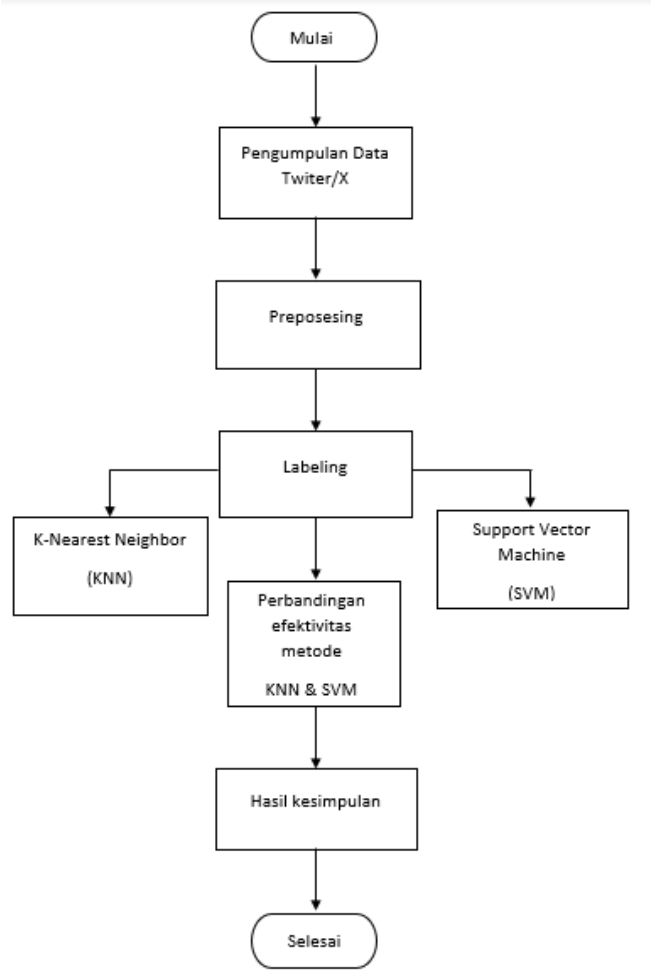
BAB III

METODOLOGI PENELITIAN

Penelitian ini akan melakukan sebuah analisis sentiment dengan menklasifikasikan data pada twitter. Data yang telah dipereloh dari Twitter akan diproses dengan text mining untuk menghindari data yang tidak diinginkan, setelah itu dikelompokan data tweet atau komentar berdasarkan klasifikasi positif dan negatif . Pengelompokan ini menggunakan perbandingan metode KNN dan SVM



Gambar 1. Tahapan Metodologi Penelitian

3.1 Tahapan metodologi penelitian terbagi menjadi beberapa tahapan sebagai berikut:

1. **Pengumpulan Data** : Pada tahap ini dilakukan pengambilan data dengan cara crawling data dari twiter/x
2. **Preposesing** : diatas telah didapatkan hasil dari crawling data dari Twitter, maka selanjutnya kita akan ketahap preprocessing agar data-data yang telah crawling tersebut dapat menjadi terstrTextur. Berikut ini adalah tahapan dari proses preprocessing:

    - **Tokenizing** : Tokenizing merupakan proses pemecahan teks menjadi kata tunggal dan menghapus tanda baca serta angka, sesuai dengan kamus yang telah ditentukan. Maka berdasarkan dataset ditwitter, pada umumnya pemilik akun hanya menggunakan kata-kata baku bahkan menggunakan kata-kata gaul untuk membuat sebuah ciutan atau tweet. Tugas pada Tokenizing ini adalah mengubah kata-kata baku tersebut sehingga lebih mudah dimengerti.

    - **Stopword Removing** : Stopword removing merupakan proses menghilangkan kata tidak penting dalam teks. Hal ini dilakukan untuk memperbesar akurasi dari pembobotan. Dalam penelitian ini Stopword removing digunakan untuk menghilankan kata-kata seperti: dan, atau, mungkin, ini, itu dan sebagainya merupakan kata yang dapat dihilangkan.

    - **Post Tagging Tweet** :Selanjutnya adalah melakukan pelabelan secara manual dari beberapa proses yang telah diciptakan tadi. Dengan adanya sebuah label maka tweet akan mudah dikelompokan, adapun label yang digunkan sebanyak tiga label

yakni Positif, Negatif dan Netral. Preprocessing telah dibuat berdasarkan point-point diatas ada beberapa data yang tidak terpakai juga bakalan dibuang, seperti tweet yang mengandung unsur emoticon, url, rt, gambar (biasanya menjadi tweet yang kosong pada saat terjadi crawling data)

3. **Labeling :** Labeling dengan menghitung nilai kata per kata positif dan negatif merupakan metode analisis sentimen berbasis leksikon. Metode ini bekerja dengan cara mengidentifikasi kata-kata yang bersifat positif atau negatif dalam sebuah teks, kemudian menghitung skor atau nilai sentimen berdasarkan jumlah kata-kata tersebut.

   Tujuan dari labeling adalah untuk melatih model pembelajaran mesin (machine learning) sehingga model tersebut dapat secara otomatis mengklasifikasikan data baru berdasarkan pola yang sudah dipelajari dari data yang sudah diberi label. Dengan kata lain, labeling membantu model memahami sentimen atau kategori dari data tersebut.

   Berikut adalah langkah-langkah umum dalam proses labeling ini:

   - **Pemilihan Leksikon Sentimen**: Leksikon atau kamus sentimen adalah daftar kata yang telah diberi nilai sentimen (positif atau negatif). Kata-kata dalam leksikon ini diberi bobot, misalnya, +1 untuk kata positif dan -1 untuk kata negatif.
   - **Pemisahan Kata dari Teks:** Setiap kalimat atau teks dipecah menjadi kata-kata terpisah (tokenizing).
   - **Pencocokan Kata dengan Leksikon:** Kata-kata dalam teks dibandingkan dengan daftar kata pada leksikon. Jika kata yang ditemukan ada di leksikon, maka nilai (positif atau negatif) dari kata tersebut ditambahkan ke total nilai sentimen teks.

- **Penghitungan Total Nilai Sentimen:** Setelah semua kata diidentifikasi, nilai-nilai positif dan negatif dijumlahkan. Jika jumlah totalnya:> 0, teks diberi label positif.

4. K-Nearest Neighbor (KNN): Metode pertama yang digunakan untuk analisis data adalah K-Nearest Neighbor (KNN). Metode ini adalah algoritma pembelajaran yang sederhana namun efektif, yang mengklasifikasikan data berdasarkan data tetangga terdekatnya.
5. Support Vector Machine (SVM): Metode kedua yang digunakan adalah Support Vector Machine (SVM). SVM adalah algoritma pembelajaran mesin yang lebih kompleks yang bertujuan untuk menemukan hyperplane terbaik yang memisahkan data ke dalam kelas-kelas yang berbeda.
6. Perbandingan Efektivitas Metode KNN & SVM: Setelah kedua metode (KNN dan SVM) diterapkan, tahap ini membandingkan hasil dari kedua metode tersebut untuk melihat mana yang lebih efektif dalam mengklasifikasikan data Twitter/X. Pengukuran efektivitas biasanya dilakukan dengan menghitung metrik seperti akurasi, presisi, recall, dan F1-score.
7. Hasil Kesimpulan: Berdasarkan perbandingan efektivitas, kesimpulan akhir diambil untuk menentukan metode mana yang lebih baik atau lebih cocok digunakan dalam kasus ini. Kesimpulan ini penting untuk memahami seberapa baik kedua metode bekerja dengan dataset yang dianalisis.