

Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Commuter Line

Implementation of Naive Bayes for Sentiment Analysis of Commuter Line Twitter Social Media

Sunarti¹, Frisma Handayanna^{2*}, Dewi Ayu Nur Wulandari³

^{1,3}Program Studi Sistem Informasi Kampus Kota Bogor, Universitas Bina Sarana Informatika

²Informatika, Universitas Nusa Mandiri, Jakarta, Indonesia

E-mail: ¹sunarti.sni@bsi.ac.id, ^{2*}frrisma.fha@nusamandiri.ac.id, ³dewi.dan@bsi.ac.id

*Penulis korespondensi

Abstrak

Commuter Line ialah salah satu sektor pelayanan publik dibidang transportasi banyak dipergunakan oleh masyarakat saat ini. Sesuai perkembangan teknologi banyak opini masyarakat terhadap pelayanan *Commuter Line* beredar secara *online*, untuk menyampaikan opininya masyarakat menggunakan media sosial seperti *Twitter*. Media *Twitter* dapat dijadikan sebagai bahan evaluasi peningkatan kualitas layanan. *Twitter* memudahkan pelanggan untuk menyampaikan keluhan serta pendapat terkait layanan seperti PT KAI *Commuter*. Untuk memahami komentar pelanggan, persepsi layanan publik, dan mendapatkan opini baru maka perlu dilakukan analisis sentimen terhadap pelayanan *Commuter Line*. Analisis sentimen ini bertujuan untuk mengklasifikasikan *tweets* masyarakat terhadap layanan *Commuter Line* ke dalam sentimen *complain* dan *not complain* dengan menggunakan metode *Naive Bayes*. Metode ini mempunyai tingkat akurasi paling tinggi dalam pengklasifikasian analisis sentimen. Data yang diunduh dari *Twitter* menggunakan aplikasi RapidMiner sebanyak 1.010 *tweet* dan data validasi sebanyak 1.003 *tweets*. Pada akhir tahap penelitian diperoleh *accuracy* 78,11%, *precision* 81,76%, *recall* 72,51%, dan AUC yang didapat sebesar 0,814.

Kata kunci: Analisis sentimen; Commuter Line; Twitter; Naive Bayes

Abstract

Commuter Line is one of the public service sectors in the field of transportation that is widely used by society today. In accordance with technological developments, many public opinions regarding *Commuter Line* services are circulating online, to convey their opinions, the public uses social media such as *Twitter*. *Twitter* media can be used as material for evaluating service quality improvement. *Twitter* makes it easy for customers to submit complaints and opinions regarding services such as PT KAI *Commuter*. In order to understand customer comments, public service perceptions, and obtain new opinions, it is necessary to carry out a sentiment analysis of *Commuter Line* services. This sentiment analysis aims to classify people's tweets about *Commuter Line* services into *complain* and *not complain* sentiments using the *Naive Bayes* method. This method has the highest level of accuracy in classifying sentiment analysis. The data downloaded from *Twitter* using the RapidMiner application are 1,010 tweets and validation data are 1,003 tweets. At the end of the research phase, the accuracy was 78.11%, the precision was 81.76%, the recall was 72.51%, and the AUC was 0.814.

Keywords: Sentiment analysis; Commuter Line; Twitter; Naive Bayes

1. PENDAHULUAN

Pemasar dan praktisi humas di seluruh dunia mulai menggunakan media sosial seperti *Twitter* untuk berbagai macam keperluan karena jumlah pengguna internet yang sangat besar dan terus meningkat. Cara orang berinteraksi dan berbagi ide terus berubah dengan media sosial.

Institusi atau lembaga dapat menggunakan teknologi untuk membangun identitas dan citra, bukan hanya di dunia nyata tetapi juga di dunia maya. Twitter adalah salah satu platform media sosial yang memberi peluang baru untuk berkomunikasi dan membina hubungan dengan publik[1]. Oleh karena itu perlu diketahui pendapat masyarakat pengguna KRL *Commuter Line* JABODETABEK (Jakarta, Bogor, Depok, Tangerang, dan Bekasi) tentang apa yang dirasakan saat menggunakan transportasi. Masyarakat bisa memberikannya pendapatnya dengan berbagai macam media. Salah satunya media yang efektif adalah *twitter*. *Twitter* termasuk tercepat dalam mengabarkan pengalaman yang dirasakan oleh masyarakat sebagai bahan evaluasi untuk pihak transportasi KRL *Commuter Line*. *Twitter* memberikan kemudahan pengumpulan pendapat dibandingkan survei ataupun kuesioner. Dengan banyaknya pengguna disertai kemudahan akses *twitter* dalam penyampaian opini, maka beragam ulasan terkumpul kemudian bisa menjadi peluang dimanfaatkan sebagai bahan penilaian dan evaluasi atas penggunaan KRL *Commuter Line* JABODETABEK [2].

Media *twitter* dijadikan bahan evaluasi peningkatan kualitas layanan. *Twitter* memudahkan pelanggan menyampaikan keluhan serta pendapat terkait layanan PT KAI *Commuter*[3]. Untuk memahami komentar pelanggan[4], persepsi layanan publik[5][6][7], dan mendapatkan opini baru[8][9][2] maka perlu dilakukan analisis sentimen terhadap pelayanan *Commuter Line*. Literatur permasalahan dalam penulisan ini sebagai berikut:

Tabel 1 Literatur Penelitian

Research Problem(RP)	Literature Support
RP1 Banyaknya komentar pelanggan atas penggunaan KRL <i>Commuter Line</i>	Untuk memahami komentar pelanggan[4]
RP2 Beragam persepsi untuk kualitas layanan publik	Persepsi layanan publik[5][6], Pendapat penggunaan layanan publik dengan media <i>twitter</i> [7]
RP3 Menimbulkan keluhan, terhadap isu-isu layanan publik	Mendapatkan opini baru[8][2], Perdebatan terhadap isu-isu publik[9]

Analisis sentimen menggambarkan keseluruhan persepsi masyarakat melalui mengelompokkan jenis opini yang muncul ke dalam kategori positif, negatif, dan netral[10]. Analisis sentimen mengolah data secara otomatis untuk memperoleh sentimen sebuah kalimat. Hal ini dilakukan untuk melihat pendapat atau kecenderungan opini masyarakat apakah cenderung berpandangan positif atau negatif. Terdapat beberapa penelitian sebelumnya yang membahas analisis sentimen dokumen teks. Penelitian menggunakan analisis sentimen untuk menilai kepuasan seseorang melalui opini berdasarkan dokumen teks. Dokumen dipecah berdasarkan teks yang penyusun, kemudian masing-masing teks dinilai untuk menentukan opini memiliki sentimen positif atau negatif[11][12].

Analisis sentimen yang dilakukan untuk mengklasifikasikan *tweets* masyarakat terhadap layanan *Commuter Line* kedalam sentimen *complain* dan *not complain* dengan menggunakan metode *Naive Bayes*[13]. Pendekatan *Naive Bayes* memiliki kelebihan artinya kecepatan perhitungannya cepat, dan algoritmanya sederhana dan akurat[14]. Metode ini memiliki tingkat akurasi tertinggi dalam pengklasifikasian analisis sentimen[4].

Penelitian ini dilakukan dengan menerapkan teknik analisis data untuk meningkatkan kualitas layanan publik, khususnya di bidang transportasi massal yang didukung teknologi informasi. *Twitter* digunakan sebagai sumber data untuk analisis sentimen, yang memberikan keuntungan dalam memungkinkan evaluasi yang cepat dan relevan. Hal ini sangat bermanfaat untuk pengelolaan transportasi yang lebih responsif dan berbasis kebutuhan pengguna. Tujuan penelitian ini adalah mengklasifikasikan tweet masyarakat tentang layanan *Commuter Line* ke dalam sentimen keluhan (*complain*) dan bukan keluhan (*not complain*) menggunakan metode *Naive Bayes*. Dalam penelitian ini, algoritma klasifikasi *Naive Bayes* digunakan untuk menentukan apakah tweet tersebut memiliki perspektif positif atau negatif.

Penelitian ini memusatkan perhatian pada penggunaan analisis sentimen *Twitter* untuk menilai dan meningkatkan kualitas layanan KRL *Commuter Line* Jabodetabek. Dengan metode

Naive Bayes, penelitian ini mengklasifikasikan tweet masyarakat ke dalam kategori keluhan (*complaint*) dan bukan keluhan (*non-complaint*). Perbedaan utama dengan penelitian sebelumnya terletak pada fokus yang diarahkan pada sektor transportasi massal, pemanfaatan *Twitter* sebagai platform utama, serta pendekatan klasifikasi sentimen yang lebih spesifik dan relevan untuk perbaikan layanan.

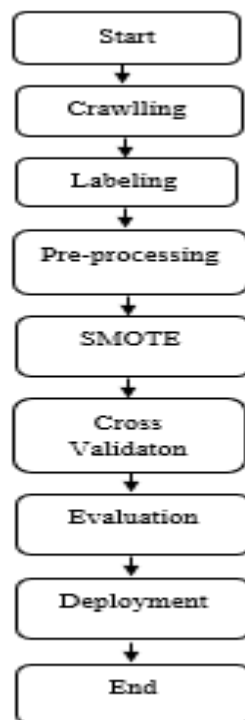
Secara keseluruhan, penelitian ini memberikan kontribusi dalam pengembangan metodologi analisis sentimen menggunakan Naive Bayes pada media sosial, khususnya di Indonesia, serta memberikan wawasan yang bermanfaat bagi pengelola transportasi publik dalam meningkatkan kualitas layanan berdasarkan opini pengguna. Dengan demikian, penelitian ini memiliki potensi untuk memperbaiki pengalaman pengguna serta memberikan pendekatan berbasis data untuk kebijakan transportasi publik.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode *Naive Bayes*. *Naive Bayes Classifier* adalah metode klasifikasi statistik bisa memprediksi probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel yang diberikan termasuk dalam kelas tertentu. Metode ini dipergunakan pada penelitian dengan klasifikasi data yang diambil dari *Twitter* dan data tersebut akan diklasifikasikan menjadi kelas *complain* dan *not complain*.

Naive Bayes classifier ialah metode klasifikasi berakar pada teorema *Bayes*, yaitu memprediksi peluang di masa akan datang berdasarkan pengalaman di masa sebelumnya. Ciri utama dari *Naive Bayes classifier* berdasarkan asumsi terkuat dari masing-masing kondisi atau kejadian[15]. *Naive Bayesian Classifier* menganggap bahwa keberadaan suatu atribut atau variabel bersifat independen dari keberadaan atribut lainnya, karena asumsi bahwa atribut-atribut tersebut tidak saling bergantung satu sama lain[16].

Berikut tahapan penelitian untuk analisa sentimen masyarakat menggunakan *Naive Bayes*:



Gambar 1 Tahapan Penelitian

Tahapan-tahapan dari penelitian yang dilakukan seperti pada gambar 1 adalah sebagai berikut:

2.1 Crawling Data Twitter

Tahapan penelitian mulai dengan:

1. Pengumpulan data berupa opini masyarakat mengenai pelayanan *Commuter Line* pada media sosial *twitter* dengan cara random dan dalam kondisi paling baru. Pengambilan *tweets* berdasarkan kata kunci yang telah ditentukan yakni *@Commuterline* diperoleh 1.010 *tweets*.
2. *Remove Duplicatet*. Setelah dilalui proses tersebut, data yang akan dianalisis sentimen adalah bagian text atau komentar terhadap *Commuter Line*. Maka dilakukan proses *Select Attribute* untuk memilih bagian *text* pada data.
3. Labeling Data. *Labeling* harus dilakukan oleh anotator ahli. Dimana label akan dibedakan kedalam sentimen *complain* untuk *tweet* yang mengandung reaksi atau opini negatif sedangkan dan sentimen *not complain* untuk *tweet* yang mengandung reaksi atau opini positif.

2.2 Pre-Processing

Pre-processing tahap 1 yang dilakukan adalah melakukan:

1. *Cleansing* data sebelum nantinya diolah oleh Rapidminer. Proses *cleansing* bertujuan untuk menormalisasi data kotor. Proses *cleansing* dilakukan dengan mengupload data pada *website* (www.gataframework.com/textmining). Untuk prosesnya dapat dilihat pada gambar 3. Data yang di upload dalam bentuk file excel. Ada 6 langkah yang harus dilakukan untuk menormalisasi data kotor pada *website gataframework* yaitu:

- a. *@Anotation Removal*: menghapus tanda *annotation* (@) yang terdapat pada *text tweet*.
- b. *#(Hashtag) removal*: menghapus kata kunci *hashtag*.
- c. *Removing UR*: menghilangkan URL biasanya terdapat pada *text tweet*.
- d. *Regexp*: menghilangkan simbol-simbol terdapat pada *text tweet*.
- e. Indonesian *steaming*: memperkecil jumlah indeks yang berbeda dari satu sehingga sebuah kata memiliki *suffix* maupun *prefix* akan kembali ke bentuk dasarnya.
- f. Indonesia *stop word removal*: menghilangkan kata penghubung didalam bahasa indonesia misalnya adalah “di” dan “yang”.

2. *Pre-processing* tahap 2(dua) dengan menggunakan aplikasi Rapidminer. Tahapan dimulai dengan mengimport data excel data yang sudah di *cleaning* pada proses sebelumnya kemudian data excel yang akan di *convert* menjadi *nominal to text* untuk mengubah sebuah tipe data yang ada di tiap kalimat, kemudian melalui tahap *Process Document From* data yang berfungsi untuk mengolah data dari *Transform Case*, *Tokenize*, *Filter Token*, dan *Stemming*. Proses selanjutnya yaitu, *Synthetic Minority Oversampling Technique* (SMOTE) agar tidak terjadi *imbalance* data atau kondisi dimana antar kelas memiliki selisih yang signifikan, Untuk mengatasi masalah ketidakseimbangan kelas, salah satu metode yang digunakan adalah tehnik *Synthetic Minority Oversampling Technique* (SMOTE) yang mengubah distribusi data antara kelas mayoritas dan kelas minoritas dalam *dataset* untuk menyeimbangkan jumlah data di setiap kelas[17]. Berikut gambar proses dokumen yang diolah *Tokenize*: proses pemotongan string *input* berdasarkan tiap kata komposisinya. *Tokenize* dilakukan untuk kata per kata dari suatu teks kalimat.

- a) *Transform Cases*: proses untuk membuat data menjadi huruf kecil semua.
- b) *Filter Token by Length (4,25)*: proses memperoleh kata-kata panjangnya antara 4 sampai 25 karakter.
- c) *Steaming Imbuhan*: proses membandingkan data distributor dihasilkan dengan menghilangkan imbuhan data seperti “di...”, “an...”, “nya”...

3. *Pre-processing* tahap 3(tiga), dimana pengujian dilakukan dengan menambahkan salah satu fungsi operator yaitu *filter stopwords*. Teknik menggunakan operator *filter stopword* (*Dictionary*) berfungsi membuang kata yang memiliki informasi yang rendah dalam sebuah teks. Data *text stopword* di *import* dari data hasil nilai pembobotan masing-masing kata. Semakin kecil nilai pembobotannya, maka kata tersebut tidak ada kolerasinya dengan objek penelitian. Proses selanjutnya yaitu *Adaboost*. Tujuannya adanya *Adaboost* yaitu untuk memberikan nilai bobot

yang lebih (*boosting*). Kemudian *Cross Validation (k-fold)* adalah metode untuk mengevaluasi model atau algoritma yang bertujuan memisahkan data menjadi data *training* dan data validasi[18]. Dari tahapan *pre-processing* tersebut akan di evaluasi menggunakan *table confusion matrix*:

- a. Nilai *accuracy* adalah presentase jumlah *record* data diklasifikasikan secara benar oleh sebuah algoritma bisa membuat klasifikasi setelah dilakukan pengujian pada hasil klasifikasi tersebut[19]
- b. Nilai *Precision* atau dikenal juga dengan nama *confidence* merupakan proporsi jumlah kasus diprediksi positif yang juga positif benar pada data yang sebenarnya[19]
- c. Nilai dari *recall* atau *sensitivity* merupakan proporsi jumlah kasus positif yang sebenarnya diprediksi positif secara benar[19]
- d. Nilai AUC menggambarkan hasil pengukuran kesesuaian model secara keseluruhan yang digunakan. Nilai AUC meningkat memiliki arti bahwa variabel yang diteliti semakin baik memprediksi kejadian[18].

2.3 Deployment

Tahap terakhir dari proses *data mining* adalah *deployment (delivery, support dan feedback)*. *Deployment*, setelah melakukan tahap *communication, planning, modeling, dan construction*, maka dilanjutkan ditahap akhir yaitu *deployment*. Pada tahap ini menghasilkan sistem siap digunakan oleh pengguna. Namun sebelum sistem bisa dipergunakan dan diimplementasikan secara langsung, diperlukan beberapa proses instalasi agar sistem bisa dijalankan. Maka diperlukan spesifikasi perangkat keras dan spesifikasi perangkat lunak[20]

3. HASIL DAN PEMBAHASAN

Naive Bayes adalah metode prediksi probabilistik sederhana didasarkan pada penerapan teorema Bayes dengan asumsi kuat akan independensi antar variabel. Keunggulan dari metode ini terletak pada kebutuhan data latih yang relatif sedikit untuk dapat memperkirakan parameter yang diperlukan dalam proses klasifikasi[21][22]. Naive Bayes adalah metode perhitungan yang mampu mengelompokkan variabel tertentu berdasarkan probabilitas dan pendekatan faktual. Metode ini memanfaatkan konsep dalam matematika yang dikenal sebagai hipotesis probabilitas untuk menentukan kemungkinan terbaik dari hasil potensial, dengan cara mengevaluasi frekuensi kemunculan setiap karakteristik dalam data pelatihan[16].

Dari metode dan tahapan penelitian yang sudah dilakukan, maka proses selanjutnya adalah pelabelan data. Dari hasil proses labeling dalam penelitian ini didapatkan 407 *tweet* dengan sentimen *complain* dan 596 dengan sentimen *not complain*.

Tabel 2 Labeling

No	Text	Class
1.	@CommuterLine TOLONG amanahnya.apa karena 24K kalian gak gubris ya.mau berapa nominal lagi yang tidak kalian refund ke pelanggan? Harus ditunggu berapa lama lagi? DM no respond, email habis respon ilang2an. CLAIM SEJAK 26 AGUSTUS LHOOOO. 24K memang kecil, tapi berapa pelanggan???? https://t.co/D5uo4XkY8o	Complain
2.	@CommuterLine pagi kak... AC nya mati di gerbong no.10 samping kursi prioritas. KA4087. Makasi https://t.co/Yng6YlK34P	Complain
3.	RT @CommuterLine: #RekanCommuters Berikut penyesuaian jadwal perjalanan CommuterLine lintas Manggarai - Angke PP pada tanggal 13-31 Oktober...	Not Complain
4.	Hidup itu seperti mengendarai sebuah sepeda. Untuk menjaga keseimbangan, kamu harus terus bergerak' (Albert Einstein). #fotosepur #wonderfullindonesia @mas_didiek @keretaapikita @KAI121 @GM_MarKA @EdanSepurBD @EdanSepurID @CommuterLine @jalur5_@jurnalrailfans @AnkerTwitter https://t.co/5SHraSlpqq	Not Complain
5.	@CommuterLine ini pengelola parkir dalam stasiun Bogor gak garcep, genangan air dari kemarin petugasnya diem aja dan aspalnya	Complain

	juga pada rusak. Jangan cari untung aja, ini foto hanya satu titik padahal 50% banyak genangan https://t.co/PONMdOOMII	
....
1.003	@icaraysha @CommuterLine Gaboleh lah yg begini mah skuy ke kak @byusniess mau beli followers, like, views subscriber youtube all sosmed needs terpercaya? cuma di aku? cus dm or https://t.co/770vtbmekB ? t. jual followers folls ig wtb folls #zonauang #zonaba #zonaBU https://t.co/CViNpCuLhh	Not Complain

3.1 Hasil Pre-Processing

Tahap *Pre-Processing* tahap 1 ini dilakukan dengan tujuan untuk menormalisasi data kotor misalnya *hashtag*, *username*, angka, simbol, tanda baca, serta kata-kata yang tidak penting.

Tabel 3 Hasil Normalisasi

No	Text	Status	@Anotation Removal	# (HashTag) removal	Transformation: Remove URL	Indonesian Stop word removal
1	@CommuterLine TOLONG amanahnya.apa karena 24K kalian gak gubris ya.mau berapa nominal lagi yang tidak kalian refund ke pelanggan? Harus ditunggu berapa lama lagi? DM no respond, email habis respon ilang2an. CLAIM SEJAK 26 AGUSTUS LHOOOO. 24K memang kecil, tapi berapa pelanggan???? https://t.co/D5uo4XkY8o	Complain	tolong amanahnya.apa karena 24k kalian gak gubris ya.mau berapa nominal lagi yang tidak kalian refund ke pelanggan? harus ditunggu berapa lama lagi? dm no respond, email habis respon ilang2an. claim sejak 26 agustus lho000. 24k memang kecil, tapi berapa pelanggan???? https://t.co/d5uo4xky8o	tolong amanahnya.apa karena 24k kalian gak gubris ya.mau berapa nominal lagi yang tidak kalian refund ke pelanggan? harus ditunggu berapa lama lagi? dm no respond, email habis respon ilang2an. claim sejak 26 agustus lho000. 24k memang kecil, tapi berapa pelanggan???? https://t.co/d5uo4xky8o	tolong amanahnya.apa karena 24k kalian gak gubris ya.mau berapa nominal lagi yang tidak kalian refund ke pelanggan? harus ditunggu berapa lama lagi? dm no respond, email habis respon ilang2an. claim sejak 26 agustus lho000. 24k memang kecil, tapi berapa pelanggan????	tolong amanah gubris nominal refund pelanggan tunggu dm no respond email habis respon ilang an claim agustus lho000 pelanggan
2	@CommuterLine pagi kak.AC nya mati di gerbong no.10 samping kursi prioritas. KA4087. Makasi https://t.co/Yng6YLK34P	Complain	pagi kak. ac nya mati di gerbong no.10 samping kursi prioritas. ka4087. makasi https://t.co/yng6ylk34p	pagi kak. ac nya mati di gerbong no.10 samping kursi prioritas. ka4087. makasi https://t.co/yng6ylk34p	pagi kak. ac nya mati di gerbong no.10 samping kursi prioritas. ka4087. makasi	pagi ac mati gerbong no samping kursi prioritas ka makasi
3	RT @CommuterLine: #RekanCommuters Berikut penyesuaian jadwal perjalanan CommuterLine lintas Manggarai - Angke PP pada tanggal 13-31 Oktober	Not Complain	rt #rekancommuters berikut penyesuaian jadwal perjalanan commuteline lintas manggarai - angke pp pada tanggal 13-31 oktober	rt berikut penyesuaian jadwal perjalanan commuteline lintas manggarai - angke pp pada tanggal 13-31 oktober	rt berikut penyesuaian jadwal perjalanan commuteline lintas manggarai - angke pp pada tanggal 13-31 oktober	rt penyesuaian jadwal jalan commuteline lintas manggarai angke pp tanggal oktober
...
1.003	@icaraysha @CommuterLine Gaboleh lah yg begini mah skuy ke kak @byusniess mau beli followers, like, views subscriber youtube all sosmed needs terpercaya? cuma di aku? cus dm or https://t.co/770vtbmekB ? t. jual followers folls ig wtb folls #zonauang #zonaba #zonaBU	Not Complain	gaboleh lah yg begini mah skuy ke kak mau beli followers, like, views subscriber youtube all sosmed needs terpercaya? cuma di aku? cus dm or https://t.co/770vtbmekB ? t. jual followers folls ig wtb folls #zonauang #zonaba #zonabu https://t.co/cvinpculhh	gaboleh lah yg begini mah skuy ke kak mau beli followers, like, views subscriber youtube all sosmed needs terpercaya? cuma di aku? cus dm or https://t.co/770vtbmekB ? t. jual followers folls ig wtb folls https://t.co/cvinpculhh	gaboleh lah yg begini mah skuy ke kak mau beli followers, like, views subscriber youtube all sosmed needs terpercaya? cuma di aku? cus dm or t. jual followers folls ig wtb folls	gaboleh mah skuy beli followers like views subscriber youtube all sosmed needs percaya cus dm or t jual followers folls ig wtb folls

https://t.co/CViNpCuLhh					
---	--	--	--	--	--

Tahapan selanjutnya adalah pengujian data yang sudah dinormlisasi kedalam aplikasi RapidMiner. Dari proses pengujian data tersebut diperoleh hasil sebagai berikut:

Tabel 4 Hasil *Pre-processing* tahap 2

<i>Technique/Feature</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>
<i>Filter Tokens by Lenth (4,25)</i>	70,89%	74,56%	78,20%	0,544
SMOTE	79,53%	80,14%	78,70%	0,739
STEMMING	78,61%	80,41%	75,84%	0,736

Pada proses *stemming* tersebut didapatkan hasil *Accuracy* 78,61%, *Precision* 80,41%, *Recall* 75,84% dan *AUC* 0.736 dari 407 sentimen *complain* dan 596 sentimen *not complain*. Kemudian dilakukan pembobotan nilai pada dataset distribusi yang dihasilkan. Dari hasil pembobotan jika semakin kecil nilai pembobotannya, maka kata tersebut tidak ada kolerasinya dengan *object* penelitian.

Tabel 5 Hasil Pembobotan

Attribute	Bobot
abai	0,4263
abanb	0,7088
abang	20,8232
abiz	0,2691
ablm	0,2605
.....
youtube	13,762

Dari tabel 5 tersebut menjelaskan hasil dari pemberian nilai bobot masing-masing kata yang didapat dari *simple distribution* data. Bobot nilai tertinggi yang didapatkan dari penelitian adalah 58,0754.

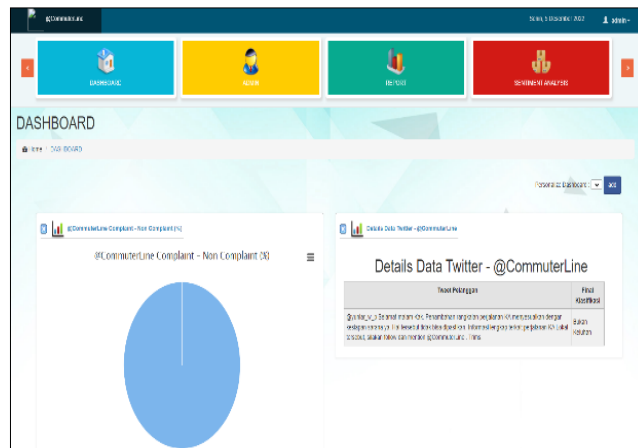
Tabel 6 Hasil *Pre-processing* tahap 3

<i>Technique/Feature</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>
<i>Stopword</i>	79,28%	80,70%	77,19%	0,743
<i>STOPWORD TERBAIK+ADABOOST</i>	78,11%	81,76%	72,51%	0,814

Pada proses *stopword* tersebut didapatkan hasil *Accuracy* 79,28%, *Precision* 80,70%, *Recall* 77,19% dan *AUC* 0.743 dari 407 sentimen *complain* dan 596 sentimen *not complain*. Dan dari hasil proses *stopword* yang sudah ditambahkan dengan *adaboost* didapatkan hasil *Accuracy* 78,11%, *Precision* 81,76%, *Recall* 72,51% dan *AUC* 0,814.

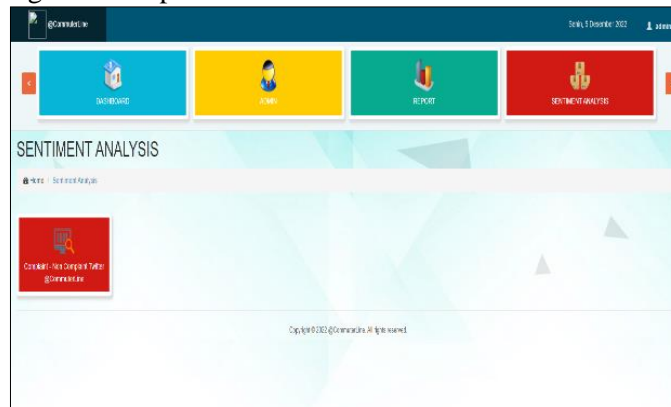
3.2 Hasil Deployment

Pada tahap *deployment* dilakukan *input database* dari *dataset* yang telah diproses terlebih dahulu pada tahap 1 dan tahap 2. *Deployment* berdasarkan hasil evaluasi proses pengujian model antara algoritma *Naive Bayes* dengan model algoritma *Naive Bayes* ditambah dengan fitur *Syntethic Minority Over Sampling Technique Method* (SMOTE). SMOTE digunakan untuk menangani masalah ketidakseimbangan kelas dalam data, dengan cara membuat sampel sintesis untuk kelas minoritas, sehingga model dapat mempelajari data dengan lebih baik dan hasil prediksi menjadi lebih akurat.



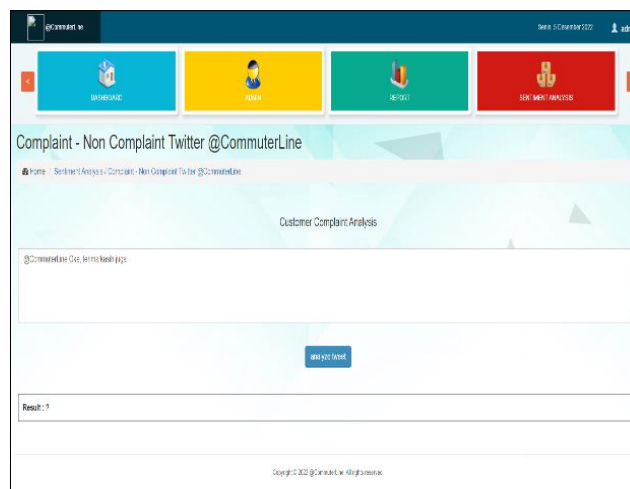
Gambar 2 Dashboard

Tahap *deployment* pada Gambar 5 ini sebagai penerapan model yang sudah dievaluasi, di mana model tersebut akan dipergunakan untuk melakukan prediksi atau pengambilan keputusan berdasarkan data yang telah diinput ke dalam sistem.



Gambar 3 Sentimen Analisis

Pada gambar 6 bahwa aplikasi ini dirancang untuk menganalisis data sentimen, dengan tujuan evaluasi berdasarkan opini pengguna.



Gambar 4 Tweet Analisis Sentimen

Pada gambar 7 menunjukkan *deployment* untuk menganalisis sentimen, dari data Twitter, untuk kemudahan penggunaan dan pengelolaan informasi.

4. KESIMPULAN DAN SARAN

Pada penelitian ini telah berhasil dilakukan mengenai Analisis Sentimen terhadap respon masyarakat pada Media Sosial *Twitter* terhadap pelayanan *Commuter Line* dengan menggunakan *algoritma Naive Bayes* ke dalam sentimen *complain* dan *not complain*. Data *tweets* yang diunduh sebanyak 1.003 melalui aplikasi RapidMiner dan digunakan untuk pelatihan model. Berdasarkan hasil klasifikasi, maka 407 merupakan *tweets complain* dan 596 merupakan *tweet not complain*. Setelah melalui beberapa tahapan penelitian maka didapatkan hasil dengan perolehan nilai *accuracy* 78,11%, *precision* 81,76%, *recall* 72,51%, dan AUC yang didapat sebesar 0,814. Untuk penelitian selanjutnya menggunakan metode data mining yang lain dengan dua metode dan dengan data lebih banyak dihitung.

DAFTAR PUSTAKA

- [1] F. Solihin, S. Awaliyah, and A. M. A. Shofa, "Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika," *J. Pendidik. Ilmu Pengetah. Sos.*, vol. 2021, no. 13, pp. 52–58, 2021.
- [2] N. Tri Romadloni, I. Santoso, and S. Budilaksono, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [3] R. I. Pratiwi, F. Adams, E. Ermatita, and N. Chamidah, "Sentimen Analisis Media Sosial Twitter Terhadap Layanan First Media Menggunakan Metode Naive Bayes," in *Senamika*, 2021, no. September, pp. 300–310.
- [4] A. D. Wulandari, R. R. Saedudin, and R. A. Andreswari, "Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada RUU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.
- [5] E. Dwianto and M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naive Bayes dan Support Vector Machine," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 1, p. 94, 2021, doi: 10.22441/format.2021.v10.i1.009.
- [6] P. P. Andreyana, Y. Pratama, E. Kharisma Krisnadi, I. Purnamasari, and D. Dwi Saputra, "Text Mining untuk Sentimen Analisis dengan Metode Naive Bayes, SMOTE, N-Gram dan AdaBoost Pada Twitter CommuterLine," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 2, pp. 961–973, 2022, doi: <http://dx.doi.org/10.30645/j-sakti.v6i2.506>.
- [7] M. Saraswati and D. Rimirasih, "Analisis Sentimen Terhadap Pelayanan Krl Commuterline Berdasarkan Data Twitter Menggunakan Algoritma Bernoulli Naive Bayes," *J. Ilm. Inform. Komput.*, vol. 25, no. 3, pp. 225–238, 2020, doi: 10.35760/ik.2020.v25i3.3256.
- [8] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, "Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [9] P. Arsi, B. A. Kusuma, and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier," *J. Inform. Upgris*, vol. 7, no. 1, pp. 1–6, 2021, doi: 10.26877/jiu.v7i1.7636.
- [10] T. Agustiranti, A. K. Izzati Kurdiana, B. Al Ghiffari, E. Dwi Juniar, and D. Gita Purnama, "Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh)," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 297–305, 2024, doi: 10.55338/jikomsi.v7i1.2946.
- [11] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020, doi:

- 10.30864/jsi.v15i1.332.
- [12] N. Wayan, A. Sekar, and A. R. Isnain, "Analisis Sentimen Terhadap Media Sosial Twitter dengan Kasus Kampanye Anti-Korupsi di Indonesia Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 8, no. April, pp. 695–703, 2024, doi: 10.30865/mib.v8i2.7582.
 - [13] S. Ernawati, E. R. Yulia, Frieyadie, and Samudi, "Implementation of the Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–5, 2019, doi: 10.1109/CITSM.2018.8674286.
 - [14] H. F. Putro, R. T. Vulandari, and W. L. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Tikomsin*, vol. 8, no. 2, 2020.
 - [15] 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, [Online]. Available: <https://core.ac.uk/download/pdf/299924402.pdf>.
 - [16] E. Martantoh, N. Yanih, T. Informatika, U. Panca, and S. Bekasi, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa ' adah Menggunakan PHP MySQL," *JTSI*, vol. 3, no. 2, pp. 166–175, 2022.
 - [17] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.
 - [18] E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 578–586, 2023, doi: 10.47065/josyc.v4i3.3414.
 - [19] D. Saputra, B. Pratama, Y. Akbar, W. G.-C. O. SPOT, and undefined 2018, "Penerapan Text Mining Untuk Assingment Complaint Handling Customer Terhadap Divisi Terkait Menggunakan Metode Decission," *Jurnal.Stikomcki.Ac.Id*, vol. 11, no. 2, pp. 207–216, 2018, [Online]. Available: <http://jurnal.stikomcki.ac.id/index.php/cos/article/view/52>.
 - [20] A. Sasmita, G. A. Pradnyana, and D. G. H. Divayana, "Sistem Analisis Sentimen Untuk Evaluasi Kinerja Dosen dengan Metode Naïve Bayes," *J. Sains Teknol.*, vol. 11, no. 2, pp. 451–462, 2022.
 - [21] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023.
 - [22] G. Firmansyah and A. Hermawan, "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Kesegaran Buah jeruk," *J. Inform.*, vol. 10, no. 2, pp. 180–184, 2023.