Analisis Sentimen Terhadap Pengguna Aplikasi Sapawarga Di Google Play Store

M.Wahidin[1*], Rahmat Gunawan[2], Alvilah Anta Wiguna[3]
[1,2,3] Sistem Informasi, STMIK Rosma, Karawang, Indonesia
Email: m.wahidin@dosen.rosma.ac.id

Abstract
The advancement of information technology has transformed the interaction between local governments and their communities. Sapawarga, an innovation from the West Java Provincial Government, serves as a platform to facilitate public participation in reporting and resolving local issues. This study aims to perform sentiment analysis on the Sapawarga application reviews in the Google Play Store using the Naive Bayes method. Review data was obtained through scraping, processed with preprocessing steps, and analyzed using the Naive Bayes algorithm. The results show that the analysis model has an accuracy of 86%, precision of 76%, recall of 89%, and F1-score of 78%. These findings are expected to provide insights for the development of a more responsive application and improve the quality of public services in West Java.

Keywords: *Naïve Bayes, Sapawarga Application, Sentiment Analysis.*

Abstrak
Kemajuan teknologi informasi telah mengubah cara interaksi antara pemerintah daerah dan masyarakat. Aplikasi Sapawarga, sebuah inovasi dari Pemerintah Provinsi Jawa Barat, berfungsi sebagai platform untuk memfasilitasi partisipasi publik dalam melaporkan dan menyelesaikan permasalahan lokal. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan aplikasi Sapawarga di Google Play Store dengan menggunakan metode Naive Bayes. Data ulasan diperoleh melalui scraping, diolah dengan tahapan preprocessing, dan dianalisis menggunakan algoritma Naive Bayes. Hasilnya menunjukkan bahwa model analisis memiliki akurasi sebesar 86%, presisi 76%, recall 89%, dan F1-score 78%. Temuan ini diharapkan dapat memberikan wawasan untuk pengembangan aplikasi yang lebih responsif dan meningkatkan kualitas pelayanan publik di Jawa Barat.

Kata Kunci: Analisis Sentimen, Aplikasi Sapawarga, Naïve Bayes

Article History :

Received 15, Oktober, 2024
Revised 19, Oktober, 2024
Accepted 31, Oktober, 2024

Corresponding Author:

Nama Penulis, Mokhamad Wahidin
Departemen, Sistem Informasi
Instansi, STMIK Rosma
Alamat. Jl. Parahiyangan, Adiarsa Barat
Email Penulis. m.wahidin@dosen.rosma.ac.id

1. Pendahuluan

Perkembangan teknologi informasi yang pesat telah mengubah paradigma interaksi antara pemerintah daerah dan masyarakatnya. Era *digital* membawa implikasi yang mendalam dalam pemahaman dan pelaksanaan praktik pemerintahan yang inklusif dan responsif. Di tengah evolusi ini, aplikasi mobile menjadi instrumen yang vital dalam memfasilitasi partisipasi publik dalam proses pembangunan dan penyelesaian masalah lokal.

P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

Berikutnya dari tahun ke tahun, menurut [1] pemanfaatan *E-Government* di Indonesia mengalami peningkatan yang sangat signifikan. Dengan wujud harapan tercipta pemerintahan yang lebih responsif dan akuntabel, yang dapat menjawab kebutuhan dan aspirasi warga dengan lebih baik. Dalam sebuah survey *United Nations E-Government Survey* pada tahun 2022 survei tersebut dilakukan untuk menemukan kelebihan dan hambatan dalam usaha untuk meningkatkan implementasi kebijakan dan strategi penerapan *e-government* [2] . Indonesia berada di peringkat 77 dari 193 negara.

Pemerintah Provinsi Jawa Barat, sebagai bagian dari perjalanan transformasi digital yang digerakkan oleh kebutuhan akan pelayanan publik yang lebih efisien dan inklusif, memperkenalkan sebuah inovasi yang menarik, yaitu aplikasi *Sapawarga*. Aplikasi ini bertujuan untuk menjadi jembatan antara masyarakat dan pemerintah dalam melaporkan serta menyelesaikan permasalahan di lingkungan sekitar.

Hadirnya aplikasi *Sapawarga* merupakan layanan publik digital jawa barat yang terintegrasi dibuat dengan memanfaatkan platform digital, yaitu *Google Play*. Setiap pengguna yang mengunduh aplikasi *Sapawarga* dapat memberikan ulasan terhadap aplikasi melalui fitur yang telah disediakan oleh *Google Play* [3] Ulasan aplikasi *Sapawarga* tersedia secara publik bagi siapa pun. Data ulasan dari aplikasi tersebut dapat sangat berharga jika dikelola dengan baik. Pengolahan data ini akan menghasilkan masukan yang penting untuk memperbaiki dan mengembangkan aplikasi, karena kritik dan saran terbaik biasanya berasal langsung dari pengguna..

Dengan mempertimbangkan betapa pentingnya ulasan pengguna untuk kelangsungan sebuah aplikasi, disarankan untuk melakukan analisis sentimen terhadap data ulasan aplikasi Sapawarga di *Google Play Store*. Menurut [4] Analisis sentimen adalah proses untuk mengklasifikasikan pendapat dalam ulasan terhadap suatu topik menjadi label positif, negatif, atau netral. Ini merupakan bagian dari penelitian text mining yang memiliki nilai komersial potensial [5].

2. Tinjauan Pustaka

Analisis Sentimen

Analisis yang juga disebut sebagai *opinion mining*, adalah salah satu aspek menarik dari text mining. Ini memungkinkan untuk menggali ke dalam dunia pendapat, evaluasi, dan emosi yang tercermin dalam teks-telusur. Dari ulasan produk hingga tanggapan terhadap berita terkini, analisis sentimen memungkinkan untuk memahami bagaimana masyarakat merespons dan bereaksi terhadap berbagai entitas dan topik pembicaraan [6].

Machine Learning

Machine learning merupakan teknik yang menggunakan faktor sintaksis dan linguistik dalam mengklasifikasikan sebuah sentimen. Teknik ini terbagi pada tiga pendekatan, yaitu supervised learning, unsupervised learning, dan semi-supervised learning. Pendekatan *supervised learning* lebih umum digunakan karena memberikan hasil yang akurat. Adapun beberapa algoritma yang digunakan yaitu *Naive Bayes Classifier (NB), Decision Tree (DT), Support Vector Machine (SVM), Logistic Reression (LR), Deep Learning-Based*, dan lain-lain.

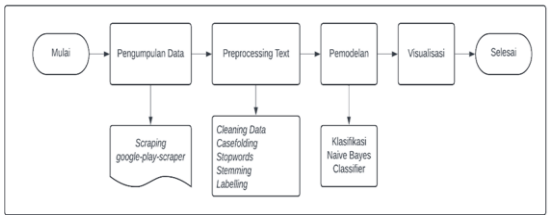
Naïve Bayes Classifier

Naïve bayes adalah salah satu algoritma klasifikasi yang populer dalam text mining

P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

dan analisis sentimen. Meskipun sederhana, algoritma ini sering digunakan karena kecepatan dan keefektifannya dalam mengklasifikasikan data teks yang besar [7]. Naive Bayes Classifier didasarkan pada teorema Bayes dengan asumsi bahwa semua fitur dalam data independen satu sama lain. Meskipun asumsi ini sering kali tidak benar dalam konteks nyata, Naive Bayes Classifier tetap efektif dan efisien dalam banyak kasus. Metode ini cocok untuk pemrosesan teks karena mampu menangani sejumlah besar fitur dengan cepat dan menghasilkan prediksi yang cukup akurat. Dalam analisis sentimen, Naive Bayes Classifier dapat digunakan untuk mengklasifikasikan teks pengguna ke dalam kategori sentimen yang berbeda, seperti positif, negatif, atau netral, berdasarkan pada kata-kata atau fitur-fitur yang terdapat dalam teks tersebut (Wardhani et al., 2023).

3. Metode



Gambar 1. Tahapan Metode

Pada gambar ini, analisis sentimen dilakukan melalui beberapa tahapan yang akan dilakukan pada penelitian ini.

Pengumpulan Data

Pengumpulan data dilakukan dengan cara melakukan scraping dari ulasan aplikasi *Sapawarga* di *Google Play Store*. Teknik dari scraping ini menggunakan *Google Colaboratory* dengan pustaka *google-play-scraper*.

a. Preprocessing Text
   Preprocessing merupakan serangkaian pada sebuah data untuk dilakukan seperti pembersihan teks, transformasi teks, juga penghapusan stopwords.
b. Naïve bayes
   Pemodelan ini menggunakan algortma naïve bayes untuk mengklasifikasikan probabilitas sentimen pada kata-kata dalam data test dan data train.
c. Confusion Matrix
   Confusion Matrix mengukur *accuracy precision*, *recall* dan *F1-score* dari model yang telah dilatih dengan mencari nilai *confusion matrix.*
d. Visualisasi
   Visualisasi ini menampilkan data dalam sebuah wourdcloud dan barchart guna memberikan gambaran awal mengenai pola distribusi dan karakteristik ulasan yang telah dikumpulkan. Proses visualisasi ini menggunakan pustaka *'matplotlib'* dan '*seaborn'* dalam bahasa pemrograman *Python.*

4. Hasil dan Pembahasan

Pengumpulan Data

```
!pip install google-play-scraper
Collecting google-play-scraper
  Downloading google_play_scraper-1.2.7-py3-none-any.whl (28 kB)

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(
    'com.sapawarga.jds',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=100000,
    filter_score_with=None
)
```

Gambar 2. Proses Pengumpulan Data

Gambar ini memuat proses *scraping* dari *google play store* dengan menggunakan *google colaboratory*, dari

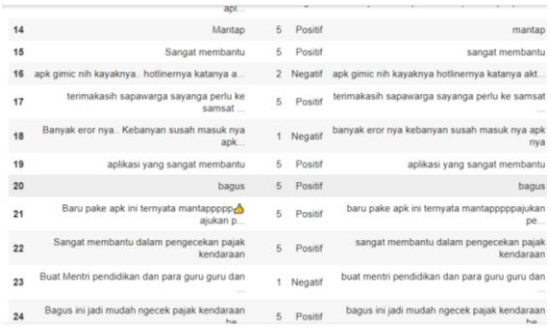
Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi
P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

gambar diatas penulis mencoba mencari data dengan sebanyak 100.000 data. Dan dari hasil scraping data terkumpul sebanyak 3.297 data.

Preprocessing Text

a. Pembersih Data

Sebuah tahap pembersihan data. Data dihindari dari berbagai karakter yang tidak diperlukan, seperti: *url*, *hashtag*, *mention*, dan *emoticon*.



14	Mantap	5	Positif	mantap
15	Sangat membantu	5	Positif	sangat membantu
16	apk gimic nih kayaknya.. hotlinernya katanya a...	2	Negatif	apk gimic nih kayaknya hotlinernya katanya akt...
17	terimakasih sapawarga sayanga perlu ke samsat ...	5	Positif	terimakasih sapawarga sayanga perlu ke samsat ...
18	Banyak eror nya.. Kebanyan susah masuk nya apk...	1	Negatif	banyak eror nya kebanyan susah masuk nya apk nya
19	aplikasi yang sangat membantu	5	Positif	aplikasi yang sangat membantu
20	bagus	5	Positif	bagus
21	Baru pake apk ini ternyata mantappppp👍 ajukan p...	5	Positif	baru pake apk ini ternyata mantappppppajukan pe...
22	Sangat membantu dalam pengecekan pajak kendaraan	5	Positif	sangat membantu dalam pengecekan pajak kendaraan
23	Buat Mentri pendidikan dan para guru guru dan ...	1	Negatif	buat mentri pendidikan dan para guru guru dan ...
24	Bagus ini jadi mudah ngecek pajak kendaraan ...	5	Positif	bagus ini jadi mudah ngecek pajak kendaraan ...

Gambar 3. Proses Cleaning Data

Gambar diatas merupakan hasil data yang telah melakukan proses cleaning data, seperti yang terlihat di gambar atas sebuah emoticon terhapus.

b. Casefolding

Proses *casefolding* perlu dalam pengolahan text karena membuat analisa text lebih konsisten. Hal ini memastikan bahwa kata yang identik, baik dalam bentuk huruf kapital maupun huruf kecil, dianggap identik, sehingga meningkatkan akurasi analisis teks.

c. Stopwords

Stopwords adalah langkah dalam *preprocessing text* yang melibatkan penghapusan kata umum yang biasa muncul dan tidak memberikan kontribusi signifikan terhadap penelitian ini. Hal ini dilakukan untuk mengurangi gangguan, menyederhanakan data, dan meningkatkan kualitas analisa.

d. Stemming

Stemming berguna untuk memproses teks dengan menghapus imbuhan atau akhiran kata, sehingga hanya tersisa kata dasar atau bentuk asli. Proses ini membantu megurangi variasi kata yang memiliki akar yang sama tetapi berbeda bentuk, sehingga mempermudah analisis teks secara keseluruhan.

e. Labelling

Tahap *labelling* atau pelabelan adalah langkah di mana data teks diberikan klasifikasi atau tag berdasarkan sentimen yang terdapat dari hasil pengolahan data. Tujuannya adalah untuk mengklasifikasikan teks atau dokumen ke dalam kategori sentimen positif dan negatif yang kemudian memudahkan analisis lebih lanjut terhadap konten yang relevan. Proses pelabelan dilakukan secara manual, dengan memberikan skor +1 untuk sentimen positif, skor -1 untuk sentimen negatif, dan skor 0 untuk sentimen netral.

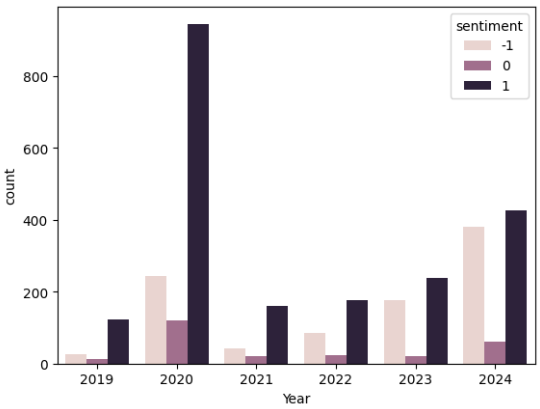
Pemodelan

Pada proses pemodelan ini dilakukan analisa algoritma *naïve bayes* dengan label sentimen positif negatif, dan netral. Pelabelan dilakukan secara manual dengan membandingkan dengan nilai pada ulasan untuk memastikan ketepatan. Hal ini berperan karena banyak data yang memiliki rating tinggi namun memiliki sentimen negatif. Berikut adalah visualisasi gambarannya:

Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi
P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

Gambar 4. Hasil Distribusi Sentimen

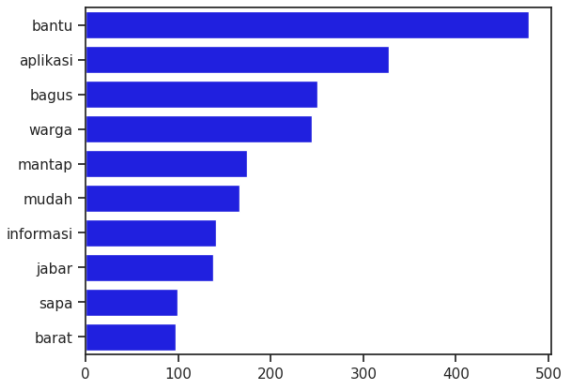
Berikut untuk presentasi ulasan per tahunnya sebagai berikut :



Gambar 5. Distribusi Sentimen Per-tahun.

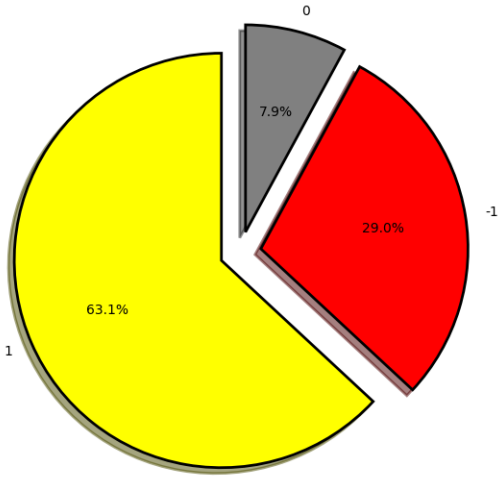
Hasil ulasan pertahun pada aplikasi sapawarga semenjak 2019 sampai 2024. Tahun 2020 menempati ulasan terbanyak dengan total 1310 ulasan dengan sebanyak 946 ulasan positif, 243 ulasan negatif, dan 121 ulasan netral.

Selanjutnya terdapat kata yang sering muncul di setiap ulasan, inilah 10 kata yang sering muncul dalam ulasan.



Gambar 6. Tabel kata yang sering

Berikutnya, evaluasi model menjadi fokus utama untuk mengukur seberapa baik akurat model dapat mengklasifikasikan ulasan. Metode Train-Test Split digunakan untuk membagi dataset menjadi dua bagian:



data pelatihan untuk melatih model dan data pengujian untuk menilai kinerja model. Rasio pembagian yang diterapkan adalah 80:20, di mana data pelatihan lebih banyak dibandingkan dengan data pengujian Implementasi model dan pembagian data dilakukan menggunakan Google Colaboratory, seperti yang dapat dilihat dalam gambar berikut.

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

clf = MultinomialNB()
clf.fit(X_train, y_train)
predicted = clf.predict(X_test)

print("MultinomialNB Accuracy:", accuracy_score(y_test,predicted))
print("MultinomialNB Precision:", precision_score(y_test,predicted, average="binary", pos_label="Negatif"))
print("MultinomialNB Recall:", recall_score(y_test,predicted, average="binary", pos_label="Negatif"))
print("MultinomialNB f1_score:", f1_score(y_test,predicted, average="binary", pos_label="Negatif"))

print(f'confusion_matrix:\n {confusion_matrix(y_test, predicted)}')
print('--------------------------------------------------------\n')
print(classification_report(y_test, predicted, zero_division=0))

# Load dataset
data_clean = pd.read_csv('hasil_TextPreProcessing.csv')
```

Gambar 6. Proses Confusion Matrix

P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

Adapun untuk hasinya sebagai berikut.

Gambar 7. Proses Confusion Matrix

Berdasarkan hasil evaluasi, nilai *precision* untuk kelas negatif adalah 0.76, menunjukkan bahwa 76% dari prediksi negatif adalah benar. Nilai recall untuk kelas negatif adalah 0.80, artinya 80% dari total sentimen negatif yang sebenarnya berhasil teridentifikasi dengan benar. F1-score untuk kelas negatif adalah 0.78, mencerminkan keseimbangan antara precision dan recall.

Untuk kelas positif, nilai precision adalah 0.91, yang menunjukkan bahwa 91% dari prediksi positif adalah benar. Nilai recall untuk kelas positif adalah 0.89, artinya 89% dari total sentimen positif yang sebenarnya berhasil teridentifikasi dengan benar. F1-score untuk kelas positif adalah 0.90, mencerminkan tingkat akurasi dan keandalan yang tinggi dalam mengidentifikasi sentimen positif.

Secara keseluruhan, model ini menunjukkan akurasi sebesar 0.86, yang berarti 86% dari total prediksi model sesuai dengan sentimen yang sebenarnya. Rata-rata precision secara makro adalah 0.84, dan rata-rata recall secara makro adalah 0.85. Rata-rata weighted untuk *precision* dan *recall* masing-masing adalah 0.87 dan 0.86, yang menunjukkan bahwa model ini memberikan performa yang konsisten untuk kedua kelas.

Hasil ini mengindikasikan bahwa pemodelan menggunakan *Naive Bayes Classifier* cukup efektif dalam mengklasifikasikan sentimen pengguna aplikasi *Sapawarga*, dengan performa yang lebih tinggi pada kelas positif dibandingkan dengan kelas negatif. Namun, peningkatan lebih lanjut mungkin diperlukan untuk mengurangi jumlah kesalahan klasifikasi pada kedua kelas tersebut..

```
MultinomialNB Accuracy: 0.8632619439868204
MultinomialNB Precision: 0.764102564102564
MultinomialNB Recall: 0.8010752688172043
MultinomialNB f1_score: 0.7821522309711285
confusion_matrix:
 [[149  37]
 [ 46 375]]
====================================================

              precision    recall  f1-score   support

     Negatif       0.76      0.80      0.78       186
     Positif       0.91      0.89      0.90       421

    accuracy                           0.86       607
   macro avg       0.84      0.85      0.84       607
weighted avg       0.87      0.86      0.86       607
```

5. Penutup

Penelitian ini melakukan analisis sentimen terhadap ulasan pengguna aplikasi Sapawarga di Google Play Store menggunakan algoritma Naive Bayes. Dari hasil analisis, ditemukan bahwa sebagian besar ulasan pengguna bersifat positif, meskipun ada beberapa kritik yang memerlukan perhatian untuk perbaikan lebih lanjut. Model analisis sentimen yang digunakan menunjukkan performa yang baik dengan akurasi tinggi dalam mengklasifikasikan ulasan. Implikasi dari penelitian ini adalah peningkatan kualitas aplikasi Sapawarga yang lebih responsif terhadap kebutuhan masyarakat, mendukung partisipasi publik yang lebih besar dalam pelaporan masalah, dan mendorong transformasi digital dalam pelayanan publik di Jawa Barat.

Penelitian ini memberikan kontribusi penting dalam meningkatkan mutu aplikasi Sapawarga dengan memberikan wawasan yang mendalam tentang persepsi dan pengalaman pengguna. Namun, untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan metode lain seperti Support Vector Machine (SVM), Decision Tree, atau pendekatan berbasis Deep Learning. Metode-metode ini dapat memberikan perspektif yang berbeda

P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

dan mungkin meningkatkan akurasi serta efektivitas analisis sentimen.

Daftar Pustaka

[1] Christian, Y., Wibowo, T., & Lyawati, M. (2024). Sentiment Analysis by Using Naïve Bayes Classification and Support Vector Machine, Study Case Sea Bank. *Sinkron*, *9*(1), 258–275. https://doi.org/10.33395/sinkron.v9i1.13141

[2] Majid, A., Nugraha, D., & Adhinata, F. D. (2023). Sentiment Analysis on Tiktok Application Reviews Using Natural Language Processing Approach. *Journal of Embedded System Security and Intelligent System*, *4*(1), 32–38. http://dx.doi.org/10.26858/jessi.v4i1.41897

[3] Sitorus, R. A., Zufria, I., & Utara, S. (2024). *Application of the Naïve Bayes Algorithm in Sentiment Analysis of Using the Shopee Application on the Play Store 1,2*. 53–66.

[4] Subbaih, K., & Kumar, B. (n.d.). *Aspect category learning and sentimental analysis using weakly supervised learning*. 1–14.

[5] Syaripah, I., Martanto, M., & Suprapti, T. (2024). Analisis Sentimen Pengguna Terhadap Aplikasi Binance Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3714–3724. https://doi.org/10.36040/jati.v7i6.8289

[6] Tri Sanudin, F., Irawan, B., & Bahtiar, A. (2024). Analisis Sentimen Ulasan Aplikasi Sapawarga Di Playstore Menggunkan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 170–175. https://doi.org/10.36040/jati.v8i1.8300

[7] Wardhani, I. P., Chandra, Y. I., & Yusra, F. (2023). Application of the Naïve Bayes Classifier Algorithm to Analyze Sentiment for the Covid-19 Vaccine on Twitter in Jakarta. *International Journal of Innovation in Enterprise System*, *7*(01), 1–18. https://doi.org/10.25124/ijies.v7i01.171

[8] Widodo, J., & Kusnan, K. (2023). Mapping and Analyzing E-Government Sentiments in Local Governments in Indonesia. *Society,*

P-ISSN : 1907-8420
E-ISSN : 2621-1106
DOI : 10.35969/interkom.v19i3.418

11(2), 434–457. https://doi.org/10.33019/society.v11i2.558