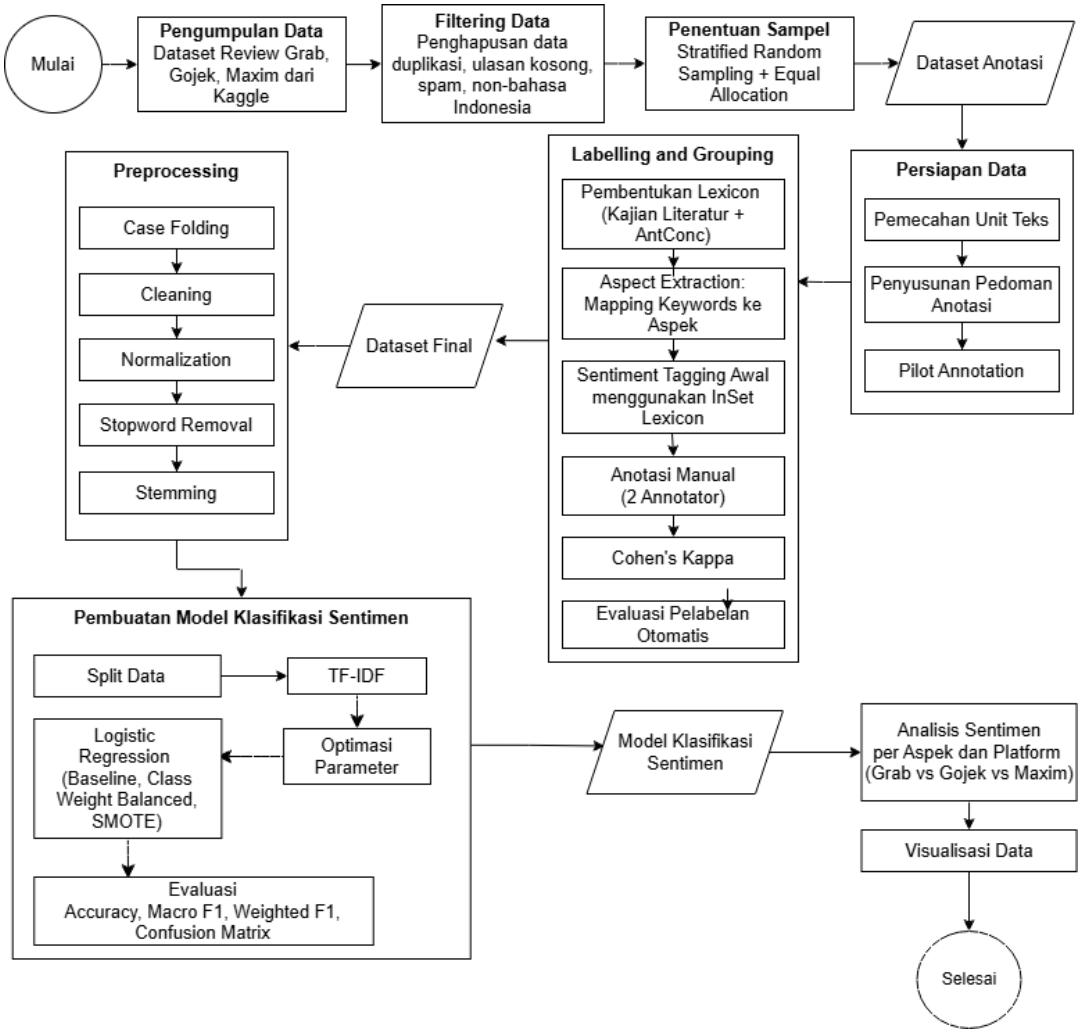
BAB III

METODOLOGI

3.1 Langkah-langkah Penelitian

Penelitian ini dilakukan untuk mengkaji sentimen pengguna berdasarkan aspek tertentu pada ulasan aplikasi Grab, Gojek, dan Maxim dengan menerapkan algoritma *Logistic Regression* serta representasi fitur TF-IDF. Tahapan penelitian dirancang secara berurutan, dimulai dari proses pengumpulan data, pembersihan dan pengolahan teks, pemberian label sentimen serta aspek melalui pendekatan semi-otomatis, pembentukan model klasifikasi, sampai pada tahap interpretasi dan penyajian hasil analisis.

Penelitian ini menggunakan pendekatan *Aspect-Based Sentiment Analysis* (ABSA) dua tahap. Tahap pertama adalah identifikasi aspek menggunakan pendekatan berbasis *lexicon* sebagai *pre-annotation*, kemudian hasil label awal diverifikasi secara manual untuk membentuk *ground truth* yang lebih andal. Tahap kedua adalah klasifikasi polaritas sentimen ke dalam kelas negatif, netral, dan positif menggunakan ekstraksi fitur TF-IDF dan algoritma *Logistic Regression*. Model yang dibangun kemudian dievaluasi untuk mengetahui kinerja klasifikasi serta distribusi sentimen pada setiap aspek layanan. Gambaran umum alur penelitian dapat dilihat pada Gambar 3. 1.



Gambar 3. 1 Alur Penelitian

3.2 Pengumpulan Data

Penelitian ini menggunakan data sekunder berupa ulasan pelanggan aplikasi Grab, Gojek, dan Maxim yang diperoleh dari platform Kaggle. Dataset tersebut merupakan hasil pengambilan ulasan pengguna dari Google Play Store oleh pihak ketiga dan disediakan secara terbuka untuk keperluan penelitian akademik.

Dataset awal dalam penelitian ini mencakup 300.000 ulasan pengguna, yang berasal dari tiga aplikasi dengan jumlah 100.000 ulasan pada setiap aplikasi, dengan periode waktu pengambilan data yang berbeda pada masing-masing

aplikasi. Data ulasan menggunakan bahasa Indonesia dan memuat informasi berupa teks ulasan, *rating* (1–5), tanggal ulasan, serta *metadata* pendukung lainnya.

Seluruh data tidak langsung digunakan secara keseluruhan, melainkan melalui tahap penyaringan (*filtering*) sebagai upaya untuk memperoleh dataset yang bersih, relevan, dan sesuai dengan kebutuhan penelitian. Proses ini meliputi penghapusan data duplikasi, ulasan kosong, serta ulasan yang terindikasi *spam*. Selain itu, dilakukan seleksi bahasa untuk memastikan hanya ulasan berbahasa Indonesia yang dibutuhkan dalam penelitian.

Setelah proses penyaringan, data yang digunakan merupakan data bersih yang siap untuk tahap analisis. Distribusi data kemudian dianalisis berdasarkan *rating* dan waktu ulasan. Data ulasan mencakup periode sekitar satu tahun, sehingga mampu merepresentasikan variasi kondisi layanan yang terjadi dalam jangka waktu tersebut. Sementara itu, distribusi *rating* menunjukkan dominasi ulasan dengan *rating* tinggi, di mana sebagian besar ulasan berada pada *rating* 5. Kondisi ini mengindikasikan kecenderungan sentimen positif yang tinggi serta adanya ketidakseimbangan kelas (*class imbalance*) yang perlu diperhatikan dalam proses pemodelan.

Selain pengumpulan data, dilakukan pula studi literatur untuk memperoleh pemahaman mengenai konsep *Aspect-Based Sentiment Analysis* (ABSA), teknik *text mining*, penggunaan *Lexicon* dalam pelabelan sentimen, serta penerapan algoritma *Logistic Regression* pada analisis sentimen ulasan aplikasi layanan *ride-hailing*.

3.3 *Preprocessing*

Pra-pemrosesan data dilakukan untuk membersihkan dan mempersiapkan teks ulasan agar dapat diolah secara maksimal oleh sistem. Pada penelitian ini, tahapan pra-pemrosesan dibatasi pada lima langkah utama untuk menjaga keseimbangan antara penyederhanaan teks dan pelestarian informasi sentimen.

1. *Case Folding*

Seluruh teks ulasan diubah menjadi huruf kecil (*lowercase*) untuk menyeragamkan representasi kata dan menghindari perbedaan makna akibat variasi penggunaan huruf kapital.

2. *Cleaning*

Proses pembersihan dilakukan dengan menghapus karakter non-alfabet, angka yang tidak relevan, simbol, URL, emoji, serta spasi berlebih yang tidak berkontribusi terhadap analisis sentimen.

3. *Normalization*

Normalisasi dilakukan untuk mengubah kata tidak baku atau slang menjadi bentuk baku, seperti “ga” menjadi “tidak”. Kamus normalisasi disusun berdasarkan eksplorasi data latih, referensi kamus *slang* Bahasa Indonesia, serta penyesuaian manual sesuai konteks ulasan aplikasi *ride-hailing*.

4. *Stopword Removal* (Selektif)

Stopword dihapus menggunakan daftar *stopword* dari NLTK yang telah dimodifikasi secara manual. Kata negasi seperti “tidak”, “kurang”, dan “bukan” tidak dihapus karena memiliki pengaruh penting terhadap makna sentimen dalam teks.

5. *Stemming* (Bahasa Indonesia)

Pada penelitian ini, proses *stemming* dilakukan menggunakan *library* Sastrawi, yaitu pustaka *stemming* Bahasa Indonesia.

3.4 Pelabelan Sentimen dan Pengelompokan Aspek

Pelabelan sentimen dan pengelompokan aspek dilakukan secara bersamaan menggunakan pendekatan semi-otomatis berbasis *Lexicon* sebagai *pre-annotation*. Selanjutnya, dilakukan verifikasi manual untuk memastikan kesesuaian label, sehingga diperoleh *ground truth* yang lebih akurat dan representatif.

3.4.1 Identifikasi Aspek

Pada tahap ini, identifikasi aspek dilakukan menggunakan pendekatan *Lexicon-based* yang disusun berdasarkan kajian literatur dan eksplorasi korpus ulasan pengguna. Eksplorasi korpus dilakukan menggunakan perangkat lunak AntConc melalui fitur *Word List* dan N-Gram untuk menemukan kata dan frasa yang sering terlihat dalam ulasan.

Sebelum analisis dilakukan, data telah melalui tahap *preprocessing* sehingga kata-kata umum (*stopwords*) tidak mendominasi hasil. Hasil eksplorasi korpus kemudian digunakan sebagai dasar dalam penyusunan aspect *Lexicon* dan penetapan aspek yang digunakan dalam penelitian.

Berdasarkan hasil kajian literatur dan eksplorasi korpus, ditetapkan lima aspek utama yang digunakan dalam penelitian ini, yaitu Sentimen Umum, Pengemudi, Layanan, Aplikasi, dan Harga/Biaya.

1. Sentimen Umum

Mencerminkan persepsi pengguna terhadap layanan secara keseluruhan tanpa merujuk pada aspek tertentu, seperti kepuasan umum atau pengalaman penggunaan secara global.

2. Pengemudi

Berkaitan dengan kualitas dan perilaku pengemudi, termasuk keramahan, profesionalitas, keterampilan berkendara, serta interaksi dengan pelanggan yang memengaruhi loyalitas pelanggan.

3. Layanan

Mencakup proses pelayanan yang diberikan, seperti kecepatan respon, ketersediaan *driver*, ketepatan waktu, serta kualitas layanan secara operasional, berperan dalam membentuk kepuasan pelanggan.

4. Aplikasi

Berkaitan dengan performa teknis aplikasi, termasuk kemudahan penggunaan, stabilitas sistem, tampilan antarmuka, serta fitur yang tersedia dalam aplikasi yang mendukung pengalaman pengguna.

5. Harga/Biaya

Mengacu pada persepsi pengguna terhadap tarif layanan, termasuk keterjangkauan harga, transparansi biaya, serta kesesuaian antara harga dengan kualitas layanan.

Tabel 3. 1 menunjukkan pemetaan kata kunci yang digunakan sebagai *aspect lexicon* pada penelitian ini.

Tabel 3. 1 *Aspect Lexicon*

Aspek	***Keyword***	**Contoh Kalimat**
Sentimen Umum	bagus, jelek, mantap, mengecewakan, memuaskan	"Secara keseluruhan bagus."
Pengemudi	driver, pengemudi, abang, sopir, ramah, kasar, sopan, amanah, baik, profesional, marah-marah	"Driver sangat ramah."
Layanan	order, cancel, pembatalan, pelayanan, respon, susah dapat driver, lama, nunggu	"Saya harus menunggu lama karena sulit mendapatkan driver."
Aplikasi	error, bug, crash, login, update, lemot, lag, app, fitur	"Aplikasinya sering error saat login."
Harga/Biaya	mahal, murah, tarif, biaya, promo, diskon, voucher, ongkir	"Tarifnya semakin mahal dibanding sebelumnya."

Keyword pada *aspect lexicon* digunakan dalam proses pelabelan otomatis dengan cara mencocokkan kata atau frasa yang muncul pada unit teks terhadap daftar *keyword* setiap aspek. Setiap unit teks diberi satu label aspek utama berdasarkan *keyword* yang paling sesuai dengan konteks opini yang terkandung di dalamnya. Apabila satu ulasan memuat lebih dari satu aspek, ulasan tersebut terlebih dahulu dipecah menjadi beberapa unit teks, sehingga setiap unit teks merepresentasikan satu opini utama dengan satu aspek dan satu polaritas sentimen. Pendekatan ini digunakan untuk menjaga konsistensi proses evaluasi aspek dan polaritas pada tahap anotasi maupun pemodelan.

3.4.2 Penentuan Sentimen Awal

Setelah aspek teridentifikasi, sistem menentukan polaritas sentimen awal dengan mencari kata-kata bermuatan sentimen di sekitar kata kunci aspek menggunakan InSet *Lexicon*. Kata dengan polaritas positif atau negatif akan memberikan skor sentimen awal pada aspek terkait.

3.4.3 Verifikasi dan Koreksi Manual

Pada tahap ini, dilakukan verifikasi manual terhadap sebagian data hasil pelabelan otomatis untuk memastikan kesesuaian antara aspek dan polaritas sentimen. Proses ini dilakukan untuk menemukan serta meminimalkan kemungkinan kesalahan yang muncul dan tidak dapat ditangkap oleh pendekatan berbasis *Lexicon*, seperti sarkasme, ambiguitas bahasa, serta perbedaan konteks penggunaan kata dalam ulasan.

Proses verifikasi dijalankan oleh dua anotator independen yang mempunyai pemahaman terhadap konteks layanan transportasi berbasis aplikasi. Masing-masing anotator melakukan pelabelan ulang terhadap data sampel secara terpisah tanpa saling berdiskusi untuk menjaga objektivitas dan independensi penilaian.

Hasil anotasi manual tersebut digunakan sebagai acuan (*ground truth*) dalam mengevaluasi kualitas pelabelan otomatis yang dihasilkan oleh sistem. Proses ini dilakukan agar data yang digunakan dalam penelitian menjadi lebih valid, konsisten, dan sesuai dengan kebutuhan analisis.

3.4.4 Pengukuran Kesepakatan Anotator

Untuk mengukur tingkat konsistensi antar anotator, digunakan metode *Inter*-Anotator *Agreement* (IAA) dengan pendekatan Cohen's Kappa. Metode ini dipilih karena sesuai untuk mengukur tingkat kesepakatan antara dua anotator dalam proses pelabelan data kategorikal.

Perhitungan Cohen's Kappa dilakukan dengan membandingkan hasil pelabelan dari kedua anotator untuk mengetahui sejauh mana tingkat kesepakatan yang terjadi tidak disebabkan oleh faktor kebetulan. Nilai Kappa yang diperoleh kemudian digunakan sebagai indikator kualitas anotasi.

Apabila nilai kesepakatan berada pada kategori baik atau sangat baik, maka hasil anotasi dianggap memiliki konsistensi yang baik dan dapat digunakan sebagai dasar evaluasi terhadap hasil pelabelan otomatis.

Sebagai ilustrasi, jika diperoleh nilai *observed agreement* (P_o) sebesar 0,8 dan *expected agreement* (P_e) sebesar 0,428, maka nilai Cohen's Kappa dapat dihitung sebagai berikut:

$$\kappa = \frac{0{,}8 - 0{,}428}{1 - 0{,}428} \approx 0{,}65$$

Nilai tersebut menunjukkan bahwa tingkat kesepakatan antar anotator berada pada kategori baik, sehingga hasil anotasi dapat dianggap konsisten.

3.5 Pembuatan Model Klasifikasi Sentimen

Pembuatan model klasifikasi sentimen menjadi tahap utama dalam penelitian ini karena berfungsi untuk menghasilkan model pembelajaran mesin yang mampu mengelompokkan sentimen pada ulasan pengguna aplikasi Grab,

Gojek, dan Maxim. Pada tahap ini dilakukan serangkaian proses yang meliputi pembagian data, ekstraksi fitur teks, penanganan ketidakseimbangan data, pelatihan model, serta validasi kinerja model.

3.5.1 Pembagian Data

Data yang telah melalui proses pelabelan selanjutnya dibagi menjadi dua subset, yaitu data latih dan data uji, dengan perbandingan 80:20. Pembagian dilakukan dengan teknik *stratified split* supaya proporsi setiap kelas sentimen pada data latih dan data uji tetap proporsional. Data pelatihan dimanfaatkan untuk membentuk dan menguji model, sedangkan data pengujian berfungsi untuk menilai kemampuan model dalam memprediksi data baru yang tidak digunakan pada tahap pelatihan.

3.5.2 Ekstraksi Fitur TF-IDF

Setelah proses pembagian data selesai, setiap unit teks perlu diubah ke bentuk yang dapat dibaca oleh model. Oleh karena itu, ulasan direpresentasikan sebagai vektor fitur melalui pembobotan TF-IDF. Teknik ini tidak hanya melihat seberapa sering sebuah term muncul dalam satu ulasan, tetapi juga memperhitungkan penyebarannya pada seluruh korpus. Dengan cara tersebut, kata yang memiliki nilai pembeda lebih kuat terhadap isi ulasan akan memperoleh bobot lebih tinggi, sedangkan kata yang bersifat umum akan ditekan bobotnya. Representasi akhir dari proses ini berupa matriks TF-IDF yang menjadi dasar masukan bagi *Logistic Regression* dalam melakukan klasifikasi sentimen.

Parameter TF-IDF ditentukan melalui proses tuning menggunakan RandomizedSearchCV pada data latih dengan kombinasi nilai max_features, ngram_range, min_df, dan max_df yang berbeda.

3.5.3 Penanganan *Imbalance Data*

Distribusi kelas sentimen pada data ulasan umumnya tidak seimbang, dengan kelas negatif sebagai kelas dominan, sedangkan kelas netral menjadi kelas paling sedikit. Berdasarkan permasalahan yang telah dijelaskan, penelitian ini membandingkan tiga pendekatan, yaitu *Baseline Logistic Regression* tanpa penanganan khusus terhadap ketidakseimbangan data, *Logistic Regression* dengan parameter class_weight='balanced', serta kombinasi *Synthetic Minority Oversampling Technique* (SMOTE) dan *Logistic Regression*.

Parameter class_weight='balanced' secara otomatis memberikan bobot lebih besar pada kelas minoritas berdasarkan frekuensi kelas, sehingga meminimalkan bias model terhadap kelas mayoritas. Dukungan atas strategi ini dapat ditemukan pada penelitian Ivanov dkk. (2023), yang menunjukkan bahwa penggunaan parameter class_weight='balanced' bersama dengan *stratified train/test split* dapat meningkatkan akurasi dan metrik performa klasifikasi pada dataset yang tidak seimbang tanpa memodifikasi struktur model secara signifikan.

Pada skenario SMOTE, proses *oversampling* diterapkan setelah data dibagi menjadi data *training* dan data *testing*. SMOTE hanya diterapkan pada data *training* yang telah direpresentasikan ke dalam bentuk numerik menggunakan TF-IDF. Data *testing* tidak dikenai proses SMOTE agar evaluasi model tetap dilakukan pada distribusi data asli dan untuk menghindari terjadinya *data leakage*. Dengan

demikian, SMOTE digunakan untuk menyeimbangkan jumlah kelas pada data *training* sebelum model *Logistic Regression* dilatih. SMOTE diterapkan untuk menambah jumlah data pada kelas minoritas melalui pembentukan sampel sintetis, sehingga komposisi data *training* menjadi lebih proporsional. Dengan adanya kondisi ketidakseimbangan kelas pada data sentimen, penelitian ini membandingkan tiga pendekatan penanganan *imbalance* untuk mengetahui metode yang paling efektif dalam meningkatkan kinerja klasifikasi.

3.5.4 Pelatihan Model *Logistic Regression*

Model klasifikasi sentimen dibangun menggunakan algoritma *Logistic Regression* dengan pendekatan klasifikasi multikelas. Proses pelatihan model dilakukan dengan memanfaatkan data latih yang sebelumnya telah diubah ke dalam representasi numerik melalui metode TF-IDF dan disesuaikan dengan bobot kelas yang telah ditentukan. *Logistic Regression* dipilih karena memiliki efisiensi komputasi yang baik, bersifat *interpretable*, serta menunjukkan kinerja yang kompetitif dalam analisis sentimen berbasis teks.

Parameter *Logistic Regression* ditentukan melalui RandomizedSearchCV dengan menguji beberapa nilai C, solver, dan class_weight menggunakan *Stratified* 5-*Fold Cross Validation*

3.5.5 Validasi Model

Validasi model dilakukan dengan menggunakan metode *Stratified* 5-*Fold Cross Validation* pada data latih untuk menilai konsistensi performa model. Metode ini membagi data latih menjadi lima bagian dengan mempertahankan perbandingan

jumlah data pada setiap kelas sentimen, sehingga distribusi kelas tetap seimbang pada masing-masing *fold*.

Dalam setiap proses validasi, satu *fold* digunakan sebagai data validasi, sementara empat *fold* lainnya digunakan untuk melatih model. Proses tersebut dilakukan secara bergantian hingga seluruh *fold* pernah menjadi data validasi. Nilai evaluasi dari setiap iterasi kemudian dihitung rata-ratanya agar diperoleh gambaran performa model yang lebih stabil dan tidak terlalu bergantung pada satu pembagian data saja.

3.6 Evaluasi Model

Evaluasi model bertujuan untuk mengetahui kemampuan algoritma *Logistic Regression* dalam memprediksi kelas sentimen berdasarkan aspek yang telah ditentukan. Pada tahap ini, kinerja model dinilai menggunakan beberapa metrik evaluasi klasifikasi, meliputi *accuracy, Macro-precision*, *Macro-recall*, *Macro*-F1, *weighted*-F1, *classification report*, dan *Confusion Matrix*.

Accuracy digunakan untuk mengetahui tingkat ketepatan prediksi model secara keseluruhan. *Macro-precision*, *Macro-recall*, dan *Macro*-F1 digunakan untuk menilai performa model pada setiap kelas dengan memberikan bobot yang sama terhadap seluruh kelas sentimen. *Weighted*-F1 digunakan sebagai metrik pendukung untuk melihat performa model dengan mempertimbangkan jumlah data pada setiap kelas. Selain itu, *classification report* digunakan untuk menampilkan nilai *precision*, *recall*, F1-*score*, dan *support* pada masing-masing kelas, sedangkan *Confusion Matrix* digunakan untuk melihat pola kesalahan prediksi model terhadap kelas negatif, netral, dan positif.

Dalam penelitian ini, *Macro*-F1 digunakan sebagai metrik utama karena distribusi kelas sentimen tidak seimbang. Pada kondisi data tidak seimbang, *accuracy* dapat terlihat tinggi apabila model lebih banyak benar dalam memprediksi kelas mayoritas, tetapi belum tentu mampu mengenali kelas minoritas dengan baik. *Macro*-F1 lebih sesuai digunakan karena menghitung rata-rata F1-*score* setiap kelas dengan bobot yang sama, sehingga performa model pada kelas negatif, netral, dan positif dapat dievaluasi secara lebih seimbang.

3.7 Analisis Hasil dan Visualisasi

Tahap analisis hasil dilakukan setelah model selesai dievaluasi. Pada tahap ini, keluaran klasifikasi sentimen digunakan untuk melihat pola persebaran sentimen pada aspek layanan yang terdapat dalam ulasan pengguna aplikasi Grab, Gojek, dan Maxim. Analisis ini bertujuan untuk mengetahui aspek mana yang lebih banyak memperoleh tanggapan positif, negatif, maupun netral dari pengguna.

Melalui analisis sentimen berbasis aspek, hasil penelitian dapat memberikan gambaran yang lebih spesifik mengenai persepsi pelanggan terhadap setiap aspek layanan pada ketiga aplikasi *ride-hailing*. Selanjutnya, hasil analisis disajikan dalam bentuk tabel dan grafik agar pola sentimen lebih mudah dipahami serta dapat mendukung proses interpretasi hasil penelitian.