



ANALISIS SENTIMEN NETIZEN TERHADAP EFISIENSI APBN MENGUNAKAN *ORANGE DATA MINING*

Anto Febria¹, Joko Triloka², Tahta Herdian Andika³

^{1,2}Fakultas Ilmu Komputer, Institut Informatika Dan Bisnis Darmajaya

³Fakultas Teknologi & Informatika, Universitas Aisyah Pringsewu

antofe.2021210010@mail.darmajaya.ac.id, joko.triloka@darmajaya.ac.id,

tahta.herdian.a@aisyahuniversity.ac.id

ABSTRACT

The government's policy on APBN efficiency in 2025 has sparked debate on social media, including the X platform (formerly Twitter). Social media users and activists express their opinions based on their respective sentiments—whether in support, opposition, or neutrality toward the policy. This study aims to analyze public sentiment on the issue using the Orange Data Mining application with the Sentiment Analysis - Multilingual Sentiment method. The Naïve Bayes algorithm is also applied to assess accuracy and prediction errors. The sentiment analysis results indicate that the majority support the APBN efficiency policy, with confidence and accuracy nearing 100%.

Keywords: *Orange Data Mining; Text Mining; APBN Efficiency; Naïve Bayes; Sentiment Analysis.*

ABSTRAK

Kebijakan pemerintah berupa efisiensi APBN TA 2025 menimbulkan perdebatan di media sosial termasuk aplikasi X (Twitter). Pengguna dan penggiat media sosial menyampaikan komentar dan opini sesuai sentimen masing-masing apakah setuju, tidak setuju, atau netral terhadap kebijakan pemerintah. Penelitian ini akan mencoba menganalisis sentimen atas opini dan pendapat netizen menggunakan aplikasi *Orange Data Mining* dengan metode *Sentiment Analysis - Multilingual Sentiment*. Pengujian juga dilakukan dengan algoritma Naïve Bayes untuk mengukur tingkat akurasi dan kesalahan prediksi. Hasil analisis sentimen pada penelitian ini adalah sebagian besar mendukung kebijakan efisiensi APBN dengan tingkat keyakinan dan akurasi mendekati 100%.

Kata Kunci: *Orange Data Mining, Text Mining, Efisiensi APBN, Naïve Bayes, Analisis Sentimen*

I. PENDAHULUAN

Anggaran Pendapatan dan Belanja Negara (APBN) adalah rencana keuangan tahunan pemerintah Indonesia yang disetujui oleh Dewan Perwakilan Rakyat [1]. APBN Tahun Anggaran (TA) 2025 telah ditetapkan dengan

Undang-Undang Nomor 62 Tahun 2024 yang disahkan pada tanggal 17 Oktober 2024. Selang waktu yang tidak terlalu lama, yaitu pada tanggal 20 Oktober 2024, Presiden dan Wakil Presiden terpilih hasil pemilu 2024 telah dilantik sebagai penanda dimulainya

pemerintahan baru yang akan menjalankan APBN TA 2025.

Awal tahun 2025, pemerintah mengeluarkan kebijakan efisiensi belanja berupa pemotongan/pemblokiran anggaran belanja kementerian/lembaga (K/L) dengan mengeluarkan Inpres Nomor 1 Tahun 2025 tentang Efisiensi Belanja dalam pelaksanaan APBN dan APBD TA 2025. Sehubungan dengan kebijakan tersebut, media sosial menjadi salah satu tempat bagi netizen untuk menyampaikan pendapat dan tanggapan bahkan keluh kesah atas kebijakan pemerintah, termasuk efisiensi APBN TA 2025. Media sosial yang banyak digunakan oleh masyarakat Indonesia salah satunya adalah aplikasi X (Twitter).

Analisis sentimen pada media sosial seperti aplikasi X (Twitter) telah menjadi metode yang semakin populer untuk memahami opini publik terhadap berbagai isu. Analisis sentimen adalah proses komputasi untuk mengidentifikasi dan mengklasifikasikan opini yang diekspresikan dalam teks, terutama untuk menentukan apakah sikap pengguna terhadap topik, produk, atau layanan tertentu bersifat positif, negatif, atau netral.

Analisis sentimen dapat diaplikasikan untuk mengekstraksi persepsi publik tentang isu-isu penting [2]. Metode ini memanfaatkan teknik kecerdasan komputasional untuk mengukur sentimen atau opini yang terkandung dalam teks [3]. Analisis sentimen melibatkan klasifikasi teks ke dalam kategori-kategori seperti positif, negatif, atau netral, yang memungkinkan peneliti dan pembuat kebijakan untuk memahami tren opini publik secara lebih mendalam [3].

Penelitian berupa analisis sentimen netizen ini memiliki tujuan untuk mengetahui bagaimana persepsi masyarakat terhadap kebijakan pemerintah khususnya efisiensi APBN diawal tahun 2025. Analisa sentimen sudah banyak dilakukan pada berbagai isu penting, seperti Covid-19 [2], [4] dan Pemilihan Umum [5]. Ide penelitian ini merupakan hasil pemikiran murni peneliti berdasarkan pada isu

dan fenomena yang terjadi di masyarakat dan belum pernah dilakukan oleh peneliti lain.

II. TINJAUAN PUSTAKA

Pengumpulan data teks dari aplikasi X (Twitter) pada dasarnya adalah proses *data mining* yaitu mengekstraksi informasi berharga yang tersembunyi di balik tumpukan data yang besar [9]. *Data mining* adalah proses analisis data untuk menemukan pola dan tren yang bermanfaat dalam pengambilan keputusan [10].

Dalam konteks analisis sentimen, *data mining* digunakan untuk mengidentifikasi opini dan sentimen yang terkandung dalam teks. analisis sentimen mengukur kecenderungan dan opini netizen melalui teks [9].

Analisis sentimen adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap suatu entitas, peristiwa, topik, atau isu [9]. Dalam hal ini, peneliti menggunakan metode *Multilingual sentiment analysis* [2] yaitu analisis dengan pendekatan *lexicon-based* yang bergantung pada kamus atau leksikon yang berisi daftar kata-kata dan frasa yang telah diberi skor sentimen.

Alasan penggunaan aplikasi *Orange Data Mining* dalam penelitian ini adalah karena aplikasi ini berbasis *GUI*, mudah digunakan dan yang utama aplikasi ini menyediakan berbagai algoritma dan teknik untuk analisis data, berupa klasifikasi, regresi, pengelompokan, dan visualisasi data [11] termasuk didalamnya analisis sentimen.

III. METODOLOGI

Secara umum penelitian ini menggunakan metode CRISP-DM atau *Cross Industry Standard Process for Data Mining* yaitu metodologi standar dalam data mining berupa kerangka kerja yang terstruktur dan iteratif yang memandu proses data mining dari awal hingga akhir [2]. CRISP-DM merupakan metodologi standar dalam data mining yang terdiri dari enam tahapan utama: pemahaman bisnis,

pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran.

Tahapan penting dalam CRISP-DM adalah pemahaman bisnis atau permasalahan yang ada. Dalam konteks penelitian ini, pemahaman permasalahan melibatkan identifikasi isu-isu terkait efisiensi APBN yang menjadi perhatian publik dan bagaimana sentimen masyarakat terhadap isu-isu tersebut dapat memengaruhi persepsi terhadap pemerintah dan kebijakan fiskal [6].

Dan, penelitian ini mengadopsi pendekatan kuantitatif untuk menganalisis sentimen masyarakat terhadap kebijakan efisiensi APBN, dengan fokus pada data yang diekstrak dari platform media sosial X (Twitter) [7].

Metode Analisis Data

1. Studi Literatur

Tahap awal kegiatan penelitian dengan peninjauan literatur yang relevan untuk memahami konsep dan metodologi analisis sentimen yang ada serta menggali penelitian sebelumnya tentang opini publik yang ada di aplikasi X (Twitter).

2. Business Understanding

Tahap ini bertujuan untuk memahami isu terkini yang mendasari penelitian ini, yaitu untuk mengetahui bagaimana sentimen netizen terhadap efisiensi APBN TA 2025.

3. Data Understanding

Pada tahap ini dilakukan pengumpulan data mentah (crawling) dari aplikasi X (Twitter) dengan menggunakan kata kunci "efisiensi" dan "APBN" dalam rentang waktu 15 - 30 Januari 2025 dengan memanfaatkan alat pihak ketiga.

4. Data Preparation

Tahap ini melibatkan pembersihan data dari *noise* (misalnya: simbol, tagar, tautan, emoji yang tidak relevan), penghapusan kata-kata yang tidak penting, dan standarisasi format teks.

5. Pengujian dan Pemodelan

Tahap akhir dengan menggunakan fitur *Sentiment Analysis* pada *Orange Data Mining* untuk melakukan analisis sentimen dari data teks yang telah diproses. Selain itu, akan dilakukan pengujian dengan menggunakan algoritma Naïve Bayes untuk melihat akurasi dan confusion matrix dari model yang dihasilkan.

Metode Pengolahan Data

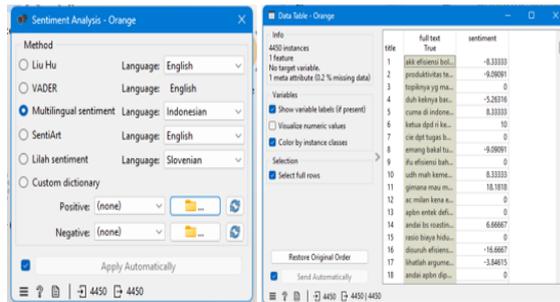
Penelitian ini memanfaatkan aplikasi perangkat lunak *Orange Data Mining* yang menyediakan berbagai algoritma dan teknik untuk analisis data [8].

1. *Crawling* Data Twitter: Proses pengumpulan data dari aplikasi X (Twitter) dilakukan dengan menggunakan alat pihak ketiga. Alat yang digunakan adalah Tweet-Harvest (perangkat lunak berbasis Python milik Helmi Satria).
2. Pembersihan Data Teks: Data teks yang diperoleh dari aplikasi X seringkali mengandung elemen-elemen yang tidak relevan untuk analisis sentimen, seperti tautan URL, tag pengguna (*username*), karakter khusus, dan angka. Dalam hal ini, digunakan Google Colab dengan skrip python.
3. Analisis Sentimen dengan *Orange Data Mining*: *Orange Data Mining* menyediakan fitur khusus untuk analisis sentimen, yang memungkinkan klasifikasi teks ke dalam kategori sentimen positif, negatif, atau netral.

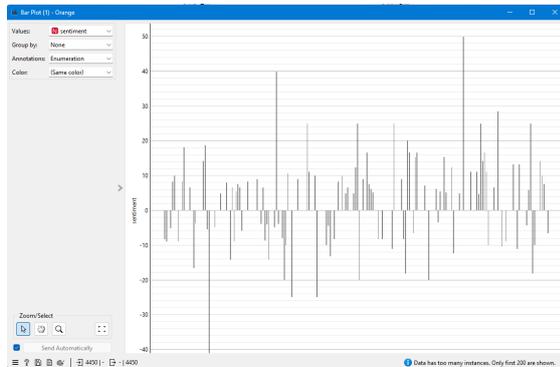
IV. HASIL DAN PEMBAHASAN

Skenario Penelitian

Kegiatan penelitian ini sebagian besar menggunakan aplikasi *Orange Data Mining* yang secara umum terdiri dari dua kegiatan utama yaitu proses analisis dan proses pengujian. Kegiatan tersebut disajikan dalam bentuk *Design Widget Text Mining* berupa alur proses yang ditunjukkan pada Gambar 1 dan Gambar 2 berurutan.



Gambar 5. Analisis Sentimen



Gambar 6. Bar Plot Sentimen

Kemudian, tabel hasil penilaian/analisis, akan disimpan dan dilakukan pelabelan sentimen. Angka negatif berarti sentimen negatif, angka nol berarti netral, dan angka positif berarti sentimen positif (Gambar 7).

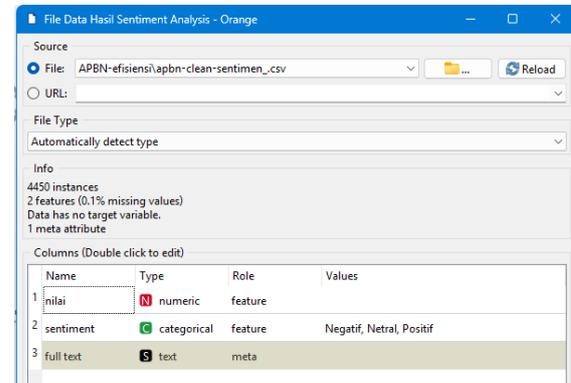
	full text	sentiment	
1	akk efisien...	-833.333	Negatif
2	produktivi...	-909.091	Negatif
3	topiknya y...	0	Netral
4	duh keny...	-526.316	Negatif
5	cuma di ir...	833.333	Positif
7	ketua dpd...	10	Positif
3	cie dpt tug...	0	Netral
3	emang ba...	-909.091	Negatif
0	ifu efisien...	0	Netral
1	udh mah l...	833.333	Positif
2	gimana m...	181.818	Positif

Gambar 7. Data Hasil Analisis Sentimen dengan Label

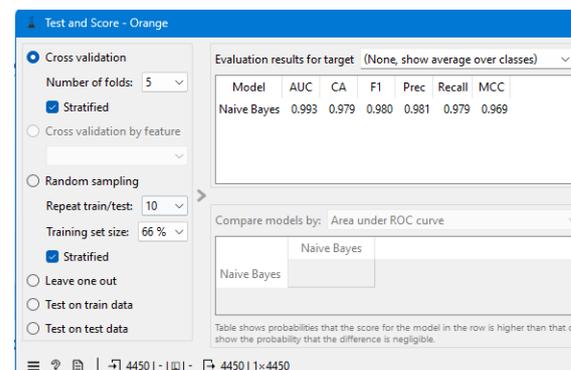
Pengujian Algoritma Naïve Bayes dan Pemodelan (Prediksi)

Pengujian hasil sentimen dilakukan dengan menggunakan algoritma Naïve Bayes. Data teks yang digunakan adalah data hasil pelabelan berdasarkan perhitungan dengan menetapkan kolom sentiment sebagai target. Gambar 8

menunjukkan pengujian file data hasil analisis sentiment. Sementara itu hasil pengujian dan nilai akurasi Naïve Bayes ditunjukkan pada Gambar 9. Hasil pengujian naïve bayes menghasilkan akurasi yang mengukur seberapa banyak prediksi model yang benar dibandingkan dengan total jumlah data. Selain itu, secara bersamaan dilakukan juga pemodelan untuk prediksi sentimen.



Gambar 8. Pengujian File Data Hasil Analisis Sentimen



Gambar 9. Hasil Pengujian dan Nilai Akurasi dengan Naïve Bayes

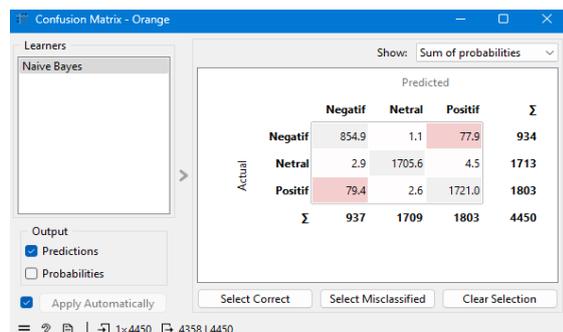
Pada Gambar 9, model dievaluasi menggunakan cross-validation dengan 5-fold. Data dibagi menggunakan *stratified sampling*, yang memastikan distribusi kelas tetap seimbang di setiap fold. Repeat train/test dilakukan 10 kali dengan rasio 66% untuk pelatihan dan 34% untuk pengujian. Hasil pengujian menunjukkan:

1. Hasil Evaluasi Model Naïve Bayes: AUC (*Area Under Curve*): 0.993 yang berarti model memiliki kemampuan diskriminasi

- yang sangat tinggi dalam membedakan kelas positif dan negatif.
2. CA (Classification Accuracy) sebesar 0.979 yang berarti model dapat mengklasifikasikan 97.9% data dengan benar.
 3. F1-Score sebesar 0.980 yang menunjukkan keseimbangan antara precision dan recall, dengan kinerja yang sangat baik.
 4. Precision sebesar 0.981 yang berarti dari semua prediksi positif, 98.1% benar.
 5. Recall sebesar 0.979 yang berarti semua data sebenarnya positif, 97.9% berhasil terdeteksi.
 6. MCC (Matthews Correlation Coefficient) sebesar 0.969 yang berarti model memiliki hubungan korelasi yang sangat kuat antara prediksi dan label sebenarnya.

Dari hasil diatas menunjukkan bahwa model Naïve Bayes memiliki performa yang sangat baik dengan akurasi tinggi, keseimbangan precision-recall yang optimal, serta kemampuan klasifikasi yang sangat baik berdasarkan AUC. Hal ini mengindikasikan bahwa model sangat efektif untuk tugas klasifikasi pada dataset ini.

Untuk mengukur kinerja model lebih detail digunakan *confusion matrix* yang menunjukkan jumlah True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN) yang dibuat oleh model. Gambar 10 menampilkan hasil *confusion matrix* dari model naïve bayes.



Gambar 10. Confusion Matrix

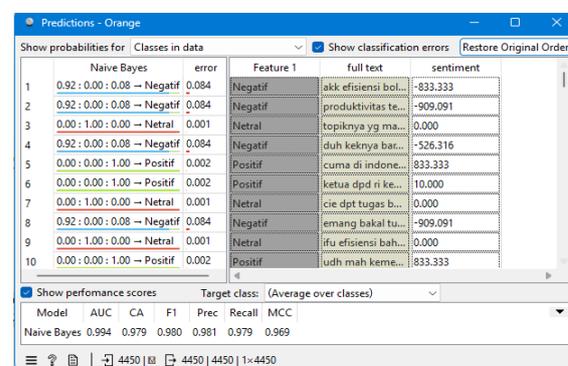
Dari Gambar 10 dapat dilihat yaitu:

1. *True Positives* (Prediksi Benar):

- Negatif: 854 dari 934 sampel diklasifikasikan dengan benar sebagai Negatif.
- Netral: 1711 dari 1713 sampel diklasifikasikan dengan benar sebagai Netral.
- Positif: 1717 dari 1803 sampel diklasifikasikan dengan benar sebagai Positif.

2. *Misclassifications* (Kesalahan Prediksi):

- 79 sampel Positif diklasifikasikan sebagai Negatif.
- 77 sampel Negatif diklasifikasikan sebagai Positif.
- 2,9 sampel Netral diklasifikasikan sebagai Negatif.
- 4,3 sampel Netral diklasifikasikan sebagai Positif.



Gambar 11. Pemodelan dan Prediksi

Hasil pemodelan dan prediksi sentimen menggunakan model Naïve Bayes di Orange Data Mining disajikan pada Gambar 11. Tujuan dari pemodelan dan prediksi ini adalah untuk menganalisis sentimen dari teks berdasarkan probabilitas yang dihasilkan oleh model. Berikut adalah beberapa aspek yang ditampilkan dalam Gambar 11:

1. Prediksi Sentimen

- Setiap baris mewakili sampel teks yang telah diproses oleh model.
- Kolom Naïve Bayes menunjukkan probabilitas untuk masing-masing kelas sentimen (Negatif, Netral, atau Positif).

- Kolom error menunjukkan Tingkat kesalahan prediksi untuk setiap sampel.
2. Fitur Teks
 - Fitur 1 menampilkan nilai fitur dari data teks yang diproses.
 - Full text berisi potongan teks asli yang digunakan untuk analisis.
 - Sentiment menunjukkan sentimen yang telah diberikan pada data.
 3. Evaluasi Model
 - AUC (0.99) menunjukkan kemampuan model membedakan antara kelas sentimen sangat baik.
 - CA (0.979) menunjukkan akurasi model cukup tinggi.
 - F1-score (0.98) memiliki arti bahwa model cukup seimbang dalam memprediksi setiap kategori.
 - Precision dan Recall (0.98 dan 0.979) bahwa model memiliki tingkat prediksi yang baik.
 - MCC (0.969) menunjukkan korelasi antara prediksi dan label asli cukup tinggi.

V. PENUTUP

Berdasarkan hasil-hasil pengujian dapat diambil kesimpulan yaitu:

1. Kebijakan efisiensi APBN TA 2025 yang dilakukan pemerintah cenderung mendapat dukungan karena cuitan yang bersentimen positif lebih banyak apalagi jika ditambah cuitan yang bersentimen netral.
2. Hasil analisis sentimen yang dihasilkan pada penelitian ini semata-mata hanya menggunakan fitur analisis yang disediakan *Orange Data Mining*. Metode yang digunakan pun hanya satu, yaitu *Multilingual Sentiment*. Penelitian ini hanya bisa menyajikan hasil analisis dengan satu metode. Untuk penelitian selanjutnya, analisis bisa menggunakan metode lain agar bisa dilakukan komparasi hasil analisis.
3. Validasi hasil prediksi model Naïve Bayes dalam klasifikasi sentimen teks memperlihatkan bahwa model memiliki akurasi tinggi dalam mendeteksi sentimen negatif, netral, dan positif.
4. Untuk meningkatkan akurasi lebih lanjut, dapat dilakukan analisis lebih dalam terhadap kesalahan prediksi, serta menggunakan teknik peningkatan data seperti penggunaan embedding atau leksikon sentimen yang lebih luas.

DAFTAR PUSTAKA

- [1] B. A. A. Uliansyah, A. Daron, and F. Febrian, "Analitika Data Dalam Pengelolaan Keuangan Negara: Use Case Data Dasar Dana Alokasi Umum," Jul. 16, 2021, Indonesian Institute of Information Technology. doi: 10.34148/teknika.v10i2.344.
- [2] F. H. Rachman and I. Imamah, "Pendekatan Data Science untuk Mengukur Empati Masyarakat terhadap Pandemi Menggunakan Analisis Sentimen dan Seleksi Fitur," Dec. 20, 2022, Tanjungpura University. doi: 10.26418/jp.v8i3.56655.
- [3] A. Arham, E. R. Swedia, M. Cahyanti, and M. R. D. Septian, "Implementasi Sentiment Analysis Pada Opini Masyarakat Indonesia Di Twitter Terhadap Virus Covid-19 Varian Omicron Dengan Algoritma Naïve Bayes, Decision Tree, Dan Support Vector Machine," Dec. 21, 2022. doi: 10.46984/sebatik.v26i2.1961.
- [4] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," Jan. 31, 2021. doi: 10.32520/stmsi.v10i1.1179.
- [5] T. Tundo and D. N. Rachmawati, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Terhadap Program Makan Siang Gratis," Sep. 15, 2024. doi: 10.35870/jimik.v5i3.978.
- [6] F. Ramadhan, "Pemanfaatan Analisis Jaringan Sosial Dalam Penentuan Centrality Dalam Pengembangan Web Berita Online," Sep. 17, 2020. doi: 10.51519/journalcisa.v1i3.43.
- [7] K. Silaen, L. D. Maduwu, and A. Zebua, "Effect of Current Ratio and Return on Asset on Company Value (Case Study on Food and Beverage Subsector

- Manufacturing Companies Listed on the Indonesia Stock Exchange in 2018-2020),” Apr. 10, 2023. doi: 10.24815/jr.v6i2.30563.
- [8] M. Axhiu, F. Veljanoska, B. Ciglovska, and M. Husejini, “The Usage of Sentiment Analysis for Hearing the Voice of the Customer and Improving Businesses,” Jun. 01, 2014, Richtmann Publishing. doi: 10.5901/jesr.2014.v4n4p401.
- [9] L. Wikarsa, A. Angdresey, and J. Kapantow, “IMPLEMENTASI METODE NAÏVE BAYES DAN LEXICON-BASED APPROACH UNTUK MENGLASIFIKASI SENTIMEN NETIZEN PADA TWEET BERBAHASA INDONESIA,” Apr. 15, 2022. doi: 10.52159/realtech.v18i1.5.
- [10] M. Harahap, F. Rozi, Y. Yennimar, and S. D. Siregar, “Analisis Wawasan Penjualan Supermarket dengan Data Science,” Nov. 21, 2021. doi: 10.47709/dsi.v1i1.1173.
- [11] W. Gunawan and M. R. Firmansyah, “Monitoring dan Evaluasi Kinerja Karyawan menggunakan Algoritma Simple Additive Weighting dan Hungarian,” Aug. 27, 2020. doi: 10.33096/ilkom.v12i2.519.87-95.
- [12] A. R. L. Teluma, “Membaca Realitas Infodemi Covid-19 di Indonesia,” Aug. 12, 2020. doi: 10.29303/jcommsci.v1i1.91.