PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN PADA MEDIA SOSIAL X TERHADAP PERFORMA TIM NASIONAL SEPAK BOLA INDONESIA DI ERA KEPEMIMPINAN SHIN TAE-YONG

Wulan Alfiyani, Doni Abdul Fatah, Firli Irhamni
Sistem Informasi, Universitas Trunojoyo Madura
Jalan Raya Telang, PO BOX 02 Kec. Kamal, Bangkalan
190441100061@student.trunojoyo.ac.id

ABSTRAK

Sepak bola merupakan olahraga yang paling diminati di Indonesia, dengan 77% penduduk menunjukkan ketertarikan terhadapnya. Selain itu, menurut laporan The Global Digital Football Benchmark mencatat enam klub sepak bola Indonesia masuk dalam daftar 200 klub terpopuler dunia. Platform media social X berperan sebagai sarana utama bagi masyarakat dalam menngekspresikan opini, termasuk opini terkait performa Tim Nasional Indonesia. Penelitian ini menganalisis sentimen masyarakat terhadap kinerja Tim Nasional Indonesia selama masa kepelatihan Shin Tae-yong. Data berupa 393 tweet diklasifikasikan ke dalam sentimen positif dan negatif menggunakan algoritma Naive Bayes. Sebelum analisis, data diproses melalui tahapan *seperti cleaning, tokenize, transform cases, stopword removal* dan *filtering* untuk meningkatkan akurasi. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mencapai tingkat akurasi yang sangat tinggi, yaitu 98,73%, dengan *precision* 100, 00% (*positive class* : positif) dan *recall* 97,50% (*positive class* : positif), menandakan bahwa naïve bayes mampu mengklasifikasikan tweet dengan sangat baik. Dari 393 tweet yang dianalisis, sentiment positif lebih dominan dengan 50,9% atau 200 tweet, sedangkan sentiment negatif sebanyak 49,1% atau 193 tweet, menunjukkan bahwa opini terhadap topik yang diteliti cukup seimbang, meskipun lebih condong ke arah yang positif. Penelitian ini memberikan pemahaman mendalam mengenai persepsi masyarakat terhadap kinerja Tim Nasional di bawah kepemimpinan Shin Tae-yong.

Kata kunci : *Timnas; Sentimen; Niave Bayes; Sosial Media X; Data Mining*

1. PENDAHULUAN

Sepak bola merupakan olahraga yang paling diminati di Indonesia dan dunia. Sepak bola juga menjadi sarana perjuangan nasionalisme yang dipelopori oleh Ir. Soeratin Sosrosoegondo melalui pendirian PSSI. Hingga saat ini, sepak bola tetap menjadi bagian penting dari kehidupan masyarakat Indonesia, dengan sekitar 77% penduduk menunjukkan minat terhadap olahraga ini [1].

Bahkan, enam klub sepak bola Indonesia tercatat masuk dalam daftar 200 klub terpopuler dunia. Berdasarkan perkembangan teknologi yang pesat, media sosial kini berfungsi sebagai salah satu platform utama bagi masyarakat dalam mengekspresikan opini dan sentiment terhadap berbagai isu, termasuk di bidang olahraga [2].

Media sosial, seperti X, memungkinkan pengguna untuk berbagi informasi, berinteraksi, serta menyampaikan pandangan pribadi mereka secara langsung. Berbagai topik, mulai dari politik hingga olahraga, menjadi bahan diskusi yang luas di media sosial, termasuk performa tim olahraga nasional [3].

Kinerja Timnas Indonesia, khususnya di bawah arahan pelatih Shin Tae Yong, telah menarik perhatian publik dengan memicu berbagai sentimen di media sosial. Performa tim yang dinamis, ekspektasi tinggi masyarakat, dan strategi pelatih yang menjadi perdebatan melahirkan opini yang beragam. Ketika Timnas meraih kemenangan, banyak netizen menyampaikan pujian dan apresiasi. Namun, kekalahan sering kali diiringi kritik tajam dan komentar negatif.

Pendekatan analisis sentiment terbukti efektif dalam memahami persepsi masyarakat terhadap suatu peristiwa atau isu tertentu [4].

Dalam konteks pertandingan sepak bola, analisis ini memungkinkan identifikasi sentiment masyaraka baik positif, negatif, maupun netral, memberikan wawasan berharga bagi pengambil keputusan, seperti klub sepak bola, penyelenggara acara, dan stakeholder lainnya. Penelitian ini bertujuan untuk menganalisis sentimen terhadap performa Tim Nasional Indonesia di era kepelatihan Shin Tae-yong menggunakan algoritma Naive Bayes. Algoritma ini dipilih berdasarkan pada beberapa alas an utama, pertama Naïve Bayes dikenal sebagai metode klasifikasi yang memiliki efisiensi tinggi dalam menangani data teks yang besar dan tidak terstruktur [5].

Kedua algoritma ini memiliki asumsi independensi antar fitur yang meskipun sederhana dapat menghasilkan klasifikasi yang sangat akurat pada analisis sentiment. Ketiga, metode ini telah terbukti efektif dalam berbagai penelitian sebelumnya untuk analisis sentiment media social. Dibandingkan dengan metode lain seperti *Support Vector Machine (SVM)* dan Random Forest, Naïve Bayes lebih cepat dalam proses pelatihan model dan lebih mudah diterapkan tanpa memerlukan parameter tuning yang kompleks. Hasil penelitian memberikan pemahaman

bagi pihak – pihak yang terlibat untuk mengevaluasi tanggapan masyarakat dan meningkatkan kinerja tim.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah metode yang menggabungkan Teknik pemrosesan Bahasa alami (NLP), *data mining*, dan pembelajaran mesin untuk mendeteksi serta mengklasifikasikan opini dalam teks [6].

Tujuan utamanya adalah memahami sikap atau perasaan seseorang terhadap suatu subjek, yang umumnya dikategorikan sebagai sentimen positif, negatif, atau netral. Pendekatan ini banyak diterapkan dalam berbagai bidang, seperti pemasaran, politik, dan penelitian media sosial, untuk memperoleh wawasan yang lebih mendalam mengenai persepsi publik. Analisis sentimen bekerja dengan menganalisis pola kata, frasa, dan struktur teks untuk mengidentifikasi emosi dan opini yang terkandung di dalamnya. Teknik yang digunakan dapat berupa pendekatan berbasis kamus (lexicon-based), pendekatan berbasis pembelajaran mesin (machine learning-based), atau kombinasi dari keduanya.

2.2. Analisis Sentimen Pada Media Sosial X

Media sosial X yang sebelumnya dikenal dengan Twitter merupakan platform berbasis teks yang memungkinkan pengguna untuk membagikan pendapat, berita, dan opini dalam bentuk posting singkat[8].

Karakteristik media sosial X yang berbasis mikroblogging menjadikannya sumber data yang kaya untuk analisis sentimen, terutama dalam melihat respons publik terhadap berbagai topik, termasuk kinerja tim sepak bola nasional Indonesia.

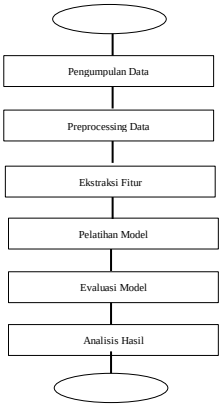
Pada media sosial X analisis sentimen dilakukan dengan mengumpulkan data dari postingan pengguna, kemudian mengklasifikasikannya ke dalam kategori sentimen tertentu. Tantangan utama dalam analisis sentimen media sosial X meliputi:

a. Kependekan teks**:** Postingan di media sosial X sering kali memiliki jumlah karakter yang terbatas, sehingga informasi yang tersedia lebih ringkas dan terkadang kurang jelas.
b. Bahasa informal**:** Pengguna sering kali menggunakan bahasa gaul, slang, singkatan, dan emotikon yang dapat mempersulit analisis sentimen.
c. Adanya ironi dan sarkasme**:** Beberapa opini yang diekspresikan dalam bentuk sarkasme sulit untuk diklasifikasikan dengan metode berbasis aturan atau kamus.
d. Perubahan tren**:** Opini publik dapat berubah dengan cepat berdasarkan peristiwa atau hasil pertandingan tertentu.

2.3. Algoritma Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang sering digunakan dalam analisis sentiment. Algoritma ini bekerja berdasarkan Teorema Bayes dengan asumsi independensi antara fitur-fitur dalam data. Naïve Bayes disebut "naïve" karena mengasumsikan bahwa setiap fitur dalam dataset bersifat independen satu sama lain, meskipun dalam kenyataannya, fitur – fitur ini sering kali memiliki keterkaitan[9].

Penerapan algoritma Naïve Bayes dalam analisis sentimen pada media sosial X terhadap performa tim nasional sepak bola Indonesia melibatkan beberapa langkah utama:



Gambar 1. Algoritma Naïve Bayes

a. Pengumpulan Data**:** Mengambil data dari postingan pengguna yang membahas performa tim nasional di media sosial X menggunakan API atau web scraping.
b. Preprocessing Data**:**
   - Menghapus karakter khusus, tanda baca, dan tautan dari data.
   - Menghilangkan stopwords.
   - Melakukan stemming dan tokenisasi.
c. Ekstraksi Fitur**:**
   Menggunakan metode TF-IDF atau Bag of Words untuk merepresentasikan teks dalam bentuk numerik.
d. Pelatihan Model**:**
   Pelatihan model Naïve Bayes dimulai dengan menghitung probabilitas setiap kata dalam kategori sentimen positif dan negatif dari data pelatihan. Model kemudian menerapkan Teorema Bayes untuk menentukan probabilitas suatu teks termasuk dalam kategori tertentu. Akhirnya, teks diklasifikasikan berdasarkan probabilitas tertinggi, sehingga setiap tweet dikategorikan sebagai sentimen positif atau negatif dengan tingkat kepercayaan tinggi.
e. Evaluasi Model**:**
   Menguji akurasi model menggunakan metrik seperti precision, recall, dan F1-score.
f. Analisis Hasil**:**
   Menafsirkan hasil analisis sentimen untuk memahami opini publik terhadap performa tim nasional di era kepemimpinan Shin Tae-Yong.

2.4. Penelitian Terdahulu

Dalam penelitian yang dilakukan oleh D. R. P. Jaya dan S. Lestari berjudul *"Analisis Sentimen Kepuasan Publik Terhadap Masa Kepemimpinan Shin Tae-yong Menggunakan Algoritma Naïve Bayes,"* hasil penelitian menunjukkan bahwa kombinasi kedua metode tersebut menghasilkan akurasi yang lebih tinggi dibandingkan dengan pendekatan tunggal [3].

Sementara itu pada penelitian M. R. Amin dan A. N. Salma dengan judul "Analisis Sentimen Opini Publik Mengenai Tragedi Kanjuruhan Pada Media Sosial Twitter Menggunakan Brand24," Penelitian ini menggunakan Brand24 untuk menganalisis sentimen publik terhadap tragedi Kanjuruhan di media sosial Twitter. Studi ini menunjukkan bagaimana opini publik dapat dikumpulkan dan diklasifikasikan menggunakan alat analisis sentimen otomatis [2].

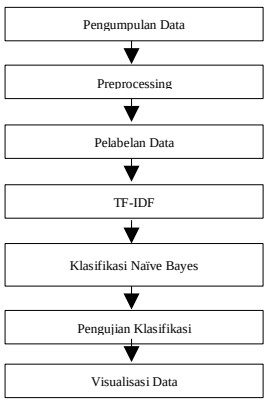
Penelitian lain Zuhrival Penelitian ini berfokus pada analisis sentimen terhadap performa Tim Nasional Indonesia selama kepemimpinan Shin Tae-yong dengan menggunakan algoritma Naïve Bayes. Studi ini memberikan wawasan tentang bagaimana metode probabilistik dapat diterapkan dalam menganalisa opini publik terhadap performa tim nasional [4].

Penelitian yang dilakukan oleh Nugraha et al. Studi ini menganalisis sentimen kepuasan publik terhadap masa kepemimpinan Shin Tae-yong menggunakan Naïve Bayes. Hasil penelitian ini menunjukkan bahwa mayoritas opini publik terhadap pelatih memiliki kecenderungan positif [5].

Penelitian lain yang dilakukan oleh T. R. P. Hermawan dan A. R. Dzikrillah Studi ini menganalisis penerapan algoritma Naïve Bayes dalam menilai sentimen ulasan pengguna terhadap aplikasi ChatGPT. Studi ini memberikan contoh penerapan Naïve Bayes pada platform digital [10].

3. METODE PENELITIAN

Diagram proses penelitian menggambarkan alur yang menjelaskan tahapan pelaksanaan penelitian, dimulai dari penetapan metode hingga tahap akhir berupa analisis. Ilustrasi proses penelitian dalam studi ini dapat dilihat pada ilustrasi berikut.



Gambar 2. Metode penelitian

Dari diagram alur di atas dapat dilihat tahapan – tahapan dalam proses penelitian ini, di antaranya adalah :

a. Pengumpulan Data – Langkah pertama dalam proses ini adalah mengumpulkan data yang akan dianalisis.
b. Preprocessing – Data yang dikumpulkan kemudian diproses untuk membersihkan dan menyesuaikan formatnya agar siap untuk analisis lebih lanjut.
c. Pelabelan Data – Data diberi label atau diklasifikasikan berdasarkan kategori tertentu.
d. TF-IDF (Term Frequency - Inverse Document Frequency) – Metode ini digunakan untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen berdasarkan frekuensi kemunculannya dan distribusinya di seluruh dokumen.
e. Klasifikasi (Naïve Bayes) – Algoritma Naïve Bayes digunakan untuk mengklasifikasikan data berdasarkan fitur yang telah diproses.
f. Pengujian Klasifikasi – Evaluasi terhadap hasil klasifikasi untuk mengukur performa model yang digunakan.
g. Visualisasi Data – Hasil akhir dari analisis ditampilkan dalam bentuk visual untuk memudahkan pemahaman.

3.1. Pengumpulan Data

Pengumpulan data merupakan proses memperoleh informasi dari berbagai sumber untuk keperluan analisis dan penelitian. Dalam konteks analisis sentimen, pengumpulan data merujuk pada langkah-langkah yang dilakukan untuk memperoleh teks yang akan dianalisis sentimennya. Penelitian ini menggunakan dataset yang diperoleh dari platform Kaggle, yang merupakan hasil crawling data dari media sosial X. Dataset tersebut dikumpulkan dengan menggunakan teknik crawling yang menggunakan kata kunci seperti “Shin Tae-yong”, “STY” dan “Timnas” untuk menjaring data relevan dari media sosial X. Proses crawling dilakukan menggunakan library tertentu dalam bahasa Python.

3.2. Preprocessing

Preprocessing adalah tahap awal yang penting dalam mempersiapkan data sebelum analisis. Tujuan utamanya adalah memastikan data bersih, terstruktur, dan bebas dari kesalahan, sehingga dapat meningkatkan akurasi hasil analisis, inkonsistensi, serta informasi yang tidak relevan. Proses ini bertujuan untuk meminimalkan potensi kehilangan atau penambahan data yang tidak perlu. Data yang dikumpulkan, khususnya terkait sentimen terhadap timnas sepakbola Indonesia, harus diproses agar siap digunakan dalam analisis lebih lanjut. Untuk melakukan preprocessing, digunakan RapidMiner Studio versi 9.10 sebagai alat utama dalam mengelola dan menguji dataset. Alat ini memungkinkan berbagai metode untuk membersihkan dan menyiapkan data,

Selanjutnya, data yang telah diproses digunakan dalam analisis sentimen. Penggunaan RapidMiner Studio memastikan bahwa dataset yang digunakan dalam analisis dapat dikelola secara optimal. memungkinkan peningkatan dan evaluasi model yang lebih baik.

Beberapa langkah utama dalam preprocessing meliputi *cleaning*, *stemming*, *case folding*, *remove stopword*, dan *tokenization* [11].

Pada tahap *cleaning* dan *case folding*, Data dibersihkan dari elemen yang tidak diperlukan, seperti tanda baca, hashtag, username, dan spasi berlebih. Selain itu, huruf diubah menjadi huruf kecil untuk mengurangi kompleksitas pada tahap analisis.

Proses ini dilakukan menggunakan operator *replace* dan *transform cases* di RapidMiner Studio. *Stemming* bertujuan mengubah kata-kata ke bentuk dasarnya dengan menghilangkan akhiran dan variasi kata. menyederhanakan analisis. Setelah itu, tahap *remove stopword* dilakukan untuk Menghilangkan kata-kata yang tidak memberikan nilai tambah dalam analisis, seperti kata penghubung dan kata sambung. Misalnya, kata-kata seperti "ada", "adalah", dan "atau" dihilangkan menggunakan kamus stopword berjumlah 758 kata yang diperoleh dari situs kaggle.com. Proses ini pada RapidMiner Studio menggunakan operator *filter stopwords (Dictionary)*. Beberapa kata, seperti "belum", "akhir", dan "lagi", dipertahankan untuk relevansi konteks. Terakhir, proses *tokenization* digunakan untuk nemecah kalimat menjadi kata-kata atau token untuk mempermudah pemrosesan dan analisis data lebih lanjut. Semua tahapan ini sangat penting dalam mempersiapkan data untuk model analisis sentimen yang lebih akurat.

3.3. Pelabelan sentimen

Setelah tahap preprocessing data, langkah selanjutnya adalah memberikan label sentimen pada setiap data, yang dilakukan baik secara otomatis maupun manual. Proses pelabelan otomatis menggunakan RapidMiner untuk memastikan klasifikasi data yang lebih efisien dan akurat. RapidMiner menyediakan berbagai alat dan teknik untuk mengotomatiskan proses pelabelan data berdasarkan fitur yang telah ditentukan.

Proses pelabelan diawali dengan pemrosesan awal data, seperti *cleaning*, *tokenization*, *case transformation*, dan *stopword removal*. Setelah data dipersiapkan, teknik *supervised learning* diterapkan dengan algoritma Naïve Bayes digunakan untuk mengklasifikasikan sentimen secara otomatis ke dalam dua kategori utama yaitu positif dan negatif. Perbandingan antara pelabelan otomatis dan manual ini dilakukan untuk memverifikasi konsistensi dan akurasi hasil pelabelan [12].

RapidMiner memanfaatkan operator *Text Processing* untuk mengekstrak fitur dari teks dan mengubahnya menjadi representasi numerik yang dapat digunakan dalam model pembelajaran mesin. Operator *Set Role* digunakan untuk menentukan label target sentimen, sedangkan *Apply Model* dan *Performance* digunakan untuk mengukur akurasi hasil pelabelan.

3.4. Pembobotan TF-IDF

Setelah data melalui tahap pra-pemrosesan, data tersebut perlu dikonversi ke bentuk numerik agar dapat diproses lebih lanjut. Proses ini melibatkan transformasi data ke dalam bentuk vektor, penentuan nilai dan bobot untuk setiap kata, serta penerapan algoritma prediksi. Salah satu metode yang digunakan dalam tahap pemberian bobot kata adalah Term-Frequency Times Inverse Document Frequency (TF-IDF). Metode ini dipilih karena kemudahan implementasi, efisiensinya, serta kemampuannya menghasilkan hasil yang akurat. TF-IDF menghitung dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF), untuk setiap kata dalam dokumen [10].

Secara sederhana, metode ini mengukur seberapa sering suatu kata muncul dalam dokumen. TF merepresentasikan frekuensi kemunculan kata dalam dokumen, di mana semakin sering kata tersebut muncul, semakin besar bobot atau tingkat relevansinya terhadap dokumen tersebut. Dengan metode ini, informasi teks dapat direpresentasikan dalam bentuk numerik yang lebih optimal, sehingga dapat diproses lebih lanjut oleh algoritma prediksi.

3.5. Klasifikasi Naïve bayes

Naïve Bayes digunakan dalam analisis sentimen untuk melakukan klasifikasi teks [13].

Proses ini dimulai dengan pengolahan data yang telah melewati tahap pra-pemrosesan, termasuk pembobotan kata menggunakan metode TF-IDF. Setelah model dilatih dengan data tersebut, dilakukan pengujian untuk menilai akurasi hasil klasifikasi. Pelatihan model adalah langkah untuk mengajarkan model cara memahami data yang telah diproses. Pada metode Naïve Bayes, data yang telah diproses digunakan untuk menghitung probabilitas kemunculan setiap kata dalam kelas positif dan negatif. Probabilitas ini kemudian digunakan untuk memprediksi kategori teks yang belum pernah dilihat sebelumnya. Setelah mempelajari pola dari data pelatihan, model Naïve Bayes siap untuk mengklasifikasikan teks baru. Teorema Bayes secara umum menyatakan bahwa peluang terjadinya suatu kejadian A, dengan asumsi B telah terjadi, dapat dihitung dengan mengalikan peluang B diberikan A, dan membagi hasilnya dengan peluang B [15].

Teorema ini memungkinkan pembaruan estimasi probabilitas berdasarkan bukti baru, dan sangat berguna dalam statistik serta pembelajaran mesin karena memungkinkan perhitungan peluang yang lebih akurat. Secara matematis, Teorema Bayes dinyatakan dalam rumus berikut :

$$P(H|E) = \frac{P(E|H) - P(H)}{P(E)} \quad (1)$$

a. $P(H|E)$ = Probabilitas kejadian H terjadi jika diketahui E telah terjadi (probabilitas posterior).
b. $P(E|H)$ = Probabilitas E terjadi jika H terjadi (probabilitas likelihood).
c. $P(H)$ = Probabilitas awal dari H (probabilitas prior).
d. $P(E)$ = Probabiliyas E terjadi (probabilitas evidence).

Pada tahap pengujian, model Naïve Bayes digunakan untuk mengklasifikasikan teks yang belum pernah dilihat. Hasil klasifikasi kemudian dibandingkan dengan kelas asli untuk mengevaluasi kinerja model menggunakan metrik seperti akurasi, presisi, dan recall.

3.6. Pengujian Klasifikasi

Tahap evaluasi bertujuan untuk menilai tingkat akurasi dan efektivitas model yang telah dikembangkan. Proses ini dilakukan menggunakan confusion matrix untuk mengukur sejauh mana model Naïve Bayes mampu mengklasifikasikan data dengan baik [16]. Selain itu, evaluasi ini juga berfungsi untuk menentukan apakah model sudah layak digunakan atau perlu dilakukan perbaikan dan pengembangan ulang jika hasilnya belum memenuhi ekspektasi. Penilaian kinerja model dilakukan dengan memanfaatkan berbagai metrik, seperti akurasi, presisi, dan recall.

3.7. Visualisai

Penelitian ini juga memanfaatkan word cloud sebagai alat visualisasi untuk mempermudah pemahaman terhadap kata-kata berdasarkan frekuensi kemunculannya. Word cloud efektif dalam menampilkan kata-kata yang paling dominan atau sering muncul, sehingga memberikan visualisasi yang jelas mengenai fokus atau tema utama dalam setiap kategori sentiment [17].

Tujuan utamanya adalah untuk mempermudah pemahaman hasil analisis data secara intuitif. Word cloud menjadi alat yang kuat untuk menyampaikan informasi.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Proses pengumpulan atau *crawling* data dari media social X menggunakan beberapa kata kunci seperti “Shin Tae yong”, “STY”, “Timnas”, “Dendy”. *Crawling* data dilakukan secara bertahap dari tanggal 25 Desember 2024 – 15 Februari 2025. Penelitian ini menggunakan sebanyak 393 tweet sebagai data analisis.

Gambar 3. Hasil *crawling* data

Gambar di atas menunjukkan hasil proses *crawling data* dari platform media sosial. Data yang diperoleh mencakup username dan teks unggahan pengguna. Informasi ini nantinya akan digunakan dalam tahap *preprocessing* untuk analisis lebih lanjut.

4.2. Preprocessing

Pada tahap ini, data yang diperoleh melalui proses *crawling* diolah menjadi format yang siap untuk dianalisis. Proses preprocessing mencakup beberapa langkah seperti *cleaning*, *tokenize*, *transform cases*, *remove stopword*, dan *filtering*. Beberapa langkah tersebut antara lain adalah:

4.2.1. Cleaning

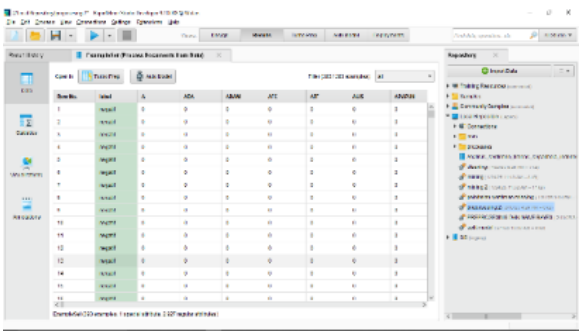
Proses ini bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dan tidak bermakna, seperti username (@), hashtag (#), tanda baca, angka, karakter tunggal, mention, retweet, dan lainnya. Penghapusan elemen-elemen tersebut dilakukan untuk mengurangi beban pemrosesan serta Memastikan data yang digunakan dalam analisis lebih relevan dan bermakna. Beberapa proses yang dilakukan pada tahapan cleaning adalah menghapus <username>, url, dan symbol yang tidak diperlukan.

Gambar 4. Hasil *cleaning*

Gambar di atas merupakan hasil dari tahap *cleaning* pada proses *preprocessing* data teks. Pada tahap ini, data yang diperoleh dari hasil *crawling* telah dibersihkan dari karakter atau simbol yang tidak diperlukan serta dilakukan normalisasi teks agar lebih terstruktur.

4.2.2. Tokenize

Langkah berikutnya adalah tokenisasi, yaitu proses memecah kalimat menjadi kata-kata atau token.”

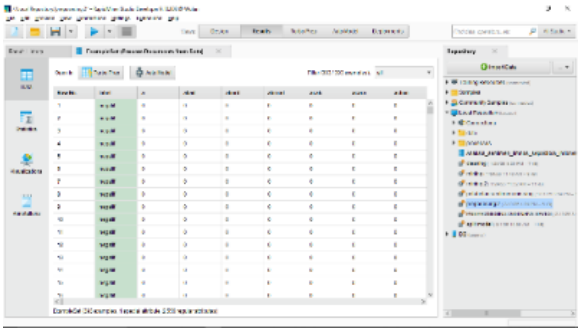


Gambar 5. Hasil *tokenize*

Gambar di atas merupakan hasil dari tahap *tokenization*, di mana setiap teks telah dipecah menjadi token atau kata-kata individual. Setiap baris mewakili satu dokumen atau komentar, dan kolom mempresentasikan frekuensi kemunculan kata dalam dokumen.

4.2.3. Transform Cases

Tahap ini bertujuan untuk mengonversi teks ke dalam format huruf kecil guna memastikan konsistensi data, sehingga meningkatkan akurasi dalam proses analisis.

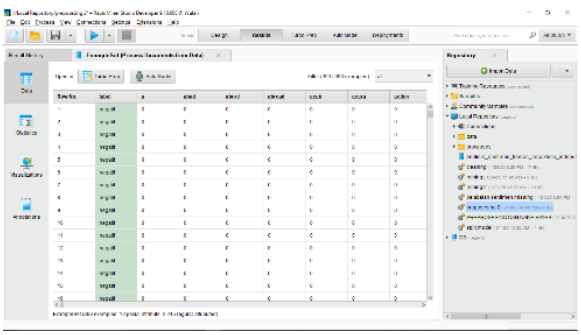


Gambar 6. Hasil *transform cases*

Gambar di atas merupakan hasil dari tahap *Transform Cases*, di mana semua teks dalam dataset telah dikonversi menjadi huruf kecil (*lowercase*). Proses ini bertujuan untuk memastikan konsistensi dalam pemrosesan teks, sehingga kata dengan huruf besar dan kecil tidak dianggap berbeda oleh sistem.

4.2.4. Stopword Removal

Tahap ini bertujuan untuk menghapus kata-kata dengan frekuensi tinggi yang tidak memiliki nilai informatif dalam analisis sentimen, sehingga meningkatkan relevansi data. Proses penghapusan stopword menggunakan daftar kata-kata umum dalam bahasa Indonesia yang diambil dari kaggle.com dan telah banyak digunakan dalam berbagai penelitian. Pada penelitian ini, beberapa kata disesuaikan untuk memenuhi kebutuhan analisis, dengan harapan dapat meningkatkan akurasi tanpa mengurangi makna data.

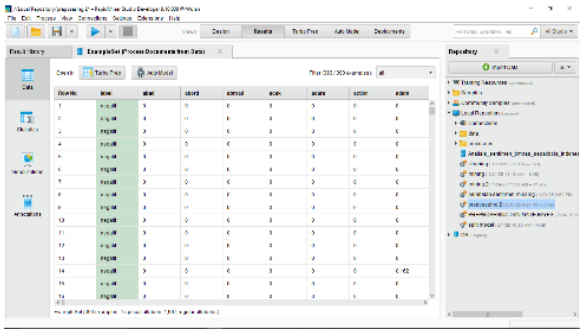


Gambar 7. Hasil *stopword removal*

Gambar di atas merupakan hasil dari tahap *Stopword Removal*, di mana kata-kata yang dianggap tidak memiliki makna penting dalam analisis teks telah dihapus. Penghapusan stopword memungkinkan analisis lebih terfokus pada kata-kata yang memiliki relevansi tinggi dan berkontribusi signifikan terhadap hasil klasifikasi.

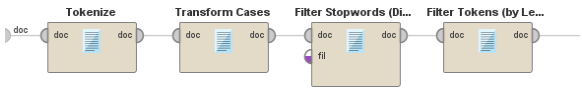
4.2.5. Filtering

Pada tahap ini, dilakukan eliminasi kata-kata yang terlalu pendek, berbentuk singkatan, atau terlalu panjang. Langkah ini bertujuan untuk menyaring kata-kata yang tidak memberikan kontribusi signifikan terhadap makna kalimat guna meningkatkan akurasi analisis data.



Gambar 8. Hasil *filtering*

Gambar di atas merupakan hasil proses filtering, di mana kata-kata yang tidak relevan atau memiliki nilai informasi rendah dalam analisis telah dieliminasi. Pada tahap ini, hanya kata-kata dengan panjang antara 3 hingga 25 karakter yang dipertahankan, sementara kata yang terlalu pendek atau terlalu panjang dihapus. Proses ini bertujuan untuk meningkatkan kualitas data dengan menyaring kata-kata yang tidak signifikan sebelum masuk ke tahap pemrosesan lebih lanjut.



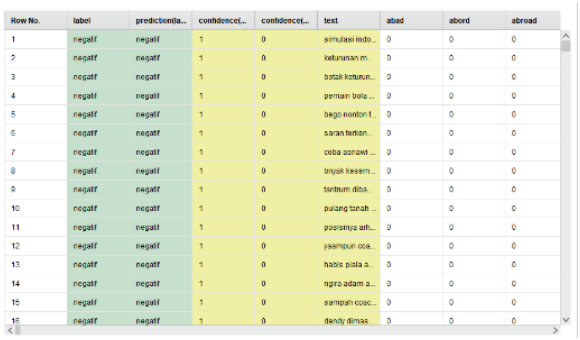
Gambar 9. Proses *tokenize, transform case, stopword removal, filtering*

Gambar di atas menunjukkan proses *preprocessing* teks, mulai dari tokenisasi, normalisasi huruf, penghapusan *stopwords*, hingga pemfilteran kata berdasarkan panjang (3-25 karakter) untuk meningkatkan kualitas data sebelum analisis lebih lanjut.

4.3. Pelabelan Kelas Sentimen

Setelah tahap preprocessing selesai, langkah berikutnya adalah pelabelan data sentimen. Proses ini bertujuan untuk mengklasifikasikan data berdasarkan sentimen yang terkandung, sehingga dapat dibedakan apakah data tersebut memiliki sentimen positif atau negatif. Pelabelan dilakukan melalui dua metode, yaitu otomatis menggunakan RapidMiner dan secara manual. Proses ini menghasilkan data yang diklasifikasikan ke dalam dua kategori, yaitu sentimen

positif dan sentimen negatif. Selain itu, pelabelan manual juga dilakukan oleh 3 validator yang membaca dan menilai sentimen teks berdasarkan pemahaman kontekstual. Pendekatan manual sering dianggap lebih akurat karena mempertimbangkan konteks dan emosi yang berpotensi tidak sepenuhnya terdeteksi oleh metode otomatis. Pelabelan manual juga digunakan untuk menghasilkan dataset referensi yang berfungsi untuk melatih dan mengevaluasi model otomatis.



Row No.	label	predictionla...	confidence(...	confidence(...	text	abad	abord	abroad
1	negatif	negatif	1	0	simulasi indo...	0	0	0
2	negatif	negatif	1	0	keturunan m...	0	0	0
3	negatif	negatif	1	0	botak keturun...	0	0	0
4	negatif	negatif	1	0	pemain bola ...	0	0	0
5	negatif	negatif	1	0	bego nonton t...	0	0	0
6	negatif	negatif	1	0	saran terken...	0	0	0
7	negatif	negatif	1	0	coba asnawi ...	0	0	0
8	negatif	negatif	1	0	bnyak kesem...	0	0	0
9	negatif	negatif	1	0	tantrum diba...	0	0	0
10	negatif	negatif	1	0	pulang tanah ...	0	0	0
11	negatif	negatif	1	0	posisinya arh...	0	0	0
12	negatif	negatif	1	0	yaampun coa...	0	0	0
13	negatif	negatif	1	0	habis piala a...	0	0	0
14	negatif	negatif	1	0	ngira adam a...	0	0	0
15	negatif	negatif	1	0	sampah coac...	0	0	0
16	negatif	negatif	1	0	dendy dimas ...	0	0	0

Gambar 10. Hasil pelabelan sentiment

Gambar di atas menunjukkan hasil pelabelan sentimen menggunakan RapidMiner, di mana setiap komentar dikategorikan sebagai sentimen negatif atau positif berdasarkan prediksi model dengan tingkat kepercayaan yang ditampilkan dalam kolom confidence.

4.4. Visualisasi

Bagian ini membahas hasil analisis data yang diperoleh melalui pembuatan word cloud. Word cloud dibuat berdasarkan data teks yang telah diproses sebelumnya, khususnya yang berkaitan dengan sentimen terhadap tim nasional sepak bola Indonesia. Analisis data menggunakan word cloud bertujuan untuk memvisualisasikan frekuensi kemunculan kata dalam dataset, di mana kata-kata yang muncul lebih sering ditampilkan dengan ukuran lebih besar. Pendekatan visual ini membantu dalam mengidentifikasi informasi penting dan menyimpulkan topik utama yang paling dominan dari data yang telah dianalisis.

Gambar 11. Visualisasi wordcloud

Gambar di atas menunjukkan hasil visualisasi data teks menggunakan WordCloud, di mana kata-kata dengan frekuensi tinggi ditampilkan dalam ukuran lebih besar dan variasi warna untuk menyoroti dominasi kata dalam dataset.

4.5. Naïve Bayes

Metode yang diterapkan dalam penelitian ini adalah Naïve Bayes Classification. Data set yang digunakan sebanyak 393 data, dengan banyaknya dataset yang digunakan dalam pengujian ini diharapkan dapat memperoleh hasil akurasi yang maksimal. Berikut merupakan proses analisis naïve bayes yang dilakukan, setelah data melalui beberapa tahapan antara lain *cleaning*, *tokenize*, *stopword*, *filtering*, barulah data dapat diproses pada tahapan analisis naïve bayes. Berikut hasil dari data uji yang telah melalui proses analisis :

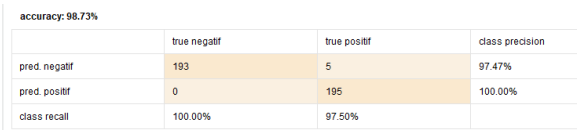
Row No.	label	prediction(la...	confidence(...	confidence(...	text
1	negatif	negatif	1	0	simulasi indo...
2	negatif	negatif	1	0	keturunan m...
3	negatif	negatif	1	0	botak keturun...
4	negatif	negatif	1	0	pemain bola ...
5	negatif	negatif	1	0	bego nonton t...
6	negatif	negatif	1	0	saran terken...
7	negatif	negatif	1	0	coba asnawi ...
8	negatif	negatif	1	0	bnyak kesem...
9	negatif	negatif	1	0	tantrum diba...
10	negatif	negatif	1	0	pulang tanah ...
11	negatif	negatif	1	0	posisinya arh...
12	negatif	negatif	1	0	yaampun coa...
13	negatif	negatif	1	0	habis piala a...
14	negatif	negatif	1	0	ngira adam a...
15	negatif	negatif	1	0	sampah coac...
16	negatif	negatif	1	0	dendy dimas ...

Gambar 12. Hasil proses data uji

Gambar di atas menampilkan hasil proses data uji, di mana setiap komentar telah dikategorikan ke dalam sentimen negatif atau positif berdasarkan prediksi model, dengan tingkat kepercayaan (confidence) yang ditampilkan dalam kolom terkait.

4.6. Pengujian

Pada tahap evaluasi dilakukan perhitungan nilai akurasi yang ditujuka untuk mengetahui tingkat ketepatan penggunaan mode.



accuracy: 98.73%

	true negatif	true positif	class precision
pred. negatif	193	5	97.47%
pred. positif	0	195	100.00%
class recall	100.00%	97.50%	

Gambar 13. Hasil nilai akurasi

Gambar di atas menunjukkan hasil pengujian model dengan akurasi 98,73%. Tabel menampilkan jumlah true negative, true positive, serta nilai precision dan recall untuk masing-masing kelas sentimen. Pada tahap evaluasi hasil yang diperoleh dari penerapan algoritma naïve bayes ini berupa nilai confussion matrix berisi nilai akurasi, presisi, dan recall yang diambil dari data test. Nilai akurasi yang didapat dari algoritma naïve bayes adalah 98,73%, artinya

sejumlah 98,73% dapat mengklasifikasikan data dengan benar. Berdasarkan penelitian yang telah dilakukan, sentiment positef terhadap timnas Indonesia lebih banyak dibandingkan sentiment negatif, yang berarti bahwa masyarakat memiliki tingkat percaya lebih kepada timnas Indonesia.

Langkah ini sangat penting untuk cepat harus diimplementasikan untuk memulihkan kepercayaan dan dukungan penggemar. Langkah-langkah ini sangat penting untuk memastikan kesuksesan jangka panjang Timnas Sepak Bola menyebabkan ketidakpuasan, meningkatkan menghadapi tantangan ini, tim manajemen perlu memprioritaskan identifikasi masalah utama yang transparansi, dan memperbaiki komunikasi dengan para penggemar. Selain itu, tindakan perbaikan yang Indonesia. Dengan pendekatan yang responsif dan proaktif, diharapkan hubungan antara tim dan penggemar dapat diperkuat, serta citra positif tim dapat dibangun kembali.

4.7. Analisa Hasil

Hasil analisis sentiment yang dilakukan terhadap opini masyarakat di media social X mengenai Timnas Sepak Bola Indonesia era kepemimpinan Shin Tae-yong diperoleh distribusi sentiment sebagai berikut :

Tabel 1. Hasil sentimen

Sentimen	Presentase
Negatif	49,1%
Positif	50,9%

Tabel 1 menunjukkan hasil analisis sentimen dari data yang telah diproses. Dari hasil yang diperoleh, sebanyak 49,1% komentar atau teks dikategorikan sebagai sentimen negatif**,** sedangkan 50,9% masuk dalam sentimen positif**.** Hal ini menunjukkan bahwa terdapat keseimbangan antara sentimen negatif dan positif dalam data yang dianalisis, dengan sentimen positif sedikit lebih dominan. Untuk mengevaluasi performa model dalam klasifikasi sentimen, beberapa metrik evaluasi digunakan, yaitu akurasi, presisi dan recall.

4.7.1. Akurasi (*accuracy)*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

a. TP (*true positive*): Jumlah data positif yang dihasilkan sebagai positif
b. TN (*true negative*): jumlah data negative yang diklasifikasikan sebagai negative
c. FP (*false positive*): jumlah data negative yang salah diklasifikasikan sebagai positif.
d. FN (*false negative*): jumlah data positif yang salah diklasifikasikan sebagai negatif.

Akurasi mengukur sejauh mana model dapat mengklasifikasikan data dengan benar. Dalam penelitian ini, model Naïve Bayes mencapai 98,73%**,** yang menunjukkan tingkat keandalan yang sangat tinggi.

4.7.2. Presisi (*Precision)*

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

a. TP (*true positive*): Jumlah data positif yang dihasilkan sebagai positif
b. FP (*false positive*): jumlah data negative yang salah diklasifikasikan sebagai positif.

Presisi mengukur proporsi prediksi positif yang benar terhadap semua prediksi positif. Model memperoleh presisi 100% untuk kelas sentimen positif, yang berarti tidak ada kesalahan dalam mengklasifikasikan tweet sebagai sentimen positif.

4.7.3. Recall

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

a. TP (*true positive*): Jumlah data positif yang dihasilkan sebagai positif
b. FN (*false negative*): jumlah data positif yang salah diklasifikasikan sebagai negatif.

Recall mengukur sejauh mana model dapat mendeteksi semua data yang benar-benar positif. Hasil penelitian menunjukkan bahwa recall model adalah 97,50%**,** yang berarti sebagian besar tweet positif dapat diklasifikasikan dengan benar.

Hasil penelitian mengindikasikan bahwa algoritma Naïve Bayes mampu mengklasifikasikan tweet dengan sangat baik. Dari 393 tweet yang dianalisis, sentiment positif lebih dominan dengan 50,9% atau 200 tweet, sedangkan sentiment negatif sebanyak 49,1% atau 193 tweet, menunjukkan bahwa opini terhadap topik yang diteliti cukup seimbang, meskipun lebih condong ke arah yang positif.

5. KESIMPULAN DAN SARAN

Penelitian ini menguunakan algoritma naïve bayes yang menghasilkan akurasi model sebesar 98,73%, dengan *precision* 100, 00% (*positive class*: positif) dan *recall* 97,50% (*positive class* : positif). Hal ini menunjukkan bahwa algoritma naïve bayes efektif dalam mengklasifikasikan sentiment pada data yang digunakan. Hasil Analisa sentiment menunjukkan bahwa sentiment positif lebih dominan dengan 50,9% atau 200 tweet, sedangkan sentiment negatif sebanyak 49,1% atau 193 tweet, menunjukkan bahwa opini terhadap topik yang diteliti cukup seimbang, meskipun lebih condong ke arah yang positif. Sebagai rekomendasi untuk penelitian selanjutnya, disarankan untuk membandingkan performa algoritma Naïve Bayes dengan pendekatan machine learning lainnya, seperti Support Vector Machine (SVM) atau Random Forest. Pendekatan ini akan memberikan perspektif yang lebih luas mengenai keunggulan dan keterbatasan masing-masing metode dalam analisis sentimen.

DAFTAR PUSTAKA

[1]. S. Pada, M. Laki, L. Penonton, U. Tentang, S. Bola, dan D. Fakultas, “Pengaruh menonton tayangan ulasan tentang sepak bola di televisi terhadap gaya hidup hedonis mahasiswa.”. 2020

[2]. M. R. Amin dan A. N. Salma, "Analisis Sentimen Opini Publik Mengenai Tragedi Kanjuruhan Pada Media Sosial Twitter Menggunakan Brand24," 2024.

[3]. D. Rizky, P. Jaya, dan S. Lestari, "Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors," 2024. [Daring]. Tersedia pada: https://journal.stmiki.ac.id

[4]. G. Zuhrival, "Analisis Sentimen Performa Tim Nasional Sepak Bola Indonesia Era Kepemimpinan Shin Tae-yong pada X Menggunakan Algoritma Naive Bayes." 2024.

[5]. P. Nugraha, F. Matheos Sarimole, dan P. Studi Sistem, "Analisis Sentimen Kepuasan Publik Terhadap Masa Kepemimpinan Shin Tae Yong Menggunakan Algoritma Naïve Bayes," 2025.

[6]. D. Purnamasari *dkk.*, "Pengantar Metode Analisis Sentimen," 2023.

[7]. F. Alim Adiyatma, S. Alam, M. Andayani Komara Teknik Informatika, S. Tinggi Teknologi Wastukancana Purwakarta Jalan Cikopak No, dan J. Barat, "Analisis sentimen masyarakat di platform x terhadap penggunaan bansos untuk memenangkan salah satu capres tertentu di pilpres 2024 menggunakan metode naïve bayes classifier," 2024.

[8]. S. M. Hudzaifah *dkk.*, "Implementasi Algoritma Naïve Bayes dalam Memprediksi Tingkat Kelulusan Siswa pada Sertifikasi Mikrotik Certified Network Associate (MTCNA)," 2024. [Daring]. Tersedia pada: https://journal.stmiki.ac.id

[9]. L. Hermawan, M. B. Ismiati, J. Bangau, N. 60, dan M. Charitas, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *TRANSFORMATIKA*, vol. 17, no. 2, hlm. 188–199, 2020.

[10]. T. Ramadhani, P. Hermawan, dan A. R. Dzikrillah, "Penerapan Metode Naïve Bayes untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi ChatGPT di Google Play Store," *Technology and Science (BITS)*, vol. 6, no. 1, hlm. 430–439, 2024, doi: 10.47065/bits.v6i1.5400.

[11]. I. Amal, "Perbandingan Pelabelan Otomatis Dan Manual Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Twitter Menggunakan Algoritma Support Vector Machine," 2023.

[12]. I. S. Wibowo, A. Witanti, dan I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF", [Daring]. Tersedia pada: http://jurnal.mdp.ac.id. 2024

[13]. A. Jh dan F. Purba, "Perbandingan Metode Bayes Dan Certenty Factor Pada Sistem Pakar Mendiagnosa Penyakit Varisela Pada Anak-Anak," 2020.

[14]. A. W. Saputra, A. I. Purnamasari, dan I. Ali, "Implementasi algoritma naïve bayes untuk memprediksi kualitas air yang dapat di konsumsi," 2024.

[15]. Irvandi, B. Irawan, dan O. Nurdiawan, "Naive Bayes dan wordcloud untuk analisis sentimen wisata halal pulau lombok," *INFOTECH journal*, vol. 9, no. 1, hlm. 236–242, Mei 2023, doi: 10.31949/infotech.v9i1.5322.