

Analisis Sentimen Pengguna X Terhadap *Live Streaming* Marapthon AAA Clan Menggunakan *Naïve Bayes*

Sentiment Analysis of X Users on AAA Clan Marapthon *Live Streaming* Using *Naïve Bayes*

Hezron Chrsiti Yohanis¹⁾, Christhian Sanda Tanan²⁾, Otto Pasorong³⁾, Beltsazar Kevin⁴⁾

^{1,2,3,4} Teknik Informatika, Fakultas Teknik, Universitas Papua

Jl. Gunung Salju Amban, Manokwari, Papua Barat, Indonesia

E-mail : hezron.c.yohanis@gmail.com¹⁾, christiansandatanan@gmail.com²⁾,
ottopasorong10@gmail.com³⁾, kevinbeltzasar@gmail.com⁴⁾

Abstract

Marapthon is a live streaming phenomenon with a Subathon format that runs for several days, where the duration of the stream is determined by the amount of audience donations. This phenomenon has been growing and attracts high levels of interaction between streamers and viewers, leading to various X user opinions, both positive and negative. This study aims to analyze user sentiment toward Marapthon live streaming on the social media platform X using the Naïve Bayes method. The data were collected through a Web Scraping technique, resulting in 1,695 tweets. The collected data then underwent several stages, including preprocessing, data labeling, data splitting, feature weighting using TF-IDF, and classification using the Naïve Bayes algorithm. The results show that there are 673 positive sentiment data and 1,016 negative sentiment data. The model evaluation achieved an accuracy of 71.89%, indicating that the model has fairly good performance in classifying X user sentiment. Overall, the findings indicate that user sentiment toward the third season of Marapthon live streaming tends to be negative, although a considerable number of users still express positive opinions.

Keywords: Sentiment Analysis, Naïve Bayes, Lexicon-Based, Live Streaming, Marapthon

Abstrak

Marapthon merupakan fenomena *live streaming* dengan format *Subathon* yang berlangsung selama beberapa hari, di mana durasi *live streaming* ditentukan oleh jumlah donasi dari penonton. Fenomena ini semakin berkembang dan menarik banyak interaksi antara *streamer* dan penonton, sehingga memunculkan berbagai opini dari pengguna X, baik bersifat positif maupun negatif. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap *live streaming* Marapthon pada media sosial X dengan menggunakan metode *Naïve Bayes*. Data dikumpulkan melalui teknik *Web Scraping* sebanyak 1.695 tweet. Data yang diperoleh kemudian melalui beberapa tahapan, yaitu *preprocessing*, *data labeling*, *split data*, pembobotan fitur menggunakan *TF-IDF*, dan proses klasifikasi menggunakan algoritma *Naïve Bayes*. Hasil penelitian menunjukkan bahwa terdapat 673 data sentimen positif dan 1.016 data sentimen negatif. Evaluasi model menghasilkan nilai *Accuracy* sebesar 71,89%, yang menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan sentimen pengguna X. Hasil ini menunjukkan bahwa penelitian analisis sentimen pengguna X cenderung negatif terhadap *live streaming* Marapthon musim ketiga, meskipun masih terdapat cukup banyak pengguna yang memberikan opini positif.

Kata kunci: Analisis Sentimen, Naïve Bayes, Lexicon-Based, Live Streaming, Marapthon

1. Pendahuluan

Era digital saat ini telah membawa perubahan dalam cara masyarakat memperoleh hiburan. Dengan berkembangnya media sosial, masyarakat memiliki berbagai pilihan dalam menikmati konten hiburan. Salah satu bentuk hiburan yang mengalami perkembangan dalam beberapa tahun terakhir adalah *live streaming* [1], seperti fenomena Marapthon yang diselenggarakan oleh AAA Clan di kanal YouTube @ybrap. Marapthon merupakan fenomena *Live Streaming* dengan format *Subathon* yang berlangsung selama sehari-hari dengan durasi acara

yang ditentukan berdasarkan jumlah donasi yang diberikan oleh penonton. Pada tanggal 24 November 2024 Marapthon tayang untuk pertama kali dan kini telah menayangkan musim ketiga pada tanggal 08 Februari 2026 dengan konten yang beragam. Marapthon menyajikan konten aktivitas keseharian, mengobrol, mengundang tamu spesial, hingga melakukan tantangan.

Media sosial kini telah menjadi platform utama bagi masyarakat untuk mengekspresikan opini, salah satu platform media sosial yang banyak digunakan adalah X, yang memungkinkan pengguna untuk berinteraksi dan berbagi opini mengenai berbagai

fenomena, termasuk *live streaming*. *Live streaming*, khususnya fenomena Marapthon, semakin populer dan menghasilkan interaksi yang tinggi antara *streamer* dan penonton. Hal ini memunculkan berbagai opini dari pengguna X, baik positif maupun negatif. Namun, belum diketahui secara jelas bagaimana kecenderungan sentimen pengguna terhadap fenomena tersebut, sehingga diperlukan suatu analisis untuk memahami opini pengguna X. Analisis sentimen merupakan salah satu pendekatan yang dapat digunakan untuk memahami opini publik terhadap suatu fenomena. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen guna memahami opini publik terhadap fenomena Marapthon pada media sosial X dengan melakukan klasifikasi menggunakan metode *Naïve Bayes*, yang dikenal memiliki kinerja yang cepat dan efisien dalam klasifikasi teks.

2. Tinjauan Pustaka

2.1 Marapthon

Marapthon adalah acara *live streaming* yang diselenggarakan oleh Reza Arap bersama AAA Clan di kanal YouTube @ybrap, istilah Marapthon merupakan kependekan dari gabungan nama "Arap" dan "marathon" yang merujuk pada durasi *live streaming* yang dilakukan secara nonstop selama sehari-hari tanpa berhenti, dengan konten yang beragam seperti menunjukkan keseharian mereka, mengobrol, mengundang tamu spesial hingga melakukan berbagai tantangan. Marapthon pertama kali tayang pada tanggal 24 November 2024 dan kini telah menayangkan musim ketiga pada tanggal 08 Februari 2026. Marapthon diselenggarakan dengan format *Subathon* yang awalnya berasal dari platform twitch, *Subathon* merupakan kependekan dari "subscription" dan "marathon". Pada dasarnya, *Subathon* merupakan tantangan yang dilakukan *streamer* untuk melangsungkan *live streaming* berdasarkan dukungan penonton dengan memberikan *subscription* yang diraih secara *real-time*, durasi *live streaming* bertambah berdasarkan seberapa banyak para penonton memberi *subscription*. Berbeda dengan platform twitch yang memiliki fitur *subscription* berbayar, Marapthon tayang di kanal YouTube maka durasi bertambah melalui seberapa besar nominal donasi yang diberikan oleh penonton.

2.2 Analisis Sentimen

Analisis sentimen merupakan bidang studi yang digunakan untuk menganalisis opini, sikap atau emosi masyarakat terhadap suatu topik atau isu [2]. Analisis sentimen diterapkan dengan mengklasifikasikan teks ke dalam kelas yang ditentukan. Analisis sentimen dapat digunakan dalam memahami opini, sikap atau emosi orang berdasarkan hasil analisis dari data berupa teks [3].

2.3 Naïve Bayes

Naïve Bayes adalah metode klasifikasi sederhana dan efisien yang menggunakan probabilitas untuk memprediksi kelas berdasarkan data yang ada. *Naïve Bayes* melakukan pengklasifikasi pembelajaran mesin probabilistik berdasarkan *teorema Bayes* dengan asumsi "naive" yaitu independensi antar fitur. Metode *Naïve Bayes* dalam analisis sentimen memiliki beberapa tahapan. Tahap awal dimulai dari representasi data, yaitu mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma. Proses ini biasanya dilakukan menggunakan pendekatan pembobotan fitur seperti *TF-IDF*. Selanjutnya masuk ke tahap pelatihan (*training*), dimana model mempelajari pola dari data latih yang telah disiapkan. Pada tahap ini dilakukan perhitungan probabilitas setiap kelas serta probabilitas kemunculan kata terhadap masing-masing kelas. Hasil dari proses ini akan membentuk model yang kemudian digunakan untuk melakukan klasifikasi. Setelah model terbentuk, dilakukan tahap klasifikasi atau pengujian (*testing*), yaitu proses untuk menentukan sentimen dari data baru yang belum pernah dilihat sebelumnya. Pada tahap ini, model akan membandingkan probabilitas dari setiap kelas, kemudian menentukan hasil akhir berdasarkan nilai probabilitas tertinggi. Dengan tahapan tersebut, metode *Naïve Bayes* dapat digunakan untuk melakukan klasifikasi teks secara sederhana dan cukup efektif [4].

2.4 Penelitian Terdahulu

Pada penelitian sebelumnya dalam analisis sentimen menggunakan metode *Naïve Bayes*, Junaidy et al. [5] menggunakan metode *Naïve Bayes* untuk menganalisis sentimen publik terhadap Rancangan Undang-Undang Kitab Undang-Undang Hukum Acara Pidana (RUU KUHP). Data yang dikumpulkan berjumlah 2.273 tweet, yang kemudian dilakukan *preprocessing* yang meliputi *Case Folding*, *Cleansing*, *Tokenizing*, *Stopword Removal*, dan *Stemming*. Hasil analisis menunjukkan bahwa sentimen publik didominasi oleh sentimen negatif sebesar 45,5%, diikuti oleh sentimen netral sebesar 32,0% dan sentimen positif sebesar 22,5%. Hasil evaluasi model menunjukkan nilai *Accuracy* sebesar 74,07%, dengan nilai *Precision*, *Recall*, dan *F1-Score* yang bervariasi pada setiap kelas, yaitu pada kelas negatif diperoleh *Precision* sebesar 0,71, *Recall* 0,93, dan *F1-Score* 0,80, pada kelas netral *Precision* 0,74, *Recall* 0,44, dan *F1-Score* 0,55, serta pada kelas positif *Precision* 0,85, *Recall* 0,80, dan *F1-Score* 0,82, yang menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik meskipun masih terdapat keterbatasan dalam mengidentifikasi sentimen netral.

Penelitian lain, Sulaeman et al. [6] menggunakan metode *Naïve Bayes* untuk menganalisis opini publik terhadap kebijakan TV digital di Indonesia pada media sosial X. Data yang digunakan berjumlah 1.855 tweet, yang kemudian diproses dan menghasilkan 60,58% sentimen positif (478 tweet) dan 39,42% sentimen negatif (311 tweet), yang menunjukkan sentimen positif lebih besar dari sentimen negatif. Hasil evaluasi memiliki nilai *Accuracy* sebesar 79,21%, yang menunjukkan model memiliki performa klasifikasi yang baik dan stabil. Penelitian lain, Akmalya et al. [7] menganalisis sentimen pelanggan terhadap layanan Kereta Cepat Whoosh menggunakan metode *Naïve Bayes* dengan jumlah data sebanyak 730 tweet yang dikumpulkan melalui teknik *Web Scraping*. Data diproses melalui tahapan *preprocessing* dan diklasifikasikan ke dalam dua kelas, yaitu positif dan negatif. Hasil penelitian menunjukkan bahwa sentimen negatif lebih dominan sebesar 69,86%, sedangkan sentimen positif sebesar 30,14%. Evaluasi model menunjukkan nilai *Accuracy* sebesar 95,07%, yang menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi dalam melakukan klasifikasi sentimen. Berdasarkan penelitian-penelitian tersebut, metode *Naïve Bayes* terbukti cukup efektif dalam analisis sentimen.

3. Metodologi Penelitian

Dalam penelitian ini diterapkan *Supervised Learning*, dengan begitu model akan dilatih dengan data yang sudah diberi label, sehingga model mampu memprediksi data baru yang belum pernah dilihat sebelumnya.



Gambar 1. Tahapan Penelitian

3.1 Pengumpulan data

Data sentimen yang digunakan untuk penelitian ini berasal dari sentimen terhadap *live streaming* Maraphthon musim ketiga oleh AAA Clan. Data ini diambil melalui teknik *Web Scraping* dari media sosial X dengan bantuan *Extensions* browser. *Dataset* ini kemudian digunakan untuk menganalisis persepsi pengguna X terhadap *live streaming* Maraphthon.

3.2 Data Preprocessing

Preprocessing merupakan tahap yang sangat penting dalam analisis sentimen. Tahap ini bertujuan untuk mempersiapkan *Dataset* sebelum dianalisis lebih lanjut [8]. *Preprocessing* memiliki beberapa tahapan dalam membuat *Dataset* lebih terstruktur dan siap diolah. *Cleaning*, Pada tahap ini, data duplikat

dihapus dan teks dibersihkan dari simbol, angka, tanda baca, dan karakter yang tidak diperlukan. Tujuan *Cleaning* adalah untuk menghilangkan noise dan memastikan hanya data relevan yang dianalisis [9]. *Case Folding*, Pada tahap ini, seluruh huruf dalam data diubah menjadi huruf kecil (*lowercase*) agar variasi kata dengan huruf besar dan kecil tidak berpengaruh terhadap hasil analisis [10]. Normalisasi kata, Pada tahap ini, data yang memiliki kata tidak baku atau slang menjadi kata baku agar teks terlihat lebih konsisten [11]. *Tokenizing*, Pada tahap ini, kalimat diubah menjadi satuan unit yang lebih kecil yang disebut "token" agar teks lebih terstruktur. *Stopword Removal*, Pada tahap ini, kata umum yang sering muncul dihilangkan karena tidak berpengaruh terhadap data sentimen. *Stemming*, Pada tahap ini, kata-kata disederhanakan ke dalam bentuk dasarnya. Dengan *Stemming* variasi kata yang memiliki makna dasar yang sama dapat dikurangi [9].

3.3 Data Labeling

Data Labeling dilakukan terhadap *Dataset* yang telah melalui tahap *preprocessing*. Data akan diberi label dengan kelas yang telah ditentukan, yaitu positif dan negatif. Proses *labeling* bertujuan untuk mengidentifikasi tweet yang mengandung sentimen positif atau negatif. *Labeling* dilakukan menggunakan metode *Lexicon-Based* berdasarkan kamus *Lexicon*. Metode *Lexicon-Based* bekerja dengan memberikan skor sentimen pada setiap kata, kemudian menentukan label positif atau negatif berdasarkan total skor. Metode *Lexicon-Based* dipilih karena mampu melakukan pelabelan data secara otomatis dengan cukup akurat, sehingga lebih efisien dalam menangani *Dataset* berukuran besar. Selain itu, metode ini juga banyak digunakan dalam penelitian analisis sentimen.

3.4 Split Data

Dataset yang telah melalui proses *preprocessing* kemudian dibagi menjadi dua bagian, yaitu: data latihan dan data uji. Jumlah data latihan lebih besar dibandingkan data uji [12]. Dalam penelitian ini, digunakan rasio pembagian 80:20. Sebanyak 80% data digunakan untuk proses pelatihan model, sedangkan 20% sisanya digunakan untuk pengujian. Pembagian ini mempunyai kelebihan karena porsi data pelatihan yang lebih besar dapat membantu model belajar dengan lebih baik, sementara data pengujian tetap cukup untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya [13].

3.5 Pembobotan Fitur TF-IDF

Pembobotan *TF-IDF* (Term Frequency–Inverse Document Frequency) merupakan metode untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen. Penilaian berdasarkan pada

seberapa sering kata tersebut muncul dalam dokumen (TF), serta disesuaikan dengan tingkat kemunculannya di seluruh kumpulan dokumen (IDF), sehingga kata yang terlalu umum akan memiliki bobot yang lebih kecil [14]. Hasil dari proses ini berupa bobot kata yang akan digunakan sebagai fitur dalam proses klasifikasi menggunakan algoritma *Naïve Bayes*. Selain itu, *TF-IDF* juga dikenal efisien, sederhana, dan akurat [15].

3.6 Klasifikasi Naïve Bayes

Dataset yang telah melalui tahap *preprocessing* dan pembobotan kemudian dibagi menjadi dua bagian, data latih dan data uji. Proses klasifikasi dilakukan menggunakan algoritma *Naïve Bayes*. Model dilatih menggunakan data latih untuk mempelajari pola dari setiap kelas sentimen. Pelatihan diperlukan sebagai dasar untuk memprediksi data yang belum dilihat sebelumnya berdasarkan pola sebelumnya. Setelah proses pelatihan selesai, model yang telah dilatih kemudian digunakan untuk mengklasifikasikan data uji. Hasil klasifikasi tersebut kemudian digunakan untuk evaluasi model.

3.7 Evaluasi Model

Model yang telah dibangun kemudian dievaluasi menggunakan data uji. Evaluasi dilakukan untuk mengetahui performa model dalam melakukan klasifikasi sentimen. Evaluasi model juga didasarkan pada *Confusion Matrix* yang menggambarkan perbandingan antara hasil prediksi dan data aktual, yang kemudian digunakan untuk menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. *Accuracy* digunakan untuk mengukur tingkat prediksi yang benar dari model dalam mengklasifikasikan data secara keseluruhan. *Precision* untuk mengukur seberapa banyak prediksi sentimen positif yang benar-benar positif. *Recall* untuk mengukur seberapa banyak data positif yang berhasil diklasifikasikan dengan benar, dan *F1-Score* yang merupakan kombinasi antara *Precision* dan *Recall* yang memberikan gambaran lebih seimbang terhadap performa model [16].

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data dikumpulkan di media sosial X dengan teknik *Web Scraping* dengan bantuan *Extensions browser* berjumlah 1.695. Data berupa teks opini atau emosi dari pengguna X terhadap live streaming Maraphthon. Data dikumpulkan dalam periode dua bulan musim ketiga dari *live streaming* Maraphthon yang sedang berlangsung, mulai dari tanggal 08 Februari sampai 08 April. Data kemudian melalui tahap *preprocessing* dan diperoleh data berjumlah 1.689.

4.2 Preprocessing

Data yang telah dikumpulkan kemudian dilakukan tahap *preprocessing* untuk meningkatkan kualitas

data agar analisis lebih baik. *Cleaning*, data yang duplikat dihapus dan teks di bersihkan dari link, emoji, dan karakter yang tidak diperlukan. *Case Folding*, teks dengan huruf yang bervariasi di ubah menjadi huruf kecil. Normalisasi kata, kata tidak baku atau slang diubah menjadi kata baku. *Tokenizing*, kalimat dipisahkan menjadi kumpulan kata-kata. *Stopword Removal*, kata umum yang sering muncul pada teks dihilangkan. *Tokenizing* dan *Stopword Removal* dilakukan dengan bantuan *Library NLTK*. *Stemming*, kata dengan awalan atau akhiran diubah menjadi bentuk dasar. *Stemming* dilakukan dengan bantuan *Library Sastrawi*. Hasil *preprocessing* dapat dilihat pada Tabel 1. Setelah melalui tahap *preprocessing* *Dataset* sudah siap disajikan untuk dianalisis lebih lanjut.

4.3 Data Labeling

Data yang telah melalui tahap *preprocessing* selanjutnya diberi label dengan 2 kelas yaitu positif dan negatif. *Data Labeling* dilakukan dengan *Library Lexicon-Based*. *Labeling* dilakukan dengan menghitung total skor dari setiap kata pada teks berdasarkan kamus *Lexicon-Based* untuk dikategori sebagai sentimen positif atau negatif. Berdasarkan proses pelabelan menggunakan metode *Lexicon-Based*, diperoleh distribusi data dengan jumlah 673 (39,9%) label positif dan 1.016 (60,1%) label negatif.

Tabel 2. *Data Labeling*

No	Hasil <i>Preprocessing</i>	Label
1	jujur seru ustad jo Maraphthon	Positif
2	Maraphthon tahun memang bener bener proper	Positif
3	rame belah brand moell campaign Maraphthon konteks kata app belah brand tentang produk bayi cara iklan masa tempat orang suka mabok rokok	Negatif
....		
1689	semua personil Maraphthon tidak lucu kalau ketawa aneh cuman ibot	Negatif

Tabel 1. Data Preprocessing

No	Tweet	Cleaning	Case folding	Normalisasi	Tokenizing	Stopword	Stemming
1	jujur seru ustad jo di Marapthon	jujur seru ustad jo di Marapthon	jujur seru ustad jo di marapthon	jujur seru ustad jo di marapthon	['jujur', 'seru', 'ustad', 'jo', 'di', 'marapthon']	['jujur', 'seru', 'ustad', 'jo', 'marapthon']	jujur seru ustad jo marapthon
2	Marapthon tahun ini memang bener bener proper	Marapthon tahun ini memang bener bener proper	marapthon tahun ini memang bener bener proper	marapthon tahun ini memang bener bener proper	['marapthon', 'tahun', 'ini', 'memang', 'bener', 'bener', 'proper']	['Marapthon , 'tahun', 'memang', 'bener', 'bener', 'proper']	marapthon tahun memang bener bener proper
3	Rame di sebelah brand moell campaign di Marapthon, konteks katanya di app sebelah: brand ini tentang produk bayi tapi cara iklannya masa di tempat orang yang suka mabok dan rokok	Rame di sebelah brand moell campaign di Marapthon konteks katanya di app sebelah brand ini tentang produk bayi tapi cara iklannya masa di tempat orang yang suka mabok dan rokok	rame di sebelah brand moell campaign di marapthon konteks katanya di app sebelah brand ini tentang produk bayi tapi cara iklannya masa di tempat orang yang suka mabok dan rokok	rame di sebelah brand moell campaign di marapthon konteks katanya di app sebelah brand ini tentang produk bayi tapi cara iklannya masa di tempat orang yang suka mabok dan rokok	['rame', 'di', 'sebelah', 'brand', 'moell', 'campaign', 'di', 'marapthon', 'konteks', 'katanya', 'di', 'app', 'sebelah', 'brand', 'ini', 'tentang', 'produk', 'bayi', 'tapi', 'cara', 'iklannya', 'masa', 'di', 'tempat', 'orang', 'yang', 'suka', 'mabok', 'dan', 'rokok']	['rame', 'sebelah', 'brand', 'moell', 'campaign', 'marapthon', 'konteks', 'katanya', 'app', 'sebelah', 'brand', 'tentang', 'produk', 'bayi', 'cara', 'iklannya', 'masa', 'tempat', 'orang', 'suka', 'mabok', 'rokok']	rame belah brand moell campaign konteks kata app belah brand tentang produk bayi cara iklan masa tempat orang suka mabok rokok
1689	Dari semua personil Marapthon yang gak lucu dan kalau ketawa aneh cuman IBOT	Dari semua personil Marapthon yang gak lucu dan kalau ketawa aneh cuman IBOT	dari semua personil Marapthon yang gak lucu dan kalau ketawa aneh cuman ibot	dari semua personil Marapthon yang tidak lucu dan kalau ketawa aneh cuman ibot	['dari', 'semua', 'personil', 'Marapthon', 'yang', 'tidak', 'lucu', 'dan', 'kalau', 'ketawa', 'aneh', 'cuman', 'ibot']	['semua', 'personil', 'Marapthon', 'tidak', 'lucu', 'kalau', 'ketawa', 'aneh', 'cuman', 'ibot']	semua personil Marapthon tidak lucu kalau ketawa aneh cuman ibot

4.4 Split Data

Dataset dibagi 80:20 dengan total data latihan 1.351 dan data uji 338. Data dibagi dengan bantuan *Library scikit-learn*. Data latihan digunakan dalam melatih model sedangkan data uji digunakan dalam klasifikasi dan mengevaluasi model.

4.5 Pembobotan Fitur *TF-IDF*

Data berupa teks diubah menjadi bentuk numerik menggunakan metode *TF-IDF*, sehingga dapat digunakan dalam proses klasifikasi. Semakin sering kata yang muncul di tweet dan seberapa jarang kata tersebut muncul di seluruh *Dataset* akan memiliki bobot yang besar. Pembobotan dilakukan dengan *Library scikit-learn*.

Tabel 3. Pembobotan *TF-IDF*

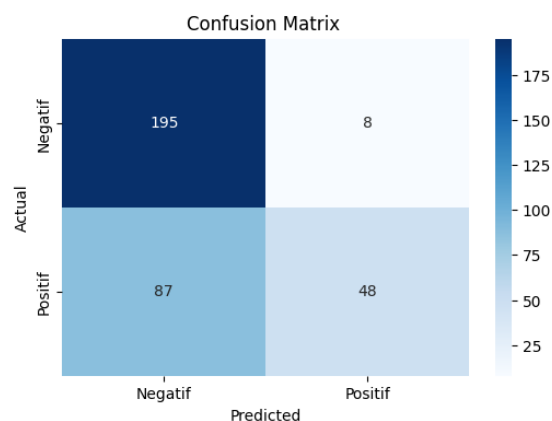
Term	TF-IDF
(0, 2507)	0.089488
(0, 130)	0.458396
(0, 4119)	0.470045
(0, 2399)	0.403556
(0, 3673)	0.249867
(0, 194)	0.579336
.....	

4.6 Klasifikasi *Naïve Bayes*

Data berupa teks yang telah diubah ke numerik dan memiliki bobot kata kemudian diproses oleh algoritma *Naïve Bayes* dengan model *Multinomial Naïve Bayes*, model ini cocok digunakan untuk data yang berbentuk frekuensi atau jumlah kemunculan kata dalam teks. Proses pelatihan dilakukan menggunakan data latihan, model yang telah dilatih kemudian diklasifikasikan oleh algoritma *Naïve Bayes* dengan data yang belum dilihat sebelumnya. Hasil dari klasifikasi menggunakan data uji kemudian digunakan untuk mengevaluasi model.

4.7 Evaluasi Model

Hasil dari model yang melalui proses klasifikasi dengan data uji kemudian dievaluasi dengan *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.



Gambar 2. Hasil *Confusion Matrix*

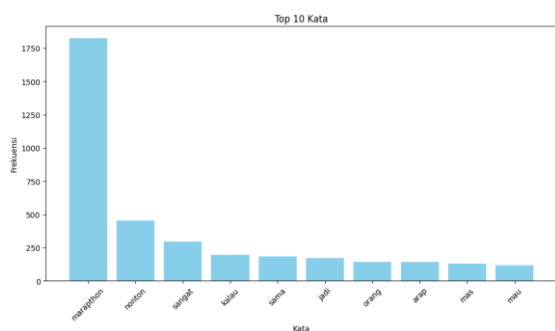
Confusion Matrix merupakan tabel evaluasi kinerja model klasifikasi untuk analisis sentimen positif dan negatif yang membandingkan hasil prediksi dengan nilai aktual.

Tabel 4. Evaluasi klasifikasi *Naïve Bayes*

<i>Accuracy</i>	71,89%		
<i>Classification Report</i>			
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Positif	0,86	0,36	0,50
Negatif	0,69	0,96	0,80

Berdasarkan hasil evaluasi model *Naïve Bayes*, diperoleh nilai *Accuracy* sebesar 71,89% , yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan cukup baik. Nilai *Precision* untuk kelas negatif sebesar 0,69 dan positif sebesar 0,86 menunjukkan bahwa model memiliki tingkat ketepatan yang lebih tinggi dalam memprediksi kelas positif dibandingkan negatif. Sementara itu, nilai *Recall* untuk kelas negatif sebesar 0,96 dan positif sebesar 0,36 mengindikasikan bahwa model sangat baik dalam mengenali data sentimen negatif, namun masih kurang optimal dalam mendeteksi data positif. Nilai *F1-Score* yang diperoleh, yaitu 0,80 untuk kelas negatif dan 0,50 untuk kelas positif, menunjukkan bahwa performa model lebih seimbang dan optimal pada kelas negatif dibandingkan kelas positif. Secara keseluruhan, hasil ini menunjukkan bahwa model *Naïve Bayes* memiliki performa yang cukup baik.

4.8 Frekuensi Kata



Gambar 3. Frekuensi Kata

Analisis frekuensi kata dilakukan sebagai analisis tambahan untuk memberikan gambaran lebih lanjut mengenai kata-kata yang sering muncul dalam *Dataset*. Analisis ini bertujuan untuk memberikan gambaran awal mengenai topik yang dibahas oleh pengguna serta pola interaksi yang terjadi pada fenomena yang diteliti. Berdasarkan Gambar 3, dapat dilihat bahwa kata “Marapthon” merupakan kata yang paling sering muncul, diikuti oleh kata “nonton” dan “sangat”. Hal ini menunjukkan bahwa sebagian besar pengguna membahas aktivitas menonton *live streaming* Marapthon. Selain itu, terdapat beberapa kata lain yang cukup sering muncul seperti “kalau”, “sama”, “jadi”, “orang”, “arap”, “mas”, dan “mau”. Kata-kata tersebut menunjukkan bentuk interaksi pengguna dalam menyampaikan pendapat, baik berupa tanggapan maupun komentar terhadap *live streaming* yang berlangsung. Secara keseluruhan, hasil frekuensi kata ini memberikan gambaran umum mengenai pembahasan yang terjadi di dalam *Dataset*.

5. Kesimpulan

5.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen pengguna X terhadap *live streaming* Marapthon menggunakan metode *Naïve Bayes*, maka dapat disimpulkan bahwa sentimen pengguna X terhadap *live streaming* Marapthon didominasi oleh sentimen negatif, dengan jumlah data 673 untuk sentimen positif dan 1.016 untuk sentimen negatif. Hasil evaluasi model memperoleh nilai Accuracy sebesar 71,89% yang menunjukkan bahwa model memiliki performa yang cukup baik. Namun, masih terdapat keterbatasan dalam model yang ditunjukkan oleh ketidakseimbangan klasifikasi pada kelas positif. Hasil ini menunjukkan bahwa penelitian analisis sentimen pengguna X cenderung negatif terhadap *live streaming* Marapthon musim ketiga, meskipun masih terdapat cukup banyak pengguna yang memberikan opini positif.

5.2 Saran

Berdasarkan hasil penelitian analisis sentimen pengguna X terhadap *live streaming* Marapthon AAA Clan menggunakan *Naïve Bayes*, terdapat beberapa saran yang dapat diberikan. Penelitian selanjutnya dapat menggunakan metode lain atau membandingkan beberapa algoritma untuk meningkatkan performa klasifikasi. Selain itu, jumlah *Dataset* yang digunakan dapat diperluas agar hasil penelitian lebih akurat dan representatif.

Daftar Rujukan

- [1] F. H. Putra, D. Andriana, dan Widarti, “Pengaruh Tontonan Live Streaming Marapthon Terhadap Gaya Komunikasi Interpersonal Para Penonton Media Sosial X,” *Media Komunikasi Efektif*, vol. 2, no. 2, hlm. 102–108, Jul 2025.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [3] T. Y. Pahtoni dan H. Jati, “ANALISIS SENTIMEN DATA TWITTER TERKAIT CHATGPT MENGGUNAKAN ORANGE DATA MINING,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 2, hlm. 329–336, Apr 2024, doi: 10.25126/jtiik.20241127276.
- [4] S. Raschka, *Naive Bayes and Text Classification I*. 2014. [Daring]. Tersedia pada: <http://arxiv.org/abs/1410.5329>
- [5] Junaidy, M. Fauzan, dan R. Kaban, “Analisis Sentimen Publik Terhadap RUU KUHP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes,” *Jurnal Dinamika Informatika*, vol. 15, no. 1, hlm. 189–206, Mar 2026.
- [6] A. A. Sulaeman, C. Naya, M. Danny, dan M. M. Effendi, “Analisis Tingkat Sentimen Opini Publik Terhadap Kebijakan TV Digital di Platform X Menggunakan Multinomial Naïve Bayes,” *BULLETIN OF COMPUTER SCIENCE RESEARCH*, vol. 6, no. 2, hlm. 753–762, Feb 2026, doi: 10.47065/bulletincsr.v6i2.951.
- [7] F. M. Akmalya, B. N. A. Silaen, L. R. P. Pamungkas, R. P. Pratama, dan W. W. Suherli, “ANALISIS SENTIMEN PELAYANAN KERETA CEPAT WHOOSH MENGGUNAKAN ALGORITMA NAIVE BAYES,” *Journal of Information System Management (JOISM)*, vol. 07, no. 02, 2026.
- [8] E. S. Y. Pandie, A. F. Fanggidae, dan R. Lona, “Perbandingan Naïve Bayes dan K-NN dalam Analisis Sentimen Aplikasi X,” *Jurnal*

- Transformatika, vol. 22, no. 2, hlm. 97–107, Jan 2025, doi: 10.26623/f4k55e04.
- [9] K. A. Sari, A. Irsyad, D. N. Aini, Islamiyah, dan S. E. Ginting, “Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif,” *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 3, no. 1, hlm. 64–73, Jun 2024, doi: 10.30872/atasi.v3i1.1373.
- [10] A. Sitanggang, Y. Umaidah, dan R. I. Adam, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN SIANG GRATIS PADA MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, hlm. 2755–2762, Agu 2024, doi: 10.23960/jitet.v12i3.4902.
- [11] M. B. R. E. Nilapaksi, T. N. Fatyanosa, dan P. P. Adikara, “Analisis Sentimen Publik Terhadap Magang Berdampak 2025 di Platform X/Twitter Menggunakan Model Indobert,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, Apr 2026, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [12] S. M. Salsabila, A. A. Murtopo, dan N. Fadhilah, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” *Jurnal Minfo Polgan*, vol. 11, no. 02, hlm. 30–35, Sep 2022.
- [13] Rahmat, A. Rahim, dan Arbansyah, “PERBANDINGAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI ALIBABA DI GOOGLE PLAY STORE,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 2, hlm. 3050–3057, Apr 2025.
- [14] M. Kholilullah, Martanto, dan U. Hayati, “ANALISIS SENTIMEN PENGGUNA TWITTER(X) TENTANG PIALA DUNIA USIA 17 MENGGUNAKAN METODE NAIVE BAYES,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, hlm. 392–398, Feb 2024.
- [15] F. A. I. Indriyani, A. Fauzi, dan S. Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, hlm. 176–184, Jul 2023, doi: 10.37373/tekno.v10i2.419.
- [16] I. G. S. D. Putra dan I. N. T. A. Putra, “IMPLEMENTASI METODE NAÏVE BAYES PADA ANALISIS SENTIMEN PENGGUNA APLIKASI MOBILE KITA BISA,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, hlm. 1202–211, Apr 2025, doi: 10.23960/jitet.v13i2.6423.