

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi digital di berbagai sektor, termasuk sektor keuangan. Transformasi tersebut melahirkan berbagai layanan *Financial Technology (FinTech)* yang memberikan kemudahan dalam melakukan transaksi keuangan secara cepat, praktis, dan efisien. Di Indonesia, perkembangan layanan *FinTech* didukung oleh meningkatnya jumlah pengguna internet serta bertambahnya penyelenggara layanan keuangan digital yang berizin. Hingga Februari 2026, terdapat 95 penyelenggara LPBBTI/*FinTech* P2PL yang telah memperoleh izin dari Otoritas Jasa Keuangan (OJK), sementara jumlah pengguna internet Indonesia mencapai sekitar 235 juta jiwa atau 81,72% dari total populasi. Kondisi tersebut menunjukkan bahwa masyarakat semakin terbiasa memanfaatkan layanan digital, termasuk layanan keuangan berbasis aplikasi (APJII, 2026; OJK, 2026).

Salah satu layanan *FinTech* yang mengalami perkembangan pesat adalah *Buy Now Pay Later (BNPL)* atau layanan pembayaran tunda. Layanan ini memungkinkan masyarakat melakukan pembelian terlebih dahulu dan membayarnya di kemudian hari melalui berbagai pilihan tenor cicilan. Kemudahan proses pendaftaran, persyaratan yang relatif sederhana, serta proses persetujuan yang cepat menjadikan layanan BNPL semakin diminati oleh masyarakat Indonesia. Perkembangan tersebut menunjukkan bahwa layanan keuangan digital telah menjadi bagian dari aktivitas transaksi masyarakat sehari-hari.

Salah satu penyedia layanan BNPL yang banyak digunakan di Indonesia adalah Kredivo. Aplikasi Kredivo menyediakan berbagai layanan, seperti pembayaran cicilan, pinjaman tunai, pembayaran tagihan, hingga pembiayaan transaksi pada berbagai platform *e-commerce*. Berdasarkan data Google Play Store, aplikasi Kredivo telah diunduh lebih dari 50 juta kali, memiliki lebih dari 4,77 juta ulasan, serta memperoleh *rating* rata-rata 4,8 dari 5. Jumlah pengguna dan ulasan

yang sangat besar tersebut menunjukkan tingginya tingkat penggunaan aplikasi sekaligus menghasilkan data opini pengguna yang dapat dimanfaatkan sebagai sumber informasi mengenai kualitas layanan (Kredivo, 2026).

Ulasan pengguna pada Google Play Store merupakan salah satu bentuk umpan balik (*feedback*) yang berisi pengalaman, kritik, saran, maupun apresiasi terhadap layanan yang diberikan. Informasi tersebut sangat penting bagi perusahaan untuk mengetahui tingkat kepuasan pelanggan serta mengidentifikasi berbagai permasalahan yang masih dirasakan pengguna, seperti kendala pada proses verifikasi akun, keterlambatan pencairan dana, gangguan sistem, maupun pelayanan pelanggan. Di sisi lain, perusahaan juga dapat mengetahui aspek layanan yang telah memberikan pengalaman positif bagi pengguna, seperti kemudahan penggunaan aplikasi, kecepatan transaksi, dan fleksibilitas pembayaran.

Meskipun demikian, jumlah ulasan yang sangat besar menyebabkan proses analisis secara manual menjadi tidak efisien, membutuhkan waktu yang lama, serta rentan terhadap subjektivitas. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data teks secara otomatis sehingga informasi yang terkandung di dalam ulasan dapat dimanfaatkan secara optimal. Salah satu metode yang banyak digunakan untuk mengatasi permasalahan tersebut adalah analisis sentimen (*sentiment analysis*).

Analisis sentimen merupakan salah satu cabang *Natural Language Processing* (NLP) dan *text mining* yang bertujuan mengidentifikasi kecenderungan opini pengguna ke dalam kategori sentimen positif maupun negatif. Dengan menerapkan analisis sentimen pada ulasan aplikasi Kredivo, perusahaan dapat mengetahui tingkat kepuasan pengguna, mengidentifikasi keluhan yang paling sering muncul, serta memperoleh informasi yang dapat dijadikan dasar dalam meningkatkan kualitas layanan maupun pengembangan aplikasi.

Berbagai algoritma telah digunakan dalam penelitian analisis sentimen, di antaranya *Naive Bayes* dan *Logistic Regression*. Algoritma *Naive Bayes* merupakan metode klasifikasi berbasis probabilitas yang memiliki keunggulan berupa proses komputasi yang sederhana, cepat, dan mampu memberikan performa yang baik pada data teks berdimensi tinggi. Sementara itu, *Logistic Regression* merupakan

algoritma klasifikasi linear yang mampu mempelajari hubungan antara fitur-fitur teks dengan kelas sentimen sehingga memiliki kemampuan generalisasi yang baik terhadap data baru.

Beberapa penelitian terdahulu telah membandingkan performa algoritma *Naive Bayes* dan *Logistic Regression* pada berbagai domain. Penelitian pada ulasan aplikasi *mobile banking* menggunakan 200.000 ulasan Google Play Store menunjukkan bahwa *Logistic Regression* memperoleh akurasi sebesar 92–93%, lebih tinggi dibandingkan *Multinomial Naïve Bayes* yang hanya mencapai 71–74% (Aryanusa et al., 2025). Penelitian lain pada ulasan produk *skincare* di Tokopedia yang menggunakan pembobotan TF-IDF juga menunjukkan bahwa *Logistic Regression* memiliki akurasi sebesar 92,3%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 75,9% (Dikusuma & Widjaja, 2025). Selain itu, penelitian pada ulasan game *Free Fire* yang menerapkan TF-IDF dan penanganan ketidakseimbangan data menunjukkan bahwa *Logistic Regression* menghasilkan akurasi sebesar 81,63% dengan *F1-score* sebesar 87,91%, lebih baik dibandingkan *Naïve Bayes* (Bachtiar et al., 2026). Hasil penelitian-penelitian tersebut menunjukkan bahwa *Logistic Regression* cenderung memberikan performa klasifikasi yang lebih baik dibandingkan *Naive Bayes* pada berbagai jenis data ulasan.

Meskipun demikian, berdasarkan penelitian-penelitian terdahulu masih terdapat beberapa celah penelitian (*research gap*). Sebagian besar penelitian dilakukan pada domain *mobile banking*, *e-commerce*, maupun permainan daring, sedangkan penelitian yang secara khusus membahas ulasan pengguna aplikasi Kredivo sebagai layanan *Buy Now Pay Later* (BNPL) masih relatif terbatas. Selain itu, sebagian besar penelitian menggunakan pelabelan sentimen berbasis *lexicon* atau pelabelan biner sederhana, sehingga belum banyak yang menerapkan pelabelan berbasis *rating* dengan verifikasi manual pada ulasan *rating* 3. Beberapa penelitian juga belum mengombinasikan tahapan *preprocessing* yang lengkap, ekstraksi fitur menggunakan TF-IDF, pembagian data *training* dan *testing*, validasi menggunakan *K-Fold Cross Validation*, serta penanganan ketidakseimbangan data menggunakan SMOTE dalam satu alur penelitian. Oleh karena itu, masih

diperlukan penelitian yang mampu memberikan evaluasi yang lebih komprehensif terhadap performa algoritma klasifikasi pada ulasan pengguna aplikasi Kredivo.

Dalam penelitian ini, data ulasan pengguna aplikasi Kredivo diperoleh dari Google Play Store selama periode Januari 2021 hingga Januari 2026. Data kemudian melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi kata menggunakan *Indo Normalizer*, *stopword removal*, dan *stemming* menggunakan *Sastrawi*. Setelah proses *preprocessing*, dilakukan pelabelan sentimen berbasis *rating*, yaitu *rating* 1–2 diberi label sentimen negatif, *rating* 4–5 diberi label sentimen positif, sedangkan *rating* 3 diverifikasi dan diberi label secara manual berdasarkan isi ulasan. Selanjutnya dilakukan ekstraksi fitur menggunakan TF-IDF, pembagian data menjadi data *training* dan *testing*, validasi model menggunakan *K-Fold Cross Validation* pada data *training*, penanganan ketidakseimbangan data menggunakan SMOTE pada setiap *fold*, serta proses klasifikasi menggunakan algoritma *Naive Bayes* dan *Logistic Regression*. Kinerja kedua algoritma kemudian dibandingkan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Kebaruan (*novelty*) penelitian ini terletak pada penerapan alur analisis sentimen yang lebih komprehensif terhadap ulasan pengguna aplikasi Kredivo. Penelitian ini tidak hanya membandingkan performa algoritma *Naive Bayes* dan *Logistic Regression*, tetapi juga menerapkan pelabelan sentimen berbasis *rating* dengan verifikasi manual pada ulasan *rating* 3, penggunaan *Indo Normalizer* dalam proses normalisasi, validasi model menggunakan *K-Fold Cross Validation*, penanganan ketidakseimbangan data menggunakan SMOTE pada setiap *fold* data *training*, serta visualisasi *word cloud* untuk mengidentifikasi kata-kata dominan pada sentimen positif dan negatif. Pendekatan tersebut diharapkan mampu menghasilkan model klasifikasi yang lebih akurat sekaligus memberikan interpretasi yang lebih mendalam mengenai kepuasan dan keluhan pengguna aplikasi Kredivo.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna aplikasi Kredivo menggunakan algoritma *Naive Bayes* dan *Logistic Regression*. Penelitian ini bertujuan mengidentifikasi sentimen positif

dan sentimen negatif pengguna, memvisualisasikan kata-kata dominan pada masing-masing kategori sentimen menggunakan *word cloud*, serta membandingkan performa kedua algoritma berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian diharapkan dapat menjadi bahan evaluasi bagi pengembang aplikasi Kredivo dalam meningkatkan kualitas layanan dan pengalaman pengguna. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya di bidang analisis sentimen pada layanan *Financial Technology* (*FinTech*) menggunakan pendekatan *machine learning*.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil analisis sentimen terhadap ulasan pengguna aplikasi Kredivo pada Google Play Store?
2. Bagaimana visualisasi *word cloud* dapat menggambarkan kata dominan pada sentimen positif dan negatif dalam ulasan pengguna aplikasi Kredivo?
3. Bagaimana perbandingan performa metode *Naive Bayes* dan *Logistic Regression* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Kredivo berdasarkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*?

1.3 Batasan Masalah

Agar pembahasan dalam penelitian ini lebih terarah dan tidak menyimpang dari tujuan utama, penulis menetapkan batasan-batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi Kredivo pada Google Play Store.
2. Data ulasan yang digunakan dalam rentang waktu Januari 2021 sampai Januari 2026.
3. Penelitian ini hanya mengklasifikasikan sentimen ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif.
4. Proses pelabelan sentimen dilakukan berdasarkan nilai rating yang diberikan pengguna pada Google Play Store.

5. Tahapan *preprocessing* data yang digunakan meliputi *cleaning*, *case folding*, *tokenizing*, *normalisasi kata*, *stopword removal*, dan *stemming*.
6. Metode pembobotan kata yang digunakan dalam penelitian ini ialah *Term Frequency-Inverse Document Frequency* (TF-IDF).
7. *Splitting* Data menggunakan 80:20.
8. Metode klasifikasi yang digunakan dalam penelitian ini hanya terbatas pada *Naive Bayes* dan *Logistic Regression*.
9. Validasi data menggunakan metode *K-Fold Cross Validation*.
10. Penanganan ketidakseimbangan data dilakukan menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE).
11. Evaluasi performa model dilakukan menggunakan *confusion matrix* berdasarkan nilai akurasi, presisi, *recall*, dan *f1-score*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis sentimen ulasan pengguna aplikasi Kredivo pada Google Play Store untuk mengidentifikasi sentimen positif dan sentimen negatif pengguna.
2. Memvisualisasikan kata-kata dominan pada sentimen positif dan sentimen negatif menggunakan *word cloud* sebagai bentuk interpretasi hasil analisis sentimen.
3. Membandingkan kinerja algoritma *Naive Bayes* dan *Logistic Regression* serta menentukan algoritma dengan performa terbaik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4.2 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat, baik secara teoretis maupun praktis, sebagai berikut:

1. Manfaat Teoritis

Secara teoretis, penelitian ini diharapkan dapat menambah wawasan dan referensi ilmiah dalam bidang *text mining* dan *Natural Language Processing*

(NLP), analisis sentimen, dan penerapan *machine learning* pada data ulasan aplikasi digital. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis sentimen pada aplikasi fintech menggunakan metode *Naive Bayes* dan *Logistic Regression*. Penelitian ini juga diharapkan dapat memberikan kontribusi terhadap pengembangan penelitian pengolahan data teks berbahasa Indonesia, khususnya dalam klasifikasi sentimen ulasan pengguna aplikasi.

2. Manfaat Praktis

Secara praktis, penelitian ini diharapkan dapat memberikan manfaat bagi pengembang aplikasi Kredivo sebagai bahan evaluasi sehingga dapat digunakan dalam meningkatkan kualitas layanan aplikasi. Bagi pengguna aplikasi Kredivo, penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi pengguna lain terhadap kualitas layanan aplikasi berdasarkan hasil analisis sentimen. Selain itu, penelitian ini juga diharapkan dapat menambah pengetahuan dan pengalaman peneliti dalam menerapkan analisis sentimen menggunakan metode *Naive Bayes* dan *Logistic Regression*.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai analisis sentimen pada ulasan aplikasi digital telah banyak dilakukan dengan menggunakan berbagai metode *machine learning*. Penelitian-penelitian tersebut umumnya bertujuan untuk mengetahui opini pengguna terhadap suatu aplikasi berdasarkan ulasan yang diberikan pada platform digital, seperti Google Play Store. Selain itu, penelitian terdahulu juga menjadi dasar dalam menentukan metode yang sesuai untuk digunakan pada penelitian ini.

Penelitian yang dilakukan oleh Aryanusa et al. (2025) bertujuan untuk menganalisis sentimen pengguna terhadap fitur-fitur aplikasi *mobile banking* dengan membandingkan performa algoritma *Logistic Regression* dan *Multinomial Naïve Bayes*. Data penelitian diperoleh dari Google Play Store yang terdiri atas 200.000 ulasan dari empat aplikasi *mobile banking*, yaitu blu by BCA, BRImo, BNI Mobile Banking, dan Livin' by Mandiri. Tahapan penelitian meliputi *preprocessing* yang terdiri dari *cleaning*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*, kemudian dilakukan pelabelan sentimen menggunakan metode *lexicon-based*, klasifikasi, serta evaluasi model. Hasil penelitian menunjukkan bahwa algoritma *Logistic Regression* memperoleh akurasi sebesar 92–93%, sedangkan *Multinomial Naïve Bayes* memperoleh akurasi 71–74%. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan karena sama-sama membandingkan algoritma *Naïve Bayes* dan *Logistic Regression*, namun berbeda pada objek penelitian serta metode pelabelan sentimen yang menggunakan pendekatan *lexicon-based* (Aryanusa et al., 2025).

Penelitian yang dilakukan oleh Dikusuma & Widjaja (2025) membandingkan performa algoritma *Naïve Bayes* dan *Logistic Regression* dalam analisis sentimen ulasan produk *skincare* di Tokopedia. Dataset yang digunakan berjumlah 973 ulasan yang diproses melalui tahapan *text preprocessing* dan ekstraksi fitur menggunakan TF-IDF. Evaluasi model dilakukan menggunakan metrik *accuracy*,

precision, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memperoleh akurasi sebesar 75,9%, sedangkan *Logistic Regression* memperoleh akurasi sebesar 92,3%. Penelitian ini menunjukkan bahwa *Logistic Regression* lebih efektif dalam mengklasifikasikan sentimen pada data ulasan *e-commerce*. Kesamaan penelitian ini dengan penelitian yang dilakukan terletak pada penggunaan algoritma *Naïve Bayes*, *Logistic Regression*, serta pembobotan TF-IDF, sedangkan perbedaannya terdapat pada objek penelitian yang digunakan (Dikusuma & Widjaja, 2025).

Penelitian yang dilakukan oleh Bachtiar et al. (2026) menganalisis sentimen ulasan pengguna aplikasi Free Fire di Google Play Store menggunakan algoritma *Naïve Bayes* dan *Logistic Regression*. Dataset terdiri atas 12.574 ulasan yang dikumpulkan melalui *web scraping* dengan pelabelan sentimen berdasarkan rating bintang. Tahapan penelitian meliputi *case folding*, normalisasi, *stopword removal*, *stemming*, ekstraksi fitur menggunakan TF-IDF *n-gram*, serta optimasi parameter menggunakan *GridSearchCV*. Hasil penelitian menunjukkan bahwa algoritma *Logistic Regression* memperoleh akurasi sebesar 81,63% dengan *F1-score* sebesar 87,91%, sedangkan *Naïve Bayes* memiliki nilai *recall* yang lebih tinggi namun akurasi lebih rendah. Penelitian ini memiliki kemiripan dengan penelitian yang dilakukan karena sama-sama menggunakan pelabelan berbasis rating dan membandingkan algoritma *Naïve Bayes* dengan *Logistic Regression* (Bachtiar et al., 2026).

Penelitian yang dilakukan oleh Sahrul & Afyati (2025) bertujuan menganalisis sentimen pengguna aplikasi Indodana di Google Play Store menggunakan algoritma *Logistic Regression*, *Naïve Bayes*, dan *Support Vector Machine* (SVM). Penelitian ini bertujuan mengidentifikasi sentimen positif dan negatif pengguna terhadap layanan *paylater* dan pinjaman daring sekaligus membandingkan performa ketiga algoritma. Hasil penelitian menunjukkan bahwa algoritma SVM memperoleh akurasi tertinggi sebesar 92,63%, diikuti oleh *Logistic Regression* sebesar 92,01%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 80,91%. Penelitian ini relevan dengan penelitian yang dilakukan karena sama-sama menggunakan objek berupa aplikasi *fintech*, namun penelitian ini hanya

memfokuskan perbandingan pada algoritma *Naïve Bayes* dan *Logistic Regression* (Sahrul & Afiyati, 2025).

Penelitian yang dilakukan oleh Ramadhani & Suryono (2024) menganalisis sentimen masyarakat terhadap teknologi *metaverse* menggunakan algoritma *Naïve Bayes* dan *Logistic Regression*. Dataset terdiri atas 6.728 komentar yang diperoleh dari media sosial X. Untuk mengatasi ketidakseimbangan data, penelitian ini menerapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). Hasil penelitian menunjukkan bahwa setelah dilakukan penyeimbangan data menggunakan *SMOTE*, algoritma *Logistic Regression* memperoleh akurasi sebesar 95%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 91%. Penelitian ini menunjukkan bahwa penerapan *SMOTE* mampu meningkatkan performa model klasifikasi sehingga menjadi salah satu referensi dalam penelitian ini yang juga menerapkan metode *SMOTE* (Ramadhani & Suryono, 2024).

Penelitian yang dilakukan oleh Monica & Purwanto (2026) menganalisis sentimen komentar YouTube berbahasa Indonesia menggunakan pendekatan *machine learning*. Tahapan penelitian meliputi *preprocessing*, pembobotan TF-IDF, optimasi parameter, dan penanganan ketidakseimbangan data menggunakan *SMOTE*. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi tertinggi sebesar 84,88%. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan karena sama-sama menggunakan tahapan *preprocessing*, TF-IDF, dan *SMOTE*, namun berbeda pada sumber data dan algoritma klasifikasi yang digunakan (Monica & Purwanto, 2026).

Penelitian yang dilakukan oleh Sari et al. (2026) melakukan analisis sentimen terhadap ulasan pengguna aplikasi Kredivo menggunakan algoritma *Multinomial Naïve Bayes* dengan pembobotan TF-IDF. Data diperoleh melalui *web scraping* dan diproses menggunakan tahapan *preprocessing* yang meliputi *cleaning*, *tokenizing*, *stopword removal*, *stemming*, serta pelabelan sentimen secara manual. Hasil penelitian menunjukkan bahwa model memperoleh akurasi sebesar 84,82%. Selain itu, hasil analisis menunjukkan bahwa sentimen positif didominasi oleh kemudahan penggunaan aplikasi dan kecepatan verifikasi, sedangkan sentimen negatif

berkaitan dengan biaya layanan, proses verifikasi, dan gangguan teknis aplikasi. Penelitian ini memiliki kesamaan pada objek penelitian, namun hanya menggunakan satu algoritma klasifikasi sehingga belum membandingkan performa dengan algoritma lain (A. Sari et al., 2026).

Penelitian yang dilakukan oleh Kusuma & Ali (2024) menganalisis sentimen pengguna aplikasi DANA menggunakan algoritma *Naïve Bayes*. Data diperoleh melalui proses *web scraping* dan dianalisis menggunakan metode *Knowledge Discovery in Database* (KDD). Hasil penelitian menunjukkan bahwa model memperoleh akurasi sebesar 74,60% dengan distribusi sentimen terdiri atas 576 ulasan positif, 205 ulasan netral, dan 219 ulasan negatif. Penelitian ini relevan karena menggunakan algoritma *Naïve Bayes* pada aplikasi *fintech*, namun belum melakukan perbandingan dengan algoritma klasifikasi lainnya (Kusuma & Ali, 2024).

Penelitian yang dilakukan oleh Prasetyo et al. (2024) membandingkan algoritma *Logistic Regression*, *Support Vector Machine* (SVM), dan *Random Forest* dalam analisis sentimen terhadap 3.000 ulasan pengguna GoPay di Google Play Store. Tahapan penelitian meliputi *preprocessing*, pelabelan menggunakan metode *lexicon-based*, serta evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *Logistic Regression* memperoleh akurasi tertinggi sebesar 88,16%, diikuti oleh SVM sebesar 87,50%, sedangkan *Random Forest* memperoleh akurasi sebesar 79,33%. Penelitian ini menjadi referensi dalam penggunaan algoritma *Logistic Regression* pada analisis sentimen aplikasi *fintech* (Prasetyo et al., 2024).

Penelitian yang dilakukan oleh Susanti et al. (2026) menganalisis sentimen pengguna aplikasi Kredivo menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) dengan kerangka kerja SEMMA. Penelitian menggunakan dua skenario pembagian data, yaitu 80:20 dan 70:30. Hasil penelitian menunjukkan bahwa pada pembagian data 80:20, algoritma SVM memperoleh akurasi sebesar 97,05%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 92,01%. Pada pembagian data 70:30, SVM tetap menunjukkan performa terbaik dengan akurasi sebesar 96,76%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 91,48%.

Penelitian ini memiliki objek yang sama dengan penelitian yang dilakukan, yaitu aplikasi Kredivo, namun menggunakan algoritma pembandingan yang berbeda. Penelitian ini menggunakan *Logistic Regression* sebagai algoritma pembandingan terhadap *Naïve Bayes* sehingga diharapkan dapat memberikan alternatif model klasifikasi yang efektif untuk analisis sentimen ulasan pengguna aplikasi Kredivo (Susanti et al., 2026).

2.2 Landasan Teori

2.2.1 Financial Technology

Financial Technology (Fintech) adalah penggabungan antara sistem keuangan dan teknologi. Perkembangan fintech yang hadir di Indonesia memunculkan berbagai inovasi aplikasi khususnya dalam layanan keuangan seperti sebagai alat transaksi pembayaran, alat penyimpanan uang, dan juga alat peminjaman uang (Oktaviani et al., 2025).

2.2.2 Kredivo

Kredivo adalah aplikasi kredit instan yang memberikan kemudahan untuk beli sekarang dan bayar nanti. Aplikasi ini juga dikenal dengan kartu kredit digital dimana peminjam dapat menggunakannya untuk belanja cicilan *online*. Selain digunakan untuk keperluan berbelanja, peminjam juga dapat mencairkan dalam bentuk uang tunai. Namun, tidak semua penggunanya merasa puas dengan pelayanan tersebut, terdapat banyak komentar yang salah satunya disampaikan melalui fitur ulasan aplikasi kredivo pada *google play store* (Agustin et al., 2023).

2.2.3 Google Play Store

Google Play Store adalah sebuah platform dari *Google* yang sangat populer untuk mengunduh aplikasi dimana di dalamnya terdapat fitur pengguna dapat memberikan ulasan terhadap aplikasi yang telah digunakannya (Astuti, 2024).

2.2.4 Ulasan

Ulasan atau *review* adalah evaluasi atau penilaian terhadap suatu produk, layanan karya atau pengalaman yang bertujuan memberikan informasi objektif dan

bermanfaat kepada calon konsumen atau pengguna lain. Dalam konteks jual beli *online*, ulasan pelanggan daring merupakan salah satu fitur yang ada pada *e-commerce* dan *marketplace* yang bertujuan untuk menjadi bahan pertimbangan bagi calon konsumen yang ingin membeli sesuatu. Ulasan ini dapat berupa positif, negatif, kritik dan saran dari pembeli yang telah mencoba produk atau layanan yang mereka beli (Suwendi et al., 2025).

2.2.5 Analisis Sentimen

Analisis sentimen atau bisa disebut juga *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif (Lukmana et al., 2025).

2.2.6 Text Mining

Text mining merupakan teknik dalam pengambilan informasi dari sejumlah data tak terstruktur dari sebuah topik tertentu yang memiliki kualitas tinggi serta dapat diperoleh data-data permasalahan dalam teks (Afdal & Elita, 2022).

2.2.7 Natural Language Processing

Natural Language Processing (NLP) merupakan pendekatan inovatif yang memungkinkan komputer untuk memahami dan menganalisis bahasa manusia secara otomatis, membuka peluang untuk mendapatkan wawasan yang lebih dalam terkait dengan persepsi dan respons pengguna terhadap aplikasi tersebut (Nurwanda et al., 2024).

2.2.8 Preprocessing

Preprocessing data merupakan tahapan penting dalam analisis sentimen yang dilakukan untuk membersihkan dan mempersiapkan data teks sebelum masuk ke proses klasifikasi. *Preprocessing* adalah proses pengolahan data awal yang digunakan untuk mengubah data mentah yang diperoleh dari berbagai sumber menjadi informasi yang lebih bersih dan dapat digunakan untuk analisis lebih lanjut (Rahman Salam et al., 2023).

Pada penelitian ini, data ulasan pengguna aplikasi Kredivo memiliki karakteristik bahasa yang tidak terstruktur karena ditulis secara bebas oleh

pengguna. Kondisi tersebut dapat memengaruhi kualitas proses klasifikasi apabila data langsung digunakan tanpa dilakukan pengolahan terlebih dahulu. Oleh karena itu, tahap *preprocessing* dilakukan untuk meningkatkan kualitas data teks agar hasil analisis sentimen menjadi lebih optimal dan akurat. Tahapan *preprocessing* yang digunakan dalam penelitian ini meliputi:

1. Cleaning: Proses pembersihan data bertujuan untuk menemukan dan memperbaiki atau menghapus data yang tidak valid atau tidak berguna dari suatu kumpulan data (Rahman Salam et al., 2023).
2. Case folding: Case folding proses pada text preprocessing yang dilakukan untuk mengubah semua huruf pada teks ke dalam huruf kecil atau huruf besar yang sama, sehingga memudahkan dalam proses perbandingan dan analisis teks (Rahman Salam et al., 2023).
3. Tokenizing: Merupakan suatu tahapan untuk memecah suatu kumpulan teks menjadi sebuah kata. Analisis kata-kata dalam teks menjadi penting karena dengan melakukan tokenisasi, makna dari teks dapat dengan mudah ditentukan (Rahman Salam et al., 2023).
4. Normalisasi: Teknik normalisasi yaitu mengubah data menjadi bentuk yang lebih standar. Normalisasi sering digunakan untuk mengubah teks menjadi bentuk yang lebih standar sebelum dianalisis (Rahman Salam et al., 2023).
5. Stopword removal: Proses melakukan reduksi kata atau token ke dasar bentuk kata (Rahman Salam et al., 2023).
6. Stemming: Kata yang tidak memiliki makna atau tidak efektif akan dihilangkan dengan menggunakan stopwords (Rahman Salam et al., 2023).

2.2.9 Pelabelan Data

Pelabelan dilakukan dengan menggunakan pendekatan berbasis rating score, yaitu metode otomatis yang memanfaatkan nilai ulasan atau penilaian yang diberikan pengguna. Data yang terkumpul akan dikategorikan menjadi sentimen positif dan sentimen negatif. Ulasan dengan rating tinggi, yaitu 4 dan 5 akan dilabeli dengan label positif, sedangkan ulasan dengan rating rendah, yaitu 1 dan 2 akan dilabeli dengan label negatif (Silitonga et al., 2025). Ulasan dengan rating 3 atau netral akan dilakukan verifikasi manual berdasar isi ulasan.

2.2.10 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) adalah suatu hubungan kata (*term*) yang berada pada dokumen yang akan diberikan suatu nilai bobot. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu yang disebut *term frequency (TF)* dan *inverse* frekuensi dokumen yang mengandung kata yang disebut *inverse document frequency (IDF)*. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut akan rendah pada kumpulan dokumen (H. Sari et al., 2021).

2.2.11 Splitting Data

Splitting Data adalah proses pembagian *dataset* menjadi data pelatihan (*training data*) dan data pengujian (*testing data*). Data pelatihan digunakan untuk melatih model agar dapat mempelajari pola dan karakteristik data, sedangkan data uji digunakan untuk menilai performa model pada data baru. Dalam penelitian ini menggunakan rasio 80:20. Rasio ini dipilih karena memberikan keseimbangan optimal antara jumlah data untuk melatih dan menguji model secara akurat. Hal ini menunjukkan bahwa rasio 80:20 efektif dalam menjaga keseimbangan antara data pelatihan dan pengujian, memungkinkan evaluasi model yang baik pada data yang belum pernah dilihat sebelumnya (Pranata et al., 2024).

2.2.12 K-Fold Cross Validation

K-Fold Cross Validation merupakan cara statistik yang digunakan dalam evaluasi performan dari model yang akan dibangun. Selain itu *k-fold cross validation* adalah cara yang efektif dalam menggabungkan atribut dan *setting parameter machine learning* dalam melatih model *prediction* yang lebih baik. Dataset akan dibagi menggunakan *5-fold* dengan ukuran yang sama pada setiap *fold*-nya, sehingga mempunyai 5 subset data. Kelebihan menggunakan *K-Fold Cross Validation* adalah untuk meningkatkan estimasi kinerja yang lebih stabil dan akurat dengan cara menghitung rata rata hasil dari “k” iterasi dan mencapai resiko

empiris yang lebih rendah, daripada hanya membaginya ke dalam data *training* dan data *testing* saja. Kelebihan lain menggunakan *K-Fold Cross Validation* dapat meminimalisasi resiko *overfitting* dengan cara melatih dan menguji model dengan berbagai kombinasi data (Wijiyanto et al., 2024).

2.2.13 SMOTE (Synthetic Minority Oversampling Technique)

Imbalanced data merupakan kondisi ketika jumlah data pada setiap kelas dalam dataset tidak seimbang. Pada analisis sentimen, kondisi ini sering terjadi karena jumlah data sentimen positif dan negatif memiliki perbedaan yang cukup besar. Ketidakseimbangan data dapat memengaruhi performa model klasifikasi karena model cenderung lebih dominan memprediksi kelas mayoritas dibandingkan kelas minoritas. *SMOTE (Synthetic Minority Oversampling Technique)* merupakan metode teknik *oversampling* yang digunakan untuk menangani masalah ketidakseimbangan data dengan pengolahan yang berbeda dari metode *oversampling* sebelumnya. *SMOTE* dapat mengatasi masalah *overfitting*, tidak menyebabkan informasi hilang, dapat meningkatkan akurasi dan prediksi kelas. Proses *SMOTE* dilakukan dengan cara menentukan kelas minoritas pada data, kemudian menentukan jarak k tetangga terdekat yang diperoleh dengan menghitung jarak *Euclidean* antar data minoritas (Anjani et al., 2023).

2.2.14 Naive Bayes

Naive Bayes adalah teknik peramalan *probabilistik* sederhana sesuai dengan pelaksanaan *teorema Bayes* (hukum *bayes*) menggunakan perkiraan independensi yang kuat. Kelebihan dari pada metode ini ialah hanya memerlukan sedikit data latih dengan tujuan memastikan parameter yang diharapkan pada langkah klasifikasi (Pebdika et al., 2023). Rumus *teorema bayes* adalah sebagai berikut:

$$P(C | X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Keterangan:

$P(C|X)$: Probabilitas kemunculan kelas C dengan syarat bukti X.

$P(X|C)$: Probabilitas bukti X jika kelas C terjadi.

$P(C)$: Probabilitas awal dari kelas C tanpa melihat bukti.

$P(X)$: Probabilitas bukti X secara keseluruhan.

2.2.15 Logistic Regression

Logistic Regression Logistic Regression merupakan jenis regresi analisis yang menjelaskan hubungan antara variabel dependen dan variabel independen. Ini menjelaskan hubungan antara satu atau lebih variabel bebas dan variabel terikat dari jenis kategori tertentu, yang dapat berkisar antara 0 dan 1 dan benar atau salah, besar atau kecil. Ini adalah alasan mengapa Logistic Regression berbeda dari multiple regression atau linear regression (Putri et al., 2025). Persamaan fungsi sigmoid pada *Logistic Regression* dapat dinyatakan sebagai berikut:

$$P(y = 1 | x) = \frac{1}{1+e^{-z}} \quad (2)$$

Dengan nilai:

$$z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (3)$$

Keterangan:

$P(y = 1 | x)$: probabilitas data termasuk ke dalam kelas positif.

e : bilangan eksponensial.

z : kombinasi linear dari fitur masukan.

w : bobot parameter model.

x : fitur data hasil pembobotan TF-IDF.

2.2.16 Confusion Matrix

Confusion matrix adalah tabel yang digunakan untuk mengevaluasi kinerja suatu model klasifikasi, dengan membandingkan hasil prediksi model dengan nilai sebenarnya dalam dataset pengujian (Hakim et al., 2025). Visualisasi *Confusion matrix* dapat dilihat pada Gambar 2.1.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. 1 Confusion Matrix

- 1) Akurasi (*Accuracy*): Akurasi adalah proporsi total prediksi yang benar (baik positif maupun negatif) dibandingkan dengan total jumlah data (Hakim et al., 2025). Secara matematis, akurasi dihitung dengan rumus:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

- 2) Presisi (*Precision*): Presisi adalah proporsi dari prediksi positif yang benar terhadap total jumlah prediksi positif. Ini mengukur seberapa banyak dari hasil yang diprediksi sebagai positif yang sebenarnya benar positif (Hakim et al., 2025). Secara matematis, presisi dihitung dengan rumus:

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

- 3) Recall: *Recall* adalah proporsi dari kelas positif yang diprediksi dengan benar dibandingkan dengan total jumlah kelas positif yang sebenarnya. Ini mengukur seberapa baik model mengidentifikasi semua kasus positif (Hakim et al., 2025). Secara matematis, recall dihitung dengan rumus:

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

- 4) *F1-Score*: *F-1 score* adalah rata-rata harmonik dari presisi dan *recall*. Ini memberikan keseimbangan antara kedua metrik ini. *F-1 score* berguna ketika kelas memiliki distribusi yang tidak seimbang (Hakim et al., 2025). Secara matematis, F-1 score dihitung dengan rumus:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Dimana:

True Positive (TP) : data positif yang berhasil diprediksi positif oleh model.

True Negative (TN) : data negatif yang berhasil diprediksi negatif oleh model.

False Positive (FP) : data negatif yang diprediksi positif oleh model.

False Negative (FN) : data positif yang diprediksi negatif oleh model.

2.2.17 *Indo-Normalizer*

Indo-Normalizer sebuah pustaka *Python* sumber terbuka yang dirancang untuk normalisasi teks bahasa Indonesia informal. Pustaka ini dibangun berdasarkan pendekatan berbasis aturan untuk menormalisasi pola informal umum: Huruf ganda, kata-kata populer, slang internet, singkatan, dan kesalahan ketik diubah menjadi bahasa formal standar (Dyasputro, 2025).

2.2.18 *Google-Play-Scraper*

Google-Play-Scraper adalah metode untuk mengambil data dari *Google Play Store* tanpa dependensi eksternal menggunakan bahasa pemrograman *python*. Data yang diambil dapat berupa informasi aplikasi seperti judul aplikasi, developer url, kategori aplikasi, keseluruhan rating dan review, deskripsi, thumbnail, rating konten dan screenshot aplikasi. Selain informasi dari aplikasi *API googleplay-scraper* juga dapat mengambil data ulasan pengguna seperti nama, foto, rating, tanggal, comment likes, dan comment (Riansyah Ramadhan & Agus Sugianto, 2024).

2.2.19 *Sastrawi*

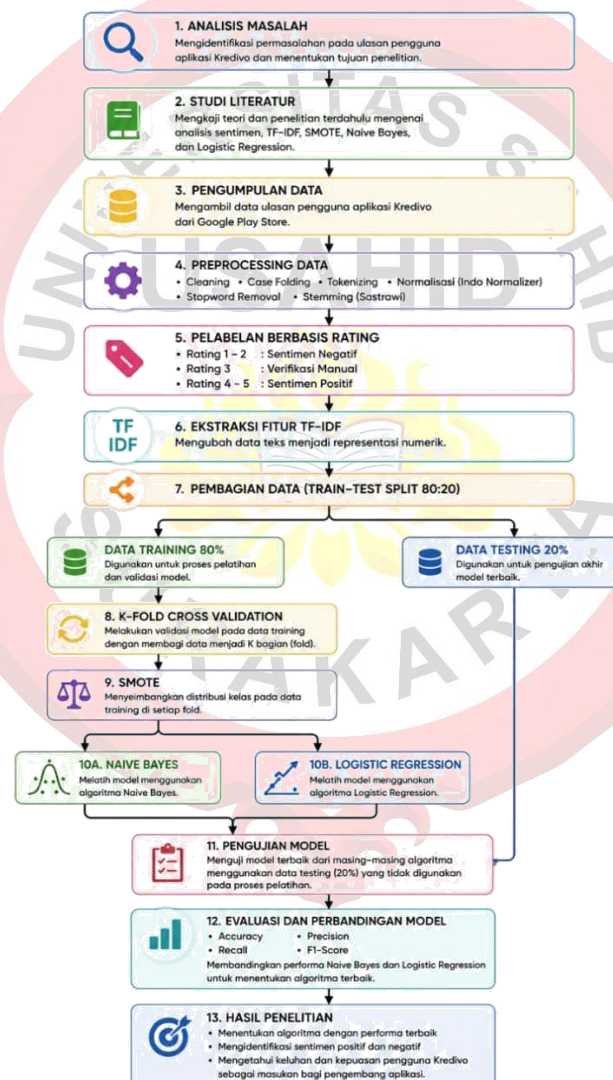
Sastrawi merupakan sebuah sumber daya yang digunakan untuk memproses kata-kata agar dapat diubah menjadi bentuk dasarnya. *Sastrawi* stemmer mengikuti prinsip yang dikembangkan oleh Nazief dan Adriani, dan telah diperbaiki dengan algoritma CS (Confix Stripping), Algoritma ECS (Enhanced Confix Stripping), serta dimodifikasi lebih lanjut dengan Modifikasi ECS (Pardede & Darmawan, 2025).

2.2.20 *Google Colaboratory*

Google Colab (singkatan dari *Google Colaboratory*) adalah platform berbasis *cloud computing* yang disediakan oleh Google. Ini memungkinkan pengguna untuk mengeksekusi kode *Python* dalam lingkungan berbasis *cloud* tanpa perlu menginstal atau mengatur lingkungan lokal mereka sendiri. *Google Colab* sering digunakan oleh para ilmuwan data, peneliti, dan pengembang untuk melakukan berbagai jenis pekerjaan, termasuk pemrosesan data, pengembangan model kecerdasan buatan, analisis data, dan pelatihan model mesin (Andarsyah & Yanuar, 2024).

2.3 Kerangka Pemikiran

Kerangka pemikiran ini disusun untuk menggambarkan alur logis penelitian dalam mengolah data ulasan tidak terstruktur menjadi informasi sentimen yang terklasifikasi. Secara esensial, kerangka ini dirancang untuk menjembatani kesenjangan antara ketersediaan data ulasan yang masif pada aplikasi Kredivo dengan kebutuhan akan informasi strategis mengenai sentimen pelanggan secara akurat dan otomatis. Adapun tahapan sistematis dalam kerangka pemikiran ini dapat dilihat pada Gambar 2.2.



Gambar 2. 2 Kerangka Pemikiran

2.3.1 Analisis Masalah

Tahap pertama dalam penelitian adalah melakukan analisis terhadap permasalahan yang menjadi dasar penelitian. Aplikasi Kredivo sebagai salah satu layanan *Buy Now Pay Later* (BNPL) memiliki jutaan ulasan pengguna pada Google Play Store yang berisi berbagai pengalaman, kritik, saran, maupun apresiasi terhadap layanan yang diberikan. Banyaknya jumlah ulasan menyebabkan proses analisis secara manual menjadi tidak efisien, membutuhkan waktu yang lama, serta rentan terhadap subjektivitas. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data teks secara otomatis sehingga dapat mengidentifikasi sentimen positif maupun negatif dari ulasan pengguna secara lebih cepat dan objektif.

2.3.2 Studi Literatur

Setelah permasalahan diidentifikasi, dilakukan studi literatur untuk memperoleh landasan teori dan referensi ilmiah yang mendukung penelitian. Studi literatur mencakup teori mengenai *Financial Technology* (FinTech), analisis sentimen, *Natural Language Processing* (NLP), *text mining*, tahapan *preprocessing*, TF-IDF, *Synthetic Minority Oversampling Technique* (SMOTE), algoritma *Naive Bayes*, algoritma *Logistic Regression*, serta metode evaluasi model. Selain itu, dilakukan kajian terhadap penelitian terdahulu untuk mengetahui perkembangan penelitian, membandingkan metode yang digunakan, serta menemukan *research gap* yang menjadi dasar pelaksanaan penelitian ini.

2.3.3 Pengumpulan Data

Tahap selanjutnya adalah pengumpulan data ulasan pengguna aplikasi Kredivo yang diperoleh dari Google Play Store menggunakan teknik *web scraping*. Proses pengambilan data dilakukan menggunakan bantuan pustaka *Google Play Scraper* pada bahasa pemrograman Python sebanyak 50.000 data ulasan. Data yang dikumpulkan kemudian dilakukan filtering selama periode Januari 2021 hingga Januari 2026, sehingga mendapatkan 17.212 data ulasan. Dataset tersebut menjadi sumber data utama dalam penelitian karena berisi opini pengguna yang dapat dimanfaatkan untuk analisis sentimen aplikasi Kredivo.

2.3.4 *Preprocessing Data*

Data hasil *web scraping* umumnya masih mengandung berbagai karakter yang tidak diperlukan sehingga perlu dilakukan proses *preprocessing*. Tahapan ini bertujuan meningkatkan kualitas data agar lebih bersih, seragam, dan siap digunakan pada proses klasifikasi. Adapun tahapan *preprocessing* yang dilakukan meliputi:

- a) *Cleaning*, yaitu menghapus URL, angka, tanda baca, emoji, simbol, dan karakter lain yang tidak memiliki pengaruh terhadap analisis.
- b) *Case Folding*, yaitu mengubah seluruh huruf menjadi huruf kecil (*lowercase*) agar penulisan kata menjadi seragam.
- c) *Tokenizing*, yaitu memisahkan kalimat menjadi kumpulan kata (*token*) sehingga lebih mudah diproses.
- d) Normalisasi, menggunakan *Indo Normalizer* untuk mengubah kata tidak baku, singkatan, atau bahasa gaul menjadi kata baku.
- e) *Stopword Removal*, yaitu menghapus kata-kata umum yang tidak memberikan informasi penting terhadap proses klasifikasi.
- f) *Stemming*, menggunakan *Sastrawi* untuk mengubah setiap kata menjadi bentuk dasarnya.

2.3.5 *Perlabelan Berbasis Rating*

Setelah data selesai diproses, dilakukan pelabelan sentimen berdasarkan nilai *rating* yang diberikan oleh pengguna pada Google Play Store. Metode pelabelan berbasis *rating* dipilih karena lebih efisien dibandingkan pelabelan manual seluruh data serta dapat menggambarkan kecenderungan penilaian pengguna terhadap aplikasi. Pendekatan ini bertujuan menghasilkan label sentimen yang lebih akurat sehingga dapat meningkatkan kualitas model klasifikasi. Aturan pelabelan yang digunakan adalah sebagai berikut:

- a) *Rating* 1 dan 2 diberi label sentimen negatif.
- b) *Rating* 4 dan 5 diberi label sentimen positif.
- c) *Rating* 3 dilakukan verifikasi manual berdasarkan isi ulasan karena bersifat ambigu dan tidak selalu mencerminkan sentimen tertentu.

2.3.6 Ekstraksi Fitur TF-IDF

Setelah proses pelabelan selesai, data teks diubah menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Algoritma klasifikasi tidak dapat memproses data dalam bentuk teks secara langsung sehingga diperlukan proses ekstraksi fitur.

TF-IDF dipilih karena mampu memberikan bobot yang lebih tinggi pada kata-kata yang memiliki tingkat kepentingan tinggi dalam suatu dokumen serta mengurangi bobot kata-kata yang terlalu sering muncul pada seluruh dokumen. Dengan demikian, representasi data menjadi lebih informatif dan mampu meningkatkan performa proses klasifikasi.

2.3.7 Pembagian Data (*Train-Test Split*)

Dataset yang telah direpresentasikan menggunakan TF-IDF kemudian dibagi menjadi dua bagian, yaitu 80% data pelatihan (*training*) dan 20% data pengujian (*testing*). Pembagian data ini bertujuan agar model dapat dilatih menggunakan sebagian besar data, sedangkan sebagian data lainnya digunakan sebagai data pengujian yang benar-benar baru. Penggunaan pembagian data 80:20 merupakan salah satu proporsi yang umum digunakan dalam penelitian *machine learning* karena mampu memberikan keseimbangan antara proses pelatihan dan evaluasi model.

2.3.8 *K-Fold Cross Validation*

Pada data *training* dilakukan proses *K-Fold Cross Validation* untuk memperoleh model yang lebih stabil dan mengurangi risiko *overfitting*. Dalam metode ini, data *training* dibagi menjadi beberapa bagian (*fold*). Selanjutnya, setiap *fold* digunakan secara bergantian sebagai data validasi, sedangkan *fold* lainnya digunakan sebagai data pelatihan. Pendekatan ini memungkinkan seluruh data *training* memiliki kesempatan menjadi data validasi sehingga hasil pelatihan lebih representatif dibandingkan hanya menggunakan satu kali pembagian data.

2.3.9 *Synthetic Minority Oversampling Technique* (SMOTE)

Distribusi data sentimen pada penelitian ini tidak seimbang karena jumlah ulasan positif lebih banyak dibandingkan ulasan negatif. Ketidakseimbangan data tersebut dapat menyebabkan model cenderung mempelajari kelas mayoritas

sehingga menurunkan kemampuan dalam mengenali kelas minoritas. Untuk mengatasi permasalahan tersebut, diterapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini menghasilkan data sintesis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. SMOTE diterapkan hanya pada data pelatihan di setiap *fold* agar tidak terjadi *data leakage* serta menjaga objektivitas proses evaluasi model.

2.3.10 Pelatihan Model

Tahap berikutnya adalah membangun model klasifikasi menggunakan dua algoritma, yaitu *Naive Bayes* dan *Logistic Regression*. Algoritma *Naive Bayes* dipilih karena merupakan metode klasifikasi berbasis probabilitas yang memiliki proses komputasi sederhana, cepat, serta banyak digunakan dalam analisis sentimen berbasis teks. Sementara itu, *Logistic Regression* dipilih karena merupakan algoritma klasifikasi linear yang mampu mempelajari hubungan antara fitur TF-IDF dengan kelas sentimen serta memiliki kemampuan generalisasi yang baik terhadap data baru. Pelatihan kedua algoritma dilakukan menggunakan data *training* yang telah melalui proses validasi dan penyeimbangan data sehingga performa model dapat dibandingkan secara objektif.

2.3.11 Pengujian Model

Setelah proses pelatihan selesai, model terbaik dari masing-masing algoritma diuji menggunakan data *testing* sebesar 20% yang sebelumnya tidak pernah digunakan pada proses pelatihan maupun validasi. Penggunaan data *testing* yang terpisah bertujuan memperoleh gambaran mengenai kemampuan model dalam mengklasifikasikan data baru. Selain itu, pemisahan data *testing* juga bertujuan menghindari terjadinya *data leakage* sehingga hasil evaluasi model lebih objektif.

2.3.12 Evaluasi

Tahap evaluasi dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

- a) *Accuracy* digunakan untuk mengukur tingkat ketepatan klasifikasi secara keseluruhan.
- b) *Precision* mengukur ketepatan model dalam memprediksi suatu kelas.

- c) *Recall* mengukur kemampuan model dalam menemukan seluruh data pada kelas tertentu.
- d) *F1-score* merupakan nilai harmonis antara *precision* dan *recall* sehingga memberikan gambaran performa model secara lebih seimbang.

2.3.13 Hasil Penelitian

Tahap akhir penelitian menghasilkan informasi mengenai algoritma yang memiliki performa terbaik berdasarkan hasil evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, penelitian ini juga menghasilkan identifikasi sentimen positif dan sentimen negatif pengguna aplikasi Kredivo, serta visualisasi *word cloud* yang menggambarkan kata-kata dominan pada masing-masing kategori sentimen. Hasil visualisasi tersebut memberikan gambaran mengenai berbagai bentuk keluhan dan kepuasan yang paling sering disampaikan oleh pengguna dalam ulasan aplikasi Kredivo.

Hasil penelitian diharapkan dapat menjadi bahan evaluasi bagi pengembang aplikasi Kredivo dalam meningkatkan kualitas layanan, memperbaiki fitur yang masih memiliki kendala, serta meningkatkan kepuasan pengguna. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis sentimen pada layanan *Financial Technology (FinTech)* menggunakan metode *machine learning*, khususnya yang membandingkan performa algoritma klasifikasi pada data ulasan pengguna.