JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3S1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3S1.7539

# ANALISIS SENTIMEN PENGGUNA KRL TERHADAP KAI COMMUTER LINE MELALUI SOSIAL MEDIA X DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER

**Risa Nur Fitriyani[1*], Mohamad Jajuli[2], Yuyun Umaidah[3]**

[1,2,3]Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

**Keywords:**
Sentiment Analysis, Naïve Bayes, KRL.

**Corespondent Email:**
risanur.fitriyani2720@gmail.com

**Abstrak.** *KRL merupakan sistem transportasi yang diandalkan berbagai kalangan untuk bepergian, tiap tahunnya pengguna KRL mengalami kenaikan penumpang. Dengan bertambahnya penumpang tiap tahun, opini yang muncul terkait KAI Commuter Line semakin beragam, baik opini positif maupun negatif, sehingga diperlukan analisis terkait opini pengguna KRL terhadap KAI Commuter Line. Penting untuk memahami sentimen pengguna bagi pemegang kebijakan sehingga dapat menanggulangi persepsi negatif yang muncul. Penelitian ini bertujuan untuk menganalisis sentimen pengguna KRL dan mengukur tingkat akurasi algoritma Naïve Bayes dalam mengklasifikasikan teks. Penelitian ini menggunakan metodologi penelitian Knowledge Discovery in Database (KDD) untuk menemukan informasi yang penting yang berguna. Pengumpulan data dilakukan dengan teknik crawling pada sosial media X dengan bantuan tools harvest-tweet. Pada analisis dengan 909 data bersih dari 6.175 data yang terkumpul, terdapat 485 sentimen positif dan 424 sentimen negatif. Pada hasil evaluasi akurasi, ditemukan bahwa* pembagian data dengan rasio 80:20 mendapatkan hasil akurasi tertinggi sebesar 79%.



**Abstract.** *The KRL is a transportation system relied upon by various groups for travel, and each year the number of KRL users increases. With the increase in passengers each year, opinions regarding the KAI Commuter Line have become increasingly diverse, both positive and negative, making it necessary to analyze KRL users' opinions of the KAI Commuter Line. It is important for policy makers to understand user sentiment so that they can address any negative perceptions that arise. This study aims to analyze KRL users' sentiments and measure the accuracy of the Naïve Bayes algorithm in classifying text. The study employs the Knowledge Discovery in Database (KDD) research methodology to identify important and useful information. Data collection was conducted using crawling techniques on social media X with the help of the harvest-tweet tool. In the analysis of 909 clean data from 6,175 collected data, there were 485 positive sentiments and 424 negative sentiments. In the accuracy evaluation results, it was found that data division with a ratio of 80:20 achieved the highest accuracy of 79%.*

## 1. PENDAHULUAN

Saat ini pertumbuhan teknologi tidak dapat dipungkiri, baik teknologi informasi, kesehatan bahkan kendaraan. Salah satu kendaraan yang terus mengalami perkembangan adalah kereta, dari kereta uap sampai kereta listrik yang saat ini banyak digunakan di berbagai negara. Indonesia sendiri memiliki beberapa jenis kereta listrik, seperti KRL, MRT, LRT, dan JCJB. Diantara kereta listrik tersebut, yang saat ini banyak dipilih masyarakat Indonesia adalah KRL. KRL merupakan kereta non-lokomotif yang bergerak dari adanya energi yang dihasilkan oleh listrik [1]. Berdasarkan informasi yang didapatkan dari laman resminya KAI Commuter Line pada Juli 2024, Commuter Line Jabodetabek melayani sebanyak 1.048 perjalanan dengan 87 rangkaian tiap harinya . Pada semester pertama tahun 2024, KAI Commuter Line mencatat rata-rata pengguna di hari kerja sebanyak 961.051 dan pada hari libur sebanyak 709.730 [2]. Bahkan diprediksi pada tahun 2027 akan ada sebanyak 410 juta penumpang yang dilayani oleh KRL Commuter Line [3]. Peningkatan penumpang tiap tahunnya merupakan indikasi efektivitas angkutan umum dalam mengurangi kemacetan sebagai transportasi umum [4].

Dengan adanya, peningkatan penumpang, maka layanan yang diberikan KRL Commuter Line juga akan berpengaruh terhadap persepsi penggunanya, banyak portal berita yang menyalurkan opini pengguna KRL terhadap KAI Commuter Line. Salah satu opini positif yang muncul adalah biaya yang tetap murah meskipun saat periode waktu liburan [5]. Sedangkan pada berita lain disampaikan bahwa karena jam kantor yang sama, penumpang harus berdesak-desakan untuk masuk ke kereta [6]. Disamping banyaknya persepsi positif yang muncul, pembuat kebijakan harus mampu menangani persepsi negatif yang berperan penting terhadap efektivitas sistem transportasi berkelanjutan [7].

Untuk memahami opini yang dimiliki oleh pengguna KRL Commuter Line terhadap KAI Commuter Line dapat dilakukan analisis sentimen. Analisis sentimen sudah banyak dilakukan pada penelitian terdahulu untuk memahami sentimen terhadap suatu topik sehingga hasilnya dapat dijadikan bahan evaluasi bagi pihak terkait. Pada penelitian ini akan dilakukan hal serupa, yakni menganalisis sentimen pengguna KRL terhadap KAI Commuter Line. Penelitian ini menggunakan Algoritma Naïve Bayes karena tingkat akurasinya yang cukup tinggi pada penelitian terdahulu. Pada penelitian tahun 2022 yang menganalisis opini publik terhadap PT PAL Indonesia, Algoritma Naïve Bayes mendapatkan nilai akurasi sebesar 84.08%, hasil ini lebih baik dibandingkan Algoritma K-NN dengan akurasi 83.38% dan Algoritma Decision Tree dengan akurasi 81.09% [8].

Penelitian ini bertujuan untuk menganalisis sentimen pengguna KRL terhadap KAI Commuter Line serta menganalisis tingkat akurasi Algoritma Naïve Bayes dalam mengklasifikasikan teks. Data yang dikumpulkan pada penelitian ini didapatkan melalui proses *crawling* sosial media X. Pada penelitian ini *pre-processing* berbasis *Natural Language Processing* (NLP) akan diterapkan. *Pre-processing* berbasis NLP terbukti penting untuk membantu model memahami teks [9]. Proses pelabelan pada penelitian ini akan dilakukan dengan dua langkah, yaitu pelabelan dengan *lexicon* serta pelabelan oleh ahli.

## 2. TINJAUAN PUSTAKA

### *2.1. KAI Commuter Line*

KAI Commuter Line merupakan anak perusahaan PT KAI yang melayani rute di wilayah Jakarta, Bogor, Depok, Tangerang, Bekasi (Jabodetabek) serta daerah lain seperti Yogyakarta-Solo. KAI Commuter Line memberikan kemudahan bagi penggunanya dalam mengatur jadwal bepergian dengan adanya aplikasi KAI Access, C-Access, dan *website* resmi untuk mengetahui jadwal keberangkatan kereta. Aplikasi C-Access memberikan pengaruh positif dan signifikan terhadap kepuasan pelanggan KAI Commuter line [10].

### *2.2. X*

Informasi yang mudah disebar, interaktivitas, dan partisipasi publik merupakan tiga dari alasan mengapa anak muda memilih sosial media [11]. Salah satu platform sosial media yang digunakan anak muda adalah X. X merupakan platform sosial media yang memungkinkan penggunanya untuk membagikan pemikiran, foto, serta video, sebelum tanggal 24 Juli 2023, X disebut dengan

Twitter. X sebagai platform sosial media sudah banyak digunakan dalam penelitian, salah satunya adalah analisis sentimen dengan tujuan untuk mengetahui sentimen pengguna X terkait suatu topik.

### *2.3. Crawling*

Teknik untuk mengumpulkan informasi dari berbagai sumber digital disebut dengan *crawling* [12]. *Crawling* merupakan proses mengunduh dan mengekstrak informasi yang dibutuhkan. *Crawling* memungkinkan pengumpulan data berjumlah besar dalam waktu yang singkat.

### *2.4. Analisis Sentimen*

Bidang untuk menganalisis pendapat atau sentimen terhadap suatu topik disebut dengan analisis sentimen [13]. Ulasan, opini sosial media, dan survei dapat dijadikan sebagai sumber data dalam analisis sentimen. Data yang dikumpulkan selanjutkan dikategorikan ke dalam kategori tertentu sesuai dengan kebutuhan analisis, seperti positif, negatif, dan netral. Hasil analisis sentimen dapat memberikan wawasan mengenai persepsi terhadap suatu topik, hasil dari analisis sentimen juga dapat dijadikan acuan dalam pengambilan keputusan berdasarkan kecenderungan sentimen yang didapatkan.

### *2.5. Algoritma Naïve Bayes*

Algoritma Naïve Bayes merupakan algoritma yang bekerja dengan menghitung probabilitas suatu teks untuk dikategorikan ke dalam kategori tertentu [14]. Algoritma Naïve Bayes memiliki karakteristik yaitu efisien dan akurasi yang tinggi, sehingga cocok untuk penelitian ini. Pada penelitian tahun 2025 mengenai sentimen pengguna sosial media X terhadap acara Clash Of Champions, Algoritma Naïve Bayes mendapatkan tingkat akurasi sebesar 97% pada perbandingan data 90:10 [14]. Berikut adalah persamaan Algoritma Naïve Bayes.

$$P(X) = \frac{P(X)P(H)}{P(X)} \quad (1)$$

Selain analisis teks, Algoritma Naïve Bayes juga mampu bekerja dalam kasus klasifikasi lain seperti sistem rekomendasi, klasifikasi gamabr, dan prediksi.

### *2.6. Term Frequency (TF)-Inverse Document Frequency (IDF)*

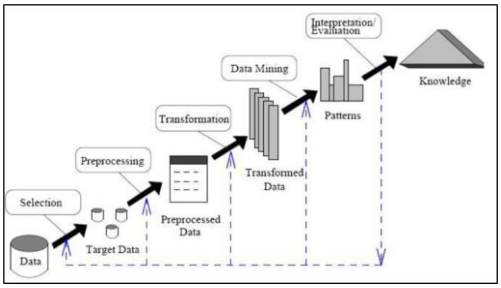
*Term Frequency* digunakan untuk mengukur tingkat seringnya suatu kata muncul dalam dokumen. Sedangkan *Inverse Document Frequency* (IDF) adalah teknik pembobotan token [15]. Penggabungan antara TF dan IDF dapat memberikan bobot yang lebih tinggi pada kata-kata yang relevan dan penting dalam proses analisis. TF-IDF membantu dalam mengukur seberapa penting suatu kata dalam dokumen terhadap keseluruhan dokumen yang ada.

### *2.7. Confusion Matrix*

*Confusion Matrix* merupakan *tools* yang diimplementasikan dalam *machine learning* untuk evaluasi [16]. *Confusion Matrix* memberikan informasi mengenai kinerja model dalam mengklasifikasikan data ke kelas yang berbeda [17]. *True Positive* merepresentasikan jumlah prediksi data positif yang benar. *True Negative* merepresentasikan jumlah prediksi data negatif yang benar. *False Positive* merepresentasikan prediksi positif yang salah. *False Negative* merepresentasikan prediksi negatif yang salah. Hasil dari *confusion matrix* selanjutnya digunakan pada tahapan evaluasi. Penelitian ini menggunakan tiga metrik untuk mengukur evaluasi Algoritma Naive Bayes dalam mengklasifikasikan teks. Metrik yang digunakan adalah *accuracy*, *precision*, dan *recall*. *Accuracy* mengukur seberapa baik model dalam memprediksi label yang benar. *Precision* untuk mengukur seberapa baik model dalam mengklasifikasikan label positif terhadap keseluruhan prediksi positif. *Recall* berfungsi untuk mengukur seberapa baik model dalam mengidentifikasi kelas positif yang benar.

### *2.8. Knowledge Discovery in Database*

*Knowledge Discovery in Database* adalah metodologi penelitian yang dapat membantu dalam menemukan pengetahuan dari data yang ada. Metodologi KDD membantu dalam penemuan pola atau ilmu dari data yang dipilih untuk diteliti [18]. KDD terdiri dari lima tahapan yang saling berkesinambungan, dimulai dari *data selection*, *pre-processing*, *transformation*, *data mining*, dan *evaluation*.




**Gambar 2.1** Metodologi Naïve Bayes (Sumber: Rizaldi, Alam, & Kurniawan, 2023)

### *2.9. Lexicon*

Metode *lexicon* merupakan pendekatan untuk menganalisis teks dengan mengidentifikasi dan memetakan kata-kata ke dalam kategori tertentu. Untuk memberikan label pada kata-kata tersebut diperlukan kamus yang berisi kata-kata yang mengungkapkan sentiment . Metode *lexicon* bekerja dengan penggunaan kamus yang berisi daftar kata, selanjutnya, kata-kata yang ada di dalam kamus dikategorikan ke dalam kategori yang dibutuhkan untuk analisis. Jika dalam suatu dokumen memiliki banyak frekuensi kata yang termasuk ke dalam suatu kategori, maka dokumen akan diklasifikasikan sebagai kategori tersebut. Metode ini sudah diterapkan pada banyak penelitian klasifikasi teks. Penelitian oleh Putra dan Wijaya menggunakan metode *lexicon* untuk mengklasifikasikan sentimen pengguna X terhadap *Artificial Intelligence* [20]. Pada penelitian tersebut, metode *lexicon* mengkategorikan 714 *tweet* ke dalam tiga sentimen, dan didapatkan 64% *tweet* dikategorikan sebagai positif, 25% *tweet* dikategorikan sebagai netral, dan 11% *tweet* dikategorikan sebagai sentimen positif.

### *2.10. Python*

Python merupakan bahasa pemrograman tingkat tinggi yang populer dalam bidang *data science* dan pemrosesan bahasa alami. Bahasa pemrograman ini memiliki *library* yang luas, kesederhanaan dan fleksibilitas menjadi alasan mengapa Python banyak digunakan. Untuk analisis sentimen, Python memberikan berbagai *tools* yang sangat berguna. Python juga mendukung berbagai format data, seperti CSV, JSON, dan SQL sehingga memudahkan integrasi dari berbagai sumber. Python merupakan bahasa pemrograman yang mudah bagi pemula karena sintaksnya yang sederhana, fleksibel, serta dukungan komunitas bahasa Python yang luas [21].

## 3. METODE PENELITIAN

Penelitian ini menerapkan metode penelitian Knowledge Discovery in Database (KDD). KDD adalah pendekatan yang dapat membantu dalam menemukan pola yang berharga atau informasi yang ada pada data yang diolah dalam analisis [18]. Metode penelitian KDD terdiri dari lima tahapan, yaitu *data selection, pre-processing, transformation, data mining, dan evaluation*.

### *3.1. Data Selection*

Pada tahap ini, dilakukan dengan crawling sosial media X dengan keyword ‘@CommuterLine’, ‘KRL’, dan ‘Commuter Line’. Periode pengumpulan data pada penelitian ini adalah 1 November 2024 sampai dengan 31 Januari 2025.

### *3.2. Pre-processing*

Tahap ini merupakan langkah awal dalam analisis sentimen yang bertujuan untuk menyaring dan menyeleksi informasi relevan dari dokumen. Proses ini dirancang untuk menghapus elemen yang tidak diperlukan agar data siap untuk dianalisis lebih lanjut. Pada penelitian ini, proses pre-processing mencakup *data cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan pelabelan.

### *3.3. Transformation*

Pada tahap ini, data opini yang telah melalui pre-processing diubah menjadi representasi numerik menggunakan metode TF-IDF. Setiap dokumen teks dikonversi menjadi vektor berdasarkan frekuensi kata.

### *3.4. Data Mining*

Tahapan data mining dilakukan dua proses, yaitu data splitting dan data mining. Pada proses data splitting data dibagi menjadi data latih dan data uji dengan lima skenario, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Untuk tahapan data mining, dilakukan modeling dengan Algoritma Naïve Bayes Classifier.

### *3.5. Evaluation*

Tahapan evaluasi dilakukan untuk mengetahui performa model dalam mengklasifikasikan label sentimen. Penelitian ini menggunakan *confusion matrix*. Hasil dari *confusion matrix* selanjutnya digunakan untuk menghitung *accuracy, precision,* dan *recall*.

## 4. HASIL DAN PEMBAHASAN

Berikut merupakan hasil dan pembahasan dari penelitian analisis sentimen pengguna KRL terhadap KAI Commuter Line. Hasil dan pembahasan didapatkan dengan menerapkan lima tahapan pada metodologi penelitian yang meliputi *data selection*, *pre-processing*, *transformation*, *data mining*, dan *evaluation*.

### *4.1. Data Selection*

Data yang dikumpulkan pada penelitian ini didapatkan melalui proses *crawling* sosial media X dengan menggunakan *tweet-harvest* sebagai *tools*. Data yang dikumpulkan merupakan cuitan yang relevan untuk penelitian. Data *tweet* berbahasa Indonesia dikumpulkan berdasarkan kata kunci '@CommuterLine', 'KRL', dan 'Commuter Line' dengan periode waktu 1 November 2024 sampai dengan 31 Januari 2025. Pemilihan periode waktu didasarkan pada prediksi BMKG terkait puncak musim hujan untuk Indonesia Bagian Barat [22]. Serta peningkatan penumpang pada liburan akhir tahun seperti yang disebutkan pada *website* resmi KAI Commuter Line [23]. Pada proses pengumpulan data berhasil didapatkan sebanyak 6.175 data bersih dengan 14 kolom yang disimpan dengan anam krl.csv.

### *4.2. Pre-processing*

Tahapan *pre-processing* pada penelitian ini terdiri dari beberapa proses, yaitu *cleaning*, *case folding*, *normalization*, *tokenizing*, *stemming*, dan pelabelan. Berikut adalah tiap proses dari tahapan *pre-processing*.

#### *4.2.1. Cleaning*

*Cleaning* merupakan proses untuk membersihkan dataset dari dokumen yang tidak relevan, tidak konsisten, dan tidak berguna untuk analisis lebih lanjut. Proses ini bertujuan untuk memastikan bahwa data yang digunakan sudah bersih sehingga dapat meningkatkan kualitas data. *Cleaning* pada penelitian ini dilakukan dengan tiga proses, yaitu menghapus dokumen dengan nama pengguna yang berasal dari akun resmi, menghapus kolom yang tidak dibutuhkan dalam analisis selanjutnya, dan menghapus elemen-elemen yang tidak diinginkan dalam analisis selanjutnya.

Hasil dari proses ini adalah dokumen yang berasal dari akun resmi yang bukan pengguna KRL sehingga total dokumen yang tersisa adalah sebanyak 909 dari yang sebelumnya 6.175. Penghapusan kolom yang tidak dibutuhkan pada analisis selanjutnya menyisakan satu kolom, yaitu kolom *full_text* yang berisi cuitan pengguna X. Penghapusan elemen yang tidak diinginkan seperti URL, *mention*, *hashtag*, angka, simbol dan spasi berlebih menghasilkan data yang lebih bersih. Gambar 4.1 merupakan hasil dari proses *cleaning*. Hasil disimpan pada kolom *text_cleaning*.

| | full_text | text_cleaning |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | Min yg bener aja kereta PSE paling pagi jam ke... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | min gap waktunya jauh banget menit line tpkjak... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | Gw pernah dong mau naik commuter line dari Pad... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | Syaiful Huda Soroti Penambahan Perjalanan Comm... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | Hiruk piruk commuter line |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | Ni line tpk ngapa jadi setengah jam sekali KOC... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | AH KRL LAMA BANGET DEH MANGGARAI LINE UDAH INI... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | Min tanggal februari commuter line jam dong mi... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | Ingat Mulai Besok Jumlah Perjalanan Commuter L... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | Para Pejabat ini pada tinggal di mana sih Mere... |

**Gambar 4.1** Hasil *Cleaning*

#### *4.2.2. Case Folding*

Pada proses *case folding* huruf kapital pada tiap dokumen diubah menjadi huruf kecil atau *lowercase*. *Casefolding* membantu dalam menyeragamkan data sehingga isinya konsisten. Penelitian ini menggunakan fungsi *.str.lower()* yang disediakan di Pandas pada kolom *text_cleaning*. Hasil dari proses ini disimpan pada kolom *text_folding* yang hasilnya dapat dilihat pada gambar 4.2. Pada gambar 4.2, dapat dilihat bahwa setiap kata dengan huruf kapital sudah diubah menjadi huruf kecil.

| | full_text | text_folding |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | min yg bener aja kereta pse paling pagi jam ke... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | min gap waktunya jauh banget menit line tpkjak... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | gw pernah dong mau naik commuter line dari pad... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | syaiful huda soroti penambahan perjalanan comm... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | hiruk piruk commuter line |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | ni line tpk ngapa jadi setengah jam sekali koc... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | ah krl lama banget deh manggarai line udah ini... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | min tanggal februari commuter line jam dong mi... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | ingat mulai besok jumlah perjalanan commuter l... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | para pejabat ini pada tinggal di mana sih mere... |

**Gambar 4.2** Hasil *Case Folding*

#### *4.2.3. Normalization*

Data yang didapatkan dari sosial media seringkali mengandung kata informal, kata gaul, bahkan kata yang memiliki interpretasi

kasar. Untuk menyamakan bentuk kata tidak baku, kata gaul dengan makna sama, maka dilakukan proses normalisasi. Proses normalisasi membantu meningkatkan konsistensi teks. Proses normalisasi pada penelitian ini dilakukan dengan pendekatan berbasis kamus, yaitu mencocokan setiap kata atau frasa yang tidak baku ke dengan padanan kata Bahasa Indonesia yang lebih formal. Kamus yang digunakan pada penelitian ini sudah divalidasi oleh ahli Bahasa Indonesia. Untuk frasa yang terdiri dari dua makna seperti kata "gajelas", digunakan metode pengenalan pola atau *regex* agar kata tidak terpecah menjadi kata yang tidak sesuai. Setelah frasa diubah, teks dipisahkan dengan fungsi *split()*. Kata yang ditemukan dalam kamus akan diganti, sedangkan jika tidak, maka akan tetap seperti asalnya. Proses ini diterapkan pada kolom *text_folding* dan hasilnya disimpan pada kolom *text_normalized*. Gambar 4.3 merupakan hasil dari proses normalisasi.

| | full_text | text_normalized |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | admin yang benar saja kereta stasiun pasar sen... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | admin jarak waktunya jauh sangat menit line ta... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | saya pernah ingin naik commuter line dari pa... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | syaiful huda soroti penambahan perjalanan comm... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | hiruk piruk commuter line |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | ini line stasiun tanjung priuk mengapa jadi se... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | ah krl lama sangat manggarai line sudah ini ... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | admin tanggal februari commuter line jam adm... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | ingat mulai besok jumlah perjalanan commuter l... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | para pejabat ini pada tinggal di mana mereka... |

**Gambar 4.3** Hasil *Normalization*

*4.2.4. Stopword Removal*

Proses *stopword removal* akan menghapus kata-kata umum yang tidak memiliki arti signifikan terhadap analisis. Tahapan ini menggunakan daftar kata *stopword* dari *library* Sastrawi. Selain itu, karena data yang digunakan berasal dari sosial media dengan gaya bahasa informal, penelitian ini menambahkan kamus tambahan untuk *stopword* yang sudah divalidasi oleh ahli Bahasa Indonesia. Proses ini diterapkan pada kolom *text_normalized* dan hasilnya disimpan pada kolom *stopword_removal*. Gambar 4.4 menampilkan hasil dari proses ini.

| | full_text | stopwords_removal |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | admin benar kereta stasiun pasar senen paling ... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | admin jarak waktunya jauh sangat menit line ta... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | naik commuter line padalarang bandung kira bel... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | syaiful huda soroti penambahan perjalanan comm... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | hiruk piruk commuter line |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | line stasiun tanjung priuk setengah jam sekali... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | krl lama sangat manggarai line tangerang line ... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | admin tanggal februari commuter line jam admin... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | ingat mulai besok jumlah perjalanan commuter l... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | pejabat tinggal kantornya senayan berarti naik... |

**Gambar 4.4** Hasil *Stopword Removal*

*4.2.5. Tokenizing*

Setelah data dibersihkan pada proses sebelumnya, data dipecah menjadi token pada proses *tokenizing*. Proses *tokenizing* diterapkan pada kolom *stopword_removal* dengan memanfaatkan fungsi *split()*. Hasil proses ini disimpan pada kolom *tokenized*. Gambar 4.5 merupakan hasil dari proses tokenisasi.

| | full_text | tokenized |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | [admin, benar, kereta, stasiun, pasar, senen, ... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | [admin, jarak, waktunya, jauh, sangat, menit, ... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | [naik, commuter, line, padalarang, bandung, ki... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | [syaiful, huda, soroti, penambahan, perjalanan... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | [hiruk, piruk, commuter, line] |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | [line, stasiun, tanjung, priuk, setengah, jam,... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | [krl, lama, sangat, manggarai, line, tangerang... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | [admin, tanggal, februari, commuter, line, jam... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | [ingat, mulai, besok, jumlah, perjalanan, comm... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | [pejabat, tinggal, kantornya, senayan, berarti... |

**Gambar 4.5** Hasil *Tokenizing*

*4.2.6. Stemming*

*Stemming* merupakan tahapan mengubah suatu teks ke bentuk dasarnya dengan tujuan menyamakan berbagai variasi kata yang ada untuk meningkatkan akurasi dalam analisis teks. Pada penelitian ini proses *stemming* dilakukan dengan memanfaatkan *library* Sastrawi dengan fungsi *stemming()* yang diterapkan pada kolom *tokenized* dan hasilnya disimpan pada kolom *stemmed*. Gambar 4.6 merupakan hasil dari proses *stemming*.

| | full_text | stemmed |
|---|---|---|
| 0 | Min yg bener aja kereta PSE paling pagi jam 6.... | [admin, benar, kereta, stasiun, pasar, senen, ... |
| 1 | @CommuterLine min gap waktunya jauh banget 30 ... | [admin, jarak, waktu, jauh, sangat, menit, lin... |
| 2 | Gw pernah dong mau naik commuter line dari Pad... | [naik, commuter, line, padalarang, bandung, ki... |
| 3 | Syaiful Huda Soroti Penambahan Perjalanan Comm... | [syaiful, huda, sorot, tambah, jalan, commuter... |
| 4 | Hiruk piruk commuter line https://t.co/WAwx7uK5zl | [hiruk, piruk, commuter, line] |
| 5 | @CommuterLine Ni line tpk ngapa jadi setengah ... | [line, stasiun, tanjung, priuk, tengah, jam, s... |
| 6 | AH KRL LAMA BANGET DEH? MANGGARAI LINE UDAH 2×... | [krl, lama, sangat, manggarai, line, tangerang... |
| 7 | Min tanggal 8-9 februari commuter line 24 jam ... | [admin, tanggal, februari, commuter, line, jam... |
| 8 | Ingat! Mulai Besok Jumlah Perjalanan Commuter ... | [ingat, mulai, besok, jumlah, jalan, commuter,... |
| 9 | @hrdbacot Para Pejabat ini pada tinggal di man... | [jabat, tinggal, kantor, senayan, arti, naik, ... |

**Gambar 4.6** Hasil *Stemming*

*4.2.7. Pelabelan*

Proses pelabelan pada penelitian ini diawali dengan pelabelan dengan *lexicon* berbahasa Inggris menggunakan *opinion_lexicon* yang terdapat di *library* nltk.corpus. Untuk kamus berbahasa Indonesia disusun secara manual berdasarkan kata-kata sentimen yang ditemukan pada dataset. Kamus berbahasa Indonesia sudah divalidasi oleh ahli. Gabungan dari kedua kamus diterapkan pada kolom *stemmed* dan hasilnya disimpan pada kolom *sentiment*.

Data yang sudah diberi label dengan metode *lexicon* selanjutnya divalidasi oleh ahli Bahasa Indonesia sehingga dapat memberikan hasil evaluasi yang lebih akurat terhadap data

yang sudah diberi label. Pelabelan hasil validasi ahli disimpan dalam file valid_sent.csv. Terdapat perbedaan pada hasil pelabelan oleh *lexicon* dengan pelabelan oleh ahli. Penelitian ini menggunakan hasil pelabelan oleh *lexicon* yang selanjutnya divalidasi oleh ahli untuk tahapan selanjutnya. Tabel 4.1 menampilkan hasil pelabelan oleh *lexicon* serta pelabelan oleh ahli.

**Tabel 4.1** Hasil *Stopword Removal*

| Langkah Pelabelan | Label Positif | Label Negatif |
|---|---|---|
| ***Lexicon*** | 630 | 279 |
| **Ahli** | 485 | 424 |

### *4.3. Transformation*

Tahapan transformation atau Transformasi pada penelitian ini dilaksanakan dengan TF-IDF. Transformasi bertujuan untuk mengubah data yang sudah di stemming menjadi representasi numerik agar dapat diproses pada tahapan selanjutnya. Penelitian ini menggunakan TF-IDF (*Term Frequency – Inverse Document Frequency*) untuk proses transformasi. TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul di suatu dokumen tetapi jarang muncul di dokumen lain. Term Frequency menghitung kemunculan kata dalam dokumen sedangkan *Inverse Document Frequency* mengukur seberapa jarang suatu kata muncul, semakin jarang maka nilai IDF-nya semakin tinggi. Gambar 4.7 menampilkan kode transformasi dengan TF-IDF.

```
# Inisialisasi TF-IDF Vectorizer untuk masing-masing kolom
tfidf_data = TfidfVectorizer()

# Lakukan fit_transform untuk masing-masing kolom
tfidf_data_matrix = tfidf_data.fit_transform(data['stemmed'].apply(lambda x: ' '.join(x)))

# Konversi hasil transformasi ke DataFrame terpisah
data_tfidf_data = pd.DataFrame(tfidf_data_matrix.toarray(), columns=tfidf_data.get_feature_names_out())

print("TF-IDF dataframe :\n", data_tfidf_data.head())
```

**Gambar 4.7** Kode Transformasi TF-IDF

Setelah proses transformasi dengan TF-IDF, kata-kata yang ada diubah ke dalam bentuk numerik. Hasil dari transformasi dapat dilihat pada Gambar 4.8 di bawah ini.

```
TF-IDF dataframe :
   aamiin  abang  abar  abu  acara  acau  acces    access  acsses  acu  ... \
0     0.0    0.0   0.0  0.0    0.0   0.0    0.0  0.000000     0.0  0.0  ...
1     0.0    0.0   0.0  0.0    0.0   0.0    0.0  0.000000     0.0  0.0  ...
2     0.0    0.0   0.0  0.0    0.0   0.0    0.0  0.246438     0.0  0.0  ...
3     0.0    0.0   0.0  0.0    0.0   0.0    0.0  0.000000     0.0  0.0  ...
4     0.0    0.0   0.0  0.0    0.0   0.0    0.0  0.000000     0.0  0.0  ...

   young  youngk  your  yuji  yup  yuuu  zalim  zayedtirtonadisky  zolim \
0    0.0     0.0   0.0   0.0  0.0   0.0    0.0                0.0    0.0
1    0.0     0.0   0.0   0.0  0.0   0.0    0.0                0.0    0.0
2    0.0     0.0   0.0   0.0  0.0   0.0    0.0                0.0    0.0
3    0.0     0.0   0.0   0.0  0.0   0.0    0.0                0.0    0.0
4    0.0     0.0   0.0   0.0  0.0   0.0    0.0                0.0    0.0

   zombie
0     0.0
1     0.0
2     0.0
3     0.0
4     0.0
```

**Gambar 4.8** Hasil Transformasi TF-IDF

Pada Gambar 4.8 dapat dilihat hasil dari proses transformasi yang berhasil dijalankan. Pada indeks 3 kata ‘access’ mendapatkan nilai tinggi, yang artinya kata itu sering muncul pada dokumen tersebut namun tidak pada dokumen lainnya. Untuk kata-kata yang bernilai 0.0 seperti kata ‘zombie’ dan ‘yuji’ maknanya kata-kata tersebut tidak muncul di dokumen yang ditampilkan namun muncul di dokumen lain.

### *4.4. Data Mining*

Pada tahap data mining, dilakukan proses pembagian data menjadi lima skenario serta pembangunan model klasifikasi untuk menemukan pola dari data opini yang telah melalui tahap transformasi. Berikut adalah pembahasan pada tahapan data mining.

#### *4.4.1. Data Splitting*

Setelah data dikonversi ke dalam bentuk numerik pada tahapan transformasi, tahapan selanjutnya adalah Membagi data atau data splitting. Proses ini bertujuan untuk membagi data menjadi data latih (*train*) dan data uji (*test*). Data latih digunakan untuk membangun model, Sedangkan data uji untuk mengukur hasil pelatihan terhadap model. Pembagian data dilakukan dengan menggunakan fungsi *train_test_split()* dari library scikit-learn. Penelitian ini menguji lima skenario, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Parameter *test_size* menentukan jumlah proporsi data yang digunakan sebagai data uji, Sedangkan *random_state* digunakan untuk memastikan hasil pembagian data tetap konsisten setiap kali dijalankan, pada penelitian ini *random_state* yang digunakan adalah 64. Gambar 4.9 menampilkan kode untuk *data splitting*.

```
from sklearn.model_selection import train_test_split

# ===== SPLIT 50:50 =====
X_train_50, X_test_50, y_train_50, y_test_50 = train_test_split(
    data_tfidf_data, data['sentiment'], test_size=0.5, random_state=64)

# ===== SPLIT 60:40 =====
X_train_60, X_test_40, y_train_60, y_test_40 = train_test_split(
    data_tfidf_data, data['sentiment'], test_size=0.4, random_state=64)

# ===== SPLIT 70:30 =====
X_train_70, X_test_30, y_train_70, y_test_30 = train_test_split(
    data_tfidf_data, data['sentiment'], test_size=0.3, random_state=64)

# ===== SPLIT 80:20 =====
X_train_80, X_test_20, y_train_80, y_test_20 = train_test_split(
    data_tfidf_data, data['sentiment'], test_size=0.2, random_state=64)

# ===== SPLIT 90:10 =====
X_train_90, X_test_10, y_train_90, y_test_10 = train_test_split(
    data_tfidf_data, data['sentiment'], test_size=0.1, random_state=64)
```

**Gambar 4.9** Kode Data Splitting

#### *4.4.2. Modelling*

Tujuan dari tahap ini adalah untuk membangun model yang mampu mengklasifikasikan sentimen secara otomatis

ke dalam kategori positif dan negatif, berdasarkan data pelatihan yang tersedia. Teknik yang digunakan adalah klasifikasi dengan Algoritma Naive Bayes, tepatnya jenis Multinomial Naive Bayes, yang umum digunakan untuk klasifikasi teks atau analisis sentimen. Model dilatih menggunakan dataset yang telah dibagi ke dalam beberapa skenario rasio pembagian data latih dan data uji, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Gambar 4.10 menampilkan kode pelatihan model.

```
# ==================== 50:50 ====================
model_sentiment_50 = MultinomialNB()
model_sentiment_50.fit(X_train_50, y_train_50)

# ==================== 60:40 ====================
model_sentiment_60 = MultinomialNB()
model_sentiment_60.fit(X_train_60, y_train_60)

# ==================== 70:30 ====================
model_sentiment_70 = MultinomialNB()
model_sentiment_70.fit(X_train_70, y_train_70)

# ==================== 80:20 ====================
model_sentiment_80 = MultinomialNB()
model_sentiment_80.fit(X_train_80, y_train_80)

# ==================== 90:10 ====================
model_sentiment_90 = MultinomialNB()
model_sentiment_90.fit(X_train_90, y_train_90)
```

**Gambar 4.10** Kode Modelling

Untuk analisis lebih dalam digunakan *confusion matrix* pada data uji di tahapan evaluasi.

### *4.5. Evaluation*

Model yang sudah dilatih dengan lima skenario pembagian data selanjutnya dievaluasi performanya. Evaluasi dengan *confusion matrix* berguna untuk mengevaluasi seberapa baik model membandingkan label sentimen yang diprediksi dengan label yang sebenarnya. *Confusion matrix* ditampilkan dengan menggunakan library seaborn dan scikit-learn.

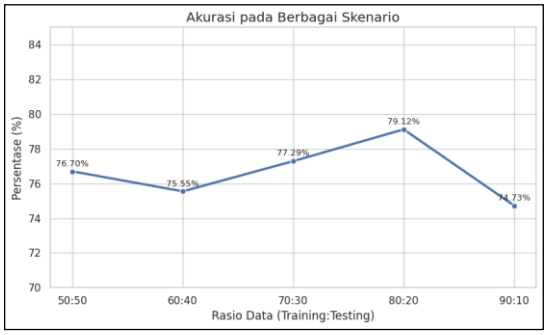
```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

def plot_conf_matrix(y_true, y_pred, title):
    cm = confusion_matrix(y_true, y_pred)
    plt.figure(figsize=(6, 5))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
                xticklabels=["Positif", "Negatif"],
                yticklabels=["Positif", "Negatif"])
    plt.xlabel('Prediksi')
    plt.ylabel('Aktual')
    plt.title(f'Confusion Matrix - {title}')
    plt.tight_layout()
    plt.show()
```

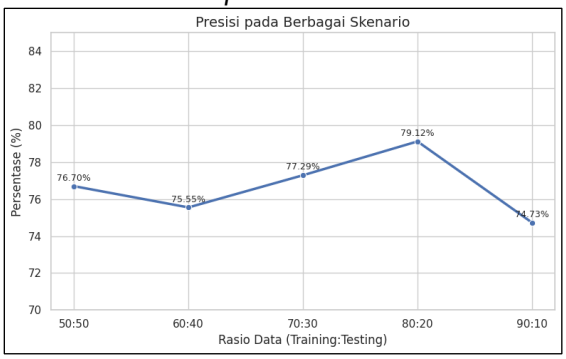
**Gambar 4.11** Kode Confusion Matrix

Hasil dari *confusion matrix* selanjutnya dihitung dalam evaluasi. Evaluasi pada penelitian ini menggunakan tiga metrik umum, yaitu akurasi (*accuracy*), presisi (*precision*), dan recall. Akurasi mengukur proporsi prediksi yang benar dibandingkan dengan total jumlah data uji. Presisi mengukur ketepatan model dalam memprediksi satu kelas tertentu. Jika hasil presisi sentimen positif tinggi maknanya prediksi sentimen itu memang benar-benar positif, begitu juga dengan sebaliknya. Penelitian ini menggunakan parameter *average='weighted'* untuk menghitung presisi rata-rata dari seluruh kelas. *Recall* mengukur seberapa baik model dalam menemukan semua data positif yang sebenarnya. Nilai *recall* yang tinggi menunjukan bahwa model tidak melewatkan terlalu banyak data yang seharusnya diklasifikasikan sebagai sentimen positif.

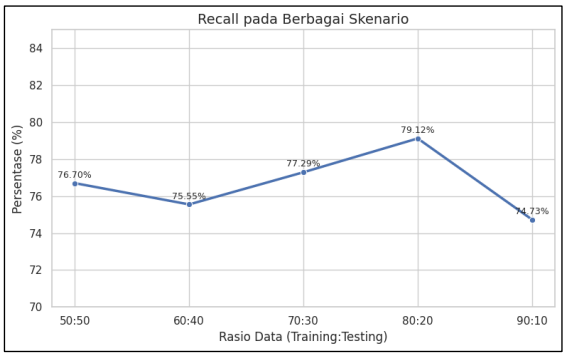
Berikut adalah hasil dari confusion matrix dari lima skenario pembagian data. Dalam Gambar 4.12, 4.13, dan 4.14 ditampilkan nilai akurasi dari tiap skenario. Hasil evaluasi tiap skenario ditampilkan dengan grafik garis untuk memudahkan memahami perbedaan nilai akurasi pada setiap skenario. Hasil evaluasi untuk setiap metrik evaluasi pada setiap skenario pembagian data konsisten dan stabil.




**Gambar 4.12** Hasil Evaluasi Akurasi dengan *Confusion Matrix*




**Gambar 4.13** Hasil Evaluasi Presisi dengan *Confusion Matrix*




**Gambar 4.14** Hasil Evaluasi *Recall* dengan *Confusion Matrix*

Pada Gambar 4.12, 4.13, dan 4.14 ditampilkan hasil akurasi, presisi, dan *recall* pada tiap skenario pembagian data. Skenario 50:50 mendapatkan hasil *accuracy*, *precision*, dan *recall* di angka 76,70%. Pada skenario 60:40 hasil evaluasi untuk accuracy, precision, dan recall adalah 75,55%. Pada skenario 70:30, nilai *accuracy*, *precision*, dan *recall* yang didapatkan adalah 77,29%. Untuk skenario 80:20 didapatkan hasil *accuracy*, *precision*, dan *recall* sebesar 79,12%. Dan untuk skenario 90:10, nilai *accuracy*, *precision*, dan *recall* adalah 74,73%. Pada gambar yang disajikan dapat disimpulkan bahwa skenario 80:20 memiliki nilai tertinggi dengan nilai akurasi, presisi, dan *recall* sebesar 79,12%. Sedangkan untuk nilai evaluasi akurasi, presisi, dan *recall* terendah ditemukan pada pembagian data 90:10 dengan nilai 74,73%. Nilai akurasi, presisi, dan *recall* cenderung stabil dengan rentang 75,55% - 77,70% pada skenario pembagian data 50:50 sampai dengan 70:30.

Penelitian dengan judul "Analisis Sentimen Pengguna KRL Terhadap KAI Commuter Line melalui Sosial Media X dengan Algoritma Naïve Bayes Classifier" bertujuan untuk menganalisis opini atau sentimen pengguna KRL pada sosial media X serta mengukur tingkat akurasi Algoritma Naïve Bayes. Penelitian dengan metode KDD ini menganalisis sentimen yang didapatkan dari hasil *crawling* sosial media X dengan memanfaatkan t*ool harvest-tweet* dengan kata kunci @CommuterLine", "KRL", dan "Commuter Line". Data yang sudah dikumpulkan selanjutnya di proses pada tahapan *pre-processing* yang terdiri dari *cleaning*, *case folding*, *normalization*, *stopword removal*, *stemming*, serta pelabelan data. Pada proses pelabelan data dengan *lexicon* didapatkan sebanyak 630 data berlabel positif dan 279 data berlabel negatif. Selanjutnya label divalidasi oleh ahli dan didapatkan bahwa total data dengan label positif adalah sebanyak 485 atau 53,36% dari total data dan sebanyak 424 data atau 46,64% diberi label negatif. Data yang sudah bersih dan sudah diberi label selanjutnya ditransformasi dengan TF-IDF untuk mengkonversi data teks ke bentuk vektor numerik sehingga mudah untuk dianalisis oleh mesin. Selanjutnya, dilakukan *data splitting* untuk membagi data latih (*train*) dan data uji (*test*) menjadi lima skenario, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Data yang sudah dibagi selanjutnya di proses pada tahapan *modelling* dengan Algoritma Naïve Bayes Classifier. Hasil evaluasi menunjukan bahwa skenario pembagian data 80:20 memberikan hasil yang paling optimal dengan tingkat akurasi sebesar 79,12%, presisi sebesar 79,12%, dan *recall* sebesar 79,12%. Dengan tingkat akurasi sebesar 79,12%, menunjukan bahwa dari seluruh data uji, 79,12% prediksi yang dilakukan oleh model sudah sesuai dengan labelnya, menunjukan performa yang cukup baik.

Keberhasilan model dalam mengklasifikasikan cuitan pengguna KRL tidak luput dari tahapan *pre-processing* berbasis NLP yang mampu meningkatkan performa model. Jumlah sentimen positif yang dominan merepresentasikan bahwa pengguna X umumnya puas dengan layanan KRL. Namun, munculnya 424 cuitan dengan label negatif menunjukan bahwa ada persepsi negatif dari pengguna seperti masalah keterlambatan pada cuitan berbunyi "Dear @CommuterLine kenapa hari ini jadwal KRL line Cikarang or Bekasi ke Kp Bandan or Angke Via Manggarai terlambatnya g karuan sih harusnya ada yg jam 5 baru ada jam 5.20. 20 menit saya kebuang weeeh cuman minta maaf doank g kasih penjelasan apa² ih." serta masalah padatnya kereta pada cuitan yang berbunyi "@CommuterLine hopefully in the future bisa tambah frekuensi kereta line rangkasbitung - tanahabang yaa agak jomplang line bogor - manggarai 5 menit sekali meanwhile rangkasbitung line 10 menit sekali. keretanya jadi penuh bgtt sampe sesek napas gencetnya jam 6-9 pagi real".

## 5. KESIMPULAN

Dataset yang digunakan dalam penelitian ini diperoleh dengan proses *crawling* menggunakan *tool harvest-tweet*. Data yang diambil adalah cuitan berbahasa Indonesia yang mengandung kata kunci '@CommuterLine', 'KRL', dan 'Commuter Line' dengan periode waktu 1 November 2024 sampai 31 Januari 2025. Total data yang dikumpulkan adalah 6175, dengan data bersih yang digunakan dalam analisis sebanyak 909 data. Setelah data diberi pelabelan dengan pendekatan *lexicon* serta validasi ahli, dari 909 data, 485 diberi label positif dan 424 data diberi label negatif. Persentase dari data dengan sentimen positif adalah 53,36% dan data dengan sentimen negatif sebanyak 46,64%. Tingkat akurasi yang didapatkan oleh Algoritma Naïve Bayes adalah 79,12%. Berdasarkan hasil penelitian sentimen yang dilakukan, mayoritas pengguna KRL memiliki sentimen cenderung positif, namun tidak sedikit juga sentimen negatif terhadap KAI Commuter Line. Sentimen negatif seperti keterlambatan dan sesaknya kereta dapat dijadikan landasan untuk pihak terkait melakukan evaluasi. Rekomendasi evaluasi yang dapat dilakukan adalah dengan mengatur jadwal lebih baik untuk mengatasi keterlambatan kereta, juga dapat dilakukan penambahan armada kereta untuk mengurangi kepadatan penumpang.

## UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih kepada semua pihak yang telah memberikan dukungan, semangat, serta doa yang terus mengalir selama proses penyelesaian penelitian ini. Diharapkan penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan menjadi acuan bagi penelitian-penelitian berikutnya.

## DAFTAR PUSTAKA

[1] A. W. Utomo dan U. Umar, "Analisis Efektivitas Kapasitas Daya Gardu Traksi Terhadap Kebutuhan KRL Jalur Yogyakarta - Solo," *Emit. J. Tek. Elektro*, vol. 22, no. 1, hal. 1–7, 2022, doi: 10.23917/emitor.v22i1.14861.

[2] K. A. I. Commuter, "Catat Rekor Volume Tertinggi, KAI Commuter Layani 1,13 Juta Pengguna pada Awal Minggu Juli 2024, KAI Commuter Layani 156.816.151 Orang pada Semester 1 Tahun 2024." [Daring]. Tersedia pada: https://commuterline.id/informasi-publik/berita/catat-rekor-volume-tertinggi-kai-commuter-layani-1-13-juta-pengguna-pada-awal-minggu-juli-2024-ini-kai-commuter-layani-156-816-151-orang-pada-semester-1-tahun-2024

[3] A. Sukmawijaya, G. Rahmatika, dan K. Bisnis, Ed., "Jumlah Penumpang KRL Diprediksi Membeludak, Tembus 410 Juta Orang hingga 2027." [Daring]. Tersedia pada: https://kumparan.com/kumparanbisnis/jumlah-penumpang-krl-diprediksi-membeludak-tembus-410-juta-orang-hingga-2027-20mE7Vb3sGy

[4] S. Sahara dan B. N. A. Nugroho, "Efektivitas Penggunaan Kereta Listrik (KRL) Commuter Line Jabodetabek Untuk Mengurangi Kemacetan Di DKI Jakarta," *Ekon. J. Ilm. Manajemen, Ekon. Bisnis, Kewirausahaan*, vol. 10, no. 2, hal. 415–426, 2023, doi: 10.30640/ekonomika45.v11i1.1926.

[5] Kompas.com, "Makin Nyaman, Kereta Commuter Line Masih Jadi Pilihan Masyarakat Selama Masa Angkutan Lebaran 2025." [Daring]. Tersedia pada: https://nasional.kompas.com/read/2025/04/06/14475891/makin-nyaman-kereta-commuter-line-masih-jadi-pilihan-masyarakat-selama-masa?page=all

[6] Jakarta.nu, "Kepadatan Penumpang KRL di Jam Sibuk Masih Menjadi Masalah," 2024. [Daring]. Tersedia pada: https://jakarta.nu.or.id/jakarta-raya/kepadatan-penumpang-krl-di-jam-sibuk-masih-menjadi-masalah-vPUPC

[7] I. Hidayati, "Beyond traffic jams: public perceptions of Jabodetabek Commuter Line (KRL) System for sustainable urban development," in *IOP Conference Series: Earth and Environmental Science*, Jakarta: Institute of Physics, 2023. doi: 10.1088/1755-1315/1263/1/012027.

[8] F. S. Pattiiha dan H. Hendry, "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, hal. 506, 2022, doi: 10.30865/jurikom.v9i2.4016.

[9] D. R. Wijaya, G. M. A. Sasmita, dan W. O. Vihikan, "Sentiment Analysis of Indonesian Citizens on Electric Vehicle Using FastText and BERT Method," vol. 6, no. 3, hal. 1360–1372, 2024, doi: 10.51519/journalisi.v6i3.784.

[10] D. Azzahra dan B. Priyono, "Informasi KRL Access dan Kualitas Pelayanan Terhadap Kepuasan Pelanggan KRL Commuter Line di Jabodetabek," *J. Bus. Adm. Econ. Entrep.*, vol. 5, no. 2, hal. 95–106, 2023, [Daring]. Tersedia

pada: https://jurnal.stialan.ac.id/index.php/jbest/article/view/709

[11] T. Andriyani, "Medsos Jadi Sarana Penyampaian Pesan dan Kritik Sosial Kalangan Anak Muda." [Daring]. Tersedia pada: https://ugm.ac.id/id/berita/medis-sosial-jadi-sarana-penyampaian-pesan-dan-kritik-sosial-kalangan-anak-muda/

[12] M. Riswan, A. Primajaya, dan A. Y. S. Irawan, "ANALISIS SENTIMEN TERHADAP PEMBERITAAN HASIL REKAPITULASI PEMILU PRESIDEN 2024 PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN NAIVE," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, hal. 203–2011, 2025, [Daring]. Tersedia pada: https://journal.eng.unila.ac.id/index.php/jitet/article/view/5559

[13] J. Supriyanto, D. Alita, dan A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, hal. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.

[14] D. E. Ratnawati, N. Y. Setiawan, P. Studi, S. Informasi, F. I. Komputer, dan U. Brawijaya, "ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TWITTERX TERHADAP ACARA CLASH OF CHAMPIONS MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES," vol. 9, no. 3, hal. 1–10, 2025.

[15] O. I. Gifari, M. Adha, F. Freddy, dan F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, hal. 36–40, 2022.

[16] A. Ridhovan dan A. Suharso, "Penerapan Metode Residual Network (Resnet) Dalam Klasifikasi Penyakit Pada Daun Gandum," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 1, hal. 58–65, 2022, doi: 10.29100/jipi.v7i1.2410.

[17] D. Bona dan B. N. Sari, "Implementasi Jaringan Hierarki Attention Untuk Klasifikasi Basis Data Multimodal Biometrik," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 3, hal. 632–638, 2022, doi: 10.29100/jipi.v7i3.2879.

[18] Khoirunnisa Hamidah dan A. Voutama, "Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier.," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 15, no. 1, hal. 1–7, 2023, doi: 10.35891/explorit.v15i1.3874.

[19] S. A. Rizaldi, S. Alam, dan I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, hal. 109–117, 2023, doi: 10.55123/storage.v2i3.2334.

[20] S. A. Putra dan A. Wijaya, "Analisis Sentimen Artificial Intelligence (Ai) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 7, no. 1, hal. 21–28, 2023, doi: 10.32524/jusitik.v7i1.1042.

[21] M. H. Maulana, "Python Bahasa Pemrograman Yang Ramah Bagi Pemula," *JISCO (Journal Inf. Syst. Comput.*, vol. 2, no. 2, hal. 73–78, 2024, [Daring]. Tersedia pada: https://jurnal.fst.uinjambi.ac.id/index.php/jisco/index

[22] BMKG, *Prediksi Musim Hujan 2024/2025*. Jakarta: BMKG, 2024. [Daring]. Tersedia pada: https://www.bmkg.go.id/iklim/prediksi-musim/prakiraan-musim-hujan-2024-2025-di-indonesia

[23] K. A. I. Commuter, "KAI Commuter Catat Rapor Positif, Layani 1,2 Juta Pengguna Commuter Line Jelang Tahun Baru 2025." [Daring]. Tersedia pada: https://www.commuterline.id/id/informasi-publik/berita/kai-commuter-catat-rapor-positif-layani-1-2-juta-pengguna-commuter-line-jelang-tahun-baru-2025