

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Gambaran Umum Objek Penelitian

Tujuan riset ini adalah menginvestigasi dinamika layanan pada Adakami, penyelenggara pinjaman daring yang menawarkan akses pembiayaan instan tanpa agunan untuk nasabah individu. Didirikan pada tahun 2018, platform ini telah mengamankan registrasi dari Otoritas Jasa Keuangan (OJK). Melalui kasus sebelumnya, ditemukan bahwa terdapat sejumlah masalah yang dihadapi oleh para peminjam, seperti pembayaran angsuran yang tidak sesuai dengan kesepakatan dan perilaku intimidatif dari penagih utang yang telah menimbulkan kekhawatiran di kalangan peminjam. Adakami, yaitu:

1. Angsuran yang harus dibayar tidak sesuai dengan tarif yang disepakati
2. Peminjam Adakami mengeluhkan bahwa angsuran yang harus mereka bayarkan tidak sesuai dengan tarif yang disepakati saat pengajuan pinjaman. Hal ini disebabkan oleh adanya biaya-biaya tambahan yang tidak dijelaskan secara transparan oleh Adakami.
3. Peminjam Adakami juga mengeluhkan perlakuan debt collector yang meresahkan. Debt collector Adakami sering kali melakukan intimidasi dan ancaman kepada peminjam yang menunggak angsuran.

Objek utama dalam penelitian ini adalah Adakami, sebuah platform layanan pinjam meminjam uang berbasis teknologi informasi (*fintech peer-to-peer lending*) yang dioperasikan oleh PT Pembiayaan Digital Indonesia. Platform yang didirikan pada tahun 2018 ini menawarkan akses pembiayaan instan tanpa agunan bagi nasabah individu dan telah resmi terdaftar serta diawasi oleh Otoritas Jasa Keuangan (OJK):

1. Sebagai salah satu entitas dominan dalam ekosistem *fintech* di Indonesia, Adakami menyasar segmen masyarakat yang belum terjangkau layanan perbankan konvensional (*unbanked*). Penelitian ini secara spesifik menyoroti aspek pengalaman pengguna aplikasi Adakami dengan menjadikan ulasan publik (public reviews) di Google Play Store sebagai sumber data primer. Ulasan-ulasan tersebut dipilih karena merepresentasikan pengalaman otentik pengguna terkait berbagai aspek layanan, mulai dari proses pengajuan, transparansi biaya, hingga etika penagihan.
2. Data yang diperoleh dari objek penelitian ini selanjutnya akan diolah menggunakan kerangka kerja metodologi *Knowledge Discovery in Databases* (KDD) untuk mengklasifikasikan sentimen dan mengekstraksi pola-pola yang mempengaruhi persepsi nasabah.

Signifikansi penelitian ini terletak pada kontribusinya dalam memberikan wawasan kepada masyarakat terkait risiko yang mungkin dihadapi saat mengakses layanan pinjaman digital Adakami . Selain itu, temuan studi ini diharapkan menjadi landasan rekomendasi bagi pihak manajemen Adakami untuk mereformasi sistem dan prosedur operasional, guna meningkatkan kualitas layanan bagi nasabah. Adapun seluruh rangkaian penelitian ini dijadwalkan berlangsung pada periode tahun 2025.

### **3.2. Metode Penelitian**

Kerangka kerja *Knowledge Discovery in Databases* (KDD) diadopsi sebagai fondasi metodologis dalam studi ini. Mengingat fokus penelitian pada pengolahan korpus data ulasan berskala besar dan kuantifikasi sentimen, pendekatan kuantitatif menjadi pilihan utama. Struktur analisis terbagi menjadi dua: metode deskriptif untuk memetakan proporsi sentimen, dan metode komparatif untuk mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) versus *Random Forest*. Validitas perbandingan ini dijamin melalui penerapan *Optimizing* guna memaksimalkan performa klasifikasi kedua model tersebut.

Kerangka kerja KDD akan memandu seluruh proses penelitian melalui tahapan:

1. Data Collection
2. Data Selection
3. Preprocessing
4. Transformation
5. Data Mining
6. Evaluation/Interpretation

Tabel 3.1 Perbandingan KDD Dan CRISP DM

	KDD	CRISP-DM
Fokus Utama	Pendekatan ini menitikberatkan pada prosedur teknis guna mengekstraksi pola serta wawasan baru yang tersimpan di dalam data. Penggunaannya sangat lazim ditemukan dalam ranah akademis maupun lingkungan riset ilmiah.	Berorientasi pada pemecahan masalah bisnis. Dimulai dengan pemahaman tujuan bisnis dan diakhiri dengan implementasi solusi praktis di industri..
Sifat Proses	Cenderung linear dan sekuensial . Tahapan dijalankan secara berurutan dari seleksi data hingga interpretasi hasil.	Bersifat siklus (cyclical) dan iteratif . Memungkinkan untuk kembali ke tahap sebelumnya jika diperlukan penyesuaian berdasarkan temuan baru.
Tujuan Akhir	Menghasilkan pengetahuan (knowledge) atau wawasan baru yang valid dan bermanfaat dari data. Penemuan itu sendiri adalah	Menghasilkan solusi yang dapat diimplementasikan (deployment) untuk menjawab kebutuhan bisnis, seperti meningkatkan

	hasil utamanya..	penjualan atau efisiensi.
Fleksibilitas	Cenderung lebih kaku karena alurnya yang terstruktur dan sekvensial. Perubahan di tengah proses lebih sulit diakomodasi.	Sangat fleksibel. Dirancang untuk beradaptasi dengan perubahan dan pemahaman baru yang muncul selama proyek berlangsung.
Konteks dan Orientasi	Fokus pada validitas dan signifikansi pola yang ditemukan dari sudut pandang data. Kurang menekankan pada implementasi langsung.	Fokus pada relevansi solusi terhadap tujuan bisnis. Setiap langkah, mulai dari data hingga model, harus sejalan dengan kebutuhan bisnis awal.

### 3.2.1. Data Collection

Penelitian ini mendayagunakan data sekunder yang bersumber dari ulasan publik pengguna aplikasi Adakami di Google Play Store. Mekanisme pengumpulan data dijalankan melalui teknik web scraping, yakni metode ekstraksi informasi otomatis dari laman web. Secara spesifik, alat yang digunakan adalah pustaka Python google-play-scraper. Pemilihan pustaka ini didasarkan pada keunggulannya dalam menghasilkan data terstruktur secara efisien, serta kemudahan implementasinya yang tidak memerlukan konfigurasi otomatisasi peramban (browser automation) yang rumit.

Dalam proses akuisisi data, ditetapkan batasan jumlah dataset sebanyak 5.000 ulasan terbaru. Pembatasan ini bukan dilakukan secara sembarang, melainkan didasarkan pada tiga pertimbangan logis utama:

1. Relevansi Waktu (Kebaruan Data): Penelitian ini bertujuan memotret sentimen masyarakat terhadap kondisi layanan Adakami

yang paling aktual (saat ini). Jika mengambil data dalam jumlah yang sangat besar (misalnya ratusan ribu), risiko masuknya ulasan-ulasan lawas (dari tahun-tahun sebelumnya) akan sangat tinggi. Ulasan lama tersebut mungkin sudah tidak relevan karena adanya perubahan kebijakan aplikasi atau regulasi OJK, sehingga dapat menyebabkan bias dalam analisis sentimen terkini.

2. Efisiensi Sumber Daya Komputasi: Metode *Support Vector Machine* (SVM) yang dikombinasikan dengan teknik optimasi *Grid Search Cross-Validation* dikenal membutuhkan sumber daya komputasi yang sangat intensif karena melakukan pelatihan model berulang kali untuk mencari akurasi terbaik. Jumlah 5.000 data dinilai sebagai titik keseimbangan yang paling ideal; cukup besar untuk melatih model menjadi cerdas, namun tetap realistik untuk diolah menggunakan perangkat komputer standar tanpa memakan waktu proses yang berlebihan.
3. Titik Jenuh Informasi (Data Saturation): Dalam ulasan aplikasi pinjaman *online*, variasi keluhan atau puji yang disampaikan nasabah cenderung memiliki pola yang berulang (misalnya kata-kata seputar "bunga", "penagihan", "cepat cair"). Secara logis, sampel 5.000 ulasan dianggap telah mencapai titik jenuh informasi, di mana penambahan data lebih banyak cenderung hanya akan mengulang pola kata yang sama tanpa memberikan wawasan baru yang signifikan, justru berpotensi menambah data sampah (*noise*) yang tidak perlu.

### 3.2.2. Data Selection

Sebagai tindak lanjut dari pengumpulan data, dilakukan mekanisme seleksi data (data selection) untuk menyaring input yang akan diolah. Tujuan utamanya adalah memastikan hanya data berkualitas dan relevan yang tersisa, sehingga akurasi hasil penelitian dapat

dipertanggungjawabkan. Proses purifikasi ini esensial untuk mereduksi distorsi (noise) yang kerap terdapat pada dataset mentah.

Untuk mencapai tujuan tersebut, beberapa kriteria seleksi yang ketat akan diterapkan pada dataset hasil scraping:

1. Relevansi Bahasa: Hanya ulasan yang ditulis dalam Bahasa Indonesia yang akan dipertahankan. Ulasan yang terdeteksi menggunakan bahasa asing akan dihapus dari dataset untuk menjaga konsistensi linguistik pada tahap pemrosesan selanjutnya.
2. Keberadaan Teks Ulasan: Ulasan yang tidak memiliki konten teks (hanya memberikan rating bintang) akan dieksklusi. Fokus penelitian ini adalah analisis sentimen berbasis teks, sehingga ulasan kosong tidak memberikan informasi yang dapat diolah.
3. Penghapusan Duplikasi Data: Seluruh dataset akan diperiksa untuk mengidentifikasi dan menghapus entri ulasan yang identik atau terduplikasi. Langkah ini penting untuk mencegah adanya bias pada model dan memastikan setiap data poin bersifat unik.
4. Panjang Minimum Ulasan: Untuk memastikan ulasan memiliki substansi yang cukup untuk dianalisis, akan ditetapkan batasan panjang minimum, misalnya ulasan harus terdiri dari minimal 3 kata. Ulasan yang terlalu pendek (contoh: "Ok", "Sip", "Bagus") akan dihilangkan karena tidak memberikan konteks sentimen yang memadai untuk analisis mendalam.

### **3.2.3. Data Preprocessing**

Setelah tahap seleksi data selesai, langkah selanjutnya dalam kerangka kerja penelitian adalah pra-pemrosesan data. Tahap ini sangat krusial karena data teks yang diperoleh dari ulasan pengguna bersifat tidak terstruktur, bervariasi, dan mengandung banyak noise. Tahapan pra-pemrosesan menjadi sangat vital. Fokus

utamanya adalah melakukan purifikasi, standarisasi, serta simplifikasi data agar dapat dikonversi menjadi representasi numerik yang valid. Kualitas input yang prima ini pada akhirnya akan mengoptimalkan akurasi dan kinerja model klasifikasi secara signifikan

Proses ini akan menerapkan serangkaian teknik Natural Language Processing (NLP) yang dijalankan secara berurutan berikut tertera :

1. Case Folding. Bertujuan untuk menyeragamkan segala huruf dalam teks ulasan menjadi format huruf kecil (lowercase). Hal ini penting untuk memastikan bahwa kata yang sama namun ditulis dengan kapitalisasi berbeda (misalnya, "Bagus", "bagus", dan "BAGUS") dianggap sebagai satu kata yang sama oleh model.

Contoh: "BUNGA TERLALU TINGGI" → "bunga terlalu tinggi"

2. Cleansing (Pembersihan Data). Tahap ini berfokus pada penghapusan karakter-karakter yang tidak memiliki nilai informatif untuk analisis sentimen. Karakter yang akan dihilangkan mencakup tanda baca (misalnya, !, ?, ., ,), angka, karakter spesial (misalnya, @, #), emoji, dan URL.

Contoh: "@adakami bunganya 100%!! tolong" → "adakami bunganya tolong"

3. Tokenizing (Tokenisasi). Setelah teks bersih dari karakter yang tidak relevan, proses tokenisasi akan memecah setiap kalimat dalam ulasan menjadi unit-unit kata individual yang disebut token. Hasil dari tahap ini adalah daftar (list) token untuk setiap ulasan, yang memungkinkan setiap kata diolah secara terpisah.

Contoh: "penagihan sangat kasar" → ['penagihan', 'sangat', 'kasar']

4. Stopword Removal (Penghapusan Stopword). Tahapan ini didesain untuk mengeliminasi *stopwords* atau kata henti, yakni kosa kata umum dalam Bahasa Indonesia yang memiliki frekuensi kemunculan tinggi namun minim kontribusi terhadap nilai sentimen. Kata-kata seperti "yang", "di", "dan", "adalah", "ini", "saya" akan dihapus menggunakan daftar stopwords yang telah ditentukan. Dengan mengeliminasi gangguan tersebut, model dapat bekerja lebih optimal dalam mengidentifikasi kata kunci yang menjadi indikator utama sentimen.

Contoh: ['penagihan', 'yang', 'sangat', 'kasar'] → ['penagihan', 'sangat', 'kasar']

5. Stemming. Sebagai langkah pemungkas dalam prosedur persiapan data teks, diterapkan teknik *stemming*, yaitu proses mengubah kata-kata berimbuhan ke bentuk kata dasarnya (root word). Tujuannya adalah untuk mengurangi jumlah variasi kata dengan makna yang sama, sehingga dimensi fitur menjadi lebih ringkas. Proses ini akan menggunakan algoritma stemmer untuk Bahasa Indonesia (Nazief & Adriani) yang diimplementasikan melalui pustaka Python seperti PySastrawi[25].

Contoh: ['penagihan', 'mengerikan'] → ['tagih', 'ngeri']

### 3.2.4. Data Transformation

Algoritma *machine learning* tidak dapat memproses data teks mentah secara langsung, sehingga diperlukan konversi data menjadi format numerik yang terstruktur. Oleh karena itu, penelitian ini menerapkan metode pembobotan CountVectorizer untuk mentransformasi teks ulasan yang telah bersih menjadi representasi vektor. Teknik ini dipilih karena kemampuannya dalam memberikan bobot nilai berdasarkan signifikansi kemunculan kata dalam sebuah dokumen relatif terhadap

keseluruhan kumpulan data (*corpus*) . Matriks vektor yang dihasilkan dari proses ini selanjutnya digunakan sebagai data input standar untuk melatih dan menguji performa kedua model klasifikasi, yakni Support Vector Machine (SVM) dan Random Forest.

### 3.2.5. Data Mining

Tahap ini adalah inti penggalian pola untuk membangun model klasifikasi sentimen. Penelitian ini melakukan komparasi performa antara SVM dan Random Forest melalui langkah-langkah berikut:

1. Pelabelan Otomatis (*Lexicon-Based Labeling*): Mengingat volume data yang besar, proses pelabelan sentimen dilakukan menggunakan pendekatan berbasis kamus (*lexicon-based*) dengan memanfaatkan kamus sentimen InSet (*Indonesia Sentiment Lexicon*). Setiap ulasan dihitung skor sentimennya berdasarkan kecocokan kata dengan daftar kata positif dan negatif di dalam kamus. Ulasan dengan skor  $> 0$  dilabeli Positif, skor  $< 0$  dilabeli Negatif, dan skor 0 dilabeli Netral . 2. Pembagian Data (*Splitting*): Dataset berlabel kemudian dibagi menjadi dua *subset* dengan rasio 80:20, di mana 80% data digunakan untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*) secara acak berstrata (*stratified*) guna menjaga proporsi kelas sentimen yang seimbang.
2. Optimasi Hyperparameter (*Hyperparameter Tuning*): Untuk mendapatkan performa terbaik, penelitian ini menerapkan teknik *Grid Search Cross-Validation* (GridSearchCV). Teknik ini menguji kombinasi parameter secara sistematis untuk menemukan konfigurasi paling optimal. Pada SVM, parameter yang diuji meliputi kernel, C, dan gamma. Pada Random Forest, parameter yang diuji meliputi *n\_estimators*, *max\_depth*, dan *min\_samples\_split* .

3. Klasifikasi: Model dengan konfigurasi parameter terbaik kemudian dilatih kembali menggunakan seluruh data latih untuk membangun model klasifikasi final.

### 3.2.6. Evaluation

Pengukuran performa model dilakukan pada tahap evaluasi dengan memanfaatkan dataset pengujian yang terpisah, guna menjamin objektivitas hasil prediksi pada data baru.

1. **Metrik Performa (*Confusion Matrix*):** Penilaian utama kinerja model didasarkan pada parameter *Confusion Matrix*. Metrik yang diukur meliputi:
  - **Akurasi (Accuracy):** Seberapa sering model memprediksi secara benar secara keseluruhan.
  - **Presisi (Precision):** Tingkat ketepatan prediksi positif model.
  - **Recall:** Kemampuan model menemukan kembali informasi positif. \* ***F1-Score:*** Rata-rata harmonik antara presisi dan *recall*, yang sangat krusial untuk mengevaluasi data yang tidak seimbang (*imbalanced data*).
2. **Analisis Komparatif:** Dilakukan perbandingan langsung (*direct comparison*) antara nilai akurasi dan *F1-Score* terbaik yang dihasilkan oleh SVM dan Random Forest setelah proses optimasi. Model dengan skor tertinggi pada data uji ditetapkan sebagai model terbaik (*Champion Model*).
3. **Interpretasi Pengetahuan:** Hasil prediksi dari model terbaik digunakan untuk memetakan distribusi sentimen secara keseluruhan. Selanjutnya, dilakukan analisis konten (seperti *Word Cloud*) pada ulasan negatif untuk menjawab rumusan masalah

mengenai aspek layanan yang menjadi pemicu ketidakpuasan nasabah.

### **3.2.7. Implementasi**

Tahap implementasi merupakan fase akhir dari metodologi penelitian, di mana model klasifikasi sentimen yang telah dibangun dan dievaluasi akan diterapkan untuk mencapai tujuan utama penelitian. Dalam konteks penelitian akademis, implementasi tidak diartikan sebagai pembuatan aplikasi komersial, melainkan sebagai proses penggunaan model untuk menganalisis keseluruhan dataset dan penyajian temuan dalam format yang dapat dipahami.

## **3.3. Teknik Pengumpulan Data**

Inisiasi penelitian dilakukan melalui tahapan pengumpulan data yang krusial. Objek data yang diolah merupakan data sekunder, bersumber dari jejak digital ulasan pengguna aplikasi Adakami . Guna menangani volume data daring yang signifikan tersebut secara efektif, teknik *web scraping* diterapkan sebagai instrumen ekstraksi utama .

Sumber data utama untuk penelitian ini adalah halaman aplikasi Adakami di platform distribusi digital Google Play Store. Platform ini dipilih sebagai sumber data primer karena beberapa alasan strategis:

1. Otentisitas: Ulasan yang tersedia merupakan umpan balik langsung dan otentik dari pengguna yang telah memiliki pengalaman nyata dalam menggunakan layanan Adakami .
2. Kekayaan Konteks: Ulasan seringkali berisi detail spesifik mengenai pengalaman pengguna, baik positif maupun negatif, yang sangat relevan untuk menjawab rumusan masalah.

- Volume Data: Sebagai platform distribusi aplikasi utama, Google Play Store menyediakan volume data ulasan yang sangat besar, yang esensial untuk melatih model *machine learning* yang andal.

### 3.4. Teknik Analisis Data

Penelitian ini mengadopsi pendekatan analisis data kuantitatif yang menggabungkan metode komparatif dan deskriptif. Alur analisis dirancang dalam dua fase sistematis guna menjawab persoalan penelitian secara komprehensif.

Tahap analisis komparatif didedikasikan untuk menyeleksi model klasifikasi paling superior. Fokus evaluasi tertuju pada dua algoritma pembelajaran mesin, yakni *Support Vector Machine* (SVM) dan *Random Forest*. Guna menjamin komparasi yang objektif, kedua model menjalani prosedur **Optimasi Hyperparameter** melalui teknik *Grid Search Cross-Validation* (GridSearchCV) untuk menentukan konfigurasi terbaik. Kinerja model kemudian diukur pada data uji berdasarkan parameter *Confusion Matrix* (Akurasi, Presisi, *Recall*, *F1-Score*) [22]. Dalam proses optimasi menggunakan *Grid Search Cross-Validation* (GridSearchCV), penelitian ini menguji berbagai kombinasi parameter untuk menemukan konfigurasi model yang paling optimal. Rentang nilai parameter yang diuji untuk masing-masing algoritma dirincikan pada tabel berikut:

Tabel 3.2 Skenario Pengujian Hyperparameter (Grid Search Space)

Algoritma	Parameter	Fungsi	Nilai yang Diuji (Search Space)
Support Vector Machine (SVM)	Kernel	Menentukan jenis fungsi pemisah data (linear atau non-linear).	['linear', 'rbf']

	C (Regularization)	Mengontrol <i>trade-off</i> antara kesalahan klasifikasi data latih dan margin pemisah. Nilai C kecil membuat margin lebih lebar (lebih general), C besar membuat margin lebih sempit (lebih akurat di training).	[0.1, 1, 10, 100]
	Gamma	Menentukan seberapa jauh pengaruh satu data latih.	['scale', 'auto']
Random Forest	n_estimators	Jumlah pohon keputusan ( <i>decision trees</i> ) yang akan dibangun. Semakin banyak pohon, semakin stabil hasil prediksi.	[50, 100, 200]
	max_depth	Kedalaman maksimal setiap pohon. Membatasi kedalaman mencegah <i>overfitting</i> .	[None, 10, 20]
	min_samples_split	Jumlah minimum sampel yang diperlukan untuk memecah simpul internal ( <i>node</i> ).	[2, 5, 10]

Model yang menghasilkan nilai Akurasi dan *F1-Score* tertinggi pada data uji (Test Set) akan dipilih sebagai metode terbaik. Selanjutnya, analisis dilanjutkan ke tahap deskriptif menggunakan visualisasi data dan *Word Cloud* untuk menginterpretasikan makna di balik label sentimen yang dihasilkan.

Setelah model dengan performa terbaik diidentifikasi pada analisis komparatif, model tersebut akan digunakan untuk mengklasifikasikan data ulasan. Menggunakan analisis deskriptif & interpretasi pengetahuan hasil dari klasifikasi

ini kemudian akan dianalisis secara deskriptif untuk mengekstrak pengetahuan dan temuan kunci. Proses analisis ini mencakup:

1. Analisis Distribusi Sentimen: Guna memperoleh gambaran komprehensif terkait persepsi pengguna Adakami , dilakukan analisis deskriptif terhadap hasil klasifikasi. Data dikonsolidasi untuk mengukur distribusi frekuensi dan persentase pada masing-masing kelas sentimen. Temuan statistik ini selanjutnya direpresentasikan secara visual melalui instrumen grafis guna memudahkan interpretasi data.
2. Analisis Konten pada Sentimen Negatif: Untuk mengidentifikasi pemicu utama ketidakpuasan nasabah, analisis mendalam akan difokuskan secara spesifik pada ulasan yang terkласifikasi sebagai sentimen negatif. Teknik yang digunakan adalah analisis frekuensi kata (*word frequency analysis*) untuk mengidentifikasi kata kunci dan frasa yang paling dominan muncul dalam keluhan, seperti "bunga", "teror", "penagihan", atau "data". Hasil ini akan divisualisasikan menggunakan *Word Cloud* untuk memberikan representasi masalah yang intuitif.
3. Interpretasi dan Penarikan Kesimpulan: Tahap terakhir adalah interpretasi holistik. Data kuantitatif dari distribusi sentimen (poin 1) dan temuan kualitatif dari analisis konten (poin 2) akan disintesis untuk menjawab rumusan masalah. Pola-pola yang ditemukan akan dihubungkan kembali dengan konteks studi kasus yang diangkat dalam latar belakang masalah. Berdasarkan interpretasi ini, kesimpulan dan rekomendasi strategis akan dirumuskan bagi perusahaan maupun regulator.

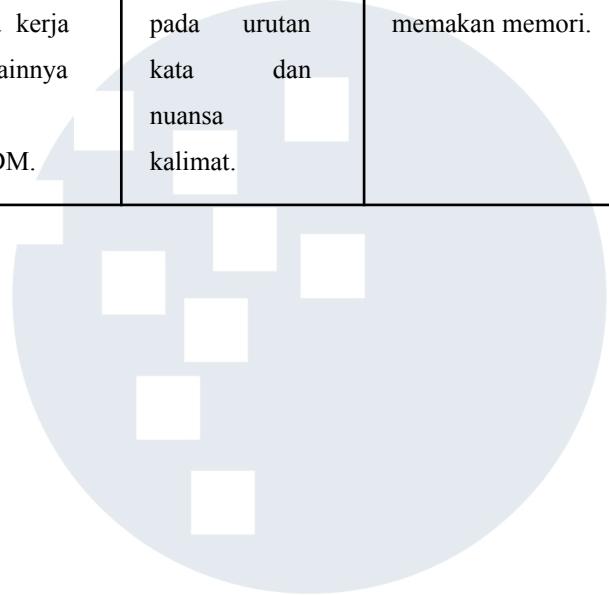
Tabel 3.3 Perbandingan KDD,SVM dan Random Forest

	KDD	SVM	Random Forest
--	-----	-----	---------------

Peran dalam Penelitian	<p>KDD didefinisikan sebagai prosedur sistematis untuk mengekstraksi wawasan bernilai dari basis data. Kerangka kerja ini mencakup serangkaian tahapan terintegrasi, mulai dari akuisisi dan purifikasi data, konversi format data, hingga analisis mendalam</p>	<p>Algoritma <i>Data Mining</i> diterapkan pada tahap ke-5 KDD untuk melakukan klasifikasi sentimen</p>	<p>Algoritma <i>Ensemble Learning</i> yang digunakan sebagai pembanding SVM untuk menemukan model dengan performa terbaik setelah proses optimasi.</p>
Konsep Utama	<p>Mekanisme sistematis yang berfungsi mengonversi data mentah menjadi wawasan bernilai melalui serangkaian prosedur yang logis dan tertata.</p>	<p>Operasional metode ini didasarkan pada pencarian <i>hyperplane</i> yang mampu mempartisi data ke dalam kategori berbeda, dengan fokus utama pada pencapaian margin pemisahan yang maksimal antar kelas</p>	<p>Algoritma ini mengonstruksi sejumlah besar struktur pohon keputusan, kemudian mengagregasi seluruh prediksi individual melalui mekanisme suara terbanyak (<i>majority vote</i>) guna menghasilkan luaran yang memiliki tingkat akurasi dan stabilitas superior</p>

Input & Output	<p>Input: Data mentah (ulasan dari Google Play Store).</p> <p>Output: Pengetahuan yang dapat ditindaklanjuti (misal, aspek pemicu ketidakpuasan).</p>	<p>Input: Representasi numerik berupa Count Vectorizer.</p> <p>Output: Prediksi kelas sentimen (Positif/Negatif/Netral).</p>	<p>Input: Representasi numerik berupa Count Vectorizer.</p> <p>Output: Prediksi kelas sentimen (Positif/Negatif/Netral).</p>
Kemampuan Memahami Konteks	Tidak relevan (KDD adalah prosesnya, bukan modelnya).	Kurang sensitif terhadap urutan kata dan konteks kalimat karena pendekatannya berbasis <i>bag-of-words</i> .	Sama seperti SVM, berbasis fitur kata (Count Vectorizer). Namun, mampu menangkap hubungan non-linear yang kompleks antar kata dengan lebih baik daripada model linear tunggal.
Kelebihan	Menyediakan alur kerja yang logis, terstruktur, dan komprehensif untuk penelitian berbasis data.	Cepat untuk dilatih, efisien secara komputasi, dan memiliki performa yang kuat bahkan dengan dataset yang tidak terlalu besar.	Keunggulan signifikannya terletak pada kemampuan memitigasi risiko <i>overfitting</i> yang kerap menjadi kelemahan pada <i>Decision Tree</i> tunggal. Selain itu, algoritma ini menunjukkan resistensi yang kuat terhadap gangguan ( <i>noise</i> ) data, sehingga mampu menghasilkan tingkat akurasi yang sangat

			kompetitif
Kekurangan	Prosesnya cenderung linear dan lebih kaku dibandingkan dengan kerangka kerja iteratif lainnya seperti CRISP-DM.	Performanya bisa lebih rendah pada tugas yang sangat bergantung pada urutan kata dan nuansa kalimat.	Waktu pelatihan bisa menjadi lama jika jumlah pohon ( <i>n_estimators</i> ) sangat banyak. Model yang dihasilkan bisa berukuran besar dan memakan memori.



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA