BAB III
METODE PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan **kuantitatif** dengan metode **analisis sentimen**, yang bertujuan untuk mengklasifikasikan opini masyarakat terhadap program "Rencana Makan Gratis" di media sosial X. Data dikumpulkan dalam bentuk komentar pengguna dan dianalisis menggunakan teknik **Machine Learning** berbasis **Naïve Bayes**.

3.2 Sumber dan Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh dari media sosial **X (Twitter)** melalui proses web scraping. Beberapa tahapan dalam pengumpulan data adalah sebagai berikut:

1. **Menentukan Kata Kunci**
   - Menggunakan kata kunci yang berkaitan dengan "Rencana Makan Gratis" seperti **"makan gratis Prabowo"**, **"program makan Prabowo-Gibran"**, dan istilah lain yang relevan.
2. **Pengambilan Data**
   - Menggunakan **API X** untuk mengambil komentar terkait dengan program kebijakan tersebut.
   - Data yang dikumpulkan mencakup **501 komentar** yang memiliki tingkat interaksi tinggi (retweet dan like).
3. **Penyaringan Data**
   - Menghapus data yang tidak relevan, seperti spam atau komentar yang tidak berhubungan dengan topik penelitian.

3.3 Proses Text Preprocessing

Data teks yang diperoleh perlu diproses sebelum dianalisis. **Text Preprocessing** dilakukan melalui beberapa tahap berikut:

1. **Case Folding** – Mengubah seluruh teks menjadi huruf kecil untuk menjaga konsistensi.
2. **Cleansing** – Menghapus simbol, angka, tanda baca, dan karakter yang tidak diperlukan.
3. **Tokenizing** – Memecah kalimat menjadi kata-kata individual.
4. **Stopword Removal** – Menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap analisis.
5. **Stemming** – Mengubah kata menjadi bentuk dasar untuk mengurangi variasi kata yang berbeda tetapi memiliki makna yang sama.

3.4 Metode Pembobotan TF-IDF

Untuk mengekstrak fitur dari teks, digunakan teknik **Term Frequency-Inverse Document Frequency (TF-IDF)**, yang bertujuan memberikan bobot pada kata-kata berdasarkan frekuensinya dalam dataset. Perhitungan TF-IDF dilakukan dengan rumus berikut:

$$TF = \frac{Jumlah\ kemunculan\ kata}{Total\ kata\ dalam\ dokumen}$$

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

Dimana:

- *N* adalah jumlah total dokumen dalam dataset.
- *df* adalah jumlah dokumen yang mengandung kata tertentu.
- Hasil perkalian *TF* dan *IDF* menghasilkan nilai **TF-IDF** yang digunakan sebagai fitur dalam model klasifikasi.

3.5 Algoritma Klasifikasi Naïve Bayes

Model klasifikasi yang digunakan dalam penelitian ini adalah **Multinomial Naïve Bayes**, yang bekerja dengan menghitung probabilitas suatu komentar termasuk dalam kategori **positif** atau **negatif**. Formula utama dari **Teorema Bayes** yang digunakan dalam algoritma ini adalah:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Dimana:

- $P(C|X)$ adalah probabilitas suatu komentar tergolong dalam kelas **positif** atau **negatif**.
- $P(X|C)$ adalah probabilitas fitur muncul dalam kelas tertentu.
- $P(C)$ adalah probabilitas awal dari kelas tersebut.
- $P(X)$adalah probabilitas keseluruhan dari data yang dianalisis.

3.6 Pengujian dan Evaluasi Model

Model yang dikembangkan diuji menggunakan **Confusion Matrix** untuk mengukur kinerjanya dalam klasifikasi sentimen. Metode evaluasi yang digunakan meliputi:

1. **Akurasi** – Mengukur seberapa banyak prediksi model yang benar dibandingkan dengan total data uji.
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
2. **Precision** – Mengukur ketepatan model dalam mengklasifikasikan data positif dan negatif.
$$Precision\ Negative = \frac{TN}{TN + FN} \qquad Precision\ Positive = \frac{TP}{TP + FP}$$
3. **Recall** – Mengukur sejauh mana model dapat menangkap semua data positif atau negatif.
$$Recall\ Negative = \frac{TN}{TN + FN} \qquad Recall\ Positive = \frac{TP}{TP + FP}$$
4. **F1-Score** – Rata-rata harmonik antara Precision dan Recall.
$$F1 - Score\ Negative = \frac{2 \times PN \times RN}{PN + RN}$$

$$F1 - Score\ Positive = \frac{2 \times PP \times RP}{PP + RP}$$

3.7 Pembagian Data

Dataset yang dikumpulkan dibagi menjadi dua bagian utama untuk memastikan model dapat diuji dengan baik:

- **Training Data (80%)** – Digunakan untuk melatih model agar memahami pola dalam data.
- **Testing Data (20%)** – Digunakan untuk menguji model dengan data yang belum pernah dipelajari sebelumnya.

3.8 Implementasi Sistem

Setelah proses preprocessing, pembobotan, dan pelatihan model dilakukan, langkah terakhir adalah **implementasi sistem**. Model yang dikembangkan akan menerima input berupa teks komentar, kemudian mengklasifikasikan apakah komentar tersebut memiliki **sentimen positif atau negatif**. Hasil klasifikasi ditampilkan dalam bentuk visualisasi grafik untuk mempermudah analisis lebih lanjut.