

**ANALISIS SENTIMEN PUBLIK TERHADAP ISU SEPAK BOLA INDONESIA  
MENGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

**Oleh:**

**KI AGENG NASROKH MANGKUNEGARA PUTRA**  
**NIM. 220605110109**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**ANALISIS SENTIMEN PUBLIK TERHADAP ISU SEPAK BOLA  
INDONESIA MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

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

Oleh :  
**KI AGENG NASROKH MANGKUNEGARA PUTRA**  
**NIM. 220605110109**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

## HALAMAN PERSETUJUAN

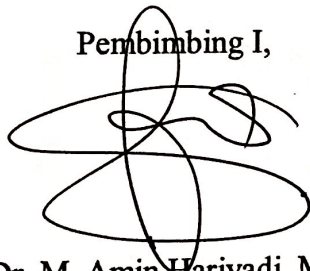
### ANALISIS SENTIMEN PUBLIK TERHADAP ISU SEPAK BOLA INDONESIA MENGGUNAKAN *SUPPORT VECTOR MACHINE*

#### SKRIPSI

Oleh :  
**KI AGENG NASROKH MANGKUNEGARA PUTRA**  
NIM. 220605110109

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 05 Mei 2026

Pembimbing I,



Dr. M. Amin Hariyadi, M.T  
NIP. 19670118 200501 1 001

Pembimbing II,



Shoffin Nahwa Utama, M.T  
NIP. 19860703 202012 1 003

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



Supriyono, M.Kom  
NIP. 19841010 201903 1 012

## HALAMAN PENGESAHAN

### ANALISIS SENTIMEN PUBLIK TERHADAP ISU SEPAK BOLA INDONESIA MENGGUNAKAN *SUPPORT VECTOR MACHINE*

#### SKRIPSI

Oleh :

**KI AGENG NASROKH MANGKUNEGARA PUTRA**  
**NIM. 220605110109**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 08 Juni 2026

#### Susunan Dewan Penguji

|                     |                                                                   |
|---------------------|-------------------------------------------------------------------|
| Ketua Penguji       | : <u>Dr. Cahyo Crysdian, M.Cs</u><br>NIP. 19740424 200901 1 008   |
| Anggota Penguji I   | : <u>Khadijah F.H. Holle, M.Kom</u><br>NIP. 19900626 202203 2 002 |
| Anggota Penguji II  | : <u>Dr. M. Amin Hariyadi, M.T</u><br>NIP. 19670118 200501 1 001  |
| Anggota Penguji III | : <u>Shoffin Nahwa Utama, M.T</u><br>NIP. 19860703 202012 1 003   |

(  )  
(  )  
(  )  
(  )

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



Simkyono, M.Kom  
NIP. 19841010 201903 1 012

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : KI AGENG NASROKH MANGKUNEGARA  
PUTRA  
NIM : 220605110109  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Analisis Sentimen Publik Terhadap Isu Sepak Bola  
Indonesia Menggunakan *Support Vector Machine*

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

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

Malang, 18 Juni 2026

Yang membuat pernyataan,



KI AGENG NASROKH MANGKUNEGARA PUTRA  
NIM. 220605110109

## **MOTTO**

*“Mungkin kita sampai, mungkin saja tidak, tugas kita hanyalah Berjalan”*

(The Jeblogs-Bersandarlal)

## HALAMAN PERSEMBAHAN

*Alhamdulillah rabbil 'alamin*, segala puji bagi Allah Swt. atas segala rahmat, nikmat, dan karunia-Nya. Atas izin dan pertolongan Allah Swt., penulis diberikan kemudahan dan kekuatan hingga dapat menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada kedua orang tua tercinta sebagai wujud bakti, cinta, kasih sayang, dan rasa terima kasih atas segala doa, pengorbanan, serta dukungan yang telah diberikan. Kehadiran dan doa mereka menjadi alasan utama yang menguatkan penulis untuk terus berjuang hingga mampu menyelesaikan skripsi ini.

Skripsi ini juga penulis persembahkan kepada diri sendiri sebagai bentuk apresiasi atas segala perjuangan, kesabaran, dan keteguhan dalam melewati setiap proses yang penuh tantangan. Semoga pencapaian ini menjadi awal dari langkah-langkah yang lebih baik serta menjadi pengingat bahwa setiap usaha yang dilakukan dengan sungguh-sungguh akan membuahkan hasil pada waktunya.

## KATA PENGANTAR

*Alhamdulillahillāhirabbil ‘ālamīn*, segala puji bagi Allah *Subḥānahu wa Ta ‘ālā*. yang telah melimpahkan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Analisis Sentimen Publik terhadap Isu Sepak Bola Indonesia Menggunakan Support Vector Machine*” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shollallohu ‘Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Bapak Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Dr. M. Amin Hariyadi, M.T., selaku dosen pembimbing I dan Bapak Shoffin Nahwa Utama, M.T., selaku dosen pembimbing II, yang telah memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas waktu, kesabaran,

dan perhatian yang telah diberikan sehingga penelitian dan penulisan skripsi ini dapat terselesaikan dengan baik dan tepat waktu.

5. Bapak Dr. Cahyo Crysdian, M.Cs., selaku dosen penguji I dan Ibu Khadijah F.H. Holle, M.Kom., selaku dosen penguji II, yang telah memberikan evaluasi, saran, serta berbagai masukan yang berharga sehingga penelitian dan penulisan skripsi ini dapat tersusun dengan lebih baik. Terima kasih atas arahan dan perhatian yang diberikan dalam proses penyusunan skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, penulis menyampaikan ucapan terima kasih atas ilmu yang telah di berikan selama masa perkuliahan. Semoga segala ilmu dan kebaikan yang telah diberikan menjadi amal jariyah serta memberikan manfaat yang berkelanjutan bagi penulis di masa yang akan datang.
7. Ibu tercinta Sri Nuryani, yang senantiasa memberikan kasih sayang, dukungan, motivasi, serta pengorbanan yang tiada henti kepada penulis. Keindahan dan ketulusan doamu yang tidak pernah putus menjadi penyerta dalam setiap perjuangan penulis, sehingga berbagai proses dan tantangan dapat dilalui hingga skripsi ini terselesaikan dengan baik.
8. Ayah tercinta Kamim Famuji Efendi, yang senantiasa menjadi teladan dalam kerja keras, keteguhan, dan tanggung jawab. Terima kasih atas doa, dukungan, serta pengorbanan yang telah diberikan selama ini. Setiap nasihat dan perjuangan beliau selalu menjadi sumber motivasi bagi penulis untuk

terus berusaha hingga mampu menyelesaikan skripsi ini dengan baik dan tepat waktu.

9. Kakak Alif Mushollina Bintang Sakti dan Adik Muhammad Ali Hasan Rohmatulloh Syarifuddin, yang selalu hadir memberikan semangat, doa, dan dukungan kepada penulis dalam setiap proses perjalanan pendidikan.
10. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu per satu, terima kasih atas doa, dukungan, perhatian, dan semangat yang telah diberikan kepada penulis selama menempuh pendidikan.
11. Para masyayikh, jajaran guru, serta teman-teman di Yayasan PPM Daruttauhid Al-Hasaniyyah, penulis menyampaikan terima kasih atas doa, ilmu, nasihat, dan dukungan yang telah diberikan.
12. Sahabat seperjuangan yang senantiasa kebersamai penulis selama masa perkuliahan, yaitu Muhammad Andrean Hidayatulloh, Ahmad Fadhli Dzilikrom, Muhammad Alif Atallah, Muhammad Zidan Alvaro, dan Muhhammad Annas Darunnaja. Terima kasih atas dukungan, kebersamaan, serta berbagai pengalaman yang telah dibagikan sehingga perjalanan akademik ini terasa lebih ringan dan dapat diselesaikan tepat waktu.
13. Keluarga besar Teknik Informatika angkatan 2022 Infinity, yang telah memberikan kebersamaan, dukungan, dan pengalaman berharga selama masa perkuliahan.
14. Seseorang dengan ukuran sepatu 37, terima kasih atas segala doa, perhatian, dan ketulusan yang telah diberikan. Terima kasih karena telah memilih untuk tetap kebersamai setiap proses, mendengarkan setiap keluh kesah,

serta menjadi penyemangat di tengah berbagai tantangan yang dihadapi penulis hingga skripsi ini dapat terselesaikan.

15. Persebaya Surabaya, tim kebanggaan penulis yang selalu menghadirkan semangat, harapan, dan cerita dalam setiap perjalanannya. Dari lapangan hijau, penulis belajar bahwa perjuangan tidak selalu berakhir dengan kemenangan, tetapi tidak pernah boleh berhenti diperjuangkan. Semangat *“Salam Satu Nyali, Wani!”* menjadi pengingat bagi penulis untuk tetap teguh, berani, dan pantang menyerah dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat berbagai keterbatasan yang memerlukan penyempurnaan di masa mendatang. Namun demikian, penulis meyakini bahwa proses belajar tidak memiliki batas akhir, sehingga setiap kritik dan saran yang membangun akan menjadi bekal berharga untuk terus berkembang. Bagi penulis, skripsi yang baik bukanlah skripsi yang sempurna, melainkan skripsi yang dapat diselesaikan dengan penuh tanggung jawab dan tepat pada waktunya. Oleh karena itu, penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang membutuhkan sebagai referensi untuk penelitian selanjutnya.

Malang, 18 Juni 2026

Penulis

## DAFTAR ISI

|                                                                          |              |
|--------------------------------------------------------------------------|--------------|
| <b>HALAMAN PENGAJUAN .....</b>                                           | <b>ii</b>    |
| <b>HALAMAN PERSETUJUAN.....</b>                                          | <b>iii</b>   |
| <b>HALAMAN PENGESAHAN .....</b>                                          | <b>iv</b>    |
| <b>PERNYATAAN KEASLIAN TULISAN .....</b>                                 | <b>v</b>     |
| <b>MOTTO .....</b>                                                       | <b>vi</b>    |
| <b>HALAMAN PERSEMBAHAN.....</b>                                          | <b>vii</b>   |
| <b>KATA PENGANTAR.....</b>                                               | <b>viii</b>  |
| <b>DAFTAR ISI.....</b>                                                   | <b>xii</b>   |
| <b>DAFTAR GAMBAR.....</b>                                                | <b>xiv</b>   |
| <b>DAFTAR TABEL.....</b>                                                 | <b>xv</b>    |
| <b>ABSTRAK .....</b>                                                     | <b>xvi</b>   |
| <b>ABSTRACT .....</b>                                                    | <b>xvii</b>  |
| <b>البحث مستخلص.....</b>                                                 | <b>xviii</b> |
| <b>BAB I PENDAHULUAN.....</b>                                            | <b>1</b>     |
| 1.1 Latar Belakang .....                                                 | 1            |
| 1.2 Pernyataan Masalah .....                                             | 7            |
| 1.3 Tujuan Penelitian .....                                              | 8            |
| 1.4 Batasan Masalah .....                                                | 8            |
| 1.5 Manfaat Penelitian .....                                             | 8            |
| <b>BAB II STUDI PUSTAKA .....</b>                                        | <b>9</b>     |
| 2.1 Analisis Sentimen Publik .....                                       | 9            |
| 2.2 <i>Support Vector Machine</i> untuk Klasifikasi Teks .....           | 15           |
| <b>BAB III DESAIN DAN IMPLEMENTASI SISTEM .....</b>                      | <b>17</b>    |
| 3.1 Pengumpulan Data .....                                               | 17           |
| 3.2 Desain Sistem.....                                                   | 21           |
| 3.3 <i>Preprocessing Text</i> .....                                      | 22           |
| 3.4 Pembobotan Kata .....                                                | 27           |
| 3.4.1 Nilai <i>Term-Frequency</i> (TF).....                              | 28           |
| 3.4.2 Nilai <i>Inverse Document Frequency</i> (IDF).....                 | 29           |
| 3.4.3 Perhitungan <i>Term Frequency-Inverse Document Frequency</i> ..... | 30           |
| 3.5 <i>Support Vector Machine</i> (SVM).....                             | 32           |
| <b>BAB IV UJI COBA DAN PEMBAHASAN .....</b>                              | <b>36</b>    |
| 4.1 Persiapan Data .....                                                 | 36           |
| 4.2 Skenario Uji Coba.....                                               | 37           |
| 4.3 Hasil Uji Coba.....                                                  | 41           |
| 4.3.1 Skenario Uji Coba 1 .....                                          | 41           |
| 4.3.2 Skenario Uji Coba 2 .....                                          | 44           |
| 4.3.3 Skenario Uji Coba 3 .....                                          | 46           |
| 4.4 Pembahasan.....                                                      | 49           |
| <b>BAB V KESIMPULAN DAN SARAN .....</b>                                  | <b>54</b>    |

|                       |    |
|-----------------------|----|
| 5.1 Kesimpulan .....  | 54 |
| 5.2 Saran .....       | 55 |
| <b>DAFTAR PUSTAKA</b> |    |

## DAFTAR GAMBAR

|                                                                                                          |    |
|----------------------------------------------------------------------------------------------------------|----|
| Gambar 2.1 Ilustrasi Support Vector Machine (Sumber: <a href="http://astroml.org">astroml.org</a> )..... | 15 |
| Gambar 3. 1 Desain Sistem.....                                                                           | 22 |
| Gambar 3.2 Visualisasi Persebaran Awal Data.....                                                         | 33 |
| Gambar 4.1 Confusion Matrix Skenario 1 .....                                                             | 43 |
| Gambar 4.2 Confusion Matrix Skenario 2 .....                                                             | 45 |
| Gambar 4.3 Confusion Matrix Skenario 3 .....                                                             | 48 |

## DAFTAR TABEL

|                                                     |    |
|-----------------------------------------------------|----|
| Tabel 2.1 Perbandingan Penelitian yang Relevan..... | 13 |
| Tabel 3.1 Pengumpulan Data .....                    | 17 |
| Tabel 3.2 Jumlah Data.....                          | 18 |
| Tabel 3.3 Sampel Data .....                         | 19 |
| Tabel 3.4 Data Terlabel.....                        | 20 |
| Tabel 3.5 Tahap Cleaning .....                      | 23 |
| Tabel 3.6 Tahap Case Folding .....                  | 24 |
| Tabel 3.7 Tahap Normalization.....                  | 25 |
| Tabel 3.8 Tahap Tokenizing .....                    | 25 |
| Tabel 3.9 Tahap Stopword Removal.....               | 26 |
| Tabel 3.10 Tahap <i>Stemming</i> .....              | 27 |
| Tabel 3.11 Nilai Parameter .....                    | 35 |
| Tabel 4.1 Jumlah Data.....                          | 36 |
| Tabel 4.2 Skenario Pengujian .....                  | 37 |
| Tabel 4.3 Confusion Matrix .....                    | 38 |
| Tabel 4.4 Pembagian Data Skenario 1 .....           | 41 |
| Tabel 4.5 Parameter Skenario 1 .....                | 42 |
| Tabel 4.6 Evaluasi Model Skenario 1 .....           | 42 |
| Tabel 4.7 Pembagian Data Skenario 2 .....           | 44 |
| Tabel 4.8 Parameter Skenario 2 .....                | 45 |
| Tabel 4.9 Evaluasi Model Skenario 2 .....           | 45 |
| Tabel 4.10 Pembagian Data Skenario 3 .....          | 47 |
| Tabel 4.11 Parameter Skenario 3 .....               | 47 |
| Tabel 4.12 Evaluasi Model Skenario 3 .....          | 47 |

## ABSTRAK

Putra, Ki Ageng Nasrokh Mangkunegara. 2026. **Analisis Sentimen Publik Terhadap Isu Sepak Bola Indonesia Menggunakan *Support Vector Machine***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T (II) Shoffin Nahwa Utama, M.T

**Kata Kunci:** Analisis Sentimen, *Support Vector Machine*, Pembobotan TF-IDF, Media Sosial X, Sepak Bola Indonesia.

Media sosial menjadi ruang utama bagi masyarakat untuk menyampaikan opini terkait berbagai isu sepak bola Indonesia, seperti pergantian pelatih Tim Nasional, performa tim dalam kompetisi internasional, hingga kebijakan federasi. Besarnya jumlah opini yang tersebar dalam bentuk teks menyebabkan analisis secara manual menjadi kurang efektif sehingga diperlukan metode otomatis untuk mengidentifikasi sentimen publik. Penelitian ini bertujuan menganalisis performa algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen publik terhadap isu sepak bola Indonesia ke dalam kategori positif dan negatif. Data penelitian diperoleh dari platform X melalui proses *crawling* terhadap enam isu sepak bola nasional dengan total 1.487 *tweet* berbahasa Indonesia yang telah diberi label oleh ahli bahasa. Tahapan pengolahan data meliputi *preprocessing*, pembobotan fitur menggunakan TF-IDF, serta klasifikasi menggunakan SVM kernel *Radial Basis Function* (RBF). Pengujian dilakukan pada tiga skenario pembagian data, yaitu 90:10, 80:20, dan 70:30. Hasil penelitian menunjukkan bahwa model terbaik diperoleh pada rasio 80:20 dengan nilai *accuracy* sebesar 86,91% dan *Macro F1-score* sebesar 68,04%. Temuan ini menunjukkan bahwa SVM mampu mengklasifikasikan sentimen publik secara efektif serta berpotensi digunakan untuk memahami respons masyarakat terhadap dinamika sepak bola Indonesia.

## ABSTRACT

Putra, Ki Ageng Nasrokh Mangkunegara. 2026. **Public Sentiment Analysis on Indonesian Football Issues Using Support Vector Machine**. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Supervisors: (I) Dr. M. Amin Hariyadi, M.T. (II) Shoffin Nahwa Utama, M.T.

**Key words:** Sentiment Analysis, Support Vector Machine, TF-IDF Weighting, X Social Media, Indonesian Football.

Social media has become a primary platform for the public to express opinions regarding various issues in Indonesian football, such as changes in the national team coaching staff, team performance in international competitions, and federation policies. The large volume of opinions distributed in textual form makes manual analysis less effective, thereby requiring automated methods to identify public sentiment. This study aims to analyze the performance of the Support Vector Machine (SVM) algorithm in classifying public sentiment toward Indonesian football issues into positive and negative categories. The research data were collected from the X platform through a crawling process on six national football issues, resulting in a dataset of 1,487 Indonesian-language tweets that were manually labeled by language experts. The data processing stages consisted of text preprocessing, feature weighting using Term Frequency–Inverse Document Frequency (TF-IDF), and classification using SVM with the Radial Basis Function (RBF) kernel. The experiments were conducted using three data-splitting scenarios: 90:10, 80:20, and 70:30. The results showed that the best model was achieved with the 80:20 split ratio, obtaining an accuracy of 86.91% and a Macro F1-score of 68.04%. These findings indicate that SVM is capable of effectively classifying public sentiment and has the potential to be utilized for understanding public responses to the dynamics of Indonesian football.

## البحث مستخلص

بوترا، كي أغينغ ناصروخ مانغكونيغارا. 2026. تحليل المشاعر العامة تجاه قضايا كرة القدم الإندونيسية باستخدام آلة المتجهات الداعمة (Support Vector Machine) رسالة جامعية. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرفان: (الأول) الدكتور محمد أمين هريادي، ماجستير تقني، (الثاني) شوفين نهاوا أوتاما، ماجستير تقني.

الكلمات المفتاحية: تحليل المشاعر، آلة المتجهات الداعمة (Support Vector Machine)، أوزان TF-IDF، منصة X، كرة القدم الإندونيسية.

أصبحت وسائل التواصل الاجتماعي فضاءً رئيسيًا يعبر من خلاله المجتمع عن آرائه تجاه مختلف قضايا كرة القدم الإندونيسية، مثل تغيير مدرب المنتخب الوطني، وأداء الفريق في المنافسات الدولية، وسياسات الاتحاد. ونظرًا إلى الكم الكبير من الآراء المتداولة في شكل نصوص، فإن تحليلها يدويًا يُعد أقل فاعلية، مما يستلزم استخدام أساليب آلية للتعرف على المشاعر العامة. تهدف هذه الدراسة إلى تحليل أداء خوارزمية آلة المتجهات الداعمة (SVM) في تصنيف المشاعر العامة تجاه قضايا كرة القدم الإندونيسية إلى فئتين: إيجابية وسلبية. جُمعت بيانات الدراسة من منصة X من خلال عملية جمع بيانات لست قضايا تتعلق بكرة القدم الوطنية، حيث بلغ عدد التغريدات باللغة الإندونيسية 1487 تغريدة، وقد تم تصنيفها يدويًا من قبل خبراء لغويين. وشملت مراحل معالجة البيانات المعالجة المسبقة للنصوص، واستخراج الأوزان باستخدام أسلوب التردد المصطلحي-معكوس تردد المستندات (TF-IDF)، ثم التصنيف باستخدام خوارزمية SVM ذات نواة الدالة الشعاعية الأساسية (RBF). أُجريت التجارب باستخدام ثلاثة سيناريوهات لتقسيم البيانات، وهي 90:10 و 80:20 و 70:30. وأظهرت النتائج أن أفضل نموذج تحقق عند نسبة تقسيم 80:20، حيث بلغت الدقة 86.91% وبلغت قيمة Macro F1-score نسبة 68.04%. وتشير هذه النتائج إلى أن خوارزمية SVM قادرة على تصنيف المشاعر العامة بكفاءة، ولها إمكانات كبيرة في فهم استجابات المجتمع تجاه تطورات كرة القدم الإندونيسية.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Sebagai salah satu cabang olahraga yang mempunyai jumlah penggemar terbesar secara global, sepak bola menempati posisi cukup krusial pada kehidupan masyarakat, termasuk di Indonesia. Popularitasnya bukan hanya terlihat sebagai aktivitas olahraga semata, tetapi juga berkembang menjadi fenomena sosial yang melekat dalam kehidupan sehari-hari. Pengaruh sepak bola menjangkau berbagai kelompok masyarakat dari beragam latar belakang dan rentang usia.(Eko Alwinasirin & Rizky Putra Santosa, 2025). Tingginya minat publik terhadap olahraga ini tercermin dari besarnya perhatian yang diberikan pada setiap pertandingan, baik yang melibatkan klub maupun Tim Nasional Indonesia. Antusiasme tersebut mampu menarik perhatian jutaan penonton dan menghadirkan keterlibatan emosional yang kuat antara masyarakat dengan sepak bola. Kondisi ini menjadikan sepak bola bukan hanya menjadi media hiburan, tetapi juga menjadi media pemersatu yang memiliki daya tarik luas di tengah (Djuyandi *et al.*, 2021).

Besarnya basis penggemar di Indonesia tidak hanya terlihat dari keramaian di tribun stadion, tetapi juga dari dominasi topik pembicaraan mengenai dinamika sepak bola nasional yang selalu hangat dibicarakan di berbagai ruang publik, seiring dengan perkembangan industri sepak bola, peran publik, khususnya penggemar sepak bola, kini telah mengalami transformasi yang signifikan, mereka tidak lagi dipandang hanya sebagai penonton pasif yang hadir untuk mengisi tribun stadion.

Suporter telah menjadi elemen penting yang secara aktif mengamati, menganalisis, dan memberikan respon terhadap berbagai dinamika yang terjadi dalam sepak bola nasional (Dina Aulia, 2024).

Di era transformasi digital, aspirasi dan kritik tajam dari supporter atau publik penggemar sepak bola tersebut menemukan ruang ekspresi yang sangat luas melalui beragam kanal jejaring sosial. Kehadiran media sosial kini berperan sebagai arena diskusi terbuka di mana ribuan *tweet* mengenai isu-isu sepak bola terkini, seperti evaluasi kinerja pelatih atau manajemen tim yang diunggah setiap harinya secara *real-time* (Eka Wahyudhi *et al.*, 2023). Kumpulan opini tersebut menciptakan arus data yang sangat besar dan dinamis, yang merepresentasikan sentimen kolektif masyarakat, baik itu berupa dukungan positif maupun kritik negatif.

Namun, di balik besarnya antusiasme tersebut, sepak bola Indonesia dihadapkan pada berbagai kebijakan strategis yang acap kali tidak sesuai dengan ekspektasi serta harapan publik. Hal ini mencakup isu krusial seperti manajemen federasi, ketidakpastian masa depan pelatih, hingga ketidakefektifan agenda internasional. Ketidaksinkronan antara kebijakan pemangku kepentingan dengan ekspektasi masyarakat memicu gejolak opini dan kritik keras di ruang publik. Akibatnya, muncul polarisasi sentimen yang tajam di media sosial sebagai bentuk kekecewaan terhadap dinamika yang dianggap menghambat kemajuan prestasi nasional (Asshiddiqi & Lhaksana, 2020).

Berdasarkan dinamika tersebut, penelitian ini memfokuskan perhatian pada beberapa isu sentral yang paling banyak memicu polarisasi opini di kalangan publik, khususnya pengguna media sosial yang mengikuti isu sepak bola dalam kurun waktu terakhir. Isu utama yang diangkat meliputi dinamika posisi pelatih kepala Tim Nasional Shin Tae-yong, yang terbukti menjadi pemantik debat masif melalui platform media digital. Di samping itu, penelitian ini turut mengumpulkan berbagai isu krusial lainnya, seperti evaluasi performa Tim Nasional dalam berbagai turnamen internasional, tidak adanya agenda FIFA *Matchday*, hingga naiknya ketua umum PSSI (Erick Thohir) menjadi ketua Menpora yang banyak mendapat sorotan dari para pecinta sepak bola. Seluruh isu tersebut dikumpulkan menjadi satu kesatuan data untuk memberikan gambaran yang utuh mengenai bagaimana pola sentimen, dukungan, maupun kritik suporter terbentuk terhadap berbagai fenomena penting yang terjadi di dalam sepak bola Indonesia.

Banyaknya isu yang beragam terkait sepak bola Indonesia memicu opini yang beragam juga dari publik penggemar sepak bola, ada yang beropini positif dengan memberikan apresiasi dan dukungan penuh terhadap perkembangan tim nasional, namun tidak sedikit pula yang memberikan opini negatif berupa kritik tajam serta tuntutan perubahan. Reaksi beragam ini seringkali diluapkan secara masif dan terbuka di platform media sosial. Untuk memahami spektrum sentimen publik tersebut secara mendalam, penelitian ini akan fokus menganalisis sentimen publik terhadap sepak bola Indonesia di platform sosial media X.

Penentuan platform X didasari sifatnya yang interaktif dan berbasis teks sebagai wadah opini publik secara langsung. Ekosistem diskusinya yang terbuka

dan padat informasi terkait isu sepak bola nasional menyediakan data masif yang representatif untuk menggambarkan kondisi sentimen kolektif suporter di tanah air secara akurat (Fardi & Akbar, 2026).

Berbagai ekspresi ini tersebar luas di media sosial dalam bentuk data teks yang sangat masif namun tidak terstruktur. Penggunaan bahasa yang informal, penggunaan singkatan yang kompleks, istilah-istilah *slang* khas suporter, hingga gaya bahasa sarkasme membuat data tersebut sulit untuk dipetakan secara konvensional. Melakukan analisis secara manual terhadap ribuan *tweet* tersebut tidak hanya memerlukan waktu yang sangat lama, tetapi juga memiliki risiko subjektivitas yang tinggi dalam menentukan kategori sentimen. Berdasarkan kondisi tersebut, diperlukan pendekatan komputasional yang dapat mengenali serta mengelompokkan opini-opini tersebut secara otomatis, objektif, dan akurat agar dapat dihasilkan informasi yang valid mengenai persepsi publik.

Untuk menjawab tantangan pengolahan data bervolume besar tersebut, penelitian ini menerapkan metode analisis sentimen. Metode ini termasuk dalam kajian *Natural Language Processing* (NLP) yang bertujuan untuk mengenali dan mengelompokkan opini maupun sikap afektif yang terdapat pada data tekstual. Fokus utamanya adalah mengidentifikasi kecenderungan sentimen atau polaritas dari opini tersebut, apakah ekspresi dari publik cenderung positif atau negatif (Fitriyah *et al.*, 2020). Dengan menggunakan analisis sentimen, ribuan komentar yang tidak terstruktur dapat diolah secara otomatis dan objektif, mengubahnya menjadi data terstruktur yang dapat diukur dan dipahami. Pendekatan komputasional ini memungkinkan peneliti untuk memetakan pandangan publik

secara luas, sesuatu yang tidak mungkin dicapai secara efisien melalui analisis manual (Julianti *et al.*, 2024).

Sebagai penelitian yang memanfaatkan data dari media sosial, objektivitas dan ketelitian menjadi prinsip utama dalam proses analisis. Setiap informasi yang diperoleh harus ditafsirkan berdasarkan pendekatan ilmiah agar terhindar dari subjektivitas dan asumsi yang tidak berdasar. Keharusan bersikap cermat dalam menerima informasi juga memiliki landasan dalam ajaran Islam. Allah Swt. berfirman dalam Surah Al-Isra [17] ayat 36:

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهِ عِلْمٌ إِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ أُولَٰئِكَ كَانَ عَنْهُ مَسْئُولًا

*“Dan janganlah kamu mengikuti apa yang kamu tidak mempunyai pengetahuan tentangnya. Sesungguhnya pendengaran, penglihatan dan hati, semuanya itu akan diminta pertanggungjawaban.”* (QS. Al-Isra’:36)

Menurut penafsiran Ibnu Katsir, QS. Al-Isra ayat 36 menekankan larangan bagi seseorang untuk menyampaikan informasi atau memberikan penilaian yang tidak didasarkan pada pengetahuan yang jelas. Pendengaran, penglihatan, dan hati nurani dipahami sebagai perangkat untuk memperoleh informasi yang nantinya akan dimintai pertanggungjawaban (Al-Sheikh, 2003a). Prinsip ini menegaskan pentingnya bersikap objektif dan berhati-hati dalam menyimpulkan suatu informasi (Fauziah & Nurulita, 2024). Dalam penelitian ini, pengklasifikasian sentimen dilakukan secara sistematis menggunakan data dan metode ilmiah agar hasilnya dapat dipertanggungjawabkan secara akademik.

Klasifikasi sentimen dalam penelitian ini dilakukan dengan menerapkan

algoritma *Support Vector Machine* (SVM) sebagai model *machine learning* pilihan. Keunggulan SVM untuk mengelola data tekstual berdimensi tinggi (*high-dimensional data*) serta reputasi akurasinya yang tinggi pada berbagai studi terdahulu menjadi alasan utama pemilihannya sebagai metode klasifikasi yang andal.(Munandar *et al.*, 2023). Di samping itu, SVM memiliki kemampuan untuk membentuk *decision boundary* atau *hyperplane* yang paling optimal dalam membedakan setiap kelas, bahkan untuk data yang kompleks dan memiliki banyak *noise*, kondisi yang relevan dengan sifat data komentar pada platform media sosial yang cenderung bervariasi dan belum terstruktur dengan baik.(Arsi & Waluyo, 2021).

Keandalan algoritma *Support Vector Machine* (SVM) dalam domain olahraga telah dibuktikan oleh berbagai penelitian sebelumnya. Dalifah dkk. (2024) mengkaji tanggapan pengguna Twitter mengenai berbagai informasi seputar Liga 1 Indonesia tahun 2022 dan memperoleh tingkat akurasi sebesar 91% dalam proses klasifikasi sentimen suporter terhadap penyelenggaraan Liga 1 Indonesia musim 2022 (Dalifah *et al.*, 2024). Dukungan terhadap penggunaan SVM juga ditunjukkan oleh penelitian Asshiddiqi dan Lhaksana (2020) yang membandingkan performa SVM dengan *Decision Tree* untuk mengklasifikasikan persepsi masyarakat terhadap kinerja PSSI melalui komentar di Instagram. Berdasarkan hasil pengujian yang dilakukan, integrasi antara metode *Support Vector Machine* (SVM) dan skema pembobotan TF-IDF mampu memberikan akurasi mencapai 94,36%. Capaian tersebut mengungguli performa algoritma *Decision Tree* yang hanya mencatatkan akurasi sebesar 87,45%. (Asshiddiqi & Lhaksana, 2020). Temuan tersebut

membuktikan keandalan performa metode SVM dalam mengidentifikasi dan mengklasifikasikan kecenderungan polaritas opini publik secara akurat, khususnya dalam studi kasus sepak bola Indonesia.

Merujuk pada permasalahan yang dipaparkan pada bab latar belakang, fokus utama dari penelitian ini adalah untuk menganalisis sentimen publik terkait isu sepak bola nasional melalui pemanfaatan data tekstual opini yang bersumber dari media sosial X. Algoritma *Support Vector Machine* (SVM) digunakan dalam penelitian ini guna untuk mengelompokkan data *tweet* menjadi dua orientasi sentimen yang berbeda, yakni sentimen positif serta sentimen negatif. Selanjutnya, fokus penelitian diarahkan pada pengukuran akurasi untuk menilai kinerja algoritma tersebut. Penelitian ini diharapkan berkontribusi dalam menyajikan data empiris mengenai kemampuan SVM dalam menganalisis respons dan sentimen masyarakat terhadap isu-isu krusial sepak bola tanah air. Temuan yang diperoleh diharapkan dapat menjadi sumber informasi awal mengenai pandangan masyarakat serta memberikan masukan bagi pemangku kepentingan, seperti PSSI dan manajemen tim, dalam memahami respons suporter secara lebih objektif guna mendukung peningkatan kualitas dan prestasi sepak bola Indonesia pada masa mendatang.

## **1.2 Pernyataan Masalah**

Bagaimana efektivitas dan kinerja metode *Support Vector Machine* (SVM) dalam melakukan klasifikasi sentimen publik ke dalam kelas positif dan negatif terkait isu sepak bola di Indonesia?

### 1.3 Tujuan Penelitian

Mengukur tingkat performa dan kapabilitas algoritma *Support Vector Machine* (SVM) dalam mengategorikan opini masyarakat terkait dinamika isu sepak bola Indonesia ke dalam kelas sentimen positif serta negatif bersumber dari data *tweet* media sosial X.

### 1.4 Batasan Masalah

- a. Dataset yang digunakan dalam studi ini diekstrak dari platform X, yang mencakup dokumen *tweet* serta *retweet* orisinal dalam bahasa Indonesia mengenai dinamika isu sepak bola nasional.
- b. Isu yang dikaji dalam penelitian ini mencakup enam isu krusial sepak bola Indonesia, antara lain dinamika pergantian kepelatihan Tim Nasional, performa Tim Nasional dalam kompetisi internasional (mencakup kualifikasi Piala Asia U-23, ajang SEA Games, dan kualifikasi Piala Dunia), agenda FIFA Matchday, serta kebijakan kepemimpinan federasi sepak bola Indonesia.

### 1.5 Manfaat Penelitian

Memberikan kontribusi terhadap pengembangan literatur *Natural Language Processing* (NLP) dan klasifikasi sentimen teks berbahasa Indonesia dalam domain olahraga, serta menjadi sumber rujukan bagi penelitian selanjutnya yang membahas penerapan algoritma *Support Vector Machine* dalam konteks sepak bola.

## BAB II

### STUDI PUSTAKA

#### 2.1 Analisis Sentimen Publik

Evaluasi terhadap opini publik mengenai pemberitaan Liga 1 Indonesia tahun 2022 di platform Twitter pernah dilakukan oleh Dalifah, Suarna, dan Prihartono (2024) menggunakan algoritma SVM. Studi tersebut memanfaatkan 509 data teks hasil penarikan (*crawling*) yang dikerjakan melewati fase *preprocessing* serta pembobotan kata TF-IDF. Guna mengidentifikasi arah kecenderungan opini publik, model *Support Vector Machine* (SVM) melakukan kategorisasi data ke dalam tiga label, yakni kelas negatif, positif, dan netral. Hasil pengujian memperlihatkan bahwa capaian akurasi maksimal berada pada angka 91%, yang mana hasil tersebut diperoleh pada alokasi data dengan proporsi 70:30. Menariknya, respons kritis masyarakat sangat mendominasi dengan persentase sentimen negatif sebesar 89,19% terkait berbagai dinamika kompetisi sepak bola tersebut (Dalifah *et al.*, 2024).

Dalam riset yang dikerjakan oleh Apryani dkk. (2025), klasifikasi sentimen dimanfaatkan untuk mengevaluasi tanggapan publik pada kinerja Timnas sepak bola Indonesia melalui komentar pada platform TikTok. Penelitian tersebut memanfaatkan 1.143 komentar yang dikumpulkan dari diskusi terkait pertandingan kualifikasi Piala Dunia. Proses analisis pada penelitian tersebut dilakukan melalui rangkaian tahapan *preprocessing* dan pembobotan nilai TF-IDF, yang dilanjutkan dengan pengujian terhadap tiga algoritma, yaitu KNN, SVM, dan Random Forest.

Menurut hasil evaluasi yang diperoleh, model RF berhasil mencatatkan tingkat *accuracy* tertinggi mencapai 97,30%. Performa maksimal tersebut kemudian disusul oleh KNN dengan raihan akurasi sebesar 94,30%, sementara algoritma SVM menempati posisi terakhir dengan perolehan akurasi sebesar 58,77%. Temuan tersebut menunjukkan bahwa karakteristik bahasa pada TikTok yang lebih santai, tidak baku, dan kaya akan *noise* menjadi tantangan tersendiri bagi kinerja SVM dalam melakukan klasifikasi sentimen (Apriyani *et al.*, 2025).

Farhan dkk. (2025) melakukan analisis sentimen terhadap komentar suporter Sriwijaya FC di Instagram yang membahas manajemen klub. Sebanyak 6.601 komentar digunakan sebagai dataset dengan rasio pembagian data 80:20. Hasil memperlihatkan bahwa mayoritas data didominasi oleh kelas sentimen netral, dengan persentase 50,92%, sementara itu, proporsi sentimen positif dan negatif tercatat hanya mencapai 25,07% dan 24,01%. Untuk memperoleh model terbaik, penelitian tersebut membandingkan empat algoritma klasifikasi yang bervariasi, meliputi *Random Forest*, SVM, *Naïve Bayes*, serta KNN. Dari hasil evaluasi yang dilakukan, SVM mencatat akurasi terbaik, yaitu sebesar 89%, melampaui *Random Forest* yang memperoleh 82%, *Naïve Bayes* sebesar 74%, dan KNN sebesar 65%. Hasil tersebut memperlihatkan bahwa algoritma SVM memiliki kapabilitas yang sangat andal dalam mengategorikan arah sentimen dari data *tweet* yang bersumber dari platform X (Farhan & Puspasari, 2025).

*Support Vector Machine* (SVM) juga dimanfaatkan pada studi yang dilakukan oleh Ferdiyawan dkk. (2025) untuk mengkaji sentimen publik pada isu pemain diaspora Tim Nasional Indonesia yang berkembang di platform X (Twitter),

khususnya terkait Mees Hilgers dan Eliano Reijnders. Penelitian tersebut menerapkan metode SEMMA dengan memanfaatkan 917 *tweet* yang sudah melewati proses seleksi dan *preprocessing*. Untuk proses pemodelan, Implementasi algoritma SVM ini memanfaatkan fungsi kernel *Radial Basis Function* (RBF), di mana pengujian dilakukan dengan menyetel nilai konstanta hiperparameter  $C$  pada angka 100 dan parameter *gamma* pada angka 0,1. Model berhasil mencatat tingkat akurasi sebesar 86% berdasarkan hasil pengujian, lebih baik dibandingkan menggunakan *Naïve*. Selain itu, penelitian tersebut menemukan bahwa sentimen positif mendominasi data dengan persentase 71,2%, yang mencerminkan tingginya dukungan dan harapan masyarakat terhadap perkembangan prestasi sepak bola Indonesia (Ferdiyawan *et al.*, 2025).

Fauzan, Anraeni, dan Imawan (2025) melakukan penelitian mengenai komentar publik pada Timnas sepak bola Indonesia pada masa kepelatihan Shin Tae-Yong memanfaatkan data postingan dari platform X. Tahapan klasifikasi dijalankan menggunakan pendekatan model *Support Vector Machine* (SVM) yang bertujuan mengkategorikan opini pada menjadi dua kelompok sentimen, berupa sentimen positif dan negatif. Hasil klasifikasi data mengungkapkan bahwa kebanyakan sampel merupakan kelas kategori positif, dengan jumlah 200 data (51,15%). Kontras dengan hal tersebut, sampel yang merepresentasikan sentimen negatif tercatat sebanyak 191 data dengan persentase 48,85%. Model SVM yang dikembangkan memperoleh akurasi sebesar 63,29%, dengan nilai *precision* dan *recall* masing-masing sebesar 65%. Temuan tersebut menunjukkan bahwa kebanyakan masyarakat memberikan tanggapan yang cenderung positif terhadap

Timnas ketika di latih *coach* STY, meskipun sejumlah kebijakan dan keputusan yang diambil masih memunculkan perbedaan pandangan di kalangan netizen (Fauzan *et al.*, 2025).

Pada studi yang dilaksanakan oleh Imaddudin dkk. (2025), melakukan penelitian untuk menganalisis pandangan masyarakat terhadap PSSI pada masa kepemimpinan Erick Thohir melalui data yang diperoleh dari platform X. Sebanyak 3.900 *tweet* dianalisis menggunakan algoritma SVM dengan pendekatan CRISP-DM. Konfigurasi terbaik didapatkan pada proporsi data 80:20 yang memperoleh *accuracy* sebanyak 76,6% dengan menerapkan kernel linear. Di samping itu, sebaran data menunjukkan bahwa kategori sentimen positif tampil dominan dengan memegang persentase sebesar 66%, sedangkan sentimen netral dan negatif masing-masing tercatat sebesar 27% dan 7,1%.. Hasil dari penelitian menandakan bahwa respons masyarakat terhadap kebijakan PSSI selama periode kepemimpinan Erick Thohir cenderung positif (Imaddudin *et al.*, 2025).

Dalam penelitiannya, Malik dkk. (2024) menerapkan algoritma SVM untuk menganalisis tanggapan publik pada penampilan Timnas Indonesia pada kompetisi Piala Asia U-23 AFC 2024 berdasarkan komentar pengguna YouTube. Sebanyak 912 komentar digunakan sebagai dataset penelitian dan dievaluasi melalui beberapa konfigurasi pembagian data. Hasil temuan memperlihatkan bahwa pengujian pertama mendapat hasil akurasi diangka 65,41% pada rasio pembagian data 60:40. Adapun model kedua memperoleh akurasi 63,76%, namun memperoleh nilai *precision* dan *recall* lebih tinggi. Temuan tersebut memperlihatkan bahwa SVM mempunyai kemampuan cukup baik dalam menjalankan klasifikasi teks pada data

komentar, meskipun dihadapkan pada distribusi kelas yang tidak merata sebagaimana karakteristik umum data komentar YouTube (Pangestu *et al.*, 2024).

Tabel 2.1 Perbandingan Penelitian yang Relevan

| No | Sumber                            | Judul                                                                                                                                                        | Metode                                               | Akurasi                                                                                                                                |
|----|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| 1  | (Dalifah <i>et al.</i> , 2024)    | Analisis Data Sentimen Negatif pada Opini Pengguna Twitter terhadap Berita Sepak Bola Liga 1 Tahun 2022 dengan Penerapan <i>Support Vector Machine</i>       | <i>Support Vector Machine</i> (SVM)                  | Tingkat akurasi maksimal (91%) diperoleh melalui skenario data 70:30, didukung oleh dominasi sentimen negatif sebesar 89,19%.          |
| 2  | (Apriyani <i>et al.</i> , 2025)   | Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi TikTok Menggunakan <i>Algoritma Machine Learning</i>                     | KNN, SVM, <i>Random Forest</i>                       | Performa tertinggi dibuktikan oleh <i>Random Forest</i> (97,30%), sementara akurasi paling rendah dihasilkan oleh model SVM (58,77%).  |
| 3  | (Farhan & Puspasari, 2025)        | Analisis Sentimen Supporter Sriwijaya FC Berbasis <i>Machine Learning</i>                                                                                    | SVM, <i>Random Forest</i> , <i>Naive Bayes</i> , KNN | SVM meraih akurasi tertinggi sebesar 89%. Sentimen Netral mendominasi hasil klasifikasi sebesar 50,92%.                                |
| 4  | (Ferdiyawan <i>et al.</i> , 2025) | Analisis Sentimen pada Media Sosial Twitter(X) terhadap Pemain Diaspora di Tim Nasional Sepak Bola Indonesia (Studi Kasus Mees Hilgers dan Eliano Reijnders) | <i>Naive Bayes</i> , SVM, SEMMA                      | SVM (akurasi 86%) memiliki performa lebih baik dibanding <i>Naive Bayes</i> (74%). Mayoritas sentimen publik bersifat Positif (71,2%). |
| 5  | (Fauzan <i>et al.</i> , 2025)     | Analisis Sentimen Tweet Netizen Terhadap Timnas Sepak Bola Indonesia di Era Shin Tae-Yong Menggunakan                                                        | <i>Support Vector Machine</i> (SVM)                  | Mencapai nilai <i>accuracy</i> 63,29%. Sentimen Positif (51,15%) sedikit lebih unggul                                                  |

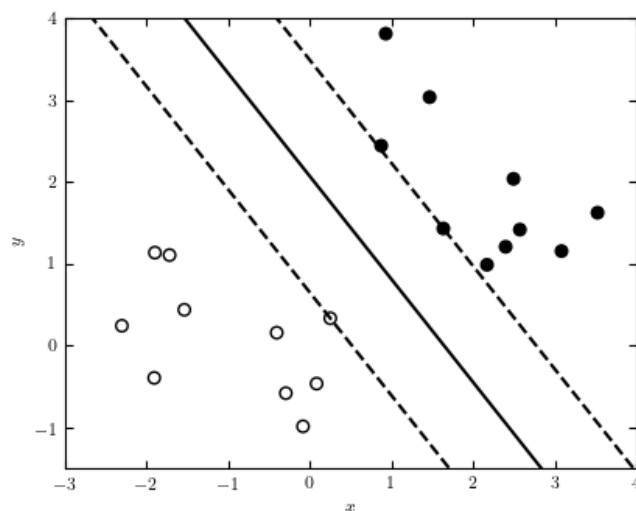
| No | Sumber                           | Judul                                                                                                                                                              | Metode                              | Akurasi                                                                                                                 |
|----|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
|    |                                  | Algoritma <i>Support Vector Machine</i> (SVM)                                                                                                                      |                                     | dibanding Negatif (48,85%).                                                                                             |
| 6  | (Imaddudin <i>et al.</i> , 2025) | Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan <i>Algoritma Support Vector Machine</i> (SVM) pada Media Sosial X                          | SVM (Kernel Linear & RBF), CRISP-DM | Akurasi terbaik sebesar 76,6% diperoleh menggunakan kernel Linear. Sentimen Positif mendominasi percakapan sebesar 66%. |
| 7  | (Pangestu <i>et al.</i> , 2024)  | Analisis Sentimen Hasil Pertandingan Sepakbola Timnas Indonesia di Piala Asia U-23 pada Platform Youtube menggunakan Algoritma <i>Support Vector Machine</i> (SVM) | <i>Support Vector Machine</i> (SVM) | <i>Accuracy</i> terbaik yang dicapai adalah 65,41% pada skenario penggunaan 40% data uji.                               |

Data literatur pada Tabel 2.1 menunjukkan bahwa integrasi skema pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) dengan algoritma *Support Vector Machine* (SVM) merupakan pendekatan yang sangat sering digunakan untuk melakukan klasifikasi opini atau analisis sentimen berbasis teks dari platform sosial media. Pembobotan TF-IDF ini membantu mengonversi data tekstual menjadi representasi matriks numerik berdasarkan tingkat kepentingan atau urgensi suatu kata di seluruh kumpulan dokumen. Di sisi lain, penggunaan SVM sebagai pengklasifikasi didasari oleh kapabilitasnya dalam mengelola data multidimensi serta konsistensi performanya di berbagai riset terdahulu. Berlandaskan efisiensi tersebut, studi ini mengimplementasikan model SVM untuk mengategorikan opini masyarakat mengenai dinamika sepak bola

nasional dengan memanfaatkan dataset yang diekstrak dari platform X. Pendekatan tersebut diharapkan mampu menghasilkan proses klasifikasi yang akurat dan konsisten dalam mengidentifikasi sentimen publik.

## 2.2 *Support Vector Machine* untuk Klasifikasi Teks

Sebagai algoritma *supervised learning*, *Support Vector Machine* (SVM) memiliki efektivitas tinggi dalam klasifikasi data, termasuk data teks. Prinsip kerja SVM adalah menemukan *hyperplane* optimal sebuah batas pemisah berbentuk garis atau bidang multidimensi yang membagi antar-kelas dengan jarak (*margin*) maksimal. Letak *hyperplane* ini ditentukan secara langsung oleh *support vectors*, yakni sampel data yang posisinya paling rapat dengan garis pembatas batas kelas (Fahrezi *et al.*, 2024).



Gambar 2.1 Ilustrasi *Support Vector Machine* (Sumber: astroml.org)

Visualisasi tersebut menggambarkan mekanisme *Support Vector Machine* (SVM) untuk memisahkan data sentimen teks ke beberapa jenis memanfaatkan *hyperplane*. Objek data disimbolkan oleh setiap titik, sementara batas pemisah

antarkelas ditunjukkan oleh garis *hyperplane* yang terbentuk. Posisi dan sudut kemiringan bidang pembatas ini sangat dipengaruhi yakni sampel data yang letaknya paling dekat dengan batas pemisah antar-kelompok. Berdasarkan titik-titik krusial tersebut, SVM akan menentukan *hyperplane* dengan *margin* terbesar guna memaksimalkan jarak antara pembatas dan data terdekat. (Husada & Paramita, 2021).

Pendekatan *Support Vector Machine* (SVM) menerapkan fungsi kernel guna mentransformasikan data ke ruang dimensi yang lebih tinggi. Proses proyeksi ini bertujuan untuk memfasilitasi pencarian *hyperplane* pemisah yang ideal pada karakteristik dataset yang tidak linear. Penentuan tipe kernel yang akan diterapkan—baik linear maupun *Radial Basis Function* (RBF)—disesuaikan secara spesifik berdasarkan sifat serta karakteristik distribusi data dari dataset yang dianalisis. Dalam eksperimen ini, SVM digunakan untuk kebutuhan klasifikasi dua kelas (biner), yakni kelas positif dan negatif. Mengoptimalkan jarak pembatas (*margin*) pada *hyperplane*, metode ini mampu mengklasifikasikan data teks sentimen secara akurat meskipun distribusinya bersifat kompleks.

## BAB III

### DESAIN DAN IMPLEMENTASI SISTEM

#### 3.1 Pengumpulan Data

Pengumpulan data penelitian ini memanfaatkan metode *scrapping* atau *crawling* langsung dari platform X. Dalam proses pengambilan data tersebut, perangkat lunak Tweet Harvest digunakan sebagai instrumen utamanya. Autentikasi untuk masuk ke sistem pencarian platform secara legal dilakukan dengan memanfaatkan *Auth Token* milik akun X. Kredensial tersebut bertindak sebagai kunci akses utama dalam seluruh rangkaian pengambilan data. Tweet Harvest bekerja dengan cara melakukan pencarian (*searching*) otomatis berdasarkan kata kunci (*keywords*) yang telah ditentukan untuk menangkap dinamika opini publik terhadap isu-isu spesifik dalam sepak bola nasional. Skenario pengumpulan data ini dilakukan berdasarkan enam kategori isu utama dengan rincian kata kunci dan rentang waktu pada tabel berikut:

Tabel 3.1 Pengumpulan Data

| No | Isu/Permasalahan                               | Kata Kunci         | Rentang Waktu         |
|----|------------------------------------------------|--------------------|-----------------------|
| 1  | Pemecatan Pelatih Shin Tae-Yong                | "STY"              | 5 -13 Januari 2025    |
| 2  | Performa Timnas U-23 di Kualifikasi Piala Asia | "Timnas U23"       | 3 -9 September 2025   |
| 3  | Tidak Lolosnya Timnas Senior ke Piala Dunia    | "Timnas Indonesia" | 9 -16 Oktober 2025    |
| 4  | Performa Timnas U-23 di SEA Games              | "Timnas U23"       | 8 -15 Desember 2025   |
| 5  | Naiknya Erick Thohir menjadi Kepala Kemenpora  | "Erick Thohir"     | 15 -21 September 2025 |

| No | Isu/Permasalahan                                         | Kata Kunci   | Rentang Waktu        |
|----|----------------------------------------------------------|--------------|----------------------|
| 6  | Tidak Adanya FIFA <i>Matchday</i> di Bulan November 2025 | "Timnas FMD" | 11 -21 November 2025 |

Berdasarkan parameter pencarian yang sudah ditetapkan pada Tabel 3.1, Rangkaian proses *crawling* yang dilakukan berhasil mengumpulkan dataset berupa *tweet* yang terbagi ke dalam enam fokus isu.. Setiap isu menghasilkan jumlah data yang berbeda, yang dipengaruhi oleh tingkat intensitas pembahasan publik, volume interaksi pengguna, serta dinamika perhatian masyarakat terhadap masing-masing isu pada platform sosial media X. Perbedaan jumlah data ini memperlihatkan bahwa respons publik terhadap setiap isu tidak bersifat seragam. Distribusi volume data yang diperoleh untuk masing-masing topik pembahasan dipaparkan secara rinci oleh Tabel berikut.

Tabel 3.2 Jumlah Data

| No         | Isu/Permasalahan                                  | Jumlah Data |
|------------|---------------------------------------------------|-------------|
| 1          | Pemecatan Pelatih Shin Tae-Yong                   | 649         |
| 2          | Performa Timnas U-23 di Kualifikasi Piala Asia    | 138         |
| 3          | Tidak Lolosnya Timnas Senior ke Piala Dunia       | 500         |
| 4          | Performa Timnas U-23 di SEA Games                 | 140         |
| 5          | Naiknya Erick Thohir menjadi Kepala Kemenpora     | 23          |
| 6          | Tidak Adanya FIFA Matchday di Bulan November 2025 | 37          |
| Total Data |                                                   | 1487        |

Berdasarkan sebaran jumlah data pada setiap isu, seluruh data hasil *crawling* kemudian dikompilasi menjadi satu himpunan dataset yang merepresentasikan berbagai opini publik terhadap dinamika sepak bola Indonesia. Dataset yang diperoleh masih berbentuk teks mentah yang memuat beragam bentuk ekspresi

bahasa alami, seperti opini, kritik, dukungan, singkatan, serta ragam bahasa nonformal yang lazim dijumpai pada platform media sosial X. Sebelum memasuki tahapan pelabelan dan pengolahan lebih lanjut, diperlukan gambaran awal mengenai karakteristik data yang dikumpulkan. Oleh karena itu, untuk menunjukkan bentuk dan struktur data mentah hasil *crawling*, beberapa contoh sampel data ditunjukkan oleh Tabel 3.3 berikut.

Tabel 3.3 Sampel Data

| No | <i>full_text</i>                                                                                                                                                                                                                                                              |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | @SiaranBolaLive Ngga semangat ngga ada STY STY di pecat badan terasa lemas                                                                                                                                                                                                    |
| 2  | @idextratime kenapa pemain timnas senior yang masih u-23 gak ada yg ditarik kesini selain struick...udah numpuk di bbrp posisi                                                                                                                                                |
| 3  | Seharusnya Timnas U22 Seagames jadi pelipur lara prestasi Timnas tahun ini eh malah nambah kesedihan                                                                                                                                                                          |
| 4  | Terlepas kegagalan Timnas U-23 ke Piala Asia U-23 edisi kali ini semoga saja Timnas U-23 bisa lebih maksimal dalam mencoba pemain terutama pemain kelahiran tahun 2003-2004 berstatus non Timnas U-18/19/20 tahun 2021-2023 ke Skuad Timnas U-23 di Sepakbola Sea Games 2025. |
| 5  | Yang bikin kesel dari galolos ini sebenarnya di gaada nya fmd timnas senior dan memilih friendly match buat timnas u22 fokus seagames dan hasilnya bapuk anying                                                                                                               |

Setelah data-data diperoleh, langkah berikutnya difokuskan tahap pelabelan yang dimaksudkan untuk menentukan polaritas atau kategori sentimen dari setiap *tweet* yang sudah dikumpulkan. Guna mengidentifikasi respons emosional masyarakat terhadap beragam isu sepak bola nasional yang sedang diteliti, data dikelompokkan ke dalam dua kelas utama, yakni kelas positif dan negatif. Sebagai bagian dari metode *supervised learning*, fase pemberian label pada data merupakan tahapan yang sangat krusial. Data yang telah terlabel ini berfungsi sebagai data latih

bagi *Support Vector Machine* (SVM) untuk mengekstraksi karakteristik dan mengidentifikasi pola sentimen publik dengan lebih sistematis. Dengan adanya label yang valid, model dapat mengevaluasi performa klasifikasi melalui metrik akurasi, presisi, dan *recall* untuk memastikan hasil analisis yang objektif.

Proses pelabelan data secara manual pada penelitian ini dilakukan oleh para praktisi ahli di bidang bahasa Indonesia. Validasi teks tersebut dipercayakan kepada Ibu Anggraeny Budi Utami, S.Pd. selaku pengajar dari SMPN 1 Baureno, bersama Bapak Dwi Musthofa, S.Pd. yang merupakan tenaga pendidik di SMP Negeri 1 Kedewan. Setiap *tweet* ditinjau dengan cermat untuk dikategorikan ke dalam salah satu dari tiga label sentimen terhadap berbagai dinamika dan isu sepak bola Indonesia yang menjadi objek penelitian. Setelah seluruh data selesai dilabeli oleh kedua ahli, hasil pelabelan tersebut kemudian digabungkan menjadi satu dataset dan dilakukan proses validasi ulang oleh Ibu Anggraeny Budi Utami, S.Pd. guna memastikan konsistensi dan akurasi label. Kategori yang digunakan dalam pelabelan adalah Positif untuk ekspresi yang mengandung dukungan, pujian, atau optimisme terhadap performa tim nasional maupun kebijakan federasi dan Negatif untuk ekspresi yang mengandung kritik tajam, ketidakpuasan, atau kekecewaan terhadap kinerja pelatih, manajemen PSSI, maupun hasil pertandingan.

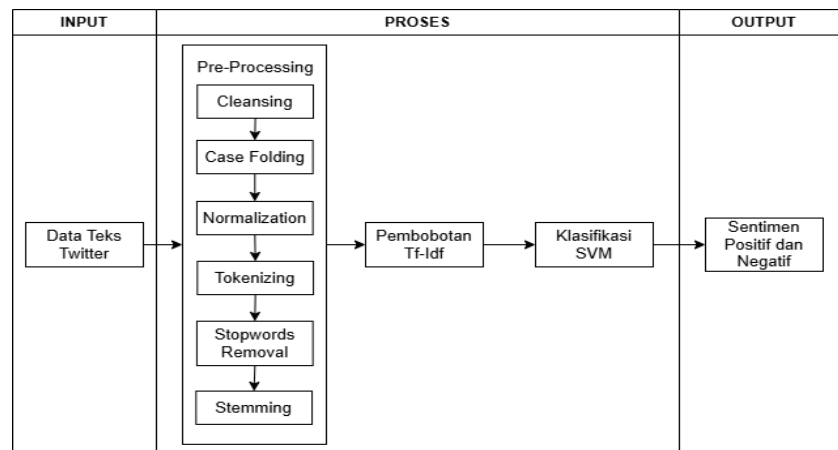
Tabel 3.4 Data Terlabel

| full_text                                                                                                                                                                            | Label   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| @SiaranBolaLive Ngga semangat ngga ada STY STY di pecat badan terasa lemas                                                                                                           | Negatif |
| Gak papa udah gak lolos pildun kali ini ya cuma bisa berdoa pemain yang bermain untuk timnas bisa up level terus kita bangga punya bang Jay Emil verdonk ole semua lah mereka hebat. | Positif |

| full_text                                                                                                                                                                                                                                                                     | Label   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Seharusnya Timnas U22 Seagames jadi pelipur lara prestasi Timnas tahun ini eh malah nambah kesedihan                                                                                                                                                                          | Negatif |
| Terlepas kegagalan Timnas U-23 ke Piala Asia U-23 edisi kali ini semoga saja Timnas U-23 bisa lebih maksimal dalam mencoba pemain terutama pemain kelahiran tahun 2003-2004 berstatus non Timnas U-18/19/20 tahun 2021-2023 ke Skuat Timnas U-23 di Sepakbola Sea Games 2025. | Positif |
| Yang bikin kesel dari galolos ini sebenarnya di gaada nya fmd timnas senior dan memilih friendly match buat timnas u22 fokus seagames dan hasilnya bapuk anying                                                                                                               | Negatif |

### 3.2 Desain Sistem

Penelitian ini mengimplementasikan desain sistem berbasis model *Input-Proses-Output* guna membangun sistem klasifikasi sentimen yang valid. Struktur utama pada blok proses terbagi ke dalam tiga fase, yaitu fase *preprocessing*, pembobotan fitur TF-IDF, serta pemodelan menggunakan SVM. Fase pertama berfokus pada pengerjaan pra-pemrosesan teks secara sistematis guna menyaring data mentah menjadi data bersih. Data yang telah bersih tersebut kemudian dialihkan ke fase kedua untuk dihitung bobot kepentingannya melalui pendekatan TF-IDF. Akhir dari alur proses ini direalisasikan lewat pengklasifikasian dengan *Support Vector Machine* (SVM) demi menghasilkan label kelas yang sesuai. Melalui integrasi ketiga fase ini, data teks mentah yang semula tidak terstruktur bisa dikonversi menjadi angka numerik yang siap dikomputasi. Keseluruhan rangkaian proses ini dirancang sedemikian rupa agar sistem dapat melakukan prediksi kelas sentimen secara otomatis dan efisien. Output akhir dari pemodelan ini nantinya disajikan dalam bentuk visualisasi performa untuk mempermudah penarikan kesimpulan penelitian.



Gambar 3. 1 Desain Sistem

Proses klasifikasi dengan algoritma SVM dimulai ketika model menerima masukan berupa vektor numerik yang telah melewati tahapan *preprocessing* serta ekstraksi fitur TF-IDF. Model SVM selanjutnya melakukan fase pembelajaran dengan tujuan menemukan *hyperplane* optimal yang memisahkan kelas-kelas sentimen dengan jarak (*margin*) paling maksimal. Sepanjang proses pembelajaran tersebut, sistem mengidentifikasi *support vector*, yakni sampel data di sekitar batas pemisah yang krusial bagi pembentukan model. Begitu model selesai dikonstruksikan menggunakan data latih, sistem siap memprediksi dan mengkategorikan data baru ke dalam jenis sentimen yang sesuai. Penentuan label akhir ini ditentukan secara langsung oleh posisi koordinat data terhadap bidang pembatas (*hyperplane*), sehingga menghasilkan output prediksi kelas sentimen yang mencerminkan kecenderungan opini publik.

### 3.3 *Preprocessing Text*

Data mentah yang diperoleh melalui proses *crawling* pada platform digital biasanya berbentuk tidak terstruktur serta dipenuhi oleh komponen *noise*. seperti

simbol, singkatan, kesalahan ketik, serta elemen lain yang kurang sesuai untuk analisis sentimen. Maka dari itu, sebelum data dapat diolah lebih lanjut menggunakan metode TF-IDF dan SVM, diperlukan tahapan *Preprocessing* terhadap data teks. Sebagai tahapan awal, pra-pemrosesan melibatkan sekumpulan metode untuk menyaring, menyelaraskan, serta mengonversi data teks mentah ke dalam format terstruktur yang siap dianalisis. Langkah ini krusial untuk mengeliminasi gangguan (*noise*), menstandardisasi bentuk kata, sekaligus memangkas redundansi kosakata. Hal ini membuat proses komputasi pada fase berikutnya berjalan lebih efisien, sehingga dapat meningkatkan performa model klasifikasi dalam mencapai tingkat akurasi yang paling optimal (Fahrezi *et al.*, 2024). Langkah *preprocessing* dalam penelitian ini dijalankan secara sistematis melalui urutan fase berikut:

1. *Cleaning*: Tahapan pertama ini dikhususkan untuk mengeliminasi elemen atau karakter teks yang dianggap sebagai *noise* dan tidak memberikan kontribusi pada makna sentimen. Proses ini mencakup penghapusan tanda baca (seperti titik, koma, tanda tanya), angka, simbol khusus (seperti #, @), *emoticon*, URL/link, dan karakter non-alfanumerik lainnya.

Tabel 3.5 Tahap *Cleaning*

| Sebelum <i>Cleaning</i>                                                                                                                                    | Setelah <i>Cleaning</i>                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| @seniman_bola Aku mah support @erickthohir full. Pergantian pelatih wajar. Apalagi dlm sepakbola. Di semua pekerjaan sama Byk kekurangan STY jg so move on | Aku mah support full Pergantian pelatih wajar Apalagi dlm sepakbola Di semua pekerjaan sama Byk kekurangan STY jg so move on    |
| @mario_an19 @garudaballers @idextratime Pengelolaan dari akar rumput sejak dini disini jelek sekali pemain asli                                            | Pengelolaan dari akar rumput sejak dini disini jelek sekali pemain asli didikan lokal maen di timnas habis dihujat sama netizen |

| Sebelum <i>Cleaning</i>                                                                                   | Setelah <i>Cleaning</i>                                                                     |
|-----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| didikan lokal maen di timnas habis dihujaat sama netizen                                                  |                                                                                             |
| @kompascom Timnas memang jos...ayo solidkan lagi 2030 masuk Piala Dunia...                                | Timnas memang jos ayo solidkan lagi 2030 masuk Piala Dunia                                  |
| @idextratime Gelagat gak lolos semakin menguat ketika Erick lebih milih jadi Menpora daripada fokus PSSI. | Gelagat gak lolos semakin menguat ketika Erick lebih milih jadi Menpora daripada fokus PSSI |

2. *Case Folding*: Tahap ini bermaksud untuk menyesuaikan kapitalisasi huruf dalam seluruh teks *tweet*. Semua huruf dikonversi menjadi huruf kecil (*lowercase*). Hal ini penting agar kata yang sama namun ditulis dengan huruf kapital berbeda (misalnya, "MENANG" dan "menang") dianggap sebagai kata yang sama oleh sistem.

Tabel 3.6 Tahap *Case Folding*

| Sebelum <i>Case Folding</i>                                                                                                         | Setelah <i>Case Folding</i>                                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Aku mah support full Pergantian pelatih wajar Apalagi dlm sepakbola Di smua pekerjaan sama Byk kekurangan STY jg so move on         | aku mah support full pergantian pelatih wajar apalagi dlm sepakbola di smua pekerjaan sama byk kekurangan sty jg... |
| Pengelolaan dari akar rumput sejak dini disini jelek sekalinya pemain asli didikan lokal maen di timnas habis dihujaat sama netizen | pengelolaan dari akar rumput sejak dini disini jelek sekalinya pemain asli didikan lokal maen di timnas...          |
| Timnas memang jos ayo solidkan lagi 2030 masuk Piala Dunia                                                                          | timnas memang jos ayo solidkan lagi 2030 masuk piala dunia                                                          |
| Gelagat gak lolos semakin menguat ketika Erick lebih milih jadi Menpora daripada fokus PSSI                                         | gelagat gak lolos semakin menguat ketika erick lebih milih jadi menpora daripada fokus pssi                         |

3. *Normalization*: Langkah *normalization* diterapkan untuk menyelaraskan format penulisan kata agar menghasilkan representasi data yang seragam. Proses ini mencakup perbaikan kata tidak baku, singkatan, serta variasi penulisan yang umum ditemukan pada *tweet* media sosial, seperti “gk” menjadi “tidak” atau

“bgt” menjadi “banget”. Selain itu, *normalization* juga dapat menghilangkan pengulangan huruf berlebihan. Langkah ini berfungsi untuk meminimalkan keberagaman bentuk kata yang sebenarnya memiliki esensi makna serupa, sehingga meningkatkan kualitas data teks.

Tabel 3.7 Tahap *Normalization*

| Sebelum <i>Normalization</i>                                                                                        | Setelah <i>Normalization</i>                                                                                       |
|---------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| aku mah support full pergantian pelatih wajar apalagi dlm sepakbola di smua pekerjaan sama byk kekurangan sty jg... | aku mah support full pergantian pelatih wajar apalagi dalam sepakbola di semua pekerjaan sama banyak kekurangan... |
| pengelolaan dari akar rumput sejak dini disini jelek sekalinya pemain asli didikan lokal maen di timnas...          | pengelolaan dari akar rumput sejak dini di sini jelek sekalinya pemain asli didikan lokal main di timnas...        |
| timnas memang jos ayo solidkan lagi 2030 masuk piala dunia                                                          | timnas memang jos ayo solidkan lagi 2030 masuk piala dunia                                                         |
| gelagat gak lolos semakin menguat ketika erick lebih milih jadi menpora daripada fokus pssi                         | gelagat tidak lolos semakin menguat ketika erick lebih memilih jadi menpora daripada fokus pssi                    |

4. *Tokenizing*: Setelah teks dibersihkan dan diseragamkan hurufnya, fase berikutnya yaitu memisahkan kalimat berperan menjadi unit-unit kata individu yang dinamakan token. Pemisahan ini biasanya dilakukan berdasarkan spasi antar kata.

Tabel 3.8 Tahap *Tokenizing*

| Sebelum <i>Tokenizing</i>                                                                                          | Setelah <i>Tokenizing</i>                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| aku mah support full pergantian pelatih wajar apalagi dalam sepakbola di semua pekerjaan sama banyak kekurangan... | ['aku', 'mah', 'support', 'full', 'pergantian', 'pelatih', 'wajar', 'apalagi', 'dalam', 'sepakbola', 'di', 'semua', 'pekerjaan', 'sama', 'banyak', 'kekurangan'] |
| pengelolaan dari akar rumput sejak dini di sini jelek sekalinya pemain asli didikan lokal main di timnas...        | ['pengelolaan', 'dari', 'akar', 'rumput', 'sejak', 'dini', 'di', 'sini', 'jelek', 'sekalinya', 'pemain', 'asli', 'didikan', 'lokal', 'main', 'di', 'timnas']     |

| Sebelum <i>Tokenizing</i>                                                                             | Setelah <i>Tokenizing</i>                                                                                                                           |
|-------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| timnas memang jos ayo solidkan lagi<br>2030 masuk piala dunia                                         | ['timnas', 'memang', 'jos', 'ayo', 'solidkan',<br>'lagi', '2030', 'masuk', 'piala', 'dunia']                                                        |
| gelagat tidak lolos semakin menguat<br>ketika erick lebih memilih jadi menpora<br>daripada fokus pssi | ['gelagat', 'tidak', 'lolos', 'semakin',<br>'menguat', 'ketika', 'erick', 'lebih',<br>'memilih', 'jadi', 'menpora', 'daripada',<br>'fokus', 'pssi'] |

5. *Stopword Removal*: Langkah ini dimaksudkan untuk menyingkirkan kata umum (*stopwords*) yang sering muncul namun tidak memiliki makna sentimen signifikan. Kata-kata seperti "yang", "di", "dan", "adalah", "ke", "atau" akan dihapus berdasarkan daftar *stopword* Bahasa Indonesia yang telah ditentukan, proses *stopword removal* dilakukan menggunakan *library* NLTK.

Tabel 3.9 Tahap *Stopword Removal*

| Sebelum <i>Stopword Removal</i>                                                                                                                     | Setelah <i>Stopword Removal</i>                                                                                                        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| ['aku', 'mah', 'support', 'full', 'pergantian',<br>'pelatih', 'wajar', 'apalagi', 'dalam',<br>'sepakbola', 'di', 'semua', 'pekerjaan', ...]         | ['support', 'full', 'pergantian', 'pelatih',<br>'wajar', 'sepakbola', 'pekerjaan', 'banyak',<br>'kekurangan', 'sty', 'move']           |
| ['pengelolaan', 'dari', 'akar', 'rumpun',<br>'sejak', 'dini', 'di', 'sini', 'jelek', 'pemain',<br>'asli', 'didikan', 'lokal', ...]                  | ['pengelolaan', 'akar', 'rumpun', 'dini',<br>'jelek', 'pemain', 'asli', 'didikan', 'lokal',<br>'main', 'timnas', 'dihujat', 'netizen'] |
| ['timnas', 'memang', 'jos', 'ayo', 'solidkan',<br>'lagi', '2030', 'masuk', 'piala', 'dunia']                                                        | ['timnas', 'jos', 'solidkan', '2030', 'masuk',<br>'piala', 'dunia']                                                                    |
| ['gelagat', 'tidak', 'lolos', 'semakin',<br>'menguat', 'ketika', 'erick', 'lebih',<br>'memilih', 'jadi', 'menpora', 'daripada',<br>'fokus', 'pssi'] | ['gelagat', 'lolos', 'menguat', 'erick',<br>'memilih', 'menpora', 'fokus', 'pssi']                                                     |

6. *Stemming*: Fase terakhir yaitu mengganti kata-kata imbuhan menjadi bentuk kata dasarnya (*root word*). Misalnya, kata "keputusan", "memutuskan", dan "diputuskan" akan diubah menjadi kata dasar "putus". *Stemming* membantu mengurangi jumlah variasi kata unik dan mengelompokkan kata-kata yang

memiliki akar makna sama. Pada penelitian ini, proses stemming dilakukan menggunakan *library* Sastrawi yang menerapkan algoritma Nazief-Adriani untuk stemming Bahasa Indonesia (Sinaga & Nainggolan, 2023).

Tabel 3.10 Tahap *Stemming*

| Sebelum <i>Stemming</i>                                                                                                          | Setelah <i>Stemming</i>                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| ['support', 'full', 'pergantian', 'pelatih', 'wajar', 'sepakbola', 'pekerjaan', 'banyak', 'kekurangan', 'sty', 'move']           | ['support', 'full', 'ganti', 'latih', 'wajar', 'sepakbola', 'kerja', 'banyak', 'kurang', 'sty', 'move']               |
| ['pengelolaan', 'akar', 'rumput', 'dini', 'jelek', 'pemain', 'asli', 'didikan', 'lokal', 'main', 'timnas', 'dihujat', 'netizen'] | ['kelola', 'akar', 'rumput', 'dini', 'jelek', 'main', 'asli', 'didik', 'lokal', 'main', 'timnas', 'hujat', 'netizen'] |
| ['timnas', 'jos', 'solidkan', '2030', 'masuk', 'piala', 'dunia']                                                                 | ['timnas', 'jos', 'solid', '2030', 'masuk', 'piala', 'dunia']                                                         |
| ['gelagat', 'lolos', 'menguat', 'erick', 'memilih', 'menpora', 'fokus', 'pssi']                                                  | ['gelagat', 'lolos', 'kuat', 'erick', 'pilih', 'menpora', 'fokus', 'pssi']                                            |

### 3.4 Pembobotan Kata

Meskipun telah menempuh fase *preprocessing* hingga data bersih, data tweet berbentuk teks tidak langsung diolah oleh algoritma SVM. Karakteristik algoritma machine learning pada umumnya mensyaratkan input data bertipe numerik agar proses pemodelan dapat berjalan. Oleh sebab itu, diperlukan tahap *term weighting* atau pembobotan kata untuk mentransformasikan teks menjadi representasi vektor numerik. Melalui pendekatan ini, kata dalam dokumen akan dikonversi menjadi nilai kuantitatif yang merepresentasikan tingkat signifikansi atau kontribusinya. Penelitian ini memanfaatkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) mengingat kemampuannya untuk mengidentifikasi kata kunci signifikan melalui pertimbangan frekuensi kemunculan sebuah kata terhadap tingkat kelangkaannya di seluruh kumpulan komentar.

Guna menghindari timbulnya masalah kebocoran informasi (*data leakage*), implementasi metode TF-IDF dilakukan secara sistematis melalui pemisahan prosedur komputasi antara dataset pelatihan dan dataset pengujian. Pada jalur pelatihan, dilakukan proses *fit* dan *transform* secara bersamaan untuk membangun representasi fitur serta menghitung nilai *Inverse Document Frequency* (IDF) berdasarkan distribusi data latih. Sebaliknya, pada jalur pengujian, data teks hanya melewati proses *transform* menggunakan model pembobotan yang telah terbentuk sebelumnya tanpa melakukan perhitungan ulang terhadap distribusi fitur. Pendekatan ini menegaskan bahwa data uji diberlakukan sebagai informasi baru yang bersifat independen, sehingga evaluasi performa model yang dihasilkan benar-benar objektif dan valid sesuai dengan kaidah penelitian ilmiah.

#### **3.4.1 Nilai *Term-Frequency* (TF)**

Komponen *Term Frequency* (TF) pada pembobotan ini dapat dimanfaatkan untuk menghitung intensitas atau jumlah kemunculan sebuah kata di pada satu teks (*tweet*). Perhitungan ini adalah cara sederhana untuk mengidentifikasi kata-kata yang mungkin menjadi fokus utama atau topik dalam sebuah teks spesifik.

Fase awal perhitungan metrik *Term Frequency* (TF) dilakukan berdasarkan pengukuran intensitas suatu kosakata yang muncul pada satu kesatuan teks (*tweet*). Angka kemunculan ini kemudian dinormalisasi terhadap jumlah keseluruhan kata yang menyusun *tweet* tersebut. Langkah normalisasi ini sangat penting dilakukan agar *tweet* dengan volume kata yang lebih banyak tidak mendapatkan keuntungan bobot yang lebih tinggi secara sepihak akibat panjangnya teks.

Penggunaan TF didasarkan pada asumsi bahwa kata yang sering diulang dalam sebuah *tweet* cenderung lebih penting untuk menggambarkan isi atau sentimen komentar tersebut. Akan tetapi, pemrosesan yang hanya mengandalkan aspek TF belum cukup memadai karena kosakata yang bersifat umum—misalnya kata hubung "yang" atau "dan"—sering memiliki kuantitas kemunculan yang dominan tanpa memberikan kontribusi makna yang signifikan. Oleh sebab itu, nilai TF harus diintegrasikan dengan pendekatan IDF. Rumus perhitungan *Term Frequency* (TF) ditunjukkan sebagai berikut:

$$tf_{t,d} = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah term } t \text{ dalam dokumen } d} \quad (3.1)$$

Keterangan :

- $tf_{t,d}$  : Nilai bobot kemunculan kata (*term*)  $t$  di dalam dokumen  $d$ .

### 3.4.2 Nilai *Inverse Document Frequency* (IDF)

Sebagai elemen kedua pada skema pembobotan ini, *Inverse Document Frequency* (IDF) berguna untuk melengkapi perhitungan frekuensi kata. Berbeda dari metrik TF yang analisisnya bersifat lokal pada satu dokumen, IDF mengukur bobot informatif dan kekhasan sebuah kata dalam cakupan keseluruhan dataset. Korpus dokumen dalam studi ini dibangun dari data teks *tweet* yang didapat dari platform X. Implementasi fungsi IDF ini krusial untuk menormalisasi dan menyeimbangkan kalkulasi bobot yang telah dihasilkan oleh tahap *Term Frequency*.

Secara teknis, IDF mengevaluasi bobot berdasarkan invers proporsi jumlah dokumen yang memuat kata tertentu. Kata-kata yang jamak ditemukan pada

mayoritas *tweet* dipandang tidak memiliki nilai diskriminatif yang cukup untuk membedakan isi dokumen, sehingga menghasilkan nilai IDF yang rendah. Di sisi lain, kata-kata yang eksklusif muncul pada sedikit *tweet* dinilai lebih representatif dalam menunjukkan kekhasan suatu dokumen, sehingga berhak mendapatkan bobot IDF yang jauh lebih tinggi.

Faktor penting dari penerapan IDF adalah perannya dalam meminimalkan pengaruh istilah-istilah yang mendominasi secara kuantitas pada banyak arsip data tanpa memberikan kontribusi makna yang berarti. Penggabungan TF dan IDF ini memastikan bahwa istilah yang dominan pada data teks tertentu, namun memiliki frekuensi rendah pada kumpulan *dataset* keseluruhan, akan mendapatkan skor fitur yang lebih tinggi dan dianggap lebih representatif. Dengan demikian, metode TF-IDF mampu menekankan kosakata yang paling representatif. Rumus perhitungan *Inverse Document Frequency* (IDF) untuk sebuah *term*  $t$  pada koleksi dokumen  $d$  adalah sebagai berikut:

$$idf_t = \log \left( \frac{N}{df_t} \right) \quad (3.2)$$

Keterangan :

- $idf_t$  : Nilai bobot *Inverse Document Frequency* dari suatu kata  $t$ .
- $N$  : Ukuran total dari keseluruhan korpus atau koleksi dokumen.
- $df_t$  : Frekuensi dokumen yang memuat kata  $t$  di dalamnya.

### 3.4.3 Perhitungan *Term Frequency-Inverse Document Frequency*

Tahap finalisasi pada pembobotan diselesaikan melalui perhitungan integratif TF-IDF (*Term Frequency-Inverse Document Frequency*). Prosedur ini dirancang untuk mengekstrak nilai bobot numerik akhir bagi tiap istilah (*term*) pada

tiap dokumen teks. Nilai yang diperoleh menjadi indikator tingkat kepentingan relatif suatu kata, yang analisisnya mengombinasikan frekuensi kemunculan dalam satu *tweet* tertentu beserta pola sebarannya di seluruh kumpulan data *tweet*.

Metode TF-IDF menutupi kelemahan inheren dari penggunaan tunggal komponen *Term Frequency* maupun *Inverse Document Frequency*. Secara parsial, kalkulasi TF rentan memprioritaskan kosakata fungsional yang mendominasi secara kuantitas, sementara kalkulasi IDF tidak mengakomodasi volume kata dalam dokumen tertentu. Dengan mengombinasikan kedua variabel ini, nilai TF-IDF yang dihasilkan mampu merelevansikan kedua dimensi penilaian tersebut secara lebih proporsional.

Penilaian final dalam metode TF-IDF ditentukan dengan menghitung matriks *Term Frequency* (indikator intensitas kata pada sebuah *tweet*) dengan nilai *Inverse Document Frequency* (indikator eksklusifitas kata dalam seluruh kumpulan data). Hasil komputasi dari perkalian ini memberikan bobot yang signifikan pada istilah yang memiliki frekuensi tinggi di suatu dokumen spesifik, tetapi cenderung langka dalam skala keseluruhan data, sehingga mampu mengidentifikasi kata kunci yang paling representatif secara optimal. Persamaan nilai TF-IDF adalah sebagai berikut:

$$tfidf_{t,d} = tf_{t,d} \times idf_t \quad (3.3)$$

Keterangan :

- $tfidf_{t,d}$  : Nilai bobot akhir TF-IDF dari kata  $t$  pada dokumen  $d$ .

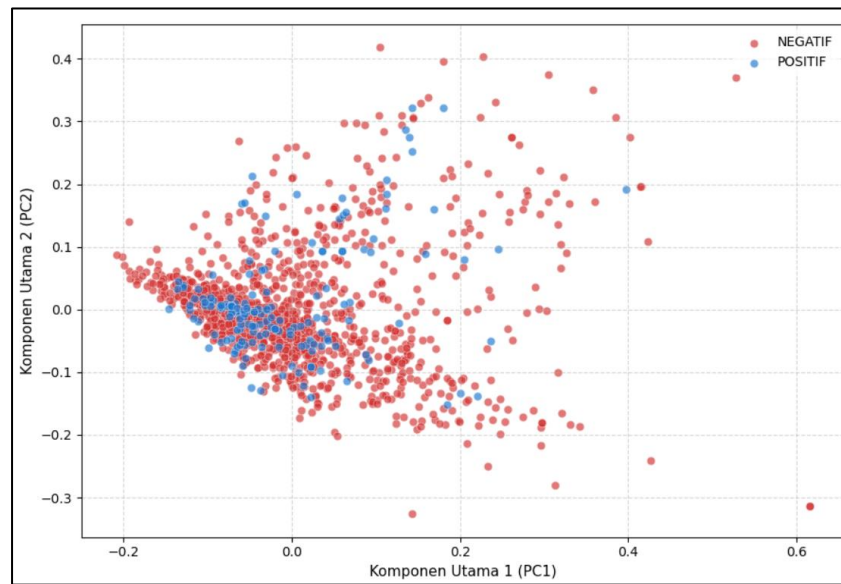
### 3.5 Support Vector Machine (SVM)

Setelah representasi teks berhasil dikonversi menjadi bentuk matriks atau vektor angka menggunakan metode TF-IDF, tahapan lanjutan yang dilakukan adalah mengklasifikasikan orientasi sentimen dari data hasil pemrosesan. Pada eksperimen ini. Pada penelitian ini, klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) guna memisahkan polaritas sentimen positif dan negatif. Prinsip kerja SVM berpusat pada pencarian bidang pemisah (*hyperplane*) paling optimal yang dapat membagi data dari kedua kelas tersebut secara maksimal di dalam ruang fitur (Rabbani *et al.*, 2023). *Hyperplane* ini dipilih dengan mempertimbangkan nilai *margin* maksimum, yang merepresentasikan jarak geometris antara *hyperplane* dengan titik data terdekat dari masing-masing kelas, di mana titik-titik tersebut dikenal sebagai *support vector*. Dengan memaksimalkan *margin*, diharapkan model dapat meningkatkan performa generalisasi model secara signifikan. Dengan demikian, sistem akan lebih andal dan presisi saat mengeksekusi proses klasifikasi terhadap data yang belum pernah dikenali sebelumnya. Setelah proses pelatihan selesai, model SVM akan menghasilkan fungsi *hyperplane* yang digunakan sebagai batas keputusan (*decision boundary*) dalam menentukan kelas sentimen suatu data. Secara matematis, fungsi *hyperplane* pada SVM dapat dinyatakan sebagai berikut:

$$f(x) = w^T x + b \quad (3.4)$$

Dengan :

- $f(x)$  = skor keputusan (*decision function*).
- $w$  = vektor bobot (*weight vector*).
- $x$  = representasi fitur hasil pembobotan TF-IDF.
- $b$  = bias atau *intercept*.



Gambar 3.2 Visualisasi Persebaran Awal Data

Untuk memahami karakteristik data yang digunakan dalam proses klasifikasi, dilakukan visualisasi persebaran data sebagaimana ditunjukkan pada Gambar 3.2 untuk mengetahui karakteristik persebaran data. Berdasarkan visualisasi tersebut, terlihat bahwa data sentimen positif dan negatif memiliki tingkat tumpang tindih (*overlapping*) yang cukup tinggi pada ruang fitur hasil reduksi dimensi. Keadaan tersebut memberikan petunjuk bahwa struktur persebaran data tidak bersifat *linearly separable*. Oleh karena itu, pemisahan kelas data tidak dapat dicapai secara optimal apabila hanya menggunakan satu *hyperplane* atau garis lurus konvensional. Adanya sebaran data yang saling beririsan mengindikasikan bahwa hubungan antar kelas bersifat non-linear sehingga memerlukan pendekatan yang mampu membentuk batas keputusan yang lebih kompleks. Untuk itu, algoritma *Support Vector Machine* pada penelitian ini diimplementasikan dengan menggunakan kernel *Radial Basis Function* (RBF),

yang berfungsi mentransformasikan data ke ruang fitur berdimensi lebih tinggi. Transformasi ini bertujuan agar terbentuk hyperplane yang lebih optimal sehingga sentimen positif dan negatif dapat dipisahkan dengan lebih baik. Fungsi matematis kernel RBF yang diterapkan dalam penelitian ini dapat dilihat sebagai berikut:

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (3.5)$$

Keterangan :

- $K(x_i, x_j)$  : Nilai kemiripan antar data.
- $\gamma$  (gamma) : Pengatur jangkauan pengaruh titik data.
- $|x_i - x_j|^2$  : Jarak kuadrat antara dua vektor fitur.

Kernel Radial Basis Function (RBF) adalah jenis kernel yang paling sering digunakan dalam SVM untuk menangani data yang bersifat non-linear. Cara kerja kernel ini adalah dengan menghitung kedekatan atau jarak geometris antara dua titik data pada ruang fitur. Apabila jarak antar titik data tersebut kecil, maka nilai kernel akan cenderung mendekati 1, sebaliknya jika jaraknya besar, nilai kernel akan mendekati 0. Adapun parameter  $\gamma$  berperan dalam mengatur sejauh mana jarak tersebut memengaruhi hasil transformasi data. Melalui pemetaan data ke ruang berdimensi yang lebih tinggi, kernel RBF mampu menghasilkan batas keputusan (*decision boundary*) yang lebih fleksibel.

Setelah kernel *Radial Basis Function* (RBF) ditetapkan, langkah selanjutnya adalah melakukan optimasi *hyperparameter* dengan menerapkan metode *GridSearchCV* guna mendapatkan konfigurasi parameter yang menghasilkan performa klasifikasi paling baik. Cara kerja metode ini adalah dengan mencoba semua kombinasi nilai *hyperparameter* yang sudah ditetapkan secara sistematis, lalu mengevaluasi masing-masing kombinasi untuk menentukan konfigurasi yang

paling optimal (Maisat Eka Darmawan & Fauzan Dianta, 2023). *GridSearchCV* dipilih sebagai metode optimasi karena tidak hanya berfungsi mencari parameter terbaik, melainkan juga menyertakan teknik *cross-validation*, sehingga tiap kombinasi parameter diuji pada beberapa subset data latih yang berbeda-beda. Hal ini menyebabkan parameter yang akhirnya dipilih bukan hanya unggul pada satu skema pembagian data tertentu saja, tetapi juga menghasilkan model yang lebih stabil dan memiliki generalisasi yang lebih baik ketika diterapkan pada data baru (Toha et al., 2022).

Pada tahap ini, peneliti menerapkan proses optimasi terhadap algoritma SVM yang menggunakan kernel RBF. Pemilihan kernel ini didasarkan pada karakteristik data yang bersifat non-linear, sehingga tidak memungkinkan untuk dipisahkan menggunakan garis lurus sederhana. Dua *hyperparameter* utama yang menjadi fokus optimasi adalah Cost (C) dan gamma ( $\gamma$ ). Parameter C berperan dalam menyeimbangkan antara lebar margin klasifikasi dengan toleransi terhadap kesalahan, sementara gamma menentukan seberapa besar pengaruh suatu titik data terhadap pembentukan garis pemisah (*decision boundary*). Rincian kombinasi nilai C dan gamma yang diujikan melalui proses *GridSearchCV* dapat dilihat pada tabel di bawah ini.

Tabel 3.11 Nilai Parameter

| Kernel | Parameter | Nilai yang Diuji |
|--------|-----------|------------------|
| RBF    | C         | 0.01, 0.1, 1, 10 |
|        | $\gamma$  | 1, 0.1, 0.01     |

## BAB IV

### UJI COBA DAN PEMBAHASAN

Bab ini menyajikan secara menyeluruh hasil pengolahan data beserta pembahasan terkait opini publik yang diperoleh. Data yang telah melalui tahapan *preprocessing* dan ekstraksi fitur dengan metode TF-IDF selanjutnya diklasifikasikan ke dalam kategori sentimen positif dan negatif menggunakan model *Support Vector Machine*. Guna mengukur performa serta tingkat akurasi model yang telah dibangun, bab ini juga menampilkan matriks evaluasi berupa *Confusion Matrix* beserta sejumlah parameter pendukung, antara lain *accuracy*, *precision*, *recall*, *F1-score*, dan *F1-score Macro*.

#### 4.1 Persiapan Data

Dataset penelitian ini terdiri atas *tweet* dan *retweet* berbahasa Indonesia dari media sosial X yang mengulas berbagai isu sepak bola dalam negeri. Pengumpulan dokumen teks ini mengandalkan metode *crawling* dengan kata kunci yang parameternya telah dipaparkan pada sub-bab 3.1 Pengumpulan Data. Data mentah yang didapatkan selanjutnya diseleksi dan dikelompokkan ke dalam sentimen positif dan negatif melalui bantuan ahli bahasa sebagai acuan komputasi pada tahap klasifikasi. Untuk melihat bagaimana sebaran jumlah data pada setiap kelompok sentimen, detailnya disajikan dalam tabel dibawah ini.

Tabel 4.1 Jumlah Data

| No | Isu/Permasalahan                               | Jumlah Data Negatif | Jumlah Data Positif |
|----|------------------------------------------------|---------------------|---------------------|
| 1  | Pemecatan Pelatih Shin Tae-Yong                | 561                 | 88                  |
| 2  | Performa Timnas U-23 di Kualifikasi Piala Asia | 122                 | 16                  |

| No         | Isu/Permasalahan                                  | Jumlah Data Negatif | Jumlah Data Positif |
|------------|---------------------------------------------------|---------------------|---------------------|
| 3          | Tidak Lolosnya Timnas Senior ke Piala Dunia       | 459                 | 41                  |
| 4          | Performa Timnas U-23 di SEA Games                 | 137                 | 3                   |
| 5          | Naiknya Erick Thohir menjadi Kepala Kemenpora     | 15                  | 8                   |
| 6          | Tidak Adanya FIFA Matchday di Bulan November 2025 | 34                  | 3                   |
| Total Data |                                                   | 1329                | 159                 |

## 4.2 Skenario Uji Coba

Penyusunan skenario pengujian dilakukan untuk mengevaluasi performa model klasifikasi sentimen dalam penelitian ini, sekaligus untuk menentukan rasio pembagian data yang paling optimal. Pengujian tersebut menerapkan algoritma *Support Vector Machine* (SVM) berbasis kernel *Radial Basis Function* (RBF), sebuah pilihan yang didasarkan pada sifat data yang non-linear. Pengaruh proporsi data terhadap akurasi model diuji menggunakan tiga variasi rasio data latih dan data uji, yakni 90:10, 80:20, dan 70:30, sehingga menghasilkan tiga skenario pengujian yang berbeda. Pembagian data dilakukan dengan teknik *stratified sampling* agar distribusi kelas sentimen positif dan negatif pada data latih maupun data uji tetap mencerminkan proporsi pada dataset asli. Selain itu, proses ini disertai pengacakan terkontrol untuk menjamin konsistensi dan akuntabilitas hasil pengujian apabila direproduksi kembali. Rincian dari ketiga skenario tersebut dapat dilihat pada Tabel 4.2 berikut ini.

Tabel 4.2 Skenario Pengujian

| Skenario   | Rasio Data |
|------------|------------|
| Skenario 1 | 90:10      |

| Skenario   | Rasio Data |
|------------|------------|
| Skenario 2 | 80:20      |
| Skenario 3 | 70:30      |

Berdasarkan skenario pengujian yang telah ditetapkan, tiap model terlebih dahulu melewati fase pelatihan menggunakan data latih agar mampu memprediksi orientasi sentimen pada data uji. Komputasi performa model dievaluasi dengan membandingkan tingkat kecocokan antara hasil prediksi model dengan label yang sebenarnya. Melalui pemanfaatan *Confusion Matrix*, tingkat keandalan model dapat diukur secara kuantitatif berdasarkan parameter *accuracy*, *precision*, *recall*, *F1-score*, serta *F1-score Macro*. Penggunaan nilai *F1-score Macro* sengaja dipilih sebagai parameter kunci guna memberikan penilaian yang akurat terhadap kemampuan model saat memproses dataset dengan sebaran kelas yang tidak seimbang.

Tabel 4.3 *Confusion Matrix*

|            |            |            |
|------------|------------|------------|
|            | <b>pos</b> | <b>neg</b> |
| <b>pos</b> | TP         | FN         |
| <b>neg</b> | FP         | TN         |

*Confusion matrix* merupakan matriks evaluasi berbentuk tabel yang digunakan untuk membandingkan kesesuaian antara hasil prediksi model dengan label aktual pada data uji. Matriks ini mengelompokkan hasil pengujian ke dalam empat indikator performa utama, yaitu:

- *True Positive (TP)*: Total data berlabel positif yang secara akurat berhasil diklasifikasikan sebagai kelas positif oleh model.

- *True Negative* (TN): Total data berlabel negatif yang dengan tepat diidentifikasi sebagai kelas negatif oleh model.
- *False Positive* (FP): Total di mana data yang sebenarnya berlabel negatif, namun salah diprediksi sebagai kelas positif oleh model.
- *False Negative* (FN): Total di mana data yang sebenarnya berlabel positif, namun secara keliru dikategorikan sebagai kelas negatif oleh model.

Berdasarkan nilai-nilai dari *Confusion Matrix*, akan dihitung beberapa metrik evaluasi standar untuk menilai performa Model SVM, yaitu:

1. *Accuracy* : Mengukur proporsi jumlah prediksi yang benar terhadap keseluruhan data uji. Nilai *accuracy* menunjukkan tingkat ketepatan model dalam mengklasifikasikan sentimen positif dan negatif berdasarkan data yang diuji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

2. *Precision* : Mengukur proporsi data yang diprediksi sebagai suatu kelas dan benar-benar termasuk ke dalam kelas tersebut. Metrik ini menunjukkan tingkat ketepatan model dalam memberikan prediksi pada masing-masing kelas sentimen.

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

3. *Recall* : Mengukur proporsi data aktual dari suatu kelas yang berhasil diidentifikasi dengan benar oleh model. Metrik ini menunjukkan kemampuan model dalam mengenali seluruh data yang benar-benar termasuk ke dalam suatu kelas sentimen.

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

4. *F1-score* : Merupakan rata-rata harmonik dari nilai *Precision* dan *Recall*. Metrik ini digunakan untuk mengukur keseimbangan antara kemampuan model dalam memberikan prediksi yang tepat dan kemampuannya dalam mengenali seluruh data aktual. Nilai *F1-score* yang tinggi menunjukkan bahwa model memiliki tingkat *precision* dan *recall* yang sama-sama baik.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4.4)$$

5. *Macro F1-score* : Merupakan metrik evaluasi yang diperoleh dengan menghitung rata-rata nilai *F1-score* dari setiap kelas tanpa mempertimbangkan jumlah data pada masing-masing kelas. Metrik ini digunakan untuk memberikan penilaian yang lebih seimbang terhadap performa model, terutama pada kondisi distribusi data sentimen positif dan negatif yang tidak seimbang, sehingga hasil evaluasi tidak didominasi oleh kelas dengan jumlah data yang lebih banyak.

$$Macro F1-Score = \frac{F1_{Positif} + F1_{Negatif}}{2} \quad (4.5)$$

### 4.3 Hasil Uji Coba

Bagian ini memaparkan hasil eksperimen model *Support Vector Machine* (SVM) yang menggunakan kernel *Radial Basis Function* (RBF) pada data sentimen yang telah diproses melalui *preprocessing* dan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Pengujian dilakukan dengan tiga variasi rasio pembagian dataset, yaitu 90:10, 80:20, dan 70:30, untuk melihat sejauh mana perubahan proporsi data latih dan data uji berpengaruh terhadap efektivitas model. Setiap skenario dievaluasi secara mendalam menggunakan nilai accuracy, precision, recall, F1-score, dan Macro F1-score yang diperoleh dari confusion matrix. Berdasarkan hasil perhitungan tersebut, dilakukan analisis untuk menentukan konfigurasi yang paling optimal dalam proses klasifikasi sentimen.

#### 4.3.1 Skenario Uji Coba 1

Skenario pengujian awal ini menerapkan pemisahan korpus data dengan rasio 90:10 untuk kebutuhan *training data* dan *testing data*. Dari total 1.487 data yang diteliti, sebanyak 1.338 data ditetapkan sebagai data latih dan 149 data diposisikan sebagai data uji. Porsi data latih dibentuk oleh 1.195 sampel berlabel negatif serta 143 sampel berlabel positif. Di sisi lain, porsi data uji mencakup 133 sampel bersentimen negatif dan 16 sampel bersentimen positif. Rincian lengkap mengenai sebaran distribusi data pada skenario pertama ini dapat dicermati pada tabel berikut.

Tabel 4.4 Pembagian Data Skenario 1

| Data       | Negatif | Positif |
|------------|---------|---------|
| Data Latih | 1195    | 143     |
| Data Uji   | 133     | 16      |

Setelah proses pembagian data dilakukan, tahap berikutnya adalah pelatihan model *Support Vector Machine* (SVM) menggunakan kernel *Radial Basis Function* (RBF). Pada tahap ini dilakukan optimasi *hyperparameter* menggunakan metode *GridSearchCV* untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik pada data latih. Berdasarkan hasil proses *training*, diperoleh kombinasi *hyperparameter* terbaik yaitu nilai Cost (C) sebesar 10 dan gamma ( $\gamma$ ) sebesar 0,01. Selain itu, proses *cross-validation* menghasilkan nilai *Macro F1-score* sebesar 0,6344. Nilai tersebut menunjukkan performa rata-rata model dalam mengklasifikasikan kedua kelas sentimen pada data latih selama proses validasi silang. Hasil optimasi *hyperparameter* ini kemudian digunakan untuk membangun model akhir yang selanjutnya dievaluasi menggunakan data uji pada skenario pengujian pertama.

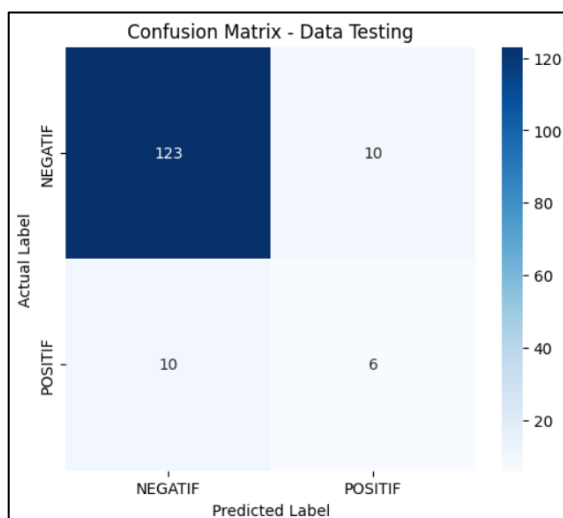
Tabel 4.5 Parameter Skenario 1

| Parameter | Nilai |
|-----------|-------|
| C         | 10    |
| Gamma     | 0,01  |

Kombinasi parameter terbaik yang diperoleh dari tahap optimasi selanjutnya diterapkan pada model SVM untuk proses pengujian menggunakan data *testing*. Performa model kemudian dievaluasi berdasarkan metrik yang telah ditentukan guna mengetahui kemampuan klasifikasi terhadap data baru. Hasil pengujian pada skenario pertama disajikan pada tabel berikut.

Tabel 4.6 Evaluasi Model Skenario 1

| Kelas   | Evaluasi Model   |               |                 |
|---------|------------------|---------------|-----------------|
|         | <i>Precision</i> | <i>Recall</i> | <i>F1-score</i> |
| NEGATIF | 0,92             | 0,92          | 0,92            |
| POSITIF | 0,38             | 0,38          | 0,38            |

Gambar 4.1 *Confusion Matrix* Skenario 1

Berdasarkan visualisasi *Confusion Matrix* Gambar 4.1 dan rangkuman evaluasi pada Tabel 4.6, pengujian skenario awal menggunakan SVM kernel RBF menunjukkan capaian akurasi senilai 0,8658 dengan *Macro F1-score* sebesar 0,6499. Ditinjau dari ketepatan prediksi, model sukses memetakan 123 data sentimen negatif dan 6 data sentimen positif secara akurat, walaupun terdapat deviasi berupa 10 data negatif yang terprediksi sebagai positif dan 10 data positif yang masuk dalam kelas negatif. Evaluasi pada kelas negatif menunjukkan performa yang kokoh dengan perolehan skor *precision*, *recall*, dan *F1-score* yang stabil di angka 0,92. Pembalikan situasi terjadi pada kelas positif, di mana ketiga metrik evaluasi tersebut mandek di angka 0,38. Kesenjangan efektivitas klasifikasi ini mengonfirmasi bahwa arsitektur SVM lebih adaptif terhadap pola sentimen negatif. Hal ini dipicu oleh tidak meratanya sebaran kelas data, di mana volume data bersentimen negatif memiliki keunggulan kuantitas yang signifikan atas sentimen positif.

### 4.3.2 Skenario Uji Coba 2

Skenario pengujian kedua menerapkan rasio pemisahan data sebesar 80:20 untuk alokasi *training data* dan *testing data*. Dari total 1.487 data yang digunakan, sebanyak 1.189 data difungsikan sebagai data latih dan 298 data diposisikan sebagai data uji. Kelompok data latih tersebut dibentuk oleh 1.062 sampel berlabel negatif serta 127 sampel berlabel positif. Sementara itu, kelompok data uji mencakup 266 sampel bersentimen negatif dan 32 sampel bersentimen positif. Rincian lengkap mengenai sebaran distribusi data pada skenario kedua ini dapat dicermati melalui tabel berikut.

Tabel 4.7 Pembagian Data Skenario 2

| <b>Data</b> | <b>Negatif</b> | <b>Positif</b> |
|-------------|----------------|----------------|
| Data Latih  | 1062           | 298            |
| Data Uji    | 266            | 32             |

Setelah proses pembagian data dilakukan, tahap berikutnya adalah pelatihan model menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF). Pada tahap ini, dilakukan optimasi *hyperparameter* menggunakan metode *GridSearchCV* untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik pada data latih. Berdasarkan hasil proses *training*, diperoleh kombinasi *hyperparameter* terbaik yaitu nilai Cost (C) sebesar 1 dan gamma ( $\gamma$ ) sebesar 0,1 dengan nilai *Macro F1-score* sebesar 0,6537 pada proses *cross-validation*. Kombinasi parameter tersebut selanjutnya digunakan untuk membangun model SVM yang akan dievaluasi menggunakan data *testing* pada skenario pengujian kedua.

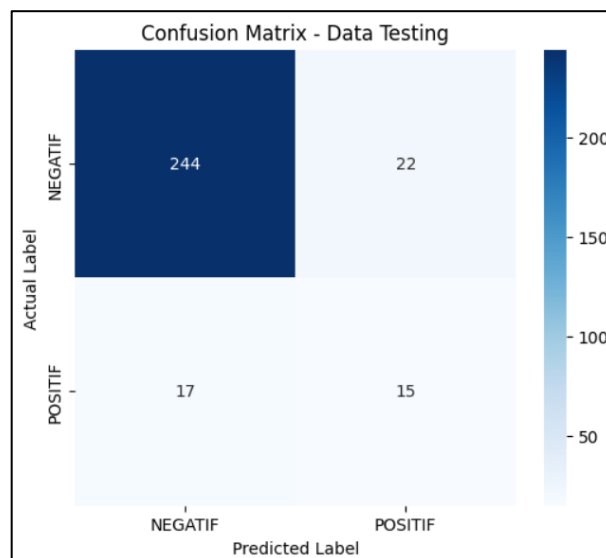
Tabel 4.8 Parameter Skenario 2

| Parameter | Nilai |
|-----------|-------|
| C         | 1     |
| Gamma     | 0,1   |

Kombinasi parameter terbaik yang diperoleh dari tahap optimasi selanjutnya diterapkan pada model *Support Vector Machine* (SVM) untuk proses pengujian menggunakan data testing. Performa model kemudian dievaluasi berdasarkan metrik yang telah ditentukan guna mengetahui kemampuan klasifikasi terhadap data baru. Hasil pengujian pada skenario pertama disajikan pada tabel berikut.

Tabel 4.9 Evaluasi Model Skenario 2

| Kelas   | Evaluasi Model   |               |                 |
|---------|------------------|---------------|-----------------|
|         | <i>Precision</i> | <i>Recall</i> | <i>F1-score</i> |
| NEGATIF | 0,93             | 0,92          | 0,93            |
| POSITIF | 0,41             | 0,47          | 0,43            |

Gambar 4.2 *Confusion Matrix* Skenario 2

Berdasarkan representasi *Confusion Matrix* di Gambar 4.2 dan ringkasan evaluasi pada Tabel 4.9, pengujian skenario kedua menggunakan algoritma SVM

kernel RBF menghasilkan tingkat akurasi senilai 0,8691 dengan *Macro F1-score* sebesar 0,6804. Ditinjau dari ketepatan klasifikasinya, model sukses memetakan 244 data sentimen negatif dan 15 data sentimen positif secara akurat, walaupun terdapat deviasi berupa 22 data negatif yang masuk dalam kelas positif dan 17 data positif yang terprediksi sebagai negatif. Evaluasi pada aspek sentimen negatif menunjukkan performa yang kokoh dengan perolehan skor *precision* 0,93, *recall* 0,92, dan *F1-score* 0,93, yang merepresentasikan kapabilitas klasifikasi yang sangat optimal. Di sisi lain, kelas sentimen positif memperoleh nilai *precision* sebesar 0,41, *recall* sebesar 0,47, dan *F1-score* sebesar 0,43. Karakteristik ini membuktikan bahwa arsitektur model masih cenderung lebih sensitif terhadap sentimen negatif. Kendati demikian, tren kenaikan pada skor *Macro F1-score* mengonfirmasi bahwa skenario ini lebih andal dalam meminimalkan ketimpangan prediksi antarkelas jika dibandingkan dengan skenario pengujian awal.

#### 4.3.3 Skenario Uji Coba 3

Skenario pengujian ketiga menerapkan rasio pemisahan data sebesar 70:30 untuk alokasi *training data* dan *testing data*. Dari total 1.487 data yang digunakan, sebanyak 1.040 data difungsikan sebagai data latih dan 447 data diposisikan sebagai data uji. Kelompok data latih tersebut dibentuk oleh 929 sampel berlabel negatif serta 111 sampel berlabel positif. Sementara itu, kelompok data uji mencakup 399 sampel bersentimen negatif dan 48 sampel bersentimen positif. Rincian lengkap mengenai sebaran distribusi data pada skenario ketiga ini dapat dicermati melalui tabel berikut.

Tabel 4.10 Pembagian Data Skenario 3

| <b>Data</b> | <b>Negatif</b> | <b>Positif</b> |
|-------------|----------------|----------------|
| Data Latih  | 929            | 111            |
| Data Uji    | 399            | 48             |

Setelah proses pembagian data dilakukan, tahap berikutnya adalah pelatihan model menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF). Pada tahap ini, dilakukan optimasi *hyperparameter* menggunakan metode *GridSearchCV* untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik pada data latih. Berdasarkan hasil proses *training*, diperoleh kombinasi *hyperparameter* terbaik yaitu nilai Cost (C) sebesar 1 dan gamma ( $\gamma$ ) sebesar 0,1 dengan nilai *Macro F1-score* sebesar 0,6181 pada proses *cross-validation*.

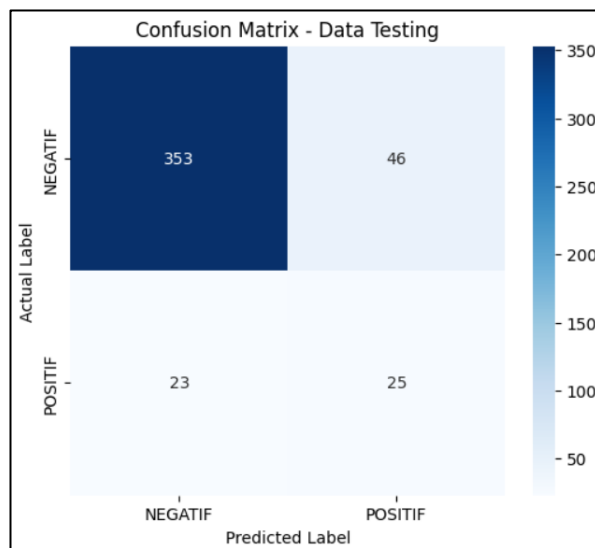
Tabel 4.11 Parameter Skenario 3

| <b>Parameter</b> | <b>Nilai</b> |
|------------------|--------------|
| C                | 1            |
| Gamma            | 0,1          |

Kombinasi parameter terbaik yang diperoleh dari tahap optimasi selanjutnya diterapkan pada model Support Vector Machine (SVM) untuk proses pengujian menggunakan data testing. Performa model kemudian dievaluasi berdasarkan metrik yang telah ditentukan guna mengetahui kemampuan klasifikasi terhadap data baru. Hasil pengujian pada skenario pertama disajikan pada tabel berikut.

Tabel 4.12 Evaluasi Model Skenario 3

| <b>Kelas</b> | <b>Evaluasi Model</b>   |                      |                        |
|--------------|-------------------------|----------------------|------------------------|
|              | <b><i>Precision</i></b> | <b><i>Recall</i></b> | <b><i>F1-score</i></b> |
| NEGATIF      | 0,94                    | 0,88                 | 0,91                   |
| POSITIF      | 0,35                    | 0,52                 | 0,42                   |

Gambar 4.3 *Confusion Matrix* Skenario 3

Berdasarkan representasi *Confusion Matrix* di Gambar 4.3 dan ringkasan evaluasi pada Tabel 4.11, pengujian skenario ketiga menggunakan algoritma SVM kernel RBF menghasilkan tingkat akurasi senilai 0,8456 dengan *Macro F1-score* sebesar 0,6656. Ditinjau dari ketepatan klasifikasinya, model sukses memetakan 353 data sentimen negatif dan 25 data sentimen positif secara akurat, walaupun terdapat deviasi berupa 46 data negatif yang masuk dalam kelas positif dan 23 data positif yang terprediksi sebagai negatif. Evaluasi pada aspek sentimen negatif menunjukkan performa yang kokoh dengan perolehan skor *precision* 0,94, *recall* 0,88, dan *F1-score* 0,91, yang merepresentasikan kapabilitas klasifikasi yang sangat optimal. Di sisi lain, kelas sentimen positif memperoleh nilai *precision* sebesar 0,35, *recall* sebesar 0,52, dan *F1-score* sebesar 0,42. Karakteristik ini membuktikan bahwa arsitektur model masih cenderung lebih sensitif terhadap sentimen negatif. Kendati demikian, penurunan tingkat akurasi secara umum pada skenario ini diimbangi oleh tren positif berupa kenaikan skor *recall* pada kelas minoritas, yang

menandakan bahwa model mampu mengidentifikasi lebih banyak data bersentimen positif secara lebih luas.

#### 4.4 Pembahasan

Berdasarkan pengujian skenario awal dengan proporsi 90:10, diperoleh nilai akurasi sebesar 0,8658 dan *Macro F1-score* sebesar 0,6499 menggunakan algoritma SVM. Representasi metrik tersebut mengindikasikan kelayakan model dalam memprediksi data tekstual dengan benar, yang ditunjukkan oleh tingginya nilai *precision*, *recall*, dan *F1-score* pada kelas sentimen negatif (mencapai 0,92). Kendati demikian, performa berbanding terbalik ditunjukkan pada kelas sentimen positif yang hanya memperoleh *F1-score* sebesar 0,38. Ketidakseimbangan distribusi dataset, di mana volume sentimen negatif jauh melampaui sentimen positif, menjadi faktor utama yang mendasari kecenderungan model untuk lebih mengenali pola dari kelas mayoritas dibanding kelas minoritas.

Pengujian skenario kedua yang menerapkan rasio data 80:20 menghasilkan tingkat akurasi mencapai 0,8691 dengan *Macro F1-score* sebesar 0,6804. Capaian ini melampaui hasil dari skenario pengujian lainnya. Kenaikan metrik *Macro F1-score* tersebut mengindikasikan bahwa model pengklasifikasi kini lebih kapabel dalam memprediksi kelas mayoritas maupun minoritas secara berimbang. Selain itu, nilai *F1-score* untuk kategori sentimen positif mengalami perbaikan hingga menyentuh angka 0,43, yang merepresentasikan peningkatan akurasi prediksi model pada kelas tersebut. Hal ini mengarah pada kesimpulan bahwa alokasi data latih dan data uji sebesar 80% berbanding 20% mampu memfasilitasi proses

pembelajaran dan evaluasi secara lebih komprehensif guna memaksimalkan kinerja model SVM.

Pengujian skenario ketiga yang menerapkan rasio data 70:30 membukukan tingkat akurasi sebesar 0,8456 dengan capaian *Macro F1-score* sebesar 0,6656. Meskipun performa akurasi secara umum merosot dibanding skenario sebelumnya, nilai *recall* pada kelas sentimen positif terbukti meningkat hingga mencapai 0,52. Karakteristik ini memberikan petunjuk bahwa model lebih sensitif dalam mengidentifikasi sampel sentimen positif yang ada. Kendati demikian, tingginya nilai *recall* tersebut berbanding lurus dengan peningkatan tingkat kesalahan klasifikasi pada sampel bersentimen negatif. Kondisi ini menggambarkan fenomena *trade-off* akademis, di mana upaya model untuk mengenali kelas minoritas harus mengorbankan ketepatan prediksi secara menyeluruh akibat berkurangnya kuantitas data latih sebagai basis pembelajaran model.

Berdasarkan hasil penelitian, performa terbaik dicapai oleh kernel RBF pada skenario pembagian data 80:20, yang menghasilkan Macro F1-score tertinggi dibandingkan skenario pengujian lainnya. Temuan ini mengindikasikan bahwa pada rasio tersebut, model mampu menghasilkan klasifikasi yang lebih seimbang antara kelas sentimen positif dan negatif. Temuan ini sejalan dengan konsep *maqāṣid al-syari‘ah*, terutama pada aspek *hifz al-‘aql* (menjaga akal), yang menekankan urgensi penggunaan akal secara tepat, objektif, dan bertanggung jawab dalam proses memperoleh pengetahuan. Pemilihan *Macro F1-score* sebagai metrik evaluasi turut menjadi representasi dari prinsip tersebut, mengingat metrik ini menilai performa model pada setiap kelas secara adil tanpa terpengaruh oleh

ketidakseimbangan jumlah data. Di samping itu, usaha untuk meminimalkan kesalahan klasifikasi serta menghasilkan analisis yang lebih akurat mencerminkan pentingnya ketelitian dalam mengolah data. Dengan demikian, penelitian ini tidak sekadar memberikan kontribusi pada bidang analisis sentimen, tetapi juga memperlihatkan penerapan nilai-nilai *maqasid al-syari'ah* dalam pemanfaatan teknologi dan pengambilan keputusan yang berbasis data.

Prinsip penggunaan nalar dan ketelitian dalam penelitian ini juga sejalan dengan ajaran Islam yang mendorong manusia untuk berpikir secara kritis dan tidak mengambil keputusan secara sembarangan. Dalam Al-Qur'an, Allah SWT berfirman dalam Surah Al-Hashr ayat 18:

يَا أَيُّهَا الَّذِينَ ءَامَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ ۖ وَاتَّقُوا اللَّهَ ۚ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ ١٨

*“Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan.”.*

Menurut tafsir Ibnu Katsir, ayat ini mengandung perintah bagi manusia untuk melakukan *muhasabah* (evaluasi diri) atas setiap tindakan serta mempersiapkan langkah yang lebih baik untuk masa depan (Al-Sheikh, 2003b). Dalam konteks penelitian yang berjudul analisis sentimen publik terhadap isu sepak bola di Indonesia, hal ini relevan dengan bagaimana masyarakat memberikan penilaian terhadap kebijakan dan keputusan yang diambil oleh pihak berwenang seperti PSSI dan tokoh seperti Erick Thohir. Proses analisis sentimen dalam penelitian ini dapat dipahami sebagai bentuk evaluasi publik terhadap keputusan tersebut, sehingga

penggunaan metode yang objektif dan sistematis menjadi penting agar hasil yang diperoleh benar-benar mencerminkan kondisi opini masyarakat secara adil dan bertanggung jawab.

Berdasarkan hasil penelitian ini, penggunaan analisis sentimen tidak hanya dapat dimanfaatkan untuk memahami kecenderungan opini publik, tetapi juga dapat menjadi dasar dalam membangun ruang diskusi sepak bola yang lebih sehat dan kondusif di media digital. Temuan mengenai dominasi sentimen negatif menunjukkan pentingnya upaya meningkatkan kesadaran masyarakat dalam menyampaikan kritik secara santun, konstruktif, dan bertanggung jawab. Dalam perspektif Islam, etika berkomunikasi merupakan bagian dari akhlak yang harus dijaga dalam setiap interaksi, termasuk di media sosial. Rasulullah SAW bersabda:

مَنْ كَانَ يُؤْمِنُ بِاللَّهِ وَالْيَوْمِ الْآخِرِ فَلْيُكُنْ خَيْرًا أَوْ لِيَصُمْتُ

*"Barang siapa yang beriman kepada Allah dan hari akhir, maka hendaklah ia berkata baik atau diam."* (HR. Bukhari dan Muslim).

Hadist tersebut mengajarkan bahwa seseorang yang beriman harus menyadari bahwa setiap ucapan yang disampaikan akan dipertanggungjawabkan di hadapan Allah. Oleh karena itu, seorang muslim dianjurkan untuk hanya menyampaikan perkataan yang baik dan bermanfaat, sedangkan apabila tidak mampu menyampaikan hal yang baik maka lebih utama untuk diam. Dalam kajian etika komunikasi Islam dijelaskan bahwa setiap perkataan memiliki dampak yang besar terhadap individu maupun masyarakat. Kesalahan dalam bertutur kata dapat memicu perpecahan, permusuhan, dan kebencian, terutama di media sosial yang

sering menjadi sarana penyebaran ujaran kebencian, fitnah, maupun informasi yang tidak bertanggung jawab (Muchlis et al., 2025). Selain itu, komunikasi yang baik harus mengedepankan kata-kata yang membangun, penuh pengertian, dan tidak menyakiti pihak lain sehingga tercipta hubungan yang harmonis dan saling menghormati. Dalam konteks penelitian ini, hasil analisis sentimen dapat menjadi rekomendasi bagi pengguna media sosial, komunitas suporter, maupun pemangku kepentingan sepak bola Indonesia untuk mendorong budaya diskusi yang lebih beretika, objektif, dan saling menghormati. Dengan menerapkan adab dalam berkomunikasi serta memanfaatkan media digital secara bijak, diharapkan tercipta wacana sepak bola yang lebih konstruktif, edukatif, dan mendukung kemajuan sepak bola Indonesia secara berkelanjutan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Merujuk pada hasil analisis dan pembahasan yang telah dipaparkan, penelitian ini menyimpulkan bahwa metode *Support Vector Machine* (SVM) terbukti efektif dalam mengeksekusi klasifikasi sentimen masyarakat terhadap sepak bola nasional ke dalam representasi kelas positif dan negatif dengan kinerja yang cukup akurat. Guna mengidentifikasi pengaruh signifikansi dari perbandingan porsi *training data* dan *testing data* terhadap efektivitas model, pengujian dikondisikan ke dalam tiga variasi skenario, yaitu perbandingan 90:10, 80:20, serta 70:30. Hasil dari pengujian tersebut menegaskan bahwa seluruh skenario mampu mempertahankan capaian akurasi di atas 84%. Persentase metrik evaluasi yang tinggi ini memberikan petunjuk kuat bahwa model memiliki daya generalisasi yang baik untuk mengategorikan orientasi opini publik berdasarkan basis data tekstual yang diproses.

Berdasarkan beberapa skema dari seluruh skenario pengujian, performa optimal diraih oleh pemisahan dataset dengan rasio 80:20 melalui perolehan akurasi sebesar 0,8691 dan *Macro F1-score* di angka 0,6804. Nilai metrik tersebut mengindikasikan bahwa alokasi data pelatihan dan data pengujian sebesar 80% berbanding 20% dapat mengondisikan interaksi yang optimal antara proses ekstraksi pola dan validasi model. Lebih lanjut, perolehan *Macro F1-score* tersebut menegaskan keandalan model yang tidak hanya terpaku pada akurasi menyeluruh,

tetapi juga mencerminkan kemampuan sistem dalam menekan ketimpangan pengklasifikasian untuk kelas mayoritas maupun minoritas.

Berdasarkan tinjauan menyeluruh terhadap hasil eksperimen, dapat disimpulkan bahwa metode *Support Vector Machine* (SVM) menunjukkan performa yang kompeten dalam mengeksekusi klasifikasi sentimen masyarakat terhadap sepak bola Indonesia. Sistem secara akurat mampu mengidentifikasi pola-pola opini publik guna dikategorikan ke dalam label sentimen yang sesuai. Namun, analisis evaluasi mencatat bahwa performa pengklasifikasi masih sensitif terhadap distribusi dataset yang tidak merata, yang ditandai dengan dominasi jumlah data bersentimen negatif atas sentimen positif. Implikasinya, model memperlihatkan tingkat akurasi yang lebih tinggi pada kelas mayoritas dibandingkan dengan kemampuannya mendeteksi kelas minoritas. Meskipun terdapat keterbatasan distribusi dataset tersebut, bukti faktual dari riset ini menegaskan bahwa algoritma SVM tetap memegang peranan yang sangat efektif untuk memetakan tren sentimen dan opini publik di sektor olahraga nasional.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, masih terdapat beberapa keterbatasan yang dapat menjadi bahan pengembangan pada penelitian selanjutnya. Oleh karena itu, diperlukan upaya lanjutan untuk meningkatkan performa model serta memperbaiki hasil klasifikasi agar lebih optimal, khususnya dalam menangani data yang bersifat tidak seimbang dan kompleks.

1. Menambah jumlah data penelitian serta meningkatkan keseimbangan distribusi antara kelas sentimen positif dan negatif agar model dapat mempelajari karakteristik setiap kelas secara lebih optimal dan menghasilkan performa klasifikasi yang lebih baik.
2. Mengembangkan penelitian dengan mengeksplorasi algoritma klasifikasi lain, seperti *Random Forest*, *Naive Bayes*, LSTM, dan BERT, untuk mengetahui metode yang memiliki performa terbaik dalam melakukan klasifikasi sentimen terhadap isu sepak bola Indonesia.
3. Mengembangkan representasi fitur teks dengan memanfaatkan teknik *word embedding*, seperti Word2Vec, FastText, atau BERT, sehingga model dapat menangkap konteks dan makna kalimat secara lebih mendalam.
4. Melakukan optimasi *hyperparameter* yang lebih luas, baik pada parameter Cost (C) maupun gamma ( $\gamma$ ), serta mengeksplorasi kombinasi kernel non-linear lainnya untuk memperoleh model klasifikasi dengan performa yang lebih optimal.

## DAFTAR PUSTAKA

- Al-Sheikh, A. bin M. bin A. bin I. (2003a). *Tafsir Ibnu Katsir Jilid 5* (Cetakan Pertama, Vol. 5). Pustaka Imam Asy-Syafi'i. (Original work published Mu-assasah Daar al-Hilaal)
- Al-Sheikh, A. bin M. bin A. bin I. (2003b). *Tafsir Ibnu Katsir Jilid 8* (Cetakan Pertama, Vol. 5). Pustaka Imam Asy-Syafi'i. (Original work published Mu-assasah Daar al-Hilaal)
- Apriyani, I., Fauzi, A., Kusumaningrum, D. S., & Handayani, H. H. (2025). Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi TikTok Menggunakan Algoritma Machine Learning. *Jurnal Accounting Information System (AIMS)*, 8(1), 76–89. <https://doi.org/10.32627/aims.v8i1.1160>
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Asshiddiqi, M. F., & Lhaksana, K. M. (2020). *Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI*.
- Dalifah, N., Suarna, N., & Prihartono, W. (2024). ANALISIS DATA SENTIMEN NEGATIF PADA OPINI PENGGUNA TWITTER TERHADAP BERITA SEPAK BOLA LIGA 1 TAHUN 2022 DENGAN PENERAPAN SUPPORT VECTOR MECHINE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 209–214. <https://doi.org/10.36040/jati.v8i1.8303>
- Dina Aulia. (2024). Semangat Suporter Timnas Indonesia: Wujud Nasionalisme Di Arena Sepak Bola Dan Pengaruh Di Kancah Internasional. *WISSEN : Jurnal Ilmu Sosial dan Humaniora*, 3(1), 157–169. <https://doi.org/10.62383/wissen.v3i1.504>
- Djuyandi, Y., Siregar, M. M., & Muradi, M. (2021). PERAN SUPORTER SEPAKBOLA SEBAGAI KELOMPOK PENEKAN. *JISIP UNJA (Jurnal Ilmu Sosial Ilmu Politik Universitas Jambi)*, 5(1), 1–14. <https://doi.org/10.22437/jisipunja.v5i1.17219>
- Eka Wahyudhi, A. S. B. S., Nasrulloh, A., & Nugroho, S. (2023). *Sisi Positif Suporter Sepakbola Sebagai Konsumen Olahraga*. <https://doi.org/doi.org/10.24036/Sporta%20Saintika/vol8-iss2/326>
- Eko Alwinasirin & Rizky Putra Santosa. (2025). Pengaruh Fanatisme Terhadap Agresi Verbal Suporter Timnas Indonesia di Media Sosial: Penelitian. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, 4(1), 4056–4062. <https://doi.org/10.31004/jerkin.v4i1.2008>
- Fahrezi, I. A., Rudiman, & Nauval Azmi Verdikha. (2024). Analisis Sentimen Twitter Atas Isu Hak Angket Menggunakan Pembobotan TF-IDF dan Algoritma SVM. *Sci-tech Journal*, 3(2). <https://doi.org/10.56709/stj.v3i2.526>

- Fardi, M. D., & Akbar, M. (2026). Analisis Sentimen Supporter Pss Sleman Terhadap Peringkat Klub Di Twitter/X Menggunakan Naive Bayes. *JEKIN - Jurnal Teknik Informatika*, 5(3), 1146–1154. <https://doi.org/10.58794/jekin.v5i3.1616>
- Farhan, M., & Puspasari, S. (2025). *Analisis Sentimen Supporter Sriwijaya FC Berbasis*. 6(1).
- Fauzan, M. R., Anraeni, S., & Budiman Ilmuwan, L. (2025). Analisis Sentimen Tweet Netizen Terhadap Timnas Sepak Bola Indonesia di Era Shin Tae-Yong Menggunakan Algoritma Support Vector Machine (SVM). *LINIER: Literatur Informatika dan Komputer*, 2(3), 329–340. <https://doi.org/10.33096/linier.v2i3.3144>
- Fauziah, A. H., & Nurulita, D. A. (2024). *INTEGRASI NILAI PENDIDIKAN DALAM SURAH AL-ISRA' AYAT 36 UNTUK MENCEGAH PERILAKU IMPULSIVITAS DI MEDIA SOSIAL*. 2.
- Ferdiyawan, A. B., Fauzy, I., & Basuki, S. (2025). *ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER(X) TERHADAP PEMAIN DIASPORA DI TIM NASIONAL SEPAK BOLA INDONESIA STUDI KASUS PESEPAKBOLA MEES HILGERS DAN ELIANO REINJEDERS MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)*.
- Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM). *Jurnal Gaussian*, 9(3), 376–390. <https://doi.org/10.14710/j.gauss.v9i3.28932>
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Imaddudin, S., Astuti, I., & Ruhama, S. (2025). Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan Algoritma Support Vector Machine (SVM) pada Media Sosial X. *Jurnal Penelitian Multidisiplin Bangsa*, 1(8), 1003–1013. <https://doi.org/10.59837/jpnmb.v1i8.193>
- Julianti, O. N., Suarna, N., & Prihartono, W. (2024). PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2936–2941. <https://doi.org/10.36040/jati.v8i3.9613>
- Maisat Eka Darmawan, Z., & Fauzan Dianta, A. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM. *Teknologi*, 13(1), 8–15. <https://doi.org/10.26594/teknologi.v13i1.3098>
- Muchlis, K. M. C., Muhajirin, & Nor Afif Hamid, A. R. (2025). Etika Bermedia Sosial dalam Perspektif Hadis: Studi Tematik Terhadap Nilai-Nilai Komunikasi Islam. *Journal of Innovative and Creativity*.
- Munandar, A. A., Farikhin, F., & Widodo, C. E. (2023). Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM. *JOINTECS (Journal of Information Technology and Computer Science)*, 8(2), 77. <https://doi.org/10.31328/jointecs.v8i2.4747>

- Pangestu, D., Malik, M., & Pribadi, M. R. (2024). Analisis Sentimen Hasil Pertandingan Sepakbola Timnas Indonesia di Piala Asia U-23 pada Platform Youtube menggunakan Algoritma Support Vector Machine (SVM). *Applied Information Technology and Computer Science (AICOMS)*, 3(1), 38–48. <https://doi.org/10.58466/aicoms.v3i1.1528>
- Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 153–160. <https://doi.org/10.57152/malcom.v3i2.897>
- Sinaga, A., & Nainggolan, S. P. (2023). ANALISIS PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA STEMMING ARIFIN-SETIONO DAN NAZIEF-ADRIANI PADA DOKUMEN TEKS BAHASA INDONESIA. *Sebatik*, 27(1), 63–69. <https://doi.org/10.46984/sebatik.v27i1.2072>
- Toha, A., Purwono, P., & Gata, W. (2022). Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV. *Buletin Ilmiah Sarjana Teknik Elektro*, 4(1), 12–21. <https://doi.org/10.12928/biste.v4i1.6079>