

Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes

Johanes Florensus Sianipar*, Yudhi Raymond Ramadhan, Irsan Jaelani

Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakarta, Indonesia

Email: ^{1,*}johanesflorensus46@wastukencana.ac.id, ²yudhi.raymond@wastukencana.ac.id, ³irsan@wastukencana.ac.id

Email Penulis Korespondensi: johanesflorensus46@wastukencana.ac.id

Abstrak—Perkembangan dunia teknologi dan komunikasi yang pesat tidak akan lepas dari penyedia layanan web yang selalu menyediakan informasi yang beragam. Salah satu contohnya adalah data text yang diambil dari twitter. Twitter adalah sebuah sosial media dan layanan micro blogging yang mengizinkan penggunaannya untuk mengirimkan pesan realtime. Pesan ini populer dengan sebutan tweet, dengan menggunakannya, pengguna akan mudah untuk mengikuti tren, cerita, informasi dan berita dari seluruh penjuru dunia. Permasalahan mengenai pembangunan kereta cepat Jakarta-Bandung (KCJB) yang membutuhkan dana yang besar, belum lagi mengenai pemindahan rute yang awalnya Jakarta-Surabaya menjadi Jakarta-Bandung yang mengakibatkan pembekakan dana yang besar. Sehingga menimbulkan suatu permasalahan seperti perbedaan pendapat dan pernyataan setuju dan tidak setujunya pembangunan kereta cepat Jakarta-Bandung. Untuk metode yang digunakan yaitu Naive Bayes karena mempunyai nilai probabilitas atau peluang tinggi untuk pengklasifikasian data, untuk pembobotan menggunakan perhitungan TF-IDF, dan pengujian data menggunakan confusion matrix. Solusi untuk permasalahannya yaitu dengan analisis sentimen, mengelompokkan sekaligus membantu data dari memprediksi data tweet tersebut. Tujuan penelitian ini untuk menganalisis sentimen pengguna twitter terhadap pembangunan kereta cepat Jakarta-Bandung dengan menggunakan metode *naive bayes* dengan menghitung nilai dari data tweet KCJB yang memiliki jumlah data 2390 data yang setelah melalui proses cleaning menjadi 2007 data. Setelah melalui pemrosesan dengan hasil sentimen negatif sebanyak 673, hasil sentimen positif sebanyak 668, dan hasil sentimen netral sebanyak 665 hasil *accuracy* 71%, *precision* 73%, *recall* 89%. Dari hasil penelitian mendapatkan hasil tanggapan tergolong negatif terhadap pembangunan kereta cepat Jakarta-Bandung.

Kata Kunci: KCJB; Naive Bayes; Analisis Sentimen; Data Text, Twitter

Abstract—The rapid development of the world of technology and communication cannot be separated from web service providers who always provide various information. One example is text data taken from Twitter. Twitter is a social media and micro blogging service that allows its users to post realtime messages. This message is popularly known as a tweet, by using it, users will find it easy to follow trends, stories, information and news from all corners of the world. Problems regarding the construction of the Jakarta-Bandung fast train (KCJB) which required large funds, not to mention the transfer of the route from Jakarta-Surabaya to Jakarta-Bandung which resulted in a large increase in funds. So that it raises a problem such as differences of opinion and statements of approval and disapproval of the construction of the Jakarta-Bandung fast train. The method used is Naive Bayes because it has a high probability or opportunity value for classifying data, for weighting using TF-IDF calculations, and testing data using a confusion matrix. The solution to the problem is sentiment analysis, grouping and at the same time helping data from predicting the tweet data. The purpose of this study is to analyze the sentiments of Twitter users towards the construction of the Jakarta-Bandung fast train using the Naive Bayes method by calculating the value of the KCJB tweet data which has a total of 2390 data which after going through the cleaning process becomes 2007 data. After processing with 673 negative sentiment results, 668 positive sentiment results, and 665 neutral sentiment results, 71% accuracy, 73% precision, recall.

Keywords: KCJB; Naive Bayes; Sentiment Analysis; Text Data; Twitter

1. PENDAHULUAN

Perkembangan dunia teknologi dan komunikasi yang pesat tidak akan lepas dari penyedia layanan web yang selalu menyediakan informasi yang beragam. Salah satu contohnya adalah data text yang diambil dari Twitter [1]. Sebagai media komunikasi masyarakat, jejaring sosial semakin populer digunakan. Salah satu jejaring sosial yang cukup populer yaitu Twitter. Twitter dapat dimanfaatkan sebagai saran promosi produk, iklan, kampanye politik maupun sebagai sarana menyampaikan pendapat terkait kritik, saran, isu-isu dan opini publik [2]. Transportasi adalah salah satu faktor utama yang harus di mulai oleh kota-kota besar. Kereta api adalah salah satu bentuk transportasi yang sangat penting bagi negara berkembang seperti Indonesia karena kemampuannya untuk mengangkut beban dengan kapasitas besar, seperti mengangkut orang dan barang untuk mengatasi masalah lalu lintas dan meningkatkan pembangunan ekonomi [3]. Tahun 2008, Bappenas bersama Kementerian Perhubungan telah merencanakan pembangunan kereta cepat yang dapat melayani penumpang dari Jakarta menuju Surabaya. Namun akibat terkendala dana dan hal lainnya, rencana tersebut mengalami perubahan yaitu rute yang berubah menjadi Jakarta menuju Bandung. Saat ini kereta cepat Jakarta-Bandung atau disebut juga dengan High Speed Railway Jakarta-Bandung (HSRJB) sedang dalam proses pembangunan [4]. Dibalik kemegahan dari proyek Kereta Cepat Jakarta-Bandung (KCJB) ini, tentu tidak luput dan tidak pernah berhenti menimbulkan kontroversi. Beberapa pihak berpendapat bahwa KCJB belum dibutuhkan di tengah upaya pemerintah untuk mengembangkan perkeretaapian nasional di wilayah Sumatera, Kalimantan, Sulawesi dan wilayah lainnya. Namun, pemerintah tampaknya menaruh perhatian yang sangat serius sehingga proyek ini dapat berlangsung sangat cepat, terutama menyangkut perizinan. Salah satunya adalah proses perizinan lingkungan, izin dari KCJB yang tampaknya terburu-buru dan tidak selaras dengan RTRW hanya dapat menghasilkan lisensi yang dikeluarkan tidak sesuai dengan ketentuan perundang-undangan sehingga menimbulkan kekhawatiran merusak rezim perizinan lingkungan yang merupakan instrumen penting dalam melindungi lingkungan hidup guna mendukung pembangunan berkelanjutan [5].

Twitter adalah sebuah sosial media dan layanan *micro blogging* yang mengizinkan penggunanya untuk mengirimkan pesan realtime. Pesan ini populer dengan sebutan tweet, dengan menggunakannya, pengguna akan mudah untuk mengikuti tren, cerita, informasi dan berita dari seluruh penjuru dunia [6].

Analisis sentimen merupakan sebuah pemrosesan bahasa alami yang mampu mendeteksi opini dan emosi seseorang terhadap suatu hal atau topik tertentu. Analisis sentimen melakukan pengolahan pada data untuk mengetahui kecenderungan opini dari sebuah objek yang nantinya mengarah pada kelas positif dan negatif [7].

Algoritma Naive Bayes adalah sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang di berikan [8].

Python merupakan salah satu dari bahasa pemrograman yang sering digunakan oleh programmer atau pembuat program dalam membuat program mereka. Python memiliki karakteristik sintaks yang tidak terlalu rumit. Sehingga Python menjadi salah satu bahasa pemrograman tingkat tinggi yang mudah untuk digunakan. Dalam menulis sebuah kode program menggunakan bahasa pemrograman Python, terdapat beberapa aturan yang harus di penuhi. Hal ini untuk mengantisipasi terjadinya error atau masalah pada program yang dibuat. Aturan sintaks Python yang pertama adalah dalam penulisan statement atau perintah [9].

Ada beberapa artikel yang berkaitan dengan penelitian ini seperti dilakukan oleh Sitti Nurul, Fitriyah, N Safriadi dengan judul “Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes” menggunakan algoritma naive bayes melalui perhitungan skor sentimen. Algoritma naive bayes yang di dapatkan nilai akurasi yaitu 64,6% untuk paslon 01, dan 58% untuk paslon 02 pada pengujian 3 kelas. Sedangkan pengujian 2 kelas 77,7% untuk paslon 01 dan 84% untuk paslon 02 [10].

Text Mining adalah proses inovasi akan informasi atau isu terkini yang sebelumnya tidak terungkap menggunakan mekanisme dan menganalisa data pada jumlah besar. Dalam Menganalisa keseluruhan atau sebagian unstructured text, text mining membuat asosiasi satu bagian text dengan lainnya berdasarkan aturan tertentu. Hasil yang diharapkan adalah kata baru yang tidak terungkap jelas sebelumnya [11].

Pada artikel ini dengan judul “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier” dan penulis data yang adalah komentar dari twitter yang membahas jasa transportasi online dan penerapan algoritma naive bayes dalam pengelompokkan sentimen positif dan negatif. Berdasarkan data ulasan komentar jasa transportasi online yang mendapatkan kesimpulan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil menunjukkan tingkat sentimen positif lebih besar dibandingkan dengan tingkat sentimen negatif. Dengan hasil penelitian ini bisa dilihat bahwa banyak masyarakat yang merasa puas dengan layanan jasa transportasi online di Indonesia berdasarkan data dari Twitter [12].

Pada artikel lainnya yang dilakukan dengan judul “Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naive Bayes Berbasis PSO” menggunakan metode naive bayes dan model Particle Swarm Optimization (PSO) dengan mengolah data di Rapidminer dan perhitungan manual memiliki perbedaan nilai accuracy, precision dan recall, pada naive bayes nilai accuracy untuk perusahaan jasa ekspedisi Anteraja yaitu memiliki accuracy 87,77%, precision 76,67%, recall 52,27%. JNE dengan accuracy 81,48%, precision 71,43%, dan recall 62,50%. JNT dengan accuracy 91,46%, precision 48,15%, recall 86,67%. Pada model Particle Swarm Optimization (PSO) perusahaan Anteraja memiliki accuracy 91,70%, precision 82,05%, recall 72,73%. JNE memiliki accuracy 93,83%, precision 88,00%, recall 91,67%. JNT memiliki accuracy 92,18%, precision 70,97%, recall 81,48% [13].

Hasil dari artikel yang berkaitan dengan penelitian ini menunjukkan bahwa algoritma *naive bayes* dapat mengklasifikasikan data dan mendapat hasil akurasi yang baik. Berdasarkan permasalahan yang terjadi bahwa peneliti ingin melakukan cara memanfaatkan pendapat pengguna Twitter terhadap pembangunan kereta cepat jakarta-bandung untuk analisis sentimen. Tujuan penelitian ini yaitu untuk mengetahui menganalisis sentimen pengguna Twitter terhadap pembangunan kereta cepat jakarta-bandung dengan menggunakan metode naive bayes.

2. METODOLOGI PENELITIAN

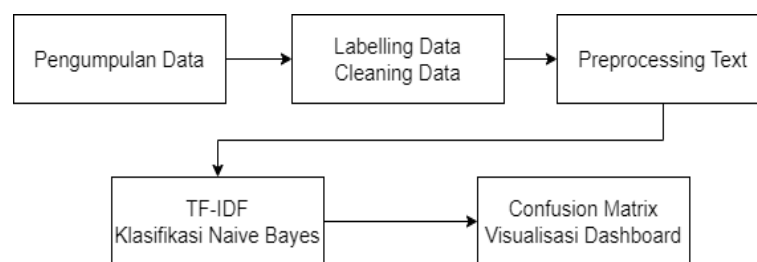
2.1 Tahapan Penelitian

Tahapan penelitian merupakan susunan kerja dari penelitian ini secara sistematis agar mencapai tujuan yang diinginkan. Adapun metodologi Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [14]. Metodologi penelitian ini yang dapat dilihat sebagai berikut:

2.2 Pengumpulan Data

Scrapping bisa digunakan dengan beberapa metode dan teknik, *Scrapping* juga dapat mengubah data yang tidak terstruktur menjadi terstruktur sehingga dapat disimpan kedalam database. Inti dari *scrapping* ini adalah mengumpulkan data, menyimpan data dan memvalidasi data, sehingga dapat diinterpretasikan menjadi informasi yang baik [15].

2.3 Labelling Data



Gambar 1. Metodologi Penelitian

Sentiment analysis atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang. Sentiment analysis dapat diklasifikasikan kedalam kelas sentimen bersifat positif, negatif dan netral [16]. Setelah proses labelling, data akan dicleaning merupakan data tweet yang duplikat secara manual menggunakan aplikasi microsoft excel.

2.4 Preprocessing Text

Preprocessing adalah teknik maupun strategi yang bertujuan untuk membuat suatu data lebih mudah untuk dikelola atau cocok untuk digunakan pada text mining yang tentunya bertujuan agar meningkatkan hasil dari analisis text mining [17] dengan melakukan *Transformation*, *Tokenization*, *Filtering*.

2.4.1 Transformation

Pertama yang dilakukan *preprocessing text* adalah *transformation* merupakan proses mengubah *input* untuk transformasi huruf kecil secara default. *Case folding* berfungsi untuk mengubah semua teks menjadi huruf kecil

2.4.2 Tokenization

Operasi memisahkan teks menjadi potongan-potongan berupa token, bisa berupa potongan huruf, kata, atau kalimat, sebelum dianalisis lebih lanjut. Entitas yang bisa disebut sebagai token misalnya kata, angka, simbol, tanda baca, dan lain sebagainya [18].

2.4.1 Filtering

Tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting) yang terdiri dari normalisasi bahasa, *stopword*, *stemming* [19].

2.5 TF-IDF

Klasifikasi terdapat beberapa tahap yaitu melakukan TF-IDF dan melakukan klasifikasi *naive bayes*. TF-IDF bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapan ini terdapat dua bagian proses yaitu TF (Term Frequency) dan IDF (Inverse Document Frequency), TF adalah jumlah kemunculan tiap kata pada sebuah dokumen semakin banyak kata muncul pada tiap dokumen maka semakin besar nilai TF, IDF adalah jumlah nilai dokumen pada tiap kata yang berbanding terbalik yaitu apabila suatu kata jarang muncul pada sebuah dokumen maka nilai IDF lebih besar daripada kata yang sering muncul. Hasil dari pembobotan kata dengan TF-IDF ini adalah perkalian dari nilai TF dan IDF yang akan menghasilkan bobot lebih kecil apabila kata tersebut sering muncul pada setiap dokumen dalam koleksi, sebaliknya bobot TF-IDF akan lebih besar apabila kata tersebut jarang muncul pada setiap dokumen dalam koleksi[20].

$$IDF(\text{Word}) = \log \frac{tf}{df} \quad (1)$$

Keterangan :

IDF = Inverse Document Frequency

tf = Term Frequency

df = Document Frequency

Naive Bayes adalah sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang di berikan

$$P(X|H) = \frac{P(X|H) \times P(H)}{P(X)} \quad (2)$$

Keterangan :

X = Data Sample dengan kelas (label) yang tidak diketahui.

H = Hipotesa bahwa X adalah data dengan kelas (label) C.

P(H|X) = Peluang bahwa hipotesa benar (valid) untuk data sampel X yang diamati.

P(X|H) = Peluang data sampel X, bila diasumsikan bahwa hipotesa benar (valid).

P(H) = Peluang dari hipotesa H.

P(X) = Peluang data sampel yang diamati.

2.6 Naive Bayes

Tahap klasifikasi ini melakukan terhadap dokumen yang mengandung kategori tertentu, kata-kata dalam dokumen tersebut yang dapat dilakukan pengklasifikasian, sekumpulan dari kata pada dokumen tersebut dapat digunakan untuk menentukan kategori dikarenakan memiliki nilai makna tertentu. Untuk menentukan sentimen ini dari dokumen *tweet* yaitu menggunakan metode *Naive Bayes*.

2.7 Evaluasi

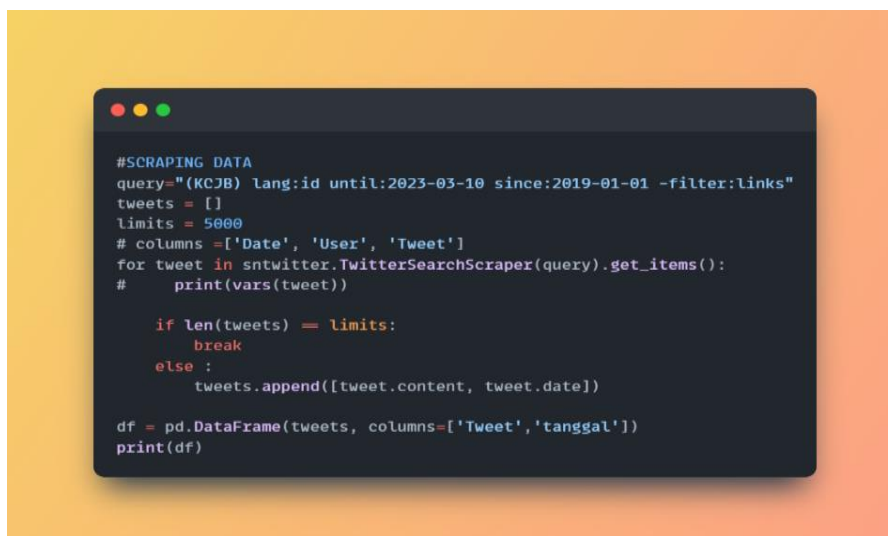
Klasifikasi terdapat tahap untuk melakukan evaluasi terdiri dari *confusion matrix* dan visualisasi *dashboard*. *Confusion matrix* adalah tabel yang berguna untuk dapat menentukan seberapa akurat proses klasiifikasi yang dilakukan [21]. Visualisasi dashboard dilakukan untuk melihat data yang sudah diolah berupa grafik.

3. HASIL DAN PEMBAHASAN

Analisis sentimen yang terdapat pada data twitter melalui beberapa tahapan pengolahan data, berikut tahapan dalam mengolah data dan hasil.

3.1 Pengumpulan Data

Pengambilan data terhadap twitter menggunakan *scrapping* dan menggunakan bahasa pemrograman python dengan *module snsrape*, dan pengambilan data Twitter di ambil pada tahun 2020 sampai tahun 2023. Pada proses *scrapping* data ini dilakukan dengan *query* KCJB. Dapat dilihat pada gambar 2

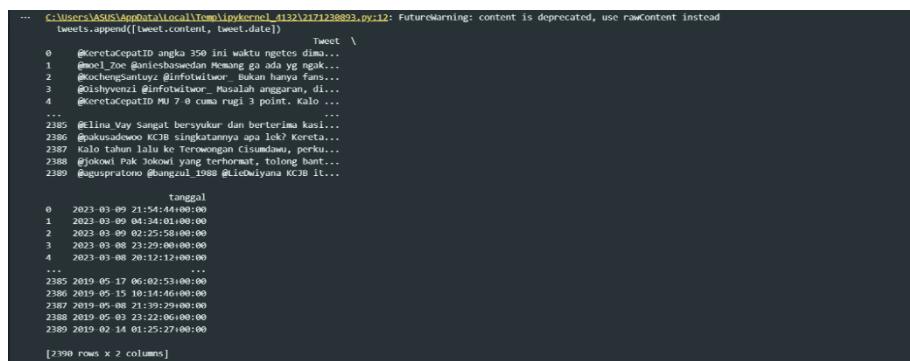


```
#SCRAPING DATA
query="(KCJB) lang:id until:2023-03-10 since:2019-01-01 -filter:links"
tweets = []
limits = 5000
# columns =['Date', 'User', 'Tweet']
for tweet in sntwitter.TwitterSearchScrapper(query).get_items():
    print(vars(tweet))
    #
    if len(tweets) == limits:
        break
    else :
        tweets.append([tweet.content, tweet.date])

df = pd.DataFrame(tweets, columns=['Tweet','tanggal'])
print(df)
```

Gambar 1. Scrapping Data

Peneliti berhasil mengambil 2390 data, seperti pada gambar 3. Hasil *scrapping* data twitter.



```
FutureWarning: content is deprecated, use rawContent instead
tweets.append([tweet.content, tweet.date])
Tweet \
0 @keretaCepatID angka 300 ini waktu ngetos dima...
1 @weel_zoe @anishabandan Penang ga ada yg agak...
2 @kohergsantuyz @infotwitwor Bukan hanya Fars...
3 @oishyvenzi @infotwitwor Masalah anggaran, di...
4 @keretaCepatID MU 7-0 cuma rugi 3 point. Kalo ...
...
2385 @Lina.Vay Sangat bersyukur dan berterima kasi...
2386 @pakusadewo KC3B singkatannya apa lek? Kereta...
2387 Kalo tahun lalu ke Terowongan Cisumdawa, perku...
2388 @jakowi Pak Jokowi yang terhormat, tolong bant...
2389 @agusprutoro @Bangsul_1988 @Lidwiyana KC3B It...
...
tanggal
0 2023-03-09 21:54:44+00:00
1 2023-03-09 04:34:01+00:00
2 2023-03-09 02:25:58+00:00
3 2023-03-08 23:29:00+00:00
4 2023-03-08 20:12:12+00:00
...
2385 2019-05-17 06:02:53+00:00
2386 2019-05-15 10:14:46+00:00
2387 2019-05-08 21:39:29+00:00
2388 2019-05-03 23:22:06+00:00
2389 2019-02-14 01:25:27+00:00

[2390 rows x 2 columns]
```

Gambar 2. Hasil Scrapping Data Twitter

3.2 Labelling Data

Pada tahap *labelling* data yang sudah dapat menggunakan tahap *scrapping* lalu akan dilakukan pelabelan data. Pelabelan data dilakukan dengan manual langsung menggunakan aplikasi microsoft excel dengan melihat 1 persatu data tweet. Lalu memberi label dari angka 0 adalah netral, 1 adalah positif dan, 2 adalah negatif. Berikut sampel labelling yang dapat dilihat pada tabel 1.

Tabel 1. Proses Labelling

Tweet	Tanggal	Label
@KeretaCepatID Bagaimana anggaran KCJB bisa membengkak 2x? Perencanaanya amburadul ya? Merengek minta bantuan APBN pula.	08/03/2023	2
@naninunano7 @kadrunmampos Dan kalian dukung CHINA? NGACA!!!! Arab kasih HIBAH masjid yg dibanggakan JKW! CHINA kasih tambahan UTANG BENGGAK KCJB!	07/03/2023	2

3.3 Preprocessing Text

Preprocessing Text membagi teks menjadi komponen-komponen yang lebih kecil seperti *transformation*, *tokenization*, *filtering*. Dalam preprocessing text data penelitian ini akan melalui tahap-tahap berikut.

3.3.1 Transformation

Tahap *Transformation* melakukan *case Folding*, yang akan mengubah semua teks menjadi huruf kecil, menghapus angka, dan menghapus tanda baca. Berikut sampel *case folding* yang dapat dilihat pada tabel 2.

Tabel 2.Proses Case Folding

Sebelum	Sesudah
Menurut Jodi ke-4 fokus tersebut sudah mendapat perhatian Menko Luhut sejak dilibatkan dalam Proyek KCJB yang dijalankan KCIC pada November 2019 Ketua Komite Kereta Cepat #LBPTerbuktiMumpuni	menurut jodi ke fokus tersebut sudah mendapat perhatian menko luhut sejak dilibatkan dalam proyek kcjb yang dijalankan kcic pada november ketua komite kereta cepat lbpterbuktimumpuni

3.3.2Tokenization

Teks yang sudah di transformation lalu memisahkan teks berupa potongan-potongan berupa token yang berupa potongan huruf, kata, atau kalimat. Berikut sampel yang dapat dilihat pada tabel 3.

Tabel 3. Proses Tokenization

yang	komen	miring	biasanya	gak
punya	duit	buat	naek	kcjb

3.3.3 Filtering

Tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpang kata penting). Pada penelitian ini penulis menggunakan *filtering stopwords* dan *stemming* yang membantu penulis dalam proses filtering. Pada tabel 4. Sampel data *stopword* sebagai berikut.

Tabel 4. Proses Stopword

Sebelum	Sesudah
Layanan tersebut adalah LRT Jabodebek dan KA Feeder KCJB, hadirnya kereta api ini akan mempermudah konektivitas masyarakat untuk menuju stasiun KCJB melalui sistem transportasi yang bebas kemacetan dan tepat waktu.	layan lrt jabodebek ka feeder kcjb hadir kereta api mudah konektivitas masyarakat stasiun kcjb sistem transportasi bebas macet

Proses *Stemming* mengubah kata imbuhan menjadi kata dasar. Berikut sampel *stemming* yang dapat dilihat pada tabel 5.

Tabel 5. Data sesudah Stemming

Sebelum	Sesudah
jodi fokus perhatian menko luhut dilibatkan proyek kcjb dijalankan kcic november ketua komite kereta cepat lbpterbuktimumpuni	jodi fokus perhati menko luhut libat proyek kcjb jalan kcic november ketua komite kereta cepat lbpterbuktimumpuni

3.4 Naive Bayes

Data tabel dengan format Excel yang diberi labelling selanjutnya melakukan perhitungan manual *term frequency* (tf) mengambil data sampel untuk *training* pada tabel 6.

Tabel 6. Data Training

Data Tweet
jadi nyaman aman prioritas kcjb

Berdasarkan tabel diatas probabilitas pada data tweet yang menggunakan *term frequency* mendapatkan 5 kata yang terdapat pada tabel 7.

Tabel 7. Kemunculan Term Frequency

No	Kata	Positif	Negatif	Netral
1	jadi	1	0	0
2	nyaman	1	0	0
3	aman	1	0	0
4	prioritas	1	0	0
5	kcjb	1	0	0
		5	0	0

Pada proses klasifikasi selanjutnya akan dilakukan perhitungan jumlah dari kemunculan kata unik dan jumlah pada setiap kategori. Yang terdapat pada tabel 8.

Tabel 7. Kemunculan Kata Unik

Kategori	Jumlah dokumen di setiap kategori	jumlah dokumen seluruhnya	P(Kategori)	Jumlah Kata Setiap Kategori
positif	4	12	0,33	43
negatif	4	12	0,33	29
netral	4	12	0,33	22
Jumlah kata unik			68	

Kata unik yang merupakan kata sering muncul disetiap dokumen. Setelah mendapatkan jumlah kata unik dan jumlah kata pada setiap kategori, yang dilakukan selanjutnya adalah menghitung probabilitas pada setiap kategori

$$P(\text{Kata}) = \frac{\text{jumlah dari satu kata di satu kategori} + 1}{\text{jumlah kata tiap kategori} + \text{jumlah keseluruhan kata unik}} \quad (3)$$

Setelah didapatkan jumlah kata unik dan jumlah kata pada setiap kategori, selanjutnya adalah menghitung probabilitas pada setiap kategori.

$$(kcjb | \text{positif}) \frac{4+1}{43+68} = \frac{5}{111} = 0,0450 \quad (4)$$

$$(kcjb | \text{negatif}) \frac{4+1}{29+68} = \frac{5}{97} = 0,0515 \quad (5)$$

$$(kcjb | \text{netral}) \frac{4+1}{22+68} = \frac{5}{90} = 0,0555 \quad (6)$$

Hasil perhitungan probabilitas dengan sampel kata kcjb mendapat kesimpulan nilai probabilitas tertinggi yaitu netral.

3.5 Evaluasi

Pada tahap evaluasi *confusion matrix* di python dengan algoritma *naive bayes* mengambil data sampel tweet menghasilkan nilai *accuracy* 71%, *precision* 73%, *recall* 89% dijabarkan pada gambar 4.

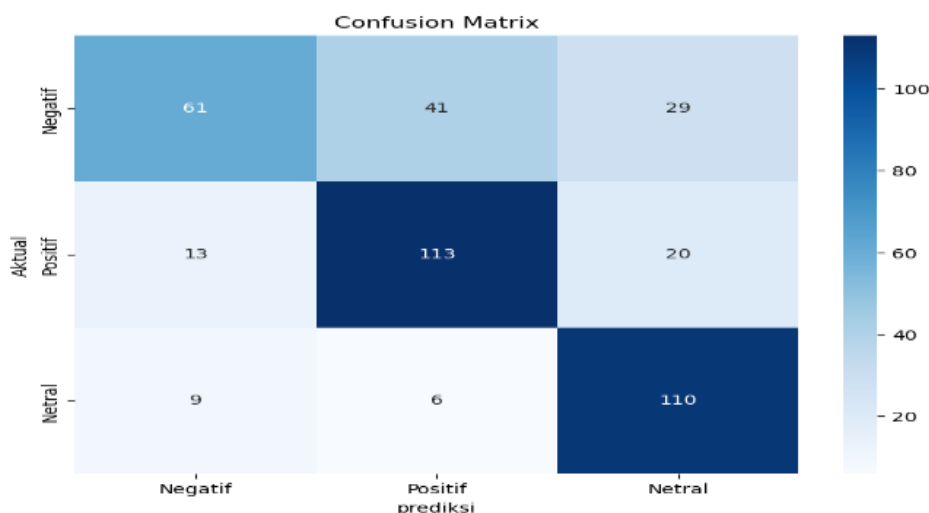
```

... prediksi benar: 284 data
prediksi salah: 118 data
Akurasi Algoritme: 70.64676616915423 %
TRUE NEGATIVE (TN): 61
FALSE NEGATIVE (FN): 13
TRUE POSITIVE (TP): 113
FALSE POSITIVE (FP): 41
PRECISION: 73.37662337662337 %
RECALL: 89.68253968253968 %
      precision  recall  f1-score  support
0         0.73    0.47    0.57    131
1         0.71    0.77    0.74    146
2         0.69    0.88    0.77    125

accuracy          0.71    402
macro avg         0.71    0.71    0.69    402
weighted avg      0.71    0.71    0.69    402
    
```

Gambar 3. Hasil Pengujian

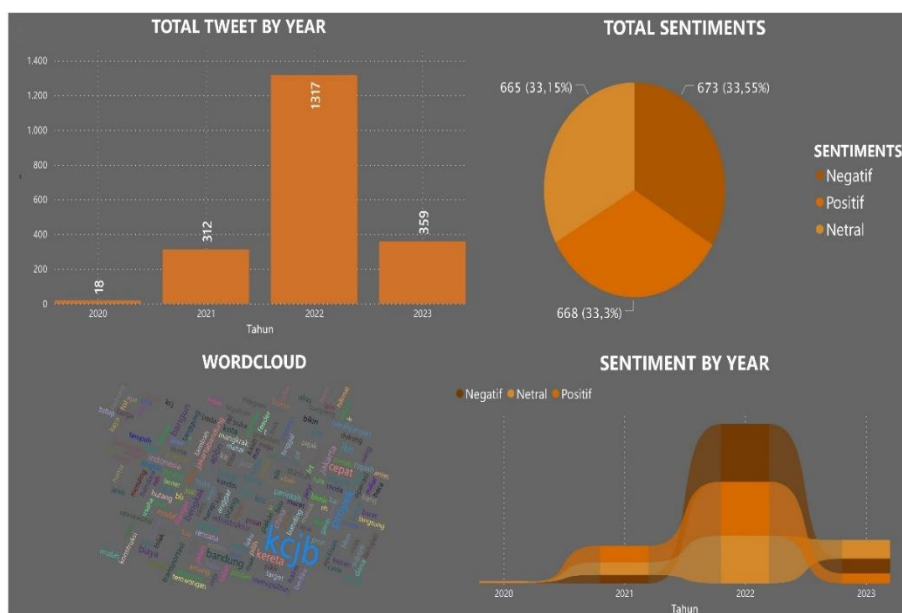
Hasil perhitungan klasifikasi *naive bayes* tersebut dijabarkan kedalam bentuk *confusion matrix* terhadap data training 80% dan data testing 20% disajikan pada gambar 5.



Gambar 4. Confusion Matrix

3.6 Visualisasi

Visualisasi dilakukan untuk mendapatkan data dengan menyederhanakan data yang sudah diolah ke dalam bentuk grafik pada gambar 6. Menampilkan 4 informasi yaitu total tweet pertahun, total sentimen, word cloud, dan sentimen pertahun.



Gambar 5. Visualisasi Dashboard

4. KESIMPULAN

Dari hasil penelitian ini tentang analisis sentimen pembangunan Kereta Cepat Jakarta Bandung (KCJB) pada media sosial twitter. Setelah melalui pemrosesan dengan hasil sentimen negatif sebanyak 673, hasil sentimen positif sebanyak 668, dan hasil sentimen netral sebanyak 665. Dapat disimpulkan bahwa hasil dari tanggapan masyarakat mengenai pembangunan Kereta Cepat Jakarta Bandung tergolong negatif. Pemodelan menggunakan Naive Bayes menghasilkan nilai akurasi (accuracy) sebesar 71%, nilai precision 73%, dan nilai tingkat keberhasilan (recall) 89%.

REFERENCES

- [1] R. W. Samsir, Ambiyar, Unung Verawardina, Firman Edi, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [2] N. Hardi, Y. Alkahfi, P. Handayani, W. Gata, and M. R. Firdaus, "Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier," *Sistemasi*, vol. 10, no. 1, p. 131, 2021, doi: 10.32520/stmsi.v10i1.1118.
- [3] T. K. Salaatsa and Y. Sibaroni, "Sentiment Analysis on the Construction of the Jakarta-Bandung High-Speed Train on Twitter Social Media Using Recurrent Neural Networks Method," vol. x, no. x, pp. 102–110, 2022.

- [4] R. Fajriati, S. Hapsoro, and T. Utomo, "Analisis Standar Perancangan Geometri Rel Kereta Cepat (Studi Kasus : Kereta Cepat Jakarta-Bandung) Analysis of High Speed Railway Geometric Design Standard (Case Study: Jakarta-Bandung High Speed Railway)," *J. Manaj. Aset Infrastruktur Fasilitas*, vol. 4, no. 3, pp. 249–260, 2020.
- [5] F. Sanjaya and V. Puspitasari, "Analisis Mengenai Dampak Lingkungan Pembangunan Kereta Cepat Jakarta-Bandung Dalam Perspektif Kritis Environmentalisme," *Padjadjaran J. Int. Relations*, vol. 2, no. 2, p. 170, 2020, doi: 10.24198/padjir.v2i2.26044.
- [6] R. Rosdiana, T. Eddy, S. Zawiyah, and N. Y. U. Muhammad, "Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar," *Proceeding SNTEI*, no. June 2020, pp. 87–93, 2019.
- [7] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5120–5127, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] I. Loelianto, M. S. S. Thayf, and H. Angriani, "Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru Stmik Kharisma Makassar," *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 110–117, 2020, doi: 10.31598/sintechjournal.v3i2.651.
- [9] P. P. O. Mahawardana, I. A. P. F. Imawati, and I. W. Dika, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *J. Manaj. dan Teknol. Inf.*, vol. 12, no. 2, pp. 50–56, 2022, [Online]. Available: <https://ojs.mahadewa.ac.id/index.php/jmti/article/view/2111>
- [10] S. Nurul, J. Fitriyyah, N. Safriadi, E. Eisyudha, and P. #3, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *(Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279–285, 2019, [Online]. Available: <http://dx.doi.org/10.26418/jp.v5i3.34368>
- [11] U. Kusnia and F. Kurniawan, "Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes," *Explor. IT*, vol. 14, no. 36, pp. 24–28, 2022.
- [12] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [13] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, "Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naive Bayes Berbasis PSO," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 930, 2022, doi: 10.30865/jurikom.v9i4.4629.
- [14] N. S. Devi Sari Oktavia Panggabean, Efori Buulolo, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," vol. 7, no. 1, pp. 56–62, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [15] M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.
- [16] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [17] A. M. F. Hulu and K. M. Lhaksmana, "Analisis Sentimen Politik pada Twitter Menggunakan Metode Support Vector Machine (Studi Kasus : Pilpres 2019)," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 2017–2020, 2019.
- [18] A. Rizki, "Operasi 'Tokenizing' pada Teks Berbahasa Indonesia," *adityarizki.net*, 2020. [https://adityarizki.net/belajarpython-9-operasi-tokenizing-pada-teks-berbahasa-indonesia/#:~:text=Tokenizing adalah operasi memisahkan teks,tanda baca%2C dan lain sebagainya](https://adityarizki.net/belajarpython-9-operasi-tokenizing-pada-teks-berbahasa-indonesia/#:~:text=Tokenizing%20adalah%20operasi%20memisahkan%20teks,tanda%20baca%20dan%20lain%20sebagaimana%20ini).
- [19] K. S. Nugroho, "Dasar Text Preprocessing dengan Python," *ksnugroho.medium.com*, 2019. [https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe#:~:text=Filtering adalah tahap mengambil kata,wordlist \(menyimpan kata penting\)](https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe#:~:text=Filtering%20adalah%20tahap%20mengambil%20kata,wordlist%20(menyimpan%20kata%20penting)).
- [20] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [21] K. Yan, D. Arisandi, and T. Tony, "Analisis Sentimen Komentar Netizen Twitter Terhadap Kesehatan Mental Masyarakat Indonesia," *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 1, 2022, doi: 10.24912/jiksi.v10i1.17865.