

PUBLIC SENTIMENT ANALYSIS AGAINST IDENTITAS KEPENDUDUKAN DIGITAL APPLICATION USING DECISION TREES

Fernanda Al Khafid Faridian^{1,*}, Sulistiyasni², Lutvi Riyandari³

^{1,2,3}Informatics Engineering STMIK Widya Utama Jl Sunan Kalijaga, Berkoh, South Purwokerto, 53146, Indonesia

^{1,*}Corresponding email: fernandaalkhafid@gmail.com,
²sulistiyasnipwt@swu.ac.id,
³lutviriyandari@gmail.com

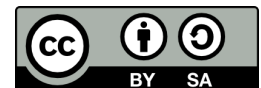
Received: September 05, 2024; Revised: August 12, 2025; Accepted: December 02, 2025.

Abstract

The Identitas Kependudukan Digital (IKD) application is an innovation in the government sector that is expected to facilitate public administration through electronic means. The implementation of the Identitas Kependudukan Digital application has sparked a public debate on its application and security. This study aims to analyze public sentiment towards the Identitas Kependudukan Digital application using the Decision Tree algorithm and to present the performance of the Decision Tree algorithm in terms of positive and negative sentiment scores. The author collected review data from Google Play for the Identitas Kependudukan Digital application using web scraping techniques with Google Collaboration. Based on the analysis results, the Decision Tree algorithm predicted 240 negative and 4 positive data points out of 244 test data points from the review data of the Identitas Kependudukan Digital application. From the results of the tests and evaluation using the Decision Tree algorithm, the precision value is 89.34%, the precision value is 90.00% and the recall value is 99.08%. Based on the results of word frequency on WordCloud, it is found that the words "Aplikasi" and "Data" are the words that have the highest frequency of words, where these words are widely reviewed by the community, so that they represent the cause of the high negative percentage in the analysis of community sentiment towards the Identitas Kependudukan Digital application.

Keywords: *Data Mining, Decision Tree, Identitas Kependudukan Digital, Sentiment Analysis*

This is an open access article under the CC BY-SA license.



Corresponding Author:

*Fernanda Al Khafid Faridian

Power Electronics and Renewable Energy Research Laboratory (PEAR-L), Informatics Engineering STMIK Widya Utama

Jl Sunan Kalijaga, Berkoh, South Purwokerto, 53146, Indonesia

Email: fernandaalkhafid@gmail.com

I. INTRODUCTION

Residents aged 17 years and above are required to have an Electronic Identity Card (KTP-el), which is valid in Indonesia. The rules for ownership of the Electronic Identity Card (KTP-el) are regulated in Law Number 24 of 2013 concerning Amendments to Law Number 23 of 2006 concerning Population Administration in Article 63 of the Law on Population Administration [1]. To realize the laws and regulations in population administration, good population administration services are needed so that every resident in Indonesia can have an Electronic Identity Card (KTP-el). But currently, there are still obstacles in population administration services. One of the constraining factors is the limited number of KTP-el blanks in population administration services. As happened at the DKI Jakarta Population and Civil Registration Office (Dukcapil) based on a Circular Letter from the Ministry of Home Affairs of the Republic of Indonesia Number 471.13/17740 / Dukcapil in 2022, a certificate (suket) and Identitas Kependudukan Digital (IKD) is issued as a substitute for an Electronic Identity Card (KTP-el) that has not been printed and the document is proof that the resident concerned has recorded the KTP-el and has been recorded in the population database [2]. In line with technological developments, Identitas Kependudukan Digital comes as a solution to the limitations and the absence of additional stock on KTP-el blanks [3]. Identitas Kependudukan Digital (IKD) is an innovation in the field of government that is expected to facilitate the existence of Identitas Kependudukan Digital with electronic-based public administration to improve service quality and provide space for the community to actively participate in the process of providing good services. The existence of Identitas Kependudukan Digital also raises debates in the community regarding the application and security of the Identitas Kependudukan Digital application. We can see the assessment of the Identitas Kependudukan Digital application on social media or digital distribution services such as the Google Play site to convey opinions, opinions, or assessments of the Identitas Kependudukan Digital application.

In this study, several previous studies were used as supporting research material. Sentiment analysis on Shopee application ratings using the SMOTE-based Decision Tree method. This research was conducted to determine the level of accuracy and opinions of Shopee application users. The Decision Tree algorithm with SMOTE yields an accuracy of 99.91%, an AUC of 0.999, a recall of 99.88%, and a precision of 99.98%. The Decision Tree algorithm without SMOTE produces an accuracy value of 99.89%, AUC 0.950, recall 99.88%, and a precision value of 99.98%. The results of the existing accuracy and AUC values, SMOTE can affect the accuracy and AUC values, while the precision and recall values have no effect even whether using or without SMOTE. The difference in the accuracy value obtained is 0.02%, and the AUC is 0.049 [4]. Application of the Support Vector Machine algorithm in sentiment analysis of Identitas Kependudukan Digital. This research aims to present information about people's opinions and assessments of Identitas Kependudukan Digital. Testing and evaluation using the Support Vector Machine algorithm produces an accuracy value of 77.0%, an average precision value is 92%, an average recall value is 36%, and an average f1-score value is 35% [5]. Comparison of classification methods in analyzing public sentiment towards Identitas Kependudukan Digital (IKD). This research tries to compare Naive Bayes, K-NN, SVM, and Neural Network methods, and find out the best sentiment and performance of each method used. This research shows that the K-NN method is the best method for analyzing sentiment related to Identitas Kependudukan Digital with 100% accuracy, precision, and recall. There is little difference in performance between the Neural Network method and the Naive Bayes method, but the SVM method has lower performance [6].

Based on the background that has been explained, the authors conducted research on analyzing public sentiment toward the Identitas Kependudukan Digital application by applying the Decision Tree algorithm. The research data was taken from the reviews in the comments column of the Identitas Kependudukan Digital application on the Google Play site. This research was conducted to provide information about people's opinions and assessments of the Identitas Kependudukan Digital application and to determine the level of achievement generated by the Decision Tree algorithm method so that it can become actionable insight for policymakers.

II. LITERATURE REVIEW

A. Sentiment Analysis and Decision Trees

Data mining is the process of extracting or mining large amounts of data and information from large databases that were previously unknown but can be understood, useful, and used to make very important business decisions. Data mining is part of the KDD (Knowledge Discovery in Databases) process and consists of several stages, such as data selection, preprocessing, transformation, data mining, and evaluation of results. Knowledge Discovery in Databases is also commonly referred to as a database [7]. Based on the

Regulation of the Minister of Home Affairs Number 72 of 2022 concerning Standards and Specifications for Hardware, Software, and Electronic Population Identity Card Stamps and the Implementation of Identitas Kependudukan Digital. Identitas Kependudukan Digital (IKD) is information in electronic form to display population documents and feedback data on digital applications via smartphones that can be downloaded on the App Store and Play Store applications [8]. Identitas Kependudukan Digital comes as a solution to the limitations and absence of additional stock on the KTP-el stamp [3]. However, the existence of Identitas Kependudukan Digital also raises debates in the community regarding the application and security of the Identitas Kependudukan Digital application.

We can see the assessment of the Identitas Kependudukan Digital application on social media or on digital distribution services such as the Google Play site to convey opinions, opinions, or assessments of the Identitas Kependudukan Digital application. Sentiment analysis is carried out to determine the public's assessment of the Identitas Kependudukan Digital application. Sentiment analysis is a text mining technique used to identify and evaluate the emotions of text. Sentiment analysis is used to determine whether a text is positive, negative, or neutral. Generally, this technique involves the use of classification algorithms that have been trained using text data that has been manually labeled so that they can categorize new text as positive, negative, or neutral [9]. The decision tree algorithm is a commonly used method for decision-making. Decision trees find solutions to problems by converting criteria into nodes that are connected to form a tree-like structure. Decision trees also represent a predictive decision-making model with a hierarchical or tree structure, where each tree has branches. These branches symbolize the attributes that must be met to move to the next branch and end up in the leaves (no branches left) [10]. Therefore, the Decision tree algorithm is used in this research because it is easy to interpret and easy to understand.

B. Confusion Matrix

Confusion Matrix is a matrix containing correct and incorrect prediction values. When testing the accuracy of search results, recall, precision, and accuracy values will be evaluated. Precision evaluates the system's performance in finding the most relevant rankings and is defined as the portion of retrieved documents that are relevant to the query. Recall evaluates the system's performance in finding all relevant items from the document set and is defined as the portion of documents that are relevant to the query. Accuracy is the ratio of correctly identified cases to the total number of cases, and error rate is the ratio of incorrectly identified cases to the total number of cases [11]. Confusion Matrix consists of 4 parts, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [12]. An example of a Confusion Matrix table is in Table I.

TABLE I. CONFUSION MATRIX

Classification	Predicted Positives	Predicted Negatives
Actual Positive Cases	Number of True Positive Cases (TP)	Number of False Negative Cases (FN)
Actual Negative Cases	Number of False Positive Cases (FP)	Number of True Negative Cases (TN)

III. RESEARCH METHOD

This research uses 2 variables. Variable X is the Identitas Kependudukan Digital application, while variable Y in this study is the public sentiment contained in the Identitas Kependudukan Digital application review on the Google Play site.

A. Data Collection

Data Collection is carried out on Identitas Kependudukan Digital application reviews on the Google Play site using the Web Scrapping technique utilizing Google Collaboration. The review data was downloaded and will be processed using the Identitas Kependudukan Digital application review data from January 2023 to January 2024 as much as 1000 review data.

B. Text Preprocessing

1) Cleansing

Cleansing is done to clean or remove some characters outside of the a-z alphabet in the research dataset, such as emojis, numbers, and punctuation marks [13]

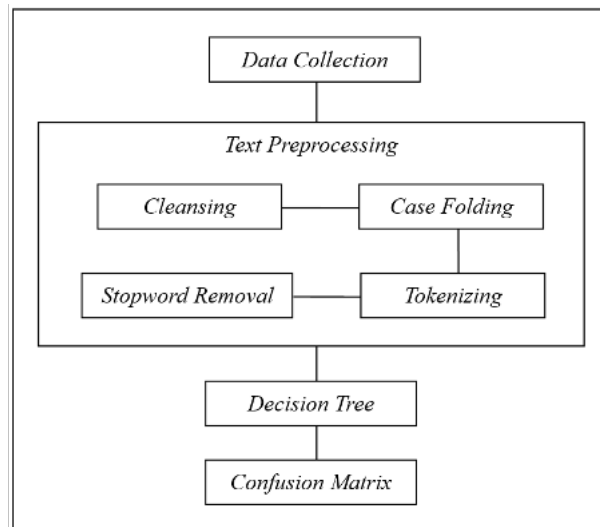


Fig. 1. Research Stages

TABLE II. TEXT PREPROCESSING

Text Preprocessing	Before	After
Cleansing	Kuota saya full, tapi ga bisa masuk tulisannya kesalahan koneksi Pak Ibu kalau aplikasi belum siap jangan di launching dulu kecewa euy	Kuota saya full tapi ga bisa masuk tulisannya kesalahan koneksi Pak Ibu kalau aplikasi belum siap jangan di launching dulu kecewa euy
Case Folding	Kuota saya full tapi ga bisa masuk tulisannya kesalahan koneksi Pak Ibu kalau aplikasi belum siap jangan di launching dulu kecewa euy	kuota saya full tapi ga bisa masuk tulisannya kesalahan koneksi pak ibu kalau aplikasi belum siap jangan di launching dulu kecewa euy
Tokenizing	kuota saya full tapi ga bisa masuk tulisannya kesalahan koneksi pak ibu kalau aplikasi belum siap jangan di launching dulu kecewa euy	['kuota', 'saya', 'full', 'tapi', 'ga', 'bisa', 'masuk', 'tulisannya', 'kesalahan', 'koneksi', 'pak', 'ibu', 'kalau', 'aplikasi', 'belum', 'siap', 'jangan', 'di', 'launching', 'dulu', 'kecewa', 'euy']
Stopword Removal	['kuota', 'saya', 'full', 'tapi', 'ga', 'bisa', 'masuk', 'tulisannya', 'kesalahan', 'koneksi', 'pak', 'ibu', 'kalau', 'aplikasi', 'belum', 'siap', 'jangan', 'di', 'launching', 'dulu', 'kecewa', 'euy']	kuota full masuk tulisannya kesalahan koneksi aplikasi launching dulu kecewa

2) *Case Folding*

At this stage, the author converts uppercase (capital) letters into lowercase letters so that all word forms in the dataset become the same [5].

3) *Tokenizing*

Tokenizing aims to separate text in the form of sentences into words [5].

4) *Stopword Removal*

In the Stopword Removal stage, a list of common words that have no significance and are not used will be removed to reduce the number of words stored in the system [13].

C. *Split Data*

At this stage, the data is split into training data and test data with a ratio of 70:30. Training data is the data used to run the Decision Tree algorithm training process, and test data is used to run tests on the Decision

Tree after training is run. The selection of the 70:30 ratio is based on the rule of thumb principle widely used in machine learning practice, as explained by Andrew Ng, where this ratio is considered quite balanced for small to medium-sized datasets (around 100 to 10,000 data) [14]. The 70:30 split was chosen in this study because it considers computational efficiency, considering that the dataset size used is in the medium scale and the training time is relatively fast. This data division process is carried out using Rapid Miner on the Identitas Kependudukan Digital application review text data taken from January 2023 to January 2024.

D. Decision Tree

This Decision Tree classification predicts labels based on training that has been done with training data that has been labeled. In determining the root of the tree, you can use the maximum gain value of each attribute or the minimum entropy index value. The step taken to find the entropy value is to use the equation in Formula One. After that, calculate the gain value using the equation in formula two [4]. Classification is done using Rapid Miner, in the parameter settings with Gain_Ratio as the criterion and maximal depth set at 10.

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (1)$$

n = Number of attributes A
 S = Set of cases
 \log_2 = Logarithm base two
 p_i = The proportion S_i to S

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S) \quad (2)$$

A = Attributes A
 S = Set of cases
 n = Number of attributes A
 $|S|$ = Number cases in S
 $|S_i|$ = Number of Cases of the i^{th} partition

E. Confusion Matrix

The Identitas Kependudukan Digital application review data was tested using a division of 70% training data and 30% test data. Accuracy, precision, and recall values are determined from the Confusion matrix. The rating scale is usually given in percentages from 1% to 100%. A system is said to be good if the accuracy, precision, and recall values have a high percentage level. The accuracy score is an assessment of the data classification accuracy of the algorithm model used. Accuracy is achieved by comparing correctly classified positive and negative data to the overall data [15]. The following is the formula for finding the accuracy value.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (3)$$

Precision is the correspondence between the collected data and the required information. The precision value is the number of correctly classified positive data divided by the total number of positively classified data [15]. The following is the formula for finding the precision value.

$$Precision = \frac{TP}{FP + TP} * 100\% \quad (4)$$

Recall is the success rate of the system in retrieving information. The recall is a performance measure expressed in terms of the percentage of positive category data that is classified as negative [15]. The following is the formula for finding the Recall value.

$$Recall = \frac{TP}{FN + TP} * 100\% \quad (5)$$

IV. RESULTS & DISCUSSION

A. Data Collection

Data Collection is carried out on the Identitas Kependudukan Digital application review on the Google Play site using the Web Scraping technique utilizing Google Collaboration. The first step is to download and install the Google Play scraper package, and then import the required package. Find the ID of the Identitas Kependudukan Digital application on Google Play, and enter the reviews into the Pandas DataFrame as shown in Figure II. The review criteria taken in the scrapping process are taking review data as much as one thousand data points, the review data stored is only the username, review rating, review date, and review content, and only uses relevant data.

```

!pip install google-play-scraper

from google_play_scraper import app
import pandas as pd
import numpy as np

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    "gov.dinasagril.mobilis_id",
    lang="id",
    country="id",
    sort=Sort.HIGHEST_RELEVANT,
    count=1000,
    filter_score_with=None
)

df_busu = pd.DataFrame(np.array(result), columns=["review"])
df_busu = df_busu.join(pd.DataFrame(df_busu.pop("review").tolist()))
df_busu.head()

len(df_busu.index)

df_busu[["username", "score", "at", "content"]].head()

new_df = df_busu[["username", "score", "at", "content"]]
sorted_df = new_df.sort_values(by="at", ascending=False)
sorted_df.head()

my_df = sorted_df[["username", "score", "at", "content"]]

my_df.head()

my_df.to_csv("scraped_dataikd.csv", index = False)
    
```

Fig. 2. Scrapping Process

The data successfully taken from the Identitas Kependudukan Digital application reviews will be stored in Excel with the Comma Separated Values File (CSV) format to make it easier to label the data. The data that was successfully taken was 1000 data, but this study only used review data from January 2023 to January 2024 so that the amount of data outside that period would be deleted manually and the remaining data was 920 data. The example of the review data that has been saved can be seen in Figure 3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	username	score	at	content																
2	Rahwadi Putraluwu	1	1/31/2024 10:19	Sudah login berkali-kali tetap saja yg muncul jaringan tidak stabil, padahal jaringan yg kita gunakan stabil, apk ap in saat di butuhkan enggak mau kebuka, harus menggunakan jaringan WiFi.																
3	Sri Wahyuni	5	1/31/2024 7:55	Sangat membantu cuma masih bingung apa KTP digital perlu diprint atau difotocopy soalnya suka butuh kalau ada keperluan terimakasih																
4	Heru OB	1	1/31/2024 7:43	ini data nya ga sinkron sama kementerian kesehatan atau gimana sih? di aplikasi peduli lindungi aku dah vaksin ke 3 atau vaksin Booster pertama tapi di aplikasi IKD vaksin nya cuma ada 2																
5	Rivan Prasetya	2	1/31/2024 2:35	Pelajaran pertama kenapa dan untuk apa aplikasi berbasis digital dibuat.. Kalo kata dosen saya, aplikasi di buat untuk mempermudah user . Di sini saya mau daftar tapi harus ada kode QR																
6	Afifah Yolonda febrina Sari	1	1/30/2024 12:45	Kirain cuma aku aja yang gk bisa buka aplikasi nya, ternyata hampir semua pengguna aplikasi 0% ... Masa iya jaringan 4g full gk bisa buka, memuat gagal suruh coba lagi. Ehh busoect, gir																
7	Melati Indri wahayuni	1	1/30/2024 9:42	Aplikasi ga guna padahal wili data lancar buak yang laen tapi buak nih apk memuat trussas ga ada rimbanya ,tolong yang jadi developer ny yang benar lah di permudah make apk ny blun																
8	Ravyan Ranja	1	1/30/2024 3:13	Mau buka lihat ktp si ga bisa. Pdhli butuh buat di bank. Krn ktpku ketinggalan. Dulu aplikasi apa sih ini! Tolong dong,min di perbaiki ini menyusahkan sekali.																
9	tb arief Rahman	1	1/30/2024 2:53	ketika perbalikan data KTP karna pindah, sy tidak d cetakkan KTP fisik barunya hanya d beri aplikasi ini. pada saat urgon k rumah sakit, untuk menggunakan asuransi hrs menunjukan kartu a																
10	Yessy Kurnale sary	1	1/30/2024 2:51	Sinyal stabil tapi pas mau di buka apk sering "koneksi terputus" apk bukan nyh memudahkan kan tapi malah menyusahkan kan untuk membuat ktp																
11	Dunia Jenal	1	1/30/2024 2:33	Mau di buka engga bisa...bacaanya kesalahan jaringan padahal mah buka YouTube aja lancar....tolong di perbaiki secepatnya																

Fig. 3. After the "username" and "at" Attributes are Removed

B. Text Preprocessing

At the text preprocessing stage using Rapid Miner software. Before doing text preprocessing in Rapid Miner, data deletion is carried out in excel documents at the "username" and "at" attributes manually because the data needed in this study is only at the "score" and "content" attributes. Furthermore, manually deleting review data on data that has a score of 3 (neutral) because this study only uses positive and negative data so that the remaining amount of data becomes 820 data. An example of deleting Excel document attributes is shown in Figure 4.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	userName	score	at	content																	
2	Rahwadi Putraluwu	1		1/31/2024 10:19	Sudah login berkali-kali tetap saja yg muncul jaringan tidak stabil, padahal jaringan yg kita gunakan stabil, apk ap in saat di butuhkan enggak mau kebuka, harus menggunakan jaringan WiFi, tolong untuk di sesuaikan lagi																
3	Sri Wahyuni	5		1/31/2024 7:55	Sangat membantu cuma masih bingung apa KTP digital perlu diprint atau difotocopy soalnya suka butuh kalau ada keperluan terimakasih																
4	Hieru GB	1		1/31/2024 7:43	Ini data nya ga sinkron sama kementerian kesehatan atau gimana sih? di aplikasi peduli lindungi aku dah vaksin ke 3 atau vaksin Booster pertama tapi di aplikasi IKD vaksin nya cuma ada 2																
5	Rivan Prasetya	2		1/31/2024 2:35	Pelajaran pertama kenapa dan untuk apa aplikasi berbasis digital dibuat... Kalo kata dosen saya, aplikasi di buat untuk mempermudah user... Di sini saya mau daftar tapi harus ada kode QR																
6	Affiah Yolonda febrina Sari	1		1/30/2024 12:45	Kirain cuma aku aja yang gak bisa buka aplikasi nya, ternyata hampir semua pengguna aplikasi... Masa iya jaringan g full gak bisa buka, memuat gagal suruh coba lagi. Ehh busecet, gir																
7	Melati Indri wahayuni	1		1/30/2024 9:42	Aplikasi ga guna padahal wifi data lancar bukak yang Leen tapi bukak nih apk memuat trussas ga ada rimbanya tolong yang jadi developer ny yang benar lah di permudah make apk ny blum																
8	Rayyan Raga	1		1/30/2024 3:13	Mau buka lihat ktp sj ga bisa. Pdhl butuh buat di bank. Krn ktpku ketinggalan. Duh aplikasi apa sih ini! Tolong dong min di perbaiki ini menyusahkan sekali.																
9	tb arief Rahman	1		1/30/2024 2:53	ketika perbaikan data KTP karna pindah, sy tidak d cetakkan KTP fisik barunya hanya d beri aplikasi ini pada saat urgen k rumah sakit, untuk menggunakan asuransi hrs menunjukan kartu as																
10	Yessy Kumala sary	1		1/30/2024 2:51	Sinyal stabil tapi pas mau di buka apk sering "koneksi teputus" apk bukan nyh memudahkan kan tapi malah menyusahkan kan untuk membuat ktp																
11	Dunia Jenal	1		1/30/2024 2:33	Mau di buka engga bisa...bacaannya kesalahan jaringan padahal mah buka YouTube aja lancar....tolong di perbaiki secepatnya																

Fig. 4. After the "username" and "at" Attributes are Removed

The next stage is to import the review data in Excel form into Rapid Miner for text preprocessing. The first stage of text preprocessing is data cleansing. The first step is to import the prepared dataset into Rapid Miner. Next, delete some characters outside of the a-z alphabet such as punctuation marks, numbers, and emoticons using the replace operator. Then add the trim operator which aims to remove excess spaces. Next, add the Filter Examples operator to filter out empty review data (errors) so that only data that is not missing remains. After the cleansing process is complete, save the dataset into an Excel file in CSV format as shown in Figure 5.

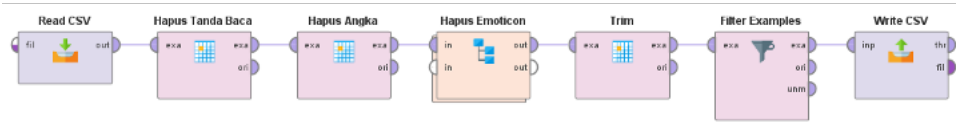


Fig. 5. Cleansing Process

After the data is saved, the next step is to manually label the dataset based on the score level of each review and only label positive and negative. An example of a dataset that has been labeled is in Figure 6.

	A	B	C
1	content		label
2	Sudah login berkali-kali tetap saja yg muncul jaringan tidak stabil padahal jaringan yg kita gunakan stabil, apk ap in saat di butuhkan enggak mau kebuka harus menggunakan jaringan WiFi tolong untuk di sesuaikan lagi		negatif
3	Sangat membantu cuma masih bingung apa KTP digital perlu diprint atau difotocopy soalnya suka butuh kalau ada keperluan terimakasih		positif
4	Ini data nya ga sinkron sama kementerian kesehatan atau gimana sih? di aplikasi peduli lindungi aku dah vaksin ke 3 atau vaksin Booster pertama tapi di aplikasi IKD vaksin nya cuma ada jadi gimana nih ga ada tanggapan kah saya mau		negatif
5	Pelajaran pertama kenapa dan untuk apa aplikasi berbasis digital dibuat... Kalo kata dosen saya aplikasi di buat untuk mempermudah user... Di sini saya mau daftar tapi harus ada kode QR dari dukcapil Otomatis saya harus ke dukcapil d		negatif
6	Kirain cuma aku aja yang gak bisa buka aplikasi nya, ternyata hampir semua pengguna aplikasi... Masa iya jaringan g full gak bisa buka, memuat gagal suruh coba lagi. Ehh busecet, gir		negatif
7	Aplikasi ga guna padahal wifi data lancar bukak yang Leen tapi bukak nih apk memuat trussas ga ada rimbanya tolong yang jadi developer ny yang benar lah di permudah make apk ny blum aja di perbaiki ini menyusahkan sekali.		negatif
8	Mau buka lihat ktp sj ga bisa. Pdhl butuh buat di bank. Krn ktpku ketinggalan. Duh aplikasi apa sih ini! Tolong dong min di perbaiki ini menyusahkan sekali.		negatif
9	ketika perbaikan data KTP karna pindah, sy tidak d cetakkan KTP fisik barunya hanya d beri aplikasi ini pada saat urgen k rumah sakit, untuk menggunakan asuransi hrs menunjukan kartu asuransi dan KTP alhamdulillah aplikasinya saat		negatif
10	Sinyal stabil tapi pas mau di buka apk sering koneksi teputus apk bukan nyh memudahkan kan tapi malah menyusahkan kan untuk membuat ktp		negatif
11	Mau di buka engga bisa...bacaannya kesalahan jaringan padahal mah buka YouTube aja lancar....tolong di perbaiki secepatnya		negatif

Fig. 6. Data Labeling

Datasets that have been labeled are then imported back into Rapid Miner. Next, convert the nominal dataset into text and preprocessing is done again in the Process Documents from Data operator as shown in Figure 7

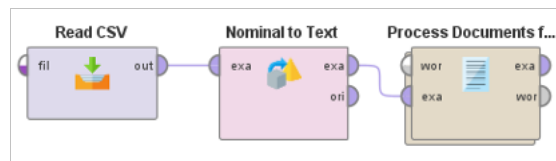


Fig. 7. Importing Data, Converting Nominal Data to Text, and Data Preprocessing

In Data Preprocessing, there are several operators used, namely transform cases, Tokenize, filter stop-words (dictionary), and filter tokens (by length) as shown in Figure 8.

The filter tokens (by length) operator is used to filter tokens that are too short or too long from the text, and will be removed because they are less relevant. The filter tokens (by length) operator is set to 4 for the minimum number of characters (letters) and 25 for the maximum number of characters (letters).

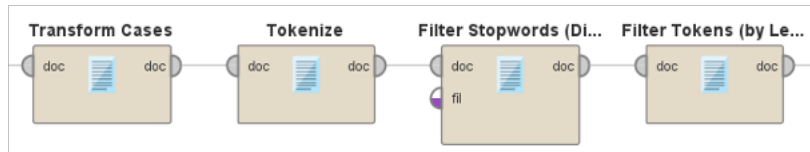


Fig. 8. Data Preprocessing

C. Split Data

After the data is cleaned at the text preprocessing stage, the next stage is to divide the data. Of the 813 data processed, it will be divided into 70% as training data and 30% as test data, so that 569 data will be obtained for training data and 244 data for test data. The operator used to divide the data is the Split Data operator, as shown in Figure 9.

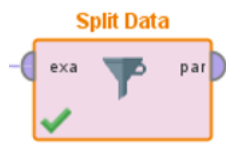


Fig. 9. Split Data Operator

D. Data Analysis

At this stage, the training data is analyzed, and the data pattern is studied using the Decision Tree algorithm. Training data totaling 569 data will be used to train the Decision Tree algorithm. The operator used is the Decision Tree operator as shown in Figure 10.

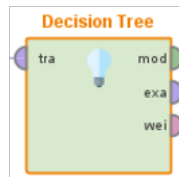


Fig. 10. Decision Tree Operator

After the training data is analyzed using the Decision Tree algorithm, it will produce a classification model from the Decision Tree algorithm. The following model of the classification results of the Decision Tree algorithm is displayed in the form of a Tree, as shown in Figure 11.

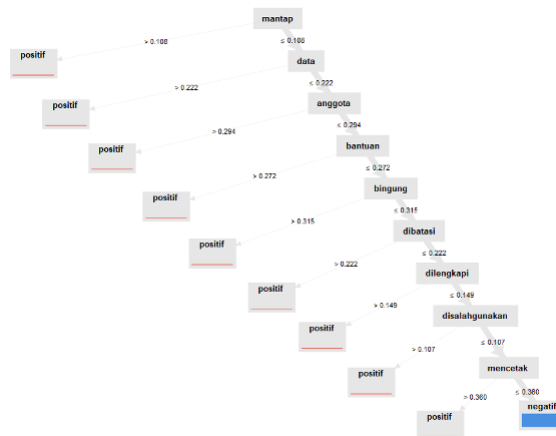


Fig. 11. Classification Result Model of the Decision Tree Algorithm

After getting the Decision Tree algorithm classification model, then the model is tested using the test data that has been prepared. In testing the classification model of the Decision Tree algorithm, it uses the Apply Model operator. The Apply Model operator is shown in Figure 12.

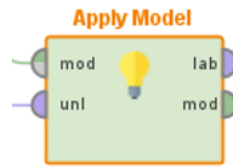


Fig. 12. Apply Model Operator

Based on the test data totaling 244 data consisting of 218 negative data and 26 positive data. The Decision Tree algorithm predicts the test data to be 240 negative data and 4 positive data. Next, add the Performance operator to see the performance results of the Decision Tree algorithm model that has been tested and then connect all operators so that it can be processed (Run) as shown in Figure 13.

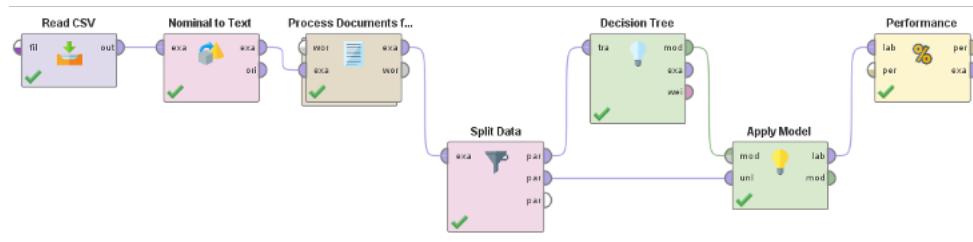


Fig. 13. Adding a Performance Operator

The results of testing the confusion matrix using the Decision Tree algorithm are shown in Figure 14.

accuracy: 89.34%

	true negatif	true positif	class precision
pred. negatif	216	24	90.00%
pred. positif	2	2	50.00%
class recall	99.08%	7.69%	

Fig. 14. Data Testing Result

The accuracy of data measurement results obtained from the analysis process reached 89.34%, the classification performance in the negative class shows much better results than the positive class. The number of negative predictions classified as negative amounted to 216 data, and the number of negative predictions classified as positive amounted to 24 data. The resulting achievement of the precision class on the sentiment analysis dataset is 90.00%. While the number of positive predictions classified as negative amounted to 2 data and the number of positive predictions classified as positive amounted to 2. The resulting achievement of the precision class is 50.00%. The measurement of the recall class resulted in true negatives reaching 99.08% and true positives reaching 7.69%. This comparison shows that the model is more capable of accurately recognizing negative sentiments, but still has difficulty in detecting and classifying positive sentiments.

E. Data Visualization With Wordcloud

Based on the dataset obtained from the Identitas Kependudukan Digital application review data. The word "Aplikasi" has the most words that appear in the dataset, which is 582 words. After that there is the word "Data", which appears in as many as 214 words, and the word "Digital" appears in as many as 199 words, displayed in the form of a wordcloud in Figure 15.

- [9] R. S. Wahono, "Data mining," 2013. Online.
- [10] A. T. Octa Nuryawan, M. Hasbullah, M. Rizal, M. F. Rajab, and N. Agustina, "Decision tree algorithm for public sentiment analysis towards marketplace in indonesia," *Naratif Jurnal Nasional Riset, Aplikasi dan Teknologi Informasi*, vol. 5, no. 1, pp. 18–25, 2023.
- [11] G. K. Locarso, "Sentiment analysis of pedulilindungi application review on google play store using nbc," *J. Tek. Inform. Kaputama*, vol. 6, no. 2, pp. 353–361, 2022.
- [12] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Sentiment analysis of airline opinions on twitter documents using support vector machine (svm) algorithm," *J. Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, 2019. Online.
- [13] A. Mukminin, "Analysis of public sentiment towards covid-19 swab-pcr test services in indonesia using the support vector machine algorithm," 2021.
- [14] R. Adinugroho, "Comparison of split ratio of training data and testing data using the lstm method in predicting asia stock price index," 2022. Online.
- [15] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Sentiment analysis of villa reviews in ubud using naive bayes, decision tree, and k-nn methods," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 205–215, 2022.