

Hilmy Zhafran Muflih - Analisis Sentimen Terhadap Pelayanan TransJakarta Berdasarkan Tweets Menggunakan Algoritma Naive Bayes

by Layanan Perpustakaan UHAMKA

Submission date: 21-May-2024 01:53PM (UTC+0700)

Submission ID: 2384692961

File name: lih_Teknik_Informatika_Turnitin_ke-1_-_Hilmy_Zhafran_Muflih.docx (1.16M)

Word count: 3887

Character count: 24550

Analisis Sentimen Terhadap Pelayanan TransJakarta Berdasarkan Tweets Menggunakan Metode Naïve Bayes Classifier

Absatrak

Tingginya penggunaan transportasi pribadi di Indonesia khususnya di daerah Jakarta menyebabkan beberapa dampak salah satunya yaitu kemacetan. Kondisi kemacetan ini bisa dikurangkan dengan transportasi umum. Transportasi umum saat ini diharapkan dapat menurunkan tingkat kemacetan di Jakarta. Salah satu transportasi umum di Jakarta yaitu TransJakarta. TransJakarta merupakan transportasi yang dapat mengangkut jumlah penumpang yang terbilang banyak dan TransJakarta menawarkan beragam fasilitas kepada pengguna, seperti tersedianya tempat duduk prioritas, halte yang terbilang cukup nyaman, kondisi di dalam bis yang nyaman ditambah harganya yang murah sehingga mendapatkan berbagai *respons* dari pengguna yang membuat peneliti untuk melakukan penelitian terhadap pandangan para pengguna TransJakarta mengenai layanan TransJakarta, apakah *respons* pengguna TransJakarta bersifat positif atau negatif. Tujuan dilakukannya penelitian ini untuk memahami para pengguna merasa puas atau tidak atas layanan yang sudah diberikan oleh TransJakarta. Metode yang digunakan dalam penelitian yaitu dengan algoritma *Naïve Bayes Classifier* yang digunakan untuk melakukan proses analisis sentimen mengenai pelayanan TransJakarta dengan bantuan aplikasi RapidMiner. Data yang dikumpulkan oleh peneliti sebanyak 773 data *tweet* yang diperoleh melalui media sosial X untuk dijadikan dataset. Hasil analisis sentimen dari algoritma *Naïve Bayes Classifier* memperoleh sebanyak 80,6% atau 623 sentimen negatif dan 19,4% atau 150 sentimen positif dari 773 dataset. Hasil evaluasi *confusion matrix* mendapatkan nilai akurasi sebesar 73,96%.

Kata Kunci: *Analisis Sentimen; TransJakarta; Naïve Bayes*

Abstract

The high use of private transportation in Indonesia, especially in the Jakarta area, causes several impacts, one of which is traffic jams. This congestion condition can be reduced by public transportation. It is hoped that public transportation can now reduce the level of congestion in Jakarta. One of the public transportation in Jakarta is TransJakarta. TransJakarta is a form of transportation that can carry a relatively large number of passengers and TransJakarta offers various facilities to users, such as the availability of priority seating, stops that are quite comfortable, comfortable conditions on the bus plus low prices so that it gets various responses from users who led researchers to conduct research on the views of TransJakarta users regarding TransJakarta services, whether TransJakarta users' responses were positive or negative. The purpose of this research is to understand whether users are satisfied or not with the services provided by TransJakarta. The method used in the research is the *Naïve Bayes Classifier* algorithm which is used to carry out the sentiment analysis process regarding TransJakarta services with the help of the RapidMiner application. The data collected by researchers was 773 tweet data obtained via social media X to be used as a dataset. The results of sentiment analysis from the *Naïve Bayes Classifier* algorithm obtained 80.6% or 623 negative sentiments and 19.4% or 150 positive sentiments from 773 datasets. The results of the *confusion matrix* evaluation obtained an accuracy value of 73.96%.

Keywords: *Sentiment Analysis; TransJakarta; Naïve Bayes*

1. PENDAHULUAN

saat ini transportasi adalah suatu bentuk kendaraan yang digunakan oleh seseorang untuk melakukan perpindahan dari sebuah tempat ke tempat lain [1]. Jenis transportasi dapat dikategorikan menjadi transportasi pribadi dan transportasi umum [2]. Transportasi pribadi merupakan transportasi yang dimiliki secara perorangan, sebaliknya transportasi umum adalah transportasi yang digunakan untuk kepentingan masyarakat umum. Daerah Jabodetabek adalah daerah di Indonesia yang mempunyai tingkat penduduk tertinggi, sehingga memiliki tingkat intensitas transportasinya yang tinggi [3]. Akibat dari tingginya pergerakan transportasi membuat daerah Jabodetabek khususnya di DKI Jakarta menjadi macet. Untuk mengurangi jumlah kendaraan di jalan raya yang menyebabkan kemacetan, upaya yang dilakukan yaitu sangat diperlukannya optimalisasi penggunaan transportasi umum [4]. Transportasi umum yang memenuhi kebutuhan dan mobilitas di Jakarta yaitu TransJakarta.

Transjakarta adalah bus yang memiliki sistem yang serupa dengan transportasi umum massal, namun beroperasi secara cepat serta menggunakan jalur khusus [5]. TransJakarta menawarkan beragam fasilitas kepada pengguna, seperti tersedianya tempat duduk prioritas, halte yang terbilang cukup nyaman, kondisi di dalam bus yang nyaman ditambah harganya yang murah. Teknologi informasi sedang mengalami kemajuan pada saat ini, sehingga sebagian besar kegiatan transaksi dilakukan menggunakan sistem kartu elektronik [6]. Saat ini pembelian tiket transjakarta sudah menerapkan sistem kartu elektronik sebagai pengganti pembayaran tunai. Pembelian kartu elektronik (*e-ticketing*) bisa dilakukan di setiap *vending machine* di beberapa halte transjakarta. Mesin isi ulang kartu elektronik dalam *vending machine* dapat memudahkan penggunaan kartu untuk memeriksa saldo, melakukan isi ulang, serta membeli kartu terbitan JAKLINGKO seperti Brizzi dari BRI, Tap Cash dari BNI, dan Mandiri E-Money [7].

Akan tetapi Transjakarta tidak menutup kemungkinan memiliki banyak kekurangan dalam faktor pelayanannya. Jika terjadinya ketidakpuasan yang dialami oleh para pengguna TransJakarta dan kurangnya terhadap layanan tidak bisa diatasi, maka akan muncul dampak seperti menurunnya minat untuk menggunakan Transjakarta. Untuk meningkatkan kualitas layanan, maka perlu adanya analisis mengenai opini masyarakat terhadap penggunaan Transjakarta. Untuk mendapatkan opini yang beredar di masyarakat, maka diperlukannya sosial media seperti media sosial X [8]. Media sosial X adalah media sosial terkenal yang berguna sebagai wadah komunikasi di masyarakat [9]. Setiap harinya, pengguna aktif twitter dapat menulis lebih dari 400 juta *tweet* terkenal [10]. Para pengguna media sosial X dapat berbagi hal dalam bentuk tulisan status atau *tweet* seperti mengungkapkan kebahagiaan, informasi, inspirasi, dan mendiskusikan topik tertentu [11]. Semua pendapat, saran, dan kritik yang diperoleh dari pengguna sosial media yaitu media sosial X dapat disebut sebagai sentimen [12].

Pada penelitian ini menjelaskan terkait bagaimana proses analisis sentimen dari sentimen atau opini yang diposting oleh masyarakat Indonesia di media sosial X, sentimen tersebut nantinya akan diklasifikasi menjadi data sentimen yang bernilai negatif dan positif [13]. Analisis sentimen adalah teknik untuk mengolah data opini dan teks secara otomatis, dengan tujuan memperoleh gambaran sentimen terdapat di sebuah opini [14]. Analisis sentimen merupakan proses *text mining* untuk melakukan klasifikasi terhadap data yang tidak beraturan untuk mendapatkan informasi mengenai sentimen dengan menggunakan *data mining* [15]. Diharapkan bahwa data mining bisa menghubungkan penggunaannya dan komunikasi antara data, karena data mining merupakan tahapan untuk menemukan informasi baru dari sejumlah besar data [16]. Diantara beberapa penelitian sentimen bisa dikelompokkan menjadi big data karena data teks menjadi semakin besar, sehingga semakin banyak dan makna dalam berbagai konteks akan semakin beragam, dan data *tweet* pada media sosial X merupakan sumber sentimen yang sering digunakan [17]. Sentimen bisa diilustrasikan dalam bentuk perasaan, keputusan,

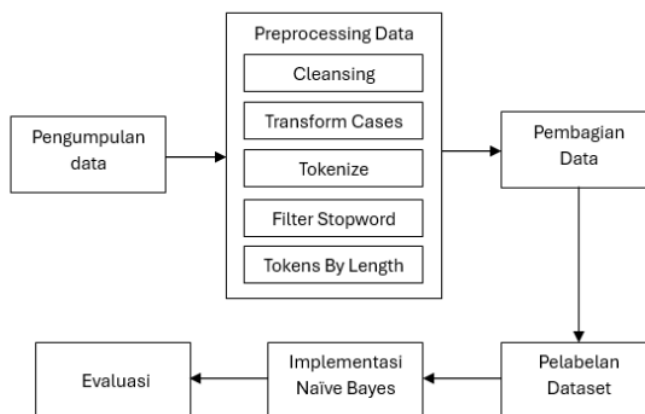
serta pendapat mengenai ide-ide untuk sumber informasi [18]. Teknik pembelajaran yang digunakan untuk melakukan analisis sentimen salah satunya adalah *Naïve Bayes Classifier* [19]. *Naïve Bayes* adalah sebuah algoritma yang menerapkan teori Bayes dengan memakai teknik *data mining* [20]. Selanjutnya proses *crawling* data pada X dan pengolahan data *tweet* dengan algoritma *Naïve Bayes* akan dilakukan pada *software RapidMiner*. *RapidMiner* adalah sebuah aplikasi data ilmu pengetahuan yang dikembangkan untuk menyediakan pembelajaran *deep learning*, *machine learning*, *predictive analytics*, dan *text mining* [21].

Tujuan penelitian ini, yaitu untuk mengetahui respon para pengguna TransJakarta terhadap pelayanan yang telah diberikan oleh TransJakarta apakah sudah memuaskan atau kurang memuaskan atas layanan Transjakarta yang telah diberikan. Jika banyak pengguna TransJakarta merasa tidak puas atas layanan yang sudah diberikan, TransJakarta diharapkan bisa meningkatkan kualitas layanan. Berbeda dengan penelitian terdahulu, penelitian ini menggunakan algoritma *machine learning Naïve Bayes Classifier* dan aplikasi *RapidMiner* untuk mengumpulkan dan menganalisis data dari *tweet* di media sosial X yang berisi ulasan atau sentimen dari para pengguna TransJakarta. Sebuah media sosial yaitu X yang digunakan oleh pengguna untuk berbagi video, foto, dan memberikan pendapat tentang topik aktual. Pada aplikasi *RapidMiner*, terdapat algoritma *Naïve Bayes Classifier* yang bisa digunakan untuk melakukan *data mining*, *text mining*, dan analisis sentimen pada *tweet* yang dikumpulkan. Dengan menggunakan aplikasi ini, peneliti dapat melakukan penelitian ini dengan lebih efektif dan efisien.

Analisis sentimen telah banyak dilakukan di penelitian sebelumnya. Penelitian sebelumnya membahas mengenai topik terkait Covid-19 dengan menganalisis sentimen yang didapat dari media sosial X dan diolah menggunakan algoritma *Naïve Bayes*, sehingga hasil akurasi yang di dapat sebesar 63.21% [22]. Penelitian terdahulu lainnya yaitu membahas mengenai respon atau opini terhadap adanya program kampus merdeka dengan cara menganalisis opini yang didapat dari media sosial X dan diolah menggunakan *Naïve Naves Classifier*. Nilai akurasi diperoleh dari alogritma *Naïve Bayes Classifier* yaitu 60% [23].

2. METODOLOGI

Metode pada penelitian ini menggunakan *tweet* yang diperoleh melalui media sosial X dan *tweet* tersebut akan dijadikan dataset. Dataset tersebut akan diproses pada aplikasi *RapidMiner* untuk memberikan prediksi sentimen terhadap data *tweet*, sehingga dapat memudahkan untuk mengetahui sentimen negatif atau positif pada data *tweet*. Berikut tahapan proses penelitian.



Gambar 1. Tahapan Proses Penelitian

A. Pengumpulan Data

Tahap ini merupakan proses pengumpulan data yang dilakukan dengan menggunakan web scraping. web scraping mempunyai kemampuan untuk merubah data tidak beraturan menjadi beraturan yang kemudian data bisa disimpan dalam basis data. Tujuan dari tahap scraping data yaitu memvalidasi, menyimpan, dan mengumpulkan data supaya dapat dipahami sebagai informasi yang berkualitas [24]. Data yang akan digunakan untuk analisis sentimen yaitu berupa data dalam bentuk *tweet* yang diperoleh dari media sosial X. Data *tweet* yang akan digunakan untuk analisis sentimen yaitu mengenai TransJakarta dengan jumlah data *tweet* yang diperoleh sebanyak 827 data *tweet* mulai dari maret 2024 hingga april 2024. Proses pengumpulan data ini dilakukan di *google colab* dengan menggunakan *python*

B. Preprocessing Data

Preprocessing data bertujuan untuk memproses data yang diperoleh dalam bentuk yang tidak beraturan. Selain itu, *preprocessing data* juga untuk menghapus karakter yang tidak diperlukan dan merubah kata asing ke dalam Bahasa Indonesia [25]. Ada beberapa langkah yang berlangsung dalam tahap *preprocessing*, yaitu mencakup

- 1) *Cleansing*, tahap ini dilakukan untuk membersihkan dataset yang telah diperoleh dengan cara menghapus simbol-simbol/karakter, *hashtag*, *retweet*, *URL*, dan *mention* yang terkandung *tweet*.
- 2) *Transform cases*, tahap ini dilakukan untuk mengganti setiap huruf yang ada pada dataset dalam berbentuk kapital menjadi huruf yang tidak kapital.
- 3) *Tokenize*, tahap ini bertujuan untuk memisahkan setiap teks yang ada pada dataset menjadi sebuah bagian-bagian/token.
- 4) *Stopword*, tahap ini dilakukan dengan tujuan untuk menghapus setiap kata mempunyai arti pada dataset yang telah diperoleh.
- 5) *Tokens by length*, tujuan dilakukannya tahap ini untuk menghapus setiap kata yang mengandung jumlah huruf yang sudah ditetapkan.

C. Pembagian Data

Pada tahap ini dataset akan dibagi menjadi 2 data, yaitu data latih dan data uji. Pembagian data ini dilakukan dengan menggunakan perbandingan 70 : 30, yaitu 70% dari total jumlah dataset akan dijadikan data latih dan 30% dari total jumlah dataset akan dijadikan data uji. Data latih berfungsi sebagai data yang akan memberikan latihan/training kepada algoritma *Naïve Bayes*, sehingga algoritma *Naïve Bayes* dapat beradaptasi terhadap pola yang ada di data latih. Sedangkan untuk data uji untuk dijadikan sebagai bahan uji untuk melakukan klasifikasi oleh algoritma *Naïve Bayes*.

D. Pelabelan Data

Pelabelan dataset dilakukan untuk mengklasifikasikan menjadi dua kategori sentimen, yaitu sentimen negatif dan positif. Sentimen berkategori positif menggambarkan emosi positif. Sedangkan Sentimen yang berkategori negatif digunakan untuk emosi negatif. Proses klasifikasi sentimen ini dilakukan manual [26]. Pelabelan dataset ini bertujuan untuk membuat data latih.

E. Implementasi *Naïve Bayes*

Naïve Bayes Classifier merupakan metode dengan konsep probabilitas untuk melakukan klasifikasi. Konsep probabilitas ini berdasarkan dari *Teorema Bayes*. Algoritma *Naïve Bayes* Ini menjadi salah satu metode klasifikasi yang sangat efisien. [27]. *Naïve Bayes* memiliki keunggulan menciptakan kinerja yang baik pada saat klasifikasi teks, walaupun memiliki kemampuannya yang sederhana [28].

Tahap Implementasi *Naïve bayes* ini merupakan tahap dilakukannya proses klasifikasi dengan algoritma *Naïve Bayes* terhadap dataset. Hasil yang akan diperoleh dari proses klasifikasi ini berupa sebuah prediksi sentimen terhadap *tweet* yang belum memiliki label sentimen. Hal yang dilakukan

dalam tahap ini yaitu membuat model algoritma *Naïve Bayes* dan juga data latih. Pembuatan model algoritma *Naïve Bayes* dan juga data latih menggunakan dataset yang telah diberikan sebagian label sentimen pada *tweet*, dataset tersebut nantinya akan difilter untuk mengambil data yang sudah memiliki label sentimen.

Setelah memiliki model algoritma *Naïve Bayes* dan juga data latih, maka akan dilanjut proses klasifikasi dengan menggunakan dataset yang akan difilter terlebih dahulu untuk diambil data *tweet* yang belum memiliki label sentimen. Setelah proses klasifikasi selesai, maka setiap *tweet* yang belum ada label sentimennya akan memiliki label sentimen secara otomatis dari hasil kalsifikasi yang berupa nilai prediksi sentimen

F. Evaluasi

Tahap evaluasi dilakukan untuk menguji suatu metode/algoritma dalam penelitian. Pada tahap ini metode/teknik *confusion matrix* akan digunakan untuk mengevaluasi algoritma *Naïve Bayes* dengan cara mengukur performa dari algoritma *naïve bayes* yang digunakan untuk melakukan proses klasifikasi. Hasil yang didapat dari *confusion matrix*, yaitu berupa nilai *precision*, *acuracy*, dan *recall*.

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada tahap awal peneliti melakukan proses pengumpulan data TransJakarta di media sosial X terlebih dahulu dengan teknik *web scraping*. Pengumpulan data dilakukan di Google Colab dengan menggunakan *Python*. Data yang diperoleh oleh peneliti sebanyak 827 data *tweet* yaitu mulai dari maret 2024 hingga april 2024. Peneliti menggunakan kata “Transjakarta” sebagai *keyword* untuk menyaring *tweet* pada media sosial X yang akan diambil untuk dijadikan dataset. Setelah data berhasil diambil, Data yang telah didapatkan akan disimpan ke sebuah file berformat CSV. Gambar 2 berikut dapat dilihat alur dari pengumpulan *tweet* untuk dataset.

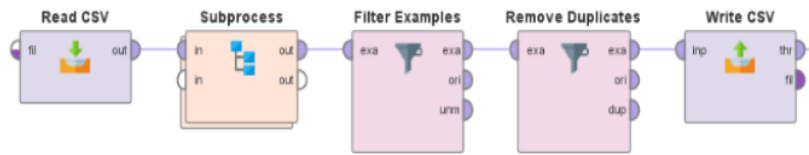


Gambar 2. Alur Pengumpulan Tweet untuk Dataset

B. Preprocessing Data

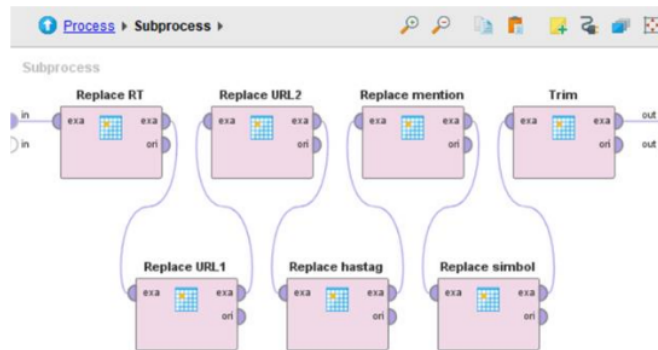
1) *Cleansing*

Proses selanjutnya peneliti melakukan preprocessing pada data *tweet* yang dilakukan di aplikasi *RapidMiner*. Pada saat preprocessing, teks yang tidak teratur pada dataset ulasan TransJakarta akan diubah menjadi teratur sehingga nantinya tidak akan terjadi kesalahan. *Cleansing* akan menjadi proses pertama yang dilakukan pada *preprocessing*. Berikut merupakan alur dari proses *cleansing*.



Gambar 3. alur dari proses cleansing

Proses cleansing terdiri dari 2 bagian yaitu menghapus data yang duplicated dan juga menghapus berbagai karakter atau simbol seperti : .?@#,#!-/ dan seterusnya. Data *tweet* akan dicleansing melalui operator *subprocess* dan *remove duplicated*. Berikut merupakan operator yang ada di subprocess



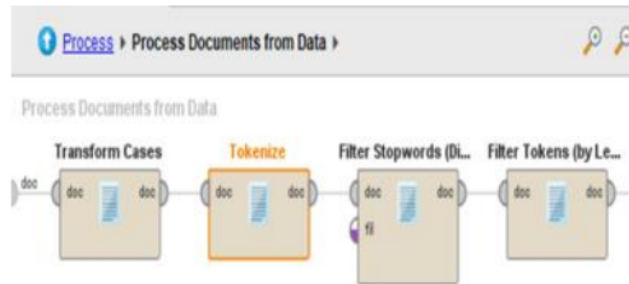
Gambar 4. Operator Dalam Subprocess

Di dalam *subprocess* terdapat operator *replace* dan *trim*. Operator *replace* berguna untuk menghilangkan URL, *retweet*, mention, simbol, dan hastag, kemudian dilanjutkan dengan operator *remove duplicated* untuk menghapus data yang duplikat. Pembersihan *tweet* ini dilakukan agar saat proses analisis sentimen mendapatkan hasil yang sesuai. Berikut dapat dilihat contoh *tweet* pada saat *cleansing*.

TABEL I
CLEANSING

SEBELUM CLEANSING	Sesudah cleansing
@PT_Transjakarta sudah 20menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya	sudah 20menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya
@PT_Transjakarta Ini 6T lagi kenapa sih, semalem nunggu 45 menit, sekarang nunggu dah mau setengah jam belum nongol juga. Hadeh.	Ini 6T lagi kenapa sih semalem nunggu 45 menit sekarang nunggu dah mau setengah jam belum nongol juga Hadeh
@PT_Transjakarta Tolong dong pelayanan untuk 5C nya. Sudah 3 bus transjakarta yang tidak melayani sampai PGC. Padahal belum jam 10 malam. Mengecewakan sama sekali pelayanan bus transjakarta nya	Tolong dong pelayanan untuk 5C nya Sudah 3 bus transjakarta yang tidak melayani sampai PGC Padahal belum jam 10 malam Mengecewakan sama sekali pelayanan bus transjakarta nya

Setelah dilakukan proses cleansing maka akan dilanjut ke proses *preprocessing* selanjutnya. *preprocessing* ini memiliki beberapa tahap yang harus dilewati yang ada di dalam *process document* yaitu *transform cases*, *tokenize*, *stopword*, *filter tokens (by length)*. Peneliti juga menggunakan operator operator *process document* di aplikasi RapidMiner untuk menyesuaikan tahapan *transform cases*, *tokenize*, *stopword*, *filter tokens (by length)*. Gambar 5 merupakan beberapa operator yang digunakan di dalam *process document*.



Gambar 5. Operator Dalam Process Document

2) *Transform cases*

Tahap *transform cases* ini dilakukan untuk mengubah huruf kapital pada dataset untuk dirubah menjadi huruf yang tidak kapital [29]. Hal ini dilakukan supaya mencegah terjadinya kesalahan saat tokenize. Berikut adalah contoh dari *tweet* setelah melalui proses *transform cases* pada Tabel 2.

TABEL II
TRANSFORM CASES

SEBELUM TRANSFORM CASES	Sesudah <i>transform cases</i>
sudah 20menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya	sudah 20menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya
Ini 6T lagi kenapa sih semalem nunggu 45 menit sekarang nunggu dah mau setengah jam belum nongol juga Hadeh	Ini 6t lagi kenapa sih semalem nunggu 45 menit sekarang nunggu dah mau setengah jam belum nongol juga hadeh
Tolong dong pelayanan untuk 5C nya Sudah 3 bus transjakarta yang tidak melyani sampai PGC Padahal belum jam 10 malam Mengecewakan sama sekali pelayanan bus transjakarta nya	tolong dong pelayanan untuk 5c nya Sudah 3 bus transjakarta yang tidak melyani sampai pgc padahal belum jam 10 malam mengecewakan sama sekali pelayanan bus transjakarta nya

3) *Tokenize*

Tokenize adalah proses untuk memisahkan teks menjadi sebuah token untuk analisis yang akan dilakukan nanti[30]. Berikut adalah contoh *tweet* yang telah melalui proses tokenize pada Tabel 3.

TABEL III
TOKENIZE

SEBELUM <i>TOKENIZING</i>	Sesudah <i>tokenizing</i>
sudah 20menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya	['sudah','menit','tunggu','di','halte','bunderan','senayan','6m','ke','manggarai','kenapa','belum','lewat','lewat','ya']
Ini 6t lagi kenapa sih semalem nunggu 45 menit sekarang nunggu dah mau setengah jam belum nongol juga hadeh	['Ini','6t','lagi','kenapa','sih','semalem','nunggu','45','menit','sekarang','nunggu','dah','mau','setengah','jam','belum','nongol','juga','hadeh']
tolong dong pelayanan untuk 5c nya Sudah 3 bus transjakarta yang tidak melyani sampai pgc Padahal belum jam 10 malam mengecewakan sama sekali pelayanan bus transjakarta nya	['tolong','dong','pelayanan','untuk','5c','nya','sudah','3bus','transjakarta','yang','tidak','melyani','sampai','pgc','padahal','belum','jam','10','malam','mengecewakan','sama','sekali','pelayanan','bus','transjakarta','nya']

4) *Stopwords*

Stopwords adalah proses menghapus kata-kata yang tidak mempunyai arti [31]. *Stopwords* yang digunakan peneliti adalah *stopwords* Bahasa Indonesia yang diperoleh dari www.kaggle.com. Tabel 4 adalah *tweet* yang telah melewati proses *stopwords*.

TABEL IV
TOKENIZE

SEBELUM <i>STOPWORD</i>	Sesudah <i>stopword</i>
sudah 20 menit tunggu di halte bunderan senayan 6m ke manggarai kenapa belum lewat lewat ya	20 menit tunggu halte bunderan senayan 6m manggarai belum lewat lewat ya
Ini 6t lagi kenapa sih semalem nunggu 45 menit sekarang nunggu dah mau setengah jam belum nongol juga hadeh	6t lagi kenapa semalem nunggu 45 menit sekarang nunggu setengah jam belum nongol hadeh
tolong dong pelayanan untuk 5c nya Sudah 3 bus transjakarta yang tidak melyani sampai pgc Padahal belum jam 10 malam mengecewakan sama sekali pelayanan bus transjakarta nya	tolong pelayanan 5c 3 bus transjakarta melyani pgc jam 10 malam mengecewakan pelayanan bus transjakarta butuh deadline 3 hari mah

5) *Tokens by length*

Tokens by length adalah proses untuk menghilangkan kata-kata yang jumlah hurufnya telah ditetapkan. Tabel 5 merupakan contoh *tweet* melewati proses *filter_tokens by length*.

TABEL V
TOKENS BY LENGTH

SEBELUM <i>FILTER BY LENGTH</i>	Sesudah <i>filter by length</i>
20 menit tunggu halte bunderan senayan 6m manggarai belum lewat lewat ya	menit tunggu halte bunderan senayan manggarai belum lewat lewat

6t lagi kenapa semalem nunggu 45 menit sekarang nunggu setengah jam belum nongol hadeh	lagi kenapa semalem nunggu menit sekarang nunggu setengah jam belum nongol hadeh
tolong pelayanan 5c 3 bus transjakarta melyani pge jam 10 malam mengecewakan pelayanan bus transjakarta butuh deadline 3 hari mah	tolong pelayanan bus transjakarta melyani pge jam 10 malam mengecewakan pelayanan bus transjakarta butuh deadline 3 hari mah

C. Pelabelan Data

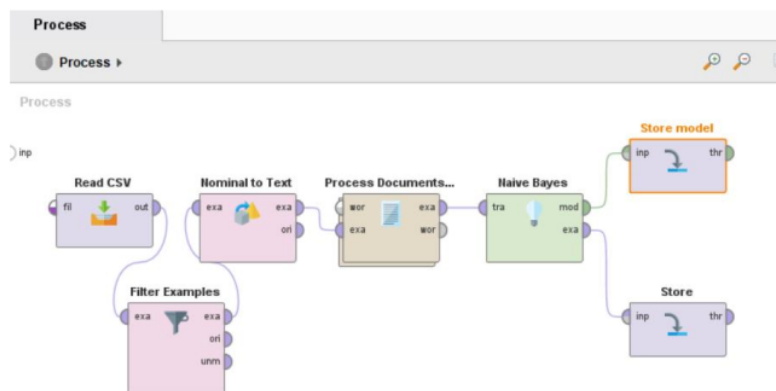
Proses berikutnya yaitu melakukan proses pelabelan sentimen pada data *tweet* oleh peneliti. Pada proses pelabelan sentimen, peneliti akan membagi dataset terlebih dahulu, yaitu data latih dan uji dengan perbandingan 70:30 dengan jumlah data *tweet* berjumlah 563 data latih dan 210 data uji. Peneliti memberikan label sentimen negatif atau positif pada setiap *tweet* yang telah ditetapkan sebagai data latih. Proses pelabelan sentimen pada data *tweet* akan dilakukan di *Miscrosoft Excel*. Berikut pada gambar 6 menunjukan proses pelabelan dataset.

1	text	sentimen
2	setiap mau kerja tj selalu penuh mana nunggunya lama lagi	NEGATIF
3	iya nih kemaren saldo gue juga gitu bro kepotong dobel	NEGATIF
4	masih berusaha sabar walaupun tarif ongkos masih suka ke potong 2 kali da	NEGATIF
5	trus kabar baik refund yang kepotong 2 kali dari 1 perjalanan kapan	NEGATIF
6	udah gak bales dm saya bales pengaduan wa saya dan bales reply saya	NEGATIF
7	terus karena gua naik transjakarta jadi di jalan bisa 2 jam	NEGATIF
8	makasih banyak kak 10h ngadet trs ahahaha lama bgt bwt ke pademangan	NEGATIF
9	lama ga naik busway kok rasanya enjoy kwwk	POSITIF
10	naek busway kalo gak mau macet	POSITIF

Gambar 6. Pelabelan Dataset

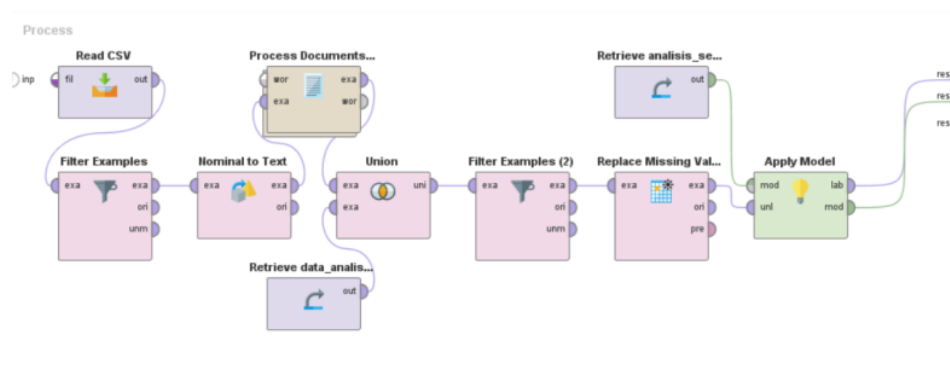
D. Implementasi Naïve Bayes

Langkah berikutnya adalah mengimplementasikan algoritma *Naïve Bayes*. Implementasi ini dilakukan melalui aplikasi RapidMiner. Tahap pertama yaitu membuat model *Naïve Bayes* dan data latih. Pembuatan model *Naïve Bayes* dan data latih, yaitu dari dataset sudah melewati operator *process document (preprocessing)* yang sudah terhubung dengan operator algoritma *Naïve Bayes*. Berikut alur pemrosesan pembuatan data latih dan model *Naïve Bayes*.



Gambar 7. Proses Pembuatan Data Latih dan Model *Naïve Bayes*.

Langkah berikutnya setelah mempunyai data latih dan model *Naïve Bayes*, maka akan dilanjut ke tahap klasifikasi dengan algoritma *Naïve Bayes*. Dataset yang berisikan *tweet* ulasan TransJakarta akan difilter untuk mengambil data yang belum memiliki label sentimen, data yang telah difilter akan dijadikan sebagai data uji. Kemudian dilanjut dengan *proses preprocessing*. Hasil dari proses sebelumnya akan digabung dengan data latih menggunakan operator union sebab terdapat perbedaan kata-kata di antara keduanya. Setelah digabungkan, data akan melalui proses filter kembali untuk mengambil data yang belum ada label sentimennya. Setelah data uji siap digunakan, maka akan dilanjut ke proses klasifikasi dengan menggunakan model *Naïve Bayes*. Peneliti akan melakukan proses klasifikasi pada data uji dengan menghubungkan model *Naïve Bayes* dengan operator *apply model*. Hasil yang diperoleh dari proses klasifikasi berupa prediksi sentimen pada data *tweet* yang belum ada sentimennya. Proses pemberian label otomatis digambarkan pada Gambar 8.



Gambar 8. Proses pelabelan secara otomatis dengan *Naïve Bayes*

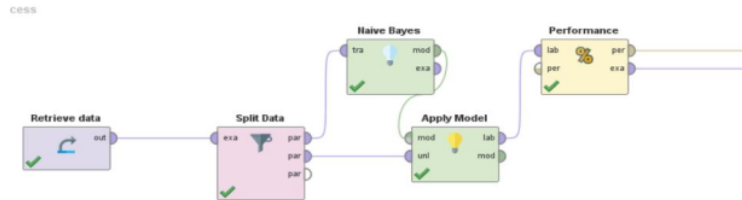
Dalam tahap pemberian label sentimen oleh algoritma *Naïve Bayes* secara otomatis. Algoritma *Naïve Bayes* telah beradaptasi dengan data latih yang telah diberi label sentimen secara manual oleh peneliti, sehingga *Naïve Bayes* dapat memprediksi data kedalam dua kelas sentimen, yaitu negatif dan positif. Berikut adalah hasil dari pemberian label sentimen menggunakan *Naïve Bayes* yang ditampilkan dalam gambar 9.

text	prediction(S...
duduk prioritas penuh sampe hamil dapet duduk	NEGATIF
murah menguji kesabaran	POSITIF
sebel nunggu jarak armadanya banget menit	NEGATIF
petugasnya kalo datang slang mulu doang bikin emosi	NEGATIF
sayangnya masuk terminal pasar minggu biar jalan stasiun pasar minggu kejauhan	NEGATIF
payah udah tunggu sbntr sdng tapping tinggal kalo giliran cewek tungguin parah parahh bn...	POSITIF
selamat pagi kemarin maafin maklum pramudi main lapangan pondok gede langsung giat ...	POSITIF
paham anjir udah berkali kali dibilang koridor arah kalideres suka ngetem harmoni pagi ate...	NEGATIF
pengerjaan pengguna kesal jalan halte mohon pengertiannya ditunggu balasan pastinya	POSITIF
astaga bener sekalinya udah nunggu gaboleh udah full penumpangnya ganti armada gede ...	NEGATIF
gimana ngatur jadwal lucu banget kaya ngomong temen lola	NEGATIF
udah sibuk datangnya jarak sejam tolonglah ditambah dikjakarta	NEGATIF
negur petugasnya aware plus bapaknya bebeulll	POSITIF
capek banget protes mulu	NEGATIF
tolong atasi kali pancoran baratpuribeta bermasalah berhenti pintu berbeda halte berhenti h...	POSITIF
antriannya udah sampe pintu masuk stasiun tebet good	POSITIF
dear dibentak diusir membantu solutif	POSITIF

Gambar 9. Hasil sentimen anlisris terhadap TransJakarta

E. Evaluasi

Setelah klasifikasi dengan algoritma *Naïve Bayes* telah selesai dilakukan, maka akan dilanjutkan dengan proses evaluasi dari metode yang digunakan, yaitu algoritma *Naïve Bayes*. proses evaluasi ini dilakukan oleh peneliti dengan menggunakan *confusion matrix* untuk mengevaluasi metode yang digunakan. Berikut alur evaluasi algoritma *Naïve Bayes*.



Gambar 10. Alur *confusion matrix* untuk evaluasi algoritma *Naïve Bayes*

Peneliti proses evaluasi peneliti menggunakan operator *split data* yaitu untuk menetapkan perbandingan data *latih* dan data *uji*. peneliti akan menggunakan perbandingan yaitu sebesar 70:30. Hasil yang diperoleh dari evaluasi algoritma *Naïve bayes* ini berupa nilai *Precision*, *accuracy*, dan *recall*. Gambar 11 merupakan hasil evaluasi *confusion matrix*.

accuracy: 73.96%

	true NEGATIF	true POSITIF	class precision
pred. NEGATIF	116	23	83.45%
pred. POSITIF	21	9	30.00%
class recall	84.67%	28.12%	

Gambar 11. Hasil evaluasi algoritma *Naïve Bayes* dengan *confusion matrix*

Hasil evaluasi dengan *confusion matrix* terhadap algoritma *Naïve Bayes* memperoleh nilai *accuracy*, yaitu 73,92%. Kemudian untuk sentimen positif dari hasil evaluasi memperoleh nilai *precision* 30% dan *recall* 28,12%, sedangkan untuk sentimen negatif didapatkan nilai *precision* 83,45% dan *recall* 84,67%. Berikut visualisasi *confusion matrix*.

		True Label	
		0	1
Predicted Label	0	True Negatif 212	False positif 23
	1	False Negatif 21	True Positif 9

Gambar 12. Visualisasi *confussion matrix*

Pada visualisasi *confussion matrix* terdapat warna biru pudar hingga pekat, warna biru pudar menandakan jumlah data yang sedikit sedangkan warna biru tua menandakan jumlah data yang banyak. Hasil *confussion matrix* dari visualisasi yang telah dibuat yaitu terdapat TP, FP, FN, dan TN. TP merupakan true positif, true positif pada hasil visualisasi diatas yaitu berjumlah 9 data. Kemudian, FP merupakan false positif, false positif berjumlah 23 data. Setelah itu ada FN yang merupakan false negatif yang berjumlah 21 data. Terakhir ada TN yaitu true negatif yang berjumlah 212 data.

4. KESIMPULAN

Berdasarkan hasil pemrosesan dengan klasifikasi *Naïve Bayes* dari dataset yang telah dilakukan terhadap ulasan penggunaan Transjakarta, peneliti mendapatkan 150 sentimen positif dan 623 sentimen negatif dari 773 dataset. Dari hasil yang didapat terdapat banyak sekali sentimen negatif yang diperoleh dibandingkan dengan sentimen positif yang diperoleh, sehingga dapat ditarik kesimpulan bahwa TransJakarta kurang memberikan pelayanan yang baik kepada para pengguna TransJakarta. Hasil evaluasi dari algoritma *Naïve Bayes Classifier* menunjukkan bahwa pada tingkat akurasi dari algoritma *Naïve Bayes Classifier* yang diperoleh dari *confusion matrix* sebesar 73,96%. Kemudian hasil lain dari *confusion matrix*, yaitu pada sentimen positif mendapat nilai *precision* sebesar 30% dan *recall* 28,12%. Sedangkan sentimen negatif mendapat nilai *precision* sebesar 83,45% dan *recall* 84,67%. Kekurangan yang ada di penelitian ini yaitu dari jumlah dataset yang kurang banyak, kemudian sampel data latih yang digunakan juga bisa lebih banyak lagi. Peneliti hanya memberi label pada dataset sebanyak 563 data untuk dijadikan data *latih*, sehingga dengan begitu masih ada beberapa data dalam proses klasifikasi *Naïve Bayes* yang tidak akurat. Sebagai saran untuk penelitian selanjutnya, data yang akan digunakan dalam proses analisis bisa lebih banyak dan saat proses pelabel dataset oleh peneliti bisa lebih banyak dilakukan agar hasil yang didapat pada proses klasifikasi *Naïve Bayes* bisa semakin akurat, karena pada penelitian ini tidak sepenuhnya sesuai dengan yang diharapkan oleh peneliti terbukti hasil dari akurasi yang didapat masih relatif tidak tinggi.

Hilmy Zhafran Muflih - Analisis Sentimen Terhadap Pelayanan TransJakarta Berdasarkan Tweets Menggunakan Algoritma Naive Bayes

ORIGINALITY REPORT

5%

SIMILARITY INDEX

3%

INTERNET SOURCES

4%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	jurnal.polibatam.ac.id Internet Source	1%
2	text-id.123dok.com Internet Source	1%
3	journal.umg.ac.id Internet Source	1%
4	Nurfazriah Attamami, Agung Triayudi, Rima Tamara Aldisa. "Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan", Jurnal JTIC (Jurnal Teknologi Informasi dan Komunikasi), 2023 Publication	<1%
5	Salman Alfaris, Kusnawi. "Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee", Indonesian Journal of Computer Science, 2023 Publication	<1%

6	Ahmad Soimun, Arinda Leliana. "Analisis Fasilitas dan Aksesibilitas Halte dan Stasiun Kereta Commuter Surabaya Porong", Jurnal Perkeretaapian Indonesia (Indonesian Railway Journal), 2021 Publication	<1 %
7	ejournal3.undip.ac.id Internet Source	<1 %
8	es.scribd.com Internet Source	<1 %
9	repositori.unsil.ac.id Internet Source	<1 %
10	Wafirotul Laila, Wisnu Widiarto, Ardhi Wijayanto, Esti Suryani. "Rekomendasi Makanan Pasien Hiperlipidemia Berdasarkan Hasil Klasifikasi Menggunakan Metode Naïve Bayes dan Decision Tree", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2022 Publication	<1 %
11	journal.trunojoyo.ac.id Internet Source	<1 %
12	Firnanda Zuhad, Nori Wilantika. Jurnal Linguistik Komputasional (JLK), 2022 Publication	<1 %
13	Fitri Wulandari, Elin Haerani, Muhammad Fikry, Elvia Budianita. "Analisis sentimen	<1 %

larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier", Jurnal CoSciTech (Computer Science and Information Technology), 2023

Publication

14

Submitted to Konsorsium Perguruan Tinggi Swasta Indonesia II

Student Paper

<1 %

15

Ilyas Jayanto, Benisius. "Analisis Perbandingan Algoritma Decision Tree untuk Prediksi Karyawan dengan Potensi Atrisi di PT. XYZ", Jurnal Informatika Komputer, Bisnis dan Manajemen, 2024

Publication

<1 %

16

Rizqia Lestika Atimi, Enda Esyudha Pratama. "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia", Jurnal Sains dan Informatika, 2022

Publication

<1 %

17

Yohana Tri Widayati, Yani Prihati, Stephanus Widjaja. "ANALISIS DAN KOMPARASI ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI LOYALITAS PELANGGAN MNC PLAY KOTA SEMARANG", Jurnal Transformatika, 2021

Publication

<1 %

18

repository.uin-suska.ac.id

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off