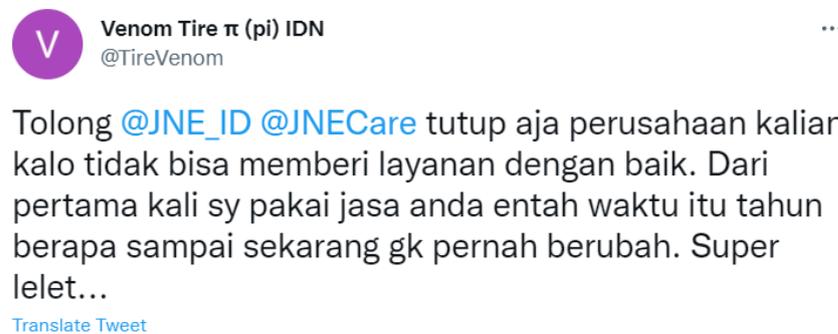


BAB III METODE PENELITIAN

3.1 Objek Penelitian

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



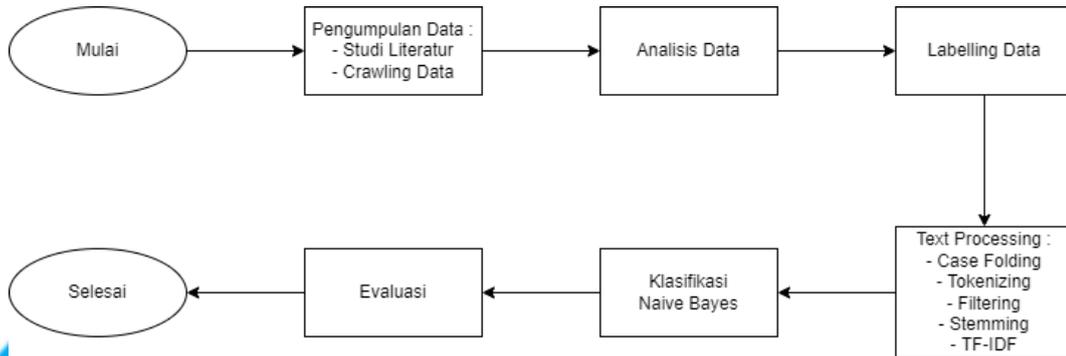
Gambar 3. 1 Contoh *Tweet Negative*



Gambar 3. 2 Contoh *Tweet Positif*

Berdasarkan yang ditunjukkan oleh gambar 3.1 terdapat kalimat “Tolong @JNE_ID @JNECare tutup aja perusahaan kalian kalo tidak bisa memberi layanan dengan baik. Dari pertama kali saya pakai jasa anda entah waktu itu tahun berapa sampai sekarang gak pernah berubah. Istilah “super lelet” mengandung konotasi kelesuan, yang mungkin bukan sifat paling disukai “lelet”. Selanjutnya pada gambar 3.2 terdapat kalimat “percaya deh JNE sangat netral dan sudah terbukti selama 30 tahun JNE berkiprah di layanan masyarakat #BahagiaBersama JNE” mengandung makna positif pada kata “Percaya”.

3.2 Prosedur Penelitian



Gambar 3. 3 Alur Prosedur Penelitian

3.2.1 Pengumpulan Data

Penelitian ini dimulai dengan melakukan pengumpulan data melalui sosial yaitu *twitter*. Pengumpulan data dari penelitian ini dibagi menjadi dua bagian sebagai berikut :

1. Studi Literatur

Studi literatur adalah sebagai pencari landasan teori dan untuk mendapatkan bahan referensi dari jurnal-jurnal atau buku terhadap pengertian teknik data mining untuk tinjau pustaka, penelitian terdahulu, dan mengambil metode yang akan di pakai.

2. *Crawling Data*

Pada tahapan *crawling data* ini dapat dilakukan dengan pengambilan data pada opini *twitter* lalu dikumpulkan menggunakan *python*, pencarian dengan kata kunci “Kurir”. Dan untuk data diambil dengan total 30768 *tweet* yaitu sejak tanggal 01 bulan Januari 2019 sampai tanggal 26 bulan Desember 2022. Kemudian pada data tersebut akan di klasifikasi dengan menggunakan Algoritma *Naïve Bayes*.

	Tanggal waktu	Id Tweet	Teks	Username	Label
0	2022-12-25 23:51:57+00:00	1607162261038338048	kurir paket tu skrg klo paket dh sampe emg ga ...	palatired	0
1	2022-12-25 23:43:02+00:00	1607160017626738689	@GlgarumonCh Kurir bayi haram	RGRTZ_	0
2	2022-12-25 23:33:01+00:00	1607157498615861248	@RadunBranko E ovo je malo sumnjivo, iste vest...	djolegsx	-1
3	2022-12-25 23:28:17+00:00	1607156305357307904	@poslednjilet Kurir ima televiziju👀	TKjjestan	0
4	2022-12-25 23:24:21+00:00	1607155315027607555	@Tr0ovble @Kedirifess @oemjly Lelaki kurir	eszcanorr	0
...
30763	2022-09-23 04:37:15+00:00	1573169599893438467	Dari tgl 10 sampe hari ini udah 5 kali kurir a...	paseoswift	0
30764	2022-09-23 04:11:30+00:00	1573163118787960832	dih parah sih @anteraja_id udah paket dari tgl...	putridailys	0
30765	2022-09-23 04:06:31+00:00	1573161866414653440	Meredam situasi dan suasana agar tercipta kond...	maung_tempur	0
30766	2022-09-23 04:04:30+00:00	1573161357993967616	@natrislavisha ngga dia ga nunggu paket lain c...	Teddypratara	0
30767	2022-09-23 03:56:46+00:00	1573159413040676864	Pakeeeeeeitttt... (bak gaya jasa kurir). Jumat 2...	eddyutyono991	0

30768 rows × 5 columns

Gambar 3. 4 Contoh Kata Kunci Kurir

3.2.2 Analisis Data

Pada tahap pertama penelitian 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 *positif* dan *negatif*.

3.2.3 Labelling Data

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

Tabel 3. 1 Tabel *Labelling* Data

Tanggal Waktu	Teks	Label
2021-06-29	terima kasih pelayanan jasa pengiriman.	1
2020-11-09	jasa pengiriman barang dari Medan ke Bengkulu dengan tarif yang murah cepat dan aman dari layanan yang profesional.	1
2020-12-02	@jne_id gue dukung #BoikotJNE karena pelayanan hancur, apalagi ditambah intoleran.	-1
2020-06-16	jne makin hari pelayanan barang makin minus, nah jasa pengiriman yang tidak menjamin keamanan barang.	-1
2020-11-16	@jnecare mulai hari ini gue boikot ini jasa pengiriman. Sudah mahal tapi pelayanan tidak memuaskan.	-1

3.2.4 Text Preprocessing

Pada tahap bagian *Preprocessing*, data yang telah diperoleh akan dilakukan proses penyeleksian data dengan tujuan menjadikan data tidak terstruktur menjadi data terstruktur. Terdapat beberapa tahapan-tahapan pada *Text Preprocessing* yaitu *Case Folding*, *Tokenizing*, *Filtering*, dan *TF-IDF* sebagai berikut :

1. Case Folding

Case Folding yaitu mengubah huruf menjadi *lowercase* atau huruf kecil.

Tabel 3. 2 Tabel *Case Folding*

<i>Input</i>	<i>Output</i>
tolong Tutup Aja Perusahaan Kalian Kalo Kalo Tidak Bisa Memberi Layanan Dengan Baik. Dari Pertama Saya Pakai Jasa Anda Entah Waktu Itu Tahun Berapa Sampai Sekarang Gak Pernah Berubah. Super Lelet.	tolong tutup aja perusahaan kalian kalo tidak bisa memberi layanan dengan baik dari pertama saya pakai jasa anda entah waktu itu tahun berapa sampai sekarang gak pernah berubah super lelet

2. *Tokenizing*

Tokenizing untuk proses yang dilakukan menghilangkan delimiter yaitu tanda baca dan simbol yang ada pada teks seperti @, \$, &, tanda titik (.), tanda koma (,), tanda seru (!), tanda tanya (?).

Tabel 3. 3 Tabel *Tokenizing*

<i>Input</i>	<i>Ouput</i>
tolong tutup aja perusahaan kalian kalo tidak bisa memberi layanan dengan baik. dari pertama saya pakai jasa anda entah waktu itu tahun berapa sampai sekarang gak pernah berubah. Super lelet.	'tolong' 'tutup' 'aja' 'perusahaan' 'kalian' 'kalo' 'tidak' 'bisa' 'memberi' 'layanan' 'dengan' 'baik'. 'Dari' 'pertama' 'saya' 'pakai' 'jasa' 'anda' 'entah' 'waktu' 'itu' 'tahun' 'berapa' 'sampai' 'sekarang' 'gak' 'pernah' 'berubah'. 'Super' 'lelet'

3. *Filtering*

Pada tahapan ini melakukan mengambil kata-kata penting yang telah dipotong kata per kata dilakukan *filter*. Dengan dilakukannya pembuangan kata-kata dalam sebuah paragraf, kalimat atau halaman ke dalam kata tunggal.

Tabel 3. 4 Tabel *Filtering*

<i>Input</i>	<i>Output</i>
tolong tutup aja perusahaan kalian kalo tidak bisa memberi layanan dengan	tolong tutup aja kalo tidak bisa memberi layanan dengan baik. Saya

baik. dari pertama saya pakai jasa anda pakai jasa anda sampai sekarang gak
 entah waktu itu tahun berapa sampai pernah berubah
 sekarang gak pernah berubah. Super
 lelet.

4. *Stemming*

Pada tahapan ini melakukan proses menghilangkan kata-kata dengan yang
 diubah menjadi kata dasar atau imbuhan kata dihilangkan. Tujuannya
 adalah untuk mengurangi kata-kata tersebut dapat dianggap sebagai satu
 entitas, sehingga memudahkan dalam pemrosesan untuk analisis teks

Tabel 3. 5 Tabel *Stemming*

<i>Input</i>	<i>Output</i>
tolong tutup aja perusahaan kalian kalo tidak bisa memberi layanan dengan baik. dari pertama saya pakai jasa anda entah waktu itu tahun berapa sampai sekarang gak pernah berubah. Super lelet.	tolong tutup aja usaha kali kalo tidak bisa beri layan dengan baik dari pertama saya pakai jasa anda entah waktu itu tahun berapa sampai sekarang gak pernah ubah super lelet

5. *TF-IDF*

TF-IDF sebuah tahapan untuk melakukan pembobotan kata pada sebuah
 dokumen. *TF-IDF* sebuah statistik numerik yang dapat menunjukkan kata
 kunci dengan kata tertentu. Selain dari itu, *TF-IDF* dapat mengetahui kata
 apa yang sering muncul pada suatu dokumen.

Adapun rumus dari pembobotan kata *TF-IDF* adalah sebagai berikut :

$$Wt = tf \times idf \quad (2)$$

$$idf = \log \frac{N}{df} \quad (3)$$

Keterangan :

tf = Jumlah kata dalam setiap dokumen

df = Jumlah total kata dari seluruh dokumen

N = Jumlah total dataset

$tf-IDF$ = Jumlah dokumen dalam koleksi dokumen yang mengandung kosakata.

3.2.5 Klasifikasi Naïve Bayes

Setelah melakukan *pre-processing* dalam kategori yang berbeda, selanjutnya akan di implementasi dengan algoritma *Naïve Bayes*. Proses klasifikasi melibatkan pengelompokan dan identifikasi *itemset* berdasarkan nilai bobot masing-masing. Metode canggih untuk mencapai ini adalah *Naïve Bayes*, yang memanfaatkan perhitungan *probabilitas* untuk mengklasifikasikan secara efektif. Dalam penelitian ini dengan kategori menjadi ada 2 kelas bagian yaitu *positif* dan *negatif*. Adapun perhitungan dari klasifikasi algoritma *Naïve Bayes* adalah sebagai berikut :

$$P(H|X) = \frac{P(X) \cdot P(H)}{P(x)} \quad (4)$$

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 data H .

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

3.2.6 Evaluasi

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

1. *Accuracy*

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

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

2. *Precision*

Precision faktor penting dalam menentukan keakuratan prediksi sistem. Ini melibatkan 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 :

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

6. *Recall*

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

$$Recall = \frac{TP}{TP+FP} \times 100\% \quad (7)$$

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).