

Analisis Sentimen Opini Supporter Pengguna Youtube terhadap Sistem Pembelian Tiket Pertandingan Persib menggunakan Metode Naïve Bayes

Adam Arifian A¹, Rini Astuti², Fadhil Muhamad Basysyar³

^{1,3}Program Studi Teknik Informatika, STMIK IKMI Cirebon

²Program Studi Manajemen Informatika, STMIK IKMI Cirebon

*Email: adamarifianalamsyah@gmail.com

Abstrak

Pemberitaan mengenai Persib tidak terlepas dari peranan media dari era perserikatan hingga masa kini Persib pemberitaan pertama kali di media setidaknya pada november tahun 1904, pada perserikatan priangan (PVB) tercatat sebagai perserikatan pertama di Bandung. Menggunakan Analisis Deskriptif, berbentuk word cloud, yang digunakan dalam penelitian ini untuk mengidentifikasi dan membentuk pola kata yang dapat berasosiasi dengan kata lain yang dianggap penting. Metode Naïve Bayes Classifier, yang digunakan dalam penelitian ini untuk mengidentifikasi dan membentuk pola kata yang dapat berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting. YouTube telah menjadi salah satu platform terbesar untuk berbagi konten visual di internet. Salah satu topik yang ramai diperbincangkan yaitu tentang sistem pembelian tiket pertandingan Persib Bandung. Hal ini banyak mengundang reaksi terutama pada masyarakat khususnya warga Jawa Barat. Sehingga mengakibatkan kontroversi menjadi polemik. Oleh karena itu, diperlukan suatu metode untuk mengklasifikasikan ulasan secara otomatis dengan melakukan analisis sentimen. Dalam penelitian ini, 2129 data komentar dalam beberapa konten yang membahas mengenai sistem tiket pertandingan Persib Bandung. Tujuan dari penelitian ini adalah untuk mengklasifikasikan analisis ulasan terhadap polemik sistem tiket pertandingan dengan memanfaatkan algoritma Naïve Bayes.

Kata kunci: Sentimen Analisis; youtube; Naïve Bayes; Persib Bandung

Abstract

Reporting about Persib cannot be separated from the role of the media from the era of the union until today. The first news about Persib in the media was at least in November 1904, when the Priangan Association (PVB) was recorded as the first association in Bandung. Using Descriptive Analysis, in the form of a word cloud, which is used in this research to identify and form word patterns that can be associated with other words that are considered important. Naïve Bayes Classifier Method, used in this research to identify and form word patterns that can be associated with other words to obtain information that is considered important. YouTube has become one of the largest platforms for sharing visual content on the internet. One of the topics that is being widely discussed is the ticket purchasing system for Persib Bandung matches. This has invited a lot of reactions, especially from the community, especially residents of West Java. This causes the controversy to become a polemic. Therefore, a method is needed to classify reviews automatically by conducting sentiment analysis. In this research, 2129 comment data in several contents discussed the Persib Bandung match ticket system. The aim of this research is to classify the analysis review of the polemic of the match ticket system using the Naïve Bayes algorithm.

Keywords: Sentiment Analysis; YouTube; Naïve Bayes; Persib Bandung

PENDAHULUAN

Saat ini, media sosial banyak digunakan tidak hanya sebagai sarana silaturahmi dan mencari teman, namun juga untuk promosi produk, jual beli, kampanye politik, dan aktivitas lainnya. Media sosial telah membantu masyarakat untuk tetap mendapatkan informasi terkini tentang peristiwa dan kejadian di wilayah mereka atau wilayah yang lebih luas. Berdasarkan hal tersebut, banyak pemerintah daerah yang memanfaatkan media sosial sebagai layanan e-Government (e-Government) untuk

menyampaikan informasi kepada Masyarakat. (Karim, 2020). Info tentang Persib tidak lain serta tidak bukan karena media berasal era perserikatan sampai sekarang. gosip Persib Bandung pertama muncul pada media cetak di tahun 1994 pada surat kabar Pemandangan. sampai ditahun 2000-an mendirikan pusat berita online yang diberi nama Simamaung pada tahun 2001 menjadi pusat semua berita buat Bobotoh. Sesuai dengan pengoptimalan dari perkembangan teknologi, tim Persib memberitahukan dan menginformasikan seluruh

kegiatannya melalui websitewww.persib.co.id. Hal tadi adalah cara yang tepat buat menjawab kebutuhan akan informasi serta menggambarkan keberadaan klub Persib Bandung kepada fans nya. Selain itu Persib mempunyai akun media sosial instagram resmi yaitu @persib. sesuai data terakhir yang ditemukan pada 18 September 2023, akun @persib telah berhasil menjaring followers sebesar lebih dari 1,86 juta pengguna. Berkaitan menggunakan sikap Masyarakat yang mengomentari tentang sistem tiket pertandingan persib bandung. (Jon, 2022).

TINJAUAN PUSTAKA

Dalam upaya untuk memahami persepsi masyarakat terhadap aplikasi Shopee, penelitian ini mengadopsi klasifikasi sentimen menggunakan algoritma Naive Bayes. Dapat dijabarkan sebagai berikut:

Penelitian pertama membahas mengenai analisis sentimen pada media sosial Instagram klub Persija Jakarta menggunakan metode Naive Bayes Classifier. Penelitian ini mencakup pengumpulan data dari komentar-komentar pada postingan akun resmi Persija Jakarta selama satu bulan, proses analisis yang meliputi tahap filtering input, pelabelan kelas sentimen, dan visualisasi data. Hasilnya menunjukkan adanya komentar positif dan negatif, serta kata-kata yang paling sering muncul dalam komentar. Selain itu, jurnal juga membahas penggunaan metode analisis sentimen lainnya seperti Support Vector Machine, Convolutional Neural Network, dan Maximum Entropy dalam konteks analisis sentimen pada media sosial (Jonathan and Nataliani, 2022).

Penelitian kedua membahas mengenai implementasi algoritma Naive Bayes dan metode stemming Sastrawi untuk analisis sentimen terkait kondisi "New Normal" di Indonesia. Penelitian bertujuan untuk membantu pemerintah memperoleh umpan balik terkait kebijakan terkait Covid-19. Proses melibatkan pengumpulan data, pra-pemrosesan data, ekstraksi fitur, dan klasifikasi menggunakan Naive Bayes. Hasilnya menunjukkan akurasi, presisi, dan nilai recall yang menjanjikan. Studi ini menggunakan data Twitter dan perpustakaan Sastrawi untuk stemming. Artikel juga memberikan tinjauan singkat tentang penelitian sebelumnya terkait analisis sentimen terkait Covid-19. (Lisangan, Gormantara and Carolus, 2022).

Penelitian ketiga membahas mengenai analisis sentimen terhadap layanan Transjakarta di Instagram menggunakan algoritma Naive Bayes dan seleksi fitur Information Gain. Penelitian ini bertujuan untuk mengklasifikasikan opini publik sebagai positif atau negatif. Hasilnya menunjukkan bahwa model Naive Bayes dengan seleksi fitur Information Gain memiliki akurasi, recall, presisi, dan spesifisitas yang lebih tinggi dibandingkan dengan model Naive Bayes saja. Metode penelitian meliputi pengumpulan data, preprocessing, seleksi fitur, dan klasifikasi. Studi ini menggunakan 700 data komentar dan merekomendasikan pengumpulan data lebih lanjut dari berbagai media sosial serta pembahasan tambahan mengenai layanan dan rute Transjakarta. (Ein, Ernawati and Widiastiwi, 2022).

Penelitian keempat membahas mengenai tentang analisis sentimen terhadap opini masyarakat mengenai vaksinasi Covid-19 di Twitter menggunakan metode Naive Bayes Classifier dan Support Vector Machine. Data berupa tweet diambil menggunakan crawling dari Twitter, lalu diberi labeling positif atau negatif. Data kemudian diproses menggunakan TF-IDF untuk ekstraksi fitur. Kedua metode diuji menggunakan 10-fold cross validation untuk mengukur akurasi, presisi, dan ingatannya. Hasilnya menunjukkan metode Naive Bayes Classifier lebih baik dengan rata-rata kinerja 57,10% dibandingkan Support Vector Machine 40,77%. Sehingga dapat disimpulkan bahwa untuk menganalisis sentimen Twitter, metode Naive Bayes Classifier lebih efisien dibandingkan Support Vector Machine (Alita and Shodiqin, 2023).

Penelitian kelima membahas tentang sentimen pelanggan toko online Tokopedia dengan menggunakan metode Naive Bayes Classifier. Tujuannya adalah untuk mengetahui hasil analisis sentimen opini publik di Tokopedia dan mengetahui hasil akurasi dari pengklasifikasian menggunakan metode Naive Bayes. Metode yang digunakan meliputi pengumpulan data ulasan pelanggan Tokopedia di Google Play Store, preprocessing data, pembagian data menjadi training dan testing, klasifikasi menggunakan Naive Bayes, dan pengujian akurasi. Hasilnya menunjukkan bahwa metode Naive Bayes mampu menganalisis secara otomatis dengan akurasi

95,1%, recall 91,46%, dan precision 100% ketika diuji pada 143 data pengujian. Kesimpulannya metode Naive Bayes terbukti efektif untuk analisis sentimen dan klasifikasi ulasan pelanggan Tokopedia. Namun disarankan menambah metode lain dan data lebih banyak untuk meningkatkan akurasi. (Salsabila, Murtopo and Fadhillah, 2022)

Penelitian keenam membahas mengenai untuk menganalisis sentimen ulasan produk atasan wanita di Shopee dengan menggunakan algoritma Naive Bayes dan mengklasifikasikannya kedalam 3 kelas (positif, negatif, netral). Data yang digunakan berupa 1000 ulasan produk yang dilabeli. Prosesnya meliputi pengumpulan data, pelabelan, pra-proses, klasifikasi sentimen, dan evaluasi. Pra-proses mencakup pembersihan emoticon dan simbol, folding kasus, normalisasi kata, tokenisasi, penghapusan stopword, dan stemming. Penimbangan kata menggunakan TF-IDF. Klasifikasi menggunakan algoritma Naive Bayes. Hasilnya menunjukkan akurasi sebesar 89%, yang menunjukkan ulasan produk atasan wanita cenderung positif. Analisis frekuensi kata menunjukkan kata positif seperti "sesuai", "baik", sedangkan negatif "tipis", "tidak sesuai". Kesimpulannya layanan dan produk atasan wanita di Shopee dinilai baik berdasarkan ulasan positif yang dominan (Subarkah *et al.*, 2023).

Penelitian ketujuh membahas mengenai analisis sentimen terhadap komentar video instansi pemerintah di Youtube guna mengetahui tingkat kepuasan masyarakat terhadap pelayanan pemerintah. Metode yang digunakan adalah algoritma Naive Bayes Classifier untuk mengklasifikasikan komentar menjadi kategori positif, negatif, dan netral. Data komentar diambil menggunakan YouTube API, kemudian dilakukan preprocessing seperti pembersihan data, penghapusan kata stopword, etc. Prosesnya meliputi crawling data, preprocessing, hitung term frequency, training model NBC, lalu testing untuk menguji akurasi model. Hasil pengujian untuk Kemen PUPR menghasilkan akurasi 69,23% dan Kemenkeu 64,10%. Penelitian ini bertujuan untuk memberikan tolak ukur tingkat kepuasan masyarakat terhadap pelayanan pemerintah berdasarkan analisis sentimen komentar di YouTube. (Yoga Saputra, Hanifudin Subhi and Zain Afif Winatama, 2019)

Penelitian kedelapan membahas mengenai menganalisis sentimen pengguna YouTube terhadap tayangan #MataNajwaMenantiTerawan

dengan mewawancarai kursi kosong yang mewakili Menteri Kesehatan Terawan. Data berupa komentar diambil dari YouTube melalui scrapping sebanyak 1400 data. Data kemudian diproses melalui preprocessing, vektorisasi, pembelian data, pemodelan, dan klasifikasi menggunakan metode Naive Bayes. Hasil klasifikasi Naive Bayes menghasilkan akurasi sebesar 90,36%. Komentar dikelompokkan menjadi sentimen positif, negatif, dan netral. Didominasi oleh sentimen netral sebanyak 1232 data. Metode Naive Bayes mampu melakukan klasifikasi komentar secara otomatis dan dapat digunakan untuk menganalisis sentimen komentar YouTube. Analisis sentimen dapat memberikan pemahaman tentang opini publik terhadap tayangan tersebut berdasarkan komentar yang direkam dalam jumlah besar. (Hudha, Supriyati and Listyorini, 2022)

Penelitian kesembilan ini menganalisis sentimen publik dalam komentar Youtube tersebut menjadi kategori positif dan negatif menggunakan klasifikasi Naive Bayes. Dilakukan juga pembobotan kata menggunakan TF-IDF. Dataset berupa 1000 komentar Youtube yang dilabel manual menjadi positif dan negatif. Dilakukan preprocessing teks, klasifikasi, dan validasi model 10-fold cross validation. Hasil terbaik 97% akurasi dicapai dengan 900 data latih, 100 data uji, pembobotan TF-IDF, dan 10-fold cross validation. Rata-rata akurasi 91,8% dengan presisi 90,35%, recall 93,6%, f-measure 91,95%. (Zhafira, Rahayudi and Indriati, 2021)

Penelitian kesepuluh membahas mengenai penerapan algoritma Naive Bayes Classifier (NBC) untuk melakukan klasifikasi sentimen masyarakat terhadap penyakit depresi yang didapatkan dari komentar di youtube Data komentar diperoleh dari beberapa akun media berita youtube seperti tvOneNews, Kompas TV, dan lainnya sebanyak 4783 data dengan 3826 data latihan dan 957 data uji. Dilakukan pra-pemrosesan data seperti case folding, cleaning, filtering, dan stemming. Kemudian dilakukan ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Dilakukan klasifikasi menggunakan NBC dengan pembagian data latih dan uji sebesar 80:20. Didapatkan akurasi sebesar 84,11% Hasil prediksi sentimen didominasi oleh sentimen positif sebesar 93,31%. Sedangkan sentimen negatif dan netral masing-masing 6,68% dan 0%. (Yelvita, 2022).

2.1 Perbedaan utama dalam penelitian Naive Bayes dari dulu hingga sekarang.

Perbedaan utama dalam penelitian Naive Bayes dari dulu hingga sekarang dapat mencakup:

1. Peningkatan Kinerja: Dulu, implementasi Naive Bayes mungkin lebih sederhana dengan asumsi independensi fitur yang lebih kuat, sementara sekarang ada upaya untuk meningkatkan kinerja dengan mempertimbangkan ketergantungan antarfitur.
2. Penyesuaian Model: Model Naive Bayes dulu mungkin lebih statis, sementara sekarang ada upaya untuk menyesuaikan model dengan data yang lebih dinamis dan rumit.
3. Manajemen Data yang Lebih Baik: Dengan kemajuan dalam manajemen data dan teknik pengolahan data, penelitian sekarang mungkin lebih fokus pada penggunaan Naive Bayes dengan data yang besar dan beragam.
4. Integrasi dengan Teknologi Terkini: Sekarang, ada penelitian yang lebih banyak tentang integrasi Naive Bayes dengan teknologi terkini seperti deep learning untuk meningkatkan kinerja dan ketepatan model.

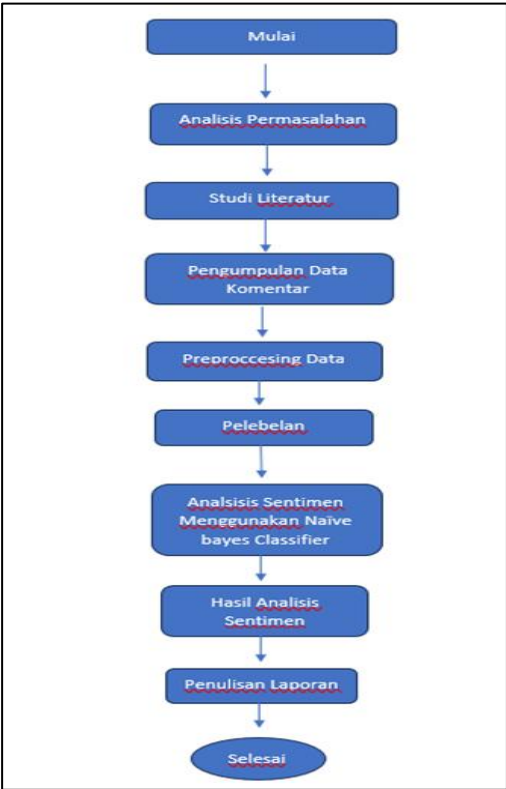
METODE PENELITIAN

Pada penelitian ini melakukan pemodelan topik analisis sentimen opini supporter pengguna youtube terhadap sistem pembelian tiket pertandingan persib menggunakan Naive Bayes agar dapat mengklasifikasi komentar positif atau negatif baik dari supporter ataupun Masyarakat penikmat sepak bola dan dapat digunakan untuk menilai kualitas pembelian tiket. Metode penelitian Naive Bayes classifier merupakan teknik klasifikasi yang berdasarkan teorema probabilitas Bayes dengan asumsi independensi antara fitur-fitur yang digunakan dalam klasifikasi.

3.1 Tahapan Metode Penelitian

Pemahaman data mulai dengan memahami data yang akan digunakan untuk klasifikasi. Data ini harus terstruktur dengan fitur-fitur yang dapat diambil untuk melakukan prediksi. Perhitungan Probabilitas: hitung probabilitas setiap kelas berdasarkan data training yang ada. Perhitungan Probabilitas fitur hitung probabilitas dari masing-masing fitur terhadap setiap kelas.

Perhitungan Probabilitas Posterior Gunakan teorema Bayes untuk menghitung probabilitas posterior dari setiap kelas berdasarkan fitur-fitur yang diberikan. Klasifikasi tentukan kelas dari data baru berdasarkan probabilitas posterior tertinggi. Metode Naive Bayes sering digunakan dalam klasifikasi teks seperti klasifikasi spam email, analisis sentimen, atau klasifikasi dokumen berdasarkan topik



Gambar 1. Tahapan Penelitian

Tabel 1. Aktifitas Tahapan Penelitian

Tahapan	Deskripsi Aktivitas
1. Persiapan	Melakukan tinjauan pustaka mendalam tentang Naive bayes classifier komentar sistem pembelian tiket pertandingan analisis data
2. Implementasi	Menyiapkan Software dan tools yang dibutuhkan, termasuk perangkat keras, perangkat lunak data mining, mengumpulkan data youtube dan sistem penyimpanan data yang aman.
3. Proses Data	Langkah-langkah yang dilakukan untuk membersihkan, merapikan, melakukan labeling agar dapat digunakan dengan lebih efektif dalam proses analisis.
4. Analisis	Melakukan analisis sentimen dari data yang ada menggunakan metode Naive classifier
5. Penyusunan	Membuat laporan yang merinci temuan, kesimpulan, rekomendasi, dan implikasi hasil penelitian.

3.2 Sumber Data

Dalam sentimen analisis yang dilakukan memerlukan data yang berasal dari aplikasi YouTube. Data yang dibutuhkan merupakan data teks berupa komentar yang berkaitan Sitem pembelian tiket pertandingan Persib. Data teks ini memiliki 2 nilai sentimen, yaitu positif dan negatif. Serta dibuat dengan tabel yang terdiri dari teks dan sentimen. Data yang dikumpulkan sebanyak 2131 data teks, setelah itu dilakukan sortir data dengan pelabelan otomatis dengan naïve bayes classifier.

3.3 Teknik Analisis Data

Teknik ini akan mempresentasikan temuan dan analisisnya dari tahap penelitian. Temuan ini merinci langkah-langkah mulai dari proses pengumpulan data hingga pemrosesan awal data, transformasi data, dan evaluasi data. Selain itu, analisis data terperinci yang didukung oleh berbagai jenis visualisasi seperti statistik, tabel, dan bagan data dapat sangat membantu dalam meningkatkan pemahaman pembaca. Temuan yang diperoleh juga menjadi dasar rekomendasi untuk penelitian selanjutnya.

Pada penelitian ini melakukan pemodelan topik analisis sentimen opini supporter pengguna youtube terhadap sistem pembelian tiket pertandingan persib menggunakan Naïve Bayes agar dapat mengklasifikasi komentar positif atau negatif baik dari supporter ataupun Masyarakat penikmat sepak bola dan dapat digunakan untuk menilai kualitas pembelian tiket. Metode penelitian Naive Bayes clasifier merupakan teknik klasifikasi yang berdasarkan teorema probabilitas Bayes dengan asumsi independensi antara fitur-fitur yang digunakan dalam klasifikasi.

True Negative (TN), dan False Negative (FN), untuk setiap kelas (Negatif, Netral, dan Positif).

Tabel 2. Kelas Prediksi

Kelas	Kelas Negatif Prediksi	Kelas Neutral Prediksi	Kelas Positif Prediksi
Kelas Negatif	TN	FNt	FP
Kelas Neutral	FN	TNt	FP
Kelas Positif	FN	FNt	TP

HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian performa algoritma Naive Bayes pada data latih dan uji ini akan dibahas mengenai performa algoritma

beserta faktor-faktor yang mempengaruhi akurasi algoritma pada analisis sentimen komentar video pengguna YouTube terhadap sistem pembelian tiket pertandingan pada data latih dan data uji dapat dilihat dari nilai akurasi, presisi, recall, dan F1-Score yang dihasilkan dari Confusion Matrix setiap model yang sebelumnya telah diproses menjadi representasi vektor menggunakan CountVectorizer dan dilakukan pembobotan kata menggunakan TF-IDF.

Sebelum digunakan untuk memprediksi data uji, dilakukan terlebih dahulu pelatihan model Naive Bayes pada data latih agar model dapat mempelajari pola yang terdapat pada data berupa kumpulan fitur teks dari komentar pengguna YouTube. Proses pelatihan ini penting agar model dapat belajar dan membaca pola yang membedakan sentimen positif dan negatif. Data latih digunakan untuk menyesuaikan bobot dan parameter sehingga akurasi sentimen oleh model Naive Bayes dapat lebih optimal. Data latih yang telah melalui beberapa tahapan Preprocessing text dan ekstrasi fitur menggunakan CountVectorizer dan TF-IDF, tahapan ini bertujuan untuk mengubah data tekstual menjadi data numerik, agar dapat diolah oleh model Naive Bayes. Hasil vektor TF-IDF mewakili bobot masing-masing kata pada setiap dokumen teks komentar. Hasil modeling data latih dapat dilihat pada Tabel 3.

Tabel 3. Hasil Modeling Data Latih

MODEL	AKURASI
MODEL PERTAMA	0,902

Hasil modeling data uji dapat dilihat pada Tabel 4.

Tabel 4. Hasil Modeling Data Uji

MODEL	AKURASI
MODEL PERTAMA	0,966

Berdasarkan Tabel 4, dapat dilihat bahwa akurasi model *Naive Bayes* pada data uji adalah 96,6% , angka akurasi ini lebih tinggi dibandingkan dengan akurasi model pada data latih yang mencapai 90,2% (Tabel 3). Peningkatan akurasi ini menunjukkan bahwa model *Naive Bayes* memiliki tantangan dalam menyimpulkan pola dari data latih ke data uji.

Selanjutnya akan dibahas lebih rinci mengenai performa model pada masing-masing kelas sentimen dengan melihat dengan melihat nilai presisi, *recall* dan *F1-Score* yang dihitung dari *Confusion Matrix*, dapat dihitung dengan nilai *TP*, *FP*, *TN*, dan *FN* pada masing masing kelas yang akan digunakan untuk menghitung metrik evaluasi.

Tabel 5. Metrik Evaluasi Kelas

Model Pertama	Kelas Negatif Prediksi	Kelas Neutral Prediksi	Kelas Positif Prediksi
Kelas Negatif	22	0	0
Kelas Neutral	0	92	0
Kelas Positif	0	0	64

Sebelum mendapatkan nilai Akurasi, Presisi, Recall dan F1-Score, dilakukan perhitungan terlebih dahulu disetiap kelasnya.

Perhitungan tersebut menggunakan rumus:

$$Akurasi = \frac{62 + 22 + 92}{22 + 0 + 0 + 92 + 0 + 62} \times 100\% = 1 = 100\%$$

a) Kelas Negatif

$$Precision = \frac{62}{64 + 0} = 1 = 100\%$$

$$Recall = \frac{64}{64 + 0} = 1 = 100\%$$

$$F1-Score = 2x \frac{1 \times 1}{1 + 1} = 1 = 100\%$$

b) Kelas Netral

$$Precision = \frac{64}{64 + 0} = 1 = 100\%$$

$$Recall = \frac{64}{64 + 0} = 1 = 100\%$$

$$F1-Score = 2x \frac{1 \times 1}{1 + 1} = 1 = 100\%$$

c) Kelas Positif

$$Precision = \frac{64}{64 + 0} = 1 = 100\%$$

$$Recall = \frac{64}{64 + 0} = 1 = 100\%$$

$$F1-Score = 2x \frac{1 \times 1}{01 + 1} = 1 = 100\%$$

Setelah melakukan perhitungan pada setiap kelasnya, selanjutnya bisa melakukan perhitungan rata-rata dari setiap kelasnya:

$$Precision = \frac{97\% + 95\% + 97\%}{3} = 96\%$$

$$Recall = \frac{100\% + 99\% + 100\%}{3} = 99\%$$

$$F1-Score = 2x \frac{98\% + 96\% + 98\%}{3} = 99\%$$

Dari semua total ada 176 data uji, pada kelas negatif dari 154 data sebanyak 22 data yang diprediksi dengan benar sebagai negatif (True Negatif), dan tidak ada data data yang salah diprediksi, dengan kelas lain (False Negative). Sementara itu dari 176 data, pada kelas Netral dari 92 data netral, yang diprediksi sebagai netral ada 92 data yang berhasil diprediksi sebagai netral (True Netral), tidak ada diantaranya merupakan data bukan netral yang salah diprediksi (False Negative), Sementara itu dari 176 data, pada kelas Positif ada 64 data netral yang berhasil di prediksi (True Positive). Oleh karena itu dapat disimpulkan bahwa akurasi 100% ini menunjukkan bahwa model memiliki tingkat keakuratan sebesar 100%, yang merupakan presentase dari data yang berhasil di prediksi, Presisi 100% ini menunjukkan rata-rata dari kelas adalah 100%, presisi mengukur sejauh mana prediksi yang di buat oleh model itu berhasil terprediksi. Recall 100%, ini menunjukkan bahwa sejauh mana model dapat memprediksi semua kelas yang seharusnya positif. F1-Score 100% ini menunjukkan rata-rata F1-Score dari semua kelas adalah 100%, F1-Score ini adalah metrik gabungan yang mengukur antara presisi dan recall.

Performa algoritma Naive Bayes dalam melakukan analisis sentimen komentar pengguna YouTube terhadap sistem pertandingan tiket Persib ini menunjukkan model mampu dengan baik dalam mengklasifikasikan sentimen pada komentar, dengan akurasi tertinggi 100% pada data uji. Meskipun secara umum Naive Bayes terbukti efektif untuk

analisis sentimen, performanya perlu ditingkatkan agar lebih baik dan lebih optimal. Oleh karena itu teknik yang lebih canggih masih diperlukan untuk mengatasi permasalahan generalisasi model Naive Bayes kedepannya, walaupun demikian, tantangan generalisasi tetap menjadi pekerjaan rumah yang harus diselesaikan pada riset mendatang supaya performa algoritma Naive Bayes bisa semakin ditingkatkan.

Data komentar berasal dari media sosial YouTube. Data komentar dikumpulkan menggunakan kunci API yang disediakan oleh YouTube, perpustakaan Snsrape. Selanjutnya data yang diambil dalam format file .json dan kita ubah menjadi file .csv. Data yang dikumpulkan YouTube merupakan data teks yang tidak diketahui memiliki nilai sentimen positif atau negatif. Kata kunci untuk mengumpulkan data komentar dari YouTube tahun 2019-2023 menggunakan kata kunci Sistem Pembelian Tiket Pertandingan Persib.

SIMPULAN

Algoritma Naive Bayes mampu melakukan klasifikasi sentimen komentar pengguna YouTube terhadap sistem pembelian tiket pertandingan Persib ini dengan akurasi tertinggi 100% pada data uji, akurasi ini lebih tinggi dibandingkan dengan akurasi pada data latih yang mencapai akurasi 100%. Hal ini menunjukkan bahwa model mampu dengan baik dalam mengklasifikasikan sentimen pada komentar. Nilai presisi, recall, dan F1-Score yang tidak seragam pada masing-masing kelas sentimen menunjukkan ketidakseimbangan performa model Naive Bayes. Masalah pada kelas tertentu perlu diperbaiki agar akurasi dapat ditingkatkan. Sentimen positif yang mendominasi komentar pengguna Youtube terhadap sistem pembelian tiket pertandingan ini dengan proporsi 50,39 %, yang menunjukkan bahwa komentar pengguna ini netral terhadap adanya kebijakan baru, dan cukup menerima sistem tiket online Persib Sementara itu pada sentimen negatif sebesar 15,06% atau sebanyak 589 data dari total 2.129 data komentar, dari sentimen negatif ini menunjukkan bahwa tidak sedikit juga yang tidak setuju atas kebijakan baru dan tidak banyak juga yang mengeluh atas pelayanan kurang baik dari sistem tiket Persib. Sedangkan dalam sentimen netral sedikit lebih tinggi dari sentimen negatif, sebesar 34,55% atau sebanyak 694 data dari total 2.129 data

komentar, dari sentimen netral ini menunjukkan bahwa kebijakan dan sistem tiket pertandingan Persib telah diterima tanpa masalah dan bisa juga menunjukkan kurangnya ketertarikan atau keterlibatan masyarakat terhadap sistem pembelian tiket pertandingan Persib. Penelitian ini telah berhasil menerapkan algoritma Naive Bayes untuk klasifikasi sentimen pada kasus analisis sentimen komentar pengguna YouTube terhadap sistem pembelian tiket pertandingan Persib.

Untuk penelitian selanjutnya yaitu dengan menyelidiki dan membandingkan performa algoritma klasifikasi sentimen lainnya untuk memastikan pemilihan metode yang paling efektif. Penggunaan ensemble learning atau pendekatan gabungan dari beberapa algoritma mungkin dapat meningkatkan akurasi klasifikasi sentiment.

DAFTAR PUSTAKA

- Alita, D. and Shodiqin, R.A. (2023) 'Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine', *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), pp. 1–12. Available at: <https://doi.org/10.58602/jaiti.v1i1.20>.
- Ein, I., Ernawati, I. and Widiastiwati, Y. (2022) 'Analisis Sentimen Terhadap Layanan Transjakarta Pada Media Sosial Instagram Menggunakan Naive Bayes dan Seleksi Fitur Information Gain', *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 442–451.
- Hudha, M., Supriyati, E. and Listyorini, T. (2022) 'Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naive Bayes Classifier', *JIKO (Jurnal Informatika dan Komputer)*, 5(1), pp. 1–6. Available at: <https://doi.org/10.33387/jiko.v5i1.3376>.
- Jonathan, M. and Nataliani, Y. (2022) 'Analisis Sentimen Penilaian Masyarakat Indonesia terhadap GeNose pada Komentar Youtube Menggunakan Metode Naive Bayes', *Jurnal Matematika dan Aplikasi*, 11(01). Available at: <https://ejournal.unsrat.ac.id/index.php/decartesian/article/view/38339>.

- Lisangan, E.A., Gormantara, A. and Carolus, R.Y. (2022) 'Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia', *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 2(1), pp. 23–32. Available at: <https://doi.org/10.24002/konstelasi.v2i1.5609>.
- Salsabila, S.M., Murtopo, A.A. and Fadhilah, N. (2022) 'Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier', 11(September), pp. 30–35.
- Subarkah, P. *et al.* (2023) 'Sentiment Analysis on Reviews of Women ' S Tops on Shopee', 9(1), pp. 126–133. Available at: <https://doi.org/10.33480/jitk.v9i1.4179>.
- INTRODUCTION.
- Yelvita, F.S. (2022) 'הכי קשה לראות את הציורים', *מה שבאמת לנגד העיניים*, 8(5.2017), pp. 2003–2005.
- Yoga Saputra, P., Hanifudin Subhi, D. and Zain Afif Winatama, Z. (2019) 'Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naive Bayes', *Jurnal Informatika Polinema*, 5(3), pp. 209–213. Available at: <http://jip.polinema.ac.id/ojs3/index.php/jip/article/view/259>.
- Zhafira, D.F., Rahayudi, B. and Indriati, I. (2021) 'Zhafira, Dhaifa Farah Rahayudi, Bayu Indriati, Indriati', *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1), pp. 55–63. Available at: <https://doi.org/10.25126/justsi.v2i1.24>.