

## **Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan Algoritma *Support Vector Machine* (SVM) pada Media Sosial X**

**Satrio Imaddudin<sup>1</sup>, Ida Astuti<sup>2</sup>, Syamsi Ruhama<sup>3</sup>**

*<sup>1,2,3</sup> Sistem Informasi, Universitas Gunadarma, Indonesia*

### **Corresponding Author**

**Nama Penulis:** Ida Astuti

**E-mail:** [astuti@staff.gunadarma.ac.id](mailto:astuti@staff.gunadarma.ac.id)

### **Abstrak**

PSSI sebagai organisasi induk sepak bola di Indonesia, memiliki peran strategis dalam pengelolaan dan pengembangan sepak bola Nasional. Penelitian ini bertujuan untuk melakukan studi sentiment masyarakat tentang PSSI di era kepemimpinan Erick Thohir melalui media sosial X dengan menerapkan metode *Support Vector Machine* (SVM). Data penelitian diperoleh melalui teknik data crawling, yang berhasil mengumpulkan 3.900 tweet dari platform media sosial X. Proses analisis dilakukan berdasarkan metodologi CRISP-DM, yang meliputi tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Pada tahap pengujian model, dilakukan tiga skenario pembagian data latih dan data uji dengan rasio 80:20, 70:30, dan 60:40, menggunakan dua bahasa, yaitu bahasa Indonesia dan bahasa Inggris. Hasil pengujian menunjukkan bahwa skenario terbaik tercapai pada rasio 80:20 dengan bahasa Inggris, menghasilkan akurasi sebesar 76,6%. Sentimen yang diidentifikasi dalam analisis ini diklasifikasikan menjadi tiga kategori: positif, netral, dan negatif. Dari hasil analisis, sentimen positif mendominasi dengan persentase 66%, diikuti oleh sentimen netral sebesar 27%, dan sentimen negatif sebesar 7,1%.

**Kata kunci** – PSSI, Erick Thohir, Studi, Sentimen Masyarakat, SVM

### **Abstract**

The Football Association of Indonesia (PSSI), as the governing body of football in Indonesia, plays a crucial role in managing and advancing national football. This study aims to analyze public opinion on PSSI during Erick Thohir's era using data from the social media platform X and employing the *Support Vector Machine* (SVM) method. The data for this research was gathered from X through data crawling techniques, collecting 3,900 tweets. The analysis process follows the CRISP-DM methodology, which includes business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In the model testing phase, three data split scenarios were applied with training-to-testing ratios of 80:20, 70:30, and 60:40, using two languages, English and Indonesian. The best-performing scenario was the 80:20 ratio with English, achieving an accuracy of 76.6%. The sentiment analysis categorized the results into positive, neutral, and negative sentiments. The outcome revealed that positive sentiment dominated with 66%, followed by neutral sentiment at 27%, and negative sentiment at 7.1%.

**Keywords** – PSSI, Erick Thohir, Study, Public Opinion, SVM

## PENDAHULUAN

Sepakbola adalah olahraga yang paling digemari dan berhasil menarik perhatian masyarakat di seluruh dunia hingga saat ini, tanpa memandang usia, gender, maupun status sosial. Di Indonesia, sepakbola tidak hanya berfungsi sebagai hiburan semata, tetapi juga menjadi elemen penting dalam budaya serta simbol nasionalisme. Namun, di balik euforia dan antusiasme yang melingkupinya, terdapat berbagai tantangan dan polemik, terutama dalam hal pengelolaan serta pemberitaan sepakbola di tanah air. (Prasetyo, 2023)

Persatuan Sepak Bola Seluruh Indonesia (PSSI), sebagai organisasi induk sepakbola di Indonesia, memegang peranan krusial dalam mengatur, mengelola, dan memajukan sepakbola di tingkat nasional (Zulhidayat, 2022). Sebagai lembaga yang bertanggung jawab atas manajemen sepakbola di Indonesia, PSSI memiliki kewajiban besar untuk menciptakan ekosistem yang mendukung pertumbuhan dan perkembangan olahraga sepakbola di tanah air.

Ketua Umum PSSI memegang peran strategis dalam menentukan arah dan kebijakan organisasi. Di bawah kepemimpinan Erick Thohir, berbagai langkah signifikan telah dilakukan untuk meningkatkan kualitas wasit dan memperbaiki standar pengambilan keputusan di lapangan. Salah satu langkah terpenting adalah penerapan teknologi *Video Assistant Referee* (VAR), yang memungkinkan wasit meninjau kembali situasi-situasi krusial atau kontroversial dalam pertandingan melalui rekaman video. Tujuan utama dari penerapan VAR adalah memastikan keputusan yang lebih akurat dan adil (Salam, 2021). Bapak Erick Thohir, yang dikenal memiliki pengalaman luas serta jaringan yang kuat di bidang olahraga dan bisnis, diharapkan mampu membawa perubahan positif bagi sepakbola Indonesia (Nayottama, 2023). Latar belakang yang baik dan visi yang jelas, ia diyakini dapat menjalankan reformasi mendalam pada struktur dan manajemen PSSI, sekaligus meningkatkan pencapaian sepakbola Indonesia di level regional maupun internasional.

Era digital saat ini, media sosial telah menjadi platform utama bagi masyarakat untuk menyampaikan pendapat, komentar, dan opini terkait berbagai aspek sepakbola, termasuk evaluasi terhadap kinerja para pemimpin di PSSI. Media sosial X menjadi ruang bagi publik untuk mengekspresikan pandangan, baik berupa apresiasi maupun kritik, terhadap kinerja Ketua Umum PSSI Erick Thohir serta kebijakan-kebijakan yang diambil oleh PSSI (Yanuarti, 2021). Analisis sentimen merupakan proses komputasi yang bertujuan untuk mengevaluasi teks digital guna mengidentifikasi makna atau emosi tertentu yang terkandung di dalamnya. Metode ini berfokus pada pemahaman dan pengelompokan emosi ke dalam tiga kategori utama: positif, negatif, dan netral (Kurniawati, 2023). Oleh karena itu, analisis respons publik di platform X dapat memberikan wawasan berharga mengenai arah sentimen masyarakat terhadap PSSI di bawah kepemimpinan Erick Thohir, apakah cenderung positif, negatif, atau netral.

*Support Vector Machine* (SVM) merupakan salah satu metode analisis yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini masyarakat dalam berbagai konteks, termasuk analisis opini terkait kinerja tokoh publik di media sosial (He, 2024). Penggunaan metode ini, dapat dilakukan evaluasi untuk menentukan apakah opini masyarakat terhadap PSSI di era kepemimpinan Erick Thohir cenderung positif, negatif, atau netral. Beberapa penelitian yang menggunakan SVM diantaranya penelitian (Muhammad Fadli Asshiddiqi, 2020) menggunakan Metode *Support Vector Machine* (SVM) dan *Decision Tree* untuk analisis sentimen komentar Instagram PSSI, akurasi terbesar diperoleh pada metode *Support Vector Machine* (SVM) yaitu sebesar 94.36%. Penelitian Analisa Sentimen tentang pembatalan Piala Dunia U20 di Indonesia menggunakan *Support Vector Machine* yang dilakukan oleh (Adelia Damayanti, 2023) diperoleh akurasi sebesar 74,4%. Penggunaan algoritma SVM pada kasus klasifikasi Bahasa menggunakan Naïve Naves dan SVM, didapatkan bahwa SVM mampu menghasilkan performa yang baik dengan akurasi sebesar 96.34% (Tuhenay Deglorians, 2021).

Berdasarkan latar belakang di atas, maka akan dilakukan studi analisis sentimen masyarakat terhadap PSSI di bawah pimpinan Erick Thohir dengan mengimplementasikan metode *Support Vector*

## TINJAUAN PUSTAKA

### *Machine Learning*

Salah satu metode yang dapat digunakan untuk menganalisis informasi dalam sebuah kalimat adalah melalui penerapan kecerdasan buatan, khususnya *machine learning*. *Machine learning* sendiri adalah cabang ilmu yang memungkinkan komputer untuk belajar secara mandiri tanpa harus diprogram secara langsung. Berdasarkan definisi ini, tujuan utama *machine learning* adalah mengembangkan sistem komputer yang mampu belajar, berpikir, dan membuat keputusan secara mandiri (Sofana Iwan, 2021). Istilah *Machine Learning* pertama kali diperkenalkan oleh beberapa ilmuwan matematika seperti Adrien Marie Legendre, Thomas Bayes dan Andrey Markov pada tahun 1920- an, dan sejak itu, teknologi ini terus berkembang.

### *Natural Language Processing*

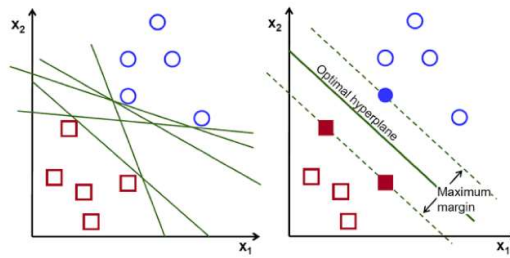
*Natural Language Processing* (NLP) merupakan bagian dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia alami. NLP melibatkan berbagai teknik dan algoritma yang bertujuan untuk memahami, menganalisis, dan menghasilkan teks dalam bahasa manusia secara otomatis (Sihombing, 2022). *Natural Language Processing* (NLP) mencakup berbagai teknik untuk menganalisis bahasa manusia, yang semakin penting dalam berbagai bidang. Berikut ini adalah beberapa teknik NLP yang sering digunakan Analisis Sentimen, Peringkasan Teks, Ekstraksi Kata Kunci, Tokenisasi dll.

### *Analisis Sentimen*

Analisis Sentimen adalah sebuah proses komputasi yang bertujuan untuk mengidentifikasi, mengekstraksi, dan menganalisis sentimen, opini, atau emosi yang terkandung dalam teks. Tujuannya utamanya adalah untuk memahami sikap atau pandangan orang terhadap suatu topik atau entitas tertentu, seperti produk, layanan, organisasi, atau peristiwa, berdasarkan pada teks yang ditulis tentangnya. Proses Analisis Sentimen terdiri dari beberapa langkah, termasuk pengumpulan data teks, pembersihan dan preprocessing data, ekstraksi fitur, pembuatan model, dan evaluasi hasil. (Kurniawati, 2023)

### *Support Vector Machine (SVM)*

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane di dalam ruang fitur yang memisahkan data menjadi kelas-kelas yang berbeda dengan margin maksimal (Esa Anindika Sari, 2020). Dalam konteks klasifikasi, SVM mencoba menemukan *hyperplane* yang memaksimalkan jarak antara titik data dari dua kelas yang berbeda, yang dikenal sebagai margin. Algoritma SVM dapat digunakan untuk masalah klasifikasi linier maupun non-linier. Untuk data yang tidak dapat dipisahkan secara linier, SVM menggunakan teknik kernel untuk memetakan data ke ruang berdimensi lebih tinggi di mana data tersebut dapat dipisahkan secara linier. Beberapa kernel yang umum digunakan adalah linear, polynomial, *Radial Basis Function* (RBF), dan sigmoid.



Gambar 1.

Support Vector Machine (SVM), (Sumber: towardsdatascience.com)

### TextBlob

TextBlob adalah library Python open source yang menyediakan berbagai metode bawaan untuk berbagai tugas pemrosesan *Natural Language Processing* (NLP). Beberapa fitur yang dapat dimanfaatkan dari TextBlob termasuk koreksi ejaan, penandaan entitas, dan klasifikasi teks. Namun, perlu dicatat bahwa labeling dalam TextBlob hanya tersedia dalam bahasa Inggris. TextBlob mempunyai kemampuan untuk mengukur polaritas dan subjektivitas teks. Polaritas mengacu pada seberapa positif atau negatif suatu teks, dengan nilai berkisar dari -1.0 (sangat negatif) hingga 1.0 (sangat positif). Di sisi lain, subjektivitas menunjukkan sejauh mana teks tersebut mencerminkan opini subjektif atau fakta objektif, dengan nilai berkisar dari 0.0 (sangat objektif) hingga 1.0 (sangat subjektif). Dengan fitur-fitur tersebut, TextBlob menjadi alat yang sangat berguna untuk analisis dan pemrosesan teks (Lerekoqholosha, 2024).

### Preprocessing

*Preprocessing* adalah proses mengubah data mentah menjadi bentuk yang lebih terstruktur dan mudah dipahami. Proses ini penting untuk memperbaiki kesalahan pada data yang sering kali tidak lengkap dan memiliki format yang tidak konsisten (Dianda Rifaldi, 2023). Proses *preprocessing* mencakup validasi dan imputasi data. Validasi bertujuan untuk mengevaluasi sejauh mana data yang telah diproses lengkap dan akurat. Imputasi, di sisi lain, berfungsi untuk memperbaiki kesalahan dan mengisi nilai yang hilang, baik secara manual atau otomatis menggunakan program otomatisasi proses bisnis. Untuk memastikan efektivitas proses *preprocessing* data, tahapan ini dibagi menjadi empat langkah utama: *Data Cleaning*, *Data Integration*, *Data Transformation*, dan *Data Reduction*.

Kualitas data secara langsung mempengaruhi keberhasilan proyek yang melibatkan analisis data. Dalam *machine learning*, data *preprocessing* memastikan bahwa *big data* telah diproses dan diformat dengan benar, sehingga algoritma dapat memahami dan memproses informasi dengan lebih akurat (Adam, 2022).

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

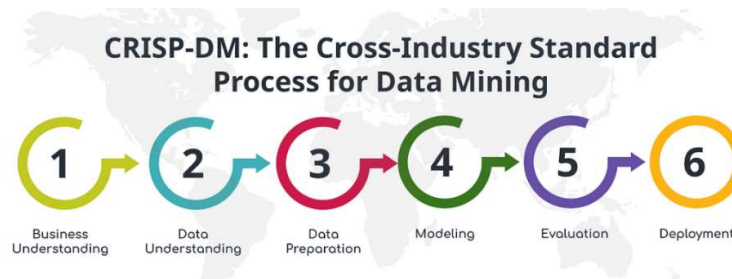
Metode TF-IDF adalah cara untuk memberikan bobot pada hubungan antara suatu kata (term) dan dokumen. Metode ini menggabungkan dua konsep dalam perhitungan bobot yaitu frekuensi kemunculan sebuah kata di dalam dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut.

Frekuensi kemunculan kata di dalam dokumen menunjukkan seberapa penting kata tersebut dalam dokumen tersebut, sedangkan frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Dengan demikian, bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi jika frekuensi kata tersebut tinggi dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah dalam kumpulan dokumen (Musfiroh Nurjannah, 2013).

## METODE

Penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) yaitu metodologi yang banyak digunakan untuk memandu proyek data mining. Dikembangkan pada akhir 1999 oleh sekelompok industri dan akademisi, CRISP-DM menyediakan kerangka kerja yang terstruktur untuk memahami, merencanakan, dan melaksanakan proyek data mining secara efektif. Metodologi ini terdiri dari enam fase yang saling berhubungan, yang dapat diulang dan diiterasi sesuai kebutuhan proyek (Afika Rianti, 2023).

Terdiri dari enam tahapan utama untuk melakukan studi analisis sentimen masyarakat terhadap PSSI di era kepemimpinan Erick Thohir menggunakan media sosial X. Adapun tahapan-tahapan yang dilakukan pada penelitian ini seperti tampak pada Gambar 2.



Gambar 2.

*Cross Industry Standard Process for Data Mining* (CRISP-DM)

(Sumber: medium.com)

Berikut adalah tahapan-tahapan pada CRISP-DM :

1. *Business Understanding* (Tahap pemahaman bisnis)  
Kepemimpinan Erick Thohir di PSSI telah menjadi topik hangat di kalangan masyarakat Indonesia, terutama di media sosial. Pengelolaan persepsi publik dan respons terhadap kebijakan yang diambil sangat penting bagi keberhasilan PSSI dalam meningkatkan kualitas sepak bola nasional. Analisis sentimen ini bertujuan untuk mengidentifikasi sentimen positif, negatif, dan netral yang muncul terkait kebijakan dan tindakan Erick Thohir sebagai ketua PSSI. Dengan pemahaman ini, PSSI dapat mengambil keputusan yang lebih baik dan strategi komunikasi yang lebih efektif untuk meningkatkan dukungan publik.
2. *Data Understanding* (Tahap pemahaman data)  
Pada tahap ini, data dikumpulkan melalui media sosial X. Data yang dikumpulkan terdapat emot, URL dan tag pengguna (@username). Proses ini dikenal sebagai crawling data, di mana data mentah dalam bentuk tweet diambil tanpa proses pembersihan.
3. *Data Preparation* (Tahap persiapan data)  
Data yang telah dikumpulkan kemudian melalui proses *preprocessing* untuk menghasilkan data yang lebih akurat. Tahapan *preprocessing* meliputi *Cleaning*, *Case folding*, *Normalization*, *Stopword Removal*, *Tokenizing*, *Stemming*, *Translation*, dan *Labelling*. Kemudian, data dibagi menjadi data latih dan data uji. Selanjutnya tahap pembobotan kata menggunakan metode TF-IDF.
4. *Modeling* (Tahap pemodelan)  
Pada tahap ini, algoritma *Support Vector Machine* (SVM) digunakan untuk analisis sentimen. SVM dipilih karena kemampuannya dalam mengklasifikasikan data dengan akurasi tinggi, terutama dalam pengolahan teks. Model dibangun menggunakan data latih dan diuji dengan data uji untuk menghasilkan *Accuracy* SVM.
5. *Evaluation* (Tahap Evaluasi)  
Evaluasi dilakukan dengan menggunakan metode *confusion matrix* untuk mengukur *Accuracy model*, *Precision*, *Recall*, *f1-Score*, dan *Support*.



## 6. Penerapan

Pada tahap penerapan dilakukan visualisasi hasil klasifikasi *Support Vector Machine* (SVM) dalam bentuk dashboard yang menunjukkan jumlah sentimen positif, negatif, dan netral dari tweet mengenai PSSI di era kepemimpinan Erick Thohir.

## PEMBAHASAN

### *Data Preperation (Persiapan Data)*

Tahap awal yang dilakukan pada penelitian ini adalah pengumpulan data untuk digunakan dalam proses klasifikasi. Proses *crawling* dilakukan dengan menggunakan program python. Variabel *search\_keyword* mengandung kata kunci "erick thohir" dan "pssi" untuk mencari tweet berbahasa Indonesia yang diposting antara 16 Februari 2023 dan 23 Juni 2024. Hasil *crawling* data disimpan dalam bentuk file dalam bentuk csv.

Pada tahap persiapan data dilakukan proses pengolahan data yang meliputi *preprocessing*, *split* data, dan pembobotan kata. Pada tahap *preprocessing*, dataset dibersihkan dari kata-kata yang tidak sesuai dengan indeks dan karakter lain yang tidak diperlukan dalam analisis sentimen. Tahap ini melibatkan beberapa proses pembersihan, di antaranya:

- Cleaning data*, yaitu tahapan membersihkan teks dari elemen seperti tanda baca, simbol, dan angka. Pembersihan data melibatkan penghapusan karakter khusus seperti @mentions, #hashtag, dan URL dalam tweet. Dilanjutkan dengan menghapus tweet duplikat untuk memastikan setiap tweet yang dianalisis tidak ada yang sama.
- Case folding*, yaitu mengubah semua huruf kapital menjadi huruf kecil untuk konsistensi.
- Normalization*, yaitu mengubah kata ke bentuk standar dan memperbaiki kesalahan pengetikan, setiap kata dalam teks disesuaikan dengan bentuk yang tercantum dalam Kamus Besar Bahasa Indonesia (KBBI). Proses ini melibatkan perbandingan setiap kata dalam dokumen teks dengan kata-kata yang ada dalam kamus bahasa baku.
- Stopword removal* yaitu, menghapus kata-kata umum yang tidak penting dalam analisis, *Stopword* mencakup kata-kata yang sering muncul tetapi tidak menambah makna pada teks. Kata-kata umum dalam data yang telah dinormalisasi akan dihapus untuk mengurangi jumlah kata yang disimpan oleh sistem. Kamus *stopword* yang digunakan dari library *sastrawi* pada Python.
- Tokenize*, yaitu memecah teks menjadi unit-unit kecil menjadi kata-kata individu dan menyusunnya menjadi potongan-potongan terpisah. Kata-kata dalam dokumen dianggap terpisah jika terdapat spasi di antara kata.
- Stemming*, yaitu mengubah kata ke bentuk dasar. suatu proses untuk mengubah kata menjadi suatu kata dasar dengan menghapus imbuhan awal dan akhiran dari kata dalam teks. Tujuan dari tahap *stemming* ini adalah untuk mengubah kata ke bentuk dasarnya, sehingga membantu mengurangi variasi kata dalam analisis data.
- Translate*, yaitu menerjemahkan tweet dari bahasa Indonesia ke bahasa Inggris. Kode ini menggunakan library *deep-translator*. Penerjemahan ini diperlukan karena pada tahap *labelling* menggunakan library *TextBlob* yang berfungsi dalam bahasa Inggris. Data yang telah diterjemahkan dibersihkan kembali dan di bentuk menjadi lowercase
- Labelling*, yaitu memberikan label kategori pada teks, seperti "positif", "negatif", dan "netral" menggunakan *TextBlob*.

Setelah dilakukan *preprocessing* data selanjutnya dilakukan *splitting* data yang bertujuan untuk memisahkan data menjadi data latih dan data uji. Kedua set data ini kemudian akan digunakan untuk melatih dan menguji model menggunakan algoritma *Support Vector Machine* (SVM). *Splitting* data dilakukan dengan menggunakan rasio data latih : data uji yaitu 80:20, 70:30, 60:40. Data yang sudah di *split* selanjutnya dilakukan pembobotan dengan TF-IDF yang bertujuan untuk menentukan bobot masing-masing kata dalam data tweet, sehingga program dapat mengukur sejauh mana kata tersebut

penting dalam keseluruhan data. Proses pembobotan ini memanfaatkan *library* sklearn, khususnya modul `feature_extraction.text` dan `TfidfVectorizer`.

### Modelling (Tahap Pemodelan)

Setelah proses TF-IDF selanjutnya data akan diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) dengan menggunakan kernel linear dan kernel RBF, 3 skenario pembagian data yang sudah dibuat dan menggunakan 2 bahasa yaitu Bahasa Indonesia dan Bahasa Inggris sehingga nantinya akan mendapatkan skenario yang memiliki nilai akurasi terbaik.

Tabel 1. Hasil Modelling SVM kernel *Linear*

| Akurasi Support Vector Machine Kernel Linear |                  |                |
|----------------------------------------------|------------------|----------------|
| Rasio Data<br>(data latih: data uji)         | Bahasa Indonesia | Bahasa Inggris |
| 80:20                                        | 65,5%            | 76,6%          |
| 70:30                                        | 64,9%            | 76,1%          |
| 60:40                                        | 64,2%            | 73,5%          |

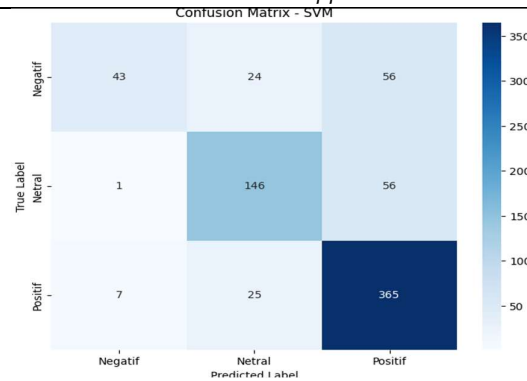
Tabel 2. Hasil Modelling SVM kernel RBF

| Akurasi Support Vector Machine Kernel RBF |                  |                |
|-------------------------------------------|------------------|----------------|
| Rasio Data<br>(data latih: data uji)      | Bahasa Indonesia | Bahasa Inggris |
| 80:20                                     | 61,2%            | 71,5%          |
| 70:30                                     | 61%              | 71,2%          |
| 60:40                                     | 60,1%            | 68,6%          |

Berdasarkan table 1 dan 2, diketahui bahwa nilai akurasi tertinggi diperoleh pada skenario data 80:20, dilatih dengan kernel *Linear* dan menggunakan bahasa Inggris, dengan nilai akurasi sebesar 76,6%. Oleh karena itu model ini dianggap model yang terbaik dan selanjutnya dilakukan evaluasi terhadap model tersebut.

### Evaluasi Hasil

Pada tahap evaluasi ini, *confusion matrix* akan digunakan untuk menilai model yang terbaik yaitu pada skenario data latih : data uji 80:20 dengan kernel Linier dan menggunakan Bahasa Inggris. Tahap ini diperlukan untuk mencoba keakuratan model klasifikasi *Support Vector Machine* yang telah dibuat. *Confusion matrix* akan ditampilkan dalam bentuk heatmap, seperti yang ditunjukkan pada Gambar 3.



**Gambar 3.**  
Grafik Confusion Matrix

Berdasarkan Gambar 3, didapatkan nilai-nilai pada gambar *confusion matrix* yang akan digunakan untuk menghitung kinerja *matrix* seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score* sebagai berikut:  
Test *Accuracy* : 0.766 atau sekitar 76.6%.

**Berikut adalah perhitungan untuk kelas Negatif :**

$$Precision\ Negatif = \frac{43}{43 + 8} = \frac{43}{51} \approx 0.84$$

$$Recall\ Negatif = \frac{43}{43 + 80} = \frac{43}{123} \approx 0.35$$

$$F1 - Score\ Negatif = 2 \times \frac{0.84 \times 0.35}{0.84 + 0.35} \approx 0.49$$

**Berikut adalah perhitungan untuk kelas Netral :**

*True Positif* (TP) = 146 (Prediksi Netral dan Sebenarnya Netral)

*False Positif* (FP) = 24 + 25 = 49 (Prediksi Netral tapi sebenarnya Negatif atau Positif)

*False Negatif* (FN) = 1 + 56 = 57 (Sebenarnya Netral tapi diprediksi sebagai Negatif atau Positif)

$$Precision\ Netral = \frac{146}{146 + 49} = \frac{146}{195} \approx 0.75$$

$$Recall\ Netral = \frac{146}{146 + 57} = \frac{146}{203} \approx 0.72$$

$$F1 - Score\ Netral = 2 \times \frac{0.75 \times 0.72}{0.75 + 0.72} \approx 0.73$$

**Berikut adalah perhitungan untuk kelas Positif :**

*True Positif* (TP) = 365 (Prediksi Positif dan Sebenarnya Positif)

*False Positif* (FP) = 56 + 56 = 112 (Prediksi Positif tapi sebenarnya Negatif atau Netral)

*False Negatif* (FN) = 7 + 25 = 32 (Sebenarnya Positif tapi diprediksi sebagai Negatif atau Netral)

$$Precision\ Positif = \frac{365}{365 + 112} = \frac{365}{477} \approx 0.77$$

$$Recall\ Positif = \frac{365}{365 + 32} = \frac{365}{397} \approx 0.92$$

$$F1 - Score\ Positif = 2 \times \frac{0.77 \times 0.92}{0.77 + 0.92} \approx 0.84$$

Output yang dihasilkan berdasarkan perhitungan melalui program dapat dilihat pada Gambar 4.



SVM Training Accuracy: 0.9536652835408023  
SVM Test Accuracy: 0.7662517289073306

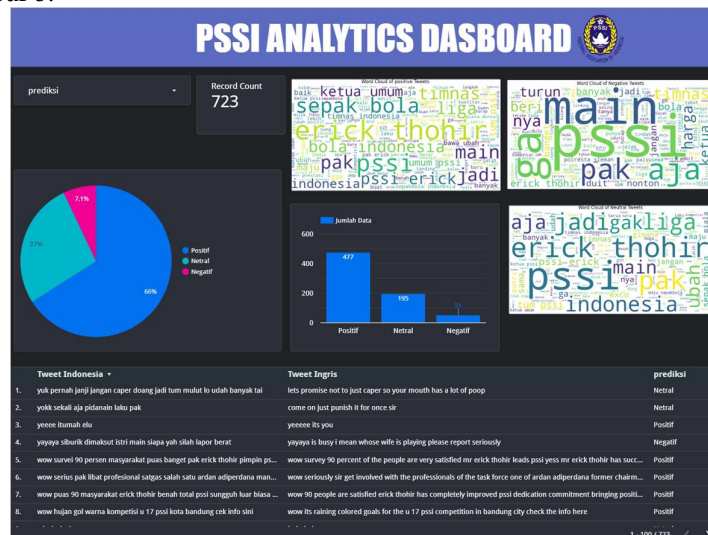
SVM Classification Report:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Negatif      | 0.84      | 0.35   | 0.49     | 123     |
| Netral       | 0.75      | 0.72   | 0.73     | 203     |
| Positif      | 0.77      | 0.92   | 0.84     | 397     |
| accuracy     |           |        | 0.77     | 723     |
| macro avg    | 0.79      | 0.66   | 0.69     | 723     |
| weighted avg | 0.77      | 0.77   | 0.75     | 723     |

**Gambar 4.**  
Klasifikasi Report SVM

## Dashboard Visualisasi

Pada Tahap terakhir dalam studi analisis sentimen ini adalah visualisasi data menggunakan *dashboard*. Dimana hasil keseluruhan dari data yang sudah diklasifikasikan menggunakan *Support vector machine* (SVM) akan divisualisasikan berupa diagram batang , diagram pie, dan kata yang sering muncul pada kelas positif, negatif, dan netral menggunakan *wordcloud*. Hasil dari visualisasi data dapat dilihat pada Gambar 5.



**Gambar 5.**  
Dashboard Hasil Klasifikasi

Pada dashboard tersebut berisi hasil dari klasifikasi *Support vector machine* (SVM). Terdapat beberapa diagram dalam dashboard tersebut seperti diagram batang , diagram pie, *wordcloud*, dan hasil dari *tweet*. Dalam diagram tersebut dapat disimpulkan sentimen positif mendapat 66% yaitu 477 *tweet*, sentimen netral mendapat 27% yaitu 195 *tweet*, dan sentimen negatif mendapatkan 7% yaitu 51 *tweet*.

## KESIMPULAN

Penerapan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) terbukti efektif dalam mengelola analisis sentimen secara menyeluruh, mulai dari tahap perencanaan hingga evaluasi akhir. Pendekatan ini terbukti efektif dalam mengidentifikasi serta memahami pola-pola sentimen masyarakat terkait PSSI di era kepemimpinan Erick Thohir. Pada tahap pengujian model algoritma *Support Vector Machine* (SVM), dilakukan skenario pembagian data latih dan data uji, dimana terdapat 3 skenario yaitu rasio 80:20, 70:30, dan 60:40. Penelitian ini menggunakan 2 bahasa yaitu bahasa Inggris dan Indonesia dan menggunakan dua jenis kernel yaitu Linier dan RBF. Dari hasil

pelatihan model didapatkan rasio tertinggi adalah 80:20 dengan menggunakan bahasa Inggris dan kernel Linier, mendapatkan akurasi sebesar 76,6%. Studi analisis sentimen ini juga melakukan klasifikasi tweet ke dalam sentimen positif, negatif, netral. Hasil sentimen pada data uji menunjukkan bahwa sentimen positif mendapatkan hasil 66%, lalu netral sebesar 27%, dan negatif sebesar 7,1%. Berdasarkan hasil klasifikasi menyatakan bahwa mayoritas tanggapan masyarakat pada platform X cenderung bersifat positif terhadap PSSI di era kepemimpinan Erick Thohir selama dari masa menjabat 16 Januari 2023 sampai dengan 23 juni 2024 saat penelitian ini dibuat.

Saran yang dapat direkomendasikan untuk PSSI dalam era kepemimpinan Erick Thohir adalah penting bagi PSSI untuk mempertahankan kinerjanya dan lebih mendengarkan aspirasi dari para penggemar sepak bola Indonesia. Penelitian selanjutnya adalah dapat menambah data set yang diperoleh dari berbagai media sosial agar model klasifikasi lebih stabil, selain itu juga pada proses pelabelan dapat menggunakan jenis library lain seperti *vader* agar model yang dihasilkan lebih optimal.

## DAFTAR PUSTAKA

- Adam, A. (2022). Data Preprocessing: Pengertian, Manfaat, dan Tahapan Kerjanya. Retrieved from <https://accurate.id/teknologi/data-preprocessing/>
- Adelia Damayanti, A. K. (2023). Analisis Sentimen Masyarakat Terhadap Pembatalan Piala Dunia U20 di Indonesia Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Ilmiah Terapan, Sains dan Teknologi (JITSI)*, 1(2), 97-103.
- Afika Rianti, N. W. (2023). CRISP-DM: Metodologi Proyek Data Science. *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*. e-ISSN 2962-1968. UNIVERSITAS DUTA BANGSA SURAKARTA.
- Coursera, S. (2024). Apa itu Pemrosesan Bahasa Alami? Pengertian dan Contohnya. Retrieved from [https://www-coursera-org.translate.goog/articles/natural-language-processing?\\_](https://www-coursera-org.translate.goog/articles/natural-language-processing?_)
- Dianda Rifaldi, A. F. (2023). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet "Mental Health". *DECODE: Jurnal Pendidikan Teknologi Informasi*, 3(2). doi:<https://doi.org/10.51454/decode.v3i2.131>
- Esa Anindika Sari, M. T. (2020). Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vector Machine. *JIKO (Jurnal Informatika dan Komputer)*, 3(3). doi:<https://doi.org/10.33387/jiko.v3i3.2364>
- He, X. (2024). Sentiment Classification of Social Media User Comments Using SVM Models," 2024. *5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, (pp. 1755-1759). Nanjing, China.
- Kurniawati, Y. (2023). Analisis Sentimen dan Jenisnya. Retrieved from <https://sis.binus.ac.id/2023/11/24/analisis-sentimen-dan-jenisnya/>
- Lerekoqholosha. (2024). TextBlob In Python.". Retrieved from <https://lerekoqholosha9.medium.com/textblob-in-python-c4f60038a65a>
- Muhammad Fadli Asshiddiqi, K. M. (2020). Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI. *eProceedings of Engineering*, 7(3).
- Musfiroh Nurjannah, H. H. (2013). Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 8(3), 110-113. doi:<http://dx.doi.org/10.30872/jim.v8i3.113>
- Nayottama, F. R. (2023). ANALISIS FRAMING ERICK THOHIR SEBAGAI KETUA UMUM PSSI DALAM PEMBERITAAN MEDIA REPUBLIKA ONLINE DAN KOMPAS. COM. 6(2). *Jurnal Ilmu Komunikasi*, 6(2), 201-222.
- Prasetyo, S. M. (2023). Aplikasi Statistik Perbandingan Sepak Bola Yang Bermain Di Champions League. *OKTAL: Jurnal Ilmu Komputer dan Sains*, 2(07).
- Salam, F. A. (2021). Aksiologi Penggunaan Var Dalam Industri Olahraga. *Jurnal Penjakora*, 8(2), 106-113.

- Sihombing, D. O. (2022). Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(2), 396-406.
- Sofana Iwan, P. B. (2021). *Implementasi Artificial Intelligence dan Machine Learning*. Bandung: Penerbit INFORMATIKA.
- Tuhenay Deglorians, M. E. (2021). “ Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (Nbc) Dan Support Vector Machine (SVM)”. *JIKO (Jurnal Informatika dan Komputer)*, 4(2), 105 – 111.
- Yanuarti, R. (2021). Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 6(2), 121-130.
- Zulhidayat, M. (2022). Disfungsi Peran Satgas Anti Mafia Bola Dalam Memberantas Match Fixing Sepakbola Indonesia. *Jurnal Ilmiah Dunia Hukum*, 6(2), 93. doi:<https://doi.org/10.35973/jidh.v6i2.3240>