

## **BAB III**

### **METODE PENELITIAN**

Bab ini secara rinci menguraikan berbagai aspek terkait dengan penelitian analisis sentimen terhadap hasil debat calon presiden yang dilakukan di platform Twitter menggunakan metode Naive Bayes. Pertama-tama, bab ini membahas jenis penelitian yang digunakan, yang mencakup penjelasan tentang pendekatan kuantitatif yang digunakan untuk menganalisis data dari platform media sosial. Pendekatan ini memungkinkan peneliti untuk mengumpulkan dan menganalisis data dengan menggunakan alat dan teknik statistik untuk memahami sentimen secara luas.

Selanjutnya, bab ini juga membahas tentang tempat dan waktu penelitian. Tempat penelitian mengacu pada lingkungan atau platform di mana data dikumpulkan, yaitu Twitter, yang merupakan salah satu platform media sosial yang paling populer digunakan untuk menyampaikan opini dan sentimen secara real-time. Sedangkan waktu penelitian menunjukkan periode waktu ketika data dikumpulkan, yang mungkin meliputi periode sebelum, selama, dan setelah debat calon presiden dilakukan.

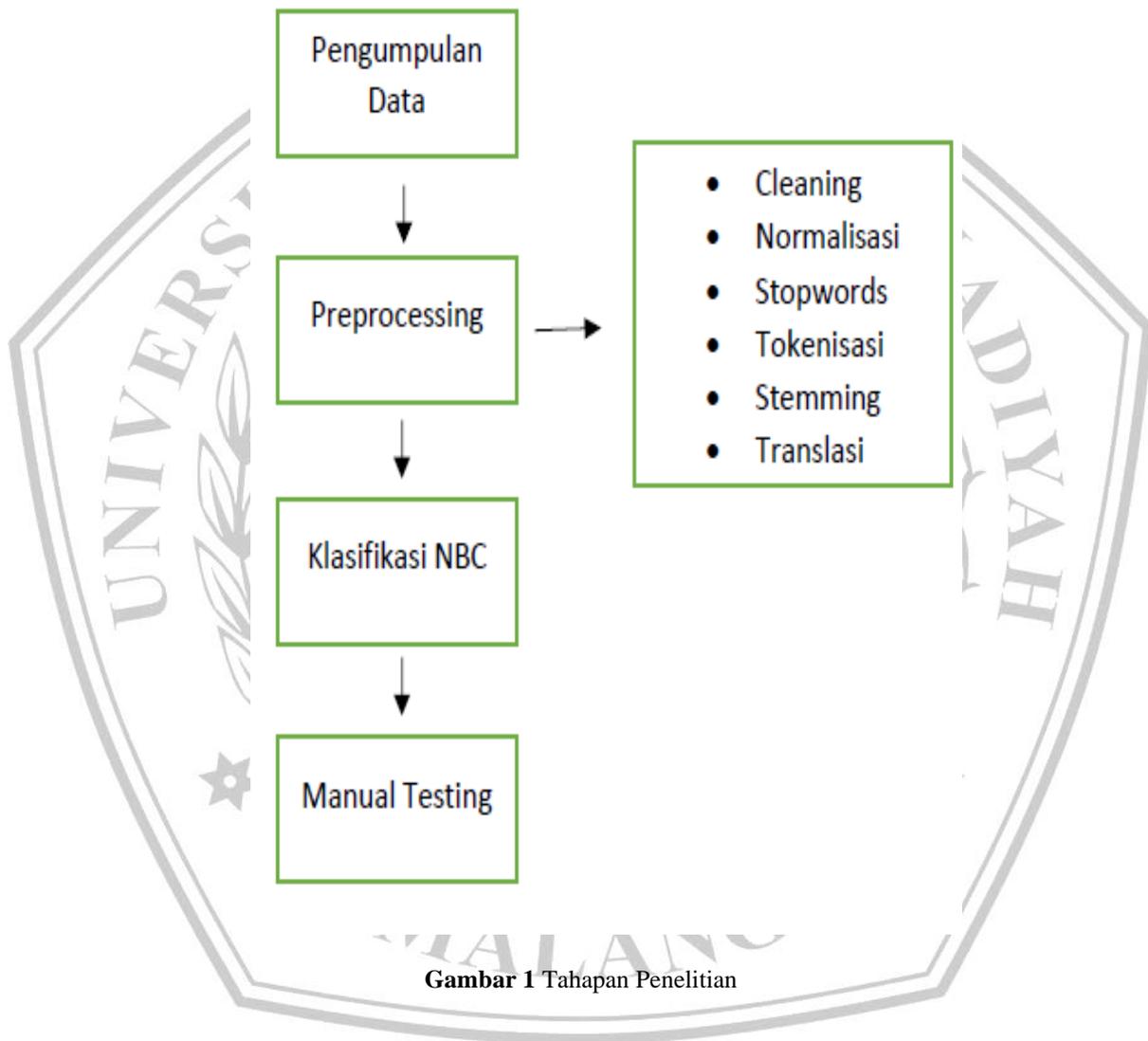
#### **Jenis Penelitian**

Penelitian ini difokuskan pada analisis sentimen masyarakat terhadap sebuah topik, dengan fokus pada opini yang diposting di linimasa Twitter. Penelitian ini menggunakan jenis penelitian kuantitatif, karena membutuhkan pengumpulan data dalam jumlah besar, mengingat penelitian ini mendasarkan analisis pada kamus yang umum digunakan oleh masyarakat untuk menyatakan sentimen mereka. Pendekatan kuantitatif dilakukan pada tahap penghitungan dan pemrosesan data, di mana tweet yang dianalisis akan melalui serangkaian proses dan pengklasifikasian sentimen dilakukan menggunakan metode Naive Bayes. Dengan demikian, penelitian ini akan memberikan pemahaman yang mendalam

tentang pandangan dan opini masyarakat yang terungkap di platform Twitter terkait dengan topik yang diteliti.

### Metode Penelitian

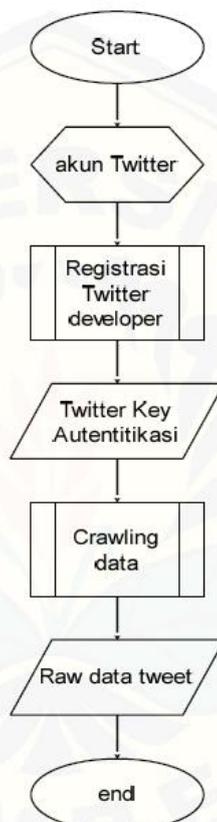
Dalam penelitian ini, terdapat beberapa tahapan yang digunakan, yang dapat dilihat pada Gambar 1 sebagai berikut:



Gambar 1 Tahapan Penelitian

## Pengumpulan Data

Penelitian ini memanfaatkan data dari tweet yang diambil dari pengguna aplikasi media sosial Twitter. Proses pengumpulan data tweet dilakukan melalui metode Crawling pada platform Twitter dengan menggunakan *tools* Tweet Harvest. Sebanyak 1444 tweet berhasil diperoleh dengan menggunakan kata kunci Anies, Ganjar, dan Prabowo. Data yang digunakan dalam penelitian ini adalah data berisi isi dari tweet, dengan beberapa atribut sebagai berikut: *created\_at*, *id\_str*, *full\_text*, *quote\_count*, *reply\_count*, *retweet\_count*, *favorite\_count*, *lang*, *user\_id\_str*, *conversation\_id\_str*, *username*, dan *tweet\_url*. Pencarian dan pengumpulan data tweet ini dimulai pada tanggal 12 Desember 2023, setelah terjadinya debat calon presiden. Berikut alur flowchart proses pengambilan data.



**Gambar 2** Tahapan Crawling data Twitter

Dari diagram alur di atas, dapat disimpulkan bahwa dalam melakukan proses pengumpulan data (*crawling data*) dari Twitter, diperlukan penggunaan kunci otentikasi (*authentication key*) yang diberikan oleh platform Twitter. Kunci otentikasi ini harus dimasukkan ke dalam program Tweet Harvest untuk memungkinkan program tersebut mengakses data dari Twitter. Hal ini menunjukkan pentingnya penggunaan kunci otentikasi dalam menjalankan proses pengumpulan data dari platform Twitter untuk keperluan penelitian atau analisis sentimen.

### Variable Penelitian

Berdasarkan hasil pengumpulan data, telah ditetapkan beberapa variabel penentu yang menjadi kunci dalam analisis sentimen ini. Variabel atau atribut utama yang digunakan melibatkan *full\_text*, *username*, dan *created\_at*. Beberapa data dan variabel yang terdapat dalam dataset dapat diidentifikasi sebagai berikut:

<i>full_text</i>	<i>username</i>	<i>created_at</i>
pastilah anies tidak punya prestasi	MarahIchsan	Fri Dec 15 04:04:44 +0000 2023
suasana pedesaan memang selalumu menyenangkan ...	Agung_wiyo	Fri Dec 15 04:04:33 +0000 2023
program anies hanyalah mengubah nama...	Alfakton_i	Fri Dec 15 04:04:31 +0000 2023

**Table 1** Twitter Dataset

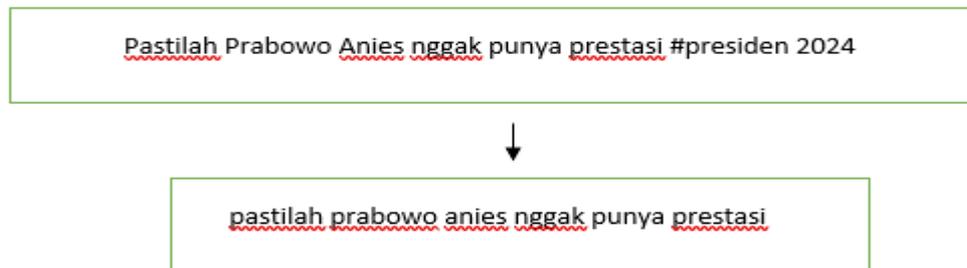
### Preporcessing

Pada tahap pra-pemrosesan teks, data akan diolah menjadi lebih terstruktur sebelum menjalani proses klasifikasi. Tujuan dari pra-pemrosesan adalah untuk menyiapkan data agar siap digunakan dengan menghilangkan unsur-unsur yang dianggap tidak relevan. Terdapat

beberapa proses pra-pemrosesan teks, seperti pembersihan teks, pengubahan huruf kecil (case folding), normalisasi, tokenisasi, penghapusan kata-kata umum (stopword removal), dan stemming. [10]

### *Cleansing*

Tahap pertama dalam proses adalah "Cleansing," yang merujuk pada proses pembersihan data teks dengan mengeliminasi informasi yang tidak relevan dan tidak konsisten. Cleansing bertujuan untuk menghapus karakter yang tidak penting, seperti hashtag (#), angka, username (@), URL, tanda baca (punctuation), dan emotikon. Langkah selanjutnya adalah "Case Folding," yang merupakan proses mengubah bentuk kata atau karakter ke dalam format seragam, umumnya menjadi huruf kecil (*lower case*). [11] Berikut merupakan contoh teks hasil *cleansing*.



**Gambar 3** Teks hasil *cleansing*

### *Normalisasi*

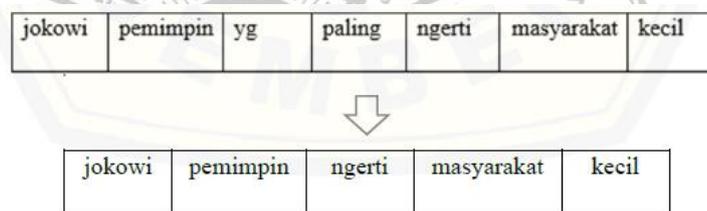
Normalisasi memiliki peranan penting dalam upaya mengurangi keragaman kata-kata yang mungkin memiliki arti atau makna serupa, sehingga memfasilitasi proses klasifikasi dan analisis data. Tujuan dari normalisasi ini adalah untuk menciptakan keseragaman dalam representasi kata-kata, sehingga perbedaan minor dalam penyajian kata tidak menghambat pemahaman atau interpretasi secara keseluruhan. Dengan cara ini, normalisasi menjadi instrumen efektif dalam mendukung keakuratan dan

kemudahan interpretasi dalam konteks pengklasifikasian dan analisis data.

[12]

### ***StopWords***

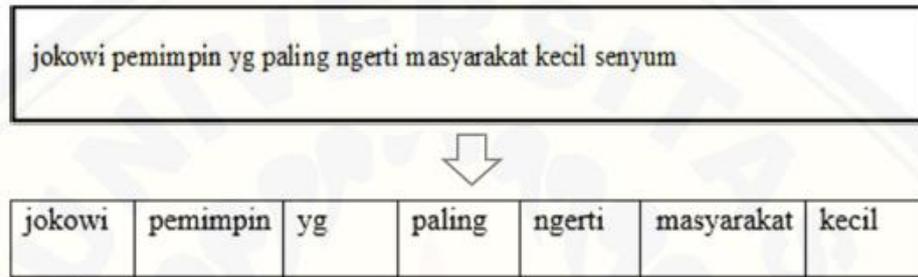
Stopword merupakan salah satu tahapan dalam pemrosesan teks yang bertujuan untuk menghapus kata-kata yang sering muncul namun tidak memiliki arti penting dan tidak berpengaruh pada analisis sistem, seperti "oh", "di", "pada", dan sebagainya. Tahapan ini penting dalam membersihkan teks dari kata-kata yang tidak memberikan kontribusi signifikan terhadap pemahaman konten atau makna teks yang sedang diproses. Dengan menghapus stopwords, kita dapat meningkatkan kualitas analisis teks dengan fokus pada kata-kata kunci yang lebih relevan dan informatif.[13]



**Gambar 4** Hasil Stopword Removal

### ***Tokenize***

Tokenisasi merupakan tahap krusial dalam rangkaian pra-pemrosesan teks, yang bertujuan untuk memecah teks menjadi unit-unit yang lebih kecil. Proses ini menggunakan alat atau perangkat khusus yang mampu memisahkan teks menjadi bagian-bagian yang lebih terdefinisi, termasuk dalam skala kata, kalimat, dan bigram. Tahap awal ini sangat relevan dalam mempersiapkan teks untuk analisis lebih lanjut, karena menyediakan dasar yang kuat dalam memahami struktur dan makna dari setiap elemen teks yang diolah.[14]. Berikut merupakan hasil dari proses Stopword.



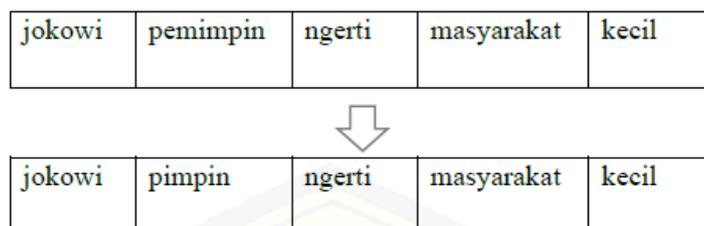
**Gambar 4** Hasil Stopword Removal

**Stemming**

Stemming adalah proses pemotongan (pembuangan) imbuhan (affix), baik prefix maupun suffix, dari sebuah term untuk mendapatkan kata dasar (root atau stem) dari kata berimbuhan.[15]

Penerapan stemming memiliki manfaat signifikan dalam mengurangi jumlah kata yang bervariasi namun memiliki arti atau makna yang serupa, sehingga memudahkan proses klasifikasi dan analisis data. Meskipun demikian, perlu diingat bahwa penggunaan stemming dalam analisis sentimen harus disesuaikan dengan kriteria dan metode yang digunakan. Hal ini dikarenakan pilihan stemming yang tepat dapat memberikan kontribusi positif terhadap akurasi pengklasifikasian, sementara pada sisi lain, pemilihan yang kurang tepat dapat mempengaruhi akurasi tersebut secara negatif. Dengan demikian, kebijakan yang cermat terkait dengan penggunaan stemming menjadi kunci dalam mendukung hasil analisis sentimen yang lebih akurat dan reliabel.[10]

Berikut merupakan hasil dari stemming.



**Gambar 5** Hasil Stemming

### *Klasifikasi Naïve Bayes*

Dalam penelitian ini, digunakan metode pengelompokan yang didasarkan pada teori Bayes yang mampu memanfaatkan pengalaman masa lalu untuk memperkirakan peluang di masa mendatang, yang biasa dikenal sebagai Naive Bayes Classifier. Melalui metode ini, peneliti akan mendapatkan beberapa keuntungan, seperti tingkat keakuratan data yang tinggi, proses yang cepat, dan hasil yang sederhana. Naive Bayes Classifier merupakan salah satu algoritma yang efektif dalam klasifikasi data, karena dapat mengasumsikan independensi antara fitur-fitur yang ada dalam data, meskipun dalam kenyataannya fitur-fitur tersebut mungkin terkait. Karena itu, metode ini sering digunakan dalam berbagai aplikasi, termasuk analisis sentimen, di mana akurasi dan efisiensi dalam pengelompokan data menjadi sangat penting. Dengan menerapkan Naive Bayes Classifier, penelitian ini diharapkan dapat memberikan hasil analisis yang handal dan efisien untuk memahami sentimen masyarakat terhadap topik yang diteliti.[16]

$$P(H|X) = P(X|H) P(H) / P(X) \quad (3)$$

Keterangan :

P(X): Peluang pada X

P(X|H): Peluang pada X menurut keadaannya

P(H): Peluang Hipotesis prior prob (H)

P(H|X): Peluang Hipotesis posteriori prob (H sesuai kondisi X)

H: Hipotesis kelas spesifik (X)

X: Kelas data yang belum diketahui.