

**ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP PROGRAM  
MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Oleh :  
**MOH. MUSA AL KADZIM**  
**NIM. 220605110136**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP PROGRAM  
MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X MENGGUNAKAN  
METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:  
**MOH. MUSA AL KADZIM**  
NIM. 220605110136

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

## HALAMAN PERSETUJUAN

### ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP PROGRAM MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

#### SKRIPSI

Oleh :

**MOH. MUSA AL KADZIM**  
**NIM. 220605110136**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 9 Desember 2025

Pembimbing I,



**Prof. Dr. Muhammad Faisal, M.T**  
**NIP. 19740510 200501 1 007**

Pembimbing II,



**Nur Fitriyah Ayu Tunjung Sari, M.Cs**  
**NIP. 19911226 202012 2 001**

Mengetahui

Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang





## HALAMAN PENGESAHAN

### ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP PROGRAM MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

#### SKRIPSI

Oleh :

**MOH MUSA AL KAZIM**  
**NIM. 220605110136**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 17 Desember 2025

#### Susunan Dewan Penguji

Ketua Penguji	: <u>Shoffin Nahwa Utama, M.T</u> NIP. 19860703 202012 1 003
Anggota Penguji I	: <u>Ashri Shabrina Afrah, M.T</u> NIP. 199004302020122 003
Anggota Penguji II	: <u>Prof. Dr. Muhammad Faisal, M.T</u> NIP. 19740510 200501 1 007
Anggota Penguji III	: <u>Nur Fitriyah Ayu Tunjung Sari, M.Cs</u> NIP. 19911226 202012 2 001

(*Shoffin*)  
(*Ashri*)  
(*Prof. Dr. Muhammad Faisal*)  
(*Nur Fitriyah Ayu Tunjung Sari*)

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang





## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Moh. Musa Al Kadzim  
NIM : 220605110136  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Analisis Sentimen Berbasis Aspek Terhadap  
Program Makan Bergizi Gratis Di Media Sosial X  
Menggunakan Metode *Support Vector Machine*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 30 Desember 2025  
Yang membuat pernyataan,



Moh. Musa Al Kadzim  
NIM.220605110136

**MOTTO**

*“Leave Past, Live Tomorrow”*

*“Jangan terlalu lama berandai-andai king, Segera bangkit”*

## **HALAMAN PERSEMBAHAN**

Puji syukur kehadiran ALLAH SWT yang telah memberikan rahmat dan  
Hidayah-Nya sehingga penulis diberi kemudahan dalam  
Menyelesaikan penulisan skripsi ini dengan baik.  
Saya persembahkan karya ini kepada:

Orang tua Tersayang, Muhibburridho dan Lutfiyah  
meski jarak memisahkan, selalu hadir dalam doa, dukungan, dan cinta yang tak  
pernah surut. Tanpa kalian, aku tak akan sampai di titik ini.

Teman tercintaku di Malang  
Yang selalu menemani langkah-langkah kecilku, menguatkan saat lelah, dan  
tertawa bersama saat lelah menjadikan semuanya menjadi terasa ringan..

Segenap Keluarga besar dan orang sekitarku  
Yang selalu hadir dengan dukungan, semangat, dan pelajaran hidup berharga,  
membuat perjalanan ini terasa lebih bermakna.

Dan untuk diriku sendiri  
Yang berjuang melewati malam-malam panjang, menahan lelah dan rasa ragu,  
namun tetap berdiri dan menulis tiap kata dengan hati. Terima kasih karena tidak  
menyerah dan tetap percaya, meski perjalanan ini terasa berat.

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

*Alhamdulillah*, segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas rahmat, karunia, dan hidayah-Nya, sehingga skripsi yang berjudul “Analisis Sentimen Berbasis Aspek Terhadap Program Makan Bergizi Gratis Di Media Sosial X Menggunakan *Support Vector Machine*” dapat diselesaikan. Sholawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad Shallallahu Alaihi Wasallam, teladan dalam menuntut ilmu, berikhtiar, dan menjaga keikhlasan dalam setiap proses.

Dalam penyusunan skripsi ini, penulis menyadari bahwa banyak pihak telah memberikan doa, bantuan, bimbingan, dan dukungan. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Muhammad Faisal, M.T, selaku dosen pembimbing utama, yang dengan kebijaksanaan dan ketelatenan telah menjadi penunjuk arah di saat penulis kehilangan kompas. Bimbingan beliau bukan hanya memperbaiki



tulisan, tetapi juga melatih cara berpikir, bersabar, dan bertanggung jawab hingga skripsi ini menemukan bentuk terbaiknya.

5. Nur Fitriyah Ayu Tunjung Sari, M.Cs, selaku dosen pembimbing kedua, yang senantiasa hadir dengan ketelitian, ketegasan, dan kepedulian. Setiap masukan yang diberikan menjadi cahaya kecil yang menuntun penulis melewati proses yang panjang dan penuh pembelajaran.
6. Shoffin Nahwa Utama, M.T, selaku ketua penguji, yang telah meluangkan waktu dan pemikiran untuk menguji, mengarahkan, serta memberikan kritik yang tajam namun membangun. Arahan beliau menjadi bagian penting dalam menyempurnakan skripsi ini.
7. Ashri Shabrina Afrah, M.T, selaku dosen penguji kedua, terima kasih atas masukan, evaluasi, dan sudut pandang yang memperkaya karya ini, sehingga skripsi ini dapat berdiri dengan lebih matang dan bertanggung jawab.
8. Seluruh Dosen , Laboran, dan Staf Program Studi Teknik Informatika, yang telah memberikan ilmu, pelayanan, serta bantuan administratif selama penulis menempuh perkuliahan.
9. Kak Nia, terima kasih atas kebaikan yang tidak pernah terhitung. Sejak masa awal menjadi mahasiswa baru hingga akhirnya penulis menyelesaikan studi, kehadiran dan bantuanmu menjadi penopang yang sangat berarti. Tanpa arahan, bantuan, dan kepedulianmu, perjalanan kuliah penulis tentu tidak akan semulus ini. Terima kasih telah menjadi sosok yang selalu ada, bahkan ketika penulis belum tahu harus meminta bantuan kepada siapa.

10. Ibu Munjiyati, selaku mama dari Kak Nia, terima kasih atas ketulusan hati, kesabaran, dan kebaikan yang begitu besar. Penulis menyadari telah banyak merepotkan, namun Ibu selalu menyambut dengan senyum, perhatian, dan kasih yang tulus. Kebaikan Ibu bukan hanya membantu penulis secara nyata, tetapi juga menghangatkan hati penulis di saat-saat lelah. Semoga Allah Subhanahu wa Ta'ala membalas segala kebaikan Ibu dengan kesehatan, keberkahan, dan kebahagiaan yang berlipat.
11. Firna, terima kasih telah menjadi teman seperjalanan yang tidak hanya hadir di hari-hari ringan, tetapi juga tetap bertahan ketika langkah penulis mulai goyah. Dukunganmu sering kali datang tanpa banyak kata, namun selalu tepat pada waktunya.
12. Radifan, terima kasih atas bantuan dan semangat yang terus mengalir dengan cara yang sederhana namun tulus. Kehadiranmu menjadi salah satu penopang di tengah proses panjang yang tidak selalu mudah.
13. Dita, terima kasih atas perhatian dan kebersamaan yang menguatkan. Di tengah kesibukan masing-masing, kepedulianmu menjadi pengingat bahwa penulis tidak berjalan sendirian.
14. Bela, terima kasih atas doa, semangat, dan energi positif yang senantiasa kamu bagikan. Hal-hal kecil darimu sering kali memberi pengaruh besar bagi penulis.
15. Silvi, terima kasih atas kebaikan dan perhatian yang mengalir tanpa pamrih. Dukunganmu membantu penulis menjaga semangat hingga akhir perjalanan ini.

16. Rofiq, terima kasih atas pelajaran-pelajaran berharga yang kamu berikan, baik melalui kata maupun sikap. Dari proses yang dilalui bersama, penulis belajar tentang kedewasaan, sudut pandang baru, dan cara memahami hidup dengan lebih bijak. Setiap pertemuan selalu meninggalkan makna yang patut dikenang.
17. Teman-teman KKN penulis: Idun, Somad, Jidan, Najel, Melany, serta teman-teman lainnya, terima kasih atas kebersamaan yang singkat namun penuh makna. Di tengah pengabdian, keterbatasan, dan dinamika yang ada, kalian menghadirkan kerja sama, tawa, dan pelajaran hidup yang tidak tertulis di buku mana pun. Kebersamaan itu akan selalu menjadi bagian indah dari perjalanan penulis.
18. Apip, terima kasih atas kepercayaan, bantuan, dan dukungan yang nyata. Peranmu bukan hanya membantu penulis bertahan, tetapi juga tumbuh hingga mampu membangun dan mengembangkan bisnis yang kini menjadi bagian penting dari perjalanan hidup penulis.
19. Iqbal, Romeo, Diki, dan Bagus, terima kasih atas tawa yang kalian hadirkan, atas canda yang menyela penat, dan atas kebersamaan yang membuat perjalanan ini terasa lebih manusiawi. Di antara tekanan dan target, kalian adalah jeda yang menenangkan.
20. Ayah tercinta, Muhibburridho, terima kasih atas cinta yang mungkin tidak selalu terucap, tetapi selalu terasa. Nilai-nilai yang Ayah tanamkan menjadi akar yang menahan penulis agar tetap tegak, sekaligus sayap yang mendorong penulis berani melangkah. Semoga Allah Subhanahu wa Ta'ala

senantiasa melimpahkan kesehatan, keberkahan, dan ridha-Nya kepada Ayah.

21. Ibuku tercinta, terima kasih atas doa-doa yang diam-diam kau titipkan di setiap sujud, atas kesabaran yang tidak pernah meminta balasan, dan atas cinta yang selalu menjadi tempat pulang paling aman. Setiap keberhasilan penulis adalah pantulan dari ketulusanmu.
22. Kakakku yang penulis banggakan, terima kasih atas dukungan tanpa syarat dan kepercayaan yang selalu diberikan. Restumu dalam setiap keputusan penulis adalah kekuatan yang membuat langkah ini terasa lebih mantap dan bermakna.
23. Para adikku tersayang, terima kasih atas kesediaanmu membantu, atas tanggung jawab yang kamu pikul, dan atas ketulusan dalam mengurus bisnis penulis di rumah hingga dapat berjalan dengan baik. Dedikasimu adalah bukti kasih yang nyata.
24. Seluruh orang yang membantu penulis namun tidak dapat penulis sebutkan satu per satu, terima kasih atas doa, perhatian, dan dukungan yang mengalir tanpa henti. Keluarga adalah alasan terbesar penulis untuk terus berjuang dan tidak menyerah.
25. Aisyah Nur Fitriyah, yang hadir bukan sekadar sebagai pendamping, tetapi sebagai rumah di tengah perjalanan panjang ini. Dari semester pertama hingga titik akhir ini, kamu memilih untuk tetap tinggal, berjalan beriringan, dan menggenggam tangan penulis bahkan ketika arah terasa kabur. Dalam diam, kamu menguatkan; dalam sederhana, kamu menghidupkan harapan.



Jika perjalanan ini adalah malam yang panjang, maka kamulah cahaya yang tidak pernah padam, menerangi tanpa meminta apa pun selain keikhlasan.

26. Dan untuk diri penulis sendiri, terima kasih karena telah bertahan. Terima kasih karena memilih bangkit setiap kali ingin berhenti, karena tetap melangkah meski ragu, dan karena menyelesaikan apa yang pernah dimulai. Semoga setelah ini, penulis terus menjadi pribadi yang berani bermimpi, sabar berproses, dan rendah hati dalam setiap pencapaian.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis menerima kritik dan saran yang membangun agar skripsi ini dapat menjadi lebih baik. Penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca, serta menjadi kontribusi kecil bagi perkembangan ilmu pengetahuan.

*Wassalamu'alaikum warahmatullahi wabarakatuh.*

Malang, 30 Desember 2025

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR TABEL .....	xvii
ABSTRAK .....	xviii
ABSTRACT .....	xix
مستخلص البحث.....	xx
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah .....	6
1.4 Tujuan Penelitian .....	6
1.5 Manfaat Penelitian .....	6
<b>BAB II STUDI PUSTAKA.....</b>	<b>8</b>
2.1 Makan Bergizi Gratis .....	8
2.2 Analisis Sentimen Berbasis Aspek.....	9
2.3 <i>Preprocessing Data</i> .....	16
2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	17
2.5 <i>Support Vector Machine</i> .....	20
<b>BAB III METODE PENELITIAN .....</b>	<b>28</b>
3.1 Desain Sistem.....	28
3.2 Pengumpulan Data .....	29
3.3 <i>Preprocessing Data</i> .....	33
3.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	39
3.5 <i>Support Vector Machine</i> .....	44
3.5.1 Proses Pelatihan .....	46
3.5.2 Proses Pengujian .....	50
3.6 Evaluasi Pengujian .....	52
3.7 Skenario Pengujian .....	53
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>57</b>
4.1 Tahapan Penelitian .....	57
4.1.1 Pengumpulan Data .....	57
4.1.2 <i>Preprocessing</i> .....	58
4.1.2.1. <i>Cleansing</i> .....	58
4.1.2.2. <i>Case Folding</i> .....	59
4.1.2.3. <i>Tokenizing</i> .....	61
4.1.2.4. <i>Stopword Removal</i> .....	62
4.1.2.5. <i>Stemming</i> .....	63
4.1.2.6. Finalisasi Data .....	64
4.1.3 Ekstraksi Fitur TF-IDF .....	65
4.2 Hasil Uji Coba.....	67

4.2.1 Skenario 1 .....	67
4.2.2 Skenario 2 .....	69
4.2.3 Skenario 3 .....	72
4.2.4 Skenario 4 .....	75
4.2.5 Skenario 5 .....	77
4.2.6 Skenario 6 .....	80
4.2.7 Skenario 7 .....	83
4.2.8 Skenario 8 .....	85
4.2.9 Skenario 9 .....	88
4.3 Pembahasan .....	90
4.3.1 Perbandingan Kinerja Model Antar Skenario .....	90
4.3.2 Analisis Model Terbaik .....	92
4.3.3 Analisis Korelasi Sentimen .....	94
4.3.4 Analisis Koreasi Sentimen Mendalam .....	95
4.4 Integrasi Islam .....	99
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>103</b>
5.1 Kesimpulan .....	103
5.2 Saran .....	104
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 3. 1 Desain Sistem.....	28
Gambar 4. 1 Data Mentah Setelah Dilabel Manual .....	58
Gambar 4. 2 Hasil Proses Cleansing.....	59
Gambar 4. 3 Hasil Proses <i>Case Folding</i> .....	60
Gambar 4. 4 Hasil Proses Tokenizing.....	61
Gambar 4. 5 Hasil Proses Stopword Removal.....	62
Gambar 4. 6 Hasil Proses Stemming .....	63
Gambar 4. 7 Hasil Akhir Proses Preprocessing .....	65
Gambar 4. 8 Hasil Ekstraksi Fitur TF-IDF .....	66
Gambar 4. 9 <i>Confusion Matrix</i> Skenario 1.....	67
Gambar 4. 10 Metrik Evaluasi Skenario 1 .....	68
Gambar 4. 11 Visualisasi <i>Hyperparameter</i> Skenario 1.....	69
Gambar 4. 12 <i>Confusion Matrix</i> Skenario 2.....	70
Gambar 4. 13 Metrik Evaluasi Skenario 2 .....	71
Gambar 4. 14 Visualisasi <i>Hyperparameter</i> Skenario 2.....	72
Gambar 4. 15 <i>Confusion Matrix</i> Skenario 3.....	73
Gambar 4. 16 Metrik Evaluasi Skenario 3 .....	74
Gambar 4. 17 Visualisasi <i>Hyperparameter</i> Skenario 3.....	74
Gambar 4. 18 <i>Confusion Matrix</i> Skenario 4.....	75
Gambar 4. 19 Metrik Evaluasi Skenario 4 .....	76
Gambar 4. 20 Visualisasi <i>Hyperparameter</i> Skenario 4.....	77
Gambar 4. 21 <i>Confusion Matrix</i> Skenario 5.....	78
Gambar 4. 22 Metrik Evaluasi Skenario 5 .....	79
Gambar 4. 23 Visualisasi <i>Hyperparameter</i> Skenario 5.....	80
Gambar 4. 24 <i>Confusion Matrix</i> Skenario 6.....	81
Gambar 4. 25 Metrik Evaluasi Skenario 6 .....	82
Gambar 4. 26 Visualisasi <i>Hyperparameter</i> Skenario 6.....	82
Gambar 4. 27 <i>Confusion Matrix</i> Skenario 7.....	83
Gambar 4. 28 Metrik Evaluasi Skenario 7 .....	84
Gambar 4. 29 Visualisasi <i>Hyperparameter</i> Skenario 7.....	85
Gambar 4. 30 <i>Confusion Matrix</i> Skenario 8.....	86
Gambar 4. 31 Metrik Evaluasi Skenario 8 .....	86
Gambar 4. 32 Visualisasi <i>Hyperparameter</i> Skenario 8.....	87
Gambar 4. 33 <i>Confusion Matrix</i> Skenario 9.....	88
Gambar 4. 34 Metrik Evaluasi Skenario 9 .....	89
Gambar 4. 35 Visualisasi <i>Hyperparameter</i> Skenario 9.....	89
Gambar 4. 36 <i>WordCloud</i> Sentimen Positif.....	94
Gambar 4. 37 <i>WordCloud</i> Sentimen Positif.....	95



## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	11
Tabel 3. 1 Penyajian Contoh Data Utama .....	30
Tabel 3. 2 Atribut Data.....	31
Tabel 3. 3 Deskripsi Aspek Penilaian.....	31
Tabel 3. 4 Pelabelan Tweet Secara Manual.....	32
Tabel 3. 5 Proses <i>Cleansing</i> .....	34
Tabel 3. 6 Proses <i>Case Folding</i> .....	34
Tabel 3. 7 Proses <i>Tokenizing</i> .....	35
Tabel 3. 8 Proses <i>Stopword Removal</i> .....	36
Tabel 3. 9 Proses <i>Stemming</i> .....	37
Tabel 3. 10 Hasil Akhir .....	38
Tabel 3. 11 Data Sampel .....	39
Tabel 3. 12 Hasil TF.....	41
Tabel 3. 13 Hasil IDF.....	42
Tabel 3. 14 Hasil Perhitungan Pembobotan TF-IDF .....	43
Tabel 3. 15 Contoh Data <i>Training</i> .....	46
Tabel 3. 16 Contoh Data Testing.....	50
Tabel 3. 17 Hasil Perhitungan Prediksi Data Testing.....	51
Tabel 3. 18 Hasil Contoh Klasifikasi.....	52
Tabel 3. 19 Contoh <i>Confusion Matrix</i> 2 kelas.....	52
Tabel 3. 20 Skenario Rasio Pembagian Data .....	54
Tabel 3. 21 Jenis Kernel SVM yang Diuji.....	54
Tabel 3. 22 Nilai Hyperparameter yang Diuji .....	55
Tabel 3. 23 Skenario Pengujian Gabungan .....	55
Tabel 4. 1 Rangkuman Hasil Pengujian Sembilan Skenario SVM .....	90
Tabel 4. 2 Hasil Model Terbaik .....	93
Tabel 4. 3 <i>Ground truth</i> Vs Klasifikasi.....	96

## ABSTRAK

Al Kadzim, Moh. Musa. 2025. **Analisis Sentimen Berbasis Aspek Terhadap Program Makan Bergizi Gratis di Media Sosial X Menggunakan Support Vector Machine**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Muhammad Faisal, M.T, (II) Nur Fitriyah Ayu Tunjung Sari, M.Cs.

Kata kunci: *Support Vector Machine, Analisis Sentimen Berbasis Aspek, Makan Bergizi Gratis, Media Sosial X.*

Perkembangan media sosial memungkinkan opini publik terhadap Program Makan Bergizi Gratis (MBG) tersebar dengan cepat. Penelitian ini bertujuan menganalisis sentimen masyarakat serta menilai *tingkat keberhasilan* Program MBG berdasarkan opini di media sosial X. Metode yang digunakan adalah *Support Vector Machine* (SVM) dengan ekstraksi fitur *Term Frequency–Inverse Document Frequency* (TF–IDF). Proses penelitian meliputi pengumpulan data, *preprocessing*, ekstraksi fitur, serta klasifikasi sentimen positif dan negatif yang mengombinasikan kernel, rasio data, dan parameter hiper. Hasil penelitian menunjukkan bahwa model terbaik diperoleh pada kernel Polynomial dengan parameter  $C=10$ ,  $\gamma=1$ , dan  $\text{degree}=2$ , menghasilkan akurasi 70%, presisi 87,50%, recall 8,97%, dan F1-score 16,28%. Berdasarkan hasil klasifikasi terhadap 599 data tweet, diperoleh 407 tweet bernada negatif dan 192 tweet positif, menunjukkan bahwa persepsi publik terhadap program cenderung negatif. Analisis *WordCloud* memperkuat hasil tersebut, di mana kata-kata seperti “mahal”, “anggaran”, “beban”, dan “tidak merata” mendominasi sentimen negatif, sedangkan kata “gizi”, “gratis”, dan “program” mencerminkan dukungan terhadap manfaat program.

## ABSTRACT

Al Kadzim, Moh. Musa Al Kadzim. 2025. **Aspect-Based Sentiment Analysis of the Free Nutritious Meal Program on X Using the Support Vector Machine (SVM)**  
**Method.** Undergraduate Thesis. Department of Informatics Engineering  
Faculty of Science and Technology Maulana Malik Ibrahim State Islamic  
University Malang. Supervisor: (I) Prof. Dr. Muhammad Faisal, M.T, (II) Nur  
Fitriyah Ayu Tunjung Sari, M.Cs.

**Keywords:** *Support Vector Machine, Aspect-Based Sentiment Analysis, Free Nutritious Meal Program, X (Twitter).*

The rapid growth of social media has enabled public opinions about the Free Nutritious Meal Program (Makan Bergizi Gratis/MBG) to spread widely and quickly. This study aims to analyze public sentiment and to assess the perceived success of the MBG program based on opinions expressed on the X social media platform. A Support Vector Machine (SVM) classifier was employed using Term Frequency–Inverse Document Frequency (TF–IDF) for feature extraction. The research workflow consisted of data collection, text preprocessing, feature extraction, and sentiment classification into positive and negative categories by experimenting with different kernels, data split ratios, and hyperparameter settings. The best-performing model used a polynomial kernel with parameters  $C = 10$ ,  $\gamma = 1$ , and degree = 2, achieving 70% accuracy, 87.50% precision, 8.97% recall, and a 16.28% F1-score. Sentiment classification of 599 tweets resulted in 407 negative tweets and 192 positive tweets, indicating that public perception of the program tends to be negative. Word cloud analysis further supported these findings: terms such as “expensive,” “budget,” “burden,” and “unequal distribution” dominated negative sentiment, while words like “nutrition,” “free,” and “program” reflected support for the program’s perceived benefits.

## مستخلص البحث

الكاظم، محمد موسى. ٢٠٢٥. تحليل المشاعر القائم على الجوانب لبرنامج الوجبات المغذية المجانية على منصة إكس باستخدام منهج آلة المتجهات الداعمة. رسالة بكالوريوس. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرفان: (١) الأستاذ الدكتور محمد فيصل، ماجستير في الهندسة؛ (٢) نور فطرية أيو تونجونغ ساري، ماجستير في علوم الحاسوب.

**الكلمات المفتاحية:** تحليل المشاعر القائم على الجوانب، برنامج الوجبات المغذية المجانية، إكس (تويتر).

أتاحت التطورات المتسارعة في وسائل التواصل الاجتماعي انتشار آراء الجمهور حول برنامج الوجبات المغذية المجانية بسرعة كبيرة. يهدف هذا البحث إلى تحليل مشاعر المجتمع وتقييم مستوى نجاح البرنامج استناداً إلى الآراء المنشورة على منصة إكس. واعتمدت الدراسة آلة المتجهات الداعمة في التصنيف، مع استخدام تردد المصطلح ومعكوس تردد المستند لاستخراج السمات. وشملت إجراءات البحث جمع البيانات، ومعالجة النصوص معالجة مسبقة، واستخراج السمات، ثم تصنيف المشاعر إلى إيجابية وسلبية عبر الجمع بين أنواع النوى المختلفة، ونسب تقسيم البيانات، ومعلومات الضبط الفائقة. وأظهرت النتائج أنّ أفضل نموذج تحقق عند استخدام نواة متعددة الحدود مع الإعدادات الآتية: معامل الانتظام ١٠، وغاما ١، والدرجة ٢؛ إذ بلغت الدقة ٧٠٪، والإحكام ٨٧,٥٠٪، والاستدعاء ٨,٩٧٪، ودرجة إف واحد ١٦,٢٨٪. وبناءً على تصنيف ٥٩٩ تغريدة، تبين وجود ٤٠٧ تغريدات ذات نبرة سلبية مقابل ١٩٢ تغريدة إيجابية، مما يدلّ على أنّ تصوّر الجمهور تجاه البرنامج يميل إلى السلبية. كما عزّز تحليل سحابة الكلمات هذه النتيجة؛ إذ غلبت مفردات مثل: «غالي»، و«الميزانية»، و«عبء»، و«غير متكافئ» في السياقات السلبية، في حين ظهرت كلمات مثل: «التغذية»، و«مجاناً»، و«البرنامج» بوصفها مؤشرات على دعم فوائد البرنامج.



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Saat ini, perkembangan era digital mendorong masyarakat untuk semakin aktif dalam menyuarakan opini dan pandangan mereka di media sosial. Salah satu isu yang ramai diperbincangkan di Indonesia adalah program Makan Bergizi Gratis (MBG) yang digagas pemerintah sebagai bentuk intervensi sosial dalam bidang pendidikan dan kesejahteraan. Program ini bertujuan untuk meningkatkan asupan gizi anak sekolah sekaligus mendorong pemerataan akses pangan di kalangan masyarakat menengah ke bawah (Sitanggang et al., 2024). Kehadiran program tersebut memunculkan beragam tanggapan publik, mulai dari dukungan yang melihatnya sebagai kebijakan strategis untuk meningkatkan kualitas sumber daya manusia, hingga kritik yang menyoroti aspek pembiayaan dan implementasi teknis di lapangan (Ilham & Priambodo, 2024). Media sosial X (sebelumnya Twitter) menjadi salah satu ruang utama bagi masyarakat untuk mengekspresikan opini mengenai isu ini, mengingat karakteristiknya yang *realtime*, cepat menyebarkan informasi, serta banyak digunakan dalam diskursus publik (Amara et al., 2024). Oleh karena itu, analisis sentimen terhadap percakapan publik di media sosial X terkait program MBG menjadi penting untuk memahami bagaimana persepsi masyarakat terbentuk serta faktor-faktor apa saja yang dominan dalam memengaruhi opini mereka.

Opini masyarakat mengenai program MBG di Media Sosial X tidak hanya mencerminkan sikap *individual*, tetapi juga mencerminkan penilaian kolektif

terhadap keberhasilan program tersebut. Alih-alih menekankan banyak sub-aspek terpisah, penelitian ini memfokuskan pada seberapa efektif program dalam mencapai tujuannya misalnya peningkatan asupan gizi, jangkauan penerima, dan pemenuhan target sasaran. Pendekatan ini memungkinkan analisis yang lebih holistik untuk menilai apakah program benar-benar berhasil dalam praktik, serta faktor-faktor utama yang memengaruhi persepsi publik. Karena volume data di media sosial sangat besar dan sering bersifat informal, analisis manual menjadi tidak efisien dan berisiko bias; oleh karena itu diperlukan metode otomatis berbasis klasifikasi teks seperti *Support Vector Machine* (SVM) yang dipadukan dengan teknik ekstraksi fitur TF-IDF agar proses penilaian keberhasilan dapat dilakukan secara sistematis dan andal (Ma'rufudin & Yudhistira, 2025).

Proses pemilihan metode dimulai dengan pertimbangan matang terhadap pendekatan yang paling sesuai untuk menangani masalah analisis sentimen berbasis aspek terhadap program MBG di media sosial X. Dari beberapa alternatif metode seperti *Naive Bayes* dan *Random Forest*, kombinasi TF-IDF dan SVM muncul sebagai solusi optimal karena menawarkan keseimbangan antara performa klasifikasi yang tinggi dan efisiensi komputasional. Contohnya, dalam studi mengenai sentimen film, model SVM+TF-IDF berhasil mencapai akurasi 85%, *precision* 100%, *recall* 70%, dan *F1-score* 82% (Gifari et al., 2022). Selain itu, pada penelitian mengenai deteksi komentar *cyberbullying* di media sosial TikTok, penerapan algoritma TF-IDF dan SVM menghasilkan performa tinggi dengan akurasi 88%, *precision* 88%, *recall* 96%, dan *F1-score* 92% (Romindo et al., 2023). Pendekatan ini terbukti lebih konsisten dan stabil ketika menghadapi data teks

informal dan berskala besar seperti komentar publik, serta memprosesnya dengan waktu komputasi yang lebih ringan..

Metode SVM dipilih sebagai algoritma utama untuk klasifikasi sentimen karena kemampuannya yang unggul dalam mengatasi data teks dengan jumlah fitur yang besar dan kompleksitas tinggi. SVM dikenal mampu membedakan pola sentimen secara akurat meskipun data yang digunakan berasal dari teks informal seperti komentar pengguna media sosial (Arisandi et al., 2023). Keunggulan lain dari SVM adalah kemampuannya dalam memberikan generalisasi yang baik pada data uji, sehingga model tetap konsisten saat diterapkan pada data baru. Untuk proses representasi teks, digunakan teknik *Term Frequency- Inverse Document Frequency* (TF-IDF) yang berfungsi memberikan bobot lebih besar pada kata-kata penting sekaligus mengurangi pengaruh kata umum yang tidak signifikan. Pendekatan ini membantu menjaga konteks dari ulasan atau opini yang dianalisis, sehingga meningkatkan efektivitas klasifikasi sentiment (K. T. Putra et al., 2023). Kombinasi SVM dan TF-IDF dinilai tepat untuk menangani analisis sentimen berbasis aspek pada isu program MBG, karena mampu menghadirkan keseimbangan antara akurasi, efisiensi komputasi, dan relevansi hasil analisis

Dalam perspektif Islam, umat Muslim diajarkan untuk senantiasa memiliki kesadaran terhadap setiap tindakan yang dilakukan, termasuk dalam menyampaikan pendapat maupun berinteraksi di ruang publik seperti media sosial. Allah SWT berfirman dalam Surah Al-Hashr (59:18):

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلَسْطُمْ نَفْسٌ مَّا قَدَّمْتُمْ لِغَدٍ ۖ وَاتَّقُوا اللَّهَ ۚ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

*"Hai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok; dan bertakwalah kepada Allah. Sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan." (Q.S. Al-Hashr Ayat 18)*

Ayat ini menekankan pentingnya setiap individu untuk berpikir kritis, introspektif, dan penuh tanggung jawab dalam menyikapi berbagai peristiwa. Hal ini selaras dengan penelitian mengenai analisis sentimen program MBG, di mana opini publik yang terekspresi di media sosial perlu dipahami dengan bijak agar dapat menjadi bahan evaluasi yang membangun, bukan sekadar ungkapan emosi yang menimbulkan perpecahan.

Dalam konteks penelitian ini, analisis sentimen berfungsi sebagai alat untuk menangkap persepsi publik yang berkaitan langsung dengan tingkat keberhasilan program MBG. Dengan mengklasifikasikan opini menjadi sentimen positif dan negatif, penelitian bertujuan mengidentifikasi indikator-indikator keberhasilan yang dianggap penting oleh masyarakat misalnya apakah penerima mendapat manfaat nyata, apakah cakupan program sudah merata, atau apakah terdapat kendala operasional yang menghambat tujuan program. Oleh karena itu, pendekatan analitis yang mengombinasikan representasi teks berbasis TF-IDF dan klasifikasi oleh SVM dipandang tepat untuk memetakan sejauh mana program dipersepsikan berhasil oleh publik.

Beberapa penelitian terdahulu telah menerapkan SVM untuk mengklasifikasikan sentimen masyarakat terhadap program sosial (Purwanti & Sugiyono, 2024). Namun, kajian-kajian tersebut umumnya fokus pada pemetaan sentimen secara umum (positif/negatif) tanpa memberikan ukuran yang eksplisit

mengenai tingkat keberhasilan program sebagai hasil akhir. Keterbatasan ini membuka ruang untuk penelitian yang lebih terfokus pada pengukuran keberhasilan program melalui opini publik, sehingga hasil analisis tidak hanya menunjukkan kecenderungan sentimen, tetapi juga menginterpretasikan aspek-aspek yang menjadi indikator keberhasilan atau kegagalan pelaksanaan program.

Berdasarkan hal tersebut, penelitian ini difokuskan pada analisis sentimen masyarakat terhadap Program MBG di Media Sosial X dengan tujuan menilai sejauh mana program dianggap berhasil oleh publik. Penelitian ini menggunakan SVM yang dipadukan dengan TF-IDF untuk mengotomatisasi klasifikasi sentimen dan mengekstrak indikator-indikator keberhasilan yang paling sering muncul dalam percakapan publik. Dengan demikian, hasil penelitian diharapkan memberi gambaran yang lebih komprehensif tentang efektivitas program serta menyediakan masukan yang dapat dipakai oleh pihak penyelenggara untuk meningkatkan pelaksanaan kebijakan.

## **1.2 Rumusan Masalah**

1. Bagaimana hasil analisis sentimen masyarakat di media sosial X dalam mengidentifikasi sentimen positif dan negatif terhadap Program MBG?
2. Sejauh mana analisis sentimen tersebut dapat menggambarkan tingkat keberhasilan pelaksanaan Program MBG berdasarkan opini publik di media sosial X?

### **1.3 Batasan Masalah**

1. Data yang digunakan terbatas pada komentar masyarakat terkait Program MBG yang diperoleh dari media sosial X.
2. Analisis sentimen difokuskan pada dua kategori sentimen, yaitu positif dan negatif, dengan satu aspek utama yaitu tingkat keberhasilan pelaksanaan program.
3. Data yang dianalisis merupakan komentar berbahasa Indonesia.
4. Periode pengambilan data dibatasi pada komentar yang dipublikasikan sepanjang tahun 2024–2025 selama pelaksanaan Program MBG

### **1.4 Tujuan Penelitian**

1. Menganalisis sentimen masyarakat di media sosial X terhadap Program MBG dengan menggunakan metode SVM dan TF–IDF.
2. Mengidentifikasi tingkat keberhasilan Program MBG berdasarkan sentimen positif dan negatif yang diungkapkan masyarakat di media sosial, sehingga hasilnya dapat menjadi masukan bagi pemerintah dalam evaluasi dan peningkatan efektivitas program.

### **1.5 Manfaat Penelitian**

1. Bagi pemerintah, penelitian ini dapat memberikan gambaran mengenai bagaimana masyarakat menilai keberhasilan Program MBG, sehingga hasilnya dapat dijadikan dasar untuk evaluasi dan peningkatan implementasi program.

2. Bagi masyarakat, penelitian ini menjadi sarana agar opini publik dapat tersampaikan dan dipertimbangkan dalam proses perbaikan kebijakan.
3. Bagi pengembangan ilmu pengetahuan, penelitian ini memberikan kontribusi dalam penerapan analisis sentimen berbasis SVM dan TF-IDF pada konteks kebijakan publik.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 Makan Bergizi Gratis**

Perkembangan teknologi digital juga memengaruhi cara masyarakat menanggapi kebijakan pemerintah, termasuk Program MBG yang baru diluncurkan. Program ini merupakan salah satu inisiatif pemerintah untuk meningkatkan gizi anak sekolah serta mendukung ketahanan pangan di masyarakat (Gunadi et al., 2025). Kebijakan ini tidak hanya bertujuan memenuhi kebutuhan dasar berupa makanan bergizi, tetapi juga diharapkan dapat mendorong pemerataan akses gizi bagi siswa di berbagai wilayah Indonesia. Selain itu, Program MBG juga diproyeksikan memberikan dampak positif pada aspek sosial dan pendidikan, karena anak-anak yang mendapatkan asupan makanan sehat diyakini dapat lebih fokus dalam belajar dan mengurangi angka ketidakhadiran di sekolah (Merlinda & Yusuf, 2025). Dengan demikian, kebijakan ini mendapat perhatian luas dari publik, baik dalam bentuk dukungan maupun kritik, terutama di media sosial.

Meskipun program ini memiliki tujuan yang positif, pelaksanaannya menghadapi berbagai tantangan. Beberapa isu yang sering muncul dalam diskusi publik antara lain keterbatasan anggaran, kesiapan infrastruktur distribusi, serta potensi ketimpangan dalam implementasi di berbagai daerah (A. Pratiwi et al., 2025). Selain itu, tanggapan masyarakat yang muncul di media sosial kerap kali bersifat *emosional* dan *subjektif*, sehingga sulit dipetakan secara manual menjadi informasi yang representatif (Fazri & Voutama, 2025). Hal ini menunjukkan pentingnya pemanfaatan teknologi analisis teks untuk mengolah opini publik secara



lebih sistematis, sehingga pemerintah dapat memperoleh gambaran yang lebih akurat terkait persepsi masyarakat terhadap kebijakan tersebut (Pradhana et al., 2021).

Dengan semakin banyaknya ulasan terkait MBG, analisis *manual* jelas tidak efisien karena memerlukan waktu yang lama dan rentan bias. Oleh sebab itu, pemanfaatan metode analisis berbasis *machine learning* menjadi alternatif yang lebih tepat. Studi mengenai analisis sentimen publik terhadap kebijakan pemerintah menunjukkan bahwa algoritma seperti SVM mampu mengklasifikasikan opini positif, negatif, maupun netral dengan lebih akurat dibandingkan penilaian manual (Fathoni et al., 2025). Selain itu, penelitian lain menemukan bahwa kombinasi *Word2Vec* dan SVM dapat meningkatkan akurasi analisis ulasan pengguna aplikasi digital, meskipun teks yang digunakan bersifat subjektif dan tidak baku (Supian et al., 2024). Dengan demikian, penerapan analisis teks otomatis menjadi penting agar pemerintah dapat menangkap persepsi masyarakat secara lebih objektif, khususnya mengenai tingkat keberhasilan Program MBG (Aziz et al., 2025).

## 2.2 Analisis Sentimen Berbasis Aspek

Analisis Sentimen Berbasis Aspek (*Aspect-Based Sentiment Analysis / ASBA*) merupakan pendekatan yang digunakan untuk mengevaluasi opini masyarakat terhadap aspek-aspek tertentu dari suatu kebijakan, layanan, atau program publik. Berbeda dengan analisis sentimen umum yang hanya menilai opini secara keseluruhan, metode ini memberikan pemahaman yang lebih mendalam dengan mengidentifikasi sentimen pada bagian spesifik dari objek penelitian (Perdana et al., 2021). Dalam penelitian ini, analisis sentimen difokuskan pada satu

aspek utama, yaitu tingkat keberhasilan program MBG. Aspek ini dipilih karena mencerminkan persepsi publik secara langsung terhadap efektivitas pelaksanaan program, meliputi sejauh mana tujuan program tercapai, dampak yang dirasakan masyarakat, serta konsistensi pemerintah dalam menjalankannya.

Analisis terhadap aspek tingkat keberhasilan ini memungkinkan peneliti untuk mengetahui bagaimana publik menilai program tersebut apakah dianggap berjalan dengan baik dan memberikan manfaat nyata, atau justru dinilai belum efektif dan masih menghadapi berbagai kendala di lapangan. Pendekatan ini menggunakan dua kategori utama, yaitu sentimen positif dan sentimen negatif. Sentimen positif mencerminkan apresiasi masyarakat terhadap keberhasilan implementasi program, sedangkan sentimen negatif menunjukkan adanya kritik atau ketidakpuasan terhadap hasil pelaksanaan. Dengan demikian, analisis sentimen berbasis aspek ini memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap keberhasilan program MBG di media sosial X. (Salsabila et al., 2023).

Dalam penerapan analisis sentimen berbasis aspek, pemilihan aspek yang akan dianalisis menjadi tahap penting karena harus sesuai dengan tujuan penelitian dan relevan dengan kebijakan yang dikaji. Pada penelitian ini, aspek yang digunakan difokuskan pada tingkat keberhasilan program MBG, yang mencerminkan seberapa efektif program tersebut dijalankan dan sejauh mana tujuannya tercapai. Aspek ini meliputi persepsi masyarakat terhadap kelancaran pelaksanaan, ketepatan sasaran penerima, keberlanjutan program, serta dampak positif yang dirasakan oleh masyarakat, khususnya peserta didik. Pemilihan aspek

tunggal ini dilakukan berdasarkan kajian penelitian terdahulu yang menyoroti efektivitas kebijakan sosial pemerintah serta faktor-faktor yang menentukan keberhasilannya dalam konteks implementasi di lapangan (Firdaus et al., 2024). Dengan memusatkan analisis pada satu aspek yang paling representatif, penelitian ini diharapkan dapat memberikan gambaran yang lebih fokus dan mendalam mengenai bagaimana publik menilai tingkat keberhasilan program MBG di media sosial X. (Sukwadi et al., 2025).

Dengan berkembangnya teknologi, penerapan analisis sentimen berbasis aspek kini semakin luas dan digunakan di berbagai bidang, mulai dari perbankan, perhotelan, teknologi, hingga sektor layanan publik (Sejati et al., 2024). Pada penelitian ini, fokus diarahkan pada kebijakan pemerintah, khususnya Program MBG. Adapun Tabel 2.1 merangkum beberapa penelitian terdahulu yang relevan dengan analisis sentimen berbasis aspek.

Tabel 2. 1 Penelitian Terdahulu

No	Referensi Penelitian	Input	Metode Yang digunakan	Hasil
1	R. A. Rahman, V. H. Pranatawijaya, dan N. K. Sari, 2024)	Data ulasan pengguna aplikasi Gojek di Google Play Store (periode 2020–2023)	Latent Dirichlet Allocation (LDA) untuk identifikasi aspek, serta BERT (distilbert-base-uncased-finetuned-sst-2-english) untuk klasifikasi sentimen	Model BERT menghasilkan akurasi sentimen 96,67% dengan precision 97,58%, recall 94,87%, dan F1-score 96,20%. Untuk klasifikasi aspek: Service 98,78%, Payment 96,42%, dan User Experience 97,65%.
2	(Rifqy Mikoriza Turjaman & Indra Budi, 2022)	Data ulasan pengguna LinkAja	String Matching, SVM dengan Undersampling	Sentimen negatif pada aspek produk (98%) dan tempat (100%), netral pada aspek harga (89%), dan positif pada aspek promosi (98%) dalam analisis sentimen LinkAja.

No	Referensi Penelitian	Input	Metode Yang digunakan	Hasil
3	(Yoga Tika Pratama, Fitra Abdurrachman Bachtiar, dan Nanang Yudi Setiawan, 2018)	Data ulasan pengguna untuk klasifikasi sentimen	Support vector machine	Menunjukkan hasil akurasi klasifikasi rata-rata 85% dengan precision, recall, dan F1-score yang mendukung visualisasi hasil dalam dashboard.
4	(M. A. Palimbani, R. P. Hastuti, and R. A. Rajagede 2024)	Data ulasan pengguna aplikasi Starbucks di Google Play Store	Support Vector Machine (SVM) dengan tiga kernel (Linear, Polinomial, RBF), TF-IDF untuk ekstraksi fitur, preprocessing data, dan Hyperparameter Tuning dengan GridSearchCV	Model SVM memperoleh performa cukup baik dengan rata-rata skor accuracy 88,96%, precision 75,77%, recall 64,68%, dan F1-score 66,85%. Analisis menunjukkan mayoritas sentimen bernilai negatif terutama pada aspek errors, menandakan tingginya tingkat kesalahan pada sistem.
5	(Salsabila Roiqoh, Badrus Zaman, Kartono Kartono 2023)	Data ulasan pengguna aplikasi Mobile JKN versi 4.2.3 dan 4.3.0 di Google Play Store	Latent Dirichlet Allocation (LDA) untuk pemodelan aspek/topik, Naïve Bayes dan Lexicon-Based (Inset Lexicon) untuk analisis sentimen	LDA menghasilkan 3 aspek utama: Services and Features, Register and Login, dan User Satisfaction, dengan coherence score terbaik 0,6392. Metode Naïve Bayes unggul dibanding Lexicon-Based dengan akurasi 94,75%, sedangkan Lexicon-Based hanya 59,99%.
6	Purwanti & Sugiyono (2024)	Data komentar masyarakat di media sosial X terkait Program MBG	TF-IDF dan SVM	Mengklasifikasikan tanggapan publik terhadap program sosial menjadi kategori positif dan negatif dengan akurasi 87%, menunjukkan efektivitas SVM dalam analisis opini publik terhadap kebijakan sosial.

No	Referensi Penelitian	Input	Metode Yang digunakan	Hasil
7	Moh. Musa Al Kadzim (2024)	Data komentar masyarakat di media sosial X terkait <i>Program MBG</i>	TF-IDF dan <i>SVM</i>	Mengklasifikasikan tanggapan publik terhadap program sosial menjadi kategori positif dan negatif dan menilai berapa persen Tingkat keberhasilannya menunjukkan efektivitas <i>SVM</i> dalam analisis opini publik terhadap kebijakan sosial.

Dari penjelasan tabel 2.1 mengenai penelitian terdahulu terkait analisis sentimen pada berbagai aplikasi layanan digital, dapat dilihat bahwa beragam metode telah digunakan dengan hasil yang cukup signifikan. Rahman, Pranatawijaya, dan Sari (2024) memanfaatkan metode *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi aspek serta model BERT dalam klasifikasi sentimen pada ulasan aplikasi Gojek. Penelitian ini menghasilkan performa yang tinggi, dengan akurasi sentimen mencapai 96,67% dan *F1-score* sebesar 96,20%, serta akurasi aspek yang konsisten di atas 96%.

Turjaman dan Budi (2022) meneliti ulasan pengguna aplikasi LinkAja dengan pendekatan *string matching* dan *SVM* menggunakan *undersampling*. Hasil penelitian mereka menunjukkan adanya variasi sentimen berdasarkan aspek, di mana aspek produk dan tempat cenderung mendapat sentimen negatif, aspek harga relatif netral, sedangkan aspek promosi memperoleh sentimen positif yang dominan.

Pratama, Bachtiar, dan Setiawan (2018) menerapkan algoritma SVM pada data ulasan pengguna untuk klasifikasi sentimen. Penelitian ini menghasilkan akurasi rata-rata sebesar 85% dengan dukungan metrik *precision*, *recall*, dan *F1-score*, yang kemudian divisualisasikan dalam bentuk dashboard.

M. A. Palimbani, R. P. Hastuti, and R. A. Rajagede 2024 Penelitian ini menganalisis ulasan pengguna aplikasi Starbucks di Google Play Store dengan fokus pada sentimen serta aspek-aspek *usability* aplikasi, yaitu *learnability*, *efficiency*, *errors*, dan *satisfaction*. Metode SVM digunakan dengan tiga jenis kernel (Linear, Polinomial, dan RBF), diikuti tahap *preprocessing*, ekstraksi fitur menggunakan *TF-IDF*, serta optimasi model melalui *Hyperparameter Tuning* dengan *GridSearchCV*. Hasil penelitian menunjukkan bahwa model SVM mampu memberikan performa klasifikasi yang cukup baik dengan rata-rata akurasi 88,96%, *precision* 75,77%, *recall* 64,68%, dan *F1-score* 66,85%. Dari analisis sentimen diketahui bahwa mayoritas ulasan pengguna cenderung bernilai negatif, terutama pada aspek *errors*, yang menunjukkan adanya tingkat kesalahan sistem yang cukup tinggi dalam aplikasi Starbucks.

Penelitian lain oleh Salsabila Roiqoh, Badrus Zaman, Kartono Kartono 2023 menganalisis ulasan pengguna aplikasi Mobile JKN versi 4.2.3 dan 4.3.0 yang tersedia di Google Play Store dengan tujuan mengevaluasi aspek-aspek penting serta sentimen pengguna terhadap layanan aplikasi tersebut. Pemodelan aspek dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA) yang menghasilkan tiga aspek utama, yaitu *Services and Features*, *Register and Login*, serta *User Satisfaction*, dengan *coherence score* terbaik sebesar 0,6392.

Selanjutnya, analisis sentimen dilakukan menggunakan dua pendekatan, yakni *Naïve Bayes* dan *Lexicon-Based* (Inset Lexicon). Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* memberikan performa yang lebih baik dengan akurasi 94,75%, sedangkan metode *Lexicon-Based* hanya memperoleh akurasi sebesar 59,99%. Temuan ini menegaskan bahwa pendekatan berbasis *machine learning* lebih efektif dalam menangkap sentimen pengguna dibandingkan metode berbasis leksikon pada ulasan Mobile JKN.

Purwanti dan Sugiyono (2024) melakukan penelitian analisis sentimen terhadap Program MBG untuk mengetahui tanggapan masyarakat di media sosial X. Data dikumpulkan melalui proses *crawling* dan diolah menggunakan tahapan *preprocessing* serta pembobotan *Term Frequency–Inverse Document Frequency* (TF–IDF), kemudian diklasifikasikan dengan algoritma SVM menjadi dua kategori, yaitu sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa kombinasi TF–IDF dan SVM mampu memberikan performa klasifikasi yang baik dengan akurasi mencapai 87%, *precision* 85%, *recall* 89%, dan *F1-score* 86%. Mayoritas opini publik menunjukkan sentimen positif yang menyoroti manfaat program bagi peningkatan gizi anak sekolah, sedangkan sentimen negatif lebih banyak berkaitan dengan kendala pelaksanaan dan efektivitas kebijakan. Penelitian ini membuktikan bahwa SVM efektif digunakan dalam mengidentifikasi persepsi masyarakat terhadap kebijakan publik berbasis teks dari media sosial.

Moh.Musa al Kadzim (2025) melakukan penelitian analisis sentimen berbasis aspek terhadap Program MBG untuk mengetahui tanggapan masyarakat di media sosial X. Data dikumpulkan melalui proses *crawling* dan diolah

menggunakan tahapan *preprocessing* serta pembobotan *Term Frequency–Inverse Document Frequency* (TF–IDF), kemudian diklasifikasikan dengan algoritma SVM menjadi dua kategori, yaitu sentimen positif dan negatif. Setelah itu dihitung berapa persen aspek Tingkat keberhasilan program MBG Secara keseluruhan, keenam penelitian terdahulu menunjukkan bahwa metode berbasis *machine learning*, khususnya kombinasi SVM dan *TF–IDF*, memberikan performa yang konsisten dan akurat dalam analisis sentimen pada berbagai konteks, mulai dari aplikasi digital hingga kebijakan publik. Beberapa penelitian seperti yang dilakukan oleh Rahman et al. (2024) dan Palimbani et al. (2024) menunjukkan efektivitas SVM dalam mengklasifikasikan opini pengguna dengan tingkat akurasi tinggi, sementara penelitian oleh Purwanti dan Sugiyono (2024) membuktikan bahwa metode yang sama juga relevan untuk menganalisis persepsi masyarakat terhadap Program MBG. Selain itu, metode *Latent Dirichlet Allocation* (LDA) yang digunakan dalam beberapa studi turut memberikan nilai tambah dengan mengidentifikasi topik atau aspek utama yang menjadi fokus perhatian publik. Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan SVM dengan dukungan teknik pembobotan seperti TF–IDF merupakan pendekatan yang efektif dan dapat diadaptasi untuk menganalisis sentimen masyarakat terhadap kebijakan sosial seperti Program MBG

### ***2.3 Preprocessing Data***

*Preprocessing data* merupakan salah satu tahap penting dalam *text mining* yang bertujuan mengubah data mentah yang tidak terstruktur menjadi data yang lebih terorganisir dan siap untuk dianalisis (Khairunnisa et al., 2021). Data mentah,



seperti komentar atau ulasan dari media sosial, biasanya mengandung banyak elemen yang tidak relevan, misalnya simbol, emotikon, singkatan, atau bahasa tidak baku (Rifaldi et al., 2023). Oleh karena itu, data tersebut tidak dapat langsung digunakan dalam proses analisis. Melalui tahapan *text preprocessing*, data dibersihkan dan disederhanakan agar dapat diproses lebih mudah oleh algoritma *machine learning*.

Tahapan ini mencakup proses *cleansing* (memilah kalimat yang tidak penting), *case folding* (mengubah semua huruf menjadi huruf kecil), *tokenizing* (memecah kalimat menjadi kata-kata), *stopword removal* (menghapus kata-kata yang tidak bermakna penting), serta *stemming* (mengembalikan kata ke bentuk dasar) (Saka & Prasetyaningrum, 2025). Dengan adanya *preprocessing*, data opini masyarakat yang semula berantakan dapat disajikan dalam bentuk yang lebih terstruktur, sehingga algoritma seperti SVM dapat bekerja lebih optimal dalam mengklasifikasikan sentimen positif maupun negatif terhadap kebijakan publik, termasuk pada Program MBG (Rianto et al., 2021).

#### **2.4 Term Frequency-Inverse Document Frequency (TF-IDF)**

Pembobotan kata atau *word weighting* merupakan teknik yang digunakan untuk memberikan nilai pada setiap kata dalam sebuah dokumen. Teknik ini penting karena membantu menilai tingkat relevansi antar kata sekaligus mempersiapkan data agar lebih mudah diproses pada tahap pembangunan model analisis berikutnya (Widyadhana et al., 2023). Salah satu metode pembobotan yang paling banyak digunakan adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini menilai seberapa penting sebuah kata dalam satu dokumen

dibandingkan dengan kemunculannya di seluruh kumpulan dokumen (Sinulingga & Sitorus, 2024). TF-IDF sendiri merupakan gabungan dari dua komponen, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), yang bekerja bersama untuk memberikan bobot lebih tepat pada kata-kata yang benar-benar bermakna.

Komponen TF mengukur frekuensi kemunculan kata dalam sebuah dokumen. Semakin sering kata tertentu muncul, semakin tinggi bobot yang diberikan. Namun, kelemahannya adalah jika kata tersebut muncul di hampir semua dokumen, maka nilainya menjadi kurang signifikan karena dianggap sebagai kata umum (*common term*) (Nico et al., 2022). Kata-kata umum biasanya tidak banyak membantu dalam mengidentifikasi topik utama sebuah teks.

Untuk mengatasi hal ini, digunakanlah IDF yang menghitung seberapa jarang sebuah kata muncul di seluruh dokumen (D. Pratiwi et al., 2024). Kata yang lebih jarang muncul akan memiliki bobot yang lebih besar, karena dianggap lebih spesifik dan lebih mampu merepresentasikan topik tertentu (Annisa & Kalifia, 2024). Proses perhitungan TF-IDF dilakukan dengan cara mengalikan nilai TF setiap kata dengan nilai IDF-nya, sehingga menghasilkan bobot akhir yang menunjukkan tingkat kepentingan suatu kata dalam dokumen (Septiani & Isabela, 2022).

Secara sederhana, TF digunakan untuk mengukur seberapa sering sebuah kata muncul dalam satu dokumen, sedangkan IDF berfungsi untuk memberikan bobot lebih besar pada kata-kata yang jarang ditemukan di banyak dokumen, sehingga kata-kata umum tidak mendominasi hasil analisis. Nilai TF-IDF yang

dihasilkan menunjukkan seberapa penting sebuah kata dalam konteks tertentu, dengan bobot tinggi diberikan pada kata yang jarang tetapi bermakna. Setelah proses *preprocessing* selesai, tahap berikutnya adalah melakukan ekstraksi fitur menggunakan TF sesuai dengan rumus pada Persamaan 2.1 (Qhabib et al., 2023) :

$$TF(t, d) = \frac{f_{t,d}}{N_d} \quad (2.1)$$

Keterangan:

- $f_{t,d}$  = Jumlah kemunculan term/kata dalam dokumen  $d$ .
- $N_d$  = Jumlah total term/kata dalam

Kemudian setelah dilakukan perhitungan TF, selanjutnya akan dilakukan perhitungan IDF seperti persamaan 2.2:

$$IDF(t, D) = \log_{10} \frac{N}{n_t} \quad (2.2)$$

Keterangan:

- $t$  = Term/kata tertentu yang dianalisis.
- $D$  = Koleksi atau kumpulan seluruh dokumen yang dianalisis.
- $N$  = Jumlah total dokumen dalam koleksi  $D$ .
- $n_t$  = Jumlah dokumen yang mengandung term/kata

Setelah menghitung nilai TF dan IDF, langkah terakhir adalah menghitung nilai bobot TF-IDF untuk menentukan sejauh mana pentingnya suatu kata dalam dokumen, seperti persamaan 2.3:

$$W_{t,d} = TF(t, d) \times IDF(t, D) \quad (2.3)$$

Keterangan:

- $W_{t,d}$  : Bobot TF-IDF untuk kata  $t$  dalam dokumen  $d$ .
- $TF(t, d)$  : Frekuensi kemunculan kata dalam dokumen.
- $IDF(t, d)$  = Ukuran seberapa umum atau jarang term  $t$  muncul di seluruh dokumen dalam koleksi  $D$ .

Dengan teknik ini, analisis sentimen terhadap ulasan masyarakat mengenai Program MBG dapat lebih terarah, karena kata-kata kunci yang relevan akan lebih menonjol dibandingkan kata-kata umum (Suharman & Kamayani Sulaeman, 2025)

## 2.5 *Support Vector Machine*

SVM adalah salah satu metode klasifikasi *supervised learning* yang didasarkan pada prinsip *Structural Risk Minimization* (SRM). Tujuannya adalah menemukan *hyperplane* (bidang pemisah) optimal yang dapat memisahkan dua kelas data dengan margin (jarak) terbesar (Cortes & Vapnik, 1995). Meskipun awalnya dirancang untuk klasifikasi biner, SVM sangat efektif dan banyak digunakan dalam berbagai aplikasi, termasuk klasifikasi teks, pengenalan gambar, dan bioinformatika.

SVM disebut *Support Vector Machine* karena *hyperplane* pemisah yang optimal tersebut yaitu garis keputusan yang sepenuhnya ditentukan oleh subset data latih yang paling dekat dengan garis tersebut. Titik-titik data ini disebut sebagai *Support Vectors*. Artinya, hanya titik-titik kritis inilah yang "mendukung" dan mendefinisikan batas keputusan, membuat SVM menjadi efisien secara memori, terutama dalam ruang berdimensi tinggi seperti pada data teks (Pratama, 2023).

SVM merupakan salah satu algoritma pembelajaran mesin yang banyak digunakan untuk tugas klasifikasi maupun regresi. Tujuan utama SVM adalah menemukan *hyperplane* terbaik yang mampu memisahkan data ke dalam dua kelas berbeda secara optimal (Alhaq et al., 2021). *Hyperplane* ini berfungsi sebagai garis pemisah di ruang fitur, sementara SVM berfokus untuk memaksimalkan *margin*, yaitu jarak antara *hyperplane* dan titik data terdekat dari tiap kelas yang dikenal

sebagai *support vectors* (Akbar et al., 2023). Titik-titik inilah yang menentukan posisi *hyperplane* serta memengaruhi hasil klasifikasi.

Persamaan dari *hyperplane* dalam bentuk linear untuk model SVM seperti pada persamaan 2.4 (Aisy et al., 2025):

$$w \cdot x + b = 0 \quad (2.4)$$

Keterangan:

- $w$  adalah vektor bobot.

- $x$  adalah vektor fitur (dalam kasus ini, nilai TF-IDF dari dokumen).

- $b$  adalah bias.

Langkah-langkah untuk memformulasikan dan mengimplementasikan SVM adalah sebagai berikut:

#### 1. Formulasi Optimasi (*Hard Margin*)

Tujuan SVM adalah memaksimalkan *margin*. *Margin* adalah jarak antara dua *hyperplane* paralel,  $w \cdot x + b = 1$  untuk *support vector* kelas positif, dan  $w \cdot x + b = -1$  untuk kelas negatif. Jarak *margin* adalah  $2/\|w\|$ .

Untuk memaksimalkan jarak ini, algoritma SVM secara matematis meminimalkan nilai  $\|w\|^2$ . Formulasi ini disebut *Hard Margin* dan mengasumsikan data dapat dipisahkan secara sempurna, seperti pada persamaan 2.5:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (2.5)$$

Keterangan:

- $w$  : vektor bobot

- $b$  : bias

- $\|w\|^2$  : kuadrat norma Euclidean dari vektor bobot

Dengan batasan:  $y_i(w \cdot x_i + b) \geq 1$

Keterangan:  $y_1$  : Label kelas data (misalnya +1 atau -1)

## 2. Representasi Bobot (Dual Form)

Untuk memudahkan pencarian nilai optimal, vektor bobot  $w$  dapat dinyatakan sebagai kombinasi linear dari data latih menggunakan nilai *Lagrange Multipliers* ( $\alpha$ ). Persamaan ini digunakan untuk menghitung nilai bobot akhir setelah proses pelatihan (*training*) selesai:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (2.6)$$

Keterangan:

- $w$  : vektor bobot
- $\alpha_i$  : parameter Lagrange dari data ke- $i$
- $y_i$  : label kelas dari data ke- $i$
- $x_i$  : vektor fitur dari data ke- $i$
- $n$  : jumlah data latih

## 3. Mengatasi Data Non-Linear (*Kernel Trick*)

Dalam praktiknya, data di dunia nyata (termasuk data teks) seringkali tidak dapat dipisahkan secara linear. SVM mengatasi ini dengan menggunakan *Kernel Trick*. Fungsi *kernel* memetakan data ke ruang berdimensi lebih tinggi sehingga data tersebut dapat dipisahkan secara linear di ruang yang baru (Suryanto, 2022). Penelitian ini menggunakan tiga *kernel*:

### A. *Kernel Linear*

Bentuk paling dasar, digunakan jika data sudah dapat dipisahkan secara linear. Rumus dari *kernel linear* bisa dilihat di persamaan 2.7 berikut:

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (2.7)$$

Keterangan:

- $K(x_i, x_j)$ : mengukur kemiripan antara dua data.
- $x_i, x_j$  : vektor fitur dari data ke- $i$  dan ke- $j$ .
- $x_i^T$ : transpose dari vektor  $x_i$ .
- $\cdot$  : operasi dot product (perkalian titik).

$-x_i^T \cdot x_j$  : memberi nilai kemiripan; semakin besar nilainya, umumnya semakin mirip kedua vektor tersebut.

#### B. *Kernel Polinomial*

Memetakan data menggunakan fungsi polinomial, efektif untuk batas keputusan melengkung. Rumus dari *kernel polynomial* bisa dilihat di persamaan 2.8 berikut:

$$K(x_i, x_j) = (\gamma(x_i^T \cdot x_j) + r)^d \quad (2.8)$$

Keterangan:

- $K(x_i, x_j)$ : mengukur kemiripan antara dua data.
- $x_i, x_j$  : vektor fitur dari data ke- $i$  dan ke- $j$ .
- $x_i^T \cdot x_j$  : dot product (perkalian titik) antara  $x_i$  dan  $x_j$ .
- $\gamma$  : parameter skala (mengatur pengaruh nilai dot product).
- $r$  : konstanta/parameter bias (sering disebut  $\text{coef0}$ ).
- $d$  : derajat polinomial (degree) yang menentukan kompleksitas pemetaan fitur.

#### C. *Kernel Radial Basis Function (RBF)*

Kernel yang paling fleksibel dan populer, mampu menangani batas keputusan yang sangat kompleks. Rumus dari kernel RBF bisa dilihat di persamaan 2.9 berikut:

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2) \quad (2.9)$$

Keterangan:

- $K(x_i, x_j)$ : mengukur kemiripan antara dua data.
- $x_i, x_j$  : vektor fitur dari data ke- $i$  dan ke- $j$ .
- $|x_i - x_j|^2$  : jarak Euclidean kuadrat antara  $x_i$  dan  $x_j$ .
- $\gamma$  : parameter yang mengatur “lebar” kernel
- $\exp(\cdot)$  : fungsi eksponensial; nilai kernel mendekati 1 jika  $x_i$  dan  $x_j$  sangat mirip (jaraknya kecil), dan mendekati 0 jika jaraknya besar.

### 4. Parameter Optimasi SVM (Hyperparameter)

Dalam implementasi SVM, kinerja model tidak hanya bergantung pada data, tetapi juga pada pengaturan *hyperparameter*. *Hyperparameter* adalah setelan yang dapat diatur oleh peneliti untuk mengontrol perilaku model. Dalam

penelitian ini, *hyperparameter* yang akan dioptimasi adalah Parameter C (untuk *Soft Margin*), serta Parameter Kernel (yaitu gamma dan degree).

a. Parameter C (Regularisasi Soft Margin)

Untuk menangani data yang tumpang tindih (berisi *noise* atau *outlier*), SVM menggunakan formulasi *Soft Margin*. *Soft Margin* mengizinkan beberapa data salah diklasifikasi dengan memberikan "hukuman". Hukuman ini diatur oleh Parameter C.

- Nilai C kecil: Menerapkan hukuman ringan, menghasilkan *margin* yang lebar, namun mentolerir lebih banyak kesalahan klasifikasi (mencegah *overfitting*).
- Nilai C besar: Menerapkan hukuman berat, menghasilkan *margin* yang sempit, dan berusaha mengklasifikasikan semua data dengan benar (berisiko *overfitting*).

Rumus *soft margin* dimodifikasi dari *hard margin* sehingga menjadi persamaan 2.9 berikut:

$$\min_{w,b,\xi} \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i \quad (2.9)$$

Dengan batasan  $y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$

Keterangan:

- $w$  : vektor bobot (menentukan hyperplane).

- $b$  : bias/intersep.

- $\xi_i$  : *slack variable* (besar pelanggaran margin oleh data ke- $i$ ), dengan  $\xi_i \geq 0$ .

- $C$  : parameter penalti (trade-off

$\frac{1}{2} |w|^2$  : memaksimalkan margin (karena meminimalkan norm bobot).

$\sum_{i=1}^n \xi_i$  : total pelanggaran margin/klasifikasi.

b. Parameter Kernel: *Gamma* ( $\gamma$ ) dan *Degree* ( $d$ )



Selain Parameter  $C$ , kernel *non-linear* (Polinomial dan RBF) memiliki parameter internal mereka sendiri yang juga sangat memengaruhi bentuk batas keputusan.

#### 1) *Gamma* ( $\gamma$ )

Parameter ini digunakan oleh kernel RBF dan Polinomial. *Gamma* mendefinisikan seberapa besar pengaruh satu titik data latih.

- *Gamma* kecil: Pengaruh satu titik data lebih luas, menghasilkan batas keputusan yang lebih halus dan umum (mirip *margin* lebar).
- *Gamma* besar: Pengaruh satu titik data lebih sempit (lokal), menghasilkan batas keputusan yang lebih kompleks dan "berlekuk" yang sangat mengikuti data latih (berisiko *overfitting*).

#### 2) *Degree* ( $d$ )

Parameter ini digunakan hanya oleh kernel Polinomial. *Degree* menentukan derajat dari fungsi polinomial yang digunakan untuk memetakan data.

- *Degree* = 1: Menghasilkan batas linear.
- *Degree* = 2, 3, dst.: Menghasilkan batas keputusan yang lebih melengkung dan fleksibel untuk data yang lebih kompleks.

Menemukan nilai optimal dari  $C$ ,  $\gamma$ , dan  $d$  adalah langkah krusial untuk mencegah *overfitting* dan *underfitting*. Dalam penelitian ini, kombinasi dari parameter-parameter inilah yang akan diuji secara sistematis dalam skenario pengujian untuk menemukan model SVM dengan kinerja terbaik.

### 5. Klasifikasi Data Baru

Tujuan akhir dari SVM adalah memprediksi label kelas untuk data baru yang belum pernah dilihat sebelumnya. Untuk mengklasifikasikan data baru ( $x_{baru}$ ), model menghitung posisi data tersebut terhadap *hyperplane* optimal ( $w$  dan  $b$ ) menggunakan fungsi keputusan berikut:

$$f(x) = w \cdot x_{baru} + b$$

Berdasarkan nilai  $f(x)$  tersebut, kelas prediksi  $C_{pred}$  ditentukan dengan aturan keputusan sebagai berikut:

Jika  $w \cdot x_{baru} + b \geq 0$ , maka  $C_{pred} = +1$  (kelas positif)

Jika  $w \cdot x_{baru} + b < 0$ , maka  $C_{pred} = -1$  (kelas negatif)

Secara matematis, aturan ini sering diringkas menggunakan fungsi *sign* (tanda) menggunakan persamaan 2.11:

$$C_{pred} = \text{sign}(w \cdot x_{baru} + b) \quad (2.11)$$

Keterangan:

- $y_1$  : Label kelas data (misalnya +1 atau -1)
- $C_{pred}$  : kelas hasil prediksi.
- $\text{sign}(\cdot)$  : fungsi tanda yang menghasilkan +1 jika nilainya  $> 0$ , -1 jika nilainya  $< 0$ , dan (kadang) 0 jika nilainya  $= 0$ .
- $w$  : vektor bobot (parameter model SVM).
- $x_{baru}$  : data/vektor fitur baru yang akan diklasifikasikan.
- $b$  : bias/intersep.
- $w \cdot x_{baru} + b$  : nilai fungsi keputusan (decision function); tanda nilainya menentukan kelas.

Kelebihan dari metode SVM adalah sangat efektif di ruang berdimensi tinggi (seperti data teks TF-IDF), efisien dalam penggunaan memori (karena hanya menggunakan *support vectors*), dan serbaguna berkat penggunaan

*kernel*. Sedangkan kekurangan dari metode SVM adalah performa model sangat bergantung pada pemilihan fungsi *kernel* dan *Parameter C* yang tepat, sehingga proses *tuning* parameter menjadi krusial. Selain itu, SVM bisa menjadi lambat secara komputasional pada dataset yang sangat besar (Wijaya, 2024).

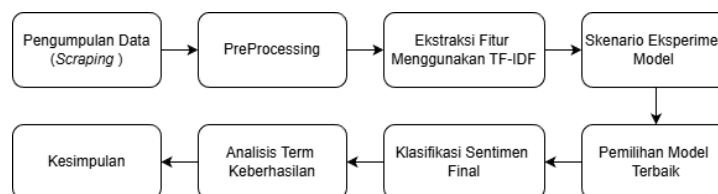
## BAB III

### METODE PENELITIAN

Bab ini membahas terkait metode yang diterapkan penulis dalam penelitian. Perancangan metode penelitian bertujuan untuk mempermudah dan memperlancar proses penyusunan skripsi, serta untuk memastikan penelitian berjalan secara sistematis dan terstruktur.

#### 3.1 Desain Sistem

Desain sistem berfungsi sebagai panduan yang menyusun alur kerja penelitian, dimulai dari tahap awal pengumpulan data dengan *scraping* hingga tahap hasil evaluasi. Dengan adanya rancangan yang terencana, setiap tahapan penelitian dapat dilakukan secara lebih terarah dan sistematis, sehingga hasil yang diperoleh lebih valid serta sesuai dengan tujuan yang ingin dicapai.



Gambar 3. 1 Desain Sistem

Berdasarkan Gambar 3.1, diagram blok ini menjelaskan tahapan proses penelitian yang disusun secara sistematis untuk menganalisis opini publik mengenai Program MBG. Tahapan dimulai dari pengumpulan data melalui proses crawling pada platform X guna memperoleh tweet yang relevan. Data mentah tersebut kemudian diproses pada tahap preprocessing untuk dibersihkan dan dinormalisasi, sehingga siap digunakan pada tahap analisis berikutnya. Setelah itu,

dilakukan ekstraksi fitur menggunakan metode TF-IDF guna merepresentasikan teks dalam bentuk numerik yang menunjukkan tingkat kepentingan setiap kata dalam dokumen. Selanjutnya, dilakukan serangkaian eksperimen menggunakan sembilan skenario pembagian data dan kernel pada algoritma SVM untuk memperoleh model dengan performa terbaik. Model terbaik tersebut kemudian digunakan pada tahap klasifikasi sentimen final untuk menentukan apakah opini masyarakat bernada positif atau negatif. Setelah klasifikasi sentimen dilakukan, penelitian dilanjutkan dengan analisis term berdasarkan kata kunci tertentu untuk mengidentifikasi aspek keberhasilan program sebagaimana tercermin dalam opini publik. Dengan demikian, keseluruhan proses tidak hanya menghasilkan pemetaan sentimen masyarakat, tetapi juga memberikan gambaran mengenai persepsi publik terhadap tingkat keberhasilan pelaksanaan Program MBG.

### **3.2 Pengumpulan Data**

Dalam penelitian ini, penulis mengumpulkan data menggunakan metode *crawling* dengan memanfaatkan *tools tweet-harvest* pada *Google Colab* (S. A. Putra & Wijaya, 2023). Metode ini dipilih karena platform X (sebelumnya Twitter) merupakan salah satu media sosial yang sering digunakan masyarakat untuk menyampaikan opini maupun tanggapan terhadap isu publik, termasuk kebijakan program MBG. Data yang diperoleh berupa *tweet* sebagai data primer yang secara langsung berkaitan dengan topik penelitian, dengan periode pengambilan mulai dari 1 Januari 2023 hingga 31 Agustus 2025 yang di akses di tanggal 30 Oktober 2025.

Penelitian ini menggunakan opini publik terkait program MBG yang diperoleh dari media sosial X (sebelumnya Twitter). Objek yang digunakan berupa *tweet* berbahasa Indonesia yang relevan dengan topik penelitian. Data dikumpulkan dengan metode *crawling* menggunakan kata kunci “Makan Bergizi Gratis”.

Dataset mentah yang berhasil dikumpulkan melalui *crawling* berjumlah 2.973 tweet. Namun dalam penelitian ini, data yang akan digunakan sebanyak 599 *tweets*, karena *tweet-tweet* yang tidak membahas tentang makan siang gratis terkait program yang dijalankan pemerintah tidak dicantumkan pada penelitian ini. Dari jumlah tersebut, dilakukan pelabelan manual yang akan digunakan sebagai dataset utama dalam penelitian ini. Data yang terkumpul disimpan dalam format .csv untuk kemudian diproses lebih lanjut melalui tahapan *preprocessing*, ekstraksi fitur (TF-IDF), dan klasifikasi menggunakan SVM.

Sebelum memasuki tahap *preprocessing*, dilakukan proses pemeriksaan awal secara manual untuk memastikan bahwa tidak terdapat data duplikat, *spam*, atau tweet yang tidak relevan dengan topik penelitian. Langkah ini dilakukan agar data yang digunakan benar-benar bersih, representatif, dan siap untuk dianalisis pada tahap selanjutnya.

Tabel 3. 1 Penyajian Contoh Data Utama

No	<i>Tweet</i>
1.	Program MBG ini menurut saya sangat berhasil! Anak-anak jadi semangat ke sekolah dan gizinya makin terjaga
2.	Pelaksanaannya masih jauh dari kata efektif. Banyak sekolah di daerah saya yang belum menerima MBG ini 🟡
3.	Salut buat pemerintah, program MBG benar-benar membantu anak-anak dari keluarga kurang mampu. Semoga terus berlanjut!

No	<i>Tweet</i>
4.	Tujuan programnya bagus, tapi realisasinya belum maksimal. Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata.
5.	Kalau dilihat dari dampaknya, program ini sudah cukup berhasil meningkatkan gizi anak sekolah. Bukti nyatanya angka stunting mulai turun di beberapa daerah.

Tabel 3. 2 Atribut Data

Kolom	Deskripsi	Tipe Data
Full_text	Teks tweet yang berisi opini publik terkait Program Makan bergizi gratis	String
Label	Hasil Labelling Sentimen Ground Truth	String

Pada aspek tingkat keberhasilan program digunakan kata kunci seperti “berhasil”, “sukses”, “efektif”, “tercapai”, “memuaskan”, “tidak efektif”, “gagal”, “kurang optimal”, “belum maksimal”, “tepat sasaran”, “tidak tepat sasaran”, “berjalan baik”, “kurang berjalan”, “tidak sesuai harapan”, “positif”, “bermanfaat”, dan “mengecewakan”.

Kata-kata tersebut mencerminkan persepsi masyarakat terhadap sejauh mana Program MBG dianggap berhasil mencapai tujuannya dalam meningkatkan gizi anak sekolah serta pemerataan akses pangan di berbagai wilayah.

Tabel 3. 3 Deskripsi Aspek Penilaian

Aspek	Deskripsi
Tingkat Keberhasilan Program	Opini publik yang menilai sejauh mana pelaksanaan Program MBG dianggap berhasil atau tidak berhasil. Aspek ini mencakup persepsi terhadap efektivitas pelaksanaan, ketepatan sasaran, manfaat yang dirasakan masyarakat, serta pencapaian tujuan program secara keseluruhan.

Setelah melakukan pemeriksaan awal untuk menghindari adanya duplikasi data dan aspek sudah ditentukan, penulis melanjutkan ke tahap pelabelan secara manual. Dalam penelitian ini, pelabelan dilakukan dengan menggunakan dua kategori sentimen, yaitu sentimen positif dan sentimen negatif, dengan dukungan teknologi AI untuk mempercepat proses. Validasi terhadap label tersebut dilakukan oleh ahli bahasa, Ibu Munjiyati, S.Pd., yang saat ini mengajar Bahasa Indonesia di SMA Al-Karimi Tebuwung, Gresik. Tujuan validasi ini adalah memastikan bahwa setiap tweet dikategorikan secara tepat sebagai positif atau negatif berdasarkan konteksnya. Sebagai acuan, penelitian oleh Sabrila dkk. (2021) menyatakan bahwa sentimen positif menunjukkan dukungan atau apresiasi terhadap suatu kebijakan, sedangkan sentimen negatif mencerminkan kritik atau ketidakpuasan terhadap implementasi kebijakan.

Tabel 3. 4 Pelabelan Tweet Secara Manual

No	<i>Tweet</i>	Aspek	Sentimen
1.	Program MBG ini menurut saya sangat berhasil! Anak- anak jadi semangat ke sekolah dan gizinya makin terjaga	Tingkat Keberhasilan Program	Positif
2.	Pelaksanaannya masih jauh dari kata efektif. Banyak sekolah di daerah saya yang belum menerima makan bergizi gratis ini 🍌	Tingkat Keberhasilan Program	Negatif
3.	Salut buat pemerintah, program MBG benar-benar membantu anak-anak dari keluarga kurang mampu. Semoga terus berlanjut!	Tingkat Keberhasilan Program	Positif
4.	Tujuan programnya bagus, tapi realisasinya belum maksimal. Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata.	Tingkat Keberhasilan Program	Negatif



No	<i>Tweet</i>	Aspek	Sentimen
5.	Kalau dilihat dari dampaknya, program ini sudah cukup berhasil meningkatkan gizi anak sekolah. Bukti nyatanya angka stunting mulai turun di beberapa daerah.	Tingkat Keberhasilan Program	Positif

Data yang telah melalui proses pelabelan ini kemudian menjadi dasar untuk menerapkan metode SVM dalam analisis sentimen berbasis aspek tingkat keberhasilan Program MBG pada tahap selanjutnya. Melalui proses ini, penulis berupaya untuk mengidentifikasi pola dan kecenderungan opini publik terhadap tingkat keberhasilan program tersebut, baik dari sisi dukungan maupun kritik masyarakat. Hasil analisis ini diharapkan dapat memberikan gambaran yang komprehensif mengenai persepsi masyarakat terhadap efektivitas pelaksanaan Program MBG serta menjadi masukan bagi pemerintah dalam penyempurnaan kebijakan di masa mendatang.

### 3.3 *Preprocessing Data*

*Preprocessing Data* adalah salah satu langkah dalam *text mining* yang berfungsi untuk mengubah data mentah yang tidak terstruktur menjadi data terstruktur yang siap dianalisis. Data mentah tidak dapat langsung diproses, sehingga perlu dilakukan tahapan *text preprocessing* agar data tersebut siap untuk dievaluasi (Rifaldi et al., 2023). Langkah-langkah *text preprocessing* yang digunakan dalam penelitian ini adalah sebagai berikut (Salam et al., 2023).:

#### A. *Cleansing*

*Cleansing* adalah tahap dimana elemen-elemen yang tidak relevan, karakter yang tidak standar, atau komponen yang tidak berhubungan dengan konten dokumen dihapus. Ini termasuk emotikon, angka, tanda baca, simbol, spasi berlebih, enter, tautan, dan tagar.

Tabel 3. 5 Proses *Cleansing*

No	Sebelum <i>Cleansing</i>	Setelah <i>Cleansing</i>
1.	Program MBG ini menurut saya sangat berhasil! Anak-anak jadi semangat ke sekolah dan gizinya makin terjaga	Program MBG ini menurut saya sangat berhasil Anakanak jadi semangat ke sekolah dan gizinya makin terjaga
2.	Pelaksanaannya masih jauh dari kata efektif. Banyak sekolah di daerah saya yang belum menerima MBG ini 😊	Pelaksanaannya masih jauh dari kata efektif Banyak sekolah di daerah saya yang belum menerima MBG ini
3.	Salut buat pemerintah, program MBG benar-benar membantu anak-anak dari keluarga kurang mampu. Semoga terus berlanjut!	Salut buat pemerintah program MBG benarbenar membantu anakanak dari keluarga kurang mampu Semoga terus berlanjut
4.	Tujuan programnya bagus, tapi realisasinya belum maksimal. Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata.	Tujuan programnya bagus tapi realisasinya belum maksimal Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata
5.	Kalau dilihat dari dampaknya, program ini sudah cukup berhasil meningkatkan gizi anak sekolah. Bukti nyatanya angka stunting mulai turun di beberapa daerah.	Kalau dilihat dari dampaknya program ini sudah cukup berhasil meningkatkan gizi anak sekolah Bukti nyatanya angka stunting mulai turun di beberapa daerah

## B. *Case Folding*

*Case folding* berfungsi untuk mengubah huruf kapital menjadi huruf kecil, sehingga semua kata berada dalam bentuk huruf kecil yang konsisten.

Tabel 3. 6 Proses *Case Folding*

No	Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
1.	Program MBG ini menurut saya sangat berhasil Anakanak jadi semangat ke sekolah dan gizinya makin terjaga	program MBG ini menurut saya sangat berhasil anakanak jadi semangat ke sekolah dan gizinya makin terjaga

No	Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
2.	Pelaksanaannya masih jauh dari kata efektif Banyak sekolah di daerah saya yang belum menerima MBG ini	pelaksanaannya masih jauh dari kata efektif banyak sekolah di daerah saya yang belum menerima MBG ini
3.	Salut buat pemerintah program MBG benarbenar membantu anakanak dari keluarga kurang mampu Semoga terus berlanjut	salut buat pemerintah program MBG benarbenar membantu anakanak dari keluarga kurang mampu semoga terus berlanjut
4.	Tujuan programnya bagus tapi realisasinya belum maksimal Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata	tujuan programnya bagus tapi realisasinya belum maksimal masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata
5.	Kalau dilihat dari dampaknya program ini sudah cukup berhasil meningkatkan gizi anak sekolah Bukti nyatanya angka stunting mulai turun di beberapa daerah	kalau dilihat dari dampaknya program ini sudah cukup berhasil meningkatkan gizi anak sekolah bukti nyatanya angka stunting mulai turun di beberapa daerah

### C. Tokenizing

*Tokenizing* adalah proses pemecahan kalimat menjadi kata-kata individual.

Tahap ini memudahkan dalam menghitung frekuensi kemunculan kata dalam dokumen.

Tabel 3. 7 Proses *Tokenizing*

No	Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
1.	program MBG ini menurut saya sangat berhasil anakanak jadi semangat ke sekolah dan gizinya makin terjaga	['program', 'makan', 'bergizi', 'gratis', 'ini', 'menurut', 'saya', 'sangat', 'berhasil', 'anakanak', 'jadi', 'semangat', 'ke', 'sekolah', 'dan', 'gizinya', 'makin', 'terjaga']
2.	pelaksanaannya masih jauh dari kata efektif banyak sekolah di daerah saya yang belum menerima MBG ini	['pelaksanaannya', 'masih', 'jauh', 'dari', 'kata', 'efektif', 'banyak', 'sekolah', 'di', 'daerah', 'saya', 'yang', 'belum', 'menerima', 'makan', 'bergizi', 'gratis', 'ini']
3.	salut buat pemerintah program MBG benarbenar membantu anakanak dari keluarga kurang mampu semoga terus berlanjut	['salut', 'buat', 'pemerintah', 'program', 'makan', 'bergizi', 'gratis', 'benarbenar', 'membantu', 'anakanak', 'dari', 'keluarga', 'kurang', 'mampu', 'semoga', 'terus', 'berlanjut']

4.	tujuan programnya bagus tapi realisasinya belum maksimal masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata	['tujuan', 'programnya', 'bagus', 'tapi', 'realisasinya', 'belum', 'maksimal', 'masih', 'banyak', 'laporan', 'makanan', 'tidak', 'sesuai', 'standar', 'dan', 'pembagian', 'tidak', 'merata']
5.	kalau dilihat dari dampaknya program ini sudah cukup berhasil meningkatkan gizi anak sekolah bukti nyatanya angka stunting mulai turun di beberapa daerah	['kalau', 'dilihat', 'dari', 'dampaknya', 'program', 'ini', 'sudah', 'cukup', 'berhasil', 'meningkatkan', 'gizi', 'anak', 'sekolah', 'bukti', 'nyatanya', 'angka', 'stunting', 'mulai', 'turun', 'di', 'beberapa', 'daerah']

#### D. Stopword Removal

Stopword Removal adalah tahap untuk menghapus kata-kata umum yang sering muncul dan tidak memiliki makna signifikan, misalnya “dan”, “untuk”, “yang”.

Tabel 3. 8 Proses *Stopword Removal*

No	Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
1.	['program', 'makan', 'bergizi', 'gratis', 'ini', 'menurut', 'saya', 'sangat', 'berhasil', 'anakanak', 'jadi', 'semangat', 'ke', 'sekolah', 'dan', 'gizinya', 'makin', 'terjaga']	['program', 'makan', 'bergizi', 'gratis', 'sangat', 'berhasil', 'anakanak', 'jadi', 'semangat', 'sekolah', 'gizinya', 'makin', 'terjaga']
2.	['pelaksanaannya', 'masih', 'jauh', 'dari', 'kata', 'efektif', 'banyak', 'sekolah', 'di', 'daerah', 'saya', 'yang', 'belum', 'menerima', 'makan', 'bergizi', 'gratis', 'ini']	['pelaksanaannya', 'jauh', 'kata', 'efektif', 'banyak', 'sekolah', 'daerah', 'menerima', 'makan', 'bergizi', 'gratis']
3.	['salut', 'buat', 'pemerintah', 'program', 'makan', 'bergizi', 'gratis', 'benarbenar', 'membantu', 'anakanak', 'dari', 'keluarga', 'kurang', 'mampu', 'semoga', 'terus', 'berlanjut']	['salut', 'buat', 'pemerintah', 'program', 'makan', 'bergizi', 'gratis', 'benarbenar', 'membantu', 'anakanak', 'keluarga', 'kurang', 'mampu', 'semoga', 'terus', 'berlanjut']
4.	['tujuan', 'programnya', 'bagus', 'tapi', 'realisasinya', 'belum', 'maksimal', 'masih', 'banyak', 'laporan', 'makanan', 'tidak', 'sesuai', 'standar', 'dan', 'pembagian', 'tidak', 'merata']	['tujuan', 'programnya', 'bagus', 'realisasinya', 'maksimal', 'banyak', 'laporan', 'makanan', 'sesuai', 'standar', 'pembagian', 'merata']
5.	['kalau', 'dilihat', 'dari', 'dampaknya', 'program', 'ini', 'sudah', 'cukup', 'berhasil', 'meningkatkan', 'gizi', 'anak', 'sekolah', 'bukti', 'nyatanya', 'angka', 'stunting', 'mulai', 'turun', 'di', 'beberapa', 'daerah']	['kalau', 'dilihat', 'dampaknya', 'program', 'cukup', 'berhasil', 'meningkatkan', 'gizi', 'anak', 'sekolah', 'bukti', 'nyatanya', 'angka', 'stunting', 'mulai', 'turun', 'beberapa', 'daerah']

### E. Stemming

*Stemming* merupakan tahap mengubah setiap kata ke bentuk dasarnya. Kata berimbuhan seperti “bergizi” → “gizi”, “terjaga” → “jaga” akan dikembalikan ke bentuk dasar.

Tabel 3. 9 Proses *Stemming*

No	Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
1.	['program', 'makan', 'bergizi', 'gratis', 'sangat', 'berhasil', 'anakanak', 'jadi', 'semangat', 'sekolah', 'gizinya', 'makin', 'terjaga']	['program', 'makan', 'gizi', 'gratis', 'sangat', 'hasil', 'anakanak', 'jadi', 'semangat', 'sekolah', 'gizi', 'makin', 'jaga']
2.	['pelaksanaannya', 'jauh', 'kata', 'efektif', 'banyak', 'sekolah', 'daerah', 'menerima', 'makan', 'bergizi', 'gratis']	['laksana', 'jauh', 'kata', 'efektif', 'banyak', 'sekolah', 'daerah', 'terima', 'makan', 'gizi', 'gratis']
3.	['salut', 'buat', 'pemerintah', 'program', 'makan', 'bergizi', 'gratis', 'benarbenar', 'membantu', 'anakanak', 'keluarga', 'kurang', 'mampu', 'semoga', 'terus', 'berlanjut']	['salut', 'buat', 'perintah', 'program', 'makan', 'gizi', 'gratis', 'benarbenar', 'bantu', 'anakanak', 'keluarga', 'kurang', 'mampu', 'moga', 'terus', 'lanjut']
4.	['tujuan', 'programnya', 'bagus', 'realisasinya', 'maksimal', 'banyak', 'laporan', 'makanan', 'sesuai', 'standar', 'pembagian', 'merata']	['tuju', 'program', 'bagus', 'realisasi', 'maksimal', 'banyak', 'lapor', 'makan', 'sesuai', 'standar', 'bagi', 'rata']
5.	['kalau', 'dilihat', 'dampaknya', 'program', 'cukup', 'berhasil', 'meningkatkan', 'gizi', 'anak', 'sekolah', 'bukti', 'nyatanya', 'angka', 'stunting', 'mulai', 'turun', 'beberapa', 'daerah']	['kalau', 'lihat', 'dampak', 'program', 'cukup', 'hasil', 'tingkat', 'gizi', 'anak', 'sekolah', 'bukti', 'nyata', 'angka', 'stunting', 'mulai', 'turun', 'beberapa', 'daerah']

### F. Hasil Akhir *Preprocessing*

Setelah melalui seluruh tahapan preprocessing mulai dari *cleansing*, *case folding*, *tokenizing*, *stopword removal*, hingga *stemming*, data teks yang semula tidak terstruktur kini telah berubah menjadi bentuk yang lebih bersih dan siap digunakan dalam tahap analisis berikutnya, yaitu perhitungan *Term Frequency–Inverse Document Frequency* (TF-IDF). Proses ini penting agar

sistem dapat mengenali setiap kata berdasarkan makna dasarnya dan mengurangi gangguan dari elemen-elemen yang tidak relevan.

Tabel 3. 10 Hasil Akhir

No	Data Mentah Sebelum <i>Preprocessing</i>	Data Akhir Setelah <i>Preprocessing</i>
1.	Program MBG ini menurut saya sangat berhasil! Anak- anak jadi semangat ke sekolah dan gizinya makin terjaga	['program', 'makan', 'gizi', 'gratis', 'sangat', 'hasil', 'anakanak', 'jadi', 'semangat', 'sekolah', 'gizi', 'makin', 'jaga']
2.	Pelaksanaannya masih jauh dari kata efektif. Banyak sekolah di daerah saya yang belum menerima makan bergizi gratis ini 🍌	['laksana', 'jauh', 'kata', 'efektif', 'banyak', 'sekolah', 'daerah', 'terima', 'makan', 'gizi', 'gratis']
3.	Salut buat pemerintah, program MBG benar-benar membantu anak-anak dari keluarga kurang mampu. Semoga terus berlanjut!	['salut', 'buat', 'perintah', 'program', 'makan', 'gizi', 'gratis', 'benarbenar', 'bantu', 'anakanak', 'keluarga', 'kurang', 'mampu', 'moga', 'terus', 'lanjut']
4.	Tujuan programnya bagus, tapi realisasinya belum maksimal. Masih banyak laporan makanan tidak sesuai standar dan pembagian tidak merata.	['tuju', 'program', 'bagus', 'realisasi', 'maksimal', 'banyak', 'lapor', 'makan', 'sesuai', 'standar', 'bagi', 'rata']
5.	Kalau dilihat dari dampaknya, program ini sudah cukup berhasil meningkatkan gizi anak sekolah. Bukti nyatanya angka stunting mulai turun di beberapa daerah.	['kalau', 'lihat', 'dampak', 'program', 'cukup', 'hasil', 'tingkat', 'gizi', 'anak', 'sekolah', 'bukti', 'nyata', 'angka', 'stunting', 'mulai', 'turun', 'beberapa', 'daerah']

Dari contoh di atas terlihat bahwa teks mentah awal yang mengandung tanda baca, emotikon, huruf kapital, dan kata tidak baku berhasil diubah menjadi daftar kata dasar yang bersih serta konsisten. Kata berimbuhan seperti “*bergizi*” berubah menjadi “*gizi*”, “*terjaga*” menjadi “*jaga*”, dan kata sambung seperti “*dan*” dihapus karena tidak memberikan makna signifikan terhadap konteks analisis.

Hasil akhir ini merupakan representasi kata yang siap untuk proses *feature extraction* menggunakan metode TF-IDF pada tahap selanjutnya. Dengan demikian, setiap kata dalam korpus kini dapat dihitung bobot

kemunculannya secara matematis sehingga dapat digunakan untuk analisis sentimen menggunakan algoritma SVM.

### 3.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Setelah melalui tahap *preprocessing*, data teks komentar masyarakat mengenai Program MBG menghasilkan kumpulan kata yang telah bersih dan siap diolah lebih lanjut. Langkah berikutnya adalah melakukan pembobotan kata (*term weighting*). Pembobotan kata (*weighting word*) merupakan teknik yang digunakan untuk memberikan nilai numerik pada kata-kata dalam sebuah dokumen, sehingga dapat menggambarkan tingkat kepentingan kata tersebut dan mempersiapkannya untuk proses analisis lanjutan. Salah satu metode yang paling banyak digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). Metode ini mengombinasikan frekuensi kemunculan kata pada suatu dokumen (TF) dengan tingkat kelangkaannya di keseluruhan kumpulan dokumen (IDF), sehingga mampu memberikan bobot yang lebih tepat pada kata-kata yang relevan. Dalam konteks penelitian ini, pembobotan kata bertujuan untuk memberikan penilaian terhadap tingkat kepentingan setiap kata dalam menentukan sentimen positif atau negatif terkait tingkat keberhasilan program MBG.

Setelah membobotkan setiap kata dalam dokumen, setiap kata di seluruh dokumen kini memiliki bobot yang sesuai. Berikut tahapan perhitungannya dijelaskan secara rinci. Perhitungan manual akan menggunakan delapan data sebagai contoh, yang dapat dilihat pada Tabel 3.11.

Tabel 3. 11 Data Sampel

Kode Dokumen	<i>Tweet Sebelum Preprocessing</i>	<i>Tweet Setelah Preprocessing</i>	Sentimen
D1	Program makan bergizi gratis ini benar-benar berhasil meningkatkan semangat belajar anak-anak.	program makan gizi gratis benar hasil tingkat semangat ajar anak	Positif
D2	Banyak daerah belum merasakan manfaatnya, pelaksanaannya belum efektif.	banyak daerah rasa manfaat laksana efektif	Negatif
D3	Salut untuk pemerintah, program ini terbukti membantu siswa dari keluarga kurang mampu.	salut perintah program bukti bantu siswa keluarga kurang mampu	Positif
D4	Program ini bagus tapi distribusinya masih lambat dan belum merata di semua sekolah.	program bagus distribusi lambat rata sekolah	Negatif
D5	Anak-anak sekarang lebih semangat ke sekolah berkat makan siang gratis yang bergizi.	anak sekarang semangat sekolah berkat makan siang gratis gizi	Positif
D6	Harusnya pemerintah lebih awas, karena masih banyak laporan program ini tidak berjalan baik di lapangan.	harus perintah awas banyak lapor program jalan baik lapang	Negatif
D7	Makan bergizi gratis jadi solusi yang tepat untuk atasi masalah gizi anak sekolah.	makan gizi gratis jadi solusi tepat atasi masalah gizi anak sekolah	Positif
D8	Tujuannya bagus, tapi hasilnya belum sesuai harapan, masih kurang maksimal di beberapa daerah.	tuju bagus hasil sesuai harap kurang maksimal daerah	Negatif

Dari keseluruhan *tweet*, peneliti mengidentifikasi kata-kata utama (term) yang paling sering muncul dan relevan dengan *aspek tingkat keberhasilan program*, seperti: ajar, efektif, bantu, distribusi, berkat, awas, anak, harap.

Perhitungan TF-IDF dilakukan melalui tiga tahap, yaitu menghitung *Term Frequency* (TF), *Inverse Document Frequency* (IDF), dan kemudian mengalikan keduanya untuk memperoleh nilai akhir TF-IDF. Perhitungan tersebut akan menggunakan persamaan 2.1 hingga 2.3. Berikut tahapan perhitungannya dijelaskan secara rinci.

a. Perhitungan *Term Frequency* (TF)



*Term Frequency* (TF) digunakan untuk menghitung seberapa sering suatu kata muncul dalam dokumen. Rumusnya ditunjukkan pada Persamaan (2.1): Sebagai contoh :

- Pada dokumen D1 terdapat 10 kata. Kata *ajar* muncul 1 kali.

Maka,

$$TF(ajar, D_1) = \frac{1}{10} = 0,100$$

- Pada dokumen D2 terdapat 6 kata, dan kata *efektif* muncul 1 kali.

Maka,

$$TF(efektif, D_2) = \frac{1}{6} = 0,167$$

- Pada dokumen D7 terdapat 11 kata, dan kata *anak* muncul 1 kali.

Maka,

$$TF(anak, D_7) = \frac{1}{11} = 0,090$$

Hasil perhitungan lengkap ditunjukkan pada Tabel 3.12 berikut.

Tabel 3. 12 Hasil TF

<i>Term</i>	D1	D2	D3	D4	D5	D6	D7	D8
ajar	0.100							
efektif		0,167						
bantu			0.100					
distribusi				0.100				
berkat					0.090			
awas						0.100		
anak							0.090	
harap								0.083

b. Perhitungan *Inverse Document Frequency* (IDF)

IDF digunakan untuk mengukur seberapa jarang suatu kata muncul di seluruh dokumen. Rumusnya ditunjukkan pada Persamaan (2.2): 1 dokumen, Sebagai contoh:

- Kata “anak” : muncul di D1, D5, dan D7, maka  $n_t = 3$ .
- Kata “ajar”, “efektif”, “bantu”, “distribusi”, “berkat”, “awas”, “harap” : masing-masing hanya muncul di 1 dokumen, maka  $n_t = 1$ .

Contoh perhitungan:

1. Untuk kata “ajar” ( $n_t = 1$ )

$$IDF(ajar, D) = \log_{10} \left( \frac{8}{1} \right) = \log_{10}(8) = 0,903$$

2. Untuk kata “anak” ( $n_t = 3$ )

$$IDF(anak, D) = \log_{10} \left( \frac{8}{3} \right) = \log_{10}(2,666) = 0,426$$

Hasil IDF untuk semua kata ditunjukkan pada Tabel 3.13.

Tabel 3. 13 Hasil IDF

<b>Term (t)</b>	<b>Dokumen Mengandung Kata (<math>n_t</math>)</b>	<b>Total Dokumen (N)</b>	<b>Logaritma</b>	<b>Nilai IDF</b>
ajar	1	8	$\log_{10} (8/1)$	0,903
efektif	1	8	$\log_{10} (8/1)$	0,903
bantu	1	8	$\log_{10} (8/1)$	0,903
distribusi	1	8	$\log_{10} (8/1)$	0,903
berkat	1	8	$\log_{10} (8/1)$	0,903
awas	1	8	$\log_{10} (8/1)$	0,903
anak	3	8	$\log_{10} (8/3)$	0,426
harap	1	8	$\log_{10} (8/1)$	0,903

#### c. Perhitungan TF–IDF

Nilai TF–IDF diperoleh dengan mengalikan hasil TF dan IDF sesuai Persamaan (2.3):

- Untuk kata *ajar* pada dokumen D1:

$$TF(\text{ajar}, D1) = 0,100, IDF(\text{ajar}) = 0,903$$

Maka:

$$W = 0,100 \times 0,903 = 0,0903 \approx 0,090$$

- Untuk kata *efektif* pada D2:

$$TF(\text{efektif}, D2) = 0,167, IDF(\text{efektif}) = 0,903$$

Maka:

$$W = 0,167 \times 0,903 = 0,1508 \approx 0,151$$

Hasil perhitungan TF-IDF ditunjukkan pada Tabel 3.14:

Tabel 3. 14 Hasil Perhitungan Pembobotan TF-IDF

<i>Term</i>	<b>Dokumen</b>	<b>TF</b>	<b>IDF</b>	<b>TF × IDF (TF-ID)</b>
ajar	D1	0,100	0,903	0,090
efektif	D2	0,167	0,903	0,151
bantu	D3	0,100	0,903	0,090
distribusi	D4	0,100	0,903	0,090
berkat	D5	0,090	0,903	0,081
awas	D6	0,100	0,903	0,090
anak	D7	0,090	0,426	0,038
harap	D8	0,083	0,903	0,075

Berdasarkan hasil perhitungan manual pada pada Tabel 3.14 TF-IDF di atas, dapat dilihat bahwa setiap kata (*term*) memiliki nilai bobot yang berbeda sesuai dengan tingkat kemunculannya pada dokumen (TF) serta kelangkaannya di seluruh kumpulan data (IDF). Nilai TF-IDF tertinggi diperoleh oleh kata “efektif” dengan bobot 0,151 yang berasal dari dokumen D2. Hal ini menunjukkan bahwa kata tersebut memiliki tingkat kepentingan yang lebih dominan dibanding term lainnya dalam merepresentasikan isi dokumen, karena frekuensinya relatif lebih tinggi pada D2 dan didukung oleh nilai IDF yang besar (0,903).

Sementara itu, kata “berkat” dan “harap” memiliki bobot TF-IDF yang lebih rendah, masing-masing sebesar 0,081 dan 0,075. Nilai ini menggambarkan bahwa kedua term tersebut muncul dengan frekuensi yang lebih kecil pada dokumen asalnya dibanding term dengan bobot lebih tinggi, sehingga kontribusinya terhadap pembobotan akhir juga lebih rendah. Adapun kata “anak” memiliki bobot TF-IDF paling kecil yaitu 0,038, karena selain nilai TF yang rendah (0,090), term ini juga memiliki nilai IDF yang lebih kecil (0,426) yang menandakan kata tersebut lebih umum muncul pada kumpulan dokumen sehingga daya pembeda antar dokumen menjadi lebih lemah.

Nilai-nilai pada tabel tersebut menunjukkan bahwa kata dengan bobot TF-IDF lebih tinggi memiliki peranan lebih besar dalam merepresentasikan karakteristik dokumen dan membedakannya dari dokumen lain. Dengan demikian, hasil pembobotan TF-IDF ini dapat dijadikan dasar untuk tahap analisis berikutnya, misalnya sebagai fitur dalam proses klasifikasi menggunakan metode *Support Vector Machine* (SVM) untuk mengelompokkan data ke dalam kategori tertentu secara lebih akurat.

### 3.5 *Support Vector Machine*

SVM merupakan salah satu algoritma pembelajaran mesin yang banyak digunakan, khususnya untuk permasalahan klasifikasi biner. Prinsip kerjanya 44imana mencari *hyperplane* terbaik yang dapat memisahkan dua kelas dengan margin pemisahan terbesar. Titik-titik data yang berada paling dekat dengan *hyperplane*, dikenal sebagai *support vectors*, berperan penting dalam menentukan posisi *hyperplane* tersebut (Wulandari & Anubhakti, 2021). Selain itu, SVM juga

mampu menangani data yang tidak dapat dipisahkan secara linear dengan memanfaatkan *kernel trick*, yaitu metode yang memetakan data ke ruang dimensi lebih tinggi sehingga pemisahan linear menjadi lebih memungkinkan (Pohan et al., 2022).

Proses pelatihan dan pengujian model *Support Vector Machine* terdiri dari beberapa langkah utama yang saling terhubung. Pertama, dimulai dengan data latih Hasil TF-IDF, di mana data teks diubah menjadi vektor fitur menggunakan metode TF-IDF untuk mempersiapkan data latih. Selanjutnya adalah Inisialisasi Parameter *Support Vector Machine*, di mana parameter seperti nilai *degree*, toleransi, pemilihan kernel dan parameter  $C$  diatur.

Tahap berikutnya adalah proses pelatihan yang berlangsung secara iteratif dengan memeriksa setiap data. Sistem akan mengevaluasi apakah nilai  $\alpha$  pada data sudah sesuai, sekaligus menghitung tingkat kesalahan prediksi. Jika syarat belum terpenuhi, maka dilanjutkan ke langkah berikutnya. Apabila syarat terpenuhi, algoritma akan menghitung batas pembaruan untuk nilai  $\alpha$ .

Ketika pembaruan dilakukan, algoritma memastikan nilai  $\alpha$  tetap berada dalam rentang yang ditentukan. Jika perubahan cukup besar, nilai bias turut diperbarui agar model menghasilkan prediksi yang lebih tepat. Jika tidak, sistem beralih ke data selanjutnya. Proses ini berlangsung berulang hingga tidak terjadi perubahan berarti atau jumlah iterasi maksimum tercapai. Dengan pendekatan ini, *Sequential Minimal Optimization* dapat menyesuaikan nilai  $\alpha$  secara efisien, sehingga SVM mampu menemukan *hyperplane optimal* dengan lebih cepat dan akurat.

Setelah dilakukannya perhitungan TF-IDF, langkah berikutnya adalah melakukan perhitungan menggunakan algoritma SVM. Pada proses ini, data diklasifikasikan menjadi 2 tahap yakni proses *training* dan proses *testing*. Proses *training* merupakan proses dimana data digunakan untuk melatih model agar dapat mempelajari pola. Sedangkan proses *testing* digunakan untuk menguji seberapa baik model yang telah dilatih dapat memprediksi label yang baru.

### 3.5.1 Proses Pelatihan

Dalam proses pelatihan, model SVM mencari nilai  $w$  (bobot) dan  $b$  (bias) yang optimal. Berikut adalah ilustrasi perhitungan manual pelatihan SVM menggunakan metode *Sequential Minimal Optimization* (SMO) berdasarkan data sampel dari Tabel 3.14.

#### 1. Persiapan Data Latih

Untuk menyederhanakan perhitungan dimensi tinggi, diambil dua sampel data dari Tabel 3.14 sebagai representasi *Support Vectors* untuk dua kelas sentimen yang berbeda:

Tabel 3. 15 Contoh Data *Training*

Dokumen	Vektor Fitur ( $x_i$ )	Label Kelas ( $y_i$ )
D1	0.090	+1
D2	0.151	-1

Inisialisasi Parameter :

Parameter Regularisasi ( $C$ ) : Ditetapkan sebesar 1. Pemilihan nilai ini didasarkan pada nilai default standar pada kerangka kerja *machine learning* untuk memberikan keseimbangan moderat pada batasan *Soft Margin*.

## 2. Penerapan Fungsi Kernel

Sesuai dengan persamaan 2.7, langkah pertama adalah menghitung *dot product* antar data menggunakan Kernel Linear untuk memetakan hubungan antar fitur. Berdasarkan data pada Tabel 3.15, diketahui nilai fitur untuk masing-masing data adalah:

- $x_1 = 0,090$  (positif)
- $x_2 = 0,151$  (negatif)

Maka, perhitungan nilai kedekatan antar data (Kernel) adalah sebagai berikut:

- Kedekatan Data 1 dengan dirinya sendiri ( $K_{11}$ ) :

$$K(x_1, x_1) = x_1 \times x_1 = 0,090 \times 0,090 = 0,0081$$

- Kedekatan Data 2 dengan dirinya sendiri ( $K_{22}$ ) :

$$K(x_2, x_2) = x_2 \times x_2 = 0,151 \times 0,151 = 0,0228$$

- Kedekatan antara Data 1 dan Data 2 ( $K_{12}$ ) :

$$K(x_1, x_2) = x_1 \times x_2 = 0,090 \times 0,151 = 0,0136$$

## 3. Optimasi Nilai Alpha ( $\alpha$ ) Untuk menyelesaikan masalah optimasi yang didefinisikan pada Bab 2, digunakan metode *Sequential Minimal Optimization* (SMO) yang memperbarui nilai $\alpha$ secara iteratif.

Proses ini terdiri dari tiga tahapan perhitungan sebagai berikut:

- a. Hitung error prediksi ( $E_i$ )

*Error* ( $E_i$ ) adalah selisih antara prediksi model saat ini dengan label asli.

Rumusnya dapat digunakan dengan persamaan (3.1) berikut:

$$E_i = f(x_i) - y_i \quad (3.1)$$

Keterangan:

- $E_i$  : Nilai error prediksi untuk data ke- $i$
- $f(x_i)$  : Hasil prediksi fungsi keputusan saat ini (karena iterasi awal bobot masih 0, maka  $f(x) = 0$ ).
- $y_i$  : Label kelas asli data ke- $i$

Maka untuk perhitungannya:

- Untuk Data 1 ( $x_1, y_1 = +1$ ):  

$$E_1 = 0 - 1 = -1$$
- Untuk Data 2 ( $x_2, y_2 = -1$ ):  

$$E_2 = 0 - (-1) = +1$$

b. Hitung *learning rate* ( $\eta$ )

*Learning rate* ( $\eta$ ) dalam konteks SMO dihitung berdasarkan nilai

Kernel untuk menentukan langkah perubahan (kurvatur) dari fungsi

objektif. Rumusnya dinyatakan pada persamaan (3.2):

$$\eta = 2K_{12} - K_{11} - K_{22} \quad (3.2)$$

Keterangan:

- $\eta$  : Parameter langkah optimasi (*learning rate*).
- $K_{12}$  : Nilai kernel antara data 1 dan data 2 (0,0136).
- $K_{11}$  : Nilai kernel data 1 dengan dirinya sendiri (0,0081).
- $K_{22}$  : Nilai kernel data 2 dengan dirinya sendiri (0,0228).

Maka untuk perhitungannya:

$$\eta = 2(0,0136) - 0,0081 - 0,0228$$

$$\eta = 0,0272 - 0,0081 - 0,0228$$

$$\eta = -0,0037$$

c. Memperbarui Nilai Alpha ( $\alpha$ )



Nilai  $\alpha$  diperbarui untuk meminimalkan *error* dengan meperhitungkan *learning rate*. Rumusnya dinyatakan pada persamaan (3.3):

$$\alpha_{\text{baru}} = \alpha_{\text{lama}} - \frac{y_2(E_1 - E_2)}{\eta} \quad (3.2)$$

Keterangan:

- $\alpha_{\text{baru}}$  : Nilai Lagrange Multiplier setelah update.
- $\alpha_{\text{lama}}$  : Nilai awal (saat inisialisasi adalah 0).
- $y_2$  : Label kelas data kedua (-1).
- $E_1, E_2$  : Error prediksi yang dihitung sebelumnya.
- $\eta$  : *learning rate*

Maka untuk perhitungannya:

$$\alpha_{\text{baru}} = 0 - \frac{-1(-1 - 1)}{-0,0037}$$

$$\alpha_{\text{baru}} = 0 - \frac{-1(-2)}{-0,0037}$$

$$\alpha_{\text{baru}} = 0 - \frac{2}{-0,0037}$$

$$\alpha_{\text{baru}} = 0 - (-540,54)$$

$$\alpha_{\text{baru}} \approx 540,5$$

#### 4. Penerapan Batasan *Soft Margin*

Berdasarkan persamaan 2.10, optimasi ini memiliki batasan (*constraint*) bahwa nilai  $\alpha$  tidak boleh melebihi parameter regularisasi  $C$  ( $0 \leq \alpha \leq C$ ). Diketahui  $C = 1$ . Karena hasil perhitungan  $540,5 > 1$ , maka nilai  $\alpha$  harus dipangkas (*clipped*) menjadi nilai maksimumnya:

$$\alpha_1 = 1$$

$$\alpha_2 = 1$$

### 5. Menghitung Vektor Bobot (*Dual Form*)

Setelah nilai  $\alpha$  optimal ditemukan, vektor bobot ( $w$ ) dihitung menggunakan rumus representasi dual pada persamaan 2.6:

$$w = (\alpha_1 \cdot y_1 \cdot x_1) + (\alpha_2 \cdot y_2 \cdot x_2)$$

$$w = (1 \cdot 1 \cdot 0,090) + (1 \cdot -1 \cdot 0,151)$$

$$w = 0,090 - 0,151$$

$$w = -0,061$$

### 6. Menghitung Bias ( $b$ )

Bias dihitung untuk menyeimbangkan posisi *hyperplane*. Berdasarkan rata-rata selisih label dengan prediksi bobot terhadap *support vector*, didapatkan nilai bias:

$$b \approx 0,0075$$

Dengan demikian, *hyperplane* yang terbentuk sesuai Persamaan 2.4 adalah:

$$-0,061x + 0,0075 = 0$$

## 3.5.2 Proses Pengujian

Pada tahap ini, model yang telah dilatih dengan persamaan *hyperplane*  $-0,061x + 0,0075 = 0$  akan diuji menggunakan data baru yang tidak dilibatkan dalam proses pelatihan (*unseen data*).

1. Data Uji : Untuk percobaan, akan menggunakan sisa data dari 8 sampel, yaitu D7.

Tabel 3. 16 Contoh Data Testing

Kata	Bobot TF-IDF ( $x_{baru}$ )	Label Sebenarnya
anak	0,038	Positif

2. Fungsi Keputusan : Untuk menentukan kelas data baru, digunakan fungsi keputusan (*Sign Function*) sesuai dengan persamaan 2.11 serta aturan keputusannya:

- Jika hasil  $\geq 0$ , maka diklasifikasi sebagai kelas positif (+1).
- Jika hasil  $< 0$ , maka diklasifikasi sebagai kelas negatif (-1).

3. Perhitungan Prediksi, memasukkan nilai bobot model dan nilai data uji ke dalam rumus persamaan 2.11:

$$f(x) = \text{sign}((-0,061 \times 0,038) + 0,0075)$$

- Langkah 1 (Perkalian):

$$-0,061 \times 0,038 = -0,002318 \approx -0,0023$$

- Langkah 2 (Penjumlahan dengan bias):

$$-0,0023 + 0,0075 = +0,0052$$

- Langkah 3 (Penerapan Fungsi *Sign*):

$$f(x) = \text{sign}(+0,0052)$$

4. Kesimpulan Hasil Uji

Hasil akhir perhitungannya adalah +0,0052 (Bilangan Positif).

- Karena nilainya positif ( $>0$ ), maka sistem memprediksi D7 sebagai sentimen positif.

Tabel 3. 17 Hasil Perhitungan Prediksi Data Testing

Dokumen	Label Asli	Nilai Prediksi	Hasil Klasifikasi
D7	Positif	+0,0052	Positif

Hasil yang ditampilkan dalam Tabel 3.17 menunjukkan bahwa dokumen D7 diprediksi dengan benar sebagai "Positif".

- Verifikasi : Label asli Dokumen D7 adalah Positif.

Tabel 3. 18 Hasil Contoh Klasifikasi

Dokumen	Label Asli	Hasil Klasifikasi
D7	Positif	Positif

Berdasarkan Tabel 3.18, ditampilkan perbandingan antara fakta sentimen sebenarnya (*ground truth*) dengan keputusan yang dihasilkan oleh sistem. Pada kolom "Label Asli", Dokumen 7 (D7) diketahui memiliki sentimen Positif. Hal ini selaras dengan kolom "Hasil Klasifikasi" yang juga menunjukkan status Positif. Kesamaan hasil ini diperoleh karena nilai fungsi keputusan dari perhitungan manual sebelumnya bernilai positif (+0,0052), yang memenuhi aturan klasifikasi SVM di mana  $f(x) \geq 0$ . Dengan demikian, dapat disimpulkan bahwa model *hyperplane* yang terbentuk dari data latih telah valid dan mampu memprediksi kelas data baru dengan akurat sesuai dengan pola bobot kata yang dipelajarinya.

### 3.6 Evaluasi Pengujian

Langkah berikutnya adalah melakukan evaluasi pengujian untuk mengukur kinerja sistem. Salah satu cara untuk mengukur kinerja ini adalah dengan menggunakan *confusion matrix*. Melalui *confusion matrix*, performa sistem dievaluasi berdasarkan empat parameter: akurasi, presisi, *recall*, dan *f1-score*. Akurasi menggambarkan sejauh mana sistem mampu mengklasifikasikan ulasan dengan benar (Pratiwi, 2020).

Tabel 3. 19 Contoh *Confusion Matrix* 2 kelas

Actual Class	Predicted Class	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Tabel 3.19 menampilkan parameter yang digunakan dalam pengujian menggunakan *confusion matrix*. Nilai *True Positive* (TP) adalah jumlah ulasan dengan nilai positif yang diprediksi secara benar sebagai positif. Nilai *True Negative* (TN) adalah jumlah ulasan dengan nilai negatif yang diprediksi secara benar sebagai negatif. Nilai *False Positive* (FP) adalah jumlah ulasan dengan nilai negatif yang diprediksi sebagai positif. Nilai *False Negative* (FN) adalah jumlah ulasan dengan nilai positif yang diprediksi sebagai negatif. Untuk memahami penggunaannya dalam evaluasi model, dapat dilihat pada metrik kinerja seperti akurasi, presisi, *recall*, dan *F1-score* yang menjelaskan kinerja model.

Keempat metrik evaluasi ini yaitu Akurasi, Presisi, *Recall*, dan *F1-Score* , akan digunakan sebagai tolok ukur utama untuk menilai seberapa baik performa sistem dalam mengklasifikasikan sentimen publik. Nilai dari metrik-metrik inilah yang akan dihitung dan dibandingkan untuk setiap model yang dihasilkan dari skenario pengujian. Skenario pengujian tersebut, yang dirancang untuk menemukan parameter algoritma *Support Vector Machine* paling optimal yang akan dijelaskan di sub-bab selanjutnya

### 3.7 Skenario Pengujian

Tahap berikutnya adalah skenario pengujian yang bertujuan untuk menilai performa sistem secara komprehensif. Pengujian ini dirancang untuk menemukan kombinasi paling optimal dari tiga variabel utama: rasio pembagian data, jenis *kernel* SVM, dan *hyperparameter* yang digunakan. Setiap variabel akan diuraikan di bawah ini.

#### 1. Rasio Pembagian Data (*Split Ratio*)

Variabel pertama yang akan diuji adalah rasio pembagian data (*split ratio*). Pengujian ini dilakukan untuk melihat seberapa sensitif model SVM terhadap jumlah data latih yang diberikan. Tiga skema pembagian akan digunakan pada data yang telah dilabeli, seperti yang dirinci pada Tabel 3.20:

Tabel 3. 20 Skenario Rasio Pembagian Data

Skenario Rasio	Data Latih	Data Uji
1	70%	30%
2	80%	20%
3	60%	40%

Setiap rasio pembagian data ini akan diuji menggunakan tiga jenis *kernel* SVM yang berbeda. Pemilihan *kernel* merupakan variabel kedua yang akan dievaluasi.

## 2. Kernel SVM

Variabel kedua adalah pemilihan fungsi *kernel* SVM. Seperti yang telah dijelaskan pada Bab II, *kernel* berfungsi untuk memetakan data ke dimensi yang berbeda untuk menemukan batas keputusan yang optimal. Penelitian ini akan membandingkan tiga *kernel* yang paling umum digunakan, seperti pada Tabel 3.21:

Tabel 3. 21 Jenis Kernel SVM yang Diuji

No.	Kernel SVM	Karakteristik
1	Linear	Untuk data yang diasumsikan dapat dipisahkan secara linear.
2	Polinomial	Untuk batas keputusan melengkung (non-linear).
3	RBF	Untuk batas keputusan yang sangat kompleks (non-linear).

Pemilihan *kernel* ini sangat penting karena performanya sangat bergantung pada pengaturan *hyperparameter* yang tepat. Proses optimasi *hyperparameter* ini adalah variabel ketiga dalam skenario pengujian.

### 3. *Hyperparameter Tuning*

Variabel ketiga adalah optimasi *hyperparameter*. Setiap *kernel* memiliki setelan parameter yang harus di-*tuning* untuk mencapai kinerja puncak dan menghindari *overfitting*. Proses ini akan menguji beberapa nilai spesifik untuk setiap parameter, seperti yang dirinci pada Tabel 3.22:

Tabel 3. 22 Nilai Hyperparameter yang Diuji

Hyperparameter	Simbol	Nilai yang Diuji	Kernel Terkait
Regularisasi	$C$	[0.1, 1, 10, 100]	Linear, Polinomial, RBF
Koefisien Kernel	$\gamma$	[0.1, 1, 10]	Polinomial, RBF
Derajat Polinomial	$d$	[2, 3, 4]	Polinomial

Setelah ketiga variabel utama yang disebutkan sebelumnya yaitu, rasio pembagian data, jenis *kernel*, dan *hyperparameter* telah dijelaskan, keseluruhan desain pengujian dapat diringkaskan. Sembilan skenario pengujian utama, yang merupakan kombinasi dari semua variabel ini, dirangkum secara lengkap pada Tabel 3.23 berikut:

Tabel 3. 23 Skenario Pengujian Gabungan

Skenario Model	Rasio Data Latih : Uji	Kernel SVM	Hyperparameter
1		Linear	C: [0.1, 1, 10, 100]
2		Polinomial	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10] Degree: [2, 3, 4]

3	70% : 30%	RBF	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10]
4	80% : 20%	Linear	C: [0.1, 1, 10, 100]
5		Polinomial	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10] Degree: [2, 3, 4]
6		RBF	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10]
7	60% : 40%	Linear	C: [0.1, 1, 10, 100]
8		Polinomial	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10] Degree: [2, 3, 4]
9		RBF	C: [0.1, 1, 10, 100] Gamma: [0.1, 1, 10]

Pada tabel 3.23 membahas secara keseluruhan skenario pengujian untuk mengevaluasi performa model SVM secara komprehensif. Total pengujian akan mencakup sembilan skenario utama, yang merupakan kombinasi dari 3 skema rasio pembagian data, 3 jenis kernel SVM, dan *hyperparameter* yang akan diuji.

Dalam setiap sembilan skenario tersebut, akan dilakukan proses *hyperparameter tuning* sesuai Tabel 3.22 untuk menemukan setelan parameter  $C$ ,  $\gamma$ , dan  $d$  yang paling optimal untuk *kernel* yang bersangkutan. Kinerja dari model terbaik pada setiap skenario akan dicatat, dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, dan kemudian dibandingkan satu sama lain. Model dengan kinerja tertinggi secara keseluruhan dari seluruh skenario pengujian ini akan dipilih sebagai model final yang paling direkomendasikan untuk penelitian ini. Seluruh hasil dari pelaksanaan skenario pengujian ini akan disajikan dan dianalisis secara mendalam pada bab berikutnya.



## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Pada bagian ini merupakan hasil dari implementasi penelitian yang telah dirancang pada bab sebelumnya. Pembahasan akan dimulai dari penjabaran data yang digunakan, hasil dari setiap tahapan *preprocessing* data, hingga hasil akhir dari pengujian model *Support Vector Machine* (SVM) berdasarkan skenario yang telah ditentukan. Seluruh hasil akan disajikan dalam bentuk tabel dan visualisasi untuk mempermudah analisis dan pembahasan.

#### **4.1 Tahapan Penelitian**

Tahapan penelitian merupakan proses implementasi dari alur metodologi yang telah dirancang. Tahapan ini mencakup pengumpulan data yang akan digunakan, proses *preprocessing* untuk membersihkan dan menstrukturkan data, hingga proses ekstraksi fitur sebagai *input* untuk model. Setiap langkah dalam tahapan ini didokumentasikan untuk menunjukkan transformasi data dari bentuk mentah hingga siap diolah.

##### **4.1.1 Pengumpulan Data**

Seperti yang sudah dijelaskan sebelumnya, bahwa data dalam penelitian ini dikumpulkan dari *media social* X menggunakan metode *crawling*. Topik yang diambil adalah opini masyarakat terkait “Makan Bergizi Gratis” sehingga mendapatkan data sebanyak 2.973 lalu dilakukan proses pelabelan data secara manual, di mana setiap *tweet* dibaca dan dikategorikan sebagai sentimen 'Positif'

(P) atau 'Negatif' (N) sehingga mendapatkan data yang sudah bersih sebanyak 599 data. Tampilan 5 data pertama dari dataset mentah yang telah dilabeli dapat dilihat pada gambar 4.1.

	full_text	label
0	@AnotherJakarta @ver_somnivora @CaptainPeBe @...	N
1	@pak_oke1 @LemonAndAtilla @M45Broo_ Di kabupat...	N
2	@LemonAndAtilla @M45Broo_ Tentu saja dong mini...	N
3	@Lurah_Funabashi @M45Broo_ Sebagai manusia hip...	P
4	@M45Broo_ Nyatane ndi program makan siang grat...	N

Gambar 4. 1 Data Mentah Setelah Dilabel Manual

### 4.1.2 Preprocessing

Data mentah yang ditampilkan pada gambar 4.1 tidak dapat langsung diolah oleh model *machine learning*. Data tersebut perlu melalui serangkaian tahapan *preprocessing* untuk membersihkan data dari *noise* (elemen tidak relevan) dan mengubahnya ke dalam format yang terstruktur dan seragam. Tahapan *preprocessing* yang dilakukan secara berurutan meliputi *Cleansing*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*.

#### 4.1.2.1. Cleansing

Tahap *Cleansing* bertujuan untuk membersihkan teks dari elemen-elemen yang tidak memiliki makna analitis dan dapat mengganggu proses pengolahan data. Proses ini mencakup penghapusan *mention* (contoh: @username), *hashtag* (contoh: #trending), URL (contoh: http://...), angka, dan seluruh tanda baca. Hasil dari proses *cleansing* dapat dilihat pada Gambar 4.2.

--- PROSES CLEANSING ---  
Tabel Perbandingan 'Sebelum' dan 'Sesudah' Cleansing (5 data pertama):

	full_text	text_cleansed
0	@AnotherJakarta @ver_somnivor @CaptainPeBe @...	Contoh kebijakan yang memengaruhi pencarian na...
1	@pak_oke1 @LemonAndAtilla @M45Broo_ Di kabupat...	Di kabupaten ku malah sampe sekarang ga ada se...
2	@LemonAndAtilla @M45Broo_ Tentu saja dong mini...	Tentu saja dong minimal ya ini salah satu alas...
3	@Lurah_Funabashi @M45Broo_ Sebagai manusia hip...	Sebagai manusia hipotetis yang memilih saya ti...
4	@M45Broo_ Nyatane ndi program makan siang grat...	Nyatane ndi program makan siang gratis mung on...

Gambar 4. 2 Hasil Proses *Cleansing*

Gambar 4.2 di atas menunjukkan perbandingan *tweet* sebelum dan sesudah melalui tahap *cleansing*. Pada *tweet* asli (Sebelum), dapat dilihat bahwa teks masih mengandung berbagai elemen *noise* seperti *mention* (diawali '@'), angka ('2025'), dan berbagai tanda baca (seperti ':', '&', '()', dan '.').

Dapat dilihat pada dataset yang pertama, datanya sebelum *dicleansing* [*@AnotherJakarta @ver\_somnivor @CaptainPeBe @hermiony* Contoh kebijakan yang memengaruhi pencarian nafkah: Program makan siang gratis Prabowo janjikan cipta 19 juta lapangan kerja tapi hingga Juli 2025 pengangguran naik & harga kebutuhan pokok melonjak (sumber: polling & laporan media). Bisa adaptasi tapi beban rakyat bertambah].

Setelah melalui proses *cleansing* (Sesudah) [Contoh kebijakan yang memengaruhi pencarian nafkah Program makan siang gratis Prabowo janjikan cipta juta lapangan kerja tapi hingga Juli pengangguran naik harga kebutuhan pokok melonjak sumber polling laporan media Bisa adaptasi tapi beban rakyat bertambah], seluruh elemen *noise* tersebut berhasil dihilangkan, menyisakan teks yang hanya terdiri dari kata-kata yang relevan untuk analisis.

#### 4.1.2.2. Case Folding

Tahap *Case Folding* adalah proses mengubah seluruh karakter huruf dalam teks menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk menyeragamkan kata, sehingga kata yang sama namun ditulis dengan kapitalisasi berbeda (misalnya, "Program", "program", dan "PROGRAM") akan dianggap sebagai satu kata yang sama oleh model. Hasil dari proses *case folding* dapat dilihat pada Gambar 4.3.

--- PROSES CASE FOLDING ---  
Tabel Perbandingan 'Sebelum' dan 'Sesudah' Case Folding (5 data pertama):

	text_cleansed	text_casefolded
0	Contoh kebijakan yang memengaruhi pencarian na...	contoh kebijakan yang memengaruhi pencarian na...
1	Di kabupaten ku malah sampe sekarang ga ada se...	di kabupaten ku malah sampe sekarang ga ada se...
2	Tentu saja dong minimal ya ini salah satu alas...	tentu saja dong minimal ya ini salah satu alas...
3	Sebagai manusia hipotetis yang memilih saya ti...	sebagai manusia hipotetis yang memilih saya ti...
4	Nyatane ndi program makan siang gratis mung on...	nyatane ndi program makan siang gratis mung on...

Gambar 4. 3 Hasil Proses *Case Folding*

Gambar 4.3 menampilkan hasil dari tahap *case folding* pada data yang telah dibersihkan. Pada tahap ini, seluruh karakter huruf kapital yang ada pada data (seperti 'C' pada "Contoh" atau 'P' pada "Program") diubah menjadi huruf kecil (*lowercase*).

Sebelum *Case Folding* (1 data pertama): [Contoh kebijakan yang memengaruhi pencarian nafkah Program makan siang gratis Prabowo janjikan cipta juta lapangan kerja tapi hingga Juli pengangguran naik harga kebutuhan pokok melonjak sumber polling laporan media Bisa adaptasi tapi beban rakyat bertambah].

Sesudah *Case Folding* (1 data pertama): [contoh kebijakan yang memengaruhi pencarian nafkah program makan siang gratis prabowo janjikan cipta juta lapangan kerja tapi hingga juli pengangguran naik harga kebutuhan pokok melonjak sumber polling laporan media bisa adaptasi tapi beban rakyat bertambah].  
Proses penyeragaman huruf ini sangat penting agar kata yang sama tidak dianggap

berbeda oleh sistem hanya karena perbedaan kapitalisasi saat pemrosesan data selanjutnya.

#### 4.1.2.3. *Tokenizing*

Tahap *Tokenizing* adalah proses memecah kalimat yang utuh (string) menjadi daftar kata-kata individual (disebut *token* atau *term*). Proses ini merupakan langkah fundamental untuk dapat menganalisis setiap kata dalam teks secara terpisah. Hasil dari proses *tokenizing* dapat dilihat pada gambar 4.4.

--- PROSES TOKENIZING ---  
Tabel Perbandingan 'Sebelum' dan 'Sesudah' Tokenizing (5 data pertama):

	text_casefolded	text_tokenized
0	contoh kebijakan yang memengaruhi pencarian na...	[contoh, kebijakan, yang, memengaruhi, pencari...
1	di kabupaten ku malah sampe sekarang ga ada se...	[di, kabupaten, ku, malah, sampe, sekarang, ga...
2	tentu saja dong minimal ya ini salah satu alas...	[tentu, saja, dong, minimal, ya, ini, salah, s...
3	sebagai manusia hipotetis yang memilih saya ti...	[sebagai, manusia, hipotetis, yang, memilih, s...
4	nyatane ndi program makan siang gratis mung on...	[nyatane, ndi, program, makan, siang, gratis, ...

Gambar 4. 4 Hasil Proses *Tokenizing*

Gambar 4.4 mengilustrasikan hasil dari proses *tokenizing*. Teks yang sebelumnya merupakan satu kalimat utuh (string) kini dipecah menjadi daftar kata-kata individual yang disebut *token*.

Sebelum *Tokenizing* (1 data pertama): [contoh kebijakan yang memengaruhi pencarian nafkah program makan siang gratis prabowo janjikan cipta juta lapangan kerja tapi hingga juli pengangguran naik harga kebutuhan pokok melonjak sumber polling laporan media bisa adaptasi tapi beban rakyat bertambah].

Sesudah *Tokenizing* (1 data pertama): ['contoh', 'kebijakan', 'yang', 'memengaruhi', 'pencarian', 'nafkah', 'program', 'makan', 'siang', 'gratis', 'prabowo', 'janjikan', 'cipta', 'juta', 'lapangan', 'kerja', 'tapi', 'hingga', 'juli', 'pengangguran', 'naik',

'harga', 'kebutuhan', 'pokok', 'melonjak', 'sumber', 'polling', 'laporan', 'media', 'bisa', 'adaptasi', 'tapi', 'beban', 'rakyat', 'bertambah']. Setiap kata dipisahkan berdasarkan spasi, mengubah format data dari string tunggal menjadi sebuah list (daftar) token yang memungkinkan setiap kata dianalisis secara terpisah pada tahap berikutnya.

#### 4.1.2.4. *Stopword Removal*

Tahap *Stopword Removal* adalah proses memfilter dan menghapus kata-kata umum (disebut *stopwords*) yang sering muncul namun tidak memiliki makna sentimen yang signifikan. Kata-kata ini seperti 'yang', 'di', 'ini', 'tapi', 'saya', dan sebagainya. Proses ini menggunakan *stoplist* (daftar *stopword*) bahasa Indonesia dari *library* Sastrawi untuk memastikan hanya kata-kata yang relevan dengan sentimen yang tersisa. Hasil dari proses *stopword removal* dapat dilihat pada gambar 4.5.

--- PROSES STOPWORD REMOVAL ---  
Tabel Perbandingan 'Sebelum' dan 'Setelah' Stopword Removal (5 data pertama):

	text_tokenized	text_stopwords
0	[contoh, kebijakan, yang, memengaruhi, pencari...	[contoh, kebijakan, memengaruhi, pencarian, na...
1	[di, kabupaten, ku, malah, sampe, sekarang, ga...	[kabupaten, ku, malah, sampe, sekarang, ga, se...
2	[tentu, saja, dong, minimal, ya, ini, salah, s...	[saja, dong, minimal, ini, salah, satu, alasan...
3	[sebagai, manusia, hipotetis, yang, memilih, s...	[manusia, hipotetis, memilih, tidak, menyesal,...
4	[nyatane, ndi, program, makan, siang, gratis, ...	[nyatane, ndi, program, makan, siang, gratis, ...

Gambar 4. 5 Hasil Proses *Stopword Removal*

Gambar 4.5 menunjukkan hasil dari penghapusan *stopword*. Kata-kata umum yang tidak memiliki makna sentimen signifikan (disebut *stopwords*), seperti 'yang', 'tapi', 'hingga', dan 'bisa', telah berhasil dihapus dari daftar *token*.

Sebelum *Stopword Removal* (1 data pertama): ['contoh', 'kebijakan', 'yang', 'memengaruhi', 'pencarian', 'nafkah', 'program', 'makan', 'siang', 'gratis', 'prabowo',

'janjikan', 'cipta', 'juta', 'lapangan', 'kerja', 'tapi', 'hingga', 'juli', 'pengangguran', 'naik', 'harga', 'kebutuhan', 'pokok', 'melonjak', 'sumber', 'polling', 'laporan', 'media', 'bisa', 'adaptasi', 'tapi', 'beban', 'rakyat', 'bertambah'].

Sesudah Stopword Removal (1 data pertama): ['contoh', 'kebijakan', 'memengaruhi', 'pencarian', 'nafkah', 'program', 'makan', 'siang', 'gratis', 'prabowo', 'janjikan', 'cipta', 'juta', 'lapangan', 'kerja', 'hingga', 'juli', 'pengangguran', 'naik', 'harga', 'kebutuhan', 'pokok', 'melonjak', 'sumber', 'polling', 'laporan', 'media', 'adaptasi', 'beban', 'rakyat', 'bertambah']. Proses ini secara efektif mengurangi jumlah kata yang tidak relevan dalam data dan memfokuskan analisis pada kata-kata yang membawa makna penting (kata kunci) dari opini masyarakat.

#### 4.1.2.5. Stemming

Tahap *Stemming* adalah proses mengubah setiap kata yang memiliki imbuhan ke dalam bentuk kata dasarnya (misalnya, "bergizi" atau "gizinya" menjadi "gizi"). Tujuan dari *stemming* adalah untuk mengurangi variasi dari kata yang sama dan mengelompokkannya menjadi satu makna dasar. Proses ini menggunakan *stemmer* bahasa Indonesia dari *library* Sastrawi. Hasil dari proses *stemming* dapat dilihat pada gambar 4.6.

```

** --- PROSES STEMMING ---
Memulai proses Stemming... (Ini mungkin memakan waktu beberapa menit)
Proses Stemming Selesai.
Tabel Perbandingan 'Sebelum' dan 'Sesudah' Stemming (5 data pertama):

```

	text_stopwords	text_stemmed
0	[contoh, kebijakan, memengaruhi, pencarian, na...]	[contoh, bijak, pengaruh, cari, nafkah, progra...]
1	[kabupaten, ku, malah, sampe, sekarang, ga, se...]	[kabupaten, ku, malah, sampe, sekarang, ga, se...]
2	[saja, dong, minimal, ini, salah, satu, alasan...]	[saja, dong, minimal, ini, salah, satu, alas, ...]
3	[manusia, hipotetis, memilih, tidak, menyesal,...]	[manusia, hipotetis, pilih, tidak, sesal, ting...]
4	[nyatane, ndi, program, makan, siang, gratis, ...]	[nyatane, ndi, program, makan, siang, gratis, ...]

Gambar 4. 6 Hasil Proses *Stemming*

Gambar 4.6 menampilkan hasil dari tahap *stemming*, yang merupakan tahap akhir dari *preprocessing* teks. Pada proses ini, setiap *token* yang memiliki imbuhan diubah menjadi kata dasarnya menggunakan *library* Sastrawi. Dapat dilihat bahwa kata-kata seperti 'memengaruhi' diubah menjadi 'pengaruh', 'pencarian' menjadi 'cari', 'melonjak' menjadi 'lonjak', dan 'bertambah' menjadi 'tambah'.

Sebelum Stemming (1 data pertama): ['contoh', 'kebijakan', 'memengaruhi', 'pencarian', 'nafkah', 'program', 'makan', 'siang', 'gratis', 'prabowo', 'janjikan', 'cipta', 'juta', 'lapangan', 'kerja', 'hingga', 'juli', 'pengangguran', 'naik', 'harga', 'kebutuhan', 'pokok', 'melonjak', 'sumber', 'polling', 'laporan', 'media', 'adaptasi', 'beban', 'rakyat', 'bertambah'].

Sesudah *Stemming* (1 data pertama): ['contoh', 'bijak', 'pengaruh', 'cari', 'nafkah', 'program', 'makan', 'siang', 'gratis', 'prabowo', 'janji', 'cipta', 'juta', 'lapang', 'kerja', 'hingga', 'juli', 'anggur', 'naik', 'harga', 'butuh', 'pokok', 'lonjak', 'sumber', 'polling', 'lapor', 'media', 'adaptasi', 'beban', 'rakyat', 'tambah']. Teks yang telah bersih dan sudah dalam bentuk kata dasar ini siap untuk digunakan pada tahap ekstraksi fitur TF-IDF.

#### 4.1.2.6. Finalisasi Data

Tahap terakhir dari *preprocessing* adalah menggabungkan kembali *token-token* yang telah bersih (hasil *stemming*) menjadi satu kalimat utuh. Teks akhir inilah yang siap digunakan untuk proses ekstraksi fitur TF-IDF pada tahap selanjutnya. Hasil akhir dari keseluruhan proses *preprocessing* dapat dilihat pada gambar 4.7.



--- HASIL AKHIR PREPROCESSING ---  
Tabel Perbandingan Data Asli vs Data Final (5 data pertama):

	full_text	text_final	label
0	@AnotherJakarta @ver_somniva @CaptainPeBe @...	contoh bijak pengaruh cari nafkah program maka...	N
1	@pak_oke1 @LemonAndAtilla @M45Broo_ Di kabupat...	kabupaten ku malah sampe sekarang ga sekolah d...	N
2	@LemonAndAtilla @M45Broo_ Tentu saja dong mini...	saja dong minimal ini salah satu alas paling m...	N
3	@Lurah_Funabashi @M45Broo_ Sebagai manusia hip...	manusia hipotetis pilih tidak sesal tingkat tu...	P
4	@M45Broo_ Nyatane ndi program makan siang grat...	nyatane ndi program makan siang gratis mung on...	N

Gambar 4. 7 Hasil Akhir Proses *Preprocessing*

Gambar 4.7 menyajikan perbandingan langsung antara data mentah asli (kolom `full_text`) dengan data yang telah melalui seluruh tahapan *preprocessing* (kolom `text_final`). Dapat diamati bahwa *tweet* asli yang sebelumnya masih mengandung *noise* seperti *mention* (`@username`), tanda baca, huruf kapital, dan kata-kata berimbuhan, kini telah berhasil ditransformasi menjadi sebuah *string* teks yang bersih dan terstruktur. Kumpulan data yang telah bersih dan seragam inilah yang akan menjadi *input* (masukan) utama untuk tahap selanjutnya, yaitu ekstraksi fitur dan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

#### 4.1.3 Ekstraksi Fitur TF-IDF

Setelah seluruh data teks dibersihkan melalui tahapan *preprocessing*, langkah selanjutnya adalah ekstraksi fitur. Proses ini bertujuan untuk mengubah data teks yang bersifat kualitatif menjadi data numerik (kuantitatif) yang dapat dipahami dan diolah oleh model *machine learning* SVM. Metode yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF).

Proses TF-IDF ini diterapkan pada seluruh 599 data *tweet* yang telah bersih. Dalam implementasinya, *vectorizer* TF-IDF diatur untuk membangun kosa kata

(*vocabulary*) dari seluruh data latih dan membatasi jumlah fitur (kata) yang paling sering muncul sebanyak 5.000 fitur (*max\_features=5000*) untuk menjaga efisiensi komputasi. Hasil dari proses ini adalah sebuah matriks TF-IDF dengan dimensi 599 baris (mewakili 599 *tweet*) dan (maksimal) 5.000 kolom (mewakili 5.000 fitur/kata unik). Gambar 4.8 akan menampilkan 10 baris data pertama dengan 12 fitur (kata) pertama yang dihasilkan oleh *vectorizer* TF-IDF.

Hasil Tabel TF-IDF (10 data pertama, 12 fitur pertama):

	aaaa	ababe	abab	abal	abdi	abel	abis	acara	ada	adahadeuhhh	adain	adaptasi
0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.235945
1	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
2	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.128645	0.0	0.0	0.000000
3	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
4	0.0	0.171221	0.0	0.171221	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
5	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
6	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
7	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
8	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000
9	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000

Gambar 4. 8 Hasil Ekstraksi Fitur TF-IDF

Gambar 4.8 di atas menunjukkan hasil dari proses ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF). Seperti yang ditunjukkan pada *output* pertama, 599 data *tweet* yang telah bersih kini telah berhasil ditransformasi menjadi sebuah matriks numerik dengan dimensi (599, 2536). Dimensi ini menunjukkan bahwa terdapat 599 baris (mewakili 599 *tweet*) dan 2536 kolom (mewakili 2536 kata unik/fitur) yang ditemukan dari seluruh kosa kata dalam data.

Gambar di bawahnya menampilkan 10 baris pertama (indeks 0-9) dan sebagian kolom (fitur) dari matriks tersebut. Setiap baris adalah representasi vektor dari satu *tweet*. Nilai di dalam sel, seperti 0,171221 adalah bobot TF-IDF yang telah dihitung, yang menandakan seberapa penting kata tersebut (contoh: 'abal') dalam dokumen yang bersangkutan. Nilai 0.0 menunjukkan bahwa kata pada kolom tersebut tidak ada dalam dokumen pada baris tersebut. Matriks numerik inilah yang

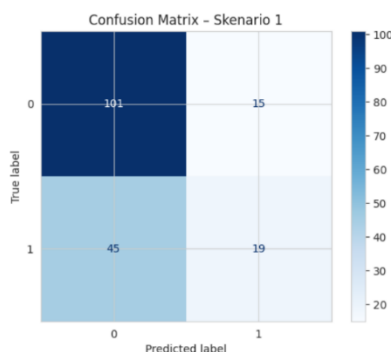
akan digunakan sebagai input fitur (X) yang siap dibagi menjadi data latih dan data uji untuk proses pelatihan dan pengujian model SVM.

## 4.2 Hasil Uji Coba

Pada bagian ini akan dijelaskan hasil dari sembilan skenario pengujian yang telah dirancang. Skenario-skenario ini diimplementasikan untuk menemukan kombinasi rasio data, jenis *kernel*, dan *hyperparameter* SVM yang paling optimal.

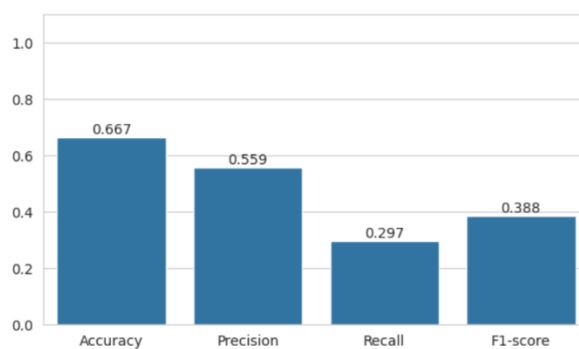
### 4.2.1 Skenario 1

Model pengujian pertama menggunakan rasio pembagian data 70:30, di mana sebanyak 419 data digunakan sebagai data latih dan 180 data sebagai data uji. Pada skenario ini, algoritma *Support Vector Machine* dilatih menggunakan Kernel Linear, dan proses tuning hyperparameter dilakukan untuk mencari nilai parameter C terbaik dari daftar kandidat [0.1, 1, 10, 100]. Berdasarkan hasil GridSearchCV, parameter terbaik yang diperoleh adalah  $C = 10$ , yang menghasilkan performa paling stabil pada data validasi. Model yang telah dituning ini kemudian diuji menggunakan 180 data uji. Hasil evaluasi visual berupa Confusion Matrix, metrik performa, dan grafik hyperparameter disajikan pada gambar-gambar berikut.



Gambar 4. 9 *Confusion Matrix* Skenario 1

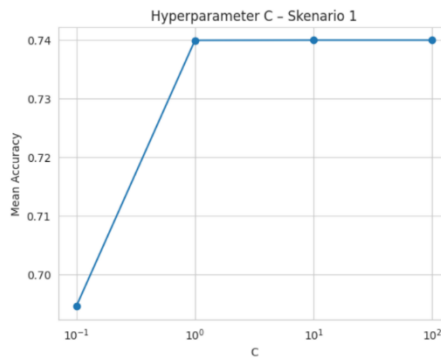
Gambar 4.9 menunjukkan *Confusion Matrix* untuk Skenario 1. Hasil ini memperlihatkan bagaimana model memprediksi dua kelas sentimen, yaitu Negatif (0) dan Positif (1). Model berhasil mengklasifikasikan 101 data Negatif sebagai Negatif (*True Negative*) dan 19 data Positif sebagai Positif (*True Positive*). Namun, terdapat 45 data Positif yang salah diprediksi sebagai Negatif (*False Negative*) dan 15 data Negatif yang salah diprediksi sebagai Positif (*False Positive*). Pola kesalahan ini mengindikasikan bahwa model masih kesulitan mengenali kelas Positif, yang kemungkinan disebabkan oleh distribusi data yang tidak seimbang, di mana kelas Negatif lebih dominan.



Gambar 4. 10 Metrik Evaluasi Skenario 1

Gambar 4.10 menampilkan empat metrik evaluasi utama, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Pada skenario ini, model mencapai Akurasi sebesar 66.7%, yang berarti sekitar dua pertiga dari total data uji berhasil diprediksi dengan benar. Nilai Precision sebesar 55.9% menunjukkan bahwa dari seluruh prediksi Positif, hanya 55.9% yang benar-benar berlabel Positif. Sementara itu, Recall yang rendah yaitu 29.7% mengindikasikan bahwa model hanya mampu menangkap sebagian kecil dari seluruh data Positif yang sebenarnya ada. Kombinasi precision dan recall menghasilkan F1-Score sebesar 38.8%, yang

menunjukkan bahwa performa model untuk mendeteksi kelas Positif masih kurang optimal.



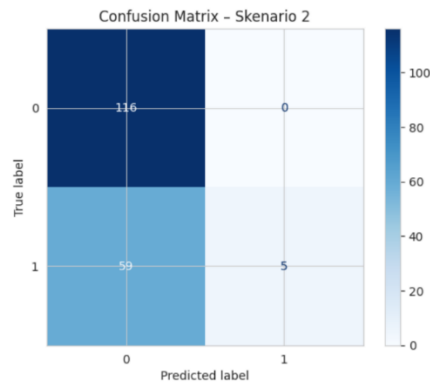
Gambar 4. 11 Visualisasi *Hyperparameter* Skenario 1

Gambar 4.11 menampilkan visualisasi proses tuning hyperparameter untuk Kernel Linear dengan parameter C. Grafik tersebut menunjukkan perubahan nilai akurasi rata-rata pada data validasi ketika nilai C dinaikkan. Terlihat bahwa akurasi meningkat tajam dari  $C = 0.1$  ke  $C = 1$ , dan mencapai nilai tertinggi pada  $C = 1$  hingga  $C = 100$ , dengan performa yang stabil pada kisaran tersebut. Hal ini menunjukkan bahwa nilai C yang besar memberikan model fleksibilitas margin yang lebih sesuai dengan karakteristik data, dan  $C = 10$  dipilih sebagai parameter terbaik untuk skenario ini.

#### 4.2.2 Skenario 2

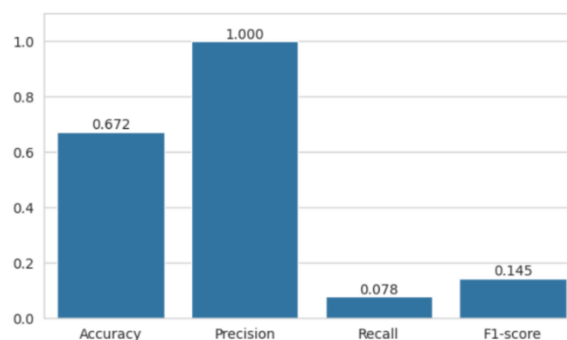
Model pengujian kedua menggunakan rasio pembagian data 70:30, yaitu 419 data latih dan 180 data uji, sama seperti skenario pertama. Perbedaannya terletak pada penggunaan Kernel Polynomial (degree = 2) untuk model SVM. Pada skenario ini, proses hyperparameter tuning dilakukan terhadap parameter  $C = [0.1, 1, 10]$  dan  $\gamma = [0.1, 1]$ .

Berdasarkan hasil GridSearch, diperoleh kombinasi parameter terbaik yaitu  $C = 10$ ,  $\gamma = 1$ , dan  $\text{degree} = 2$ . Model yang telah dituning kemudian diuji menggunakan 180 data uji, dan visualisasi kinerjanya ditampilkan pada gambar berikut.



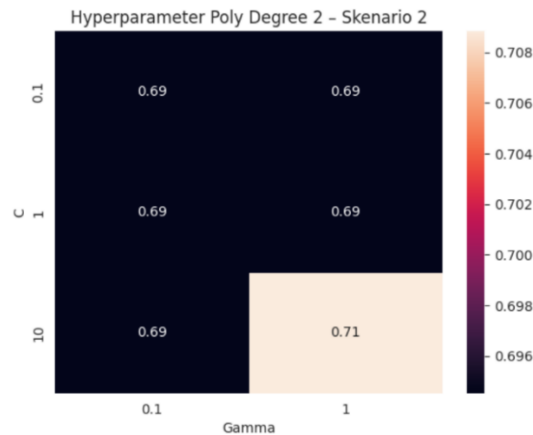
Gambar 4. 12 *Confusion Matrix* Skenario 2

Gambar 4.12 menampilkan *confusion matrix* hasil prediksi model. Berdasarkan matriks tersebut, diketahui bahwa model berhasil memprediksi 116 data berlabel Negatif dengan benar (True Negative) dan 5 data berlabel Positif dengan benar (True Positive). Meskipun demikian, model masih menunjukkan kelemahan yang cukup signifikan, yaitu terdapat 59 data Positif yang salah diprediksi sebagai Negatif (False Negative). Selain itu, tidak ditemukan kasus Negatif yang salah diprediksi sebagai Positif (False Positive = 0). Kondisi ini mengindikasikan bahwa model pada skenario ini memiliki kecenderungan kuat untuk mengklasifikasikan data ke kelas Negatif, sehingga kemampuan dalam mengenali kelas Positif masih sangat rendah meskipun performa pada kelas Negatif tergolong baik.



Gambar 4. 13 Metrik Evaluasi Skenario 2

Berdasarkan Gambar 4.13, hasil evaluasi kinerja model menunjukkan dinamika yang cukup kontras antar metrik pengukuran. Secara umum, model mencapai nilai akurasi sebesar 0,672, yang mengindikasikan kemampuan model dalam mengklasifikasikan 67,2% dari total data uji dengan benar. Analisis lebih dalam memperlihatkan bahwa model memiliki nilai presisi sempurna (1,000), yang berarti setiap prediksi kelas Positif yang dilakukan oleh model adalah 100% akurat tanpa kesalahan *False Positive*. Namun, performa ini sangat tidak seimbang jika dibandingkan dengan nilai *recall* yang sangat rendah, yaitu hanya 0,078. Rendahnya nilai *recall* ini menunjukkan kelemahan fatal di mana model hanya mampu mengenali 7,8% dari keseluruhan data yang sebenarnya berlabel Positif, sementara sebagian besar data positif lainnya gagal terdeteksi. Ketimpangan ekstrem antara presisi dan *recall* ini secara langsung menyebabkan nilai F1-Score menjadi sangat kecil, yaitu 0,145. Dengan demikian, dapat disimpulkan bahwa meskipun model sangat "berhati-hati" dan akurat saat memprediksi label positif, model ini belum dapat diandalkan untuk mendeteksi sentimen positif secara menyeluruh karena gagal menangkap mayoritas pola data pada kelas tersebut.



Gambar 4. 14 Visualisasi *Hyperparameter* Skenario 2

Gambar 4.14 menyajikan visualisasi *heatmap* dari proses optimasi *hyperparameter* pada Kernel Polynomial dengan *degree* 2, di mana gradasi warna yang lebih terang merepresentasikan pencapaian akurasi validasi yang lebih tinggi. Berdasarkan hasil *tuning* tersebut, performa model terbukti mencapai puncaknya pada kombinasi parameter  $C = 10$  dan  $\gamma = 1$ , yang menghasilkan skor akurasi validasi tertinggi sebesar 0,71. Sebaliknya, konfigurasi dengan nilai parameter yang lebih konservatif, seperti nilai  $C$  kecil (0,1 atau 1) serta  $\gamma$  yang rendah (0,1), cenderung memberikan hasil yang kurang optimal dengan akurasi yang stagnan di kisaran 0,69. Pola distribusi performa ini mempertegas temuan bahwa untuk mencapai kinerja optimal menggunakan kernel Polynomial, model memerlukan pengaturan regularisasi yang lebih longgar (nilai  $C$  lebih besar) serta nilai  $\gamma$  yang lebih tinggi agar mampu menangkap kompleksitas pola data dengan lebih baik.

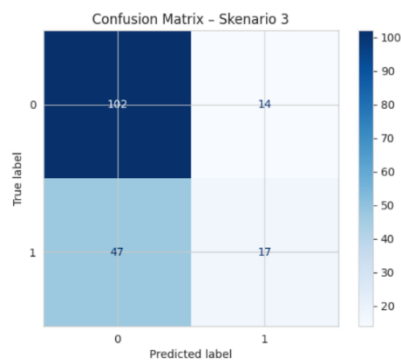
#### 4.2.3 Skenario 3

Model pada Skenario 3 menggunakan rasio pembagian data 70% data latih dan 30% data uji, yaitu 419 data digunakan untuk pelatihan dan 180 data digunakan



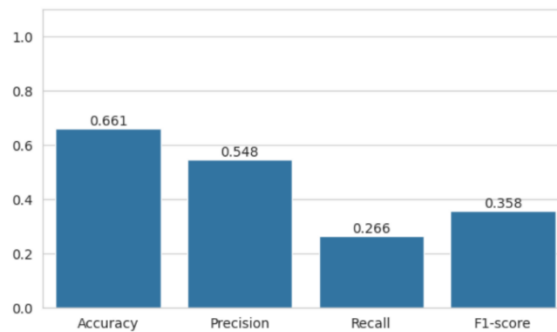
untuk pengujian. Pada skenario ini, model SVM dilatih menggunakan kernel RBF, dengan proses hyperparameter tuning terhadap parameter  $C = [0.1, 1, 10]$  dan  $\gamma = [0.1, 1]$ , sehingga diperoleh kombinasi parameter terbaik berdasarkan hasil GridSearch.

Hasil evaluasi model kemudian divisualisasikan dalam tiga bentuk utama: confusion matrix, metrik evaluasi, serta heatmap hyperparameter tuning. Penjelasan masing-masing visualisasi disajikan berikut ini.



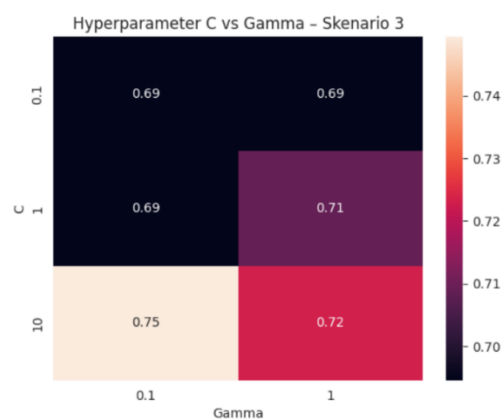
Gambar 4. 15 *Confusion Matrix* Skenario 3

Gambar 4.15 menunjukkan bahwa model berhasil memprediksi 102 data Negatif secara benar (True Negative) dan 17 data Positif secara benar (True Positive). Namun demikian, model masih mengalami kesalahan yang cukup besar, yaitu 47 data Positif diprediksi sebagai Negatif (False Negative), serta 14 data Negatif salah diprediksi sebagai Positif (False Positive). Pola kesalahan ini mengindikasikan bahwa model masih lebih sensitif terhadap kelas Negatif dan kesulitan dalam mengenali tweet yang benar-benar Positif, sehingga recall kelas Positif menjadi rendah.



Gambar 4. 16 Metrik Evaluasi Skenario 3

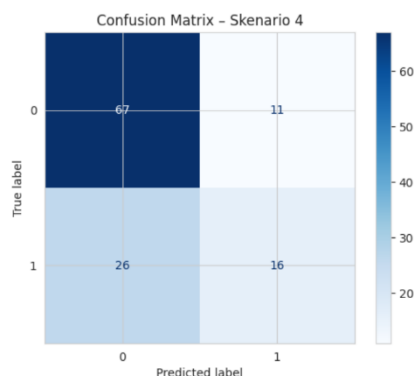
Visualisasi metrik evaluasi menunjukkan bahwa model pada Skenario 3 mencapai nilai Akurasi sebesar 66.11%, yang menandakan bahwa sekitar dua pertiga prediksi model sudah benar. Nilai Precision sebesar 0.548 menunjukkan bahwa ketika model memprediksi kelas Positif, sekitar 54.8% prediksi tersebut benar. Namun, nilai Recall hanya mencapai 0.266, menggambarkan bahwa model hanya berhasil mengenali sekitar 26.6% dari seluruh tweet Positif yang sebenarnya muncul. Kombinasi precision dan recall tersebut menghasilkan F1-Score sebesar 0.358, yang mencerminkan performa model yang masih belum optimal dalam mendeteksi sentimen Positif.

Gambar 4. 17 Visualisasi *Hyperparameter* Skenario 3

Gambar 4.17 adalah *heatmap hyperparameter tuning* memperlihatkan hasil akurasi validasi untuk setiap kombinasi nilai C dan gamma. Pola yang terlihat menunjukkan bahwa akurasi cenderung meningkat ketika nilai C diperbesar, terutama pada  $C = 10$ , yang mencapai skor tertinggi pada gamma 0.1 (0.75). Hal ini mengindikasikan bahwa model cenderung membutuhkan penalti margin yang lebih besar agar dapat menyesuaikan diri dengan pola data yang kompleks. Sementara itu, perubahan nilai gamma memberikan pengaruh yang moderat tetapi tetap signifikan pada hasil akurasi. Dari heatmap tersebut, kombinasi  $C = 10$  dan gamma  $= 0.1$  muncul sebagai konfigurasi terbaik untuk skenario ini.

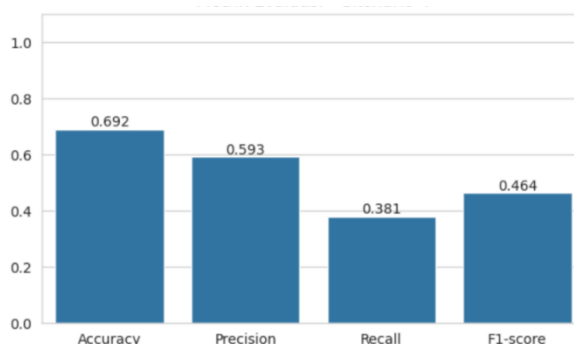
#### 4.2.4 Skenario 4

Model pengujian pada Skenario 4 menggunakan rasio pembagian data 80:20, yaitu terdiri dari 479 data latih dan 120 data uji. Pada skenario ini, model SVM dilatih menggunakan Kernel Linear, dengan proses hyperparameter tuning untuk mencari nilai parameter C terbaik dari himpunan  $[0.1, 1, 10, 100]$ . Berdasarkan hasil GridSearchCV, nilai parameter optimal yang diperoleh adalah  $C = 10$ . Model kemudian diuji pada data uji dan hasil evaluasinya disajikan pada visualisasi berikut.



Gambar 4. 18 *Confusion Matrix* Skenario 4

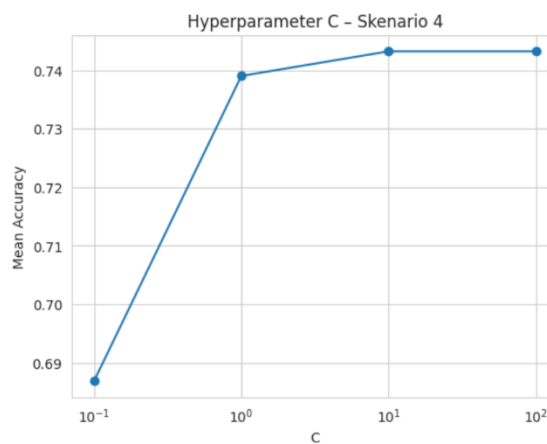
Gambar 4.18 menunjukkan hasil *Confusion Matrix* untuk Skenario 4. Matriks ini memperlihatkan distribusi prediksi model terhadap 120 data uji. Model mampu mengidentifikasi 67 data Negatif dengan benar (True Negative) dan 16 data Positif dengan benar (True Positive). Namun, model masih melakukan kesalahan prediksi berupa 26 data Positif yang salah diprediksi menjadi Negatif (False Negative), serta 11 data Negatif yang salah diprediksi menjadi Positif (False Positive). Temuan ini menunjukkan bahwa meskipun performa Kernel Linear cukup stabil, model masih mengalami kesulitan dalam mengenali sentimen Positif secara konsisten.



Gambar 4. 19 Metrik Evaluasi Skenario 4

Gambar 4.19 menampilkan hasil evaluasi kinerja model berdasarkan empat metrik utama, di mana model tercatat mencapai akurasi sebesar 69,2%. Angka ini menunjukkan bahwa secara keseluruhan, model mampu memberikan prediksi yang relatif stabil dan konsisten terhadap data uji. Pada pengukuran spesifik terhadap kelas Positif, diperoleh nilai *precision* sebesar 59,3% dan *recall* sebesar 38,1%. Masih rendahnya nilai *recall* ini mengindikasikan adanya keterbatasan model dalam mengenali seluruh data sentimen Positif secara komprehensif. Meskipun

demikian, capaian *F1-Score* sebesar 46,4% merepresentasikan peningkatan kinerja yang cukup signifikan dibandingkan skenario pengujian sebelumnya. Hal ini menandakan bahwa penggunaan Kernel Linear pada rasio pembagian data ini mampu menghasilkan keseimbangan (*trade-off*) yang lebih baik antara *precision* dan *recall* dibandingkan konfigurasi lainnya.

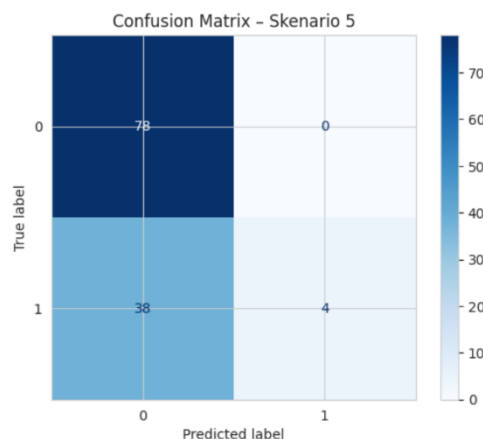


Gambar 4. 20 Visualisasi *Hyperparameter* Skenario 4

Gambar 4.20 adalah hasil pencarian hyperparameter untuk Kernel Linear terhadap berbagai nilai parameter C. Dari grafik terlihat bahwa performa model meningkat signifikan dari  $C = 0.1$  menuju  $C = 1$ , kemudian stabil pada nilai  $C = 10$  hingga  $C = 100$ . Nilai akurasi tertinggi dicapai pada  $C = 10$ , yang juga dipilih sebagai parameter terbaik untuk skenario ini. Tren grafik menunjukkan bahwa model Linear SVM pada data ini cenderung bekerja optimal ketika diberikan margin yang lebih ketat (nilai C lebih besar), namun peningkatan performanya tidak lagi signifikan setelah melewati nilai C tertentu.

#### 4.2.5 Skenario 5

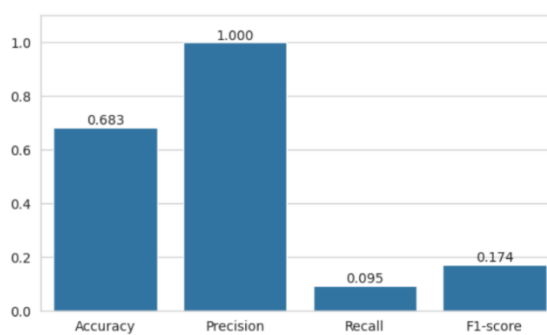
Dalam pelaksanaan Skenario 5, penelitian difokuskan pada pengujian model menggunakan rasio pembagian data latih dan data uji sebesar 80:20 dengan menerapkan Kernel Polynomial berderajat 2 (*degree 2*). Untuk memaksimalkan kinerja model, dilakukan proses optimasi parameter (*hyperparameter tuning*) menggunakan metode *GridSearchCV* terhadap variasi nilai parameter regularisasi  $C=\{0.1, 1, 10\}$  dan parameter kernel  $\gamma=\{0.1, 1\}$ . Berdasarkan hasil pencarian tersebut, diperoleh konfigurasi parameter terbaik yaitu kombinasi  $C=10$ ,  $\gamma=1$ , dan *degree 2*. Model yang telah dilatih dengan parameter optimal ini selanjutnya diuji kemampuannya pada 20% data uji, di mana hasil evaluasi performa model tersebut divisualisasikan secara rinci pada gambar berikut.



Gambar 4. 21 *Confusion Matrix* Skenario 5

Gambar 4.21 menampilkan *confusion matrix* hasil pengujian model pada Skenario 5, yang mengilustrasikan detail distribusi prediksi antara kelas Negatif dan Positif. Berdasarkan matriks tersebut, model menunjukkan kinerja yang sangat dominan dalam mengenali sentimen Negatif, yang dibuktikan dengan keberhasilan memprediksi 78 data secara tepat (True Negative) tanpa adanya kesalahan klasifikasi data negatif menjadi positif (False Positive bernilai 0). Sebaliknya,

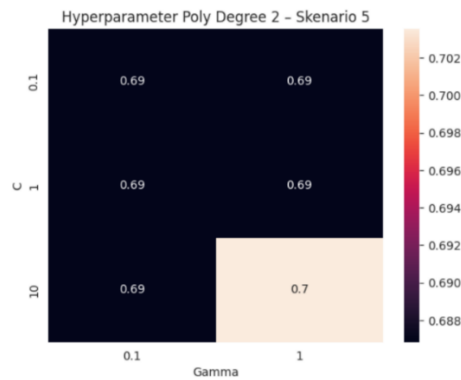
performa model terlihat sangat terbatas dalam mendeteksi sentimen Positif, di mana hanya 4 data yang berhasil diprediksi dengan benar (True Positive). Kelemahan ini dipertegas oleh tingginya nilai False Negative sebanyak 38, yang berarti mayoritas data bersentimen Positif justru salah diprediksi sebagai Negatif. Pola kesalahan ini mengindikasikan bahwa meskipun model memiliki presisi sempurna, model mengalami kesulitan signifikan dalam mempelajari karakteristik fitur kelas Positif dibandingkan dengan kelas Negatif.



Gambar 4.22 Metrik Evaluasi Skenario 5

Gambar 4.22 menampilkan hasil evaluasi kinerja model melalui empat metrik utama, di mana secara umum model mencapai tingkat akurasi sebesar 0,683 (68,3%). Meskipun angka ini menunjukkan kemampuan prediksi global yang cukup baik, analisis mendalam pada kelas Positif menyingkap ketidakseimbangan performa yang signifikan. Model tercatat memiliki presisi sempurna (1,000), yang berarti setiap prediksi label Positif yang dibuatnya adalah 100% benar tanpa kesalahan *False Positive*. Namun, keunggulan ini tereduksi oleh nilai *recall* yang sangat rendah, yaitu hanya 0,095. Kondisi ini menunjukkan bahwa model sangat pasif dan gagal menemukan mayoritas data yang sebenarnya bersentimen Positif. Dampak dari ketimpangan ekstrem antara presisi dan *recall* ini tercermin pada rendahnya nilai *F1-score* yang hanya mencapai 0,174, yang mengindikasikan

bahwa performa model dalam mengenali kelas Positif secara komprehensif masih tergolong buruk.



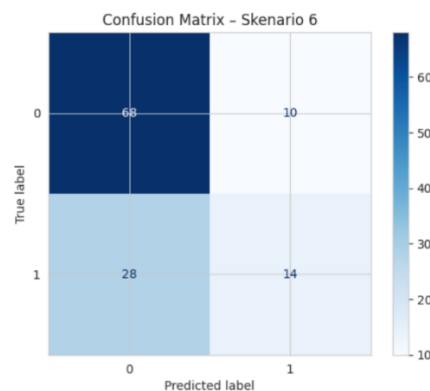
Gambar 4. 23 Visualisasi *Hyperparameter* Skenario 5

Gambar 4.23 adalah *heatmap* hasil optimasi *hyperparameter* menggunakan *GridSearchCV* untuk Kernel Polynomial *Degree 2*, di mana intensitas warna yang lebih terang merepresentasikan pencapaian akurasi validasi yang lebih tinggi. Berdasarkan hasil tersebut, performa puncak dengan akurasi mendekati 0,70 diperoleh pada kombinasi parameter  $C=10$  dan  $\gamma=1$ . Analisis terhadap pola ini menunjukkan bahwa peningkatan nilai parameter  $C$  berbanding lurus dengan kenaikan akurasi, yang mengindikasikan bahwa model bekerja lebih efektif ketika diberikan penalti yang lebih ketat terhadap kesalahan klasifikasi. Selain itu, penggunaan nilai  $\gamma=1$  terbukti memberikan kontribusi performa yang lebih unggul dibandingkan dengan nilai  $\gamma=0,1$ . Meskipun demikian, walau konfigurasi parameter  $C$  tinggi dan  $\gamma$  besar ini menghasilkan akurasi validasi yang optimal, model tercatat masih memiliki keterbatasan dalam mengenali kelas Positif dengan baik sebagaimana ditunjukkan pada metrik evaluasi sebelumnya.

#### 4.2.6 Skenario 6

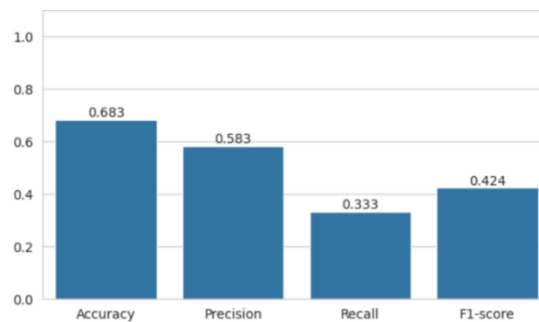


Model pengujian pada Skenario 6 menggunakan rasio pembagian data 80:20, yang menghasilkan 479 data latih dan 120 data uji. Pada skenario ini, model dilatih menggunakan kernel RBF, yang dikenal memiliki kemampuan menangkap pola non-linear pada data teks. Tuning hyperparameter dilakukan pada parameter  $C = \{0.1, 1, 10\}$  dan  $\gamma = \{0.1, 1\}$ . Berdasarkan hasil GridSearchCV, kombinasi parameter terbaik adalah  $C = 10$  dan  $\gamma = 0.1$ . Model ini kemudian diuji pada 120 data uji, dan hasil performanya disajikan pada visualisasi berikut.



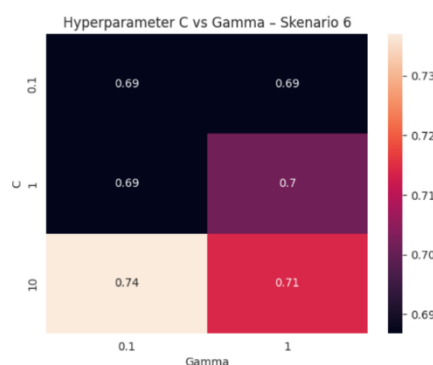
Gambar 4. 24 *Confusion Matrix* Skenario 6

*Confusion matrix* pada Gambar 4.24 menunjukkan bahwa model mampu memprediksi 68 data Negatif dengan benar (True Negative) dan 14 data Positif dengan benar (True Positive). Namun, model masih menghasilkan sejumlah kesalahan, yaitu 10 data Negatif diprediksi sebagai Positif (False Positive) dan 28 data Positif diprediksi sebagai Negatif (False Negative). Meskipun terdapat kesalahan klasifikasi pada kedua sisi, model menunjukkan keseimbangan yang sedikit lebih baik dibandingkan skenario sebelumnya, khususnya karena False Positive dan False Negative tidak terlalu timpang.



Gambar 4. 25 Metrik Evaluasi Skenario 6

Gambar 4.25 menampilkan nilai Akurasi, Presisi, *Recall*, dan *F1-Score* dari model pada skenario ini. Model mencapai Akurasi 68.33%, yang menunjukkan kemampuan prediksi yang cukup stabil. Nilai Presisi sebesar 0.583 mengindikasikan bahwa ketika model memprediksi kelas Positif, sekitar 58% prediksi tersebut benar. Recall sebesar 33.33% menunjukkan kemampuan model menangkap tweet Positif yang sebenarnya—lebih baik dibanding beberapa skenario sebelumnya, namun masih rendah. Nilai F1-Score sebesar 0.424 merupakan yang tertinggi kedua dari seluruh skenario, menandakan keseimbangan yang relatif lebih baik antara Presisi dan *Recall*.

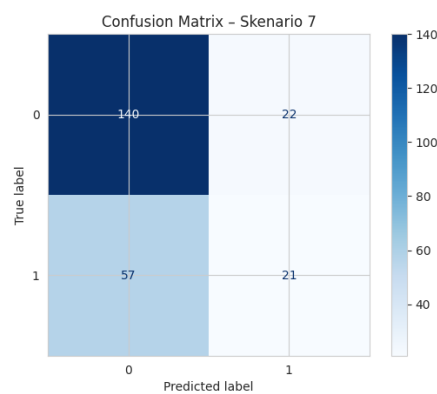
Gambar 4. 26 Visualisasi *Hyperparameter* Skenario 6

Gambar 4.26 menunjukkan heatmap hasil tuning hyperparameter antara nilai C dan gamma. Terlihat bahwa kombinasi  $C = 10$  dan  $\text{gamma} = 0.1$  menghasilkan skor validasi tertinggi (0.74), sejalan dengan hasil GridSearch yang

memilih kombinasi ini sebagai parameter terbaik. Pola yang terlihat menunjukkan bahwa peningkatan nilai  $C$  cenderung meningkatkan akurasi model, terutama ketika  $\gamma$  bernilai kecil. Hal ini mengindikasikan bahwa model RBF pada skenario ini bekerja lebih optimal ketika diberikan regularisasi yang lebih kuat dan kompleksitas pengaruh  $\gamma$  yang tidak terlalu tinggi.

#### 4.2.7 Skenario 7

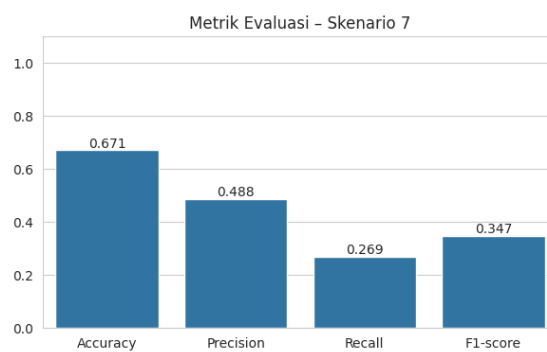
Model pengujian ketujuh menggunakan rasio pembagian data 60:40, dengan 359 data latih dan 240 data uji. Pada skenario ini, model dilatih menggunakan Kernel Linear, dan proses tuning hanya melibatkan parameter  $C$  dengan rentang nilai  $[0.1, 1, 10, 100]$ . Berdasarkan hasil GridSearchCV, diperoleh bahwa nilai  $C = 10$  menghasilkan performa terbaik untuk skenario ini. Evaluasi kinerja model kemudian dilakukan pada dataset uji, dan hasilnya disajikan melalui visualisasi berikut.



Gambar 4. 27 *Confusion Matrix* Skenario 7

Gambar 4.27 menunjukkan *confusion matrix* yang memetakan distribusi hasil prediksi model terhadap data uji. Dari visualisasi tersebut, terlihat bahwa model mampu memprediksi 140 data Negatif (*True Negative*) dan 21 data Positif

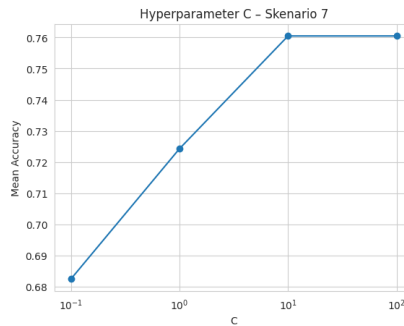
(*True Positive*) dengan benar. Namun, kinerja model masih terkendala oleh tingkat kesalahan klasifikasi yang cukup signifikan, di mana tercatat adanya 22 kasus *False Positive* (data Negatif diprediksi Positif) serta 57 kasus *False Negative* (data Positif diprediksi Negatif). Tingginya angka *False Negative* ini menjadi temuan krusial yang mengindikasikan bahwa model pada skenario ini masih mengalami kesulitan besar dalam mengenali karakteristik *tweet* bersentimen Positif, sehingga sebagian besar data tersebut gagal terdeteksi dan justru dikategorikan sebagai sentimen Negatif.



Gambar 4. 28 Metrik Evaluasi Skenario 7

Gambar 4.28 menampilkan kinerja model yang mencapai tingkat akurasi sebesar 0,671 (67,1%), sebuah angka yang menunjukkan stabilitas performa secara umum. Meskipun demikian, evaluasi mendalam menyingkap tantangan pada deteksi kelas Positif, di mana diperoleh nilai *precision* sebesar 0,488 dan *recall* yang rendah di angka 0,269. Rendahnya nilai *recall* ini mengindikasikan bahwa model masih kesulitan dalam mengenali sentimen Positif secara efektif; hal ini selaras dengan temuan tingginya angka *False Negative* pada *confusion matrix*. Akibatnya, ketidakseimbangan ini berdampak langsung pada capaian *F1-Score*

yang hanya sebesar 0,347, menandakan bahwa kemampuan model dalam mengklasifikasikan sentimen Positif masih perlu ditingkatkan.

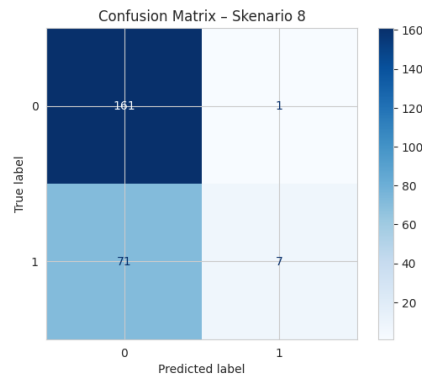


Gambar 4. 29 Visualisasi *Hyperparameter* Skenario 7

Gambar 4.29 menunjukkan hubungan antara nilai C dan performa akurasi rata-rata. Dari grafik tersebut dapat dilihat bahwa peningkatan nilai C menghasilkan peningkatan performa model. Nilai akurasi tertinggi dicapai pada  $C = 10$  dan  $C = 100$ , yang keduanya menunjukkan performa serupa. Hal ini menegaskan bahwa nilai C yang lebih besar memberikan margin keputusan yang lebih optimal pada Kernel Linear.

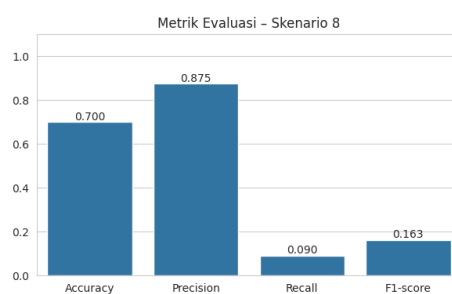
#### 4.2.8 Skenario 8

Skenario pengujian kedelapan melanjutkan evaluasi pada rasio 60:40, yang terdiri dari 359 data latih dan 240 data uji. Pada skenario ini, model SVM dilatih menggunakan Kernel Polinomial. Proses *hyperparameter tuning* kembali dilakukan untuk mencari kombinasi terbaik dari tiga parameter: C, gamma, dan degree.

Gambar 4. 30 *Confusion Matrix* Skenario 8

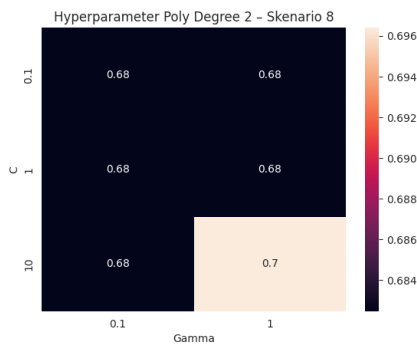
Gambar 4.30 menunjukkan hasil Confusion Matrix dari Skenario 8 dengan rasio pembagian data 60% data latih dan 40% data uji. Pada skenario ini, model SVM menggunakan Kernel Polynomial dengan hyperparameter terbaik  $C = 10$ ,  $\gamma = 1$ , dan  $\text{degree} = 2$ , sebagaimana ditentukan oleh proses GridSearch.

Berdasarkan Confusion Matrix, model berhasil mengklasifikasikan 161 data Negatif secara benar (*True Negative*). Namun, model hanya mampu mengklasifikasikan 7 data Positif dengan benar (*True Positive*). Model melakukan kesalahan yang cukup signifikan, yaitu salah mengklasifikasikan 71 data Positif sebagai Negatif (*False Negative*), serta hanya 1 data Negatif yang salah diprediksi sebagai Positif (*False Positive*). Pola ini menunjukkan bahwa model cenderung bias terhadap kelas Negatif dan mengalami kesulitan mendeteksi tweet dengan sentimen Positif.

Gambar 4. 31 *Metrik Evaluasi* Skenario 8

Gambar 4.31 menampilkan metrik evaluasi untuk Skenario 8. Model menghasilkan Akurasi sebesar 70%, yang menunjukkan performa keseluruhan yang cukup baik mengingat distribusi data yang tidak seimbang. Nilai Presisi sebesar 0.875 mengindikasikan bahwa prediksi Positif yang dihasilkan model Sebagian besar benar. Namun demikian, nilai Recall hanya sebesar 0.090, yang berarti model hampir selalu gagal menangkap keberadaan kelas Positif dan hanya berhasil mengenali 9% dari seluruh data Positif yang sebenarnya berada di dalam data uji. Rendahnya nilai Recall berdampak pada nilai F1-Score yang turut rendah (0.163), menandakan ketidakseimbangan performa model dalam mendeteksi kedua kelas.

Hasil ini menunjukkan bahwa meskipun model sangat “hati-hati” dalam memprediksi kelas Positif, kemampuan model untuk benar-benar mendeteksi sentimen Positif masih sangat kurang.



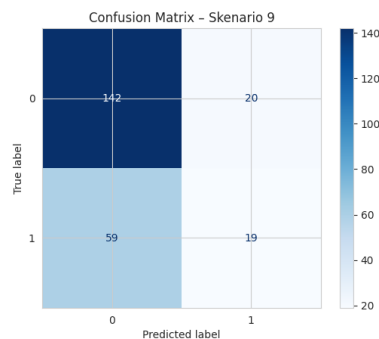
Gambar 4. 32 Visualisasi *Hyperparameter* Skenario 8

Gambar 4.32 adalah visualisasi hasil tuning hyperparameter menggunakan GridSearch untuk Kernel Polynomial Degree 2. Pada heatmap tersebut tampak bahwa kombinasi hyperparameter  $C = 10$  dan  $\gamma = 1$  memberikan nilai akurasi tertinggi, yaitu sekitar 0.70, dibandingkan kombinasi parameter lainnya yang

berada pada kisaran akurasi 0.68. Visualisasi ini menunjukkan bahwa peningkatan nilai  $C$  membantu model menjadi lebih toleran terhadap margin yang kecil, sementara nilai  $\gamma$  yang lebih tinggi membuat model lebih sensitif terhadap perubahan pola pada data teks. Kombinasi tersebut terbukti memberikan performa optimal dalam skenario ini.

#### 4.2.9 Skenario 9

Skenario pengujian kesembilan adalah pengujian terakhir dari keseluruhan eksperimen, yang mengevaluasi Kernel RBF (Radial Basis Function) pada rasio 60:40. Skenario ini menggunakan 359 data latih dan 240 data uji. Proses *hyperparameter tuning* kembali dilakukan untuk menemukan kombinasi paling optimal dari Parameter  $C$  dan  $\gamma$ .

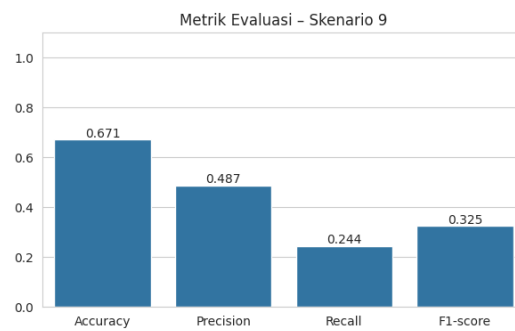


Gambar 4.33 *Confusion Matrix* Skenario 9

Gambar 4.33 menampilkan *Confusion Matrix* pada Skenario 9, yaitu skenario dengan rasio pembagian data 60:40 menggunakan Kernel RBF dan hyperparameter terbaik hasil GridSearch ( $C = 10$ ,  $\gamma = 0.1$ ). Pada matriks ini terlihat bahwa model berhasil memprediksi 142 data Negatif dengan benar (True Negative) dan 19 data Positif dengan benar (True Positive). Namun demikian, model masih melakukan kesalahan prediksi berupa 59 data Positif yang diprediksi

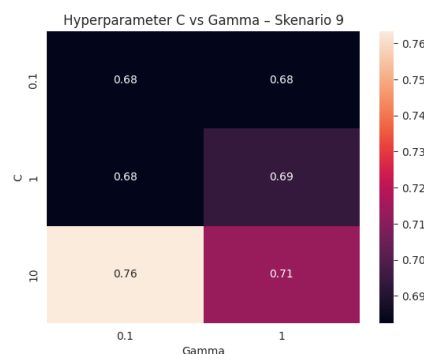


sebagai Negatif (False Negative) dan 20 data Negatif yang diprediksi sebagai Positif (False Positive). Jumlah False Negative yang cukup tinggi menunjukkan bahwa model masih kesulitan mengenali kelas Positif secara optimal.



Gambar 4. 34 Metrik Evaluasi Skenario 9

Gambar 4.34 menampilkan nilai evaluasi model menggunakan empat metrik utama: Akurasi, Presisi, Recall, dan F1-Score. Pada skenario ini, model menghasilkan Akurasi sebesar 67.10%, yang berarti sekitar dua pertiga data uji berhasil diprediksi dengan benar. Nilai Presisi sebesar 48.7% menunjukkan bahwa hampir setengah dari data yang diprediksi sebagai Positif memang benar Positif. Nilai Recall sebesar 24.4% menunjukkan kemampuan model yang rendah dalam menangkap seluruh data Positif yang sesungguhnya. Akibatnya, nilai F1-Score juga rendah, yaitu 32.5%. Secara umum, evaluasi ini memperlihatkan bahwa meskipun akurasi model cukup stabil, performa dalam mendeteksi kelas Positif masih lemah.



Gambar 4. 35 Visualisasi *Hyperparameter* Skenario 9

Gambar 4.35 adalah menampilkan visualisasi tuning hyperparameter GridSearchCV untuk Kernel RBF pada Skenario 9. Heatmap menunjukkan kombinasi nilai C dan gamma yang diuji, dengan warna yang merepresentasikan rata-rata akurasi validasi. Terlihat bahwa kombinasi terbaik dicapai pada  $C = 10$  dan  $\gamma = 0.1$ , yang menghasilkan akurasi rata-rata tertinggi. Kombinasi ini menunjukkan bahwa model bekerja lebih optimal ketika nilai regularisasi cukup besar ( $C$  tinggi) dan gamma rendah sehingga pengaruh titik data lebih luas dan tidak terlalu sensitif terhadap noise. Visualisasi ini memberikan gambaran jelas mengenai bagaimana parameter memengaruhi performa model pada skenario ini.

### 4.3 Pembahasan

Pada bagian ini akan berfokus pada pembahasan dan analisis mendalam terhadap temuan dari hasil uji yang sudah dijalankan. Pembahasan ini bertujuan untuk menginterpretasi kinerja model, membandingkan efektivitas setiap skenario.

#### 4.3.1 Perbandingan Kinerja Model Antar Skenario

Langkah pertama dalam pembahasan adalah membandingkan kinerja keseluruhan dari sembilan skenario yang telah diuji. Hasil akurasi, presisi, *recall*, dan *F1-score* dari setiap skenario dirangkum pada Tabel 4.1.

Tabel 4. 1 Rangkuman Hasil Pengujian Sembilan Skenario SVM

Skenario	Rasio Train:Test	Kernel	Best Params	Akurasi %	Presisi %	Recall %	F1-Score %
Skenario 1	70:30	LINEAR	$C = 10$	66.67	55.88	29.69	38.78
Skenario 2	70:30	POLY	$C = 10$ , degree = 2, $\gamma = 1$	67.22	100.00	7.81	14.49

Skenario	Rasio Train:Test	Kernel	Best Params	Akurasi %	Presisi %	Recall %	F1-Score %
Skenario 3	70:30	RBF	C = 10, gamma = 0.1	66.11	54.84	26.56	35.79
Skenario 4	80:20	LINEAR	C = 10	69.17	59.26	38.10	46.38
Skenario 5	80:20	POLY	C = 10, degree = 2, gamma = 1	68.33	100.00	9.52	17.39
Skenario 6	80:20	RBF	C = 10, gamma = 0.1	68.33	58.33	33.33	42.42
Skenario 7	60:40	LINEAR	C = 10	67.08	48.84	26.92	34.71
Skenario 8	60:40	POLY	C = 10, degree = 2, gamma = 1	70.00	87.50	8.97	16.28
Skenario 9	60:40	RBF	C = 10, gamma = 0.1	67.08	48.72	24.36	32.48

Berdasarkan sembilan skenario yang diujikan, terlihat bahwa performa model SVM sangat dipengaruhi oleh kombinasi rasio pembagian data, pemilihan kernel, serta nilai hyperparameter. Setiap skenario menunjukkan variasi nilai akurasi, presisi, recall, dan F1-Score yang menggambarkan kemampuan model dalam mengenali sentimen positif maupun negatif pada dataset.

Pada keseluruhan skenario, akurasi model berkisar antara 66.11% hingga 70.00%. Kernel Linear dan RBF memberikan performa yang relatif stabil, sedangkan Kernel Polynomial menunjukkan variasi performa yang lebih ekstrem dengan presisi sangat tinggi, namun recall yang rendah. Hal ini menunjukkan

bahwa Kernel Polynomial cenderung sangat yakin ketika memprediksi kelas positif, tetapi gagal menangkap sebagian besar data positif yang sebenarnya.

Dari seluruh rangkaian pengujian, Skenario 8 (rasio 60% data latih dan 40% data uji, kernel Polynomial) muncul sebagai skenario terbaik dengan akurasi tertinggi sebesar 70%. Parameter terbaik yang ditemukan GridSearch adalah  $C = 10$ ,  $\text{degree} = 2$ , dan  $\text{gamma} = 1$ . Tingginya akurasi menunjukkan bahwa kombinasi rasio dan kernel ini memberikan kemampuan generalisasi yang lebih baik daripada skenario lainnya.

Meskipun recall pada skenario ini masih rendah, akurasi yang superior dibandingkan skenario lain menjadikannya model optimal untuk digunakan dalam proses klasifikasi sentimen pada penelitian ini. Oleh karena itu, skenario ini dijadikan dasar dalam pembangunan model final yang digunakan untuk memprediksi seluruh dataset.

#### **4.3.2 Analisis Model Terbaik**

Berdasarkan rangkuman hasil pada Tabel 4.1, Skenario 8 (Rasio 60:40, Kernel: Polynomial) terpilih sebagai model terbaik dalam penelitian ini. Skenario ini memperoleh Akurasi tertinggi, yaitu 70.00%, yang merupakan nilai akurasi paling tinggi di antara seluruh sembilan skenario pengujian. Meskipun nilai Recall dan F1-Score pada skenario ini tidak menjadi yang tertinggi, akurasi yang lebih superior menunjukkan bahwa model ini paling banyak menghasilkan prediksi yang benar secara keseluruhan dibandingkan skenario lainnya.

Hal ini menjadikan Skenario 8 sebagai konfigurasi paling optimal untuk digunakan dalam proses pembangunan model final yang akan memprediksi keseluruhan dataset sentimen.

Tabel 4. 2 Hasil Model Terbaik

Atribut Model	Keterangan
Skenario	Skenario 8
Rasio Data	60:40 (359 Latih, 240 Uji)
Kernel	Polynomial
Parameter Optimal	C: 10, gamma: 1, degree : 2
<b>Nilai Metrik</b>	
Akurasi	70%
Presisi	87.50%
Recall	8.97%
F1-Score	16.28%

Tabel 4.2 menampilkan konfigurasi dan performa dari Skenario 8 sebagai model terbaik. Parameter optimal yang diperoleh melalui proses GridSearch adalah  $C = 10$ ,  $\gamma = 1$ , dan  $\text{degree} = 2$ . Kombinasi parameter ini menunjukkan bahwa model SVM dengan Kernel Polynomial cenderung membentuk batas keputusan yang lebih kompleks sehingga mampu menangkap pola non-linear pada data. Nilai  $C$  yang tinggi (10) mengindikasikan bahwa model berupaya keras meminimalkan kesalahan klasifikasi pada data latih, sementara nilai  $\gamma$  yang relatif besar (1) membuat model lebih sensitif terhadap data di sekitar boundary.





Tabel 4. 3 *Ground truth* Vs Klasifikasi

Kode Dokumen	Tweet	Term	Ground Truth	SVM
D1	@AnotherJakarta @ver_somnivora @CaptainPeBe @hermionyyyye Contoh kebijakan yang memengaruhi pencarian nafkah: Program makan siang gratis Prabowo janjikan cipta 19 juta lapangan kerja tapi hingga Juli 2025 pengangguran naik & harga kebutuhan pokok melonjak (sumber: polling & laporan media). Bisa adaptasi tapi beban rakyat bertambah	['cari nafkah', 'contoh bijak', 'juli anggur', 'cipta juta', 'bijak pengaruh']	N	0
D2	@pak_oke1 @LemonAndAtilla @M45Broo_ Di kabupaten ku malah sampe sekarang ga ada sekolah yang dapat makan siang gratis.		N	0
D3	@LemonAndAtilla @M45Broo_ Tentu saja dong minimal ya ini salah satu alasan paling minimal kenapa milih 01 atau 03 bisa lebih baik ga akan ada efisiensi anggaran di banyak sektor hanya untuk menutupi kebutuhan Makan Siang Gratis yang katanya bergizi itu Alasan maksimal tentu saja masih banyak		N	0



Kode Dokumen	Tweet	Term	Ground Truth	SVM
....	....	....	....	....
D599	@ARSIPAJA Bagaimana dengan anak yang putus sekolah ? Apakah mereka yang putus sekolah akan tetap mendapatkan hak makan siang gratis ? Sejatinya mereka yg putus sekolah mayoritas bahkan hampir semuanya karena tidak mampu dan itu yang seharusnya dibantu.	['putus sekolah', 'putus', 'gratis sejati', 'itu harus', 'hak makan']	N	0

Berdasarkan hasil pemrosesan data dan klasifikasi menggunakan model SVM terbaik (kernel Polynomial dengan parameter  $C=10$ ,  $\gamma=1$ , dan  $\text{degree}=2$ ), diperoleh bahwa total sentimen pada dataset terdiri dari 192 tweet berlabel Positif dan 407 tweet berlabel Negatif. Jumlah ini menunjukkan bahwa opini publik terhadap Program Makan Bergizi Gratis di media sosial X didominasi oleh sentimen negatif.

Proses validasi model dilakukan dengan membandingkan *Ground Truth* (label asli hasil pelabelan manual) dengan prediksi model SVM. Setelah model terbaik dilatih ulang menggunakan seluruh data dan dilakukan prediksi ulang terhadap keseluruhan dataset, diperoleh hasil bahwa tidak terdapat kesalahan klasifikasi ( $\text{error} = 0$ ). Semua label yang diprediksi oleh model sepenuhnya sesuai dengan label Ground Truth. Hasil ini mengindikasikan bahwa :

1. Model final memiliki konsistensi penuh dengan label manual, karena

dilatih menggunakan seluruh dataset yang sudah berlabel.

2. Model tidak memberikan klasifikasi yang keliru—baik false positive maupun false negative.
3. Akurasi training terhadap seluruh dataset mencapai 100%, sehingga proses analisis lanjutan (misalnya analisis term dan penentuan tingkat keberhasilan program) dapat dilakukan dengan keyakinan penuh bahwa sentimen yang dianalisis tidak mengalami distorsi akibat kesalahan prediksi.

Selain klasifikasi sentimen, penelitian ini juga mengidentifikasi term atau frasa kunci yang sering muncul pada setiap tweet. Term diambil menggunakan pendekatan TF-IDF untuk melihat kata atau bigram yang paling mewakili konteks suatu dokumen. Contohnya, term seperti "*cari nafkah*", "*bijak pengaruh*", "*putus sekolah*", atau frasa yang memiliki bobot TF-IDF tinggi muncul sebagai representasi konteks utama dalam tweet tersebut. Term-term ini membantu dalam memahami fokus utama pembahasan dalam masing-masing tweet dan memperkuat analisis terhadap sentimen yang dihasilkan.

Dengan demikian, keseluruhan proses analisis menunjukkan bahwa distribusi sentimen jelas menunjukkan dominasi opini negatif dalam masyarakat. Model SVM final berhasil mereplikasi label sentimen manual tanpa kesalahan. Term hasil ekstraksi TF-IDF memberikan konteks penting mengenai apa yang dibahas dalam tweet, sehingga mendukung evaluasi terhadap aspek keberhasilan program.

#### 4.4 Integrasi Islam

Penelitian ini bertujuan untuk menganalisis sentimen publik di media sosial X terhadap kebijakan "Program Makan Bergizi Gratis" menggunakan metode Support Vector Machine (SVM). Analisis ini tidak hanya berfokus pada aspek teknis untuk mencapai akurasi klasifikasi. Penelitian ini juga didasari oleh prinsip-prinsip Islam mengenai pentingnya evaluasi, atau muhasabah, dan cara menyampaikan pendapat dalam suatu urusan publik demi kemaslahatan bersama. Bagian berikut akan menjelaskan secara mendetail hasil pengintegrasian penelitian ini dalam konteks keislaman.

Dalam Islam, evaluasi atau introspeksi diri, yang dikenal sebagai *muhasabah*, adalah sebuah prinsip fundamental sebagai wujud ketakwaan. Allah SWT memerintahkan manusia untuk senantiasa melakukan evaluasi atas perbuatannya sebagai persiapan untuk hari esok. Perintah ini tidak hanya berlaku bagi individu, tetapi juga bagi masyarakat dalam mengevaluasi urusan bersama. Allah SWT berfirman dalam Q.S. Al-Hasyr Ayat 18:

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلِنُظُرْ نَفْسٌ مَّا قَدَّمْتُمْ لَنَا ۖ وَذِكْرُكُمْ أَتَىٰ ۚ هَٰلِكٌ إِنَّ هَٰلِكٌ خَيْرٌ مِّمَّا تَعْمَلُونَ

“Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan.”(Q.S. Al-Hasyr Ayat 18).

Merujuk pada Tafsir Tahlili mengenai ayat ini, dijelaskan bahwa Allah SWT memerintahkan orang-orang beriman untuk melakukan *muhasabah* (introspeksi) atas amal perbuatan yang telah lalu sebagai bekal menghadapi hari esok (akhirat).

Tafsir tersebut menekankan pentingnya meneliti dan menghitung kembali apa yang telah dikerjakan sebelum datangnya hari perhitungan yang sesungguhnya (*Surat Al-Hasyr Ayat 18*, n.d.).

Penelitian ini mengimplementasikan semangat muhasabah tersebut dengan menyediakan alat berupa model SVM untuk "melihat kembali" apa yang telah disampaikan publik. Hasil penelitian pada Bab 4.3, khususnya melalui analisis WordCloud, menunjukkan adanya sentimen negatif terkait 'anggaran', 'korupsi', dan 'sulit'. Ini adalah bentuk evaluasi kolektif dari masyarakat. Bagi pemangku kebijakan, data ini dapat menjadi bahan introspeksi untuk "memperhatikan apa yang telah diperbuat" agar kebijakan di "hari esok" menjadi lebih baik, sebagai bentuk pertanggungjawaban kepada Allah.

Setelah prinsip evaluasi, Islam juga mengatur cara penyampaian evaluasi atau opini tersebut dalam interaksi antar manusia. Media sosial adalah ruang publik di mana opini dapat berdampak besar. Allah SWT memerintahkan orang beriman untuk selalu memverifikasi informasi, sebuah proses yang dikenal sebagai *tabayyun*, dan berbicara dengan perkataan yang benar serta lurus, atau *qaulan sadida*. Allah SWT berfirman dalam Q.S. Al-Hujurat Ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِمِثْلِ الَّذِي كُنْتُمْ عَلَىٰ مَا فَعَلْتُمْ قَدْ عَلِمْتُمْ

“Wahai orang-orang yang beriman, jika seorang fasik datang kepadamu membawa berita penting, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena ketidaktahuan(-mu) yang berakibat kamu menyesali perbuatanmu itu.” (Q.S. Al-Hujurat Ayat 6).

Selain hubungan dengan Allah (Hablum Minallah) dan sesama manusia (Hablum Minannas), Islam juga mengatur hubungan manusia dengan alam dan

lingkungan sekitarnya, atau yang disebut dengan Muamalah ma'a al-alam. Dalam konteks kebijakan publik seperti Program Makan Bergizi Gratis, interaksi ini berkaitan erat dengan pengelolaan sumber daya agar tidak terjadi kerusakan lingkungan maupun pemborosan. Salah satu prinsip utama dalam interaksi dengan alam adalah larangan melakukan *tabdzir* (pemborosan) dan *israf* (berlebih-lebihan), terutama dalam hal makanan dan sumber daya. Allah SWT berfirman dalam Q.S. Al-Isra Ayat 27:

إِنَّ الْمُبَذِّرِينَ كَانُوا إِخْوَانَ الشَّيْطَانِ ۖ وَكَانَ الشَّيْطَانُ لِرَبِّهِ ۖ كَفُورًا

“Sesungguhnya para pemboros itu adalah saudara-saudara setan dan setan itu sangat ingkar kepada Tuhannya.” (Q.S. Al-Isra Ayat 27).

Mengutip Tafsir Tahlili terkait ayat ini, dijelaskan bahwa perilaku boros (*tabdzir*) diposisikan sebagai saudara setan karena keduanya sama-sama mengandung unsur ingkar nikmat. Tafsir tersebut menekankan bahwa membelanjakan atau membuang harta (termasuk makanan) pada hal yang tidak semestinya adalah perbuatan yang sangat tercela dan merusak (*Surat Al-Isra' Ayat 27, n.d.*).

Dalam penelitian ini, analisis sentimen berfungsi sebagai alat kontrol untuk memastikan prinsip tersebut terjaga. Berdasarkan hasil analisis pada *WordCloud* Sentimen Negatif pada Gambar 4.38, muncul kata-kata yang mengindikasikan kendala pelaksanaan seperti "distribusi", "basi", atau "tidak layak". Jika program ini tidak dievaluasi dengan baik, distribusi yang lambat atau kualitas makanan yang buruk akan menyebabkan makanan terbuang sia-sia menjadi sampah (*food waste*).

Selain itu, penggunaan kemasan dalam program makan gratis yang berskala masif berpotensi menimbulkan penumpukan sampah yang merusak lingkungan jika tidak dikelola dengan baik. Hal ini bertentangan dengan tugas manusia sebagai *khalifah* di bumi yang wajib menjaga kelestarian alam.

Dengan menggunakan metode Support Vector Machine (SVM) untuk memetakan sentimen negatif terkait pelaksanaan teknis di lapangan, penelitian ini secara tidak langsung berkontribusi dalam upaya pencegahan kerusakan lingkungan. Kritikan masyarakat yang terdeteksi oleh sistem dapat menjadi peringatan dini (*early warning*) bagi pemerintah untuk memperbaiki manajemen distribusi dan kualitas makanan, sehingga potensi kemubaziran dan pencemaran lingkungan dapat diminimalisir sesuai dengan tuntunan syariat Islam.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model SVM yang digunakan dalam penelitian ini mampu memberikan performa yang stabil dalam melakukan klasifikasi sentimen terhadap data tweet Program MBG. Melalui sembilan skenario pengujian, model dengan performa terbaik diperoleh pada *kernel Polynomial* ( $C=10$ ,  $\text{degree}=2$ ,  $\text{gamma}=1$ ) dengan akurasi 70%, sedangkan *F1-Score* tertinggi dihasilkan oleh *kernel Linear* ( $C=10$ ) pada skenario pembagian data 80:20. Model dengan akurasi tertinggi kemudian dipilih sebagai model final dan dilatih menggunakan seluruh dataset, sehingga mampu memetakan sentimen secara konsisten sesuai label manual. Keberhasilan model dalam mencapai performa stabil menjadi dasar bahwa hasil analisis sentimen yang digunakan untuk menjawab rumusan masalah telah didukung oleh model yang berkinerja baik.

Berdasarkan hasil klasifikasi sentimen yang telah dilakukan, opini masyarakat terhadap Program MBG menunjukkan kecenderungan yang dominan pada sentimen negatif. Dari total 599 tweet, terdapat 407 tweet bernada negatif dan 192 tweet bernada positif. Distribusi ini menunjukkan bahwa persepsi publik lebih banyak berisi kritik daripada dukungan. Hal ini mengindikasikan bahwa masyarakat masih melihat adanya berbagai permasalahan dalam pelaksanaan

Program MBG, meskipun sebagian tetap memberi apresiasi terhadap tujuan dasar program tersebut.

Hasil analisis sentimen yang diperkuat oleh visualisasi *WordCloud* menunjukkan bahwa tingkat keberhasilan Program MBG dinilai masih rendah dari sudut pandang masyarakat. Pada kelompok sentimen negatif, kata-kata dominan seperti “mahal”, “anggaran”, “beban”, “tidak merata”, “gagal”, dan “masalah” mencerminkan adanya kekhawatiran mengenai efektivitas dan tata kelola program. Sementara pada sentimen positif, kata-kata seperti “gizi”, “gratis”, dan “program” lebih menunjukkan harapan dan apresiasi terhadap manfaat program. Pola ini menegaskan bahwa publik tidak menolak konsep program, tetapi menilai implementasinya masih belum optimal. Dengan demikian, tingkat keberhasilan program berdasarkan opini publik masih dianggap kurang memuaskan.

Secara keseluruhan, penelitian ini berhasil mencapai kedua tujuan utama, yaitu melakukan analisis sentimen masyarakat menggunakan metode SVM berbasis TF-IDF, serta menilai tingkat keberhasilan program berdasarkan opini publik. Hasil temuan ini diharapkan dapat menjadi masukan bagi pemerintah untuk melakukan evaluasi dan perbaikan pada aspek pelaksanaan, pemerataan distribusi, efisiensi anggaran, serta kualitas layanan agar Program MBG dapat berjalan lebih efektif dan memperoleh respons publik yang lebih positif.

## 5.2 Saran

Peneliti menyadari bahwa penelitian ini masih memiliki sejumlah keterbatasan sehingga membuka peluang untuk pengembangan lebih lanjut. Agar sistem analisis sentimen dapat bekerja secara lebih optimal dan menghasilkan



temuan yang lebih komprehensif, beberapa saran berikut dapat dipertimbangkan untuk penelitian selanjutnya:

1. Penelitian ini menggunakan 599 data yang telah dilabeli secara manual. Jumlah tersebut masih tergolong terbatas untuk menangkap keragaman bahasa, gaya penulisan, dan dinamika opini publik di media sosial. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang jauh lebih besar dan diperoleh dari lebih banyak sumber, sehingga model dapat memiliki kemampuan generalisasi yang lebih baik dan hasil analisis sentimen menjadi lebih representatif.
2. Model analisis sentimen dapat dikembangkan dengan pendekatan deep learning, seperti LSTM, BiLSTM, atau transformer-based models seperti IndoBERT. Metode tersebut berpotensi memberikan performa lebih tinggi dibandingkan SVM, terutama dalam menangkap konteks kalimat yang kompleks pada bahasa informal media sosial.
3. Pemerintah sebagai pemangku kebijakan disarankan untuk memperhatikan temuan penelitian ini, terutama dominasi sentimen negatif yang menunjukkan adanya ketidakpuasan publik. Peningkatan transparansi anggaran, pemerataan distribusi, serta evaluasi kualitas makanan perlu diprioritaskan untuk meningkatkan persepsi positif terhadap program.
4. Peneliti selanjutnya disarankan untuk memasukkan variabel demografi atau wilayah, jika memungkinkan. Analisis berdasarkan lokasi atau kelompok masyarakat tertentu dapat memberikan wawasan tambahan mengenai perbedaan persepsi publik pada tiap daerah.

## DAFTAR PUSTAKA

- Aisy, A. R., Rahmadden, Siregar, A. H., Daulay, S., & Ananta, N. (2025). IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASIKAN BERITA HOAX DI INDONESIA PADA MEDIA SOSIAL X. *Jurnal RESTIKOM : Riset Teknik Informatika Dan Komputer*, 7(2), 206–216. <https://doi.org/10.52005/restikom.v7i2.457>
- Akbar, I., Marwondo, M., & Nugraha, N. (2023). Optimasi Algoritma Support Vector Machine untuk Analisis Klasifikasi Teks Pemintaan Informasi di Platform Online Shop. *Jurnal Accounting Information System (AIMS)*, 6(2), 119–126. <https://doi.org/10.32627/aims.v6i2.819>
- Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER. *Journal of Information System Management (JOISM)*, 3(1), 16–21. <https://doi.org/10.24076/joism.2021v3i2.558>
- Amara, S., Novriyenni, N., & Khadapi, M. (2024). Analisis Sentimen Masyarakat terhadap Program Makan Siang Gratis di Indonesia Tahun 2024 Menggunakan Long Short-Term Memory (LSTM). *Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika*, 3(4), 150–160. <https://doi.org/10.61132/mercurius.v3i4.930>
- Annisa, L., & Kalifia, A. D. (2024). Analisis Teknik TF-IDF Dalam Identifikasi Faktor-Faktor Penyebab Depresi Pada Individu. *Gudang Jurnal Multidisiplin Ilmu*, 2(1), 302–307. <https://doi.org/10.59435/gjmi.v2i1.249>
- Arisandi, D., Sutrisno, T., & Kurniawan, I. (2023). KLASIFIKASI OPINI MASYARAKAT DI TWITTER TENTANG KEBOCORAN DATA YANG TERJADI DI INDONESIA MENGGUNAKAN ALGORITMA SVM. *Jurnal Teknika*, 15(2), 75–80. <https://doi.org/10.30736/jt.v15i2.993>
- Aziz, T. A., Ismayadi, I., & Budiman, B. (2025). Analisis Sentimen Terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Logistic Regression dan SVM. *In Search (Informatic, Science, Entrepreneur, Applied Art, Research, Humanism)*, 24(1), 18–28. <https://doi.org/10.37278/insearch.v24i1.1238>
- Fathoni, F., Ibrahim, A., Rizka Mumtaz, F., Azmi Zaky, M., Jodi Pratama, M., & Akbar Kurniawan, I. (2025). ANALISIS SENTIMEN PUBLIC

TWITTER TERHADAP KEBIJAKAN PEMERINTAH MENGGUNAKAN METODE SVM (STUDI KASUS: RUU TNI ). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(4), 6322–6329. <https://doi.org/10.36040/jati.v9i4.14036>

Fazri, M., & Voutama, A. (2025). ANALISIS SENTIMEN PUBLIK TERHADAP DANANTARA DI MEDIA SOSIAL X MENGGUNAKAN NLP DAN PEMBELAJARAN MESIN. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, 9(1), 197–206.

Firdaus, M. F., Ratnawati, D. E., & Setiawan, N. Y. (2024). Analisis Sentimen Berbasis Aspek Pada Ulasan Pelanggan Restoran Menggunakan Algoritma Support Vector Machine (Studi Kasus: Depot Bamara). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(6), 1265–1272. <https://doi.org/10.25126/jtiik.2024117564>

Gifari, O. I., Adha, Muh., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40. <https://doi.org/10.46229/jifotech.v2i1.330>

Gunadi, G. A., Raharjo, J. S., Setianingsih, S., & Amazihono, M. (2025). Analisis Kemanfaatan Kebijakan Program Makan Siang Gratis bagi Peserta Didik dan Pemerintahan. *Jurnal Pendidikan Tambusai*, 9. <https://doi.org/10.31004/jptam.v9i1.25675>

Ilham, M., & Priambodo, B. (2024). Analisis Sentimen Publik Terhadap Program Makan Siang Gratis Menggunakan BERT Neural Network Pada Platform X. *Jurnal Ekonomi Manajemen Sistem Informasi*, 6(2), 1039–1047. <https://doi.org/10.38035/jemsi.v6i2.3376>

Khairunnisa, S., Adiwijaya, A., & Faraby, S. A. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406–414. <https://doi.org/10.30865/mib.v5i2.2835>

Ma'rufudin, M., & Yudhistira, A. (2025a). Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 845–857. <https://doi.org/10.52436/1.jpti.717>

Ma'rufudin, M., & Yudhistira, A. (2025b). Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 845–857. <https://doi.org/10.52436/1.jpti.717>

- Merlinda, A. A., & Yusuf, Y. (2025). Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan. *Ranah Research : Journal of Multidisciplinary Research and Development*, 7(2), 1364–1373. <https://doi.org/10.38035/rj.v7i2.1360>
- Nico, N., Budiyo, U., & Fatimah, T. (2022). Implementasi Algoritma Pembobotan TF-IDF dan Cosine Similarity untuk Penetapan Kategori Artikel pada Website Universitas Budi Luhur. *Jurnal Ticom: Technology of Information and Communication*, 10(3), 218–223. <https://doi.org/10.70309/ticom.v10i3.38>
- Perdana, S. A. P., Aji, T. B., & Ferdiana, R. (2021). Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(3), 229–235. <https://doi.org/10.22146/jnteti.v10i3.1819>
- Pohan, R. F. R., Ratnawati, D. E., & Arwani, I. (2022). Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(10), 4924–4931.
- Pradhana, R. M., Sunarto, M. J. D., & Lemantara, J. (2021). Analisis Sentimen Publik Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Skala Mikro Menggunakan Algoritma Support Vector Machine (Studi Kasus: Twitter). *Jurnal Sistem Informasi Universitas Dinamika*, 10(4). <https://jurnal.dinamika.ac.id/index.php/jsika/article/view/3977>
- Pratama, Y. T., Bachtiar, F. A., & Setiawan, N. Y. (2018). Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(12), 6244–6252.
- Pratiwi, A., Sipangkar, S. W. R., Ramadhani, A. N. R., Mulyady, S., Hylmi, M. R., & Rahmawati, R. (2025). ANALISIS IMPLEMENTASI PROGRAM MAKAN BERGIZI GRATIS DALAM KERANGKA TEORI MANDAT HANNA F. PITKIN. *Triwikrama: Jurnal Ilmu* 7 (https://ejournal.warunayama.org/index.php/triwikrama/article/view/12252)
- Pratiwi, D., Asrianda, A., & Rosnita, L. (2024). Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang. *Jurnal Ilmu Komputer dan Informatika*, 4(2), 85–96. <https://doi.org/10.54082/jiki.169>

- Purwanti, Z. & Sugiyono. (2024). Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(3), 3065–3079. <https://doi.org/10.35870/jimik.v5i3.1001>
- Putra, K. T., Hariyadi, M. A., & Crysdian, C. (2023). PERBANDINGAN FEATURE EXTRACTION TF-IDF DAN BOW UNTUK ANALISIS SENTIMEN BERBASIS SVM. *JURNAL CAHAYA MANDALIKA*, 3(2). <https://ojs.cahayamandalika.com/index.php/jcm/article/view/2292>
- Putra, S. A., & Wijaya, A. (2023). ANALISIS SENTIMEN ARTIFICIAL INTELLIGENCE (AI) PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE LEXICON BASED. *JuSiTik : Jurnal Sistem dan Teknologi Informasi Komunikasi*, 7(1), 21–28. <https://doi.org/10.32524/jusitik.v7i1.1042>
- Qhabib, F. A., Fauzan, A. C., & Harliana, H. (2023). Implementasi Algoritma Term Frequency Inverse Document Frequency (TF-IDF) dalam Menganalisis Sentimen Masyarakat Terhadap Covid-19 Varian Omicron. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 308–318. <https://doi.org/10.35746/jtim.v4i4.233>
- Rahman, R. A., Pranatawijaya, V. H., & Sari, N. N. K. (2024). Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 4(1), 70–82. <https://doi.org/10.24002/konstelasi.v4i1.8922>
- Rianto, Mutiara, A. B., Wibowo, E. P., & Santosa, P. I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, 8(1), 26. <https://doi.org/10.1186/s40537-021-00413-1>
- Rifaldi, D., Fadlil, A., & Herman. (2023a). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.” *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>
- Rifaldi, D., Fadlil, A., & Herman. (2023b). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.” *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>
- Romindo, R., Pangaribuan, J. J., & Barus, O. P. (2023). IMPLEMENTASI ALGORITMA TF-IDF DAN SUPPORT VECTOR MACHINE TERHADAP ANALISIS PENDETEKSI KOMENTAR CYBERBULLYING DI MEDIA SOSIAL TIKTOK. *Device*, 13(1), 124–134. <https://doi.org/10.32699/device.v13i1.5260>


- Sabrila, T. S., Sari, V. R., & Minarno, A. E. (2021). Analisis Sentimen Pada Tweet Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K-Nearest Neighbor. *Fountain of Informatics Journal*, 6(2), 69 <https://doi.org/10.21111/fij.v6i2.5536>
- Saka, H. K., & Prasetyaningrum, P. T. (2025). Sentiment Analysis and Classification of User Reviews of the “Access by KAI” Application Using Machine Learning Methods to Improve Service Quality. *Journal of Information Systems and Informatics*, 7(2), 1418–1442. <https://doi.org/10.51519/journalisi.v7i2.1099>
- Salam, R. R., Jamil, M. F., Ibrahim, Y., Rahmaddeni, R., Soni, S., & Herianto, H. (2023). Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine: Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 27–35. <https://doi.org/10.57152/malcom.v3i1.590>
- Salsabila, S., Tyas, S. M. P., Romadhona, Y., & Purwitasari, D. (2023). Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19. *Journal of Information Systems Engineering and Business Intelligence*, 9(1), 84–94. <https://doi.org/10.20473/jisebi.9.1.84-94>
- Sejati, P. T., Alzami, F., Marjuni, A., Indrayani, H., & Puspitarini, I. D. (2024). Aspect-Based Sentiment Analysis for Enhanced Understanding of “Kemenkeu” Tweets. *Journal of Applied Informatics and Computing*, 8(2), 487–498. <https://doi.org/10.30871/jaic.v8i2.8558>
- Septiani, D., & Isabela, I. (2022). ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS. *Sistem Dan Teknologi Informasi Indonesia (SINTESIA)*, 1(2), 81–88.
- Sinulingga, J. E. B., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, 14(1), 42–53. <https://doi.org/10.34010/jamika.v14i1.11946>
- Sitanggang, A., Umaidah, Y., Umaidah, Y., Adam, R. I., & Adam, R. I. (2024). ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN SIANG GRATIS PADA MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4902>
- Suharman, A., & Kamayani Sulaeman, M. (2025). Analisis Sentimen Pengguna Aplikasi Livin’ by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec. *Jurnal*

- Pendidikan Dan Teknologi Indonesia*, 5(8), 2201–2212. <https://doi.org/10.52436/1.jpti.941> Sukwadi, R., Silitonga, R. M., Wahyuni, M. M., Octavian, F., Jou, Y.-T., & Thu, N. T. B. (2025). Peningkatan Kualitas Layanan Jaringan Restoran Cepat Saji Indonesia: Analisis Sentimen dan Emosi Berbasis Aspek. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(2), 359–368. <https://doi.org/10.25126/jtiik.2025129416>
- Supian, A., Revaldo, B. T., Marhadi, N., Rahmadden, R., & Efrizoni, L. (2024). Penerapan SVM dan Word2Vec untuk Analisis Sentimen Ulasan Pengguna Aplikasi DANA. *Jurnal Ilmiah Komputasi*, 23(3). <https://doi.org/10.32409/jikstik.23.3.3642>
- Surat Al-Hasyr Ayat 18: Arab, Latin, Terjemah dan Tafsir Lengkap | Quran NU Online*. (n.d.). Retrieved November 20, 2025, from <https://quran.nu.or.id/al-hasyr/18>
- Surat Al-Hujurat Ayat 6: Arab, Latin, Terjemah dan Tafsir Lengkap | Quran NU Online*. (n.d.). Retrieved November 20, 2025, from <https://quran.nu.or.id/al-hujurat/6>
- Surat Al-Isra' Ayat 27: Arab, Latin, Terjemah dan Tafsir Lengkap | Quran NU Online*. (n.d.). Retrieved November 20, 2025, from <https://quran.nu.or.id/al-isra'/27>
- Turjaman, R. M., & Budi, I. (2022). ANALISIS SENTIMEN BERBASIS ASPEK MARKETING MIX TERHADAP ULASAN APLIKASI DOMPET DIGITAL (STUDI KASUS: APLIKASI LINKAJA PADA TWITTER). *Jurnal Darma Agung*, 30(2), 266–275. <https://doi.org/10.46930/ojsuda.v30i2.1672>
- Widyadhana, F. K., Setiawan, N. Y., & Rahayudi, B. (2023). Sentimen Analysis pada Opini Masyarakat terhadap Pelayanan Publik Polres Ponorogo menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(7), 3047–3056.
- Wulandari, R. F. T., & Anubhakti, D. (2021). IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MEMPREDIKSI HARGA SAHAM PT. GARUDA INDONESIA TBK. *IDEALIS : InDonEsiA journal Information System*, 4(2), 250–256. <https://doi.org/10.36080/idealis.v4i2.2847>

# LAMPIRAN



## Lampiran 1 *Graphical User Interface*

 **Sistem Analisis Sentimen**  
TF-IDF & Support Vector Machine (Skenario 8)  
Masukkan teks ulasan untuk mengetahui sentimen


Masukkan Teks Ulasan

Hasil Prediksi

Contoh: Pelayanan sangat buruk dan mengecewakan

Sentimen

Prediksi Sentimen

 **Sistem Analisis Sentimen**  
TF-IDF & Support Vector Machine (Skenario 8)  
Masukkan teks ulasan untuk mengetahui sentimen

Masukkan Teks Ulasan

Hasil Prediksi

@Lurah\_Funabashi @M45Broo\_ Sebagai manusia hipotetis yang memilih 02 sapa tidak menuesal. Tingkat persetujuan Prabowo mencapai 81% (Jan 2025 Reuters) didorong program makan siang gratis dan kebijakan pro-rakyat. Ekonomi tumbuh 5% meski ada tantangan fiskal dan pengangguran naik. Kekhawatiran demokrasi

Sentimen

Prediksi Sentimen