



ANALISIS SENTIMEN ULASAN AKADEMI CRYPTO DI PLAY STORE MENGGUNAKAN ALGORITMA XGBOOST

Ryan Gunawan

Program Studi Sistem Informasi, Universitas Muhammadiyah Sumatera Utara

Email: ryangunawan040532@gmail.com

Abstrak — Ulasan pengguna pada Google Play Store mengandung informasi mengenai pengalaman, kepuasan, serta kendala yang dirasakan selama menggunakan suatu aplikasi. Jumlah ulasan yang cukup banyak dan bentuk teks yang tidak terstruktur menyebabkan informasi tersebut sulit dianalisis secara manual. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Akademi Crypto di Google Play Store dan mengevaluasi kemampuan algoritma XGBoost dalam mengklasifikasikan sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Data penelitian terdiri atas 569 ulasan yang dikumpulkan pada periode Januari hingga Juni 2026. Tahapan penelitian meliputi pengumpulan data, pelabelan berdasarkan rating, preprocessing, pembobotan TF-IDF, pembagian data 80:20, klasifikasi menggunakan XGBoost, serta evaluasi model. Hasil pelabelan menunjukkan 317 ulasan positif, 230 ulasan negatif, dan 22 ulasan netral. Dari proses pembobotan TF-IDF, term dengan bobot tinggi antara lain “aplikasi”, “crypto”, “bagus”, “login”, dan “daftar”. Berdasarkan hasil evaluasi model yang digunakan dalam penelitian, XGBoost memperoleh accuracy 74,07%, precision 58,59%, recall 58,35%, dan F1-score 57,68%. Hasil tersebut menunjukkan bahwa XGBoost dapat digunakan untuk klasifikasi sentimen pada ulasan Akademi Crypto, dengan keterbatasan terutama pada pengenalan kelas netral yang memiliki jumlah data paling sedikit.

Kata kunci: analisis sentimen; Akademi Crypto; Google Play Store; TF-IDF; XGBoost

Abstract — User reviews on the Google Play Store contain information about user experiences, satisfaction, and problems encountered while using an application. The unstructured nature of review texts makes manual analysis inefficient. This study analyzes user sentiment toward the Akademi Crypto application on the Google Play Store and evaluates XGBoost for classifying reviews into positive, negative, and neutral categories. The dataset consists of 569 reviews collected from January to June 2026. The research stages include data collection, rating-based labeling, preprocessing, TF-IDF feature weighting, an 80:20 train-test split, XGBoost classification, and model evaluation. The dataset contains 317 positive, 230 negative, and 22 neutral reviews. High-weight TF-IDF terms include “aplikasi”, “crypto”, “bagus”, “login”, and “daftar”. Based on the evaluation results used in this study, XGBoost achieved 74.07% accuracy, 58.59% precision, 58.35% recall, and 57.68% F1-score. The results indicate that XGBoost can be applied to sentiment classification of Akademi Crypto reviews, although the neutral class remains difficult to identify because it contains substantially fewer samples.

Keywords: sentiment analysis; Akademi Crypto; Google Play Store; TF-IDF; XGBoost

Article History

Received: Oktober 2026

Reviewed: Oktober 2026

Published: Oktober 2026

Plagiarism Checker

No 237.u873247.fh8374

Copyright : Author

Publish by : Kohesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong penggunaan aplikasi sebagai sarana penyediaan informasi dan layanan di berbagai bidang. Pada aplikasi yang berkaitan dengan edukasi dan informasi cryptocurrency, pengalaman pengguna dapat terlihat melalui ulasan yang diberikan pada platform distribusi aplikasi. Ulasan tersebut tidak hanya berupa penilaian berbintang, tetapi juga memuat komentar mengenai kemudahan penggunaan, akses akun, materi, layanan, dan kendala teknis yang dialami pengguna.

Ulasan pada Google Play Store memiliki bentuk yang tidak terstruktur dan menggunakan variasi bahasa yang beragam. Kondisi tersebut membuat proses identifikasi kecenderungan opini secara manual menjadi kurang efisien ketika jumlah data bertambah. Analisis sentimen dapat digunakan untuk mengubah data teks menjadi informasi terstruktur berdasarkan polaritas atau kategori sentimen. Pendekatan ini umum digunakan untuk memperoleh gambaran opini pengguna terhadap produk atau layanan digital (Yang et al., 2022).

Beberapa penelitian sebelumnya menunjukkan bahwa XGBoost dapat diterapkan pada analisis sentimen berbasis teks. Setiyono et al. (2024) menerapkan XGBoost pada ulasan aplikasi Threads dengan penguatan SMOTE, sedangkan Wardana et al. (2024) mengombinasikan XGBoost dengan Word2Vec pada ulasan MyPertamina. Penelitian lain menggunakan XGBoost pada konteks komentar media sosial dan ulasan pengguna. Perbedaan penelitian ini terletak pada objek yang dikaji, yaitu ulasan aplikasi Akademi Crypto, serta penggunaan TF-IDF sebagai representasi fitur.

Berdasarkan kondisi tersebut, penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Akademi Crypto di Google Play Store menggunakan algoritma XGBoost. Penelitian juga bertujuan mengetahui distribusi sentimen, karakteristik kata berdasarkan bobot TF-IDF, serta performa model berdasarkan accuracy, precision, recall, dan F1-score.

2. METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif deskriptif. Alur penelitian terdiri atas pengumpulan data ulasan, pelabelan sentimen, preprocessing, ekstraksi fitur TF-IDF, pembagian data, klasifikasi XGBoost, dan evaluasi model. Selain model analisis, penelitian juga menghasilkan sistem berbasis web untuk membantu proses analisis sentimen.

2.1 Pengumpulan Data

Data berupa ulasan pengguna aplikasi Akademi Crypto yang diperoleh dari Google Play Store melalui teknik web scraping menggunakan library Google Play Scraper pada Python. Dataset yang digunakan berjumlah 569 ulasan dengan periode pengumpulan Januari sampai Juni 2026. Atribut data yang digunakan meliputi rating, tanggal, dan isi ulasan.

2.2 Pelabelan Sentimen

Pelabelan dilakukan berdasarkan rating pengguna. Rating 1 dan 2 dikategorikan sebagai negatif, rating 3 sebagai netral, sedangkan rating 4 dan 5 sebagai positif. Hasil pelabelan menghasilkan 317 ulasan positif, 230 ulasan negatif, dan 22 ulasan netral.

Kategori	Jumlah	Persentase
Negatif	230	40,42%
Netral	22	3,87%
Positif	317	55,71%
Total	569	100%

Tabel 1. Distribusi data berdasarkan label sentimen.



2.3 Preprocessing Data

Preprocessing dilakukan untuk mengurangi unsur yang tidak diperlukan dan menyeragamkan bentuk teks. Tahapannya meliputi cleaning untuk menghapus karakter atau simbol yang tidak relevan, case folding untuk mengubah huruf menjadi huruf kecil, tokenizing untuk memecah teks menjadi token, normalisasi untuk menyeragamkan bentuk kata, stopwords removal untuk mengurangi kata umum, dan stemming menggunakan Sastrawi untuk memperoleh bentuk dasar kata.

2.4 Ekstraksi Fitur TF-IDF

Teks hasil preprocessing dikonversi menjadi representasi numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Pembobotan digunakan untuk merepresentasikan tingkat kepentingan suatu term pada dokumen relatif terhadap keseluruhan dokumen. Hasil TF-IDF kemudian digunakan sebagai fitur masukan pada model XGBoost.

No.	Term	Bobot TF-IDF
1	aplikasi	32,54
2	crypto	22,42
3	bagus	22,34
4	login	22,32
5	daftar	18,57
6	bantu	16,09
7	gak	15,14
8	mudah	14,82
9	masuk	14,35
10	banget	14,32

Tabel 2. Sepuluh term dengan bobot TF-IDF tertinggi.

2.5 Pembagian Data dan Klasifikasi

Dataset dibagi menggunakan perbandingan 80:20 sehingga 455 data digunakan sebagai data latih dan 114 data sebagai data uji. Algoritma Extreme Gradient Boosting (XGBoost) digunakan untuk membangun model klasifikasi tiga kelas sentimen berdasarkan fitur TF-IDF.

2.6 Evaluasi Model

Kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score. Confusion matrix digunakan untuk melihat distribusi hasil prediksi pada setiap kelas. Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data uji, sedangkan precision, recall, dan F1-score digunakan untuk melihat kinerja klasifikasi secara lebih rinci.

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Sentimen

Dari 569 ulasan yang dianalisis, sebanyak 317 ulasan (55,71%) termasuk kategori positif, 230 ulasan (40,42%) termasuk kategori negatif, dan 22 ulasan (3,87%) termasuk kategori netral. Distribusi tersebut memperlihatkan bahwa data positif merupakan kategori terbesar, sedangkan kategori netral merupakan kategori dengan jumlah paling sedikit. Ketidakseimbangan ini perlu

diperhatikan karena jumlah contoh pada suatu kelas dapat memengaruhi kemampuan model dalam mempelajari pola kelas tersebut.

3.2 Karakteristik Ulasan Berdasarkan TF-IDF

Hasil pembobotan TF-IDF menunjukkan bahwa kata “aplikasi” memiliki bobot tertinggi, yaitu 32,54. Term lain yang memiliki bobot tinggi adalah “crypto”, “bagus”, “login”, “daftar”, dan “bantu”. Kemunculan term tersebut menggambarkan topik yang sering muncul dalam ulasan, terutama pembahasan mengenai aplikasi, materi atau informasi cryptocurrency, proses akses akun, pendaftaran, dan permintaan bantuan.

Pada ulasan negatif, kata-kata seperti “daftar”, “member”, “masuk”, “email”, “akun”, “password”, dan “lambat” muncul sebagai bagian dari pembahasan mengenai kendala akses dan penggunaan aplikasi. Sementara itu, ulasan positif banyak memuat kata yang berkaitan dengan kualitas atau manfaat aplikasi. Temuan tersebut menunjukkan bahwa isi ulasan tidak hanya berkaitan dengan penilaian umum, tetapi juga dengan aspek fungsional aplikasi.

3.3 Hasil Klasifikasi XGBoost

Metrik	Hasil
Accuracy	74,07%
Precision	58,59%
Recall	58,35%
F1-score	57,68%

Tabel 3. Hasil evaluasi model XGBoost.

Berdasarkan hasil evaluasi yang digunakan dalam penelitian, model XGBoost memperoleh accuracy sebesar 74,07%. Nilai tersebut menunjukkan bahwa model mampu menghasilkan prediksi yang benar pada sebagian besar data uji. Precision sebesar 58,59%, recall sebesar 58,35%, dan F1-score sebesar 57,68% menunjukkan bahwa kinerja model masih berbeda antar kelas sentimen.

Hasil evaluasi per kelas menunjukkan bahwa kelas positif dan negatif lebih mudah dikenali dibandingkan kelas netral. Kelas netral memiliki jumlah data paling sedikit sehingga pola yang tersedia untuk dipelajari model juga terbatas. Kondisi ini dapat menjelaskan mengapa model masih mengalami kesulitan dalam mengenali ulasan yang tidak menunjukkan kecenderungan sentimen secara jelas.

3.4 Pembahasan

Hasil penelitian memperlihatkan bahwa kombinasi preprocessing, TF-IDF, dan XGBoost dapat digunakan untuk mengklasifikasikan ulasan Akademi Crypto ke dalam tiga kategori sentimen. Penggunaan TF-IDF memberikan representasi numerik terhadap teks sehingga dapat diproses oleh model klasifikasi. XGBoost kemudian memanfaatkan fitur tersebut untuk membangun pola klasifikasi.

Dibandingkan pendekatan yang menggunakan representasi berbeda, seperti Word2Vec atau pendekatan hibrida berbasis RoBERTa yang dibahas dalam penelitian terdahulu, penelitian ini menggunakan pipeline yang lebih sederhana, yaitu TF-IDF dan XGBoost. Hasil yang diperoleh dapat menjadi gambaran awal mengenai karakteristik sentimen pengguna pada objek penelitian, tetapi belum dapat digunakan untuk menyimpulkan bahwa XGBoost merupakan metode terbaik karena penelitian ini tidak melakukan perbandingan eksperimental dengan beberapa algoritma lain pada dataset yang sama.



4. IMPLEMENTASI SISTEM

Penelitian menghasilkan sistem analisis sentimen berbasis web yang mengintegrasikan proses preprocessing, TF-IDF, dan klasifikasi XGBoost. Sistem menyediakan halaman beranda, informasi mengenai teknologi yang digunakan, serta halaman analisis. Pada halaman analisis, pengguna dapat memasukkan teks tunggal maupun mengunggah data dalam format CSV atau Excel untuk diproses.

Pengujian fungsional menggunakan metode Black Box dilakukan terhadap fungsi utama sistem. Hasil pengujian pada skripsi menunjukkan bahwa fungsi utama yang diuji dapat berjalan sesuai kebutuhan. Dengan demikian, sistem dapat digunakan sebagai media implementasi pipeline analisis sentimen yang dikembangkan dalam penelitian.

5. KESIMPULAN

Penelitian ini menganalisis 569 ulasan aplikasi Akademi Crypto di Google Play Store yang dikumpulkan pada periode Januari sampai Juni 2026. Berdasarkan pelabelan rating, diperoleh 317 ulasan positif, 230 ulasan negatif, dan 22 ulasan netral. Tahapan preprocessing yang meliputi cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming digunakan untuk menyiapkan teks sebelum pembobotan TF-IDF.

Model XGBoost yang menggunakan fitur TF-IDF menghasilkan accuracy 74,07%, precision 58,59%, recall 58,35%, dan F1-score 57,68% berdasarkan hasil evaluasi yang digunakan dalam penelitian. Kelas netral menjadi bagian yang paling sulit dikenali karena jumlah datanya jauh lebih sedikit daripada kelas positif dan negatif. Penelitian selanjutnya dapat memperluas jumlah dan periode data serta melakukan perbandingan dengan algoritma lain untuk memperoleh gambaran performa yang lebih komprehensif.

DAFTAR PUSTAKA

- Chandradev, V., Suarjaya, I. M. A. D., & Bayupati, I. P. A. (2023). Analisis sentimen review hotel menggunakan metode deep learning BERT. *Jurnal Buana Informatika*, 14(2), 107-116.
- Gilbert, G., Manurung, R., & Saputra, A. (2023). Analisis sentimen berdasarkan ulasan pengguna aplikasi MyPertamina pada Google Playstore menggunakan metode Naïve Bayes. *Jurnal Sistem Penyimpanan dan Teknologi Informasi*, 4(1), 21-29.
- Kurniawanda, M. R., & Tobing, F. A. T. (2022). Analysis sentiment cyberbullying in Instagram comments with XGBoost method. *IJNMT*, 9(1), 28-34.
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis sentimen ulasan aplikasi MyPertamina pada Google Play Store menggunakan algoritma NBC. *Jurnal Teknologi Terpadu*, 9(1), 42-48.
- Munandar, A. A., Farikhin, F., & Widodo, C. E. (2023). Sentimen analisis aplikasi belajar online menggunakan klasifikasi SVM. *JOINTECS*, 8(2), 77.
- Prasetyo, A. W., & Putra, R. E. (2024). Analisis sentimen Twitter pada cryptocurrency menggunakan BERT dan RCNN. *JINACS*, 6(1), 11-21.
- Prayitno, E., Suprawoto, T., & Riyanto, B. F. (2021). Optimasi hasil pencarian pada web scrapping menggunakan pembobotan TF-IDF. *Journal of Innovation Research and Knowledge*, 1(7), 241-246.
- Setiyono, B. N., Prasetyo, A., & Ramadhan, M. (2024). Analisis sentimen ulasan pengguna aplikasi Threads di Google Play menggunakan algoritma XGBoost dengan penguatan SMOTE. *Jurnal Informatika dan Teknologi Informasi*, 5(2), 112-120.
- Wardana, D. A. B., Nugroho, S., & Firdaus, M. (2024). Analisis sentimen ulasan pengguna aplikasi MyPertamina pada Google Play Store menggunakan metode XGBoost dan Word2Vec embedding. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(1), 155-163.
- Widiarta, I. P. A. P., Dwiyanaputra, R., & Aranta, A. (2023). Analisis sentimen masyarakat terhadap kebijakan penerapan PPKM di media sosial Twitter dengan menggunakan metode XGBoost. *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya*, 5(2), 154-163.



- Yang, Y., Siau, K., Xie, W., & Sun, Y. (2022). A systematic review of sentiment analysis: Approaches, applications, and challenges. *Frontiers in Psychology*, 13, 918447. <https://doi.org/10.3389/fpsyg.2022.918447>
- Zain, A. F., Rahman, F., & Hidayat, T. (2025). Analisis sentimen ulasan pengguna iPhone dengan pendekatan hibrida RoBERTa dan XGBoost. *Jurnal Algoritma*, 22(1), 45-53.
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis sentimen kebijakan Kampus Merdeka menggunakan Naive Bayes dan pembobotan TF-IDF berdasarkan komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1). <https://doi.org/10.25126/justsi.v2i1.24>