

**IMPLEMENTASI *HYBRID LEXICON-BASED* DAN SVM UNTUK  
KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN  
BBPSDMP KOMINFO MAKASSAR**

**SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Mendapat Gelar Sarjana Komputer  
(S.Kom) Program Studi Informatika



**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MAKASSAR  
2025**

**IMPLEMENTASI HYBRID LEXICON-BASED DAN SVM UNTUK  
KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN  
BBPSDMP KOMINFO MAKASSAR**

Diajukan Sebagai Salah Satu Syarat Untuk Mendapat Gelar Sarjana Komputer  
(S.Kom) Program Studi Informatika



**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MAKASSAR  
2025**



MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH MAKASSAR

## FAKULTAS TEKNIK



### PENGESAHAN

Skripsi atas nama NUR ALAM dengan nomor induk Mahasiswa 105841103621, dinyatakan diterima dan disahkan oleh Panitia Ujian Tugas Akhir/Skripsi sesuai dengan Surat Keputusan Dekan Fakultas Teknik Universitas Muhammadiyah Makassar Nomor : 0004/SK-Y/55202/091004/2025, sebagai salah satu syarat guna memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar pada hari Sabtu, 30 Agustus 2025.

Panitia Ujian :

1. Pengawas Umum

a. Rektor Universitas Muhammadiyah Makassar

Dr. Ir. H. Abd. Rakhim Nanda, S.T., M.T., IPU.

Makassar,

17 Rabi'ul Awal 1447 H

10 September 2025 M

b. Dekan Fakultas Teknik Universitas Hasanuddin

Prof. Dr. Eng. Muhammad Jeran Ramli, S.T., M.T., ASEAN., Eng.

2. Penguji

a. Ketua : Dr. Ir. Zahir Zainuddin, M.Sc.

b. Sekertaris : Darniati, S.Kom., M.T.

3. Anggota

: 1. Rinal Rezkiawan B., S.Kom., M.T.

2. Muhyiddin A.M Hayati, S.Kom., M.T.

3. Desi Anggreani, S.Kom., M.T.

Mengetahui,

Pembimbing I

Pembimbing II

Ir. Muhammad Faisal, S.SI., M.T., Ph.D., IPM.

Rizki Yusliana Bakti, S.T., M.T.

Dekan





MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH MAKASSAR

## FAKULTAS TEKNIK



### HALAMAN PENGESAHAN

Tugas Akhir ini diajukan untuk memenuhi syarat ujian guna memperoleh gelar Sarjana Komputer (S.Kom) Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.

Judul Skripsi : **IMPLEMENTASI HYBRID LEXICON-BASED DAN SVM UNTUK KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN BBPSDMP KOMINFO MAKASSAR**

Nama : NUR ALAM  
Stambuk : 105 84 11036 21

Makassar, 10 September 2025



Mengetahui,

Ketua Prodi Informatika



## MOTTO DAN PERSEMBAHAN

### *Motto*

“Aku lahir dengan mempertaruhkan nyawa ibuku, maka mustahil jika hidupku tidak berarti. Aku tumbuh dari lelah ayahku, dan aku pastikan setiap lelahnya tidak akan pernah sia-sia.”

### *Persembahan*

Dengan penuh rasa syukur kepada Allah SWT, karya sederhana ini penulis persembahkan kepada:

Ayahanda tercinta, M. Rusdi, dan Ibunda tersayang, Kasma, yang selalu merawat dengan penuh kasih sayang dan memberikan dukungan tanpa henti, baik dalam bentuk material maupun nonmaterial. Terima kasih atas doa-doa yang tidak pernah putus, yang senantiasa menyelimuti langkah penulis dengan ketenangan dan kekuatan. Di saat penulis menghadapi masa-masa sulit, cinta dan doa Ayah serta Ibu menjadi pelita yang menuntun agar tetap tegar dan tidak menyerah. Tanpa kasih, cinta, dan pengorbanan Ayah serta Ibu, penulis mungkin tidak akan mampu melewati setiap cobaan yang ada. Segala upaya dan kerja keras ini, dengan sepenuh hati dan penuh pengorbanan, penulis persembahkan sebagai bentuk rasa hormat dan terima kasih yang mendalam.

Untuk diri sendiri, sebagai pengingat bahwa perjuangan ini adalah bukti bahwa kesabaran, ketekunan, dan doa dapat mengantarkan pada keberhasilan. Untuk keluarga, sahabat, dan semua pihak yang turut memberi semangat serta doa tulus, sehingga penulis mampu menapaki perjalanan panjang ini hingga akhir.

Dedikasi ini adalah bukti kecil betapa besar peran Ayahanda dan Ibunda dalam setiap langkah yang penulis jalani. Semoga karya ini menjadi kebanggaan sederhana bagi orang tua tercinta dan menjadi doa terbaik dari seorang anak untuk kedua orang tuanya.

## ABSTRAK

**NUR ALAM**, Implementasi *Hybrid Lexicon-Based* dan SVM Untuk Klasifikasi Analisis Sentimen Terhadap Pelatihan BBPSDMP KOMINFO Makassar. (Dibimbing Oleh Ir. Muhammad Faisal, S.SI., M.T., Ph.D., IPM dan Rizki Yusliana Bakti, S.T., M.T

Evaluasi program pelatihan pada instansi BBPSDMP Kominfo Makassar menghasilkan umpan balik kualitatif dari peserta, namun pengolahannya secara manual tidak efisien dan rentan terhadap bias. Penelitian ini bertujuan untuk mengimplementasikan metode *hybrid lexicon-based* dan *Support Vector Machine* untuk mengklasifikasikan sentimen secara otomatis terhadap umpan balik peserta. Selain itu, penelitian ini juga membandingkan kinerja model SVM standar dengan model SVM yang parameternya dioptimasi menggunakan *Particle Swarm Optimization*. Data yang digunakan adalah 2.313 ulasan kritik dan saran dari peserta pelatihan *Vocational School Graduate Academy* yang mencakup empat aspek: fasilitas, makanan, pengajar, dan penguji. Proses penelitian meliputi tahapan pra-pemrosesan teks, pelabelan sentimen awal menggunakan kamus *InSet Lexicon*, pembobotan fitur dengan TF-IDF, dan klasifikasi menggunakan SVM dengan kernel RBF. Hasil penelitian menunjukkan bahwa penerapan optimasi PSO secara konsisten meningkatkan performa model di seluruh kategori. Peningkatan akurasi paling signifikan tercatat pada kategori Pengajar yang naik dari 84,71% menjadi 89,02% dan kategori Penguji yang meningkat dari 87,44% menjadi 91,46%. Keunggulan utama model SVM yang dioptimasi terletak pada kemampuannya yang lebih andal dalam mendekripsi kelas minoritas. Hal ini terbukti dari peningkatan signifikan pada nilai recall. Disimpulkan bahwa metode *hybrid* dengan optimasi PSO efektif dalam membangun model klasifikasi sentimen yang akurat dan sensitif.

**Kata Kunci:** Analisis Sentimen, *Hybrid Method*, *Lexicon-Based*, *Support Vector Machine*, *Particle Swarm Optimization*, Evaluasi Pelatihan.

## ***ABSTRACT***

**NUR ALAM**, *Implementation of a Hybrid Lexicon-Based and SVM Method for Sentiment Analysis Classification of Training at BBPSDMP KOMINFO Makassar.* (Supervised by Ir. Muhammad Faisal, S.SI., M.T., Ph.D., IPM and Rizki Yusliana Bakti, S.T., M.T)

*The evaluation of training programs at the BBPSDMP Kominfo Makassar institution generates qualitative participant feedback, but its manual processing is inefficient and prone to bias. This research aims to implement a hybrid lexicon-based and Support Vector Machine method to automatically classify sentiment from participant feedback. Furthermore, this study compares the performance of a standard SVM model with an SVM model whose parameters were optimized using Particle Swarm Optimization. The dataset consists of 2,313 critique and suggestion reviews from participants of the Vocational School Graduate Academy, covering four aspects: facilities, food, instructors, and assessors. The research process involves text preprocessing, initial sentiment labeling using the InSet Lexicon, feature weighting with TF-IDF, and classification using an SVM with an RBF kernel. The results demonstrate that the application of PSO optimization consistently improves model performance across all categories. The most significant accuracy improvements were observed in the Instructor category, rising from 84.70% to 89.01%, and the Assessor category, which increased from 87.43% to 91.45%. The primary advantage of the optimized SVM model lies in its more reliable ability to detect the minority class, as evidenced by a significant increase in recall values. It is concluded that the hybrid method with PSO optimization is effective in building an accurate and sensitive sentiment classification model.*

**Keywords:** *Sentiment Analysis, Hybrid Method, Lexicon-Based, Support Vector Machine, Particle Swarm Optimization, Training Evaluation.*

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji senantiasa kita panjatkan kepada Allah SWT yang senantiasa melimpahkan rahmat, ridha, dan karunianya. Shalawat serta salam semoga tercurahkan kepada junjungan kita Nabi Muhammad SAW yang menjadi suri teladan seluruh umat. Sehingga penulis dapat menyelesaikan proposal skripsi dengan judul **"IMPLEMENTASI HYBRID LEXICON-BASED DAN SVM UNTUK KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN BBPSDMP KOMINFO MAKASSAR"**.

Penulisan skripsi ini disusun sebagai bagian dari persyaratan untuk menyelesaikan program sarjana (S1) di program studi Informatika, Fakultas Teknik Universitas Muhammadiyah Makassar. Pada kesempatan ini, penulis ingin menyampaikan rasa syukur dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Ir. H. Abd. Rakhim Nanda, S.T., M.T., IPU. Sebagai Rektor Universitas Muhammadiyah Makassar
2. Bapak Ir. Muhammad Syafaat S Kuba, ST., MT. Sebagai Dekan Fakultas Teknik Universitas Muhammadiyah Makassar.
3. Ibu Rizki Yusliana Bakti, S.T., M.T, selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar.
4. Bapak Ir. Muhammad Faisal, S.SI., M.T., Ph.D., IPM selaku Dosen Pembimbing I.
5. Ibu Rizki Yusliana Bakti, S.T., M.T, juga selaku Dosen Pembimbing II.
6. Dosen dan Staff Fakultas Teknik Universitas Muhammadiyah Makassar.

7. Kedua orang tua tercinta yang selalu memberikan semangat dan dukungan baik berupa moral, materi, serta spiritual agar terselesaikannya penulisan skripsi ini.
8. Teman-teman Informatika Kelas B Fakultas Teknik Universitas Muhammadiyah Makassar. Khususnya, Syahril Akbar, Muhammad Asygar Faeruddin, Muh Al Iqram Marzah, Gempar Perkasa Tahir, Muh Ulil Amri dan Muhammad Adil Syaputra.
9. Kepada semua pihak yang sudah membantu, Penulis mengucapkan terima kasih yang sebesar-besarnya.

Semoga Tuhan Yang Maha Esa melimpahkan segala nikmat yang lebih besar kepada beliau. Akhir kata, semoga laporan ini dapat memberikan manfaat dan barokah kepada pembaca secara umum dan terkhusus bagi penulis.

**Billahi fisabililhaq, fastabiqul khairat.**

**Wassalamualaikum Warahmatullahi Wabarakatuh.**

Makassar, 30 Agustus 2025

Nur Alam

## DAFTAR ISI

ABSTRAK .....	iii
<i>ABSTRACT</i> .....	iv
KATA PENGANTAR.....	v
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	ix
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN .....	xiii
DAFTAR ISTILAH .....	xiv
BAB I PENDAHULUAN .....	1
A. Latar Belakang .....	1
B. Rumusan Masalah .....	3
C. Tujuan Penelitian.....	4
D. Manfaat Penelitian .....	4
E. Ruang Lingkup Penelitian.....	5
F. Sistematika Penulisan .....	6
BAB II TINJAUAN PUSTAKA .....	8
A. Landasan Teori .....	8
B. Penelitian Terkait.....	19
C. Kerangka Pikir .....	22
BAB III METODE PENELITIAN.....	23
A. Tempat dan Waktu Penelitian.....	23

B.	Alat dan Bahan.....	23
C.	Perancangan Sistem .....	23
D.	Teknik Pengujian Sistem.....	36
E.	Teknik Analisis Data .....	38
BAB VI HASIL DAN PEMBAHASAN.....		40
A.	Deskripsi Dataset .....	40
B.	Pra-Pemrosesan .....	42
C.	Pelabelan Sentimen .....	62
D.	Model Klasifikasi <i>Support Vector Machine</i> .....	70
E.	Perhitungan Manual Model Klasifikasi SVM.....	96
F.	Analisis Perbandingan Kinerja Model .....	99
G.	Interpretasi Hasil Analisis Sentimen .....	102
BAB V KESIMPULAN DAN SARAN.....		112
A.	Kesimpulan .....	112
B.	Saran.....	113
DAFTAR PUSTAKA.....		115
LAMPIRAN.....		119

## DAFTAR GAMBAR

Gambar 1. <i>Hyperplane Illustration of SVM</i> (Taqiuddin et al., 2021) .....	14
Gambar 2. Kerangka Berpikir .....	22
Gambar 3. Rancangan Penelitian .....	24
Gambar 4. Pra-pemrosesan Data.....	27
Gambar 5. Pelabelan Sentimen .....	29
Gambar 6. Model Klasifikasi SVM .....	32
Gambar 7. Klasifikasi SVM dengan Optimasi PSO .....	34
Gambar 8. <i>Confusion Matrix</i> .....	36
Gambar 9. Distribusi Sentimen .....	68
Gambar 10. Jumlah Data Latih dan Uji per Kategori .....	72
Gambar 11. <i>Heatmap Confusion Matrix</i> SVM Ulasan Fasilitas .....	74
Gambar 12. <i>Heatmap Confusion Matrix</i> SVM Ulasan Makanan .....	77
Gambar 13. <i>Heatmap Confusion Matrix</i> SVM Ulasan Pengajar .....	79
Gambar 14. <i>Heatmap Confusion Matrix</i> SVM Ulasan Pengaji .....	82
Gambar 15. <i>Heatmap Confusion Matrix</i> SVM+PSO Ulasan Fasilitas .....	87
Gambar 16. <i>Heatmap Confusion Matrix</i> SVM+PSO Ulasan Makanan.....	89
Gambar 17. <i>Heatmap Confusion Matrix</i> SVM+PSO Ulasan Pengajar .....	92
Gambar 18. <i>Heatmap Confusion Matrix</i> SVM+PSO Ulasan Pengaji .....	95
Gambar 19. Perbandingan Kinerja Model .....	101
Gambar 20. <i>Word cloud</i> Sentimen Positif Fasilitas.....	103
Gambar 21. <i>Word cloud</i> Sentimen Negatif Fasilitas .....	104
Gambar 22. <i>Word cloud</i> Sentimen Positif Makanan .....	105
Gambar 23. <i>Word cloud</i> Sentimen Negatif Makanan.....	106
Gambar 24. <i>Word cloud</i> Sentimen Positif Pengajar .....	107
Gambar 25. <i>Word cloud</i> Sentimen Negatif Pengajar .....	108
Gambar 26. <i>Word cloud</i> Sentimen Positif Pengaji.....	109

Gambar 27. *Word cloud* Sentimen Negatif Pengaji ..... 110



## DAFTAR TABEL

Tabel 1. Fungsi Kernel .....	15
Tabel 2. Penelitian Terkait.....	19
Tabel 3. Data Survey Evaluasi .....	41
Tabel 4. <i>Cleaning</i> Data Ulasan Fasilitas .....	43
Tabel 5. <i>Cleaning</i> Data Ulasan Makanan.....	43
Tabel 6. <i>Cleaning</i> Data Ulasan Pengajar.....	44
Tabel 7. <i>Cleaning</i> Data Ulasan Penguji .....	45
Tabel 8. <i>Case Folding</i> Ulasan Fasilitas.....	46
Tabel 9. <i>Case Folding</i> Ulasan Makanan .....	46
Tabel 10. <i>Case Folding</i> Ulasan Pengajar .....	47
Tabel 11. <i>Case Folding</i> Ulasan Penguji .....	48
Tabel 12. Normalisasi Data Ulasan Fasilitas.....	49
Tabel 13. Normalisasi Data Ulasan Makanan .....	50
Tabel 14. Normalisasi Data Ulasan Pengajar .....	51
Tabel 15. Normalisasi Data Ulasan Penguji.....	51
Tabel 16. Tokenisasi Data Ulasan Fasilitas .....	53
Tabel 17. Tokenisasi Data Ulasan Makanan .....	53
Tabel 18. Tokenisasi Data Ulasan Pengajar .....	54
Tabel 19. Tokenisasi Data Ulasan Penguji .....	55
Tabel 20. <i>Stopword Removal</i> Data Ulasan Fasilitas.....	56
Tabel 21. <i>Stopword Removal</i> Data Ulasan Makanan .....	57
Tabel 22. <i>Stopword Removal</i> Data Ulasan Pengajar .....	58
Tabel 23. <i>Stopword Removal</i> Data Ulasan Penguji.....	58
Tabel 24. <i>Stemming</i> Data Ulasan Fasilitas .....	60
Tabel 25. <i>Stemming</i> Data Ulasan Makanan.....	60
Tabel 26. <i>Stemming</i> Data Ulasan Pengajar .....	61

Tabel 27. <i>Stemming</i> Data Ulasan Pengaji .....	62
Tabel 28. Label Sentimen Ulasan Fasilitas .....	65
Tabel 29. Label Sentimen Ulasan Makanan.....	65
Tabel 30. Label Sentimen Ulasan Pengajar.....	66
Tabel 31. Label Sentimen Ulasan Pengaji .....	67
Tabel 32. Distribuasi Sentimen .....	68
Tabel 33. Rincian Pembagian Dataset Pelatihan dan Pengujian .....	72
Tabel 34. <i>Classification Report</i> Model SVM Ulasan Fasilitas .....	73
Tabel 35. <i>Classification Report</i> Model SVM Ulasan Makanan.....	76
Tabel 36. <i>Classification Report</i> Model SVM Ulasan Pengajar.....	78
Tabel 37. <i>Classification Report</i> Model SVM Ulasan Pengaji .....	81
Tabel 38. Parameter Optimal Hasil Pencarian PSO .....	84
Tabel 39. <i>Classification Report</i> Model SVM dengan PSO Ulasan Fasilitas .....	86
Tabel 40. <i>Classification Report</i> Model SVM dengan PSO Ulasan Makanan .....	88
Tabel 41. <i>Classification Report</i> Model SVM dengan PSO Ulasan Pengajar.....	91
Tabel 42. <i>Classification Report</i> Model SVM dengan PSO Ulasan Pengajar.....	94
Tabel 43. Data Latih.....	96
Tabel 44. Data Uji .....	97
Tabel 45. Perbandingan Kinerja Model .....	100
Tabel 46. Perbandingan Klasifikasi Sentimen Ulasan Fasilitas .....	104
Tabel 47. Perbandingan Klasifikasi Sentimen Ulasan Makanan .....	106
Tabel 48. Perbandingan Klasifikasi Sentimen Ulasan Pengajar .....	108
Tabel 49. Perbandingan Klasifikasi Sentimen Ulasan Pengaji .....	110

## **DAFTAR LAMPIRAN**

Lampiran 1. Dataset .....	119
Lampiran 2. <i>Source Code</i> .....	123
Lampiran 3. Surat Permohonan Penelitian dari Program Studi .....	136
Lampiran 4. Surat Permohonan Izin Pelaksanaan Penelitian dari LP3M .....	137
Lampiran 5. Surat Izin Penelitian dari Dinas Penanaman Modal dan PTSP Provinsi Sulawesi Selatan.....	138
Lampiran 6. Surat Balasan Izin Penelitian dari BBPSDMP Kominfo Makassar.....	139
Lampiran 7. Surat Keterangan Bebas Plagiat .....	140
Lampiran 8. Hasil Turnitin.....	141



## DAFTAR ISTILAH

<b>Analisis Sentimen</b>	Cabang dari <i>Natural Language Processing</i> yang berfokus pada identifikasi, ekstraksi, dan kuantifikasi opini, sentimen, atau polaritas emosional dari data teks.
<b>BBPSDMP</b>	Singkatan dari Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian, instansi pemerintah di bawah Kominfo yang menjadi lokasi dan penyedia data dalam penelitian ini.
<b>Confusion Matrix</b>	Tabel yang digunakan untuk memvisualisasikan kinerja model klasifikasi dengan menyajikan jumlah prediksi benar dan salah untuk setiap kelas.
<b>Metode Hybrid</b>	Pendekatan dalam penelitian ini yang mengintegrasikan dua metode berbeda, yaitu pendekatan <i>lexicon-based</i> untuk pelabelan awal dan machine learning SVM untuk klasifikasi akhir.
<b>Hyperplane</b>	Batas keputusan yang dibuat oleh algoritma SVM untuk memisahkan data ke dalam kelas-kelas yang berbeda.
<b>InSet Lexicon</b>	Kamus sentimen berbahasa Indonesia yang berisi daftar kata-kata beserta skor polaritasnya, yang digunakan dalam tahap pelabelan berbasis leksikon.
<b>Kernel</b>	Fungsi matematis yang digunakan oleh SVM untuk memetakan data ke ruang berdimensi

		lebih tinggi, memungkinkan pemisahan data yang tidak dapat dipisahkan secara linear. Dalam penelitian ini, kernel yang digunakan adalah <i>Radial Basis Function</i> .
<b>Lexicon-Based</b>		Pendekatan analisis sentimen yang mengandalkan kamus yang telah diberi label polaritas untuk menentukan sentimen suatu teks.
<b>Particle Swarm Optimization (PSO)</b>		Algoritma optimasi yang terinspirasi dari perilaku sosial kawanan. Dalam penelitian ini, PSO digunakan untuk mencari kombinasi parameter C dan gamma yang optimal bagi model SVM.
<b>Pra-pemrosesan Teks</b>		Serangkaian langkah awal untuk membersihkan dan mempersiapkan data teks mentah agar siap diolah oleh model, meliputi <i>cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming</i> .
<b>TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)</b>		Metode pembobotan statistik yang digunakan untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen relatif terhadap keseluruhan korpus data.
<b>VSGA</b>		Singkatan dari <i>Vocational School Graduate Academy</i> , nama program pelatihan yang diselenggarakan oleh BBPSDMP Kominfo dan menjadi objek evaluasi dalam penelitian ini.

<b>Akurasi (Accuracy)</b>	Metrik evaluasi yang mengukur proporsi total prediksi yang benar dari keseluruhan data uji.
<b>F1-Score</b>	Rata-rata harmonik dari <i>Precision</i> dan <i>Recall</i> . Metrik ini memberikan gambaran kinerja yang seimbang, terutama pada kasus dataset dengan distribusi kelas yang tidak seimbang.
<b>Presisi (<i>Precision</i>)</b>	Metrik evaluasi yang mengukur tingkat ketepatan prediksi.
<b>Recall</b>	Metrik evaluasi yang mengukur kemampuan model dalam menemukan kembali semua data positif yang sebenarnya.



## **BAB I**

### **PENDAHULUAN**

#### **A. Latar Belakang**

Perkembangan teknologi informasi dan komunikasi telah membawa transformasi signifikan dalam berbagai aspek kehidupan, tidak terkecuali pada sektor penyelenggaraan program-program pemerintah yang bertujuan meningkatkan kualitas sumber daya manusia. Menurut Hendrawati et al. (2024), kemajuan teknologi dan informasi secara umum telah memberikan dampak positif bagi masyarakat Indonesia, yang salah satunya termanifestasi dalam upaya peningkatan kapabilitas individu melalui program-program strategis.

Salah satu program yang diinisiasi pemerintah oleh Kementerian Komunikasi dan Informatika melalui Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian (BBPSDMP) Kominfo Makassar yaitu program pelatihan *Vocational School Graduate Academy* (VSGA). Program tersebut merupakan bagian dari inisiatif strategis nasional, yaitu *Digital Talent Scholarship* yang dirancang untuk membekali angkatan kerja muda dengan keterampilan digital yang relevan sebagai respons terhadap percepatan transformasi digital nasional (Kominfo, 2021). Dalam pelaksanaannya, evaluasi program menjadi sebuah komponen penting untuk mengukur efektivitas pelaksanaan dan dampak yang dihasilkan bagi peserta.

Berdasarkan observasi selama melaksanakan magang pada BBPSDMP Kominfo Makassar, penulis menemukan bahwa proses pengolahan data *monitoring* dan evaluasi masih menghadapi kendala signifikan, diantaranya kritik dan saran yang bersifat teks tidak terstruktur. Umpatan balik tersebut mencerminkan persepsi peserta terhadap empat aspek penting pelatihan antara lain fasilitas, konsumsi, pengajar, dan penguji. Namun, karena format kritik dan saran yang naratif dan informal, menjadi tidak efisien dan rawan bias interpretasi jika dikelola melalui analisis manual (Syahputra, 2021).

Data kualitatif berupa kritik dan saran dari peserta pelatihan biasanya bersifat tidak terstruktur, menggunakan ragam bahasa yang bervariasi, dan tidak jarang mengandung noise atau elemen-elemen yang tidak relevan seperti kata-kata tidak baku atau ungkapan emosional. Peneliti terdahulu menyatakan bahwa variasi bentuk bahasa yang tinggi dan kurangnya struktur formal, memperumit proses pengolahan dan pemaknaan informasi (Nurmalasari et al., 2024). Hal tersebut menyebabkan proses analisis manual menjadi kurang efisien, serta berpotensi menimbulkan bias atau subjektivitas dari pihak penilai. Faktanya, kritik dan saran yang disampaikan peserta umumnya mencerminkan evaluasi terhadap pelatihan VSGA yang merupakan indikator penting dalam mengukur tingkat kepuasan peserta. Dalam rangka mempertimbangkan peran penting kritik dan saran sebagai sumber informasi evaluatif, maka dibutuhkan pendekatan komputasional yang mampu mengotomatisasi proses analisis dan klasifikasi sentimen terhadap data teks secara objektif dan terukur.

Salah satu solusi yang dapat diimplementasikan untuk mengatasi permasalahan pengelolaan dan analisis data kritik serta saran yang tidak terstruktur adalah melalui teknik analisis sentiment, berupa pendekatan komputasional yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengkaji keadaan afektif serta informasi subjektif melalui analisis teks (Savira et al., 2023). Pendekatan *lexicon-based* merupakan metode yang memanfaatkan kamus sentimen yang telah diberi label polaritas, seperti positif, negatif, atau netral, untuk memberikan bobot pada setiap kata atau frasa dalam suatu teks. Skor yang diperoleh dari masing-masing kata kemudian diakumulasi untuk menentukan kecenderungan sentimen secara keseluruhan, sehingga memungkinkan pelabelan yang cepat dan konsisten (I Kadek Arya Budi Artana et al., 2023). Sementara, SVM merupakan algoritma *machine learning* yang dirancang untuk menemukan *hyperplane* optimal yang mampu memisahkan kelas-kelas data dengan margin maksimum khususnya pada data berdimensi tinggi (Nauli et al., 2025).

Dengan keunggulan karakteristik masing-masing metode, integrasi antara *lexicon-based* dan SVM dinilai relevan untuk diterapkan dalam analisis sentimen. Penggunaan *lexicon-Based* memungkinkan pelabelan data secara lebih cepat dan konsisten, sementara SVM memiliki kemampuan yang baik dalam klasifikasi data, khususnya untuk data dengan dimensi tinggi (Rizky Pratama et al., 2023). Pendekatan analisis sentimen dengan menggunakan gabungan metode *Lexicon-Based* dan *Support Vector Machine* (SVM) telah terbukti efektif dalam mengolah opini publik dari media sosial (Muhammad et al., 2022).

Mempertimbangkan urgensi permasalahan yang ditemukan, penelitian ini bertujuan untuk mengembangkan sebuah model analisis sentimen menggunakan metode *hybrid Lexicon-Based* dan SVM. Model yang dibangun akan diaplikasikan pada proses klasifikasi terhadap kritik dan saran peserta pelatihan VSGA. Adapun aspek yang terlibat pada proses penerapan model antara lain kritik dan saran terkait Ruangan dan Fasilitas Pelatihan, Makanan, Pengajar, dan Penguji. Proses penelitian akan mencakup beberapa tahapan penting, mulai dari pengumpulan data, pra-pemrosesan teks untuk membersihkan dan menstandarisasi data, pelabelan sentimen menggunakan kamus *lexicon*, hingga pelatihan model SVM dan evaluasi performa klasifikasinya. Hasil penelitian ini tidak hanya mampu menyajikan gambaran tingkat kepuasan peserta pelatihan secara akurat dan komprehensif, tetapi juga dapat memberikan rekomendasi berbasis data kepada pihak penyelenggara guna merancang strategi perbaikan dan peningkatan kualitas program pelatihan di masa mendatang.

## B. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat dirumuskan beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana implementasi metode *hybrid lexicon-based* dan SVM terhadap proses klasifikasi sentimen peserta program pelatihan VSGA BBPSDMP Kominfo Makassar?

2. Bagaimana perbedaan hasil klasifikasi sentimen antara model *hybrid lexicon-based* dan SVM menggunakan kernel RBF dengan model SVM yang dioptimasi menggunakan *Particle Swarm Optimization*?

### C. Tujuan Penelitian

Sejalan dengan rumusan masalah yang telah diajukan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *hybrid lexicon-SVM* dalam klasifikasi sentimen pada data kritik dan saran peserta pelatihan VSGA BBPSDMP Kominfo Makassar.
2. Menganalisis hasil klasifikasi sentimen yang dihasilkan oleh model *hybrid lexicon-based* dan SVM menggunakan kernel RBF dengan model SVM yang dioptimasi menggunakan *Particle Swarm Optimization*.

### D. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang berarti, baik secara praktis maupun secara teoritis, sebagai berikut:

1. Terhadap Peneliti

Menjadi sarana untuk mengaplikasikan pengetahuan dan keterampilan teoritis yang diperoleh selama perkuliahan ke dalam penyelesaian masalah praktis di lapangan, khususnya dalam bidang analisis data teks dan machine learning.

2. Terhadap instansi BBPSDMP KOMINFO Makassar

Menyediakan sebuah model sistem analisis sentimen yang dapat digunakan sebagai penunjuk praktek pengambilan keputusan yang lebih efektif dalam mengolah umpan balik peserta pelatihan VSGA. Memberikan hasil analisis konkret mengenai tingkat kepuasan peserta terhadap berbagai aspek pelatihan, yang dapat langsung dimanfaatkan untuk merumuskan rekomendasi perbaikan program.

3. Terhadap Universitas Muhammadiyah Makassar

Hasil penelitian dapat menjadi salah satu referensi bagi mahasiswa lain yang berminat melakukan penelitian dengan topik serupa dan menunjukkan kontribusi universitas dalam memberikan solusi berbasis teknologi bagi instansi mitra.

#### E. Ruang Lingkup Penelitian

Agar penelitian ini tetap terfokus pada jalur yang telah ditentukan dan mencapai tujuan secara efektif, maka ditetapkan batasan-batasan sebagai berikut:

1. Data yang digunakan dalam penelitian ini terbatas pada data monev kritik dan saran dari peserta pelatihan *Vocational School Graduate Academy* (VSGA) yang diselenggarakan oleh BBPSDMP Kominfo Makassar pada periode 2024. Umpulan yang dianalisis hanya mencakup aspek fasilitas, makanan, pengajar, dan pengujian.
2. Penelitian ini secara spesifik akan mengimplementasikan dan mengevaluasi pendekatan *hybrid* yang menggabungkan metode berbasis *lexicon* dengan algoritma SVM untuk klasifikasi sentimen.
3. Analisis difokuskan pada data teks berbahasa Indonesia. Teks dalam bahasa lain yang mungkin muncul dalam dataset tidak akan menjadi fokus utama pengolahan.
4. Klasifikasi sentimen akan menghasilkan tiga kategori utama, yaitu: positif, negatif, dan netral.
5. Penelitian ini menghasilkan model klasifikasi sentimen, analisis performanya, serta interpretasi hasil klasifikasi terhadap tingkat kepuasan peserta. Visualisasi data akan digunakan untuk mendukung analisis, namun pembuatan dashboard interaktif yang kompleks tidak termasuk dalam lingkup utama.

## **F. Sistematika Penulisan**

Penelitian ini disusun dalam beberapa bab dengan sistematika penulisan sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini menguraikan dasar penelitian yang meliputi latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup penelitian, dan sistematika penulisan

### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi teori-teori yang mendasari penelitian, seperti analisis sentimen, metode *Lexicon-Based*, algoritma SVM, serta studi-studi terdahulu yang relevan.

### **BAB III METODE PENELITIAN**

Bab ini menjelaskan metode penelitian yang digunakan, mencakup jenis dan pendekatan penelitian, teknik pengumpulan data, lokasi dan waktu, alat dan bahan, tahapan pra-pemrosesan data, implementasi metode *hybrid lexicon-SVM*, serta teknik analisis dan evaluasi model.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil penelitian yang diperoleh serta pembahasan secara mendalam. Bagian ini mencakup deskripsi dataset, hasil dari setiap tahapan pra-pemrosesan, hasil pelabelan sentimen, implementasi dan pengujian model SVM standar maupun model SVM yang dioptimasi dengan PSO. Lebih lanjut, bab ini juga menyajikan analisis perbandingan kinerja kedua model serta interpretasi kontekstual terhadap hasil klasifikasi sentimen pada setiap aspek yang dievaluasi.

## BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bagian penutup yang merangkum keseluruhan hasil penelitian. Bab ini terbagi menjadi dua bagian utama, yaitu kesimpulan yang ditarik berdasarkan hasil analisis data untuk menjawab rumusan masalah, serta saran yang bersifat praktis bagi instansi terkait dan saran akademis untuk pengembangan penelitian di masa mendatang.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **A. Landasan Teori**

##### **1. Analisis Sentimen**

Analisis sentimen yang juga dikenal sebagai opinion mining, merupakan sebuah pendekatan komputasional yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengolah data tekstual untuk memperoleh informasi sentimen yang terkandung di dalamnya (Damanik dan Setyohadi, 2021). Tujuan utama analisis sentimen adalah melakukan klasifikasi polaritas opini yang umumnya dikelompokkan kedalam kategori positif, negatif, atau netral yang diekspresikan dalam suatu teks (Muttaqin dan Andrika, 2025). Dengan demikian, analisis sentimen memungkinkan pengungkapan pandangan subjektif yang tersembunyi dalam volume besar data tidak terstruktur.

Sebagai sebuah bidang yang bersinggungan erat dengan *Natural Language Processing* (NLP) dan *text mining*, analisis sentimen telah mengalami perkembangan signifikan, seiring dengan meningkatnya volume data tekstual dari berbagai platform digital. Kemampuannya dalam mengekstraksi pandangan subjektif dari data tidak terstruktur menjadikannya instrumen analitis yang strategis dalam berbagai domain, mulai dari pemantauan reputasi merek hingga analisis dinamika pasar (Hadju dan Jayadi, 2021).

Pada konteks evaluasi program pelatihan, analisis sentimen berfungsi sebagai pendekatan metodologis untuk mengukur persepsi dan tingkat kepuasan peserta secara lebih objektif dan sistematis (Fadlil et al., 2024). Melalui analisis terhadap umpan balik kualitatif berupa kritik dan saran, BBPSDMP kominfo makassar dapat memperoleh pemahaman mendalam mengenai aspek-aspek program yang dinilai positif maupun yang memerlukan perbaikan. Informasi kualitatif yang telah diolah menjadi landasan berharga dalam memberikan dasar bagi peningkatan mutu layanan secara berkelanjutan dan pengambilan keputusan

berbasis data, sebagaimana relevansinya juga ditemukan dalam analisis terhadap opini pengguna aplikasi digital (Madyatmadja et al., 2024).

## 2. Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahapan awal dalam analisis teks yang bertujuan untuk mengubah data mentah menjadi format yang lebih terstruktur dan siap diolah oleh model komputasional. Tujuan utamanya adalah untuk membersihkan, menstandarisasi, dan menyiapkan data mentah sehingga dapat diolah secara efektif, agar meningkatkan performa model klasifikasi, termasuk akurasi pelabelan sentimen (Ulya et al., 2023).

Tahapan-tahapan utama dalam pra-pemrosesan data teks umumnya meliputi:

### a. *Cleaning Data*

*Cleaning data* adalah proses awal dalam pra-pemrosesan teks yang bertujuan untuk menghapus elemen-elemen yang tidak relevan, seperti tanda baca, angka, emoji, serta karakter non-alfabetik lainnya. Tahapan ini penting untuk memastikan bahwa teks yang dianalisis bersih dari gangguan sintaksis dan bebas dari noise yang dapat menurunkan akurasi pemodelan. Pembersihan data membantu menyederhanakan representasi teks dan memperbaiki kualitas ekstraksi fitur.

### b. *Case Folding*

*Case Folding* merupakan proses konversi seluruh huruf dalam teks menjadi huruf kecil. Langkah ini bertujuan untuk menyeragamkan kata-kata yang memiliki perbedaan kapitalisasi agar dianggap sebagai entitas yang sama. Sebagai contoh, kata “Pelatihan”, “pelatihan”, dan “PELATIHAn” akan diperlakukan sebagai satu token identik.

### c. Normalisasi Kata

Normalisasi kata bertujuan untuk mengubah bentuk variasi kata, seperti kata tidak baku atau slang, ke dalam bentuk standar yang sesuai dengan kaidah bahasa yang berlaku. Langkah ini diperlukan untuk mengurangi keragaman

leksikal yang tidak relevan dan memastikan konsistensi data teks dalam proses analisis.

d. *Tokenisasi*

*Tokenisasi* merupakan proses pemecahan teks atau kalimat menjadi unit-unit yang lebih kecil, yang disebut token, biasanya berupa kata individual. Tahapan ini sangat penting karena mengubah aliran teks yang kontinu menjadi segmen-segmen diskrit yang dapat dianalisis lebih lanjut oleh sistem komputasi linguistik.

e. *Stopword Removal*

*Stopword Removal* adalah proses penghapusan kata-kata umum yang sering muncul dalam teks, seperti “yang”, “di”, atau “dan”, yang dinilai tidak memberikan kontribusi substansial terhadap makna sentimen. Penghapusan ini bertujuan untuk mengurangi kompleksitas data dan meningkatkan fokus analisis terhadap kata-kata yang memiliki bobot informasi yang lebih tinggi.

f. *Stemming*

*Stemming* adalah proses mereduksi kata-kata menjadi bentuk dasarnya dengan menghilangkan imbuhan, baik awalan, akhiran, maupun sisipan. Tujuannya adalah untuk menyatukan berbagai bentuk morfologis dari suatu kata agar dapat dianalisis sebagai satu kesatuan makna yang representatif dalam proses klasifikasi.

Dalam data berbahasa Indonesia, tahapan pra-pemrosesan memiliki tantangan tersendiri. Penggunaan bahasa yang sangat variatif, termasuk kata tidak baku, singkatan, dan kosakata slang yang terus berkembang, menuntut penggunaan kamus normalisasi yang komprehensi (Alfauzan et al., 2023). Oleh karena itu, untuk proses *Stemming*, penggunaan stemmer yang dirancang khusus untuk Bahasa Indonesia, seperti Sastrawi, menjadi sangat penting untuk menghasilkan bentuk dasar kata yang akurat (Ependi dan Ahmad, 2024).

Data yang dihasilkan dari pra-pemrosesan akan sangat menentukan efektivitas proses analisis selanjutnya, seperti pelabelan sentimen menggunakan kamus *lexicon* dan ekstraksi fitur untuk pelatihan model SVM. Dengan data yang bersih dan terstandarisasi, model klasifikasi diharapkan mampu mengenali pola sentimen dengan lebih baik dan menghasilkan akurasi yang lebih tinggi, sebagaimana pentingnya tahapan ini sebelum pemodelan dengan SVM dan algoritma lainnya (Kristiyanti dan Sri Hardani, 2023).

### 3. Teks Tidak Terstruktur dan Data Kualitatif

Data kualitatif dalam bentuk teks, seperti kritik dan saran dari peserta pelatihan, umumnya bersifat sebagai data tidak terstruktur. Karakteristik utama dari teks tidak terstruktur adalah kebebasan dalam penyampaian informasi, biasanya dalam format naratif dan memuat opini, serta sudut pandang subjektif pengguna (Syahputra, 2021). Sebagaimana ditemukan pada platform digital, termasuk media sosial, yang di dalamnya mengandung berbagai elemen seperti identitas pengguna, percakapan, dan perasaan yang diekspresikan secara lugas. Kompleksitas data teks tidak terstruktur semakin meningkat akibat penggunaan bahasa yang variatif, termasuk singkatan, tagar, *emoticon*, dan tanda baca yang tidak standar, yang menjadikannya sebuah tantangan tersendiri dalam pemrosesan otomatis (Al-Khowarizmi et al., 2023).

Keberagaman bentuk dari teks tidak terstruktur dan potensi ketidakkonsistenan dalam penyampaian kritik dan saran peserta pelatihan memerlukan tahapan pra-pemrosesan yang cermat. Tahapan pra-pemrosesan bertujuan untuk mengubah data mentah menjadi format yang lebih terstruktur dan siap dianalisis, meliputi proses seperti konversi ke huruf kecil, penghapusan simbol, *Stemming* (pengembalian kata ke bentuk dasar), dan tokenisasi (Priadana dan Rizal, 2021). Dengan demikian, pendekatan khusus dalam analisis sentimen diperlukan untuk dapat mengekstraksi makna dan polaritas opini secara efektif dari kumpulan data kualitatif.

#### 4. *Lexicon-Based*

Pendekatan *lexicon-based* merupakan salah satu metode dasar dalam analisis sentimen yang mengandalkan kamus sentimen dengan kosakata yang telah dilabeli pada polaritas tertentu seperti positif, negatif, atau netral. Sebagaimana diilustrasikan dalam penelitian mengenai analisis sentimen terhadap pembelajaran tatap muka terbatas, metode *lexicon* dapat secara langsung memberikan bobot pada tweet atau teks berdasarkan kamus yang telah ada (I Kadek Arya Budi Artana et al., 2023). Proses analisis dilakukan dengan mencocokan kosakata dalam teks terhadap entri dalam kamus, kemudian dilakukan penjumlahan skor untuk menentukan kecenderungan sentimen keseluruhan. Pendekatan *lexicon* seringkali menjadi pilihan awal karena kemampuannya untuk melakukan pelabelan otomatis tanpa memerlukan korpus data latih yang besar dan teranotasi secara manual.

Salah satu keunggulan utama berbasis *lexicon* terletak pada efisiensi dan transparansi dalam proses pelabelan. Karena tidak memerlukan tahap pelatihan model yang kompleks, implementasinya cenderung lebih cepat dan sumber daya komputasi yang dibutuhkan relatif lebih rendah. Penerapan *Valence Aware Dictionary and sEntiment Reasoner* (VADER), sebagai *lexicon* yang dirancang khusus untuk menganalisis sentimen pada teks media sosial, yang kemampuannya dalam pelabelan otomatis menjadi pembanding dalam evaluasi kinerja sistem lain seperti Grok AI (Agustin et al., 2025).

Meskipun demikian, pendekatan berbasis *lexicon* juga memiliki keterbatasan yang perlu dipertimbangkan. Ketergantungan pada kamus yang bersifat statis seringkali menjadi kendala utama, terutama dalam menghadapi dinamika bahasa yang terus berkembang. Sebagaimana penelitian mengenai klasifikasi tingkat stres yang menggunakan NRCLex untuk menentukan emosi dalam teks menunjukkan bahwa kualitas kamus sangat mempengaruhi hasil akurasi, metode *lexicon* mungkin kurang fleksibel dalam menangani kata-kata baru, istilah slang, atau ungkapan ironi dan sarkasme yang maknanya sangat bergantung pada konteks (Fathirachman Mahing et al., 2023). Selain itu, variasi dan

penggunaan bahasa tidak standar dalam platform digital, seperti yang sering ditemui dalam ulasan di *marketplace*, dapat mengurangi akurasi pelabelan jika kata-kata tersebut tidak terdaftar atau memiliki makna yang berbeda dalam *lexicon* yang digunakan (Fadlil et al., 2024).

Dalam penerapan analisis sentimen terhadap teks berbahasa Indonesia, *InSet Lexicon* terbukti efektif untuk mengkategorikan opini menjadi sentimen positif atau negatif. *InSet lexicon* terdiri dari 3.609 kata positif dan 6.609 kata negatif dalam Bahasa Indonesia, yang masing-masing memiliki bobot nilai atau skor polaritas (Muttakin dan Andrika, 2025). Dasar utama dari metode *lexicon* adalah pelabelan otomatis berdasarkan skor akumulatif dari kata-kata yang ditemukan dalam teks.

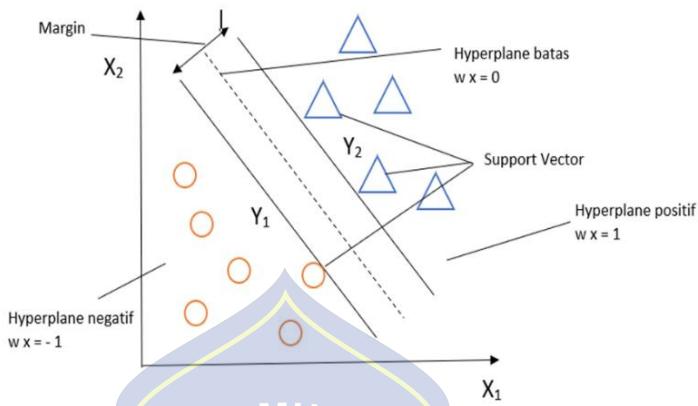
Penerapan pendekatan *Lexicon-Based* terhadap konteks kalimat yang kompleks serta variasi bahasa tidak baku dibutuhkan penyesuaian *lexicon* terhadap konteks atau topik tertentu. Oleh karena itu, pendekatan *lexicon-based* membutuhkan peran metode lain untuk mengatasi kelemahan yang ditemukan. Dalam penelitian ini menggabungkan metode *lexicon-based* dan SVM untuk menghasilkan klasifikasi yang lebih adaptif dan presisi.

## 5. Algoritma *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) merupakan salah satu algoritma *supervised learning* yang paling fundamental dalam bidang *machine learning*, dengan penerapan utama pada tugas klasifikasi. Prinsip dasar dari SVM adalah membangun sebuah *hyperplane* yang optimal untuk memisahkan dua kelas atau lebih dalam *feature space* dengan margin maksimum, yaitu jarak terjauh yang dapat dicapai antara *hyperplane* dan titik data terdekat dari masing-masing kelas yang dikenal sebagai *support vectors* (Gede et al., 2025).

Dengan memaksimalkan margin, SVM bertujuan untuk menghasilkan model yang memiliki kemampuan generalisasi yang tinggi, sehingga dapat memberikan prediksi yang akurat terhadap data baru yang sebelumnya belum

pernah dilibatkan dalam proses pelatihan. Ilustrasi konsep *Hyperplane* pada algoritma SVM dapat dilihat pada Gambar 1.



Gambar 1. *Hyperplane Illustration of SVM* (Taqiuddin et al., 2021)

Salah satu keunggulan utama yang menjadikan SVM banyak digunakan dalam analisis teks adalah kemampuannya dalam menangani data berdimensi tinggi secara efisien. Data berbasis teks, setelah melalui tahap pra-pemrosesan dan ekstraksi fitur seperti metode *Term Frequency-Inverse Document Frequency* (TF-IDF), umumnya direpresentasikan dalam bentuk vektor dengan ribuan dimensi. SVM telah terbukti mampu mengatasi kompleksitas tersebut tanpa mengalami penurunan kinerja yang signifikan. Hokijuliandy et al. (2023) menunjukkan bahwa penerapan SVM dalam klasifikasi ulasan pengguna layanan publik dapat menghasilkan performa klasifikasi yang optimal, terlebih saat dikombinasikan dengan teknik seleksi fitur seperti *Chi-Square*.

Proses kerja SVM dimulai dari tahap pelatihan, di mana model mempelajari parameter optimal dari data berlabel untuk membentuk hyperplane pemisah. Setelah *hyperplane* terbentuk, model dapat mengklasifikasikan data baru dengan menentukan pada sisi mana data tersebut berada terhadap *hyperplane*, sebagaimana diterapkan dalam penelitian klasifikasi kalimat perundungan pada data media sosial

(Nauli et al., 2025). Proses klasifikasi pada algoritma SVM dilakukan dengan menghitung nilai fungsi keputusan  $f(x)$ , dimana  $x$  merepresentasikan data yang akan diklasifikasikan. Perumusan fungsi keputusan dalam algoritma SVM dapat dilihat dalam Persamaan berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \quad (1)$$

Keterangan :

- $f(x)$  : Fungsi klasifikasi untuk data uji  $x$
- $n$  : jumlah data latih
- $\alpha_i$  : Lagrange multipliers untuk data ke- $i$
- $y_i$  : Label kelas dari data latih ke- $i$  (+1 atau -1)
- $K(x_i, x)$  : Fungsi kernel yang mengukur kesamaan antara data latih  $x_i$  dan data uji  $x$
- $b$  : Bias dari hyperplane

Algoritma SVM dapat menggunakan berbagai fungsi kernel untuk memproyeksikan data ke dalam ruang berdimensi lebih tinggi tanpa menghitungnya secara eksplisit, melalui teknik yang dikenal sebagai kernel *trick* (Gede et al., 2025). Fungsi-fungsi kernel yang umum digunakan ditampilkan pada Tabel 1.

Tabel 1. Fungsi Kernel

Nama Kernel	Fungsi Kernel
Linear	$K(x_i, x_j) = (x_i x_j)$ <span style="float: right;">(2)</span>
Polynomial	$K(x_i, x_j) = (x_i x_j)^d$ <span style="float: right;">(3)</span>
Gaussian RBF	$K(x_i, x_j) = \exp\left(-\frac{ x_i - x_j ^2}{2\sigma^2}\right)$ <span style="float: right;">(4)</span>
Sigmoid	$K(x_i, x_j) = \tanh(\sigma(x_i x_j) + c)$ <span style="float: right;">(5)</span>

Dalam penelitian ini, algoritma SVM dipilih karena kemampuannya yang kuat dalam mengklasifikasikan data teks hasil kritik dan saran peserta pelatihan, yang telah melalui proses ekstraksi fitur menggunakan TF-IDF. Penggunaan SVM diharapkan mampu mengidentifikasi polaritas sentimen secara akurat, sekaligus mengatasi tantangan dari data tidak terstruktur dengan dimensi fitur yang tinggi, sebagaimana ditunjukkan dalam penelitian klasifikasi teks menggunakan kombinasi SVM dan TF-IDF (Arifin et al., 2021).

## 6. *Particle Swarm Optimization* (PSO)

*Particle Swarm Optimization* adalah salah satu algoritma optimasi yang didasarkan pada kecerdasan kolektif. Algoritma ini terinspirasi dari perilaku sosial kawanan burung atau kelompok ikan dalam mencari sumber makanan (Darmawan et al., 2022). Dalam PSO, setiap solusi potensial direpresentasikan sebagai sebuah “partikel” yang bergerak dalam ruang pencarian. Setiap partikel memiliki posisi dan kecepatan, serta menyimpan posisi terbaik yang pernah dicapainya dan mengetahui posisi terbaik yang pernah dicapai oleh keseluruhan kawanan.

Prinsip kerja PSO adalah setiap partikel akan terus memperbarui posisi dan kecepatannya berdasarkan pengalaman  $p_{best}$  dan  $g_{best}$  dalam setiap iterasi. Proses ini berlangsung hingga kriteria berhenti terpenuhi, misalnya jumlah iterasi maksimum atau konvergensi solusi. Tujuan utamanya adalah mengarahkan seluruh kawanan partikel menuju titik optimal dalam ruang pencarian (Risawati et al., 2020). Keunggulan PSO terletak pada implementasinya yang relatif mudah, kebutuhan parameter yang sedikit, dan efisiensi komputasi yang baik (Permana & Panudju, 2020).

Dalam konteks penelitian ini, PSO diterapkan untuk mengoptimalkan parameter pada model SVM. Fungsi objektif atau *fitness function* yang digunakan untuk mengevaluasi setiap partikel adalah performa akurasi dari model SVM yang dihasilkan. Dengan kata lain, PSO bertugas mencari kombinasi nilai parameter C dan gamma yang mampu menghasilkan akurasi klasifikasi tertinggi. Pendekatan

ini memungkinkan model SVM untuk mencapai kinerja yang lebih optimal dibandingkan dengan penentuan parameter secara konvensional.

Meskipun demikian, perlu dicatat bahwa seperti algoritma optimasi lainnya, PSO juga memiliki potensi untuk terjebak pada titik optimum lokal, meskipun kemampuannya dalam pencarian global dinilai kuat (Tuerxun et al., 2021). Oleh karena itu, penerapan PSO dalam penelitian ini diharapkan dapat menemukan parameter SVM yang lebih baik untuk meningkatkan akurasi klasifikasi sentimen terhadap umpan balik peserta pelatihan.

## 7. *Hybrid Lexicon* dan SVM

Pendekatan *hybrid* dalam analisis sentimen, khususnya dalam menggabungkan metode *Lexicon-Based* dengan algoritma *machine learning* seperti SVM, menawarkan solusi yang menyeluruh dalam mengatasi keterbatasan masing-masing metode. Dengan mengintegrasikan kedua pendekatan ini, metode *hybrid* berupaya memanfaatkan kekuatan *lexicon* dalam memberikan pembobotan sentimen awal, yang kemudian disempurnakan dengan kemampuan klasifikasi SVM. Penerapan model *hybrid* terbukti efektif, sebagaimana ditunjukkan dalam penelitian terkait analisis opini publik terhadap kebijakan larangan impor pakaian bekas, dimana kombinasi metode mampu menangkap keragaman ekspresi sentimen masyarakat secara lebih menyeluruh (Hendrawati et al., 2024).

Secara teknis, penerapan metode *hybrid Lexicon–SVM* diawali dengan pemberian skor polaritas awal terhadap kata atau kalimat berdasarkan kamus *lexicon*, seperti *InSet* atau VADER. Skor-skor ini kemudian difungsikan sebagai fitur tambahan yang dikombinasikan dengan representasi fitur statistik lainnya seperti yang diperoleh melalui metode TF-IDF dan kemudian dijadikan input untuk algoritma SVM. Dengan demikian, model klasifikasi tidak hanya memanfaatkan frekuensi kata atau bobot term dalam dokumen, tetapi juga diperkaya oleh informasi semantik yang mencerminkan kecenderungan sentimen sejak tahap awal. Alfauzan et al. (2023) menunjukkan bahwa penggunaan skor leksikal sebagai fitur pelengkap dalam klasifikasi sentimen terhadap isu kenaikan harga bahan bakar

berdampak secara signifikan terhadap peningkatan akurasi model, terutama pada konteks ekonomi yang sarat makna implisit. Setelah seluruh fitur dikompilasi, model SVM dilatih menggunakan data yang telah diperluas dengan informasi semantik dari *lexicon*, sehingga menghasilkan klasifikasi sentimen yang lebih presisi.

Keberhasilan implementasi pendekatan dengan menggabungkan metode *Lexicon* dan SVM telah banyak diimplementasikan dalam berbagai penelitian, yang menegaskan potensinya dalam meningkatkan akurasi klasifikasi serta ketahanan model terhadap variasi data. Penelitian oleh Rizky Pratama et al. (2023) menemukan bahwa integrasi fitur berbasis *lexicon* dalam sistem klasifikasi sentimen terhadap ulasan aplikasi perbankan digital mampu meningkatkan presisi prediksi dibandingkan penggunaan SVM atau *lexicon* secara terpisah. Temuan serupa diungkapkan oleh NurmalaSari et al. (2024) dalam penelitian mengenai persepsi pengguna terhadap layanan perpustakaan, yang menyimpulkan bahwa pendekatan *hybrid* mampu mengungkap pola sentimen yang lebih mendalam pada data yang tidak terstruktur. Keunggulan pendekatan ini menjadi sangat relevan dalam pengolahan teks naratif yang kaya makna seperti ulasan produk, komentar media sosial, atau survei terbuka di mana *lexicon* dapat membantu dalam pelabelan awal yang efisien, sementara SVM berperan menyempurnakan klasifikasi berdasarkan pola yang lebih kompleks.

Dalam konteks analisis kritik dan saran peserta pelatihan BBPSDMP Kominfo Makassar, pendekatan *hybrid Lexicon-SVM* menjadi pilihan metodologis yang strategis. Data umpan balik dari peserta umumnya bersifat bebas struktur, memuat ekspresi subjektif, serta beragam dalam pilihan bahasa dan intensitas emosional. Pendekatan metode dengan menggabungkan *lexicon* dan SVM memungkinkan terwujudnya keseimbangan antara efisiensi pemrosesan awal pendekatan *lexicon* dengan kedalaman klasifikasi SVM, sehingga diharapkan dapat menghasilkan hasil analisis yang tidak hanya akurat tetapi juga mampu merepresentasikan persepsi peserta secara utuh dan terukur.

## B. Penelitian Terkait

Sejumlah penelitian terdahulu telah membahas penerapan analisis sentimen terhadap data teks tidak terstruktur dengan memanfaatkan pendekatan *Lexicon-Based* maupun algoritma klasifikasi seperti *Support Vector Machine*. Tinjauan terhadap penelitian-penelitian terkait memberikan landasan konseptual dalam merancang pendekatan metodologis yang lebih adaptif dan relevan terhadap karakteristik data kritik dan saran peserta pelatihan VSGA. Beberapa penelitian terkait yang menjadi acuan dalam penelitian ini ditampilkan pada Tabel 2.

Tabel 2. Penelitian Terkait

Judul Penelitian	Kontribusi Penelitian
1. Combination of <i>Support Vector Machine</i> and <i>Lexicon-Based Algorithm</i> in Twitter Sentiment Analysis (Muhammad et al., 2022).	Menunjukkan efektivitas kombinasi <i>Lexicon-Based</i> dan SVM, dengan SVM. Metode <i>Lexicon-Based</i> efektif melabeli 519 cuitan Twitter secara otomatis. Kernel RBF mencatatkan <i>recall</i> tertinggi (81.73%), kernel <i>Sigmoid</i> menghasilkan presisi tertinggi (78.68%), dan F1-score keduanya 79.60%. Mengidentifikasi tantangan dalam prediksi kelas netral akibat dataset yang tidak seimbang.
2. Analisis Sentimen terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan <i>Lexicon Based Features</i> dan <i>Support Vector Machine</i> (Oktaviana et al., 2022).	Penggunaan <i>Lexicon Based Features</i> berpengaruh signifikan, model mencapai akurasi 0.6, presisi 0.56, <i>recall</i> 0.75, dan F1-measure 0.64. Metode gabungan <i>Lexicon Based Features</i> dan SVM meningkatkan kinerja evaluasi sebesar 12% dibandingkan SVM saja.

Judul Penelitian	Kontribusi Penelitian
3. <i>Sentiment Classification of Fuel Price Rise in Economic Aspects Using Lexicon and SVM Method</i> (Alfauzan et al., 2023).	Model SVM dengan ekstraksi fitur TF-IDF dengan akurasi rata-rata 0.85, F1-score 0.66 lebih unggul secara signifikan dibandingkan model <i>Lexicon Inset</i> saja dengan akurasi rata-rata 0.68, F1-score 0.50. Menunjukkan efektivitas pendekatan berbasis <i>machine learning</i> SVM dengan TF-IDF dalam menangani data teks tidak terstruktur.
4. <i>Discovering User Sentiment Patterns in Libraries with a Hybrid Machine Learning and Lexicon-Based Approach</i> (Nurmalasari et al., 2024).	Algoritma SVM dan <i>Random Forest</i> mencapai akurasi sangat tinggi (99%) serta presisi, <i>recall</i> , dan F1-score 100%. <i>Naïve Bayes</i> menunjukkan akurasi 98% dengan recall 100%, efektif menangkap semua umpan balik negatif. Menggaris bawahi potensi pendekatan <i>hybrid</i> untuk mendukung pengambilan keputusan berbasis data dalam pengelolaan layanan informasi.
5. Perbandingan Pelabelan Rating-based dan <i>Inset Lexicon-Based</i> dalam Analisis Sentimen Menggunakan SVM (Studi Kasus: Ulasan Aplikasi GoBiz di <i>Google Play Store</i> ) (Firda et al., 2025).	Metode pelabelan <i>Inset Lexicon</i> dengan akurasi 89.7% lebih unggul dibandingkan metode <i>Rating-based</i> dengan akurasi 87%. Kombinasi pelabelan berbasis <i>Inset Lexicon</i> dan SVM lebih efektif dalam mengklasifikasikan sentimen ulasan menggunakan bahasa Indonesia.

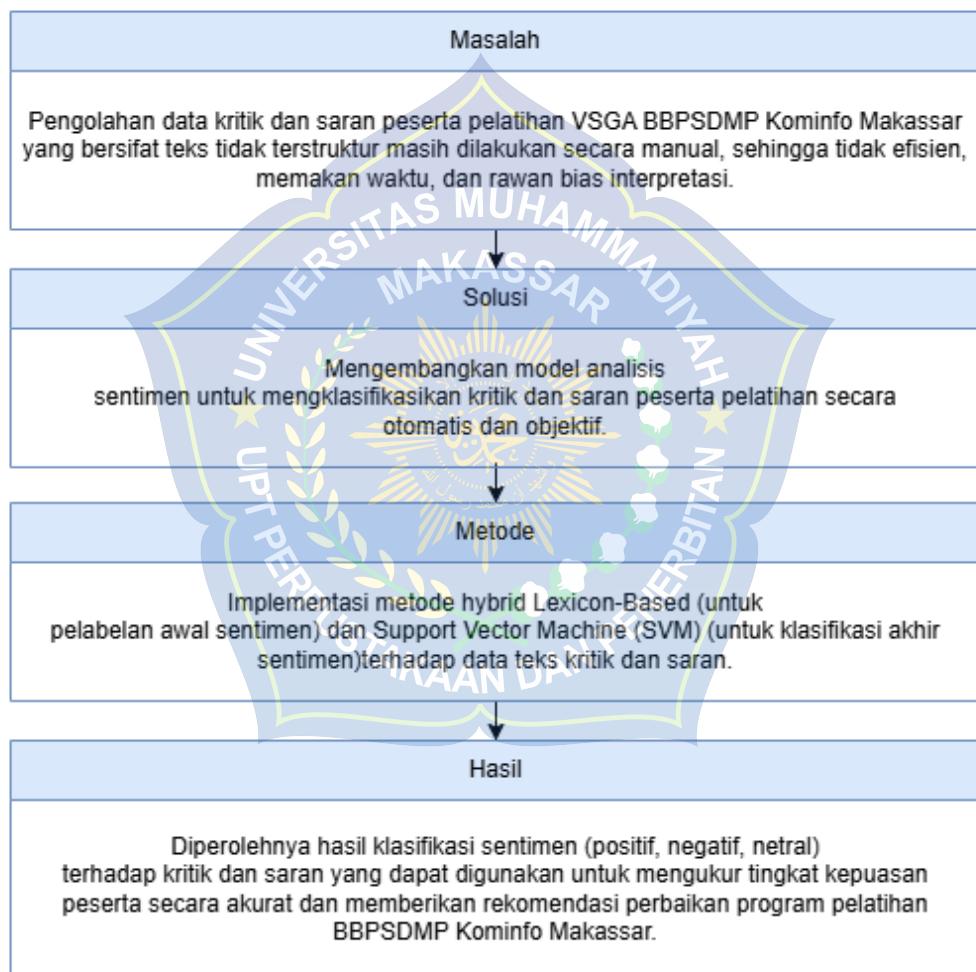
Sejumlah penelitian terdahulu telah menunjukkan keberhasilan penerapan pendekatan hybrid dalam analisis sentimen, khususnya kombinasi antara metode *Lexicon-Based* dan algoritma SVM, pada data teks berbahasa Indonesia. Muhammadi et al. (2022) dan Oktaviana et al. (2022) membuktikan efektivitas pendekatan tersebut dalam menganalisis opini publik dari media sosial, dengan capaian akurasi yang baik dan konsistensi pelabelan yang didukung oleh kamus sentimen. Sementara itu, NurmalaSari et al. (2024) memfokuskan penerapannya pada data survei, yang meskipun lebih terstruktur, tetap mengandung unsur subjektif dan variatif dalam penyampaiannya. Penelitian Firda et al. (2025) secara khusus merekomendasikan penggunaan *InSet Lexicon* sebagai sumber leksikal yang relevan untuk Bahasa Indonesia, terutama ketika digunakan bersama algoritma SVM, menunjukkan performa yang lebih unggul dibanding metode pelabelan lainnya. Adapun penelitian Alfauzan et al. (2023) menegaskan kapabilitas SVM sebagai klasifikator yang adaptif dan efektif, bahkan saat diterapkan pada data yang kompleks dan berdimensi tinggi.

Berdasarkan temuan yang dijelaskan, maka penelitian ini memperluas kontribusi pada bidang analisis sentimen dengan mengaplikasikan pendekatan *Lexicon-Based* dan SVM secara kontekstual pada domain yang lebih spesifik, yaitu kritik dan saran peserta pelatihan *Vocational School Graduate Academy* (VSGA) di bawah naungan BBPSDMP Kominfo Makassar. Berbeda dengan studi sebelumnya yang umumnya berfokus pada data media sosial atau survei publik secara umum, penelitian ini memanfaatkan hasil analisis sentimen untuk mendukung evaluasi program pelatihan secara terukur dan aplikatif. Melalui penggabungan pendekatan *Lexicon* dan algoritma SVM, model yang dikembangkan diharapkan tidak hanya mampu meningkatkan keakuratan analisis terhadap data tidak terstruktur, tetapi juga menghasilkan rekomendasi strategis

berbasis data yang dapat digunakan sebagai dasar peningkatan mutu program pelatihan pemerintah secara berkelanjutan.

### C. Kerangka Pikir

Kerangka pikir disusun untuk menggambarkan alur logis antara perumusan masalah, solusi yang ditawarkan, serta pendekatan metode yang digunakan dalam penelitian ini. Adapun kerangka pikir penelitian ditampilkan pada gambar 2.



Gambar 2. Kerangka Berpikir

## **BAB III**

### **METODE PENELITIAN**

#### **A. Tempat dan Waktu Penelitian**

##### **1. Tempat penelitian**

Penelitian ini dilaksanakan di Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian (BBPSDMP) Kominfo Makassar, Jl. Prof. Abdurrahman Basalamah II No.25, Karampuang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231.

##### **2. Waktu penelitian**

Pelaksanaan kegiatan penelitian direncanakan berlangsung dalam rentang waktu bulan Mei hingga Juli 2025, sebagaimana dijabarkan secara terstruktur pada Tabel 3.

#### **B. Alat dan Bahan**

##### **1. Kebutuhan *Hardware***

- a. Laptop Asus RAM 8
- b. Prosesor Intel Core i5-1035G1

##### **2. Kebutuhan *Software***

- a. Visual Studio Code
- b. Jupyter notebook
- c. Python 3.12.5

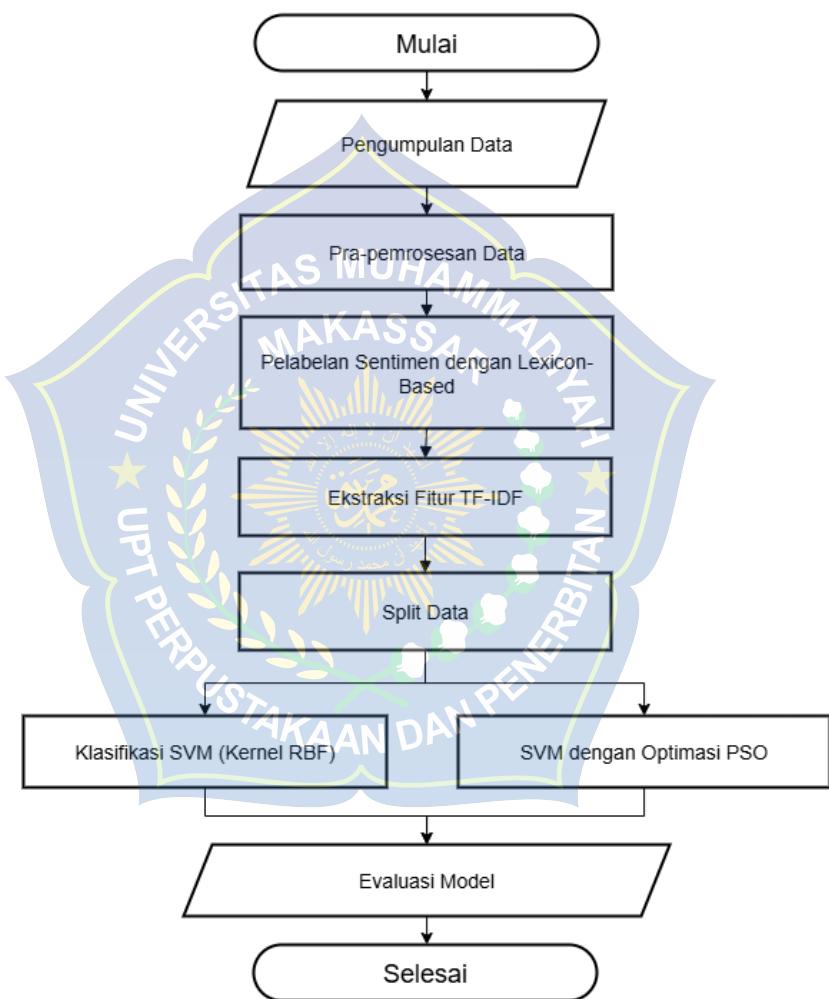
#### **C. Perancangan Sistem**

Perancangan sistem dalam penelitian ini bertujuan untuk membangun model analisis sentimen berbasis metode *hybrid* antara pendekatan *lexicon-based* dan algoritma SVM. Sistem dirancang untuk mengklasifikasikan data kritik dan saran peserta pelatihan BBPSDMP Kominfo Makassar ke dalam kategori sentimen

positif, negatif dan netral. Adapun tahapan perancangan dibagi ke dalam beberapa bagian sebagai berikut.

### 1. Rancangan Penelitian

Rancangan penelitian menjelaskan keseluruhan kerangka kerja penelitian yang dimulai dari pengumpulan data hingga evaluasi model klasifikasi sentimen, rancangan penelitian diilustrasikan dalam Gambar 3.



Gambar 3. Rancangan Penelitian

Rancangan penelitian pada Gambar 3. Rancangan Penelitian menggambarkan tahapan utama, mulai dari pengumpulan data hingga evaluasi

model klasifikasi sentimen menggunakan metode SVM. Penjelasan tiap tahapan dijabarkan sebagai berikut:

a. Mulai

Proses dimulai dari identifikasi kebutuhan dan tujuan penelitian, yaitu menganalisis sentimen peserta pelatihan VSGA berdasarkan data kritik dan saran yang dikumpulkan.

b. Input Data Kritik dan Saran

Data diperoleh dari hasil monitoring dan evaluasi pelatihan VSGA BBPSDMP Kominfo Makassar tahun 2024. Data berupa teks naratif pada empat aspek: fasilitas, konsumsi, pengajar, dan penguji.

c. Pra-Pemrosesan Data

Tahapan ini bertujuan membersihkan dan menyiapkan data teks agar dapat diolah secara komputasional. Proses meliputi:

- 1). *Cleaning*: menghapus karakter non-teks, angka, dan simbol.
- 2). *Case Folding*: mengubah semua huruf menjadi huruf kecil.
- 3). Normalisasi: mengubah kata tidak baku menjadi kata standar.
- 4). *Tokenisasi*: memecah kalimat menjadi satuan kata.
- 5). *Stopword removal*: menghapus kata umum yang tidak bermakna sentimen.
- 6). *Stemming*: mengembalikan kata ke bentuk dasar.

d. Pelabelan Sentimen dengan *Lexicon-Based*

Setelah data dibersihkan, dilakukan pelabelan otomatis menggunakan pendekatan lexicon-based. Setiap kata diberi nilai polaritas, dan dari hasil akumulasi skor, setiap kalimat diklasifikasikan menjadi sentimen positif, negatif, atau netral.

e. Ekstraksi Fitur TF-IDF

Data teks dilanjutkan ke proses vektorisasi menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi.

f. *Split Data*

Dataset dibagi menjadi dua bagian menggunakan metode *hold-out* untuk pengujian model:

- 1). Data latih (80%) digunakan untuk melatih model SVM
- 2). Data uji (20%) digunakan untuk menguji performa model yang telah dilatih

g. Model Klasifikasi SVM (Kernel RBF)

Pada alur pertama, dilakukan pembangunan model klasifikasi menggunakan SVM dengan kernel *Radial Basis Function* (RBF). Kernel RBF dipilih karena kemampuannya dalam menangani data non-linear dan fleksibel dalam membentuk garis pemisah pada ruang fitur berdimensi tinggi.

h. Model Klasifikasi SVM dengan Optimasi PSO

Pada alur kedua, model SVM ditingkatkan dengan penambahan optimasi parameter menggunakan *Particle Swarm Optimization* (PSO). Algoritma PSO digunakan untuk mencari nilai optimal dari parameter C dan gamma yang berperan penting dalam pembentukan margin dan fungsi kernel.

i. Evaluasi Model

Hasil dari kedua model, baik SVM tanpa optimasi maupun SVM dengan optimasi PSO, kemudian dibandingkan dari segi performanya.

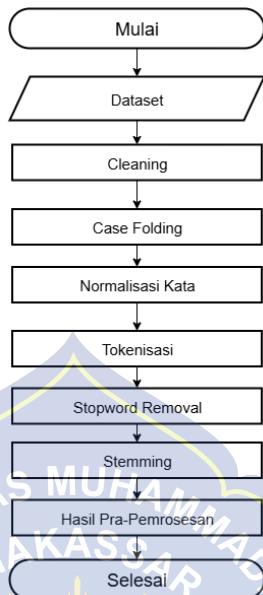
j. Selesai

Proses klasifikasi sentimen selesai

## 2. *Pre-processing Data*

Tahapan pra-pemrosesan data berfokus pada persiapan dan pembersihan teks agar siap dianalisis secara komputasional. Tahapan pra-pemrosesan data

disusun secara terstruktur dalam Gambar 4, yang menunjukkan langkah-langkah pembersihan dan normalisasi data sebagai fondasi analisis sentimen.



Gambar 4. Pra-pemrosesan Data

Tahapan pra-pemrosesan data dalam penelitian bertujuan untuk membersihkan dan menyiapkan teks dari kritik dan saran agar dapat dianalisis secara komputasional. Setiap langkah dalam proses ini dirancang untuk mengurangi noise dan meningkatkan kualitas representasi data sebelum memasuki tahap analisis lebih lanjut.

a). *Mulai*

Proses dimulai dengan menyiapkan dataset.

b). *Dataset*

Proses dimulai dengan data mentah berupa teks kritik dan saran dari peserta pelatihan VSGA dan data masih dalam bentuk tidak terstruktur.

c). *Cleaning*

Langkah pertama adalah *cleaning*, yaitu pembersihan data dari elemen non-teks. Proses ini meliputi penghapusan emoji, angka, simbol, dan

karakter khusus lainnya yang tidak memiliki nilai semantik dalam konteks analisis sentimen. Pembersihan ini penting untuk mengurangi noise dalam data dan meningkatkan kualitas analisis.

d). *Case Folding*

Selanjutnya, dilakukan *Case Folding*, yaitu konversi seluruh huruf menjadi huruf kecil. Tahapan ini bertujuan untuk menjaga konsistensi kata sehingga variasi penulisan huruf kapital tidak menimbulkan pengulangan makna dalam proses selanjutnya.

e). Normalisasi Kata

Langkah berikutnya adalah normalisasi, yaitu mengganti kata-kata tidak baku, slang, atau kata dengan ejaan tidak standar menjadi bentuk baku sesuai dengan kamus kata baku. Proses ini sangat penting dalam konteks Bahasa Indonesia yang memiliki banyak bentuk variasi dalam penulisan informal.

f). *Tokenisasi*

Selanjutnya, teks diproses melalui *tokenisasi*, yaitu pemecahan kalimat menjadi unit-unit kata atau token. Tahapan ini diperlukan agar sistem dapat menganalisis struktur kata secara individual.

g). *Stopword Removal*

Langkah selanjutnya adalah *stopword removal*, yaitu menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap identifikasi sentimen. Contoh stopword antara lain “yang”, “di”, “dan”, serta kata-kata fungsional lainnya. Penghapusan ini membantu dalam memfokuskan analisis pada kata-kata bermakna utama.

h). *Stemming*

Tahapan *Stemming* bertujuan untuk mengembalikan kata ke bentuk dasarnya dengan menghilangkan imbuhan. Misalnya, kata “mengajar”, “pengajar”, dan “diajarkan” akan dikembalikan ke bentuk dasar “ajar”.

Proses ini meningkatkan efisiensi representasi fitur dan mengurangi redundansi.

i). Hasil Pra-Pemrosesan

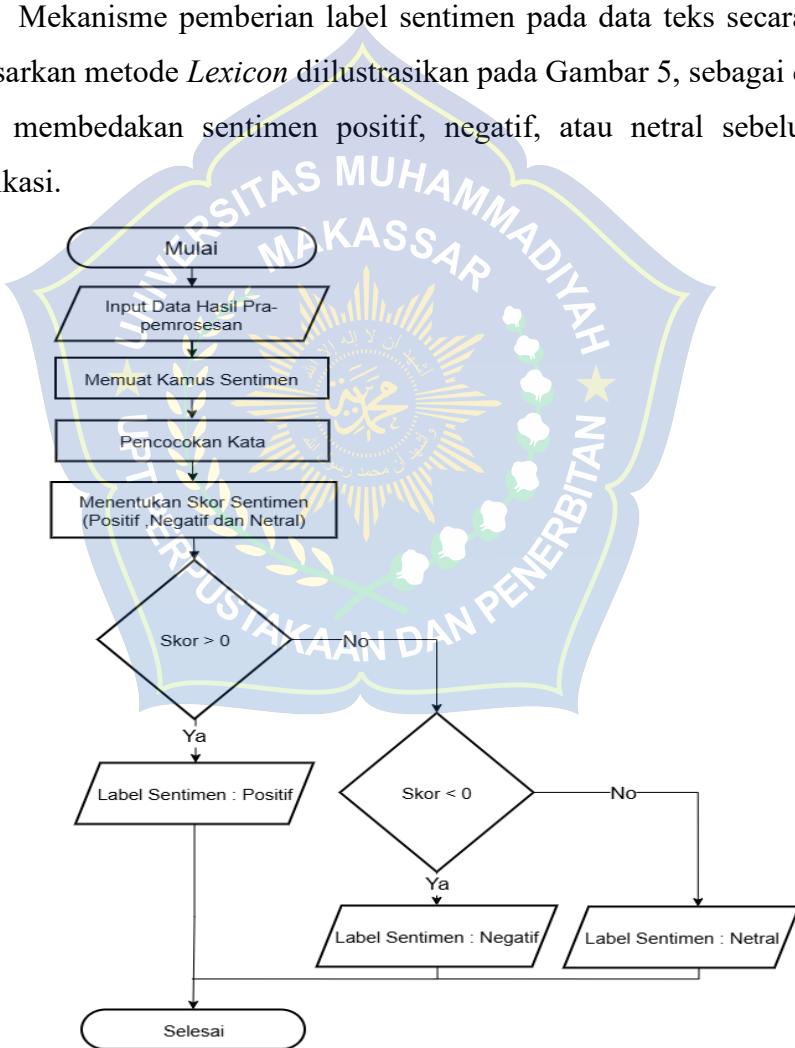
Setelah seluruh tahapan pra-pemrosesan selesai, diperoleh data teks yang bersih, konsisten, dan telah dinormalisasi.

j). Selesai

Data ini siap untuk digunakan dalam proses selanjutnya.

### 3. Pelebelan Sentimen

Mekanisme pemberian label sentimen pada data teks secara otomatis berdasarkan metode *Lexicon* diilustrasikan pada Gambar 5, sebagai dasar awal untuk membedakan sentimen positif, negatif, atau netral sebelum proses klasifikasi.



Gambar 5. Pelabelan Sentimen

Mekanisme pemberian label sentimen secara otomatis ditunjukkan dalam Gambar 5. Pelabelan Sentimen, yang menggambarkan bagaimana sistem menetapkan polaritas sentimen berdasarkan pendekatan *lexicon* sebelum tahap klasifikasi dilakukan.

a). Mulai

Tahapan ini diawali ketika data hasil pra-pemrosesan telah siap untuk dianalisis lebih lanjut.

b). Input data hasil pra-pemrosesan

Proses pelabelan diawali dengan mengambil data teks yang telah melalui tahapan pra-pemrosesan.

c). Memuat kamus sentimen

Sistem memuat kamus sentimen Bahasa Indonesia yang berisi daftar kata berpolaritas positif dan negatif.

d). Pencocokan kata

Setiap kata dalam data teks dibandingkan dengan entri pada kamus sentimen. Proses ini bertujuan untuk mengenali dan menghitung jumlah kata bermuatan positif dan negatif.

e). Perhitungan skor sentimen

Sistem menghitung selisih antara jumlah kata positif dan negatif untuk menghasilkan skor sentimen. Nilai ini digunakan untuk menentukan polaritas teks.

$$\text{Skor Sentimen} = \text{Jumlah Kata Positif} - \text{Jumlah Kata Negatif} \quad (7)$$

f). Penentuan label sentimen

- 1). Jika skor lebih besar dari nol, maka teks diklasifikasikan sebagai positif.
- 2). Jika skor kurang dari nol, maka teks diklasifikasikan sebagai negatif.
- 3). Jika skor sama dengan nol, maka terkласifikasikan sebagai netral.

g). Selesai

Proses pelabelan otomatis selesai dan hasilnya siap untuk digunakan dalam pembangunan model klasifikasi.

#### 4. Pembobotan Fitur TF-IDF

Pada tahap pembobotan fitur, sistem menggunakan pendekatan *Term Frequency–Inverse Document Frequency* untuk mengubah teks hasil praprosesan menjadi representasi numerik. TF-IDF merupakan metode pembobotan statistik yang mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh korpus. Representasi numerik ini diperoleh dengan menerapkan *TfidfVectorizer* pada data hasil *Stemming*, sehingga setiap dokumen dikonversi menjadi vektor yang mencerminkan distribusi kata-kata penting. Vektor TF-IDF tersebut selanjutnya digunakan sebagai input dalam pelatihan dan pengujian model klasifikasi menggunakan algoritma Support Vector Machine (SVM).

Proses perhitungan bobot kata dengan metode TF-IDF, menggunakan tiga rumus utama sebagai berikut:

- a. TF mengukur seberapa sering sebuah kata muncul dalam satu dokumen relatif terhadap total kata dalam dokumen itu sendiri.

$$TF(t, d) = \frac{f(t, d)}{\sum_{k \in d} f(k, d)} \quad (8)$$

Keterangan :

$t$  : kata yang sedang di evaluasi

$d$  : dokumen

$f(t, d)$  : frekuensi kemunculan kata  $t$  dalam dokumen  $d$

$\sum_{k \in d} f(k, d)$  : jumlah total kemunculan semua kata dalam dokumen

- b. IDF mengukur seberapa penting atau jarang kata tersebut di seluruh kumpulan dokumen. Kata yang sering muncul di banyak dokumen akan memiliki nilai IDF yang rendah.

$$IDF(t, D) = \log \left( \frac{N}{DF(t)} \right) \quad (9)$$

Keterangan:

$N$  : jumlah total dukumen dalam korpus

$DF(t)$  : jumlah dokumen yang mengandung kata t

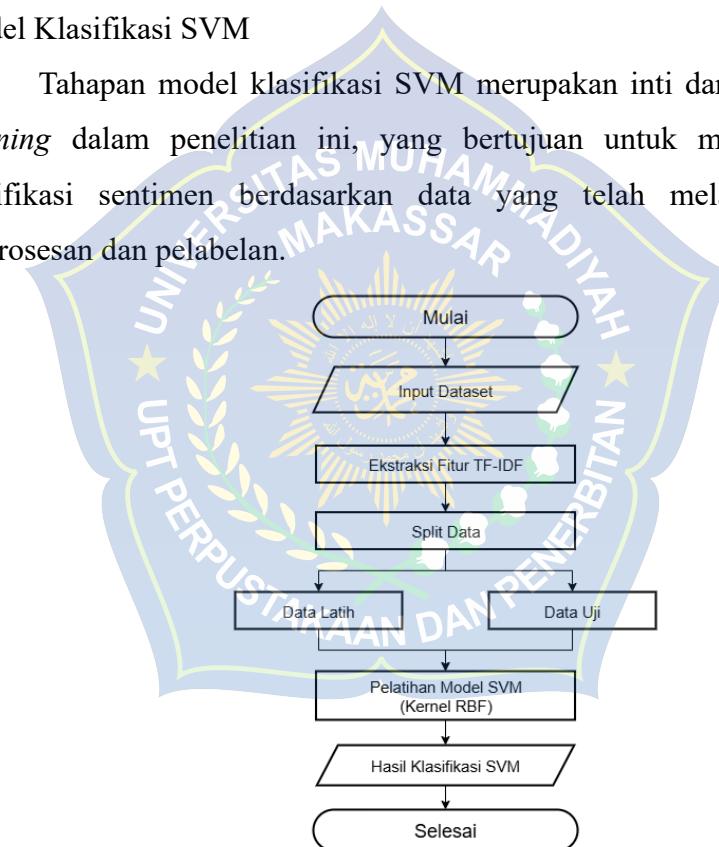
- c. Nilai TF-IDF diperoleh dengan mengalikan TF dan IDF dari suatu kata.

Hasil ini memberikan bobot akhir yang menunjukkan seberapa penting kata tersebut dalam dokumen dan seluruh korpus.

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (10)$$

## 5. Model Klasifikasi SVM

Tahapan model klasifikasi SVM merupakan inti dari proses *machine learning* dalam penelitian ini, yang bertujuan untuk membentuk model klasifikasi sentimen berdasarkan data yang telah melalui proses pra-pemrosesan dan pelabelan.



Gambar 6. Model Klasifikasi SVM

Pada Gambar 6. Klasifikasi SVM memperlihatkan alur proses klasifikasi sentimen dengan menggunakan SVM berbasis kernel RBF sebagai berikut:

a). Mulai

Proses klasifikasi dimulai setelah data siap digunakan, yaitu setelah data teks kritik dan saran peserta pelatihan VSGA selesai diproses dan diberi label sentimen.

b). Input Hasil Pelabelan

Dataset yang digunakan dalam proses ini adalah hasil pelabelan sentimen terhadap data teks. Data dimuat dengan dua atribut utama, yakni data teks hasil *Stemming* dan label sentimen.

c). Ekstraksi Fitur TF-IDF

d). Data teks dilanjutkan ke proses vektorisasi menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi.

e). *Split* Data

Dataset dibagi menjadi dua bagian menggunakan metode *hold-out* untuk pengujian model. Pembagian ini menggunakan rasio 80:20 untuk memastikan bahwa model memperoleh cukup informasi dalam proses pelatihan sekaligus dapat diuji pada data yang belum pernah dilihat sebelumnya.

f). Data Latih dan Data Uji

- 1). Data latih (80%) digunakan untuk melatih model SVM
- 2). Data uji (20%) digunakan untuk menguji performa model yang telah dilatih

g). Pelatihan Model Klasifikasi

Model dilatih menggunakan algoritma SVM dengan kernel RBF. Kernel RBF digunakan untuk menangani data yang tidak dapat dipisahkan secara linear, karena kernel RBF mampu memetakan data ke ruang berdimensi lebih tinggi sehingga margin antar kelas dapat ditentukan dengan optimal.

#### h). Hasil Klasifikasi SVM

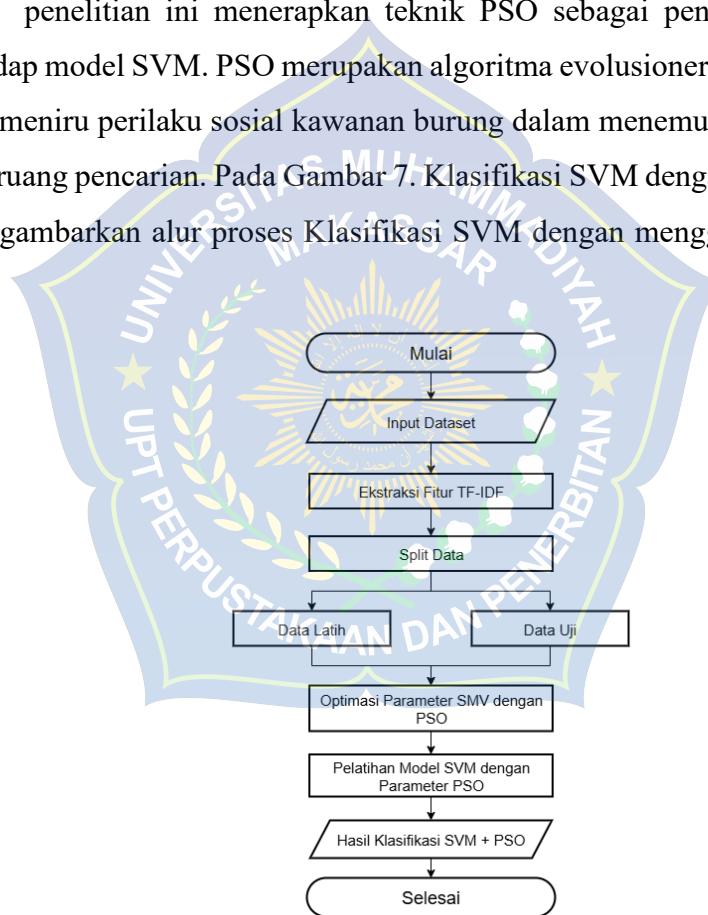
Model yang telah dilatih kemudian diterapkan pada data uji untuk menghasilkan prediksi sentimen. Hasil klasifikasi ini selanjutnya dianalisis untuk menilai tingkat akurasi dan efektivitas model dalam mengidentifikasi sentimen dari data kritik dan saran pelatihan.

#### i). Selesai

Proses Klasifikasi selesai

### 6. Optimasi Model dengan *Particle Swarm Optimization* (PSO)

penelitian ini menerapkan teknik PSO sebagai pendekatan optimasi terhadap model SVM. PSO merupakan algoritma evolusioner berbasis populasi yang meniru perilaku sosial kawanan burung dalam menemukan posisi terbaik pada ruang pencarian. Pada Gambar 7. Klasifikasi SVM dengan Optimasi PSO, manggambarkan alur proses Klasifikasi SVM dengan menggunakan optimasi PSO.



Gambar 7. Klasifikasi SVM dengan Optimasi PSO

Penjelasan tiap tahapan dalam alur klasifikasi SVM dengan menggunakan optimasi PSO sebagai berikut:

- a). Mulai
- b). Proses klasifikasi dimulai setelah data siap digunakan, yaitu setelah data teks kritik dan saran peserta pelatihan VSGA selesai diproses dan diberi label sentimen.
- c). Input Dataset

Dataset yang digunakan dalam proses ini adalah hasil pelabelan sentimen terhadap data teks. Data dimuat dengan dua atribut utama, yakni data teks hasil *Stemming* dan label sentimen.
- d). Ekstraksi Fitur TF-IDF

Data teks dilanjutkan ke proses vektorisasi menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi.
- e). *Split* Data

Dataset dibagi menjadi dua bagian menggunakan metode *hold-out* untuk pengujian model. Pembagian ini menggunakan rasio 80:20 untuk memastikan bahwa model memperoleh cukup informasi dalam proses pelatihan sekaligus dapat diuji pada data yang belum pernah dilihat sebelumnya.
- f). Data Latih dan Data Uji
  - 1). Data latih (80%) digunakan untuk melatih model SVM
  - 2). Data uji (20%) digunakan untuk menguji performa model yang telah dilatih
- g). Optimasi Parameter SVM dengan PSO

Selanjutnya optimasi PSO digunakan untuk menemukan kombinasi parameter terbaik, yaitu C dan gamma, dengan meminimalkan fungsi objektif berupa *negative cross-validation Accuracy*. Setiap partikel PSO mewakili satu solusi kandidat, yang diperbarui secara iteratif mengikuti pengalaman terbaik lokal dan global dalam ruang solusi.
- h). Pelatihan Model SVM dengan Parameter PSO

Setelah parameter optimal diperoleh dari PSO, model SVM kemudian dilatih ulang menggunakan parameter tersebut.

i). Hasil Klasifikasi SVM + PSO

Model yang telah dioptimasi digunakan untuk melakukan prediksi terhadap data uji.

j). Selesai

Proses klasifikasi SVM dengan optimasi PSO selesai

## D. Teknik Pengujian Sistem

Teknik pengujian dalam penelitian ini bertujuan untuk mengevaluasi kinerja sistem klasifikasi sentimen berbasis pendekatan metode *hybrid lexicon-based* dan algoritma SVM. Pengujian dilakukan secara komputasional dan kuantitatif, dengan fokus pada akurasi klasifikasi sentimen terhadap data kritik dan saran peserta pelatihan BBPSDMP Kominfo Makassar.

### 1. Evaluasi Model

Tahap evaluasi model bertujuan untuk mengukur kinerja model klasifikasi SVM dalam mengidentifikasi sentimen berdasarkan data kritik dan saran. Evaluasi dilakukan dengan menghitung empat metrik utama yaitu akurasi, presisi, *recall*, dan F1-score yang diperoleh dari *Confusion Matrix* dan divisualisasikan dalam bentuk *heatmap*.

		Predicted		
		Class	Positif	Netral
Aktual	Positif	TP	FN	
	Netral	FP	TN	

Gambar 8. Confusion Matrix

Hasil dari perhitungan ini dianalisis untuk menyimpulkan seberapa efektif dan andal model yang telah dibangun dalam konteks klasifikasi sentimen. Perhitungan metrik evaluasi dituliskan dalam rumus sebagai berikut:

- a. *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

- b. *Presisi (Precision)* Mengukur ketepatan model dalam mengidentifikasi kelas positif.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (12)$$

- c. *Recall* Mengukur sejauh mana model dapat mendekripsi kelas positif secara benar.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

- d. *F1-score* menggabungkan rata-rata harmonis antara presisi dan recall, memberikan gambaran yang seimbang atas kedua aspek tersebut.

$$\text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (14)$$

Keterangan:

- 1). *TP (True Positive)*: Jumlah kasus aktualnya positif dan diprediksi sebagai positif.
- 2). *TN (True Negative)*: Jumlah kasus aktualnya negatif dan diprediksi sebagai negatif.
- 3). *FP (False Positive)*: Jumlah kasus aktualnya negatif tetapi salah diprediksi sebagai positif
- 4). *FN (False Negative)*: Jumlah kasus aktualnya positif tetapi salah diprediksi sebagai negatif

Keempat metrik tersebut dihitung berdasarkan *Confusion Matrix* dan divisualisasikan untuk interpretasi lebih lanjut. Pemahaman mendalam terhadap masing-masing metrik sangat penting untuk mengidentifikasi kekuatan dan kelemahan spesifik dari model yang dikembangkan. Dengan demikian, hasil evaluasi ini tidak hanya mengukur performa model klasifikasi,

tetapi juga berfungsi sebagai dasar untuk menginterpretasikan tingkat kepuasan peserta pelatihan berdasarkan persebaran opini sentimen yang diklasifikasikan.

## E. Teknik Analisis Data

Teknik analisis data dalam penelitian bertujuan untuk menginterpretasikan hasil klasifikasi sentimen berdasarkan data kritik dan saran peserta pelatihan BBPSDMP Kominfo Makassar. Analisis dilakukan melalui beberapa tahapan yang mencerminkan keseluruhan prosedur penelitian, dimulai dari hasil klasifikasi hingga evaluasi kinerja model. Setiap tahap dirancang untuk mendeskripsikan data secara kuantitatif dan mendalam untuk menjawab rumusan masalah yang telah ditetapkan.

### 1. Analisis Distribusi Sentimen

Langkah awal dalam analisis data adalah memetakan distribusi sentimen yang diperoleh setelah proses pelabelan dan klasifikasi. Sentimen dibagi ke dalam dua atau tiga kategori: positif, negatif, dan netral. Hasil distribusi divisualisasikan dalam bentuk grafik batang untuk memberikan gambaran umum persebaran opini peserta pada aspek fasilitas, makanan, pengajar, dan Penguji. Tujuannya adalah untuk mengidentifikasi kecenderungan umum dan area yang memerlukan perhatian lebih lanjut.

### 2. Evaluasi Kinerja Model

Tahap evaluasi kinerja model mencakup analisis metrik evaluasi dari model SVM, yang meliputi akurasi, presisi, *recall*, dan *F1-score*. Keempat metrik dihitung menggunakan *Confusion Matrix* sebagai dasar perhitungan. Metrik tersebut digunakan untuk menilai efektivitas model dalam mengenali dan mengklasifikasikan sentimen pada data tidak terstruktur secara obyektif.

### 3. Interpretasi Hasil Analisis

Pada tahap ini hasil klasifikasi sentimen dianalisis lebih lanjut untuk memberikan pemahaman kontekstual terhadap kritik dan saran yang masuk. Interpretasi dilakukan terhadap kelompok sentimen dominan pada aspek

pengajar, fasilitas, makanan, dan Pengudi. Pendekatan ini memungkinkan eksplorasi mendalam terhadap tema atau isu yang paling sering diangkat peserta pelatihan.

#### 4. Visualisasi Data

Visualisasi data dilakukan untuk mendukung penafsiran hasil analisis secara intuitif dan komunikatif. Teknik visualisasi yang digunakan meliputi Diagram batang untuk persebaran sentimen per aspek, *word cloud* untuk kata-kata dominan pada tiap kelas sentimen, serta *heatmap Confusion Matrix* untuk mengevaluasi performa klasifikasi. Visualisasi ini tidak hanya memperjelas interpretasi, tetapi juga memudahkan pembaca dalam memahami fokus perbaikan yang diperlukan.



## **BAB VI**

### **HASIL DAN PEMBAHASAN**

#### A. Deskripsi Dataset

Data yang menjadi objek dalam penelitian ini diperoleh dari umpan balik kualitatif berupa kritik dan saran peserta pelatihan *Vocational School Graduate Academy* yang diselenggarakan oleh Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian (BBPSDMP) Kominfo Makassar selama periode pelaksanaan tahun 2024. Pengumpulan data dilakukan melalui formulir digital *Monitoring* dan Evaluasi yang diisi oleh peserta pada akhir sesi pelatihan.

Secara keseluruhan, dataset yang berhasil dikumpulkan berjumlah 2313 ulasan. Data tersebut bersifat tidak terstruktur, mencerminkan kebebasan peserta dalam menyampaikan persepsi dan pengalaman mereka secara naratif. Karakteristik data yang variatif, mengandung istilah informal, singkatan, dan terkadang kesalahan ketik, menjadi tantangan utama yang diatasi melalui tahapan pra-pemrosesan. Fokus analisis dalam penelitian ini diarahkan pada empat aspek utama penyelenggaraan pelatihan, yaitu: Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan, Kritik dan Saran terkait Makanan, Kritik dan Saran terkait Pengajar, dan Kritik dan Saran terkait Assessor (Penguji). Keempat aspek ini dipilih karena merupakan komponen krusial yang secara langsung memengaruhi tingkat kepuasan dan efektivitas pengalaman belajar peserta.

Untuk memberikan gambaran yang lebih jelas mengenai format data mentah yang digunakan, beberapa contoh ulasan disajikan pada Tabel 3. Data Survey Evaluasi berikut.

Tabel 3. Data Survey Evaluasi

No.	Kritik dan Saran terkait Ruangan dan Fasilitas	Kritik dan Saran terkait Makanan	Kritik dan Saran terkait Pengajar	Kritik dan Saran terkait Assessor (Pengaji)
1	ACnya belum di cuci	Thanks	thanks	thanks
2	ruangan dan fasilitas yang digunakan sangat bagus, nyaman, dan bersih.	menu makanan selalu enak setiap hari.	Pengajar sangat baik dalam menjelaskan dan memberikan materi. Sehingga saya mudah paham.	-
....	....	....	....	....
2312	Lebih mempersiapkan instalasi listrik dengan baik pemakaian UPS sangat diperlukan apabila terjadi pemadaman listrik secara tiba-tiba	Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya	Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	Saran saya kepada pengaji(Assesor) agar membalsas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai

No.	Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan	Kritik dan Saran terkait Makanan	Kritik dan Saran terkait Pengajar	Kritik dan Saran terkait Assessor (Penguji)
		mohon untuk sendo sedikit kuat		orang yg lebih tua.
2313	Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	Tidak ada	Tidak ada	Tidak ada

#### B. Pra-Pemrosesan

Setelah dataset berhasil dikumpulkan, tahapan selanjutnya adalah pra-pemrosesan data. Proses ini merupakan langkah fundamental dalam analisis teks yang bertujuan untuk mengubah data mentah yang tidak terstruktur menjadi format yang bersih, konsisten, dan dapat diproses oleh algoritma *machine learning*. Kualitas hasil dari tahapan ini sangat menentukan efektivitas ekstraksi fitur dan akurasi model klasifikasi yang akan dibangun. Setiap langkah dalam pra-pemrosesan dirancang untuk mengurangi noise atau gangguan, menyeragamkan format teks, dan mempertahankan informasi semantik yang relevan.

Berikut adalah uraian dari setiap tahapan pra-pemrosesan yang dilakukan dalam penelitian ini:

### 1. *Cleaning Data*

Tahap awal ini berfokus pada eliminasi karakter-karakter yang tidak memiliki nilai informatif untuk analisis sentimen. Proses pembersihan mencakup penghapusan tanda baca, angka, simbol, tautan URL, serta karakter non-alfabetik lainnya. Tujuannya adalah untuk menyederhanakan teks dan memastikan hanya kata-kata yang berpotensi mengandung sentimen yang akan diproses lebih lanjut. Penerapan tahap *cleaning* pada keempat kategori ulasan dapat diamati pada Tabel 4, Tabel 5, Tabel 6, dan Tabel 7.

Tabel 4. *Cleaning* Data Ulasan Fasilitas

Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan	<i>cleaning</i>
ACnya belum di cuci	ACnya belum di cuci
.....	.....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti

Tabel ini secara spesifik mengilustrasikan proses pembersihan data pada ulasan terkait aspek Fasilitas. Tabel 4 menunjukkan bagaimana tanda baca dan karakter non-alfabetik dihilangkan, menghasilkan teks yang lebih fokus pada konten inti ulasan.

Tabel 5. *Cleaning* Data Ulasan Makanan

Kritik dan Saran terkait Makanan	<i>cleaning</i>
menu makanan selalu enak setiap hari.	menu makanan selalu enak setiap hari

<b>Kritik dan Saran terkait</b>	<i>cleaning</i>
<b>Makanan</b>	
.....	.....
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat

Tabel 5 menampilkan penerapan proses pembersihan pada ulasan terkait aspek Makanan.

Tabel 6. *Cleaning* Data Ulasan Pengajar

<b>Kritik dan Saran terkait Pengajar</b>	<i>cleaning</i>
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	Pengajar sudah sangat kompleks membagikan dan menjelaskan materi
.....	.....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi

Tabel 6 menunjukkan proses pembersihan data pada ulasan mengenai Pengajar. Proses ini memastikan bahwa hanya kata-kata yang relevan yang dipertahankan, sehingga meningkatkan kualitas data untuk analisis sentimen.

Tabel 7. *Cleaning* Data Ulasan Pengaji

<b>Kritik dan Saran terkait Assessor <i>cleaning</i> (Pengaji)</b>	
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta
.....	.....
Saran saya kepada pengaji(Assesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	Saran saya kepada pengaji Assesor agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua

Tabel 7 menunjukkan hasil pembersihan data pada ulasan terkait aspek Pengaji. Dengan menghilangkan karakter non-informatif, teks menjadi lebih terstruktur dan siap untuk diproses pada tahap berikutnya.

## 2. *Case Folding*

Proses ini bertujuan untuk menyeragamkan seluruh huruf dalam teks menjadi format huruf kecil atau *lowercase*. *Case Folding* sangat penting untuk menghindari ambiguitas, di mana kata yang sama namun dengan kapitalisasi berbeda misalnya, “Pengajar”, “pengajar”, dan “PENGAJAR” akan dianggap sebagai token yang identik. Dengan demikian, konsistensi data dapat terjaga dan dimensi fitur tidak membengkak secara tidak perlu. Penerapan tahap *case folding* pada keempat kategori ulasan dapat diamati pada Tabel 8, Tabel 9, Tabel 10, dan Tabel 11.

Tabel 8. *Case Folding* Ulasan Fasilitas

<b>Kritik dan Saran</b>	<b><i>cleaning</i></b>	<b><i>case_folding</i></b>
<b>terkait Ruangan dan Fasilitas Pelatihan</b>		
ACnya belum di cuci	ACnya belum di cuci	acnya belum di cuci
.....	.....	....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti

Tabel 8 menampilkan hasil konversi seluruh teks ulasan Fasilitas menjadi huruf kecil. Proses ini memastikan bahwa variasi kapitalisasi tidak lagi menjadi faktor pembeda, sehingga kata seperti “Ruangan” dan “ruangan” dianggap identik.

Tabel 9. *Case Folding* Ulasan Makanan

<b>Kritik dan Saran</b>	<b><i>cleaning</i></b>	<b><i>case_folding</i></b>
<b>terkait Makanan</b>		
menu makanan selalu enak setiap hari.	menu makanan selalu enak setiap hari	menu makanan selalu enak setiap hari
.....	.....	.....

<b>Kritik dan Saran</b>	<b><i>cleaning</i></b>	<b><i>case_folding</i></b>
<b>terkait Makanan</b>		
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat

Tabel 9 menampilkan hasil penerapan *case folding* pada ulasan Makanan. Penyeragaman huruf ini merupakan langkah penting untuk menjaga konsistensi data sebelum diolah lebih lanjut.

Tabel 10. *Case Folding* Ulasan Pengajar

<b>Kritik dan Saran</b>	<b><i>cleaning</i></b>	<b><i>case_folding</i></b>
<b>terkait Pengajar</b>		
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	pengajar sudah sangat kompleks membagikan dan menjelaskan materi
....	....	....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi	semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi

Tabel 10 menunjukkan hasil konversi teks pada ulasan Pengajar menjadi format *lowercase*. Proses ini menyederhanakan data dan mempersiapkannya untuk tahapan analisis selanjutnya.

Tabel 11. *Case Folding* Ulasan Pengaji

<b>Kritik dan Saran terkait Assessor (Pengaji)</b>	<b>cleaning</b>	<b>case_folding</b>
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta
....	.....	.....
Saran saya kepada pengiji(Assesor) agar membala senyum dan membala sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	Saran saya kepada pengijiAssesor agar membala senyum dan membala sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua	saran saya kepada pengijiassesor agar membala senyum dan membala sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua

Tabel 1 menunjukkan hasil konversi teks pada ulasan Pengaji menjadi format *lowercase*. Proses ini menyederhanakan data dan mempersiapkannya untuk tahapan analisis selanjutnya.

### 3. Normalisasi

Bahasa Indonesia dalam konteks ulasan informal sering kali mengandung kata-kata tidak baku, singkatan, atau istilah slang. Tahap normalisasi berfungsi untuk mengubah variasi kata tersebut ke dalam bentuk standar sesuai kaidah bahasa Indonesia yang baku. Proses ini dilakukan dengan memanfaatkan kamus normalisasi yang telah disusun sebelumnya, misalnya mengubah “yg” menjadi “yang”, “tdk” menjadi “tidak”, atau “keren bgt” menjadi “keren sekali”. Langkah ini krusial untuk memastikan makna kata dapat dikenali secara akurat oleh kamus sentimen dan model. Hasil proses normalisasi pada keempat kategori ulasan dapat diamati pada Tabel 12, Tabel 13, Tabel 14, dan Tabel 15.

Tabel 12. Normalisasi Data Ulasan Fasilitas

Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan	case_folding	hasil_normalisasi
ACnya belum di cuci	acnya belum di cuci	acnya belum di cuci
....	.....	.....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yang diberikan juga dapat di mengerti	semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yang diberikan juga dapat di mengerti

Tabel 12 menunjukkan penerapan tahap normalisasi pada ulasan Fasilitas, di mana kata-kata tidak baku atau singkatan diubah ke bentuk standarnya.

Tabel 13. Normalisasi Data Ulasan Makanan

<b>Kritik dan Saran terkait Makanan</b>	<b>case_folding</b>	<b>hasil_normalisasi</b>
menu makanan selalu enak setiap hari.	menu makanan selalu enak setiap hari	menu makanan selalu enak setiap hari
.....	.....	.....
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat

Tabel 13 menunjukkan penerapan tahap normalisasi pada ulasan Makanan, di mana kata-kata tidak baku atau singkatan diubah ke bentuk standarnya.

Tabel 14. Normalisasi Data Ulasan Pengajar

<b>Kritik dan Saran terkait Pengajar</b>	<i>case_folding</i>	<i>hasil_normalisasi</i>
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	pengajar sudah sangat kompleks membagikan dan menjelaskan materi	pengajar sudah sangat kompleks membagikan dan menjelaskan materi
....	.....	.....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi	semoga para pengajar semakin sukses dan dapat memberikan materi dengan lebih baik lagi

Tabel 14 menunjukkan penerapan tahap normalisasi pada ulasan Pengajar, di mana kata-kata tidak baku atau singkatan diubah ke bentuk standarnya.

Tabel 15. Normalisasi Data Ulasan Penguji

<b>Kritik dan Saran terkait Assessor (Penguji)</b>	<i>case_folding</i>	<i>hasil_normalisasi</i>
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta
....	.....	.....

<b>Kritik dan Saran</b> <b>terkait Assessor</b> <b>(Penguji)</b>	<b>case_folding</b>	<b>hasil_normalisasi</b>
Saran saya kepada pengiji(Assesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	saran saya kepada pengijiassesor agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua	saran saya kepada pengijiassesor agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yang lebih tua

Tabel 15 menunjukkan penerapan tahap normalisasi pada ulasan Penguji, di mana kata-kata tidak baku atau singkatan diubah ke bentuk standarnya.

#### 4. Tokenisasi

Setelah teks bersih dan ternormalisasi, proses tokenisasi dilakukan untuk memecah setiap kalimat atau ulasan menjadi unit-unit kata individual yang disebut token. Sebagai contoh, kalimat “materi mudah dipahami” akan dipecah menjadi tiga token: [“materi”, “mudah”, “dipahami”]. Tahapan ini mengubah aliran teks kontinu menjadi segmen-segmen diskrit yang menjadi dasar untuk analisis lebih lanjut. Hasil dari proses *tokenisasi* pada setiap kategori ulasan dapat dilihat pada Tabel 16, Tabel 17, Tabel 18, dan Tabel 19.

Tabel 16. Tokenisasi Data Ulasan Fasilitas

<b>Kritik dan Saran</b>	<b>hasil_normalisasi</b>	<b>tokenize</b>
<b>terkait Ruangan dan Fasilitas Pelatihan</b>		
ACnya belum di cuci	acnya belum di cuci	['acnya', 'belum', 'di', 'cuci']
....	....	....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yang diberikan juga dapat di mengerti	['semua', 'fasilitas', 'sangat', 'memuaskan', 'dan', 'saya', 'merasa', 'nyaman', 'saat', 'pelatihan', 'dan', 'menteri', 'yang', 'diberikan', 'juga', 'dapat', 'di', 'mengerti']

Tabel 16 menunjukkan bagaimana kalimat-kalimat dari ulasan Fasilitas dipecah menjadi unit-unit kata individual atau token. Representasi dalam bentuk daftar ini merupakan format dasar yang diperlukan untuk analisis pada tahap selanjutnya.

Tabel 17. Tokenisasi Data Ulasan Makanan

<b>Kritik dan Saran</b>	<b>hasil_normalisasi</b>	<b>tokenize</b>
<b>terkait Makanan</b>		
menu makanan selalu enak setiap hari.	menu makanan selalu enak setiap hari	['menu', 'makanan', 'selalu', 'enak', 'setiap', 'hari']
....	....	....

<b>Kritik dan Saran</b>	<b>hasil_normalisasi</b>	<b>tokenize</b>
<b>terkait Makanan</b>		
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	['manurut', 'saya', 'rasa', 'dan', 'lauknya', 'beefariasi', 'tetapi', 'untuk', 'sendoknya', 'agak', 'susah', 'untuk', 'membelah', 'lauknya', 'mohon', 'untuk', 'sendo', 'sedikit', 'kuat']

Tabel 17 menunjukkan bagaimana kalimat-kalimat dari ulasan Makanan dipecah menjadi unit-unit kata individual atau token.

Tabel 18. Tokenisasi Data Ulasan Pengajar

<b>Kritik dan Saran</b>	<b>hasil_normalisasi</b>	<b>tokenize</b>
<b>terkait Pengajar</b>		
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	pengajar sudah sangat kompleks membagikan dan menjelaskan materi	['pengajar', 'sudah', 'sangat', 'kompleks', 'membagikan', 'dan', 'menjelaskan', 'materi']
.....	.....	.....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	semoga para pengajar semakin sukses dan dapat memberikan materi dengan lebih baik lagi	['semoga', 'para', 'pengajar', 'semakin', 'sukses', 'dan', 'dapat', 'memberikan', 'materi', 'dengan', 'lebih', 'baik', 'lagi']

Tabel 18 menunjukkan bagaimana kalimat-kalimat dari ulasan Pengajar dipecah menjadi unit-unit kata individual atau token.

Tabel 19. Tokenisasi Data Ulasan Pengaji

Kritik dan Saran terkait Assessor (Pengaji)	hasil_normalisasi	<i>tokenize</i>
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	['assessor', 'juga', 'sudah', 'sangat', 'kompleks', 'dan', 'totalitas', 'dalam', 'menguji', 'peserta']
.....	.....	.....
Saran saya kepada pengiji(Asesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	saran saya kepada pengijiassesor agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yang lebih tua	['saran', 'saya', 'kepada', 'pengijiassesor', 'agar', 'membalas', 'senyum', 'dan', 'membalas', 'sapaan', 'itu', 'bentuk', 'rasa', 'hormat', 'dan', 'rasa', 'menghargai', 'orang', 'yang', 'lebih', 'tua']

Tabel 19 menunjukkan bagaimana kalimat-kalimat dari ulasan Pengaji dipecah menjadi unit-unit kata individual atau token.

##### 5. Stopword Removal

Tidak semua kata dalam sebuah kalimat memberikan kontribusi signifikan terhadap makna sentimen. Stopword adalah kata-kata umum yang

sering muncul namun memiliki bobot informasi rendah, seperti kata hubung (“dan”, “atau”, “tetapi”), kata depan (“di”, “ke”, “dari”), dan kata ganti (“yang”, “ini”). Proses *Stopword Removal* bertujuan untuk menghapus kata-kata tersebut dari daftar token. Dengan demikian, analisis dapat lebih terfokus pada kata-kata kunci yang secara langsung mencerminkan opini positif atau negatif. Penerapan tahap *stopword removal* setelah *tokenisasi* dari setiap kategori ulasan disajikan pada Tabel 20, Tabel 21, Tabel 22, dan Tabel 23.

Tabel 20. *Stopword Removal* Data Ulasan Fasilitas

Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan	<i>tokenize</i>	<i>stopword_removal</i>
ACnya belum di cuci	['acnya', 'belum', 'di', 'cuci']	['acnya', 'cuci']
.....	.....	.....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	['semua', 'fasilitas', 'sangat', 'memuaskan', 'dan', 'saya', 'merasa', 'nyaman', 'saat', 'pelatihan', 'dan', 'menteri', 'yang', 'diberikan', 'juga', 'dapat', 'di', 'mengerti']	['fasilitas', 'memuaskan', 'nyaman', 'pelatihan', 'menteri', 'mengerti']

Tabel 20 menunjukkan hasil penghapusan kata-kata umum dari daftar token ulasan Fasilitas. Proses ini secara efektif mengurangi noise dengan

menyisakan kata-kata yang lebih padat makna dan relevan untuk analisis sentimen.

Tabel 21. *Stopword Removal* Data Ulasan Makanan

<b>Kritik dan Saran</b>	<b>tokenize</b>	<b>stopword_removal</b>
<b>terkait Makanan</b>		
menu makanan selalu enak setiap hari.	['menu', 'makanan', 'selalu', 'enak', 'setiap', 'hari']	['menu', 'makanan', 'enak']
.....	.....	.....
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	['manurut', 'saya', 'rasa', 'dan', 'lauknya', 'beefariasi', 'tetapi', 'untuk', 'sendoknya', 'agak', 'susah', 'untuk', 'membelah', 'lauknya', 'mohon', 'untuk', 'sendo', 'sedikit', 'kuat']	['manurut', 'lauknya', 'beefariasi', 'sendoknya', 'susah', 'membelah', 'lauknya', 'mohon', 'sendo', 'kuat']

Tabel 21 menunjukkan hasil penghapusan kata-kata umum dari daftar token ulasan Makanan.

Tabel 22. *Stopword Removal* Data Ulasan Pengajar

<b>Kritik dan Saran</b>	<i>tokenize</i>	<i>stopword_removal</i>
<b>terkait Pengajar</b>		
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	['pengajar', 'sudah', 'sangat', 'kompleks', 'membagikan', 'dan', 'menjelaskan', 'materi']	['pengajar', 'kompleks', 'membagikan', 'materi']
.....	.....	....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	['semoga', 'para', 'pengajar', 'semakin', 'sukses', 'dan', 'dapat', 'memberikan', 'materi', 'dengan', 'lebih', 'baik', 'lagi']	['semoga', 'pengajar', 'sukses', 'materi']

Tabel 22 menunjukkan hasil penghapusan kata-kata umum dari daftar token ulasan Pengajar.

Tabel 23. *Stopword Removal* Data Ulasan Penguji

<b>Kritik dan Saran</b>	<i>tokenize</i>	<i>stopword_removal</i>
<b>terkait Assessor (Penguji)</b>		
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	['assessor', 'juga', 'sudah', 'sangat', 'kompleks', 'totalitas', 'dan', 'menguji', 'peserta']	['assessor', 'kompleks', 'totalitas', 'menguji', 'peserta']

<b>Kritik dan Saran terkait Assessor (Penguji)</b>	<i>tokenize</i>	<i>stopword removal</i>
	'totalitas', 'dalam', 'menguji', 'peserta']	
....	....	....
Saran saya kepada pengiji(Assesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	['saran', 'saya', 'kepada', 'pengijiassesor', 'agar', 'membalas', 'senyum', 'dan', 'menghargai', 'orang', 'yang', 'lebih', 'tua']	['saran', 'pengijiassesor', 'membalas', 'senyum', 'membalas', 'sapaan', 'bentuk', 'hormat', 'menghargai', 'orang', 'tua']

Tabel 23 menunjukkan hasil penghapusan kata-kata umum dari daftar token ulasan Penguji.

#### 6. Stemming

Tahap akhir dari pra-pemrosesan adalah *Stemming*, yaitu proses mereduksi kata-kata berimbuhan ke bentuk kata dasarnya. Dalam konteks Bahasa Indonesia, proses ini menghilangkan awalan, akhiran, dan sisipan. Sebagai contoh, kata “penyampaian”, “menyampaikan”, dan “disampaikan” akan diubah menjadi kata dasar “sampai”. Untuk memastikan akurasi, penelitian ini menggunakan stemmer yang dirancang khusus untuk Bahasa Indonesia seperti pustaka Sastrawi. *Stemming* bertujuan untuk menyatukan berbagai variasi morfologis dari sebuah kata ke dalam satu representasi makna

tunggal, sehingga dapat mengurangi redundansi fitur secara efektif. Hasil akhir dari proses *stemming* untuk keempat kategori ulasan dapat dilihat pada Tabel 24, Tabel 25, Tabel 26, dan Tabel 27.

Tabel 24. *Stemming* Data Ulasan Fasilitas

<b>Kritik dan Saran</b>	<i>stopword_removal</i>	<i>stemming_data</i>
<b>terkait Ruangan dan Fasilitas Pelatihan</b>		
ACnya belum di cuci	['acnya', 'cuci']	acnya cuci
....	.....	.....
Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	['fasilitas', 'memuaskan', 'nyaman', 'pelatihan', 'menteri', 'mengerti']	fasilitas muas nyaman latih menteri erti

Tabel 24 menampilkan hasil akhir dari tahap pra-pemrosesan untuk ulasan Fasilitas, di mana setiap kata direduksi ke bentuk dasarnya.

Tabel 25. *Stemming* Data Ulasan Makanan

<b>Kritik dan Saran</b>	<i>stopword_removal</i>	<i>stemming_data</i>
<b>terkait Makanan</b>		
menu makanan selalu enak setiap hari.	['menu', 'makanan', 'enak']	menu makan enak
....	....	....

<b>Kritik dan Saran</b>	<i>stopword_removal</i>	<i>stemming_data</i>
<b>terkait Makanan</b>		
Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	['manurut', 'lauknya', 'beefariasi', 'sendoknya', 'susah', 'membelah', 'lauknya', 'mohon', 'sendo', 'kuat']	manurut lauk beefariasi sendok susah belah lauk mohon sendo kuat

Tabel 25 menampilkan hasil akhir dari tahap pra-pemrosesan untuk ulasan Makanan, di mana setiap kata direduksi ke bentuk dasarnya.

Tabel 26. Stemming Data Ulasan Pengajar

<b>Kritik dan Saran</b>	<i>stopword_removal</i>	<i>stemming_data</i>
<b>terkait Pengajar</b>		
Pengajar sudah sangat kompleks membagikan dan menjelaskan materi	['pengajar', 'kompleks', 'membagikan', 'materi']	ajar kompleks bagi materi
....	....	....
Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	['semoga', 'pengajar', 'sukses', 'materi']	moga ajar sukses materi

Tabel 26 menampilkan hasil akhir dari tahap pra-pemrosesan untuk ulasan Pengajar, di mana setiap kata direduksi ke bentuk dasarnya.

Tabel 27. *Stemming* Data Ulasan Penguji

Kritik dan Saran terkait Assessor (Penguji)	<i>stopword_removal</i>	<i>stemming_data</i>
Assessor juga sudah sangat kompleks dan totalitas dalam menguji peserta	['assessor', 'kompleks', 'totalitas', 'menguji', 'peserta']	assessor kompleks totalitas uji serta
..... Saran saya kepada pengiji(Assesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.	['saran', 'pengijiassesor', 'membalas', 'senyum', 'membalas', 'sapaan', 'bentuk', 'hormat', 'menghargai', 'orang', 'tua']	saran pengijiassesor balas senyum balas sapa bentuk hormat harga orang tua

Tabel 27 menampilkan hasil akhir dari tahap pra-pemrosesan untuk ulasan Penguji, di mana setiap kata direduksi ke bentuk dasarnya.

### C. Pelabelan Sentimen

Setelah data teks melalui seluruh tahapan pra-pemrosesan, langkah selanjutnya adalah memberikan label sentimen awal pada setiap ulasan. Tujuannya adalah untuk mengklasifikasikan setiap ulasan secara otomatis ke dalam tiga kategori: positif, negatif, atau netral. Penelitian ini memanfaatkan *InSet Lexicon*, sebuah kamus sentimen yang telah dikembangkan secara khusus untuk Bahasa Indonesia. Kamus ini berisi daftar kosakata yang telah diberi bobot polaritas positif dan negatif. Mekanisme pelabelan bekerja secara sistematis sebagai berikut:

## 1. Input Data Pra-pemrosesan

Data teks yang telah melalui proses *cleaning*, *case folding*, normalisasi, *tokenisasi*, *stopword removal*, dan *stemming* dimasukkan sebagai input utama.

```
dataFasilitas =  
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_  
Fasilitas.csv")  
dataMakanan =  
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_  
Makanan.csv")  
dataPengajar =  
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_  
Pengajar.csv")  
dataPenguji =  
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_  
Penguji.csv")  
dataFasilitas =  
pd.DataFrame(dataFasilitas[['steming_data']])  
dataMakanan = pd.DataFrame(dataMakanan[['steming_data']])  
dataPengajar =  
pd.DataFrame(dataPengajar[['steming_data']])  
dataPenguji = pd.DataFrame(dataPenguji[['steming_data']])
```

## 2. Memuat Kamus Sentimen

Sistem mengimpor dua daftar kosakata dari InSet Lexicon, yaitu daftar kata positif dan negatif yang masing-masing dimuat sebagai set koleksi kata.

```
url_negatif =  
'https://raw.githubusercontent.com/fajri91/InSet/master/n  
egative.tsv'  
url_positif =  
'https://raw.githubusercontent.com/fajri91/InSet/master/p  
ositive.tsv'  
positif_lexicon = set(pd.read_csv(url_positif, sep='\t',  
header=None) [0])  
negatif_lexicon = set(pd.read_csv(url_negatif, sep='\t',  
header=None) [0])
```

### 3. Pencocokan Kata dan Penghitungan Skor

Setiap kata dalam ulasan dibandingkan dengan entri yang terdapat dalam kamus sentimen. Apabila kata ditemukan di kamus positif, maka skor positif bertambah dan jika ditemukan di kamus negatif, skor negatif bertambah.

```
def hitung_sentimen(teks):
    if isinstance(teks, str):
        teks = teks.lower()
        positif_count = sum(1 for word in teks.split() if
word in positif_lexicon)
        negatif_count = sum(1 for word in teks.split() if
word in negatif_lexicon)
```

### 4. Perhitungan Skor Total

Skor sentimen dihitung dengan mengurangkan jumlah kata negatif dari jumlah kata positif:

```
sentimen_skor = positif_count - negatif_count
```

### 5. Penentuan Label Sentimen

Berdasarkan nilai skor total, sistem secara otomatis menetapkan label sentimen sebagai berikut:

- Positif: Jika skor total > 0.
- Negatif: Jika skor total < 0.
- Netral: Jika skor total = 0

```
if sentimen_skor > 0:
    sentimen = "Positif"
elif sentimen_skor < 0:
    sentimen = "Negatif"
else:
    sentimen = "Netral"
return sentimen_skor, sentimen
return 0, "Netral"
```

Mekanisme pelabelan otomatis ini diterapkan secara sistematis pada keseluruhan data ulasan yang telah melalui tahap pra-pemrosesan. Dengan mengalkulasi skor sentimen berdasarkan kamus *InSet Lexicon*, setiap ulasan

kemudian diklasifikasikan ke dalam kategori positif, negatif, atau netral. Hasil dari pelabelan awal ini memberikan gambaran umum mengenai persepsi peserta pelatihan. Distribusi sentimen yang diperoleh disajikan pada Tabel 28, Tabel 29, Tabel 30, dan Tabel 31.

Tabel 28. Label Sentimen Ulasan Fasilitas

No.	steming_data	Score	Sentimen
1.	meja sesuai kebutuhan desain ruang mouse gerak	1	Positif
2.	latih jgd cocok kursi kuliah mouse	0	Netral
3.	ac ruang ith kelas okm a matimati bikin ruang panas ganggu kurang nyaman	-4	Negatif
4.	saran ac ya dingin	1	Positif
....	.....	....	....
1461.	instalasi listrik pakai ups diperlukan madam listrik tibatiba	0	Netral
1462.	fasilitas muas nyaman latih menteri erti	2	Positif

Tabel 28 menunjukkan hasil pelabelan sentimen untuk ulasan Fasilitas. Tabel ini memperlihatkan bagaimana skor sentimen dihitung berdasarkan kata-kata dalam teks, yang kemudian menentukan klasifikasi akhir sebagai positif, negatif, atau netral.

Tabel 29. Label Sentimen Ulasan Makanan

No.	steming_data	Score	Sentimen
1.	makan enak	0	Netral
2.	makan ayam asin bu	0	Netral
3.	makan enak	0	Netral
4.	makan enak	0	Netral

No.	<i>stemming_data</i>	Score	Sentimen
....	.....	....	....
1249.	muas	1	Positif
1250.	manurut lauk beefariasi sendok susah belah lauk mohon sendo kuat	-2	Negatif

Tabel 29 menunjukkan hasil pelabelan sentimen untuk ulasan Makanan. Tabel ini memperlihatkan bagaimana skor sentimen dihitung berdasarkan kata-kata dalam teks, yang kemudian menentukan klasifikasi akhir sebagai positif, negatif, atau netral.

Tabel 30. Label Sentimen Ulasan Pengajar

No.	<i>stemming_data</i>	Score	Sentimen
1.	saran materi langsung praktik layar monitor	1	Positif
2.	materi rinci	0	Netral
3.	puas latih uji kompetensi	3	Positif
4.	materi humble mudah paham mogah latih materi	1	Positif
....	.....	....	....
1270.	ajar	0	Netral
1271.	moga ajar sukses materi	2	Positif

Tabel 30 menunjukkan hasil pelabelan sentimen untuk ulasan Pengajar. Tabel ini memperlihatkan bagaimana skor sentimen dihitung berdasarkan kata-kata dalam teks, yang kemudian menentukan klasifikasi akhir sebagai positif, negatif, atau netral.

Tabel 31. Label Sentimen Ulasan Penguji

No.	steming_data	Score	Sentimen
1.	asesor	0	Netral
2.	uji cukup efisien	1	Positif
3.	mantab	0	Netral
4.	assesor ya	1	Positif
....	.....	....	....
993.	asesor support saran uji kompetensi	3	Positif
994.	saran pengijiassesor balas senyum balas sapa bentuk hormat harga orang tua	1	Positif

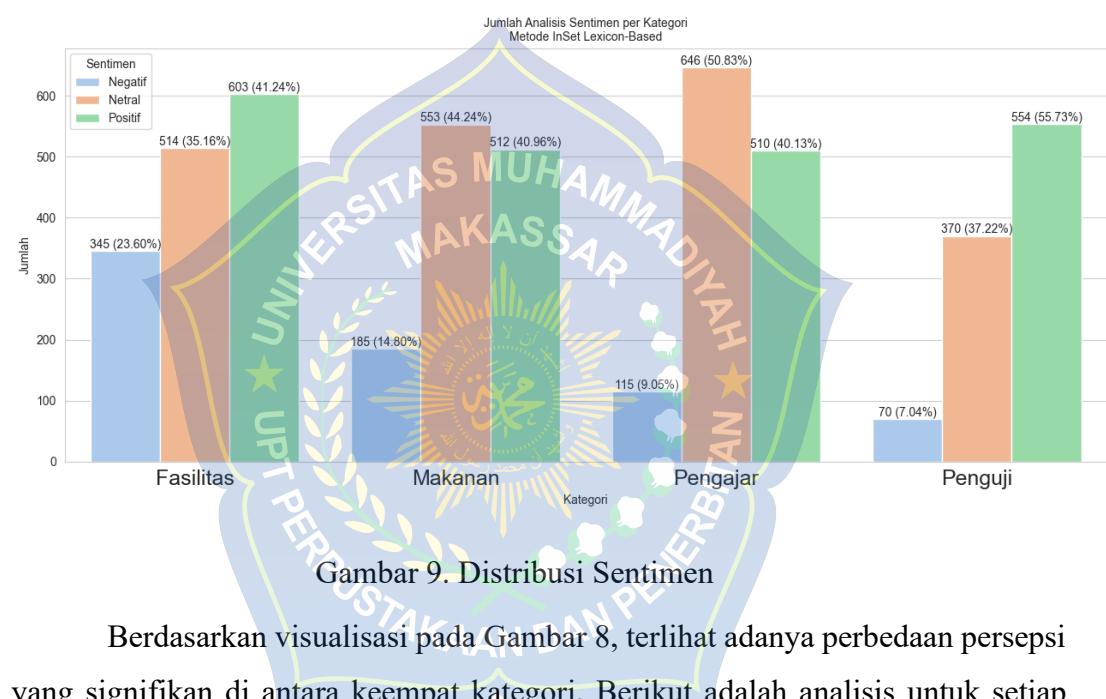
Tabel 31 menunjukkan hasil pelabelan sentimen untuk ulasan Penguji. Tabel ini memperlihatkan bagaimana skor sentimen dihitung berdasarkan kata-kata dalam teks, yang kemudian menentukan klasifikasi akhir sebagai positif, negatif, atau netral.

#### 1. Distribusi Sentimen Hasil Pelabelan

Untuk mendapatkan pemahaman yang lebih komprehensif, hasil pelabelan sentimen diklasifikasikan lebih lanjut ke dalam empat kategori spesifik yang merepresentasikan aspek-aspek utama pelatihan. Komponen tersebut mencakup empat aspek utama pelatihan, yaitu: Ruanagan dan Fasilitas, Makanan, Pengajar, serta, Penguji. Pendekatan ini bertujuan untuk memetakan secara jelas area yang menjadi kekuatan dan area yang memerlukan perbaikan dalam penyelenggaraan program VSGA. Distribusi sentimen untuk setiap komponen tersebut ditampilkan pada Tabel 32 dan divisualisasikan pada Gambar 4.2

Tabel 32. Distribusi Sentimen

Ulasan	Sentimen	Sentimen	Sentimen	Jumlah
	Positif	Netral	Negatif	Ulasan
Fasilitas	41.24%	35.16%	23.60%	1462
Makanan	40.96%	44.24%	14.80%	1250
Pengajar	40.13%	50.83%	9.05%	1271
Penguji	55.73%	37.22%	7.04%	994



Gambar 9. Distribusi Sentimen

Berdasarkan visualisasi pada Gambar 8, terlihat adanya perbedaan persepsi yang signifikan di antara keempat kategori. Berikut adalah analisis untuk setiap aspek:

a. Ruangan dan Fasilitas

Ulasan Fasilitas secara jelas menonjol sebagai area dengan tantangan terbesar. Aspek ini mencatatkan jumlah sentimen negatif tertinggi, yaitu sebanyak 345 ulasan (23.60%). Sentimen positif yang juga cukup tinggi, sebanyak 603 ulasan (41.24%), mengindikasikan adanya polarisasi persepsi. Artinya, meskipun sebagian fasilitas dinilai memadai, terdapat masalah

signifikan pada fasilitas lainnya yang sering dikeluhkan. Sentimen positif, seperti pada ulasan “fasilitas muas nyaman latih menteri erti”, menunjukkan apresiasi terhadap kondisi yang mendukung. Namun, jumlah sentimen negatif yang tinggi didorong oleh keluhan spesifik seperti “ac ruang ith kelas okm a matimati bikin ruang panas ganggu kurang nyaman” dan kritik terhadap ketersediaan alat seperti “instalasi listrik pakai ups dibutuhkan madam listrik tiba” yang dilabeli netral namun mengandung saran perbaikan yang krusial. Ulasan netral seperti “latih jgd cocok kursi kuliah mouse” menunjukkan bahwa banyak umpan balik bersifat deskriptif, namun tidak menutupi adanya isu-isu teknis yang perlu diperbaiki.

b. Makanan

Persepsi terhadap ulasan makanan terlihat lebih beragam. Kategori ini mencatatkan jumlah sentimen negatif yang lebih signifikan dibandingkan ulasan pengajar dan penguji, yaitu sebanyak 185 ulasan (14.80%). Tingginya jumlah ulasan netral sebanyak 553 ulasan (44.24%) sering kali bersifat deskriptif namun cenderung mengandung kepuasan implisit, seperti pada ulasan “makan enak” yang secara makna bernilai positif. Di sisi lain, sentimen positif yang eksplisit juga kuat, seperti pada ulasan “muas”. Namun, adanya kritik seperti “manurut lauk beefariasi sendok susah belah lauk mohon sendo kuat” menunjukkan adanya ruang untuk perbaikan dalam hal variasi menu dan kelengkapan alat makan. Hal ini menandakan bahwa meskipun banyak peserta merasa puas, terdapat sekelompok peserta yang secara aktif menyuarakan ketidakpuasan, menjadikan aspek makanan sebagai area yang memerlukan perhatian lebih lanjut.

c. Pengajar

Ulasan pengajar menerima umpan balik yang sangat positif, dengan 510 ulasan (40.13%) positif dan hanya 115 ulasan (9.05%) negatif. Tingginya jumlah sentimen netral, yaitu sebanyak 646 ulasan (50.83%), kemungkinan besar disebabkan oleh ulasan yang bersifat faktual atau deskriptif, misalnya

“materi rinci”, yang menyatakan fakta tanpa opini kuat. Kepuasan peserta secara jelas tergambar dalam ulasan seperti “puas latih uji kompetensi” dan “materi humble mudah paham moga latih materi”. Ulasan netral sederhana seperti “ajar” juga memperkuat kesan bahwa tidak ada keluhan signifikan. Secara keseluruhan, data ini menunjukkan bahwa kualitas pengajaran dinilai sebagai kekuatan utama program.

d. Penguji

Ulasan penguji menunjukkan proporsi sentimen yang paling positif di antara semua aspek. Dari total umpan balik, sebanyak 554 ulasan (55.73%) berlabel positif, sementara hanya 70 ulasan (7.04%) yang bersifat negatif. Tingginya sentimen positif ini mengindikasikan persepsi yang sangat baik terhadap profesionalisme dan kompetensi penguji, seperti yang tecermin pada ulasan “asesor support saran uji kompetensi” dan “uji cukup efisien”. Bahkan ulasan netral seperti “mantab” dan “asesor” secara kontekstual memperkuat citra positif. Jumlah sentimen negatif yang sangat rendah menjadikan aspek penguji sebagai salah satu keunggulan utama dari program pelatihan yang harus dipertahankan.

#### D. Model Klasifikasi Support Vector Machine

Tahapan implementasi merupakan fase inti dalam penelitian ini, di mana model klasifikasi dibangun untuk memprediksi sentimen dari data umpan balik peserta. Berdasarkan analisis awal yang menunjukkan adanya perbedaan karakteristik teks antar kategori, penelitian ini menerapkan pendekatan dengan membangun empat model SVM, masing-masing dilatih secara spesifik untuk ulasan Fasilitas, Makanan, Pengajar, dan Penguji. Pendekatan ini bertujuan untuk menangkap nuansa dan pola linguistik yang unik pada setiap aspek, sehingga diharapkan dapat menghasilkan klasifikasi yang lebih akurat.

## 1. Vektorisasi Fitur Menggunakan TF-IDF

Algoritma SVM bekerja dengan data numerik, sehingga data teks yang telah melalui tahap pra-pemrosesan perlu diubah menjadi representasi vektor. Proses ini dikenal sebagai vektorisasi atau ekstraksi fitur. Dalam penelitian ini, digunakan metode pembobotan TF-IDF, yang diimplementasikan menggunakan *TfidfVectorizer* dari pustaka *Scikit-learn*.

TF-IDF merupakan prasyarat teknis yang sangat penting karena kemampuannya dalam mengukur relevansi sebuah kata dalam sebuah ulasan. Metode ini memberikan bobot tinggi pada kata-kata yang signifikan untuk ulasan tertentu namun tidak umum di seluruh ulasan, sehingga model dapat lebih fokus pada kata-kata kunci yang membedakan sentimen. Hasil dari proses ini adalah sebuah matriks numerik di mana setiap baris merepresentasikan satu ulasan dan setiap kolom merepresentasikan bobot TF-IDF dari sebuah kata unik dalam korpus.

## 2. Model SVM dengan Kernel RBF

Model pertama yang dibangun untuk setiap kategori adalah model SVM standar dengan fungsi kernel *Radial Basis Function*. Kernel ini dipilih karena kemampuannya yang efektif dalam menangani hubungan *non-linear* antar fitur data. Pada tahap awal ini, model dikonfigurasi menggunakan parameter standar dari pustaka *Scikit-learn*, yaitu dengan nilai  $C=1.0$  dan  $\text{gamma}='scale'$ . Penggunaan parameter default ini bertujuan untuk membangun sebuah model dasar yang kinerjanya akan menjadi tolok ukur sebelum tahap optimasi parameter dilakukan.

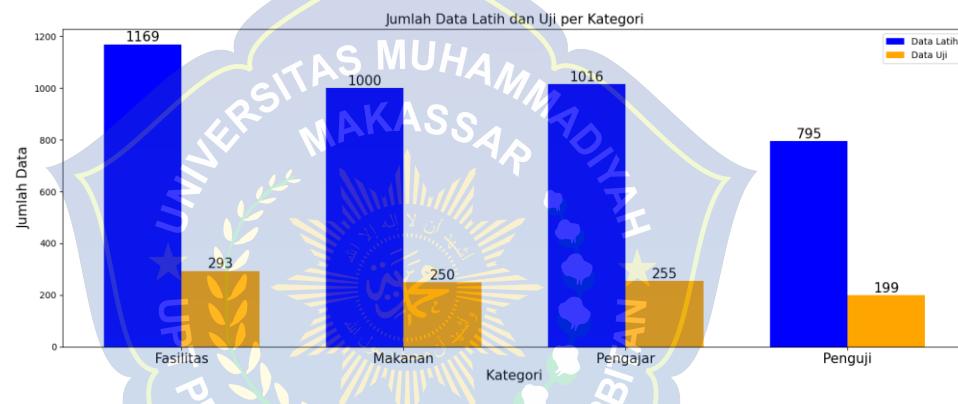
### a. Proses Pelatihan

Untuk setiap kategori, dataset dibagi menjadi dua himpunan menggunakan metode *hold-out validation* dengan proporsi 80% data latih dan 20% data uji. Penggunaan parameter  $\text{random\_state}=42$  memastikan bahwa pembagian data bersifat konsisten dan dapat direproduksi. Rincian jumlah data untuk setiap data ulasan disajikan pada Tabel 33.

Tabel 33. Rincian Pembagian Dataset Pelatihan dan Pengujian

Kategori	Jumlah Data Latih	Jumlah Data Uji	Total Data
Fasilitas	1169	293	1462
Makanan	1000	250	1250
Pengajar	1016	255	1271
Penguji	795	199	994

Setiap model SVM dilatih secara terpisah menggunakan data latih dari kategorinya masing-masing untuk mempelajari pola dan membentuk hyperplane pemisah sentimen.



Gambar 10. Jumlah Data Latih dan Uji per Kategori

#### b. Hasil Evaluasi Kinerja SVM

Untuk memperoleh pemahaman yang lebih mendalam terkait kinerja masing-masing model, dilakukan analisis menggunakan *classification report* yang memuat tiga metrik utama, yaitu *precision*, *recall*, dan *f1-score* untuk setiap kelas sentimen. Ketiga metrik tersebut memberikan perspektif yang lebih detail dalam menilai keunggulan dan keterbatasan model, baik dari segi ketepatan prediksi maupun kelengkapan deteksi sentimen.

Selain itu, untuk mengidentifikasi kesalahan klasifikasi secara lebih spesifik, digunakan *confusion matrix* yang divisualisasikan dalam bentuk *heatmap*. Matriks ini menampilkan perbandingan antara label aktual dan label prediksi model, sehingga pola kesalahan dapat diamati secara jelas dan terukur. Dengan demikian, proses evaluasi tidak hanya mengukur tingkat akurasi, tetapi juga memberikan wawasan mendalam mengenai area yang memerlukan perbaikan.

Hasil evaluasi kinerja untuk setiap model disajikan dalam bentuk *Classification Report* dan *headmap confusion matrix* berikut:

### 1). Kinerja Model Ulasan Fasilitas

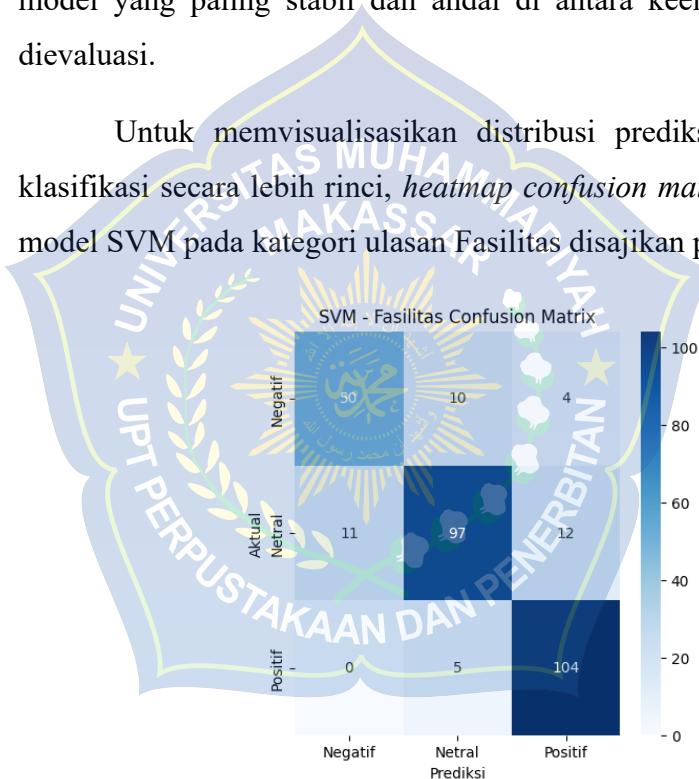
Model SVM yang dilatih pada data ulasan kategori Fasilitas berhasil mencapai akurasi keseluruhan sebesar 85.66%. Hasil ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan sentimen secara umum. Analisis lebih rinci pada Tabel 34 *Classification Report*.

Tabel 34. *Classification Report* Model SVM Ulasan Fasilitas

	Precision	Recall	F1 Score	Support
Negatif	0.82	0.78	0.80	64
Netral	0.87	0.81	0.84	120
Positif	0.87	0.95	0.91	109
<i>Accuracy</i>			0.86	293
<i>Macro</i>	0.85	0.85	0.85	293
<i>Average</i>				
<i>Weighted</i>	0.86	0.86	0.86	293
<i>Average</i>				

Analisis pada Tabel 33, menunjukkan kemampuan terbaik model ini terlihat pada klasifikasi sentimen positif, yang mencapai F1-score 0.91. Nilai ini didukung nilai recall yang tinggi 0.95, yang berarti model mampu mengidentifikasi 95% dari seluruh ulasan positif yang ada dengan benar. Untuk kelas netral dan negatif, model juga menunjukkan kinerja yang solid dengan F1-score masing-masing sebesar 0.84 dan 0.80. Keseimbangan antara precision dan recall pada semua kelas menjadikan model untuk kategori Fasilitas ini sebagai model yang paling stabil dan andal di antara keempat model yang dievaluasi.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM pada kategori ulasan Fasilitas disajikan pada Gambar 11.



Gambar 11. *Heatmap Confusion Matrix* SVM Ulasan Fasilitas

- Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{50+97+104}{293} = \frac{251}{293} \times 100 = 85,66$$

- Kelas Negatif

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{50+(97+12+5+104)}{50+(97+12+5+104)+11+(10+4)} \\
 &= \frac{50+218}{50+218+11+14} = \frac{268}{293} \times 100 = 91,46 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{50}{50+11} = 0,8196 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{50}{50+14} = 0,78125 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{82 \times 78}{82+78} = 2 \times \frac{6396}{160} \\
 &= 79,60
 \end{aligned}$$

### c). Kelas Netral

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{97+(50+4+104)}{97+(50+4+104)+(10+5)+(11+12)} \\
 &= \frac{97+158}{97+158+15+23} = \frac{255}{293} \times 100 = 87,03 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{97}{97+15} = 0,8660 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{97}{97+23} = 0,8083 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{87 \times 81}{87+81} = 2 \times \frac{7047}{168} \\
 &= 83,89
 \end{aligned}$$

### d). Kelas Positif

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{104+(50+10+11+97)}{104+(50+10+11+97)+(4+12)+5} \\
 &= \frac{104+168}{104+168+16+5} = \frac{272}{293} \times 100 = 92,83 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{104}{104+16} = 0,8666 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{104}{104+5} = 0,9541 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{87 \times 95}{87+95} = 2 \times \frac{8265}{182} \\
 &= 90,82
 \end{aligned}$$

## 2). Kinerja Model Ulasan Makanan

Model SVM yang dilatih pada data ulasan kategori Makanan berhasil mencapai akurasi keseluruhan sebesar 84.40%. Hasil ini

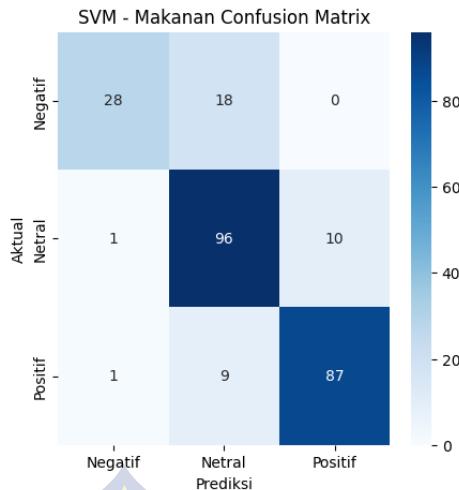
mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan sentimen secara umum. Analisis lebih rinci pada Tabel 35 *Classification Report* menunjukkan adanya kinerja yang berbeda secara signifikan di antara ketiga kelas sentimen.

Tabel 35. *Classification Report* Model SVM Ulasan Makanan

	Precision	Recall	F1 Score	Support
Negatif	0.93	0.61	0.74	46
Netral	0.78	0.90	0.83	107
Positif	0.90	0.90	0.90	97
<i>Accuracy</i>			0.84	250
<i>Macro</i>	0.87	0.80	0.82	250
<i>Average</i>				
<i>Weighted Average</i>	0.85	0.84	0.84	250

Analisis pada Tabel 34 menunjukkan bahwa model ini sangat andal dalam mengenali ulasan positif, dengan F1-score mencapai 0.90. Namun, tantangan utama terletak pada kelas negatif, di mana nilai recall yang rendah yaitu 0.61, mengindikasikan model kesulitan mendeteksi semua ulasan negatif, meskipun precision-nya tinggi 0.93. Hal ini menghasilkan F1-score sebesar 0.74 untuk kelas negatif dan 0.83 untuk kelas netral. Secara keseluruhan, model ini efektif dalam mengonfirmasi sentimen positif dan netral, namun kurang sensitif dalam menangkap seluruh kritik terkait makanan.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM pada kategori ulasan Makanan disajikan pada Gambar 12.



Gambar 12. *Heatmap Confusion Matrix* SVM Ulasan Makanan

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{28+96+87}{250} = \frac{211}{250} \times 100 = 84,4$$

- b). Kelas Negatif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{28+(96+10+9+87)}{28+(96+10+9+87)+2+18} = \frac{28+202}{28+202+2+18} = \frac{230}{250} \times 100 = 92$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{28}{28+2} = 0,9333$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{28}{28+18} = 0,6086$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{93 \times 61}{93 + 61} = 2 \times \frac{5673}{154} = 73,67$$

- c). Kelas Netral

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{96+(28+1+87)}{96+((28+1+87))+(18+9)+(1+10)} = \frac{96+116}{96+116+27+11} = \frac{212}{250} \times 100 = 84,0$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{96}{96+27} = 0,7804$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{96}{96+11} = 0,8971$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{78 \times 90}{78+90} = 2 \times \frac{7020}{168} = 83,57$$

d). Kelas Positif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{87+(28+18+1+96)}{87+(28+18+1+96)+10+10} = \frac{87+143}{87+143+10+10} = \frac{230}{250} \times 100 = 92$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{87}{87+10} = 0,8969$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{87}{87+10} = 0,8969$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{90 \times 90}{90+90} = 2 \times \frac{8100}{180} = 90$$

3). Kinerja Model Ulasan Pengajar

Model SVM yang dilatih pada data ulasan kategori Pengajar berhasil mencapai akurasi keseluruhan sebesar 84.70%. Hasil ini mengindikasikan bahwa model memiliki kemampuan yang cukup baik dalam membedakan sentimen secara umum. Analisis lebih rinci pada Tabel 36 *Classification Report* menyoroti adanya tantangan signifikan pada salah satu kelas sentimen.

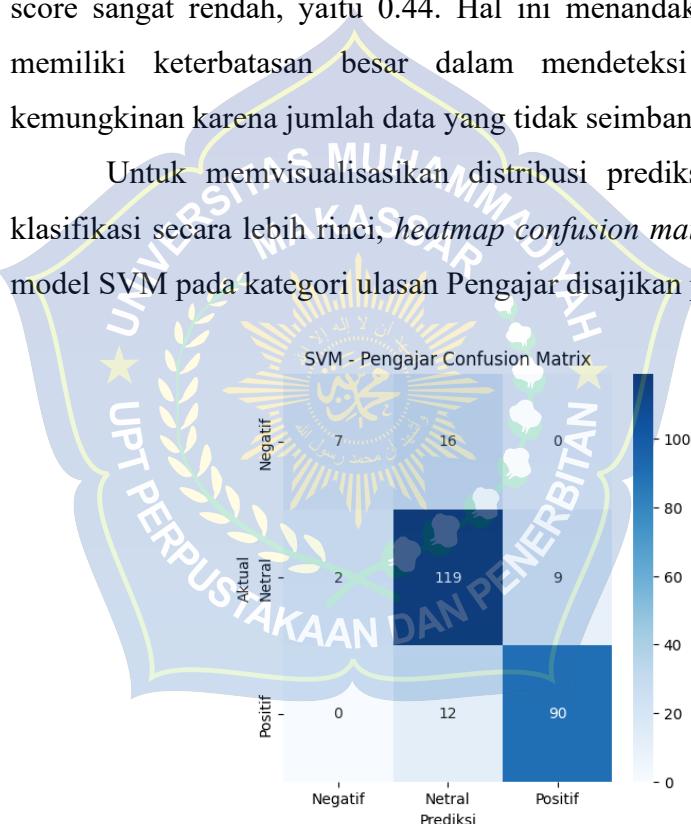
Tabel 36. *Classification Report* Model SVM Ulasan Pengajar

	Precision	Recall	F1 Score	Support
Negatif	0.78	0.30	0.44	23
Netral	0.81	0.92	0.86	130
Positif	0.91	0.88	0.90	102
<i>Accuracy</i>			0.85	255
<i>Macro</i>	0.83	0.70	0.73	255
<i>Average</i>				

	Precision	Recall	F1 Score	Support
<i>Weighted</i>	0.85	0.85	0.84	255
<i>Average</i>				

Analisis pada Tabel 35 menunjukkan kemampuan model yang sangat baik dalam mengklasifikasikan sentimen positif dengan F1-score 0.90 dan netral 0.86. Namun, performa model menurun drastis pada kelas negatif, dengan nilai recall hanya 0.30 yang menghasilkan F1-score sangat rendah, yaitu 0.44. Hal ini menandakan bahwa model memiliki keterbatasan besar dalam mendeteksi ulasan negatif, kemungkinan karena jumlah data yang tidak seimbang.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM pada kategori ulasan Pengajar disajikan pada Gambar 13.



Gambar 13. *Heatmap Confusion Matrix* SVM Ulasan Pengajar

- Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{7+119+90}{255} = \frac{216}{255} \times 100 = 84,70$$

b). Kelas Negatif

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{7+(119+9+12+90)}{7+(119+9+12+90)+2+16} \\
 &= \frac{7+230}{7+230+2+16} = \frac{237}{255} \times 100 = 92,94 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{7}{7+2} = 0,7777 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{7}{7+16} = 0,3043 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{78 \times 30}{78+30} = 2 \times \frac{2340}{108} \\
 &= 43,33
 \end{aligned}$$

c). Kelas Netral

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{119+(7+90)}{119+(7+90)+(16+12)+(2+9)} \\
 &= \frac{119+97}{119+97+28+11} = \frac{216}{255} \times 100 = 84,70 \\
 (2). \text{Presisi} &= \frac{216}{255} \times 100 = 84,70 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{119}{119+11} = 0,9153 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{81 \times 91}{81+91} = 2 \times \frac{7371}{172} \\
 &= 85,70
 \end{aligned}$$

d). Kelas Positif

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{90+(7+16+2+119)}{90+(7+16+2+119)+9+12} \\
 &= \frac{90+144}{90+144+9+12} = \frac{234}{255} \times 100 = 91,76 \\
 (2). \text{Presisi} &= \frac{234}{255} \times 100 = 91,76 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{90}{90+12} = 0,8823 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{91 \times 88}{91+88} = 2 \times \frac{8008}{179} \\
 &= 89,47
 \end{aligned}$$

#### 4). Kinerja Model Ulasan Pengaji

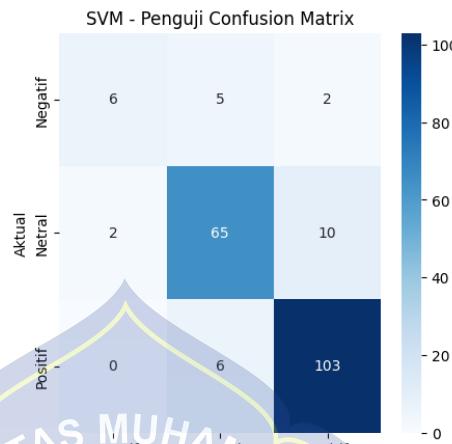
Model SVM yang dilatih pada data ulasan kategori Pengaji berhasil mencapai akurasi keseluruhan tertinggi, yaitu sebesar 87.43%. Hasil ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam membedakan sentimen secara umum. Analisis lebih rinci pada Tabel 37 *Classification Report* menunjukkan performa yang kuat, meskipun masih terdapat kelemahan pada kelas minoritas.

Tabel 37. *Classification Report* Model SVM Ulasan Pengaji

	Precision	Recall	F1 Score	Support
Negatif	0.75	0.46	0.57	13
Netral	0.86	0.84	0.85	77
Positif	0.90	0.94	0.92	109
<i>Accuracy</i>			0.87	199
<i>Macro</i>	0.83	0.75	0.78	199
<i>Average</i>				
<i>Weighted</i>	0.87	0.87	0.87	199
<i>Average</i>				

Analisis pada Tabel 36 menunjukkan keunggulan utama model ini dalam mengidentifikasi sentimen positif, yang mencapai F1-score 0.92 dengan recall tinggi 0.94. Kinerja pada kelas netral juga sangat solid dengan F1-score 0.85. Serupa dengan model Pengajar, model ini masih kesulitan mendeteksi sentimen negatif dengan F1-score 0.57, namun akurasi keseluruhannya yang tinggi menjadikan model ini sangat andal, terutama dalam mengonfirmasi persepsi positif dan netral dari peserta terhadap pengaji.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM pada kategori ulasan Penguji disajikan pada Gambar 14.



Gambar 14. *Heatmap Confusion Matrix* SVM Ulasan Penguji

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{6+65+103}{199} = \frac{174}{199} \times 100 = 87,43$$

- b). Kelas Negatif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{6+(65+10+6+103)}{6+(65+10+6+103)+2+7} \\ = \frac{6+184}{6+184+2+7} = \frac{190}{199} \times 100 = 95,47$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{6}{6+2} = 0,75$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{6}{6+7} = 0,4615$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{75 \times 46}{75+46} = 2 \times \frac{3450}{121} \\ = 57,02$$

- c). Kelas Netral

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{65+(6+2+103)}{65+(6+2+103)+(6+5)+(2+10)}$$

$$\begin{aligned}
&= \frac{65+111}{65+111+11+12} = \frac{176}{199} \times 100 = 88,44 \\
(2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{65}{65+11} = 0,8552 \\
(3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{65}{65+12} = 0,8441 \\
(4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{86 \times 84}{86+84} = 2 \times \frac{7224}{170} \\
&= 84,98
\end{aligned}$$

#### d). Kelas Positif

$$\begin{aligned}
(1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{103+(6+5+2+65)}{103+(6+5+2+65)+12+6} \\
&= \frac{103+78}{103+78+12+6} = \frac{181}{199} \times 100 = 90,95 \\
(2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{103}{103+12} = 0,8956 \\
(3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{103}{103+6} = 0,9449 \\
(4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{90 \times 94}{90+94} = 2 \times \frac{8460}{184} \\
&= 91,95
\end{aligned}$$

### 3. Model SVM dengan *Optimasi Particle Swarm Optimization*

Untuk mengatasi keterbatasan yang teridentifikasi pada model dasar, terutama dalam mengenali kelas sentimen minoritas, penelitian ini menerapkan tahap optimasi parameter. Proses ini bertujuan untuk menemukan kombinasi *hyperparameter* C dan gamma yang paling optimal untuk setiap model SVM, sehingga dapat meningkatkan kinerja klasifikasi secara keseluruhan.

#### a. Optimasi Parameter

Optimasi dilakukan menggunakan algoritma *Particle Swarm Optimization*. Algoritma ini merupakan sebuah metode pencarian optimasi yang terinspirasi dari perilaku sosial kawanan burung atau ikan dalam mencari sumber makanan. Dalam konteks ini, “makanan” adalah kombinasi parameter yang menghasilkan akurasi tertinggi.

Dua *hyperparameter* utama pada SVM dengan kernel RBF yang dioptimasi adalah:

- 1). Parameter C, Mengontrol *trade-off* antara mencapai margin yang lebar dan meminimalkan kesalahan klasifikasi pada data latih.
- 2). Parameter Gamma, Menentukan jangkauan pengaruh dari satu sampel data latih.

Proses PSO ini bekerja dengan melepaskan sejumlah "partikel" untuk bergerak di dalam sebuah ruang pencarian yang telah ditentukan batas-batasnya. Dalam penelitian ini, ruang pencarian untuk parameter C dan gamma dibatasi dengan nilai terendah [0.1, 0.0001] dan nilai tertinggi [1000, 10]. Penetapan rentang ini bertujuan agar proses pencarian dapat menjelajahi berbagai kemungkinan nilai, baik yang kecil maupun yang besar, untuk menemukan kombinasi terbaik.

Setiap partikel akan dinilai kinerjanya berdasarkan kemampuannya memaksimalkan rata-rata akurasi dari proses cross-validation 5-fold pada data latih. Penggunaan cross-validation bertujuan untuk mendapatkan estimasi kinerja model yang lebih stabil dan menghindari overfitting terhadap satu set data validasi tunggal. Dengan pengaturan swarmsize=20 dan maxiter=30, algoritma secara berulang mencari kombinasi C dan gamma terbaik.. Hasil parameter optimal yang ditemukan untuk setiap kategori disajikan pada Tabel 38.

Tabel 38. Parameter Optimal Hasil Pencarian PSO

Kategori	Parameter C	Parameter Gamma
Fasilitas	696.3381	0.0125
Makanan	915.1470	0.0570
Pengajar	309.7231	0.2062
Penguji	563.4800	0.4049

Parameter-parameter optimal inilah yang kemudian digunakan untuk melatih ulang keempat model SVM, yang kinerjanya akan dievaluasi pada bagian evaluasi kinerja model dengan optimasi PSO.

#### b. Evaluasi Kinerja Model SVM dengan Optimasi PSO

Untuk memperoleh pemahaman yang lebih mendalam terkait kinerja model SVM yang telah dioptimasi menggunakan PSO, dilakukan analisis menggunakan *classification report* yang memuat tiga metrik utama, yaitu *precision*, *recall*, dan f1-score pada setiap kelas sentimen. Ketiga metrik ini memberikan gambaran yang lebih terperinci mengenai performa model, baik dari segi ketepatan prediksi maupun kelengkapan deteksi sentimen, serta keseimbangan antara keduanya. Melalui metrik tersebut, dapat diidentifikasi keunggulan model dalam mengenali pola sentimen tertentu sekaligus mengetahui keterbatasan yang masih ada pada kelas sentimen lain.

Selain itu, untuk mendalami analisis terhadap kesalahan klasifikasi, digunakan *confusion matrix* yang divisualisasikan dalam bentuk *heatmap*. Visualisasi ini memetakan secara jelas perbandingan antara label aktual dan hasil prediksi model, sehingga pola kesalahan dapat diidentifikasi dengan lebih spesifik. Informasi dari *confusion matrix* memungkinkan peneliti mengamati seberapa besar tingkat kesalahan antar kelas dan menentukan area yang memerlukan perbaikan lebih lanjut.

Hasil evaluasi kinerja untuk setiap model disajikan dalam bentuk *Classification Report* dan *headmap confusion matrix* berikut:

##### 1). Kinerja Model Ulasan Fasilitas dengan Optimasi

Setelah melalui optimasi parameter menggunakan PSO, model SVM untuk kategori Fasilitas menunjukkan peningkatan kinerja yang signifikan, dengan akurasi keseluruhan mencapai 88.05%. Capaian ini menegaskan bahwa optimasi berhasil meningkatkan kemampuan model

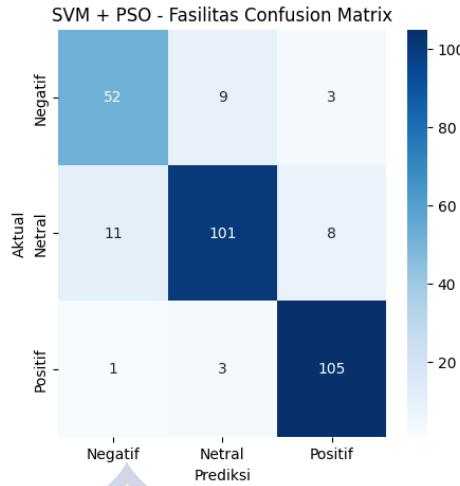
dalam membedakan sentimen secara umum. Analisis lebih rinci pada Tabel 39 *Classification Report*.

Tabel 39. *Classification Report* Model SVM dengan PSO Ulasan Fasilitas

	Precision	Recall	F1 Score	Support
Negatif	0.81	0.81	0.81	64
Netral	0.89	0.84	0.87	120
Positif	0.91	0.96	0.93	109
<i>Accuracy</i>			0.88	293
<i>Macro</i>	0.87	0.87	0.87	293
<i>Average</i>				
<i>Weighted</i>	0.88	0.88	0.88	293
<i>Average</i>				

Analisis pada Tabel 38 menunjukkan bahwa performa terbaik tetap pada klasifikasi sentimen positif, dengan F1-score yang sangat tinggi, yaitu 0.93. Nilai ini didukung oleh recall 0.96, yang berarti model mampu mengidentifikasi 96% ulasan positif dengan benar. Peningkatan signifikan juga terlihat pada kelas negatif, yang kini mencapai F1-score 0.81, menunjukkan kemampuan deteksi yang jauh lebih seimbang. Kinerja pada kelas netral juga tetap kuat dengan F1-score 0.87, menjadikan model hasil optimasi ini sangat andal untuk analisis pada kategori Fasilitas.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM dengan Optimasi PSO pada kategori ulasan Fasilitas disajikan pada Gambar 15.



Gambar 15. *Heatmap Confusion Matrix* SVM+PSO Ulasan Fasilitas

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{52+101+105}{293} = \frac{258}{293} \times 100 = 88,05$$

- b). Kelas Negatif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{52+(101+8+3+105)}{52+(101+8+3+105)+12+12}$$

$$= \frac{52+217}{52+217+12+12} = \frac{269}{293} \times 100 = 91,80$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{52}{52+12} = 0,8125$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{52}{52+12} = 0,8125$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{81 \times 81}{81+81} = 2 \times \frac{6561}{162} = 81$$

- c). Kelas Netral

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{101+(52+3+105+1)}{101+(52+3+105+1)+12+19}$$

$$= \frac{101+161}{101+161+12+19} = \frac{262}{293} \times 100 = 89,41$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{101}{101+12} = 0,8938$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{101}{101+19} = 0,8416$$

$$(4). F1\text{-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{89 \times 84}{89 + 84} = 2 \times \frac{7476}{173} = 86,42$$

d). Kelas Positif

$$\begin{aligned}(1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{105+(52+9+11+101)}{105+(52+9+11+101)+11+4} \\&= \frac{105+173}{105+173+11+4} = \frac{278}{293} \times 100 = 94,88 \\(2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{105}{105+11} = 0,9051 \\(3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{105}{105+4} = 0,9633 \\(4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{91 \times 96}{91 + 96} = 2 \times \frac{8736}{187} \\&= 93,43\end{aligned}$$

2). Kinerja Model Ulasan Makanan dengan Optimasi

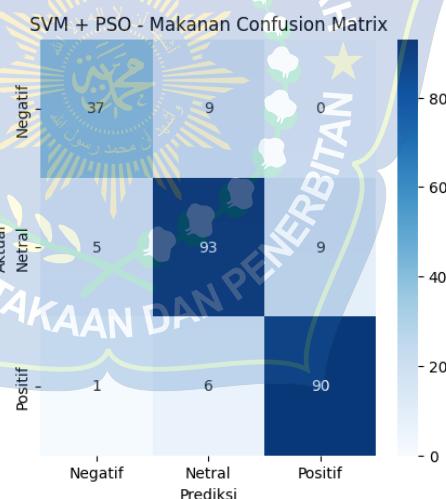
Penerapan optimasi PSO pada model SVM kategori Makanan berhasil meningkatkan akurasi secara substansial menjadi 88.00%. Hasil ini menunjukkan bahwa penyesuaian parameter C dan gamma secara efektif mengatasi keterbatasan yang ada pada model dasar. Analisis lebih rinci pada Tabel 40 *Classification Report*.

Tabel 40. *Classification Report* Model SVM dengan PSO Ulasan Makanan

	Precision	Recall	F1 Score	Support
Negatif	0.86	0.80	0.83	46
Netral	0.86	0.87	0.87	107
Positif	0.91	0.93	0.92	97
<i>Accuracy</i>			0.88	250
<i>Macro</i>	0.88	0.87	0.87	250
<i>Average</i>				
<i>Weighted</i>	0.88	0.88	0.88	250
<i>Average</i>				

Analisis pada Tabel 39 menunjukkan bahwa model yang telah dioptimasi kini memiliki performa yang jauh lebih seimbang. Kelas positif tetap menjadi yang terkuat dengan F1-score 0.92. Perbaikan paling signifikan terjadi pada kelas negatif, di mana F1-score meningkat menjadi 0.83 berkat kenaikan recall ke angka 0.80. Kelas netral juga menunjukkan kinerja yang sangat baik dengan F1-score 0.87. Peningkatan ini membuktikan bahwa model hasil optimasi mampu mengenali sentimen positif, netral, dan negatif dengan tingkat keandalan yang tinggi.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM dengan Optimasi PSO pada kategori ulasan Makanan disajikan pada Gambar 16.



Gambar 16. *Heatmap Confusion Matrix* SVM+PSO Ulasan Makanan

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{37+93+90}{250} = \frac{220}{250} \times 100 = 88$$

b). Kelas Negatif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{37+(93+9+6+90)}{37+(93+9+6+90)+6+9} \\ = \frac{37+198}{37+198+6+9} = \frac{235}{250} \times 100 = 94$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{37}{37+6} = 0,8604$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{37}{37+9} = 0,8043$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{86 \times 80}{86+80} = 2 \times \frac{6880}{166} \\ = 82,89$$

c). Kelas Netral

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{93+(37+1+90)}{93+(37+1+90)+(9+6)+(5+9)} \\ = \frac{93+128}{93+128+15+14} = \frac{221}{250} \times 100 = 88,4$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{93}{93+15} = 0,8611$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{93}{93+14} = 0,8691$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{86 \times 87}{86+87} = 2 \times \frac{7482}{173} \\ = 86,49$$

d). Kelas Positif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{90+(37+9+5+93)}{90+(37+9+5+93)+9+7} \\ = \frac{90+144}{90+144+9+7} = \frac{234}{250} \times 100 = 93,6$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{90}{90+9} = 0,9090$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{90}{90+7} = 0,9278$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{91 \times 93}{91+93} = 2 \times \frac{8463}{184} \\ = 91,98$$

### 3). Kinerja Model Ulasan Pengajar dengan Optimasi

Model SVM untuk kategori Pengajar menunjukkan peningkatan akurasi yang paling signifikan setelah optimasi PSO, mencapai 89.01%. Peningkatan ini secara jelas menunjukkan efektivitas PSO dalam memperbaiki kemampuan klasifikasi, terutama pada dataset dengan kelas minoritas yang sulit dideteksi. Analisis lebih rinci pada Tabel 41 *Classification Report*.

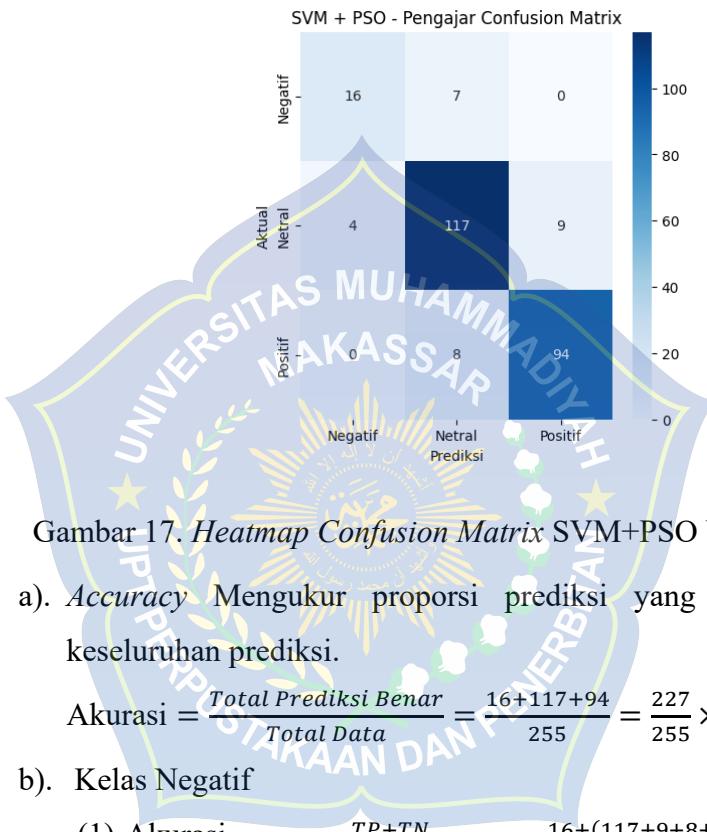
Tabel 41. *Classification Report* Model SVM dengan PSO Ulasan Pengajar

	Precision	Recall	F1 Score	Support
Negatif	0.80	0.70	0.74	23
Netral	0.89	0.90	0.89	130
Positif	0.91	0.92	0.92	102
Accuracy			0.89	255
Macro	0.87	0.84	0.85	255
Average				
Weighted	0.89	0.89	0.89	255
Average				

Analisis pada Tabel 40 menunjukkan bahwa kelemahan utama pada model dasar berhasil diatasi secara efektif. Nilai recall untuk kelas negatif melonjak dari 0.30 menjadi 0.70, yang meningkatkan F1-scorenya menjadi 0.74. Meskipun masih menjadi kelas dengan performa terendah, peningkatannya sangat signifikan. Sementara itu, kelas positif dan netral menunjukkan kinerja yang luar biasa dengan F1-score masing-masing 0.92 dan 0.89. Hasil ini membuktikan bahwa optimasi

PSO berhasil membuat model lebih sensitif terhadap kritik, sekaligus mempertahankan keunggulannya pada kelas lain.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM dengan Optimasi PSO pada kategori ulasan Pengajar disajikan pada Gambar 17.



Gambar 17. *Heatmap Confusion Matrix* SVM+PSO Ulasan Pengajar

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{16+117+94}{255} = \frac{227}{255} \times 100 = 89,01$$

- b). Kelas Negatif

$$(1). \text{ Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+(117+9+8+94)}{16+(117+9+8+94)+4+7} = \frac{16+228}{16+228+4+7} = \frac{244}{255} \times 100 = 95,68$$

$$(2). \text{ Presisi} = \frac{TP}{TP+FP} = \frac{16}{16+4} = 0,8$$

$$(3). \text{ Recall} = \frac{TP}{TP+FN} = \frac{16}{16+7} = 0,6956$$

$$(4). \text{ F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{0,8 \times 0,6956}{0,8 + 0,6956} = 2 \times \frac{0,5600}{1,4956} = 2 \times 0,3818 = 0,7636$$

c). Kelas Netral

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{117+(16+94)}{117+(16+94)+(7+8)+(4+9)} \\
 &= \frac{117+110}{117+110+15+13} = \frac{227}{255} \times 100 = 89,01 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{117}{117+15} = 0,8863 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{117}{117+13} = 0,9 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{89 \times 90}{89+90} = 2 \times \frac{8010}{179} \\
 &= 89,49
 \end{aligned}$$

d). Kelas Positif

$$\begin{aligned}
 (1). \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{94+(16+7+4+117)}{94+(16+7+4+117)+9+8} \\
 &= \frac{94+144}{94+144+9+8} = \frac{238}{255} \times 100 = 93,33 \\
 (2). \text{Presisi} &= \frac{TP}{TP+FP} = \frac{94}{94+9} = 0,9126 \\
 (3). \text{Recall} &= \frac{TP}{TP+FN} = \frac{94}{94+8} = 0,9215 \\
 (4). \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{91 \times 92}{91+92} = 2 \times \frac{8372}{183} \\
 &= 91,49
 \end{aligned}$$

4). Kinerja Model Ulasan Penguci dengan Optimasi

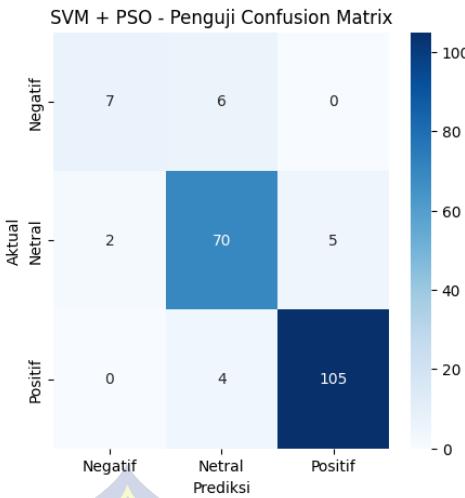
Setelah optimasi PSO, model SVM untuk kategori Penguci mencatatkan akurasi tertinggi di antara semua model, yaitu sebesar 91.45%. Hasil ini mengukuhkan bahwa model yang telah dioptimasi memiliki kapabilitas prediksi yang sangat andal dan presisi. Analisis lebih rinci pada Tabel 42 *Classification Report*.

Tabel 42. *Classification Report* Model SVM dengan PSO Ulasan Pengajar

	Precision	Recall	F1 Score	Suppot
Negatif	0.78	0.54	0.64	13
Netral	0.88	0.91	0.89	77
Positif	0.95	0.96	0.96	109
<i>Accuracy</i>			0.91	199
<i>Macro</i>	0.87	0.80	0.83	199
<i>Average</i>				
<i>Weighted</i>	0.91	0.91	0.91	199
<i>Average</i>				

Analisis pada Tabel 41 menunjukkan dominasi model dalam mengenali sentimen positif, dengan F1-score mencapai 0.96. Kinerja pada kelas netral juga sangat kuat dengan F1-score 0.89. Meskipun optimasi berhasil meningkatkan recall pada kelas negatif menjadi 0.54 dan F1-score menjadi 0.64, kelas ini tetap menjadi tantangan. Namun, akurasi keseluruhan yang sangat tinggi menjadikan model ini sangat efektif untuk analisis sentimen pada kategori Pengujii, dengan tingkat kepercayaan yang luar biasa pada prediksi positif dan netral.

Untuk memvisualisasikan distribusi prediksi dan kesalahan klasifikasi secara lebih rinci, *heatmap confusion matrix* dari performa model SVM dengan Optimasi PSO pada kategori ulasan Pengujii disajikan pada Gambar 18.



Gambar 18. Heatmap Confusion Matrix SVM+PSO Ulasan Pengujian

- a). *Accuracy* Mengukur proporsi prediksi yang benar terhadap keseluruhan prediksi.

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} = \frac{7+70+105}{199} = \frac{182}{199} \times 100 = 91,45$$

- b). Kelas Negatif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{7+(70+5+4+105)}{7+(70+5+4+105)+2+6} \\ = \frac{7+184}{7+184+2+6} = \frac{191}{199} \times 100 = 95,97$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{7}{7+2} = 0,7777$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{7}{7+6} = 0,5384$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{78 \times 54}{78 + 54} = 2 \times \frac{4212}{132} \\ = 63,81$$

- c). Kelas Netral

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{70+(7+105)}{70+(7+105)+(6+4)+7} \\ = \frac{70+112}{70+112+10+7} = \frac{182}{199} \times 100 = 91,45$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{70}{70+10} = 0,875$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{70}{70+7} = 0,9090$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{87 \times 91}{87+91} = 2 \times \frac{7917}{178} = 88,95$$

d). Kelas Positif

$$(1). \text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{105+(7+6+2+70)}{105+(7+6+2+70)+5+4} = \frac{105+85}{105+85+5+4} = \frac{190}{199} \times 100 = 91,45$$

$$(2). \text{Presisi} = \frac{TP}{TP+FP} = \frac{105}{105+5} = 0,9545$$

$$(3). \text{Recall} = \frac{TP}{TP+FN} = \frac{105}{105+4} = 0,9633$$

$$(4). \text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \times \frac{95 \times 96}{95+96} = 2 \times \frac{9120}{191} = 95,49$$

E. Perhitungan Manual Model Klasifikasi SVM

Untuk memperkuat hasil eksperimen yang dilakukan dengan perangkat lunak, pada bagian ini ditampilkan perhitungan manual sederhana menggunakan algoritma SVM dengan kernel RBF. Tujuannya adalah memberikan gambaran bagaimana proses klasifikasi dilakukan pada data teks yang telah direpresentasikan ke dalam bentuk nilai numerik TF-IDF.

1. Data Latih dan Data Uji

Sebagai contoh, digunakan dua data latih dan dua data uji dari ulasan fasilitas sebagaimana ditunjukkan pada Tabe 43 dan Tabel 44.

Tabel 43. Data Latih

Teks	Sentimen	TF-IDF Value	$\alpha$	$b$
aman nice	Positif	0.001487	-0.66391	-0.48801
terima kasih bimbing	Positif	0.002752	-0.41936	-0.48801
latih uji tes sanagat				
.....	.....	.....	....	....

Teks	Sentimen	TF-IDF Value	$\alpha$	$b$
ac dingin depan perhati fasilitas	Netral	0.002361	-0.2762	-0.48801

Tabel 44. Data Uji

Teks	Sentimen	TF-IDF Value	$b$
puas lengkap fasilitas berterimakasih bimbing ya latih bawah pulng bekal ilmu terimakasih	Positif	0.003063	-0.48801
mohon tingkat	Positif	0.001514	-0.48801
....	....	....	...
alhamdulillah ruang bagus pakai ruang kelas	Negatif	0.002352	-0.48801
oke	Positif	0.001075	-0.48801

## 2. Rumus Perhitungan

Algoritma SVM dengan kernel RBF menggunakan fungsi keputusan sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b$$

Dengan:

$\alpha_i$  : bobot *support vector* ke-i

$y_i$  : label kelas dari support vector ke-i (+1 untuk Positif, -1 untuk Negatif)

$K(x_i, x)$  : fungsi kernel RBF antara data latih ke-i dan data uji

$b$  : bias (*intercept*) model

Kernel RBF didefinisikan dengan persamaan:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$$

dengan parameter  $\sigma$  diperoleh dari  $\gamma = \frac{1}{2\sigma^2}$  pada model yang digunakan  $\gamma = scale$ , sehingga nilai dihitung otomatis oleh library SVM

### 3. Langkah Perhitungan

#### a. Data Uji 1

- 1). Hitung jarak kuadrat antara data uji dan setiap *support vector*

$$\|x - x_1\|^2 = (0.003063 - 0.001487)^2 = 0.000002483776$$

$$\|x - x_2\|^2 = (0.003063 - 0.002752)^2 = 0.000000096721$$

- 2). Hitung nilai kernel dengan asumsi  $\sigma^2 = 1$

$$K(x_1, x) = \exp(-0.000002483776) \approx 0.999997516224$$

$$K(x_2, x) = \exp(-0.000000096721) \approx 0.999999903279$$

- 3). Subtitusi ke fungsi keputusan

$$f(x) = (\alpha_1 \cdot y_1 \cdot K(x_1, x)) + (\alpha_2 \cdot y_2 \cdot K(x_2, x)) + b$$

$$f(x) = (-0.66391(+1)(0.999997516224)) +$$

$$(-0.41936(+1)(0.999999903279)) + (-0.48801)$$

$$f(x) \approx -0.663908350996 - 0.419359959439 - 0.48801 = \\ -1.571278310435$$

- 4). Keputusan

Karena  $f(x) < 0$ , maka prediksi = negatif

#### b. Data Uji 2

- 1). Hitung jarak kuadrat antara data uji dan setiap *support vector*

$$\|x - x_1\|^2 = (0.001514 - 0.001487)^2 = 0.000027$$

$$\|x - x_2\|^2 = (0.001514 - 0.002752)^2 = -0.001238$$

- 2). Hitung nilai kernel dengan asumsi  $\sigma^2 = 1$

$$K(x_1, x) = \exp(-0.000027) \approx 0.999973$$

$$K(x_2, x) = \exp(-(-0.001238)) \approx 1.001238$$

3). Subtitusi ke fungsi keputusan

$$f(x) = (\alpha_1 \cdot y_1 \cdot K(x_1, x)) + (\alpha_2 \cdot y_2 \cdot K(x_2, x)) + b$$
$$f(x) = (-0.66391(+1)(0,999973)) + (-0.41936(+1)(1.001238) + (-0.48801)$$
$$f(x) \approx -0.66389207443 - 0.41987916768 - 0.488011 = -1.57178224211$$

4). Keputusan

Karena  $f(x) < 0$ , maka prediksi = negatif

#### 4. Kesimpulan Perhitungan Manual

Hasil perhitungan manual menunjukkan bahwa kedua data uji diprediksi sebagai sentimen negatif, meskipun label aktualnya adalah positif. Perbedaan ini terjadi karena perhitungan manual hanya menggunakan sebagian kecil representasi vektor yaitu rata-rata dari proses TF-IDF, serta nilai parameter kernel yang disederhanakan, sehingga keputusan klasifikasi tidak sepenuhnya identik dengan implementasi SVM yang dilakukan pada program. Walaupun demikian, proses ini memberikan gambaran jelas mengenai bagaimana nilai  $\alpha$ , bias, serta kernel RBF berkontribusi dalam menentukan hasil klasifikasi pada SVM.

#### F. Analisis Perbandingan Kinerja Model

Bagian ini membahas secara mendalam hasil perbandingan performa dua pendekatan klasifikasi sentimen, yakni SVM dengan *kernel Radial Basis Function* dan SVM yang telah dioptimasi menggunakan algoritma *Particle Swarm Optimization*. Perbandingan dilakukan untuk menilai efektivitas optimasi parameter terhadap peningkatan performa klasifikasi sentimen berdasarkan empat kategori data ulasan, yaitu fasilitas, makanan, pengajar, dan penguji. Analisis ini menjawab rumusan masalah kedua dalam penelitian, yaitu mengidentifikasi model mana yang menunjukkan performa lebih unggul dalam tugas klasifikasi sentimen.

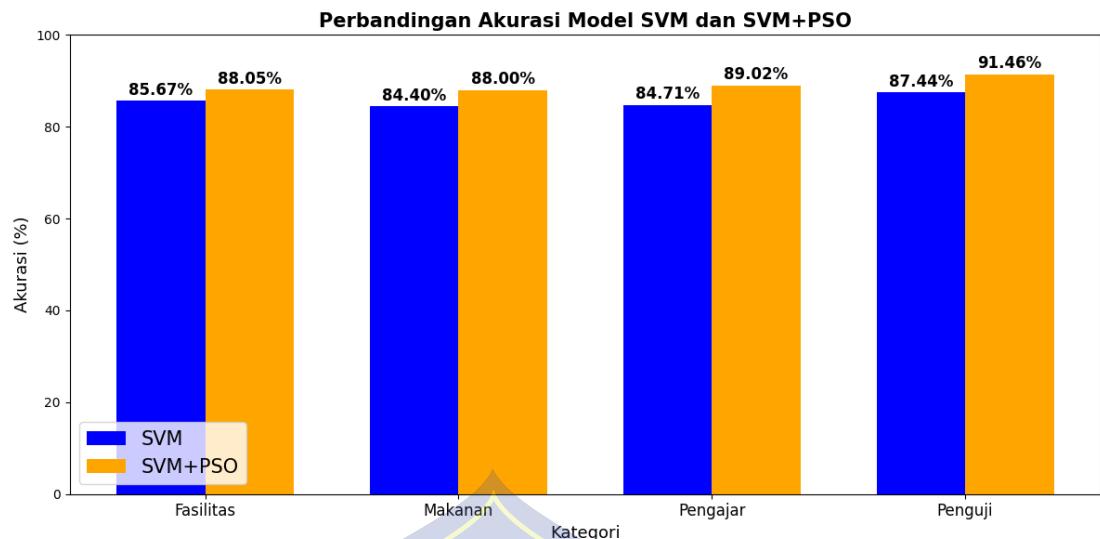
## 1. Perbandingan Metrik Evaluasi

Evaluasi performa model dilakukan dengan mengacu pada empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Keempat metrik ini dihitung dari *confusion matrix* hasil klasifikasi terhadap data uji. Nilai yang digunakan untuk perbandingan ini adalah nilai *weighted average* dari *Classification Report* setiap model. Penggunaan *weighted average* dipilih karena metrik ini memberikan gambaran kinerja yang seimbang dengan memperhitungkan jumlah sampel pada setiap kelas sentimen, sehingga relevan untuk dataset yang memiliki potensi ketidakseimbangan kelas.

Untuk memvisualisasikan perbedaan kinerja secara langsung, hasil evaluasi dari kedua model pada keempat kategori penilaian, dirangkum dalam tabel perbandingan. Tabel 45 menyajikan perbandingan dari setiap metrik dan divisualisasikan pada Gambar 19.

Tabel 45. Perbandingan Kinerja Model

Kategori	Model	Akurasi	Presisi	Recall	F1-Score
Fasilitas	SVM	85.66%	0.86	0.86	0.86
	SVM + PSO	88.05%	0.88	0.88	0.88
	SVM	84.40%	0.85	0.84	0.84
Makanan	SVM + PSO	88.00%	0.88	0.88	0.88
	SVM	84.70%	0.85	0.85	0.84
	SVM + PSO	89.01%	0.89	0.89	0.89
Pengajar	SVM	87.43%	0.87	0.87	0.87
	SVM + PSO	91.45%	0.91	0.91	0.91



Gambar 19. Perbandingan Kinerja Model

Secara keseluruhan, data pada tabel 45 dan gambar 19 menunjukkan bahwa penerapan optimasi PSO secara konsisten memberikan peningkatan performa pada model SVM di seluruh kategori yang dianalisis. Peningkatan ini terlihat merata pada semua metrik evaluasi, yang mengindikasikan bahwa pencarian *hyperparameter* C dan *gamma* yang optimal melalui PSO berhasil menemukan konfigurasi model yang secara fundamental lebih baik daripada konfigurasi standar.

## 2. Analisis Peningkatan Performa

Data pada Tabel 42 secara konsisten menunjukkan bahwa penerapan optimasi PSO memberikan peningkatan kinerja yang signifikan pada seluruh metrik evaluasi di keempat kategori yang diuji. Peningkatan akurasi yang paling menonjol tercatat pada kategori Pengajar, dengan kenaikan sebesar 4.31 poin persentase, dari 84.70% menjadi 89.01%. Peningkatan substansial juga terlihat pada kategori Pengugi yang naik sebesar 4.02 poin dan kategori Makanan yang naik 3.60 poin.

Peningkatan performa yang paling signifikan pada kategori Pengajar dan Pengaji dapat diatribusikan pada kemampuan PSO dalam mengatasi kelemahan utama yang teridentifikasi pada model dasar. Sebagaimana dibahas sebelumnya, model SVM standar menunjukkan kesulitan dalam mengenali kelas minoritas, khususnya sentimen negatif yang jumlah datanya relatif sedikit dan ekspresi kritiknya cenderung lebih halus. Hal ini tercermin dari nilai recall yang sangat rendah untuk kelas negatif pada model SVM dasar, yaitu hanya 0.30 untuk kategori Pengajar dan 0.46 untuk kategori Pengaji.

Optimasi PSO berhasil menemukan kombinasi parameter C dan gamma yang lebih optimal, sehingga mampu membentuk *hyperplane* pemisah yang lebih sensitif dan akurat. Dengan parameter yang lebih baik, model SVM dengan optimasi PSO tidak hanya lebih unggul dalam klasifikasi secara umum, tetapi juga secara spesifik lebih andal dalam mendekripsi sampel dari kelas minoritas. Implikasinya adalah peningkatan nilai recall dan F1-score untuk kelas negatif, yang membuat model menjadi lebih seimbang dan dapat dipercaya.

Pada kategori Fasilitas dan Makanan, meskipun peningkatannya tidak sedramatis dua kategori lainnya, optimasi PSO tetap membuktikan efektivitasnya. Hal ini menunjukkan bahwa bahkan pada dataset dengan distribusi kelas yang lebih seimbang, penentuan parameter yang tepat tetap menjadi faktor krusial untuk mencapai performa maksimal.

Secara keseluruhan, hasil analisis ini mengonfirmasi bahwa optimasi parameter menggunakan PSO merupakan langkah yang sangat efektif dan relevan. Pendekatan ini berhasil meningkatkan ketahanan dan akurasi model, terutama dalam menghadapi tantangan data yang tidak seimbang dan variasi ekspresi linguistik yang kompleks dalam data teks ulasan.

#### G. Interpretasi Hasil Analisis Sentimen

Pada bagian ini, membahas makna di balik hasil klasifikasi sentimen yang telah diperoleh. Tujuannya adalah untuk memahami secara mendalam apa yang

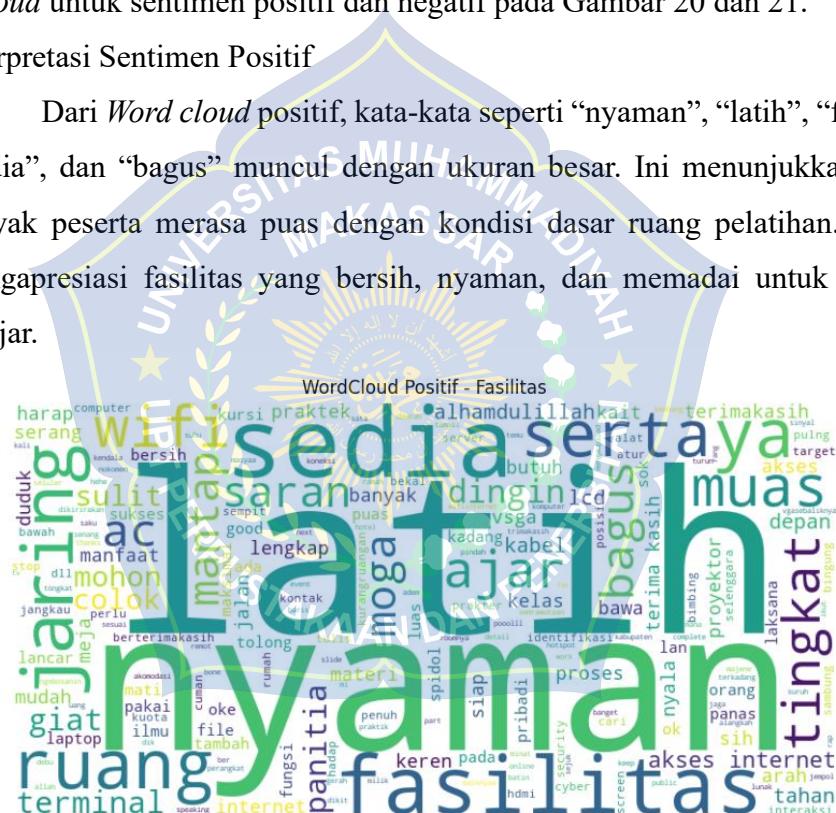
dirasakan dan disampaikan oleh para peserta pelatihan mengenai empat aspek utama yaitu fasilitas, makanan, pengajar, dan penguji. Analisis ini didukung oleh visualisasi *word cloud* untuk melihat kata-kata yang paling dominan serta tabel perbandingan untuk menunjukkan bagaimana model yang dioptimasi memberikan hasil yang lebih akurat.

#### 1. Analisis Sentimen Terhadap Ulasan Ruangan dan Fasilitas

Aspek fasilitas merupakan salah satu komponen dengan dinamika sentimen yang paling terlihat. Untuk memahami persepsi peserta secara visual, disajikan *word cloud* untuk sentimen positif dan negatif pada Gambar 20 dan 21.

a. Interpretasi Sentimen Positif

Dari *Word cloud* positif, kata-kata seperti “nyaman”, “latih”, “fasilitas”, “sedia”, dan “bagus” muncul dengan ukuran besar. Ini menunjukkan bahwa banyak peserta merasa puas dengan kondisi dasar ruang pelatihan. Mereka mengapresiasi fasilitas yang bersih, nyaman, dan memadai untuk kegiatan belajar.



Gambar 20. *Word cloud* Sentimen Positif Fasilitas

#### a. Interpretasi Sentimen Negatif

Di sisi lain, *Word cloud* negatif menyoroti masalah-masalah yang sangat spesifik dan teknis. Kata “fasilitas”, “ruang”, dan “latih” juga besar di sini,

namun diikuti oleh keluhan seperti “panas”, “desak”, “kurang”, dan “sampah”. Kata “panas” jelas merujuk pada kinerja pendingin ruangan yang tidak optimal.



Gambar 21. *Word cloud* Sentimen Negatif Fasilitas

Untuk menunjukkan peningkatan kemampuan model dalam mengidentifikasi sentimen kritis, Tabel 46 menyajikan perbandingan klasifikasi pada beberapa ulasan representatif.

Tabel 46. Perbandingan Klasifikasi Sentimen Ulasan Fasilitas

Teks	Labelin Lexicon-based	Prediksi SVM Kernel RBF	Prediksi SVM + PSO
puas lengkap fasilitas berterimakasih bimbing ya latih bawah pulng bekal ilmu terimakasih	Positif	Positif	Positif
mohon tingkat	Positif	Positif	Positif
fasilitas ruang	Negatif	Negatif	Negatif
....	....	....	....
alhamdulillah ruang bagus pakai ruang kelas	Negatif	Negatif	Negatif

<b>Teks</b>	<b>Labelin</b>	<b>Prediksi SVM</b>	<b>Prediksi</b>
	<b>Lexicon-based</b>	<b>Kernel RBF</b>	<b>SVM + PSO</b>
oke	Positif	Positif	Positif

## 2. Analisis Sentimen Terhadap Ulasan Makanan

Persepsi peserta terhadap makanan selama pelatihan juga bervariasi, dengan pujian pada rasa dan kritik pada variasi menu.

a. Interpretasi Sentimen Positif

*Word cloud* positif didominasi oleh kata “makan”, “nasi”, “enak”, “saji”, dan “muas”. Ini jelas bahwa dari segi rasa dan porsi, makanan yang disajikan sudah memenuhi harapan banyak peserta. Kata “terimakasih”, “bagus”, “sehat” dan “gizi” semakin memperkuat kesan positif.



Gambar 22. *Word cloud* Sentimen Positif Makanan

#### b. Interpretasi Sentimen Negatif

Kritik utama terlihat pada *Word cloud* negatif. Kata “kritik”, “saran”, “kurang”, “menu”, dan “beda” menunjukkan bahwa keluhan utama peserta adalah variasi makanan yang monoton atau “itu-itu saja”. Selain itu, munculnya

kata “basi” merupakan kritik serius terhadap kualitas penyajian yang perlu menjadi perhatian utama penyelenggara.



Gambar 23. Word cloud Sentimen Negatif Makanan

Untuk menunjukkan peningkatan kemampuan model dalam mengidentifikasi sentimen kritis, Tabel 47 menyajikan perbandingan klasifikasi pada beberapa ulasan representatif.

Tabel 47. Perbandingan Klasifikasi Sentimen Ulasan Makanan

Teks	Labelin Lexicon-based	Prediksi SVM Kernel RBF	Prediksi SVM + PSO
kualitas makan	Positif	Positif	Positif
milik kritik apa makan tiap hari berbedabeda makan enak	Negatif	Netral	Netral
makan enak gizi	Netral	Netral	Netral
....	....	....	....
muas	Positif	Positif	Positif
cukup	Negatif	Netral	Negatif

### 3. Analisis Sentimen Terhadap Ulasan Pengajar

Aspek pengajar secara umum menerima ulasan yang sangat positif, menunjukkan bahwa kualitas instruktur adalah kekuatan utama dari program ini.

#### a. Interpretasi Sentimen Positif

Kata-kata seperti “materi”, “ajar”, “paham”, “mudah”, “jelas”, dan “ramah” mendominasi *Word cloud* positif. Ini berarti peserta merasa bahwa pengajar mampu menyampaikan materi dengan cara yang efektif dan mudah dimengerti



Gambar 24. *Word cloud* Sentimen Positif Pengajar

#### b. Interpretasi Sentimen Negatif

Kritik yang muncul pada *Word cloud* negatif lebih bersifat masukan untuk perbaikan, bukan keluhan terhadap kemampuan pengajar. Kata “cepat”, “terburu-buru”, dan “kurang” menunjukkan bahwa beberapa peserta merasa kesulitan mengikuti ritme pengajaran.



Gambar 25. *Word cloud* Sentimen Negatif Pengajar

Untuk menunjukkan peningkatan kemampuan model dalam mengidentifikasi sentimen kritis, Tabel 48 menyajikan perbandingan klasifikasi pada beberapa ulasan representatif.

Tabel 48. Perbandingan Klasifikasi Sentimen Ulasan Pengajar

Teks	Labelin	Prediksi SVM	Prediksi
	Lexicon-based	Kernel RBF	SVM + PSO
mudah erti	Netral	Netral	Netral
ajar ahli materi	Netral	Netral	Netral
ajar alam bantu dalam materi	Positif	Positif	Positif
.....	.....	.....	.....
terimakasih pakk	Positif	Positif	Positif
kak herman asesor ajar materi	Netral	Netral	Netral

#### 4. Analisis Sentimen Terhadap Ulasan Pengaji

Aspek Penguji mendapatkan sentimen paling positif di antara semua kategori, menandakan proses uji kompetensi berjalan dengan sangat baik.

a. Interpretasi Sentimen Positif

*Word cloud* positif dipenuhi dengan kata-kata apresiasi seperti “assessor”, “bagus”, “terimakasih”, “sesuai”, dan “profesional”. Ini menunjukkan bahwa peserta memandang para asesor atau penguji sebagai figur yang ramah, adil, dan kompeten. Kata “mudah” dan “sesuai” juga menandakan bahwa instruksi selama ujian tersampaikan dengan baik, sehingga mengurangi kebingungan peserta.



Gambar 26. *Word cloud* Sentimen Positif Pengujian

#### b. Interpretasi Sentimen Negatif

Jumlah kritik untuk asesor sangat sedikit. Kata-kata seperti “kritik” dan “saran” memang muncul, namun tidak diikuti oleh keluhan yang spesifik dan berulang. Munculnya kata “sebal” lebih menggambarkan perasaan subjektif peserta saat diuji daripada kritik langsung terhadap perilaku pengaji.



Gambar 27. *Word cloud* Sentimen Negatif Penguij

Untuk menunjukkan peningkatan kemampuan model dalam mengidentifikasi sentimen kritis, Tabel 49 menyajikan perbandingan klasifikasi pada beberapa ulasan representatif.

Tabel 49. Perbandingan Klasifikasi Sentimen Ulasan Penguij

Teks	Labelin	Prediksi SVM		Prediksi SVM + PSO
		Lexicon-based	Kernel RBF	
assesor tenang hadap serta tekan serta latih lambat	Positif	Positif	Positif	Positif
uji bagus banget	Positif	Positif	Positif	Positif
semangat	Netral	Netral	Netral	Netral
.....	.....	.....	.....	.....
assesor ya ramah	Positif	Positif	Positif	Positif
moga depan lebiih	Positif	Positif	Positif	Positif

Secara keseluruhan, interpretasi hasil analisis sentimen yang didukung oleh visualisasi *Word cloud* dan tabel perbandingan memberikan bukti kuat mengenai

area yang telah unggul seperti aspek Pengajar dan Penguji serta area yang memerlukan perhatian dan perbaikan strategis seperti pada aspek Fasilitas dan Makanan.



## **BAB V**

### **KESIMPULAN DAN SARAN**

#### A. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan pada bab-bab sebelumnya, maka dapat ditarik beberapa kesimpulan sebagai jawaban atas rumusan masalah penelitian:

1. Implementasi metode *hybrid lexicon-based* dan SVM untuk klasifikasi sentimen terhadap umpan balik peserta pelatihan VSGA BBPSDMP Kominfo Makassar telah berhasil dilakukan melalui serangkaian tahapan yang sistematis. Proses ini diawali dengan pengumpulan dan pra-pemrosesan data teks yang meliputi *cleaning*, *case folding*, normalisasi kata tidak baku, *tokenisasi*, dan *stemming* untuk menghasilkan data yang bersih dan terstruktur. Selanjutnya, pendekatan *lexicon-based* menggunakan kamus *InSet* digunakan untuk pelabelan sentimen awal. Data yang telah dilabeli kemudian diubah menjadi representasi numerik menggunakan pembobotan fitur TF-IDF dan digunakan untuk melatih model klasifikasi SVM dengan kernel RBF. Model ini mampu mengklasifikasikan sentimen pada empat aspek utama pelatihan, yaitu fasilitas, makanan, pengajar, dan penguji.
2. Terdapat perbedaan kinerja yang signifikan antara model SVM standar dengan model SVM yang dioptimasi menggunakan *Particle Swarm Optimization*. Penerapan optimasi PSO secara konsisten berhasil meningkatkan performa klasifikasi di seluruh metrik evaluasi pada keempat kategori aspek. Peningkatan akurasi tertinggi tercatat pada kategori Pengajar yaitu naik dari 84.70% menjadi 89.01% dan Penguji naik dari 87.43% menjadi 91.45%. Keunggulan utama model SVM dengan penggunaan optimasi PSO terletak pada kemampuannya yang lebih andal dalam mendeteksi kelas minoritas, khususnya sentimen negatif. Hal ini terbukti dari peningkatan signifikan pada nilai recall untuk kelas negatif, yang menunjukkan bahwa model hasil optimasi lebih sensitif dan

efektif dalam mengidentifikasi kritik dari peserta, yang merupakan informasi krusial untuk perbaikan program.

## B. Saran

Berdasarkan hasil penelitian yang diperoleh serta keterbatasan yang ada, berikut adalah beberapa saran yang dapat dipertimbangkan:

### 1. Terhadap Instansi BBPSDMP Kominfo Makassar

Model analisis sentimen yang telah dikembangkan dapat diadopsi sebagai alat bantu untuk mengotomatisasi proses evaluasi umpan balik pelatihan. Ini akan memungkinkan instansi untuk mengolah data kualitatif secara lebih cepat, objektif, dan terukur, sehingga dapat merespons kebutuhan peserta secara lebih efektif.

Berdasarkan interpretasi hasil sentimen, disarankan agar instansi memprioritaskan perbaikan pada aspek fasilitas, terutama terkait konektivitas internet dan kenyamanan pendingin ruangan (AC). Untuk aspek makanan, disarankan untuk meningkatkan variasi menu guna mengurangi monotonitas. Sementara itu, kinerja yang sangat positif pada aspek pengajar dan penguji perlu dipertahankan sebagai keunggulan utama program.

### 2. Terhadap Penelitian Selanjutnya

Terhadap penelitian selanjutnya, terdapat beberapa jalur pengembangan yang dapat dieksplorasi untuk memperdalam hasil analisis. Salah satunya adalah penerapan *Aspect-Based Sentiment Analysis* (ABSA) yang memungkinkan identifikasi sentimen pada entitas yang lebih spesifik dalam satu ulasan, tidak hanya di tingkat kalimat. Selain itu, mengingat tingginya volume data berlabel netral yang seringkali mengandung saran implisit, penelitian mendatang dapat memanfaatkan teknik seperti topic modeling untuk mengekstrak wawasan tersembunyi dari kategori ini. Dari sisi metodologi, untuk meningkatkan akurasi dan kemampuan pemahaman konteks linguistik yang lebih kompleks, disarankan untuk menguji implementasi model *deep*

*learning* seperti *Long Short-Term Memory* (LSTM) atau model berbasis Transformer yang dirancang khusus untuk Bahasa Indonesia, seperti IndoBERT, guna menghasilkan analisis yang lebih presisi dan kontekstual.



## DAFTAR PUSTAKA

- Agustin, Y. H., Kurniadi, D., Julianto, I. T., Jr, B. B. B., & City, L. (2025). *Sentiment Analysis Using Grok AI as an Auto-Labeling Tool in The Text Processing*. 9(2), 700–708.
- Alfauzan, M. F., Sibaroni, Y., & Fitriyani, F. (2023). Sentiment Classification of Fuel Price Rise in Economic Aspects Using Lexicon and SVM Method. *Sinkron*, 8(4), 2526–2536. <https://doi.org/10.33395/sinkron.v8i4.12851>
- Al-Khowarizmi, Sari, I. P., & Maulana, H. (2023). Detecting Cyberbullying on Social Media Using Support Vector Machine: A Case Study on Twitter. *International Journal of Safety and Security Engineering*, 13(4), 709–714. <https://doi.org/10.18280/ijssse.130413>
- Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 6(2), 129. <https://doi.org/10.30998/string.v6i2.10133>
- Damanik, F. J., & Setyohadi, D. B. (2021). Analysis of public sentiment about COVID-19 in Indonesia on Twitter using multinomial naive bayes and support vector machine. *IOP Conference Series: Earth and Environmental Science*, 704(1). <https://doi.org/10.1088/1755-1315/704/1/012027>
- Darmawan, R., Indra, I., & Surahmat, A. (2022). Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru di Twitter. *Jurnal Kajian Ilmiah*, 22(2), 143–152. <https://doi.org/10.31599/jki.v22i2.1130>
- Ependi, U., & Ahmad, N. A. (2024). A Novel Hybrid Classification on Urban Opinion Using ROS-RF: A Machine Learning Approach. *Jurnal Penelitian Pendidikan IPA*, 10(8), 5816–5824. <https://doi.org/10.29303/jppipa.v10i8.8042>
- Fadlil, A., Riadi, I., & Andrianto, F. (2024). Improving Sentiment Analysis in Digital Marketplaces through SVM Kernel Fine-Tuning. *International Journal of Computing and Digital Systems*, 16(1), 159–171. <https://doi.org/10.12785/ijcds/160113>
- Fathirachman Mahing, N., Lazuardi Gunawan, A., Foresta Azhar Zen, A., Abdurrachman Bachtiar, F., & Agung Wicaksono, S. (2023). Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector

- Machine (SVM) dan Random Forest. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(7), 1527–1536. <https://doi.org/10.25126/jtiik.1078010>
- Firda, H., Putra, P., Oktadini, N. R., Sevtiyuni, P. E., Sevtiyuni, P. E., Studi, P., Informasi, S., Komputer, F. I., Sriwijaya, U., Ilir, K. O., & Selatan, S. (2025). *Perbandingan Pelabelan Rating - based dan Inset Lexicon - based dalam Analisis Sentimen Menggunakan SVM ( Studi Kasus : Ulasan Aplikasi GoBiz di Google Play Store ) Comparison of Rating - based and Inset Lexicon - based Labeling in Sentiment Analysis usin.* 14, 516–528.
- Gede, I. P., Suputra, H., Sukadarmika, I. G., Sastra, N. P., Matematika, F., Alam, P., & Udayana, U. (2025). *KLASIFIKASI JUDUL BERITA BAHASA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE DAN SELEKSI FITUR MUTUAL INFORMATION.* 22(1), 69–79.
- Hadju, S. F. N., & Jayadi, R. (2021). Sentiment analysis of Indonesian e-commerce product reviews using support vector machine based term frequency inverse document frequency. *Journal of Theoretical and Applied Information Technology*, 99(17), 4316–4325.
- Hendrawati, T., Ginantra, N. L. W. S. R., & Saiman, C. M. (2024). Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine dan Lexicon Based. *Tematik*, 11(1), 56–64. <https://doi.org/10.38204/tematik.v11i1.1890>
- Hokijuliandy, E., Napitupulu, H., & Firdaniza. (2023). Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application. *Mathematics*, 11(17). <https://doi.org/10.3390/math11173765>
- I Kadek Arya Budi Artana, Gede Aditra Pradnyana, & I Gede Mahendra Darmawiguna. (2023). Analisis Sentimen Twitter Untuk Menilai Kesiapan Pembelajaran Tatap Muka Terbatas Dengan Inset Lexicon Dan Levenshtein Distance. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 20(2), 200–209. <https://doi.org/10.23887/jptkundiksha.v20i2.64579>
- Kominfo, B. (2021). Peluncuran Program Literasi Digital Nasional. In *Badan Pengembangan Sumber Daya Manusia Kominfo*. <https://bpsdm.komdigi.go.id/berita-peluncuran-program-literasi-digital-nasional-44-777>
- Kristiyanti, D. A., & Sri Hardani. (2023). Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector

- Machine, and Long Short-Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(3), 722–732. <https://doi.org/10.29207/resti.v7i3.4737>
- Madyatmadja, E. D., Candra, H., Nathaniel, J., Jonathan, M. R., & Rudy. (2024). Sentiment Analysis on User Reviews of Threads Applications in Indonesia. *Journal Europeen Des Systemes Automatises*, 57(4), 1165–1171. <https://doi.org/10.18280/jesa.570423>
- Muhammad, R. H., Laksana, T. G., & Arifa, A. B. (2022). Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 8(1), 59–71. <https://doi.org/10.23917/khif.v8i1.15213>
- Muttakin, F., & Andrika, N. (2025). *Sentiment Analysis of Shoe Product Reviews on Indonesian E-Commerce Platform Using Lexicon Based and Support Vector Machine*. 6(2), 839–854.
- Nauli, S., Berutu, S. S., Budiati, H., & Maedjaja, F. (2025). KLASIFIKASI KALIMAT PERUNDUNGAN PADA TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(1), 107–122. <https://doi.org/10.29100/jipi.v10i1.5749>
- Nurmalasari, D., Hidayatul, D., Authors, Q., Chairani, N., & Yuliantoro, H. R. (2024). *Discovering User Sentiment Patterns in Libraries with a Hybrid Machine Learning and Lexicon-Based Approach*. 13, 311–317.
- Oktaviana, N. E., Sari, Y. A., & Indriati, I. (2022). Analisis Sentimen terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features dan Support Vector Machine. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(2), 357–362. <https://doi.org/10.25126/jtiik.2022925625>
- Permana, B. R. S., & Panudju, A. T. (2020). Comparison and Performance Analysis of SVM and PSO-SVM Algorithms (Case Study Classification of Senior High School). *Researchgate.Net*, V(Viii), 81–88. [https://www.researchgate.net/profile/Andreas-Tri-Panudju/publication/370866625\\_Comparison\\_and\\_Performance\\_Analysis\\_of\\_SVM\\_and\\_PSO-SVM\\_Algorithms\\_Case\\_Study\\_Classification\\_of\\_Senior\\_High\\_School/links/6466ce699533894cac7c64a9/Comparison-and-Performance-Anal](https://www.researchgate.net/profile/Andreas-Tri-Panudju/publication/370866625_Comparison_and_Performance_Analysis_of_SVM_and_PSO-SVM_Algorithms_Case_Study_Classification_of_Senior_High_School/links/6466ce699533894cac7c64a9/Comparison-and-Performance-Anal)

- Priadana, A., & Rizal, A. A. (2021). Sentiment Analysis on Government Performance in Tourism During The COVID-19 Pandemic Period With Lexicon Based. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, 7(1), 28–39. <https://doi.org/10.18860/ca.v7i1.12488>
- Risawati, R., Ernawati, S., & Maryani, I. (2020). OPTIMASI PARAMETER PSO BERBASIS SVM UNTUK ANALISIS SENTIMEN REVIEW JASA MASKAPAI PENERBANGAN BERBAHASA INGGRIS. *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(2), 64–71. <https://doi.org/10.31294/evolusi.v8i2.9248>
- Rizky Pratama, M., Ramadhan, Y. R., & Komara, M. A. (2023). Analisis sentimen BRImo dan BCA Mobile menggunakan Support Vector Machine dan Lexicon Based. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 12(3), 1439–1450.
- Savira, R., Solichin, A., & Syafrullah, M. (2023). Analisis Sentimen Pada Twitter Terhadap Kenaikan Bbm 2022 Dengan Lexicon Dan Support Vector Machine. *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2(1), 211–218. <https://senafti.budiluhur.ac.id/index.php/senafti/article/view/564>
- Syahputra, H. (2021). Sentiment Analysis of Community Opinion on Online Store in Indonesia on Twitter using Support Vector Machine Algorithm (SVM). *Journal of Physics: Conference Series*, 1819(1). <https://doi.org/10.1088/1742-6596/1819/1/012030>
- Taqiuddin, R., Bachtiar, F. A., & Purnomo, W. (2021). Opinion Spam Classification on Steam Review using Support Vector Machine with Lexicon-Based Features. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4. <https://doi.org/10.22219/kinetik.v6i4.1323>
- Tuerxun, W., Chang, X., Hongyu, G., Zhijie, J., & Huajian, Z. (2021). Fault Diagnosis of Wind Turbines Based on a Support Vector Machine Optimized by the Sparrow Search Algorithm. *IEEE Access*, 9, 69307–69315. <https://doi.org/10.1109/ACCESS.2021.3075547>
- Ulya, D. I., Kunaefi, A., Rolliawati, D., & Nugroho, B. A. (2023). Unpacking Public Perceptions of Qris with Twitter Data: A Vader And LDA Methodology. *Jurnal ELTIKOM*, 7(2), 145–159. <https://doi.org/10.31961/eltikom.v7i2.742>

## LAMPIRAN

### Lampiran 1. Dataset

No.	Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan	Kritik dan Saran terkait Makanan	Kritik dan Saran terkait Pengajar	Kritik dan Saran terkait Assessor (Penguji)
1	-	-	-	-
2	Bagus	Mantap	Sangat baik	Sangat sangat baik
3	acnya kurang dingin	.	.	.
4	Semoga pelatihnya di perluas lagi lingkupnya di Kota Parepare	Baik	Pengalaman yang luar biasa	Sangat efesien
5	lumayan dinginn dan lumayan bersih	ENAKK	untuk bapak pengajarnyaa Semangatt ki bapak	kepadaa ibu/ bapak asesor terimaaksihh ibu bapak
6	sudah bagus	bagus	bagus	bagus
7	Baik	Enak	Baik	Baik
8	tidak ada kritik dan saran	tidak ada	tidak ada	tidak ada
9	Tidak ada	Tidak ada	Tidak ada	Tidak ada
10	-	-	-	-
11	Pencahayaan ruangan kurang	.	Keren	baik
12	Tidak ada	tidak ada	tidak ada	tidak ada

13	thank's for all the knowledge	lebih di permantap lagi ajaaa	nyamann	lebih permantap lagi ya abangku
14	Bagus	Enak	Tdk ada	Tdk ada
.....	.....	.....	.....	.....
76	Terlalu dingin	Makanan ada yang terlalu asin	Tidak ada kritik	Tidak ada kritik
77	Mantap sekali mi	Enak sekali mi	Keren sekali mi	Mantap
78	Ruangan nyaman	Makanan yang enak enak	Sangat puas dengan materi yang disampaikan	Sangat baik
79	Semoga pelatiannya bisaaa terus terlaksana dengan baik	Sangat baik	Luar biasa	Semoga tambah baik kedepanya
80	#NAME?	Semua baik, enak dan mantap	Metode mengajar yang sangat baik, seru dan mudah dipahami	Semua baik
81	Fasilitas hingga kenyamanan tempat/ruang pelatihan sudah sangat baik	Enak	Pengajar sudah sangat baik untuk memberi pemahaman terhadap materi yg dibawakan	Assesor telah memberikan pengujian dengan cara yang baik dan memberi peserta ketersediaan waktu yang memadai

76	Terlalu dingin	Makanan ada yang terlalu asin	Tidak ada kritik	Tidak ada kritik
....	....	....	....	....
2308	Semoga kedepannya ada kegiatan yang bermanfaat lagi dari vgsa untuk memotivasiikan untuk semangat dan bersaing dalam dunia kerja.	Sangat komplit	Penilainy tenaga pengajar sangat memuaskan Sudah lebih dari cukup untuk kami mengikuti kegiatan saya hanya bilang terima Untuk vgsa Kominfo yg sudah buat kegiatan untuk kami lulusan SMK	Untuk asesor sangat support dan memberi saran untuk kita dalam mengerjakan ujian kompetensi dengan baik ..
2309	Karna kebanyakan pelatihan digitalent sifatnya berhubungan dengan internet, alangkah lebih baik diberi akses internet tambahan yg memadai untuk peserta sebagai cadangan agar bisa mengikuti intruksi dan kegiatan dengan lebih maksimal.	Mantap no complain	Mungkin waktu pelatihan bisa diperpanjang agar para peserta yang pemula lebih bisa mengikuti dan tidak kebingungan. dan pengajar diberi pendamping agar bisa membantu dalam hal materi,karna rata" kelas digitalent sifatnya praktek, sehingga lebih bisa menakomodir dan	.

			mendampingi para peserta	
2310	fasilitas lengkap dan bagus kebersihan bagus	memuaskan	penjelasan mudah dipahami materinya lengkap sehingga	memuaskan
2311	Bagus	Enak	pengajarnya sangat baik	Pengujinya sangat baik
2312	Lebih mempersiapkan instalasi listrik dengan baik pemakaian UPS sangat diperlukan apabila terjadi pemadaman listrik secara tiba-tiba	Manurut saya rasa dan lauknya beefariasi tetapi untuk sendoknya agak susah untuk membelah lauknya mohon untuk sendo sedikit kuat	Semoga para pengajar semakin sukses dan dapat memberikan materi dgn lebih baik lagi.	Saran saya kepada pengaji(Assesor) agar membalas senyum dan membalas sapaan itu bentuk rasa hormat dan rasa menghargai orang yg lebih tua.
2313	Semua fasilitas sangat memuaskan dan saya merasa nyaman saat pelatihan dan menteri yg diberikan juga dapat di mengerti	Tidak ada	Tidak ada	Tidak ada

## Lampiran 2. Source Code

### PEMROSESAN DATA

```
import pandas as pd

data = pd.read_csv("Data Survey Evaluasi1.csv", on_bad_lines='warn')

dataFasilitas = pd.DataFrame(data[['Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan']])

dataMakanan = pd.DataFrame(data[['Kritik dan Saran terkait Makanan']])

dataPengajar = pd.DataFrame(data[['Kritik dan Saran terkait Pengajar']])

dataPenguji = pd.DataFrame(data[['Kritik dan Saran terkait Assessor (Penguji)']])

print(dataFasilitas.head(3),)

print(dataMakanan.head(3),)

print(dataPengajar.head(3),)

print(dataPenguji.head(3))

dataFasilitas.drop_duplicates(subset=["Kritik dan Saran terkait Ruangan dan Fasilitas Pelatihan"], keep='first', inplace=True)

dataMakanan.drop_duplicates(subset=["Kritik dan Saran terkait Makanan"], keep='first', inplace=True)

dataPengajar.drop_duplicates(subset=["Kritik dan Saran terkait Pengajar"], keep='first', inplace=True)

dataPenguji.drop_duplicates(subset=["Kritik dan Saran terkait Assessor (Penguji)"], keep='first', inplace=True)

#### CLEANING

import re

import string

import nltk

# Fungsi menghapus emoji

def remove_emoji(text):

    if text is not None and isinstance(text,str):

        emoji_pattern = re.compile("["

                                    u"\U0001F600-\U0001F64F" # emoticons

                                    u"\U0001F300-\U0001F5FF" # symbols &

pictographs
```

```

u"\U0001F680-\U0001F6FF" # transport &
map symbols

u"\U0001F700-\U0001F77F" # alchemical
symbols

u"\U0001F780-\U0001F7FF" # Geometric
Shapes Extended

u"\U0001F800-\U0001F8FF" # Supplemental
Arrows-C

u"\U0001F900-\U0001F9FF" # Supplemental
Symbols and Pictographs

u"\U0001FA00-\U0001FA6F" # Chess Symbols
u"\U0001FA70-\U0001FAFF" # Symbols and
Pictographs Extended-A

u"\U0001F004-\U0001F0CF" # Playing Cards
u"\U0001F1E0-\U0001F1FF" # Flags

(iOS)

    return emoji_pattern.sub(r'', text)
else:
    return text

# Fungsi untuk menghapus simbol
def remove_symbols(text):
    if text is not None and isinstance(text,str):
        text = re.sub(r'[^a-zA-Z\s]', '', text)
    return text

# Fungsi untuk menghapus angka
def remove_numbers(text):
    if text is not None and isinstance(text,str):
        text = re.sub(r'\d+', '', text)
    return text

dataFasilitas['cleaning'] = dataFasilitas['Kritik dan Saran terkait Ruangan
dan Fasilitas Pelatihan'].apply(lambda x: remove_emoji(x))

dataFasilitas['cleaning'] = dataFasilitas['cleaning'].apply(lambda x:
remove_symbols(x))

```



```

dataFasilitas['cleaning'] = dataFasilitas['cleaning'].apply(lambda x:
remove_numbers(x))

##### CASE FOLDING

def case_folding(text):

    if isinstance(text, str):

        lowercase_text = text.lower()

        return lowercase_text

    else:

        return text

dataFasilitas['case_folding'] =
dataFasilitas['cleaning'].apply(case_folding)

dataMakanan['case_folding'] = dataMakanan['cleaning'].apply(case_folding)

dataPengajar['case_folding'] = dataPengajar['cleaning'].apply(case_folding)

dataPenguji['case_folding'] = dataPenguji['cleaning'].apply(case_folding)

##### NORMALISASI KATA

import pandas as pd

# Fungsi penggantian kata tidak baku

def replace_taboo_words(text, kamus_tidak_baku):

    if isinstance(text, str):

        words = text.split()

        replace_words = []

        kalimat_baku = []

        kata_diganti = []

        kata_tidak_baku_hash = []

        for word in words:

            if word in kamus_tidak_baku:

                baku_word = kamus_tidak_baku[word]

                if isinstance(baku_word, str) and all(char.isalpha() for
char in baku_word):

                    replace_words.append(baku_word)

                    kalimat_baku.append(baku_word)

                    kata_diganti.append(word)

```

```

        kata_tidak_baku_hash.append(hash(word))

    else:
        replace_words.append(word)

    replace_text = ' '.join(replace_words)

else:
    replace_text = ''
    kalimat_baku = []
    kata_diganti = []
    kata_tidak_baku_hash = []

return replace_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash

kamus_data = pd.read_excel('kamuskatabaku.xlsx')

kamus_tidak_baku = dict(zip(kamus_data['tidak_baku'],
kamus_data['kata_baku']))

kamus_data.head()

# kamus_data.info()

# Penerapan fungsi penggantian kata tidak baku

dataFasilitas['hasil_normalisasi'], dataFasilitas['kata_baku'],
dataFasilitas['kata_tidak_baku'], dataFasilitas['kata_tidak_baku_hash'] =
zip(*dataFasilitas['case_folding'].apply(lambda x: replace_taboo_words(x,
kamus_tidak_baku)))

dataFasilitas = pd.DataFrame(dataFasilitas[['Kritik dan Saran terkait
Ruangan dan Fasilitas
Pelatihan','cleaning','case_folding','hasil_normalisasi']])

#### TOKENIZATION

def tokenize(text):
    tokens = text.split()
    return tokens

dataFasilitas['tokenize'] =
dataFasilitas['hasil_normalisasi'].apply(tokenize)

dataMakanan['tokenize'] = dataMakanan['hasil_normalisasi'].apply(tokenize)

dataPengajar['tokenize'] = dataPengajar['hasil_normalisasi'].apply(tokenize)

dataPenguji['tokenize'] = dataPenguji['hasil_normalisasi'].apply(tokenize)

#### STOPWORD REMOVAL

from nltk.corpus import stopwords

```

```

nltk.download('stopwords')

stop_words = set(stopwords.words('indonesian'))

def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

dataFasilitas['stopword_removal'] = dataFasilitas['tokenize'].apply(lambda
x:remove_stopwords(x))

dataMakanan['stopword_removal'] = dataMakanan['tokenize'].apply(lambda
x:remove_stopwords(x))

dataPengajar['stopword_removal'] = dataPengajar['tokenize'].apply(lambda
x:remove_stopwords(x))

dataPenguji['stopword_removal'] = dataPenguji['tokenize'].apply(lambda
x:remove_stopwords(x))

#### STEAMING DATA

import pandas as pd

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

from nltk.stem import PorterStemmer

from nltk.stem.snowball import SnowballStemmer

factory = StemmerFactory()

stemmer = factory.create_stemmer()

def stem_text(text):
    return [stemmer.stem(word) for word in text]

dataFasilitas['stemming_data'] =
dataFasilitas['stopword_removal'].apply(lambda x: ' '.join(stem_text(x)))

dataMakanan['stemming_data'] = dataMakanan['stopword_removal'].apply(lambda
x: ' '.join(stem_text(x)))

dataPengajar['stemming_data'] = dataPengajar['stopword_removal'].apply(lambda
x: ' '.join(stem_text(x)))

dataPenguji['stemming_data'] = dataPenguji['stopword_removal'].apply(lambda
x: ' '.join(stem_text(x)))

HAPUS DATA BERNILAI KOSONG

dataFasilitas = dataFasilitas.dropna()

dataMakanan = dataMakanan.dropna()

dataPengajar = dataPengajar.dropna()

dataPenguji = dataPenguji.dropna()

dataPenguji.info()

```

```

dataFasilitas.to_csv('./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Fasilitas.csv', encoding='utf8', index=False)

dataMakanan.to_csv('./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Makanan.csv', encoding='utf8', index=False)

dataPengajar.to_csv('./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Pengajar.csv', encoding='utf8', index=False)

dataPenguji.to_csv('./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Penguji.csv', encoding='utf8', index=False)

LABELING DATA

import pandas as pd

dataFasilitas =
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Fasilitas.csv")

dataMakanan =
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Makanan.csv")

dataPengajar =
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Pengajar.csv")

dataPenguji =
pd.read_csv("./Hasil_Pemrosesan_Data/Hasil_Preprocessing_Penguji.csv")

dataFasilitas = pd.DataFrame(dataFasilitas[['steming_data']])
dataMakanan = pd.DataFrame(dataMakanan[['steming_data']])
dataPengajar = pd.DataFrame(dataPengajar[['steming_data']])
dataPenguji = pd.DataFrame(dataPenguji[['steming_data']])

import pandas as pd
import requests
# Unduh kamus lexicon positif dan negatif di github

url_negatif =
'https://raw.githubusercontent.com/fajri91/InSet/master/negative.tsv'

url_positif =
'https://raw.githubusercontent.com/fajri91/InSet/master/positive.tsv'

positif_lexicon = set(pd.read_csv(url_positif, sep='\t', header=None)[0])
negatif_lexicon = set(pd.read_csv(url_negatif, sep='\t', header=None)[0])

# Fungsi untuk menentukan sentimen dan menghitung skor sentimen

def hitung_sentimen(teks):
    if isinstance(teks, str):
        teks = teks.lower()

```

```

        positif_count = sum(1 for word in teks.split() if word in
positif_lexicon)

        negatif_count = sum(1 for word in teks.split() if word in
negatif_lexicon)

        sentimen_skor = positif_count - negatif_count

        if sentimen_skor > 0:
            sentimen = "Positif"
        elif sentimen_skor < 0:
            sentimen = "Negatif"
        else:
            sentimen = "Netral"

        return sentimen_skor, sentimen

    return 0, "Netral"

# Menentukan sentimen dan skor sentimen untuk setiap teks

dataFasilitas[['Score','Sentimen']] = dataFasilitas['steming_data'].apply(lambda x: pd.Series(hitung_sentimen(x)))
dataMakanan[['Score','Sentimen']] = dataMakanan['steming_data'].apply(lambda x: pd.Series(hitung_sentimen(x)))
dataPengajar[['Score','Sentimen']] = dataPengajar['steming_data'].apply(lambda x: pd.Series(hitung_sentimen(x)))
dataPenguji[['Score','Sentimen']] = dataPenguji['steming_data'].apply(lambda x: pd.Series(hitung_sentimen(x)))

dataPenguji.head(10)

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("whitegrid")

# Tambahkan kolom 'Kategori' pada masing-masing DataFrame

dataFasilitas['Kategori'] = 'Fasilitas'
dataMakanan['Kategori'] = 'Makanan'
dataPengajar['Kategori'] = 'Pengajar'
dataPenguji['Kategori'] = 'Penguji'

```

```

# Gabungkan semua data

dataGabungan = pd.concat([dataFasilitas, dataMakanan, dataPengajar,
dataPengaji], ignore_index=True)

# Hitung jumlah total per kategori untuk hitung persentase

total_per_kategori =
dataGabungan.groupby('Kategori')['Sentimen'].count().to_dict()

# Hitung jumlah sentimen per kategori

sentimen_per_kategori = dataGabungan.groupby(['Kategori',
'Sentimen']).size().reset_index(name='Jumlah')

# Hitung persentase

sentimen_per_kategori['Persentase'] = sentimen_per_kategori.apply(
    lambda row: f'{100 * row["Jumlah"] / total_per_kategori[row["Kategori"]]:.2f}%', axis=1
)

# Visualisasi

plt.figure(figsize=(13, 6))
ax = sns.barplot(
    data=sentimen_per_kategori,
    x='Kategori',
    y='Jumlah',
    hue='Sentimen',
    palette='pastel'
)

# Tambahkan jumlah dan persentase di atas setiap batang

for i in range(len(sentimen_per_kategori)):

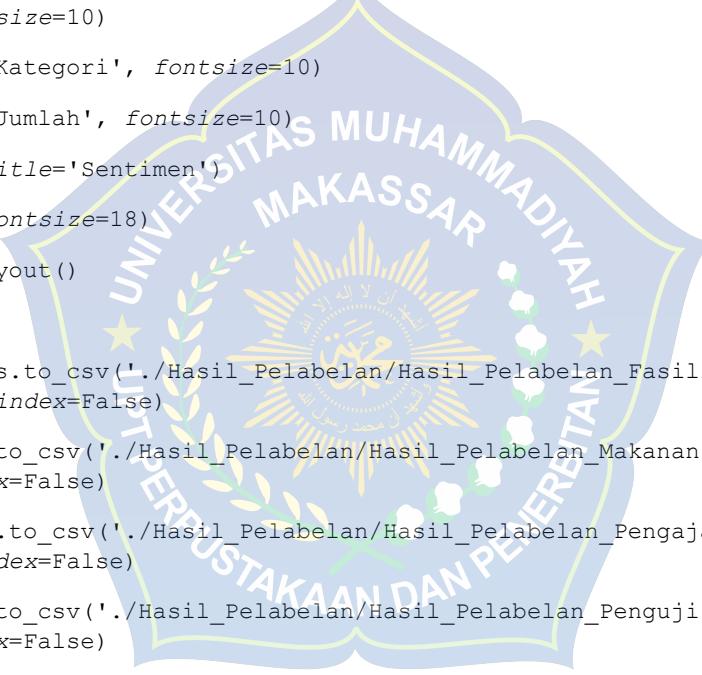
    row = sentimen_per_kategori.iloc[i]

    kategori = row['Kategori']
    sentimen = row['Sentimen']
    jumlah = row['Jumlah']
    persentase = row['Persentase']

    # Cari posisi batang

    x = list(sentimen_per_kategori['Kategori'].unique()).index(kategori)

```



```

hue_offset =
list(sentimen_per_kategori['Sentimen'].unique()).index(sentimen)

bars = [bar for bar in ax.containers[hue_offset]]

if bars and len(bars) > x:

    bar = bars[x]

    height = bar.get_height()

    ax.text(bar.get_x() + bar.get_width() / 2, height + 0.5, f'{jumlah} {persentase}',

            ha='center', va='bottom', fontsize= 10)

# Finalisasi

plt.title('Jumlah Analisis Sentimen per Kategori\nMetode InSet Lexicon-Based', fontsize=10)

plt.xlabel('Kategori', fontsize=10)

plt.ylabel('Jumlah', fontsize=10)

plt.legend(title='Sentimen')

plt.xticks(fontsize=18)

plt.tight_layout()

plt.show()

dataFasilitas.to_csv('../Hasil_Pelabelan/Hasil_Pelabelan_Fasilitas.csv', encoding='utf8', index=False)

dataMakanan.to_csv('../Hasil_Pelabelan/Hasil_Pelabelan_Makanan.csv', encoding='utf8', index=False)

dataPengajar.to_csv('../Hasil_Pelabelan/Hasil_Pelabelan_Pengajar.csv', encoding='utf8', index=False)

dataPenguji.to_csv('../Hasil_Pelabelan/Hasil_Pelabelan_Penguji.csv', encoding='utf8', index=False)

```

## MODEL KLASIFIKASI

```

import pandas as pd


dataFasilitas =
pd.read_csv("../Hasil_Pelabelan/Hasil_Pelabelan_Fasilitas.csv")

dataMakanan = pd.read_csv("../Hasil_Pelabelan/Hasil_Pelabelan_Makanan.csv")

dataPengajar = pd.read_csv("../Hasil_Pelabelan/Hasil_Pelabelan_Pengajar.csv")

dataPenguji = pd.read_csv("../Hasil_Pelabelan/Hasil_Pelabelan_Penguji.csv")

```

### **Split Data**

```
from sklearn.model_selection import train_test_split

# Fasilitas

X_train_fasilitas, X_test_fasilitas, y_train_fasilitas, y_test_fasilitas =
train_test_split(
    dataFasilitas['steming_data'], dataFasilitas['Sentimen'], test_size=0.2,
    random_state=42)

# Simpan data latih ke file

train_set_fasilitas = pd.DataFrame({'text': X_train_fasilitas, 'Sentimen':
y_train_fasilitas})

train_set_fasilitas.to_csv('./DataLatih_DataUji/train_set_fasilitas.csv',
index=False)

# fimpang data uji ke file

test_set_fasilitas = pd.DataFrame({'text': X_test_fasilitas, 'Sentimen':
y_test_fasilitas})

test_set_fasilitas.to_csv('./DataLatih_DataUji/test_set_fasilitas.csv',
index=False)

# menampilkan jumlah data latih dan data uji

print(f"[Fasilitas] Jumlah data latih: {len(X_train_fasilitas)}")
print(f"[Fasilitas] Jumlah data uji: {len(X_test_fasilitas)}")

import matplotlib.pyplot as plt

# Hitung ukuran data latih dan uji untuk masing-masing kategori

train_sizes = [len(X_train_fasilitas), len(X_train_makanan),
len(X_train_pengajar), len(X_train_penguji)]

test_sizes = [len(X_test_fasilitas), len(X_test_makanan),
len(X_test_pengajar), len(X_test_penguji)]

labels = ['Fasilitas', 'Makanan', 'Pengajar', 'Penguji']

x = range(len(labels))

width = 0.35

plt.figure(figsize=(15, 6))

bars1 = plt.bar([i - width/2 for i in x], train_sizes, width, label='Data
Latih', color='blue')
```

```

bars2 = plt.bar([i + width/2 for i in x], test_sizes, width, label='Data
Uji', color='orange')

# Tambahkan label jumlah di atas setiap bar

for bars in [bars1, bars2]:
    for bar in bars:
        height = bar.get_height()

        plt.text(bar.get_x() + bar.get_width()/2, height + 0.5, f'{height}',

ha='center', va='bottom', fontsize=15)

plt.title('Jumlah Data Latih dan Uji per Kategori', fontsize=15)
plt.xlabel('Kategori', fontsize=15)
plt.ylabel('Jumlah Data', fontsize=15)
plt.xticks(ticks=x, labels=labels, fontsize=15)
plt.legend()
plt.tight_layout()
plt.show()

### SUPPORT VECTOR MACHINE

import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.svm import SVC
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

# Mendefenisikan vectorizer

vectorizer_fasilitas = TfidfVectorizer()

X_train_vectorized_fasilitas =
vectorizer_fasilitas.fit_transform(X_train_fasilitas)

X_test_vectorized_fasilitas =
vectorizer_fasilitas.transform(X_test_fasilitas)

# Menampilkan hasil vectorizer

print("Hasil Vectorizer untuk data latih Fasilitas:")

```

```

print(X_train_vectorized_fasilitas.toarray())

# Model SVM dengan kernel RBF

svm_rbf_fasilitas = SVC(kernel='rbf', C=1.0, gamma='scale')

svm_rbf_fasilitas.fit(X_train_vectorized_fasilitas, y_train_fasilitas)

# Prediksi dengan model SVM RBF

y_pred_rbf_fasilitas =
svm_rbf_fasilitas.predict(X_test_vectorized_fasilitas)

# Menampilkan akurasi model

accuracy_svm_fasilitas = accuracy_score(y_test_fasilitas,
y_pred_rbf_fasilitas)

print(f"Akurasi Model SVM untuk Fasilitas: {accuracy_svm_fasilitas:.4f}")

print('SVM Accuracy: {:.2f}%'.format(accuracy_svm_fasilitas * 100))

# Menampilkan classification report

accuracy_svm_fasilitas = accuracy_score(y_test_fasilitas,
y_pred_rbf_fasilitas)

print(f"Akurasi Model SVM untuk Fasilitas: {accuracy_svm_fasilitas:.4f}")

print('SVM Accuracy: {:.2f}%'.format(accuracy_svm_fasilitas * 100))

print("Classification Report - Fasilitas:")

print(classification_report(y_test_fasilitas, y_pred_rbf_fasilitas))

# Menampilkan hasil confusion matrix

print("Hasil Confusion Matrix - Fasilitas:")

print(confusion_matrix(y_test_fasilitas, y_pred_rbf_fasilitas))

# Fungsi untuk menampilkan visualisasi Confusion Matrix

def plot_confusion_matrix(model_name, y_true, y_pred):

    cm = confusion_matrix(y_true, y_pred)

    plt.figure(figsize=(5, 5))

    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
                xticklabels=['Negatif','Netral','Positif'],
                yticklabels=['Negatif','Netral','Positif'])

    plt.title(f'{model_name} Confusion Matrix')

    plt.xlabel('Prediksi')

    plt.ylabel('Aktual')

```

```

plt.show()

# Visualisasi Confusion Matrix untuk setiap kategori

plot_confusion_matrix('SVM - Fasilitas', y_test_fasilitas,
y_pred_rbf_fasilitas)

### SVM DENGAN OPTIMASI PSO

from pyswarm import pso

from sklearn.model_selection import cross_val_score

import numpy as np

# Batas parameter

lb = [0.1, 0.0001] # batas bawah (C, gamma)

ub = [1000, 10] # batas atas

def svm_pso_eval_fasilitas(params):

    C, gamma = params

    model = SVC(kernel='rbf', C=C, gamma=gamma)

    scores = cross_val_score(model, X_train_vectorized_fasilitas,
y_train_fasilitas, cv=5)

    return -scores.mean()

best_params, _ = pso(svm_pso_eval_fasilitas, lb, ub, swarmsize=20,
maxiter=30)

C_opt, gamma_opt = best_params

print(f'[Fasilitas] Parameter optimal: C={C_opt:.4f},\n
gamma={gamma_opt:.4f}')

svm_pso = SVC(kernel='rbf', C=C_opt, gamma=gamma_opt)

svm_pso.fit(X_train_vectorized_fasilitas, y_train_fasilitas)

y_pred_pso_fasilitas = svm_pso.predict(X_test_vectorized_fasilitas)

# Classification report PSO

acc_pso_fasilitas = accuracy_score(y_test_fasilitas, y_pred_pso_fasilitas)

print(f"[Fasilitas] Akurasi SVM RBF + PSO: {acc_pso_fasilitas * 100:.2f}%")

print("[Fasilitas] Classification Report PSO:")

print(classification_report(y_test_fasilitas, y_pred_pso_fasilitas))

# Confusion matrix PSO

plot_confusion_matrix("SVM + PSO - Fasilitas", y_test_fasilitas,
y_pred_pso_fasilitas)

```

### Lampiran 3. Surat Permohonan Penelitian dari Program Studi

**MAJELIS PENDIDIKAN TINGGI PIMPINAN PUSAT MUHAMMADIYAH  
UNIVERSITAS MUHAMMADIYAH MAKASSAR  
FAKULTAS TEKNIK  
PROGRAM STUDI INFORMATIKA**

\_\_\_\_\_  
*بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ*

Nomor : 055/05/IF/C.4-VI/VII/47/2025 Makassar, 20 Muharram 1447 H  
Lamp. : - 15 Juli 2025 M  
Hal : Permohonan Data Penelitian

Kepada yang Terhormat,  
**Ketua LP3M Unismuh Makassar**  
Di –  
Tempat

*Assalamu 'Alaikum Warahmatullahi Wabarakatuh*

Dengan Rahmat Allah SWT, Semoga aktivitas kita bermula ibadah di Sisi-Nya. Dalam rangka penyelesaian Tugas Akhir pada Program Studi Informatika dengan judul “Implementasi Hybrid Lexicon-Based dan SVM untuk Klasifikasi Analisis Sentimen Terhadap Pelatihan BBPSDMP Kominfo Makassar”. Bersama ini kami sampaikan mahasiswa:

Stambuk	Nama
105 84 11036 21	Nur Alam

Sehubungan dengan hal tersebut, maka kami memohon dibuatkan surat pengantar pada instansi di bawah ini:

Nama Instansi : Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian (BBPSDMP) Kominfo Makassar  
Alamat : Jln. Prof. Abdurrahman Basalamah II No. 25 Karampuang, Kec. Panakkukang, Kota Makassar

Demikian surat kami atas perhatian dan kerja samanya kami haturkan banyak terima kasih.  
*Jazakumullah Khairan Katsiran*  
*Wassalamu 'Alaikum warahmatullahi Wabarakatuh*

*Muhyiddin A.M Havat, S.Kom., M.T.*  
NBMT 104/577

Tembusan:  
1. Dekan Fakultas Teknik  
2. Arsip

Gedung Menara Iqra Lantai 3  
Jl. Sultan Alauddin No. 259 Telp. (0411) 866 972 Fax (0411) 865 588 Makassar 90221  
Web: <https://teknik.unismuh.ac.id/>, e-mail: [teknik@unismuh.ac.id](mailto:teknik@unismuh.ac.id)

## Lampiran 4. Surat Permohonan Izin Pelaksanaan Penelitian dari LP3M



UNIVERSITAS MUHAMMADIYAH MAKASSAR  
LEMBAGA PENELITIAN PENGEMBANGAN DAN PENGABDIAN KEPADA MASYARAKAT

Jl. Sultan Alauddin No. 259 Telp. 866972 Fax. (0411) 865588 Makassar 90221 e-mail: lp3m@unismuh.ac.id



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Nomor : 132/LP3M/05/C.4-VIII/VII/1447/2025

Lampiran : 1 (satu) rangkap proposal

Hal : Permohonan Izin Pelaksanaan Penelitian

Kepada Yth:

Bapak Gubernur Provinsi Sulawesi Selatan

Cq. Kepala Dinas Penanaman Modal & PTSP Provinsi Sulawesi Selatan  
di-

Makassar

Assalamu Alaikum Wr. Wb

Berdasarkan surat Dekan Fakultas Teknik Program Studi Informatika, nomor: 055 tanggal: 16 Juli 2025, menerangkan bahwa mahasiswa dengan data sebagai berikut.

Nama : NUR ALAM

Nim : 105841103621

Fakultas : Teknik

Prodi : Informatika

Bermaksud melaksanakan penelitian/pengumpulan data dalam rangka penulisan laporan tugas akhir Skripsi dengan judul :

"IMPLEMENTASI HYBRID LEXICON-BASED DAN SVM UNTUK KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN BBPSDMP KOMINFO MAKASSAR"

Yang akan dilaksanakan dari tanggal 23 Juli 2025 s/d 23 September 2025.

Sehubungan dengan maksud di atas, kiranya Mahasiswa tersebut diberikan izin untuk melakukan penelitian sesuai ketentuan yang berlaku.

Demikian, atas perhatian dan kerjasamanya diucapkan jazakumullah khaeran katziraa.  
Billahi Fii Sabili Haq, Fastabiqul Khaerat.

Wassalamu Alaikum Wr. Wb

20 Muharram 1447

16 Juli 2025

Ketua LP3M Unismuh Makassar,

Dr. Muh. Arief Muhsin, M.Pd.  
NBM. 112 7761



Kampus  
Merdeka  
INDONESIA-JAYA



Jl. Sultan Alauddin No. 259 Telp. (0411) 866972 Fax (0411) 865588 Makassar 90221  
E-mail: lp3m@unismuh.ac.id Official Web: https://lp3m.unismuh.ac.id

## Lampiran 5. Surat Izin Penelitian dari Dinas Penanaman Modal dan PTSP Provinsi Sulawesi Selatan



PEMERINTAH PROVINSI SULAWESI SELATAN  
DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU SATU PINTU  
Jl.Bougenville No.5 Telp. (0411) 441077 Fax. (0411) 448936  
Website : <http://simap-new.sulselprov.go.id> Email : [ptsp@sulselprov.go.id](mailto:ptsp@sulselprov.go.id)  
Makassar 90231

Nomor	: 16038/S.01/PTSP/2025	Kepada Yth.
Lampiran	: -	Kepala Balai Besar Pengembangan Sumber Daya Manusia dan Penelitian (BBPSDMP) Kominfo Makassar
Perihal	: <u>Izin penelitian</u>	

di-  
Tempat

Berdasarkan surat Ketua LP3M UNISMUH Makassar Nomor : 132/LP3M/05/C.4-VIII/VII/1447/2025 tanggal 16 Juli 2025 perihal tersebut diatas, mahasiswa/peneliti dibawah ini:

Nama : NUR ALAM  
Nomor Pokok : 105841103621  
Program Studi : Informatika  
Pekerjaan/Lembaga : Mahasiswa (S1)  
Alamat : Jl. Sultan Alauddin No. 259 Makassar

Bermaksud untuk melakukan penelitian di daerah/kantor saudara dalam rangka menyusun SKRIPSI, dengan judul :

" IMPLEMENTASI HYBRID LEXICON-BASED DAN SVM UNTUK KLASIFIKASI ANALISIS SENTIMEN TERHADAP PELATIHAN BBPSDMP KOMINFO MAKASSAR "

Yang akan dilaksanakan dari : Tgl. 19 Juli s/d 17 Oktober 2025

Sehubungan dengan hal tersebut diatas, pada prinsipnya kami **menyetujui** kegiatan dimaksud dengan ketentuan yang tertera di belakang surat izin penelitian.

Demikian Surat Keterangan ini diberikan agar dipergunakan sebagaimana mestinya.

Diterbitkan di Makassar  
Pada Tanggal 19 Juli 2025

KEPALA DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU  
SATU PINTU PROVINSI SULAWESI SELATAN



ASRUL SANI, S.H., M.Si.  
Pangkat : PEMBINA UTAMA MUDA (IV/c)  
Nip : 19750321 200312 1 008

Tembusan Yth

1. Ketua LP3M UNISMUH Makassar di Makassar;
2. Pertinggal.

## Lampiran 6. Surat Balasan Izin Penelitian dari BBPSDMP Kominfo Makassar



Nomor : B-390/BBPSDMP.73/UM.01.01/08/2025 Makassar, 1 Agustus 2025  
Sifat : Biasa  
Hal : Balasan Surat Izin Penelitian

Kepada Yth.  
Kepala Dinas Penanaman Modal dan Pelayanan Terpadu  
Satu Pintu Provinsi Sulawesi Selatan  
di Tempat

Sehubungan dengan Surat Permohonan Izin Penelitian No. 16038/S.01/PTSP/2025  
yang diajukan kepada kami oleh mahasiswa Bapak/Ibu atas nama :

No.	Nomor Pokok	NAMA	Program Studi
1	105841103621	Nur Alam	Informatika

Bersama surat ini kami memberikan izin kepada mahasiswa tersebut untuk melakukan Penelitian Skripsi dengan judul Implementasi Hybrid Lexicon-Based dan SVM untuk Klasifikasi Analisis Sentimen Terhadap Pelatihan BBPSDMP Kominfo Makassar. Kontak Person yang bisa dihubungi atas nama Mardiana (0859196484559) atau Dhilah (085395808110).

Selanjutnya dalam upaya mendukung Zona Integritas Wilayah Bebas dari Korupsi, Apabila dalam Pelayanan kami terdapat penyimpangan atau pelanggaran kode etik oleh pegawai kami, mohon perkenan Bapak/Ibu melaporkan dengan disertai bukti otentik (identitas pelapor akan dijamin kerahasiaannya).

Demikian surat balasan ini dibuat agar dapat dipergunakan sebagaimana mestinya.

Kepala BBPSDMP Makassar



Baso Saleh

Catatan :

- UU ITE No 11 Tahun 2008 Pasal 5 ayat 1

- "Informasi Elektronik dan/atau Dokumen Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah."

- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSRI

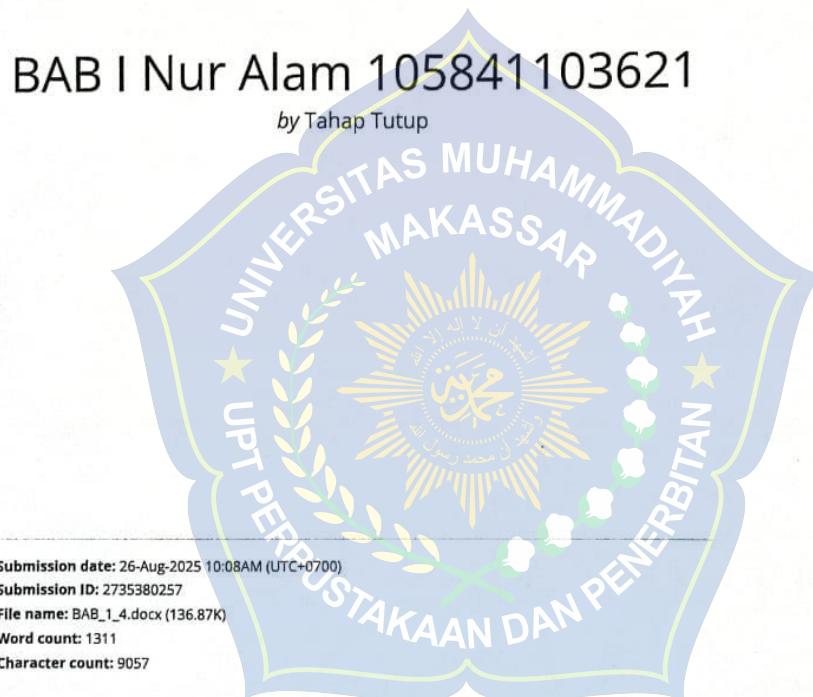
## Lampiran 7. Surat Keterangan Bebas Plagiat



## Lampiran 8. Hasil Turnitin

BAB I Nur Alam 105841103621

by Tahap Tutup



Submission date: 26-Aug-2025 10:08AM (UTC+07:00)

Submission ID: 2735380257

File name: BAB\_1\_4.docx (136.87K)

Word count: 1311

Character count: 9057

BAB I Nur Alam 105841103621

ORIGINALITY REPORT

2% 2% 2% %  
SIMILARITY INDEX INTERNET SOURCES PUBLICATIONS STUDENT PAPERS

PRIMARY SOURCES

1 etheses.uinmataram.ac.id 2%  
Internet Source

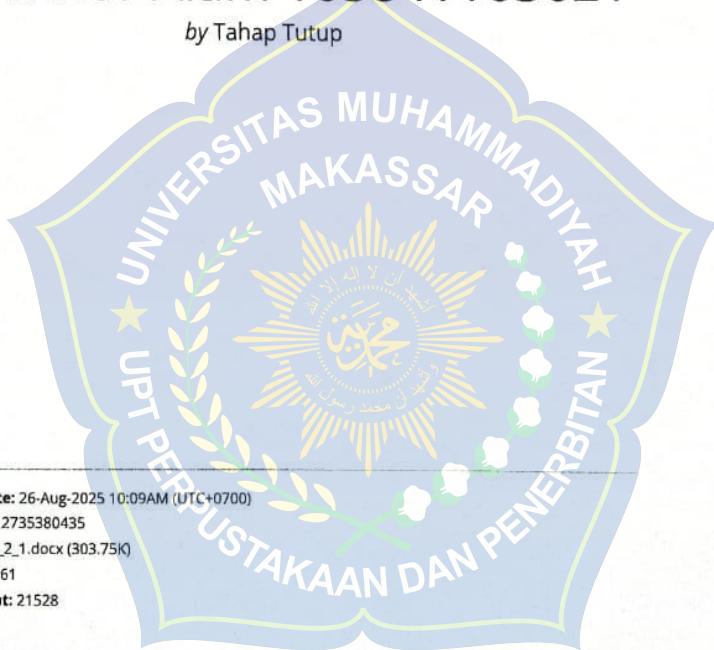
Exclude quotes Off  
Exclude bibliography Off

Exclude matches < 2%



## BAB II Nur Alam 105841103621

by Tahap Tutup



**Submission date:** 26-Aug-2025 10:09AM (UTC+0700)

**Submission ID:** 2735380435

**File name:** BAB\_2\_1.docx (303.75K)

**Word count:** 3161

**Character count:** 21528

## BAB II Nur Alam 105841103621

### ORIGINALITY REPORT



- 10** Ivan Aditya Nugraha, Irving Vlitra Paputungan. "Analisis Sentimen Video Review Starlink Indonesia Menggunakan Pendekatan Lexicon Dictionary Bahasa Indonesia Dan Algoritma Random Forest", Innovative: Journal Of Social Science Research, 2025  
Publication <1 %
- 11** 123dok.com  
Internet Source <1 %
- 12** pemrogramanmatlab.com  
Internet Source <1 %
- 13** Ferry Darmawan, Michael Joe, Yogiek Indra Kurniawan, Lasmedi Afuan. "Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada Twitter Menggunakan Support Vector Machine", Jurnal Eksplora Informatika, 2023  
Publication <1 %
- 14** adoc.pub  
Internet Source <1 %
- 15** library.universitaspertamina.ac.id  
Internet Source <1 %
- 16** www.researchgate.net  
Internet Source <1 %
- 17** Hendry Cipta Husada, Adi Suryaputra Paramita. "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)", Teknika, 2021  
Publication <1 %
- 18** djournals.com  
Internet Source <1 %

19	jurnal.unimus.ac.id Internet Source	<1 %
20	media.neliti.com Internet Source	<1 %
21	publikasi.uyelindo.ac.id Internet Source	<1 %

Exclude quotes Off      Exclude matches Off  
Exclude bibliography Off



# BAB III Nur Alam 105841103621

by Tahap Tutup



Submission date: 26-Aug-2025 10:09AM (UTC+0700)

Submission ID: 2735380736

File name: BAB\_3\_2.docx (327.9K)

Word count: 2532

Character count: 16645

### BAB III Nur Alam 105841103621

#### ORIGINALITY REPORT

9%  
SIMILARITY INDEX  
5%  
INTERNET SOURCES  
9%  
PUBLICATIONS  
%  
STUDENT PAPERS

#### PRIMARY SOURCES

- |   |   |    |
|---|---|----|
| 1 | Albert Genawan, Andika Bayu Saputra, M. Abu Amar Al Badawi. "Analisis Sentimen Opini Masyarakat Tentang Penggunaan Aplikasi Bimbingan Belajar Online di Masa Pandemi Covid-19 Menggunakan Metode Support Vector Machine (SVM)", Teknomatika: Jurnal Informatika dan Komputer, 2022<br>Publication | 3% |
| 2 | Dedi Irawan, Yayu Sri Rahayu, Mohamad Farozi. "KLASIFIKASI SENTIMEN PELANGGAN RESTORAN KOKI SUNDA MENGGUNAKAN ALGORITMA NAIVE BAYES", Jurnal Manajemen Informatika dan Sistem Informasi, 2025<br>Publication  | 2% |
| 3 | ejournal.gunadarma.ac.id<br>Internet Source   | 2% |
| 4 | Saiful Nur Budiman, Sri Lesanti, Erwan. "Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes", JAMI: Jurnal Ahli Muda Indonesia, 2024<br>Publication   | 2% |

Exclude quotes Off  
Exclude bibliography Off

Exclude matches Off

BAB IV Nur Alam  
105841103621

by Tahap Tutup

**Submission date:** 26-Aug-2025 10:10AM (UTC+0700)  
**Submission ID:** 2735381020  
**File name:** BAB\_4\_2.docx (2.18M)  
**Word count:** 9308  
**Character count:** 63284

BAB IV Nur Alam 105841103621

ORIGINALITY REPORT



6	pt.scribd.com Internet Source	<1 %
7	e-journal.rosma.ac.id Internet Source	<1 %
8	journal.uta45jakarta.ac.id Internet Source	<1 %
9	id.123dok.com Internet Source	<1 %
10	Anand Nayyar, Ajantha Devi Vairamani, Kuldeep Kaswan. "Evaluation and refinement of prompt engineering", Elsevier BV, 2025 Publication	<1 %
11	Mohammed Alarfaj, Mosleh Hmoud Al-Adhaileh, M. Irfan Uddin, Muhammad Adnan, Theyazn H. H. Aldhyani. "Enhancing Seizure Detection Accuracy in Wearable EEG Devices Using Deep Learning Algorithms", Journal of Disability Research, 2025 Publication	<1 %
12	Putu Jeevallucas Jnanamaitriya Surya Gautama, Argo Wibowo, Jong Jek Siang. "Analisis Sentimen Pengguna Terhadap Akun X/Twitter Resmi "DANA" Dengan Algoritma Indobert", INTECOMS: Journal of Information Technology and Computer Science, 2025 Publication	<1 %
13	Muhammad I. Rosadi, Agus Z. Arifin, Anny Yuniarti. "KLASIFIKASI MASSA PADA CITRA MAMMOGRAM MENGGUNAKAN KOMBINASI SELEKSI FITUR F-SCORE DAN LS-SVM", Teknologi, 2016 Publication	<1 %

- 14 Nuri Cahyono, Anggista Oktavia Praneswara. "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes", Indonesian Journal of Computer Science, 2023 Publication <1 %
- 15 Muhammad Reyan, Franindya Purwaningtyas. "ANALISIS SENTIMEN ULASAN APLIKASI IBI LIBRARY PADA GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER", Djtechno: Jurnal Teknologi Informasi, 2025 Publication <1 %
- 16 www.sulselsatu.com Internet Source <1 %
- 17 repository.unika.ac.id Internet Source <1 %
- 18 Rizky Saputra Cendikiawan, Ali Ibrahim, Mira Afrina, Rizka Dhini Kurnia. "Sentiment Analysis of Zalora Products on Google Play Store Using Random Forest Method", Indonesian Journal of Innovation Studies, 2025 Publication <1 %
- 19 amslaurea.unibo.it Internet Source <1 %
- 20 jurnal.untan.ac.id Internet Source <1 %
- 21 mafiadoc.com Internet Source <1 %
- 22 erepo.unud.ac.id Internet Source <1 %

- 23** repo.itera.ac.id **<1 %**  
Internet Source
- 
- 24** Azzikra Ramadhanti Aksan, Deviana Dyah Anggraini, Muhamad Fikry Maulana Ridwan, Herdiesel Santoso. "Sistem Deteksi Dini Anemia pada Anak Usia 0-59 Bulan Menggunakan Naïve Bayes dan Optimasi Particle Swarm Optimization", Teknematika: Jurnal Informatika dan Komputer, 2025 Publication **<1 %**
- 
- 25** adoc.pub **<1 %**  
Internet Source
- 
- 26** eprints.uny.ac.id **<1 %**  
Internet Source
- 
- 27** jurnal.umrah.ac.id **<1 %**  
Internet Source
- 
- 28** repository.ub.ac.id **<1 %**  
Internet Source
- 
- 29** www.elrocomputer.com **<1 %**  
Internet Source
- 
- 30** Mohammad Firos Anggarda, Nurul Fahmi Arief Hakim. "Pengembangan Sistem Prediksi Waktu Penyiraman Optimal pada Perkebunan: Pendekatan Machine Learning untuk Peningkatan Produktivitas Pertanian", JURNAL BUDIDAYA PERTANIAN, 2023 Publication **<1 %**
- 
- 31** Rizal Dwi Rizki Apriliansyah, Rini Astuti, Willy Prihartono, Ryan Hamonangan. "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN PENGUNJUNG DI PANTAI" **<1 %**

KEJAWANAN", Jurnal Informatika dan Teknik  
Elektro Terapan, 2025  
Publication

Exclude quotes Off  
Exclude bibliography Off

Exclude matches Off



BAB V Nur Alam 105841103621



**Submission date:** 26-Aug-2025 10:10AM (UTC+0700)

**Submission ID:** 2735381085

**File name:** BAB\_5\_2.docx (109.58K)

**Word count:** 222

**Character count:** 1498

## BAB V Nur Alam 105841103621

### ORIGINALITY REPORT

5% SIMILARITY INDEX 5% INTERNET SOURCES 0% PUBLICATIONS % STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ repository.stikeswirahusada.ac.id  
Internet Source

Exclude quotes  
Exclude bibliography

On  
On

Exclude matches

< 2%

