BAB III
METODE PENELITIAN

3.1. Objek Penelitian

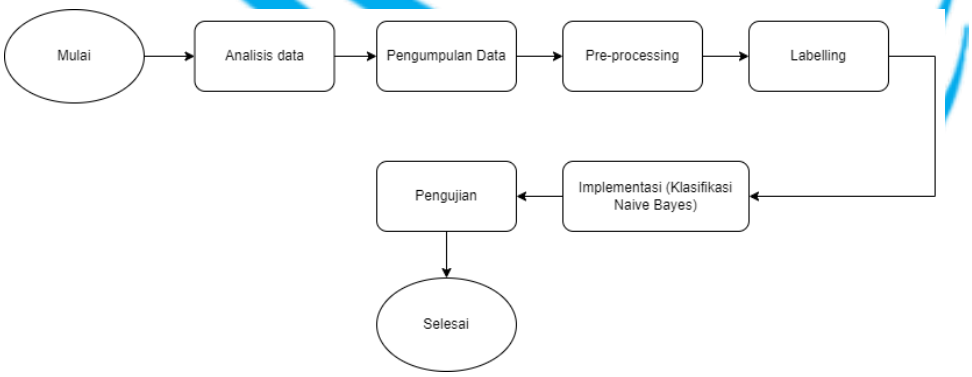
Penelitian ini adalah opini masyarakat tentang kebijakan larangan penjualan rokok batangan, Adapun sumber data yang diambil dari media sosial *Twitter*. Pelaksanaan penelitian dilakukan selama 8 bulan dari bulan Desember 2022 hingga Juli 2023. Dilaksanakan di Universitas Buana Perjuangan Karawang. Jadwal penelitian dapat dilihat pada tabel 3.1 dibawah ini :

Tabel 3. 1 Jadwal Penelitian

No	Kegiatan	Bulan							
		1	2	3	4	5	6	7	8
1	Analisis Data								
1	Pengumpulan Data								
2	*Pre-processing*								
3	Labeling								
4	Implementasi (Naive Bayes)								
5	Pengujian								

3.2. Prosedur Penelitian

Tahapan prosedur pada penelitian ini dapat dilihat pada Gambar 1 dibawah:



Gambar 3. 1 Proses Metode Penelitian

Gambar 3.1 merupakan penelitian yang melibatkan serangkaian proses. Data yang dikumpulkan (*crowling data*) merupakan opini masyarakat mengenai kebijakan pemerintah tentang larangan penjualan rokok batangan melalui media sosial twitter. Setelah dataset terkumpul, Langkah berikutnya adalah proses *Pre-processing* yang terdiri dari cleaning, case folding, tokenizing, normalize, stopword, dan Stemming. Kalimat opini yang telah mengalami pre-processing akan disajikan dalam bentuk bahasa Indonesia yang mudah dipahami. Setelah fase pra-pemrosesan, langkah selanjutnya memerlukan proses pemberian label (labelling) pemberian label dilakukan secara manual pada dataset, dan akan melewati tahapan Implementasi (klasifikasi) menggunakan metode algoritma *Naïve Bayes*. Proses terakhir yaitu tahap Pengujian.

3.2. 1. Analisis Data

Dalam penelitian ini dilakukan analisis data untuk mengetahui permasalahan yang terdapat pada media sosial *twitter* yaitu opini masyarakat tentang larangan penjualan rokok batangan. Analisis data dilakukan untuk menemukan dan memperoleh sentimen positif dan sentimen negatif dapat digambarkan sebagai berikut.

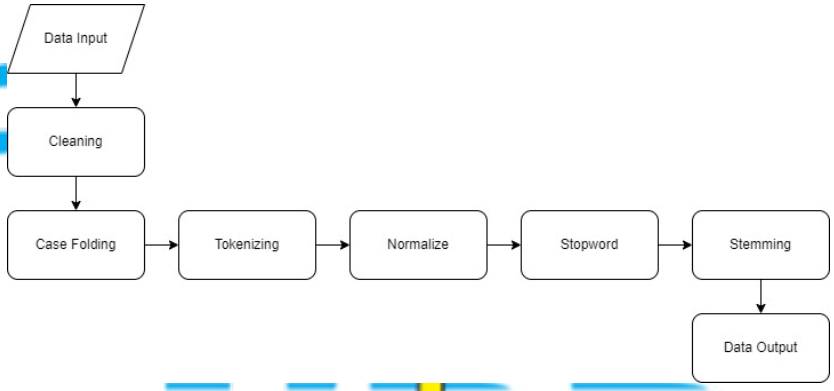
1. Sentimen positif dapat didefinisikan sebagai aktivitas atau reaksi yang dilakukan tanpa menghadapi penolakan atau yang memberikan dukungan.
2. Emosi negatif mengacu pada tindakan atau reaksi yang dilakukan dalam konteks penolakan, dan ditandai dengan menjadi negatif atau tidak mendukung..

3.2. 2. Pengumpulan Data

Penelitian ini diawali dengan melakukan pengumpulan data melalui salah satu media sosial *twitter*. Data dari *twitter* yang terkait dengan opini masyarakat mengenai larangan penjualan rokok batangan menggunakan metode algoritma *Naïve Bayes* untuk pengumpulan *twitter* disediakan oleh *API*, kemudian disimpan dalam format csv.

3.2. 3. Data *Pre-processing*

Pada tahap *pre-processing*, data yang diperoleh akan dilakukan proses penyelesaian data dengan tujuan menjadikan data tidak terstruktur menjadi data terstruktur. Terdapat beberapa tahapan pada *pre-processing* penelitian ini yaitu sebagai berikut:



Gambar 3. 2 Tahapan Data *Pre-processing*

a. *Cleaning*

Cleaning data Praktik pembersihan data berkaitan dengan tindakan pemurnian dan pemurnian data yang baru saja diperoleh dari platform media sosial, Twitter. Selama prosedur ini, data yang baru diperoleh dari tindakan *crawling* dapat berisi beberapa elemen tekstual yang dianggap tidak sesuai untuk digunakan, seperti tagar, URL, dan @, maka dari itu perlu dilakukan proses *cleansing*.

Sebelum	Sesudah
@CNNIndoonesia Bener Pak. Mending bisnis rokok ditutup aja Pak	bener pak mending bisnis rokok ditutup aja pak

Tabel 3. 2 Proses Tahap *Cleaning*

b. *Case folding*

Selama fase ini, proses normalisasi huruf melibatkan konversi huruf besar menjadi huruf kecil. Berikut adalah contoh Case Folding.

Sebelum	Sesudah
Meskipun aku nggk ngrokok kalau lihat harga rokok di Indomaret atau Alfa hargane ngeri banget sekarang Temen2 bnyak yg pindah rokok yg 10r dapat 1bngkus	meskipun aku gak ngerokok kalua lihar harga rokok di indomaret atau alfa ngeri banget sekarang temen-temen banyak yang pindah rokok yang 10rb dapat 1 bungkus

Tabel 3. 3 Proses Tahap *Case Folding*

c. *Tokenizing*

pada tahapan ini *tokenizing* prosesnya melibatkan segmentasi frasa menjadi kata-kata individual. Akibatnya, dimungkinkan untuk memeriksa setiap kata secara independen selama tahap selanjutnya.

Sebelum	Sesudah
agar menjaga kesehatan anak yang di bawah umur	agar, menjaga, kesehatan, anak yang di bawah umur

Tabel 3. 4 Proses Tahap *Tokezining*

d. *Normalize*

Normalize yaitu untuk mengubah kata yang tidak baku menjadi baku pada pengetikan (typo) menjadi kata yang mudah di pahami.

Sebelum	Sesudah
agar lebih menghindari anak di bawah umur	agar menghindari anak di bawah umur

Tabel 3. 5 proses Tahap *Normalize*

e. *Stopword*

Pada titik ini, prosesnya melibatkan identifikasi dan pemilihan istilah-istilah penting. Prosedur ini menghilangkan kata-kata yang berlebihan, seperti contoh yang diberikan “yang”, “di”, “ke” dan lain sebagainya.

Sebelum	Sesudah
agar menjaga kesehatan anak yang di bawah umur	agar menjaga kesehatan anak bawah umur

Tabel 3. 6 Proses Tahap *Stopword*

f. *Stemming*

pada tahapan *stemming* yaitu pencarian kata dasar dari hasil *stopword* yaitu bekerja dengan membuang kata-kata seperti contoh “me-kan”, “ber-kan” dan lain sebagainya.

Sebelum	Sesudah
kebijakan, melarang, penjualan, rokok	bijak, larang, jual, rokok

Tabel 3. 7 Proses Tahap *Stemming*

3.2. 4. Labeling

Dalam penelitian ini, diberikan dua label sebagai parameter, label negatif dan positif, untuk mengklasifikasikan sentiment pada *Twitter*. Langkah-langkah dalam proses pemberian label pada data *twitter* yaitu sebagai berikut.

a. Kata atau gtweet yang memiliki sentimen negatif dan positif akan dipilih.
b. Jika jumlah kata dengan sentimen positif lebih banyak daripada jumlah kata dengan sentimen negatif, maka label sentimennya akan ditetapkan sebagai positif (1).
c. Jika kata dengan sentimen positif lebih sedikit daripada jumlah kata dengan sentimen negatif, maka label sentimennya adalah negatif (-1).

3.2. 5. Implementasi

1. TF-IDF

TF-IDF dilakukan setelah tahapan *pre-processing* dan pelabelan data kemudian akan dilakukan pembobotan data menggunakan TF-IDF berikut tahapan-tahapan TF-IDF.

1) Hitung Term Frequency (TF) total kemunculan istilah dalam dokumen di bagi dengan jumlah total kata dalam dokumen.
2) Hitung frekuensi dokumen terbalik (IDF) logaritma dari jumlah total dokumen dibagi dengan jumlah dokumen yang berisi term.
3) Hitung TF-IDF hasil perkalian TF dan IDF. Nilai menunjukkan pentingnya istilah dalam dokumen.
4) Normalisasi TF-IDF pastikan nilai TF-IDF berada pada skala yang sama sehingga dapat dibandingkan di seluruh dokumen.

2. Pembagian Data

Tahapan ini dilakukan setelah melakukan tahapan TF-IDF. Pada proses ini data akan dibagi menjadi dua yaitu data latih dan data uji secara acak.

3. Klasifikasi Naïve Bayes

Setelah melakukan tahapan TF-IDF data akan diimplementasikan dengan menggunakan algoritma *Naïve Bayes*. Tahapan Algoritma *Naïve Bayes* dimulai dengan mempersiapkan data yang sudah di dapatkan dari media sosial *twitter*. Dalam penelitian ini data yang dikelompokan dibagi menjadi dua bagian yaitu positif dan negatif.

3.2. 6. Pengujian

Pengujian dalam penelitian ini digunakan untuk mengukur kinerja algoritma yang digunakan. Proses evaluasi dimulai dengan menggunakan matriks konfusi untuk menilai tingkat akurasi algoritma. Selanjutnya, dilakukan pengujian algoritma menggunakan metode *Confusion Matrix* untuk mendapatkan gambaran lebih lengkap tentang performa algoritma tersebut.

Tabel 3. 8 Pengujian *Confusion Matrix*

Manual	Sistem	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Pengujian dilakukan dengan menghitung nilai *accuracy*, *precision*, dan *recall* dan ditampilkam dalam bentuk persentase.

1. *Accuracy*

   Akurasi adalah persentase dari jumlah sentimen yang benar diketahui. Menghitung nilai akurasi dilakukan dengan cara membagi jumlah data sentimen yang benar dengan total data dan di uji. Untuk menghitung nilai akurasi persamaan dibawah ini:

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} x\ 100\% \ldots\ldots (3.1)$$

   Keterangan :

   TP : Kuantitas positif yang dikategorikan positif.

   TN : Kuantitas negatif yang dikategorikan sebagai negatif.

   FP : Kuantitas negatif yang dikategorikan positif.

   FN : Kuantitas contoh bagus yang secara keliru dikategorikan sebagai negatif.

2. *Precision*

   Precision merupakan metrik yang mengukur perbandingan antara jumlah data relevan yang berhasil ditemukan pada jumlah data yang ada. Menghitung *precision* dilakukan dengan membagi jumlah data benar yang terklasifikasikan sebagai positif oleh model (*True Positive*) dengan jumlah data yang benar terklasifikasi sebagai positif ditambah

data yang salah yang terklasifikasi sebagai positif (*False Positive*). Untuk menghitung nilai *precision* digunakan persamaan dibawah ini:

$$Precision = \frac{TP}{FP + TP} \; x \; 100\% \ldots\ldots (3.2)$$

Keterangan :

TP : Kuantitas positif yang dikategorikan positif.

FP : Kuantitas negatif yang dikategorikan positif.

3. *Recall*

Recall merupakan metrik yang mengukur perbandingan jumlah data yang relevan yang berhasil ditemukan oleh model terhadap jumlah data yang sebenarnya relevan dalam dataset. Perhitungan recall dilakukan dengan membagi jumlah data yang benar yang terklasifikasikan sebagai positif oleh model (*True Positive*) dengan hasil penjumlahan antara jumlah data yang benar yang terklasifikasikan sebagai positif dan data yang seharusnya bernilai positif namun terklasifikasikan sebagai negatif oleh model (*False Negative*). Untuk menghitung nilai *recall* digunakan persamaan dibawah ini:

$$Recall = \frac{TP}{FN + TP} \; x \; 100\% \ldots\ldots (3.3)$$

Keterangan :

TP : Kuantitas positif yang dikategorikan positif

FN : Kuantitas contoh bagus yang secara keliru dikategorikan sebagai negatif.