

Affandi, Hasan - cek.01

by Irfan Ricky Affandi, Firman Noor Hasan

Submission date: 19-Aug-2022 05:56AM (UTC+0700)

Submission ID: 1884107338

File name: Affandi,_Hasan_-_cek.01.doc (901.5K)

Word count: 4160

Character count: 26334

Sentiment Analysis Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan 10-Fold Cross Validation dan Metode Naive Bayes

Irfan Ricky Afandi¹, Firman Noor Hasan^{#2}, Nunik Pratiwi³, Mia Kamayani⁴, Zuhri Halim⁵

^{1,2,3,4,5}Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka
Jl. Tanah Merdeka No.6, Kec. Pasar Rebo, DKI Jakarta 13830, Indonesia

¹richky99@gmail.com

²firman.noorhasan@uhamka.ac.id

³npratiwi@uhamka.ac.id

⁴mia.kamayani@uhamka.ac.id

⁵zuhri@uhamka.ac.id

Abstrak— Peningkatan jumlah transaksi pada berbagai platform e-commerce **5** mempunyai implikasi terhadap peningkatan penggunaan jasa ekspedisi. Salah satu jasa ekspedisi yang ada di Indonesia yaitu anteraja, dimana perusahaan tersebut harus bisa memiliki inovasi untuk mempunyai hubungan serta memberikan pelayanan yang terbaik bagi penggunanya. Saat ini banyak pengguna layanan Anteraja mempunyai pendapat yang beragam terhadap layanan mereka pada media sosial twitter. Penelitian ini menerapkan teknik *sentiment analysis* untuk membantu mengevaluasi, menganalisis, menilai, serta mengetahui sikap masyarakat terhadap pelayanan Anteraja. Metode untuk mengkategorikan sentimen yang digunakan oleh peneliti yaitu menerapkan algoritma *naive bayes* yang mempunyai akurasi tinggi, serta prosesnya sederhana dan cepat. Peneliti juga menggunakan bantuan perangkat lunak *python* untuk proses pengambilan *dataset* pada *twitter* serta *rapidminer studio* untuk pengolahan data serta penerapan algoritma. Hasil dari proses pengolahan data yang dilakukan oleh peneliti didapatkan dari 1180 data, jumlah kategori yang paling banyak yaitu kategori sentimen positif sebesar 638 lalu kategori sentimen negatif sebesar 493 sedangkan paling sedikit yaitu kategori sentimen netral sebanyak 49. Hal ini menunjukkan dari 1180 data bahwa banyak masyarakat yang menyukai pelayanan yang diberikan oleh jasa ekspedisi Anteraja, namun tidak sedikit masyarakat yang masih kurang puas terhadap pelayan yang diberikan. Nilai akurasi penerapan algoritma *naive bayes* dalam penelitian ini diper **16** persentase sebesar 85.06% yang menunjukkan bahwa data tersebut dapat digunakan sebagai dasar bagi perusahaan untuk pertimbangan pengambilan keputusan.

Kata kunci— analisis sentimen, naive bayes, pelayanan, twitter, python, rapidminer studio, anteraja

Abstract— The increase in the number of transactions on various e-commerce platforms has implications for increasing the use of shipping services. One of the expedition services in Indonesia, namely Anteraja, is where the company must be able to have innovation to have a relationship and provide the best service for its users. Currently, many users of Anteraja services have diverse opinions on their services on Twitter. This research applies

*sentiment analysis techniques to help evaluate, analyze, assess, and know people's attitudes towards service delivery. The sentiment categori method used by the researcher is using the Naive Bayes algorithm which has high accuracy, and the process is simple and fast. Researchers also use python software for the dataset retrieval process on Twitter and rapidminer studio for data processing and implementation algorithm. The results of the data processing were obtained 1180 data, the most number of categories was positive at 638 and negative sentiment at 493 while the least was neutral sentiment as much as 49. This shows from 1180 data that many people like the services provided by the anteraja expedition service, but not a few people are still not satisfied with the service provided. The accuracy value of the application of the Naive Bayes algor **6** in this study obtained a percentage of 85.95% which indicates that the data can be used as the basis for companies to make decisions.*

Keywords— sentiment analysis, naive bayes, service, twitter, python, rapidminer studio, anteraja

I. PENDAHULUAN

Kemajuan teknologi internet sangat membantu manusia untuk mendukung kegiatannya sehari-hari [1]. Hal ini disebabkan oleh kebutuhan penggunaan teknologi internet terus mengalami peningkatan karena dapat digunakan sebagai media untuk berkomunikasi serta dapat meningkatkan efisiensi kinerja manusia seperti dalam pekerjaan jasa ekspedisi [2]. Pada tahun 2019 APJII melakukan survei terhadap penggunaan internet di Indonesia dimana hasilnya menunjukkan adanya kenaikan sebesar 73%, tentunya hal ini akan terus mengalami peningkatan [3]. Peluang tersebut dapat dimanfaatkan oleh pelaku bisnis eksp **10** untuk melakukan pengembangan terhadap bisnisnya [4]. Salah satu perusahaan di Indonesia yang bergerak pada bidang jasa ekspedisi pengiriman barang yaitu perusahaan Anteraja [5].

Persaingan bisnis di zaman industri 4.0 juga terus meningkat dan membuat suatu perusahaan ekspedisi harus dapat mengikuti arus dari perkembangan teknologi agar bisnisnya dapat terus berkembang [6]. Hal terpenting bagi

sebuah perusahaan ekspedisi yaitu bagaimana perusahaan tersebut dapat menjalin hubungan yang baik dengan penggunanya [7]. Salah satu caranya yaitu dengan memberikan pelayanan yang sangat ramah serta yang terbaik kepada pengguna, dimana apabila pengguna merasa puas terhadap layanan tersebut maka akan sangat membantu perusahaan untuk mencapai target laba secara maksimal [8]. Apabila pelayanan yang diberikan kepada pengguna kurang baik maka akan menjadi pertimbangan kembali oleh pengguna apakah akan menggunakan jasa ekspedisi tersebut atau tidaknya [9].

Suatu perusahaan ekspedisi pasti mempunyai bagian *customer service* untuk memberikan pelayanan serta menerima keluhan ataupun masalah yang dihadapi oleh penggunanya dalam menggunakan jasa tersebut [10]. Pengguna dapat memberikan bagaimana tentang pendapat terhadap penggunaan pelayanan jasa ekspedisi tersebut melalui platform sosial media *twitter* [11]. Sosial media *twitter* yaitu sebuah platform aplikasi yang dapat digunakan serta dijadikan sebagai wadah untuk seseorang memberikan sebuah penilaian terhadap suatu pelayanan [12]. Perusahaan jasa ekspedisi Anteraja juga mempunyai akun *official twitter @anteraja_id* sebagai *customer service* yang disediakan oleh perusahaan untuk menerima kritik, pendapat, keluhan serta saran dari penggunanya. Pendapat yang beragam dari pengguna layanan pada suatu perusahaan jasa ekspedisi dapat dijadikan sebagai bahan untuk melakukan proses peningkatan pelayanan kepada penggunanya [13].

Semua kritik, pendapat, keluhan serta saran dari pengguna layanan jasa pada sosial media *twitter* dapat di kumpulkan menjadi satu yang disebut dengan sentimen [14]. Setiap sentimen dapat dikategorikan menjadi kategori sentimen positif, kategori sentimen netral, serta kategori sentimen negatif yang merupakan representasi dari sudut pandang pengguna terhadap layanan yang diberikan oleh suatu perusahaan [15]. Kumpulan dari opini tersebut nantinya dapat dijadikan sebagai sumber data untuk melakukan proses *sentiment analysis* [16]. *Sentiment Analysis* yaitu teknik yang cukup umum digunakan dalam hal yang bersangkutan untuk melakukan perbandingan terhadap suatu sudut pandang, penilaian, sentimen, sikap serta emosi seseorang terkait topik dimana nantinya akan diekspresikan dalam bentuk teks [17].

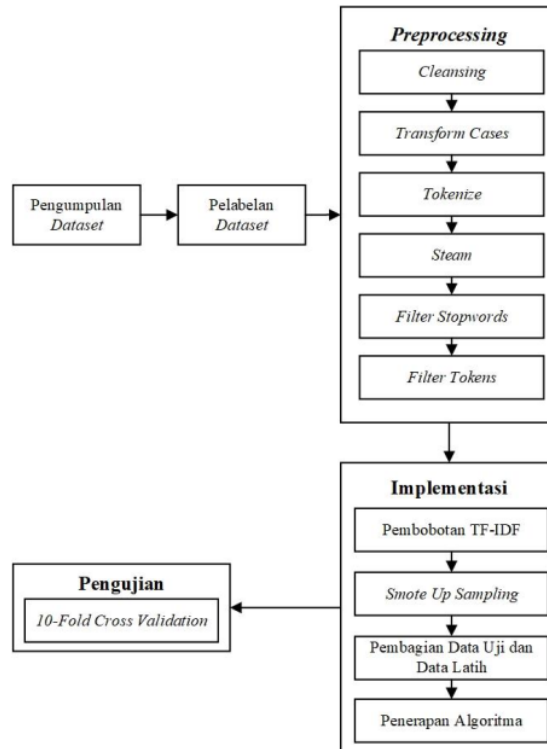
Pada kegiatan bisnis proses *sentiment analysis* dapat digunakan oleh perusahaan pelayanan jasa untuk melakukan peningkatan layanan jasa [18]. Penerapan algoritma klasifikasi *naive bayes* juga cukup umum digunakan dalam proses *sentiment analysis* karena memiliki kelebihan yang dimana algoritma tersebut mempunyai akurasi tinggi, serta prosesnya sederhana dan cepat [19]. Berdasarkan uraian diatas dalam pembuatan penelitian ini peneliti mendapatkan *dataset* menggunakan perangkat lunak python tentang opini masyarakat mengenai pelayanan jasa ekspedisi yang ditujukan kepada akun *official twitter* perusahaan Anteraja yaitu @anteraja_id.

Selanjutnya *tweet* yang sudah dikumpulkan kemudian diolah menggunakan perangkat lunak *rapidminer studio* dan akan dikategorikan menjadi kategori sentimen positif, kategori sentimen netral, serta kategori sentimen negatif dengan

menerapkan algoritma *naive bayes* untuk mengetahui tingkat nilai akurasinya.

II. METODE PENELITIAN

Proses pengolahan data yang dilakukan oleh peneliti melalui beberapa tahapan yang ditampilkan pada gambar diagram berikut.



Gambar. 1 Diagram pada tahapan penelitian

A. Pengumpulan Dataset

Pada tahap ini peneliti mendapatkan *dataset* pada media sosial *twitter* terhadap *tweet* yang berisi opini pengguna dengan kata kunci pencarian “pelayanan” yang ditujukan pada akun *official twitter* perusahaan Anteraja yaitu @anteraja_id menggunakan *library twint*.



Gambar. 2 Proses penarikan data *tweet* pada akun @anteraja_id

Peneliti melakukan proses penarikan data dengan menggunakan kata kunci “pelayanan” yang ditujukan kepada akun *official twitter* perusahaan Anteraja yaitu @anteraja_id dimana data tersebut diambil pada tanggal 1 Januari 2021 sampai pada tanggal 16 Agustus 2022 dengan teks berbahasa

Indonesia dan didapatkan sebanyak 1194 data *tweet* yang dimana data tersebut di simpan dalam format *file* yaitu *xlsx*.

	tweet
0	@anteraja_id dear team, mohon konfirmasinya ke...
1	@anteraja_id pelayanan super buruk
2	@anteraja_id @OfficialKT48 Sayang sekali pelay...
3	@anteraja_id no respon? Pelayanan mengecewakan...
4	@anteraja_id woi...gmn neh paket nya...gak nya...
...	...
1189	@anteraja_id Gimana sih min pelayanan nya. Dar...
1190	@anteraja_id Sungguh sangat kecewa menggunakan...
1191	@anteraja_id mhn konfirmasi donk pengiriman 10...
1192	@anteraja_id saya kecewa dengan pelayanan anda...
1193	@anteraja_id min, bilang ke management anteraj...

Gambar. 3 Hasil proses penarikan data *tweet*

Gambar diatas menampilkan beberapa contoh hasil proses penarikan data *tweet* yang berkaitan dengan opini masyarakat terhadap layanan Anteraja yang ditunjukkan pada akun @anteraja_id menggunakan bahasa pemrograman *python* serta *library twint*.

B. Pelabelan Dataset

Tahap berikutnya yaitu melakukan proses pelabelan secara manual oleh lima orang peneliti dengan melakukan pengambilan suara terbanyak pada setiap *tweet* dengan menggunakan bantuan aplikasi *microsoft excel 365* seperti pada gambar berikut.

tweet	kategori
kota bekasi harus muter2 ke gudang sortir jakarta utara? tolong dibantu no resi 10005463378565. apa tidak ada pembenahan setelah banyak konsumen komplain dgn pelayanan kalian?	negatif
@anteraja_id pelayanan super buruk	negatif
schedule waktu pick up dg alasan "parsel belum siap" padahal semua sudah beres sebelum jam yg dipilih	negatif
@anteraja_id no respon? Pelayanan mengecewakan ini. Email no respon jg.	negatif
minggu...konsumen udah pada komplek.komplek ke antar aja gak ada respon..smua dicobaim gak ada respon..jgn iklan doank di gedin tp pelayanan ke konsumen nol besar..	negatif
https://t.co/u0eKUG68bO	negatif

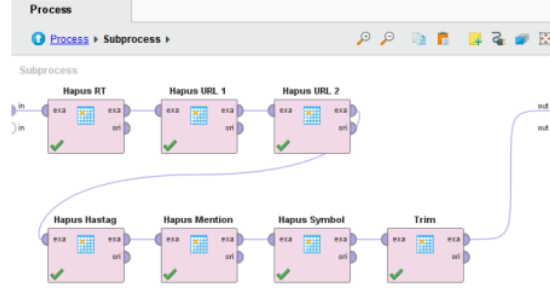
Gambar. 4 Hasil proses pelabelan data *tweet*

Setelah dilakukan proses pelabelan serta menghilangkan *tweet* yang sama didapatkan total keseluruhan data yaitu sebanyak 1180 data, dimana untuk jumlah kategori yang paling banyak yaitu kategori sentimen positif sebesar 638 lalu kategori sentimen negatif sebesar 493 sedangkan paling sedikit yaitu kategori sentimen netral sebanyak 49.

C. Preprocessing Data

Selanjutnya yaitu melakukan *preprocessing dataset* terhadap *tweet* yang sudah diberikan pelabelan menggunakan perangkat lunak *rapidminer studio*. *Preprocessing* yaitu tahapan mengganti teks tidak teratur supaya menjadi teks yang teratur tentunya hal ini dapat membantu tahap pengolahan data selanjutnya [20]. Terdapat beberapa tahap dalam *preprocessing dataset* seperti *cleansing*, *transform case*, *tokenize*, *steam*, *filter stopwords*, serta *filter tokens by length* [21].

Tahapan *cleansing* yaitu proses untuk menghapus semua *retweet*, *URL*, *hashtag*, *mention*, serta *symbol* pada *dataset* [22]. Peneliti menggunakan operator *replace* untuk menghilangkan *retweet*, *URL*, *hashtag*, *mention*, serta *symbol* seperti pada gambar berikut.



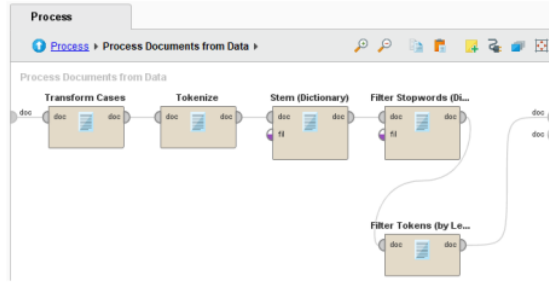
Gambar. 5 Operator *cleansing* yang digunakan peneliti

Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *cleansing*.

TABEL I
HASIL *CLEANSING*

Sebelum	Sesudah	Kategori
@anteraja_id anteraja emang bagus pelayanan ramah ☺☺	anteraja emang bagus pelayanan ramah	Positif
@anteraja_id min untuk idul fitri apa ada libur pelayanan dan pengiriman? Kalau iya berapa hari ya?	min untuk idul fitri apa ada libur pelayanan dan pengiriman Kalau iya berapa hari ya	Netral
@anteraja_id pelayanan super buruk 🙄	pelayanan super buruk	Negatif
@anteraja_id Semenjak adanya https://t.co/2HADwg1drL semakin mudah melakukan apapun. Pelayanan sangat cepat, aman dan sangat ramah. Terima kasih https://t.co/2HADwg1drL	Semenjak adanya semakin mudah melakukan apapun Pelayanan sangat cepat aman dan sangat ramah Terima kasih	Positif
@anteraja_id @anteraja_id kecewa bgt si sama pelayanan nya	kecewa bgt si sama pelayanan nya	Negatif
@anteraja_id Pelayanan super cepat banget sih, kurirnya juga responsif 🙏	Pelayanan super cepat banget sih kurirnya juga responsif	Positif

Setelah *dataset* sudah dihilangkan *retweet*, *URL*, *hashtag*, *mention*, serta *symbol* selanjutnya peneliti menggunakan beberapa operator *process document* pada *rapidminer studio* untuk menampung tahapan *transform case*, *tokenize*, *steam*, *filter stopwords*, serta *filter tokens by length* seperti pada gambar berikut.



Gambar. 6 Operator yang digunakan dalam *process document*

Transform case yaitu proses untuk mengganti kata yang berkapital pada *dataset* supaya kate tersebut menjadi tidak kapital [23]. Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *transform case*.

TABEL II
HASIL *TRANSFORM CASE*

Sebelum	Sesudah	Kategori
anteraja emang bagus pelayanan ramah	anteraja emang bagus pelayanan ramah	Positif
min untuk idul fitri apa ada libur pelayanan dan pengiriman Kalau iya berapa hari ya	min untuk idul fitri apa ada libur pelayanan dan pengiriman kalau iya berapa hari ya	Netral
pelayanan super buruk	pelayanan super buruk	Negatif
Semenjak adanya semakin mudah melakukan apapun Pelayanan sangat cepat aman dan sangat ramah Terima kasih	semenjak adanya semakin mudah melakukan apapun pelayanan sangat cepat aman dan sangat ramah terima kasih	Positif
kecewa bgt si sama pelayanan nya	kecewa bgt si sama pelayanan nya	Negatif
Pelayanan super cepat banget sih kurirnya juga responsif	pelayanan super cepat banget sih kurirnya juga responsif	Positif

Selanjutnya merupakan tahap *tokenize* yaitu proses untuk memisahkan suatu kalimat menjadi beberapa kata untuk memudahkan proses *steam* serta *stopword* [24]. Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *tokenize*.

TABEL III
HASIL *TOKENIZE*

Sebelum	Sesudah	Kategori
anteraja emang bagus pelayanan ramah	anteraja, emang, bagus, pelayanan, ramah	Positif
min untuk idul fitri apa	min, untuk, idul,	Netral

Sebelum	Sesudah	Kategori
ada libur pelayanan dan pengiriman kalau iya berapa hari ya	fitri, apa, ada, libur, pelayanan, dan, pengiriman, kalau, iya, berapa, hari, ya	
pelayanan super buruk	pelayanan, super, buruk	Negatif
semenjak adanya semakin mudah melakukan apapun pelayanan sangat cepat aman dan sangat ramah terima kasih	semenjak, adanya, semakin, mudah, melakukan, apapun, pelayanan, sangat, cepat, aman, dan, sangat, ramah, terima, kasih	Positif
kecewa bgt si sama pelayanan nya	Kecewa, bgt, si, sama, pelayanan, nya	Negatif
pelayanan super cepat banget sih kurirnya juga responsif	pelayanan, super, cepat, banget, sih, kurirnya, juga, responsif	Positif

Tahap selanjutnya melakukan *steam* yaitu proses untuk menghilangkan semua kata imbuhan dan mengubah kata menjadi bentuk dasarnya pada *dataset* [25]. Proses ini dilakukan secara manual dimana peneliti membuat kamus *steam* yang berpedoman terhadap kamus besar bahasa Indonesia (KBBI). Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *steam*.

TABEL IV
HASIL *STEAM*

Sebelum	Sesudah	Kategori
anteraja, emang, bagus, pelayanan, ramah	anteraja, memang, bagus, layanan, ramah	Positif
min, untuk, idul, fitri, apa, ada, libur, pelayanan, dan, pengiriman, kalau, iya, berapa, hari, ya	Min, untuk, idul, fitri, apa, ada, libur, layanan, dan, kirim, kalau, iya, berapa, hari, ya	Netral
pelayanan, super, buruk	layanan, super, buruk	Negatif
semenjak, adanya, semakin, mudah, melakukan, apapun, pelayanan, sangat, cepat, aman, dan, sangat, ramah, terima, kasih	sejak, ada, makin, mudah, lakukan, apa, layanan, sangat, cepat, aman, dan, sangat ramah, terima, kasih	Positif
kecewa, bgt, si, sama, pelayanan, nya	kecewa, bgt, si, sama, layanan, nya	Negatif
pelayanan, super, cepat, banget, sih, kurirnya, juga, responsif	layanan, super, cepat, banget, sih, kurir, juga, responsif	Positif

Selanjutnya merupakan tahap *filter stopwords* yaitu proses untuk menghilangkan semua penghubung serta kata-kata yang

tidak diperlukan pada *dataset* [26]. Kamus *stopwords* yang digunakan peneliti didapatkan dari situs www.keaja.com. Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *filter stopwords*.

TABEL V
HASIL FILTER STOPWORD

Sebelum	Sesudah	Kategori
anteraja, memang, bagus, layanan, ramah	anteraja, bagus, layanan, ramah	Positif
min, untuk, idul, fitri, apa, ada, libur, layanan, dan, kirim, kalau, iya, berapa, hari, ya	min, apa, ada, libur, layanan, kirim, iya, hari, ya	Netral
layanan, super, buruk	layanan, super, buruk	Negatif
sejak, ada, makin, mudah, lakukan, apa, layanan, sangat, cepat, aman, dan, sangat ramah, terima, kasih	ada, mudah, apa layanan, cepat, aman, ramah, terima, kasih	Positif
kecewa, bgt, si, sama, layanan, nya	kecewa, bgt, si, sama, layanan, nya	Negatif
layanan, super, cepat, banget, sih, kurir, juga, responsif	layanan, super, cepat, banget, sih, kurir, responsif	Positif

Tahap selanjutnya melakukan *filter tokens by length* yaitu proses untuk menghilangkan kata dengan jumlah huruf yang sudah ditentukan [27]. Dimana dalam penelitian ini peneliti menggunakan minimal panjang huruf yaitu sebanyak empat dan maksimal panjang huruf yaitu dua puluh dua. Tabel berikut memperlihatkan beberapa contoh *tweet* sebelum dan sesudah dilakukan proses *filter tokens by length*.

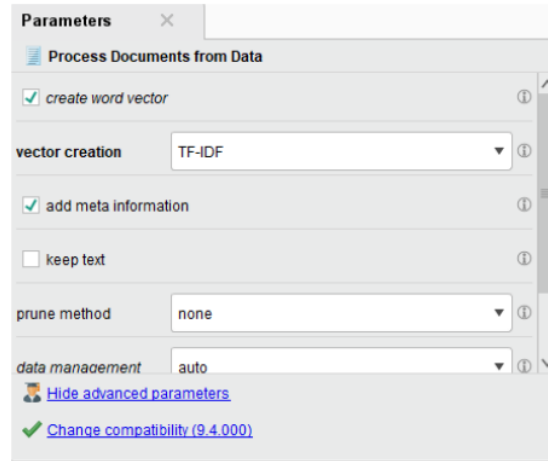
TABEL VI
HASIL FILTER BY LENGTH

Sebelum	Sesudah	Kategori
anteraja, bagus, layanan, ramah	anteraja, bagus, layanan, ramah	Positif
min, apa, ada, libur, layanan, kirim, iya, hari, ya	libur, layanan, kirim, hari	Netral
layanan, super, buruk	layanan, super, buruk	Negatif
apa, mudah, apa layanan, cepat, aman, ramah, terima, kasih	mudah, layanan, cepat, aman, ramah, terima, kasih	Positif
kecewa, bgt, si, sama, layanan, nya	kecewa, sama, layanan	Negatif
layanan, super, cepat, banget, sih, kurir, responsif	layanan, super, cepat, banget, kurir, responsif	Positif

III. HASIL PEMBAHASAN

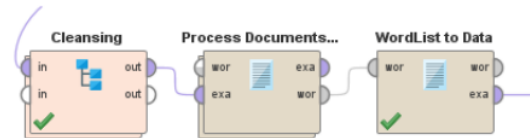
A. Pembobotan TF-IDF

Tahapan ini dilakukan untuk memberikan nilai bobot kepada kata-kata yang ada pada *tweet* setelah melewati tahap *preprocessing*. Penggunaannya dengan caranya yaitu memilih *vector creation* TF-IDF pada operator *process document* yang sebelumnya digunakan untuk menampung tahapan *transform case*, *tokenize*, *stem*, *filter stopwords*, serta *filter tokens by length* seperti pada gambar berikut.



Gambar. 7 Pembobotan TF-IDF

Langkah selanjutnya yaitu menghubungkan operator *process document* ke operator *WordList to Data* untuk menghitung nilai bobot pada kata serta frekuensi kemunculan dari kata tersebut pada *dataset* seperti gambar berikut.



Gambar. 8 Menghubungkan operator *process document* dengan operator *wordlist to data*

Berikut merupakan gambar yang menunjukkan hasil dari proses pembobotan TF-IDF yang menampilkan jumlah frekuensi kemunculan sebuah kata pada *dataset*.

Open in Turbo Prep Auto Model Filter (83)

Row No.	word	in docs...	total	in class (ne...	in class (net...	in class (po...
1	layanan	1166	1211	510	48	653
2	cepat	497	518	36	9	473
5	ramah	361	372	5	2	365
3	paket	354	404	205	13	185
4	kirim	335	395	126	3	266
7	aman	323	343	11	7	325
6	anteraja	307	349	144	12	193
9	pakai	239	255	89	2	164
10	respon	231	240	44	3	193
8	banget	211	263	79	5	179
11	kurir	193	223	114	4	105
12	barang	151	179	84	4	91
14	puas	115	117	16	2	99
13	jasa	113	119	39	1	79
15	online	106	106	2	1	103

ExampleSet (833 examples, 0 special attributes, 6 regular attributes)

Gambar. 9 Hasil proses pembobotan TF-IDF

Berdasarkan gambar diatas menunjukkan bahwa kata yang paling banyak muncul yaitu layanan sebanyak 1166 kata, dilanjutkan dengan kata cepat sebanyak 497 kata, kemudian kata ramah sebanyak 361 kata. Setelah melakukan pembobotan TF-IDF peneliti juga melakukan *data visualisation* kepada setiap kata yang sudah dilakukan pembobotan kedalam bentuk *wordcloud* untuk menampilkan 20 kata yang memiliki frekuensi kemunculan sering digunakan seperti pada gambar berikut.



Gambar. 10 *Data visualisation* dalam *wordcloud*

Berdasarkan gambar hasil *data visualisation* dalam bentuk *wordcloud* diatas menunjukkan bahwa semakin besar suatu ukuran kata maka kata tersebut merupakan jumlah kata yang paling banyak muncul. Kata yang sering muncul pada dataset yaitu kata layanan, dilanjutkan dengan kata cepat, ramah, paket, kirim, aman, anteraja, pakai respon, banget kemudian kata kurir.

B. Implementasi Algoritma

Dalam penelitian ini peneliti menggunakan pemodelan algoritma *naive bayes* terhadap *dataset* yang sudah dilakukan tahapan pembobotan TF-IDF. Dimana dalam hal ini data uji serta data latih pada *dataset* secara otomatis akan dibagi menggunakan perangkat lunak *rapidminer studio* yang terbagi menjadi tiga kategori sentimen yaitu kategori sentimen positif, kategori sentimen netral serta kategori sentimen negatif.

Berikut merupakan contoh proses perhitungan algoritma *naive bayes* secara manual terhadap *tweet* yang diambil secara acak pada *dataset* yang sudah melewati tahap *preprocessing* dan pembobotan TF-IDF. Contoh *tweet* yang digunakan yaitu "layanan super buruk" dimana sebelumnya *tweet* tersebut sudah diberikan label secara manual dengan kategori negatif berikut tabel frekuensi katanya.

TABEL VII
FREKUENSI KATA PADA TWEET

Kata	Kelas Positif	Kelas Netral	Kelas Negatif	Total
layanan	653	48	510	1211
super	12	0	6	18
buruk	0	2	100	102

Nilai kelas positif :

$$= \left(\frac{653}{1211} \times \frac{12}{18} \times \frac{0}{102} \right) \times \frac{638}{1180} = 0$$

Nilai kelas netral :

$$= \left(\frac{48}{1211} \times \frac{0}{18} \times \frac{2}{102} \right) \times \frac{49}{1180} = 0$$

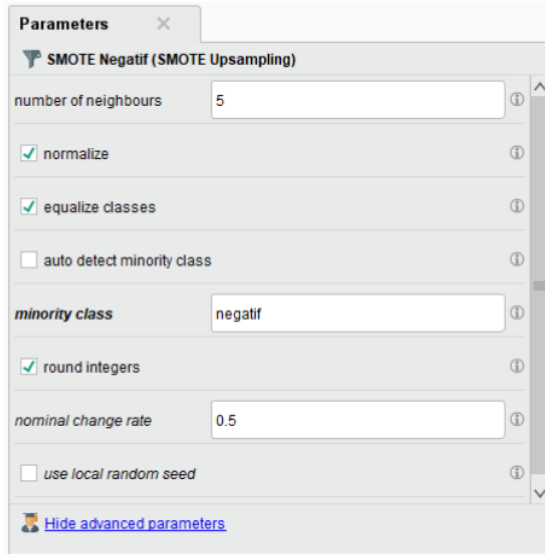
Nilai kelas negatif :

$$= \left(\frac{510}{1211} \times \frac{6}{18} \times \frac{100}{102} \right) \times \frac{493}{1180} = 0,057$$

Kesimpulan bahwa *tweet* tersebut memiliki kategori "negatif", hal ini menunjukkan bahwa hasil pelabelan secara manual dan pelabelan menggunakan perhitungan algoritma *naive bayes* didapatkan kategori "negatif".

C. Smote Up Sampling

Pada saat melakukan pelabelan secara manual didapatkan bahwa kategori sentimen yang didapatkan mempunyai label yang tidak sama dengan label terbanyak yaitu kategori positif. Sebelum menuju tahap pengujian *10-fold cross validation* label tersebut perlu di seimbangkan terlebih dahulu dengan menggunakan operator *smote up sampling* pada *rapidminer studio*. Penggunaan operator *smote* dilakukan dua kali untuk menyesuaikan pada label dengan kategori netral serta label dengan kategori negatif. Berikut gambar tampilan parameter *smote up sampling* yang digunakan.

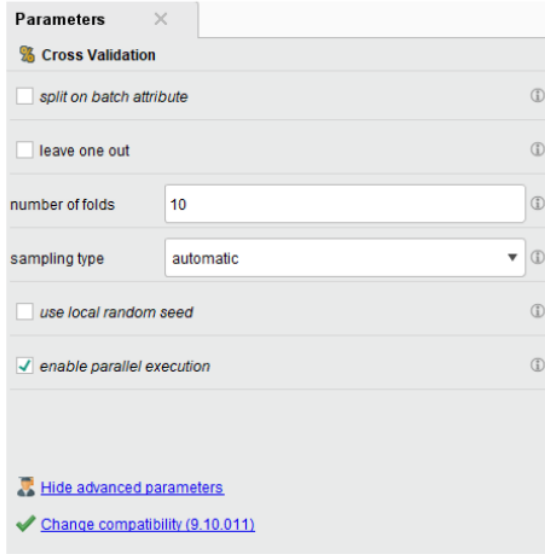


Gambar. 11 Parameter *smote up sampling* yang digunakan

Gambar diatas merupakan parameter *smote up sampling* yang digunakan untuk melakukan penyesuaian pada label dengan kategori negatif. Peneliti tidak mencentang pada bagian *auto detect minority class* karena pada bagian kolom *minority class* peneliti mengisikannya dengan nama kategori yang belum seimbang sebagai contoh pada gambar diatas peneliti menuliskannya untuk kategori negatif.

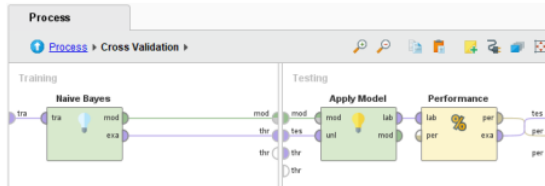
D. Pengujian Evaluasi

Setelah *dataset* melalui tahap *smote up sampling* selanjutnya peneliti melakukan pengujian *10-fold cross validation* untuk melihat hasil klasifikasi serta keakuratan data dari tahapan penerapan algoritma *naive bayes*. Operator yang digunakan yaitu *Cross Validation* dan berikut merupakan tampilan parameter dari operator tersebut yang digunakan oleh peneliti.



Gambar. 12 Parameter *cross validation* yang digunakan

Setelah menetapkan parameter yang digunakan selanjutnya menghubungkan antara data latih serta data uji ke algoritma *naive bayes* untuk melakukan evaluasi penerapan algoritma tersebut dengan memakai operator *performance* seperti pada gambar berikut.



Gambar. 13 Tampilan tahapan pada operator *cross validation*

Setelah menjalankan semua operator maka luaran yang dihasilkan dari proses ini dengan menggunakan aplikasi *rapidminer studio* seperti pada gambar dibawah ini yang menunjukkan nilai akurasi dari penerapan algoritma *naive bayes* dan pengujian *10-fold cross validation*.

	true negatf	true netral	true positif	class precision
pred. negatf	528	0	52	91.03%
pred. netral	54	638	124	78.19%
pred. positif	56	0	452	89.19%
class recall	92.70%	100.00%	72.41%	

Gambar. 14 Tampilan hasil *performance vector*

Pada gambar diatas memperlihatkan hasil dari pengujian *dataset* dengan menggunakan *10-fold cross validation* serta algoritma *naive bayes* pada aplikasi *rapidminer studio* didapatkan nilai akurasinya sebesar 85,06%.

IV. KESIMPULAN

Dari hasil pemrosesan *dataset* sebanyak 1180 data *tweet* untuk mengetahui bagaimana pandangan pengguna terhadap pelayanan yang diberikan oleh jasa ekspedisi Anteraja yang ditujukan pada akun @anteraja_id dari tanggal 1 Januari 2022 sampai 16 Agustus 2022 didapatkan kategori sentimen yang paling banyak yaitu kategori sentimen positif yaitu sebanyak 638 data kemudian untuk kategori sentimen negatif yaitu sebanyak 493 data, sedangkan untuk kategori sentimen netral yaitu hanya 49 data. Diman saat ini banyak masyarakat yang merasa puas terhadap pelayanan yang diberikan oleh perusahaan jasa anteraja terlihat dari banyaknya jumlah tweet yang memiliki kategori positif. Namun masih ada beberapa masyarakat yang merasa kurang puas terhadap pelayanan yang diberikan yang terlihat dari jumlah kategori sentimen negatif yang cukup lumayan banyak. Nilai akurasi dari proses penerapan algoritma *naive bayes* serta pengujian *10-fold cross validation* didapatkan nilai akurasi sebesar 85.06%. Tentunya dari pengolahan *dataset* yang sudah dilakukan dapat dijadikan sebagai bahan pertimbangan pengambilan keputusan untuk melakukan inovasi peningkatan pelayanan yang diberikan perusahaan terhadap penggunanya.

REFERENSI

- [1] I. R. Afandi, Y. Handika, I. F. Hanif, and D. Ismail, "Sistem Informasi Penjualan Online Hasil Tani Desa Blukbuk Berbasis Aplikasi Website," *Pros. Semin. Nas. Teknoka*, vol. 5, no. 2502, pp. 65–72, 2020, doi: 10.22236/teknoka.v5i1.334.
- [2] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.
- [3] A. K. Dewi and Sulastri, "Analisis Sentimen Ekspedisi Sicepat Dari Ulasan Google Play Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informatika)*, vol. 9, no. 2, pp. 796–805, 2022, doi: 10.35957/jatisi.v9i2.1802.
- [4] M. T. Nitami and H. Februriyanti, "ANALISIS SENTIMEN ANALISIS SENTIMEN ULASAN EKSPEDISI J&T EXPRESS MENGGUNAKAN ALGORITMA NAIVE BAYES," *J. Manaj. Inform. Sist. Inf.*, vol. 5, p. 29, 2022, [Online]. Available: https://www.google.com/search?q=jurnal+misi+stmik+lombok&rlz=C1CHBF_enID993ID993&oq=jurn&aqs=chrome.69i59j69i57j0i13li433i512j0i433i512i2j69i60i2j69i61.1.2564j0j7&sourceid=chrome&ie=UTF-8
- [5] L. N. Rahman and Wahyudin, "Optimalisasi Penugasan Karyawan Jasa Ekspedisi Menggunakan Metode Hungarian (Studi Kasus CV. Anteraja Cabang Mekarmukti)," *J. Serambi Eng.*, vol. 6, no. 3, pp. 2120–2127, 2021, doi: 10.32672/jse.v6i3.3245.
- [6] B. Nugraha, "Metode Klasifikasi Analisis Sentimen pada Media Sosial," *Syntax J. Inform.*, vol. 9, no. 2, pp. 109–118, 2020, doi: 10.35706/syji.v9i2.3593.
- [7] M. N. Akbar, Darmatiasia, and Y. Ardana, "Analisis Sentimen Terhadap Jasa Ekspedisi Pos Indonesia Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier," *J. SHIFT*, vol. 2, pp. 42–51, 2022.
- [8] W. Rohmatiana and A. Sadiqin, "Analisis Tingkat Kepuasan Pelanggan Terhadap Pelayanan Online Shop Kara Project," *J. Ekon. Manajemen, Bisnis dan Sos.*, vol. 2, no. November, pp. 62–67, 2021, [Online]. Available: <https://www.embiss.com/index.php/embiss/article/view/63>
- [9] F. Prasetyawan, S. Widiyanesti, and T. Widarmanti, "Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Barang Sicepat Di Media Sosial Twitter," in *e-Proceeding of Management*, 2022, vol. 9, no. 2, pp. 1–14.
- [10] D. P. Isnarwaty and Irahmah, "Text Clustering pada Akun TWITTER Layanan Ekspedisi JNE, J&T, dan Pos Indonesia Menggunakan Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dan K-Means," *J. Sains dan Seni ITS*, vol. 8, no. 2, pp. 2–9, 2019, doi: 10.12962/j23373520.v8i2.49094.
- [11] T. A. Lorosae, B. D. Prakoso, Saifudin, and Kusriani, "Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan Naive Bayes," *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 25–30, 2018.
- [12] F. Rizqi Irawan, A. Jazuli, and T. Khotimah, "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.v5i1.4267.
- [13] A. R. Rizqullah, A. Wedhasmara, R. I. Heroza, A. Putra, F. Fathoni, and P. Putra, "Analisis Masalah Pada Data Review Aplikasi Terhadap Layanan E-Commerce Menggunakan Metode Text Classification," *J. Tekno Kompak*, vol. 16, no. 1, p. 186, 2022, doi: 10.33365/jtk.v16i1.1448.
- [14] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [15] J. J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naive Bayes and K-Nearest Neighbor Meth," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2019, doi: 10.25126/jtiik.202294960.
- [16] R. Akmalia, I. Slamet, and H. Pratiwi, "Analisis Sentimen Twitter Berbahasa Indonesia Terhadap Aplikasi," 2022.
- [17] J. D. C. Aruan, B. Rahyudi, and A. Ridok, "Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan Term Frequency – Inverse Document Frequency," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 5, pp. 2072–2078, 2022.
- [18] A. R. Prananda and I. Thalib, "Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 1, p. 1, 2020, doi: 10.20473/jisebi.6.1.1-8.
- [19] N. I. P. Kalingara, O. N. Pratiwi, and H. D. Anggana, "Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naive Bayes," in *e-Proceeding of Engineering*, 2021, vol. 8, no. 5, pp. 9035–9048.
- [20] I. Verawati and B. S. Audit, "Algoritma Naive Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By . u," *J. Media Inform. Budidarma*, vol. 6, pp. 1411–1417, 2022, doi: 10.30865/mib.v6i3.4132.
- [21] Asyharudin *et al.*, "COMPARING ALGORITHM FOR SENTIMENT ANALYSIS IN HEALTHCARE AND SOCIAL SECURITY AGENCY (BPJS KESEHATAN)," *J. Comput. Inf. Technol.*, vol. 19, no. 1, pp. 31–37, 2022.
- [22] A. Asro'i and F. Hery, "Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor," *J. Khatulistiwa Inform.*, vol. 10, no. 1, pp. 17–24, 2022.
- [23] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [24] R. Slamet, W. Gata, A. Novtarianty, K. Hilyati, and F. A. Jariyah, "ANALISIS SENTIMEN TWITTER TERHADAP PENGGUNAAN ARTIS KOREA SELATAN SEBAGAI BRAND AMBASSADOR PRODUK KECANTIKAN LOKAL," *J. Inf. Technol. Comput. Sci.*, vol. 5, pp. 9–25, 2022.
- [25] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [26] A. Aziz and Fauziah, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 115–125, 2022.
- [27] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

ORIGINALITY REPORT

5%

SIMILARITY INDEX

5%

INTERNET SOURCES

2%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	Prista Afikah, Irfan Ricky Affandi, Firman Noor Hasan. "Implementasi Business Intelligence Untuk Menganalisis Data Kasus Virus Corona di Indonesia Menggunakan Platform Tableau", Pseudocode, 2022 Publication	1%
2	journal.lppmunindra.ac.id Internet Source	<1%
3	polgan.ac.id Internet Source	<1%
4	seputar-labuhanbatu.blogspot.com Internet Source	<1%
5	blog.uad.ac.id Internet Source	<1%
6	dokumen.tips Internet Source	<1%
7	shift.sin.fst.uin-alauddin.ac.id Internet Source	<1%

8

Muhammad Ezar Al Rivan, Molavi Arman, Hafiz Irsyad, Reynald Dwika Prameswara. "Klasifikasi Hewan Mamalia Berdasarkan Bentuk Wajah Menggunakan Fitur Histogram of Oriented dan Metode Support Vector Machine", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2022

Publication

<1 %

9

ejurnal.teknokrat.ac.id

Internet Source

<1 %

10

repository.ipb.ac.id

Internet Source

<1 %

11

repository.uhamka.ac.id

Internet Source

<1 %

12

vdocuments.site

Internet Source

<1 %

13

Oman Somantri, Dairoh Dairoh. "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2019

Publication

<1 %

14

media.neliti.com

Internet Source

<1 %

15

repository.upi.edu

Internet Source

<1 %

Exclude quotes Off
Exclude bibliography On

Exclude matches Off