

Analisis Sentimen Publik terhadap Perkembangan Topik Program Makan Bergizi Gratis (MBG) Menggunakan Algoritma Support Vector Machine (SVM) dan K-Means Clustering

Rahma Yuni Simanullang*, Zulham Sitorus, Muhammad Iqbal

Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email : ^{1,*}rahmayunisimanullang2009@gmail.com, ²zulhamsitorus@dosen.pancabudi.ac.id,

³muhammadiqbal@dosenpancabudi.ac.id

Email Penulis Korespondensi: rahmayunisimanullang2009@gmail.com*

Submitted: **18/08/2026**; Accepted: **26/09/2026**; Published: **30/09/2026**

Abstrak–Program Makan Bergizi Gratis (MBG) merupakan salah satu program prioritas pemerintah yang mulai diterapkan secara bertahap pada tahun 2025 untuk membantu memenuhi kebutuhan gizi peserta didik. Pelaksanaan program tersebut menimbulkan beragam tanggapan masyarakat melalui media sosial sehingga menghasilkan data komentar yang besar dan tidak terstruktur. Kondisi tersebut menyulitkan identifikasi kecenderungan sentimen dan topik yang berkembang apabila dilakukan secara manual. Penelitian ini bertujuan menganalisis sentimen publik serta mengidentifikasi kelompok topik pembahasan terkait Program MBG menggunakan algoritma Support Vector Machine (SVM) dan K-Means Clustering. Data penelitian terdiri dari 2.783 komentar pada salah satu video Chanel YouTube yaitu COMPAS.TV yang membahas MBG, sedangkan survei dan wawancara terhadap guru dan siswa digunakan sebagai bahan perbandingan bagi penulis. Tahapan penelitian meliputi preprocessing, pelabelan sentimen, pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), klasifikasi menggunakan SVM, dan clustering menggunakan K-Means. Hasil klasifikasi SVM terhadap 557 data testing menghasilkan accuracy 81,51%, precision 81,48%, recall 81,51%, dan F1-score 81,29%. Selanjutnya, hasil Elbow Method menunjukkan jumlah cluster optimal sebanyak empat cluster (K=4). Dari 2.783 komentar, Cluster 1 terdiri atas 2.133 komentar (76,64%), Cluster 2 sebanyak 263 komentar (9,45%), Cluster 3 sebanyak 143 komentar (5,14%), dan Cluster 4 sebanyak 244 komentar (8,77%). Berdasarkan kata dominan, cluster merepresentasikan pandangan umum terhadap MBG, kritik terhadap manfaat dan pengelolaan MBG, keraguan dan kritik terhadap kebijakan pemerintah, serta kritik terhadap pemerintah dan pelaksanaan MBG. Hasil penelitian menunjukkan bahwa kombinasi SVM dan K-Means dapat memberikan gambaran mengenai kecenderungan sentimen dan topik pembahasan publik terhadap Program MBG.

Kata Kunci : Program Makan Bergizi Gratis; Analisis Sentimen; Support Vector Machine; K-Means Clustering; TF-IDF

1. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) mulai diperkenalkan sebagai salah satu program prioritas pemerintah pada tahun 2024 dan mulai diterapkan secara bertahap pada awal tahun 2025 di berbagai daerah di Indonesia. Pelaksanaan program ini ditujukan kepada peserta didik tingkat pendidikan dasar hingga menengah dengan tujuan membantu memenuhi kebutuhan gizi harian siswa selama proses pembelajaran berlangsung. Tahap awal implementasi program dilakukan secara bertahap dengan memperhatikan kesiapan sarana dan prasarana, sistem distribusi logistik, serta koordinasi antar pihak terkait sebelum diterapkan secara lebih luas. Program ini tidak hanya bertujuan untuk memenuhi kebutuhan gizi siswa, tetapi juga diharapkan mampu mendukung peningkatan kualitas sumber daya manusia serta mengurangi angka stunting di Indonesia dalam jangka panjang. Setelah ditetapkan nya Program MBG tersebut, program MBG secara cepat menjadi perhatian luas di tengah masyarakat dan memicu berbagai diskusi di ruang publik, terutama melalui media sosial. Diberbagai Platform digital seperti YT, Instagram, Tiktok, Facebook , Twitter dll menjadi sarana utama bagi masyarakat untuk menyampaikan pendapat, baik dalam bentuk dukungan, kritik, maupun saran terhadap pelaksanaan program tersebut.

Banyaknya komentar dan opini yang diperoleh, Terutama dari media sosial menyebabkan data yang terkumpul menjadi sangat besar dan tidak terstruktur. Kondisi ini menimbulkan tantangan dalam proses analisis karena sulit dilakukan secara manual untuk mengetahui kecenderungan sentimen serta topik yang sedang banyak dibahas. Selain itu, opini yang tersebar juga sangat beragam, tidak konsisten, dan sering mengandung informasi yang tidak relevan sehingga diperlukan tehnik maupun metode yang tepat agar hasil analisis lebih akurat. Oleh karena itu, dibutuhkan metode berbasis komputasi yang mampu mengolah data teks secara efektif dan efisien. Teknik yang dapat digunakan yaitu analisis sentimen untuk mengetahui kecenderungan opini masyarakat, serta analisis clustering untuk mengelompokkan topik-topik yang paling banyak dibahas.

Berdasarkan Perkembangan dari Program MBG tersebut, Penulis tertarik melakukan penelitian ini yang dimana Penulis akan memperoleh Data melalui opini Publik yg ada di media Sosial, data ulasan tersebut penulis ambil dari Data komentar dari salah satu Chanel Youtube tentang MBG yaitu COMPAS.TV . Selain memperoleh data dari media sosial, Penulis juga melakukan survei langsung ke beberapa sekolah untuk mendapatkan pendapat, komentar Atau ulasan, serta permasalahan terkait MBG. Dalam proses pengumpulan data tersebut, peneliti melakukan wawancara dan diskusi dengan guru serta siswa mengenai pelaksanaan MBG saat ini di lingkungan sekolah. Peneliti juga menyebarkan link kuesioner kepada guru dan beberapa siswa guna memperoleh data komentar

secara real dan penilaian secara lebih terstruktur. Data hasil survei langsung ini kemudian digunakan sebagai bahan perbandingan dengan ulasan publik yang diperoleh dari media sosial, sehingga hasil analisis dapat memberikan gambaran yang lebih objektif dan sesuai dengan kondisi nyata di lapangan. Hasil survei secara langsung menunjukkan bahwa komentar dan pendapat yang diperoleh dari guru maupun siswa tidak jauh berbeda dengan opini yang berkembang di media sosial, di mana terdapat berbagai tanggapan yang beragam, baik berupa komentar positif, negatif, maupun netral terkait pelaksanaan MBG saat ini.

Analisis sentimen merupakan salah satu teknik dalam data mining yang digunakan untuk mengenali, mengelompokkan, dan memahami opini atau tanggapan masyarakat terhadap suatu topik tertentu ke dalam kategori positif, negatif, maupun netral. Oleh karena itu, pada penelitian ini Penulis menggunakan metode Support Vector Machine (SVM) untuk melakukan proses klasifikasi sentimen. SVM adalah algoritma machine learning berbasis supervised learning yang bekerja dengan menentukan hyperplane terbaik sebagai pemisah antar kelas data. Metode ini juga dikenal memiliki performa dan tingkat akurasi yang baik dalam pengolahan serta klasifikasi data teks [1],[2].

Selain metode SVM, penelitian ini juga menggunakan metode K-Means Clustering untuk mengelompokkan data berdasarkan kesamaan karakteristik yang dimiliki. K-Means merupakan salah satu metode unsupervised learning yang bekerja dengan membagi data ke dalam beberapa kelompok (cluster) berdasarkan kedekatan antar data dalam kelompok tersebut. Metode ini dapat membantu menemukan pola maupun topik tertentu yang tersembunyi di dalam data. Dengan menggabungkan kedua metode tersebut, diharapkan penelitian dapat memberikan hasil analisis yang lebih lengkap terkait persepsi masyarakat terhadap Program MBG serta mendukung proses evaluasi kebijakan yang lebih berdasarkan data.

Berdasarkan Penelitian yang dilakukan oleh Chen. dalam jurnal Decision Support Systems berjudul Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast menunjukkan bahwa kombinasi metode Support Vector Machine (SVM) dan K-Means Clustering mampu digunakan secara efektif untuk menganalisis data teks pada forum online. Hasil penelitian tersebut membuktikan bahwa SVM dapat mengklasifikasikan sentimen dengan baik, sementara K-Means mampu mengelompokkan topik yang sedang berkembang, sehingga menghasilkan analisis yang lebih komprehensif [3].

Selanjutnya Penelitian yang dilakukan oleh Mukhlisin dan Berliana. dalam artikel tersebut berjudul Literature Review: Comparison of Machine Learning Algorithms for Sentiment Analysis of Free Nutritious Meals mengkaji 18 studi terkait analisis sentimen pada Program Makan Bergizi Gratis (MBG). Hasil penelitian tersebut menunjukkan bahwa berbagai algoritma machine learning memiliki tingkat efektivitas yang berbeda dalam menganalisis sentimen, di mana Support Vector Machine (SVM) terbukti menjadi salah satu metode yang memiliki kinerja baik dalam klasifikasi sentimen pada data teks [4].

2. METODOLOGI PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari text mining yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam data teks [5],[6],[7],[8]. Analisis ini umumnya digunakan untuk mengetahui kecenderungan pendapat seseorang terhadap suatu objek, yang dapat berupa sentimen positif, negatif, atau netral [9],[10],[11]. Dalam konteks media sosial, analisis sentimen menjadi penting karena mampu mengolah data opini publik dalam jumlah besar dan tidak terstruktur menjadi informasi yang lebih bermakna [12],[13]. Proses analisis sentimen biasanya melibatkan beberapa tahapan utama, yaitu pengumpulan data, preprocessing, ekstraksi fitur, serta klasifikasi. Dengan memanfaatkan teknik machine learning, analisis sentimen dapat dilakukan secara otomatis dengan tingkat akurasi yang cukup tinggi [14],[15],[16],[17],[18],[19]. Oleh karena itu, metode ini banyak digunakan dalam berbagai bidang, termasuk evaluasi kebijakan publik, pemasaran, dan analisis perilaku pengguna [20],[21],[22],[23],[24],[25].

2.2 Algoritma SVM

Support Vector Machine (SVM) merupakan salah satu algoritma dalam bidang machine learning yang termasuk ke dalam kategori supervised learning. Algoritma ini digunakan untuk melakukan proses klasifikasi dengan cara menentukan suatu batas pemisah optimal (optimal hyperplane) yang mampu memaksimalkan jarak (margin) antar kelas data dalam ruang berdimensi tinggi (N-dimensi) [26]. Dengan prinsip tersebut, SVM memiliki kemampuan yang baik dalam memisahkan data secara efektif, khususnya pada kasus dengan jumlah dimensi yang besar. Metode SVM pertama kali dikembangkan pada dekade 1990-an oleh Vladimir N. Vapnik bersama para peneliti lainnya [27],[28].

2.3 Algoritma K-Means

Algoritma K-Means merupakan salah satu metode clustering yang tertua dan paling banyak digunakan dalam menganalisis data [29]. Algoritma K-Means pertama kali dikenalkan oleh Stuart Lloyd pada tahun 1957 dalam konteks pemrosesan sinyal di Bell Labs [30], [31],[32]. Adapun rumus pengerjaan metode K-Means dapat dilihat sebagai berikut :

- Menentukan jumlah nilai K sebagai total cluster
- Mengambil sejumlah titik acak sebanyak nilai K
- Melakukan perhitungan jarak menggunakan rumus Euclidean Distance antara setiap data input dan setiap centroid, sehingga ditemukan jarak terdekat antara data dan centroid. Berikut rumus persamaan *euclidean Distance*:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan:

x_i : data kriteria

μ_j : centroid pada cluster ke-j

- Mengelompokkan setiap data berdasarkan kedekatannya dengan centroid yang memiliki jarak terkecil.
- Mengganti nilai centroid dengan nilai baru yang dihitung sebagai rata-rata dari data dalam cluster yang bersangkutan, menggunakan rumus yang sesuai:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad (2)$$

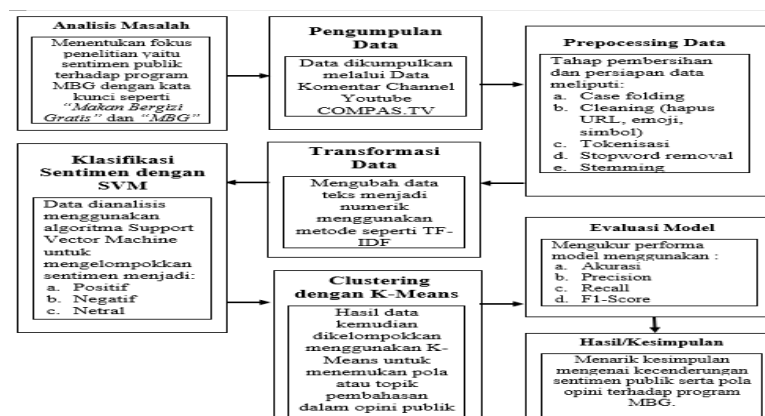
- Teruskan perulangan dari langkah 2 hingga 5 sampai tidak ada perubahan dalam anggota setiap cluster.

2.4 MBG

Program Makan Bergizi Gratis (MBG) adalah kebijakan penyediaan makan siang tanpa biaya yang dijalankan pada masa pemerintahan Prabowo Subianto dan mulai dilaksanakan secara bertahap sejak Januari 2025 [33]. Program ini menasar peserta didik dari jenjang PAUD hingga SMA/SMK, serta ibu hamil dan menyusui. Meskipun bertujuan untuk meningkatkan asupan gizi masyarakat dan menekan angka stunting, implementasinya memunculkan berbagai kritik dan penolakan di sejumlah daerah, terutama akibat munculnya kasus keracunan massal [34].

2.5 Tahapan Penelitian

Adapun Tahapan penelitian yg dilakukan penulis untuk menyelesaikan Penelitian ini dapat dilihat pada Gambar 1 berikut :



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Penelitian ini dilaksanakan secara daring (online) dengan memanfaatkan media sosial sebagai sumber data penelitian. Data yang digunakan berupa ulasan, komentar, dan opini publik terkait Program Makan Bergizi Gratis (MBG) yang diperoleh melalui 1 channel platform Youtube yaitu COMPAS.TV dengan link sebagai berikut : <https://youtu.be/BkZwWNtcJko?si=K5yjpILnEcNwjsn>. Berikut Contoh data yg sudah di scrapping :

Tabel 1. Data Ulasan/komentar Hasil Scrapping pada Postingan di Youtube

No	Author	Comment	Published Time
1	@Redmic-lx4xo	yang di samping Prabowo kebanyakan makan telur 🤔🤔🤔🤔🤔🤔	1 month ago
2	@diwansetyawan705	MBG sdh di bilang tidak bermanfaat msih ja dipaksakan	2 months ago

3	@cacasramdani2253	Hentikan mbg	2 months ago
4	@enanenan1421	KALAU PEMIMPIN SEORANG PENGUSAHA YG DI UTAMAKAN BURUH APA PENGUSAHA ?? MUDAH2AN TIDAK OMON2.MBG TIDAK TEPAT SASARAN YG MEMBUTUHKAN ITU YG KURANG GIZI,SEKOLAHAN ELITE KO DAPAT BISA2 NURUNIN GIZINYA 😊	3 months ago (edited)
5	@ho750418	5:15 MBG bermanfaat tidak? tidak? bermanfaat dong. bagi pegang bendera. pemasukan dana di kantong	3 months ago
....			
2.783	@benibeni7457	Mubajir	3 months ago

3.2 Tahapan *Preprocessing* Data

3.2.1 Proses Tahapan *Cleaning*

Tahapan *cleaning* bertujuan untuk menghilangkan berbagai elemen yang tidak relevan dari data komentar YouTube, seperti tautan (*URL*), simbol, tanda baca, angka, serta karakter khusus lainnya yang berpotensi mengganggu proses pengolahan data. Kegiatan ini dilakukan untuk menghasilkan data teks yang lebih rapi, konsisten, dan siap digunakan dalam proses analisis sentimen menggunakan algoritma *machine learning*. Pada tahap ini, komentar yang mengandung data duplikat, komentar kosong, emoji, simbol, URL, serta karakter yang tidak diperlukan dihapus dari dataset. Berikut hasil dari proses *cleaning* tersebut dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cleaning*

No	Comment Asli	Cleaning
1	yang di samping Prabowo kebanyakan makan telur 😊😊😊😊😊😊😊😊	yang di samping Prabowo kebanyakan makan telur
2	MBG sdh di bilang tidak bermanfaat msih ja dipaksakan	MBG sdh di bilang tidak bermanfaat msih ja dipaksakan
3	Hentikan mbg	Hentikan mbg
4	KALAU PEMIMPIN SEORANG PENGUSAHA YG DI UTAMAKAN BURUH APA PENGUSAHA ?? MUDAH2AN TIDAK OMON2.MBG TIDAK TEPAT SASARAN YG MEMBUTUHKAN ITU YG KURANG GIZI,SEKOLAHAN ELITE KO DAPAT BISA2 NURUNIN GIZINYA 😊	KALAU PEMIMPIN SEORANG PENGUSAHA YG DI UTAMAKAN BURUH APA PENGUSAHA MUDAH AN TIDAK OMON MBG TIDAK TEPAT SASARAN YG MEMBUTUHKAN ITU YG KURANG GIZISEKOLAHAN ELITE KO DAPAT BISA NURUNIN GIZINYA
5	5:15 MBG bermanfaat tidak? tidak? bermanfaat dong. bagi pegang bendera. pemasukan dana di kantong	MBG bermanfaat tidak tidak bermanfaat dong bagi pegang bendera pemasukan dana di kantong
...	...	...
2.873	Mubajir	Mubajir

3.2.2 Case Folding dan Normalisasi

Setelah tahapan *cleaning* selesai dilakukan, proses selanjutnya adalah *case folding* dan normalisasi teks. Tahap *case folding* bertujuan untuk menyeragamkan format penulisan dengan mengonversi seluruh huruf dalam komentar menjadi huruf kecil (*lowercase*). Langkah ini dilakukan agar kata yang memiliki arti sama tetapi ditulis dengan variasi huruf besar dan kecil, seperti "Program", "PROGRAM", dan "program", dapat dikenali sebagai satu istilah yang sama oleh sistem. Selanjutnya, dilakukan proses normalisasi untuk mengubah kata-kata yang tidak sesuai dengan kaidah bahasa baku, termasuk singkatan, istilah gaul, maupun bentuk bahasa informal lainnya menjadi bentuk yang lebih standar. Berikut pada Tabel 3 ditampilkan Hasil Case Folding dan Normalisasi

Tabel 3. Hasil Case folding dan Normalisasi

No	Case Folding	Normalisasi
1	yang di samping prabowo kebanyakan makan telur	yang di samping prabowo kebanyakan makan telur

2	mbg sdh di bilang tidak bermanfaat msih ja dipaksakan	mbg sudah di bilang tidak bermanfaat msih ja dipaksakan
3	hentikan mbg	hentikan mbg
4	kalau pemimpin seorang pengusaha yg di utamakan buruh apa pengusaha mudah an tidak omon mbg tidak tepat sasaran yg membutuhkan itu yg kurang gizisekolahan elite ko dapat bisa nurunin gizinya	kalau pemimpin seorang pengusaha yang di utamakan buruh apa pengusaha mudah an tidak omon mbg tidak tepat sasaran yang membutuhkan itu yang kurang gizisekolahan elite ko dapat bisa nurunin gizinya
5	mbg bermanfaat tidak tidak bermanfaat dong bagi pegang bendera pemasukan dana di kantong	mbg bermanfaat tidak tidak bermanfaat dong bagi pegang bendera pemasukan dana di kantong
.....	...	...
2.783	mubajir	mubajir

3.2.3 Tokenization, Stopword removal, dan Stemming

Setelah proses case folding dan normalisasi selesai dilakukan, tahapan berikutnya meliputi tokenization, stopwords removal, dan stemming. Ketiga tahapan tersebut bertujuan untuk menyederhanakan serta mempersiapkan data teks agar lebih efektif digunakan dalam proses ekstraksi fitur dan klasifikasi sentimen. Tokenization dilakukan dengan memecah kalimat menjadi unit-unit kata yang lebih kecil yang disebut token. Selanjutnya, proses stopwords removal digunakan untuk menghilangkan kata-kata yang sering muncul tetapi tidak memberikan kontribusi penting terhadap penentuan sentimen, seperti kata penghubung maupun kata depan. Berikut dapat dilihat pada Tabel 4.

Tabel 4. Tokenization, Stopword removal, dan Stemming

No.	Tokenizing	Stopword Removal	Stemming
1	yang di samping prabowo kebanyakan makan telur	samping prabowo kebanyakan makan telur	samping prabowo banyak makan telur
2	mbg sudah di bilang tidak bermanfaat msih ja dipaksakan	mbg bilang tidak bermanfaat msih ja dipaksakan	mbg bilang tidak manfaat msih ja paksa
3	hentikan mbg	hentikan mbg	henti mbg
4	kalau pemimpin seorang pengusaha yang di utamakan buruh apa pengusaha mudah an tidak omon mbg tidak tepat sasaran yang membutuhkan itu yang kurang gizisekolahan elite ko dapat bisa nurunin gizinya	kalau pemimpin seorang pengusaha utamakan buruh apa pengusaha mudah an tidak omon mbg tidak tepat sasaran membutuhkan kurang gizisekolahan elite ko nurunin gizinya	kalau pimpin orang usaha utama buruh apa usaha mudah an tidak omon mbg tidak tepat sasar butuh kurang gizisekolahan elite ko nurunin gizi
5	mbg bermanfaat tidak tidak bermanfaat dong bagi pegang bendera pemasukan dana di kantong	mbg bermanfaat tidak tidak bermanfaat dong pegang bendera pemasukan dana kantong	mbg manfaat tidak tidak manfaat dong pegang bendera pasu dana kantong
.....			
2.783	mubajir	mubajir	mubajir

3.3 Pelabelan Sentimen

Setelah melalui tahap preprocessing, tahap selanjutnya adalah melakukan pelabelan sentimen terhadap 2.783 komentar publik mengenai Program Makan Bergizi Gratis (MBG) menggunakan metode Lexicon-Based dengan InSet Lexicon yg diolah atau diproses menggunakan Python. Proses pelabelan dilakukan dengan mencocokkan setiap kata pada teks hasil stemming dengan kata yang terdapat dalam lexicon, kemudian menjumlahkan nilai bobot polaritas untuk memperoleh skor sentimen pada setiap komentar. Berdasarkan skor yang diperoleh, komentar dikategorikan menjadi tiga kelas, yaitu Positif apabila skor > 0, Negatif apabila skor < 0, dan Netral apabila skor = 0. Hasil pelabelan sentimen terhadap data komentar ditunjukkan pada Tabel 5.

Tabel 5. Hasil Labeling

No	Stemming	Score	Label
1	samping prabowo banyak makan telur	0	Netral
2	mbg bilang tidak manfaat msih ja paksa	-4	Negatif
3	henti mbg	-4	Negatif
4	kalau pimpin orang usaha utama buruh apa usaha mudah an tidak omon mbg tidak tepat sasar butuh kurang gizisekolahan elite ko nurunin gizi	-21	Negatif
5	mbg manfaat tidak tidak manfaat dong pegang bendera	0	Netral

...	pasu dana kantong	...	...
2.783	mubajir	0	Netral

Berdasarkan Tabel 5, terlihat bahwa setiap komentar memperoleh skor sentimen yang digunakan untuk menentukan kelas sentimennya. Sebagai contoh, komentar “henti | mbg” memperoleh skor -4 sehingga dikategorikan sebagai Negatif, sedangkan komentar “mbg | banyak | copet | nya” memperoleh skor 3 sehingga dikategorikan sebagai Positif. Komentar yang memperoleh skor 0, seperti “samping | prabowo | banyak | makan | telur”, dikategorikan sebagai Netral. Hasil pelabelan terhadap seluruh 2.783 komentar ini selanjutnya digunakan sebagai label kelas dalam proses klasifikasi menggunakan algoritma Support Vector Machine (SVM).

3.4 TF-IDF

Setelah proses pelabelan sentimen selesai, tahap selanjutnya adalah melakukan pembobotan teks menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Tahap ini bertujuan untuk mengubah data komentar yang berbentuk teks menjadi representasi numerik sehingga dapat digunakan sebagai masukan pada proses klasifikasi menggunakan algoritma Support Vector Machine (SVM). Pembobotan TF-IDF dilakukan dengan mempertimbangkan frekuensi kemunculan suatu kata dalam dokumen serta tingkat kemunculan kata tersebut pada keseluruhan dokumen. Semakin sering suatu kata muncul dalam dokumen tetapi semakin jarang ditemukan pada dokumen lainnya, maka kata tersebut memiliki bobot TF-IDF yang semakin tinggi. Berdasarkan hasil pembobotan terhadap 2.783 komentar, diperoleh sejumlah fitur yang selanjutnya digunakan sebagai representasi numerik dalam proses klasifikasi sentimen. Berikut ditampilkan pada Tabel 6 yaitu Hasil TF-IDF.

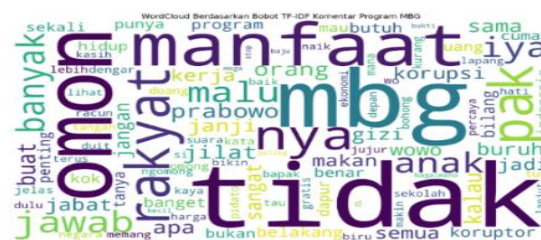
Tabel 6. Hasil TF-IDF

No	Kata	DF	IDF	Bobot TF-IDF
1	tidak	564	2.371585	0.066219
2	mbg	517	2.458435	0.05414
3	omon	163	3.608544	0.044679
4	manfaat	283	3.059436	0.040601
5	rakyat	190	3.456137	0.02113
6	pak	158	3.639506	0.021097
7	nya	130	3.833213	0.01767
8	jawab	101	4.083438	0.016319
9	malu	71	4.431745	0.014691
....				
1221	saat	2	7.609798	0.00016

Berdasarkan hasil pengolahan data, dari total 2.783 komentar yang digunakan dalam penelitian, sebanyak 2.226 data digunakan sebagai data training dan 557 data digunakan sebagai data testing.

3.5 Visualisasi WordCloud berdasarkan TF-IDF

Setelah proses pembobotan TF-IDF dilakukan, tahap selanjutnya adalah melakukan visualisasi kata menggunakan WordCloud. Visualisasi ini digunakan untuk memberikan gambaran mengenai kata-kata yang memiliki bobot TF-IDF dominan pada komentar publik mengenai Program Makan Bergizi Gratis (MBG). WordCloud dibentuk berdasarkan nilai bobot TF-IDF yang diperoleh dari data training, sehingga ukuran kata pada visualisasi merepresentasikan tingkat kepentingan relatif kata dalam kumpulan dokumen.



Gambar 2. Wordcloud Hasil TF-IDF

3.6 Klasifikasi Sentimen Menggunakan Support Vector Machine (SVM)

Setelah proses preprocessing, pelabelan sentimen, pembagian data, dan pembobotan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) selesai dilakukan, tahap selanjutnya adalah melakukan klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM). Pada tahap ini, data yang telah diberikan label sentimen digunakan untuk membangun model klasifikasi. Representasi numerik berupa bobot TF-IDF digunakan sebagai masukan bagi model SVM untuk mempelajari pola dari data komentar yang telah dikelompokkan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Data dibagi menjadi data training dan data testing dengan rasio

80:20. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data yang belum pernah digunakan selama proses pelatihan.

3.6.1 Pembentukan Model SVM

Pada proses pembentukan model, algoritma SVM menggunakan kernel linear dengan nilai parameter C sebesar 1,0. Pemilihan kernel linear digunakan karena data komentar telah direpresentasikan dalam bentuk fitur TF-IDF yang memiliki dimensi tinggi, sehingga pendekatan linear sesuai untuk menemukan batas pemisah antar kelas sentimen. Model SVM dilatih menggunakan 2.226 data training beserta label sentimennya. Hasil proses training menunjukkan bahwa model SVM berhasil dibentuk dengan tiga kelas sentimen, yaitu Negatif, Netral, dan Positif. Ketiga kelas tersebut menjadi dasar bagi model untuk melakukan klasifikasi pada 557 data testing. Dengan demikian, model yang telah terbentuk selanjutnya dapat digunakan untuk melakukan prediksi sentimen terhadap data testing.

Tabel 7. Informasi Pembentukan Model SVM

Parameter	Hasil
Algoritma	Support Vector Machine (SVM)
Kernel	Linear
Nilai C	1,0
Data Training	2.226
Data Testing	557
Jumlah Kelas	3
Kelas Sentimen	Negatif, Netral, Positif

3.6.2 Data Testing

Setelah model SVM selesai dilatih, kita gunakan model tersebut untuk memprediksi 557 data testing. Setelah model Support Vector Machine (SVM) berhasil dibentuk menggunakan 2.226 data training, tahap selanjutnya adalah melakukan prediksi terhadap 557 data testing. Data testing digunakan untuk mengetahui kemampuan model dalam mengklasifikasikan komentar yang belum digunakan pada proses pelatihan ke dalam tiga kelas sentimen, yaitu Negatif, Netral, dan Positif. Berdasarkan hasil prediksi, dari 557 data testing diperoleh 315 komentar diprediksi sebagai Negatif, 136 komentar diprediksi sebagai Netral, dan 106 komentar diprediksi sebagai Positif. Hasil prediksi tersebut selanjutnya dibandingkan dengan label aktual untuk mengetahui tingkat ketepatan model melalui confusion matrix dan classification report.

Tabel 8. Hasil Prediksi Data Testing SVM

No	Teks	Label	Prediksi SVM
1	ni orang buta tuli percuma mau ngomong tidak dekat kuping tidak dengar	Negatif	Negatif
2	tidak ira dulu pilih malah	Negatif	Negatif
3	tolong sembako bahan bahan turun tidak beres	Negatif	Negatif
4	mbg tidak tumbuh sel anak anak paktapi tumbuh perut buncit tri bgn kawan kawan nya	Negatif	Negatif
5	kalian situ cuma dengerin cerita	Negatif	Negatif
....			
558	tidak manfaat program basi	Negatif	Negatif

Tabel di atas hanya menampilkan beberapa contoh data testing, sedangkan keseluruhan 557 data dapat disimpan dalam file hasil prediksi sebagai data pendukung penelitian. Hasil distribusi prediksi dapat disajikan sebagai:

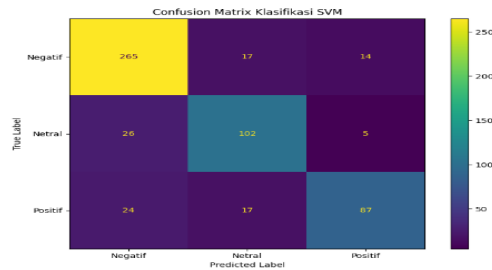
Tabel 9. Hasil Distribusi Prediksi Data Training

Kelas Sentimen	Jumlah Prediksi
Negatif	315
Netral	136
Positif	106
Total	557

3.6.3 Confusion Matrix Model SVM

Setelah dilakukan prediksi terhadap 557 data testing, tahap selanjutnya adalah mengevaluasi hasil klasifikasi menggunakan confusion matrix. Confusion matrix digunakan untuk mengetahui tingkat ketepatan model Support Vector Machine (SVM) dalam mengklasifikasikan komentar ke dalam tiga kelas sentimen, yaitu Negatif, Netral, dan Positif. Matriks pada Gambar 5 menunjukkan perbandingan antara label aktual (True Label) dengan label hasil prediksi SVM (Predicted Label). Nilai pada diagonal utama menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar, sedangkan nilai di luar diagonal menunjukkan data yang mengalami kesalahan klasifikasi.

Berdasarkan hasil confusion matrix pada Gambar 5, dari 342 data yang sebenarnya berlabel Negatif, sebanyak 265 data berhasil diprediksi sebagai Negatif, 17 data diprediksi sebagai Netral, dan 14 data diprediksi sebagai Positif. Pada kelas Netral, dari 133 data aktual, sebanyak 102 data berhasil diprediksi sebagai Netral, sedangkan 26 data diprediksi sebagai Negatif dan 5 data sebagai Positif. Sementara itu, dari 128 data aktual Positif, sebanyak 87 data berhasil diprediksi sebagai Positif, 24 data diprediksi sebagai Negatif, dan 17 data diprediksi sebagai Netral. Secara keseluruhan, confusion matrix menunjukkan bahwa model SVM mampu melakukan klasifikasi dengan jumlah prediksi benar sebanyak 454 data dari 557 data testing, sedangkan 103 data lainnya mengalami kesalahan klasifikasi. Hasil ini selanjutnya digunakan untuk menghitung nilai precision, recall, F1-score, dan accuracy pada tahap evaluasi kinerja model SVM.



Gambar 3. Confusion Matrix Klasifikasi Model SVM

3.6.4 Classification Report

Setelah diperoleh hasil klasifikasi melalui confusion matrix, tahap selanjutnya adalah melakukan evaluasi kinerja model menggunakan classification report. Classification report digunakan untuk mengetahui kemampuan model Support Vector Machine (SVM) dalam mengklasifikasikan masing-masing kelas sentimen berdasarkan nilai precision, recall, F1-score, dan support. Evaluasi dilakukan terhadap 557 data testing yang terdiri dari kelas Negatif, Netral, dan Positif. Hasil classification report ditunjukkan pada Tabel 4.9. Berdasarkan Tabel 10, kelas Negatif memperoleh nilai precision sebesar 0,8413, recall sebesar 0,8953, dan F1-score sebesar 0,8674 dengan support sebanyak 296 data. Kelas Netral memperoleh precision sebesar 0,7500, recall sebesar 0,7669, dan F1-score sebesar 0,7584 dengan support sebanyak 133 data. Sementara itu, kelas Positif memperoleh precision sebesar 0,8208, recall sebesar 0,6797, dan F1-score sebesar 0,7436 dengan support sebanyak 128 data.

Model SVM memperoleh accuracy sebesar 0,8151 atau 81,51%. Nilai macro average menghasilkan precision sebesar 0,8040, recall sebesar 0,7806, dan F1-score sebesar 0,7898, sedangkan weighted average menghasilkan precision sebesar 0,8148, recall sebesar 0,8151, dan F1-score sebesar 0,8129. Hasil tersebut menunjukkan bahwa model SVM mampu melakukan klasifikasi sentimen terhadap komentar Program Makan Bergizi Gratis (MBG) dengan tingkat akurasi sebesar 81,51%. Selanjutnya, hasil evaluasi ini digunakan untuk membahas kemampuan model dalam mengenali setiap kelas sentimen. Berdasarkan hasil perhitungan yang dilakukan dengan menggunakan pemrograman python, kelas Negatif memperoleh precision sebesar 84,13%, recall sebesar 89,53%, dan F1-score sebesar 86,74%. Nilai tersebut sesuai dengan hasil classification report yang diperoleh dari implementasi model SVM.

Tabel 10. Classification Report

	Precision	Recall	F1-Score	Support
Negatif	0.8413	0.8953	0.8674	296
Netral	0.75	0.7669	0.7584	133
Positif	0.8208	0.6797	0.7436	128
Accuracy	0.8151	0.8151	0.8151	0.8151
Macro Avg	0.804	0.7806	0.7898	557
Weighted Avg	0.8148	0.8151	0.8129	557

3.6.5 Evaluasi Kinerja SVM

Setelah dilakukan proses klasifikasi menggunakan algoritma Support Vector Machine (SVM) dan diperoleh confusion matrix serta classification report, tahap selanjutnya adalah melakukan evaluasi terhadap kinerja model secara keseluruhan. Evaluasi dilakukan menggunakan empat metrik, yaitu Accuracy, Precision, Recall, dan F1-Score. Berdasarkan hasil pengujian terhadap 557 data testing, model SVM memperoleh nilai Accuracy sebesar 81,51%, Precision sebesar 81,48%, Recall sebesar 81,51%, dan F1-Score sebesar 81,29%. Hasil evaluasi tersebut menunjukkan kemampuan model dalam melakukan klasifikasi sentimen publik terhadap Program Makan Bergizi Gratis (MBG) ke dalam tiga kelas, yaitu Negatif, Netral, dan Positif.

Berdasarkan nilai tersebut, model SVM dapat dikatakan mampu memberikan hasil klasifikasi yang cukup baik, dengan tingkat ketepatan keseluruhan sebesar 81,51%. Nilai F1-Score sebesar 81,29% juga menunjukkan adanya keseimbangan yang relatif baik antara kemampuan model dalam menemukan data yang sesuai dengan kelasnya dan ketepatan hasil prediksi. Berdasarkan confusion matrix yang ditampilkan, Accuracy model adalah 81,51%. Perhitungan manual menunjukkan bahwa weighted recall juga sebesar 81,51%. Namun, nilai weighted

precision dan weighted F1-Score yang dihitung langsung dari confusion matrix tersebut tidak sama dengan output program sebesar 81,48% dan 81,29%.

Tabel 11. Evaluasi Kinerja SVM

Metrik	Nilai	Persentase
Accuracy	0.8151	81.51
Precision	0.8148	81.48
Recall	0.8151	81.51
F1-Score	0.8129	81.29

3.6.6 Analisis Clustering Menggunakan K-Means

3.6.6.1 Persiapan Data untuk Clustering

Pada tahap persiapan data untuk clustering, digunakan seluruh data hasil preprocessing sebanyak 2.783 komentar. Data tersebut merupakan komentar yang telah melalui tahapan cleaning, case folding, normalisasi, tokenizing, stopword removal, dan stemming. Selanjutnya, hasil pembobotan TF-IDF yang telah diperoleh pada tahap sebelumnya digunakan sebagai representasi numerik data komentar. Matriks TF-IDF tersebut kemudian dijadikan sebagai data masukan untuk proses K-Means Clustering dalam mengidentifikasi kelompok atau topik yang terdapat pada komentar publik mengenai Program Makan Bergizi Gratis (MBG).

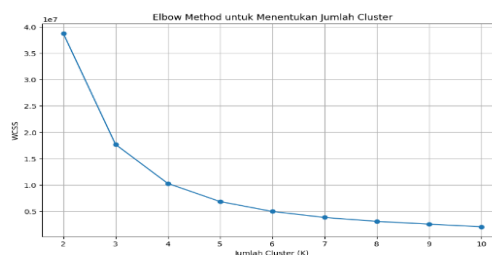
3.6.6.2 Penentuan Jumlah Cluster dengan Elbow Method

Berdasarkan hasil pengujian menggunakan Elbow Method dengan rentang K=2 sampai K=10, diperoleh nilai WCSS yang semakin menurun seiring dengan bertambahnya jumlah cluster. Penurunan WCSS terlihat sangat signifikan pada K=2, K=3, dan K=4, sedangkan setelah K=4 penurunannya mulai semakin landai. Kondisi tersebut menunjukkan bahwa penambahan jumlah cluster setelah K=4 memberikan pengurangan WCSS yang relatif lebih kecil. Dengan demikian, berdasarkan pola siku (elbow) pada grafik, jumlah cluster optimal dalam penelitian ini ditetapkan sebanyak 4 cluster (K=4). Nilai K=4 selanjutnya digunakan pada proses K-Means Clustering untuk mengelompokkan komentar publik mengenai Program Makan Bergizi Gratis (MBG) berdasarkan kemiripan karakteristik kata pada data TF-IDF. Pengujian jumlah cluster dilakukan pada rentang K=2 sampai K=10. Nilai K=1 tidak digunakan karena seluruh data akan berada dalam satu cluster sehingga tidak memberikan pengelompokan yang bermakna.

Tabel 12. Penentuan Jumlah Kluster

Jumlah Cluster K	WCSS
2	38797682.9
3	17682761.5
4	10271493.3
5	6829231.76
6	4947859.07
7	3808256.52
8	3057076.23
9	2539168.89
10	2034491.54

Nilai WCSS pada tabel berasal dari proses K-Means pada data TF-IDF penelitian yang dihitung dengan menggunakan Pemograman python.



Gambar 4. Grafik Elbew

3.6.6.3 Analisis Clustering Menggunakan K-Means

Setelah jumlah cluster optimal ditentukan menggunakan Elbow Method, tahap selanjutnya adalah menerapkan algoritma K-Means dengan jumlah K=4. Proses clustering dilakukan terhadap 2.783 komentar yang telah direpresentasikan dalam bentuk bobot TF-IDF. Algoritma K-Means mengelompokkan komentar berdasarkan tingkat kemiripan karakteristik kata sehingga setiap komentar ditempatkan ke dalam salah satu dari empat cluster. Hasil pengelompokan selanjutnya digunakan untuk mengetahui distribusi komentar dan karakteristik topik yang terdapat

pada masing-masing cluster. Berdasarkan hasil penerapan algoritma K-Means, sebanyak 2.783 komentar berhasil dikelompokkan ke dalam empat cluster. Cluster 1 merupakan cluster dengan jumlah anggota terbesar, yaitu 2.133 komentar atau 76,64%. Selanjutnya, Cluster 2 terdiri atas 263 komentar atau 9,45%, Cluster 3 terdiri atas 143 komentar atau 5,14%, dan Cluster 4 terdiri atas 244 komentar atau 8,77%. Hasil tersebut menunjukkan bahwa sebagian besar komentar dalam dataset memiliki karakteristik yang paling dekat dengan karakteristik Cluster 1.

Tabel 13. Jumlah Anggota Setiap Cluster

Cluster	Jumlah Komentar	Persentase
1	2133	76.64
2	263	9.45
3	143	5.14
4	244	8.77

Berdasarkan Tabel 13, Cluster 1 memiliki jumlah anggota paling besar, yaitu sebanyak 2.133 komentar atau 76,64% dari keseluruhan data. Selanjutnya, Cluster 2 terdiri atas 263 komentar atau 9,45%, Cluster 4 sebanyak 244 komentar atau 8,77%, sedangkan Cluster 3 merupakan cluster dengan jumlah anggota paling sedikit, yaitu 143 komentar atau 5,14%. Jumlah seluruh anggota cluster mencapai 2.783 komentar atau 100%, sehingga seluruh data penelitian telah berhasil dikelompokkan ke dalam empat cluster. Dominasi Cluster 1 menunjukkan bahwa sebagian besar komentar memiliki karakteristik yang relatif serupa dan lebih dekat dengan centroid Cluster 1. Sementara itu, jumlah anggota yang lebih kecil pada Cluster 2, Cluster 3, dan Cluster 4 menunjukkan adanya kelompok komentar dengan karakteristik yang lebih spesifik. Perbedaan jumlah anggota tersebut menjadi dasar untuk melakukan analisis lebih lanjut terhadap karakteristik masing-masing cluster berdasarkan kata-kata dominan yang dihasilkan dari nilai centroid TF-IDF.

Tabel 14. Kata Dominan

Cluster	Ranking	Kata	Nilai Centroid TF-IDF
1	1	mbg	0.04072
1	2	tidak	0.032299
1	3	rakyat	0.020416
1	4	pak	0.016328
1	5	malu	0.015046
1	6	nya	0.014349
1	7	jawab	0.013398
1	8	anak	0.013285
1	9	iya	0.012316
1	10	manfaat	0.011527
1	11	apa	0.010782
1	12	prabowo	0.009664
1	13	buat	0.009095
1	14	butuh	0.008889
1	15	gizi	0.008829
2	1	tidak	0.314831
2	2	manfaat	0.302536
2	3	mbg	0.129811
2	4	pak	0.043986
2	5	koruptor	0.035437
2	6	buat	0.03493
2	7	sangat	0.033684
2	8	jabat	0.030855
2	9	jawab	0.029976
2	10	dapur	0.027996
2	11	rakyat	0.027306
2	12	jelas	0.0221
2	13	sekali	0.019607
2	14	punya	0.018914
2	15	nya	0.017539
3	1	omon	0.67904
3	2	doang	0.036268
3	3	jangan	0.021811
3	4	tidak	0.020505
3	5	percaya	0.019668
3	6	cuma	0.019072
3	7	moga	0.018821
3	8	janji	0.015317
3	9	banyak	0.014764
3	10	si	0.011904

3	11	pak	0.011857
3	12	wo	0.011281
3	13	tok	0.011218
3	14	prabowo	0.008947
3	15	mana	0.008378
4	1	jilat	0.112877
4	2	orang	0.09961
4	3	banyak	0.096497
4	4	belakang	0.072956
4	5	tidak	0.037831
4	6	prabowo	0.026747
4	7	mbg	0.02617
4	8	terlalu	0.021836
4	9	semua	0.0201
4	10	tua	0.019281
4	11	pak	0.019211
4	12	banget	0.018687
4	13	lihat	0.018598
4	14	anak	0.017942
4	15	kerja	0.017605

Berdasarkan hasil pada Tabel 14, Cluster 1 memiliki kata dominan seperti mbg, tidak, rakyat, pak, anak, manfaat, dan gizi. Cluster 2 didominasi oleh kata tidak, manfaat, mbg, koruptor, jabat, dan dapur. Cluster 3 memiliki kata dominan seperti omon, doang, jangan, percaya, cuma, moga, dan janji, sedangkan Cluster 4 didominasi oleh kata jilat, orang, banyak, belakang, tidak, prabowo, dan mbg. Perbedaan kata dominan tersebut menunjukkan bahwa setiap cluster memiliki karakteristik pembahasan yang berbeda sehingga dapat digunakan sebagai dasar dalam mengidentifikasi dan memberikan interpretasi terhadap topik yang terbentuk. Berikut penulis berikan contoh perhitungan manualnya beberapa data. Selanjutnya jika ingin menghitung manual. Maka menyesuaikan dengan perhitungan tersebut, namun untuk seluruhnya perhitungan, penulis menggunakan python.

3.6.6.4 Identifikasi dan Interpretasi Topik Setiap Cluster

Setelah seluruh komentar dikelompokkan menggunakan algoritma K-Means, tahap selanjutnya adalah mengidentifikasi karakteristik dari setiap cluster. Identifikasi dilakukan dengan melihat kata-kata yang memiliki nilai centroid TF-IDF tertinggi pada masing-masing cluster. Kata dengan nilai centroid yang tinggi menunjukkan kata yang lebih representatif terhadap karakteristik cluster tersebut. Berdasarkan kata-kata dominan yang diperoleh, setiap cluster kemudian diberikan interpretasi topik sesuai dengan konteks pembahasan komentar mengenai Program Makan Bergizi Gratis (MBG).

Tabel 15. Identifikasi Topik Berdasarkan Kata Dominan

Cluster	Kata Dominan	Interpretasi Topik
Cluster 1	mbg, tidak, rakyat, pak, anak, manfaat, gizi	Pandangan Umum terhadap Program MBG
Cluster 2	tidak, manfaat, mbg, koruptor, jabat, rakyat, dapur	Kritik terhadap Manfaat dan Pengelolaan MBG
Cluster 3	omon, doang, jangan, percaya, cuma, moga, janji	Keraguan dan Kritik terhadap Kebijakan Pemerintah
Cluster 4	jilat, orang, banyak, belakang, tidak, prabowo, mbg	Kritik terhadap Pemerintah dan Pelaksanaan MBG

Berdasarkan Tabel 15, Cluster 1 memiliki kata dominan seperti mbg, tidak, rakyat, pak, anak, manfaat, dan gizi. Kemunculan kata-kata tersebut menunjukkan bahwa kelompok ini banyak berkaitan dengan pembahasan umum mengenai Program MBG, terutama yang berhubungan dengan masyarakat, anak-anak, manfaat program, dan pemenuhan gizi. Oleh karena itu, Cluster 1 diinterpretasikan sebagai Pandangan Umum terhadap Program MBG. Cluster 2 didominasi oleh kata tidak, manfaat, mbg, koruptor, jabat, rakyat, dan dapur. Kata-kata tersebut menunjukkan adanya pembahasan yang lebih kritis mengenai manfaat program serta aspek pengelolaan dan penyelenggaraan MBG. Kemunculan kata yang berkaitan dengan korupsi, jabatan, rakyat, dan dapur menunjukkan adanya perhatian terhadap tata kelola program. Berdasarkan karakteristik tersebut, Cluster 2 diinterpretasikan sebagai Kritik terhadap Manfaat dan Pengelolaan MBG.

Selanjutnya, Cluster 3 memiliki kata dominan seperti omon, doang, jangan, percaya, cuma, moga, dan janji. Kata-kata tersebut menunjukkan adanya ekspresi keraguan, ketidakpercayaan, serta tanggapan terhadap janji atau kebijakan pemerintah. Karakteristik tersebut menunjukkan bahwa komentar dalam kelompok ini lebih banyak merepresentasikan keraguan dan kritik terhadap kebijakan yang berkaitan dengan program pemerintah. Oleh karena itu, Cluster 3 diinterpretasikan sebagai Keraguan dan Kritik terhadap Kebijakan Pemerintah. Sementara itu, Cluster 4 ditandai oleh kata-kata seperti jilat, orang, banyak, belakang, tidak, prabowo, dan mbg. Kemunculan kata-kata tersebut menunjukkan adanya tanggapan yang bersifat kritis terhadap pemerintah dan pihak-pihak yang berkaitan

dengan pelaksanaan program MBG. Karakteristik tersebut kemudian digunakan untuk menginterpretasikan Cluster 4 sebagai Kritik terhadap Pemerintah dan Pelaksanaan MBG.

3.6.6.5 Hasil Interpretasi

Berdasarkan keseluruhan hasil identifikasi, terbentuk empat kelompok topik utama dalam komentar publik mengenai Program MBG. Cluster 1 merepresentasikan pandangan umum mengenai MBG, Cluster 2 merepresentasikan kritik terhadap manfaat dan pengelolaan program, Cluster 3 menunjukkan keraguan dan kritik terhadap kebijakan pemerintah, sedangkan Cluster 4 menunjukkan kritik terhadap pemerintah dan pelaksanaan MBG. Hasil ini menunjukkan bahwa penerapan K-Means tidak hanya mengelompokkan komentar berdasarkan kemiripan teks, tetapi juga dapat membantu mengidentifikasi pola pembahasan yang muncul dalam komentar publik.

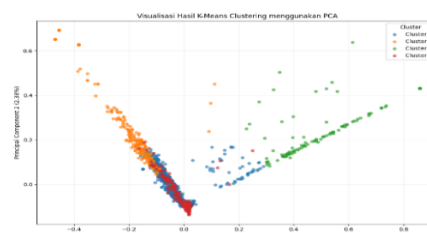
Tabel 16. Hasil Penelompokan Topik

No	Comment Asli	Cluster
1	yang di samping Prabowo kebanyakan makan telur 😊😊😊😊😊😊😊😊	4
2	MBG sdh di bilang tidak bermanfaat msih ja dipaksakan	2
3	Hentikan mbg	1
4	KALAU PEMIMPIN SEORANG PENGUSAHA YG DI UTAMAKAN BURUH APA PENGUSAHA ?? MUDAH2AN TIDAK OMONG.MBG TIDAK TEPAT SASARAN YG MEMBUTUHKAN ITU YG KURANG GIZI,SEKOLAHAN ELITE KO DAPAT BISA2 NURUNIN GIZINYA 😊	1
5	5:15 MBG bermanfaat tidak? tidak? bermanfaat dong. bagi pegang bendera. pemasukan dana di kantong	2
....		
2.783	Mubajir	1

3.6.6.6 Visualisasi

Untuk memberikan gambaran visual mengenai hasil pengelompokan komentar, dilakukan visualisasi terhadap hasil clustering menggunakan metode Principal Component Analysis (PCA). PCA digunakan untuk mereduksi dimensi matriks TF-IDF menjadi dua komponen utama tanpa mengubah label cluster yang telah dihasilkan oleh algoritma K-Means. Dengan demikian, setiap komentar dapat direpresentasikan dalam bidang dua dimensi berdasarkan posisi masing-masing terhadap komponen utama.

Visualisasi ini digunakan untuk melihat pola persebaran komentar pada Cluster 1, Cluster 2, Cluster 3, dan Cluster 4, serta mengetahui apakah kelompok-kelompok tersebut menunjukkan pola persebaran yang berbeda. Hasil visualisasi tidak digunakan untuk menentukan jumlah cluster karena jumlah cluster telah ditetapkan sebelumnya berdasarkan hasil Elbow Method, tetapi digunakan untuk memberikan gambaran visual terhadap hasil pengelompokan.



Gambar 5. Visualisasi Hasil K-Means Menggunakan PCA

Berdasarkan Gambar 5, persebaran titik menunjukkan bahwa keempat cluster memiliki pola distribusi yang berbeda, meskipun pada beberapa bagian terdapat titik yang berdekatan atau saling tumpang tindih. Cluster 2 terlihat lebih banyak tersebar pada sisi kiri grafik, sedangkan Cluster 3 cenderung berada pada sisi kanan. Cluster 1 dan Cluster 4 memiliki beberapa titik yang berdekatan pada area tengah hingga bagian bawah grafik. Kondisi tersebut menunjukkan bahwa terdapat perbedaan karakteristik antar kelompok komentar, tetapi juga terdapat sebagian komentar yang memiliki karakteristik teks yang relatif mirip. Visualisasi PCA ini digunakan sebagai pendukung untuk memberikan gambaran secara visual terhadap hasil pengelompokan yang telah diperoleh melalui algoritma K-Means.

3.7 Pembahasan

3.7.1 Pembahasan Hasil Analisis Sentimen Menggunakan SVM

Berdasarkan hasil pengujian menggunakan algoritma Support Vector Machine dengan kernel Polynomial, diperoleh data pengujian sebanyak 557 komentar. Hasil evaluasi menunjukkan nilai accuracy sebesar 81,51%, precision sebesar 81,48%, recall sebesar 81,51%, dan F1-score sebesar 81,29%. Hasil tersebut menunjukkan bahwa model SVM mampu melakukan klasifikasi sentimen dengan tingkat kinerja yang cukup baik dalam membedakan komentar

ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Berdasarkan confusion matrix, dari 296 komentar negatif, sebanyak 265 komentar berhasil diklasifikasikan dengan benar sebagai negatif, sedangkan 17 komentar diklasifikasikan sebagai netral dan 14 komentar sebagai positif. Pada kelas netral, dari 133 komentar, sebanyak 102 komentar berhasil diklasifikasikan dengan benar sebagai netral, sedangkan 26 komentar diklasifikasikan sebagai negatif dan 5 komentar sebagai positif. Sementara itu, dari 128 komentar positif, sebanyak 87 komentar berhasil diklasifikasikan dengan benar sebagai positif, sedangkan 24 komentar diklasifikasikan sebagai negatif dan 17 komentar sebagai netral.

Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang relatif baik dalam mengenali ketiga kelas sentimen. Kelas negatif memperoleh precision 84,13%, recall 89,53%, dan F1-score 86,74%, sehingga menjadi kelas dengan performa klasifikasi terbaik. Sementara itu, kelas positif memperoleh precision 82,08%, recall 67,97%, dan F1-score 74,36%. Kelas netral memperoleh precision 75,00%, recall 76,69%, dan F1-score 75,84%. Perbedaan nilai tersebut menunjukkan bahwa model lebih mudah mengenali komentar negatif dibandingkan komentar positif dan netral. Secara keseluruhan, nilai accuracy sebesar 81,51% menunjukkan bahwa sebagian besar data pengujian dapat diklasifikasikan secara tepat. Hasil ini menjadi dasar bahwa SVM dapat digunakan untuk mengidentifikasi kecenderungan sentimen publik terhadap Program MBG.

3.7.2 Pembahasan Hasil Clustering Menggunakan K-Means

Hasil klasifikasi sentimen kemudian dilengkapi dengan analisis clustering menggunakan algoritma K-Means. Berdasarkan Elbow Method, jumlah cluster yang digunakan dalam penelitian ditetapkan sebanyak K=4. Penerapan K-Means terhadap 2.783 komentar menghasilkan empat kelompok dengan distribusi yang berbeda. Cluster 1 menjadi kelompok terbesar dengan 2.133 komentar atau 76,64%, diikuti Cluster 2 sebanyak 263 komentar atau 9,45%, Cluster 4 sebanyak 244 komentar atau 8,77%, dan Cluster 3 sebanyak 143 komentar atau 5,14%. Dominasi Cluster 1 menunjukkan bahwa sebagian besar komentar memiliki karakteristik teks yang relatif serupa dan berada pada kelompok yang sama berdasarkan representasi TF-IDF.

Berdasarkan kata-kata dominan, Cluster 1 banyak memuat kata seperti mbg, tidak, rakyat, pak, anak, manfaat, dan gizi. Karakteristik tersebut menunjukkan bahwa kelompok ini berkaitan dengan pandangan umum terhadap Program MBG, terutama mengenai program, masyarakat, anak, manfaat, dan gizi. Cluster 2 memiliki kata dominan seperti tidak, manfaat, mbg, koruptor, jabat, rakyat, dan dapur. Kata-kata tersebut menunjukkan pembahasan yang lebih kritis mengenai manfaat dan pengelolaan Program MBG. Cluster 3 memiliki kata seperti omon, doang, jangan, percaya, cuma, moga, dan janji, yang menunjukkan adanya keraguan dan kritik terhadap kebijakan pemerintah. Sementara itu, Cluster 4 memiliki kata seperti jilat, orang, banyak, belakang, tidak, Prabowo, dan mbg, yang menunjukkan karakteristik kritik terhadap pemerintah dan pelaksanaan MBG. Dengan demikian, K-Means memberikan informasi tambahan yang tidak diperoleh hanya melalui klasifikasi sentimen. SVM menjawab bagaimana kecenderungan sentimen komentar, sedangkan K-Means membantu menjawab mengenai topik apa komentar tersebut disampaikan.

3.7.3 Pembahasan Integrasi Hasil SVM dan K-Means

Integrasi hasil SVM dan K-Means memberikan gambaran yang lebih komprehensif mengenai respons publik terhadap Program MBG. SVM digunakan untuk mengelompokkan komentar berdasarkan polaritas sentimen, sedangkan K-Means digunakan untuk menemukan kelompok pembahasan berdasarkan kemiripan karakteristik kata. Dengan pendekatan tersebut, penelitian tidak hanya mengetahui apakah komentar cenderung positif, negatif, atau netral, tetapi juga dapat mengetahui topik yang menjadi perhatian publik. Hasil clustering menunjukkan bahwa pembahasan publik mencakup beberapa aspek, mulai dari pandangan umum mengenai manfaat MBG dan pemenuhan gizi, kritik terhadap pengelolaan program, keraguan terhadap kebijakan pemerintah, hingga kritik terhadap pelaksanaan program dan pemerintah. Sementara itu, hasil SVM menunjukkan bahwa model mampu mengklasifikasikan komentar ke dalam tiga kategori sentimen dengan tingkat akurasi sebesar 81,51%. Kombinasi kedua metode tersebut memberikan nilai tambah dalam analisis karena klasifikasi sentimen dan identifikasi topik saling melengkapi. Klasifikasi SVM memberikan informasi mengenai kecenderungan sikap publik, sedangkan K-Means memberikan informasi mengenai isu atau topik yang menjadi perhatian dalam komentar. Dengan demikian, hasil penelitian dapat digunakan untuk memahami respons publik terhadap MBG secara lebih terstruktur berdasarkan sentimen dan topik pembahasannya.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis sentimen publik terhadap perkembangan topik Program Makan Bergizi Gratis (MBG) menggunakan algoritma Support Vector Machine (SVM) dan K-Means Clustering, yang meliputi tahapan pengumpulan data, preprocessing, pelabelan sentimen, pembobotan TF-IDF, klasifikasi SVM, dan clustering K-Means, diperoleh bahwa proses preprocessing dan pembobotan TF-IDF berhasil mengubah 2.783 komentar publik menjadi representasi teks dalam bentuk numerik yang dapat digunakan untuk proses klasifikasi dan clustering. Penerapan algoritma SVM berhasil mengklasifikasikan komentar ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif, dengan pengujian terhadap 557 data testing menghasilkan accuracy sebesar 81,51%,

precision sebesar 81,48%, recall sebesar 81,51%, dan F1-score sebesar 81,29%, dengan kelas negatif memperoleh precision 84,13%, recall 89,53%, dan F1-score 86,74%. Selanjutnya, penerapan K-Means Clustering berdasarkan Elbow Method menghasilkan jumlah optimal sebanyak empat cluster ($K=4$), yaitu Cluster 1 sebanyak 2.133 komentar (76,64%), Cluster 2 sebanyak 263 komentar (9,45%), Cluster 3 sebanyak 143 komentar (5,14%), dan Cluster 4 sebanyak 244 komentar (8,77%). Berdasarkan kata-kata dominan pada masing-masing cluster, Cluster 1 merepresentasikan pandangan umum terhadap Program MBG, Cluster 2 merepresentasikan kritik terhadap manfaat dan pengelolaan MBG, Cluster 3 merepresentasikan keraguan dan kritik terhadap kebijakan pemerintah, sedangkan Cluster 4 merepresentasikan kritik terhadap pemerintah dan pelaksanaan MBG. Dengan demikian, penggabungan metode SVM dan K-Means Clustering memberikan informasi yang saling melengkapi, di mana SVM digunakan untuk mengetahui kecenderungan sentimen publik, sedangkan K-Means digunakan untuk mengidentifikasi kelompok topik yang dibicarakan masyarakat, sehingga penelitian dapat memberikan gambaran mengenai kecenderungan sentimen sekaligus topik yang menjadi perhatian publik terhadap Program MBG.

REFERENCES

- [1] H. Rodhiy And Z. Sitorus, "Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Tarif Pajak Penghasilan Di Oeniy," *Bull. Inf. Technol.*, Vol. 4, No. 2, Pp. 198–204, 2023.
- [2] Z. Sitorus, E. Hariyanto, And F. Kurniawan, "Analysis Of Artificial Intelligence Machine Learning Technology For Mapping And Predicting Flood Locations In Pahlawan Batu Bara Village," *Int. J. Comput. Sci. Math. Eng.*, Vol. 2, No. 2, Pp. 281–288, 2023.
- [3] S. Sharifmoghammad *Et Al.*, "Rankllm: A Python Package For Reranking With Llm," In *Proceedings Of The 48th International Acm Sigir Conference On Research And Development In Information Retrieval*, 2025, Pp. 3681–3690.
- [4] M. Mukhlisin, B. Berililana, And R. E. Saputro, "Literature Review: Comparison Of Machine Learning Algorithms For Sentiment Analysis Of Free Nutritious Meals," *Sist. J. Sist. Inf.*, Vol. 15, No. 3, Pp. 1110–1120, 2026.
- [5] Z. Sitorus And M. Indra, "Analysis Of Room Allocation Based On Age In Nursing Homes Using The C4 . 5 Decision Tree Method," *J. Inf. Technol. Comput. Sci. Electr. Eng.*, Vol. 1, No. 2, Pp. 153–157, 2024, Doi: 10.61306/Jitcse.V1i2.
- [6] Z. S. Iswadi Hamzah, "Analisa Classification Decision Tree C45 Dan Naïve Bayes Pada Indikasi Penyakit Diabetes Menggunakan Rapid Miner," *J. Nas. Teknol. Komput.*, Vol. 4, No. 1, Pp. 25–33, 2024.
- [7] Z. Sitorus, I. Syahputra, C. I. Angkat, And D. Sartika, "Implementation Of K-Means Clustering For Inventory Projection," Pp. 673–678.
- [8] M. Sunjaya, Z. Sitorus, M. Iqbal, And A. P. U. Siahaan, "Analysis Of Machine Learning Approaches To Determine Online Shopping Ratings Using Naïve Bayes And Svm," *Int. J. Comput. Sci. Math. Eng.*, Vol. 3, No. 1, Pp. 7–16, 2024.
- [9] W. Wicaksono, A. Brastama Putra, And Agussalim, "Sistem Pendukung Keputusan Kelayakan Pinjaman Kredit Berbasis Web (Di Pt. Fif Group Cabang Surabaya)," *J. Inform. Dan Sist. Inf.*, Vol. 2, No. 3, Pp. 503–510, 2021, Doi: 10.33005/Jifosi.V2i3.411.
- [10] R. Damanhuri And V. A. Husein, "Analisis Sentimen Pada Ulasan Aplikasi Access By Kai Berbahasa Indonesia Menggunakan Word-Embedding Dan Classical Machine Learning," *J. Masy. Inform.*, Vol. 15, No. 2, Pp. 97–106, 2024, Doi: 10.14710/Jmasif.15.2.62383.
- [11] A. S. Manah, "Analisis Sentimen Publik Terhadap Kasus Keracunan Makanan Mbg Berbasis Data Media Sosial Menggunakan Model Indobert," *J. Inform. Dan Tek. Elektro Terap.*, Vol. 14, No. 2, 2026.
- [12] M. L. Akmal, D. Pernadi, And J. Kusumaningrum, "Analisis Sentimen Program Makan Siang Gratis Menggunakan Multinomial Naive Bayes," *J. Ilm. Sinus*, Vol. 23, No. 1, Pp. 117–126, 2025.
- [13] Z. Sitorus, M. Iqbal, D. Nasution, And R. F. Wijaya, "Penerapan Deep Learning Dan Analisis Sentimen Terhadap Gap Kompetensi Lulusan Lembaga Pendidikan Dan Pelatihan Vokasi Terhadap Dunia Kerja Dengan Metode Long Short-Term Memory (Lstm)," Vol. 6, No. 2, Pp. 0–11, 2025, Doi: 10.47065/Bit.V5i2.2029.
- [14] Y. T. Handika, S. Defit, And G. W. Nurcahyo, "Text Mining Dalam Membandingkan Metode Naïve Bayes Dengan C. 45 Dalam Mengidentifikasi Berita Hoax Pada Media Sosial," *Rang Tek. J.*, Vol. 5, No. 1, Pp. 116–123, 2022.
- [15] A. E. Widjaja, A. Fransisko, And C. A. Haryani, "Text Mining Application With K-Means Clustering To Identify Sentiments And Popular Topics : A Case Study Of The Three Largest Online Marketplaces In Indonesia," *Inform. J. Ilmu Komput.*, Vol. 4, No. 4, Pp. 441–453, 2023.
- [16] Z. Purwanti, "Pemodelan Text Mining Untuk Analisis Sentimen Terhadap Program Makan Siang Gratis Di Media Sosial X Menggunakan Algoritma Support Vector Machine (Svm)," *J. Abdinus J. Pengabd. Nusan.*, Vol. 5, No. 3, Pp. 3065–3079, 2024.
- [17] B. Rahmatullah, S. A. Saputra, P. Budiono, And D. P. Wigandi, "Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes," *J. Inf. Technol.*, Vol. 5, No. 1, 2025.
- [18] M. E. Zainuri, "Analisis Sentimen Program Makan Bergizi Gratis (Mbg) Pada Platform X Dengan Menggunakan Metode Naïve Bayes," *J. Univ. Din.*, 2026.
- [19] N. P. A. Noviyanti And I. G. N. A. C. Putra, "Penerapan Multinomial Naïve Bayes Dan Chi-Square Pada Analisis Sentimen Makan Bergizi Gratis," *J. Nas. Teknol. Inf. Dan Apl.*, Vol. 3, No. 4, Pp. 785–794, 2025.
- [20] N. Sakina, F. Wajidi, And M. R. Rasyid, "Evaluasi Algoritma Knn Dan Naive Bayes Untuk Analisis Sentimen Kebijakan Program Makan Bergizi Gratis," *Jambura J. Informatics*, Vol. 1, No. 2, Pp. 83–97, 2025.
- [21] D. C. Nugroho, A. D. Agustina, B. Maulana, F. Darussalam, M. Asyraf, And B. O. Lubis, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Publik Pada Program Makan Bergizi Gratis," *Infotech J.*, Vol. 11, No. 2, Pp. 474–479, 2025.
- [22] D. A.-T. Mukarom, G. P. Pangestu, M. F. Margadipraja, And H. M. Winata, "Analisis Sentimen Berbasis Aspek Terhadap

- Program ‘Makan Bergizi Gratis (Mbg)’ Di Twitter/X Dengan Pendekatan Efektivitas Duncan,” *J. Pendidik. Sos. Dan Hum.*, Vol. 5, No. 1, Pp. 1890–1907, 2026.
- [23] T. Yuniarti And B. Andriza, “Analisis Sentimen Dan Opini Publik: Evaluasi Komunikasi Media Sosial Instagram Tim Kepresidenan@ Pco. Ri Dalam Menangani Kontroversi Program Makan Bergizi Gratis (Mbg),” *J. Ilmu Komun. Uho J. Penelit. Kaji. Ilmu Komun. Dan Inf.*, Vol. 11, No. 1, Pp. 200–215, 2026.
- [24] R. A. Munir, A. Voutama, U. S. Karawang, T. Timur, And J. Barat, “Analisis Sentimen Cuitan Di Media Sosial X Tentang Program Makan Bergizi Gratis Dengan Metode Nlp,” *J. Inform. Dan Elektro Terap.*, Vol. 13, No. 3, Pp. 589–595, 2025.
- [25] K. Y. Pratama, S. Achmadi, And K. A. Sari, “Analisis Sentimen Terhadap Makan Bergizi Gratis Pada Sosial Media X Menggunakan Metode Decicion Tree Dengan Pembanding Naive Bayes,” *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 9, No. 6, Pp. 11061–11067, 2025.
- [26] S. A. S. Mola, Y. C. Luttu, And D. N. Rumlaklak, “Perbandingan Metode Machine Learning Dalam Analisis Sentimen Komentar Pengguna Aplikasi Indriver Pada Dataset Tidak Seimbang,” *J. Sist. Inf. Bisnis*, Vol. 14, No. 3, Pp. 247–255, 2024, Doi: 10.21456/Vol14iss3pp247-255.
- [27] C. A. Bahri And L. H. Suadaa, “Aspect-Based Sentiment Analysis In Bromo Tengger Semeru National Park Indonesia Based On Google Maps User Reviews,” *Ijccs (Indonesian J. Comput. Cybern. Syst.*, Vol. 17, No. 1, P. 79, 2023, Doi: 10.22146/Ijccs.77354.
- [28] A. Pambudi And S. Suprpto, “Effect Of Sentence Length In Sentiment Analysis Using Support Vector Machine And Convolutional Neural Network Method,” *Ijccs (Indonesian J. Comput. Cybern. Syst.*, Vol. 15, No. 1, P. 21, 2021, Doi: 10.22146/Ijccs.61627.
- [29] A. Putera, U. Siahaan, N. A. Harahap, And R. Y. Simanullang, “Analysis Of Inpatient Data Using Cluster Analysis On Simulation Dataset,” *J. Bit (Bulletin Inf. Technol.*, Vol. 6, No. 1, Pp. 33–39, 2025, Doi: 10.47065/Bit.V5i2.1830.
- [30] P. K. Berindikasi, M. A. Syawal, And B. A. Pradhana, “Implementation Of The K-Means Method For Grouping Words Indicative Of Cyberbullying In,” *Inform. J. Tek. Inform. Dan Multimed.*, Vol. 9, No. 2, Pp. 330–347, 2025, Doi: 10.52362/Jisicom.V9i2.2163.
- [31] M. Herviany, S. Putri Delima, T. Nurhidayah, And Kasini, “Comparison Of K-Means And K-Medoids Algorithms For Grouping Landslide Prone Areas In West Java Province,” *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 1, No. 1, Pp. 34–40, 2021.
- [32] D. M. Putra And F. F. Abdulloh, “Comparison Of Clustering Algorithms : Fuzzy C-Means , K-Means , And DbSCAN For House Classification Based On Specifications And Price,” *J. Appl. Informatics Comput.*, Vol. 8, No. 2, Pp. 509–515, 2024.
- [33] T. Arya Bimantoro, Y. Rahmawati, Y. Findawati, And M. A. Rosid, “Sentiment Analysis Of Youtube Comments On The East Java Grant Case Using The Support Vector Machine (Svm) Algorithm. [Analisis Sentimen Komentar Youtube Terhadap Kasus Dana Hibah Jawa Timur Menggunakan Algoritma Support Vector Machine (Svm)],” *J. Abdimas*, Vol. 1, No. 1, Pp. 1–13, 2022.
- [34] S. Howay And S. Suhirman, “Comparison Of Svm, Nbc, And Knn Classification Methods In Determining Students’ Majors At Smk N02 Manokwari,” *J. Comput. Sci. Technol. Stud.*, Vol. 5, No. 1, Pp. 15–23, 2023, Doi: 10.32996/Jcsts.2023.5.1.3.