



Analisis kinerja algoritma naïve bayes pada dataset sentimen masyarakat aplikasi NEWSAKPOLE samsat jawa tengah

Eko Budi Susanto^{*1}, Paminto Agung Christianto², Mohammad Reza Maulana³, Satriedi Wahyu Binabar⁴

Email: ¹EkoBudi.S@stmik-wp.ac.id, ²p_a_chr@stmik-wp.ac.id, ³reza@stmik-wp.ac.id, ⁴binabars@gmail.com

^{1,2,3,4}Teknik Informatika, STMIK Widya Pratama

Diterima: 30 November 2022 | Direvisi: 9 Desember 2022 | Disetujui: 16 Desember 2022

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Direktorat Lalu Lintas POLDA Jawa Tengah, BPPD Provinsi Jawa Tengah, dan PT. Jasa Raharja Jawa Tengah menciptakan terobosan pembuatan aplikasi Sistem Administrasi Kendaraan Pajak Online (NEWSAKPOLE) untuk membantu masyarakat untuk mempermudah pembayaran pajak kendaraan bermotor. NEWSAKPOLE ini telah banyak diunduh dari marketplace dan digunakan oleh masyarakat Jawa Tengah. Di dalam penggunaannya, terdapat masyarakat yang puas dan tidak puas dalam menggunakan aplikasi ini, terbukti adanya berbagai macam jenis komentar dan rating yang dituliskan pengguna di Google Play Store. Dari hal tersebut, perlu dilakukan sebuah penelitian analisis sentimen, agar instansi terkait dapat mengetahui komentar positif dan negatif dari penggunaan aplikasi NEWSAKPOLE dengan cepat. Algoritma naïve bayes digunakan pada penelitian ini untuk melakukan klasifikasi ulasan yang ada di Google Play Store. Kemudian penggunaan algoritma ini diukur tingkat kinerjanya dalam melakukan proses klasifikasi. Pertama, dataset dikumpulkan terlebih dahulu dari Google Play Store. Setelah itu, sebelum dilakukan *preprocessing* data, proses klasifikasi dilakukan untuk mengetahui ulasan positif, negatif atau netral. Dari hasil proses tersebut selanjutnya dilakukan proses pembobotan dan pengujian penggunaan algoritma naïve bayes, serta hasilnya dilakukan proses analisis. Dari 2414 komentar positif dan 2395 komentar negatif, algoritma naïve bayes memiliki tingkat akurasi sebesar 88% setelah dilakukan pengukuran K-Fold validation sebanyak 10 kali dalam melakukan klasifikasi ulasan. Hal ini dapat disimpulkan bahwa algoritma naïve bayes memiliki performa yang baik dalam melakukan klasifikasi komentar penggunaan aplikasi yang ada di Google Play Store.

Kata kunci: *naïve bayes, analisis sentimen, algoritma klasifikasi, NEWSAKPOLE*

Performance analysis of the naïve bayes algorithm on the sentiment dataset of central java samsat NEWSAKPOLE applications

Abstract

Direktorat Lalu Lintas POLDA Jawa Tengah, BPPD Provinsi Jawa Tengah, and PT. Jasa Raharja Jawa Tengah makes innovations in making it easier for the public to pay motor vehicle tax by creating an Online Vehicle Tax Administration System (NEWSAKPOLE) application. The NEWSAKPOLE application has been downloaded a lot from the marketplace and is used by the people of Central Java. There are people who are satisfied and dissatisfied with using this application, as evidenced by the existence of various types of comments and ratings written by users on the Google Play Store. There are people who are satisfied and dissatisfied with using this application, as evidenced by the existence of various types of comments and ratings written by users on the Google Play Store. It is necessary to conduct a sentiment analysis study, so that the relevant agencies can quickly classify positive and negative comments from the use of the NEWSAKPOLE application. This study uses the naïve Bayes algorithm to classify reviews on the Google Play Store. Then the use of this algorithm measured its performance level in carrying out the classification process. First, the dataset is collected first from the Google Play Store. After that, before preprocessing the data, a classification process was carried out to find out positive, negative or neutral reviews. From the results of this process, the weighting process and testing of the use of the naïve Bayes algorithm are carried out, and the results are analyzed. From 2414 positive reviews and 2395 negative reviews, the naïve Bayes algorithm has an accuracy rate of 88% after 10 times K-Fold

validation measurements in classifying reviews. This shows that the naïve Bayes algorithm has a good performance in classifying review sentiments in the Google Play Store in using the NEWSAKPOLE application.

Keywords: naïve bayes, analisis sentimen, classification algorithm, NEWSAKPOLE

1. PENDAHULUAN

Sistem Administrasi Kendaraan Pajak Online (NEWSAKPOLE) merupakan sistem pembayaran pajak kendaraan secara online, yang diciptakan oleh Direktorat Lalu Lintas POLDA Jawa Tengah, BPPD Provinsi Jawa Tengah, dan PT. Jasa Raharja Jawa Tengah. Tujuan dari aplikasi ini adalah untuk membantu masyarakat dalam membayar pajak kendaraan bermotor [1]. Saat ini aplikasi NEWSAKPOLE sudah mengalami beberapa perbaikan, dan berubah namanya menjadi newNEWSAKPOLE yang didapat di unduh di Google Play Store, aplikasi tersebut sudah diunduh lebih dari satu juta kali [2]

Pada platform Google Play Store terdapat komentar dari pengguna terkait aplikasi yang diunduh. Komentar tersebut merupakan sentimen dari masyarakat terkait pengalaman masyarakat dalam menggunakan aplikasi di Google Play Store. Tujuan dari adanya rating, ulasan atau komentar dimaksudkan untuk memberikan kepercayaan dari suatu aplikasi, memberikan komentar atau ulasan di Google Play Store merupakan salah satu cara untuk memberikan masukan yang bermanfaat serta membantu masyarakat untuk menemukan aplikasi yang memiliki layanan yang bagus.

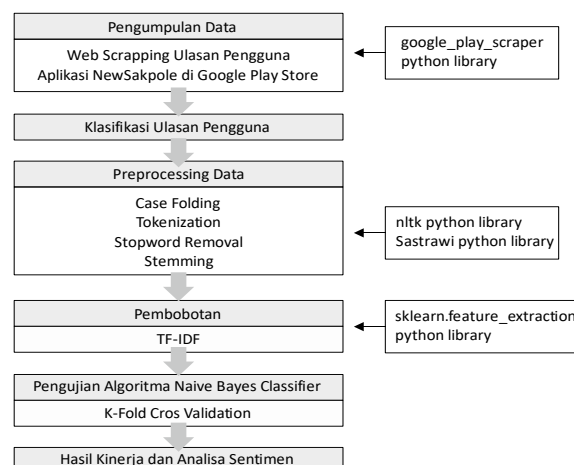
Salah satu metode yang dapat digunakan untuk mengetahui ulasan di Google Play Store yaitu dengan teknik analisis sentimen. Analisis sentimen atau opinion mining adalah suatu disiplin ilmu komputasi yang mempelajari atau menganalisa pandangan, penilaian, sikap, emosi orang tentang suatu obyek (peristiwa, topik, situasi, dsb) dan atributnya [3]. Analisis sentimen merupakan bagian dari bidang ilmu natural language processing (NLP) [4]. Natural Language Processing (NLP) merupakan bidang penelitian dan aplikasi yang mengeksplorasi bagaimana komputer dapat digunakan untuk memahami dan memproses teks atau ucapan bahasa alami untuk melakukan hal-hal yang bermanfaat. NLP mencakup beberapa bidang studi seperti terjemahan mesin, pemrosesan teks bahasa alami, *multilingual and cross language information retrieval (CLIR)*, *speechrecognition*, *artificial intelligence* dan *expert systems*, dan lain sebagainya.

Beberapa penelitian yang membahas analisis sentimen yaitu: analisis sentimen masyarakat terhadap kebijakan vaksinasi covid-19 dengan algoritma logistic regression [3], analisis sentimen terhadap kebijakan merdeka belajar kampus merdeka pada komentar twitter dengan algoritma K-Nearest Neighbor [5], [6], analisis sentimen dan information extraction pembelajaran daring menggunakan pendekatan lexicon[6], klasifikasi data review imdb berdasarkan analisis sentimen menggunakan algoritma support vector machine[7].

Pada penelitian ini akan dilakukan pengujian kinerja algoritma naive bayes classifier mengklasifikasikan ulasan pengguna di google play store pada aplikasi NEWSAKPOLE. Kinerja algoritma naive bayes classifier akan diukur tingkat akurasi. Dataset yang digunakan pada penelitian ini adalah hasil ulasan pengguna aplikasi NEWSAKPOLE di Google Play Store. Hasil analisis sentimen dapat dijadikan bahan pertimbangan dalam pengembangan aplikasi NEWSAKPOLE agar menjadi lebih baik.

2. METODE PENELITIAN

Metode penelitian pada penelitian ini terdiri dari enam tahapan, yaitu: pengumpulan data, klasifikasi ulasan pengguna, preprocessing data, pembobotan, pengujian algoritma naive bayes classifier, dan analisis hasil kinerja. Metode meliputi analisis masalah dan desain yang digunakan untuk memecahkan permasalahan. Analisis menggambarkan masalah yang ada dan akan diselesaikan dalam penelitian. Desain menggambarkan bagaimana menyelesaikan permasalahan dan disajikan dalam bentuk diagram dengan penjelasan lengkap.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Pada tahap pertama dilakukan pengumpulan data. Data yang digunakan adalah hasil usulan pengguna aplikasi NEWSAKPOLE pada Google Play Store. Metode web scrapping digunakan untuk mengambil data usulan pada Google Play Store. Web scrapping merupakan cara untuk mengunduh, mengurai, dan mengatur data dari web secara otomatis[8]. Alat yang digunakan untuk mengambil data yaitu pustaka python google_play_scrapper [9].

2.2 Klasifikasi Ulasan Pengguna

Tahap selanjutnya setelah data didapatkan yaitu: pengklasifikasian ulasan dari pengguna. Hasil evaluasi akan dikategorikan menjadi tiga kategori yaitu: positif, negatif, dan netral. Proses pelabelan berdasarkan rating dari ulasan pengguna. Rating aplikasi pada Google Play Store menunjukkan angka bintang 1, 2, 3, 4, 5. Kategori positif adalah ulasan yang memiliki bintang 4 dan 5. Kategori negatif adalah ulasan yang memiliki bintang 1 dan 2. Sedangkan netral adalah ulasan yang memiliki bintang 3.

2.3 Preprocessing data

Pada tahap ketiga dilakukan proses *preprocessing data*, yaitu tahapan case folding, tokenization, stopword removal, dan stemming. Proses perubahan semua huruf pada komentar menjadi huruf kecil, dilakukan pada tahap case folding. Penggunaan huruf besar dan kecil terkadang memiliki arti yang berbeda. Tahap selanjutnya dilakukan pemecahan kalimat yang ada pada komentar menjadi bagian-bagian kata, proses ini disebut sebagai tokenization. Stop-Word Removal merupakan proses penghapusan kata. Tujuan dari penghapusan kata adalah menghilangkan kata-kata yang tidak penting, proses ini dapat mengurangi dimensi dataset. Kata-kata umum dalam komentar seperti imbuhan depan atau belakang tidak memberikan makna sehingga perlu dihilangkan. Stemming adalah proses memetakan bentuk kata dan memecahnya menjadi bentuk kata dasar atau dapat dikatakan, proses mengubah kata berimbuhan menjadi kata dasar. Proses stemming ini merupakan proses terpenting dalam tahap text mining. Hasil stemming yang baik dapat mempengaruhi baik tidaknya aplikasi text mining [10].

2.4 Pembobotan

Tahap keempat adalah pembobotan terhadap kata yang sudah dilakukan preprocessing. Jumlah kata yang memiliki tingkat kemunculan pada suatu dokumen disebut Term Frequency (TF). Kata yang memiliki tingkat kemunculan paling tinggi diberikan bobot paling besar. Pemberian bobot untuk mengukur intensitas kemunculan dari kata disebut Document Frequency (DF). DF diperoleh dari TF dan IDF dibentuk dari hasil invers DF [5]. Berikut ini adalah persamaan untuk menghitung TF-IDF:

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

$$Wdt = TF \cdot IDF \quad (2)$$

$$TF - IDF(w, d) = \frac{TF - IDF(w, d)}{\sqrt{\sum_{w=1}^n TF - IDF(w, d)^2}} \quad (3)$$

Keterangan:

IDF(w) = pembobotan kata pada dokumen

W = kata

TF(w,d) = jumlah intensitas kemunculan kata w dalam kolom d,

IDF(w) = invers DF dari kata w

N = jumlah seluruh dokumen

DF(w) = jumlah dokumen yang mengandung kata w

2.5. Pengujian *algorithm naive bayes classifier*

Algoritma Naïve Bayes adalah metode probabilitas pengklasifikasian berdasarkan Teorema Bayes. Proses klasifikasi dilakukan dengan melatih dengan data training. Pada algoritma Naïve Bayes nilai sebuah input features pada kelas diasumsikan tidak memiliki ketergantungan nilai features lainnya. Ada beberapa jenis algoritma naïve bayes berdasarkan jenis features datanya, yaitu: Gaussian Naïve Bayes, Categorical Naïve Bayes, Bernoulli Naïve Bayes, dan Multinomial Naïve Bayes. Untuk klasifikasi pada sentimen analisis ini akan digunakan tipe algoritma Multinomial Naïve Bayes [11]. Multinomial Naïve Bayes menentukan apakah data yang dimasukkan milik kelas tertentu atau tidak dengan menghitung probabilitas posterior (prior).

$$P(C_i|d) = P(C_i) \prod_{i=1}^{n_d} P(t_k | C_i) \quad (4)$$

Formula untuk menghitung posterior ditunjukkan pada persamaan (4). Dimana $P(C_i|d)$ adalah probabilitas posterior dari C_i diberikan dokumen d, C_i adalah *class variable* ke i , $i = \{1, \dots, n\}$, $P(C_i)$ adalah probabilitas posterior kelas ke- i , $P(t_k|C_i)$ adalah probabilitas dari t ke- k , diberikan kelas C ke i . $P(t_k|C_i)$ disebut sebagai likelihood. dan n_d adalah jumlah fitur yang terdapat dalam dokumen d [12].

Teknik Validasi dan Evaluasi untuk teknik klasifikasi analisis sentimen digunakan untuk mengukur hasil penelitian. Banyak metode yang digunakan untuk memvalidasi model berdasarkan data yang ada, seperti holdout, random subsampling, cross-

validation, stratified sampling, bootstrapping, dan lain-lain [13]. Pada penelitian ini akan digunakan metode confusion matrix dan cross-validation untuk mengukur kinerja dari algoritma naive bayes classifier. Evaluasi performa algoritma dilakukan untuk mengevaluasi hasil klasifikasi dengan mengukur nilai kebenarannya. Akurasi digunakan untuk mengukur nilai kebenaran. Akurasi merupakan hasil persentase dokumen yang diklasifikasikan oleh sistem. Semua parameter untuk mendapatkan akurasi diperoleh dari Confusion Matrix menurut Christopher D Manning et al [13]. Pada tabel berikut ini:

Table 1. Confusion Matrix

	Relevant	Non Relevant
Retrieved	True Positive (TP)	False Positive (FP)
Non-Retrieved	False Negative (FN)	True Negative (TN)

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Dataset yang digunakan merupakan kumpulan ulasan pengguna aplikasi NEWSAKPOLE dari Google Play Store dengan metode web scrapping. Proses web scrapping menggunakan bahasa pemrograman python dengan memanfaatkan pustaka google_play_scrapper.

```
from google_play_scraper import Sort, app, reviews_all

g_reviews = reviews_all(
    "com.jatengprov.bapenda.newsakpole",
    sleep_milliseconds=0,
    lang='id',
    country='id',
    sort=Sort.NEWEST,
)
```

Gambar 2. Cara memanggil pustaka google_play_scrapper

Pengambilan data dilakukan pada tanggal 16 November 2022, terkumpul sebanyak 5.219 ulasan yang berhasil dilakukan scrapping. Dari data ulasan yang sudah didapatkan akan dilakukan klasifikasi sentimen berdasarkan dari rating ulasan yang ada. Dari proses klasifikasi sentimen, didapatkan sebanyak 2.496 positif, 2.407 ulasan negatif, dan 315 netral. Untuk ulasan netral akan dihilangkan atau dihapus pada saat pengujian kinerja algoritma naïve bayes. Hal ini dikarenakan untuk sentimen netral jumlahnya terlalu sedikit. Sehingga data yang digunakan hanya terdiri dari dua label yaitu: positif dan negatif. Jumlah dataset dan hasil dari klasifikasi dapat dilihat pada tabel 2, sedangkan isi datasetnya dapat dilihat pada tabel 3.

Tabel 2. Jumlah Dataset

Sentimen	Jumlah	Persen
Positif	2.495	50,91%
Negatif	2.407	49,09%
Total	4.902	100,00%

Tabel 3 Dataset Ulasan Pengguna

id	review_description	rating	sentimen
1	suwun pak Ganjar	5	Positif
2	Sangat membantu	5	Positif
3	5\$dfd& dd 1 ddd5¥ yang tidak bisa sddi^¥/\$\$\$\$1\$2\$#\$9\$gg\$esegess	5	Positif
4	Good	5	Positif
5	Sangat membantu	4	Positif
....		1	Negatif
4000	Orang yang punya sudah meninggal, bagaimana foto bareng KTP realtime? Mohon pencerahannya suhuuuu	1	Negatif
4001	Servernya lemot banget, buat masuk aplikasi gagal terus	1	Negatif
...		5	Positif
4903	Developer bocil ngabisin anggaran	1	Negatif

3.2 Preprocessing Data

Setelah dataset berhasil dikumpulkan, langkah selanjutnya adalah melakukan preprocessing data dengan tahapan case folding, tokenization, stopword removal, dan stemming. Setelah itu dilanjutkan dengan melakukan penerjemahkan ke dalam bahasa indonesia jika terdapat kata- kata dalam Bahasa Inggris dan proses data cleaning untuk menghilangkan field yang kosong (missing value) setelah adanya preprocessing data. Pada proses preprocessing data menggunakan beberapa pustaka dari python yaitu: nltk dan sastrawi. Hasil dari tahapan preprocessing data dapat dilihat pada tabel 4.

```
import nltk
import numpy as np
import pandas as pd
from googletrans import Translator
from nltk.corpus import stopwords
from sklearn.pipeline import Pipeline
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

Gambar 3. pustaka python yang digunakan untuk tahapan preprocessing data

Tabel 4. Dataset Setelah dilakukan preprocessing data

id	review_description	sentimen
1	suwun ganjar	positif
2	bantu	positif
3	bagus	positif
4	bantu	positif
...		
4000	servernya lot banget masuk aplikasi gagal	negatif
4001	njlimet sanget sakmeniko revisi	negatif
....		positif
4809	developer bocil ngabisin anggaran	negatif

Jumlah record pada dataset menjadi 4809, karena setelah dilakukan preprocessing terdapat missing value pada beberapa record sehingga dilakukan penghapusan record yang mengalami missing value.

Tabel 5. Jumlah record setelah dilakukan cleaning data

Sentimen	Jumlah	Persen
Positif	2.414	50,20
Negatif	2.395	49,80
Total	4.809	100,00%

3.3 Pembobotan

Proses pembobotan menggunakan pustaka milik pytho yaitu: TfidfVectorizer dan CountVectorizer. Tabel 6 adalah hasil pembobotan TF-IDF.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
```

Gambar 4. Pustaka TF-IDF

Tabel 6. Penggalan data hasil pembobotan TF-IDF

kata	idf_weights	kata	idf_weights
bayar	2,705372	sulit	3,861681
aplikasi	2,709959	buka	4,006182
pajak	2,71688	samsat	4,010392
bantu	2,865068	online	4,145734
ribet	2,959305	verifikasi	4,145734
mudah	3,084862	motor	4,155442

bagus	3,273895	susah	4,180135
sakpole	3,586808	stnk	4,190185
foto	3,59513	mantap	4,258097
kendara	3,74188		

3.4 Pengujian Algoritma Naïve Bayes Classifier

Implementasi dari algoritma naïve bayes classifier dilakukan dengan menggunakan pustaka python sklearn.naive_bayes, dengan cara memanggil pustaka MultinomialNB. Sedangkan untuk mengukur tingkat akurasi menggunakan pustaka sklearn.metrics, dengan memanggil pustaka accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, dan classification_report.

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import KFold, cross_val_score
```

Gambar 5. Pustaka python untuk implementasi naive bayes

Untuk pengukuran akurasi menggunakan metode confusion matrix dan cross-validation. Dataset dibagi menjadi 2 bagian 10% sebagai data testing, 90% sebagai data training. Untuk penerapan cross-validation menggunakan pustaka sklearn.model_selection dan pustaka KFold, cross_val_score.

Tabel 7. Hasil Accuracy, Precision, Recall, dan F1-Score

K-Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	87	87	87	87
2	88	88	89	88
3	87	87	88	87
4	89	90	88	89
5	90	89	90	89
6	92	92	92	92
7	87	87	87	87
8	91	91	91	91
9	88	89	87	88
10	81	82	79	80
Rata2	88	88,2	87,8	87,8

Hasil pengukuran dan validasi dapat dilihat pada tabel 7. Nilai akurasi terendah berapa pada k-10 yaitu 81%, sedangkan nilai akurasi tertinggi berada pada k-6 yaitu 96%. Dari hasil pengukuran K-Fold validation diperoleh rata-rata nilai akurasi dari algoritma naïve bayes classifier yaitu 88%. Hasil confusion matrix di setiap K-Fold validation terdapat pada tabel 8.

Tabel 8. Confusion Matrix Tiap K-Fold

K-Fold	Aktual		Prediksi
	Positif	Negatif	
1	229	20	Positif
	43	189	Negatif
2	208	17	Positif
	39	217	Negatif
3	184	17	Positif
	46	234	Negatif
4	260	14	Positif
	39	168	Negatif
5	179	20	Positif
6	183	13	Positif
	25	260	Negatif
7	180	24	Positif
	37	240	Negatif
8	259	19	Positif
	23	180	Negatif
9	263	17	Positif
	40	161	Negatif
10	265	24	Positif

	30	252	Negatif		66	125	Negatif
--	----	-----	---------	--	----	-----	---------

4. KESIMPULAN

Berdasarkan dari hasil pengukuran kinerja algoritma naïve bayes classifier pada dataset analisis sentimen aplikasi NEWSAKPOLE yang ada di Google Play Store disimpulkan bahwa algoritma naïve bayes memiliki tingkat akurasi yang baik yaitu 88%. Jumlah record pada dataset yang diuji sebanyak 4.809, yang terdiri dari 2.414 komentar sentimen positif, dan 2.395 komentar negatif. Hasil proses tahapan preprocessing pada NLP analisis sentimen dipengaruhi oleh penggunaan bahasa dari pengguna. Pada kasus ini terdapat pengguna yang memberikan ulasan menggunakan Bahasa Jawa, Bahasa Indonesia yang tidak baku, penulisan kata yang disingkat, penulisan yang kata yang dikombinasikan dengan angka semisal huruf A diganti dengan angka 4, dan campuran penggunaan Bahasa Inggris. Hal ini tentunya mempengaruhi proses tokenization, stopword removal, dan stemming yang dapat mempengaruhi kinerja dari algoritma saat dilakukan klasifikasi. Saran untuk penelitian selanjutnya perlu menambahkan tokenization, stopword removal, stemming untuk Bahasa Jawa, selain itu dapat digunakan algoritma klasifikasi lainnya selain naïve bayes untuk proses klasifikasinya. Sedangkan saran untuk Direktorat Lalu Lintas POLDA Jawa Tengah, BPPD Provinsi Jawa Tengah, dan PT. Jasa Raharja Jawa Tengah yaitu : selalu menciptakan inovasi dan meningkatkan pelayanan yang terbaik kepada masyarakat agar menjadi instansi/institusi yang lebih baik lagi dan selalu dicintai oleh masyarakat.

DAFTAR PUSTAKA

- [1] N. A. Saputri, "Pengaruh Penerapan Pelayanan SAKPOLE (Sistem Administrasi Kendaraan Pajak Online) Terhadap Kepuasan Wajib Pajak Dalam Pembayaran Pajak Kendaraan Bermotor," 2019.
- [2] "NEWSAKPOLE - Aplikasi di Google Play." <https://play.google.com/store/apps/details?id=com.jatengprov.bapenda.newsakpole> (accessed Nov. 29, 2022).
- [3] Ash Shiddicky and Surya Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, 2022, doi: 10.37859/coscitech.v3i2.3836.
- [4] P. Chitra *et al.*, "Sentiment analysis of product feedback using natural language processing," *Mater Today Proc*, Feb. 2021, doi: 10.1016/j.matpr.2020.12.1061.
- [5] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, "K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 152–158, 2022, doi: 10.37859/coscitech.v3i2.3841.
- [6] M. Hamka and D. Ratna Sari, "ANALISIS SENTIMEN DAN INFORMATION EXTRACTION PEMBELAJARAN DARING MENGGUNAKAN PENDEKATAN LEXICON," 2022.
- [7] G. Cahyani, W. Widayani, S. D. Anggita, Y. Pristyanto, I. Ikma, and A. Sidauruk, "Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1418, Jul. 2022, doi: 10.30865/mib.v6i3.4023.
- [8] S. Han and C. K. Anderson, "Web Scraping for Hospitality Research: Overview, Opportunities, and Implications," *Cornell Hospitality Quarterly*, vol. 62, no. 1, pp. 89–104, 2021, doi: 10.1177/1938965520973587.
- [9] "google-play-scraper · PyPI." <https://pypi.org/project/google-play-scraper/> (accessed Nov. 29, 2022).
- [10] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- [11] "1.9. Naive Bayes — scikit-learn 1.1.3 documentation." https://scikit-learn.org/stable/modules/naive_bayes.html (accessed Nov. 30, 2022).
- [12] A. M. K. Izzaty, M. S. Mubarak, N. S. Huda, and Adiwijaya, "A multi-label classification on topics of quranic verses in English translation using Tree Augmented Naïve Bayes," *2018 6th International Conference on Information and Communication Technology, ICoICT 2018*, vol. 0, no. 1, pp. 103–106, 2018, doi: 10.1109/ICoICT.2018.8528802.

- [13] I. Parapat Hotel -Medan, "Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter." [Online]. Available: www.twitter.com