

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

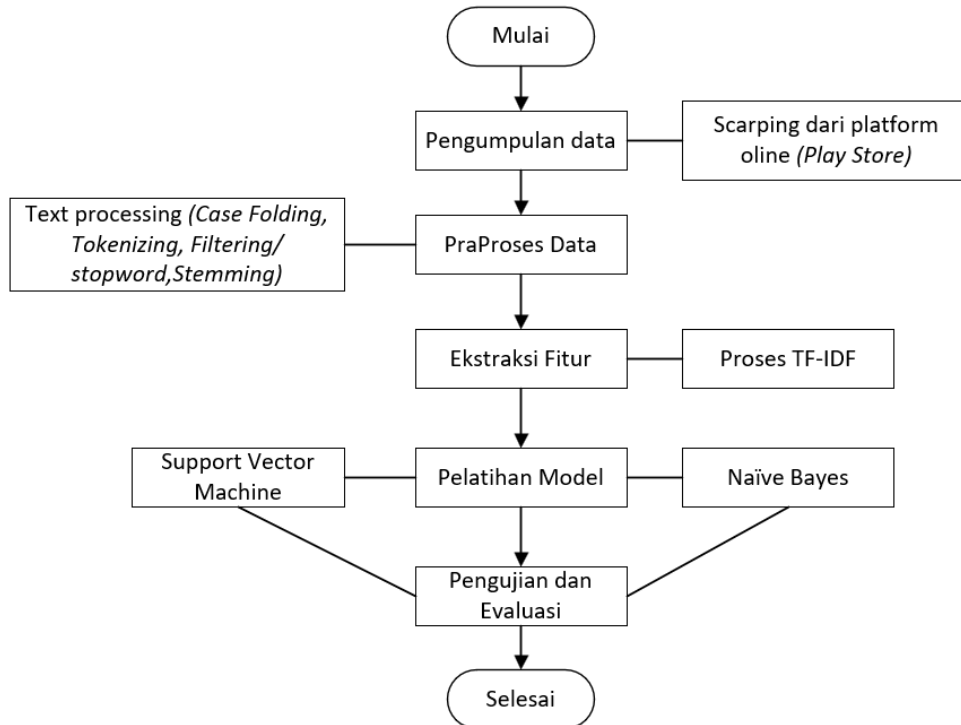
Objek penelitian ini adalah ulasan pengguna terhadap aplikasi Lalamove yang diambil dari platform Google Play Store menggunakan teknik *scraping*. Ulasan pengguna meliputi teks ulasan, skor penilaian (rating), dan informasi waktu ulasan. Data ulasan ini mencerminkan sentimen pengguna terhadap berbagai aspek layanan aplikasi Lalamove, seperti kecepatan pengiriman, layanan pelanggan, dan stabilitas aplikasi. Penelitian dilakukan secara daring karena pengambilan data bersumber dari platform digital. Penelitian ini menggunakan data ulasan dari pengguna di wilayah Indonesia.

Lingkup penelitian ini mencakup ulasan pelanggan dalam bahasa Indonesia dan Inggris yang diambil dari platform online, dengan skor penilaian 1-2 sebagai sentimen negatif dan 4-5 sebagai sentimen positif. Penelitian ini berfokus pada analisis performa algoritma Support Vector Machine (SVM) dan Naive Bayes, dengan evaluasi berdasarkan akurasi, presisi, recall, dan F1-score. Tahapan pre-processing seperti tokenisasi, stemming, dan normalisasi dilakukan untuk mempersiapkan data, yang kemudian dibagi menjadi data latih dan data uji. Hasil penelitian akan membandingkan efektivitas kedua algoritma dalam klasifikasi sentimen.

3.2 Prosedur Penelitian

Prosedur penelitian disusun secara sistematis untuk mencapai tujuan penelitian. Berikut adalah tahapan penelitian beserta diagram alur yang menggambarkan proses secara keseluruhan.

3.3 Diagram Alur Penelitian



Gambar 3. 1 Diagram Alur penelitian(Huda et al. 2024).

Gambar di atas menunjukkan *Diagram Alur Penelitian*, yang merupakan tahapan sistematis dalam melakukan penelitian berbasis data. Berikut adalah penjelasan untuk setiap langkah pada diagram:

1. Mulai: Langkah awal dalam proses penelitian, menandai dimulainya pekerjaan, termasuk identifikasi masalah dan tujuan penelitian.
2. Pengumpulan Data: Data dikumpulkan dari berbagai sumber yang relevan, seperti dataset ulasan aplikasi (*scraping* dari platform online) atau sumber lain yang sesuai dengan kebutuhan penelitian. Data ulasan pengguna diambil menggunakan teknik *scraping* dari Google Play Store menggunakan pustaka *google-play-scraper*.
 - a. Data yang diperoleh mencakup teks ulasan, skor penilaian, dan waktu ulasan.
 - b. Data mentah disimpan dalam format Excel untuk analisis lebih lanjut.

Dataset yang dibuat pada tahun 2024 merupakan kompilasi data dari 18 Juni 2020 hingga 24 Januari 2025 dan didasarkan pada ulasan pengguna aplikasi Lalamove yang dapat ditemukan di Google Play Store. Proses pengumpulan data dilakukan dengan menganalisis informasi yang baru ditemukan dan membuatnya relevan dengan tujuan penelitian. Jumlah total data yang diperoleh dari Google Play Store mencapai 10.000, yang kemudian digunakan sebagai dasar analisis untuk melihat pengalaman dan pandangan pengguna terkait aplikasi Lalamove. Salah satu teknik yang digunakan untuk pengumpulan data adalah scraping, yaitu sebuah metode yang menggunakan perangkat lunak komputer untuk mengumpulkan data secara otomatis dari situs web. Pada penelitian ini, teknik scraping digunakan untuk mengumpulkan data komentar pada website Google Play Store.

3. PraProses Data: Tahapan ini bertujuan untuk membersihkan dan menyiapkan data ulasan agar siap digunakan dalam analisis sentimen.
 - a. Pembersihan Data: Menghapus karakter khusus, angka, dan tanda baca.
 - b. Normalisasi: Mengubah kata-kata slang menjadi kata baku.
 - c. Case Folding: Mengubah semua teks menjadi huruf kecil untuk konsistensi.
 - d. Tokenisasi: Memisahkan kalimat ulasan menjadi kata-kata individu.
 - e. Stopword Removal: Menghapus kata-kata tidak bermakna seperti "dan," "atau," "di."
 - f. Stemming: Mengubah kata-kata menjadi bentuk dasar menggunakan pustaka Sastrawi.
4. Ekstraksi Fitur: Mengubah data menjadi representasi fitur yang bisa dimengerti oleh model pembelajaran mesin. Dalam penelitian berbasis teks, ini sering melibatkan metode seperti:
 - a. Menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) untuk mengekstrak fitur dari teks ulasan.

- b. Fitur-fitur yang dihasilkan adalah representasi vektor dari kata-kata yang digunakan dalam ulasan.
5. Pelatihan Model: Data ulasan yang telah diproses selanjutnya digunakan untuk melatih dua algoritma klasifikasi, yakni Support Vector Machine (SVM) dan Naive Bayes. Algoritma SVM menggunakan kernel linear untuk membangun model klasifikasi, sedangkan Naive Bayes—sebagai metode probabilistik—memanfaatkan distribusi kata dalam ulasan untuk memprediksi sentimen. Proses pelatihan dilakukan dengan membagi data ke dalam dua bagian, yaitu 70% sebagai data latih dan 30% sebagai data uji.
6. Pengujian dan Evaluasi: Model diuji menggunakan data uji untuk mengukur kinerjanya. Indikator evaluasi bisa berupa akurasi, presisi, *recall*, F1-score, atau *confusion matrix* untuk melihat seberapa baik model bekerja.

Tabel 3. 1 Confusion Matrix

	Actual Positif	Actual Negatif
Prediksi Positif	TP	FP
Prediksi Negatif	FN	TN

Dari tabel di atas, confusion matrix yang berisi True Positif (TP), True Negatif (TN), False Positif (FP) dan False Negatif (FN). TP yaitu persentase dari kelas positif yang sudah diklasifikasikan menjadi kelas positif. Ini berarti model berhasil mengidentifikasi dengan benar kasus-kasus yang benar-benar positif. Di sisi lain, TN yaitu persentase dari kelas negatif yang sudah diklasifikasikan menjadi kelas negatif, yang mengindikasikan bahwasannya model juga dapat membedakan dengan baik kasus-kasus yang sebenarnya negatif. Dengan demikian, TP dan TN memberikan gambaran mengenai kemampuan model dalam mengenali dan membedakan kelas positif serta negatif dengan akurasi yang baik. Dari capaian kalkulasi “Confusion matrix”, misalnya akurasi “accuracy”, presisi “precision”, “recall”, serta “f1-score”, dapat dihitung memakai persamaan yang disajikan.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} * 100\%$$

$$\text{Precision} = \frac{TP}{TP+FP} * 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} * 100\%$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} * 100\%$$

Pengujian

- a. Model diuji menggunakan data uji untuk mengevaluasi kinerjanya berdasarkan metrik berikut:
 - 1) Akurasi: Persentase prediksi yang benar dari seluruh data uji.
 - 2) Presisi: Seberapa banyak prediksi positif yang benar-benar positif.
 - 3) Recall: Kemampuan model untuk menemukan semua data positif.
 - 4) F1-Score: Rata-rata harmonis dari presisi dan recall.
- b. SMOTE (Synthetic Minority Oversampling Technique) digunakan untuk menangani ketidakseimbangan data.

Analisis Hasil

- a. Hasil evaluasi dibandingkan antara algoritma SVM dan Naive Bayes.
 - b. Analisis dilakukan untuk mengidentifikasi algoritma yang lebih efektif dalam klasifikasi sentimen ulasan pengguna.
 - c. Temuan hasil klasifikasi digunakan untuk memberikan rekomendasi praktis kepada pengembang aplikasi Lalamove.
7. Selesai: Penelitian diakhiri dengan analisis hasil, penarikan kesimpulan, dan dokumentasi temuan untuk keperluan publikasi atau penerapan. Hasil akhir penelitian didokumentasikan dan disusun dalam bentuk laporan ilmiah yang mencakup Kesimpulan serta rekomendasi

3.4 Alat dan Bahan

1. Perangkat Keras: Komputer dengan spesifikasi memadai untuk analisis data.

Device name : LAPTOP-65PL658O

Processor : AMD Ryzen 3 3250U with Radeon Graphics 2.60 GHz

Installed RAM: 8,00 GB (5,92 GB usable)

Device ID : 75727C1F-CF54-43D3-A7C1-7009F53E979D

Product ID : 00356-02364-71656-AAOEM

System type : 64-bit operating system, x64-based processor

Pen and touch : No pen or touch input is available for this display

2. Perangkat Lunak:

- a. Python 3.x dengan pustaka berikut:

- 1) google-play-scraper: Untuk pengambilan data ulasan.
- 2) nltk, Sastrawi: Untuk praproses teks.
- 3) scikit-learn: Untuk pelatihan dan pengujian algoritma.
- 4) pandas, matplotlib, seaborn: Untuk analisis data dan visualisasi.

- b. Microsoft Excel: Untuk penyimpanan data awal dan hasil.

- c. Toolsnya menggunakan scraping

- d. Aplikasinya menggunakan Google Collabs

- e. Bahasa emogramannya menggunakan Phyton