

# Implementasi Algoritma *Naive Bayes* Pada Analisis Sentimen Terhadap Pelayanan Jasa Pengiriman Berdasarkan Opini *Twitter*

1<sup>st</sup> Abdul Majid  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if19.abdulmajid@mhs.ubpkarawang.ac.id  
082246053395

2<sup>nd</sup> Amril Mutoi Siregar  
Universitas Buana Perjuangan  
Karawang, Indonesia  
amrilmutoi@ubpkarawang.ac.id

3<sup>rd</sup> Deden Wahiddin  
Universitas Buana Perjuangan  
Karawang, Indonesia  
dedenwahiddin@mhs.ubpkarawang.ac.id

**Abstrak**— Tujuan penelitian ini menganalisis yang hasilnya dapat di olah bahan untuk evaluasi bagi pelayanan jasa pengiriman. Dalam penelitian ini data yang di *Crawling Data* dengan berbentuk format *excel* terbagi menjadi 3 kelas yaitu kelas *positif*, *negatif*, dan *netral*. Setelah itu data mengalami tahap filter data dari *labelling data* menjadi 2 kategori yaitu kategori kelas *positif* dan kelas *negatif* yang ada di data *tweet* menggunakan metode klasifikasi dengan label *positif* ditandai angka 1, sedangkan label *negatif* ditandai dengan angka -1. Data yang sudah dilakukan pelabelan sebanyak 1325 data *tweet* yang masuk dalam kata kunci "Kurir" terdapat kelas *positif* ada 848 data *tweet*, dan kelas *negatif* ada 477 data *tweet* dengan menggunakan Algoritma *Naive Bayes*. Lalu akan melakukan tahapan *text preprocessing* yang terdiri dari *case folding*, *tokenizing*, *filtering*, *stemming*, dan *TF-IDF* dan akan menghasilkan nilai sentimen *positif* dan *negatif*. Selanjutnya akan melakukan tahap pembagian dataset dibagi menjadi data *training* sebanyak 927 data, dan data *testing* sebanyak 398 data dengan menggunakan rasio 0.3 yang artinya 70% data *training* dan 30% data *testing*. Selanjutnya melakukan tahap evaluasi model klasifikasi algoritma *Naive Bayes* dengan menggunakan *Confusion Matrix*. Dari pengujian tersebut didapatkan hasil analisis sentimen dengan kata kunci "kurir" mendapatkan nilai akurasi yang cukup baik dengan nilai akurasi sebesar 80%, *presion* 84%, dan *Recall* sebesar 87%.

**Kata kunci** — Analisis Sentimen, *Twitter*, Kurir, *Text Mining*, *Naive Bayes*, *Confusion Matrix*.

## I. PENDAHULUAN

Munculnya *e-commerce* di Indonesia berdampak pada meningkatnya permintaan akan jasa pengiriman. Perkembangan banyaknya penjual yang melakukan transaksi pada *platform e-commerce*, kebutuhan akan layanan jasa pengiriman yang pandai untuk mengantarkan pesanan pelanggan ke penerima yang dituju menjadi sangat penting. Akibatnya, lonjakan perusahaan jasa pengiriman semakin banyak. Jumlah perusahaan swasta di bidang jasa pengiriman ekspres yang terdaftar sebagai anggota dari Asosiasi Perusahaan Jasa Pengiriman Ekspres, Pos dan Logistik Indonesia (ASPERINDO) [1].

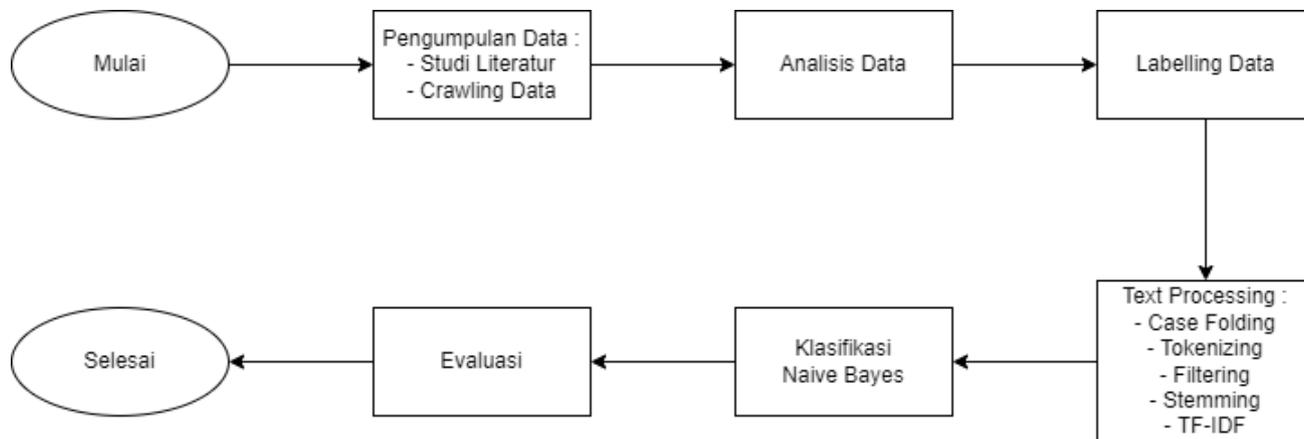
Jasa pengiriman memiliki peran yang penting bagi masyarakat Indonesia, terlebih dalam di bidang perdagangan antara produsen dan konsumen. Jasa pengiriman tersedia melalui beberapa jalur, yaitu ada jalur darat, jalur laut, dan jalur udara. Hal ini mengakibatkan terjadi perubahan kehidupan masyarakat terutama di wilayah kota besar. Mempengaruhi banyak konsumen yang berminat pada toko *online* yang memudahkan transaksi tanpa dibatasi oleh waktu (24 jam). Selain itu belanja secara *online* menyediakan perbandingan produk lebih beragam, transaksi lebih cepat, tanpa antrean, rahasia konsumen terlindungi, dan dapat menghemat waktu, tenaga, dan biaya [2]. Namun permasalahan ini memiliki beberapa kendala dalam pelayanan jasa pengiriman salah satunya yaitu hilangnya barang kiriman yang mengakibatkan kerugian terhadap konsumen, pemberitaan tentang jasa pengiriman yang tidak kurang begitu efektif terhadap layanan yang di terima oleh konsumen, dan perlunya solusi permasalahan dengan analisis sentimen yang hasilnya dapat di olah untuk evaluasi bagi penyedia jasa pengiriman. Proses mengetahui seberapa baik dan buruk pelayanan jasa pengiriman tersebut maka diperlukan opini-opini yang akan diambil dari media sosial seperti *twitter*. Pertumbuhan pengguna media sosial maka dapat berkembang pula, data dihasilkan yang berbentuk teks, saat ini telah mencapai jumlah yang sangat besar.

Analisis sentimen, juga disebut sebagai penambangan opini, adalah metode canggih untuk memahami, mengekstraksi dan proses data tekstual secara otomatis untuk memperoleh informasi sentimen berharga yang tertanam dalam sebuah kalimat, apakah itu terdiri dari opini yang disukai atau tidak disukai, sehingga dengan analisis sentimen dapat mengetahui masyarakat terhadap suatu pelayanan atau produk [3]. Adapun beberapa rujukan jurnal penelitian terkait dalam bidang analisis sentimen terhadap suatu topik dengan memanfaatkan data *twitter* yang telah dilakukan. Salah satu pada penelitian [4] tentang klasifikasi opini pengguna media sosial *twitter* terhadap jnt (jet lee dan tony) di Indonesia dengan algoritma *decision tree*, dengan penelitian ini menghasilkan akurasi sebesar 94,55% dengan rasio data pelatihan terhadap data pengujian 90 berbanding 10, akurasi 100% tidak dapat dicapai karena potensi adanya gangguan pada kata yang salah eja dan kata yang tidak ada dalam data pelatihan. Selanjutnya dalam penelitian [5] tentang analisis sentimen untuk penilaian pelayanan situs belanja *online* menggunakan algoritma *Naive Bayes*, dengan menghasilkan akurasi dari proses tercipta dua kelas data yaitu kelas data *positif* dan kelas data *negatif* dengan jumlah data total 1200 data dan hasil evaluasi menunjukkan rata-rata akurasi sebesar 93,33%. Selanjutnya dengan penelitian [6] tentang analisis sentimen data komentar sosial media *Facebook* dengan *K-Nearest Neighbor* (studi kasus pada akun jasa ekspedisi barang jnt (jet lee dan tony chen) menghasilkan rata-rata akurasi tertinggi adalah 79,21%, sedangkan akurasi terendah 70,3%. Selanjutnya pada penelitian [7] tentang analisis sentimen pada ulasan pembelian produk di *marketplace* Shopee menggunakan pendekatan *Natural Language Processing* dengan menghasilkan nilai akurasi 76,92%, presisi 80,00%, dan *recall* 74,07%. Selanjutnya dengan penelitian [8] tentang klasifikasi keluhan masyarakat

pada sosial media *twitter* terhadap pelayanan toko *online* di Indonesia menggunakan metode *cosine TF-IDF*, dengan dapat menghasilkan *persentase* proses sebesar 63,1%.

Penelitian ini menganalisis tanggapan masyarakat terhadap pelayanan jasa pengiriman berdasarkan pada opini *twitter*. Penelitian ini akan menerapkan metode klasifikasi dengan menggunakan Algoritma *Naive Bayes*. Data yang diperoleh dari hasil *Crawling* data pada media sosial seperti *twitter* akan diklasifikasikan menjadi dua kategori analisis sentimen yaitu *positif* dan *negatif*. Dalam penelitian ini diharapkan dapat melihat nilai akurasi yang tertinggi dari metode tersebut.

## II. METODE PENELITIAN



Gambar 1 Alur Penelitian

### A. Pengumpulan Data

Objek penelitian yang digunakan pada penelitian ini adalah opini publik pada sosial media *twitter* melalui pelayanan jasa pengiriman. Data diambil melalui *crawling* sebanyak 10001 data terhitung dari tanggal 01 Januari 2021 sampai 26 Desember 2022. Data *crawling* terbagi menjadi 3 kategori yaitu *positif*, *negatif*, *netral*. Setelah itu data mengalami tahap *filtering* menjadi 2 kategori kelas *positif* dan kelas *negatif*.

### B. Analisis Data

Pada tahapan ini, dimulai dari analisis data dilakukan untuk mengetahui permasalahan pada media *twitter* yaitu opini masyarakat tentang layanan jasa pengiriman. Analisis data dilakukan untuk memperoleh sentimen berupa kelas *positif* dan kelas *negatif*.

### C. Labelling Data

Setelah data dibersihkan, maka tahapan selanjutnya *labelling* data pada setiap data dalam kategori kelas yang berbeda yaitu kelas *positif* dan kelas *negatif*. Label kelas *positif* ditandai dengan angka 1, sedangkan kelas *negatif* ditandai dengan -1.

### D. Text Preprocessing

*Text preprocessing* merupakan sebuah tahapan untuk dilakukannya seleksi atau tahap *filtering* pada dokumen atau dataset, karena tidak semua teks dapat diproses orientasi sentimen. Proses *text preprocessing* adalah praktik yang penting. Ini mencakup seni pelipatan kasus, penghapusan URL, pemisahan konjungsi melalui *Tokenizing*, penghilangan *Stopwords*, dan penyempurnaan penggunaan kata melalui *stemming* dan pembobotan kata, termasuk pendekatan *TF-IDF* [9].

### E. Klasifikasi Algoritma *Naive Bayes*

Algoritma *Naive Bayes* merupakan sebuah algoritma yang termasuk dalam teknik klasifikasi data mining. Keefektifan algoritma *Naive Bayes* sangat dipengaruhi oleh proses pengelompokan setiap atribut dan kelas target dari setiap objek. Sangat penting untuk mempertimbangkan faktor-faktor ini dengan hati-hati untuk mencapai hasil yang optimal. Proses dari kemungkinan setiap kelas [10].

Adapun untuk rumusan dari algoritma *Naive Bayes* sebagai berikut:

$$P(H|X) = \frac{P(X) \cdot P(H)}{P(X)} \tag{1}$$

Keterangan:

$X$  = Data yang akan diklasifikasikan ke dalam kelas tertentu.

$H$  = Hipotesis atau dugaan bahwa data  $X$  termasuk dalam kelas tertentu.

$P(H|X)$  = Probabilitas hipotesis  $H$  atau berdasarkan kondisi (*posteriori probability*)

$P(H)$  = Probabilitas hipotesis  $H$  sebelum melihat data  $X$  atau (*prior probability*)

$P(X|H)$  = Probabilitas data  $X$  berdasarkan kondisi pada hipotesis  $H$ , yaitu probabilitas data  $X$  jika diketahui bahwa data  $X$  termasuk dalam kelas  $H$ .

$P(X)$  = Probabilitas data  $X$  secara umum, tanpa memperhatikan kelas atau hipotesis tertentu.

F. Evaluasi Algoritma *Naïve Bayes*

Pada tahap evaluasi ini yaitu melakukan pengujian model klasifikasi dengan menggunakan *Confusion Matrix*. Adapun untuk langkah-langkah menghitung *Accurary*, *Precision*, dan *Recall* dengan menggunakan *Confusion Matrix* adalah sebagai berikut:

Predicted Class	Aktual Class		
		True Positive	True Negative
	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

1. *Accuracy*

*Accurarcy* adalah untuk mengetahui jumlah data yang telah di klasifikasi secara benar. Adapun untuk rumus perhitungan *Accuracy* sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{2}$$

2. *Precison*

*Precision* adalah perhitungan rasio prediksi yang benar dengan jumlah total prediksi yang dibuat oleh sistem, sehingga memastikan tingkat *precision* dan keandalan yang tinggi dalam keluarannya. Adapun untuk rumus perhitungan *precision* sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

3. *Recall*

*Recall* adalah untuk nilai tingkat keberhasilan dalam mengenali suatu kelas yang perlu dikenali. Adapun untuk rumus perhitungan *recall* sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

Keterangan:

*TP* = True Positif (memprediksi positif dan itu benar)

*TN* = True Negatif (memprediksi negatif dan itu benar)

*FP* = False Positif (memprediksi positif dan itu salah)

*FN* = False Negatif (memprediksi negatif dan itu salah)

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan *Crawling Data* dari analisis sentimen dengan menggunakan media sosial *Twitter* adalah proses mengambil data dari *Twitter* dengan bertujuan untuk menganalisis sentimen atau pandangan *Twitter* terkait topik pelayanan terhadap jasa pengiriman. Proses awal *Crawling Data* dapat dilakukan dengan memanfaatkan *Twitter API* yang memungkinkan pengguna untuk mengakses data publik dari *Twitter*. *Data Crawling* sebanyak 10001 data terhitung dari tanggal 01 Januari 2021 sampai 26 Desember 2022.

	Tanggal Waktu	Id Tweet	Teks	Username	Label
0	2022-12-25 18:26:03+00:00	1607080245869542912	SMITH JANKERMAN Lowongan Kerja Terbaru di PT A...	PasirRinggit	0
1	2022-12-25 08:21:11+00:00	1606928027975578112	@TokopediaCare @SebutirDoang Min, saran aja ni...	ndrrrraaa	1
2	2022-12-25 05:22:48+00:00	1606883135857524992	@Mantiko_Bungo Apa kerja mereka sampe dikasih ...	attarimai	0
3	2022-12-24 05:01:04+00:00	1606515278581752064	Sobat Auto sekarang sudah tau kan? Perbedaan p...	Autocargo_id	0
4	2022-12-23 03:12:44+00:00	1606125628385890048	@punyamoza Pastikan km udh mengaktifkan tipe l...	ShopeeID	0
...	...	...	...	...	...
9996	2021-08-07 10:30:02+00:00	1423954582791627008	@williamgames Terima kasih sudah menggunakan A...	anteraja_id	0
9997	2021-08-07 10:25:42+00:00	1423953489181381120	@ApaYa90681092 Terima kasih sudah menggunakan ...	anteraja_id	0
9998	2021-08-07 10:21:14+00:00	1423952368131973120	@AbestiSelly Terima kasih sudah menggunakan An...	anteraja_id	0
9999	2021-08-07 10:11:53+00:00	1423950012883471872	@delimas72113258 Terima kasih sudah mengguna...	anteraja_id	0
10000	2021-08-07 10:11:11+00:00	1423949836919860992	@anonyme___ Terima kasih sudah menggunakan Ant...	anteraja_id	0

10001 rows x 5 columns

Gambar 2 Hasil Crawling Data

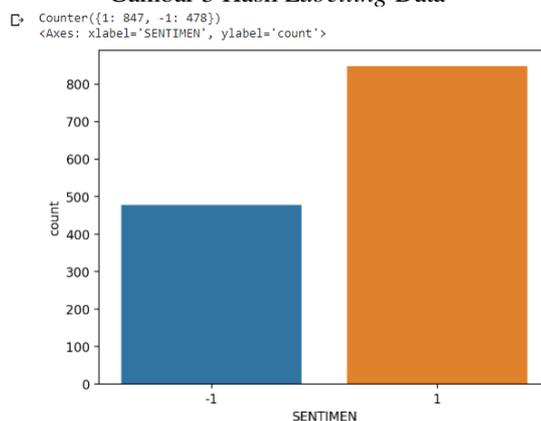
B. Labelling Data

Pada tahapan selanjutnya melakukan *labelling data* pada setiap data dalam kategori kelas yang berbeda yaitu kelas *positif* dan kelas *negatif* dan kelas netral. Proses *labelling* yaitu proses pelabelan data, *labelling* pada penelitian ini di bagi menjadi 2 kategori sentimen berdasarkan data menentukan sentimen *negatif* dan sentimen *positif*. Proses *labelling data* dapat dilakukan dengan cara manual oleh Pionika Linda Sari (Lulusan Sarjana Sastra Indonesia) dengan format *excel*. Pelabelan kalimat *positif* ditandai dengan angka 1, sedangkan pelabelan kalimat *negatif* ditandai angka -1.

	TEKS	SENTIMEN
0	sbt yg suka pick up paket pake jodante kalian ...	1
1	trs ada kurir lg dari sayangku yg sbnrnya ga s...	1
2	trs bangun ada kurir buket duitzzz dari umi ab...	1
3	me right now boxeran doang kaos kutang seringn...	1
4	aoaowkwkwkwk kurir able mazzeh	1
...	...	...
1320	teume ada yang mau di tag few tweet au treasur...	-1
1321	kurir antweaja nih emg ga ninggalin jejak ya m...	-1
1322	ekpedisi entaraja sucks paket gua mutermuter d...	-1
1323	fuck cod udahmah kalo yang beli gaada di rumah...	-1
1324	minimal kalo miskin gausah sok sokan mesen pak...	-1

1325 rows x 2 columns

Gambar 3 Hasil Labelling Data



Gambar 4 Hasil diagram labelling data tweet

Berdasarkan pada gambar diagram di atas menjelaskan bahwa yang ditandai -1 merupakan kategori kelas *negatif* yang berjumlah 478 data *tweet*, sedangkan dan yang ditandai 1 kategori kelas *positif* dengan berjumlah 847 data *tweet*.

C. Text Preprocessing

Dalam pengolahan data teks, *text preprocessing* dapat membantu meningkatkan kualitas teks serta meningkatkan akurasi dalam analisis, data teks, pada tahapan *text preprocessing* ada beberapa terdiri dari *Case Folding*, *Tokenizing*, *Stopwords*, *Stemming*, *TF-IDF* dengan menggunakan *python google colab*.

1. Case Folding

*Case folding* merupakan proses pengubah huruf menjadi *lowercase* atau huruf kecil. Fungsi dari *lower case folding* adalah untuk konsistensi, normalisasi, dan kemudahan pemrosesan dalam data teks.

Tabel 1 Hasil Case Folding

Nomor	Input	Case Folding
1	sbt! yg suka pick up paket pake jodante, kalian chat kurir nya utk pick up pake kata? apa? sender punya social anxiety jd takut chat sm orng lain huhu, tia ya 🙄	sbt yg suka pick up paket pake jodante kalian chat kurir nya utk pick up pake kata apa sender punya social anxiety jd takut chat sm orng lain huhu tia ya
2	trs ada kurir lg dari sayangku yg sbnrnya ga surprise2 bgt si soalnya kespoiler kemarin2 🤔🙄 super happy krn bunganya cantik bgt jujurrr trs wangi lg huhu lov https://t.co/rhk876pE18	trs ada kurir lg dari sayangku yg sbnrnya ga surprise bgt si soalnya kespoiler kemarin super happy krn bunganya cantik bgt jujurrr trs wangi lg huhu lov

2. *Tokenizing*

Pada tahap *Tokenizing* adalah proses pemrosesan teks ke dalam bagian-bagian yang lebih kecil yang disebut token. Token bisa berupa kata, frasa, simbol, atau elemen lainnya dalam teks. Tujuannya untuk teks menjadi lebih terstruktur sehingga memudahkan analisis.

Tabel 2 Hasil *Tokenizing*

Nomor	Case Folding	Tokenizing
1	sbt yg suka pick up paket pake jodante kalian chat kurir nya utk pick up pake kata apa sender punya social anxiety jd takut chat sm orng lain huhu tia ya	['sbt', 'yg', 'suka', 'pick', 'up', 'paket', 'pake', 'jodante', 'kalian', 'chat', 'kurir', 'nya', 'utk', 'pick', 'up', 'pake', 'kata', 'apa', 'sender', 'punya', 'social', 'anxiety', 'jd', 'takut', 'chat', 'sm', 'orng', 'lain', 'huhu', 'tia', 'ya']
2	trs ada kurir lg dari sayangku yg sbnrnya ga surprise bgt si soalnya kespoiler kemarin super happy krn bunganya cantik bgt jujurrrr trs wangi lg huhu lov	['trs', 'ada', 'kurir', 'lg', 'dari', 'sayangku', 'yg', 'sbnrnya', 'ga', 'surprise', 'bgt', 'si', 'soalnya', 'kespoiler', 'kemarin', 'super', 'happy', 'krn', 'bunganya', 'cantik', 'bgt', 'jujurrrr', 'trs', 'wangi', 'lg', 'huhu', 'lov']

3. *Filtering*

Pada tahapan *filtering* ini adalah proses mengambil kata-kata penting dari hasil token untuk bisa melakukan analisis *filtering* ini dapat digunakan algoritma *stopwords* (membuang kata yang kurang penting) atau (menyimpan kata penting).

Tabel 3 Hasil *Stopwords*

Nomor	Tokenizing	Stopwords
1	['sbt', 'yg', 'suka', 'pick', 'up', 'paket', 'pake', 'jodante', 'kalian', 'chat', 'kurir', 'nya', 'utk', 'pick', 'up', 'pake', 'kata', 'apa', 'sender', 'punya', 'social', 'anxiety', 'jd', 'takut', 'chat', 'sm', 'orng', 'lain', 'huhu', 'tia', 'ya']	['suka', 'pick', 'up', 'paket', 'pake', 'jodante', 'chat', 'kurir', 'pick', 'up', 'pake', 'social', 'takut', 'chat']
2	['trs', 'ada', 'kurir', 'lg', 'dari', 'sayangku', 'yg', 'sbnrnya', 'ga', 'surprise', 'bgt', 'si', 'soalnya', 'kespoiler', 'kemarin', 'super', 'happy', 'krn', 'bunganya', 'cantik', 'bgt', 'jujurrrr', 'trs', 'wangi', 'lg', 'huhu', 'lov']	['kurir', 'sayangku', 'surprise', 'kespoiler', 'kemarin', 'super', 'happy', 'bunganya', 'cantik', 'wangi']

4. *Stemming*

Pada tahapan proses *Stemming* digunakan untuk mengubah kata-kata dalam teks menjadi bentuk dasar atau kata dasar dengan menghilangkan imbuhan kata yang berbeda seperti awalan, dan akhiran dalam pemrosesan teks yang lebih efisien.

Tabel 4 Hasil *Stemming*

Nomor	Stopwords	Stemming
1	['suka', 'pick', 'up', 'paket', 'pake', 'jodante', 'chat', 'kurir', 'pick', 'up', 'pake', 'social', 'takut', 'chat']	['suka', 'pick', 'up', 'paket', 'pake', 'jodante', 'chat', 'kurir', 'pick', 'up', 'pake', 'social', 'takut', 'chat']
2	['kurir', 'sayangku', 'surprise', 'kespoiler', 'kemarin', 'super', 'happy', 'bunganya', 'cantik', 'wangi']	['kurir', 'sayang', 'surprise', 'spoiler', 'kemarin', 'super', 'happy', 'bunga', 'cantik', 'wangi']

5. Pembobotan Kata *TF-IDF*

*TF-IDF (Term Frequency-Inverse Document Frequency)* digunakan untuk mengekstrak informasi dari dokumen teks dan melihat seberapa umum atau tidak umum sebuah kata yang ada di antara sekumpulan teks yang terstruktur. Pada proses ekstraksi fitur yaitu mengubah data teks menjadi sebuah angka.

Tabel 5 Kalimat *TF-IDF*

Nomor	Kalimat
1	sbt yg suka pick up paket pake jodante kalian chat kurir nya utk pick up pake kata apa sender punya social anxiety jd takut chat sm orng lain huhu tia ya
2	trs ada kurir lg dari sayangku yg sbnrnya ga surprise bgt si soalnya kespoiler kemarin super happy krn bunganya cantik bgt jujurrrr trs wangi lg huhu lov

3	trs bangun ada kurir buket duitzzz dari umi abah dgn ucapan yg super singkat padat dan jelas tp gpp hadiahnya jg jelas jelas berguna haha request lo aku jam pengantarannya soalnya w mau ke kampus
4	me right now boxeran doang kaos kutang seringnya kaos dilepas pas kurir manggil gelagapan pake baju
...	...
1321	teume ada yang mau di tag few tweet au treasure as kurir paket ga drop aja yaa
1322	kurir antweaja nih emg ga ninggalin jejak ya minimal wasap kak aku taruh disini ya barangnya notif toped juga gaada tbtb pagi pagi mak marah krn ada paket semaleman diluar
1323	ekpedisi entaraja sucks paket gua mutermuter doang ini id kurir ngopi dulu
1324	fuck cod udahmah kalo yang beli gaada di rumah trus abang kurir nya nungguin lama apalagi kalo minta tolong buat stay nungguin paketnya dia like gue masih ada kesibukan lain tolong nyusahin doang anjg emg knp sih harus cod ujung juga tetep bayar dulu sebelum buka paket kan
1325	minimal kalo miskin gausah sok sokan mesen paket cod lah anj nyusahin kurir doang bgst

Adapun menghitung dari hasil TF-IDF sebagai berikut:

1. Menghitung *Term Frequency* (TF)

*Term Frequency* (TF) adalah frekuensi kemunculan term (t) pada dokumen (D).

Tabel 6 Hasil menghitung TF

Term (t)	D1	D2	D3	D4	D1321	D1322	D1323	D1324	D1325
paket	1	0	0	0	1	1	1	1	1
cod	0	0	0	0	0	0	0	1	1
kurir	1	1	1	1	1	1	1	1	1
notif	0	0	0	0	0	1	0	0	0
rumah	0	0	0	0	0	0	0	1	0

2. Menghitung *Document Frequency* (DF)

*Document Frequency* (DF) adalah banyaknya dokumen suatu term (t) muncul kata-kat term.

Tabel 7 Hasil menghitung DF

Term (t)	df
paket	6
cod	2
kurir	9
notif	1
rumah	1

3. Menghitung *Invers Document Frequency* (IDF)

*Invers Document Frequency* (IDF) dilakukan dengan membagi jumlah total dokumen dalam term (t) dengan jumlah kalima dokumen yang mengandung kata. Berikut rumus untuk menghitung IDF:

$$IDF = \log \frac{N}{df} \tag{5}$$

Maka setelah menghitung DF, lalu akan menghitung nilai IDF (dengan jumlah dokumen = N).

Tabel 8 Hasil menghitung IDF

Term (t)	DF	IDF
paket	6	$\log \left[ \frac{1325}{6} \right] = 2,34$
cod	2	$\log \left[ \frac{1325}{2} \right] = 2,82$
kurir	9	$\log \left[ \frac{1325}{9} \right] = 2,17$

notif	1	$\log \left[ \frac{1325}{1} \right] = 3,12$
rumah	1	$\log \left[ \frac{1325}{1} \right] = 3,12$

4. Menghitung *TF-IDF*

*TF-IDF* menggabungkan kedua metrik tersebut untuk menghasilkan skor yang mempresentasikan tingkat kepentingan suatu kata yang ada di dalam dokumen. Berikut rumus untuk menghitung *TF-IDF*:

$$TF \times IDF \tag{6}$$

Ketangan:

TF adalah nilai *Term Frequency* dari kata tersebut dalam dokumen.

IDF adalah nilai *Inverse Document Frequency* dari kata tersebut dalam koleksi dokumen. Berikut hasil dari menghitung *TF-IDF*:

Term (t)	TF										IDF	TF x IDF									
	D1	D2	D3	D4	D1321	D1322	D1323	D1324	D1325	D1		D2	D3	D4	D1321	D1322	D1323	D1324	D1325		
paket	1	0	0	0	1	1	1	1	1	1	$\log [1325/6]=2,34$	2,34	0	0	0	2,34	2,34	2,34	2,34	2,34	
cod	0	0	0	0	0	0	0	1	1	1	$\log [1325/2]=2,82$	0	0	0	0	0	0	0	2,82	2,82	
kurir	1	1	1	1	1	1	1	1	1	1	$\log [1325/9]=2,17$	2,17	2,17	2,17	2,17	2,17	2,17	2,17	2,17	2,17	
notif	0	0	0	0	0	1	0	0	0	0	$\log [1325/1]=3,12$	0	0	0	0	0	0	3,12	0	0	
rumah	0	0	0	0	0	0	0	1	0	0	$\log [1325/1]=3,12$	0	0	0	0	0	0	0	3,12	0	

Gambar 4 Hasil menghitung TF-IDF

D. Klasifikasi Algoritma *Naïve Bayes*

Setelah melakukan tahapan *Text Preprocessing* untuk data selanjutnya yaitu proses klasifikasi. Klasifikasi dilakukan di *Platform Google Colab* dengan bahasa pemrograman *Python* menggunakan algoritma *Naive Baye*, maka untuk pembagian data dan hasil klasifikasi Algoritma *Naive* sebagai berikut:

1. Pembagian Data:

Tahap ini pembagian data akan diklasifikasikan dibagi menjadi 2 kelas, yaitu kelas *positif* dan *negatif*. Dataset dibagi menjadi data *training* sebanyak 927 dan data *testing* sebanyak 398 dengan menggunakan rasio 0.3. Artinya, 70% data digunakan untuk data *training*, dan 30% untuk data *testing* dari total 1325 data *tweet*. Untuk proses selanjutnya, peneliti menggunakan data testing untuk tahap pengolahan data

2. Klasifikasi Algoritma *Naive Bayes*:

Pada tahapan ini, proses klasifikasi *Naive Bayes* dilakukan dengan menghitung *probabilitas* kata dalam kalimat yang memiliki banyak peluang untuk setiap kelas. Data untuk menguji model klasifikasi *Naive Bayes* yang dibuat oleh *skleran.metrics* yang menghasilkan dari perhitungan *matrix* dengan *code python* menunjukkan tingkat kerja atau akurasi sebesar 80% untuk menggunakan model klasifikasi *Naive Bayes*.

Pada tahapan evaluasi menggunakan metode *Confusion Matrix* untuk mengetahui *Accuracy*, *Precision*, dan *Recall*. Proses pengujian dengan menggunakan *Confusion Matrix* merupakan tahap implementasi untuk membandingkan hasil pengujian klasifikasi dari sebuah sistem dengan hasil pengujian data yang sebetulnya. Evaluasi pengujian pada penelitian ini dilakukan pada objek sebuah opini pengguna *twitter* terhadap pelayanan jasa pengiriman. Berikut hasil Perhitungan pada dibawah ini.

Tabel 9 Hasil Perhitungan *Confusion Matrix*

<i>Predicted Class</i>	<i>Aktual Class</i>	
	<i>True Positive</i>	<i>True Negative</i>
	<i>Positive</i>	235
<i>Negative</i>	34	83

Tabel 10 Hasil *Confusion Matrix*

<i>Algoritma Naïve Bayes</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
	80%	84%	87%

1. Perhitungan *Accuracy*:

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \\
 &= \frac{235 + 83}{235 + 83 + 46 + 34} \times 100 \\
 &= \frac{318}{398} \times 100 \\
 &= 80\%
 \end{aligned}
 \tag{7}$$

Dalam hasil tersebut terdapat perhitungan *accuracy* dari model yang menghasilkan 235 *True Positive*, 83 *True Negative*, 46 *False Positive*, dan 34 *False Negative*. Setelah melakukan perhitungan dengan rumus, didapatkan hasil akurasi sebesar 80%. Hal ini menunjukkan bahwa model tersebut dapat memprediksi dengan benar sekitar 80% data dari keseluruhan data yang diberikan.

2. Perhitungan *Precision*:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \times 100 \\ &= \frac{235}{235 + 46} \times 100 \\ &= 0,8362 \times 100 \\ &= 84\% \end{aligned} \tag{8}$$

Dalam hasil di atas terdapat nilai *True Positive* sebesar 235 dan nilai *False Positive* sebesar 46, *Precision* yang dapat diperoleh tersebut adalah 84%.

3. Perhitungan *Recall*:

$$\begin{aligned} Recall &= \frac{TP}{TP + FN} \times 100 \\ &= \frac{235}{235 + 34} \times 100 \\ &= \frac{235}{269} \times 100 \\ &= 0,8736 \times 100 \\ &= 87\% \end{aligned} \tag{9}$$

Dalam hasil di atas terdapat pada model *recall* sebesar 87% *tweet* dari data kelas *positif* dengan berlabel 1 sebenarnya *positif*.

Berdasarkan hasil pembahasan di atas menunjukkan bahwa tahap evaluasi *Confusion Matrix* digunakan dalam menghitung *Confusion Matrix* untuk klasifikasi algoritma *Naive Bayes*, yang dibuktikan dengan tahapan menunjukkan dalam efektivitas pedekatang yang digunakan "analisis sentimen pelayanan terhadap jasa pengiriman berdasarkan opini *Twitter*" yang telah menunjukkan nilai *Accuracy* 80%, lalu perhitungan model *Precision* yang menunjukkan sebesar 84%, dan ada model *Recall* sebesar 87%, maka analisis sentimen dengan menggunakan algoritma *Naive Bayes* dapat menghasilkan dengan cukup baik.

#### IV. PENUTUP

##### A. Kesimpulan

Setelah melakukan hasil dari penelitian dilakukan, dapat disimpulkan dengan beberapa yang bisa ditarik kedalam kesimpulan sebagai berikut:

1. Model yang dihasilkan yaitu berbentuk *google colab* dan *crawling* data yang paling sering muncul pada kata "kurir". Lalu menghasilkan nilai sentimen yang dihasilkan opini *tweet* dari sentimen *positif* dan *negatif*.
2. Pada penelitian ini evaluasi model klasifikasi algoritma *Naive Bayes* dengan menggunakan *Confusion Matrix*. Dari pengujian tersebut didapatkan hasil analisis sentimen dengan kata kunci "kurir" mendapatkan nilai akurasi yang cukup baik dengan nilai *Accuracy* sebesar 80%, *Precision* 84%, dan *Recall* sebesar 87%.

##### B. Saran

Dalam melakukan penelitian ini data yang digunakan melalui hasil *Crawling* Data dari analisis sentimen melalui media sosial seperti *Twitter* dengan menggunakan algoritma *Naive Bayes* sehingga data yang di proses analisis sentimen lebih akurat, dan untuk saran lanjutan pada melakukan penelitian ini bisa diharapkan dapat menerapkan algoritma yang berbeda agar untuk mendapatkan hasil yang lebih akurat dan dengan menggunakan dataset yang lebih banyak dalam analisis sentimen.

#### PENGAKUAN

Naskah Ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Abdul Majid dengan berjudul Analisis Sentimen Pelayanan Terhadap Jasa Pengiriman Berdasarkan Opini *Twitter* Menggunakan Algoritma *Naive Bayes* yang dibimbing oleh Bapak Amril Mutoi Siregar dan Bapak Deden Wahiddin.

#### DAFTAR PUSTAKA

- [1] N. Ika, P. Kalingara, O. N. Pratiwi, and H. D. Anggana, "Analisis Sentimen Review Customer Terhadap Layanan Eekspedisi JNE Dan J&T Express Menggunakan Metode Naive Bayes Sentiment Analysis Review Customer Of JNE And J&T Express Expedition Services Using Naive Bayes Method," vol. 8, no. 5, 2021.
- [2] N. Juwitasari, S. Rs, M. Junaidi, and S. Soegianto, "Perlindungan Konsumen Terhadap Pengguna Jasa Ekspedisi," *J. USM*

- Law Rev.*, vol. 4, p. 688, 2021, [Online]. Available: <https://journal.uui.ac.id/Lex->
- [3] A. V. Sudiantoro and E. Zuliarso, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," vol. 10, no. 2, pp. 69–73, 2018.
- [4] W. T. Handoko, E. Supriyanto, D. I. Purwadi, Z. Budiarmo, and H. Listiyono, "Klasifikasi Opini Pengguna Media Sosial Twitter Terhadap JNT Di Indonesia dengan Algoritma Decision Tree," 2022.
- [5] A. Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes Muljono, D. Putri Artanti, A. Syukur, A. Prihandono, and D. I. Rosal Moses Setiadi, "Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang," 2018. [Online]. Available: <http://twitter.com>
- [6] A. Salam, J. Zeniarja, R. Septiyan, and U. Khasanah, *Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)*. 2018.
- [7] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.
- [8] I. Syarif, R. Asmara, and N. U. Rusmayani, "Klasifikasi Keluhan Masyarakat pada Sosial Media Twitter terhadap Pelayanan Toko Online di Indonesia menggunakan Metode Cosine TF-IDF," *BINA Insa. ICT J.*, vol. 7, no. 1, pp. 33–42, 2020.
- [9] Y. F. Alfiansah and A. M. Siregar, "Penerapan Algoritma K-Nearest Neighbors untuk Analisis Sentimen pada Buletin APTIKOM," vol. III, no. 1, 2022, [Online]. Available: <https://medium.com/@krismasuccess98/belajar->
- [10] A. Triayudi and G. Soepriyono, "Penerapan Data Mining Untuk Mengukur Kepuasan Mahasiswa Terhadap Pembelajaran dengan Menggunakan Algoritma Naïve Bayes," 2022, doi: 10.47065/josyc.v4i1.2524.