Terbit *online* pada laman : http://teknosi.fti.unand.ac.id/

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |

Artikel Penelitian

Komparasi Kinerja Algoritma Naive Bayes, SVM, dan Random Forest dalam Analisis Sentimen Aplikasi *Crowdfunding*

Rafika Octaria Ningsih [a], Dwi Rosa Indah [b], Ardina Ariani [c], Mgs. Afriyan Firdaus [d], Rahmat Izwan Heroza [e], M. Husni Syahbani [f]*

[acbdf]*Program Studi Sistem Informasi, Universitas Sriwijaya, Kota Palembang 30129, Indonesia*
[e]*University of Essex, Colchester, United Kingdom*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 27 Februari 2026
Revisi Akhir: 13 Agustus 2026
Diterbitkan *Online*: 23 Agustus 2026

KATA KUNCI

Crowdfunding,
CRISP-DM,
Naïve Bayes,
SVM,
Random Forest

KORESPONDENSI

E-mail: indah812@unsri.ac.id *

A B S T R A C T

Ulasan pengguna pada *platform crowdfunding* merepresentasikan pengalaman, tingkat kepuasan, serta berbagai kendala yang dihadapi pengguna, sehingga memiliki nilai strategis sebagai bahan evaluasi dalam pengembangan kualitas layanan. Penelitian ini difokuskan pada analisis sentimen ulasan pengguna aplikasi Kitabisa sekaligus melakukan perbandingan kinerja algoritma Naive Bayes, Support Vector Machine (SVM), dan Random Forest dengan menerapkan *framework* CRISP-DM. Data penelitian berupa 1.624 ulasan pengguna yang dihimpun dari Google Play Store melalui metode *web scraping*. Pada tahap *data preparation*, data diolah melalui proses pembersihan, normalisasi teks, tokenisasi, penghapusan *stopword, stemming*, serta pembobotan fitur menggunakan TF-IDF. Selanjutnya, untuk mengatasi ketimpangan distribusi kelas sentimen, digunakan *Synthetic Minority Over-sampling Technique* (SMOTE) agar model mampu mempelajari pola data secara lebih seimbang. Proses pemodelan dilakukan dengan membagi dataset ke dalam data latih dan data uji dengan perbandingan 80:20, kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall, F1-score,* serta *k-Fold Cross Validation*. Hasil pengujian mengindikasikan bahwa algoritma SVM menunjukkan performa paling unggul dengan tingkat akurasi pada data uji sebesar 94,51% serta rata-rata akurasi k-Fold sebesar 95,30%, disertai nilai standar deviasi terendah dibandingkan algoritma lainnya. Selain itu, analisis topik berbasis L*atent Dirichlet Allocation* (LDA) memperlihatkan bahwa sentimen positif didominasi oleh topik manfaat aplikasi dan kemudahan dalam berdonasi, sedangkan sentimen negatif umumnya berkaitan dengan kendala teknis, seperti proses verifikasi akun dan gangguan transaksi. Hasil Penelitian ini diharapkan dapat menjadi dasar pertimbangan bagi pengembang dalam meningkatkan kualitas layanan aplikasi Kitabisa.

1. PENDAHULUAN

Perubahan lanskap sosial dan ekonomi global saat ini didorong oleh pesatnya kemajuan teknologi digital. Kemunculan teknologi seperti website, media sosial, dan juga platform online mempermudah aktivitas di berbagai sektor [1] . Salah satunya adalah sektor keuangan yang bertransformasi melalui inovasi *financial technology* (*fintech*) [2]. *Fintech* mengubah transaksi keuangan sekaligus memperluas akses layanan melalui pembayaran digital, *peer-to-peer lending,* dan *crowdfunding* [3]. Di antara model tersebut, *crowdfunding* menonjol sebagai pendanaan alternatif yang kolaboratif, karena memungkinkan individu atau organisasi menggalang dana secara terbuka melalui platform digital [4].

Popularitas global *crowdfunding* dimulai pada tahun 2008 dengan hadirnya dua platform AS, Kickstarter dan Indiegogo, yang kini menjadi yang terbesar. Di Indonesia, adopsi model ini dimulai sejak 2012, ditandai kemunculan berbagai platform lokal [5]. Di antara platform-platform tersebut, Kitabisa.com menonjol sebagai salah satu yang paling sukses dan berhasil mempertahankan eksistensinya hingga kini. Berdiri sejak tahun 2013, platform ini berfokus pada penggalangan dana untuk

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

keperluan sosial, pendidikan, kesehatan, serta bantuan bencana. Selain itu, Yayasan Kitabisa menunjukkan komitmen untuk menciptakan ekosistem gotong royong berkelanjutan dengan mengintegrasikan prinsip *Environmental, Social, and Governance* (ESG) ke dalam strateginya. Komitmen ESG ini diwujudkan melalui penerapan standar pelaporan *Global Reporting Initiative* (GRI) untuk menjamin transparansi distribusi bantuan sosial, pelaksanaan inisiatif lingkungan, dan penguatan tata kelola yang akuntabel serta berkelanjutan. Pada 2024, Yayasan Kitabisa menyalurkan lebih dari Rp830 miliar dari delapan juta donatur dan 400 mitra, dengan dampak penanaman 59.000 pohon, pengelolaan 3.600 ton sampah, serta program air bersih, layanan medis, dan pendidikan. Untuk mendukung pencapaian tersebut, Kitabisa konsisten melaksanakan program-program yang selaras dengan ESG dan mendukung SDGs [6].

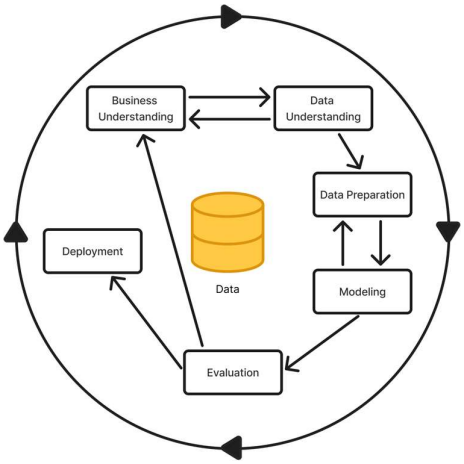
Mengingat besarnya dampak sosial dan skala donasi Kitabisa, penting memahami persepsi dan pengalaman pengguna untuk peningkatan layanan sesuai kebutuhan [7]. Untuk mengukur persepsi tersebut secara komprehensif, dibutuhkan pendekatan analisis sentimen berbasis data mining yang terstruktur [8]. *Framework* yang sesuai untuk mendukung tujuan penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu sebuah metodologi yang secara luas diadopsi sebagai standar industri dalam pelaksanaan proyek *data mining* [9]. Beberapa algoritma klasifikasi teks seperti *Naive Bayes, Support Vector Machine* (SVM), dan *Random Forest* telah banyak diterapkan dalam penelitian serupa karena keunggulannya dalam menangani data ulasan. Selain algoritma, analisis sentimen juga memanfaatkan SMOTE (*Synthetic Minority Oversampling Technique*) untuk menangani ketidakseimbangan kelas serta *k-Fold Cross Validation* guna meningkatkan reliabilitas evaluasi model.

Sejumlah penelitian terdahulu juga telah memanfaatkan algoritma tersebut dalam konteks analisis sentimen pada platform *crowdfunding* dan *fintech*. Surohman et al. [10] membandingkan metode Naive Bayes dengan K-Nearest Neighbor (KNN) pada aplikasi fintech dan menunjukkan bahwa Naive Bayes memiliki tingkat akurasi sebesar 84,76%, lebih tinggi dibandingkan KNN sebesar 82,92%. Selanjutnya, Switrayana et al. [11] menerapkan Support Vector Machine (SVM) dengan metode SMOTE–Tomek Links pada ulasan aplikasi Kitabisa dan memperoleh akurasi sebesar 98%. Indriani dan Wiranata [12] melakukan analisis sentimen pada aplikasi GoPay menggunakan algoritma SVM, Decision Tree, dan Random Forest, dengan hasil menunjukkan bahwa Random Forest memiliki akurasi tertinggi sebesar 90%. Terakhir, Amalsyah et al. [13] menerapkan *framework* CRISP-DM pada aplikasi *fintech* menggunakan algoritma SVM dan memperoleh akurasi sebesar 86%, yang menunjukkan bahwa CRISP-DM mampu mengidentifikasi persepsi pengguna secara lebih terstruktur.

Namun, sebagian besar penelitian masih berfokus pada satu algoritma tanpa membandingkan kinerja beberapa algoritma sekaligus, serta belum sepenuhnya menerapkan *framework CRISP-DM* secara komprehensif dalam konteks *crowdfunding* di Indonesia. Maka dari itu, permasalahan penelitian ini adalah belum adanya kajian komprehensif yang membandingkan beberapa algoritma klasifikasi teks dengan menerapkan CRISP-DM untuk menganalisis sentimen ulasan pengguna, sehingga persepsi dan pengalaman pengguna sebagai dasar pengembangan layanan *crowdfunding* Kitabisa belum teridentifikasi secara optimal. Dengan demikian, penelitian ini melakukan evaluasi komparatif terhadap performa klasifikasi Naive Bayes, Support Vector Machine, dan Random Forest pada ulasan pengguna aplikasi Kitabisa sebagai representasi platform *crowdfunding* di Indonesia..

Berdasarkan uraian yang telah dijelaskan, peneliti memutuskan untuk mengangkat judul "Penerapan *Framework* CRISP-DM untuk Analisis Sentimen Ulasan Aplikasi *Crowdfunding* Kitabisa Menggunakan Naive Bayes, SVM, dan Random Forest." Dengan demikian, penelitian ini diharapkan mampu memberikan kontribusi terhadap pengembangan layanan Kitabisa sebagai platform *crowdfunding* yang lebih baik.

2. METODE



Gambar 1. *Framework CRISP-DM*

2.1. Business Understanding

Tahap *Business Understanding* difokuskan pada perumusan konteks dan sasaran penelitian. Pada tahap ini, permasalahan yang diidentifikasi adalah keberadaan ulasan pengguna aplikasi Kitabisa di Google Play Store yang mengandung berbagai opini positif dan negatif, namun belum diolah secara sistematis untuk memetakan kecenderungan sentimen pengguna. Padahal, ulasan tersebut berpotensi memberikan informasi penting terkait pengalaman dan tingkat kepuasan pengguna terhadap aplikasi.

2.2. Data Understanding

Tahap *Data Understanding* bertujuan mengkaji karakteristik awal data yang diperoleh dari ulasan pengguna aplikasi Kitabisa di Google Play Store melalui teknik *web scraping*. Fitur *rating* dan ulasan pada *platform* tersebut menjadi instrumen penting dalam mengevaluasi kualitas aplikasi, karena mencakup penilaian pengguna dalam skala 1–5 serta teks ulasan yang berisi opini terhadap aplikasi [14].

Jumlah data yang berhasil dikumpulkan secara real-time sebanyak 1.624 ulasan, yang diambil dari rentang waktu tahun 2023 hingga 2025. Pemilihan periode dua tahun ini dianggap cukup untuk mengamati perubahan sentimen pengguna terhadap aplikasi Kitabisa sekaligus tetap relevan dengan kondisi dan fitur

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

aplikasi terkini, sejalan dengan pendapat Izza et al. [15] yang menyatakan bahwa periode dua tahun merupakan durasi yang memadai untuk mengamati perkembangan dan tren dalam ulasan atau opini pengguna.

2.3. Data Preparation

Tahap *Data Preparation* bertujuan mempersiapkan data hasil *web scraping* agar layak digunakan dalam proses analisis [16]. Data mentah yang diperoleh dari Google Play Store umumnya masih mengandung noise, data duplikat, serta ketidakseimbangan kelas, sehingga diperlukan tahapan pembersihan, transformasi, dan penyesuaian data. Tahapan *data preparation* yang diterapkan dalam penelitian ini meliputi:

2.3.1. Data Cleaning

Pada tahap ini dilakukan penghapusan ulasan yang tidak relevan, seperti ulasan yang kosong, duplikat, atau hanya berisi simbol dan emoji tanpa makna. Selain itu, ulasan yang terlalu singkat, misalnya hanya terdiri dari satu kata, juga dieliminasi karena dianggap tidak mampu merepresentasikan opini pengguna secara memadai. Pembersihan data dilakukan guna meningkatkan kelayakan dataset pada tahap analisis.

2.3.2. Text Normalization

Normalisasi dilakukan untuk menyeragamkan format teks ulasan. Tahapan ini meliputi konversi seluruh teks ke huruf kecil (*lowercasing*), penghapusan angka, tanda baca, URL, serta karakter khusus yang tidak relevan. Dengan adanya normalisasi, variasi penulisan yang berbeda dapat diseragamkan sehingga meminimalkan kemungkinan perbedaan makna akibat perbedaan bentuk penulisan. Pada penelitian ini, proses normalisasi dibantu dengan *library IndoNLP* yang digunakan untuk menstandarkan kata-kata informal atau slang menjadi bentuk formalnya. Selain itu, *library* ini juga diintegrasikan untuk meningkatkan akurasi normalisasi terhadap kosakata spesifik domain [17].

2.3.3. Tokenization

Pada tahap tokenisasi, teks ulasan diuraikan menjadi unit kata sebagai dasar pemrosesan lanjutan [18]. Tahapan ini diperlukan agar data teks dapat diolah pada tahap analisis selanjutnya yang berfokus pada level kata. Dengan proses tokenisasi, setiap kata dapat dievaluasi secara terpisah sehingga pengaruhnya terhadap pembentukan sentimen dapat diidentifikasi secara lebih jelas.

2.3.4. Stopword Removal

Tahap ini menghapus kata umum (*stopwords*) yang tidak berkontribusi terhadap pembentukan informasi sentimen, seperti "yang", "dan", "di", serta "ke" [19]. Kata-kata tersebut umumnya hanya berperan sebagai unsur tata bahasa dalam kalimat sehingga tidak berpengaruh terhadap penentuan arah atau polaritas sentimen. Dengan melakukan *stopword removal*, dimensi data dapat dikurangi dan fokus analisis hanya tertuju pada kata-kata yang relevan. Dalam penelitian ini, proses *stopword removal* dilakukan menggunakan *library Sastrawi* yang menyediakan daftar *stopwords* bawaan Bahasa Indonesia, sehingga memudahkan proses pembersihan data.

2.3.5. Stemming

Tahap *stemming* digunakan untuk mengubah kata berimbuhan menjadi bentuk kata dasar [19]. Sebagai contoh, kata "membeli", "dibeli", dan "pembelian" direduksi menjadi kata dasar "beli". Tahap ini menyelaraskan berbagai variasi kata bermakna serupa agar analisis sentimen dapat dilakukan secara lebih efektif. Proses *stemming* pada penelitian ini dilakukan menggunakan library *Sastrawi* yang dirancang khusus untuk Bahasa Indonesia [20]. Penggunaan *Sastrawi* memungkinkan kata-kata berimbuhan diproses secara otomatis menjadi bentuk dasarnya, sehingga menghasilkan representasi teks yang lebih konsisten untuk dianalisis.

2.3.6. Pemberian Label Sentimen

Pada tahap ini, penentuan label sentimen dilakukan berdasarkan nilai rating bintang pada ulasan aplikasi. Ulasan dengan rating 4–5 dikelompokkan ke dalam kelas sentimen positif, sedangkan rating 1–2 dimasukkan ke dalam kelas sentimen negatif. Adapun ulasan dengan rating 3 dikecualikan dari dataset karena dianggap netral dan tidak menunjukkan kecenderungan sentimen pengguna yang tegas [14].

2.3.7. Pembobotan TF-IDF

Proses pembobotan dilakukan menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk menentukan tingkat kepentingan setiap kata dalam dataset ulasan pengguna aplikasi Kitabisa. Teknik ini merepresentasikan teks dalam bentuk vektor numerik agar dapat dimanfaatkan sebagai fitur pada proses klasifikasi dan analisis teks [21]. Dengan demikian, kata yang sering muncul tetapi tidak memiliki makna penting akan memperoleh bobot yang lebih rendah dibandingkan dengan kata yang lebih informatif.

Proses pembobotan dilakukan menggunakan objek *TfidfVectorizer* dari pustaka *sklearn* pada *Python*. Objek ini berfungsi untuk mengonversi sekumpulan teks ulasan menjadi matriks TF-IDF, di mana setiap elemen merepresentasikan bobot suatu kata pada dokumen tertentu. Hasil pembobotan tersebut selanjutnya dimanfaatkan sebagai fitur input pada model klasifikasi sentimen yang diterapkan dalam penelitian ini, yaitu Naive Bayes, Support Vector Machine (SVM), dan Random Forest. Penerapan metode TF-IDF memungkinkan kata-kata yang memiliki keterkaitan kuat dengan sentimen memperoleh bobot yang lebih tinggi, sehingga dapat mendukung peningkatan kinerja model klasifikasi.

2.3.8. Penyeimbangan Data (Data Balancing)

Penelitian ini menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas dalam dataset dengan membuat data sintetis pada kelas yang lebih sedikit. Penggunaan metode ini dilakukan karena data ulasan didominasi oleh sentimen positif, yang dapat menyebabkan bias pada model klasifikasi. Penerapan SMOTE diketahui mampu meningkatkan kinerja dan akurasi model secara keseluruhan [22].

2.4. Modeling

Tahap *modeling* dilakukan dengan menerapkan algoritma machine learning pada dataset yang telah diproses melalui tahap *data preparation*. Dalam penelitian ini, tiga model klasifikasi

digunakan, yaitu Naive Bayes, Support Vector Machine (SVM), dan Random Forest.

Proses pemodelan dilakukan dengan memisahkan dataset ke dalam data latih (*training set*) dan data uji (*testing set*). Selain itu, metode *k-Fold Cross Validation* diterapkan untuk memperoleh hasil evaluasi yang lebih konsisten serta meminimalkan potensi bias akibat pembagian data. Kinerja ketiga algoritma kemudian dianalisis dan dibandingkan guna menentukan model dengan performa terbaik dalam mengklasifikasikan sentimen ulasan aplikasi Kitabisa.

2.5. Evaluation

Tahap *evaluation* bertujuan untuk mengukur kinerja model klasifikasi yang dihasilkan pada tahap *modeling*. Proses evaluasi dilakukan untuk memastikan bahwa model mampu mengklasifikasikan sentimen secara efektif serta sesuai dengan tujuan penelitian. Adapun metrik evaluasi yang digunakan dalam penelitian ini meliputi:

1. Akurasi (*Accuracy)* mengukur perbandingan antara jumlah prediksi yang tepat dengan total data uji.
2. Presisi *(Precision)* menilai kemampuan model dalam mengklasifikasikan data secara akurat ke dalam kelas tertentu dan mengurangi kesalahan prediksi pada kelas positif maupun negatif.
3. *Recall* menunjukkan kemampuan model dalam mengenali semua data yang benar-benar termasuk dalam suatu kelas.
4. Skor F1 *(F1-score)* menilai keseimbangan antara *precision* dan *recall* terutama pada dataset dengan distribusi kelas yang tidak seimbang.

2.6. Deployment

Tahap *deployment* merupakan langkah terakhir dalam metodologi CRISP-DM, yaitu penerapan hasil analisis agar dapat dimanfaatkan secara praktis. Namun, pada penelitian ini tahap deployment tidak diimplementasikan ke dalam sistem produksi karena fokus utama penelitian adalah analisis sentimen menggunakan kerangka kerja CRISP-DM, bukan penerapan model secara *real-time*. Meskipun demikian, hasil prediksi model tetap dianalisis untuk memberikan wawasan strategis terkait persepsi pengguna dan dapat menjadi landasan pengembangan sistem analisis sentimen otomatis di masa mendatang, sehingga mendukung pengambilan keputusan berbasis data [13].

Sebagai bentuk pemanfaatan hasil analisis tersebut, penelitian ini tidak hanya melihat distribusi sentimen positif dan negatif, tetapi juga menggali informasi yang lebih mendalam dari ulasan pengguna. Oleh karena itu, *topic modeling* menggunakan *Latent Dirichlet Allocation* (LDA) diterapkan untuk mengidentifikasi topik utama pada ulasan positif maupun negatif, sehingga memberikan gambaran topik serta kata kunci yang dominan. Hasil analisis topik ini dapat menjadi dasar dalam perbaikan layanan serta pengembangan fitur aplikasi.

3. HASIL

Performa Naïve Bayes, SVM, dan Random Forest berdasarkan akurasi data uji dan rata-rata *K-Fold Cross Validation* ditunjukkan pada tabel berikut:

Tabel 1. Perbandingan Akurasi Model Klasifikasi Sentimen

Model	Akurasi Data Uji	Rata-rata akurasi k-fold	Standar Deviasi K-Fold
Naïve Bayes	0.9176	0.9329	0.0216
SVM	**0.9451**	**0.9530**	**0.0106**
Random Forest	0.9176	0.9256	0.0139

Berdasarkan tabel 1, terlihat bahwa:

1. SVM memiliki akurasi tertinggi pada data uji (94,51%) dan rata-rata akurasi K-Fold (95,30%), serta standar deviasi paling kecil (0,0106). Hal ini menunjukkan bahwa SVM paling stabil dan konsisten dalam mengklasifikasikan sentimen ulasan.
2. Naive Bayes dan Random Forest memiliki akurasi data uji yang sama (91,76%), namun rata-rata K-Fold Random Forest sedikit lebih tinggi (92,56%) dibanding Naive Bayes (93,29%). Perbedaan standar deviasi menunjukkan bahwa Naive Bayes sedikit kurang stabil dibanding SVM, tetapi masih menunjukkan performa yang baik.
3. Secara keseluruhan, ketiga model mampu mengklasifikasikan sentimen ulasan dengan baik, dengan nilai F1-score di atas 0,9 pada kedua kelas. Namun, SVM menjadi model terbaik jika dilihat dari akurasi dan kestabilan performa secara keseluruhan.

Untuk memvisualisasikan kata-kata yang paling sering muncul dalam ulasan, dibuat *wordcloud* terpisah untuk sentimen positif dan negatif, seperti ditunjukkan pada Gambar 2 dan Gambar 3.

Gambar 2. *WordCloud* Sentimen Positif

Gambar 3. *WordCloud* Sentimen Negatif

4. PEMBAHASAN

4.1. Business Understanding

Tahap *Business Understanding* bertujuan memahami konteks dan tujuan analisis. Pada tahap ini, penelitian menilai kecenderungan sentimen pengguna terhadap aplikasi Kitabisa berdasarkan ulasan di Google Play Store untuk memperoleh gambaran kepuasan dan pengalaman pengguna.

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

Analisis sentimen digunakan untuk memproses ulasan pengguna yang subjektif menjadi informasi yang dapat diukur dan digunakan dalam perbaikan aplikasi. Hasilnya diharapkan dapat membantu pihak Kitabisa dalam mengenali kelebihan serta kekurangan layanan dari sudut pandang pengguna. Dari sisi bisnis, tahap ini menegaskan bahwa hasil analisis sentimen dapat dimanfaatkan untuk:

1. Menilai tingkat kepuasan pengguna secara kuantitatif berdasarkan opini di Google Play Store.
2. Mengidentifikasi masalah atau keluhan yang sering disampaikan pengguna.
3. Menjadi dasar bagi pengembang dalam pengambilan keputusan untuk meningkatkan kualitas fitur dan layanan aplikasi.

Dengan demikian, tahap *Business Understanding* memastikan bahwa proses analisis sentimen yang dilakukan relevan dengan kebutuhan pengembangan aplikasi Kitabisa secara berkelanjutan dan berorientasi pada peningkatan kepuasan pengguna.

4.2. Data Understanding

Proses pengumpulan data dilakukan dengan metode *web scraping* terhadap Google Play Store menggunakan Google Colab. Data dikumpulkan dalam rentang waktu Januari 2023 hingga Oktober 2025 dan menghasilkan sebanyak 1.624 ulasan pengguna aplikasi Kitabisa. Dataset yang diperoleh terdiri dari dua atribut utama, yaitu *rating* dan *ulasan*. Atribut *rating* merepresentasikan penilaian pengguna terhadap aplikasi dalam skala 1 sampai 5, sedangkan atribut *ulasan* berisi komentar atau pendapat pengguna terkait pengalaman mereka dalam menggunakan aplikasi Kitabisa.

Tabel 2. Hasil Scraping Data

Rating	Ulasan
3	Bagus tapi tak perlu ada biaya admin, orang ni…
3	Aplikasi yang bermanfaat bagi sesama, tetapi…
1	Saldo yang di top up tidak masuk ke kantong do..
1	Gak jelas… aku plot bayar sedekah subuh utk…
5	Selama beberapa tahun pakai kitabisa, all good

4.3. Data Preparation

4.3.1. Data Cleaning

Proses *data cleaning* dilakukan dengan dua tahapan, yaitu secara manual dan menggunakan *tools* di Google Colab. Pada tahap pertama, pembersihan dilakukan secara manual dengan meninjau kembali setiap data untuk memastikan kesesuaian antara nilai rating dan isi ulasan, serta menghapus data yang tidak relevan atau tidak sesuai konteks. Setelah melalui tahap ini, jumlah data yang tersisa sebanyak 1.497 ulasan.

Selanjutnya, tahap kedua dilakukan secara otomatis menggunakan Google Colab, dengan langkah-langkah seperti menghapus ulasan kosong, mendeteksi dan menghapus data duplikat, serta mengeliminasi ulasan yang terlalu singkat sehingga tidak memiliki makna yang signifikan. Setelah proses pembersihan otomatis ini, jumlah data akhir yang diperoleh adalah 1.116 ulasan.

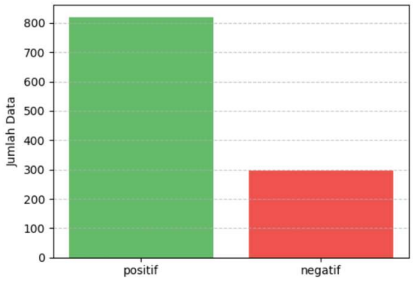
Seluruh ulasan dinormalisasi agar konsisten, mencakup konversi ke huruf kecil dan penghapusan tanda baca, angka, serta karakter non-esensial. Teks kemudian diproses melalui tokenisasi, penghapusan stopword, dan stemming menggunakan *library Sastrawi* di Google Colab untuk memperoleh kata dalam bentuk dasar. Untuk memperjelas hasil tahapan *preprocessing* yang telah dilakukan, berikut ditampilkan salah satu contoh ulasan pengguna yang melalui proses normalisasi, tokenisasi, *stopword removal*, dan *stemming*.

Tabel 3. Hasil *Preprocessing Data*

Tahapan	Hasil
Review	gak jelas... aku plot bayar sedekah subuh utk 2 Minggu, belum seminggu dah habis.... udah beberapa x.....
Text Normalization	enggak jelas aku plot bayar sedekah subuh untuk minggu belum seminggu deh habis sudah beberapa….
Tokenization	[enggak, jelas, aku, plot, bayar, sedekah, subuh, untuk, minggu, belum, seminggu, deh, habis, sudah, beberapa…]
Stopword Removal	[enggak, jelas, aku, plot, bayar, sedekah, subuh, minggu, seminggu, habis, beberapa,]
Stemming	[enggak, jelas, aku, plot, bayar, sedekah, minggu, minggu, deh, habis, beberapa…]

4.3.2. Pemberian Label Sentimen

Pelabelan sentimen berdasarkan rating yang diberikan, didapatkan dengan hasil sentiment positif sebenyak 820 dan negatif sebanyak 297, dan untuk data netral yang dihapus sebanyak 97 ulasan.



Gambar 4. Pelabelan Sentimen

4.3.3. Hasil Pembobotan TF-IDF

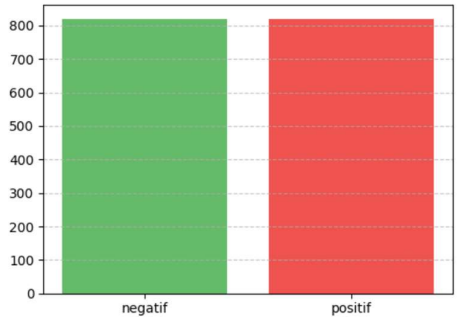
Sebelum dilakukan penyeimbangan, data teks diubah menjadi representasi numerik menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) dengan maksimum 5.000 fitur. TF-IDF mengubah setiap dokumen menjadi vektor angka yang mencerminkan bobot kata-kata dalam dokumen, sehingga model dapat memproses data teks secara lebih efektif dengan memanfaatkan *TfidfVectorizer* dari *library scikit-learn*. Hasil dari tahap ini berupa representasi data dalam bentuk matriks sparse, di mana setiap nilai menunjukkan bobot TF-IDF suatu kata pada masing-masing ulasan. Data hasil pembobotan inilah yang menjadi masukan untuk proses pelatihan model klasifikasi Naive Bayes, SVM, dan Random Forest pada tahap berikutnya.

	feature	tfidf	document
0	hamba	0.548647	0
1	saudara	0.490575	0
2	tolong	0.473732	0
3	senantiasa	0.280491	0
4	sebut	0.236866	0
...	...	...	...
5028	gbsa	0.691398	1115
5029	mulu	0.302457	1115
5030	kok	0.230438	1115
5031	dong	0.228429	1115
5032	sih	0.226503	1115

Gambar 5. Hasil Pembobotan TF-IDF

4.3.4. Penyeimbangan Data dengan SMOTE

Distribusi kelas pada label sentimen diperiksa dan ditemukan ketidakseimbangan antara kelas positif (820 sampel) dan negatif (297 sampel), yang dapat memengaruhi performa model. SMOTE kemudian diterapkan pada fitur TF-IDF untuk menambah sampel sintetis pada kelas minoritas, sehingga distribusi kelas menjadi seimbang dan siap untuk tahap pemodelan.



Gambar 6. Distribusi label setelah SMOTE

4.4. Modeling

4.4.1. Naive Bayes

Data hasil pembobotan TF-IDF dan penyeimbangan dengan SMOTE dimanfaatkan untuk membangun model klasifikasi sentimen menggunakan *Multinomial Naive Bayes*. Dataset dibagi menjadi 80% data latih dan 20% data uji, kemudian model dilatih dan diuji untuk menghasilkan prediksi sentimen.

4.4.2. Support Vector Machine

Data TF-IDF yang sudah diseimbangkan melalui SMOTE digunakan untuk membangun model prediksi sentimen dengan *LinearSVC*. *LinearSVC* merupakan salah satu varian dari *Support Vector Classification* (SVC) yang menerapkan kernel linear untuk memisahkan kelas sentimen positif dan negatif melalui sebuah *hyperplane* linear pada ruang fitur berbasis TF-IDF. Model ini digunakan untuk melakukan prediksi sentimen pada dataset yang telah melalui proses penyeimbangan kelas.

4.4.3. Random Forest

Model Random Forest dibangun dengan menerapkan pendekatan *ensemble learning* yang mengombinasikan sejumlah *decision tree* untuk meningkatkan akurasi serta meminimalkan risiko *overfitting*. Dalam penelitian ini, proses pelatihan dilakukan menggunakan 200 pohon keputusan (*n_estimators* = 200) pada data latih, kemudian model tersebut digunakan untuk menghasilkan prediksi sentimen pada data uji.

4.5. Evaluasi

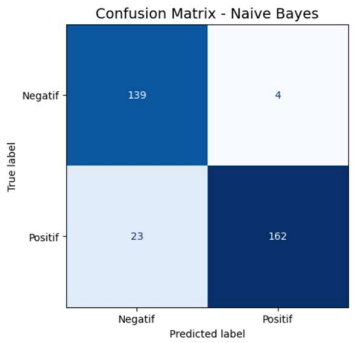
4.5.1. Naive Bayes

Hasil evaluasi model Naive Bayes pada data uji berdasarkan metrik akurasi, presisi, recall, dan F1-score dapat dilihat pada gambar berikut.

```
=== HASIL EVALUASI NAIVE BAYES ===

Akurasi: 0.9176829268292683

Classification Report:
              precision    recall  f1-score   support

     negatif       0.86      0.97      0.91       143
     positif       0.98      0.88      0.92       185

    accuracy                           0.92       328
   macro avg       0.92      0.92      0.92       328
weighted avg       0.92      0.92      0.92       328
```

Gambar 7. Hasil Pemodelan Naive bayes

Sebagaimana ditunjukkan pada Gambar 7, model mencapai *precision* dan *recall* yang tinggi pada kedua kelas, dengan F1-*score* lebih dari 0,9, menandakan kemampuan Naive Bayes dalam mengklasifikasikan ulasan positif maupun negatif secara efektif. Sebagai pelengkap evaluasi metrik, Gambar 8 menyajikan confusion matrix yang memperlihatkan distribusi prediksi benar dan salah pada kelas positif dan negatif



Gambar 8. *Confusion Matrix* Naive bayes

Dari matriks ini, dapat dilihat bahwa dari 143 ulasan negatif, 139 diklasifikasikan dengan benar, sedangkan 4 ulasan salah dikategorikan sebagai positif. Sedangkan dari 185 ulasan positif, 162 diklasifikasikan dengan benar dan 23 salah dikategorikan sebagai negatif.

Selanjutnya, kestabilan model dievaluasi melalui K-Fold Cross Validation dengan k = 10. Hasil evaluasi dapat dilihat pada gambar berikut.

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

```
=== HASIL K-FOLD CROSS VALIDATION (Naive Bayes) ===

Fold 1: 0.9207
Fold 2: 0.9146
Fold 3: 0.9573
Fold 4: 0.9451
Fold 5: 0.9329
Fold 6: 0.8902
Fold 7: 0.9512
Fold 8: 0.9390
Fold 9: 0.9634
Fold 10: 0.9146

Rata-rata Akurasi: 0.9329
Standar Deviasi  : 0.0216
```

Gambar 9. Hasil *K-fold Cross Validation* NB

Nilai akurasi tertinggi tercatat pada fold ke-9 sebesar 0,9634, sedangkan akurasi terendah ditemukan pada fold ke-6 dengan 0,8902. Rata-rata akurasi keseluruhan mencapai 93,29% dengan standar deviasi 0,0216, menandakan bahwa model bekerja secara konsisten dan mampu mengklasifikasikan sentimen ulasan pengguna secara efektif.

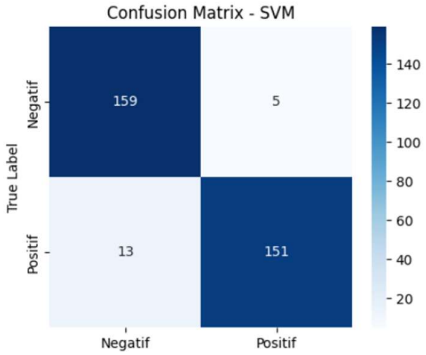
4.5.2. Support Vector Machine

Evaluasi performa model Support Vector Machine (SVM) pada data uji berdasarkan metrik akurasi, presisi, recall, dan F1-score disajikan pada Gambar 10.

```
=== HASIL EVALUASI SVM ===
Accuracy: 0.9451219512195121

Classification Report:
              precision    recall  f1-score   support

     negatif       0.92      0.97      0.95       164
     positif       0.97      0.92      0.94       164

    accuracy                           0.95       328
   macro avg       0.95      0.95      0.95       328
weighted avg       0.95      0.95      0.95       328
```

Gambar 10. Hasil Pemodelan SVM

Precision, recall, dan *F1-score* di atas 0,9 menunjukkan kemampuan model dalam mengklasifikasikan ulasan positif dan negatif dengan akurat. Untuk memvisualisasikan kinerja klasifikasi secara lebih mendetail, Gambar 11 menampilkan *confusion matrix* dari model SVM, termasuk prediksi yang tepat dan yang keliru pada masing-masing kelas.



Gambar 11. *Confusion Matrix* SVM

Matriks menunjukkan bahwa dari 164 ulasan negatif, 159 diklasifikasikan dengan benar, sementara 5 ulasan salah ditempatkan ke kelas positif. Dari 164 ulasan positif, 151 diklasifikasikan secara akurat dan 13 ulasan salah masuk kelas negatif. Secara keseluruhan, model SVM mencapai akurasi 94,51%, mencerminkan performa yang baik dengan tingkat kesalahan prediksi yang rendah pada masing-masing kelas.

```
=== HASIL K-FOLD CROSS VALIDATION (SVM

Fold 1: 0.9329
Fold 2: 0.9573
Fold 3: 0.9695
Fold 4: 0.9512
Fold 5: 0.9634
Fold 6: 0.9390
Fold 7: 0.9573
Fold 8: 0.9573
Fold 9: 0.9573
Fold 10: 0.9451

Rata-rata Akurasi: 0.953
Standar Deviasi  : 0.0106
```

Gambar 12. Hasil *K-fold Cross Validation* SVM

Nilai akurasi tertinggi tercatat pada fold ke-3 sebesar 0,9695, sedangkan fold ke-1 menghasilkan akurasi terendah sebesar 0,9329. Rata-rata akurasi sebesar 95,3% dengan standar deviasi 0,0106 menunjukkan bahwa model bekerja secara konsisten.

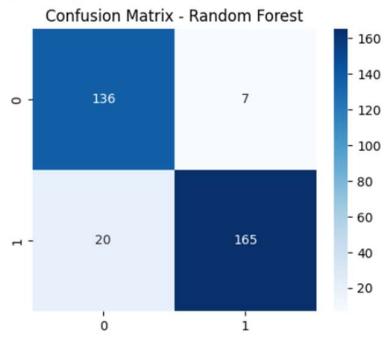
4.5.3. Random Forest

Model Random Forest dievaluasi pada data uji melalui metrik akurasi, presisi, recall, dan F1-score, sebagaimana ditunjukkan pada Gambar 13.

```
Akurasi: 0.9176829268292683

Classification Report:
              precision    recall  f1-score   support

     negatif       0.87      0.95      0.91       143
     positif       0.96      0.89      0.92       185

    accuracy                           0.92       328
   macro avg       0.92      0.92      0.92       328
weighted avg       0.92      0.92      0.92       328
```

Gambar 13. Hasil Pemodelan Random Forest

Nilai precision dan recall yang tinggi pada kedua kelas, disertai F1-score lebih dari 0,9 seperti ditunjukkan pada Gambar 13, mengindikasikan bahwa model mampu mengklasifikasikan ulasan positif dan negatif secara andal. Untuk memberikan gambaran yang lebih rinci mengenai prediksi model, *confusion matrix* ditampilkan pada Gambar 14, menunjukkan jumlah ulasan yang diklasifikasikan dengan benar maupun yang salah untuk masing-masing kelas.



Gambar 14. *Confusion Matrix* Random Forest

Matriks menunjukkan bahwa dari 143 ulasan negatif, 136 diklasifikasikan dengan benar, sementara 7 ulasan salah masuk kelas positif. Dari 185 ulasan positif, 165 diklasifikasikan secara akurat, dan 20 ulasan salah ditempatkan pada kelas negatif. Secara keseluruhan, model Random Forest mencapai akurasi 91,76%, menandakan performa yang baik dengan tingkat kesalahan prediksi yang rendah pada masing-masing kelas.

```
=== HASIL K-FOLD CROSS VALIDATION (Random Forest

Fold 1: 0.9207
Fold 2: 0.9085
Fold 3: 0.9451
Fold 4: 0.9268
Fold 5: 0.9268
Fold 6: 0.9390
Fold 7: 0.9268
Fold 8: 0.9390
Fold 9: 0.9268
Fold 10: 0.8963

Rata-rata Akurasi: 0.9256
Standar Deviasi  : 0.0139
```

Gambar 15. Hasil *K-fold Cross Validation* Random Forest

Nilai rata-rata akurasi sebesar 92,56% dengan standar deviasi 0,0139 mengindikasikan bahwa model mempertahankan kinerja yang konsisten dan stabil dalam klasifikasi sentimen ulasan pengguna.

4.6. Deployment

Pada penelitian ini, tahap *deployment* tidak dilakukan dalam bentuk penerapan model ke dalam sistem produksi, melainkan difokuskan pada pemanfaatan hasil prediksi untuk analisis. Model *Support Vector Machine* (SVM) dengan performa terbaik digunakan untuk mengklasifikasikan sentimen ulasan aplikasi Kitabisa, yang kemudian diinterpretasikan untuk memahami kecenderungan opini pengguna dan isu yang muncul.

Data uji yang digunakan sebesar 20% dari total data yang telah melalui tahap *data preparation*. Selanjutnya, analisis topik menggunakan *Latent Dirichlet Allocation* (LDA) pada setiap kelompok sentimen untuk mengidentifikasi topik utama berdasarkan kata kunci dominan, kemudian hasilnya divalidasi melalui expert judgement untuk memastikan relevansi dengan konteks ulasan.

Tabel 4. Topic Modelling Sentimen Positif

Topik	Kata Kunci	Presentase
Manfaat Aplikasi dan Kemudahan dalam Beramal	sangat, aplikasi, orang, baik, moga, bantu, manfaat, mudah, bagus, jadi	36.36%
Kemudahan Donasi dan Bantuan melalui Aplikasi	aplikasi, bantu, mudah, sedekah, alhamdulillah, salur, kasih, donasi, butuh, benar	27.27%
Pengalaman Positif Pengguna dengan Fitur Donasi	donasi, aplikasi, sedekah, baik, moga, kitabisa, jadi, waktu, banget, terimakasih	16.97%
Doa dan Harapan Pengguna terhadap Aplikasi	moga, selalu, amin, allah, berkah, semua, beri, kasih, amanah, terima	12.73%
Fitur Otomatis dan Tambahan untuk Mempermudah Donasi	jadi, otomatis, sedekah, donasi, lebih, butuh, apa, fitur, program, tambah	6.67%

Analisis topik menggunakan LDA mengelompokkan ulasan positif ke beberapa topik utama yang mencerminkan pengalaman dan pandangan pengguna terhadap aplikasi. Setiap topik merepresentasikan fokus tertentu dan disertai persentase jumlah ulasan yang termasuk di dalamnya. Temuan pengelompokan ulasan positif ini telah divalidasi melalui *expert judgement* dan disajikan sebagai berikut:

1. Manfaat Aplikasi dan Kemudahan bagi Pengguna (36.36%)
   Topik ini merupakan topik dominan pada ulasan positif pengguna. Kata kunci yang muncul mayoritas menggambarkan penilaian positif terhadap manfaat aplikasi, seperti kemudahan dan kebaikan yang dirasakan pengguna. Berdasarkan hasil validasi *expert judgement*, topik ini dinyatakan sesuai karena kata-kata yang dihasilkan relevan dengan manfaat aplikasi. Meskipun terdapat kata yang bersifat umum seperti "moga" dan "jadi", kata tersebut tetap dianggap relevan karena sering muncul dalam konteks ulasan positif pengguna.
2. Kemudahan Donasi dan Bantuan melalui Aplikasi (27.27%)
   Topik ini menggambarkan pengalaman pengguna dalam melakukan donasi melalui aplikasi. Kata kunci seperti "donasi", "sedekah", "bantu", dan "mudah" menunjukkan bahwa proses donasi dinilai praktis dan membantu. Hasil *expert judgement* menyatakan bahwa topik ini sesuai, karena kata-kata yang muncul berkaitan langsung dengan aktivitas donasi. Selain itu, keberadaan unsur religius seperti "alhamdulillah" dianggap wajar dalam konteks keberhasilan setelah berdonasi
3. Pengalaman Positif Pengguna dengan Fitur Donasi (16.97%)
   Topik ini mencerminkan kepuasan pengguna terhadap fitur donasi yang tersedia dalam aplikasi. Kata kunci seperti "terimakasih", "banget", dan "baik" menunjukkan adanya apresiasi terhadap layanan yang diberikan. Berdasarkan validasi *expert judgement*, seluruh kata kunci pada topik ini dinilai relevan dan mampu menggambarkan pengalaman positif pengguna dalam menggunakan fitur donasi pada aplikasi Kitabisa
4. Doa dan Harapan Pengguna terhadap Aplikasi (12.73%)
   Berdasarkan hasil *expert judgement*, doa dan harapan tersebut umumnya merupakan ekspresi positif yang muncul terlepas dari sentimen pengguna terhadap objek aplikasi
5. Fitur Otomatis dan Tambahan untuk Mempermudah Donasi (6.67%)
   Berdasarkan hasil validasi *expert judgement*, topik ini dinilai sebagian sesuai. Kata kunci yang dihasilkan sebagian menggambarkan topik pengembangan fitur, namun dinilai agak sulit untuk menentukan sentimen hanya berdasarkan kata kunci tersebut tanpa melihat konteks ulasan secara keseluruhan.

Tabel 5. Topic Modelling Sentimen Negatif

Topik	**Kata Kunci**	**Presentase**
Kendala Verifikasi dan Masuk Aplikasi	donasi, jadi, enggak, aplikasi, padahal, mau, verifikasi, masuk, update, buat	28.81%
Masalah Penyaluran Donasi dan Waktu Transfer	sedekah, donasi, enggak, subuh, nya, dana, buat, jam, lewat, kalo	25.42%
Kesulitan Pengguna dalam Menerima Donasi	donasi, mau, aplikasi, terima, buat, galang, pakai, dana, enggak, baru	22.03%
Gangguan dan Error Saat Transfer Dana	donasi, lewat, dana, kok, aplikasi, mau, uang, pinjam, eror, banyak	13.55%
Aplikasi Susah Digunakan dan Sering Error	aplikasi, nya, kata, kitabisa, susah, terus, error, benar, kayak, apk	10.16%

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

Berdasarkan hasil analisis topik menggunakan metode *Latent Dirichlet Allocation* (LDA), ulasan negatif pengguna dikelompokkan ke dalam beberapa topik utama. Setiap topik mencerminkan kendala, keluhan, atau pengalaman kurang memuaskan pengguna terhadap aplikasi, disertai persentase jumlah ulasan yang masuk dalam masing-masing topik. Berikut pengelompokan ulasan negatif yang telah divalidasi *melalui expert judgement*:

1. Kendala Verifikasi dan Masuk Aplikasi (28.81%)
   Topik ini menggambarkan kendala yang dialami pengguna saat melakukan verifikasi dan masuk ke aplikasi. Berdasarkan hasil *expert judgement*, topik ini dinilai sebagian sesuai. Kata kunci seperti "verifikasi", "masuk", "aplikasi", "enggak", dan "update" relevan dalam menggambarkan permasalahan login.
2. Masalah Penyaluran Donasi dan Waktu Transfer (25.42%)
   Topik ini berkaitan dengan keluhan pengguna terhadap penyaluran donasi dan waktu transfer dana. Berdasarkan hasil validasi *expert judgement*, topik ini dinilai sebagian sesuai. Kata kunci seperti "donasi", "dana", "jam", "lewat", dan "subuh" relevan dalam menggambarkan keterlambatan atau permasalahan waktu penyaluran dana.
3. Kesulitan Pengguna dalam Menerima Donasi (22.03%)
   Topik ini mencerminkan kesulitan pengguna dalam menerima donasi melalui aplikasi. Berdasarkan hasil *expert judgement*, topik ini dinyatakan sesuai, karena kata kunci seperti "terima", "galang", dan "dana" relevan dalam konteks kegagalan atau kendala menerima donasi. Adapun kata umum seperti "buat" dan "baru" tetap dapat muncul sebagai bagian dari struktur kalimat keluhan pengguna.
4. Gangguan dan Error Saat Transfer Dana (13.56%)
   Topik ini menggambarkan gangguan teknis yang dialami pengguna saat melakukan transfer dana. Berdasarkan hasil validasi *expert judgement*, topik ini dinilai sebagian sesuai. Kata kunci seperti "dana", "eror", "aplikasi", dan "lewat" relevan dalam menggambarkan masalah transfer.
5. Aplikasi Susah Digunakan dan Sering Error (10.16%)
   Topik ini berkaitan dengan pengalaman pengguna yang merasa kesulitan dalam menggunakan aplikasi serta sering mengalami error. Berdasarkan hasil *expert judgement*, topik ini dinilai sebagian sesuai. Kata kunci utama seperti "susah" dan "error" relevan dalam menggambarkan pengalaman negatif pengguna

5. KESIMPULAN

Penerapan *framework CRISP-DM* berhasil memfasilitasi proses analisis sentimen ulasan pengguna Kitabisa secara terstruktur, mulai dari *business understanding*, *data preparation, modeling* hingga evaluasi. Seluruh tahapan berjalan optimal dan menghasilkan data siap analisis setelah melalui preprocessing lengkap serta penyeimbangan data menggunakan SMOTE yang menghasilkan distribusi kelas positif 820 dan negatif 820 ulasan.

Model *Support Vector Machine* (SVM) menjadi algoritma terbaik dalam penelitian ini, dengan akurasi data uji sebesar 94,51% serta rata-rata akurasi K-Fold sebesar 95,30%, dan standar deviasi terendah (0,0106), menunjukkan performa paling stabil dan konsisten dibandingkan Naive Bayes dan Random Forest.

Analisis topik menggunakan LDA mengungkapkan beberapa topik utama pada sentimen positif, seperti manfaat aplikasi, kemudahan donasi, pengalaman positif pengguna, serta harapan agar aplikasi tetap amanah dan bermanfaat. Sementara itu, sentimen negatif didominasi oleh kendala teknis, khususnya verifikasi akun, gangguan saat proses donasi, keterlambatan transfer, serta error pada aplikasi.

DAFTAR PUSTAKA

[1] R. A. D. Zaki, A., Rukmanti, A. A., Syah, D. K. A., Yuliana, E., & Putri, "Implementasi Relationship Marketing pada Digital Crowdfunding (Studi Kasus: Kitabisa.com)," *ATRABIS J. Adm. Bisnis*, vol. 10, no. 2, pp. 249–263, 2024, doi: 10.38204/atrabis.v10i2.2078.

[2] N. F. Tsakila, M. A. Wirahadi, A. A. Fadilah, and H. Simanjuntak, "Analisis Dampak Fintech terhadap Kinerja dan Inovasi Perbankan di Era Ekonomi Digital," *Indones. J. Law Justice*, vol. 1, no. 4, p. 11, 2024, doi: 10.47134/ijlj.v1i4.2787.

[3] D. A. Aini, Q., & Fadilla, "Peran Financial Technology (FinTech) dalam Meningkatkan Inklusi Keuangan di Indonesia," *J. Kolaboratif Sains*, vol. 8, no. 1, pp. 928–936, 2025, [Online]. Available: https://jurnal.unismuhpalu.ac.id/index.php/JKS

[4] F. Shalihah, *Equity crowdfunding di Indonesia*. UAD PRESS, 2022.

[5] R. Edward, M. Y., Ismanto, H., Fu'ad, E. N., Atahau, A. D. R., *Crowdfunding di Indonesia*, vol. 32, no. 3. UNISNU Press, 2021. [Online]. Available: https://unisnupress.unisnu.ac.id/assets/media/crowdfunding-di-indonesia.pdf

[6] Y. Kitabisa, "Tentang Kita bisa." Accessed: Sep. 09, 2025. [Online]. Available: https://kitabisa.com/about-us

[7] Y. A. Putrima and R. Sayekti, "Persepsi pengguna dalam menggunakan mesin drop box mandiri di Perpustakaan Digital Universitas Negeri Medan," *Pustaka Karya J. Ilm. Ilmu Perpust. dan Inf.*, vol. 13, no. 1, pp. 147–157, 2025, doi: 10.18592/pk.v13i1.14589.

[8] D. A. Harjita, P. W., Putra, F. M., & Tyas, "Sentimen Analisis Pengguna Media Sosial Berdasarkan Metode Ekstraksi Fitur dan Klasifikasi," *J. Ilmu Komput.*, vol. 16, no. 2, pp. 88–96, 2023.

[9] M. F. Baskara, A. A., Piranti, N. M., & Romdendine, "Framework Data Mining: Sebuah Survei," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4886–4895, 2025, doi: 10.36040/jati.v9i3.13803.

[10] F. F. Surohman, Aji, S., Rousyati, & Wati, "Analisa Sentimen Terhadap Review Fintech

Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *J. Teknol. Inf. dan Komput.*, vol. 8, no. 1, pp. 93–105, 2020, doi: 10.36002/jutik.v8i1.1582.

[11] A. Switrayana, I. N., Ashadi, D., Hairani, & Aminuddin, "Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods," *Int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, pp. 81–91, 2023, doi: 10.30812/ijecsa.v2i2.3406.

[12] Indriani and A. Davy Wiranata, "Perbandingan Tingkat Akurasi Algoritma Svm, Decision Tree,Dan Random Forest Dalam Analisis Sentimen Tanggapan Pengguna Aplikasi Gopay," *J. Tek. Inform.*, vol. 5, no. 3, pp. 777–787, 2024, [Online]. Available: https://doi.org/10.52436/1.jutif.2024.5.3.1885

[13] M. R. Amalsyah, D. Kurniawan, A. Rifai, and P. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Fintech menggunakan Framework CRISP-DM dalam Penentuan Prioritas Pengembangan Produk," *Sist. J. Sist. Inf.*, vol. 14, no. 2, pp. 813–825, 2025, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[14] P. R. Sari, D. R. Indah, E. Rasywir, M. A. Firdaus, and G. Athalina, "Comparison of Naive Bayes and SVM Algorithms for Sentiment Analysis of PUBG Mobile on Google Play Store," *Sistemasi*, vol. 13, no. 6, p. 2767, 2024, doi: 10.32520/stmsi.v13i6.4814.

[15] S. Izza, N. N., Sari, M., Kahila, M., & Al-Ayubi, "A Twitter Sentimen Analysis on Islamic Banking Using Drone Emprit Academic (DEA): Evidence from Indonesia," *J. Ekon. Syariah Teor. dan Terap.*, vol. 10, no. 5, pp. 496–510, 2023, doi: 10.20473/vol10iss20235pp496-510.

[16] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, N. W. Paton, and R. Sakellariou, "Data Preparation: A Technological Perspective and Review," *SN Comput. Sci.*, vol. 4, no. 4, pp. 1–20, 2023, doi: 10.1007/s42979-023-01828-8.

[17] C. Suhaeni, L. Nissa, A. Mualifah, and H. Wijayanto, "LDA Topic Modeling Analysis of Public Discourse on Indonesia's Free Nutritious Meals Program (MBG)," *Int. J. Informatics Dev.*, vol. 14, no. 1, pp. 587–600, 2025, doi: 10.14421/ijid.2025.5211.

[18] F. N. Budiman, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Ulasan Aplikasi CapCut Menggunakan Model RoBERTa Dengan Fitur Ekstraksi Word2vec," *J. Algoritm.*, vol. 22, no. 2, pp. 358–369, 2025, doi: 10.33364/algoritma/v.22-2.2480.

[19] S. Khariroh *et al.*, "Implementation of LSA for Topic Modeling on Tweets with the Keyword 'Kemenkeu,'" *Sinkron*, vol. 9, no. 1, pp. 129–148, 2025, doi: 10.33395/sinkron.v9i1.14309.

[20] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *J. Transform.*, vol. 20, no. 2, pp. 1–12, 2023, doi: 10.26623/transformatika.v20i2.5374.

[21] G. H. Setiawan and I. M. B. Adnyana, "Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 252–257, 2023, doi: 10.30871/jaic.v7i2.6527.

[22] M. Sulistiyono, Y. Pristyanto, S. Adi, and G. Gumelar, "Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi," *Sistemasi*, vol. 10, no. 2, p. 445, 2021, doi: 10.32520/stmsi.v10i2.1303.

BIODATA PENULIS

Rafika Octaria Ningsih, Mahasiswa jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Dwi Rosa Indah, Dosen Fakultas Ilmu Komputer Universitas Sriwijaya, Program Studi Sistem Informasi, dengan disiplin konsentrasi pada bidang Sistem dan Teknologi Informasi.

Ardina Ariani, Dosen Fakultas Ilmu Komputer Universitas Sriwijaya, Program Studi Sistem Informasi, dengan disiplin konsentrasi pada bidang Data Mining.

Mgs. Afriyan Firdaus, Dosen Fakultas Ilmu Komputer Universitas Sriwijaya, Program Studi Sistem Informasi, dengan disiplin konsentrasi pada bidang Sistem dan Teknologi Informasi.

https://doi.org/10.25077/TEKNOSI.v12i2.2026.231-241

Rahmat Izwan Heroza, PhD candidate di University of Essex, United Kingdom. Fokus penelitiannya pada deep learning for medical imaging.

Husni Syahbani, Dosen Fakultas Ilmu Komputer Universitas Sriwijaya, Program Studi Sistem Informasi, dengan disiplin konsentrasi pada bidang Pemrograman Mobile dan Sistem Informasi.