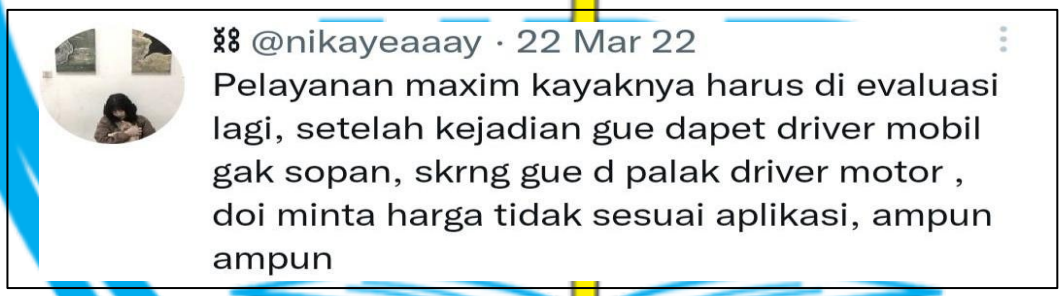
BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek pada penelitian tugas akhir ini adalah tentang sentimen masyarakat pada aplikasi Maxim, adapun sumber data yang diambil dengan mengambil sampel dari postingan yang di *tweet* oleh masyarakat ditunjukkan pada Gambar 3.1 postingan positif dan Gambar 3.2 postingan negatif.

Gambar 3.1 Postingan Positif



Gambar 3.2 Postingan Negatif

Berdasarkan yang ditunjukan oleh Gambar 3.1 terdapat sentimen positif pada kalimat "bapak maxim baik banget" mengandung makna positif pada kata "Baik", pada Gambar 3.2 terdapat sentimen negatif pada kalimat "Pelayanan maxim kayaknya harus di evaluasi lagi, setelah kejadian gue dapet driver mobil ga sopan, sekarang gue di palak driver motor, doi minta harga tidak sesuai aplikasi, ampun ampun" mengandung makna negatif pada kalimat "Driver mobil ga sopan".

3.2 Lokasi dan Waktu Penelitian

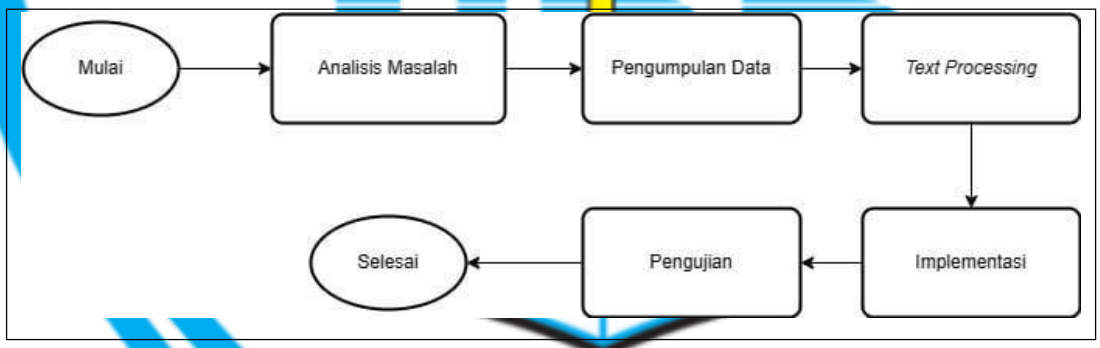
Penelitian ini berlokasi di Laboratorium Riset Universitas Buana Perjuangan Karawang, kemudian waktu penelitian berlangsung sejak bulan November dan rincian kegiatan penelitian ditunjukkan pada Tabel 3.1.

Tabel 3.1 Jadwal Penelitian

No	Kegiatan	Bulan					
		1	2	3	4	5	6
1	Analisis Masalah						
3	Pengumpulan Data						
4	*Text Processing*						
5	Implementasi						
6	Pengujian						

3.3 Prosedur Penelitian

Untuk prosedur penelitian dapat digambarkan dengan bentuk Flowchart seperti gambar dibawah ini:



Gambar 3.3 Prosedur Penelitian

Berdasarkan Gambar 3.3, penelitian dimulai dengan tahapan analisis masalah. Setelah itu, dilakukan pengumpulan data atau *Crawling* data di *Twitter*. Data yang telah terkumpul kemudian diproses pada tahap *Text Processing*. Selanjutnya, data akan diklasifikasikan menjadi dua kategori, yaitu positif dan negatif. Setelah proses klasifikasi selesai, data yang telah diklasifikasikan akan diimplementasikan oleh algoritma *Naïve Bayes*. Kemudian, dilakukan tahap pengujian untuk menguji performa dan mendapatkan hasil akurasi terbaik dari algoritma yang digunakan.

3.3.1 Analisis Masalah

Pada tahap ini, proses pencarian permasalahan data dilakukan dengan tujuan untuk mengidentifikasi dan memahami fakta-fakta sebenarnya yang terkait dengan data yang ada. Cara menemukan fakta data sentimen yang bernilai positif dan negatif dapat dijelaskan sebagai berikut:

1. Sentimen positif dapat diartikan sebagai suatu tindakan atau pendapat yang mendukung atau tidak ada penolakan terhadap suatu objek atau topik. Dalam analisis sentimen, data yang mengekspresikan perasaan senang, puas, atau pujian terhadap suatu produk, layanan, atau topik dianggap memiliki sentimen positif.
2. Sentimen negatif dapat diartikan sebagai suatu tindakan atau pendapat yang mencerminkan penolakan, ketidakpuasan, atau kritikan terhadap suatu objek atau topik. Dalam analisis sentimen, data yang menyatakan perasaan marah, kecewa, atau ketidakpuasan terhadap suatu produk, layanan, atau topik dianggap memiliki sentimen negatif.

Dengan melakukan analisis sentimen maka akan mendapatkan kesimpulan dan pemahaman yang lebih baik tentang bagaimana masyarakat atau pengguna merespons suatu objek atau topik tertentu dengan sentiment positif atau negatif berdasarkan data yang telah dianalisis.

3.3.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dibagi menjadi dua, antara lain:

a. Studi Literatur

   Pada langkah ini, tujuannya agar meninjau berbagai referensi, seperti buku, paper, media sosial, dan lain-lain, yang akan digunakan dalam menyusun penelitian ini.

b. *Crawling Data*

   Data yang dikumpulkan diambil dari media *Twitter* data tersebut diambil dari tanggapan masyarakat terhadap Maxim. Proses pengumpulan data dilakukan dengan menggunakan API dengan cara *Crawling data* dengan sebuah *query* “Maxim”, “Aplikasi Maxim”.

3.3.3 *Text Processing*

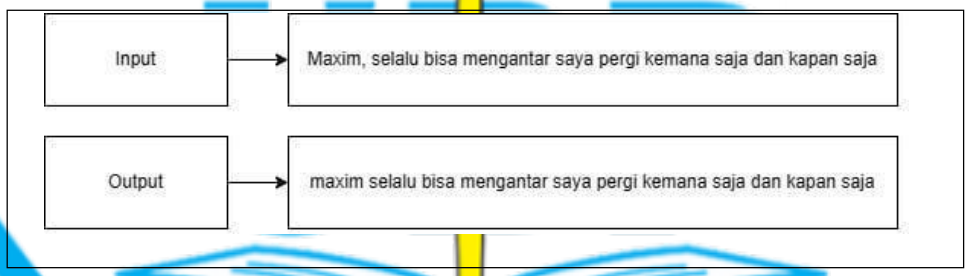
Text Processing biasanya melakukan klasifikasi, teks dokumen atau data harus disiapkan terlebih dahulu. Tahapan Pengolahan Data berguna agar teks menjadi lebih terstruktur.

1. *Cleanning* data

   Cleanning data merupakan tahapan penghapusan data duplikat dan bahasa asing dari hasil data yang telah didapatkan pada tahap pengumpulan data, *Cleanning* data akan dilakukan secara manual.

2. *Case Folding*

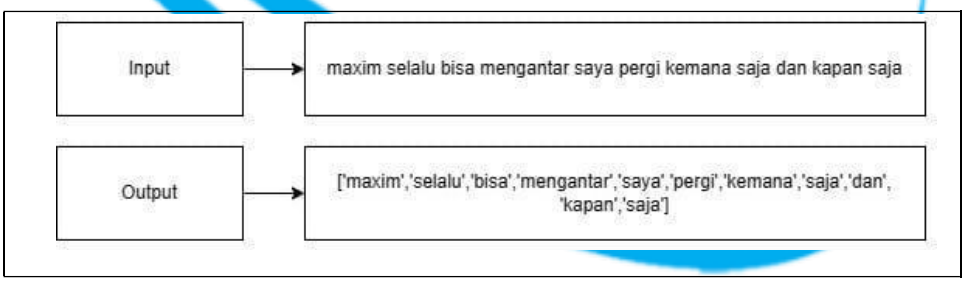
   Case Folding adalah proses untuk mengubah huruf kapital menjadi huruf kecil atau standar dalam teks. Proses ini dilakukan dengan tujuan mempermudah pencarian dan analisis teks, karena tidak semua dokumen atau data teks konsisten dalam penggunaan huruf kapital. Dengan mengubah seluruh teks menjadi huruf kecil, kita dapat menyamakan format huruf dan membuatnya lebih seragam, sehingga pencarian teks menjadi lebih efisien dan konsisten. Berikut adalah contoh *Case Folding* :



Gambar 3.4 Contoh *Case Folding*

3. *Tokenizing*

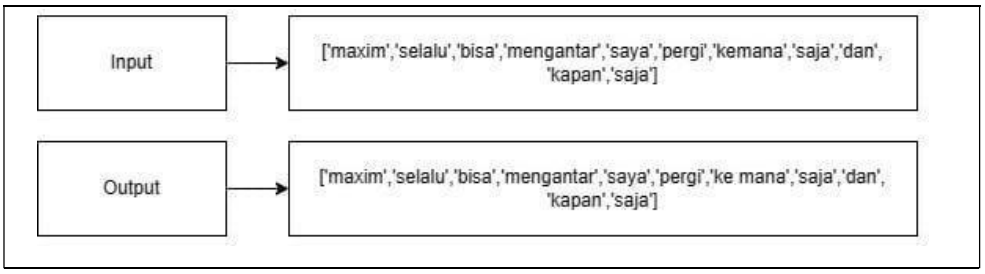
   Tokenizing merupakan sebuah proses untuk memecah kalimat menjadi kata-kata yang akan menjadikan kalimat lebih bermakna. Berikut contoh *Tokenizing* :



Gambar 3.5 Contoh *Tokenizing*

4. *Konversi Slangword*

   Konversi Slangword digunakan untuk proses mengubah kata tidak baku menjadi kata baku. Tahap ini dilakukan dengan menggunakan bantuan kamus Slangword, kata yang ada dikaus slangword nantinya akan menjadi kata baku. Berikut Contoh Konversi Slangword:

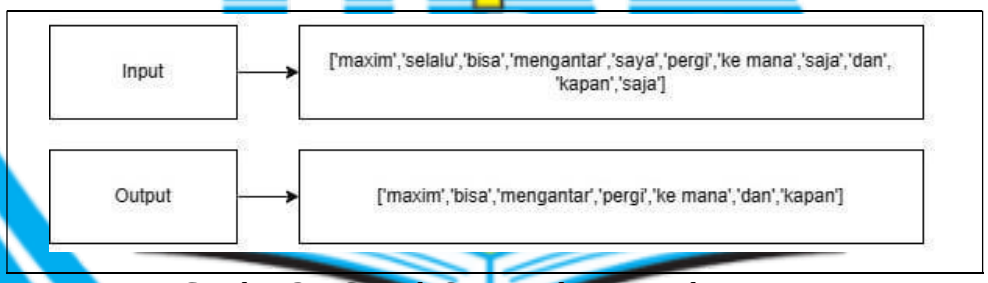


Gambar 3.6 Contoh *Konversi Slangword*

Berdasarkan tabel diatas tahapan untuk merubah kata baku menjadi tidak baku yaitu dengan menggunakan kamus perbaikan data.xlsx, kamus perbaikan data dibuat berdasarkan kata tidak baku dan baku.

5. *Stopwords Removal*

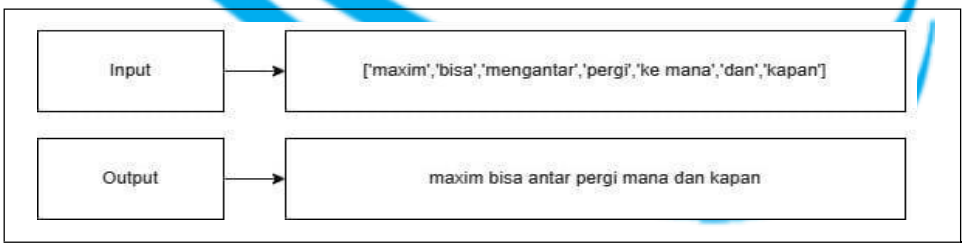
   Stopwords merupakan proses membuang kata yang tidak penting yang menyimpang dari kosa kata dan pembuangan *term* yang tidak memiliki arti atau tidak relevan. Berikut adalah contoh *Stopwords Removal* :



Gambar 3.7 Contoh Stopwords *Removal*

6. *Stemming*

   Stemming merupakan tahapan untuk memperkecil jumlah indeks yang berbeda dari satu data sehingga kata yang memiliki *suffix* maupun *prefix* akan kembali ke bentuk dasarnya. Berikut contoh *Stemming* :



Gambar 3.8 Contoh *Stemming*

7. Pelabelan Data

   Pelabelan data dilakukan secara manual, label yang diberikan ada 2 kelas, yaitu kelas positif, dan kelas negatif.

Tabel 3.2 Contoh Pelabelan Data

Kalimat	Label
Bapak maxim baik banget	Positif
Bad bgt anjir pelayanan maxim	Negatif

3.3.4 Implementasi

Setelah melakukan *Text Processing* dan pelabelan data, kemudian data tersebut diimplementasikan dengan tahapan sebagai berikut :

1. TF – IDF

   Metode algoritma yang umum digunakan untuk menghitung bobot setiap kata dalam sebuah dokumen adalah *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini memungkinkan kita untuk mengevaluasi pentingnya kata dalam sebuah dokumen atau kumpulan dokumen berdasarkan frekuensi kemunculan kata tersebut di dokumen tersebut dan seluruh kumpulan dokumen.

2. Pembagian Data

   Setelah melakukan TF-IDF kemudian akan dilakukan pembagian data, pada proses ini data akan dibagi menjadi dua yaitu data latih sebanyak 80% dan data uji sebanyak 20% secara acak.

3. Klasifikasi *Naïve Bayes*

   Metode *Naive Bayes* menggunakan teori probabilitas ilmuwan Inggris Thomas Bayes 8, yaitu. memprediksi probabilitas masa depan berdasarkan pengalaman masa lalu. Pengklasifikasi *Naive Bayes* menggunakan pendekatan teorema Bayes untuk menghitung probabilitas kelas yang diberikan dokumen yang telah diketahui.

3.3.5 Pengujian

Pengujian ini dilakukan untuk mengetahui kinerja dan mengevaluasi algoritma yang digunakan dalam penelitian ini. Proses evaluasi sendiri dilakukan dengan menggunakan metode *Confusion Matrix* untuk mengetahui nilai akurasi algoritma. Metode ini sangat berguna untuk mengevaluasi kinerja analisis klasifikasi. Adapun rumus untuk perhitungan *accuracy, precision*, dan *recall* sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} X100\% \quad (5)$$

$$Precision = \frac{TP}{TP+FP} X100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} X100\% \quad (7)$$

Accuracy adalah ukuran seberapa tepat model dalam mengklasifikasikan data dengan benar. *Precision* menjelaskan sejauh mana data yang diminta sesuai dengan hasil prediksi yang diberikan oleh model. Sementara itu, *recall* menggambarkan sejauh mana model berhasil menemukan informasi tertentu.