

SENTIMEN ANALISIS PERILAKU PENGGEMAR COLDPLAY DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES

Luthfi Abdillah Fudholi¹, Nining Rahaningsih², Raditya Dinar Dana³

¹ Teknik Informatika, STMIK IKMI Cirebon

² Komputerisasi Akuntansi, STMIK IKMI Cirebon

³ Manajemen Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon

Abdillahluthfi694@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk memahami sentimen dan perilaku penggemar Coldplay di Twitter. Data tweet dikumpulkan, diolah, dan diklasifikasikan menggunakan metode Naive Bayes. Hasil penelitian ini diharapkan memberikan manfaat bagi Coldplay, peneliti, dan industri musik dalam memahami perilaku penggemar dan meningkatkan strategi pemasaran. Dalam penelitian ini, digunakan metode Naive Bayes untuk menganalisis perilaku pendukung Coldplay di Twitter. Data awal dikumpulkan dengan mengumpulkan tweet yang mengandung kata kunci yang relevan selama periode waktu tertentu. Kemudian, dilakukan teknik pemrosesan lanjutan seperti eliminasi stopword, normalisasi kata, dan stemming. Tweet-tweet tersebut diklasifikasikan menjadi dua kategori berdasarkan sentimennya: positif dan negatif. Dengan kemampuan untuk mengelola data dalam jumlah besar, metode Naive Bayes melakukan klasifikasi sentimen dengan memprediksi kategori melalui perhitungan probabilitas berdasarkan teorema Bayes dan asumsi independensi fitur. Kumpulan data yang telah dianalisis secara manual digunakan untuk melatih model dan menentukan parameter yang diperlukan dalam perhitungan probabilitas. Hasil analisis menunjukkan distribusi sentimen penggemar Coldplay di Twitter, yang mencerminkan popularitas dan penerimaan terhadap konten yang terkait dengan band. Penelitian ini menganalisis sentimen dan perilaku penggemar Coldplay di Twitter menggunakan metode Naive Bayes. Hasilnya menunjukkan akurasi model sebesar 80.25% dalam mengidentifikasi 29 perilaku positif dan 36 perilaku negatif, namun terdapat bias terhadap prediksi negatif.

Kata kunci: Analisis Sentimen, Naive Bayes, Twitter, Coldplay.

1. PENDAHULUAN

Coldplay sebagai subjek penelitian karena band ini dikenal secara luas dengan penggemar yang sangat antusias dan loyal di seluruh dunia. Coldplay juga terkenal dengan lagu-lagu yang penuh emosi dan makna, sehingga analisis sentimen terhadap perilaku penggemar Coldplay di media sosial Twitter dapat memberikan wawasan yang berharga tentang bagaimana penggemar merespons konten musik dan aktivitas band ini. Selain itu, dengan fokus pada Coldplay, penelitian ini dapat memberikan pemahaman yang lebih mendalam tentang interaksi antara penggemar dan band musik terkenal di era digital saat ini.

Salah satu media sosial yang sangat populer saat ini yaitu Twitter. Twitter merupakan media sosial yang berbentuk microblogging atau ngeblog secara singkat dalam satu paragraf dengan maksimal 280 huruf, karena jumlah huruf dalam satu kali tweet terbatas/dibatasi. Pengguna Twitter terdiri dari berbagai kalangan, seperti pejabat, selebritis, artis, hingga masyarakat biasa pada umumnya. Akan tetapi, kemudahan yang diberikan untuk berbagi informasi melalui media sosial tak luput dari penyalahgunaan yang dilakukan oleh penggunanya. Salah satu bentuk penyalahgunaan tersebut yaitu dalam berkomentar [1].

Penelitian sebelumnya di bidang ini telah mengeksplorasi berbagai metode analisis sentimen, termasuk penggunaan pengklasifikasi Naive Bayes.

Temuan utama menunjukkan bahwa meskipun metode ini mungkin efektif, masih ada ruang untuk perbaikan dalam hal akurasi dan pemahaman konteks. Misalnya, penelitian sering kali dibatasi oleh cakupan data atau bahasa yang dianalisis. Keterbatasan ini memberikan peluang untuk penelitian lebih lanjut, seperti mengeksplorasi analisis sentimen multibahasa atau mengintegrasikan data kontekstual untuk mencapai pemahaman yang lebih beragam [2].

Tujuan utama dari penelitian ini adalah untuk meningkatkan pemahaman sentimen penggemar Coldplay di Twitter dengan menggunakan pendekatan Naive Bayes. Penelitian ini bertujuan untuk mengisi kesenjangan pengetahuan dalam literatur analisis sentimen media sosial dan berkontribusi pada bidang ilmu komputer dengan meningkatkan metode analisis. Manfaat praktis dari penelitian ini dapat mencakup pengembangan strategi pemasaran yang lebih bertarget bagi para pemangku kepentingan dan wawasan yang lebih kaya mengenai sentimen publik untuk tujuan akademis.

Penelitian mengenai analisis sentimen terhadap konser musik *Coldplay* perlu dilakukan karena adanya kebutuhan untuk memahami respons dan reaksi audiens terhadap konser tersebut. Coldplay adalah salah satu band musik terkenal dengan penggemar yang sangat antusias. Dengan melakukan analisis sentimen terhadap konser *Coldplay*, dapat memahami bagaimana reaksi penggemar merespons konser

tersebut, apakah mereka merasa senang, kecewa, atau terinspirasi [3].

Jika penelitian ini berhasil, maka akan memberikan kontribusi yang signifikan terhadap pemahaman sentimen masyarakat di bidang ilmu komputer. Praktisi dapat menggunakan temuan ini untuk mengembangkan kampanye pemasaran yang lebih bertarget, dan peneliti dapat menggunakan temuan ini untuk terus mengembangkan metode analisis sentimen. Selain itu, implikasi penelitian ini dapat berdampak pada kemajuan teknologi alat dan aplikasi analisis sentimen.

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terdahulu terkait topik sentimen analisis perilaku penggemar *Coldplay* di media sosial Twitter menggunakan metode *Naive Bayes* dapat dijabarkan sebagai berikut:

Berdasarkan penelitian yang dilakukan oleh [4] penerapan algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* untuk analisis sentimen terhadap boy band *BTS* di media sosial Twitter menunjukkan bahwa *SVM* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan *Naive Bayes*. Hasil analisis sentimen tersebut mengungkapkan bahwa sebagian besar komentar yang diterima dari pengguna Twitter terkait *BTS* bersifat positif, mencerminkan popularitas grup vokal tersebut di kalangan remaja Indonesia. Studi ini memberikan wawasan berharga tentang penerimaan positif terhadap *BTS* di kalangan penggemar Indonesia di media sosial.

Penelitian selanjutnya dari penelitian yang berjudul "Sentiment Analysis Using *Naive Bayes* Algorithm Of The Data Crawler: Twitter". metode *Naive Bayes* telah ditemukan memiliki tingkat akurasi yang lebih baik dalam analisis sentimen dibandingkan dengan metode lain seperti *KNN* dan *SVM*. Sebagai contoh, dalam satu studi, metode *Naive Bayes* memiliki tingkat akurasi sebesar 80,90%, sementara *KNN* memiliki tingkat akurasi sebesar 75,58% dan *SVM* memiliki tingkat akurasi sebesar 63,99%. Analisis sentimen telah digunakan dalam berbagai aplikasi, termasuk menganalisis kepuasan publik terhadap kinerja pejabat terpilih, menganalisis kepuasan pelanggan dalam *e-commerce*, dan menganalisis sentimen terhadap film. Proses pengumpulan data untuk analisis sentimen umumnya melibatkan web scraping atau crawling data dari platform media sosial seperti Twitter. Data yang terkumpul kemudian diproses melalui pembersihan teks, tokenisasi, dan penyaringan untuk mengambil kata-kata penting. Secara keseluruhan, analisis sentimen menggunakan algoritma *Naive Bayes* telah terbukti menjadi alat yang berguna dalam mengukur pendapat dan sentimen publik terhadap berbagai topik.

Penelitian berikutnya dari [1] terdapat beberapa hasil dari literature review terkait analisis sentimen

cyberbullying *KPOP* di media sosial Twitter menggunakan metode *Naive Bayes*:

- Penelitian sebelumnya telah menunjukkan bahwa analisis sentimen dapat digunakan untuk memprediksi kecenderungan opini atau pendapat masyarakat terhadap suatu hal, apakah cenderung positif atau negatif.
- Terdapat dua pendekatan umum dalam melakukan analisis sentimen, yaitu berbasis machine learning dan berbasis leksikal. Pendekatan berbasis machine learning melibatkan pelatihan classifier pada dataset yang telah dilabeli secara manual, sementara pendekatan berbasis leksikal mengukur polaritas suatu kalimat atau dokumen berdasarkan sentimen kata-kata dan frasa-frasa.
- Dalam konteks analisis sentimen cyberbullying, penting untuk dapat mengklasifikasikan komentar ke dalam kelas posisi negatif untuk memahami pandangan masyarakat terhadap isu tersebut.

Meskipun informasi yang diberikan tidak terlalu rinci, hasil literature review tersebut memberikan gambaran tentang pentingnya analisis sentimen dalam konteks cyberbullying, khususnya terkait dengan penggemar *K-Pop* di media sosial Twitter.

Penelitian berikutnya dari [5] berdasarkan hasil literature review pada penelitian "Analisis Sentimen Kaum *LGBT* pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes*," ditemukan beberapa temuan penting:

- Algoritma *Naive Bayes* dipilih karena memiliki performa dan akurasi yang tinggi dalam proses klasifikasi.
- Penelitian ini menggunakan algoritma *Naive Bayes* untuk mengklasifikasi komentar masyarakat di Twitter terkait kaum *LGBT*, dengan tujuan untuk melihat seberapa besar penyebaran atau imbas pengaruh bullying terhadap kaum *LGBT* di Twitter.
- Data tweet yang diambil dari Twitter sebanyak 1.579 data tweet.
- Analisis sentimen terhadap kaum *LGBT* menggunakan metode *Naive Bayes* menghasilkan tingkat akurasi sebesar 95%, dengan nilai polaritas positif sebesar 77% dan polaritas negatif sebesar 150,2%.
- Pengujian *K-fold cross validation* dengan $k=10$ menghasilkan nilai akurasi sebesar 87,58%.
- Hashtag yang digunakan dalam penelitian ini antara lain *#gayindonesia*, *#LGBT*, *#lgbtindonesia*, dan *#lgbtq* untuk mengumpulkan data tweet terkait *LGBT* di Indonesia.
- Metode pengumpulan data menggunakan teknik scraping dengan library *tweepy* yang memberikan akses pada Twitter API.
- Media sosial memiliki peran dalam meningkatkan jumlah kasus *LGBT* karena fitur-fiturnya yang memungkinkan penyebaran informasi dengan cepat dan mudah.

Dengan demikian, penelitian ini memberikan wawasan yang penting mengenai analisis sentimen

kaum LGBT pada media sosial Twitter menggunakan algoritma *Naive Bayes*, serta menyoroti peran media sosial dalam ekspresi pendapat terkait isu-isu LGBT.

Penelitian selanjutnya dari [3]. dengan judul artikel “analisis sentimen terhadap calon presiden Indonesia 2019 dari media sosial Twitter menggunakan metode *Naive Bayes*” analisis sentimen merupakan riset komputasi dari sentimen, emosi, dan opini yang diluapkan kedalam teks, dengan tujuan untuk menentukan perilaku atau opini dari seorang penulis terkait suatu topik tertentu. Twitter sebagai platform komunikasi daring memungkinkan pengguna untuk mengekspresikan pendapat secara singkat dan jelas melalui tweet yang dapat dipublikasikan kepada berbagai kelompok. Prosesnya dapat ditingkatkan dengan labelisasi otomatis, peningkatan koleksi data latih, korpus stopwords, dan korpus stemming guna meningkatkan akurasi analisis sentimen, serta dengan penambahan proses preprocessing seperti n-gram dan negation untuk meningkatkan akurasi klasifikasi sentimen. Oleh karena itu, topik ini merupakan area penelitian yang menarik dan relevan untuk dipelajari lebih lanjut dalam konteks pemilihan kepala negara.

Menurut [6] Dalam penelitian mengenai Sentimen Analisis pada Media Sosial Twitter menggunakan Metode *Naive Bayes Classifier*, literature review menyoroti pentingnya analisis sentimen dalam memahami opini masyarakat terhadap entitas tertentu. Dengan menggunakan algoritma *Naive Bayes Classifier*, penelitian ini bertujuan untuk mengklasifikasikan data tweet yang terdapat pada akun Twitter @Sandiuno ke dalam kategori opini positif, negatif, atau netral. Selain itu, Twitter sebagai platform media sosial memberikan akses bagi pengguna untuk mengekspresikan pendapat mereka dalam bentuk teks singkat, sehingga diperlukan teknik preprocessing untuk membersihkan dan mengorganisir data teks agar dapat diolah secara efektif. Dengan demikian, penelitian ini menggabungkan konsep-konsep analisis sentimen, algoritma klasifikasi, dan pengolahan data teks untuk memahami sentimen pengguna Twitter terhadap akun @Sandiuno.

Menurut [7] dengan judul Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen Review Data Twitter BMKG Nasional penelitian ini menggunakan metode pengumpulan data sekunder dari akun Twitter resmi BMKG Nasional untuk membuat kamus opini negatif dan positif. Proses klasifikasi data dilakukan dengan algoritma *Naive Bayes* untuk mengkategorikan tweet ke dalam tiga kelas kata: positif, netral, dan negatif. Penelitian ini termasuk dalam fine-grained sentiment analysis, di mana analisis dilakukan pada level kalimat komentar, dengan data diproses menggunakan text mining. Proses crawling data Twitter dilakukan dengan Python 3.74 dan data disimpan dalam format CSV. Tujuan penelitian adalah untuk mencari komentar negatif, positif, dan netral pada review data Twitter BMKG Nasional menggunakan algoritma *Naive Bayes*, yang

menghasilkan klasifikasi berdasarkan nilai dari algoritma tersebut. Dengan demikian, penelitian ini memberikan kontribusi dalam menerapkan algoritma *Naive Bayes* untuk analisis sentimen terhadap data Twitter BMKG Nasional dengan fokus pada klasifikasi sentimen tweet.

Menurut [8] dengan judul Implementasi Algoritma *Naive Bayes* Terhadap Analisis Sentimen Opini Film Pada Twitter terdapat beberapa penelitian terkait analisis sentimen opini film dan penggunaan algoritma *Naive Bayes*. Penelitian oleh Ratnawati, dkk (2017) membahas pengklasifikasian opini film berdasarkan sentimen positif, negatif, dan netral dari komentar di Twitter menggunakan metode deep learning dengan akurasi 80,99%. Selain itu, Falahah, dkk (2015) juga melakukan penelitian terkait pengklasifikasian opini publik di Twitter mengenai layanan pemerintah dengan pendekatan *Naive Bayes*. Penelitian yang dijelaskan dalam PDF ini mencakup implementasi algoritma *Naive Bayes* dalam analisis sentimen opini film pada Twitter dengan akurasi sebesar 90%. Dari literatur review tersebut, dapat disimpulkan bahwa analisis sentimen opini film di media sosial, terutama Twitter, merupakan topik yang menarik dan penting untuk dipelajari dengan berbagai metode klasifikasi yang telah dikembangkan.

Menurut [9] dengan judul sentiment analysis of the community in the Twitter to the 2020 election in Pandemic Covid-19 by method *Naive Bayes Classifier* studi ini menggunakan teknik text mining seperti *tokenization*, *vectorization*, dan *modeling* untuk menganalisis dataset 1200 tweet yang diekstrak dari Twitter. Dengan menerapkan klasifikasi *Naive Bayes*, penelitian bertujuan untuk mengelompokkan sentimen pengguna Twitter ke dalam kategori positif dan negatif terkait pemilihan umum 2020. Text mining memainkan peran penting dalam mengekstrak pengetahuan dan informasi dari data teks, memungkinkan para peneliti memahami sentimen dan pendapat masyarakat terhadap peristiwa penting seperti pemilihan umum. Langkah-langkah preprocessing yang terlibat dalam *text mining*, seperti *case folding*, menghilangkan angka atau tanda baca, dan menghilangkan *stopwords*, penting untuk mempersiapkan data untuk analisis.

Menurut [10] penelitian ini menunjukkan bahwa “analisis sentimen terhadap kinerja Dewan Perwakilan Rakyat (DPR) melalui media sosial Twitter menggunakan metode *Naive Bayes Classifier* (NBC)” telah berhasil dilakukan. Dengan memanfaatkan NBC dalam proses pelatihan dan klasifikasi teks, penelitian ini mampu mengidentifikasi polaritas sentimen positif, netral, dan negatif dari data tweet yang dianalisis. Hasil penelitian menunjukkan bahwa DPR menerima 95 tweet positif dengan sentimen positif sebesar 75%, 693 tweet netral dengan sentimen netral sebesar 79%, dan 758 tweet negatif dengan sentimen negatif sebesar 82%. Dengan akurasi sebesar 80% berdasarkan data testing sebanyak 20%, penelitian ini memberikan

kontribusi dalam memahami pandangan masyarakat terhadap kinerja DPR melalui platform Twitter.

2.2. Naïve Bayes

Metode Naïve Bayes adalah algoritma klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur dalam data adalah independen satu sama lain. Metode ini digunakan untuk memprediksi kelas atau label dari suatu data berdasarkan fitur-fitur yang ada.

Dalam Naive Bayes, rumus yang digunakan untuk menghitung probabilitas posterior $P(C|X1, X2, \dots, Xn)$ adalah sebagai berikut:

$$P(C|X1, X2, \dots, Xn) = \frac{P(X1|C) P(X2|C) \dots P(Xn|C) P(C)}{P(X1, X2, \dots, Xn)}$$

Di mana:

- $P(C|X1, X2, \dots, Xn)$ adalah probabilitas kelas C diberikan fitur-fitur $X1, X2, \dots, Xn$.

- $P(Xi|C)$ adalah probabilitas fitur ke-i diberikan kelas C.

Metode Naive Bayes digunakan untuk mengklasifikasikan data berdasarkan perhitungan probabilitas dengan asumsi bahwa setiap fitur adalah independen dari fitur lainnya. Dalam hal ini, Naive Bayes menghitung probabilitas fitur-fitur yang muncul dalam setiap kelas dan menggunakan rumus Naive Bayes untuk memprediksi kelas yang paling mungkin untuk instance baru.

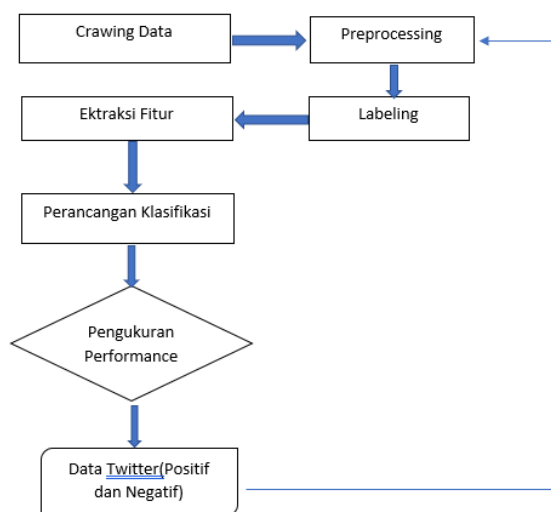
Dalam kasus ini, metode Naive Bayes digunakan untuk menganalisis sentimen perilaku penggemar Coldplay di media sosial Twitter. Akurasi sebesar 80.25% menunjukkan sejauh mana model berhasil mengklasifikasikan sentimen dengan benar. Namun, perlu diperhatikan bahwa hasil Confusion Matrix mengindikasikan adanya kesalahan dalam memprediksi beberapa data.

Sebagai catatan, rumus dan hasil yang disajikan di atas didasarkan pada informasi yang diberikan dalam Confusion Matrix, dan implementasinya dapat bervariasi tergantung pada alat atau platform yang digunakan.

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode analisis sentimen menggunakan metode *Naive Bayes*. Metode *Naive Bayes* adalah metode klasifikasi teks yang sederhana namun sangat akurat dan efektif. Metode ini bekerja dengan mengasumsikan bahwa setiap fitur teks independen satu sama lain. Penelitian ini bertujuan untuk menganalisis sentimen tweet dari penggemar *Coldplay* di media sosial Twitter. Sentimen tweet diklasifikasikan menjadi tiga kategori, yaitu positif dan negatif. Analisis sentimen ini dilakukan untuk

mengetahui pandangan penggemar *Coldplay* terhadap band *Coldplay*.



Gambar 1. Tahapan Metode Penelitian

- a. **Crawling Data**
Data *Twitter* berisi opini masyarakat yang mengandung unsur komentar positif dan negatif, pada *Twitter* dengan data tahun 2023 melalui aplikasi *Twitter*.
- b. **Pelabelan**
Pelabelan sentimen Tweet Memberikan label sentimen pada setiap tweet yang telah diambil dari *Twitter*.
- c. **Preprocessing**
Preprocessing Melakukan preprocessing terhadap data yang telah diperoleh untuk menghasilkan data yang bersih dan siap untuk diproses lebih lanjut.
- d. **Ekstraksi Fitur**
Melakukan ekstraksi fitur dari teks yang telah dipreprocessing untuk digunakan dalam proses klasifikasi.
- e. **Perancangan Klasifikasi**
Perancangan klasifikasi Menyusun rancangan klasifikasi untuk mengelompokkan tweet-tweet berdasarkan kategori sentimen yang telah ditentukan sebelumnya.
- f. **Pengukuran Performance**
Evaluasi kinerja model *Naive Bayes* untuk mengukur seberapa baik model tersebut dapat mengklasifikasikan data.

3.1. Sumber Data

Untuk melakukan “analisis sentimen perilaku penggemar *Coldplay* di media sosial *Twitter* menggunakan metode *Naive Bayes*”, langkah pertama adalah mengidentifikasi sumber data yang sesuai. Dalam konteks ini, sumber data utama adalah tweet-tweet yang terkait dengan *Coldplay* di *Twitter*. Penelitian ini dapat menggunakan *Twitter API* atau alat web scraping untuk mengumpulkan data yang mencakup tweet-tweet penggemar *Coldplay* dengan

menggunakan kata kunci atau hashtag yang relevan. Setelah data terkumpul, langkah selanjutnya adalah memproses dan menyiapkan data tersebut agar sesuai dengan format yang diperlukan untuk analisis menggunakan metode Naive Bayes, termasuk dalam hal ini pemisahan menjadi sentimen positif dan negatif. Dengan demikian, sumber data dari media sosial Twitter akan menjadi landasan untuk menjalankan analisis sentimen dengan metode Naive Bayes terhadap perilaku penggemar Coldplay. https://colab.research.google.com/drive/1gmloPzl_gf_35bd9rbVyx3aaPMSOcjn?usp=sharing Tautan yang disediakan mengarah ke sebuah notebook di Google Colab yang berfungsi sebagai titik awal untuk melakukan pengumpulan data dari platform media sosial Twitter. Notebook ini dirancang untuk mengumpulkan ulasan dan peringkat pengguna dari Twitter. Dengan menggunakan skrip atau kode Python di Google Colab, pengguna dapat mengambil data yang relevan dari Twitter untuk analisis lanjutan.

3.2. Teknik Pengumpulan Data

Penelitian ini akan menggunakan teknik web scraping untuk mengumpulkan data. Proses dimulai dengan mengidentifikasi URL pada halaman Twitter yang mencakup ulasan dan peringkat aplikasi Twitter. Selanjutnya, alat atau pustaka web scraping seperti BeautifulSoup atau Selenium akan dipakai untuk mengekstraksi informasi dari halaman tersebut. Data yang akan diambil mencakup ulasan pengguna, peringkat, tanggal ulasan, dan komentar terkait.

Setelah selesai mengumpulkan data, langkah selanjutnya adalah melakukan pra-pemrosesan. Ini melibatkan pembersihan teks dari karakter khusus, penghapusan ulasan duplikat, dan mengubah peringkat menjadi label sentimen (positif dan negatif). Data yang sudah diproses akan digunakan untuk melatih model analisis sentimen menggunakan metode Naive Bayes.

3.3. Teknik Analisis Data

Teknik analisis data dalam penelitian ini adalah sebagai berikut:

- Menggunakan metode Naive Bayes untuk mengklasifikasikan sentimen tweet.
- Menganalisis hasil klasifikasi sentimen tweet untuk mengetahui perilaku penggemar Coldplay di media sosial Twitter.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini menjelaskan tentang sentimen analisis perilaku penggemar *Coldplay* di media sosial twitter menggunakan metode *Naive Bayes*. Hasil ini diperoleh dengan mengumpulkan data review pengguna dan mengolah data serta menerapkan *RapidMiner* sebagai Algoritma Analisis Sentimen.

4.1. Crawling Data

Tahap pengumpulan data melalui Teknik Web Scraping pada *Google Colab* melalui Twitter.

```
# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list
!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v

[ ] # Crawl Data

filename = 'coldplay.csv'
twitter_thread_url = 'https://x.com/tvOneNews/status/1658261283966447616?s=20'
limit = 1000
```

Gambar 2. Crawling Data

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang
0	Fri Nov 10 14:13:59 +0000 2023	1722980927327051875	Konser Coldplay di Indonesia dihadapi dengan p...	0	0	0	2	in 1466673...
1	Fri Nov 10 06:47:32 +0000 2023	1722868577739706499	PA 212 melakukan aksi demo di depan Mabes Polr...	0	0	2	3	in 1714125...
2	Sun Nov 05 13:58:56 +0000 2023	1721165201356227056	Mana tadi dibilang bulan depan kita kumpul lag...	0	1	0	0	in
3	Mon Jul 24 01:39:48 +0000 2023	1683290855535177729	Akun ini sok nantingin konser band LGBT Coldpl...	0	0	0	0	in 1636627...

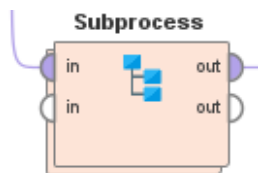
Gambar 3. Data hasil Crawling Komentar Coldplay

4.2. Preprocessing

Preprocessing adalah tahapan awal dalam analisis data yang melibatkan langkah-langkah seperti pembersihan data, normalisasi teks, tokenisasi, stemming atau lemmatization, dan filtering untuk membersihkan, mengubah, dan mempersiapkan data mentah agar siap untuk analisis lebih lanjut. Tujuannya adalah meningkatkan kualitas data, mengurangi noise, dan mempersiapkan data untuk proses analisis seperti klasifikasi sentimen atau analisis statistik.

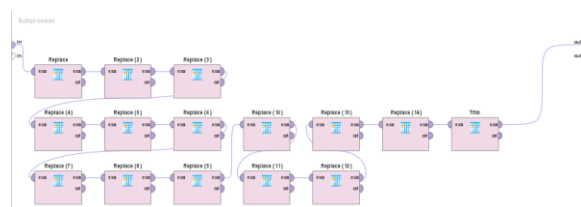
4.3. Pembersihan Data

Pembersihan data adalah proses mengidentifikasi, menghapus, atau memperbaiki data yang tidak lengkap, tidak akurat, tidak relevan, atau tidak terstruktur dalam sebuah dataset. Langkah-langkah pembersihan data meliputi deteksi dan penanganan missing values, penanganan outlier, deduplikasi data, dan penanganan data yang tidak konsisten.



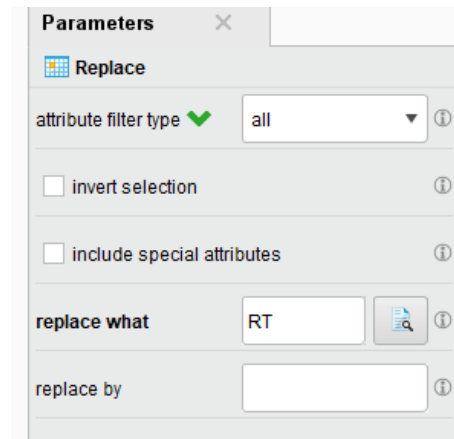
Gambar 4. Operators Subprocess

Subprocess adalah bagian kecil dari proses utama yang memiliki langkah-langkah atau aktivitas tersendiri untuk mencapai tujuan tertentu dalam proses tersebut. Subproses memungkinkan pemisahan tugas yang kompleks menjadi bagian-bagian yang lebih kecil dan terkelola dengan lebih baik, membantu dalam pemahaman yang lebih baik terhadap proses secara keseluruhan, dan memungkinkan fokus pada detail-detail yang spesifik dalam setiap subproses. Dengan adanya subproses, proses utama dapat dipecah menjadi bagian-bagian yang lebih mudah dikelola, dipantau, dan ditingkatkan, serta memungkinkan perbaikan atau penyesuaian pada bagian-bagian tertentu tanpa harus mengganggu keseluruhan proses.



Gambar 5. Operators Replace

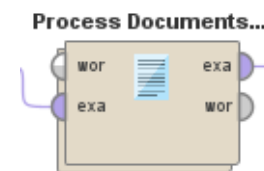
Operator "replace" digunakan dalam pemrograman untuk menggantikan nilai tertentu dalam suatu string, list, atau struktur data dengan nilai baru. Dengan "replace", pengguna dapat melakukan penggantian nilai secara spesifik sesuai dengan kebutuhan dalam program.



Gambar 6. Operators Filter Atribut

4.4. Normalisasi Teks

Normalisasi teks adalah proses mengubah teks menjadi format standar atau bentuk yang seragam. Ini melibatkan langkah-langkah seperti mengubah huruf menjadi lowercase, menghilangkan karakter khusus, mengganti sinonim, dan mengatasi variasi ejaan untuk memudahkan analisis teks.



Gambar 7. Operators Process Documents From Data

Proses dokumen data merujuk pada serangkaian langkah yang dilakukan untuk mengelola dokumen secara efisien dan efektif, termasuk pengumpulan, pengolahan, penyimpanan, dan distribusi dokumen dalam suatu organisasi atau sistem. Tujuannya adalah memastikan informasi dalam dokumen dapat diakses, digunakan, dan dikelola sesuai kebutuhan dan tujuan yang ditetapkan. Dengan menjalankan proses dokumen dengan baik, organisasi dapat meningkatkan produktivitas, efisiensi, dan akurasi dalam pengelolaan informasi.

4.5. Tokenisasi



Gambar 8. Operators Tokenize

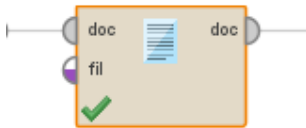
Tokenisasi adalah proses membagi teks menjadi unit-unit yang lebih kecil yang disebut token. Token bisa berupa kata, frasa, atau karakter tergantung pada kebutuhan analisis. Tokenisasi membantu dalam

memahami struktur teks dan mempersiapkan data untuk analisis lebih lanjut.

4.6. Filtering

Filtering adalah proses menghilangkan atau mempertahankan sebagian data berdasarkan kriteria tertentu. Dalam konteks text mining, filtering dapat digunakan untuk menghapus stopwords (kata-kata umum yang tidak memberikan informasi penting), menghilangkan kata-kata dengan frekuensi rendah, atau mempertahankan hanya kata-kata tertentu yang relevan untuk analisis. Filtering membantu menyaring data sehingga hanya informasi yang penting dipertahankan untuk analisis lebih lanjut.

Filter Stopwords (Dictionary)

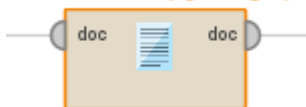


Gambar 9. Operators Filter Stopword

Filter stopword merupakan teknik penting dalam praproses teks untuk meningkatkan efektivitas analisis. Teknik ini membantu menghilangkan kata-kata umum seperti "di", "dari", "yang", dan lain-lain, sehingga fokus analisis menjadi terarah pada kata-kata yang mengandung makna substansial. Hal ini dapat membantu mempercepat proses analisis, meningkatkan performa model machine learning, dan meningkatkan akurasi analisis dengan fokus pada kata-kata yang relevan.

Filter stopword dapat ditulis dalam berbagai bahasa pemrograman dan tools, seperti library NLTK untuk Python, library stopwords untuk R, dan operator "Stop Word Removal" di RapidMiner. Meskipun sederhana, filter stopword memberikan manfaat yang signifikan dalam analisis teks. Gunakan teknik ini untuk mendapatkan hasil analisis yang lebih bermakna dan akurat.

Filter Tokens (by Length)



Gambar 10. Operators Filter Tokens

Operator filter tokens berfungsi sebagai mekanisme pembersihan data teks secara sistematis. Operator ini memungkinkan pengguna untuk menghilangkan noise berupa tanda baca, simbol, dan karakter khusus. Selain itu, operator ini juga mampu melakukan normalisasi teks dengan cara mengubah huruf besar menjadi huruf kecil atau mengubah kata menjadi bentuk tunggal.

4.7. Labeling

Labeling adalah proses pemberian kategori atau tag pada data untuk mengidentifikasi atau

mengklasifikasikan data berdasarkan karakteristik tertentu. Dalam analisis data, labeling memberikan informasi pada setiap data, seperti dalam analisis sentimen yang dapat diberi label positif atau negatif. Hal ini memungkinkan sistem machine learning mempelajari pola dari data yang telah diberi label untuk membuat prediksi atau analisis lebih lanjut.

1	Komentar	Sentimen
2	pembawa bencana koi di datangkan?	Negatif
3	hadah ign cari sensasi deh	Positif
4	Ekspresi ku saat ini	Positif
5	Dipikernya negara itu punya sekelompok organisasi ya ... bisaA'nya nolak konser musik.	Positif
6	Kalian2 hrs nonton konser coldplay. Kim di surga/ neraka toh akan ada konser. Coldplay pun kalo konser ditap negara juga lihat2 situasi. Nggak mungkin kiba?Positif	Positif
7	Setuju Mending Diagilitan aja	Negatif
8	Denger aja lagunya napa ngurisin pribadinya ? Kto ga mampu beli tiketnya ga usah ngadâ !!	Positif
9	hadah nihh orman	Positif
10	Lah apa hubungannya?Kalo masih pada doyan makan indomie dan palek hp china ign kebanyakan maen tola2an dah malu sama wiro sableng	Positif
11	ini yg dulu suka demo jllid2an ya di monas. Hmm ga boleh liat org semang yak. Pas demo aja ga putus2 kita warga DKI ga marah toh hehehe	Positif
12	Indonesia bakal dikenal sebagai negara tolah sana tolah sini	Positif
13	wah sena edi	Positif
14	Gov suka ni	Positif
15	Knpa pas blackpink, Ademi2 aja ya .. giliran Coldplay... hmmmzzz	Positif
16	ini juga yg ngasih berita sama kortletnya dengan zzz	Negatif
17	ini pendukung SWANSEA CITY yah	Positif
18	Cie pendukung tuboli pada nongol	Negatif
19	Pak zzz masih hobi? suka, msem politik, wajar kalau muncul lagi.	Positif
20	Prek zzz emang masih ada....	Positif
21	Coldplay blm konser LGBT di indo sdt ada emang gk tau ya? cb try ke jurnalis ada kasus pembunuhan LGBT yg pernah heboh	Positif
22	Waktu konser blackpink di dkt suk diam sga gik kemana	Positif
23	Cari kesempatan utk amloob dan nasi bunekus eratis. Dasar balinean Ramo. para banasat	Positif

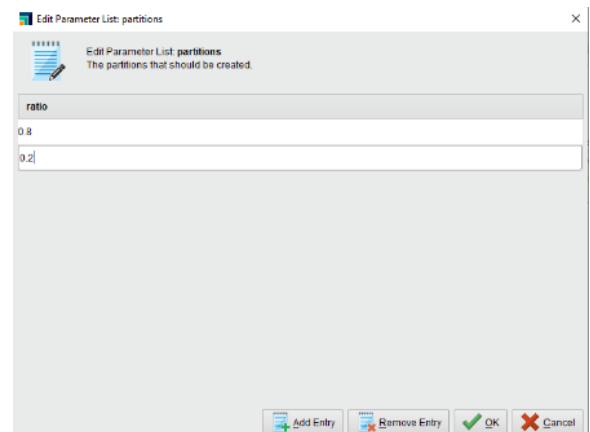
Gambar 11. Data Labeling

4.8. Ekstraksi Fitur

Ekstraksi fitur adalah proses mengidentifikasi, mengekstrak, dan menggambarkan fitur penting dari data mentah untuk digunakan dalam analisis lebih lanjut. Tujuannya adalah mengubah data menjadi representasi yang lebih kompak dan bermakna agar algoritma machine learning dapat bekerja efisien. Dalam analisis teks, ekstraksi fitur melibatkan konversi teks menjadi vektor fitur numerik. Proses ini penting dalam mempersiapkan data untuk analisis lanjutan seperti klasifikasi atau prediksi.

4.9. Perancangan Klasifikasi

Perancangan klasifikasi adalah proses merancang model klasifikasi untuk memprediksi label data baru berdasarkan fitur-fitur yang ada. Langkah-langkahnya meliputi pemisahan data training dan testing, melakukan klasifikasi dengan *Naïve Bayes* dan pengukuran performance. Tujuannya adalah menghasilkan model akurat untuk klasifikasi data dengan tepat.



Gambar 12. Ratio untuk mengukur Performance

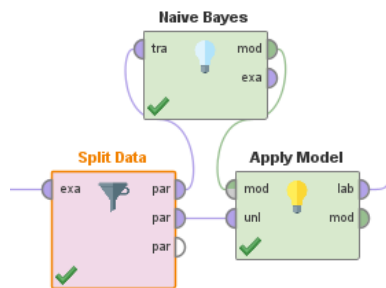
4.10. Data Training

Pengajaran data sangat penting dalam membangun model pengajaran mesin yang efektif. Data pembelajaran dapat digambarkan sebagai

kumpulan materi pembelajaran yang diberikan kepada model. Model akan mempelajari pola dan hubungan yang ada dalam data pelatihan, yang biasanya sekitar 80% dari data. Tujuan dari proses pembelajaran ini adalah untuk memberi model kemampuan untuk melakukan prediksi atau pengambilan keputusan di masa mendatang. Dengan kata lain, kualitas dan representasi data pelatihan sangat memengaruhi seberapa baik model bekerja pada akhirnya. Kemampuan model untuk generalisasi dan menyesuaikan diri dengan situasi baru sebanding dengan kualitas data pelatihan.

4.11. Data Testing

Setelah proses pembelajaran selesai, langkah selanjutnya adalah evaluasi. Untuk tahap ini, kita membutuhkan data testing. Data testing biasanya berjumlah 20% dari keseluruhan data, dan merupakan kumpulan data "baru" yang belum pernah ditemui oleh model selama proses pelatihan. Dengan menguji performa model pada data testing, kita dapat mengetahui apakah model tersebut mampu untuk bekerja dengan baik pada data yang belum pernah dilihatnya sebelumnya. Hal ini penting untuk memastikan bahwa model tidak hanya menghafal pola-pola yang ada di dalam data training, namun juga dapat menggeneralisasi pengetahuannya untuk menghadapi situasi yang baru. Pemisahan data training dan data testing menjadi langkah penting untuk membangun model machine learning yang handal dan akurat.



Gambar 13. Operators untuk perancangan Klasifikasi

4.12. Split Data

Operator Split Data di RapidMiner memiliki fungsi untuk memecah data. Bagian data yang sudah dipisah dengan operator ini bisa dilakukan secara acak untuk keperluan memvalidasi model atau kriteria tertentu, seperti rentang tanggal atau kategori pelanggan. Operator ini sangat berguna untuk mempersiapkan data latih dan data uji saat kita ingin membangun model Machine Learning dan mengamankan kandungan analisis data yang lebih fokus pada sub-bagian tertentu.

4.13. Naive Bayes

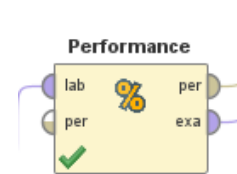
Naive Bayes ialah suatu teknik klasifikasi probabilitas yang sederhana namun memiliki performa yang tinggi di dalam pembelajaran mesin. Algoritma

ini bekerja dengan cara menghitung probabilitas bahwa suatu data tertentu termasuk ke dalam suatu kelas spesifik, berdasarkan pada Teorema Bayes. Dengan kata lain, Naive Bayes akan memprediksi kelas dari data baru dengan membandingkan kemiripannya dengan data yang sudah memiliki label pada tahap pelatihan model. Meskipun memiliki asumsi bahwa fitur-fitur data bersifat independen (tidak saling berpengaruh) dan rentan terhadap outlier (data yang sangat berbeda dengan mayoritas data lain), Naive Bayes tetap menjadi pilihan yang menarik karena kemudahan penggunaan, kecepatan dalam proses pelatihan model, serta kinerja yang baik dalam berbagai jenis data dan masalah klasifikasi.

4.14. Apply Model

Operator Apply Model dalam RapidMiner sangat penting karena dapat digunakan untuk menerapkan model yang telah dilatih pada data baru. Melalui operator ini, Anda dapat melakukan prediksi nilai atribut target, mengklasifikasikan data baru, mengevaluasi performa model, serta menemukan pola dan tren. Kemudahan penggunaan operator ini sangat bermanfaat untuk berbagai aplikasi, seperti prediksi harga rumah, klasifikasi email spam, dan penentuan kelayakan kredit.

4.15. Pengukuran Performance



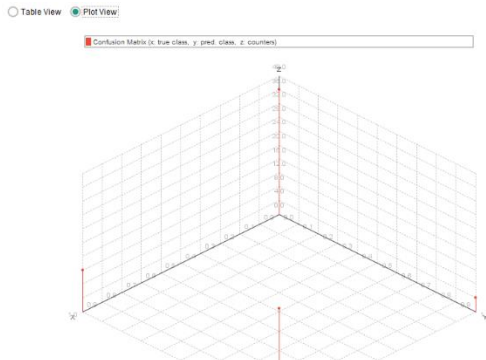
Gambar 14. Operators Performance

Operator Performance di RapidMiner membantu Anda menilai performa model machine learning dengan menghitung metrik seperti akurasi, presisi, dan recall. Operator ini mudah digunakan dan membantu Anda memilih model terbaik, mengidentifikasi area yang perlu diperbaiki, dan memastikan model siap digunakan dalam aplikasi real-world.

Model	True Positif	True Negatif	Presisi
Model A	35	12	75.00%
Model B	4	20	87.50%
Model C	99.99%	70.77%	

Gambar 15. Hasil Pengukuran Performance

Pengukuran performa adalah proses evaluasi kinerja suatu sistem atau model berdasarkan kriteria tertentu. Tujuannya adalah menilai sejauh mana entitas mencapai tujuan atau berfungsi dengan baik. Dalam analisis data, metrik performa seperti akurasi, presisi, recall digunakan. Pengukuran performa penting untuk evaluasi dan pengambilan keputusan yang lebih baik.



Gambar 16. Matriks Akurasi

4.16. Pembahasan Penelitian

Berdasarkan hasil penelitian yang telah dijelaskan pada sub bab 4.1 di atas maka dapat di deskripsikan pembahasan yang berkaitan dengan hasil penelitian yang diperoleh sebagai berikut. Evaluasi menunjukkan bahwa model ini dapat memprediksi perilaku penggemar Coldplay di media sosial dengan akurasi sebesar 80.25%. Ini berarti model berhasil mengklasifikasikan data dengan benar pada 80.25% dari keseluruhan evaluasi. Meskipun performanya cukup baik, perlu dicatat bahwa akurasi ini didasarkan pada data yang digunakan dan mungkin berbeda pada data lain. Analisa lebih lanjut diperlukan untuk pengujian yang lebih menyeluruh. Model ini berhasil mengidentifikasi 29 perilaku positif dan 36 perilaku negatif dengan benar. Namun, model salah mengklasifikasikan 4 perilaku negatif sebagai positif dan 12 perilaku positif sebagai negatif. Hasil ini menandakan bahwa model lebih akurat dalam memprediksi perilaku negatif. Kemungkinan hal ini disebabkan oleh distribusi data yang tidak seimbang, di mana data perilaku negatif lebih banyak tersedia dibandingkan data perilaku positif. Analisa lebih lanjut diperlukan untuk memahami bias ini dan meningkatkan akurasi model dalam mengklasifikasikan kedua jenis perilaku.

Algoritma Naive Bayes merupakan salah satu metode klasifikasi probabilistik yang populer digunakan dalam analisis sentimen. Algoritma ini memiliki beberapa kelebihan yang membuatnya cocok untuk memprediksi sentimen analisis perilaku penggemar Coldplay di media sosial, seperti:

- a. Kesederhanaan: Algoritma Naive Bayes relatif mudah untuk dipahami dan diimplementasikan dibandingkan dengan algoritma klasifikasi lainnya. Hal ini membuatnya mudah untuk diintegrasikan dengan sistem analisis sentimen yang ada.
- b. Efisiensi: Algoritma Naive Bayes sangat efisien dalam menghitung probabilitas kelas untuk data baru. Hal ini membuatnya ideal untuk aplikasi real-time yang membutuhkan prediksi sentimen yang cepat.
- c. Kemampuan Beradaptasi: Algoritma Naive Bayes dapat beradaptasi dengan data baru dengan mudah. Hal ini membuatnya cocok untuk menganalisis

data media sosial yang terus berubah dan berkembang.

- d. Robustness: Algoritma Naive Bayes cukup robust terhadap noise dan data yang hilang. Hal ini membuatnya cocok untuk menganalisis data media sosial yang sering kali tidak lengkap atau tidak akurat.

5. KESIMPULAN DAN SARAN

Algoritma Naive Bayes merupakan alat yang efektif untuk memprediksi sentimen analisis perilaku penggemar Coldplay di media sosial. Kesederhanaan, efisiensi, kemampuan beradaptasi, dan robustness-nya membuatnya ideal untuk aplikasi real-time. Studi menunjukkan bahwa model dapat memprediksi kinerja Coldplay di media sosial dengan akurasi sebesar 80.25%, akurasi ini bergantung pada data yang ditemukan di media sosial Twitter dari para penggemar Coldplay. Model menemukan 29 tindakan positif dan 36 tindakan negatif, tetapi juga menemukan 4 tindakan negatif sebagai positif dan 12 tindakan positif sebagai negatif, yang menunjukkan bahwa prediksinya bias. Penelitian di atas menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk mengidentifikasi sentimen positif dan negatif dari tweet penggemar Coldplay. Informasi ini kemudian dapat digunakan untuk memahami perilaku penggemar, meningkatkan strategi pemasaran, dan mengembangkan produk dan layanan yang lebih baik. Saran yang dapat diberikan adalah untuk melanjutkan penelitian ini dengan memperluas cakupan analisis sentimen terhadap penggemar Coldplay di platform media sosial lainnya, serta mempertimbangkan penggunaan metode klasifikasi yang lebih kompleks untuk meningkatkan akurasi dan keberagaman hasil analisis sentimen. Selain itu, disarankan untuk melakukan validasi hasil analisis dengan melibatkan responden atau pengguna media sosial secara langsung untuk mendapatkan masukan dan feedback yang lebih mendalam.

DAFTAR PUSTAKA

- [1] A. Putri and A. Muzakir, ‘ANALISIS SENTIMEN CYBERBULLYING KPOP DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES’, vol. 7, no. 9, 2022.
- [2] M. Wankhade, A. C. S. Rao, and C. Kulkarni, ‘A survey on sentiment analysis methods, applications, and challenges’, *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [3] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, ‘JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes’, [Online]. Available: <http://dev.twitter.com>.
- [4] R. Noviana and I. Rasal B A Jurusan, ‘PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN

- BOY BAND BTS PADA MEDIA SOSIAL TWITTER', *JTS*, vol. 2, no. 2.
- [5] A. M. Anjani, A. A. Chamid, and A. C. Murti, 'ANALISIS SENTIMEN KAUM LGBT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES', Halaman 1~8, 2022.
- [6] M. Imam Syafii, 'Sentimen analisis Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier (NBC)'.
[7] D. Darwis, N. Siskawati, and Z. Abidin, 'Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional', vol. 15, no. 1.
- [8] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, 'Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter', vol. 3, no. 1.
- [9] A. Muzaki and A. Witanti, 'SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER', *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [10] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, 'Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier', *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.