

Sentiment Analisis Opini Masyarakat Sistem Ganjil Genap di Twitter Menggunakan Algoritma Naive Bayes Classifier dan Algoritma K-NN

Acep Setiawan, Arafat Febriandirza*

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

Email: ¹aceps.me@gmail.com, ^{2,*}arafat@uhamka.ac.id

Email Penulis Korespondensi: arafat@uhamka.ac.id

Abstrak—Abstrak jurnal ini membahas analisis sentimen opini publik terkait penerapan sistem ganjil-genap di Twitter, dengan menggunakan algoritma K-NN dan Naive Bayes Classifier. Sistem ganjil-genap ini dibahas dalam tweet oleh pengguna Twitter, yang menjadi sumber data. Tweet-tweet tersebut dikategorikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa dari total tweet yang dikumpulkan, 391 dikategorikan sebagai netral, 50 sebagai negatif, dan 59 sebagai positif. Selain itu, ditemukan bahwa algoritma Naive Bayes dan algoritma K-Nearest Neighbor keduanya memiliki tingkat akurasi rata-rata sekitar 79,72%. Ini menunjukkan bahwa kedua algoritma memiliki kinerja yang serupa dalam mengklasifikasikan sentimen tweet yang dibahas. Terkait dengan analisis sentimen opini publik di platform Twitter, kesimpulan ini memperjelas perbandingan kinerja antara algoritma Naive Bayes dan K-Nearest Neighbor.

Kata Kunci: Analisis Sentimen; Opini Masyarakat; Twitter; Sistem Ganjil Genap; Algoritma Naive Bayes Classifier; Algoritma K-Nearest Neighbor; Akurasi

Abstract—This journal's abstract addresses sentiment analysis of public opinion in relation to the odd-even system's implementation on Twitter, utilizing the K-NN and Naive Bayes Classifier algorithms. The odd-even system was discussed in tweets by Twitter users, which served as the source data. The tweets were categorized into three sentiment categories: positive, negative, and neutral. The analysis's findings indicate that, of the total number of tweets gathered, 391 were categorized as neutral, 50 as negative, and 59 as positive. In addition, it was found that the Naive Bayes algorithm and the K-Nearest Neighbor algorithm both had an average accuracy rate of approximately 79.72%. This suggests that both algorithms do similarly well when it comes to classifying the sentiment of the tweets under discussion. With respect to sentiment analysis of public opinion on the Twitter platform, this conclusion clarifies the performance comparison between the Naive Bayes and K-Nearest Neighbor algorithms.

Keywords: Sentiment Analysis; Public Opinion; Twitter; Odd-Even System; Naive Bayes Classifier Algorithm; K-Nearest Neighbor Algorithm; Accuracy

1. PENDAHULUAN

Data dari Badan Pusat Statistik (BPS) DKI Jakarta menunjukkan bahwa sebanyak 11.839.921 mobil terdaftar pada tahun 2016 (BPS DKI Jakarta, 2019). Jumlah ini tidak termasuk kendaraan resmi, militer, atau polisi. Karena DKI Jakarta memiliki jumlah kendaraan yang begitu besar, kemacetan lalu lintas akan semakin parah setiap tahun jika tidak ada perluasan jalan yang signifikan di daerah tersebut[1]. Tujuan dari kebijakan sistem ganjil-genap di Jakarta adalah untuk mengurangi kemacetan lalu lintas. Jumlah orang yang menggunakan kendaraan pribadi meningkat seiring dengan bertambahnya jumlah mobil di jalan. Akibatnya, kebijakan seperti pembatasan plat nomor ganjil-genap—yang berlaku khusus untuk mobil—harus diterapkan guna mengantisipasi peningkatan jumlah kendaraan di jalan. Namun, efektivitas kebijakan ini semakin terhambat oleh penerapan disiplin dan kesadaran publik yang tidak konsisten[2].

Dari 272,1 juta penduduk Indonesia, 175,4 juta memiliki akses internet, dan 160 juta di antaranya adalah pengguna media sosial. Laporan Digital Indonesia 2020 menyatakan bahwa, selain platform lain seperti YouTube, Messenger, TikTok, dan Tumblr[3]. Jaringan media sosial terbesar berdasarkan persentase pengguna adalah 82% pengguna Facebook, 79% pengguna Instagram, dan 56% pengguna Twitter. Twitter, sebuah situs media sosial yang terkenal, digunakan oleh orang-orang di seluruh muka bumi, termasuk di Indonesia. Jaringan media sosial ini dikembangkan oleh Jack Dorsey agar orang bisa berkomunikasi satu sama lain melalui tweet. Dengan 19,5 juta pengguna, Twitter menduduki peringkat kelima di antara situs media sosial yang terkenal pada tahun 2020[4].

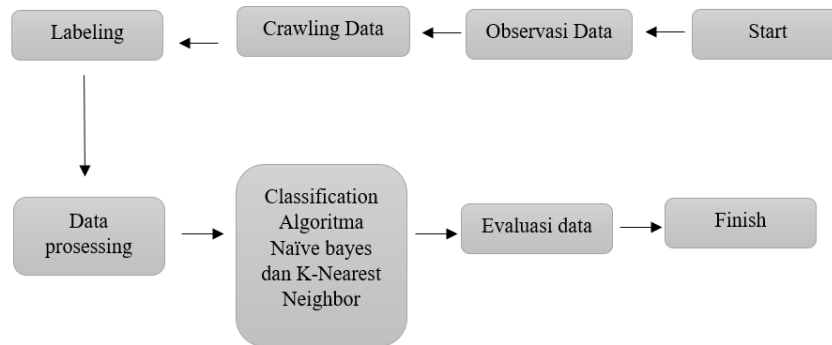
Analisis sentimen adalah interpretasi dan pemrosesan otomatis data tekstual untuk mengekstrak informasi sentimen dari pernyataan opini menggunakan model yang telah dikembangkan dengan data pelatihan[5]. Penelitian yang memeriksa opini publik di Twitter mengenai kebijakan ganjil-genap adalah salah satu contoh bagaimana analisis sentimen telah berkembang dan banyak dibahas dalam berbagai studi. Teknik K-NN dan Naive Bayes digunakan kepada klasifikasi[6]. maksud dari studi ini adalah untuk memeriksa opini publik mengenai pembelajaran daring sehubungan dengan dukungan dan penolakan yang muncul di masyarakat terhadap praktik tersebut. Teknik untuk menganalisis, menghasilkan, dan memproses data tekstual dengan cepat guna mengekstrak informasi yang terdapat dalam kalimat opini dikenal sebagai analisis sentimen (juga dikenal sebagai penambangan opini)[7]. Studi kasus memiliki dampak besar pada teknik analisis sentimen yang digunakan. Misalnya, metode K-Nearest Neighbor (KNN) digunakan oleh F. A. Rohmansyah, B. Bintoro, dan I. Santoso Untuk mengukur opini terkait penerapan sistem lalu lintas ganjil-genap. Sebaliknya, R. T. Aldisa dan P. Maulana[9] menganalisis bagaimana perasaan publik tentang vaksin booster COVID-19 menggunakan pendekatan Naive Bayes, Decision Tree, dan SVM. Selain itu, F. D. Samuel, P. D. Atika, dan S. Setiawati[8]. Menunjukkan bahwa penggunaan Analisis sentimen opini publik mengenai pembelajaran daring di Twitter menggunakan algoritma Naive Bayes dan Support Vector Machine menghasilkan akurasi 72% untuk Naive Bayes dan

66% untuk SVM. Oleh karena itu, dapat disimpulkan bahwa pendekatan Naïve Bayes lebih baik dalam mengklasifikasikan data ini. Hal ini menunjukkan bahwa tingkat akurasi, presisi, dan memori juga mempengaruhi pemilihan metode. Berdasarkan percakapan sebelumnya, penulis berencana untuk mempelajari opini publik di Twitter mengenai penerapan sistem ganjil-genap. Penelitian ini akan mengkategorikan opini yang diungkapkan dan membandingkan akurasi algoritma Naïve Bayes dan K-Nearest Neighbor (KNN).

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Gambar 1 di bawah merupakan diagram alir dimana menggambarkan proses atau metode dari penelitian ini.



Gambar 1. Diagram alir

2.2 Metode Penelitian

2.2.1 Teknik Pengumpulan Data

Di situs jejaring sosial Twitter, pengguna bertukar pesan singkat yang dikenal sebagai tweet. Sebagai sarana untuk pengembangan profesional yang berkelanjutan dan pribadi, pendidik dari seluruh dunia telah bergabung dengan Twitter dan memanfaatkannya untuk berinteraksi, berkomunikasi, dan berbagi satu sama lain[8] API Twitter digunakan untuk melakukan teknik crawling sebagai pendekatan pengumpulan data dalam penelitian ini. Teknik ini memungkinkan pengumpulan data tweet berdasarkan kata kunci yang ditentukan, dengan rentang waktu selama tahun 2021. Secara prinsip, teknik ini mirip dengan menggunakan fitur pencarian pada situs web Twitter melalui peramban internet[9]. Hasil dari tahap ini adalah sebuah file yang berisi 680 data opini.

2.2.2 Data Prossesing

Langkah penting dalam klasifikasi teks adalah pra-pemrosesan data, yang bertujuan untuk membuat data menjadi lebih bersih sebelum diproses lebih lanjut[10]. Dengan menghilangkan noise atau informasi yang tidak perlu seperti emoji, URL, dan angka, serta menstandarkan bentuk kata. Tujuan utama dari fase pra-pemrosesan adalah untuk menyiapkan data mentah agar siap untuk pemrosesan berikutnya. Pada langkah ini, data yang tidak terstruktur dengan baik akan dihapus, diubah menjadi format yang lebih mudah dianalisis dan diorganisir, atau dimasukkan ke dalam sistem[11]. Dalam data yang dikumpulkan, sering ditemukan istilah yang tidak terstruktur, yang memerlukan teknik pemrosesan atau pembersihan data seperti case folding untuk mengubah semua huruf kapital menjadi huruf kecil[12]. Dalam text mining, pra-pemrosesan adalah langkah pertama yang menyiapkan teks untuk pemrosesan lebih lanjut. Salah satu keuntungan dari pra-pemrosesan data ini adalah memperoleh daftar istilah yang signifikan dan praktis untuk analisis sentimen[13]. Sebelum melanjutkan ke tahap berikutnya, pra-pemrosesan melibatkan penerapan berbagai operasi pada data, seperti pembersihan, tokenisasi, transformasi bentuk kata, penghapusan stopword, dan penyaringan token[14].

2.2.3 Classification Algoritma Naive Bayes

Algoritma probabilistik yang populer untuk masalah klasifikasi adalah Naive Bayes. Naive Bayes Classifier adalah sekumpulan algoritma klasifikasi yang dibangun berdasarkan premis umum yang berasal dari Teorema Bayes[15]. Teknik ini, yang mengasumsikan independensi fitur dalam data[16]. Dikenal karena metode klasifikasinya yang sederhana dan kecepatan pemrosesan yang cepat. Naive Bayes cocok untuk klasifikasi biner maupun multikelas. Menggunakan teknik klasifikasi terawasi, Naive Bayes Classifier memprediksi label kelas untuk data yang belum diketahui dengan menggunakan probabilitas bersyarat[17]. Berdasarkan teorema Bayes, klasifikasi Bayesian dapat memberikan akurasi dan kecepatan klasifikasi yang sangat baik, terutama ketika digunakan dengan basis data besar[18].

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(x)} \quad (1)$$

Probabilitas posterior, yang dilambangkan sebagai $P(C_i | X)$, menggambarkan probabilitas kelas C_i dengan input X , sedangkan likelihood, yang dilambangkan sebagai $P(X|C_i)$, menunjukkan probabilitas pengamatan X diberikan kelas C_i . Probabilitas total, dilambangkan sebagai $P(X)$, mencakup probabilitas keseluruhan kejadian.

2.2.4 Algoritma K-Nearest Neighbor

Karena popularitasnya dalam data mining, teknik KNN (K-nearest Neighbors) dapat melakukan regresi dan klasifikasi, terutama ketika menangani data non-numerik. Algoritma ini memerlukan dataset yang dibagi menjadi data training dan data testing[19]. Untuk melakukan klasifikasi. Salah satu keunggulan utama KNN adalah kemampuannya untuk menangani masalah multikelas. Namun, teknik ini menghadapi kesulitan dalam menemukan tetangga terdekat dari titik query dalam dataset yang digunakan. Gorunescu menyatakan bahwa algoritma Nearest Neighbor, yang sering dikenal sebagai K-nearest Neighbor atau K-NN, mengklasifikasikan data berdasarkan seberapa dekat atau jauh data tersebut dari titik data lainnya[20].

$$\text{Cos sim (a,b)} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

A adalah Sampel Uji, B adalah Sampel Latih, dengan A_i sebagai elemen dari Sampel Uji, dan B_i sebagai elemen dari Sampel Latih.

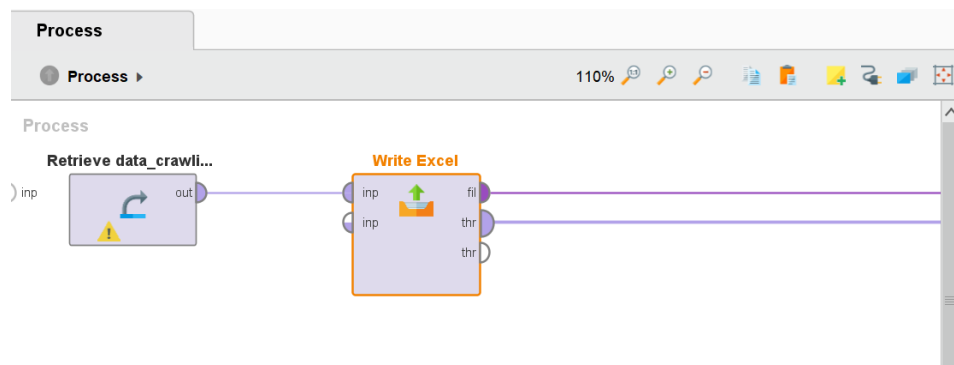
2.2.5 Rapidminer

RapidMiner adalah program sumber terbuka yang menawarkan alat untuk analisis prediktif, text mining, dan data mining. RapidMiner berupaya memberikan informasi yang dibutuhkan pengguna untuk membuat keputusan terbaik dengan memanfaatkan berbagai metodologi deskriptif dan prediktif. Program ini tersedia untuk diunduh di situs web resminya. Pertama kali dikembangkan pada tahun 2001[21].

3. HASIL DAN PEMBAHASAN

3.1 Data Crawling

Proses pengambilan data menggunakan Rapid Miner diilustrasikan dalam gambar tersebut. Pertama, digunakan beberapa operator untuk melakukan crawling data, seperti operator search Twitter untuk terkoneksi dengan Twitter melalui penyediaan kode akses token dari Twitter API. Operator remove duplicates dipakai untuk menghapus data duplikat saat proses crawling. Kemudian, operator select attributes, seperti yang diperlihatkan dalam gambar bagian pertama, dipergunakan untuk mengekstrak atribut yang penting seperti username dan teks. Untuk menyimpan data dalam format Excel, Write Excel adalah operator terakhir yang digunakan. RapidMiner digunakan untuk mendapatkan hasil pengambilan data.



Gambar 2. Proses Data Crawling

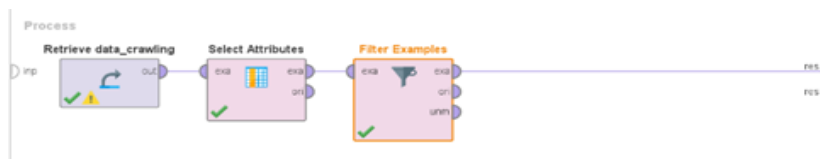
3.2 Tahap Processing Data

Dalam proses pengolahan data, peneliti melakukan berbagai langkah, seperti pembersihan, penyaringan, tokenisasi, analisis sentimen, dan langkah-langkah lainnya. Langkah ini bertujuan untuk menghasilkan data yang akurat dan mudah dibandingkan untuk analisis lebih lanjut. Pembersihan data adalah proses menghilangkan noise atau gangguan yang tidak terkait. Sedangkan, penyaringan data digunakan untuk memilih data yang memenuhi kriteria tertentu. Tokenisasi kemudian memecah teks menjadi token-token terpisah untuk mempermudah analisis. Analisis sentimen bertujuan untuk menentukan sentimen dari teks tersebut, seperti apakah bersifat positif, negatif, atau netral. Dengan menjalankan serangkaian langkah ini, diharapkan data yang diperoleh menjadi lebih bermakna dan siap digunakan sebagai perbandingan dalam penelitian. Berikut beberapa proses dalam processing data :

a. Filtering

Operator filter digunakan untuk membuang, mengubah, dan menetapkan batasan pada data dalam aliran kerja atau set data. Sebagai contoh, ketika data mengandung informasi yang tidak dapat dibaca atau tidak relevan, operator filter

dapat digunakan untuk menyingkirkan data tersebut dari hasil akhir. Dengan cara ini, hanya data yang bermutu dan relevan yang akan dimasukkan ke dalam hasil akhir dalam RapidMiner.



Gambar 3. Proses Filtering

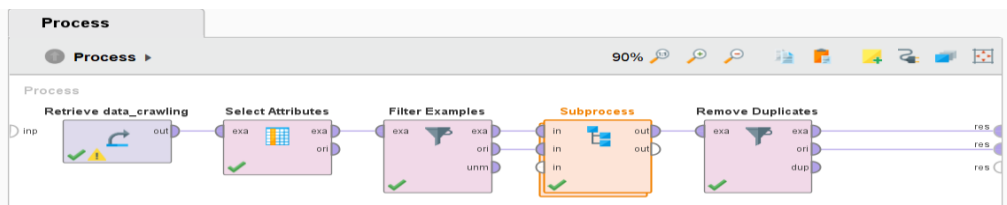
Open in Turbo Prep Auto Model Interactive Analysis Filter (513 / 613 examples): all

Row No.	Text
1	gini amat seketah jadwal bajunya ganjil genap ganjil genap 🙄
2	@beardumps_bcs di Jakardah ada ganjil genap cayank area kampus w tu kawasan ganap jd tidak bisa naik kereta kalo platnya tida...
3	@Cetstial Boleh. Tapi setiap tanggal ganjil dan genap sama gue ya? Adil kan.
4	@ucvolla @sbmptnfess ini kira-kira...kalau mah ada yang membingungkan monggo ditanyakan. Jadi sih sebenarnya penggunaa...
5	buku ganjil genap 50k. dom jkt. bisa oren https://t.co/BHcgsvycip
6	@renjunmn gaki sama kamu tanggal ganjil, sama aku tanggal genap 🙄
7	ini yt music lagi error kah atau punya aku aja ya? lagi enak2 dengerin playlist. eh pas setiap lagu kedua/ genap ke-skip & cuma pute...
8	@collegemansfess Tapi sumpah semester genap lebih berat daripada ganjil asdfghjkl
9	Dr awal dia ini juga salah kelas. Nim genap harusnya ambil kelas genap. Eh malah ambil ganjil. padahal kelas belum penuh. Akhir...
10	buat apa mobil 4 kalo ga bisa dipake gegara ganjil genap
11	@hipsstermaid ganjil genap ya
12	bebas dari ganjil genap. dan banyak keuntungan lainnya.
13	Wts want to sell buku novel fiksi preloved original

Gambar 4. Hasil Filtering

b. Cleansing

Setelah data difilter, langkah berikutnya adalah melakukan proses pembersihan. Dalam tahap ini, kita memanfaatkan operator remove duplicate untuk menghapus entri yang sama, sehingga menghindari hasil yang redundan dalam teks. Selanjutnya, dalam subproses ini, digunakan operator select untuk memilih fitur-fitur yang relevan. Setelah itu, digunakan operator replace untuk menghapus kata-kata seperti "RT", "HTTPS", "@", dan "#", yang dapat dianggap tidak relevan atau mengganggu dalam analisis data.



Gambar 5. Proses Cleansing

Tabel 1. Hasil Cleansing

	Teks Sebelum	Teks Sesudah
Cleansing proses	@AndyHuskyyy Itu udh ada aturan nya ..nah yg kaga ada aturan nya..pas tanggal genap pake plat genap,pas ganjil pake plat ganjil.....🙄jangkrikkk boss....	Itu udh ada aturan nya ..nah yg kaga ada aturan nya..pas tanggal genap pake plat genappas ganjil pake plat ganjil.....jangkrikkk boss....

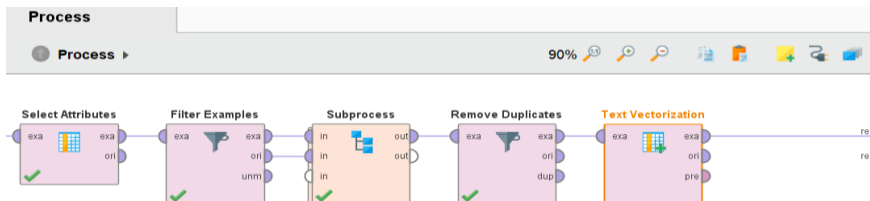
Row No.	Text
50	Pagi Ini Ganjil Genap di 25 Jalan DKI Jakarta Kembali Berlaku
51	Dengan menggunakan mobil listrik mempunyai beberapa keunggulan diantaranya hemat operasional tidak bising bebas d...
52	Sambil hitung wijennya ganjil apa genap
53	bingung sendiri kok bisa ganjil genap gini
54	[cm] galas mau nanyaa boleh ga sil ajuin cuti di tengah2 semester genap kek skrg trus klo boleh nanti itungan masuk ny lag...
55	Jika Ekonomi konfigurasi kursnya 32 A B C D E Jumlah baris 24 dan berhadapan. Nomor tempat duduk yang berhadapan ...
56	Selamat malam dan mohon maaf baru terbalaskan karena padatnya antrean interaksi. KA Tegal Bahari menggunakan rang...
57	1994 1997 tapi genap ganjil howw
58	pattern kecocokan yg kaya kamu bilang tadi tu loo. krn akupun mostly lebih cocok sama yg kelahiran taun ganjil dibanding ...
59	Blar cantik kelihatan punya duit lbh dan mah ada institusi yg jualan nya.
60	min info KA Airlangga 7050 tujuan Surabaya pasar turl yang seat nya hadap depan ke arah maju kereta seat ganjil atau ge...
61	ada keluhan apa disemester genap apa masih dipuncak komedi seperti semester ganjil
62	karena tergantung tahun kelahiran kalo biru itu genap kalo merah itu ganjil

ExampleSet (513 examples, 0 special attributes, 1 regular attribute)

Gambar 6. Hasil Setelah Proses Cleansing

c. Sentiment

Setelah tahap pembersihan data, langkah selanjutnya adalah menganalisis sentimen. Pada tahap ini, digunakan operator text vectorization untuk mengevaluasi respons dari tweet yang dihasilkan oleh masyarakat. Tiga kategori—positif, negatif, dan netral—digunakan dalam analisis sentimen ini untuk membantu mengkategorikan respons dari tweet tersebut. Setelah proses analisis selesai, hasilnya akan dipresentasikan dalam bentuk tabel baru yang dinamakan "sentiment", yang berisi informasi mengenai sentimen dari setiap tweet.



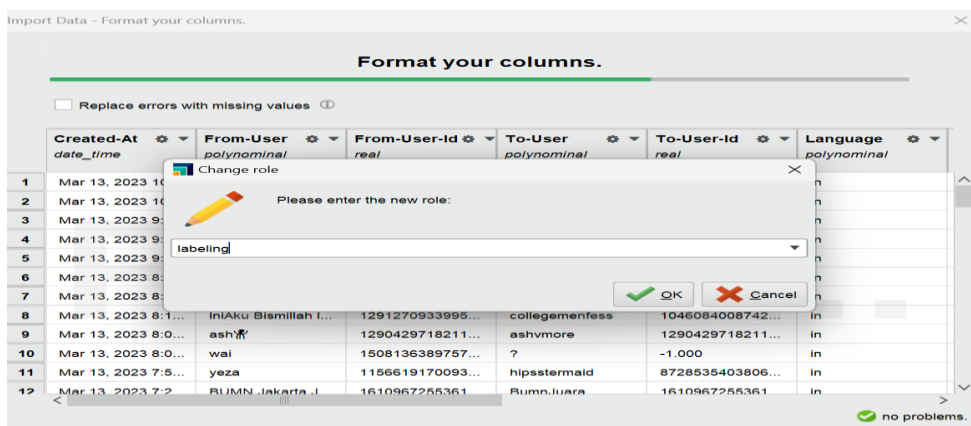
Gambar 7. Proses Sentiment

Tabel 2. Hasil Sentiment

Tweet	Sentiment
daerah jakarta yang kena ganjil genap	Neutral
Polri Tim ETLE Mobile Satlantas Jakarta Utara melaksanakan Penindakan Ganjil Genap di Traffic Light Mangga Dua Jl. Gunung Sahari Pademangan Jakut. Diimbau Bagi Pengguna Jalan agar mematuhi Rambu Lalu Lintas.	Positif
beli mobil anti ganjil genap	negative

d. Label

Setelah selesai tahap analisis sentimen, langkah berikutnya adalah tahap labeling. Pada tahap ini, atribut sentimen diberi label dan dibagi menjadi tiga kategori: positif, negatif, dan netral menggunakan operator Set Role. Ini dimaksudkan untuk memberikan klasifikasi yang jelas terhadap sentimen yang terdapat dalam data, guna mempermudah analisis dan pemahaman terhadap pola-pola yang muncul dari reaksi masyarakat.



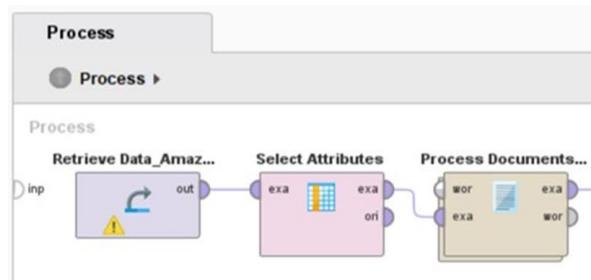
Gambar 8. Proses Labeling

Row No.	sentiment(T...	Text
1	Neutral	gini amat sekolah jadwal bajunya ganjil genap ganjil genap ?
2	Positive	bcs di jakardah ada ganjil genap cayank area kampus w tu kawasan ganap jd tidak bisa naik kereta kalo platnya tidak sesuai
3	Neutral	Boleh. Tapi setiap tanggal ganjil dan genap sama gue ya Adil kan.
4	Neutral	Ini kirakira...kalau msh ada yang membingungkan monggo ditanyakan. Jadi sih sebenarnya penggunaan k di kunci jawaban
5	Neutral	buku ganjil genap 50k dom jkt bisa oren
6	Neutral	gakl sama kamu tanggal ganjil sama aku tanggal genap ??
7	Negative	ini yt music lagi error kah atau nyaku aja ya lagi enak2 dengerin playlist eh pas setiap lagu kedua/ genap keskip & cuma p
8	Neutral	Tapi sumpah semester genap lebih berat daripada ganjil asdfghjkl
9	Neutral	Dr awal dia ini juga salah kelas. Nim genap harusnya ambil kelas genap. Eh malah ambil ganjil padahal kelas belum penuh..
10	Neutral	buat apa mobil 4 kalo ga bisa dipake gegara ganjil genap
11	Neutral	ganjil genap ya
12	?	jadi tunggu setelah sup gunakan mobil listrik mulai mengurangi emisi karbon demi masa depan anak cucu kita

Gambar 9. Hasil Labeling

e. Tokenizing

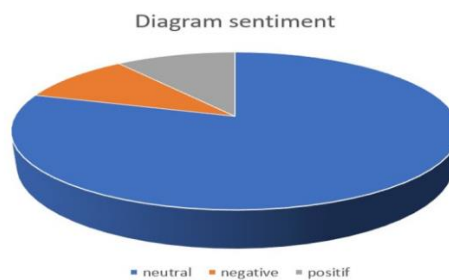
Setelah melalui langkah-langkah sebelumnya seperti penyaringan, pembersihan, analisis sentimen, dan pelabelan, langkah selanjutnya adalah tokenisasi. Tokenisasi bertujuan untuk memproses teks atau kalimat dengan cara membaginya menjadi frasa-frasa kata yang terpisah.



Gambar 10. Tokenizing

3.3 Hasil Pie Diagram

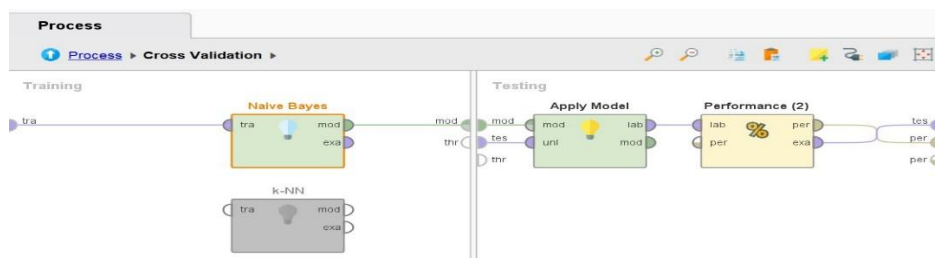
Diagram lingkaran digunakan untuk menampilkan data sentimen dari hasil yang dihasilkan oleh algoritma K-Nearest Neighbor (K-NN) dan Naïve Bayes. Data menunjukkan 385 poin data dengan sentimen netral, 50 poin data dengan sentimen negatif, dan 48 poin data dengan sentimen positif.



Gambar 11. Diagram Sentiment

3.4 Algoritma Naïve Bayes

Proses di mana data yang telah diproses kemudian diuji menggunakan algoritma Naïve Bayes untuk menilai akurasi baik dari algoritma Naïve Bayes maupun algoritma K-Nearest Neighbor.



Gambar 12. Proses uji data

Kinerja algoritma Naïve Bayes dihitung dan dipantau secara otomatis menggunakan operator Performance. Di sini, tujuannya adalah untuk menghitung tingkat akurasi algoritma Naïve Bayes, seperti yang terlihat pada Contoh gambar 12 Hasil Akurasi Algoritma Naïve Bayes.

Table View Plot View

accuracy: 79.72% +/- 1.20% (micro average: 79.71%)

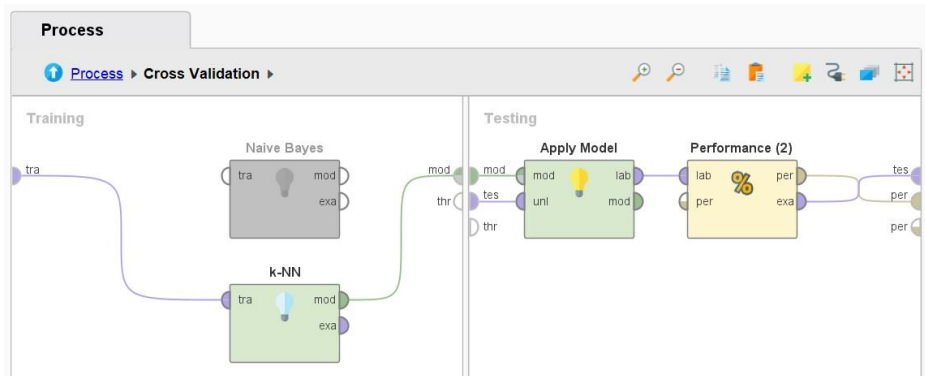
	true Neutral	true Positive	true Negative	class precision
pred. Neutral	385	48	50	79.71%
pred. Positive	0	0	0	0.00%
pred. Negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 13. Naïve Bayes Algorithm Accuracy Results.

Naive Bayes Algorithm Accuracy Results. adalah 79,72%, dengan tingkat kesalahan klasifikasi sebesar 20,28%. Data sentimen yang terklasifikasi meliputi 385 netral, 50 negatif, dan 48 positif.

3.5 Algoritma K-Nearest Neighbor

Proses di mana algoritma K-Nearest Neighbor digunakan untuk menguji data yang telah diproses dalam Performance dan menilai akurasi dari kedua algoritma, yaitu K-Nearest Neighbor dan Naive Bayes.



Gambar 14. Proses uji data

Kinerja algoritma K-Nearest Neighbor (K-NN) dihitung dan dipantau secara otomatis menggunakan operator Performance. Tujuannya adalah untuk menilai tingkat akurasi algoritma K-Nearest Neighbor (K-NN). Sebagai contoh, Gambar 13 menampilkan hasil akurasi dari algoritma K-Nearest Neighbor (K-NN).

Table View
 Plot View

accuracy: 79.72% +/- 1.20% (micro average: 79.71%)

	true Neutral	true Positive	true Negative	class precision
pred. Neutral	385	48	50	79.71%
pred. Positive	0	0	0	0.00%
pred. Negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 15. K-Nearest Neighbor (K-NN) Algorithm Accuracy Results.

Algoritma K-Nearest Neighbor memiliki tingkat kesalahan klasifikasi sebesar 20,28% dan rata-rata akurasi sebesar 79,72%. Data yang terklasifikasi mencakup 48 poin data dengan sentimen positif, 50 poin data dengan sentimen negatif, dan 385 poin data dengan sentimen netral.

4. KESIMPULAN

Hasil analisis sentimen terhadap tweet menunjukkan bahwa mayoritas dari tweet-tweet yang dianalisis memiliki sentimen netral, dengan jumlah mencapai 391 tweet, diikuti oleh tweet-tweet yang mengandung sentimen positif sebanyak 59 tweet, dan sentimen negatif hanya terdapat pada 50 tweet. Evaluasi dua algoritma klasifikasi, Dengan kata lain, rata-rata akurasi sebesar 79,72% yang dicapai oleh K-Nearest Neighbor dan Naive Bayes hampir sama. Singkatnya, Hasil studi ini menunjukkan bahwa tidak ada perbedaan yang mencolok antara Naive Bayes dan K-Nearest Neighbor dalam hal akurasi kategorisasi sentimen tweet. Perlu diingat bahwa selalu ada variabel lain yang mungkin mempengaruhi hasil, dan penelitian lanjutan mungkin diperlukan untuk mendalaminya lebih lanjut.

REFERENCES

- [1] N. Sucahyo, L. Nurlaela, and R. R. Waryono, "Analisis Sentimen Masyarakat Jakarta Terhadap Kebijakan Perluasan Dan Perpanjangan Ganjil Genap di Media Sosial Twitter," J. Teknol. Inform. dan Komput., vol. 7, no. 1, pp. 97–111, 2021, doi: 10.37012/jtik.v7i1.506.
- [2] F. A. Rohmansyah, B. Bintoro, and I. Santoso, "Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode K-Nearest Neighbor (Knn)," iKRAITH-INFORMATIKA, vol. 7, no. 2, pp. 165–169, 2022.
- [3] H. Rachmi, S. Suparni, and A. Al Kaafi, "Analisis Sentimen Sistem Ganjil Genap Kota Bogor," J. ELTIKOM, vol. 5, no. 2, pp. 92–99, 2021, doi: 10.31961/eltikom.v5i2.429.
- [4] R. Amelia, D. Darmansah, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 2, p. 338,

- 2022, doi: 10.30865/jurikom.v9i2.3895.
- [5] R. Mubarak, "Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Sosial Berskala Besar (Psbb) Dengan Metode ...," *J. Siliwangi Seri Sains dan Teknol.*, vol. 7, no. 1, pp. 19–24, 2021, [Online]. Available: <http://jurnal.unsil.ac.id/index.php/jssainstek/article/view/3726>
- [6] W. P. Anggraini and M. S. Utami, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia," *Fakt. Exacta*, vol. 13, no. 4, p. 255, 2021, doi: 10.30998/faktorexacta.v13i4.7964.
- [7] R. Pinka et al., "Sentiment Analysis of Public Opinions on the Effectiveness of Online Learning Using Naïve Bayer Algorithm," *J. Inf. Syst. Informatics Comput.*, vol. 6, no. 1, pp. 273–279, 2022, doi: 10.52362/jisicom.v6i1.822.
- [8] F. D. Samuel, P. D. Atika, and S. Setiawati, "Analisis Sentimen Masyarakat Terhadap Perkuliahan Daring Di Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *J. Students' Res. Comput. Sci.*, vol. 4, no. 2, pp. 261–272, 2023, doi: 10.31599/jsrscs.v4i2.3253.
- [9] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- [10] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [11] H. Faisal, A. Febriandirza, and F. N. Hasan, "Analisis Sentimen Terkait Ulasan Pada Aplikasi PLN Mobile Menggunakan Metode Support Vector Machine," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 5, no. 1, pp. 303–312, 2024.
- [12] M. mahrus Zain, "Analisis Sentimen Pendapat Masyarakat Mengenai Vaksin Covid-19 Pada Media Sosial Twitter dengan Robustly Optimized BERT Pretraining Approach," *J. Komput. Terap.*, vol. 7, no. 2, pp. 280–289, 2021, doi: 10.35143/jkt.v7i2.4782.
- [13] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.
- [14] D. Surya Sayogo, B. Irawan, and A. Bahtiar, "Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3314–3319, 2024, doi: 10.36040/jati.v7i6.8178.
- [15] Ernianti Hasibuan and Elmo Alistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [16] D. F. Salsabillah, D. E. Ratnawati, and N. Y. Setiawan, "Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma Support Vector Machine dengan Naive bayes (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 107–116, 2024, doi: 10.25126/jtiik.20241117584.
- [17] N. R. Siahaan, R. Y. Tiffany, and S. R. E. Sinaga, "Analisis Sentimen Ulasan Aplikasi Media Sosial Whatsapp Menggunakan Metode Naive Bayes Classifier," *J. Ilm. Betrik*, vol. 14, no. 02, pp. 343–354, 2023, [Online]. Available: <https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/104%0Ahttps://ejournal.pppmitpa.or.id/index.php/betrik/article/download/104/76>
- [18] M. A. A. A. Solichin, "Analisis Sentimen MotoGP Mandalika Pada Twitter Menggunakan Metode Naive Bayes," *J. Ticom Technol. Inf. Commun.*, vol. 11, no. Vol 11 No 1 (2022): Jurnal Ticom: Technology of Information and Communication, pp. 20–25, 2022, [Online]. Available: <https://jurnal-ticom.jakarta.aptikom.or.id/index.php/Ticom/article/view/66/55>
- [19] Q. A. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
- [20] L. Legito et al., "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 324–330, 2023, doi: 10.57152/malcom.v3i2.893.
- [21] A. H. Yuanti, "Analisis Pengaruh Covid-19 Terhadap Kesehatan Mental dengan Visualisasi Data Rapidminer," *Gudang J. Multidisiplin Ilmu*, vol. 2, pp. 183–187, 2024.