

Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decision Tree

Ihsan Zulfahmi

Universitas Negeri Medan

Alamat: Jl. Williém Iskandar Ps. V Medan Estate, Percut Sei Tuan, Deli Serdang

Korespondensi penulis: ihsanzulfahmi@mhs.unimed.ac.id

Abstract. This study utilizes Twitter data to understand user opinions and emotions towards an application. The Decision Tree method was chosen due to its ability to describe the relationship between input variables and the target. The TF-IDF method was used to weight words in the text, and the confusion matrix was used to evaluate the accuracy of the classification model. The research included the research process flow, data preprocessing, and data modeling. Word cloud visualization was used to display the frequency of words in the text. Data was collected from Twitter using Python and the Tweepy library. After preprocessing, the data was categorized into positive, negative, and neutral labels. The evaluation results using the decision tree algorithm showed an accuracy of 96%. The word cloud revealed that the word "aplikasi" (application) has the highest frequency, which shows the importance of the PLN Mobile application but also shows the need for further development. This study provides insights into user sentiment towards the PLN Mobile application and demonstrates the effectiveness of the Decision Tree method in sentiment analysis

Keywords: sentiment analysis, decision tree, pln mobile application

Abstrak. Studi ini memanfaatkan data Twitter untuk memahami opini dan emosi pengguna terhadap aplikasi. Metode Decision Tree dipilih karena kemampuannya dalam menggambarkan hubungan antara variabel input dengan target. Metode TF-IDF digunakan untuk memberikan bobot pada kata-kata dalam teks, dan matriks konfusi digunakan untuk mengevaluasi keakuratan model klasifikasi. Penelitian tersebut meliputi alur proses penelitian, preprocessing data, dan pemodelan data. Visualisasi word cloud digunakan untuk menampilkan frekuensi kata dalam teks. Data dikumpulkan dari Twitter menggunakan Python dan perpustakaan Tweepy. Setelah diolah, data dikategorikan ke dalam label positif, negatif, dan netral. Hasil evaluasi menggunakan algoritma pohon keputusan menunjukkan akurasi sebesar 96%. Kata cloud mengungkapkan bahwa kata "aplikasi" (aplikasi) memiliki frekuensi tertinggi, yang menunjukkan pentingnya aplikasi PLN Mobile namun juga menunjukkan perlunya pengembangan lebih lanjut. Studi ini memberikan wawasan mengenai sentimen pengguna terhadap aplikasi PLN Mobile dan menunjukkan efektivitas metode Decision Tree dalam analisis sentimen.

Kata kunci: analisi sentimen, decision tree, aplikasi pln mobile

LATAR BELAKANG

Seiring dengan berkembangnya teknologi banyak perusahaan yang mulai merambah ke aplikasi mobile guna mempermudah konsumen untuk mendapatkan pelayanan. PLN (Perusahaan Listrik Negara) sebagai perpanjangan tangan pemerintah dalam menyediakan tenaga listrik secara andal, aman, dan terjangkau kepada masyarakat Indonesia. Dengan adanya aplikasi pln mobile, masyarakat menjadi lebih mudah dalam melakukan transaksi pembayaran ataupun mengetahui perkembangan terbaru dari layanan PT.PLN tanpa harus datang langsung ke kantor PLN. Tetapi untuk membantu kinerja dan perkembangan dari aplikasi tersebut maka diperlukan masukan masukan dari kalangan masyarakat. Biasanya masyarakat mengemukakan pendapatnya melalui sosial media. Salah, satu, aplikasi yang sering digunakan masyarakat

Indonesia, adalah, Twitter. Twitter, adalah sebuah platform untuk menyampaikan opini atau pendapat seseorang. Pertumbuhan pengguna Twitter di Indonesia sangat pesat dan menduduki peringkat 5 di dunia. (Utami, 2020). Twitter banyak digunakan orang untuk menyampaikan keluhan kesahnya mulai dari keluhan kesah mengenai kehidupan sehari-hari ataupun keluhan kesah terhadap layanan yang diberikan baik dari pemerintah atau bidang lainnya. Oleh karena itu sangat efisien jika menggunakan Twitter sebagai media untuk mengambil data mengenai keluhan masyarakat Indonesia terkait layanan menggunakan aplikasi PLN mobile.

Analisis sentimen merupakan teknik penting untuk memahami sikap, opini, dan emosi yang diungkapkan dalam teks. Ini memiliki banyak kegunaan, mulai dari menganalisis ulasan produk hingga memantau reputasi merek. Penggunaan metode ini bertujuan untuk memberikan wawasan kepada pengguna tentang sentimen postingan diskusi dan komentar. Salah satu metode pembelajaran mesin yang efektif untuk analisis sentimen adalah pohon keputusan (Decision Tree). Pohon keputusan bekerja dengan membagi kumpulan data menjadi subkumpulan yang semakin kecil berdasarkan kondisi jika-maka yang terkait dengan fitur-fiturnya. Struktur hierarki ini memungkinkan algoritme memperoleh batasan keputusan kompleks dari data. Penulis memilih untuk menerapkan model pohon keputusan karena pohon keputusan memiliki beberapa keunggulan untuk aplikasi ini. Pertama, pohon keputusan bekerja dengan baik dengan data yang jarang dan hilang, yang umum terjadi pada postingan dan komentar berdurasi pendek. Mereka juga memberikan penjelasan logikanya dalam struktur pohon, sehingga memungkinkan interpretasi yang lebih baik dibandingkan model kotak hitam seperti jaringan saraf. Selain itu, pohon keputusan relatif mudah diterapkan dan efisien secara komputasi untuk dijalankan di perangkat seluler dibandingkan dengan algoritme pembelajaran mesin lainnya.

Untuk melatih pengklasifikasi pohon keputusan, kami mengumpulkan kumpulan data dari ribuan postingan media sosial yang diberi tag sentimen positif, negatif, dan netral. Kami mengekstrak fitur seperti jumlah kata, tanda baca, penggunaan huruf besar, dan frasa tertentu yang biasanya mengekspresikan emosi. Algoritme secara rekursif membagi data pada fitur-fitur ini untuk membuat model yang dapat mengkategorikan contoh-contoh baru secara akurat. Dengan menerapkan analisis pohon keputusan yang cepat dan dapat ditafsirkan, ini memungkinkan pengguna mendapatkan wawasan yang baik. Pola sentimen dapat mengungkapkan bagaimana berbagai topik diterima, kapan perdebatan memanasi, atau reaksi umum masyarakat terhadap konten yang dibagikan. Menganalisis sikap dan emosi ini akan memungkinkan pemahaman yang lebih dalam dan diskusi yang lebih bermakna

KAJIAN TEORITIS

1. Natural language programming (NLP)

Pemrograman bahasa alami (NLP) adalah bidang kecerdasan buatan dan linguistik komputasi yang berfokus pada interaksi antara komputer dan bahasa manusia (bahasa alami). Singkatnya, pemrograman bahasa alami dapat didefinisikan sebagai berikut: “Memberikan mesin (komputer) kemampuan untuk memahami, menafsirkan dan memanipulasi bahasa manusia untuk melakukan berbagai tugas seperti menjawab pertanyaan, meringkas teks, menerjemahkan bahasa, dll . Tujuan utama NLP adalah menciptakan program/sistem yang memungkinkan komputer berkomunikasi dengan manusia menggunakan bahasa alami, baik tertulis maupun lisan. Berikut beberapa contoh aplikasi NLP:

- Chatbots/chatbots
- Penerjemah otomatis
- Pengenalan ucapan dan text-to-speech
- Rekomendasi produk e-commerce
- Klasifikasi teks dan analisis sentimen
- Peringkasan teks
- Pemeriksaan ejaan

Dengan NLP, interaksi antara manusia dan mesin menjadi lebih mudah dan intuitif, tanpa gejala seperti interaksi manusia. NLP saat ini merupakan teknologi yang sangat penting dalam banyak aplikasi kecerdasan buatan.

2. Analisis Sentimen

Analisis sentimen adalah teknik di bidang pembelajaran mesin dan pemrosesan bahasa alami, yang digunakan untuk mengidentifikasi dan mengekstrak opini, sentimen, ulasan, sikap, dan emosi dari teks. Singkatnya, analisis sentimen melakukan:

- Menentukan apakah sebuah teks mengandung emosi positif, negatif, atau netral.
- Mengukur positif/negatifnya suatu sudut pandang dalam sebuah teks.
- Identifikasi maksud/tujuan di balik teks tersebut.
- Kelompokkan komentar ke dalam beberapa kategori seperti bingung, marah, sedih, senang.

- Memahami opini publik tentang topik, merek, produk, atau kebijakan tertentu.

Analisis sentimen banyak digunakan untuk menganalisis data teks seperti tweet, komentar, review, forum, artikel, dan lain-lain. Wawasan sentimen berguna bagi bisnis, organisasi, dan pembuat kebijakan untuk memahami suara pelanggan, reputasi merek, dan tren opini publik. Jadi, analisis sentimen pada dasarnya adalah teknik untuk memahami sikap seseorang di balik teks yang ditulisnya dengan menggunakan teknik kecerdasan buatan.

3. Decission Tree

Pohon keputusan merupakan model prediksi yang menggunakan struktur pohon (hierarki) dimana setiap cabang mewakili hasil pengujian suatu fitur/atribut. Pohon keputusan sering digunakan dalam pembelajaran yang diawasi. Cara kerja pohon keputusan adalah sebagai berikut:

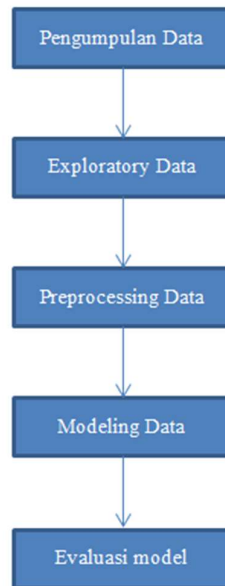
- Pohon dimulai dengan simpul akar
- Node akar kemudian membagi data menjadi pengujian berorientasi fitur.
- Hasil setiap tes akan menentukan cabang mana yang akan dipilih selanjutnya.
- Proses ini terus berulang hingga simpul daun tercapai.
- Node daun mewakili hasil klasifikasi akhir.

Contoh sederhana: Misalkan kita mempunyai data tentang buah-buahan dan ingin mengklasifikasikan enak atau tidaknya buah-buahan berdasarkan warna dan rasanya. Pohon keputusan akan memeriksa misalnya: Apakah warnanya merah? Jika ya -> apakah rasanya manis? Jika ya ---> Delicious. Keuntungan pohon keputusan: Mudah diinterpretasikan dan divisualisasikan. Dapat menangani fitur kategorikal dan numerik.: Tidak diperlukan penskalaan atau normalisasi data.

METODE PENELITIAN

a. alur proses penelitian

untuk alur proses penelitian dapat dilihat pada gambar dibawah ini



Gambar 1. Alur proses penelitian

b. Data introduction

Dataset yang digunakan berasal dari hasil crawling data yang dilakukan oleh penulis dengan menambang data dari aplikasi twiter dan berisi tentang data curhatan, ungkapan saran dan kritik dari para pengguna yang di tuangkan kedalam medio sosisa berupa aplikasi twitter tentang pengalaman mereka menggunakan aplikasi pln mobile.

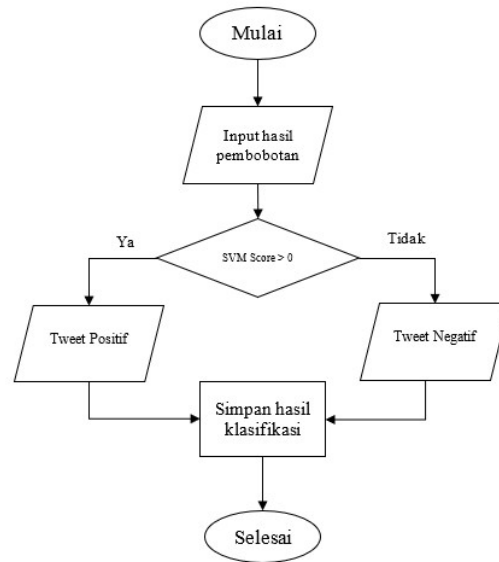
	Tanggal	Tweet Id	Teks	Username
0	2022-06-30 21:29:11+00:00	1542621292285030400	@novitamuhayani 3. Pilih ID Pelanggan yang sud...	pln_123
1	2022-06-30 21:29:08+00:00	1542621281795063808	@novitamuhayani Hi Kak, untuk informasi tsb Kk...	pln_123
2	2022-06-30 21:28:57+00:00	1542621235481563136	@novitamuhayani Hi Kak, untuk informasi tsb Kk...	pln_123
3	2022-06-30 21:23:24+00:00	1542619836844748800	@dwifortuna34 2. Kakak juga dapat lapor melalu...	pln_123
4	2022-06-30 20:24:12+00:00	1542604938282340352	@pln_123 Di aplikasi sudah tercover kak tapi t...	YBudyyy

Gambar 2. Hasil Crawling Data Twitter

Dan didapat hasil crawling data sebanyak kurang lebih sepuluh ribu data lebih.

c. Data Preprocessing

Tujuan dari prapemrosesan data adalah untuk membersihkan, memformat, dan mengatur data agar lebih siap digunakan. Setelah semua data melalui preprocessing, langkah selanjutnya adalah menerapkan algoritma prediksi. Pada penelitian ini algoritma yang akan diterapkan adalah algoritma Decision tree, yaitu algoritma yang cocok digunakan pada data prediksi.



Gambar 3. alur proses decision tree

d. Modeling Data

Penulis menggunakan Decision tree sebagai salah satu metode untuk menyelesaikan persoalan, dimana Decision tree adalah model machine learning yang memprediksi nilai target dengan membuat serangkaian keputusan. Alasan dipilihnya metode Decision tree karena kemampuannya dalam menggambarkan dan memahami hubungan antara variabel input dengan variabel target dan Decision tree biasanya meniru kemampuan berpikir manusia untuk memproses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Sehingga mudah dimengerti. berdasarkan fitur yang diberikan.

e. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah sebuah metode untuk memberikan bobot pada kata-kata dalam teks dengan cara menghitung seberapa sering kata tersebut muncul dalam suatu dokumen dan seberapa umum kemunculannya dalam seluruh koleksi dokumen teks [17]. Term Frequency mengukur jumlah kemunculan suatu kata dalam dokumen tertentu, dan semakin sering kata tersebut muncul dalam dokumen tersebut, semakin tinggi nilai Term Frequency-nya [18]. Inverse Document Frequency menghitung berapa banyak dokumen yang mengandung kata tersebut relatif terhadap seluruh dokumen dalam dataset, memberikan nilai yang lebih tinggi jika kata tersebut relatif jarang muncul dalam seluruh dataset [19].

f. Confusion Matrix

Confusion Matrix merupakan tabel yang merangkum performa model klasifikasi dengan membandingkan label prediksi untuk sekumpulan data dengan label sebenarnya untuk mengevaluasi keakuratan dan efektivitas model klasifikasi. Baris matriks konfusi mewakili label sebenarnya, sedangkan kolom mewakili label prediksi. Setiap sel dalam matriks berisi jumlah titik data yang diberi label prediksi tertentu, berdasarkan label sebenarnya.

g. Wordcloud

Wordcloud merupakan visualisasi frekuensi kata dalam suatu teks dalam bentuk daftar berbobot. Teknik ini biasanya digunakan untuk memvisualisasikan konten tematik. Dengan demikian, kata cloud dapat menjadi cara alternatif untuk menyusun dan memvisualisasikan informasi tekstual, yang sesuai untuk digunakan dalam pekerjaan Pendidikan [21]. Sebanyak 10248 baris data diambil dari berbagai sumber. Namun, setelah dilakukan pemeriksaan, ditemukan 142 baris data yang merupakan duplikat. Oleh karena itu, baris data tersebut dihapus. Sisanya, sebanyak 4858 baris data, diolah dan dibagi menjadi tiga label, yaitu positif, negatif, dan netral.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan python dimana dengan menggunakan library “tweepy” yang tersedia pada python. Tweepy adalah library Python yang memudahkan akses ke Twitter API untuk keperluan pengambilan dan pengolahan data Twitter. Kemudian dilanjutkan dengan memasukkan token yang kita dapat dari aplikasi twitter yang dilanjutkan dengan mengambil query aplikasi pln mobile dikarenakan peneliti ingin mengetahui kesan, pesan serta kritik dan saran dari pengguna aplikasi pln mobile. Setelah codingan dijalankan maka peneliti berhasil mendapatkan 10000 lebih data komentar yang berhasil di crawling dari aplikasi twitter

```
# Scraping data menggunakan API Twitter
import tweepy as tw
import pandas as pd

# Menggunakan token dari twitter
consumer_key = '2kAfWpPySnPcoToRoIqSaIM'
consumer_secret = 'hKA4wAIWqS59s30QfMrj0EKJzHVf2bjBdif1HZM5DE6iUkaSX7'
access_token = '1539625531591004163-IuSh4LKdgd9wXmyaPvuosVBKU5iKv5'
access_token_secret = '96bkZCGovhBLmLVlu48N1sJ5xSp7rSQtvPwRggcopxLf'

auth = tw.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tw.API(auth, wait_on_rate_limit=True)

# menentukan query
search_words = "aplikasi pln mobile"
new_search = search_words + " -filter:retweets"

tweets = tw.Cursor(api.search,
                    q = new_search,
                    count = 10000,
                    include_entities = True).items(10000)

users_locs = [[tweet.created_at,
                tweet.text] for tweet in tweets]

tweet_text = pd.DataFrame(data = users_locs,
                           columns = ["Tanggal",
                                       "Teks"])

tweet_text.to_csv('plnmobile.csv', index=False)
```

Gambar 4. proses penambangan data dari aplikasi twitter

Hasil dari penambangan data (data mining) sebanyak 10206 komentar berhasil di tambang

```
[ ] # Cek kolom dan baris
df.shape

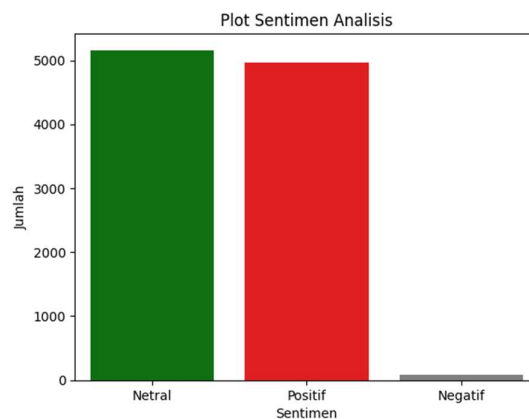
(10206, 4)

[ ] # Cek Tanggal
df.Tanggal.min(), df.Tanggal.max()

('2022-06-08 18:39:17+00:00', '2022-11-30 22:06:14+00:00')
```

Gambar 5. total data dan rentang waktu data yang berhasil di mining

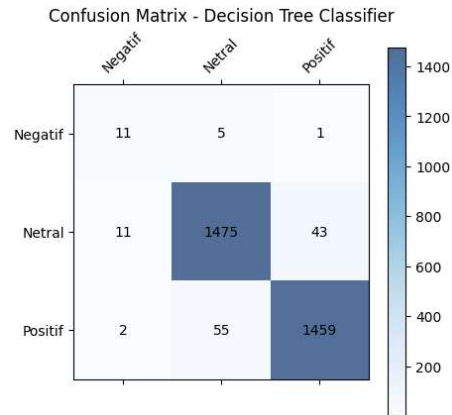
Selanjutnya didapatkan hasil dari presantase label yang dibagi kedalam 3 label yaitu postif netral dan negatif, hasilnya dapat dilihat pada data dibawah ini



Gambar 6. hasil pembagian data ke positif, netral, negatif.

Proses klasifikasi data training dan data testing menghasilkan Confusion matrix 3x3. Confusion matrix ini digunakan untuk mengevaluasi performa algoritma yang digunakan. Berdasarkan hasil perhitungan, nilai akurasi algoritma sebesar 96%. Nilai presisi algoritma

rata-rata sebesar 91%. Nilai recall algoritma rata-rata sebesar 96%. Nilai f1-score algoritma rata-rata sebesar 93%.



Gambar 7. hasil confusion matrix

Ini memberikan gambaran tentang bagaimana model memprediksi setiap kelas. Terlihat bahwa:

- Label 'Negatif' diprediksi sebagai 'Negatif' sebanyak 11 kali, sebagai 'Netral' sebanyak 5 kali, dan sebagai 'Positif' sebanyak 1 kali.
- Label 'Netral' diprediksi sebagai 'Negatif' sebanyak 11 kali, sebagai 'Netral' sebanyak 1475 kali, dan sebagai 'Positif' sebanyak 43 kali.
- Label 'Positif' diprediksi sebagai 'Negatif' sebanyak 2 kali, sebagai 'Netral' sebanyak 55 kali, dan sebagai 'Positif' sebanyak 1459 kali.

Tahap terakhir pada penelitian ini adalah menampilkan wordcloud, yaitu visualisasi yang menampilkan kata-kata dari teks yang diberikan, dengan ukuran font yang lebih besar untuk kata-kata yang lebih sering muncul dalam teks tersebut. Wordcloud dapat membantu memahami inti dari sentimen dengan menunjukkan kata-kata yang paling signifikan dalam membangun sentimen.



Gambar 8. Hasil dari word cloud positif , netral, negatif

KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan oleh penulis kita melihat bahwasanya Studi ini menggunakan ulasan twiter, memproses total 10206 baris data dan mengklasifikasikannya menjadi tiga label: positif, negatif, dan netral. Hasil pengujian dan evaluasi menggunakan algoritma decision tree menunjukkan akurasi sebesar 96%, rata-rata presisi sebesar 91%, rata-rata recall sebesar 96%, dan rata-rata skor f1 sebesar 93%. Berdasarkan hasil frekuensi kata Digital Population Identity Sentiment Analysis di Wordcloud, "Aplikasi" merupakan kata dengan frekuensi kata tertinggi, dengan total frekuensi 5558 dalam konteks positif, negatif, dan netral. Dalam konteks positif, kata "aplikasi" muncul dalam kalimat seperti "Aplikasinya bagus", "Aplikasinya mudah digunakan", dan "Aplikasinya sangat berguna".

Dalam konteks negatif, kata "aplikasi" muncul dalam kalimat seperti "Saya tidak dapat membuka aplikasi", "Aplikasi terus memberikan saya kesalahan", dan "Aplikasi terlalu rumit". Dalam konteks netral, kata "aplikasi" muncul dalam kalimat seperti "Aplikasi ini adalah aplikasi identifikasi penduduk secara digital", "Aplikasi ini merupakan inovasi pemerintah", dan "Aplikasi ini dapat digunakan untuk berbagai keperluan". Tingginya frekuensi kata "aplikasi" menunjukkan bahwa aplikasi pln mobile diakui sebagai aplikasi penting dan dapat digunakan oleh penduduk untuk berbagai tujuan. Istilah ini juga menandakan bahwa aplikasi pln mobile masih memerlukan pengembangan lebih lanjut untuk meningkatkan kualitasnya, terutama dalam hal keamanan dan kemudahan penggunaan.

DAFTAR REFERENSI

- [1] Iqbal and H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/s42979-021-00592-x.
- [2] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [3] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, "Sentiment analysis of financial news using unsupervised approach," *Procedia Comput Sci*, vol. 167, pp. 589–598, 2020, doi: 10.1016/j.procs.2020.03.325.
- [4] A. Shathik and K. Prasad, "A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques," *International Journal of Applied Engineering and Management Letters (IJAEML) A Refereed International Journal of Srinivas University*, vol. 4, no. 2, pp. 2581–7000, 2020, doi: 10.5281/zenodo.3977576.
- [5] J. Homepage, N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, "The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of

Shopee Reviews on Google Play Store,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, pp. 47–54, 2022.

- [6] J. Administrasi Publik, V. Salsa Bella, D. Widodo, and J. Semolowaru No, “Implementasi Aplikasi Identitas Kependudukan Digital (IKD) Dalam Menunjang Pelayanan Publik Masyarakat Di Kecamatan Tambaksari,” *Jurnal SARAQ OPAT*, vol. 6, pp. 14–31, 2024, doi: 10.55542/saraqopat.v6i1.833.
- [7] S. S. Sundari, C. Rahmat Hidayat, and I. Difari, “Implementasi Naïve Bayes Untuk Analisis Sentimen Twiter Terhadap Pergerakan Harga Saham,” *InfoSys Journal*, vol. 8, pp. 98–110, 2022.
- [8] T. Z. Maulani, Z. K. Simbolon, and Amirullah, “Implementasi Algoritma Naïve Bayes Classifier Dalam Menentukan Topik Tugas Akhir Mahasiswa Berbasis Web,” *Jurnal Infomedia : Teknik Informatika, Multimedia, dan Jaringan*, vol. 4, pp. 33–41, 2019.
- [9] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, “Sentiment analysis for customer review: Case study of Traveloka,” *Procedia Comput Sci*, vol. 216, pp. 682–690, 2022, doi: 10.1016/j.procs.2022.12.184.
- [10] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, “Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web,” *SOSCIED*, vol. 6, no. 1, p. p-ISSN, 2023.