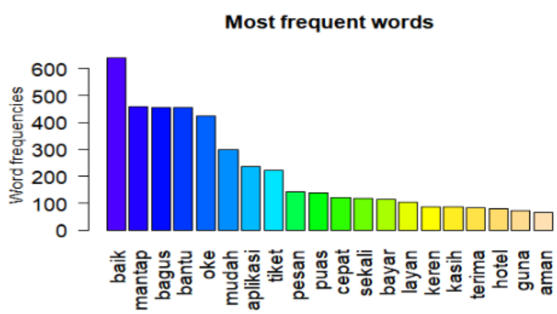


**Figure 2**. the highest frequency of term in positive class

Based on figure 2 the term is often used to give review by Traveloka user in positive class, include the term “baik” 639 times, “mantap” 457 times, “bagus” 454 times, et cetera. The highest frequency term in this above picture are the most positive topic discussed by Traveloka user. This term will be used as a basic to find association with other term.

**Table 5.** Positive association term

| Promosi | | Cepat | | Keren | | Tiket | |
|---|---|---|---|---|---|---|---|
| Gencar | 0,24 | Respon | 0,25 | Diskon | 0,21 | Pesan | 0,51 |
| Konsisten | 0,24 | Akurat | 0,18 | Spesial | 0,21 | Beli | 0,43 |
| Wajar | 0,24 | Tepat | 0,17 | | | Cari | 0,21 |
| Kupon | 0,21 | Proses | 0,15 | | | Murah | 0,21 |
| | | | | | | Elektronik | 0,15 |

*4.6.2. Negative Reviews Data.* It was identified 450 negative reviews from 4090 reviews. It shows that some Traveloka application users have deficient preception. The information extraction results were identified according to term frequency in negative reviews. If a term frequency is higher, it shows that users often talk that term.

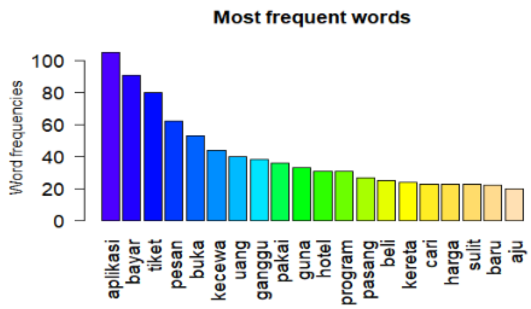


**Figure 3.** the highest frequency of term in positive class

The highest frequency term in this above picture are the most negative topic discussed by Traveloka user. This term will be used as a basic to find association with other term.

**Table 6.** Negative term association

| Aplikasi | | Bayar | | Sistem | | Pesan | |
|---|---|---|---|---|---|---|---|
| Tipu | 0,23 | Tagih | 0,44 | Pindah | 0,21 | Sial | 0,23 |
| Ampun | 0,18 | Kredit | 0,42 | Tetangga | 0,21 | Batal | 0,21 |
| Pikir | 0,18 | Hutang | 0,39 | Payah | 0,24 | Respon | 0,22 |
| Pindah | 0,18 | Denda | 0,23 | Gila | 0,21 | Balas | 0,18 |
| Program | 0,16 | Lambat | 0,20 | | | Bohong | 0,18 |
| Tetangga | 0,18 | Tolak | 0,19 | | | Kendala | 0,17 |

*4.6.3. Monthly data review.* The process of extracting informations is carried out on all the reviews based on the month of the review to see overall whether the user gave much negative or positive review on the Traveloka application.

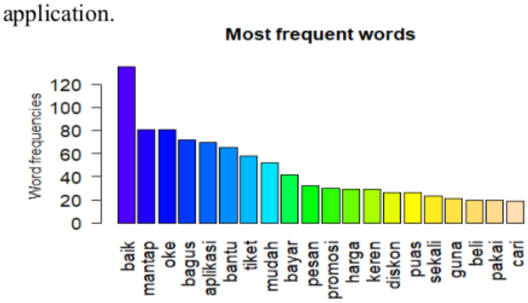


**Figure 4.** the highest frequency of term on September 2019

Based on figure 3 the term is often used to give review by Traveloka user on September, include the term “baik” 135 times, “mantap” 81 times, “oke” 81 times, et cetera. In September, user responses are much more positive review than negative one.

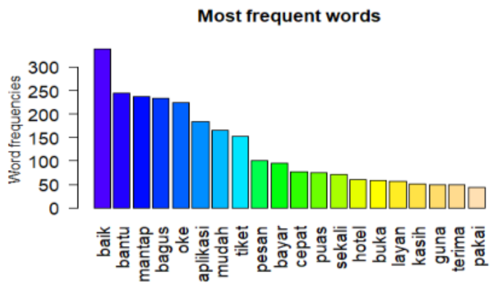


**Figure 5.** the highest frequency of term on October 2019

Based on figure 3 the term is often used to give review by Traveloka user on October, include the term “baik” 338 times, “bantu” 244 times, “mantap” 238 times, “bagus” 234 times, et cetera. In September, user responses are much more positive review than negative one.

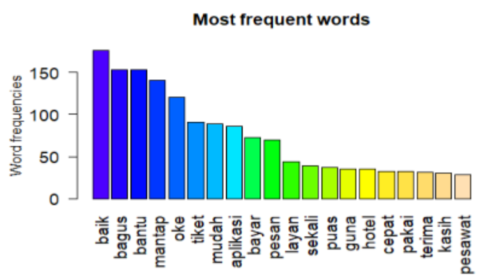


**Figure 6.** the highest frequency of term on November 2019

Based on figure 5 the term is often used to give review by Traveloka user on November, include the term “baik” 176 times, “bagus” 153 times, “bantu” 153 times, “mantap” 141 times, et cetera. In November, user responses are much more positive review than negative one. The negative review given is “disappointment” 25 times.

## 5. Conclusion

Classification sentiment from the result of data transmission using *sentiment scoring* using *Naïve Bayes Classifier* method on the Traveloka application with 80% comparison of training data and 20% of test data produces the best overall accuracy is 91,20%, and the best kappa accuracy is 59,56%. According to the overall classification and textual association is noted that Traveloka users talk mostly about baik, aplikasi, bantu, mudah, and tiket for the top 5 term on the overall review. In the Traveloka positive class, Traveloka users talk a lot about baik, mantap, oke for the top 5 term by extracting information obtained from positive association including the term promosi, cepat, keren, and tiket In the Traveloka negative class, Traveloka users talk a lot about aplikasi, bayar, tiket, pesan, and buka for the top 5 term by extracting information obtained from negative association including the term aplikasi, sistem, bayar, and pesan.

## References

[1] Asosiasi Penyelenggara Jasa Internet Indonesia 2018 *Penetrasi & Perilaku Pengguna Internet Indonesia.* https://apjii.or.id/survei. Accessed on 20 December 2019
[2] Indonesian E-Commerce Association 2019 https://www.idea.or.id/berita. Accessed on 20

December 2019
[3] Pearson, Mc Leod, and Schell, George 2008 *E-Commerce: Business, Technology, and Society: 4th Edition* (NewJersey: PrenticeHall, Inc).
[4] Feldman R 2013 *Commun. ACM* **56**
[5] Feldman R and Sanger J 2007 *The text mining handbook: advanced approaches in analyzing unstructured data* (New York : Cambridge university press)
[6] Wahid D H and Azhari S N 2016 *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* **10** 2 207-218
[7] Salton G and Buckley C 1988 *J. Inf. Proc. Manag.*
[8] Mitchell T M 2010 *Generative and Discriminative Classifiers: Naïve Bayes and Logistic Regression* (Helena : Morgan Kauffman Publisher)
[9] Haddi E, Liu X and Shi Y 2013 Procedia Comput. Sci. **17** 26-32
[10] Arisondang V, Sudarsono B and Prasetyo Y 2015 *J. Geodecy* Diponegoro University
[11] Ulwan M N 2016 *Pattern Recognition pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association*

Oktaviani_2021_J._Phys.__Conf._Ser._1943_012147_ISNPINSA...

ORIGINALITY REPORT

| 15% | 14% | 13% | 9% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Universitas Diponegoro<br>Student Paper | 5% |
| 2 | jurnal.kdi.or.id<br>Internet Source | 2% |
| 3 | Natalia Syafitri Kustanto, Intan Nurma Yulita, Indra Sarathan. "Sentiment Analysis of Indonesia's National Health Insurance Mobile Application using Naïve Bayes Algorithm", 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021<br>Publication | 2% |
| 4 | jurnal.polgan.ac.id<br>Internet Source | 2% |
| 5 | E M Akhmetshin, A V Plotnikov. "Sentiment analysis of client reviews of the Russian Agricultural Bank service and predicted rating reviews", IOP Conference Series: Earth and Environmental Science, 2020<br>Publication | 1% |

| | | | |
|---|---|---|---|
| 6 | Kornelius, Wella. "Is the Internet in Indonesia Has a Good Sentiment from Netizen?", 2021 6th International Conference on New Media Studies (CONMEDIA), 2021<br>Publication | 1 % |
| 7 | ejournal3.undip.ac.id<br>Internet Source | 1 % |
| 8 | Salvador García, José Ramón Cano, Alberto Fernández, Francisco Herrera. "Chapter 168 A Proposal of Evolutionary Prototype Selection for Class Imbalance Problems", Springer Science and Business Media LLC, 2006<br>Publication | 1 % |
| 9 | Zainab A. Khalaf, Zainab M. Jawad. "Measuring the Impact of Using Different Tools on Classification System Results", Journal of Physics: Conference Series, 2020<br>Publication | 1 % |
| 10 | iopscience.iop.org<br>Internet Source | 1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |