

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

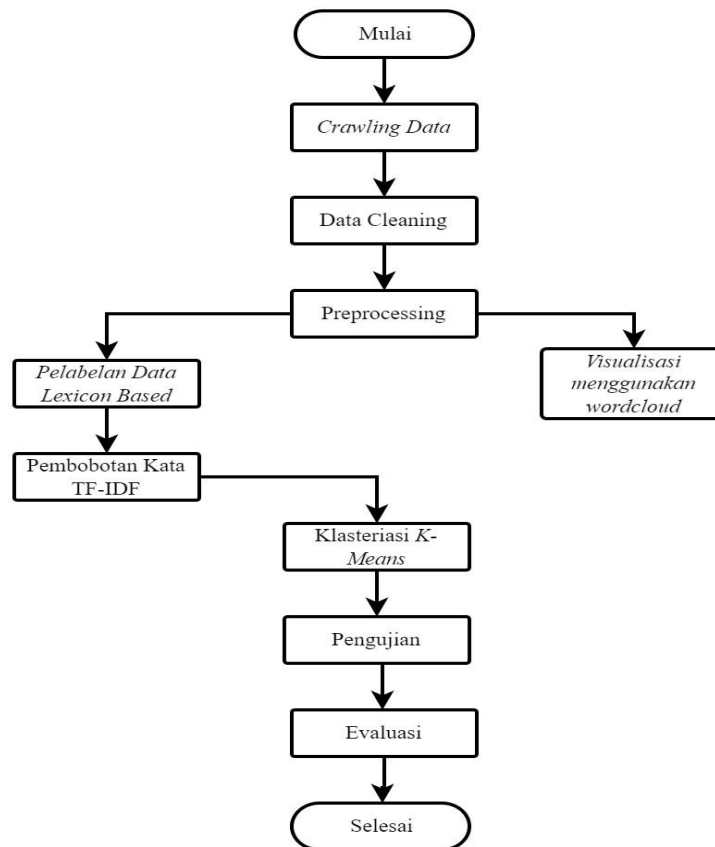
Jenis Penelitian yang digunakan pada penelitian ini ialah kuantitatif dikarenakan hasil dari penelitian ini akan menghasilkan angka untuk didapatkan statistik untuk menganalisisnya. Penelitian ini menggunakan algoritma *K-Means* yang memiliki kompleksitas waktu yang cukup rendah dan dapat bekerja dengan kumpulan data yang besar. Penulis menggunakan 1120 data dimana data yang di uji sebanyak 300 data.

3.2 Prosedur Penelitian

Prosedur penelitian merujuk pada serangkaian langkah atau tahapan sistematis yang diterapkan dalam pelaksanaan suatu penelitian. Tujuan dari prosedur penelitian ini adalah untuk menjamin bahwa penelitian dilaksanakan dengan metode yang dapat dipercaya, terorganisir, dan menghasilkan informasi yang akurat dan signifikan

3.2.1 Alur Pengolahan Data

Dalam pelaksanaan penelitian ini, diperlukan suatu gambaran mengenai cara penyelenggaraan penelitian. Agar pengamatan menjadi lebih mudah, tahapan pengolahan data dapat dipaparkan secara visual melalui gambar 3.1. (Sugiyono, 2013)



Gambar 3.1 Alur Pengolahan Data

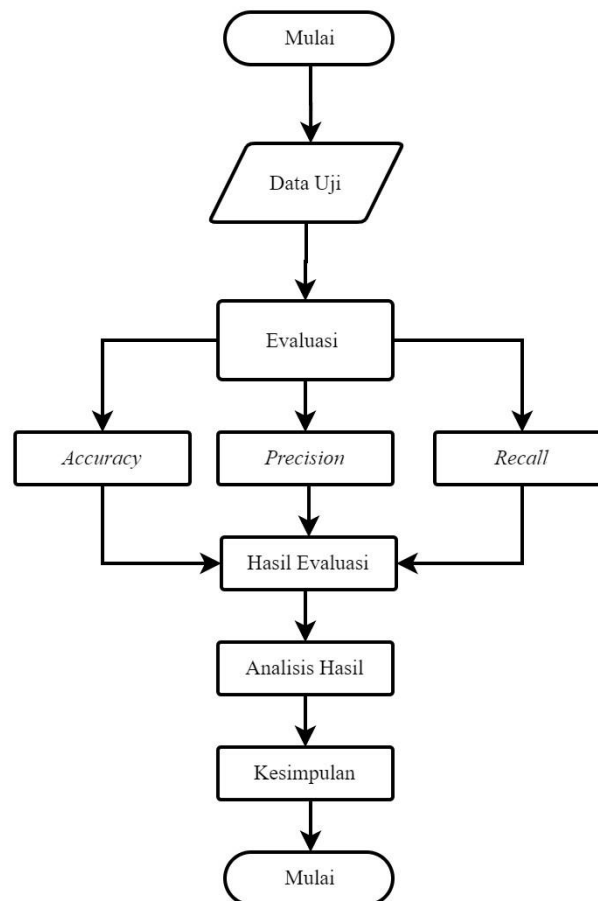
Dari Gambar 3.1 di atas dapat dijelaskan bahwa:

- a. *Crawling Data* merupakan pengumpulan dataset menggunakan *coding* bahasa *python* pada *google colab* menggunakan *library pandas* dan *library sns* untuk. Pengumpulan data diambil pada sosial media *twitter* dengan topik yang dicari 'tka dihapus' dalam rentang waktu mulai bulan januari tahun 2022 hingga bulan february tahun 2023, data yang terkumpul sebanyak 1120 data.
- b. *Data Cleaning* merupakan langkah untuk pembersihan data pada kalimat seperti *username*, tanda baca, dan URL.
- c. *Preprocessing Data* pada tahap ini dilakukan beberapa langkah yaitu :
 1. *Transform Case* proses untuk mengubah semua huruf kapital menjadi non kapital.
 2. *Tokenize* proses untuk memecahkan kata-kata yang terdapat dalam setiap kalimat.

3. *Stopword* proses untuk menghilangkan kata tidak penting yang terdapat dalam *tweet*. Pada proses tersebut penulis menggunakan kamus *Stopword* berbahasa Indonesia
 4. *Stemming* proses untuk menghilangkan imbuhan pada kata.
- d. Pelabelan Data menggunakan *Lexicon Based* pelabelan menggunakan kamus *Lexicon Based*, hasil dari proses pelabelan akan memberikan gambaran sentimen manakah yang unggul apakah positif atau negatif.
 - e. Pembobotan kata TF-IDF proses pembobotan kata menggunakan TF-IDF untuk memberikan nilai pada setiap kata dan kalimat untuk keperluan proses klusterisasi menggunakan algoritma *K-Means*.
 - f. Klusterisasi menggunakan algoritma *K-Means* setelah melewati proses TF-IDF maka dilakukan klusterisasi menggunakan algoritma *K-Means*, dimana pada proses tersebut penulis melakukan perbandingan perhitungan menggunakan perhitungan manual dengan menggunakan persamaan *Euclidean Distance* dan perhitungan menggunakan aplikasi WEKA.
 - g. Pengujian setelah melakukan proses klusterisasi tahap selanjutnya ialah melakukan pengujian pada algoritma *K-Means* pengujian dilakukan dengan membandingkan sentimen yang didapatkan menggunakan bantuan pelabelan *vadersentiment* dan *Lexicon Based*.
 - h. Evaluasi dilakukan setelah melakukan pengujian pada dua kamus pelabelan menggunakan model evaluasi *accuracy*, *precision*, dan *recall*.

3.2.2 Alur Evaluasi Data

Dalam aplikasi pemodelan prediksi atau *forecasting* menggunakan algoritma dalam machine learning, perlu dilakukan pemeriksaan dan evaluasi terhadap model yang telah dibuat untuk menilai sejauh mana efektivitasnya dalam melakukan prediksi terhadap data baru. Dalam penelitian ini, evaluasi terhadap model prediksi atau *forecasting* pada algoritma dilakukan dengan menggunakan tiga jenis penilaian model. Model yang digunakan untuk evaluasi diantaranya *accuracy*, *precision*, *recall*. Untuk memudahkan pengamatan alur evaluasi dapat dilihat pada gambar 3.2.



Gambar 3.2 Alur Evaluasi Data

Berdasarkan Gambar 3.2 dapat dijelaskan sebagai berikut :

a. Data Uji

Data Uji didapatkan setelah melakukan proses pengujian menggunakan dua kamus pelabelan yaitu *vadersentiment* dan *lexicon based*. Hasil data uji tersebut akan dilakukan evaluasi model dengan tiga skenario percobaan 50, 100, dan 150 data.

b. Evaluasi model

Evaluasi model pada penelitian ini menggunakan tiga model evaluasi yang diantaranya *accuracy*, *precision*, dan *recall*. Hasil dari evaluasi yang telah dilakukan akan menampilkan nilai evaluasi yang nantinya akan dibandingkan dan dianalisis.

c. Kesimpulan

Setelah evaluasi model dilakukan, tahap terakhir ialah penarikan kesimpulan. Pada tahapan ini akan dilakukan analisis dan perbandingan terhadap dua kamus pelabelan yang telah dilakukan evaluasi pada tahap sebelumnya.

3.3 Data Penelitian

Data penelitian yang digunakan ialah sentimen-sentimen yang diberikan oleh masyarakat pada aplikasi *twitter*. Data yang diambil merupakan *tweet* yang diambil dari bulan Januari tahun 2022 sampai dengan Februari tahun 2023 sebanyak 1120 data. Berikut pada tabel 3.1 merupakan contoh data yang digunakan dalam penelitian ini.

Tabel 3.1 Data Sentimen pada aplikasi *Twitter*

No	Tweet
1	@dapaaah @sociotalker Dgn diubah dihapus TKA nya, yg masuk SBMPTN kurang qualified dibanding yg lulus SBMPTN dgn ada TKA.
2	@edufikation True banget sih,aku gagal sbmptn 2022,dan aku udah mulai belajar TKA karena utbk kemarin aku lemah banget
3	Saya dulu sbmptn kayaknya lebih enjoy ngerjain tps daripada tka soshum deh.
4	i'm actually happy for this decision. aku anak bahasa dapet keadian karena, di kelas ga diajarin mapel soshum tp
5	mau utbk tps dihapus, tka dihapus atau yg ikut utbk sampe 1 juta org jg kalo

3.4 Analisis Data

Peneliti secara terstruktur mencari dan mengorganisasi data dari wawancara, catatan lapangan, dan dokumentasi dalam analisis data. Langkah-langkah tersebut mencakup pengelompokan data ke dalam kategori, dekomposisi dalam kelompok tertentu, sintesis, pembentukan pola, pemilihan objek yang relevan untuk studi, dan penyusunan kesimpulan. Tujuannya adalah agar peneliti dan orang lain dapat memahami informasi tersebut dengan baik. Dalam penelitian ini, teknik analisis data diterapkan melalui pengujian menggunakan tiga model evaluasi, *accuracy*, *recall*, *precision*.

Riyadh Ahmad Faridz, 2024

ANALISIS SENTIMEN TWITTER MENGGUNAKAN ALGORITMA
K-MEANS MENGENAI DAMPAK PENGHAPUSAN TKA SNBT TAHUN 2023

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

3.4.1 Accuracy (Akurasi)

Akurasi adalah pengukuran sejauh mana model dapat memprediksi dengan tepat. Model mencapai akurasi tinggi ketika rasio prediksi benar (baik positif maupun negatif) dibagi dengan jumlah total prediksi. Kinerja yang baik dalam memprediksi tercermin dari tingginya akurasi, yang memberikan gambaran umum tentang performa keseluruhan model. Berikut merupakan persamaan 3.1 untuk mencari nilai *accuracy*.

$$\frac{\text{Jumlah data yang diprediksi benar}}{\text{Total data yang diprediksi}} = \quad (3.1)$$

3.4.2 Precision (Presisi)

Presisi, atau *precision*, mengukur sejauh mana prediksi positif yang dibuat oleh model adalah benar. Presisi dihitung dengan membagi jumlah prediksi benar positif dengan total prediksi positif. Presisi memberikan ukuran sejauh mana prediksi positif model dapat diandalkan. Pertimbangan ini bermanfaat ketika tujuannya adalah untuk meminimalkan jumlah positif palsu. Berikut merupakan persamaan 3.2 untuk mencari nilai *precision*.

$$\frac{\text{Jumlah True Positif}}{\text{Jumlah True Positif} + \text{Jumlah False Positif}} = \quad (3.2)$$

3.4.3 Recall

Recall, atau yang juga dikenal sebagai sensitivitas atau "*true positive rate*" mengukur kemampuan model untuk mengidentifikasi semua kasus positif yang sebenarnya. Tingginya nilai *recall* menunjukkan bahwa model dapat menangkap sebagian besar kasus positif yang sebenarnya. Ini bermanfaat ketika tujuannya adalah untuk meminimalkan jumlah positif yang terlewat. Berikut merupakan persamaan 3.2 untuk mencari nilai *recall*.

$$\frac{\text{Jumlah True Positif}}{\text{Jumlah True Positif} + \text{Jumlah False Negatif}} = \quad (3.3)$$

3.5 Alat Penelitian

Dalam pengerjaan proses penelitian ini menggunakan perangkat yang memadai dalam pengerjaannya. Terdapat dua jenis perangkat dalam pengerjaan penelitian ini yaitu:

3.5.1 Kebutuhan Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam pengerjaan tugas akhir ini sebagai penunjang kinerja aplikasi. Adapun perangkat lunak yang digunakan ialah:

- a. Sistem Operasi Windows 11
- b. *Google Colaboratory*
- c. *Microsoft Excel*
- d. WEKA (*Waikato Environment for Knowledge Analysis*)

3.5.2 Kebutuhan Perangkat Keras (*Hardware*)

Perangkat keras digunakan untuk media implementasi aplikasi yang dirancang dan harus sesuai dengan kebutuhan aplikasi. Adapun perangkat keras yang digunakan ialah:

- a. *Processor Intel(R) Core(TM) i5-8625U CPU @ 1.60GHz*
- b. *RAM 8,00 GB*
- c. *System type 64 bit*