

Analisis Sentimen Pengguna Terhadap Kinerja Sistem Transportasi Umum Jakarta Menggunakan Algoritma Naive Bayes

Muhammad Imam Santoso, Ahmad Rizal Dzibrillah*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹muhammadimamsantoso188@gmail.com, ^{2,*}ahmadrizaldzibrillah@gmail.com

Email Penulis Korespondensi: ahmadrizaldzibrillah@gmail.com

Abstrak—Penelitian ini menggunakan algoritma Naive Bayes untuk menganalisis sentimen netizen terkait Transportasi Umum di Jakarta. Pada masa lalu, sistem transportasi umum di Jakarta didominasi oleh operator swasta atau perorangan, termasuk bus, angkot, dan taksi. Namun, terjadi berbagai tantangan seperti kurangnya koordinasi, kualitas layanan yang tidak konsisten, masalah keamanan, dan cakupan wilayah yang tidak memadai. Untuk meningkatkan kualitas dan ketersediaan transportasi umum, pemerintah daerah atau nasional telah melakukan intervensi, seperti mengambil alih layanan transportasi umum atau memberlakukan regulasi ketat pada operator swasta. Investasi besar telah dilakukan dalam mengembangkan moda transportasi umum seperti TransJakarta, KRL Commuter Line, MRT, Jaklingko, dan layanan angkutan umum lainnya. Penelitian ini bertujuan untuk menganalisis manfaat dari Transportasi Umum yang sudah banyak diambil alih oleh pemerintah supaya dapat meminimalisir kekurangan yang ada. Penelitian ini berfokus pada analisis perbedaan pendapat yang tersebar di berbagai platform media sosial. Data yang di dapatkan dari media social seperti youtube dan X. Terdapat data yang nantinya akan digunakan pada penelitian ini dengan jumlah 987 data. Kemudian, terdapat 612 data positif dan 375 data negatif. Setelah dilakukan riset, maka akan diperoleh hasil dengan akurasi 94.22%. Kemudian dari analisis ini terdapat variasi sentiment yang cukup signifikan di antara netizen terkait Transportasi Umum di Jakarta. Beberapa kelompok netizen sudah bisa merasakan hasil yang positif dari terintegrasinya Transportasi Umum saat ini. Namun juga masih ada kekurangan dalam eksekusinya. Analisis juga mengidentifikasi faktor-faktor kunci yang mempengaruhi perbedaan pendapat, seperti daerah pengguna, tidak meratanya supir dengan pribadi yang bagus, dan juga ekonomi masyarakat. Dari akan itu, dalam penelitian ini berkontribusi dalam bidang analisis sentiment dan pengolahan Bahasa alami yang menerapkan prosedur pemecahan sehingga dapat di klasifikasikan pada data komentar netizen di berbagai platform. Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk mengklasifikasikan sentimen komentar netizen Transportasi Umum di daerah Jakarta ini dengan tingkat akurasi yang cukup tinggi. Hasil ini dapat berguna pemerintah dan warga Jakarta juga dalam mencari solusi untuk masalah ini. Dengan demikian, penelitian ini dapat menjadi dasar bagi pemahaman yang lebih menyeluruh terkait respons pemerintah terhadap isu Transportasi Umum di Jakarta

Kata Kunci: Analisis Sentimen; Algoritma Naive Bayes; Transportasi Umum; Transjakarta; Jaklingko; Rapid Miner

Abstract—This study uses the Naive Bayes algorithm to analyze netizen sentiment regarding public transportation in Jakarta. In the past, Jakarta's public transportation system was dominated by private operators, including buses, angkot (minibuses), and taxis. However, various challenges arose, such as lack of coordination, inconsistent service quality, safety issues, and inadequate coverage. To improve the quality and availability of public transportation, local or national governments have intervened by taking over public transportation services or imposing stricter regulations on private operators. Significant investments have been made in developing public transportation modes such as TransJakarta (bus rapid transit), KRL Commuter Line (electric train), MRT (Mass Rapid Transit), Jaklingko, and other public transport services. This study aims to analyze the benefits of public transportation, which has largely been taken over by the government, to minimize existing shortcomings. The research focuses on analyzing the differing opinions spread across various social media platforms. Data was collected from social media platforms such as YouTube and X. A total of 987 data points were used in this study, with 612 positive and 375 negative data points. After conducting the research, an accuracy of 94.22% was achieved. The analysis revealed significant variations in sentiment among netizens regarding public transportation in Jakarta. Some groups of netizens have begun to feel positive effects from the current integration of public transportation, but there are still execution shortcomings. The analysis also identified key factors influencing differing opinions, such as user areas, the uneven distribution of drivers with good personal qualities, and the economic conditions of the community. Consequently, this study contributes to sentiment analysis and natural language processing by applying problem-solving procedures to classify netizen comments on various platforms. The results of this study indicate that the Naive Bayes algorithm can be used to classify netizen sentiment about public transportation in Jakarta with a high level of accuracy. These findings can be useful for the government and Jakarta residents in finding solutions to these issues. Thus, this study can serve as a basis for a more comprehensive understanding of the government's response to public transportation issues in Jakarta.

Keywords: Sentiment Analysis; Naive Bayes Algorithm; Public Transportation; Transjakarta; Jaklingko Fast Miner

1. PENDAHULUAN

Kota Jakarta, sebagai ibu kota Indonesia menghadapi dinamika kompleks dalam mengelola sistem transportasi umum yang efisien dan memadai untuk memenuhi kebutuhan mobilitas penduduk yang terus meningkat. Dalam beberapa dekade terakhir, Jakarta telah menyaksikan pertumbuhan pesat dalam populasi dan urbanisasi, yang telah menyebabkan peningkatan lalu lintas, kemacetan, dan masalah lingkungan seperti polusi udara [1]. Meningkatnya beban transportasi di Jakarta memberikan tantangan yang signifikan bagi otoritas kota dan pengelola transportasi umum untuk menciptakan solusi yang efektif dan berkelanjutan.

Pada masa lalu, sebagian besar sistem transportasi umum di wilayah Jabodetabek mungkin didominasi oleh operator swasta atau perorangan. Kondisi ini mungkin mencakup berbagai jenis transportasi, seperti bus, angkutan kota (angkot), dan layanan taksi. Meskipun ada operator swasta yang menyediakan layanan transportasi, sering kali terjadi berbagai tantangan, seperti kurangnya koordinasi antara operator, kualitas layanan yang tidak konsisten, masalah keamanan, dan ketidakmampuan untuk menyediakan cakupan yang memadai di seluruh wilayah.

Kurangnya koordinasi yang dimaksud seperti banyaknya operator yang merebutkan sewa. Mengupayakan segala cara supaya hanya operator dia sendiri yang untung tanpa memikirkan keselamatan penumpang. Supir yang mengendarai secara ugal-ugalan, menurunkan atau menaikkan penumpang tidak ditempatnya itu menurunkan tingkat keselamatan penumpang. Jam operasional yang sesuka hati operator membuat penumpang tidak tahu kapan beroperasi transportasi tersebut. Masih banyaknya copet atau tindak kejahatan lainnya juga di dalam moda transportasi itu.

Dalam upaya untuk mengatasi permasalahan transportasi tersebut, Pemerintah Provinsi DKI Jakarta telah mengambil langkah-langkah progresif untuk meningkatkan sistem transportasi publiknya. Pemerintahan Provinsi DKI banyak mencabut izin trayek transportasi umum yang sudah tidak layak dalam segi apapun. Peremajaan kendaraan juga menjadi syarat untuk trayek itu tetap bisa jalan. Tetapi dengan sudah banyaknya daya saing dengan transportasi online membuat operator swasta banyak yang tidak sanggup dalam hal peremajaan kendaraan. Hal itulah yang membuat pemerintahan banyak mencabut izin trayek moda transportasi umum di DKI Jakarta.

Dalam rangka meningkatkan kualitas dan ketersediaan transportasi umum bagi penduduk, pemerintah daerah atau nasional mungkin telah melakukan intervensi. Hal ini dapat mencakup pengambilalihan sebagian atau seluruh layanan transportasi umum, baik melalui kepemilikan langsung maupun melalui regulasi yang lebih ketat terhadap operator swasta. Investasi besar telah dilakukan dalam pengembangan berbagai moda transportasi umum, termasuk TransJakarta (bus rapid transit), KRL Commuter Line (kereta rel listrik), MRT (Mass Rapid Transit), Jaklingko dan penyediaan layanan angkutan umum lainnya.

Transportasi umum terintegrasi menjadi tujuan utama pemerintah saat ini, dimana nanti semua transportasi umum di Jakarta bisa saling berenergi dan terhubung. Jaklingko akan menjadi transportasi umum yang paling sibuk karena akan disebar ke seluruh wilayah Jakarta untuk membantu masyarakat menuju tempat pemberhentian transportasi umum besar yang ada di Jakarta seperti Transjakarta, KRL, MRT dan LRT[2]. Transportasi umum besar ini yang nantinya akan mengantarkan masyarakat ke tujuan utama mereka yaitu daerah pusat-pusat yang ada di Jakarta.

Ada beberapa algoritma yang dapat digunakan untuk analisis sentimen, seperti Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors, dan Decision Trees[3]. Dalam penelitian ini, peneliti memilih algoritma Naïve Bayes karena memiliki beberapa keunggulan yang membuat implementasinya lebih mudah dan perhitungannya lebih sederhana. Salah satu keunggulan Naïve Bayes adalah kemampuannya untuk beroperasi dengan baik pada data yang terdistribusi secara tidak merata atau tidak seimbang. Hal ini membuatnya efektif dalam situasi di mana jumlah sampel dalam satu kelas jauh lebih banyak atau lebih sedikit dibandingkan dengan kelas lainnya. Selain itu, Naïve Bayes dikenal mampu menangani data berukuran besar dengan efisien.

Algoritma ini adalah algoritma machine learning yang digunakan untuk masalah klasifikasi, didasarkan pada teorema probabilitas Bayes. Keberhasilan algoritma ini telah terbukti dalam berbagai aplikasi, seperti penyaringan spam, analisis sentimen, dan klasifikasi artikel berita, terutama pada dataset berdimensi tinggi seperti yang digunakan dalam penelitian ini. Dengan menerapkan algoritma ini pada data yang diambil dari Dengan menerapkan algoritma ini pada data yang diperoleh dari berbagai platform media sosial yang mencerminkan beragam pandangan, penelitian ini bertujuan untuk memberikan wawasan lebih mendalam mengenai pendapat netizen Indonesia tentang transportasi umum terintegrasi di Jakarta. Dengan hasil analisis sentimen ini, diharapkan dapat memberikan informasi yang berharga kepada para pemangku kepentingan, termasuk pengguna, pemerintah, dan petugas, sehingga mereka dapat membuat keputusan yang lebih baik terkait transportasi umum yang telah terintegrasi. Melalui kontribusinya, penelitian ini diharapkan dapat meningkatkan pemahaman tentang peran media sosial dalam membentuk opini publik dan kebijakan terkait isu-isu lingkungan dan kesehatan.

Penelitian dengan cakupan yang serupa telah dilakukan oleh Enos Dwiyanto[4]. Dalam penelitiannya, Enos Dwiyanto menggunakan algoritma Naïve Bayes dan Support Vector Machine. Dalam tahap penelitian, data yang terkumpul dikelompokkan ke dalam dua kelas, yaitu positif dan negatif. Hasil akurasi penggunaan algoritma Naïve Bayes dan SVM oleh GrabId mencapai 84.08%, sementara untuk Gojek Indonesia mencapai 69.50%.

Penelitian yang memiliki cakupan serupa telah dilakukan oleh Ismia Iwandini[5]. Dalam penelitiannya, Ismia Iwandini menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. Dalam tahap penelitian, hasil analisis sentimen menunjukkan dominasi sikap positif, yang menunjukkan bahwa masyarakat umum peduli dengan adopsi teknologi. Penyelidikan ini menunjukkan bahwa pendekatan K-Nearest Neighbor lebih efektif daripada menggunakan algoritma Naïve Bayes.

Penelitian serupa dilakukan oleh Nova Tri Romadloni[6]. Dalam penelitian ini, digunakan metode Naive Bayes Classifier, KNN, dan Decision Tree dengan beberapa tahapan yaitu Convert Emoticon, Cleansing, Case Folding, Tokenizing, dan Stemming. Hasil uji coba menunjukkan perbedaan pada masing-masing metode: Naive Bayes memiliki akurasi sebesar 80%, precision 66,67%, sensitivity 100%, dan specificity 66,67%. Metode KNN memiliki akurasi sebesar 80%, precision 100%, sensitivity 50%, dan specificity 100%. Sedangkan metode Decision Tree mencapai akurasi 100%, precision 100%, sensitivity 100%, dan specificity 100%..

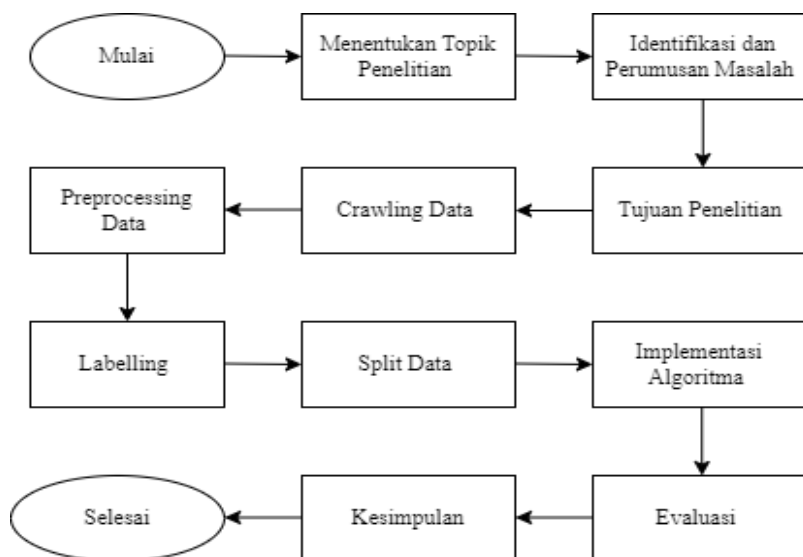
Artanti Inez[7] Tanggraeni juga telah melakukan penelitian serupa dengan menggunakan algoritma Naïve Bayes. Dalam penelitiannya, dia melakukan analisis sentimen terhadap 65 data ulasan pengguna yang dikumpulkan melalui teknik Web Scraping. Hasil analisis sentimen ini menunjukkan tingkat akurasi yang luar biasa tinggi. Dengan tingkat akurasi mencapai 87%, yang mendekati 100%, hal ini menunjukkan keberhasilan yang signifikan dalam analisis sentimen menggunakan klasifikasi Naïve Bayes dengan menerapkan pembobotan TF-IDF. Precision yang mencapai 83%, menunjukkan kinerja yang sangat baik dalam memprediksi data ulasan positif secara tepat. Selain itu, recall sebesar 87% menunjukkan sensitivitas prediksi terhadap data ulasan positif berdasarkan data yang sebenarnya.

Penelitian serupa juga dilakukan oleh Firman Noor Hasan[8] yang mengidentifikasi jumlah ulasan positif dan negatif terkait pelayanan Grab Indonesia menggunakan dataset. Tujuan dari penelitian ini adalah untuk mengevaluasi hasil analisis data dan menilai efisiensi layanan Grab Indonesia. Dari hasil yang dianalisis, terdapat 1000 dataset yang menunjukkan tingkat kepuasan pelanggan terhadap layanan Grab Indonesia, dengan 911 dataset memiliki sentimen positif dan 89 dataset memiliki sentimen negatif. Untuk sentimen negatif, precision yang dihasilkan adalah 57%, recall sebesar 67%, dan f1-score sebesar 62%. Sedangkan untuk sentimen positif, precision mencapai 97%, recall sebesar 95%, dan f1-score sebesar 96%.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Langkah-langkah yang akan dilakukan dalam penelitian ini mencakup: Pengumpulan Data (Crawling Data), Prapemrosesan Data (*Preprocessing Data*), Pemberian Label (*Labeling*), Pembagian Data Implementasi, Penerapan Algoritma Naïve Bayes, dan Evaluasi Hasil. Evaluasi akan membandingkan percobaan pertama yang tidak menggunakan operator Pembagian Data dengan percobaan kedua yang melibatkan operator Pembagian Data. Kesimpulan akan ditarik berdasarkan perbandingan hasil dari kedua percobaan tersebut. Rincian langkah-langkah penelitian dapat ditemukan dalam Gambar 1.



Gambar 1. Tahapan penelitian

a. Crawling Data

Pengambilan data dengan cara *crawler* dapat mempersatukan data secara otomatis. *Crawler* yang diperlukan dapat menyesuaikan dengan karakteristik platform masing-masing, *crawling* data dilakukan oleh peneliti dengan cara manual melalui platform media sosial seperti YouTube, , dan threads kemudian data yang telah dikumpulkan disimpan kedalam format *xlsx*[9].

b. Preprocessing Data

Dalam tahap ini, peneliti melakukan proses seperti *cleansing data* fungsi nya untuk menghilangkan data-data yang mengandung tanda baca, *emoticon* dan *numbering* kemudian *pada case folding* untuk mengubah kata yang berawalan huruf kapital ke bentuk huruf kecil selanjutnya *tokenizing* bertujuan untuk memisahkan setiap kata dalam suatu kalimat lalu *stemming* untuk merubah kata-kata singkatan menjadi kata dasar dan yang terakhir tujuan dari Filter Token (By Length) adalah untuk menghilangkan kata-kata yang terdiri dari 4 huruf.[10]

c. Labeling

Pada tahap *labeling* peneliti melakukan proses labelling di bantu oleh seorang expert dibidangnya dengan melakukan diskusi kedua belah pihak dengan Mengidentifikasi kata-kata yang mengandung sentimen positif atau negatif.[11]

d. Split Data

Dalam langkah *split* data, dilakukan pembagian data menjadi dua bagian, yaitu data latih dan data uji. Ini dilakukan dengan membagi dataset menjadi data pelatihan sebesar 80% dan data uji sebesar 20% [12].

e. Implementasi Algoritma.

Pada tahap ini peneliti melakukan implementasi algoritma dengan cara pengolahan data yang diterapkan disebuah tools. Maka dari itu peneliti menggunakan tools RapidMiner Studio Langkah dari implementasi ialah pertama dengan

menentukan algoritma machine learning yang akan digunakan seperti algoritma naïve bayes, selanjutnya melatih model menggunakan dataset pelatihan[13].

f. Evaluasi

Pada Tahap terakhir yaitu evaluasi yang di mana proses ini dilakukan evaluasi dari data yang sudah terbentuk. Tahap evaluasi ini untuk melakukan perhitungan nilai akhir dalam analisis klasifikasi mencakup *accuracy*, *precision*, dan *recall* dengan memakai *table confusion matrix* sebagai referensi, hasil perhitungan kemudian dipresentasikan kedalam tampilan *table confusion matrix*[14]

g. Kesimpulan

Pada tahap ini kita dapat mengetahui kata – kata yang sering digunakan oleh netizen untuk persoalan Transportasi Umum ini dan juga jumlah data class positif dan negatif yang di peroleh

2.2 Preprocessing

Tahap *preprocessing* memiliki tujuan untuk mengeliminasi dan Mengidentifikasi kata-kata yang kurang relevan dalam data untuk meningkatkan kualitas data dan menghasilkan teks yang lebih optimal dalam kumpulan data. Dalam tahap ini, peneliti melakukan proses seperti:

- a. *Cleansing data*: fungsi nya untuk menghilangkan data-data yang mengandung tanda baca, *emoticon* dan *numbering case folding* untuk mengubah kata yang berawalan huruf kapital ke bentuk huruf kecil sel.
- b. *Tokenizing* bertujuan untuk memisahkan setiap kata dalam suatu kalimat[15]. Operator yang digunakan adalah *Tokenizing* dengan menggunakan tabel rumus yang tersedia di web.kaggle.com[16]
- c. *Stemming* untuk merubah kata-kata singkatan menjadi kata dasar[17].
- d. *Filter Token (By Length)* bertujuan untuk menghilangkan kata-kata yang terdiri dari empat huruf. langkah ini melibatkan penggunaan perpustakaan *stopword* (Indonesia) dan formula yang berasal dari web.kaggle.com[18]

2.3 Algoritma Naïve Bayes

Tugas-tugas Data Mining melibatkan klasifikasi data, yaitu membagi atau mengklasifikasikan data ke dalam kelas atau beberapa kelas yang sebelumnya telah ditentukan. Salah satu metode yang digunakan untuk melakukan klasifikasi ini adalah Naïve Bayes. Naïve Bayes merupakan metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris bernama Thomas Bayes. Metode ini bekerja dengan mengestimasi probabilitas di masa yang akan datang berdasarkan pengalaman di masa yang sudah terjadi. Untuk rumus Naïve Bayes[19] terdapat pada persamaan rumus. Untuk menghitung probabilitas bersyarat A yang diberikan B yaitu dengan cara probabilitas bersyarat B yang diberikan A dikalikan dengan probabilitas kejadian A lalu dibagi dengan probabilitas kejadian B (1).

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1}$$

2.3 Confusion Matrix

Confusion matrix merupakan alat instrumen fungsinya itu mengkaji keefektifan pengelompokan yang sudah dilakukan dan juga untuk mengidentifikasi tupel dari kelas yang berbeda [20].

Tabel 1. Rumus confusion

Prediction Class	Actual Class	
	Pred. Positif	Pred. Negatif
Positif	TP	FP
Negatif	FN	TN

Dalam penilaian kualitas klasifikasi dengan melakukan perhitungan, terdapat beberapa metode yang dapat digunakan, namun salah satu yang umum digunakan adalah menggunakan akurasi, presisi, dan recall [19]. Akurasi adalah persentase dari prediksi model yang telah dibuat. Perhitungan nilai akurasi dilakukan dengan menggunakan persamaan rumus sebagai berikut (2).

$$Accuracy = (TP+TN) / (TP+TN+FN+FP) \times 100\% \tag{2}$$

Selain mengukur akurasi, metode untuk menilai kualitas klasifikasi melibatkan perhitungan *precision* dan *recall*. *Precision* mengindikasikan tingkat keakuratan informasi yang digunakan untuk menghasilkan jawaban oleh sistem. Penghitungan nilai *precision* dapat dilakukan dengan menggunakan persamaan rumus(3).

$$Precision = TP / (TP+FP) \times 100\% \tag{3}$$

Recall adalah perbandingan antara jumlah materi yang relevan dengan total materi yang relevan. Penghitungan nilai *recall* dapat dilakukan dengan menggunakan persamaan rumus(4).

$$Recall = TP / (TP+FN) \times 100\%. \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Crawling Data

Dalam penelitian ini, peneliti memanfaatkan metode *crawling* data dan menerapkan model Naive Bayes untuk mengklasifikasikan data dalam analisis sentimen. Pada awalnya, hanya berhasil mengumpulkan 40 data dari platform media sosial seperti YouTube. Namun, pada fase berikutnya, berhasil mengumpulkan 987 data dari platform media sosial seperti X. Pengambilan data dilakukan secara manual dari setiap platform untuk memastikan kualitas dan akurasi data yang dikumpulkan. Dalam analisis ini, penelitian memusatkan perhatian pada tiga kata kunci, yaitu Jaklingko, Transjakarta, dan MRT sebagai fokus utama untuk menilai sentimen atau respons netizen terhadap Transportasi Umum. Penggunaan Bahasa Indonesia dalam pengambilan data memungkinkan analisis untuk memberikan penafsiran yang lebih tepat mengenai respons dari pengguna Transportasi Umum, seperti yang terlihat dalam gambar 2.

1	created_at	full_text	tweet_url	username
2	Wed Jun 05 14:59:34 +0	@ezezzss @redzelnut @reeujinn Tapi emang copet hipnotis gacuma di 42 doang dinomor lain juga ada supirnya sendiri yang bilang pr	https://x.com/itsmeiffa_/s/itsmeiffa_	
3	Wed Jun 05 15:05:55 +0	@wonusaogf kayaknya naik tj/jaklingko kakak dari mana?	https://x.com/miguzips/st/miguzips	
4	Wed Jun 05 15:06:20 +0	@oxide_fa Hai Kak. Untuk KMT dapat digunakan juga di MRT Jakarta namun tap in dilakukan pada gate reader JakLingko di stasiun ya.	https://x.com/mrtjakarta/s/mrtjakarta	
5	Wed Jun 05 15:11:31 +0	@wonusaogf kalo naik krl nanti turun di st tebet kak nah di depan st tebet nanti ada minitrans/jaklingko 48a yang ke arah sana trus na	https://x.com/miguzips/st/miguzips	
6	Wed Jun 05 15:13:43 +0	@txtdrjkt Masih bersyukur bgt yaAllah kerja dr Cipete ke blok m cuma naik jaklingko cuma 13menit itupun gratis ga perlu war war krl l	https://x.com/iramaAyu/s/iramaAyu	
7	Wed Jun 05 15:15:01 +0	@miguzips kak mau barengan di jaklingko/minitrans nya nggaa?? kaya aku berangkat lebih awal nanti pas aku sampe stasiun Tebet kak	https://x.com/wonusaogf/wonusaogf	
8	Wed Jun 05 15:39:55 +0	Hari ini pas naik jaklingko aku ngescan kartuku & bantuin ibu2. Trs pas ada rombongan naik pada gabisa ngescan kartu mereka. Ké	https://x.com/jkstewart11/jkstewart11	
9	Wed Jun 05 16:04:09 +0	@thefirstake Kamu ada rekomen tempat ga? :D sjauh ini aku ada rekomen pling oke dan affordable di dket pondok indah (dan dket b)	https://x.com/ALN_Fandoi/ALN_Fandomless	
10	Wed Jun 05 16:24:48 +0	nama panjangnya sayidah nailaturahman biasa dipanggil naila. selasa pagi tanggal 4 juni naila pamil ke ortunya mau berangkat sekola	https://x.com/zoxoldyck/s/zoxoldyck	
11	Wed Jun 05 16:27:56 +0	@reeujinn tolong dong pihak terkait polisi Rt/Rw atau warga setempat cek cctv dari tempat biasa dia naik jaklingko apakah beneran n	https://x.com/tinaringosta/tinaringostar	
12	Wed Jun 05 17:22:18 +0	@ALN_Fandomless eh iyakah itu bisa naik jaklingko? kata teman ku kalau ke tadeo sana dari haltenya agak jauh	https://x.com/thefirstake/thefirstake	
13	Wed Jun 05 17:25:10 +0	@thefirstake Halte mayan tp jaklingko dket bgt:D Kalau ternyata lbh rame tar kitaa liat venue yg lbh okeh :3	https://x.com/ALN_Fandoi/ALN_Fandomless	
14	Wed Jun 05 18:10:52 +0	@cheeseyow maaf ya sudah beralih jd supir jaklingko	https://x.com/pepperonii/pepperonii	
15	Wed Jun 05 20:35:42 +0	@sasarails owh maaf yah aku soalnya tiap hari diajak fourtj naik jaklingko jadi gak tau mobil mahal	https://x.com/nekofourth/nekofourth	
16	Wed Jun 05 21:23:46 +0	@Shinjiii_Lee aku antara naik tj/jaklingko kak	https://x.com/miguzips/st/miguzips	
17	Wed Jun 05 23:04:41 +0	@RadioElshinta Semoga segera ditemukan jaklingko ada cctv harusnya bisa mempermudah pencarian	https://x.com/kura2giok/s/kura2giok	
18	Wed Jun 05 23:15:23 +0	Selalu males deh naek gojek dr rumah ke pinang ranti padahal bisa naek jaklingko gratis dan sehattt. Yuk hemat yukk woee	https://x.com/rapunz_jel/rapunz_jel	
19	Wed Jun 05 23:23:38 +0	@tanyakanrl coba di cek cctv jaklingko nyaa tuh tiap jaklingko jurusan apa aja psti ada cctv siapa tau membantu	https://x.com/abenkkkkkk/abenkkkkkkkk	
20	Thu Jun 06 00:04:36 +00	Suka bingung sm driver jak yang melipir ke jalur cepat ngapain coba. Mau stop di halte Tije gitu?? #jaklingko @PT_Transjakarta @Jakli	https://x.com/akun_buat_akun_buat_rant	

Gambar 2. Hasil Crawling data

3.2 Preprocessing

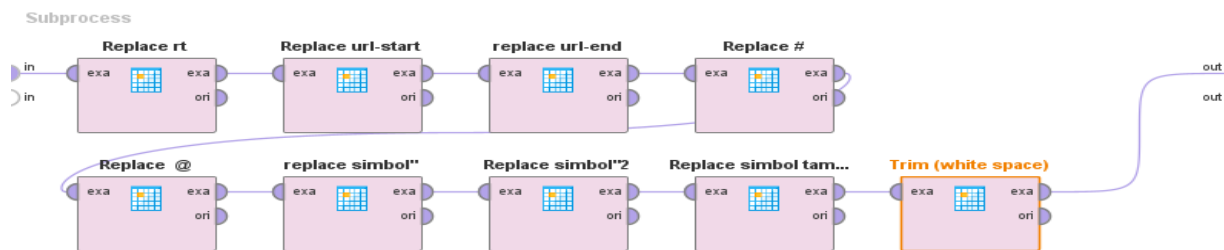
Kemudian, proses *preprocessing* data. Langkah ini bertujuan Meminimalkan penggunaan istilah yang kurang berhubungan, menyempurnakan mutu data terakhir menciptakan teks yang lebih efisien. pada kumpulan data[21]. Langkah ini melibatkan terdapat tahap, antara lain:

- a. Pada tahap *Cleansing Data*, tahap pertama dilakukan yaitu memakai operator *replace*. Langkah selanjutnya menyesuaikan parameter *replace* pada data yang memerlukan pembersihan, Untuk mengatasi hal ini, data duplikat, data yang mengandung simbol atau tanda baca, dan data yang mengandung emotikon dihilangkan. Hasil tahap *Cleansing data* dapat terlihat pada Tabel 2.

Tabel 1. Hasil Sample Cleaning Data

Input	Output
kalo mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih 😊	kalo mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih
@pttransjakarta kadang suka ada oknum min yang barbar bawa mobilnya jd bikin mual 😞	kadang suka ada oknum min yang barbar bawa mobilnya jd bikin mual
halo aku sebagai pemakai jaklingko menurut pendapat kacamata pribadi, jaklanko ini sangat amat membamntu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingan harga ojek pangkalan/online maupun pas banget lagi gapunya kendaraan, mau kemanapun juga ga mikir 😊	halo aku sebagai pemakai jaklingko menurut pendapat kacamata pribadi, jaklanko ini sangat amat membamntu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingan harga ojek pangkalan/online maupun pas banget lagi gapunya kendaraan, mau kemanapun juga ga mikir
Saya tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga👍	Saya tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga

Tahapan *Cleansing data* Data akan dihapus setelah diambil secara manual Dengan menggunakan aplikasi RapidMiner, dilakukan penghapusan data yang mengandung simbol atau tanda baca, data yang mengandung emotikon, dan data duplikat. Tahap *Cleansing data* sesuai dengan yang diperlihatkan dalam Gambar 3.



Gambar 3. Proses Cleaning data

- b. Pada tahap *Case Folding*, Langkah yang diterapkan adalah mengubah kata atau kalimat menjadi huruf kecil. menggunakan *lower case text*. Hasil tahap *Case Folding* dapat terlihat pada Tabel 3

Tabel 2. Hasil Case Folding

Input	Output
Kalo mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih @pttransjakarta Kadang suka ada oknum min yang barbar bawa mobilnya jd bikin mual	kalo mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih kadang suka ada oknum min yang barbar bawa mobilnya jd bikin mual
Halo aku sebagai pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko ini sangat amat menbammentu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingkan harga ojek pangkalan/online maupun pas banget lagi gapunya kendaraan, mau kemanapun juga ga mikir	halo aku sebagai pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko ini sangat amat menbammentu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingkan harga ojek pangkalan/online maupun pas banget lagi gapunya kendaraan, mau kemanapun juga ga mikir
Saya tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga	saya tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga
Saya beberapa kali naik jaklingko itu supir nya ngebut banget asli. Lebih ngebut daripada kalo naik angkot.	saya beberapa kali naik jaklingko itu supir nya ngebut banget asli. Lebih ngebut daripada kalo naik angkot.

Dalam langkah ini, case folding dilakukan menggunakan operator transform case pada RapidMiner.

- c. Kemudian, langkah penghapusan *Tokenizing* Tujuan utamanya adalah untuk memisahkan setiap kata dalam kalimat. Operator yang digunakan adalah *Tokenizing* dengan menggunakan tabel rumus yang tersedia di web.kaggle.com. Hasil proses *Tokenizing* dapat terlihat pada Tabel 4.

Tabel 3. Hasil Tokenizing

Input	Output
Kl mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih	kl, mau, nerapin, jam, malem, buat, jaklingko, 36, Cililitan, Cilangkap, ya, tambah, jg, armadanya, dong, nunggu, nya, masa, ,15, menit, lebih, sih
Kdg suka ada oknum min yang barbar bawa mobilnya jd bikin mual	kdg, suka, ada, oknum, min, yang, barbar, bawa, mobilnya, jd, bikin, mual
Halo aku sbg pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko ini sangat amat menbammentu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingkan Harga ojek pangkalan/online maupun pas bgt lagi gapunya kendaraan, mau kemanapun juga ga mikir	halo, aku, sbg, pemakai, jaklingko, menurut, pendapat, kacamata, pribadi, jaklinko, ini, sangat, amat, menbammentu, dalam, bidang, transportasi, buat, aku, kesana, kemari, kalo, lagi, membandingkan, harga, ojek pangkalan, online, maupun, pas, bgt, lagi, gapunya, kendaraan, mau, kemanapun, juga, ga, mikir
Sy tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga	sy, tiap, hari, turun, dari, komuterline, tanah, Abang, ke, kantor, pake, jaklingko, beneran, tertib, sih, angkotnya, bersih, di, semua, angkot, pasti, ada, cctv, nya, jadi, keamanan, juga, terjaga
Sy beberapa kali naik jaklingko itu supir nya ngebut banget asli. Lebih ngebut daripada kalo naik angkot.	sy, beberapa, kali, naik, jaklingko, itu, supir, nya, ngebut, banget, asli, Lebih, ngebut, daripada, kalo, naik, angkot.

Dalam langkah ini, *tokenizing* dilakukan menggunakan operator *tokenize* pada RapidMiner.

- d. Selanjutnya, proses *stemming* dilakukan untuk merubah kata-kata berimbuhan dan kata sambung menjadi bentuk dasarnya, sehingga dapat dipergunakan pada proses *stemming*, dapat terlihat pada Tabel 5

Tabel 4. Hasil Stemming

Input	Output
kl	Kalau
jg	Juga
sy	Saya
sbg	Sebagai
bgt	banget

Melalui *stemming*, diperoleh hasil dari proses ini yang mencakup transformasi kalimat dengan bahasa kekinian dapat disebut juga bahasa gaul di transformasikan Mereka dikonversi ke bentuk dasar dan dipisahkan berdasarkan parameter yang ditetapkan. Transformasi *Stemming* dapat terlihat pada Tabel 6.

Tabel 5. Transformasi Stemming

Input	Output
Kl mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih	kalau, mau, nerapin, jam, malem, buat, jaklingko, 36, Cililitan, Cilangkap, ya, tambah, juga, armadanya, dong, nunggu, nya, masa ,15, menit, lebih, sih
Kdg suka ada oknum min yang barbar bawa mobilnya jd bikin mual	kadang, suka, ada, oknum, min, yang, barbar, bawa, mobilnya, jd, bikin, mual
Halo aku sbg pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko ini sangat amat menbamtu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingan Harga ojek pangkalan/online maupun pas bgt lagi gapunya kendaraan, mau kemanapun juga ga mikir	halo, aku, sebagai, pemakai, jaklingko, menurut, pendapat, kacamata, pribadi, jaklinko, ini, sangat, amat, menbantu, dalam, bidang, transportasi, buat, aku, kesana, kemari, kalo, lagi, membandingan, harga, ojek, pangkalan, online, maupun, pas, banget, lagi, gapunya, kendaraan, mau, kemanapun, juga, ga, mikir
Sy tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga	saya, tiap, hari, turun, dari, komuterline, tanah, Abang, ke, kantor, pake, jaklingko, beneran, tertib, sih, angkotnya, bersih, di, semua, angkot, pasti, ada, cctv, nya, jadi, keamanan, juga, terjaga
Sy beberapa kali naik jaklingko itu supir nya ngebut banget asli. Lebih ngebut daripada kalo naik angkot.	saya, beberapa, kali, naik, jaklingko, itu, supir, nya, ngebut, banget, asli, Lebih, ngebut, daripada, kalo, naik, angkot.

- e. Kemudian, pada langkah ini, terdapat hasil dari proses *tokenizing* yang berfokus pada pengambilan kata-kata yang signifikan. Hasil *filter stopword* dapat terlihat pada Tabel 7.

Tabel 6. Hasil Filter Stopword

Input	Output
Kl mau nerapin jam malem buat jaklingko 36 Cililitan - Cilangkap ya tambah jg armadanya dong nunggu nya masa 15 menit lebih sih	kalau, mau, nerapin, jam, malem, buat, jaklingko, 36, Cililitan, Cilangkap, ya, tambah, juga, armadanya, dong, nunggu, nya, masa ,15, menit, lebih, sih
Kdg suka ada oknum min yang barbar bawa mobilnya jd bikin mual	kadang, suka, ada, oknum, min, yang, barbar, bawa, mobilnya, jd, bikin, mual
Halo aku sbg pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko ini sangat amat menbamtu dalam bidang transportasi buat aku kesana kemari kalo lagi membandingan Harga ojek pangkalan/online maupun pas bgt lagi gapunya kendaraan, mau kemanapun juga ga mikir	halo, aku, sebagai, pemakai, jaklingko, menurut, pendapat, kacamata, pribadi, jaklinko, ini, sangat, amat, menbantu, dalam, bidang, transportasi, buat, aku, kesana, kemari, kalo, lagi, membandingan, harga, ojek, pangkalan, online, maupun, pas, banget, lagi, gapunya, kendaraan, mau, kemanapun, juga, ga, mikir
Sy tiap hari, turun dari komuterline tanah Abang ke kantor pake jaklingko, beneran tertib sih, angkotnya bersih, di semua angkot pasti ada cctv nya, jadi keamanan juga terjaga	saya, tiap, hari, turun, dari, komuterline, tanah, Abang, ke, kantor, pake, jaklingko, beneran, tertib, sih, angkotnya, bersih, di, semua, angkot, pasti, ada, cctv, nya, jadi, keamanan, juga, terjaga
Sy beberapa kali naik jaklingko itu supir nya ngebut banget asli. Lebih ngebut daripada kalo naik angkot.	saya, beberapa, kali, naik, jaklingko, itu, supir, nya, ngebut, banget, asli, Lebih, ngebut, daripada, kalo, naik, angkot.

Dalam langkah ini, Proses *filter stopwords* dilakukan memanfaatkan operator *filter stopwords* (dictionary) pada RapidMiner.

- f. Dalam langkah ini, proses *Filter Token (By Length)* bertujuan untuk mengeliminasi kata-kata yang telah ditetapkan, yakni susunan kata-kata yang mencakup kurang dari empat huruf. Oleh karena itu, langkah ini melibatkan penggunaan perpustakaan *stopword (Indonesia)* dan formula yang berasal dari *web.kaggle.com*. Hasil dari *filter token (by length)* dapat terlihat pada Tabel 8.

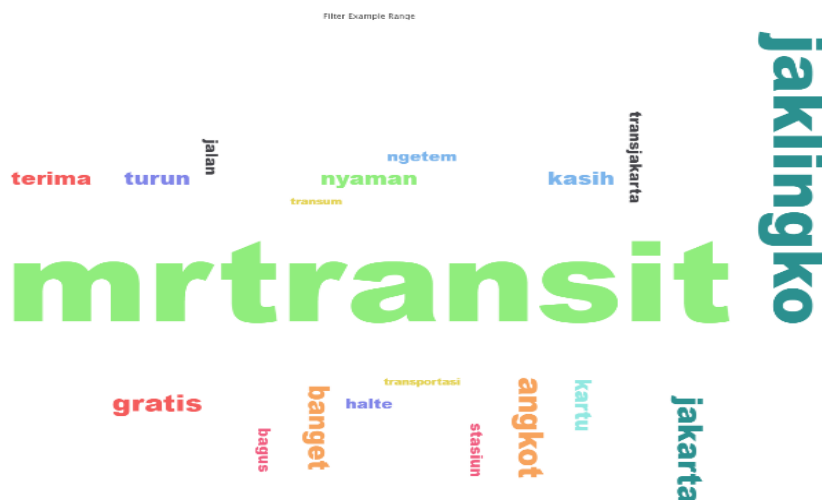
Tabel 7. Proses Filter Token

Input	Output
kalau, mau, nerapin, jam, malem, buat, jaklingko, 36, Cililitan, Cilangkap, ya, tambah, juga, armadanya, dong, nunggu, nya, masa ,15, menit, lebih, sih kadang, suka, ada, oknum, min, yang, barbar, bawa, mobilnya, jadi, bikin, mual halo, aku, sebagai, pemakai, jaklingko, menurut, pendapat, kacamata, pribadi, jaklinko, ini, sangat, amat, membantu, dalam, bidang, transportasi, buat, aku, kesana, kemari, kalo, lagi, membandingan, harga, ojek pangkalan, online, maupun, pas, banget, lagi, gapunya, kendaraan, mau, kemanapun, juga, ga, mikir saya, tiap, hari, turun, dari, komuterline, tanah, Abang, ke, kantor, pake, jaklingko, beneran, tertib, sih, angkotnya, bersih, di, semua, angkot, pasti, ada, cctv, nya, jadi, keamanan, juga, terjaga saya, beberapa, kali, naik, jaklingko, itu, supir, nya, ngebut, banget, asli, Lebih, ngebut, daripada, kalo, naik, angkot.	Kalau nerapin malem buat jaklingko Cililitan - Cilangkap tambah juga armadanya dong nunggu masa menit lebih Kadang suka oknum yang barbar bawa mobilnya jadi bikin mual Halo sebagai pemakai jaklingko menurut pendapat kacamata pribadi, jaklinko sangat amat membantu dalam bidang transportasi buat kesana kemari kalo lagi membandingan Harga ojek pangkalan/online maupun banget lagi gapunya kendaraan, mau kemanapun juga mikir Saya tiap hari, turun dari komuterline tanah Abang kantor pake jaklingko, beneran tertib sih, angkotnya bersih, semua angkot pasti cctvnya, jadi keamanan juga terjaga saya beberapa kali naik jaklingko supir, ngebut, banget, asli, Lebih, ngebut, daripada, kalo, naik, angkot.

Dalam langkah ini, Proses *Filter Token (By Length)* dilakukan memanfaatkan operator *Filter Token (By Length)* pada RapidMiner.

3.3 Labeling

Dalam tahap ini, terdapat empat langkah dalam mengklasifikasikan data. Pertama *labeling* untuk memberikan label pada data yang akan dianalisis adalah langkah pertama. Setelah itu, dilakukan pemisahan data dengan Pembagian Data menjadi dua bagian, yaitu *Data Training* (data latih) dan *Data Testing* (data uji), dilanjutkan dengan pelaksanaan algoritma menggunakan operator Naive Bayes. Pada langkah terakhir, evaluasi dilakukan, termasuk penyisipan grafik dan wordcloud sebagai hasil akhir dari pengujian data. Dari dataset yang didapatkan terbagi menjadi 338 MRT yang terdiri dari 252 positif dan 86 negatif, 213 Transjakarta yang terdiri dari 130 positif dan 92 negatif, 424 Jaklingko yang terdiri dari 228 positif dan 196 negatif. Diperoleh dengan total data class positif sejumlah 612 data dan class negatif sejumlah 375 data. Penggunaan visualisasi wordcloud digunakan untuk mengamati kalimat yang kerap terlihat pada komentar yang bersifat positif, dapat terlihat pada gambar 4.



Gambar 4. Wordcloud Positif

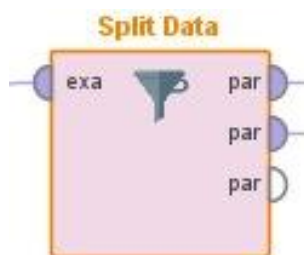
Penggunaan wordcloud visualisasi digunakan untuk kalimat yang kerap terlihat pada komentar yang bersifat negatif, dapat terlihat pada gambar 5.



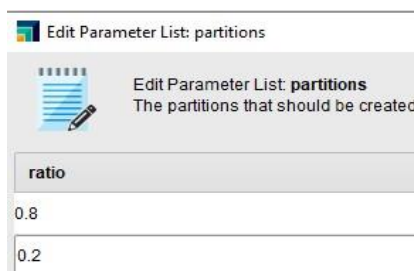
Gambar 5. Wordcloud Negatif

3.4 Split Data

Tahap selanjutnya melibatkan pembagian data ke dalam dua kelompok, yakni data *training* (data latih) dan data *testing* (data uji). Dalam penelitian ini, pengelompokan data dilakukan dengan mengalokasikan 80% atau 0,8 dari total data sebagai data training, sementara 20% atau 0,2 sisanya diterapkan ke dalam data testing. Alasan digunakannya split data dengan perbandingan 80:20 yaitu pada perbandingan rasio 80:20 masih memberikan ukuran data yang cukup besar untuk pengujian model karena jika dataset memiliki jumlah data yang kecil maka akan terjadi resiko kemungkinan overfitting, yang dimana overfitting adalah adaptasi model yang berlebihan terhadap data pelatihan, terdapat pada gambar 6 dan 7.



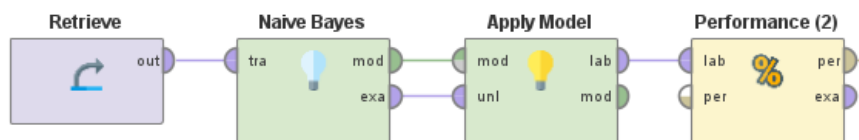
Gambar 6. Processing Operator Split Data



Gambar 6. Ration Operator *Split Data*

3.5 Implementasi Algoritma Naïve Bayes

Langkah penerapan algoritma klasifikasi Naïve Bayes melibatkan analisis dan pengklasifikasian data *training* dengan algoritma *Naïve Bayes*, sementara data testing disesuaikan dengan metode pemodelan. Tahapan tersebut dapat diamati pada gambar 8.



Gambar 7. Performance

3.6 Evaluasi

Pada tahap akhir, peneliti membandingkan dua percobaan analisis klasifikasi menggunakan tiga dataset moda transportasi umum dan 1 dataset gabungan semua transportasi umum. Hasil perhitungan disajikan dalam bentuk tabel confusion matrix. Berikut adalah data yang dihasilkan dari perbandingan ketiga percobaan tersebut, yang dapat dipresentasikan melalui tabel *confusion matrix*.

a. Percobaan pertama dataset Jaklingko.

Tabel 9. Percobaan pertama dataset Jaklingko

	True Negatif	True Positif	Class precision
Pred. positif	0	220	100.00%
Pred. negatif	196	8	98.06%
Class recall	100.00%	96.49%	
Accuracy			98.11%

Diperoleh hasil dari uji coba pertama tanpa menerapkan operator Pembagian Data, dengan tingkat akurasi sebesar 98,11%. Pada kategori positif, diperoleh Precision sebesar 100% dan Recall sebesar 94,49%, sementara pada kategori negatif, Precision mencapai 98,06% dan Recall mencapai 100%.

b. Percobaan kedua dataset Transjakarta.

Tabel 10. Percobaan kedua dataset Transjakarta

	True Negatif	True Positif	Class precision
Pred. positif	0	130	100.00%
Pred. negatif	93	0	100.00%
Class recall	100.00%	100.00%	
Accuracy			100.00%

Diperoleh hasil dari uji coba pertama tanpa menerapkan operator Pembagian Data, dengan tingkat akurasi sebesar 100,00%. Pada kategori positif, diperoleh Precision sebesar 100% dan Recall sebesar 100,00%, sementara pada kategori negatif, Precision mencapai 100,00% dan Recall mencapai 100%.

c. Percobaan ketiga dataset MRT.

Tabel 10. Percobaan ketiga dataset MRT

	True Negatif	True Positif	Class precision
Pred. positif	0	234	100.00%
Pred. negatif	86	18	82.69%
Class recall	100.00%	92.86%	
Accuracy			94.67%

Diperoleh hasil dari uji coba pertama tanpa menerapkan operator Pembagian Data, dengan tingkat akurasi sebesar 94.67%. Pada kategori positif, diperoleh Precision sebesar 100% dan Recall sebesar 92,86%, sementara pada kategori negatif, Precision mencapai 82,69% dan Recall mencapai 100%.

d. Percobaan akhir dataset semua transportasi umum.

Perhitungan akhir ini adalah hasil utama dari penelitian. Hasil yang dapat ditampilkan menggunakan tabel confusion matrix dapat dilihat pada tabel 9.

Tabel 11. Hasil Eksperimen

	True Negatif	True Positif	Class
Pred. positif	0	555	100.00%
Pred. negatif	57	375	86.81%
Class recall	100.00%	90.69%	
Accuracy			94.22%

Memanfaatkan operator *Split* Data dapat diperoleh hasil, dengan tingkat *Accuracy* sebesar 94.06%. Pada *class* positif, ditemukan tingkat *Precision* sebesar 100% dan *Recall* sebesar 90.49%, sedangkan untuk *class* negatif, *Precision* mencapai 86.33% dan *Recall* mencapai 100%

Untuk perhitungan manual nilai akurasi diperlukan rumus sebagai berikut:

$$Akurasi = \frac{555 + 375}{555 + 0 + 375 + 57}$$

$$Akurasi = \frac{930}{987}$$

$$Akurasi = 0,94$$

Maka didapatkan hasil nilai akurasi yaitu 0,9422 atau 94.22%. Selanjutnya untuk menghitung presisi digunakan rumus sebagai berikut:

$$\begin{aligned} Presisi &= \frac{TP}{TP + FP} \\ &= \frac{555}{555 + 0} \\ &= \frac{555}{555} = 1 \end{aligned}$$

Maka didapatkan hasil nilai presisi yaitu 1 atau 100%. Selanjutnya untuk menghitung *recall* digunakan rumus sebagai berikut:

$$\begin{aligned} Recall &= \frac{TP}{TP + FN} \\ &= \frac{555}{555 + 57} \\ &= \frac{555}{612} = 0,90 \end{aligned}$$

Maka didapatkan hasil nilai recall yaitu 0,90 atau 90%.

4. KESIMPULAN

Melalui analisis sentimen yang dilakukan terhadap Transportasi Umum di Jakarta, dapat diambil kesimpulan bahwa sebagian besar netizen yang juga seorang pengguna Transportasi Umum menyampaikan bahwa sudah bagusnya system Transportasi Umum yang ada di Jakarta. Dengan terintergrasinya semua moda Transportasi membuat pengguna merasakan keamanan dan kenyamanan yang berlebih saat di gunakan. kamera pengawas yang terpasang membuat rasa aman, kondisi kendaraan yang layak jalan juga membuat nyaman saat digunakan. Terlebih murah nya Harga Transportasi Umum di Jakarta membuat semua pengguna bisa menaikinya Serta menghemat biaya perjalanan. Tapi juga ada Sebagian pengguna yang tidak merasa nyaman dan aman. Memang harga yang dikasih murah tapi beberapa pendapat bilang menjadikan Transportasi Umum tidak aman dan nyaman seperti masih banyaknya supir yang ugal-ugalan, Waktu tiba kendaraan yang lama dan juga belum beroperasinya Transportasi Umum secara 24 jam. secara umum, sentimen netizen mencerminkan kesadaran akan pentingnya menaiki Transportasi Umum untuk tata Kelola kota yang lebih baik lagi. Oleh karena itu, diperlukan kerjasama bersama dan implementasi tindakan nyata untuk mengatasi permasalahan Transportasi umum di Jakarta. Dari hasil dan analisis yang telah dilakukan, dapat disimpulkan bahwa Hasil dari total data class yang diperoleh yaitu data class positif sejumlah 612 data dan class negatif sejumlah 375 data. MRT menjadi jumlah class positif paling banyak diantara moda transportasi umum yang lain. Sedangkan jaklingko menjadi class negatif terbanyak diantara moda transportasi yang lain. Hasil evaluasi, termasuk nilai *Accuracy*, presisi, dan *recall* dalam penelitian ini, dapat dirinci sebagai berikut: Akurasi *Naive Bayes* mencapai 94.22%, nilai presisi sebesar 86.81%, dan nilai *recall* sebesar 90.69%. Secara keseluruhan, hasil penelitian menegaskan bahwa algoritma Naive Bayes merupakan pilihan yang tepat untuk analisis sentimen pada topik yang diteliti, dan hasilnya menunjukkan tingkat akurasi yang memuaskan.

REFERENCES

- [1] A. D. Nurlukman *et al.*, "IMPLEMENTASI SMART ENVIRONMENT CITY DALAM TATA KELOLA LINGKUNGAN DI KOTA TANGERANG," vol. 9, no. November, pp. 769–784, 2023.
- [2] B. D. Forino *et al.*, "PERSEPSI PENGGUNA TRANSPORTASI UMUM DI JABODETABEK TERHADAP INTEGRASI TARIF PT JAKLINGKO INDONESIA," vol. 6, no. 1, pp. 71–84, 2023.
- [3] L. A. Pramesti and N. Pratiwi, "Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine," vol. 4, no. 4, 2023, doi: 10.47065/josh.v4i4.3807.
- [4] E. Dwianto and M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naive Bayes dan Support Vector Machine," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 1, p. 94, 2021, doi: 10.22441/format.2021.v10.i1.009.
- [5] I. Iwandini, A. Triayudi, and G. Soepriyono, "Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 543–550, 2023, doi: 10.47065/josh.v4i2.2937.
- [6] N. T. Romadloni, I. Santoso, S. Budilaksono, M. I. Komputer, and C. Folding, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL," vol. 3, no. 2, pp. 1–9, 2019.
- [7] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [8] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan

- Multinomial Naïve Bayes Classifier,” vol. 4, no. 2, pp. 52–58, 2021.
- [9] J. Budiarto, “Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling (Requirements Identification for NTB People in pandemic covid-19 at Social Media Using Crawling Method),” vol. 2, no. 4, pp. 244–250, 2021.
- [10] E. Kurnianto and D. Febriawan, “Analisis Sentimen Perbedaan Pendapat Netizen Indonesia Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naïve Bayes,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 404–414, 2023, doi: 10.30865/json.v5i2.7170.
- [11] A. Aziz, “Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, p. 115, 2022.
- [12] A. Puteri *et al.*, “Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstraksi Warna HSV dan Tekstur GLCM,” vol. 6, no. 1, pp. 124–137, 2021.
- [13] N. R. Fadhila, T. Sutarti, M. Patende, G. Yudha, and D. D. Saputra, “Implementasi Machine Learning Untuk Klasifikasi Narasi Informatif dan Non-Informatif Pada Media Sosial Twitter TMC Polda Metro Jaya Menggunakan Naïve Bayes Classifier,” vol. 1, pp. 11–17, 2023.
- [14] G. F. Akbar and E. Sinduningrum, “CESS Analisis Sentimen Netizen Terhadap Polusi Udara di DKI Jakarta Analysis of Netizen Sentiment Towards Air Pollution in DKI Jakarta,” vol. 9, no. 1, pp. 276–287, 2024.
- [15] S. Yogi Yunefri, Yogi Ersan Fadrial, “CHATBOT PADA SMART COOPERATIVE ORIENTED PROBLEM MENGGUNAKAN NATURAL LANGUAGE PROCESSING DAN NAIVE BAYES CLASSIFIER,” vol. 4, pp. 131–141, 2021.
- [16] RAVI SHAH, *Guide to Tokenizers and Processing*. 2022.
- [17] O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, and S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” vol. 2, no. 1, pp. 36–40, 2022.
- [18] B. K. dan M. Luca, *Packt Publishing, Limited, Packt Publishing*. 2022.
- [19] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma Naïve Bayes Classifier,” vol. 5, no. 1, pp. 34–43, 2022.
- [20] T. A. Sundara, S. Ekaputri, and S. Sotar, “Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme,” *Pros. SISFOTEK*, vol. 4, no. 1, pp. 93–98, 2020.
- [21] V. Alpiana, A. Salam, F. Alzami, I. Rizqa, and D. Aqmala, “Analisis Topic-Modelling Menggunakan Latent Dirichlet Allocation (LDA) Pada Ulasan Sosial Media Youtube,” vol. 8, pp. 332–341, 2024, doi: 10.30865/mib.v8i1.7127.