

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Data yang dianalisis dalam studi ini merupakan respon pengguna Indonesia terhadap performa ekspedisi PT. Tiki Jalur Nugraha Ekakurir (JNE) melalui aplikasi atau situs *web Google Play* dengan rentang waktu tahun 2021-2022. Pertahun 2022, terdapat 118rb komentar sejak aplikasi *My JNE* pertama kali dirilis, dan jumlahnya terus bertambah dari waktu ke waktu. Data yang diperoleh akan dianalisis dengan algoritma klasifikasi. Kemudian di simulasikan menggunakan *Google Colab/Python*.

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk meneliti dan menganalisis data terkait ulasan pelanggan pada layanan ekspedisi. Penggunaan pendekatan ini melibatkan penggunaan statistik dan metode matematika untuk meringkas, menggambarkan, dan menganalisis data numerik. Hal ini bertujuan untuk memberikan gambaran yang jelas dan terukur tentang distribusi data dan tren yang ada.

3.2 Populasi dan Sampel

Seluruh populasi data yang dianalisis dalam penelitian ini terdiri dari review data atau penilaian yang diberikan pengguna kepada PT. Tiki Jalur Nugraha Ekakurir (JNE) dari aplikasi atau situs *web Google Play* menggunakan aplikasi yang bernama *My JNE* dengan total 118.000 komentar dari pengguna. Namun, untuk keperluan penelitian, dilakukan pengambilan sampel dari populasi tersebut. Dalam konteks penelitian ini, sampel diambil dari ulasan pengguna mengenai JNE setelah diperbarui dengan versi terbaru aplikasi *My JNE*, yaitu versi 1.2.5 yang dirilis pada tanggal 10 Desember 2022. Jumlah sampel yang diambil sebanyak 10.415 komentar.

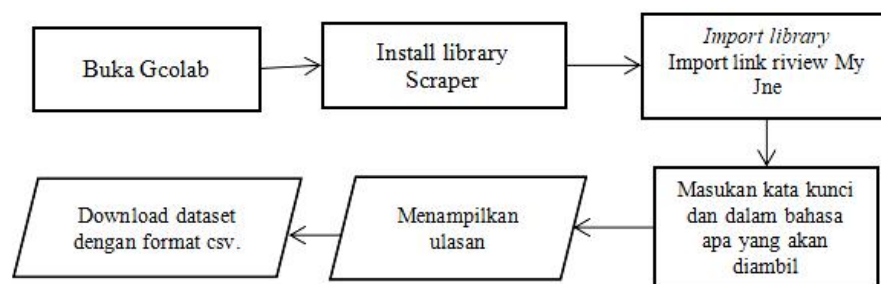
3.3 Variabel dan Definisi Operasional

Variabel penelitian ini digunakan sebagai pertimbangan saat memilih ulasan dari pengguna JNE atau ulasan *My JNE* dari *website/aplikasi Google Play Store*. Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Machine Learning , konteks machine learning mengacu pada elemen spesifik yang digunakan dalam proses pemodelan dan evaluasi. Ini mencakup berbagai parameter, metrik, dan langkah-langkah yang memungkinkan untuk mengatur, melatih, dan mengukur performa model.
2. Perbandingan data uji dan data latih, dalam penelitian adalah cara konkret untuk mengukur dan mengidentifikasi proporsi atau jumlah data yang digunakan dalam masing-masing kelompok. Pembagian data uji dan data latih adalah langkah penting dalam pengembangan model prediktif, seperti yang dilakukan dalam algoritma klasifikasi seperti *K-Nearest Neighbor* (KNN) dan *Decision Tree*.
3. Parameter Metriks dan Kriteria, faktor yang dapat dikendalikan atau dikelola dalam proses *machine learning* untuk membentuk, melatih, dan mengevaluasi model.

3.6 Teknik Pengumpulan Data

Berikut flowchart teknik pengumpulan data pada Gambar 3.1 :



Gambar 3.1 Flowchart pengumpulan data

Langkah mengambil data :

1. Buka *Google colab* di *Google Chrome*
2. Kemudian install *google-play-scraper* pada *Google Colab*
3. Setelah itu import library dan import link ulasan *My JNE* yang akan di gunakan.
4. Lalu memasukan kata kunci dan dalam bahasa apa ulasan yang akan diambil.
5. Tampilkan data ulasan untuk dicek, dan jika data sudah lengkap maka download dataset menggunakan format csv.

3.7 Teknik Analisis Data

Berikut merupakan teknik analisis data yang digunakan pada penelitian ini, antara lain :

Metode analisis deskriptif, yang membantu memahami data secara detail dan mendeskripsikan rating pengguna JNE secara keseluruhan melalui aplikasi *My JNE* di *website Google Play Store*

1. Metode analisis deskriptif, yang membantu memahami data secara detail dan mendeskripsikan rating pengguna JNE secara keseluruhan melalui aplikasi *My JNE* di *website Google Play Store*
2. *Google Colab*, yang membantu untuk mensimulasi data pada software berbasis formula *Excel* atau aplikasi spreadsheet lainnya.
3. Metode *Machine Learning* dengan pendekatan *Supervised Learning* yang digunakan untuk melakukan klasifikasi menggunakan metode *K-Nearest Neighbor* dan *Decision Tree* mengenai review yg berbentuk kelas positif juga negatif.
4. *Confusion Matrix*, membantu menghitung performa akurasi, *recall*, *precision* dan *F1-score*.
5. *Wordcloud*, yang membantu untuk memvisualisasikan kata-kata yang paling banyak digunakan atau paling banyak muncul pada komentar pengguna di aplikasi *My JNE*

3.8 Pembagian Data Uji dan Data Latih

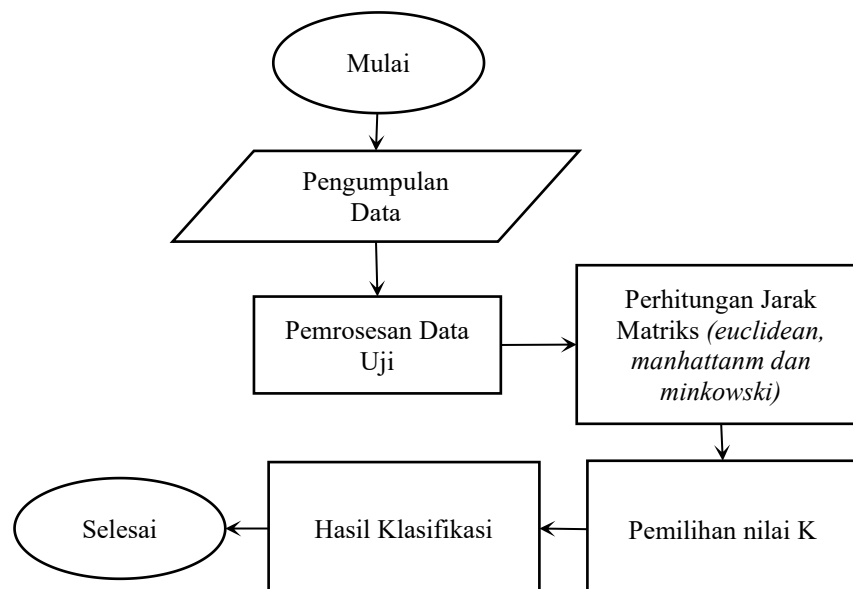
Algoritma klasifikasi menggunakan *data training* untuk membangun model klasifikasi. Model ini adalah representasi dari data yang dapat digunakan untuk memprediksi kelas data baru yang belum pernah ada sebelumnya. Sedangkan data uji adalah kumpulan data yang digunakan untuk mengetahui tingkat akurasi model yang dihasilkan kemudian *data testing* digunakan untuk memperkirakan seberapa baik model algoritma berhasil mengklasifikasikan dengan benar. Data yang sudah memiliki pengenalan kelas digunakan untuk data latih dan data uji. Untuk pembagian data pada penelitian ini menggunakan 2 data latih dan data uji yaitu yang pertama 80% data latih dan data uji 20% serta 70% data uji dan 30% data latih.

3.9 Analisis Kebutuhan

Analisis kebutuhan memainkan peran awal yang sangat penting dalam memastikan keberhasilan dan kepuasan pengguna serta pemangku kepentingan untuk solusi yang dikembangkan. Dalam konteks *machine learning*, analisis kebutuhan membantu menentukan tugas yang harus dilakukan oleh model dan juga menentukan kriteria evaluasi yang tepat untuk mengukur kinerja model. Dalam penelitian ini, akan menggunakan dua metode, yaitu *K-nearest neighbor* dan *Decision Tree*. Kedua metode ini termasuk dalam kategori *supervised learning* yang digunakan untuk melakukan klasifikasi pada dokumen teks.

3.10. Perancangan Sistem

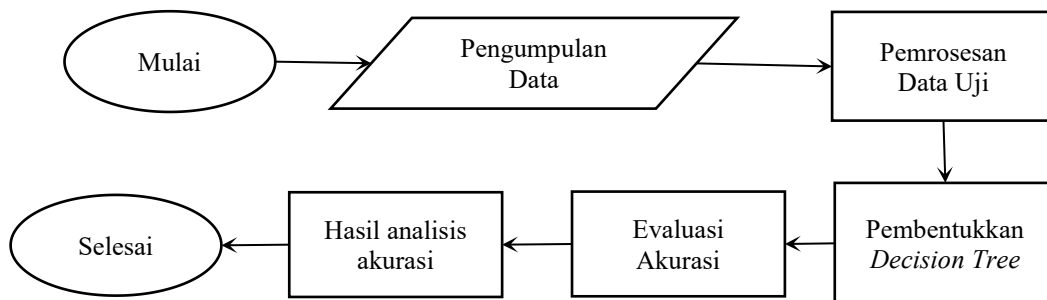
Dalam proses *K-Nearest Neighbor*, langkah yang perlu dijalankan adalah melakukan perhitungan kesamaan antara data uji dan seluruh data ulasan. Ini dilakukan dengan mengukur jarak antara data uji dan setiap data ulasan menggunakan metode matriks *Euclidean*, *Manhattan*, dan *Minkowski*. Dapat dilihat diagram alir metode K-NN pada Gambar 3.2 :



Gambar 3.2 Flowchart metode K-NN

Diagram alir ini memberikan gambaran visual tentang langkah-langkah utama dalam metode *K-Nearest Neighbor*, yang melibatkan pengumpulan data, perhitungan jarak, pemilihan tetangga terdekat, hasil klasifikasi, dan selesai. Proses ini diulang untuk setiap data uji yang perlu diklasifikasikan.

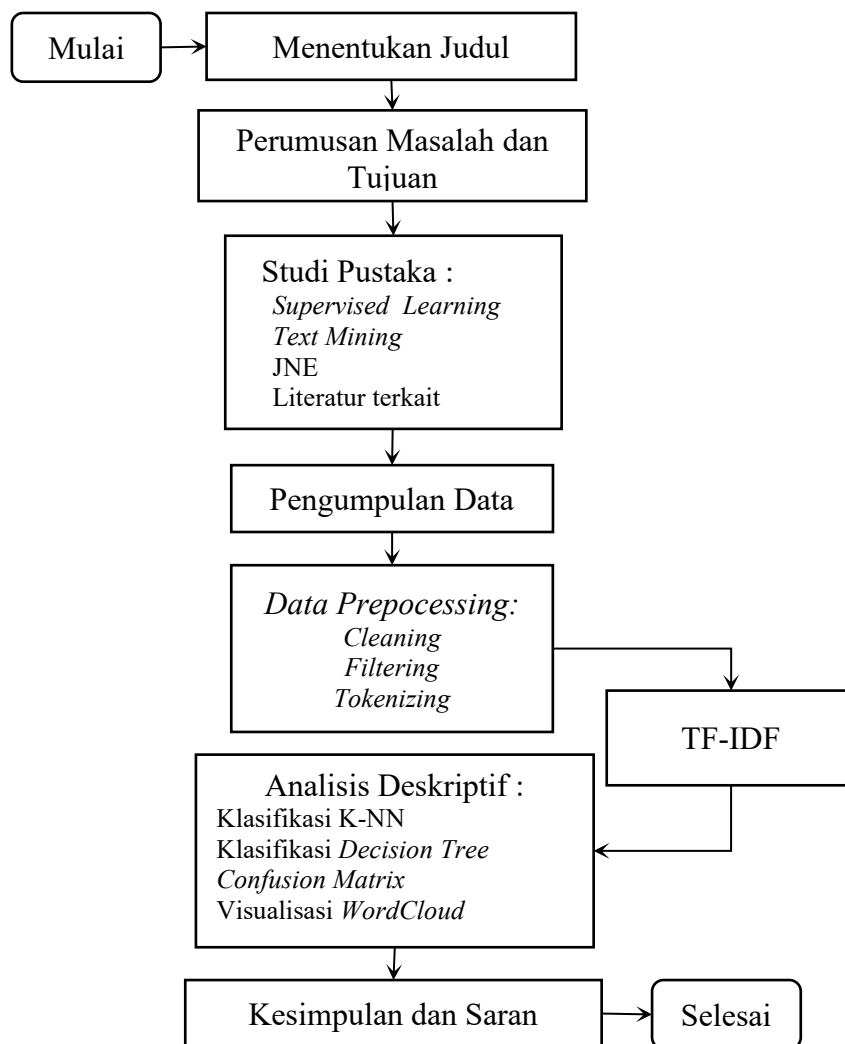
Dalam proses *Decision Tree*, tahap yang harus dilakukan adalah melakukan evaluasi karakteristik data uji dengan seluruh data ulasan yang ada. Langkah ini melibatkan pengukuran kedekatan antara data uji dan setiap data ulasan dengan menggunakan metode kriteria *Gini*, Entropi, dan *log_loss*. Berikut ini adalah diagram alir yang menggambarkan langkah-langkah dalam metode *Decision Tree* dapat dilihat pada Gambar 3.3 :



Gambar 3.3 Flowchart metode decision tree

Diagram alir ini mengilustrasikan langkah-langkah utama dalam membangun *Decision Tree*, mengukur akurasi, dan menganalisis hasil akurasi yang telah diperoleh. Proses ini dapat diulang dengan variasi parameter atau metrik evaluasi yang berbeda untuk mendapatkan pemahaman yang lebih baik tentang performa *Decision Tree*.

3.11 Alur Penelitian



Gambar 3.4 Alur penelitian

Pada Gambar 3.4 adalah alur dari penelitian ini. Penelitian dimulai dengan mengidentifikasi dan merumuskan masalah. Langkah ini melibatkan pencarian informasi dari berbagai sumber seperti artikel, jurnal, dan literatur lainnya. Permasalahan yang menjadi topik penelitian dengan pengamatan secara langsung, tetapi melihat kondisi saat ini melalui *website Google PlayStore* yaitu rendahnya rating pada aplikasi *My JNE* yaitu sekitar 1,9 pada tanggal 26 Januari 2023. Sehingga diperlukannya analisis sentimen untuk mengetahui persepsi pengguna ekspedisi JNE berdasarkan penilaian ulasan pada aplikasi *My JNE* di *website/aplikasi Google Play*, lalu bagaimana performa penerapan metode algoritma *K-NN* dan *Decision Tree*

dalam mengklasifikasikan persepsi pengguna aplikasi *My JNE*, dan bagaimana nilai parameter matriks kinerja klasifikasi algoritma *K-Nearest Neighbor* dan *Decision Tree*

Tahap selanjutnya pengumpulan data, data yang diambil melalui ulasan di aplikasi *My JNE* pada web/aplikasi Google Play dengan sampel 10.415 ulasan. Kemudian data melalui tahap preprocessing tahap ini merupakan tahap awal untuk mengubah struktur konten data ke dalam bentuk yang sesuai agar dapat diolah oleh algoritma K-NN dan *Decision Tree*. Dengan melalui tahap *Data Cleaning*, *Filtering*, dan *Tokenizing*. Selanjutnya tahap TF-IDF Hasil dari pembobotan TF-IDF selanjutnya akan menjadi masukan untuk proses pengklasifikasian K-NN. Berikut tahapan untuk menentukan pembobotan pada setiap kata yaitu :

- 1) Menghitung frekuensi kata (*Term Frequency*) dari setiap kata pada setiap ulasan. Ini mencakup menghitung berapa kali kata tertentu muncul dalam suatu ulasan.
- 2) Menghitung *Document Frequency*, yaitu jumlah dokumen yang mengandung kata tersebut. Ini mengukur seberapa umum atau jarang suatu kata muncul di seluruh dokumen.
- 3) Menghitung nilai *Inverse Document Frequency* (IDF) dari kata-kata tersebut. IDF dihitung dengan cara membagi total jumlah data ulasan dengan jumlah *Document Frequency*. Nilai IDF akan menunjukkan seberapa penting suatu kata dalam menggambarkan ulasan secara keseluruhan.
- 4) Menghitung bobot kata (TF-IDF) dengan mengalikan nilai *Term Frequency* (TF) dengan nilai IDF. Langkah ini memberikan bobot pada setiap kata berdasarkan frekuensi dan pentingnya kata dalam koleksi ulasan.
- 5) Melakukan normalisasi vektor pembobotan pada setiap ulasan menggunakan metode *Normalisasi L2*. Ini bertujuan untuk

memastikan bahwa panjang vektor pembobotan pada setiap ulasan tidak memiliki pengaruh yang signifikan pada hasil akhir.

Kemudian dilanjut dengan tahap klasifikasi menggunakan algoritma K-NN dan Decision Tree. Metode *Machine Learning* dengan pendekatan *Supervised Learning* yaitu *K-Nearest Neighbor* yang digunakan selama melakukan klasifikasi ulasan dengan bentuk kelas positif dan negatif. Metode *K-Nearest Neighbor* melewati beberapa proses selama klasifikasi untuk mendapatkan hasil klasifikasi. Lalu hasil klasifikasi dari algoritma kedua ini akan melalui tahap *Confusion Matrix*. Pada tahap *confusion matrix* ini digunakan untuk membentuk suatu kelompok masalah dengan akurasi yang tepat. *Confusion Matrix* dan matrik evaluasi membantu dalam menggambarkan kinerja model klasifikasi secara lebih detail. Dan juga membantu dalam mengidentifikasi area dimana model berhasil dan dimana model memerlukan perbaikan. Selanjutnya tahap terakhir untuk memvisualisasi menggunakan Wordcloud. Visualisasi ini dilakukan dengan mengekstraksi semua data untuk mengidentifikasi kata yang paling banyak digunakan dalam komentar pengguna. metode grafis yang digunakan untuk mewakili data teks dalam bentuk visual, di mana kata-kata yang paling sering muncul diberi bobot lebih besar dan ditampilkan dengan ukuran yang lebih besar dalam gambar. Ini memberikan gambaran visual tentang frekuensi kata-kata yang ada dalam teks dan membantu mengidentifikasi kata-kata yang paling umum atau dominan dalam kumpulan teks.