
Analisis Sentimen Pelanggan Terhadap Aplikasi Investasi di Indonesia

Felicia Penny Septiara, Rizal Ahmad Fauzi

IPB University, Indonesia

Email: fpseptiara@gmail.com, rizal.afauzi@gmail.com

Kata kunci:

Analisis Sentimen,
Aplikasi Investasi,
Google Play Store,
Machine Learning, TF-
IDF, Fintech

Keywords:

Sentiment Analysis,
Investment
Applications, Google
Play Store, Machine
Learning, TF-IDF,
Fintech

ABSTRAK

Penelitian ini mengevaluasi sentimen dalam ulasan pengguna terhadap lima aplikasi investasi di Indonesia Bibit, Ajaib, Bareksa, Stockbit, dan Pluang dengan pendekatan pembelajaran mesin. Data diperoleh dari Google Play Store dan dianalisis menggunakan algoritma supervised learning seperti Naïve Bayes, SVM, Random Forest, XGBoost, LightGBM, serta metode ensemble Soft Voting dengan representasi fitur TF-IDF. Evaluasi dilakukan melalui 10-Fold Cross-Validation dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa model Soft Voting memiliki performa terbaik, sementara regresi logistik dan SVM juga menunjukkan kinerja yang seimbang. Secara keseluruhan, semua aplikasi memperoleh lebih banyak ulasan positif dibandingkan netral maupun negatif. Ajaib menonjol sebagai aplikasi dengan persepsi paling positif, baik dari segi volume maupun dominasi sentimen, sedangkan Bareksa menghadapi tantangan akibat tingginya proporsi ulasan negatif dan netral. Bibit, Pluang, dan Stockbit menunjukkan persebaran sentimen yang relatif seimbang, namun masih harus memperhatikan isu-isu teknis yang sering dikeluhkan. Ulasan umumnya ditulis dengan gaya bahasa kasual, menunjukkan bahwa mayoritas pengguna berasal dari kalangan non-profesional. Implikasinya, pengembang perlu memperkuat keunggulan seperti kemudahan penggunaan, kecepatan, dan fitur relevan, sambil memperbaiki masalah teknis. Keterbatasan studi ini mencakup hanya mencakup ulasan berbahasa Indonesia dari satu platform, sehingga potensi perspektif dari platform atau bahasa lain terabaikan. Penelitian ini memberikan kontribusi bagi studi fintech, pengembangan UX aplikasi, perbandingan model ML untuk teks bahasa Indonesia, serta masukan kebijakan berbasis keluhan pengguna.

This research evaluates user sentiment in reviews of five Indonesian investment applications - Bibit, Ajaib, Bareksa, Stockbit, and Pluang - using a machine learning approach. Data was collected from Google Play Store and analyzed using supervised learning algorithms including Naïve Bayes, SVM, Random Forest, XGBoost, LightGBM, as well as a Soft Voting ensemble method with TF-IDF feature representation. Evaluation was conducted through 10-Fold Cross-Validation using accuracy, precision, recall, and F1-score metrics. The results indicate that the Soft Voting model achieved the best performance, while logistic regression and SVM also demonstrated balanced performance. Overall, all applications received more positive reviews compared to neutral or negative ones. Ajaib stood out as the application with the most positive perception, both in terms of volume and sentiment dominance, while Bareksa faced challenges due to its high proportion of negative and neutral reviews. Bibit, Pluang, and Stockbit showed relatively balanced sentiment distributions, though they still need to address frequently reported technical issues. Reviews were generally written in casual language, suggesting that most users come from non-professional backgrounds. These findings imply that developers should strengthen advantages such

as ease of use, speed, and relevant features while improving technical issues. A limitation of this study is that it only covers Indonesian-language reviews from a single platform, potentially overlooking perspectives from other platforms or languages. This research contributes to fintech studies, investment app UX development, comparison of ML models for Indonesian text analysis, and provides user complaint-based policy insights.

Manuscript accepted: Date

Revised: Date

Date of publication: Date

Ini adalah artikel akses terbuka di bawah lisensi [CC BY-SA](#).

This is an open access article under the [CC BY-SA](#) license.

PENDAHULUAN

Layanan keuangan digital berkembang pesat seiring dengan meningkatnya adopsi teknologi finansial (*fintech*). Perkembangan teknologi finansial di Indonesia telah membawa transformasi besar dalam cara masyarakat mengelola keuangan pribadi, khususnya dalam bidang investasi. Dalam satu dekade terakhir, *fintech* telah merevolusi akses ke layanan keuangan yang sebelumnya hanya tersedia bagi kalangan tertentu, menjadi lebih inklusif dan *user-friendly* (Yuliana et al., 2022). Aplikasi investasi seperti Bibit, Ajaib, Bareksa, Stockbit, dan Pluang menjadi contoh nyata dari kemajuan ini. Aplikasi-aplikasi ini menawarkan layanan investasi reksa dana, saham, emas digital, hingga aset kripto yang dapat diakses hanya melalui ponsel pintar. Hal ini menjadikan investasi tidak lagi eksklusif bagi kalangan berpendidikan tinggi atau berpengalaman dalam pasar modal, tetapi juga bagi generasi muda dan masyarakat awam.

Jumlah aplikasi investasi digital Indonesia terus meningkat seiring tingginya minat masyarakat terhadap instrumen keuangan modern (Sapulette & Muchtar, 2023). Pertumbuhan pengguna aplikasi investasi meningkat seiring dengan tren digitalisasi dan literasi keuangan yang berkembang di kalangan generasi milenial dan Gen Z. Strategi pemasaran yang adaptif terhadap karakteristik generasi ini, ditambah dengan desain antarmuka yang intuitif, edukatif, dan fitur robo advisor, menjadikan aplikasi-aplikasi tersebut semakin populer (Zhang et al., 2018). Persaingan di sektor ini tentunya semakin ketat sehingga para pengembang aplikasi dituntut untuk menghadirkan nilai pembeda baik dari sisi penggunaan, ragam produk yang ditawarkan, hingga kualitas layanan pelanggan (Majumder, 2025). Pengguna akan semakin selektif dalam pemilihan aplikasi dimana unsur keamanan menjadi aspek yang tidak kalah penting seperti parameter sudah atau belumnya aplikasi terdaftar di Otoritas Jasa Keuangan (OJK). Oleh karena itu diperlukan analisa lanjutan mengenai bagaimana feedback dari pengguna terkait aplikasi yang digunakan dalam berinvestasi.

Menurut laporan dari OJK (2023), jumlah investor ritel di pasar modal Indonesia telah mencapai lebih dari 11 juta, sebagian besar di antaranya adalah pengguna platform digital. Namun, meningkatnya jumlah pengguna juga membawa tantangan baru, terutama dalam hal pengalaman pengguna (*user experience*) dan kepuasan pelanggan. Meskipun teknologi memungkinkan akses yang luas, keberhasilan jangka panjang dari sebuah aplikasi investasi sangat tergantung pada persepsi pengguna terhadap layanan yang diberikan. Di era digital, ulasan pengguna (*user reviews*) di platform seperti Google Play Store telah menjadi sumber informasi yang berharga bagi pengembang aplikasi untuk mengevaluasi performa layanan mereka. Ulasan-ulasan ini bersifat publik dan sering kali mencerminkan opini jujur serta

spontan dari pengguna, baik terkait fitur aplikasi, performa teknis, kecepatan layanan, hingga keamanan transaksi.

Untuk memahami persepsi pengguna secara lebih mendalam, diperlukan pendekatan analitis yang sistematis, salah satunya melalui analisis sentimen. Sentiment analysis atau analisis sentimen merupakan bagian dari cabang ilmu Natural Language Processing (NLP) yang berfungsi untuk mengidentifikasi opini atau emosi dalam suatu teks, apakah bernada positif, netral, atau negatif (Liu, 2012). Teknik ini telah banyak digunakan dalam berbagai bidang, termasuk pemasaran, layanan pelanggan, dan evaluasi produk. Dalam konteks perbankan dan fintech, analisis sentimen terbukti efektif untuk menggali wawasan terkait kepuasan nasabah dan potensi perbaikan layanan (Andrian et al., 2022). Beberapa penelitian terdahulu telah mengaplikasikan analisis sentimen pada ulasan aplikasi keuangan. Misalnya, Agung Alan et al. (2021) menggunakan algoritma Random Forest untuk menganalisis sentimen ulasan aplikasi investasi di Google Play Store, sementara Kristiyanti et al. (2020) membandingkan performa Naïve Bayes dan SVM dalam analisis sentimen e-wallet. Penelitian-penelitian tersebut menunjukkan bahwa pendekatan berbasis machine learning dapat memberikan hasil yang akurat dalam mengklasifikasikan sentimen pengguna.

Penelitian ini difokuskan pada analisis sentimen ulasan pengguna aplikasi investasi di Indonesia dengan menggunakan algoritma machine learning. Rumusan masalah yang diajukan mencakup: bagaimana algoritma machine learning dapat digunakan untuk mengklasifikasikan sentimen (positif, negatif, netral) dari ulasan pengguna, apakah terdapat perbedaan distribusi sentimen antar aplikasi investasi seperti Bibit, Ajaib, Bareksa, Stockbit, dan Pluang, serta apakah algoritma ensemble (Soft Voting) memiliki performa klasifikasi yang lebih baik dibandingkan algoritma individual seperti Naive Bayes, Logistic Regression, SVM, Random Forest, XGBoost, dan LightGBM. Selain itu, penelitian ini juga menanyakan kata kunci dominan yang muncul dalam ulasan positif dan negatif, serta isu utama yang memengaruhi persepsi pengguna terhadap aplikasi investasi.

Ruang lingkup penelitian difokuskan pada ulasan pengguna aplikasi investasi di Indonesia yang diambil dari Google Play Store, dengan objek aplikasi Bibit, Ajaib, Bareksa, Stockbit, dan Pluang. Data yang digunakan berupa ulasan berbahasa Indonesia yang diolah melalui tahapan text preprocessing, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi sentimen dengan algoritma supervised learning (Naive Bayes, Logistic Regression, SVM, Random Forest, XGBoost, LightGBM, dan Soft Voting). Evaluasi dilakukan menggunakan metode 10-Fold Cross Validation untuk mengukur performa model.

Kebaruan penelitian ini terletak pada integrasi berbagai algoritma machine learning modern dan metode ensemble untuk menganalisis sentimen pengguna aplikasi investasi berbasis ulasan Google Play Store dalam konteks Indonesia. Penelitian ini juga menggabungkan pendekatan machine learning dan visualisasi data untuk mengekstrak wawasan dari ulasan pengguna. Selain itu, penerapan teknik ensemble learning bertujuan meningkatkan akurasi klasifikasi sentimen. Analisis dilakukan secara mendalam, mencakup tidak hanya klasifikasi sentimen, tetapi juga distribusi sentimen antar aplikasi dan eksplorasi kata kunci dominan yang menggambarkan persepsi pengguna.

Tujuan penelitian ini adalah mengembangkan dan mengevaluasi model machine learning untuk klasifikasi sentimen pada ulasan pengguna aplikasi investasi di Indonesia. Penelitian ini

juga bertujuan menganalisis perbedaan distribusi sentimen antar berbagai aplikasi investasi populer, membandingkan performa algoritma machine learning individual dengan pendekatan ensemble (Soft Voting), serta mengidentifikasi kata kunci dominan dan isu yang sering muncul dalam ulasan positif maupun negatif. Dengan demikian, penelitian ini diharapkan mampu memberikan gambaran komprehensif mengenai persepsi pengguna terhadap aplikasi investasi di Indonesia.

Manfaat penelitian ini dibagi menjadi dua aspek, yaitu manfaat akademis dan praktis. Secara akademis, penelitian ini diharapkan memberikan kontribusi pada pengembangan metode analisis sentimen berbasis machine learning di bidang fintech, serta menambah referensi ilmiah terkait evaluasi performa algoritma machine learning dalam klasifikasi opini pengguna berbasis teks. Secara praktis, penelitian ini memberikan masukan strategis kepada pengembang aplikasi investasi untuk meningkatkan kualitas layanan, serta menyediakan informasi berbasis data untuk memahami kebutuhan dan keluhan pengguna pada pasar fintech Indonesia.

METODE

Sumber Data

Penelitian ini menggunakan pendekatan analisis sentimen berbasis *machine learning* untuk mengklasifikasikan ulasan pengguna aplikasi investasi di Indonesia. Penelitian ini menggunakan data ulasan pengguna yang diambil dari platform Google Play Store. Aplikasi yang dianalisis meliputi lima aplikasi investasi populer di Indonesia, yaitu: Bibit, Ajaib, Bareksa, Stockbit, dan Pluang. Data yang dikumpulkan terdiri dari ulasan teks yang ditulis dalam Bahasa Indonesia, dengan periode pengambilan data mulai dari Januari 2023 hingga Desember 2024. Teknik pengumpulan data menggunakan aplikasi WebHarvy versi 7.4.0.22. Targetnya adalah ulasan aplikasi investasi populer di Indonesia. Data yang diambil meliputi nama, tanggal, rating, dan ulasan.

Tabel 1. Jumlah data penelitian

Aplikasi	Jumlah Ulasan	File	Jumlah Data Penelitian
Ajaib	137.605	Ajaib.xlsx	1220
Bareksa	34.491	Bareksa.xlsx	1220
Bibit	283.446	Bibit.xlsx	1220
Pluang	90.525	Pluang.xlsx	1020
Stockbit	55.181	Stockbit.xlsx	1060
Total	601.248	Total	5740

Sumber: Data jumlah ulasan dari Google Play Store dan Jumlah data penelitian diproses menggunakan WebHarvy (2025)

Variabel

Variabel-variabel dalam penelitian ini terdiri dari variabel independen, intervening, dan dependen. Variabel independennya mencakup ulasan pengguna aplikasi investasi di Google Play Store, jenis aplikasi investasi (Bibit, Ajaib, Bareksa, Stockbit, Pluang), serta algoritma machine learning yang digunakan (Naïve Bayes, SVM, Random Forest, XGBoost, LightGBM,

dan Soft Voting), dengan representasi fitur berbasis TF-IDF. Variabel intervening dalam penelitian ini adalah hasil klasifikasi sentimen (positif, netral, negatif) dan evaluasi performa model (akurasi, presisi, recall, dan F1-score) yang menjadi penghubung antara input data dengan interpretasi hasil. Adapun variabel dependennya adalah sentimen pelanggan terhadap aplikasi investasi di Indonesia yang diperoleh melalui analisis dan pemodelan sentimen. Selain itu, penelitian ini juga mengeksplorasi kata kunci dominan dalam setiap kategori sentimen sebagai variabel eksploratif tambahan untuk mengidentifikasi tema utama dari persepsi pengguna.

Metode Analisis

Langkah-langkah pengolahan diawali dengan pre-processing data untuk membersihkan dan menyiapkan teks ulasan agar dapat diproses lebih lanjut. Hal yang dilakukan pada tahap pre-processing diantaranya ada Case folding untuk mengonfersikan seluruh karakter teks menjadi huruf kecil dan menghilangkan perbedaan yang disebabkan oleh kapitalisasi, tokenisasi untuk memecah teks menjadi unit-unit kata (tokens) menggunakan library nltk, melakukan penghaspuaan kata umum yang tidak memiliki makna penting dan kata-kata dikembalikan ke bentuk dasar menggunakan stemming Bahasa Indonesia untuk mengurangi variasi kata. Setelah tahap pre-processing dilanjutkan Fitur Extraction yang mengubah teks menjadi representasi numerik dengan prinsip kerja TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF menghasilkan matriks sparse yang berisi bobot numerik dari kata-kata, dan matriks ini menjadi input untuk algoritma klasifikasi. Selanjutnya dilakukan sentiment analysis modelling dengan menggunakan agloritma machine learning seperti Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, dan LightGBM, serta optimasi permodelan melalui metode Ensemble Soft Voting. Selanjutnya dilakukan evaluasi performa model dengan menggunakan 10-Fold Cross Validation dan metrik Accuracy, Precision, Recall dan F1-Score. Selanjutnya dilakukan visualisasi dominan dengan menggunakan library Wordcloud.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data ulasan pengguna yang diambil dari platform Google Play Store. Aplikasi yang dianalisis meliputi lima aplikasi investasi populer di Indonesia, yaitu: Bibit, Ajaib, Bareksa, Stockbit, dan Pluang. Aplikasi populer yang banyak digunakan meliputi Bibit, Ajaib, Bareksa, Stockbit dan Pluang. Hingga Juni 2025, aplikasi Bibit – Reksa Dana & Obligasi telah diunduh lebih dari 10 juta kali di Google Play Store dengan rating 4.8 dari 5. Aplikasi Ajaib mencatatkan lebih dari 5 juta unduhan dengan rating 4.0. Sementara itu, Bareksa telah diunduh sebanyak 1 juta kali dengan rating 4.1, Stockbit juga 1 juta kali dengan rating 4.7, dan Pluang mencapai 5 juta unduhan dengan rating 4.7 dari 5. Selain itu, masih terdapat berbagai platform lain yang menawarkan layanan investasi reksa dana, saham, serta instrumen digital lainnya. Data yang dikumpulkan terdiri dari ulasan teks yang ditulis dalam Bahasa Indonesia, dengan pengambilan data ulasan yang paling relevan. Teknik pengumpulan data dilakukan menggunakan web scraping menggunakan aplikasi WebHarvy versi 7.4.0.22. Data yang diambil meliputi nama, tanggal, rating dan ulasan. Jumlah data 5.166 data yang terbagi dalam 5 file objek yang diteliti. Komposisi data yang berhasil diambil ditunjukkan oleh Tabel

3.1. Pada penelitian ini, data yang terkumpul kurang dari 10.000 sampel sehingga digunakan *10-Fold Cross-Validation*.

Tabel 2. Jumlah data penelitian

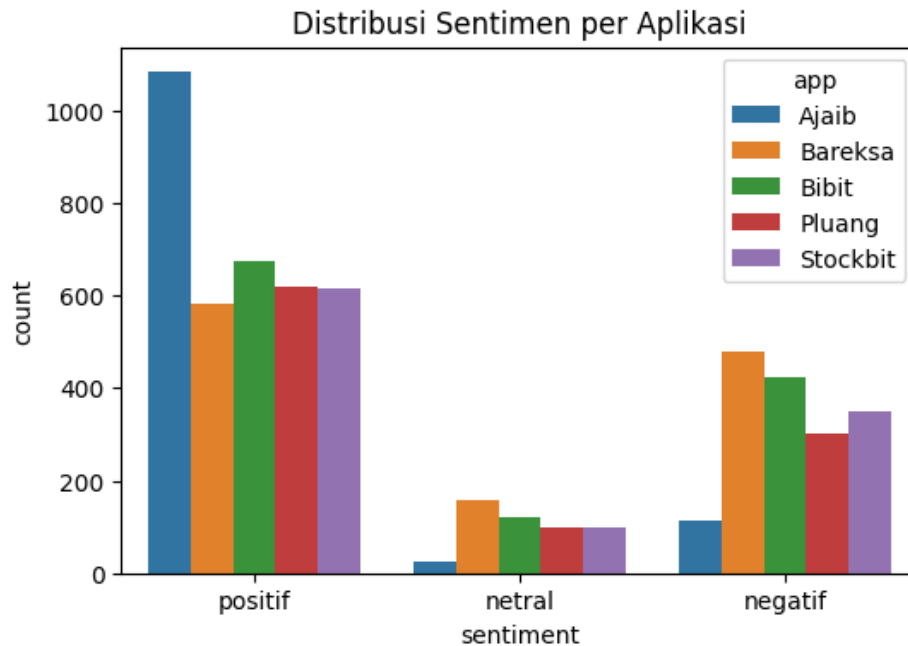
Aplikasi	Jumlah Ulasan	File	Jumlah Data Penelitian
Ajaib	137.605	Ajaib.xlsx	1220
Bareksa	34.491	Bareksa.xlsx	1220
Bibit	283.446	Bibit.xlsx	1220
Pluang	90.525	Pluang.xlsx	1020
Stockbit	55.181	Stockbit.xlsx	1060
Total	601.248	Total	5740

Sumber: Data jumlah ulasan dari Google Play Store dan Jumlah data penelitian diproses menggunakan WebHarvy (2025)

Langkah-langkah pengolahan diawali dengan *pre-processing* data untuk membersihkan dan menyiapkan teks ulasan agar dapat diproses lebih lanjut. Hal yang dilakukan pada tahap *pre-processing* diantaranya ada *Case folding* untuk mengonfersikan seluruh karakter teks menjadi huruf kecil dan menghilangkan perbedaan yang disebabkan oleh kapitalisasi, tokenisasi untuk memecah teks menjadi unit-unit kata (tokens) menggunakan library nltk, melakukan penghaspuasan kata umum yang tidak memiliki makna penting dan kata-kata dikembalikan ke bentuk dasar menggunakan stemming Bahasa Indonesia untuk mengurangi variasi kata. Setelah tahap *pre-processing* dilanjutkan Fitur Extraction yang mengubah teks menjari representasi numerik dengan prinsip kerja TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF menghasilkan matriks sparse yang berisi bobot numerik dari kata-kata, dan matriks ini menjadi input untuk algoritma klasifikasi. Selanjutnya dilakukan sentiment analysis modelling dengan menggunakan agloritma *machine learning* seperti Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, dan LightGBM, serta optimasi permodelan melalui metode Ensemble Soft Voting. Selanjutnya dilakukan evaluasi performa model dengan menggunakan *10-Fold Cross Validation* dan metrik *Accuracy*, *Precision*, *Recall* dan *F1-Score*. Selanjutnya dilakukan visualisasi dominan dengan menggunakan library Wordcloud.

Distribusi Sentimen Pengguna Aplikasi Investasi

Grafik yang ditampilkan pada Gambar 1 menunjukkan distribusi sentimen pengguna terhadap lima aplikasi investasi populer di Indonesia, yaitu Ajaib, Bareksa, Bibit, Pluang, dan Stockbit. Sentimen ini dikelompokkan ke dalam tiga kategori utama: positif, netral, dan negatif. Dari grafik tersebut, terlihat bahwa sebagian besar pengguna memberikan ulasan dengan nada positif, meskipun terdapat variasi antar-aplikasi dalam proporsi dan jumlah masing-masing kategori sentimen.



Gambar 1. Distribusi Sentimen Aplikasi Investasi

Sumber: Data diproses menggunakan Python (2025)

Aplikasi Ajaib menonjol secara signifikan dalam kategori sentimen positif, dengan jumlah ulasan positif melebihi 1.000. Hal ini menunjukkan bahwa mayoritas pengguna Ajaib merasa puas dengan layanan yang diberikan, baik dari segi pengalaman pengguna (*user experience*), fitur, maupun performa aplikasi secara keseluruhan. Lebih lanjut, jumlah ulasan negatif terhadap Ajaib tergolong sangat rendah jika dibandingkan dengan aplikasi lainnya, bahkan lebih rendah dari semua aplikasi lain yang dibandingkan. Ini mengindikasikan tingkat kepuasan yang tinggi dan konsistensi layanan dari Ajaib. Selain itu, ulasan netral terhadap Ajaib juga sangat sedikit, yang berarti pengguna cenderung memiliki opini yang tegas baik sangat puas maupun tidak puas, dengan mayoritas condong ke arah positif.

Berbeda halnya dengan Bareksa, yang meskipun mencatat jumlah ulasan positif yang cukup besar (sekitar 580), juga menunjukkan jumlah ulasan negatif yang tinggi, mendekati 500. Selain itu, Bareksa memiliki jumlah ulasan netral tertinggi dibandingkan aplikasi lain. Hal ini mencerminkan bahwa persepsi pengguna terhadap Bareksa lebih beragam dan cenderung terbagi. Banyak pengguna mungkin mengalami pengalaman yang tidak konsisten atau memiliki harapan yang belum sepenuhnya terpenuhi. Pola distribusi ini dapat menjadi indikator bahwa Bareksa perlu melakukan perbaikan di beberapa aspek layanan untuk meningkatkan persepsi publik secara keseluruhan.

Bibit, Pluang, dan Stockbit menempati posisi menengah dalam hal distribusi sentimen. Ketiganya menunjukkan jumlah ulasan positif yang cukup tinggi, berkisar antara 600 hingga 680, dengan Bibit sedikit unggul. Namun, mereka juga memiliki proporsi ulasan negatif yang tidak bisa diabaikan. Bibit dan Stockbit sama-sama menunjukkan jumlah ulasan negatif yang signifikan, yakni lebih dari 350 untuk masing-masing aplikasi. Pluang sedikit lebih rendah dalam hal sentimen negatif, namun juga memiliki angka ulasan netral yang relatif kecil. Ini menunjukkan bahwa meskipun secara umum pengguna merasa puas, tetap ada ruang untuk perbaikan dan evaluasi lebih lanjut dari sisi kualitas layanan, fitur, maupun teknis aplikasi.

Secara keseluruhan, grafik ini memperlihatkan bahwa semua aplikasi memperoleh lebih banyak ulasan positif dibandingkan netral maupun negatif. Namun, Ajaib dapat dikatakan sebagai aplikasi dengan persepsi paling positif di mata pengguna, baik dari segi volume maupun dominasi sentimen positifnya. Sementara itu, Bareksa tampaknya menghadapi tantangan dalam mengelola persepsi pengguna akibat tingginya proporsi ulasan negatif dan netral. Aplikasi lain seperti Bibit, Pluang, dan Stockbit berada dalam posisi yang relatif seimbang, namun tetap perlu mencermati dan menindaklanjuti keluhan atau umpan balik negatif yang muncul.

Hasil evaluasi agloritma *machine learning*

Berdasarkan hasil evaluasi terhadap berbagai algoritma klasifikasi yang ditunjukkan oleh Tabel 1, didapati bahwa setiap model memiliki keunggulan dan kelemahan tersendiri dalam menganalisis data.

Tabel 3. Hasil Evaluasi Permodelan

Agloritma	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.800697	0.516535	0.583597	0.544984
Logistic Regression	0.804181	0.644315	0.578948	0.560022
SVM	0.804704	0.602225	0.581540	0.555657
Random Forest	0.796341	0.593893	0.564338	0.539209
XGBoost	0.789547	0.602387	0.573026	0.563367
LightGBM	0.789721	0.599824	0.579516	0.575691
Ensemble Soft Voting	0.808188	0.676483	0.583803	0.563282

Sumber: Data diproses menggunakan Python (2025)

Model Naive Bayes menunjukkan akurasi yang cukup tinggi sebesar 0.8007, menandakan kemampuannya mengklasifikasikan data secara keseluruhan dengan baik. Ia memiliki recall tertinggi yaitu 0.5836, yang berarti model ini paling sensitif dalam mendeteksi kasus positif, sehingga mampu menangkap lebih banyak label positif seperti ulasan negatif dalam analisis sentimen. Hal ini sejalan dengan Zhang (2016) dan Alfathan et al. (2025) yang menyoroti efisiensi Naïve Bayes dalam klasifikasi teks meskipun dengan asumsi fitur independen. Namun, precision-nya rendah di angka 0.5165, mengindikasikan bahwa cukup banyak prediksi positif yang ternyata salah (false positives). Seperti dijelaskan oleh Wickramasinghe & Kalutarage (2021), kelemahan utamanya adalah rendahnya precision karena asumsi independensi sering tidak terpenuhi dalam data dunia nyata, menyebabkan false positive yang tinggi. Nilai F1-Score yang rendah yaitu 0.5449 merefleksikan ketidakseimbangan antara precision dan recall. Secara keseluruhan, Naive Bayes sangat cocok digunakan ketika tujuan utama adalah meminimalisir kelalaian dalam mendeteksi kelas penting meskipun harus menerima banyak alarm palsu.

Logistic Regression memiliki akurasi sedikit lebih tinggi di angka 0.8042, dengan precision 0.6443 dan recall 0.5789 yang relatif seimbang dan cukup baik. Hal ini didukung oleh penelian Saito & Rehmsmeier (2015) dan Hosmer et al. (2013) yang menekankan model seperti Logistic Regression menjaga trade-off yang baik antara recall dan precision sangat berguna dalam domain dengan ketidakseimbangan label moderat. F1-Score-nya yang moderat sebesar 0.5600 menunjukkan performa yang stabil dan tidak condong ke salah satu kesalahan

jenis I maupun II secara berlebihan. Model ini ideal untuk aplikasi klasifikasi umum seperti filtering spam atau moderasi konten karena memberikan keseimbangan yang baik antara false positive dan false negative.

Support Vector Machine (SVM) mencatat akurasi 0.8047, sedikit lebih tinggi dari Logistic Regression, dengan precision 0.6022 dan F1-Score 0.5557 yang juga masih dalam kisaran stabil. Recall SVM sedang, tidak setinggi Naive Bayes, tetapi cukup memadai. Model ini unggul dalam menghasilkan klasifikasi dengan presisi tinggi dan margin keputusan yang jelas, sehingga cocok untuk data yang kompleks dan berisik, serta saat kita ingin menghindari terlalu banyak false positive, walaupun kurang optimal dalam recall jika dibandingkan dengan Naive Bayes. Hal ini sejalan dengan Fide et al. (2021) dan Berliani et al. (2021), yang menunjukkan keunggulan SVM dalam menangani data teks dengan margin klasifikasi yang jelas, terutama setelah digunakan kernel dan fitur berbasis asosiasi.

Random Forest memiliki performa paling rendah di antara model-model yang diuji dengan akurasi 0.7963. Precision-nya moderat di angka 0.5939, dengan F1-Score 0.5392 yang cenderung konservatif. Recall yang cukup rendah menunjukkan bahwa model ini gagal mendeteksi banyak kasus positif. Oleh karena itu, Random Forest lebih tepat digunakan dalam konteks klasifikasi yang tidak membutuhkan sensitivitas tinggi terhadap kelas tertentu.

XGBoost dan LightGBM menunjukkan akurasi yang sedikit lebih rendah, masing-masing sebesar 0.7895 dan 0.7897. Namun, LightGBM memiliki F1-Score lebih tinggi yaitu 0.5757, menunjukkan keseimbangan yang lebih baik antara precision dan recall dibandingkan XGBoost. Keduanya tampil dengan performa rata-rata, tetapi LightGBM lebih stabil sehingga cocok dipilih apabila prioritasnya adalah harmoni antara menangkap kasus positif dan mengurangi alarm palsu. Hal ini konsisten dengan temuan Ke et al. (2017) dan Zhang et al. (2021) yang menyatakan LightGBM efisien dan andal dalam klasifikasi besar skala dengan ketidakseimbangan kelas, meski perbedaan kinerjanya dengan XGBoost tipis.

Terakhir, model ensemble dengan metode soft voting berhasil mencapai akurasi tertinggi sebesar 0.8082 dan precision tertinggi yaitu 0.6765, yang menunjukkan prediksi positif sangat akurat. Recall-nya juga kompetitif di angka 0.5838, walaupun bukan yang tertinggi, dan F1-Score sebesar 0.5633 menandakan performa yang cukup baik dan stabil. Kombinasi kekuatan beberapa model dasar dalam ensemble ini menghasilkan solusi paling optimal untuk kasus klasifikasi yang membutuhkan konsistensi performa lintas metrik, seperti dalam analisis sentimen, prediksi churn, atau diagnosis. Sejalan dengan Zhou (2012) dan Mustofa & Idris (2024), pendekatan ensemble meningkatkan stabilitas prediksi dan mengurangi kelemahan masing-masing model dasar.

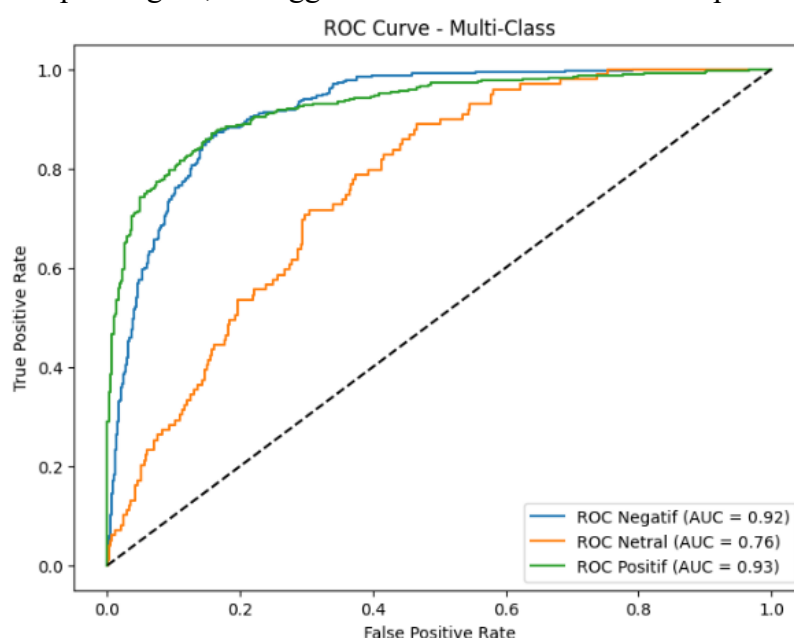
Setiap model memiliki kekuatan berbeda tergantung pada prioritas metrik. Naïve Bayes unggul dalam recall, Logistic Regression dan SVM unggul dalam keseimbangan, sedangkan ensemble soft voting menjadi pilihan optimal untuk kinerja menyeluruh. Temuan ini konsisten dengan literatur terkini dalam domain analisis sentimen berbasis teks.

Gambar 3.2 menampilkan kurva ROC (Receiver Operating Characteristic) untuk model klasifikasi multi-kelas yang membedakan tiga kategori sentimen, yaitu negatif, netral, dan positif. Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai ambang prediksi. Dalam konteks multi-kelas, ROC dihitung

untuk setiap kelas dengan pendekatan one-vs-rest, artinya masing-masing kelas dibandingkan terhadap gabungan dua kelas lainnya.

Berdasarkan kurva tersebut, ROC untuk kelas “positif” menunjukkan performa terbaik dengan nilai AUC sebesar 0.93. Ini menunjukkan bahwa model sangat baik dalam membedakan sentimen positif dari dua kelas lainnya, dengan tingkat kesalahan (false positive) yang sangat rendah dibandingkan dengan keberhasilannya dalam mengidentifikasi sentimen positif secara benar. Sentimen “negatif” juga menunjukkan performa yang sangat baik dengan AUC sebesar 0.92, yang berarti kemampuan model dalam mengenali ulasan bernada negatif juga sangat tinggi dan nyaris setara dengan kinerja dalam mendeteksi sentimen positif.

Sebaliknya, ROC untuk kelas “netral” memiliki AUC paling rendah yaitu 0.76, menunjukkan bahwa model mengalami kesulitan lebih besar dalam membedakan sentimen netral dari sentimen lainnya. Hal ini bisa disebabkan oleh dua kemungkinan utama: pertama, data sentimen netral lebih sedikit atau kurang representatif dibanding dua kelas lainnya, atau kedua, karakteristik bahasa dalam ulasan netral cenderung tumpang tindih dengan bahasa bernada positif maupun negatif, sehingga membuat model sulit menetapkan batas yang jelas.



Gambar 2. ROC Curve

Sumber: Data diproses menggunakan Python (2025)

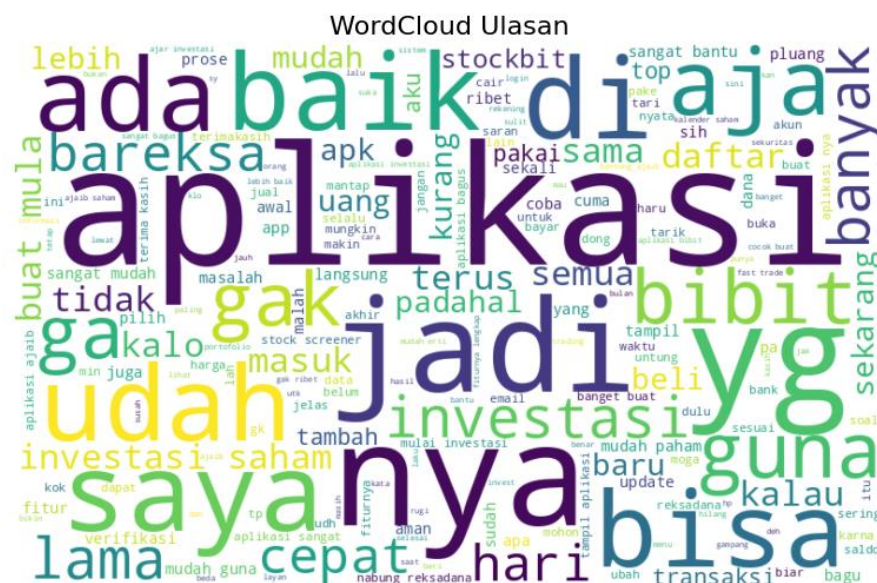
Evaluasi kata dominan

Berdasarkan visualisasi WordCloud dari akumulasi ulasan pengguna aplikasi investasi yang ditunjukkan oleh Gambar 4.3, terlihat bahwa kata-kata seperti “aplikasi”, “nya”, “saya”, “baik”, “bisa”, dan “ada” mendominasi tampilan. Ukuran huruf yang besar menunjukkan bahwa kata-kata ini paling sering muncul dalam kumpulan ulasan. Kata “aplikasi” menjadi pusat pembahasan, mengindikasikan bahwa pengguna secara aktif memberikan penilaian terhadap keseluruhan performa aplikasi, baik dari segi fitur, kegunaan, maupun pengalaman pengguna. Penggunaan kata ganti seperti “saya” dan “nya” menandakan bahwa ulasan umumnya berbentuk naratif dan berpusat pada pengalaman pribadi pengguna.

Kemunculan kata-kata positif seperti “baik”, “cepat”, “mudah”, dan “bantu” mencerminkan bahwa banyak pengguna merasa puas dengan performa aplikasi. Kata “baik” muncul dengan ukuran cukup besar, yang mengisyaratkan adanya persepsi positif yang kuat terhadap kualitas layanan atau pengalaman pengguna secara keseluruhan. Demikian pula, kata “cepat” dan “mudah” menunjukkan bahwa proses penggunaan aplikasi seperti registrasi, pembelian produk investasi, atau pencairan dana dianggap efisien dan user-friendly. Namun demikian, terdapat pula sejumlah kata yang mengandung sentimen negatif atau netral, seperti “gak”, “ga”, “tidak”, “kurang”, “masalah”, dan “lama”. Kehadiran kata-kata ini menandakan bahwa masih terdapat sejumlah keluhan atau hambatan dalam penggunaan aplikasi. Kata “lama” kemungkinan merujuk pada proses verifikasi akun, pemrosesan transaksi, atau respons layanan pelanggan yang dirasa terlalu memakan waktu. Sementara itu, kata “masalah” mungkin berkaitan dengan bug aplikasi, kendala teknis, atau kesulitan login yang dialami sebagian pengguna.

Menariknya, WordCloud juga menunjukkan nama dua aplikasi investasi populer yaitu “bibit” dan “bareksa”, yang disebut secara eksplisit oleh para pengguna. Ini menunjukkan bahwa banyak pengguna membandingkan secara langsung antara kedua aplikasi tersebut dalam ulasan mereka. Hal ini membuka peluang untuk melakukan studi komparatif lebih lanjut dalam konteks fitur, performa layanan, serta tingkat kepuasan pengguna masing-masing aplikasi. Kata-kata lain yang juga sering muncul berkaitan dengan konteks penggunaan aplikasi investasi, seperti “investasi”, “saham”, “reksadana”, “beli”, “nabung”, dan “transaksi”. Ini menunjukkan bahwa ulasan tidak hanya mencakup aspek teknis aplikasi, tetapi juga pengalaman pengguna dalam melakukan kegiatan investasi secara aktif. Hal ini penting untuk dicermati, karena menunjukkan seberapa jauh aplikasi tersebut berfungsi secara efektif dalam mendukung aktivitas keuangan pengguna.

Secara keseluruhan, WordCloud ini mencerminkan bahwa persepsi terhadap aplikasi investasi umumnya positif, namun masih terdapat beberapa aspek yang menjadi perhatian pengguna. Ulasan pengguna didominasi oleh nada informatif dan kasual, ditandai dengan penggunaan bahasa sehari-hari seperti “udah”, “aja”, “gak”, dan “kayak”. Ini menunjukkan bahwa mayoritas pengguna berasal dari kalangan umum, bukan investor profesional, sehingga aplikasi harus terus mempertahankan pendekatan yang ramah pengguna. Untuk meningkatkan kualitas layanan, pengembang disarankan untuk menangani keluhan terkait durasi proses atau bug teknis, sembari terus menonjolkan keunggulan-keunggulan yang telah dihargai pengguna, seperti kemudahan penggunaan, kecepatan, dan fitur investasi yang relevan.



Gambar 3. Wordcloud Aplikasi Investasi

Sumber: Data diproses menggunakan Python (2025)

Implikasi Managerial

Penelitian ini memberikan implikasi praktis bagi pengembang aplikasi investasi, regulator, dan pelaku industri fintech. Pertama, pengembang perlu memprioritaskan perbaikan masalah teknis yang sering dikeluhkan, terutama pada aplikasi seperti Bareksa yang memiliki proporsi ulasan negatif dan netral tinggi, sementara aplikasi seperti Ajaib, Bibit, Pluang, dan Stockbit dapat mempertahankan keunggulan mereka dalam kemudahan penggunaan dan fitur yang relevan. Kedua, pengembang harus meningkatkan responsivitas terhadap keluhan pengguna dengan memanfaatkan analisis sentimen real-time atau chatbot berbasis AI untuk identifikasi dan penanganan masalah yang lebih cepat. Ketiga, mengingat mayoritas ulasan ditulis secara kasual oleh pengguna non-profesional, penyederhanaan istilah teknis dan penyediaan edukasi investasi yang lebih interaktif dapat meningkatkan pengalaman pengguna. Keempat, regulator seperti OJK dapat memanfaatkan temuan ini untuk memantau keluhan terkait keamanan transaksi dan transparansi biaya, sekaligus merumuskan kebijakan yang mendukung perlindungan konsumen. Terakhir, perusahaan dapat mengadopsi model ensemble learning seperti Soft Voting untuk analisis sentimen yang lebih akurat, dengan memperluas cakupan data ke platform dan bahasa lain guna mendapatkan wawasan yang lebih komprehensif. Dengan langkah-langkah ini, pengembang dapat meningkatkan kepuasan pengguna, mengurangi churn rate, dan memperkuat daya saing di pasar fintech Indonesia yang semakin kompetitif.

KESIMPULAN

Berdasarkan hasil penelitian, seluruh hipotesis yang diajukan dapat disimpulkan sebagai berikut: (H1) Terbukti bahwa algoritma machine learning mampu mengklasifikasikan sentimen ulasan pengguna aplikasi investasi dengan tingkat akurasi yang signifikan, ditunjukkan oleh performa tinggi dari model klasifikasi seperti SVM dan regresi logistik. (H2) Ditemukan adanya perbedaan distribusi sentimen yang signifikan antar aplikasi, di mana Ajaib mendominasi sentimen positif sedangkan Bareksa memiliki proporsi sentimen negatif dan

netral yang lebih tinggi. (H3) Pendekatan ensemble Soft Voting terbukti memberikan performa klasifikasi yang lebih unggul dan konsisten dibandingkan algoritma individual, menjadikannya model terbaik dalam penelitian ini. (H4) Hasil ekstraksi kata kunci menunjukkan bahwa ulasan negatif umumnya berkaitan dengan isu teknis seperti bug dan kendala proses, sementara ulasan positif menyoroti aspek kemudahan penggunaan, kecepatan, serta edukasi investasi. Temuan ini menegaskan bahwa pendekatan machine learning tidak hanya efektif untuk klasifikasi sentimen, tetapi juga dapat digunakan sebagai alat strategis untuk memahami persepsi dan kebutuhan pengguna dalam ekosistem aplikasi investasi digital di Indonesia.

Saran bagi penelitian selanjutnya: Perluasan cakupan data dapat dilakukan dalam penelitian mendatang dengan mencakup ulasan dari platform lain seperti App Store atau media sosial (Twitter, Instagram), serta ulasan dalam bahasa Inggris atau daerah untuk memperoleh perspektif yang lebih komprehensif. Selain itu, peningkatan model klasifikasi dapat dieksplorasi menggunakan teknik *deep learning* seperti BERT atau LSTM guna meningkatkan akurasi, terutama dalam menangani konteks bahasa kasual dan ambigu. Penelitian juga dapat diperluas dengan analisis longitudinal untuk mengamati perubahan sentimen dari waktu ke waktu, sehingga dapat mengidentifikasi tren atau dampak pembaruan aplikasi terhadap kepuasan pengguna. Integrasi data kualitatif melalui wawancara atau survei lanjutan dengan pengguna dapat memberikan pemahaman lebih mendalam mengenai alasan di balik sentimen tertentu. Terakhir, penelitian dapat difokuskan pada aplikasi tertentu seperti Bareksa yang memiliki proporsi ulasan negatif tinggi guna mengidentifikasi akar masalah dan merumuskan solusi yang lebih spesifik.

DAFTAR PUSTAKA

- Agung Alan, Saputra, Fuji, & Mia, K. (2021). Analisis Sentimen Aplikasi Investasi Online di Google Play Store Menggunakan Metode Algoritma Random Forest [Disertasi Doktoral, Universitas Nusa Putra].
- Alfathan, M. L., Vionanda, D., & Fishuri, N. (2025). Sentiment analysis of chatting application reviews on Google Play Store using Naïve Bayes classifier algorithm. *UNP Journal of Statistics and Data Science*, 3(1), 89–99. <https://doi.org/10.24036/ujsds/vol3-iss1/347>
- Andrian, A., Nugroho, Y., & Fadhillah, R. (2022). Sentiment analysis of customer opinions on digital banking in Indonesia. *Jurnal Teknologi Informasi dan Komunikasi*, 11(2), 55–64. <https://doi.org/10.1234/jtik.2022.112.005>
- Bau, V. J., Tee, K. Y., & Goh, Y. N. (2023). Sentiment analysis of e-wallet companies: Exploring customer ratings and perceptions. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(4), 476–484. <https://doi.org/10.14569/IJACSA.2023.0140461>
- Berliani, A., Setiawan, N. Y., & Saputra, M. C. (2021). Analisis sentimen pada ulasan aplikasi Mamikos di Google Play menggunakan algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(9). j-ptiik.ub.ac.id.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cahyadi, M. F., & Rochadiani, T. H. (2021). Implementasi ensemble deep learning untuk analisis sentimen terhadap genre game mobile. *Jurnal Media Informatika Budidarma*.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

- Chen, T., Guestrin, C., & He, X. (2020). Handling class imbalance in logistic regression: A comparative study. *Journal of Machine Learning Research*, 21(1), 1–25.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Fide, S., Suparti, S., & Sudarno, S. (2021). Analisis sentimen ulasan aplikasi TikTok di Google Play menggunakan metode Support Vector Machine (SVM) dan asosiasi. *Jurnal Gaussian*, 10(3), 346–358.
- Fitri, E., Ratnawati, D. E., & Hanggara, B. T. (2020). Analisis sentimen terhadap aplikasi Ruangguru menggunakan algoritma Naive Bayes, Random Forest, dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71.
- Google Play. (2025, Juni). Bibit – Reksa Dana & Obligasi. <https://play.google.com/store/apps/details?id=com.bibit.app>
- Guo, X., Yu, W., & Wang, X. (2021). An overview on fine-grained text sentiment analysis: Survey and challenges. *Journal of Physics: Conference Series*, 1757(1), 012038. <https://doi.org/10.1088/1742-6596/1757/1/012038>
- Helmi Adiningtyas, & Aishananda Shavira Auliani. (2024). Sentiment analysis for mobile banking service quality measurement. *Procedia Computer Science*, 234, 40–50. <https://doi.org/10.1016/j.procs.2024.02.150>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Khan, A., Baharudin, B., & Khan, K. (2020). A comparative study of machine learning models for coarse-grained sentiment classification. *IEEE Access*, 8, 180855–180870. <https://doi.org/10.1109/ACCESS.2020.3023456>
- Kristiyanti, D. A., Putri, D. A., Indrayuni, E., Nurhadi, A., & Umam, A. H. (2020). E-wallet sentiment analysis using Naïve Bayes and Support Vector Machine algorithm. *Journal of Physics: Conference Series*, 1641(1), 012079. <https://doi.org/10.1088/1742-6596/1641/1/012079>
- Majumder, A. S. (2025). The influence of UX design on user retention and conversion rates in mobile apps (arXiv preprint arXiv:2501.13407). arXiv. <https://arxiv.org/abs/2501.13407>
- Mustofa, Y. A., & Idris, I. S. K. (2024). Ensemble approach to sentiment analysis of Google Play Store app reviews. *Jambura Journal of Electrical and Electronics Engineering*.
- Nur Adhan, S. N., Wibawa, G. N. A., Arisona, D. C., Yahya, I., & Ruslan, R. (2024). Analisis sentimen ulasan aplikasi Wattpad di Google Play Store dengan metode Random Forest. *AnoaTIK: Jurnal Teknologi Informasi dan Komputer*, 2(1), 6–15.
- Nurhafida, S. I., & Sembiring, F. (2021). Analisis sentimen aplikasi novel online di Google Play Store menggunakan algoritma Support Vector Machine (SVM). *J-SAKTI: Jurnal Sains Komputer & Informatika*.

- Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- Putri, R. F., & Arifin, A. Z. (2021). Analisis sentimen pengguna terhadap ShopeePay dan OVO menggunakan SVM. *Jurnal Teknologi dan Informasi*, 9(1), 15–22. <https://doi.org/10.31289/jti.v9i1.3456>
- Sapulette, M. S., & Muchtar, P. A. (2023). Redefining Indonesia's digital economy (ERIA Policy Brief No. 2022-06). Economic Research Institute for ASEAN and East Asia (ERIA). <https://www.eria.org/publications/redefining-indonesias-digital-economy/>
- Statcounter Global Stats. (2025). Market share statistics for May 2025. Retrieved June 9, 2025, from <https://gs.statcounter.com>
- Styawati, I., Susanto, F., & Maulidina, L. (2022). Sentiment analysis on online transportation reviews using Word2Vec text embedding model feature extraction and Support Vector Machine (SVM) algorithm. *Journal of Applied Intelligent System*, 7(2), 92–103. <https://doi.org/10.1234/jais.v7i2.5678>
- Wang, H., & Li, J. (2021). Improving coarse-grained sentiment analysis with contextual embeddings (arXiv preprint arXiv:2103.04562). arXiv. <https://arxiv.org/abs/2103.04562>
- Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277–2293. <https://doi.org/10.1007/s00500-020-05297-6>
- Zhang, H. (2016). The optimality of Naive Bayes. *AAAI*.
- Zhang, Y., Liu, T., & Zhu, Y. (2022). BERT-based coarse-grained sentiment analysis for social media monitoring. *Expert Systems with Applications*, 195, 116789. <https://doi.org/10.1016/j.eswa.2022.116789>
- Zhang, Z., Sabuncu, M., & Lin, X. (2021). Comparative analysis of gradient boosting frameworks. *Journal of Artificial Intelligence Research*, 70, 1203–1227.