

# Analisis Sentimen Aplikasi MPStore Menggunakan Algoritma Logistic Regression dan LDA

<sup>1</sup>Tia Arlin Dita, <sup>2</sup>Rizka Rahmadhani, <sup>3</sup>Mira Afrina, <sup>4</sup>Ali Ibrahim

<sup>1,2,3,4</sup> Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia

[tiaarlindita3@gmail.com](mailto:tiaarlindita3@gmail.com); [rizka.rahmadhani02@gmail.com](mailto:rizka.rahmadhani02@gmail.com); [miraafrina@unsri.ac.id](mailto:miraafrina@unsri.ac.id); [aliibrahim@unsri.ac.id](mailto:aliibrahim@unsri.ac.id)

## Article Info

### Article history:

Received, 2025-12-04

Revised, 2025-12-20

Accepted, 2025-12-30

### Kata Kunci:

*latent\_dirichlet\_allocation  
regresi\_logistik  
analisis\_sentimen  
ulasan\_pengguna  
kepuasan\_pengguna*

## ABSTRAK

Pertumbuhan pesat ekonomi digital menjadikan kepuasan pengguna sebagai kunci sukses inovasi aplikasi. Dalam *technopreneurship*, umpan balik pengguna yang terekam pada ulasan aplikasi menjadi sumber informasi penting untuk memahami persepsi pasar, mengidentifikasi masalah, serta merumuskan prioritas perbaikan berbasis data. Karena volume ulasan yang besar sulit dianalisis secara manual, penelitian ini memanfaatkan analisis sentimen dan pemodelan topik untuk memetakan kepuasan pengguna terhadap aplikasi MPStore berdasarkan ulasan di *Google Play Store*. Data ulasan dikumpulkan melalui scraping, kemudian diklasifikasikan menjadi sentimen positif, negatif, dan netral menggunakan *Logistic Regression* (LR), serta diekstraksi faktor-faktor kepuasan menggunakan *Latent Dirichlet Allocation* (LDA). Proporsi sentimen pengguna didominasi 79% sentimen positif, 17% negatif, dan 4% netral. Hasil menunjukkan model LR mencapai akurasi 89,2%. Analisis LDA mengungkapkan delapan topik utama yang merepresentasikan faktor kepuasan dan keluhan pengguna, termasuk kemudahan penggunaan, kecepatan transaksi, serta kendala teknis (*error, login, saldo*). Secara keseluruhan, mayoritas pengguna menunjukkan persepsi positif terhadap efisiensi dan kemudahan transaksi MPStore. Penelitian ini menyimpulkan bahwa kombinasi analisis sentimen dan pemodelan topik efektif untuk menggambarkan tingkat kepuasan pengguna serta memberikan dasar strategis bagi pengembang aplikasi digital.

## ABSTRACT

### Keywords:

*latent\_dirichlet\_allocation  
logistic\_regression  
sentiment\_analysis  
user\_review  
user\_satisfaction*

*The rapid growth of the digital economy encourages user satisfaction as the key to successful application innovation. In technopreneurship, user feedback captured in app reviews serves as an important source of information for understanding market perceptions, identifying issues, and formulating data-driven improvement priorities. Because large volumes of reviews are difficult to analyze manually, this study employs sentiment analysis and topic modeling to map user satisfaction with the MPStore application based on reviews from the Google Play Store. Review data were collected via scraping and analyzed using Logistic Regression (LR) for positive, negative, and neutral sentiment classification, and Latent Dirichlet Allocation (LDA) for satisfaction topic extraction. User sentiment was dominated by 78% positive reviews, followed by 18% negative and 4% neutral. The result shows that the LR model achieved an accuracy of 89.2%. The LDA analysis also successfully revealed eight main topics, including ease of use, transaction speed, and technical obstacles (errors, login, balance issues). Overall, a majority of users hold a positive perception of MPStore's efficiency and ease of transactions. This study concludes that the combination of sentiment analysis and topic modeling is effective for explaining the level of user satisfaction and providing a strategic foundation for digital application developers.*

*This is an open-access article under the CC BY-SA license.*



### Penulis Korespondensi:

Ali Ibrahim,  
Program Studi Magister Ilmu Komputer,  
Universitas Sriwijaya Indonesia,  
Email: [aliibrahim@unsri.ac.id](mailto:aliibrahim@unsri.ac.id)

## 1. PENDAHULUAN

Aplikasi *mobile* berperan penting dalam strategi bisnis *modern* karena memungkinkan perusahaan membangun orientasi pelanggan digital secara *real-time* melalui umpan balik dan data penggunaan [1]. Di pasar yang kompetitif ini, bisnis harus dapat memahami kepuasan pengguna yang merupakan pendorong utama loyalitas dan pertumbuhan [2]. Membangun basis pelanggan yang loyal memerlukan strategi proaktif karena jika pelanggan tidak puas, mereka akan mencari dan beralih ke kompetitor lain [3].

Pada praktiknya, *app store* seperti *Google Play Store* menjadi wadah utama dalam penghantaran nilai karena langsung menghubungkan inovasi dengan kebutuhan pasar dan pengalaman pengguna sehari-hari [4]. Pengalaman pengguna dituliskan dalam ulasan aplikasi yang juga diperkaya dengan penilaian bintang, ini merupakan sumber informasi yang dapat membantu perusahaan memahami bagaimana preferensi dan pandangan pengguna terhadap aplikasi [4]. Umpan balik yang kaya atas pengalaman itu terekam dalam ulasan pengguna di *app store*, yang terbukti bahwa ulasan tersebut dapat mempengaruhi persepsi [1] dan unduhan aplikasi [5].

Namun, pesatnya perkembangan internet menghasilkan besarnya volume dan teks dengan bentuk yang tidak terstruktur menjadikan membaca ulasan satu per satu bukanlah solusi yang tepat, sehingga diperlukan pendekatan komputasional untuk mengekstrak makna secara otomatis [6]. Terdapat dua pendekatan yang dapat saling melengkapi untuk ini, pertama adalah analisis sentimen, sebuah subtugas dari penggalian opini dalam pemrosesan bahasa alami yang mengklasifikasikan konten tekstual berdasarkan nada emosional penulis [7]. Kedua adalah pemodelan topik yang merupakan teknik otomatis untuk mengekstraksi dan mengidentifikasi tema semantik laten atau topik terpenting dalam kumpulan dokumen teks [8]. Penggunaan kedua pendekatan memungkinkan untuk menangkap tren opini publik dan mengungkap tema laten atau faktor yang dibicarakan pengguna dalam konten yang tidak terstruktur [9].

Untuk mendukung proses klasifikasi sentimen, penelitian ini menggunakan *Logistic Regression* karena relatif sederhana, mudah diinterpretasikan, dan terbukti kompetitif dalam menangani data teks berskala besar [10]. Pada sejumlah penelitian juga menunjukkan performa yang akurat dan stabil, berdampingan dan bahkan mengungguli beberapa pendekatan lain sehingga relevan digunakan sebagai model utama analisis sentimen [10], [11]. Sementara itu, untuk mengeksplorasi tema yang dibicarakan, penelitian ini menggunakan *Latent Dirichlet Allocation* (LDA) yang umum dipakai untuk memetakan topik dominan dari ulasan dan mengaitkannya dengan aspek kualitas serta pengalaman aplikasi [12]. Pendekatan kombinasi analisis sentimen dan pemodelan topik telah diterapkan di berbagai domain aplikasi, misalnya pada aplikasi *banking* untuk memetakan aspek yang perlu ditingkatkan serta mengukur sentimen pengguna [13], [14]. Riset menyatakan bahwa ulasan negatif dari pengguna seringkali mendorong inovasi produk karena memberi sinyal masalah yang perlu ditangani dengan segera [15].

Hal ini sangat penting bagi aplikasi seperti MPStore sebagai produk wirausaha digital yang bergantung pada kepuasan dan kepercayaan pengguna untuk mempertahankan keberlanjutan usaha di tengah persaingan digital agar arah inovasi produk dapat disesuaikan dengan kebutuhan pasar secara *real-time* [10]. Meski demikian, penerapan analisis ulasan berbasis sentimen dan topik pada aplikasi yang secara spesifik mendukung Usaha Mikro, Kecil, dan Menengah (UMKM) dan pedagang tradisional di Indonesia masih relatif terbatas, sehingga generalisasi temuan dari domain lain ke konteks ini belum kuat. Oleh karena itu, penelitian ini berkontribusi dalam tiga aspek. Pertama, konteks penelitian memfokuskan pada aplikasi transaksi digital yang menyasar ekosistem UMKM Indonesia. Kedua, metode penelitian yang dipakai mengintegrasikan *Logistic Regression* dan LDA untuk mengevaluasi pengalaman pengguna secara komprehensif, tidak hanya mengetahui polaritas sentimen tetapi juga mengungkap tema atau isu utama yang membentuk kepuasan dan keluhan. Ketiga, temuan analitik tersebut dihubungkan dengan kebutuhan pengambilan keputusan pengembangan produk dalam *technopreneurship*, sehingga dapat menjadi dasar prioritas perbaikan berbasis *customer insight* yang berpotensi memperkuat keunggulan kompetitif dan keberlanjutan usaha [16].

## 2. METODE PENELITIAN

Analisis dilakukan melalui dua tahap utama yaitu, analisis sentimen menggunakan algoritma *Logistic Regression* (LR), dan analisis pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA). Pada Gambar 1 berikut alur penelitiannya:



Gambar 1. Alur Penelitian

Data dikumpulkan dari ulasan aplikasi pada *Google Play Store* melalui teknik *scraping* menggunakan pustaka *google-play-scraping* pada *python*. Dalam penelitian ini, data yang digunakan berupa ulasan pengguna aplikasi MPStore. Sebanyak 10.000 ulasan dikumpulkan dengan metode pengurutan berdasarkan *most relevant*.

Selanjutnya dilakukan pra-pemrosesan pada data. Data hasil *scraping* diolah ke dalam bentuk *data frame* dengan tiga kolom utama, yaitu *content* (isi ulasan), *score* (rating pengguna dengan skala 1–5), dan *at* (tanggal ulasan). Selanjutnya, kolom *at* dipecah menjadi tiga atribut waktu yaitu *year*, *month*, dan *day* untuk mempermudah analisis temporal, sementara kolom aslinya dihapus. Selain itu dilakukan penghapusan data duplikat berdasarkan kesamaan isi ulasan. Hal ini dilakukan untuk menghindari bias yang diakibatkan oleh pengulangan opini yang sama dari pengguna.

Distribusi ulasan menunjukkan bahwa ulasan dengan rating 5 mendominasi dengan total 7.419 entri, sedangkan rating 2 menjadi yang paling sedikit, yaitu 355 entri. Untuk keperluan analisis sentimen, skor numerik selanjutnya dikonversi menjadi tiga kategori sentimen dengan kriteria: positif (1) untuk skor lebih dari 3, netral (0) untuk skor sama dengan 3, dan negatif (-1) untuk skor kurang dari 3. Selanjutnya, penelitian ini melakukan *text cleaning*, *case folding*, *tokenization* dengan *RegexpTokenizer*, *stopword removal* dari pustaka *nltk*, *stemming* dari pustaka *Sastrawi*. Selain itu juga dilakukan juga filtering Bahasa dengan ulasan yang berbahasa Indonesia saja yang digunakan serta ulasan dengan jumlah kata kurang dari 3 kata dan ulasan kosong dihapus karena tidak mengandung informasi semantic yang cukup untuk dianalisis. Tahapan terakhir Adalah dilakukan *word length filtering* yaitu menghapus kata-kata yang memiliki panjang kurang dari 4 karakter pada teks ulasan yang telah didapat. Dengan teks yang bersih dan seragam, model pembobotan dan klasifikasi yang digunakan di tahapan selanjutnya dapat bekerja lebih optimal dan menghasilkan hasil analisis yang lebih akurat [17].

Analisis sentimen dilakukan menggunakan LR dengan pendekatan multinomial. Model ini dipilih karena mampu menangani lebih dari dua kelas sentimen (positif, netral, negatif) secara efisien serta memberikan interpretasi probabilistik yang jelas terhadap hasil prediksi [18]. Adapun tahapan proses yang dilakukan adalah pertama, pembagian data untuk pelatihan dan pengujian model sebesar 80% dan 20% data. Kedua, ekstraksi fitur teks menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk merepresentasikan setiap kata sebagai vektor numerik.

Kinerja model yang telah dikembangkan pada tahap analisis akan dievaluasi untuk memastikan tingkat akurasi hasil. Pada tahap analisis sentimen dengan LR, kinerja model diukur menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*, guna menilai keseimbangan kemampuan klasifikasi secara menyeluruh [19]. Setelah model sentimen menunjukkan performa yang memadai, tahap berikutnya adalah melakukan analisis topik guna memahami konteks yang lebih luas dari opini pengguna.

Penelitian dilanjutkan dengan menerapkan LDA untuk mengekstraksi topik-topik utama yang sering muncul dalam ulasan pengguna menggunakan *topic coherence score*. Dengan dilakukannya pengujian beberapa variasi topik, maka didapatkan topik optimanya Adalah 8 topik dengan nilai *coherence score* tertinggi sebesar 0,04979. Pendekatan ini bertujuan untuk memahami konteks di balik opini pengguna, sehingga dapat memberikan wawasan yang lebih mendalam mengenai aspek-aspek yang paling sering menjadi perhatian dalam penggunaan aplikasi MPStore. Setiap topik yang terbentuk akan dianalisis secara semantik untuk mengidentifikasi faktor penentu kepuasan, seperti kinerja aplikasi, kemudahan penggunaan, layanan pelanggan, ataupun stabilitas sistem [20].

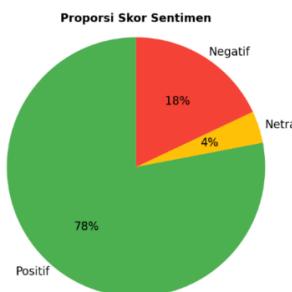
### 3. HASIL DAN ANALISIS

Tahapan awal dalam Penelitian ini adalah pra-pemrosesan data ulasan pengguna aplikasi MPStore. Berdasarkan hasil rekapitulasi skor ulasan, distribusi penilaian pengguna menunjukkan dominasi skor yang tinggi khususnya pada skor 5 dengan jumlah ulasannya Adalah 7.419 ulasan dan skor rendah seperti skor 1 dan 2 masih muncul meskipun dalam jumlah yang lebih sedikit. Hasil rekapitulasi ulasan tersebut dapat dilihat pada Tabel 1.

Tabel 1.Distribusi Skor Ulasan

Skor	Hasil
5	7419
4	367
3	417
2	355
1	1423

Setelah dilakukan konversi skor numerik menjadi kategori sentimen berdasarkan kriteria, diperoleh proporsi kategori sentimen seperti pada Gambar 2. Ini menunjukkan bahwa sebagian besar pengguna memberikan ulasan dengan nada positif terhadap aplikasi MPStore dan memberikan gambaran umum mengenai variasi pengalaman yang terekam dalam data ulasan.



Gambar 2. Proporsi Skor Sentimen

Data-data teks ulasan kemudian diproses ke dalam serangkaian tahapan pra-pemrosesan dan menghasilkan kolom *text\_string*, yang merupakan representasi teks yang bersih, padat, dan siap untuk pemodelan topik atau klasifikasi sentimen. Perbedaan teks awal dan hasil akhir dari pra-pemrosesan dapat dilihat seperti pada Tabel 2 berikut ini.

Tabel 2. Hasil Pra-Pemrosesan

Kalimat Ulasan Awal (content)	Hasil Akhir Kalimat (text_string)
aplikasi yang sangat bagus, sangat membantu untuk usaha sampingan saya, terima kasih m pulsa	bagus bantu usaha samping terima kasih pulsa
apaan, ga guna daftar aja susah	daftar susah
aplikasi cocok untuk jualan pulsa dan paket data karena harga yang terjangkau bisa ambil untung banyak	cocok jual pulsa paket data harga jangkau ambil untung
harga produknya masih mahal dri aplikasi sbelah	harga produk mahal sbelah
bikin segala nya jadi lebih mudah dan praktis saya bisa jualan dimana saja dan kapan saja, kalo bisa tambahan fitur transfer ke antar bank ke bank lain nya agar bisa lebih mempermudah pembayaran	bikin mudah praktis jual mana kalo tambah fitur transfer bank bank mudah bayar
...	...

Melalui langkah pra-pemrosesan ini, data teks yang semula tidak terstruktur berhasil dinormalisasi dan disiapkan dengan baik untuk masuk ke proses analisis selanjutnya. Tahap ini memastikan bahwa setiap ulasan telah berada dalam format yang konsisten sehingga dapat diolah secara efektif oleh model klasifikasi.

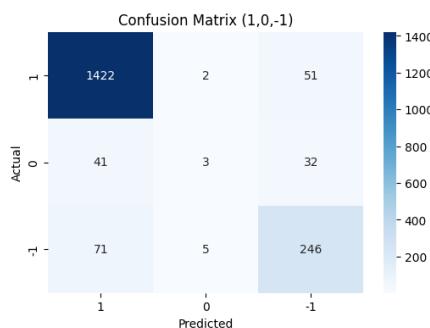
Pada tahapan analisis sentimen data dibagi menjadi 80% data latih dan 20% data uji. Model *Logistic Regression* dan TF-IDF digunakan sebagai metode ekstraksi fitur. Parameter utama yang digunakan dalam penelitian ini ditunjukkan pada Gambar 3. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 89,2% dengan nilai *precision*, *recall* dan *f1-score* untuk masing-masing kelas sentimen dapat dilihat pada Tabel 3, sedangkan untuk distribusi hasil prediksi ditampilkan pada *confusion matrix* yang dapat dilihat pada Gambar 4.

```
LogisticRegression
LogisticRegression(max_iter=2000, multi_class='multinomial', n_jobs=-1,
random_state=42, solver='saga')
```

Gambar 3. Konfigurasi Model *Logistic Regression*

Tabel 3. Hasil Evaluasi Model

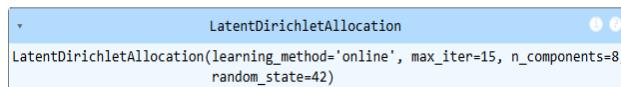
Kelas Sentimen	Precision	Recall	F1-Score	Jumlah Data
Negatif (-1)	0.74	0.76	0.75	322
Netral (0)	0.30	0.03	0.06	76
Positif (1)	0.92	0.96	0.94	1475



Gambar 4. Confusion Matrix

Hasil evaluasi ini menunjukkan bahwa kelas sentimen positif memiliki nilai *precision*, *recall* dan *f1-score* tertinggi dibandingkan dengan kelas lainnya yaitu 0,92, 0,96 dan 0,94. Sebaliknya pada kelas sentimen netral mendapatkan nilai *precision*, *recall* dan *f1-score* yang relatif rendah yaitu 0,30, 0,03 dan 0,06. Pada Gambar 4, model menunjukkan kemampuan yang sangat baik dalam mendeteksi sentimen positif, terlihat dari jumlah prediksi benar yang dominan pada kelas tersebut. Hal ini menunjukkan adanya kecenderungan model untuk mengklasifikasikan sebagian besar ulasan sebagai sentimen positif, termasuk beberapa ulasan yang sebenarnya bersifat netral atau negatif ringan.

Sebagai langkah berikutnya, dilakukan analisis topik menggunakan LDA. LDA dikonfigurasi dengan parameter utama sebagaimana ditunjukkan pada Gambar 5, yang menggambarkan pengaturan model dalam proses pembentukan topik.



Gambar 5. Konfigurasi Model LDA

Selanjutnya dilakukan pengujian beberapa variasi topik, maka didapatkan topik optimanya adalah 8 topik dengan nilai *coherence score* tertinggi sebesar 0,04979. Penggunaan *online learning* juga membuat model dapat memproses data dalam jumlah besar dengan efisien, sedangkan batas iterasi sebanyak 15 memastikan model mencapai kondisi stabil atau konvergen secara optimal. Hasil analisis menggunakan metode LDA menunjukkan bahwa ulasan pengguna aplikasi MPStore dapat dikelompokkan ke dalam topik-topik utama seperti yang ditampilkan pada Tabel 4. Setiap topik merepresentasikan tema tertentu yang sering muncul dalam ulasan pengguna, mulai dari aspek kemudahan penggunaan, harga, hingga kendala teknis yang dialami.

Tabel 4. Hasil Analisis Menggunakan Metode LDA

Topik ID	Kata Dominan Topik	Contoh Review	Interpretasi Tema Utama
Topic_0	kode, daftar, masuk, login, referral, kode referral, tolong.	agen pulsa untung inti best banget sukses trus banyak agen pulsa pm3196; kerenn update versi harga murah murah bngung jaring unlimited hasil unlimited batas sukses mpulsa.	Registrasi, login dan penggunaan kode referral dalam aplikasi
Topic_1	pulsa, moga, mpulsa, bantu, lancer, cepat, sukses.	simpel mudah tangga terimakasih pulsa sukses; aplikasi sangat membantu mudah percaya mantab.	Kemudahan penggunaan aplikasi dan kelancaran transaksi pulsa
Topic_2	pulsa, jual, banget, bagus, token, jual pulsa, listrik	kalo transaksi kalo token gagal2 mulu; proses kirim pulsa lelet.	Kendala transaksi token Listrik dan keterlambatan pengiriman pulsa
Topic_3	pulsa, jual, jual pulsa, bantu, kasih, baik, terima.	pulsa muas keran pulsa murah; cocok jual pulsa csnnya bantu alam.	Harga pulsa yang kompetitif dan respon layanan pelanggan
Topic_4	usaha, download, cocok, buru, nyesel, samping, kali.	mudah simpel cocok usaha; kecewa nyesel download.	Kesesuaian aplikasi untuk usaha dan pengalaman pengguna awal
Topic_5	bagus, cepat, harga, transaksi, murah, harga murah, bisnis	aplikasi bagus banget pulsa cepat paket muraah; bagus admin layanin biarpun harga pulsa saing bagus mantap nyesel mm1394.	Kecepatan transaksi dan persepsi nilai harga layanan
Topic_6	saldo, masuk, bintang, udah, mohon, kecewa, karna.	udah transfer saldo saldo tetep kosong; saldo indomart mohon admin jelas.	Masalah saldo dan harapan pengguna terhadap penanganan admin
Topic_7	pulsa, murah, mudah, harga, banget, lengkap, paket.	mantap fitur mudah bayar; beli paket murah rekomendet.	Kemudahan fitur dan kelengkapan produk dengan harga terjangkau

Berdasarkan hasil analisis topik menggunakan metode LDA didapatkan bahwa ulasan pengguna MPStore mencakup berbagai tema yang merepresentasikan pengalaman dan persepsi pengguna terhadap aplikasi. Ulasan positif umumnya membahas tentang kemudahan penggunaan, kecepatan transaksi, serta harga pulsa atau paket data yang tergolong murah. Beberapa pengguna juga menyoroti layanan pelanggan yang responsif dan kemudahan dalam menjalankan usaha kecil melalui aplikasi. Sementara itu, ulasan dengan sentimen negatif banyak menyoroti masalah teknis, seperti error saat pembaruan aplikasi, kendala dalam proses login, saldo yang tidak masuk, atau perbedaan harga antara tampilan dan hasil transaksi. Adapun ulasan netral cenderung bersifat deskriptif atau informatif tanpa mengekspresikan emosi kuat, seperti hanya menyebutkan fungsi aplikasi, pengalaman umum, atau saran perbaikan ringan. Temuan LDA ini juga diperkuat dan tervalidasi melalui visualisasi *word cloud* dari pengelompokan sentimen awal pada Gambar 6 dan Gambar 7.



Gambar 6. *Word Cloud Sentimen Positif*



Gambar 7. Word Cloud Sentimen Negatif

Pada Gambar 6, tampak bahwa kata-kata seperti mudah, murah, dan mantap sering muncul dalam ulasan dengan sentimen positif. Pola ini menunjukkan bahwa pengguna umumnya merasa puas terhadap kemudahan penggunaan serta performa layanan aplikasi. Sementara itu, pada Gambar 7, terlihat kemunculan kata-kata seperti kecewa, masuk, dan transaksi yang lebih banyak muncul dalam ulasan bernada negatif, menggambarkan adanya kendala teknis yang masih dialami sebagian pengguna. Pola ini menegaskan bahwa persepsi pengguna terhadap MPStore sangat dipengaruhi oleh dua faktor utama, yaitu kemudahan penggunaan dan kinerja teknis aplikasi.

Dominasi skor tinggi dan sentimen positif menunjukkan bahwa secara umum pengguna memiliki persepsi yang lebih baik terhadap aplikasi MPStore. Selain itu ulasan yang bersifat negatif dan netral juga penting untuk mempresentasikan pengalaman pengguna yang kurang optimal dan menjadi sumber informasi untuk melakukan evaluasi system. Distribusi sentimen yang tidak seimbang dengan dominasi kelas positif juga berpengaruh bagi kinerja model klasifikasi.

Model *Logistic Regression* dengan TF-IDF mendapatkan hasil akurasi yang tinggi yaitu 89,2%. Hal ini mengindikasikan bahwa kombinasi metode ini efektif dalam menangkap pola linguistic pada ulasan pengguna. Akan tetapi, rendahnya performa dalam melakukan sentimen netral menunjukkan bahwa adanya bias klasifikasi akibat ketidak seimbangan data. Model cenderung mengarahkan prediksi pada kelas mayoritas, sehingga ulasan netral sering salah diklasifikasikan sebagai positif.

Meskipun akurasi model tinggi, potensi risiko *overfitting* tetap perlu diperhatikan. Evaluasi dilakukan menggunakan skema *train test split* tanpa penerapan *k-fold cross-validation*, sehingga hasil yang diperoleh masih berpotensi dipengaruhi oleh karakteristik data uji. Penerapan validasi silang serta Teknik penanganan data tidak seimbang, seperti *class weighting* atau *resampling* berpotensi meningkatkan kemampuan generalisasi model pada penelitian selanjutnya.

Berdasarkan hasil evaluasi, model *Logistic Regression* (LR) yang dikombinasikan dengan TF-IDF dalam penelitian ini menunjukkan kinerja yang sangat baik dengan akurasi 89,2%. Ketika dibandingkan dengan penelitian terdahulu, akurasi ini terbukti sangat kompetitif dan bahkan mengungguli banyak pendekatan lain. Secara spesifik, akurasi penelitian ini lebih tinggi daripada penelitian yang menggunakan metode *machine learning* klasik lainnya pada ulasan Google Play Store. Sebagai contoh, model ini melampaui *Support Vector Machine* (SVM) dengan akurasi sebesar 85% [21], *K-Nearest Neighbor* (KNN) 84,68% [22], dan *Naive Bayes* (NB) 83% [23]. Kinerja ini juga lebih unggul dibandingkan penelitian lain yang menerapkan KNN dengan akurasi sebesar 81% dan 80% pada aplikasi Identitas Kependudukan Digital [24], SVM 80% pada aplikasi Deepseek [25], NB 80% pada aplikasi inDrive [26], dan SVM 77% pada aplikasi iGrow [27].

Perbandingan yang paling relevan adalah dengan penelitian Gupta dan Kamthania [28] yang juga menggunakan LR untuk analisis sentimen aplikasi, penelitian tersebut mencapai akurasi 81,1%. Akurasi 89,2% yang dicapai dalam penelitian ini secara signifikan lebih tinggi, yang mengindikasikan bahwa proses pra-pemrosesan dan penerapan model pada dataset ulasan MPStore sangat efektif.

Menariknya, model LR dalam penelitian ini menunjukkan kinerja unggul dengan akurasi 89,2%, melampaui *Convolutional Neural Network* (CNN) 84,67%, *Long Short-Term Memory* (LSTM) 82,19% [29], serta model berbasis *Transformer* seperti *Bidirectional Encoder Representations from Transformers* (BERT) sebesar 71,54% dan *DistilBERT* 71,68%[30]. Hasil ini membuktikan bahwa model machine learning klasik yang dioptimalkan dengan baik masih dapat mengungguli arsitektur deep learning yang lebih kompleks dalam klasifikasi sentimen ulasan aplikasi. Selain itu, pencapaian akurasi tinggi meskipun mempertahankan tiga kelas (positif, negatif, netral) menunjukkan robustisitas model, meski kelas netral tetap menjadi tantangan akibat ketidakseimbangan data. Secara keseluruhan, performa model sangat dipengaruhi oleh kualitas data, keseimbangan kelas, parameter, dan pra-pemrosesan teks, sehingga model sederhana pun dapat mencapai hasil kompetitif jika dikelola dengan baik.

Hasil analisis topik menggunakan LDA memberikan konteks yang melengkapi analisis sentimen. Topik-topik yang dihasilkan dari proses ini menunjukkan bahwa ulasan positif banyak berkaitan dengan kemudahan pengguna, harga yang terjangkau dan kecepatan dalam transaksi. Sementara itu ulasan negatif didominasi oleh permasalahan teknis seperti saldo, login dan keterlambatan dalam transaksi. Dengan dilakukannya visualisasi *word cloud* dapat memperkuat hasil ini dengan menunjukkan perbedaan kosakata yang menonjol antara sentimen *positif* dan *negatif*.

Penelitian ini memberikan kebaruan melalui penggunaan dataset ulasan yang berskala besar yaitu 10.000 data serta integrasi analisis sentimen dan analisis topik dalam satu kerangka analisis. Dengan dilakukannya pendekatan ini dapat memungkinkan pemahaman yang lebih komprehensif terhadap persepsi pengguna, baik dari sisi arah opini maupun isu yang mendasarinya. Secara praktis hasil penelitian ini dapat dimanfaatkan oleh pengembang MPStore sebagai dasar pengambilan Keputusan strategis baik untuk mempertahankan keunggulan layanan atau untuk memprioritaskan perbaikan pada aspek teknis yang paling sering dikeluhkan pengguna.

#### 4. KESIMPULAN

Penelitian ini menegaskan bahwa pemanfaatan ulasan pengguna sebagai sumber data mampu memberikan gambaran yang komprehensif mengenai tingkat kepuasan MPStore. Hasil analisis ini menunjukkan bahwa pendekatan klasifikasi berbasis *Logistic Regression* yang didukung pra-pemrosesan teks yang baik mampu menghasilkan performa yang kompetitif, bahkan dibandingkan dengan beberapa pendekatan yang lebih kompleks. Temuan ini menegaskan bahwa model *machine learning* klasik yang dioptimalkan dengan pra-pemrosesan yang baik akan menghasilkan akurasi yang baik untuk analisis teks ulasan berbahasa Indonesia. Pemodelan topik dengan menggunakan LDA berhasil mengungkap pola utama pengalaman pengguna, dimana kepuasan yang paling utama dipengaruhi oleh kemudahan penggunaan, kecepatan transaksi, dan harga yang terjangkau. Sedangkan sentimen negatif banyak dipengaruhi oleh kendala teknis seperti masalah saldo, login, dan error pada transaksi. Sehingga pola ini menunjukkan bahwa kepuasan pengguna sangat bergantung pada keseimbangan antara kemudahan operasional dan stabilitas sistem. Secara keseluruhan, penelitian ini memberikan kontribusi ilmiah dengan melalui integrasi analisis sentimen dan pemodelan topik dalam satu kerangka analisis berbasis dataset ulasan berskala besar serta kontribusi praktis sebagai dasar pengambilan keputusan bagi pengembang MPStore dalam memprioritaskan perbaikan teknis tanpa mengabaikan keunggulan layanan yang telah dinilai baik oleh pengguna. Penelitian selanjutnya disarankan untuk menitikberatkan pada penanganan ketidakseimbangan data serta pengembangan fitur semantik untuk menangkap ulasan secara lebih mendalam.

#### REFERENSI

- [1] L. Stocchi, N. Pourazad, N. Michaelidou, A. Tanusondjaja, and P. Harrigan, "Marketing research on Mobile apps: past, present and future," *J Acad Mark Sci*, vol. 50, no. 2, pp. 195–225, Mar. 2022, doi: 10.1007/s11747-021-00815-w.

- [2] P. Sun, L. Li, M. S. Hossain, S. Ray, and K. A. Law, "Predicting and explaining customer satisfaction: A deep learning and sentiment analysis of emotional impacts," *Acta Psychol (Amst)*, vol. 260, p. 105597, 2025, doi: <https://doi.org/10.1016/j.actpsy.2025.105597>.
- [3] T. E. Mofokeng, "Mechanisms of building customer loyalty: Mediation of customer satisfaction with online retailing in South Africa," *Telematics and Informatics Reports*, vol. 20, p. 100262, 2025, doi: <https://doi.org/10.1016/j.teler.2025.100262>.
- [4] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Mining and searching app reviews for requirements engineering: Evaluation and replication studies," *Inf Syst*, vol. 114, p. 102181, 2023, doi: <https://doi.org/10.1016/j.is.2023.102181>.
- [5] H. Sällberg, S. Wang, and E. Numminen, "The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and productivity apps from their initial app store launch," *Journal of Marketing Analytics*, vol. 11, no. 3, pp. 426–442, 2023, doi: 10.1057/s41270-022-00171-w.
- [6] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, p. 102048, 2024, doi: <https://doi.org/10.1016/j.jksuci.2024.102048>.
- [7] F. Ye and Z. Zhao, "Scalable vortex search-tuned intelligent adaptive boosting for sentiment analysis of social media data using natural language processing," *Systems and Soft Computing*, vol. 7, p. 200403, 2025, doi: <https://doi.org/10.1016/j.sasc.2025.200403>.
- [8] G. Ibba, R. Neykova, M. Ortú, R. Tonelli, S. Counsell, and G. Destefanis, "A machine learning approach to vulnerability detection combining software metrics and topic modelling: Evidence from smart contracts," *Machine Learning with Applications*, vol. 22, p. 100759, 2025, doi: <https://doi.org/10.1016/j.mlwa.2025.100759>.
- [9] E. C. M. Torres and L. G. de Picado-Santos, "Sentiment Analysis and Topic Modeling in Transportation: A Literature Review," *Applied Sciences*, vol. 15, no. 12, 2025, doi: 10.3390/app15126576.
- [10] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120340.
- [11] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
- [12] I. Z. Mustaqim and R. R. Suryono, "A Systematic Literature Review of Topic Modeling Techniques in User Reviews," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 2, pp. 238–253, Jul. 2025, doi: 10.20473/jisebi.11.2.238-253.
- [13] M. Eksa Permana, H. Ramadhan, I. Budi, A. Budi Santoso, and P. Kresna Putra, "Sentiment Analysis and Topic Detection of Mobile Banking Application Review," in *Fifth International Conference on Informatics and Computing*, 2020, pp. 1–6. doi: 10.1109/ICIC50835.2020.9288616.
- [14] E. S. Asyhar, S. H. Wijoyo, and N. Y. Setiawan, "Analisis Sentimen dan Pemodelan Topik Terhadap Ulasan Aplikasi Jenius Menggunakan Metode Support Vector Machine dan Latent Dirichlet Allocation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 9, Aug. 2024.
- [15] L. Sun, Y. He, and F. Fu, "Impact of online negative and positive reviews on app innovation," *PLoS One*, vol. 20, no. 5, p. e0323205, May 2025, doi: 10.1371/journal.pone.0323205.
- [16] M. Grimaldi, O. Troisi, A. Papa, and E. de Nuccio, "Conceptualizing data-driven entrepreneurship: from knowledge creation to entrepreneurial opportunities and innovation," *J Technol Transf*, 2025, doi: 10.1007/s10961-024-10176-5.
- [17] O. Efrata Madao, A. Irsyad, and M. Rivani Ibrahim, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Jamsostek Mobile Dengan Menggunakan Naïve Bayes Dan Logistic Regretion," *Djtechno: Jurnal Teknologi Informasi*, vol. 6, no. 2, pp. 458–472, Aug. 2025, doi: 10.46576/djtechno.v6i2.6775.
- [18] R. Prabowo, H. Sujaini, and T. Rismawan, "Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 366, Jul. 2023, doi: 10.26418/justin.v11i2.57449.
- [19] D. Amangeldi, A. Usmanova, and P. Shamoi, "Understanding Environmental Posts: Sentiment and Emotion Analysis of Social Media Data," *IEEE Access*, vol. 12, pp. 33504–33523, 2024, doi: 10.1109/ACCESS.2024.3371585.
- [20] P. Ç. Çaylak *et al.*, "Analysing Online Reviews Consumers' Experiences of Mobile Travel Applications with Sentiment Analysis and Topic Modelling: The Example of Booking and Expedia," *Applied Sciences*, vol. 14, no. 24, 2024, doi: 10.3390/app142411800.
- [21] C. Wulandari, L. Sunardi, and Hasbiana, "Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine," *Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 5, pp. 2588–2595, Apr. 2024.

- [22] Y. T. Gosal, O. N. Pratiwi, I. Y. Mukti, and D. Witarsyah, “Analisis Sentimen Dan Pemodelan Topik Terhadap Aplikasi Pembelajaran Online Pada Platform Google Play (Studi Kasus: Quipper),” *Jurnal Ilmiah Komputer & Sistem Informasi*, vol. 23, no. 4, pp. 469–476, Dec. 2024, doi: 10.32409/jikstik.23.4.3663.
- [23] R. R. Rismansyah, A. Sudiarjo, and T. Mufizar, “Analisis Sentimen Ulasan Shopee Pada Google Play Store Menggunakan Algoritma Naive Bayes,” *Jurnal Elektro dan Informatika Swadharma*, vol. 5, no. 1, 2025.
- [24] R. Kurniawan, H. O. L. Wijaya, and R. P. Aprisusanti, “Sentiment Analysis of Google Play Store User Reviews on Digital Population Identity App Using K-Nearest Neighbors,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 2, pp. 170–178, Jun. 2024, doi: 10.32736/sisfokom.v13i2.2071.
- [25] E. S. Lalupanda, P. A. R. L. Lede, and I. Priyastiti, “Application of Support Vector Machine Algorithm for Sentiment Analysis of Deepseek App User Reviews on Google Play Store,” *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 5, no. 1, pp. 390–396, Oct. 2025, doi: 10.59934/jaiea.v5i1.1332.
- [26] R. Pratama, “Sentiment Analysis Of The Indrive Application Reviews On Google Play Store Using The Naive Bayes Method,” *Journal of Advances in Computational Intelligence and Information Systems*, vol. 1, no. 1, pp. 10–23, Feb. 2025.
- [27] M. Idris and Mussalimun, “Sentiment Analysis of Google Play Store Reviews using Support Vector Machines,” *Int J Appl Inf Syst*, vol. 12, no. 42, pp. 48–53, Jan. 2024.
- [28] A. Gupta and D. Kamthania, “Study of Sentiment on Google Play Store Applications,” in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, SSRN, Apr. 2021. doi: 10.2139/ssrn.3833926.
- [29] M. P. Ningrum, R. Mutia, H. Azmi, and H. D. Khalifah, “Sentiment Analysis of Twitter Reviews on Google Play Store Using a Combination of Convolutional Neural Network and Long Short-Term Memory Algorithms,” *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 2, pp. 107–115, Jan. 2025, doi: 10.57152/predatecs.v2i2.1625.
- [30] G. Eser and C. Sahin, “Sentiment Analysis and Rating Prediction for App Reviews Using Transformer-based Models,” in *3rd International Conference on Engineering, Natural and Social Sciences ICENSOS*, International Journal of Advanced Natural Sciences and Engineering Researches, May 2024, pp. 372–379.