

**ANALISIS KEPUASAN PENGGUNA APLIKASI GRAMEDIA  
DIGITAL MENGGUNAKAN METODE SUPPORT VECTOR  
MACHINE (SVM)**

**SKRIPSI**



**Oleh:**

**PRIAJI DEWA SASIKIRANA**

**NIM. 220607110011**

**PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**ANALISIS KEPUASAN PENGGUNA APLIKASI GRAMEDIA  
DIGITAL MENGGUNAKAN METODE SUPPORT VECTOR  
MACHINE (SVM)**

**SKRIPSI**

**Oleh:**

**PRIAJI DEWA SASIKIRANA  
NIM. 220607110011**

**Diajukan Kepada:**

**Fakultas Sains dan Teknologi**

**Universitas Islam Negeri Maulana Malik Ibrahim Malang**

**untuk Memenuhi Salah Satu Persyaratan Dalam**

**Memperoleh Gelar Sarjana Sains dan Informasi (S.S.I)**

**PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2026**

**HALAMAN PERSETUJUAN**

**ANALISIS KEPUASAN PENGGUNA APLIKASI GRAMEDIA DIGITAL  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)**

**SKRIPSI**

**Oleh:**

**PRIAJI DEWA SASIKIRANA  
NIM. 220607110011**

Telah diperiksa dan disetujui untuk diuji :

Tanggal : 15 April 2026

**Pembimbing 1**



**Fakhri Khusnu Reza Mahfud, M.Kom**  
NIP. 199005062019031007

**Pembimbing 2**



**Ach. Nizam Rifqi, M.A.**  
NIP. 199206092022031002

**Mengetahui,**

**Ketua Program Studi Perpustakaan dan Sains Informasi  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang**



**Nita Siti Muylawamah, M.IP**  
NIP. 199002232018012001

**HALAMAN PENGESAHAN**

**ANALISIS KEPUASAN PENGGUNA APLIKASI GRAMEDIA DIGITAL  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)**

**SKRIPSI**

**Oleh:**

**PRIAJI DEWA SASIKIRANA  
NIM. 220607110011**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan Untuk Memperoleh  
Gelar Sarjana Sains Informasi (S.S.I)  
Pada Tanggal 15 April 2026

**Susunan Dewan Penguji**

**Ketua Penguji : Dr. Ir. Mokhammad Amin Hariyadi, M.T  
NIP. 196701182005011001**

**Anggota Penguji I : Nita Siti Mudawamah, M.IP  
NIP. 199002232018012001**

**Anggota Penguji II : Fakhris Khusnu Reza Mahfud, M.Kom  
NIP. 199005062019031007**

**Anggota Penguji III : Ach. Nizam Rifqi, M.A.  
NIP. 199206092022031002**

**Tanda Tangan**

  
(.....)

  
(.....)

  
(.....)

  
(.....)

**Mengetahui,**  
**Ketua Program Studi Perpustakaan dan Sains Informasi**  
**Fakultas Sains dan Teknologi**  
**Universitas Islam Negeri Maulana Malik Ibrahim Malang**

  
**Nita Siti Mudawamah, M.IP**  
**NIP. 199002232018012001**

## PERNYATAAN KEASLIAN TULISAN

Yang bertanda tangan di bawah ini:

Nama : Priaji Dewa Sasikirana

NIM : 220607110011

Program Studi : Perpustakaan dan Sains Informasi

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dalam bentuk kutipan yang sudah disebutkan sumbernya pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 April 2026

Yang membuat pernyataan



Priaji Dewa Sasikirana

NIM.220607110011

## KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Analisis Kepuasan Pengguna Aplikasi Gramedia Digital Menggunakan Metode Support Vector Machine (SVM)”**. Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar sarjana di Program Studi Perpustakaan dan Sains Informasi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Penulis menyadari bahwa proses penyusunan skripsi ini tidak lepas dari berbagai tantangan, sehingga keberhasilan penelitian ini sangat dipengaruhi oleh bantuan, arahan, serta dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua, Bapak Agung Harimurti dan Ibu Lungsi Lestari serta keluarga besar yang telah memberikan dukungan dari segi moral, materi maupun spiritual. Serta kakak penulis Aninda Eka Agustiningsih yang telah memberikan semangat kepada penulis
2. Ibu Nita Siti Mudawamah, M.IP selaku Ketua Program Studi Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang, serta menjadi Dosen Penguji II yang telah memberikan masukan dan saran pada setiap tahap pengerjaan skripsi hingga selesai
3. Bapak Fakhris Khusnu Reza Mahfud, M.Kom selaku Dosen Pembimbing I yang telah memberikan waktu, bimbingan dan nasihat selama proses pengerjaan skripsi, sehingga tulisan ini dapat terselesaikan
4. Bapak Ach. Nizam Rifqi, M.A. selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk membimbing pengerjaan skripsi ini
5. Bapak Dr. Ir. M. Amin Hariyadi, M.T selaku Dosen Penguji II yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai
6. Seluruh dosen dan staff Program Studi Perpustakaan dan Sains Informasi yang telah memberikan ilmu dan pengalamannya

7. Ibu Rektor UIN Maulana Malik Ibrahim Malang beserta staff dan Seluruh dekanat Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang
8. Seluruh mahasiswa Program Studi Perpustakaan dan Sains Informasi angkatan 2022 (GARASENA)
9. Teman-teman HMPS Perpustakaan dan Sains Informasi Periode 2024
10. Teman-teman Kontrakan Rumah Iyus, yang berisi Iyus, Zaka, Dimas, Singgi, Efendi telah memberikan tempat tinggal untuk beristirahat, berkeluh kesah dan berjuang bersama
11. Teman-teman Jemblungers yang telah memberikan ruang untuk berdiskusi dan berbagi kebahagiaan
12. Terakhir, penulis mengucapkan terima kasih kepada diri sendiri yang telah tetap semangat dan berjuang dalam menyelesaikan skripsi ini

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat berbagai keterbatasan dan kekurangan. Penulis berharap skripsi ini dapat memberikan manfaat, menambah pengetahuan, serta menjadi wawasan bagi para pembaca, termasuk bagi penulis sendiri. Amin ya Rabbal ‘Alamin.

Wassalamu’alaikum Wr.Wb.

Malang, 15 April 2026

Penulis

Priaji Dewa Sasikirana

## ABSTRAK

Sasikirana, Priaji Dewa, 2026. **Analisis Kepuasan Pengguna Aplikasi Gramedia Digital Menggunakan Metode Support Vector Machine (SVM)**. Skripsi. Program Studi Perpustakaan dan Sains Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Pembimbing: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Ach. Nizam Rifqi, M.A

Kata kunci: *analisis sentimen, kepuasan membaca pengguna, Gramedia Digital, Support Vector Machine*

Gramedia Digital merupakan platform layanan baca berbasis digital yang hadir sebagai respons terhadap kemajuan teknologi. Penelitian ini bertujuan untuk menganalisis kepuasan pengguna aplikasi Gramedia Digital melalui analisis sentimen terhadap ulasan pengguna di Google Play Store menggunakan metode Support Vector Machine (SVM). Ulasan pengguna menunjukkan adanya ketidaksesuaian antara harapan dan pengalaman, seperti kendala teknis berupa logout otomatis, keterbatasan fitur, serta gangguan stabilitas aplikasi. Data yang digunakan sebanyak 1032 ulasan dalam rentang waktu Januari 2020 hingga Mei 2025. Metode penelitian menggunakan pendekatan *text mining* dengan tahapan pengumpulan data, pelabelan oleh ahli bahasa, *preprocessing* (*case folding, cleaning, normalisasi, tokenizing, stopword removal, dan stemming*), pembobotan TF-IDF, serta klasifikasi menggunakan SVM. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik akurasi, presisi, *recall*, dan *F1-score* melalui tiga skenario pembagian data (70:30, 80:20, dan 90:10). Skenario terbaik adalah 70:30 dengan akurasi 81% yang menunjukkan kesesuaian antara hasil prediksi dan label aktual, presisi 82% yang menunjukkan ketepatan prediksi positif, *recall* 78% yang menunjukkan kemampuan model dalam mengenali data positif, serta *F1-score* 79% yang mencerminkan keseimbangan antara presisi dan *recall*. Hasil penelitian ini menunjukkan bahwa metode SVM mampu melakukan analisis sentimen dengan cukup baik. Berdasarkan analisis sentimen, kepuasan membaca pengguna menunjukkan bahwa sebagian besar pengguna merasa belum puas, ditandai oleh lebih banyak ulasan negatif dibanding ulasan positif serta keluhan terkait keterbatasan fitur, stabilitas aplikasi, dan kendala akses setelah pembayaran. Temuan ini menunjukkan bahwa analisis sentimen dapat digunakan sebagai alat untuk memahami persepsi dan kepuasan membaca pengguna, serta memberikan rekomendasi bagi Gramedia Digital dalam meningkatkan kualitas layanan aplikasi.

## ABSTRACT

Sasikirana, Priaji Dewa, 2026. **The Analysis of User Satisfaction on the Gramedia Digital Application Using the Support Vector Machine (SVM) method.** Thesis. Library and Information Science, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang Advisors: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Ach. Nizam Rifqi, M.A

Keywords: *sentiment analysis, user reading satisfaction, Gramedia Digital, Support Vector Machine*

Gramedia Digital is a digital reading service platform in response to technological advancements. This research aims to analyze user satisfaction with the Gramedia Digital application by conducting sentiment analysis on user reviews from the Google Play Store using the Support Vector Machine (SVM) method. User reviews reveal a discrepancy between expectations and actual experiences, including technical issues such as automatic logout, limited features, and application instability. The research used a dataset of 1,032 reviews between January 2020 and May 2025. It applied a text mining approach consisting of several stages: data collection, expert-based labeling, preprocessing (including case folding, cleaning, normalization, tokenization, stopword removal, and stemming), TF-IDF weighting, and classification using SVM. For model evaluation, the researcher employed a confusion matrix with accuracy, precision, recall, and F1-score metrics across three data-splitting scenarios (70:30, 80:20, and 90:10). The results identify the 70:30 scenario as the best-performing configuration, achieving an accuracy of 81%, indicating strong alignment between predicted and actual labels. The model also attains a precision of 82%, demonstrating high correctness in positive predictions, a recall of 78%, reflecting its ability to identify positive instances, and an F1-score of 79%, indicating a balanced performance between precision and recall. The research result shows that the SVM method performs sentiment analysis effectively. Based on the sentiment analysis results, user reading satisfaction remains relatively low, as indicated by a higher proportion of negative reviews compared to positive ones, along with frequent complaints regarding feature limitations, application stability, and access issues after payment. These findings indicate that sentiment analysis serves as a valuable tool to understand user perceptions and reading satisfaction, while also providing recommendations for Gramedia Digital to improve the quality of its application services.

## مستلخص البحث

ساسيكيرانا، بريا جي ديفا، 2026. تحليل رضا مستخدمي تطبيق غراميديا الرقمي باستخدام طريقة آلة المتجه الداعمة (SVM). البحث الجامعي. قسم المكتبات وعلوم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فخرس حسن ريزا محفوظ، الماجستير؛ المشرف الثاني: أحمد نظام رفقي، الماجستير.

**الكلمات الرئيسية:** تحليل مشاعر، رضا مستخدمين لقراءة، غراميديا رقمي، آلة متجه داعمة.

غراميديا الرقمي هو منصة خدمة قراءة رقمية ظهرت استجابةً للتقدم التكنولوجي. هدف هذا البحث إلى تحليل رضا المستخدمين عن تطبيق غراميديا الرقمي من خلال تحليل المشاعر للتعليقات على متجر جوجل بلاي باستخدام طريقة آلة المتجه الداعمة (SVM). تُظهر تعليقات المستخدمين وجود عدم توافق بين التوقعات والتجربة، مثل الصعوبات التقنية المتمثلة في تسجيل الخروج التلقائي، ومحدودية الميزات، وكذلك اضطرابات استقرار التطبيق. تم استخدام 1032 تعليقًا خلال الفترة من يناير 2020 حتى مايو 2025. استخدم هذا البحث منهج التنقيب النصي من خلال مراحل جمع البيانات، ووضع العلامات من قبل خبراء اللغة، والمعالجة المسبقة (تحويل الحالات، التنظيف، التطبيع، التجزئة، إزالة كلمات الوقف، والجذر)، وتوزين TF-IDF، وكذلك التصنيف باستخدام SVM. تم تقييم النموذج باستخدام مصفوفة الالتباس مع مقاييس الدقة والثبات والاسترجاع ونسبة ف1 من خلال ثلاثة سيناريوهات لتقسيم البيانات (70:30، 80:20، و90:10). أفضل سيناريو هو 70:30 مع دقة تبلغ 81%، مما يشير إلى التوافق بين نتائج التنبؤ والتسميات الفعلية، والثبات يبلغ 82% مما يشير إلى صحة التنبؤات الإيجابية، والاسترجاع يبلغ 78% مما يشير إلى قدرة النموذج على التعرف على البيانات الإيجابية، ونسبة ف1 تبلغ 79% مما يعكس التوازن بين الثبات والاسترجاع. أظهرت نتائج هذا البحث أن طريقة SVM قادرة على إجراء تحليل المشاعر بشكل جيد إلى حد ما. بناءً على تحليل المشاعر، أشار رضا المستخدمين للقراءة أن أغلب المستخدمين يشعرون بعدم الرضا، واتضح ذلك من خلال وجود مراجعات سلبية أكثر من المراجعات الإيجابية بالإضافة إلى الشكاوى المتعلقة بمحدودية الميزات، واستقرار التطبيق، ومشاكل الوصول بعد الدفع. تشير هذه النتائج إلى أن تحليل المشاعر يمكن استخدامه كأداة لفهم تصورات ورضا المستخدمين للقراءة، وكذلك لتقديم توصيات لغراميديا الرقمي في تحسين جودة خدمات التطبيق.

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN TULISAN.....	iii
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
مستخلص البحث.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan.....	6
BAB II KAJIAN PUSTAKA.....	8
2.1 Tinjauan Pustaka.....	8
2.2 Landasan Teori.....	11
2.2.1 Gramedia Digital.....	12
2.2.2 Analisis Sentimen.....	12
2.2.3 <i>Preprocessing</i> .....	13
2.2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	15
2.2.5 <i>Support Vector Machine (SVM)</i> .....	15
2.2.6 <i>Confusion Matrix</i> .....	16
BAB III METODE PENELITIAN.....	17
3.1 Jenis Penelitian.....	17
3.2 Alur Penelitian.....	17

3.3 Desain Sistem.....	18
3.3.1 <i>Preprocessing</i> .....	19
3.3.2 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	22
3.3.3 <i>Support Vector Machine (SVM)</i> .....	24
3.3.4 Confusion Matrix.....	25
3.4 Akuisisi Data.....	26
BAB IV HASIL DAN PEMBAHASAN.....	27
4.1 Hasil Penelitian.....	27
4.1.1 Pelabelan Data.....	27
4.1.2 Hasil <i>Preprocessing</i> .....	28
4.1.3 Hasil TF-IDF.....	31
4.1.4 Hasil klasifikasi SVM.....	32
4.1.5 Hasil <i>Wordcloud</i> Ulasan Pengguna Aplikasi Gramedia Digital.....	36
4.2 Pembahasan.....	37
4.2.1 Kemampuan Metode <i>Support Vector Machine (SVM)</i> dalam Melakukan Analisis Sentimen terhadap Ulasan Aplikasi Gramedia Digital.....	37
4.2.2 Analisis Kepuasan Membaca Pengguna dari Hasil Analisis Sentimen	41
BAB V PENUTUP.....	45
5.1 Kesimpulan.....	45
5.2 Saran.....	45
DAFTAR PUSTAKA.....	46
LAMPIRAN.....	50

## DAFTAR TABEL

Tabel 2.1 Confusion Matrix .....	16
Tabel 3.1 Contoh Pelabelan Data .....	19
Tabel 3.2 Contoh Case Folding .....	19
Tabel 3.3 Contoh Cleaning .....	20
Tabel 3.4 Contoh Normalisasi .....	20
Tabel 3.5 Contoh Tokenizing .....	21
Tabel 3.6 Contoh Stopword Removal .....	21
Tabel 3.7 Contoh Stemming .....	22
Tabel 4.1 Data Mentah Ulasan .....	28
Tabel 4.2 Tahap Case Folding .....	29
Tabel 4.3 Hasil Cleaning .....	29
Tabel 4.4 Hasil Normalisasi .....	30
Tabel 4.5 Hasil Tokenizing .....	30
Tabel 4.6 Stopword Removal .....	30
Tabel 4.7 Hasil Stemming .....	31
Tabel 4.8 Hasil TF-IDF .....	32
Tabel 4.9 Hasil Evaluasi SVM .....	33

## DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori .....	11
Gambar 3.1 Alur Penelitian.....	17
Gambar 3.2 Desain Sistem.....	18
Gambar 4.1 Distribusi Sentimen.....	28
Gambar 4.2 Confusion Matrix 70:30 .....	34
Gambar 4.3 Confusion Matrix 80:20 .....	35
Gambar 4.4 Confusion Matrix 90:10 .....	35
Gambar 4.5 Wordcloud Ulasan Positif.....	36
Gambar 4.6 Wordcloud Ulasan Negatif.....	36

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Saat ini, aktivitas membaca tidak lagi terbatas pada buku fisik yang tersedia di perpustakaan atau tempat baca lainnya. Peralihan ke media digital memungkinkan bahan bacaan tersedia lebih cepat, luas, dan tidak bergantung pada ruang dan waktu. Salah satu langkah strategis untuk mendorong peningkatan literasi di Indonesia, khususnya di kalangan generasi muda melalui pengembangan perpustakaan digital (Andita, 2022). Kehadiran perpustakaan digital memberi alternatif baru bagi masyarakat yang sebelumnya lebih terbiasa membaca buku fisik, sehingga muncul referensi baru dalam bentuk digital. Transformasi ini menunjukkan bahwa pengembangan layanan baca digital berperan penting dalam meningkatkan frekuensi membaca di Indonesia, terutama di kalangan generasi muda yang dekat dengan teknologi (Judijanto *et al.*, 2025).

Salah satu bentuk transformasi layanan baca di Indonesia adalah hadirnya aplikasi Gramedia Digital, yang dikembangkan oleh PT Gramedia Asri Media sebagai bagian dari ekosistem literasi nasional. Aplikasi Gramedia Digital merupakan *platform* yang menyediakan layanan baca dan langganan *e-book* secara digital yang hadir sebagai tanggapan atas perkembangan teknologi (Ratnaningsih *et al.*, 2023). Melalui aplikasi ini, masyarakat dapat mengakses koleksi yang sangat beragam, mulai dari *e-book*, majalah, hingga koran digital. Gramedia Digital menyediakan koleksi lebih dari 100.000 judul, sehingga memberi peluang bagi pengguna untuk memperluas pilihan bacaan sesuai kebutuhan. Tujuan utama adanya Gramedia Digital antara lain memperluas akses bahan bacaan yang legal dan memudahkan masyarakat untuk membaca dimana saja.

Koleksi di Gramedia Digital tidak semua dapat diakses secara bebas. Sebagian koleksi tersedia melalui pembelian satuan atau paket langganan premium, namun terdapat pula koleksi gratis. Kriteria *e-book* gratis di Gramedia Digital pada judul yang diberi penanda harga Rp0 di aplikasi dan bisa ditemukan melalui fitur filter harga Gratis. Koleksi gratis ini merupakan program promosi dari penerbit dan penulis atau bagian dari program Gramedia Digital seperti *free trial* untuk

memperkenalkan koleksi tertentu kepada publik (Melania, 2020). Penyediaan bacaan gratis untuk memperluas kesempatan membaca bagi masyarakat yang memiliki keterbatasan biaya dan mendukung upaya peningkatan minat baca melalui akses bacaan yang legal.

Pengalaman pengguna aplikasi Gramedia Digital menjadi hal yang menarik dalam pengembangan layanan. Pengguna dapat memberikan ulasan dan penilaian (*rating*) terhadap aplikasi yang telah mereka gunakan (Warow & Pandia, 2024). Ulasan menjadi sumber informasi untuk mengukur persepsi dan tingkat kepuasan yang dirasakan pengguna aplikasi tersebut (Nurian et al., 2024). Peningkatan jumlah pengguna aplikasi Gramedia Digital secara langsung berdampak pada semakin banyaknya ulasan yang tersedia. Berdasarkan data saat penelitian ini dirancang, aplikasi ini telah diunduh lebih dari satu juta kali, dengan rata-rata penilaian sebesar 4,2 dan sekitar 15.000 ulasan di Google Play Store.

Ulasan ini menjadi bentuk komunikasi modern dan gambaran kepuasan pengguna terhadap layanan baca digital. Sebagian pengguna memuji kemudahan akses dan koleksi yang lengkap, sementara sebagian lainnya mengeluhkan masalah teknis seperti ketiadaan fitur pencarian pada filter buku gratis, pengaturan ukuran *font*, kesulitan akses meskipun sudah berlangganan, aplikasi yang sering *logout* otomatis, serta belum tersedianya pada versi dekstop. Pengalaman membaca secara digital menjadi lebih optimal dengan adanya kemudahan akses serta kelengkapan fitur, seperti *bookmark*, pencarian instan, dan *highlight*. Minimnya fitur yang tersedia serta adanya kendala pada sistem berpotensi mengurangi kenyamanan pengguna dan berdampak negatif pada tingkat kepuasan pembaca (Aqilah et al., 2025). Memahami persepsi pengguna menjadi penting agar manfaat *platform* baca digital ini benar-benar mendukung minat baca masyarakat.

Ulasan pengguna ini tidak hanya cerminan kepuasan terhadap sebuah layanan, tetapi juga berkaitan dengan menyampaikan pendapat atau persepsi pengguna. Dalam perspektif Islam, hal ini telah ditegaskan dalam Al-Qur'an, sebagaimana firman Allah dalam surat *Al-Isrā'* [17]:53:

وَقُلْ لِعِبَادِي بُرْهَانَ هِيَ أَحْسَنُ إِنَّ الشَّيْطَانَ يَنْزِعُ بَيْنَهُمْ إِنَّ الشَّيْطَانَ كَانَ لِلْإِنْسَانِ عَدُوًّا مُّبِينًا

*Artinya “Katakan kepada hamba-hamba-Ku supaya mereka mengucapkan perkataan yang lebih baik (dan benar). Sesungguhnya setan itu selalu menimbulkan perselisihan di antara mereka. Sesungguhnya setan adalah musuh yang nyata bagi manusia.” Al-Isrā' [17]:53.*

Dalam Tafsir Kemenag dijelaskan bahwa Allah memerintahkan Rasulullah untuk mengarahkan manusia agar senantiasa menggunakan perkataan yang baik ketika berbicara, termasuk saat menghadapi perbedaan pendapat atau perdebatan dengan orang lain. Ucapan yang disampaikan tidak boleh berupa kata-kata kasar, hinaan, atau caci maki, karena hal tersebut dapat menimbulkan permusuhan dan kebencian. Sebaliknya, setiap perkataan hendaknya berisi kebenaran, disampaikan dengan cara yang baik, serta mampu memberi pelajaran dan membawa kebaikan bagi pihak yang mendengarnya (Kemenag, 2025).

Adapun hadits Al-Bukhari, no. 6018; Muslim, no.47, yang sehubungan dengan surat *Al-Isrā' [17]:53*.

مَنْ كَانَ يُؤْمِنُ بِاللَّهِ وَالْيَوْمِ الْآخِرِ فَلْيَقُلْ خَيْرًا أَوْ لِيَصْمُتْ

*Artinya “Barang siapa yang beriman kepada Allah dan Hari Akhir maka hendaklah ia berkata baik atau hendaklah ia diam.” (Muttafaq ‘alaih: Al-Bukhari, no. 6018; Muslim, no.47).*

Ibnu Hajar menjelaskan bahwa hadis ini merupakan ungkapan yang singkat, tetapi memiliki makna yang sangat luas. Menurut beliau, setiap perkataan pada dasarnya dapat mengandung kebaikan, keburukan, atau berada di antara keduanya. Perkataan yang termasuk baik adalah ucapan yang bernilai wajib atau sunnah untuk disampaikan, serta segala bentuk ucapan yang mengarah kepada kebaikan tersebut. Sebaliknya, ucapan yang tidak mengandung nilai kebaikan, atau justru cenderung mengarah pada keburukan, termasuk ke dalam perkataan yang tidak patut diucapkan. Karena itu, seseorang seharusnya memilih untuk diam apabila perkataan yang akan disampaikan tidak membawa manfaat atau justru berpotensi menimbulkan keburukan (Afifah, 2025).

Penjelasan yang terkandung pada tafsir kemenag surat *Al-Isrā' [17]:53* dan Ibnu Hajar hadits Al-Bukhari, no. 6018; Muslim, no. 47 sehubungan dengan penelitian ini. Ulasan pengguna pada aplikasi merupakan bentuk komunikasi modern yang berperan dalam membentuk citra serta kualitas layanan. Ulasan yang membangun dapat menjadi masukan bagi pengembang, sedangkan ulasan negatif berpotensi memengaruhi persepsi publik secara buruk. Oleh karena itu, diperlukan pendekatan yang mampu mengolah data teks ulasan agar persepsi pengguna dapat dipahami.

Pendekatan yang digunakan dalam memahami persepsi pengguna adalah analisis sentimen. Analisis sentimen merupakan proses otomatis untuk mengambil data dari teks, khususnya untuk mengidentifikasi persepsi atau opini terhadap suatu objek, termasuk mengklasifikasikan ulasan sebagai positif atau negatif. Persepsi pengguna ini dapat menjadi pertimbangan dalam menentukan keputusan terkait suatu produk atau aplikasi (Radiena & Nugroho, 2023). Metode *Support Vector Machine* (SVM) dipilih untuk mendukung analisis sentimen dari ulasan pengguna Gramedia Digital di Google Play Store. Pemilihan metode ini didasarkan pada kemampuannya yang terbukti mampu menangani klasifikasi teks dengan akurasi yang baik, terutama pada data yang memiliki dimensi tinggi (Bei & Saepudin, 2021).

Penelitian sebelumnya dilakukan oleh (Fitriyana et al., 2023) yang berjudul “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine”. Penelitian tersebut bertujuan untuk menganalisis sentimen ulasan aplikasi Jamsostek Mobile (JMO) menggunakan metode SVM. Sebanyak 279.877 ulasan dari Google Play Store periode 2018–2022 dikelompokkan ke dalam tiga jenis sentimen, yaitu positif, negatif, dan netral. Proses pengolahan data dilakukan melalui tahapan *pre-processing* (*case folding, tokenizing, stop words removal, stemming*) dan pelabelan *lexicon-based*. Hasilnya menunjukkan akurasi 96%, presisi 92%, *recall* 96% dan *f1-score* 94%.

Berdasarkan kesesuaian permasalahan dan penelitian terdahulu maka peneliti mengusulkan penelitian dengan judul “Analisis Kepuasan Pengguna Aplikasi Gramedia Digital Menggunakan Metode Support Vector Machine

(SVM)”, adapun tahapan-tahapan dalam penelitian ini yaitu pengumpulan data, pelabelan data, teks *preprocessing*, klasifikasi dan evaluasi.

## 1.2 Identifikasi Masalah

Merujuk pada latar belakang yang telah dijelaskan, penelitian ini dirancang untuk menjawab pertanyaan mengenai:

1. Bagaimana kemampuan metode *Support Vector Machine* (SVM) dalam melakukan analisis sentimen terhadap ulasan aplikasi Gramedia Digital berdasarkan akurasi, presisi, *recall*, dan *F1-score*?
2. Bagaimana analisis kepuasan membaca pengguna aplikasi Gramedia Digital berdasarkan hasil analisis sentimen?

## 1.3 Tujuan Penelitian

Adapun tujuan pada penelitian ini adalah:

1. Untuk mengetahui kemampuan metode *Support Vector Machine* (SVM) dalam melakukan analisis sentimen terhadap ulasan aplikasi Gramedia Digital berdasarkan akurasi, presisi, *recall*, dan *F1-score*.
2. Untuk menganalisis kepuasan membaca pengguna aplikasi Gramedia Digital berdasarkan hasil analisis sentimen.

## 1.4 Manfaat Penelitian

Manfaat penelitian ini dapat memberikan wawasan mengenai analisis sentimen menggunakan metode *Support Vector Machine* (SVM) terhadap aplikasi Gramedia Digital berdasarkan ulasan di Google Play Store. Serta dapat berguna bagi PT Gramedia untuk mengidentifikasi fitur aplikasi yang memerlukan perbaikan, pengembangan fitur baru, atau peningkatan kualitas layanan demi meningkatkan kepuasan pengguna secara keseluruhan.

## 1.5 Batasan Masalah

Batasan masalah penelitian ini adalah ulasan pengguna aplikasi Gramedia Digital pada Google Play Store. Data yang digunakan dibatasi mulai dari 1 Januari 2020 hingga 31 Mei 2025.

## **1.6 Sistematika Penulisan**

### **BAB I Pendahuluan**

Bab ini berisi pengantar mengenai latar belakang masalah yang mendasari pentingnya analisis sentimen terhadap aplikasi Gramedia Digital. Identifikasi masalah menjelaskan pertanyaan penelitian tentang bagaimana kemampuan metode *Support Vector Machine* (SVM) dalam analisis sentiment ulasan aplikasi Gramedia Digital. Tujuan penelitian mengetahui bagaimana kemampuan metode SVM dalam analisis sentiment ulasan aplikasi Gramedia Digital. Manfaat penelitian ada dua yaitu manfaat teoritis dan praktis. Batasan masalah menjelaskan ruang lingkup, serta fokus pada klasifikasi sentimen menggunakan metode SVM. Bagian akhir menjelaskan sistematika penulisan skripsi ini secara keseluruhan.

### **BAB II Tinjauan Pustaka**

Bab ini berisi kajian pustaka dan landasan teori yang relevan. Kajian pustaka membahas hasil-hasil penelitian sebelumnya yang membahas analisis sentimen serta penggunaan algoritma klasifikasi seperti SVM, Naïve Bayes, dan KNN, untuk menemukan perbedaan penelitian yang menjadi dasar penelitian ini. Selanjutnya, landasan teori menjelaskan konsep-konsep penting seperti analisis sentimen, *preprocessing*, TF-IDF, *Support Vector Machine* (SVM) dan *confusion matrix*.

### **BAB III Metode Penelitian**

Bab ini menjelaskan pendekatan yang digunakan, yaitu *text mining* dengan metode *machine learning* yaitu *Support Vector Machine* (SVM). Alur dan desain penelitian seperti pengumpulan data, pelabelan data, *preprocessing*, klasifikasi dan evaluasi model.

### **BAB IV Hasil dan Pembahasan**

Bab ini menyajikan hasil analisis sentimen menggunakan SVM dalam bentuk tabel, diagram, dan uraian naratif. Pembahasan mengaitkan temuan dengan penelitian terdahulu serta mengevaluasi performa model menggunakan akurasi, presisi, recall, dan F1-score. Selain itu, dianalisis pula kepuasan membaca pengguna aplikasi Gramedia Digital.

## **BAB V Penutup**

Bab ini akan memuat kesimpulan dari penelitian yang sudah dilakukan menggunakan metode SVM dan saran – saran untuk peneliti selanjutnya yang didapat pada keterbatasan penelitian ini.

## BAB II KAJIAN PUSTAKA

### 2.1 Tinjauan Pustaka

Penelitian yang berjudul “Analisis Sentimen Komentar Video Mobil Listrik di Platform YouTube dengan Metode Naïve Bayes” berfokus pada terpecahnya opini masyarakat terkait mobil listrik di Indonesia, terutama menyangkut harga yang tinggi, keterbatasan infrastruktur, serta minimnya pemahaman publik. Permasalahan tersebut mendorong dilakukannya analisis sentimen untuk mengetahui bagaimana persepsi masyarakat terhadap mobil listrik dengan menggunakan algoritma Naïve Bayes. Dengan menggunakan 1.517 komentar yang diperoleh dari platform Kaggle dengan menggunakan 2 kategori sentimen positif dan negatif. Proses analisis dilakukan melalui tahapan *data selection*, *preprocessing (cleaning, case folding, tokenizing, stopword removal)*, pembobotan menggunakan TF-IDF, dan klasifikasi menggunakan Naïve Bayes. Untuk menangani ketidakseimbangan kelas data, peneliti menerapkan teknik SMOTE *Upsampling*, yang berhasil meningkatkan akurasi model dari 50,70% menjadi 70,69%, dengan presisi sebesar 43,64% dan *recall* sebesar 39,48%. Hasil penelitian menunjukkan bahwa Naïve Bayes cukup efektif digunakan untuk klasifikasi sentimen berbasis teks, meskipun performanya masih dapat ditingkatkan melalui eksplorasi algoritma lain yang lebih kompleks (Karimah et al., 2024).

Penelitian berjudul “Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN” dilatarbelakangi oleh semakin meningkatnya minat masyarakat terhadap investasi online. Namun, banyaknya aplikasi yang tersedia seringkali membuat pengguna bingung dalam menentukan pilihan yang aman dan terpercaya. Maraknya kasus penipuan investasi online juga menambah keraguan masyarakat. Ulasan pengguna di Google Play Store sebenarnya dapat menjadi sumber informasi yang penting, tetapi jumlahnya yang besar dan tidak terstruktur menyulitkan untuk dianalisis secara manual. Penelitian tersebut menggunakan algoritma *K-Nearest Neighbor (KNN)* untuk mengklasifikasikan ulasan ke dalam dua kategori sentimen, yaitu positif dan negatif. Data yang digunakan berjumlah 998 pada aplikasi Bibit dan 1063 data pada

aplikasi Bareksa. Ulasan diperoleh melalui teknik *web scraping*, kemudian dianalisis melalui tahapan *preprocessing* (*remove duplicates, case folding, stopword* dan *stemming*). Selanjutnya, data diubah menjadi representasi numerik menggunakan TF-IDF, lalu diklasifikasikan menggunakan KNN dengan nilai parameter  $K=10$ . Evaluasi dilakukan pada lima skenario rasio pembagian data latih dan uji, dengan hasil terbaik pada rasio 60:40. Model menunjukkan performa akurasi sebesar 85.14%, presisi 91.91%, *recall* 76.44% untuk aplikasi Bibit dan akurasi sebesar 81.70%, presisi 87.15%, *recall* 75.73% untuk aplikasi Bareksa (Putra & Juanita, 2021).

Penelitian lain yang berjudul “*Sentiment Analysis of Arabic Social Media Texts: A Machine Learning Approach to Deciphering Customer Perceptions*” yang membahas tentang opini konsumen di media sosial, khususnya Twitter, yang berisi persepsi terhadap produk namun sulit dianalisis secara manual karena jumlahnya sangat besar dan tidak terstruktur. Permasalahan ini penting karena informasi berharga dari konsumen seringkali tidak dimanfaatkan secara optimal oleh perusahaan. Data yang digunakan berjumlah 10.646 tweet, yang telah dianotasi secara manual ke dalam tiga kategori sentimen: positif, negatif, dan netral. Tahapan dalam penelitian meliputi *data preprocessing* (pembersihan, penghapusan *stopword*, dan normalisasi), serta ekstraksi fitur menggunakan dua teknik, yaitu TF-IDF dan *Minimum Redundancy Maximum Relevance* (MRMR). Empat algoritma *machine learning* diterapkan dalam proses klasifikasi *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest* (RF), dan *K-Nearest Neighbor* (KNN). Selanjutnya, dikembangkan pula dua model *ensemble* menggunakan metode *hard voting* dan *soft voting* untuk meningkatkan performa klasifikasi. Hasil pengujian menunjukkan bahwa model *ensemble hard voting* memiliki kinerja terbaik dengan akurasi 95,95%, disusul *soft voting* dengan 94,51%. Di antara algoritma individu, SVM menghasilkan akurasi tertinggi sebesar 95%, menjadikannya algoritma terbaik untuk klasifikasi sentimen dalam studi ini. Penelitian ini menunjukkan bahwa kombinasi *ensemble learning* dan teknik ekstraksi fitur yang tepat dapat secara signifikan meningkatkan akurasi analisis sentimen pada teks berbahasa Arab di media sosial (Alsemaree et al., 2024).

Penelitian selanjutnya yang berjudul “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)” Fokus penelitian tersebut adalah fenomena ramainya diskusi mengenai wacana pemindahan ibu kota Indonesia di Twitter yang memicu beragam tanggapan masyarakat, baik yang mendukung maupun menolak. Perbedaan pandangan tersebut menghasilkan banyak opini publik yang menyebar di media sosial, namun sulit ditelaah secara manual karena jumlahnya sangat besar. Untuk itu, penelitian ini memanfaatkan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi sentimen guna mengelompokkan opini masyarakat ke dalam dua kategori utama, yakni positif dan negatif. Data komentar dalam Bahasa Indonesia dikumpulkan dari *platform* Twitter, lalu dilakukan pelabelan secara manual oleh dua orang annotator. Tahapan praproses mencakup *handle duplicate, labelling, case folding, cleanning, tokenize, filtering, stopword removal, stemming* dan *wighting*. Proses pembobotan kata dilakukan dengan TF-IDF. Evaluasi model dilakukan menggunakan *confusion matrix* dengan pembagian data latih dan uji menggunakan teknik *10-fold cross-validation*. Hasil klasifikasi menunjukkan bahwa SVM mampu mencapai akurasi 96.82%, presisi 95.82%, dan *recall* 94.04% (Arsi & Waluyo, 2021).

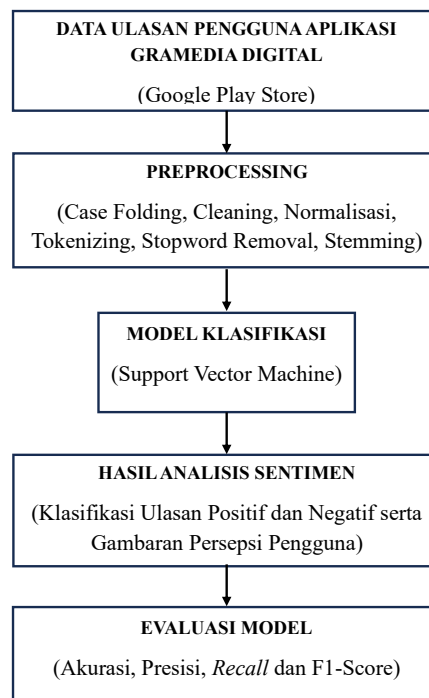
Penelitian terakhir yang berjudul “Analisis Sentimen pada Ulasan Aplikasi Threads di Google Play Store Menggunakan Algoritma *Support Vector Machine*” fokus penelitian pada tingginya popularitas aplikasi Threads yang dirilis oleh Instagram pada Juli 2023, dengan jumlah unduhan lebih dari 50 juta dan rating rata-rata 4,5 di Google Play Store. Meskipun begitu, rating tidak selalu mencerminkan isi ulasan, karena terdapat perbedaan antara skor bintang dengan opini yang dituliskan pengguna. Ulasan yang jumlahnya besar dan tidak terstruktur ini penting untuk dianalisis agar pengembang memahami sentimen publik terhadap aplikasi. Dataset yang digunakan terdiri dari 1.429 ulasan, dengan 1.103 ulasan positif dan 326 ulasan negatif. Data diperoleh melalui teknik *crawling* menggunakan library *google-play-scraper*. Tahapan dalam penelitian tersebut meliputi *preprocessing (cleaning, case folding, tokenization, normalization, stopword removal, dan stemming)* serta pelabelan data menggunakan pendekatan *lexicon-based* dengan

bantuan *VADER Sentiment Analyzer* setelah diterjemahkan ke bahasa Inggris. Data kemudian dibagi menggunakan skema 80% data latih dan 20% data uji. Proses klasifikasi digunakan algoritma SVM dengan *kernel linear*. Hasil evaluasi menunjukkan bahwa metode SVM berhasil mencapai tingkat akurasi sebesar 88%, dengan *precision* untuk sentimen positif sebesar 90% dan *recall* sebesar 96%. (Nufairi et al., 2024).

Dari lima penelitian terdahulu yang telah dijabarkan, terdapat beberapa kesamaan dengan penelitian ini, terutama pada judul yang sama-sama membahas analisis sentimen serta menggunakan dua kategori sentimen yaitu positif dan negatif. Terdapat juga perbedaannya pada metode yang digunakan, beberapa penelitian terdahulu menggunakan data dari media sosial YouTube maupun Twitter, pada penelitian ini menggunakan data ulasan dari Google Play Store. Selain itu, penelitian ini berfokus pada layanan literasi digital.

## 2.2 Landasan Teori

Berikut disajikan diagram kerangka teori yang menggambarkan hubungan antar konsep utama dalam penelitian ini.



**Gambar 2.1 Kerangka Teori**  
(Sumber: Hasil Olah Data Peneliti)

### 2.2.1 Gramedia Digital

Gramedia Digital merupakan sebuah aplikasi yang memberikan kemudahan dan kecepatan akses terhadap berbagai bahan bacaan *digital*, seperti *e-book*, majalah, serta publikasi digital lainnya (Rahma et al., 2024). Aplikasi ini tidak hanya memperluas akses terhadap sumber bacaan, tetapi juga berkontribusi dalam meningkatkan minat baca masyarakat dengan menyediakan berbagai konten yang dapat diakses kapan saja dan dimana saja melalui perangkat digital. Dalam penggunaan Gramedia Digital, pengguna perlu membuat akun terlebih dahulu agar dapat mengakses layanan yang tersedia di dalam aplikasi. Akun tersebut digunakan untuk memudahkan pengguna dalam memilih *e-book*, melakukan pembelian atau berlangganan koleksi digital, serta membaca bahan bacaan yang telah diperoleh. Koleksi digital yang dibeli melalui Gramedia Digital juga hanya dapat diakses melalui akun pribadi pengguna di aplikasi tersebut, sehingga tidak dapat dibaca pada platform lain di luar aplikasi Gramedia Digital yang tersedia di Google Play Store untuk perangkat Android.

### 2.2.2 Analisis Sentimen

Analisis sentimen merupakan cabang dari *Natural Language Processing* (NLP) yang berfokus pada pemahaman interaksi antara manusia dan komputer melalui teks. Bidang ini bertujuan untuk memahami, mengelola, menganalisis, serta mengekstrak informasi dari data teks, seperti opini, evaluasi, sikap, maupun emosi terhadap suatu topik tertentu. Analisis sentimen banyak dimanfaatkan untuk menyaring emosi dari data di media sosial yang terus meningkat, dengan cara mengidentifikasi kata atau frasa yang mencerminkan sikap positif maupun negatif. Emosi dasar manusia yang sering menjadi acuan dalam analisis ini meliputi enam kategori, yaitu takut (*fear*), marah (*anger*), jijik (*disgust*), sedih (*sadness*), terkejut (*surprise*), dan senang (*joy*) (Prakash & Aloysius, 2021).

Metode ini merupakan teknik untuk secara otomatis mengidentifikasi, mengambil, dan memproses informasi berbentuk teks guna menemukan makna sentimental dalam penyampaian suatu pemikiran (Gifari et al., 2022). Pada penelitian ini, analisis sentimen dilakukan ke dalam dua kategori, yaitu positif dan

negatif. Pemilihan dua kategori ini bertujuan untuk menjaga keseimbangan distribusi data, sehingga proses analisis dan evaluasi dapat berlangsung serta menghasilkan hasil yang akurat (Kusuma & Cahyono, 2023). Teknik ini banyak digunakan untuk memahami persepsi publik terhadap produk, layanan, atau isu tertentu, sehingga dapat mendukung pengambilan keputusan.

Analisis sentimen menjadi semakin penting karena jumlah data yang dihasilkan melalui media sosial, *e-commerce*, maupun forum daring lainnya terus meningkat dengan sangat cepat. Ulasan pengguna, komentar, dan opini publik dalam jumlah besar tidak mungkin dianalisis secara manual. Dengan adanya analisis sentimen, maka perusahaan, organisasi, maupun instansi pemerintah dapat memperoleh wawasan berharga mengenai kepuasan pengguna, tren masyarakat, serta persepsi publik terhadap suatu isu dalam waktu yang lebih singkat dan efisien.

### **2.2.3 Preprocessing**

*Preprocessing* adalah teknik yang digunakan untuk mempersiapkan data dari bentuk teks mentah agar siap dianalisis sesuai dengan metode yang diterapkan (Astari et al., 2020). *Preprocessing* teks adalah proses persiapan data sebelum dianalisis, yang mencakup serangkaian tahapan seperti mengubah ke huruf kecil, penghapusan karakter non alfabet, normalisasi kata menjadi bentuk baku, pemecahan teks menjadi per kata, pembersihan kata-kata yang tidak penting, serta pengubahan kata ke bentuk dasar. Langkah ini penting untuk memastikan data yang dianalisis lebih seragam dan akurat (Hickman et al., 2022). Tahapan *preprocessing* yaitu:

#### 1. *Case Folding*

Tahap *case folding* merupakan langkah awal dalam *preprocessing* yang bertujuan menyeragamkan semua huruf menjadi huruf kecil (*lowercase*), sehingga tidak ada perbedaan antara huruf besar dan kecil pada setiap kata (Wati & Ernawati, 2021). Misalnya, “Kenapa semua ebook yg udah didownload gak ke save di perpustakaan ya?” Menjadi “kenapa semua ebook yg udah didownload gak ke save di perpustakaan ya?”. Maka data menjadi lebih konsisten dan mudah diproses oleh algoritma.

## 2. *Cleaning*

Pada tahap ini, teks ulasan dibersihkan dari berbagai elemen yang tidak relevan, seperti angka, tanda baca, simbol khusus, tautan (*URL*), *hashtag*, hingga emoji. Elemen-elemen tersebut biasanya muncul dalam teks ulasan pengguna, tetapi tidak memberikan informasi penting untuk proses analisis sentimen. Pembersihan ini penting agar data yang digunakan lebih fokus pada kata-kata bermakna yang mencerminkan opini pengguna (Wati & Ernawati, 2021).

## 3. Normalisasi

Normalisasi dilakukan untuk menyamakan kata-kata yang ditulis secara tidak baku atau dalam bentuk singkatan menjadi kata baku sesuai kaidah bahasa Indonesia (Vonega et al., 2022). Misalnya, kata “gk” diubah menjadi “tidak”, atau “bgs” diubah menjadi “bagus”. Hal ini dilakukan karena ulasan pengguna sering menggunakan bahasa informal, singkatan, atau bahkan salah ketik. Tanpa normalisasi, kata-kata tersebut bisa dianggap berbeda oleh sistem meskipun sebenarnya memiliki makna yang sama.

## 4. *Tokenizing*

Tahap *tokenizing* adalah proses memecah kalimat ulasan menjadi potongan kata atau token (Wati & Ernawati, 2021). Misalnya, kalimat “aplikasi ini sangat bagus” akan dipecah menjadi [“aplikasi”, “ini”, “sangat”, “bagus”]. Pemecahan ini bertujuan agar setiap kata dapat diperlakukan sebagai unit analisis yang berdiri sendiri, sehingga memudahkan proses pembobotan maupun klasifikasi pada tahap selanjutnya.

## 5. *Stopword Removal*

*Stopword removal* dilakukan dengan menghapus kata-kata umum yang sering muncul dalam kalimat tetapi tidak memiliki makna penting untuk analisis sentimen (Wati & Ernawati, 2021). Contohnya adalah kata “yang”, “ini”, “dan”, atau “di”. Kata-kata tersebut tidak membawa bobot makna terhadap opini pengguna, sehingga penghapusannya dapat membantu model fokus pada kata-kata bermakna yang benar-benar merepresentasikan sentimen.

## 6. *Stemming*

Tahap stemming berfungsi untuk mengembalikan kata ke bentuk dasarnya (*root word*) (Wati & Ernawati, 2021). Misalnya, kata “mengecewakan”, “dikecewakan” akan diubah menjadi kata dasar yaitu “kecewa”. Langkah ini penting agar variasi bentuk kata yang berbeda tetapi memiliki arti yang sama dapat digabungkan menjadi satu representasi. Maka model analisis akan lebih efisien dalam mengenali pola sentimen yang terkandung dalam data.

### **2.2.4 Term Frequency-Inverse Document Frequency (TF-IDF)**

*Term Frequency-Inverse Document Frequency (TF-IDF)* merupakan metode dalam pengolahan teks yang digunakan untuk menilai tingkat pentingnya suatu kata di dalam sebuah dokumen dengan mempertimbangkan keseluruhan kumpulan dokumen. Metode ini diperkenalkan oleh Karen Spärck Jones pada tahun 1972 sebagai salah satu teknik dalam *information retrieval*. TF-IDF menunjukkan performa terbaik saat diterapkan pada klasifikasi dengan jumlah kelas dengan tingkat kemiringan data yang rendah atau cenderung seimbang (Mahfud et al., 2020). Melalui metode ini, sebuah kata diberi bobot berdasarkan seberapa sering kata tersebut muncul dalam satu dokumen dan seberapa jarang kemunculannya pada dokumen-dokumen lain dalam koleksi yang sama. Metode ini juga berfungsi untuk mengubah representasi kata menjadi bentuk numerik sehingga dapat diproses lebih lanjut dalam analisis teks maupun pembelajaran mesin (Setiawan & Adnyana, 2023).

### **2.2.5 Support Vector Machine (SVM)**

*Support Vector Machine (SVM)* dikembangkan pada akhir tahun 1960-an di AT&T Bell Laboratories oleh Vladimir Vapnik dan Alexey Chervonenkis. Metode ini dikenal sebagai salah satu metode klasifikasi yang efektif dan banyak digunakan dalam berbagai bidang seperti klasifikasi, pengenalan tulisan tangan, bioinformatika, dan analisis sentimen (Nurhayati et al., 2024). SVM bekerja dengan cara membagi data ke dalam dua kelas berbeda menggunakan *hyperplane* optimal yang memaksimalkan margin antar kelas. Dalam menjalankan algoritma ini, data

terlebih dahulu harus diberi label positif atau negatif agar model dapat mempelajari pola klasifikasinya (Suhardjono et al., 2019). Salah satu kekuatan utama SVM terletak pada kemampuannya menangani data yang tidak dapat dipisahkan secara *linear*.

### 2.2.6 Confusion Matrix

*Confusion matrix* menggambarkan perbandingan antara jumlah prediksi yang benar dan prediksi yang salah terhadap nilai sebenarnya dari data (Nufairi et al., 2024). Kemampuan yang dievaluasi ada empat, yaitu akurasi, presisi, *recall* dan *F1-score*. Dapat dilihat pada tabel 2.1 untuk skema evaluasinya.

**Tabel 2.1 Confusion Matrix**

		Prediksi	
		Negatif	Positif
Aktual	Negatif	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
	Positif	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

Berdasarkan Tabel 2.1, dapat dilihat bahwa hasil prediksi model Support Vector Machine (SVM) dibandingkan dengan data aktual menghasilkan empat kombinasi nilai, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. *TP (True Positive)* digunakan untuk menggambarkan jumlah data yang aktualnya positif dan diprediksi juga positif oleh sistem, artinya model berhasil mengidentifikasi data positif dengan benar. *FP (False Positive)* menggambarkan jumlah data yang aktualnya negatif tetapi diprediksi sebagai positif oleh sistem, sehingga menunjukkan kesalahan model dalam mengklasifikasikan data negatif. *FN (False Negative)* menunjukkan jumlah data yang aktualnya positif namun terprediksi menjadi negatif, yang berarti model gagal mengenali data positif. Sedangkan *TN (True Negative)* menggambarkan jumlah data yang aktualnya negatif dan diprediksi juga negatif oleh sistem, sehingga model dapat dikatakan benar dalam mengenali data negatif.

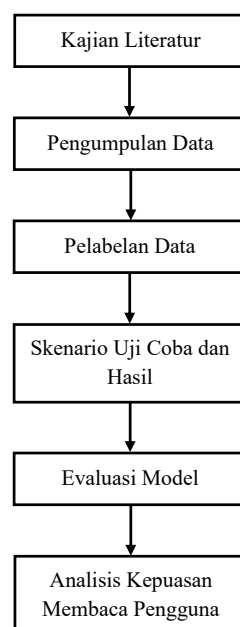
## BAB III METODE PENELITIAN

### 3.1 Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah pendekatan *text mining*. *Text Mining* merupakan proses penggalian informasi penting dari data berbentuk teks yang cenderung tidak terstruktur, dengan tujuan membentuk pola tertentu sehingga dapat menghasilkan informasi yang bermanfaat (Astuti et al., 2024). Data yang dianalisis berupa ulasan pengguna aplikasi Gramedia Digital yang diperoleh melalui Google Play Store. Penelitian ini menerapkan metode *Support Vector Machine* (SVM) sebagai algoritma klasifikasi untuk mengelompokkan ulasan ke dalam kategori sentimen positif dan negatif.

### 3.2 Alur Penelitian

Alur penelitian menggambarkan tahapan menyeluruh mengenai proses penelitian mulai dari pengumpulan data sampai evaluasi pada gambar dibawah ini.



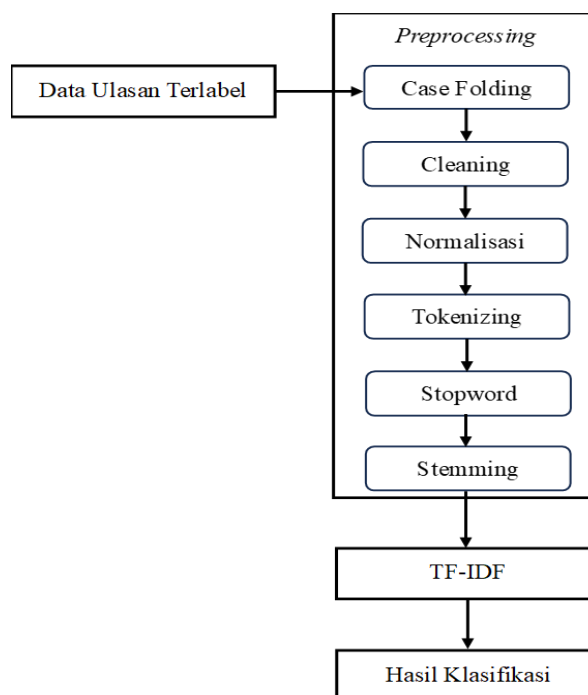
**Gambar 3.1 Alur Penelitian**  
(Sumber: Hasil Olah Data Peneliti)

Alur penelitian pada Gambar 3.1 meliputi enam tahap, Kajian literatur dilakukan untuk memahami metode analisis sentimen yang relevan dari penelitian

terdahulu. Data dikumpulkan dari ulasan Google Play Store dalam rentang waktu 1 Januari 2020 sampai 31 Mei 2025, dengan teknik *scraping* menggunakan *library* *google-play-scraper* di Python. Ulasan diberi label positif atau negatif oleh tiga ahli Bahasa Indonesia yang berprofesi sebagai guru Bahasa Indonesia, pelibatan beberapa ahli bahasa bertujuan untuk meminimalkan subjektivitas dalam pelabelan data. Menentukan data yang akan dijadikan data *training* dan data *test*, dengan tiga skenario pembagian untuk menguji performa model, yaitu 90:10, 80:20 dan 70:30. Evaluasi model dilakukan dengan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil klasifikasi sentimen dianalisis untuk melihat kepuasan membaca pengguna berdasarkan ulasan yang menggambarkan pengalaman membaca di aplikasi Gramedia Digital.

### 3.3 Desain Sistem

Desain sistem dilakukan untuk merancang sistem analisis sentimen. Google Colab dan bahasa pemrograman Python digunakan untuk membangun sistem ini. Rancangan sistem secara lengkap disajikan pada Gambar 3.2.



**Gambar 3.2 Desain Sistem**  
(Sumber: Hasil Olah Data Peneliti)

### 3.3.1 Preprocessing

Setelah proses pelabelan data, tahap berikutnya adalah *preprocessing*, yaitu langkah penting dalam pengolahan data teks yang bertujuan mempersiapkan data mentah agar lebih bersih, terstruktur, dan siap digunakan pada proses analisis selanjutnya (Arifin et al., 2021). Tahapan ini memastikan bahwa data yang digunakan memiliki kualitas yang optimal untuk mendukung analisis sentimen.

**Tabel 3.1 Contoh Pelabelan Data**

Ulasan	Label
Mohon diperbaiki. Akun sering logout jadi lumayan mengganggu sekali	Negatif
Aplikasinya sangat bagus 🤝 🤝 🤝	Positif
Perkembangannya luar biasa, readernya susah semakin baik. Meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke	Positif
Kurang update. buku yg aku cari semuanya gak ada	Negatif
sering logout otomatis, risih banget apalagi kalo lagi asik2nya baca	Negatif

#### 1. Case Folding

Tahap *case folding* berfungsi untuk mengubah huruf besar menjadi huruf kecil agar *dataset* menjadi lebih seragam dan mengurangi inkonsisten data (Jasmarizal et al., 2024).

**Tabel 3.2 Contoh Case Folding**

Data Mentah	Case Folding
Mohon diperbaiki. Akun sering logout jadi lumayan mengganggu sekali	mohon diperbaiki. akun sering logout jadi lumayan mengganggu sekali
Aplikasinya sangat bagus 🤝 🤝 🤝	aplikasinya sangat bagus 🤝 🤝 🤝
Perkembangannya luar biasa, readernya susah semakin baik. Meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke	perkembangannya luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
Kurang update. buku yg aku cari semuanya gak ada	kurang update. buku yg aku cari semuanya gak ada
sering logout otomatis, risih banget apalagi kalo lagi asik2nya baca	sering logout otomatis, risih banget apalagi kalo lagi asik2nya baca

## 2. *Cleaning*

Tahap *cleaning*, teks ulasan dibersihkan dari komponen yang tidak diperlukan, seperti angka, tanda baca, simbol khusus, link URL, tagar, serta emoji (Zidna et al., 2021). Proses pembersihan ini menggunakan *library* *Regex* (re) yang tersedia di *Phyton*.

**Tabel 3.3 Contoh *Cleaning***

<i>Case Folding</i>	<i>Cleaning</i>
mohon diperbaiki. akun sering logout jadi lumayan mengganggu sekali	mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali
aplikasinya sangat bagus 🤝 🤝 🤝	aplikasinya sangat bagus
perkembangannya luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke	perkembangannya luar biasa readernya susah semakin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
kurang update. buku yg aku cari semuanya gak ada	kurang update buku yg aku cari semuanya gak ada
sering logout otomatis, risih banget apalagi kalo lagi asik2nya baca	sering logout otomatis risih banget apalagi kalo lagi asik2nya baca

## 3. Normalisasi

Tahap normalisasi diawali dengan mengidentifikasi *dataset* secara manual oleh peneliti. Tujuan dari tahap ini adalah mengubah kata tidak baku maupun singkatan menjadi bentuk kata yang sesuai dengan kaidah ejaan Bahasa Indonesia yang benar (Vonega et al., 2022).

**Tabel 3.4 Contoh Normalisasi**

<i>Cleaning</i>	Normalisasi
mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali	mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali
aplikasinya sangat bagus	aplikasinya sangat bagus
perkembangannya luar biasa readernya susah semakin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke	perkembangannya luar biasa readernya susah makin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke

kurang update buku yg aku cari semuanya gak ada	kurang update buku yang aku cari semuanya tidak ada
sering logout otomatis risih banget apalagi kalo lagi asik2nya baca	sering logout otomatis risih banget apalagi kalo lagi asik-asiknya baca

#### 4. *Tokenizing*

Tahap *tokenizing* adalah proses pemisahan kalimat ulasan menjadi unit-unit kata atau token (Vonega et al., 2022). Proses ini memungkinkan setiap kata untuk diperlakukan secara terpisah sebagai objek analisis, sehingga mempermudah tahap pembobotan dan klasifikasi berikutnya.

**Tabel 3.5 Contoh *Tokenizing***

<b>Normalisasi</b>	<b><i>Tokenizing</i></b>
mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali	['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']
aplikasinya sangat bagus	['aplikasinya', 'sangat', 'bagus']
perkembangannya luar biasa readernya susah semakin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke	['perkembangannya', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'ada', 'beberapa', 'yang', 'perlu', 'di', 'tingkatkan', 'tapi', 'gramedia', 'digital', 'sekarang', 'sudah', 'sangat', 'oke']
kurang update buku yang aku cari semuanya tidak ada	['kurang', 'update', 'buku', 'yang', 'aku', 'cari', 'semuanya', 'tidak', 'ada']
sering logout otomatis risih banget apalagi kalo lagi asik-asiknya baca	['sering', 'logout', 'otomatis', 'risih', 'banget', 'apalagi', 'kalo', 'lagi', 'asik', 'asiknya', 'baca']

#### 5. *Stopword Removal*

Tahap *stopword removal* adalah proses menghilangkan kata-kata umum yang sering muncul dalam kalimat, namun kurang memiliki makna atau arti penting dalam analisis sentimen, seperti kata “di”, “yang” “dan” “atau” (Al-Husna et al., 2024). Proses ini dilakukan dengan menggunakan *library* sastrawi.

**Tabel 3.6 Contoh *Stopword Removal***

<b><i>Tokenizing</i></b>	<b><i>Stopword Removal</i></b>
['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']	['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']
['aplikasinya', 'sangat', 'bagus']	['aplikasinya', 'sangat', 'bagus']

<i>Tokenizing</i>	<i>Stopword Removal</i>
['perkembangannya', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'ada', 'beberapa', 'yang', 'perlu', 'di', 'tingkatkan', 'tapi', 'gamedia', 'digital', 'sekarang', 'sudah', 'sangat', 'oke']	['perkembangannya', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'beberapa', 'perlu', 'tingkatkan', 'gamedia', 'digital', 'sekarang', 'sangat', 'oke']
['kurang', 'update', 'buku', 'yang', 'aku', 'cari', 'semuanya', 'tidak', 'ada']	['kurang', 'update', 'buku', 'aku', 'cari', 'semuanya', 'ada']
['sering', 'logout', 'otomatis', 'risih', 'banget', 'apalagi', 'kalo', 'lagi', 'asik', 'asiknya', 'baca']	['sering', 'logout', 'otomatis', 'risih', 'banget', 'kalo', 'asik', 'asiknya', 'baca']

## 6. *Stemming*

Tahap *stemming* adalah proses mengubah kata-kata yang berimbuhan menjadi bentuk dasarnya atau akar kata (Sinaga & Nainggolan, 2023). Tujuannya untuk menyamakan variasi kata agar analisis teks menjadi lebih konsisten dan efektif. Tahap ini menggunakan *library* sastrawi karena dalam teks yang diproses berbahasa Indonesia.

**Tabel 3.7 Contoh *Stemming***

<i>Stopword Removal</i>	<i>Stemming</i>
['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']	['mohon', 'baik', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'ganggu', 'sekali']
['aplikasinya', 'sangat', 'bagus']	['aplikasi', 'sangat', 'bagus']
['perkembangannya', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'beberapa', 'perlu', 'tingkatkan', 'gamedia', 'digital', 'sekarang', 'sangat', 'oke']	['kembang', 'luar', 'biasa', 'reader', 'susah', 'makin', 'baik', 'meski', 'berapa', 'perlu', 'tingkat', 'gamedia', 'digital', 'sekarang', 'sangat', 'oke']
['kurang', 'update', 'buku', 'aku', 'cari', 'semuanya', 'ada']	['kurang', 'update', 'buku', 'aku', 'cari', 'semua', 'ada']
['sering', 'logout', 'otomatis', 'risih', 'banget', 'kalo', 'asik', 'asiknya', 'baca']	['sering', 'logout', 'otomatis', 'risih', 'banget', 'kalo', 'asik', 'baca']

### 3.3.2 *Term Frequency-Inverse Document Frequency (TF-IDF)*

Pembobotan dilakukan untuk memberi nilai numerik pada setiap kata dan sejauh mana suatu kata memberikan kontribusi terhadap suatu kelas sentimen, baik positif atau negatif. Metode ini mengubah teks menjadi bentuk numerik agar dapat diproses oleh algoritma klasifikasi. TF-IDF memberikan bobot lebih tinggi pada

kata yang sering muncul dalam satu ulasan tetapi jarang muncul pada ulasan lain. Kata-kata yang memiliki karakteristik kuat terhadap sentimen tertentu akan lebih menonjol dalam penyajian data. Proses pembobotan ini menjadi dasar bagi model SVM dalam membentuk batas pemisah antar kelas sentimen. Melalui tahapan pembobotan TF-IDF, data ulasan tidak lagi berupa teks mentah, melainkan telah berubah menjadi data terstruktur yang siap dianalisis. Tahapan ini dapat berkontribusi terhadap kemampuan model dalam mengenali pola sentimen secara lebih akurat (Ramadhan, 2025).

Teknik ini menggabungkan dua konsep, yaitu *Term Frequency* (TF) yang mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen, dan *Inverse Document Frequency* (IDF) yang menghitung seberapa jarang kata tersebut muncul di seluruh dokumen. TF-IDF dapat dihitung menggunakan rumus sebagai berikut.

1. Rumus perhitungan TF normalisasi

$$TF = \frac{f_{t,d}}{df} \quad (3.1)$$

Keterangan:

TF: Frekuensi kemunculan kata pada sebuah dokumen

$f_{t,d}$ : Kata yang muncul dalam dokumen

df: Jumlah kata dalam dokumen

Rumus ini menunjukkan seberapa sering sebuah kata muncul dalam sebuah dokumen. Semakin sering kata muncul, semakin tinggi nilai TF yang diperoleh. Misalnya, jika kata “aplikasi” muncul 5 kali dalam dokumen yang memiliki 100 kata, maka nilai TF untuk kata tersebut adalah 0,05.

2. Rumus perhitungan IDF

$$IDF = \log\left(\frac{N}{df_t}\right) \quad (3.2)$$

Keterangan:

IDF: *Inverse Document Frequency*

$N$ : Jumlah semua dokumen

$df_t$ : Jumlah dokumen yang mengandung kata tersebut

Nilai IDF digunakan untuk mengukur seberapa jarang suatu kata muncul di seluruh kumpulan dokumen. Kata yang sering muncul di hampir semua dokumen akan memiliki nilai IDF rendah karena dianggap kurang penting. Sebaliknya, kata yang hanya muncul di sedikit dokumen akan memiliki nilai IDF tinggi karena dianggap lebih spesifik dan bermakna.

### 3. Rumus TF-IDF

$$TF - IDF = TF \times IDF \quad (3.3)$$

Keterangan:

TF-IDF: *Term Frequency-Inverse Document Frequency*

TF: *Term Frequency*

IDF: *Inverse Document Frequency*

Hasil perkalian antara TF dan IDF menghasilkan bobot akhir sebuah kata. Kata dengan frekuensi tinggi dalam satu dokumen tetapi jarang muncul di dokumen lain akan memiliki nilai TF-IDF yang besar, sehingga dianggap penting dalam merepresentasikan dokumen tersebut. Sebaliknya, kata yang sering muncul di semua dokumen akan memiliki bobot kecil karena tidak memberikan informasi yang spesifik.

#### 3.3.3 *Support Vector Machine (SVM)*

Metode SVM menggunakan teknik *kernel trick*, yakni memetakan data dari ruang berdimensi rendah ke ruang berdimensi lebih tinggi, sehingga data tersebut menjadi lebih mudah dipisahkan secara *linear* di ruang baru tersebut (Dewi, 2020).

Secara matematis, SVM mencoba menemukan *hyperplane* terbaik yang dapat dinyatakan dengan rumus berikut:

$$f(x) = w \cdot x + b = 0 \quad (3.4)$$

Keterangan:

w: vektor bobot

x: vektor fitur

b: bias

Rumus tersebut menunjukkan persamaan sebuah garis atau bidang pemisah (*hyperplane*) yang digunakan oleh SVM untuk membagi data ke dalam dua kelas berbeda. Vektor bobot ( $w$ ) dan bias ( $b$ ) menentukan letak dan arah *hyperplane*. SVM memiliki beberapa jenis kernel, kernel yang digunakan pada penelitian ini adalah kernel linear yang dirumuskan sebagai:

$$K(x, y) = x \cdot y \quad (3.5)$$

Keterangan:

$K(x, y)$ : fungsi kernel linear yang menghitung kesamaan antara dua vektor data  $x$  dan  $y$ .

$x$ : vektor fitur dari data uji (misalnya hasil representasi TF-IDF dari sebuah ulasan).

$y$ : vektor fitur dari data latih atau support vector.

Kernel linear merupakan bentuk kernel paling sederhana yang menghitung nilai kesamaan (*similarity*) antara dua vektor dengan cara melakukan operasi dot product. Artinya, nilai kernel akan tinggi jika kedua vektor memiliki arah yang sama, dan rendah jika vektor saling berbeda. Berdasarkan nilai tersebut fungsi keputusan untuk menentukan kategori sentimen dirumuskan sebagai:

$$f(x) = \text{sign}(w \cdot x + b) \quad (3.6)$$

dengan aturan klasifikasi:

$f(x) > 0$  dikategorikan sebagai sentimen positif,

$f(x) < 0$  dikategorikan sebagai sentimen negatif.

SVM memilih *hyperplane* dengan margin maksimum, yang dirumuskan dengan meminimalkan, sehingga diperoleh pemisahan optimal antara dua kategori sentimen.

### 3.3.4 Confusion Matrix

Adapun rumus untuk menghitung akurasi, presisi, *recall* dan *F1-score*:

1. Akurasi: Akurasi mengukur proporsi prediksi yang benar dari keseluruhan data. Semakin tinggi akurasi, semakin baik model dalam melakukan klasifikasi secara umum.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.7)$$

2. *Presisi*: *Presisi* menunjukkan tingkat ketepatan prediksi positif, yaitu seberapa banyak data yang diprediksi positif benar-benar positif. Nilai *presisi* tinggi berarti model jarang memberikan prediksi positif yang salah.

$$Presisi = \frac{TP}{TP + FP} \quad (3.8)$$

3. *Recall*: *Recall* mengukur kemampuan model dalam menemukan semua data positif. Nilai *recall* tinggi berarti sebagian besar data positif berhasil diprediksi dengan benar.

$$Recall = \frac{TP}{TP + FN} \quad (3.9)$$

4. *F1-score*: *F1-Score* merupakan rata-rata antara *presisi* dan *recall*. Metrik ini berguna ketika data tidak seimbang (jumlah data positif dan negatif tidak sama).

$$F1 - score = \frac{2 \cdot presisi \cdot recall}{presisi + recall} \quad (3.10)$$

### 3.4 Akuisisi Data

Data dikumpulkan dengan teknik *scraping* menggunakan *library* *google-play-scraper* yang tersedia di Python. Periode akuisisi data mulai dari 1 Januari 2020 hingga 31 Mei 2025. Data diambil dua kali, akuisisi pertama mendapatkan 1019 data ulasan, akuisisi kedua terdapat penambahan jumlah yaitu 1032 data ulasan. Seluruh data yang diperoleh kemudian diperiksa untuk menghapus ulasan yang terduplikasi menggunakan fungsi *df.drop\_duplicates* agar tidak memengaruhi hasil analisis. Data yang terduplikat dapat menimbulkan bias dalam proses pelatihan model sehingga dapat memengaruhi hasil evaluasi (Arnap & Kusri, 2024).

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Hasil Penelitian**

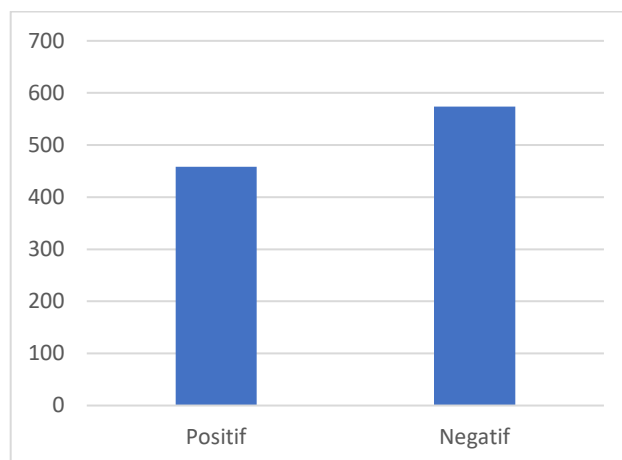
Pada BAB ini, memaparkan hasil analisis sentimen menggunakan metode Support Vector Machine (SVM) serta mengulas terkait analisis kepuasan membaca pengguna aplikasi Gramedia Digital. Terdapat tahap *preprocessing*, TF-IDF, klasifikasi dan analisis kepuasan membaca pengguna. Penelitian ini melakukan pengujian dengan menghitung akurasi, presisi, *recall* dan *F1-Score*. Analisis kepuasan membaca pengguna didapat dari *worldcloud* setiap sentimen dan ulasan yang mengandung kepuasan membaca pengguna. Seluruh proses dalam penelitian ini dilakukan menggunakan sistem yaitu Google Colab dengan Bahasa pemrograman Python kecuali pada tahap pelabelan data. Berikut ini adalah hasil dari penelitian yang dilakukan.

##### **4.1.1 Pelabelan Data**

Pelabelan dilakukan secara manual dengan melibatkan ahli di bidang Bahasa Indonesia. Data yang dilabeli merupakan data mentah hasil *scrapping* sebanyak 1.032 ulasan. Penggunaan data mentah bertujuan agar para ahli dapat memahami makna setiap kalimat secara utuh sebelum melalui tahap *preprocessing*.

Proses pelabelan dilakukan dengan membaca setiap ulasan secara menyeluruh, kemudian menentukan kategori sentimen berdasarkan makna utama yang terkandung dalam ulasan tersebut. Ulasan yang mengandung ungkapan kepuasan, apresiasi, atau penilaian positif terhadap aplikasi dikategorikan sebagai sentimen positif. Sementara itu, ulasan yang berisi keluhan, kekecewaan, atau kritik terhadap layanan dan fitur aplikasi dikategorikan sebagai sentimen negatif.

Apabila terdapat perbedaan pendapat dalam penentuan label, maka digunakan label yang memperoleh jumlah pilihan terbanyak dari para ahli. Maka dari itu penggunaan tiga ahli Bahasa agar dapat memilih dua sentimen terbanyak jika ada perbedaan. Berdasarkan proses tersebut, diperoleh hasil pelabelan yang ditampilkan pada gambar berikut.



**Gambar 4.1 Distribusi Sentimen**  
(Sumber: Hasil Olah Data Peneliti)

Gambar 4.1 merupakan visualisasi berupa diagram batang dimana garis horizontal menunjukkan jenis kelas sentimen, sedangkan garis vertikal menunjukkan jumlah pada masing-masing kelas sentimen. Sentimen positif memiliki jumlah 458 ulasan dan sentimen negatif memiliki jumlah 574 ulasan.

#### 4.1.2 Hasil *Preprocessing*

Data yang telah dilabeli oleh ahli selanjutnya dibersihkan untuk memaksimalkan proses klasifikasi. Tahapan *preprocessing* meliputi *case folding*, *cleaning* menggunakan *regular expression* (regex) untuk menghapus elemen seperti angka, URL, *emoticon*, hastag, tanda baca dan simbol lainnya, serta normalisasi untuk mengganti kata tidak baku menjadi bentuk baku. Proses *tokenizing* dilakukan dengan library NLTK, sedangkan *stopword removal* dan *stemming* menggunakan library Sastrawi. Hasil *case folding*, *cleaning*, normalisasi, *tokenizing*, *stopwords*, *stemming* sebagai berikut.

**Tabel 4.1 Data Mentah Ulasan**

No	Data Mentah
1.	smoga makin.banyak ebook yg update lagi
2.	perkembangannya luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
3.	saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook, sangat kecewa!

No	Data Mentah
4.	lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua??
5.	Senang sekali bisa baca banyak buku dalam satu aplikasi.

Tabel 4.1 berisi data mentah ulasan pengguna aplikasi Gramedia Digital yang diambil langsung dari Google Play Store sebelum melalui proses pembersihan. Data mentah ini yang kemudian dijadikan dasar dalam proses analisis pada tahap berikutnya.

**Tabel 4.2 Tahap Case Folding**

No	Case Folding
1.	smoga makin.banyak ebook yg update lagi
2.	perkembangannya luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
3.	saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook, sangat kecewa!
4.	lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua??
5.	senang sekali bisa baca banyak buku dalam satu aplikasi.

Tabel 4.2 menampilkan hasil proses *case folding* dengan mengubah huruf besar atau kapital menjadi huruf kecil semua. Selanjutnya dilakukan proses *cleaning* sebagai berikut.

**Tabel 4.3 Hasil Cleaning**

No	Cleaning
1.	smoga makin banyak ebook yg update lagi
2.	perkembangannya luar biasa readernya susah semakin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
3.	saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook sangat kecewa
4.	lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua
5.	senang sekali bisa baca banyak buku dalam satu aplikasi

Tabel 4.3 menampilkan hasil penghapusan atau pembersihan dari angka, URL, *emoticon*, hastag, tanda baca dan simbol lainnya, proses selanjutnya dilakukan tahap normalisasi dengan hasil berikut.

**Tabel 4.4 Hasil Normalisasi**

No	Normalisasi
1.	semoga makin banyak ebook yang update lagi
2.	perkembangannya luar biasa readernya susah semakin baik meski ada beberapa yang perlu di tingkatkan tapi gramedia digital sekarang sudah sangat oke
3.	saya baru update aplikasi dan semua bacaan yang pernah saya download hilang tidak bersisa padahal ada ratusan ebook sangat kecewa
4.	loh setelah diupdate koleksi yang sudah saya unduh kok hilang semua
5.	senang sekali bisa baca banyak buku dalam satu aplikasi

Tabel 4.4 menampilkan hasil normalisasi yang mengubah kalimat tidak baku atau *slang word* menjadi kata baku yang sesuai dengan kamus bahasa Indonesia, pada proses selanjutnya *tokenizing* seperti tabel berikut.

**Tabel 4.5 Hasil Tokenizing**

No	Tokenizing
1.	['semoga', 'makin', 'banyak', 'ebook', 'yang', 'update', 'lagi']
2.	['perkembangannya', 'luar', 'biasa', 'reader', 'nya', 'susah', 'semakin', 'baik', 'meski', 'ada', 'beberapa', 'yang', 'perlu', 'di', 'tingkatkan', 'tapi', 'gramedia', 'digital', 'sekarang', 'sudah', 'sangat', 'oke']
3.	['saya', 'baru', 'update', 'aplikasi', 'dan', 'semua', 'bacaan', 'yang', 'pernah', 'saya', 'download', 'hilang', 'tidak', 'bersisa', 'padahal', 'ada', 'ratusan', 'ebook', 'sangat', 'kecewa']
4.	['loh', 'setelah', 'diupdate', 'koleksi', 'yang', 'sudah', 'saya', 'unduh', 'kok', 'hilang', 'semua']
5.	['senang', 'sekali', 'bisa', 'baca', 'banyak', 'buku', 'dalam', 'satu', 'aplikasi']

Tabel 4.5 hasil *tokenizing* menampilkan pemisahan kata sehingga data yang tertokenisasi dapat dilakukan pembobotan menggunakan TF-IDF dengan maksimal, setelah itu dilakukan proses *stopword removal* dengan hasil berikut.

**Tabel 4.6 Stopword Removal**

No	Stopword Removal
1.	['semoga', 'makin', 'banyak', 'ebook', 'yang', 'update', 'lagi']

No	<i>Stopword Removal</i>
2.	['perkembangannya', 'luar', 'biasa', 'reader', 'susah', 'semakin', 'baik', 'meski', 'beberapa', 'perlu', 'tingkatkan', 'gramedia', 'digital', 'sekarang', 'sangat', 'oke']
3.	['baru', 'update', 'aplikasi', 'bacaan', 'pernah', 'download', 'hilang', 'bersisa', 'padahal', 'ratusan', 'ebook', 'kecewa']
4.	['loh', 'diupdate', 'koleksi', 'unduh', 'kok', 'hilang']
5.	['senang', 'sekali', 'baca', 'banyak', 'buku', 'aplikasi']

Tabel 4.6 menampilkan hasil *stopword* dengan menghapus kata yang tidak diperlukan dalam proses analisis, setelah itu dilakukan proses *stemming* dengan hasil berikut.

**Tabel 4.7 Hasil Stemming**

No	<i>Stemming</i>
1.	['semoga', 'makin', 'banyak', 'ebook', 'yang', 'update', 'lagi']
2.	['kembang', 'luar', 'biasa', 'reader', 'susah', 'baik', 'meski', 'beberapa', 'tingkat', 'gramedia', 'digital', 'sekarang', 'oke']
3.	['baru', 'update', 'aplikasi', 'baca', 'pernah', 'download', 'hilang', 'sisa', 'padahal', 'ratus', 'ebook', 'kecewa']
4.	['loh', 'update', 'koleksi', 'unduh', 'kok', 'hilang']
5.	['senang', 'kali', 'baca', 'banyak', 'buku', 'aplikasi']

Tabel 4.7 menampilkan hasil pengembalian kata menjadi kata dasar. Secara keseluruhan, proses *preprocessing* dilakukan terhadap 1.032 data ulasan yang diperoleh dari Google Play Store. Setelah melalui tahapan *case folding*, *cleaning*, *normalisasi*, *tokenizing*, *stopword removal*, dan *stemming*, jumlah data akhir yang digunakan dalam penelitian ini menjadi 940 ulasan. Maka, Terdapat 92 data yang tereliminasi dalam *dataset* akhir. *Dataset* hasil *preprocessing* ini kemudian disimpan dalam format *.csv* dan digunakan sebagai input pada tahap berikutnya.

#### 4.1.3 Hasil TF-IDF

Data hasil *preprocessing* selanjutnya dilakukan vektorisasi menggunakan *TfidfVectorizer* dari *library* *scikit-learn*. Hasil TF-IDF memunculkan nilai frekuensi kata berupa angka desimal, pada 940 dokumen dan terdapat 2.263 kata yang terbobot. Hasil penerapan TF-IDF disajikan pada tabel berikut.

Tabel 4.8 Hasil TF-IDF

No	Dokumen	Kata	Score Pembobotan
1.	Dokumen 1 (negatif)	makin	0,5088
2.	Dokumen 1 (negatif)	update	0,4611
3.	Dokumen 1 (negatif)	moga	0,4432
4.	Dokumen 1 (negatif)	ebook	0,4143
5.	Dokumen 1 (negatif)	banyak	0,4006

Pada Tabel 4.8 ditampilkan hasil TF-IDF terhadap sebagian data ulasan setelah melalui tahap *preprocessing*. Kolom Dokumen menunjukkan identitas ulasan beserta label sentimennya, kolom Kata berisi kata yang muncul pada ulasan tersebut, sedangkan kolom *Score* pembobotan menampilkan nilai bobot TF-IDF dalam bentuk bilangan desimal. Nilai bobot yang lebih besar menandakan bahwa kata tersebut lebih penting dalam dokumen karena relatif lebih sering muncul pada dokumen tersebut dan jarang muncul pada dokumen lain. Hasil pembobotan TF-IDF ini selanjutnya digunakan sebagai data pada proses klasifikasi menggunakan metode Support Vector Machine (SVM).

#### 4.1.4 Hasil klasifikasi SVM

Setelah diperoleh hasil TF-IDF, tahap selanjutnya adalah melakukan klasifikasi sentimen menggunakan metode Support Vector Machine (SVM). Untuk melakukan klasifikasi dilakukan pembagian data latih dan data uji terlebih dahulu.

##### 4.1.4.1 Skenario Uji Coba

Pembagian data dilakukan dengan memanfaatkan fungsi *train\_test\_split* dari *library* scikit-learn, yang digunakan untuk memisahkan data secara acak namun tetap terarah sesuai dengan rasio yang telah ditetapkan. Pada skenario 70:30 yang berarti 70% data latih dan 30% data uji, dari 940 ulasan diperoleh 658 data latih dan 282 data uji, skenario 80:20 atau 80% data latih dan 20% data uji menghasilkan 752 data latih dan 188 data uji, sedangkan skenario 90:10 atau 90% data latih dan 10% data uji, menghasilkan 846 data latih dan 94 data uji. Seluruh

data latih pada masing-masing skenario kemudian digunakan untuk melatih metode Support Vector Machine (SVM) dengan kernel linear agar mampu mempelajari hubungan antara  $x_{train}$  yang didapat dari vektor fitur TF-IDF dan  $y_{train}$  dari label sentimen, sehingga model dapat mengklasifikasikan ulasan positif dan negatif secara lebih tepat pada tahap pengujian.

Setelah data terbagi menjadi data latih dan data uji, tahap berikutnya adalah proses pelatihan model menggunakan metode Support Vector Machine (SVM). Pada tahap ini, vektor fitur hasil TF-IDF dari data latih dimasukkan ke dalam fungsi  $fit()$  bersama dengan label sentimennya sehingga model dapat mempelajari pola keterkaitan antara kata-kata yang muncul dalam ulasan dengan kategori sentimen positif maupun negatif. Selama proses pelatihan, metode SVM membentuk garis batas pemisah atau *hyperplane* yang memaksimalkan jarak antara kelas positif dan negatif. Model yang telah terbentuk dari data latih inilah yang kemudian digunakan untuk memprediksi label pada data uji dan menjadi dasar perhitungan kinerja klasifikasi pada tahap pengujian.

#### 4.1.4.2 Hasil Pengujian Data

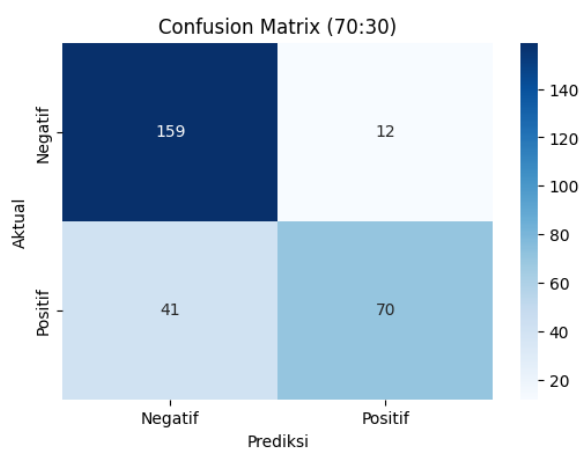
Setelah model SVM selesai dilatih pada masing-masing skenario pembagian data, tahap selanjutnya adalah melakukan pengujian menggunakan data uji yang telah dipisahkan dan dievaluasi menggunakan *Confusion Matrix*. Evaluasi ini dilakukan untuk menilai seberapa tepat hasil klasifikasi yang dihasilkan oleh sistem. Untuk memperoleh nilai *true positive*, *true negative*, *false positive*, dan *false negative*, digunakan fungsi *confusion\_matrix* dari *library* scikit-learn. Setelah itu, fungsi *classification\_report* dimanfaatkan untuk menilai kinerja model secara lebih menyeluruh dengan menyajikan metrik evaluasi berupa akurasi, presisi, *recall*, dan *F1-score* pada setiap skenario pembagian data dengan rasio sebagai berikut.

Tabel 4.9 Hasil Evaluasi SVM

Rasio Skenario	Akurasi	Presisi	Recall	F1-Score
70:30	0.81	0.82	0.78	0.79
80:20	0.80	0.81	0.76	0.77
90:10	0.79	0.79	0.76	0.77

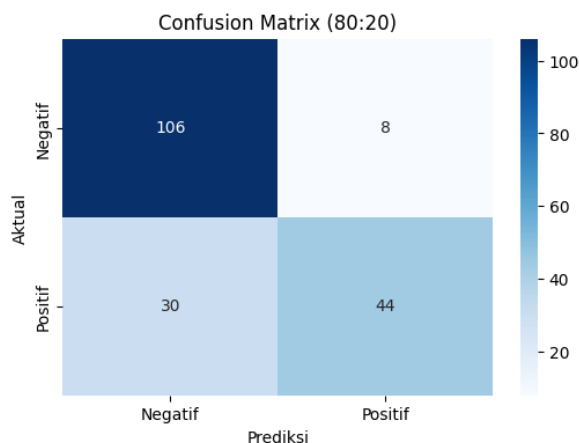
Tabel 4.8 menunjukkan hasil evaluasi kinerja model SVM pada tiga skenario pembagian data. Pada skenario 70:30, model menghasilkan akurasi sebesar 81% dengan nilai presisi 82%, *recall* 78%, dan F1-score 79%. Skenario 80:20 memberikan akurasi sedikit lebih rendah, yaitu 80% dengan presisi 81%, *recall* 76%, dan F1-score 77%. Adapun skenario 90:10 menghasilkan akurasi 79%, presisi 79%, *recall* 76%, dan F1-score 77%.

Pada skenario pembagian data 70:30, distribusi kelas pada data latih dan data uji tetap menunjukkan dominasi sentimen negatif dibandingkan sentimen positif. Data latih terdiri atas 398 ulasan negatif dan 260 ulasan positif, sedangkan data uji terdiri atas 171 ulasan negatif dan 111 ulasan positif. Kondisi ini menunjukkan bahwa komposisi kelas pada proses pelatihan dan pengujian masih mencerminkan kecenderungan distribusi sentimen pada *dataset* penelitian. Hasil evaluasi tersebut kemudian divisualisasikan menggunakan *library* scikit-learn dalam bentuk tabel matriks seperti pada gambar berikut.



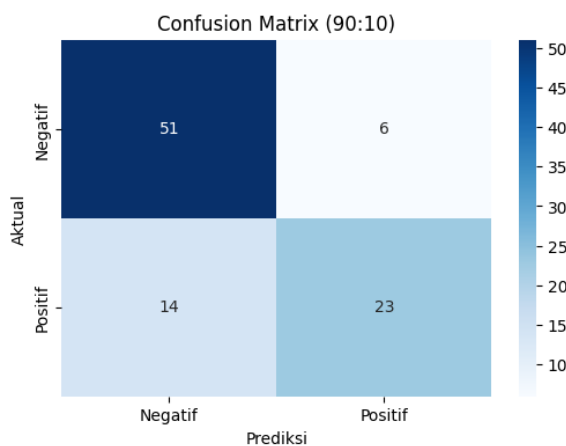
**Gambar 4.2 Confusion Matrix 70:30**  
(Sumber: Hasil Olah Data Peneliti)

Gambar 4.3 menyajikan confusion matrix untuk skenario pembagian data 70:30. Pada matriks tersebut tampak bahwa model mampu mengklasifikasikan 159 ulasan negatif sebagai negatif (*true negative*) dan 70 ulasan positif sebagai positif (*true positive*). Namun, masih ditemukan 12 ulasan yang sebenarnya negatif tetapi diprediksi sebagai positif (*false positive*), serta 41 ulasan yang sebenarnya positif namun salah terdeteksi sebagai negatif (*false negative*).



**Gambar 4.3 Confusion Matrix 80:20**  
(Sumber: Hasil Olah Data Peneliti)

Gambar 4.4 menggambarkan confusion matrix pada skenario pembagian data 80:20. Dalam skenario ini, model berhasil mengklasifikasikan 106 ulasan negatif secara benar sebagai negatif (*true negative*) dan 44 ulasan positif secara tepat sebagai positif (*true positive*). Di sisi lain, masih terdapat 8 ulasan yang sebenarnya negatif tetapi terprediksi sebagai positif (*false positive*), serta 30 ulasan positif justru terdeteksi sebagai negatif (*false negative*).



**Gambar 4.4 Confusion Matrix 90:10**  
(Sumber: Hasil Olah Data Peneliti)

Gambar 4.5 menampilkan confusion matrix untuk skenario pembagian data 90:10. Pada skenario ini, model mampu mengklasifikasikan 51 ulasan negatif sebagai negatif (*true negative*) dan 23 ulasan positif sebagai positif (*true positive*). Sementara itu, masih dijumpai 6 ulasan yang sebenarnya negatif namun diprediksi



Pada gambar 4.7 dapat dijelaskan *wordcloud* ulasan negatif didominasi oleh kata buku, baca, beli, bayar, langgan, buka, download, dan fitur. Kata buku dan baca Kembali menjadi yang sering muncul, tetapi banyak digunakan dalam memberikan keluhan. Kata bayar, beli, dan langgan berkaitan dengan ketidakpuasan terhadap proses pembayaran dan akses. Kata buka, download, dan fitur menunjukkan masalah teknis dan keterbatasan fasilitas baca.

Berdasarkan hasil dari kata yang sering muncul, terdapat kalimat asli yang dapat memperjelas makna kata-kata yang muncul dominan tersebut. Pada sentimen positif, yaitu kata baca dan buku tampak dalam ungkapan pengguna yang merasa “bisa baca banyak buku dalam satu aplikasi” dan menyebut aplikasi ini “membantu membaca jadi lebih praktis”. Pada sentimen negatif, kata baca, buku, bayar, dan langgan muncul dalam keluhan seperti buku yang sudah dibeli “tidak bisa dibuka”, bacaan terhenti karena aplikasi “sering keluar sendiri”, atau paket langganan yang sudah dibayar tetapi akses buku “Aplikasinya jujur, jelek sih. Ngga ada kategori buku, ngga ada filter buku, baik misalnya sesuai abjad, judul atau pengarang”. Ada juga kalimat yang menyinggung tampilan, yaitu pengguna merasa “tampilannya malah sulit untuk dibaca, perlu di *zoom*, dan setelah di *zoom* harus diperkecil” atau perpindahan halaman yang membuat “Lompat beberapa halaman susah”.

## **4.2 Pembahasan**

Pada pembahasan ini, menginterpretasikan dari hasil penelitian untuk mengetahui analisis sentimen dan kepuasan membaca pengguna aplikasi Gramedia Digital dan keterkaitannya dengan nilai keislaman.

### **4.2.1 Kemampuan Metode *Support Vector Machine* (SVM) dalam Melakukan Analisis Sentimen terhadap Ulasan Aplikasi Gramedia Digital**

Data yang digunakan penelitian ini merupakan ulasan pengguna aplikasi Gramedia Digital yang berisi pengalaman, penilaian dan keluhan pengguna terhadap aplikasi tersebut. Ulasan aplikasi Gramedia Digital perlu dianalisis karena jumlahnya cukup banyak dan berbentuk teks dengan gaya bahasa yang berbeda-beda atau tidak terstruktur, sehingga sulit dipahami untuk dianalisis secara manual.

Oleh karena itu, dilakukan analisis sentimen dengan metode Support Vector Machine (SVM) untuk mengklasifikasikan ulasan ke dalam sentimen positif dan negatif. Klasifikasi dilakukan dengan tiga skenario pembagian data latih dan uji, yaitu 70:30, 80:20, dan 90:10. Penggunaan beberapa rasio pembagian data bertujuan untuk memperoleh performa model yang paling baik. Setelah proses pelatihan model dilakukan, selanjutnya dievaluasi menggunakan data uji melalui confusion matrix menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*.

Berdasarkan tiga skenario pengujian, rasio 70:30 menghasilkan performa terbaik dengan akurasi 81%. Hal ini menunjukkan bahwa komposisi data latih sebesar 70% sudah cukup untuk membentuk pola pemisahan sentimen tanpa mengurangi variasi pada data uji. Peningkatan proporsi data latih hingga 80% maupun 90% tidak memberikan peningkatan performa model. Berdasarkan penelitian sebelumnya tentang analisis sentimen dengan objek aplikasi Dompot Digital dan data dari ulasan di Google Play Store, menunjukkan proporsi data 70:30 cukup seimbang antara data latih dan data uji, sehingga model mampu mengenali karakteristik yang dimiliki pada tiap kelasnya dengan baik (Putri et al., 2024). Kondisi ini menunjukkan bahwa kualitas variasi data lebih berpengaruh dibandingkan sekadar banyaknya jumlah data latih. Temuan tersebut menunjukkan bahwa penambahan proporsi data latih belum tentu dapat meningkatkan performa model, karena kemampuan klasifikasi juga dipengaruhi oleh karakteristik data dan kejelasan pola sentimen yang dipelajari model. Hasil ini sejalan dengan penelitian yang menunjukkan bahwa metode SVM juga mencapai performa terbaik pada skenario pembagian data 70:30, sehingga memperlihatkan bahwa proporsi data latih yang lebih besar tidak menjadi penentu utama peningkatan kinerja model, melainkan kecocokan distribusi data dan pola sentimen yang dipelajari model (Pratama et al., 2024). Keunggulan skenario 70:30 pada penelitian ini menunjukkan bahwa pembagian data tersebut sudah cukup untuk membantu model mempelajari pola sentimen dan tetap mempertahankan variasi data uji. Kondisi ini memperlihatkan bahwa performa model yang baik dalam penelitian ini tidak hanya ditentukan oleh jumlah data latih, tetapi juga oleh kesesuaian antara distribusi data dan kompleksitas bahasa pada ulasan pengguna.

Hasil pengujian SVM pada tiga skenario pembagian data latih dan uji menunjukkan fenomena menarik, yaitu skenario dengan proporsi data latih terkecil 70:30 menghasilkan akurasi terbaik. Kondisi ini menunjukkan bahwa ulasan pengguna Gramedia Digital relatif monoton, dengan pola komentar yang berulang-ulang, sehingga penambahan jumlah data latih pada skenario 80:20 atau 90:10 tidak meningkatkan performa model. Berulangnya ulasan menyebabkan SVM kesulitan membedakan sentimen positif yang mengandung variasi ekspresi, sementara ulasan negatif lebih mudah dikenali karena pola kata yang lebih eksplisit. Dengan kata lain, kualitas variasi data lebih memengaruhi performa klasifikasi daripada kuantitas data latih semata.

Hasil analisis menggunakan confusion matrix pada ketiga skenario pembagian data menunjukkan bahwa pada rasio 70:30, model lebih baik dalam mengidentifikasi ulasan negatif, namun masih sering melakukan kesalahan pada ulasan positif, terutama pada ulasan yang memiliki kata-kata mirip dengan ulasan negatif. Pada rasio 80:20 dan 90:10, kecenderungan serupa juga muncul, yaitu model lebih konsisten mengklasifikasikan ulasan negatif dan lebih sering salah mengklasifikasikan ulasan positif sebagai negatif. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali ulasan positif Gramedia Digital. Kesalahan prediksi tersebut dapat dipahami dari karakteristik bahasa pada data ulasan. Ulasan negatif umumnya menggunakan kata-kata yang lebih eksplisit, seperti “tidak bisa”, “error”, “jelek”, dan “buruk”, sehingga lebih mudah dikenali oleh model. Sedangkan ulasan positif dalam penelitian ini cukup sering memuat unsur apresiasi sekaligus kritik, sehingga sentimennya tidak dinyatakan secara tegas. Kondisi ini menyebabkan model cenderung memilih kelas negatif ketika kata atau pola yang ditemukan tidak cukup jelas (Herza et al., 2024). Kesalahan klasifikasi yang lebih banyak pada kelas positif menunjukkan bahwa tantangan utama model terletak pada ulasan yang mengandung sentimen campuran dan tidak sepenuhnya tegas secara semantik.

Capaian akurasi sebesar 81% pada skenario pembagian data 70:30 menunjukkan bahwa metode Support Vector Machine (SVM) mampu mengklasifikasikan sentimen ulasan aplikasi Gramedia Digital dengan tingkat

ketepatan yang cukup baik. Nilai presisi, *recall*, dan *F1-score* yang berada pada rentang yang relatif seimbang juga memperlihatkan bahwa model tidak hanya tepat dalam memprediksi kelas tertentu, tetapi juga cukup konsisten dalam mengenali kedua kategori sentimen. Hasil ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa SVM memiliki kemampuan yang baik dalam klasifikasi teks dengan akurasi yang baik sebesar 96,68% (Arsi & Waluyo, 2021).

Hasil penelitian menunjukkan bahwa metode Support Vector Machine (SVM) pada skenario 70:30 mampu mengklasifikasikan sentimen ulasan pengguna dengan cukup baik setelah dievaluasi. Evaluasi tersebut merupakan bentuk pengukuran terhadap kemampuan model dalam mengklasifikasikan sentimen secara tepat dan konsisten. Prinsip pengukuran yang objektif dan adil dalam evaluasi model ini sejalan dengan firman Allah Swt. dalam Surah Al-Muthaffifin ayat 1–3.

وَيْلٌ لِّلْمُطَفِّفِينَ. الَّذِينَ إِذَا أَكْتَالُوا عَلَى النَّاسِ يَسْتَوْفُونَ. وَإِذَا كَالُواهُمْ أَوْ وُزِنُوا لَهُمْ يُخْسِرُونَ

*Artinya: “Celakalah bagi orang-orang yang curang (dalam menakar dan menimbang, (yaitu) orang-orang yang apabila menerima takaran dari orang lain, mereka minta dicukupkan dan apabila mereka menakar atau menimbang (untuk orang lain), mereka mengurangi”.*

Tafsir Surah Al-Mutaffifin ayat 1-3 menjelaskan bahwa Allah Swt. memberikan peringatan yang sangat keras kepada orang-orang yang berbuat curang dalam takaran dan timbangan. Kecurangan itu tampak ketika seseorang menuntut haknya dipenuhi secara sempurna saat menerima takaran atau timbangan dari orang lain, tetapi dapat mengurangi takaran atau timbangan ketika ia memberikan hak orang lain. Perilaku ini menunjukkan adanya ketidakjujuran, ketidakadilan, dan sikap mementingkan diri sendiri dengan merugikan pihak lain secara sengaja. Kecurangan semacam ini bukan hanya merugikan manusia dalam urusan muamalah, tetapi juga menjadikan harta yang diperoleh bersifat haram, tidak membawa keberkahan, dan mengantarkan pelakunya kepada azab Allah (Kemenag, 2025).

Ayat tersebut menegaskan pentingnya kejujuran dan keadilan dalam setiap bentuk pengukuran. Dalam penelitian ini, proses evaluasi menggunakan *confusion*

*matrix* dapat dianalogikan sebagai bentuk timbangan ilmiah yang digunakan untuk menilai kinerja model. Metrik akurasi, presisi, *recall*, dan *F1-score* menjadi indikator yang memastikan bahwa hasil klasifikasi tidak bersifat subjektif, melainkan berdasarkan perhitungan yang terukur dan transparan.

#### 4.2.2 Analisis Kepuasan Membaca Pengguna dari Hasil Analisis Sentimen

Berdasarkan hasil prediksi pada skenario terbaik analisis sentimen, ulasan negatif lebih banyak dibandingkan dengan ulasan positif, yakni 569 ulasan negatif dan 371 ulasan positif. Meskipun aplikasi Gramedia Digital memiliki *rating* 4,2 di Google Play Store, sebagian besar pengguna masih memberikan tanggapan yang lebih cenderung negatif terhadap pengalaman membaca mereka. Hal ini menunjukkan bahwa *rating* aplikasi belum mencerminkan kepuasan pengguna secara menyeluruh. Sebagian besar pengguna menilai aplikasi ini belum sepenuhnya memenuhi harapan mereka, meskipun menawarkan berbagai macam koleksi.

Banyaknya sentimen negatif dalam penelitian ini menunjukkan bahwa pengguna cenderung masih mengalami ketidakpuasan dalam menggunakan layanan membaca pada aplikasi Gramedia Digital. Ketidakpuasan tersebut terlihat pada keluhan mengenai buku yang tidak dapat diakses setelah pembayaran atau langganan, aplikasi yang sering *logout* otomatis, kestabilan sistem yang mengganggu proses membaca, dan fitur yang belum sepenuhnya mendukung kebutuhan pengguna. Selain itu, tampilan baca yang kurang nyaman karena belum adanya, fitur untuk mengubah ukuran *font* dan keterbatasan fitur pencarian juga memperkuat munculnya persepsi negatif dari pengguna. Kesulitan pengguna untuk mengakses buku secara lancar dapat mengurangi efektivitas layanan, sementara proses akses yang lama dan gangguan teknis pada aplikasi dapat menurunkan efisiensi penggunaan (Hendrayani & Rahmi, 2025).

Hasil analisis kepuasan membaca pengguna pada penelitian ini menunjukkan bahwa pengguna Gramedia Digital masih mengalami beberapa kendala yang memengaruhi kepuasan dalam menggunakan aplikasi. Temuan tersebut terlihat dari dominasi ulasan negatif yang berkaitan dengan keterbatasan

fitur, gangguan sistem seperti aplikasi yang sering *logout* otomatis, kesulitan akses buku dan keluhan dalam melakukan pembayaran atau berlangganan. Hasil ini sejalan dengan penelitian yang membahas faktor yang memengaruhi kepuasan pengguna aplikasi Gramedia Digital. Penelitian tersebut menyatakan bahwa *usability*, *dependability*, dan *perceived value* berpengaruh terhadap kepuasan pengguna. *Usability* berkaitan dengan ketersediaan fitur dalam aplikasi, *dependability* berkaitan dengan kestabilan sistem dalam aplikasi untuk memudahkan pengguna dalam mengakses, sedangkan *perceived value* berkaitan dengan kesesuaian manfaat yang diterima pengguna dengan biaya yang telah dikeluarkan (Hapsari et al., 2025). Dalam penelitian ini, keluhan mengenai kurangnya fitur, aplikasi yang sering *logout* atau stabilitas aplikasi, proses pembayaran dan langganan, serta akses koleksi menunjukkan bahwa ketiga faktor tersebut berperan dalam membentuk kepuasan membaca pengguna Gramedia Digital. Kondisi ini menunjukkan bahwa kepuasan membaca pengguna tidak hanya ditentukan oleh ketersediaan koleksi bacaan, tetapi juga oleh kemudahan penggunaan aplikasi, keandalan sistem, dan manfaat yang dirasakan pengguna setelah menggunakan layanan.

Kondisi ini memiliki keterkaitan dengan nilai keislaman yang menekankan pentingnya kualitas dan tanggung jawab dalam memberikan layanan kepada pengguna. Sebagaimana dijelaskan dalam Surat Al-Baqarah [2]:267, Allah SWT berfirman.

يَا أَيُّهَا الَّذِينَ آمَنُوا أَنْفِقُوا مِنْ طَيِّبَاتِ مَا كَسَبْتُمْ وَمِمَّا أَخْرَجْنَا لَكُمْ مِنَ الْأَرْضِ ۖ وَلَا تَيَمَّمُوا الْخَبِيثَ مِنْهُ تُنْفِقُونَ  
وَأَلْسِنَتُمْ بِإِجْزِيهِ إِلَّا أَنْ تُغْمِضُوا فِيهِ ۗ وَاعْلَمُوا أَنَّ اللَّهَ عَنِّي حَمِيدٌ

*Artinya: “Wahai orang-orang yang beriman, infakkanlah sebagian dari hasil usahamu yang baik-baik dan sebagian dari apa yang Kami keluarkan dari bumi untukmu. Janganlah kamu memilih yang buruk untuk kamu infakkan, padahal kamu tidak mau mengambilnya, kecuali dengan memicingkan mata (enggan) terhadapnya. Ketahuilah bahwa Allah Maha Kaya lagi Maha Terpuji”.*

Dalam tafsir kemenag, surat Al-Baqarah ayat 267 menegaskan bahwa setiap bentuk pemberian, harus diberikan dalam kondisi yang baik dan layak (Kemenag, 2025). Ayat tersebut mengajarkan bahwa layanan yang diberikan kepada orang lain seharusnya merupakan layanan yang baik dan layak, bukan layanan yang kita sendiri tidak mau menerimanya. Ajaran ini dapat diterapkan dalam penyediaan layanan membaca dengan memberikan layanan yang baik berarti menghadirkan pengalaman yang nyaman dan stabil bagi pengguna. Sementara layanan yang buruk berupa kendala teknis, fitur yang tidak berjalan, atau ketidaknyamanan penggunaan yang berarti bentuk pelayanan yang kurang memenuhi standar *ṭayyib* sebagaimana yang diajarkan dalam Islam.

Selain itu, untuk menyempurnakan pelayanan dapat dipahami melalui Surat Ali ‘Imran ayat 159 yang menjelaskan pentingnya bersikap dalam berinteraksi ketika terdapat persoalan.

فَمَا رَحْمَةٌ مِّنَ اللَّهِ إِنَّكَ لَئِن كُنْتَ فَظًّا غَلِيظَ الْقَلْبِ لَانْقَضُوا مِنْ حَوْلِكَ فَاعْفُ عَنْهُمْ وَاسْتَغْفِرْ لَهُمْ وَشَاوِرْهُمْ فِي الْأَمْرِ فَإِذَا عَزَمْتَ فَتَوَكَّلْ عَلَى اللَّهِ إِنَّ اللَّهَ يُحِبُّ الْمُتَوَكِّلِينَ

*Artinya: “Maka, berkat rahmat Allah engkau (Nabi Muhammad) berlaku lemah lembut terhadap mereka. Seandainya engkau bersikap keras dan berhati kasar, tentulah mereka akan menjauh dari sekitarmu. Oleh karena itu, maafkanlah mereka, mohonkanlah ampunan untuk mereka, dan bermusyawarahlah dengan mereka dalam segala urusan (penting). Kemudian, apabila engkau telah membulatkan tekad, bertawakallah kepada Allah. Sesungguhnya Allah mencintai orang-orang yang bertawakal”.*

Rasulullah saw, tetap menunjukkan sikap lemah lembut, pemaaf, dan penuh kasih sayang meskipun menghadapi situasi yang berat, termasuk ketika sebagian kaum Muslimin melakukan kesalahan dalam Perang Uhud yang menimbulkan kerugian besar. Dalam kondisi tersebut, Nabi tidak merespons dengan kemarahan atau kekasaran, melainkan memilih memaafkan mereka, memohonkan ampun kepada Allah untuk mereka, dan tetap melibatkan mereka dalam musyawarah. Sikap ini menunjukkan bahwa pendekatan yang lembut, terbuka, dan mengedepankan kebersamaan justru mampu menjaga kedekatan serta kepercayaan

umat. Ayat ini menegaskan bahwa kekerasan sikap dan hati yang kasar hanya akan membuat orang menjauh, sedangkan kelembutan, pemaafan, dan musyawarah menjadi dasar penting dalam membina hubungan dan menyelesaikan persoalan (Kemenag, 2025).

Ulasan pengguna ini dapat dipandang sebagai bentuk interaksi modern yang menyampaikan masukan dan kritik terkait layanan. Jika keluhan pengguna diabaikan atau ditanggapi dengan layanan yang kurang baik, pengguna akan dapat meninggalkan aplikasi tersebut. Apabila masukan pengguna diperhatikan dan direspons melalui perbaikan kualitas, hal tersebut dapat meningkatkan kepuasan pengguna. Maka dari itu, perbaikan teknis dan peningkatan kualitas layanan pada aplikasi merupakan bagian dari upaya untuk menghadirkan layanan yang lebih baik dan sesuai dengan nilai ajaran Islam.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Machine (SVM) mampu melakukan analisis sentimen terhadap ulasan pengguna aplikasi Gramedia Digital dengan performa yang cukup baik. Dari tiga skenario pengujian, skenario 70:30 merupakan skenario terbaik dengan nilai akurasi 81%, presisi 82%, *recall* 78%, dan *F1-score* 79%. Hasil tersebut menunjukkan bahwa SVM mampu melakukan analisis sentimen ulasan, karena memiliki tingkat ketepatan prediksi dengan akurasi yang cukup baik. Berdasarkan hasil analisis sentimen, kepuasan membaca pengguna aplikasi Gramedia Digital menunjukkan bahwa sebagian besar pengguna merasa belum puas. Hal ini terlihat dari dominasi ulasan negatif serta keluhan yang banyak berkaitan dengan keterbatasan fitur, gangguan stabilitas aplikasi, dan kendala akses bacaan setelah pembayaran atau langganan. Oleh karena itu, kepuasan membaca pengguna Gramedia Digital ditentukan oleh kemudahan penggunaan, kestabilan layanan, dan kesesuaian manfaat yang diterima pengguna saat menggunakan aplikasi.

#### **5.2 Saran**

1. Untuk peneliti selanjutnya, pengujian dapat diperluas dengan strategi yang lebih beragam, seperti *cross-validation* dan penggunaan kernel lain pada parameter SVM agar performa tidak hanya bergantung pada satu pembagian data.
2. Pengembang Gramedia Digital perlu meningkatkan stabilitas aplikasi, memastikan akses buku setelah pembayaran berjalan lancar, serta menambahkan fitur pengaturan ukuran huruf dan filter pencarian pada buku gratis.

## DAFTAR PUSTAKA

- Afifah, U. (2025). *Penjelasan Hadis Berkata Baik atau Diam*. Muslimah.
- Al-Husna, G. S., Asmarajati, D., Ihsannuddin, I. A., & Mahmudati, R. (2024). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 3(2), 139–144. <https://doi.org/10.55123/storage.v3i2.3602>
- Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon*, 10(9), e27863. <https://doi.org/10.1016/j.heliyon.2024.e27863>
- Andita, S. S. P. (2022). Manfaat Perpustakaan Digital Dalam Meningkatkan Minat Baca Generasi Milenial di Era Globalisasi. *LIBRIA*, 14, 123–142.
- Aqilah, Z. R., Fitriyani, W., Azalia, N. D., & Lushinta, I. P. (2025). Jurnal Penelitian Nusantara Kepuasan Mahasiswa FIP UPI Terhadap Perpustakaan Digital Dan Pengaruhnya Terhadap Minat Baca. *Jurnal Penelitian Nusantara*, 1, 93–100.
- Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 6(2), 129. <https://doi.org/10.30998/string.v6i2.10133>
- Arnap, A., & Kusriani. (2024). Optimasi Model Support Vector Machine dengan Particle Swarm Optimization untuk Mendeteksi Serangan Injeksi SQL ( Studi Kasus : PT . Naisha Inspirasi Muslimah ). *Jnanaloka*.
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Astuti, K. C., Firmansyah, A., & Riyadi, A. (2024). Implementasi Text Mining Untuk Analisis Sentimen Masyarakat Terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store. *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8(1), 383–394.
- Bei, F., & Saepudin, S. (2021). Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm). *Sismatik*, 01(01), 91–97.

- Dewi, M. R. (2020). Klasifikasi Akses Internet Oleh Anak-Anak dan Remaja Dewasa di Jawa Timur Menggunakan Support Vector Machine. *Jurnal Riset Dan Aplikasi Matematika (JRAM)*, 4(1), 17. <https://doi.org/10.26740/jram.v4n1.p17-27>
- Fitriyana, V., Hakim, L., Novitasari, D. C. R., & Asyhar, A. H. (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. *Jurnal Buana Informatika*, 14(1), 40–49.
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
- Hapsari, O. N. I., Az-zahra, H. M., & Aknuranda, I. (2025). Analisis Faktor-Faktor yang Memengaruhi Kepuasan Pengguna Aplikasi Gramedia Digital. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(3), 1–10.
- Hendrayani, P. F. A., & Rahmi, R. (2025). Analisis Perbandingan Sentimen Pengguna iPusnas dan JakLITera dengan Algoritma SentiWordNet. *Jurnal Ilmu Informasi, Perpustakaan Dan Kearsipan*, 27(2). <https://doi.org/10.7454/JIPK.v27i2.1132>
- Herza, F. M., Rahmat, B., Muharrom, M., & Haromainy, A. L. (2024). Pengaruh Rfe Terhadap Logistic Regression Dan Support Vector Machine Pada Analisis Sentimen Hotel Shangri-La Surabaya. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11612–11619.
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- Jasmarizal, Junadhi, Rahmaddeni, & M. Khairul Anam. (2024). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap Produk Skincare. *Indonesian Journal of Computer Science*, 13(1), 1438–1450. <https://doi.org/10.33022/ijcs.v13i1.3654>
- Judijanto, L., Hakim, L., Utami, W. S., & Adiazmil, A. (2025). The Impact of the BookTok Phenomenon on the Transformation of Generation Z ' s Reading Habits in the Digital Age in Indonesia. *The Eastasouth Journal of Social Science and Humanities*, 2(02), 152–160. <https://doi.org/10.58812/esssh.v2i02>
- Karimah, A., Dwilestari, G., & Mulyawan, M. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–737. <https://doi.org/10.36040/jati.v8i1.8373>
- Kemenag. (2025). *Kementerian Agama RI*.
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap

- Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), 302–307. <https://doi.org/10.30591/jpit.v8i3.5734>
- Mahfud, F. K. R., Mudawamah, N. S., & Hariyanto, W. (2020). Sentiment Analysis of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. *MATICS: Jurnal Ilmu Komputer Dan ...*, 12(1), 90–93.
- Melania, A. L. (2020). *Baca E-Books Gratis dengan Free Trial Premium Gramedia Digital*. Gramedia.
- Nufairi, F., Pratiwi, N., & Herlando, F. (2024). Analisis Sentimen pada Ulasan Aplikasi Threads di Google Play Store Menggunakan Algoritma Support Vector Machine. *JUPI (Jurnal Ilmu Perpustakaan Dan Informasi)*, 9(1), 339–348. <https://doi.org/10.31539/intecom.v5i1.3708>
- Nurhayati, Setyadinsa, R., Pradana, M. G., Prasetyo, D., Adrezo, M., Pinastawa, I. W. R., & Yasmin, J. K. (2024). *Machine Learning*. PT Samudera Media.
- Nurian, A., Ma'arif, M. S., Amalia, I. N., & Rozikin, C. (2024). Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1). <https://doi.org/10.23960/jitet.v12i1.3631>
- Prakash, T. N., & Aloysius, A. (2021). Textual Sentiment Analysis using Lexicon Based Approaches. *Annals of the Romanian Society for Cell Biology*, 25(4), 9878–9885. <http://annalsofrscb.ro>
- Pratama, E. A., Primajaya, A., & Maulana, I. (2024). Penerapan Algoritma Support Vector Machine Terhadap Analisis Sentimen Pada Ulasan Aplikasi Linkaja. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 13(3), 1908–1917.
- Putra, A. D. A., & Juanita, S. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646. <https://doi.org/10.35957/jatisi.v8i2.962>
- Putri, H. M., Faisal, M., & Kurniawan, M. F. (2024). Multi Stage Analisis Sentimen Berbasis Aspek Pada Ulasan Pengguna Aplikasi Dompot Digital Menggunakan Metode Multinomial Naive Bayes. *The Indonesian Journal of Computer Science*, 13(5), 8018–8028.
- Radiena, G., & Nugroho, A. (2023). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, 6(1), 1–10. <https://doi.org/10.37792/jukanti.v6i1.836>
- Rahma, D. S., Wulandari, K., & Ridha, A. A. (2024). Analisis User Interface Dan User Experience Pada Aplikasi Gramedia Digital Menggunakan Metode Human Centered Design. *DINAMIK*, 29(2).

- Ramadhan, M. F. (2025). Klasifikasi Topik dan Sentimen Judul Berita dengan Augmentasi dan. *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 4(2), 6732–6741.
- Ratnaningsih, W. O. W., Murwani, E., & Berto, A. R. (2023). Proses Pengambilan Keputusan Adopsi Inovasi Aplikasi Gamedia Digital. *Jurnal Komunikasi Profesional*, 6(6), 573–587. <https://doi.org/10.25139/jkp.v6i6.5294>
- Setiawan, G. H., & Adnyana, I. M. B. (2023). Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models. *Journal of Applied Informatics and Computing*, 7(2), 252–257. <https://doi.org/10.30871/jaic.v7i2.6527>
- Sinaga, A., & Nainggolan, S. P. (2023). Analisis Perbandingan Akurasi Dan Waktu Proses Algoritma Stemming Arifin-Setiono Dan Nazief-Adriani Pada Dokumen Teks Bahasa Indonesia. *Sebatik*, 27(1), 63–69. <https://doi.org/10.46984/sebatik.v27i1.2072>
- Suhardjono, Ganda, W., & Abdul, H. (2019). Prediksi Kellusan Menggunakan Svm Berbasis Pso. *Bianglala Informatika*, 7(2), 97–101.
- Vonega, D. A., Fadila, A., & Kurniawan, D. E. (2022). Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024. *Journal of Applied Informatics and Computing*, 6(2), 129–135. <https://doi.org/10.30871/jaic.v6i2.4300>
- Warow, G. G., & Pandia, H. (2024). Analisis Sentimen Aplikasi Dana Menggunakan Naïve Bayes Classifier dan Support Vector Machine. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(1), 609. <https://doi.org/10.35889/jutisi.v13i1.1893>
- Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 06, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>
- Zidna, A., Ali, M., Sri, M., & Joko Dwi, S. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Journal of Information System Management (JOISM)*, 3(1), 16–21.

## LAMPIRAN

### Lampiran 1 *Preprocessing*

#### *Case Folding*

```
#casefolding/lowering
df['lower'] = df['content'].apply(lambda x: str(x).lower())
df[['content', 'lower']].head(10)
```

	content	lower
0	Kenapa semua ebook yg udah didownload gak ke s...	kenapa semua ebook yg udah didownload gak ke s...
1	kurang update. buku yg aku cari semuanya gak ada	kurang update. buku yg aku cari semuanya gak ada
2	tolonglah, kalau udah langganan premiumnya itu...	tolonglah, kalau udah langganan premiumnya itu...
3	bagus	bagus
4	ok	ok
5	halo ... saya susah masuk ke akun saya lagi pa...	halo ... saya susah masuk ke akun saya lagi pa...
6	saya mencoba mendaftar menggunakan email tapi k...	saya mencoba mendaftar menggunakan email tapi k...
7	tambahin pencarian dong buat yg udah beli prem...	tambahin pencarian dong buat yg udah beli prem...
8	kenapa belum ada novel tere liye lengkap?	kenapa belum ada novel tere liye lengkap?
9	min, tolong bantu ya. saya sudah beli paket bu...	min, tolong bantu ya. saya sudah beli paket bu...

#### *Cleaning*

```
#cleaning
def clean_text(text):
    text = re.sub(r'http\S+', '', text) # hapus URL
    text = re.sub(r'^a-zA-Z\s', '', text) # hapus karakter non-alfabet
    return text

df['cleaned'] = df['lower'].apply(clean_text)
df[['lower', 'cleaned']].head(10)
```

	lower	cleaned
0	kenapa semua ebook yg udah didownload gak ke s...	kenapa semua ebook yg udah didownload gak ke s...
1	kurang update. buku yg aku cari semuanya gak ada	kurang update buku yg aku cari semuanya gak ada
2	tolonglah, kalau udah langganan premiumnya itu...	tolonglah kalau udah langganan premiumnya itu ...
3	bagus	bagus
4	ok	ok
5	halo ... saya susah masuk ke akun saya lagi pa...	halo saya susah masuk ke akun saya lagi padah...
6	saya mencoba mendaftar menggunakan email tapi k...	saya mencoba mendaftar menggunakan email tapi k...
7	tambahin pencarian dong buat yg udah beli prem...	tambahin pencarian dong buat yg udah beli prem...
8	kenapa belum ada novel tere liye lengkap?	kenapa belum ada novel tere liye lengkap
9	min, tolong bantu ya. saya sudah beli paket bu...	min tolong bantu ya saya sudah beli paket buku...

## Normalisasi

```

▶ normalisasi_kamus = {
    'gak': 'tidak', 'gak': 'tidak', 'ga': 'tidak', 'ngga': 'tidak', 'tdk': 'tidak', 'nga': 'tidak', 'bgt': 'banget', 'dlu': 'dulu', 'tp': 'tapi',
    'udh': 'sudah', 'sdh': 'sudah', 'blm': 'belum', 'sm': 'sama', 'aja': 'saja', 'menggunakan': 'menggunakan', 'yg': 'yang',
    'makinbanyak': 'makin banyak', 'sudab': 'sudah', 'udah': 'sudah', 'emailoke': 'email oke', 'sy': 'saya', 'nggak': 'tidak', 'ok': 'oke', 'sy
    'bukunyaudahmah': 'bukunya udah mah', 'ribed': 'ribet', 'ilanggg': 'hilang', 'jgn': 'jangan', 'jngn': 'jangan', 'eror': 'error',
    'didownload': 'di unduh', 'trs': 'terus', 'smoga': 'semoga', 'dl': 'dulu', 'ajaaa': 'saja', 'semuaaaaaaaaa': 'semua', 'gilaa': 'gila', 'kr
    'baikkkkkk': 'baik', 'bacs': 'baca', 'paswd': 'password', 'dr': 'dari', 'msk': 'masuk', 'jd': 'jadi', 'ttp': 'tetap', 'jg': 'juga', 'lbh
    'g': 'tidak', 'bs': 'bisa', 'knp': 'kenapa', 'bsa': 'bisa', 'ad': 'ada', 'aplikasi': 'aplikasi', 'apk': 'aplikasi', 'app': 'aplikasi',
    'sudan': 'sudah', 'aq': 'aku', 'q': 'aku', 'lainnn': 'lain', 'malh': 'malah', 'd': 'di', 'gmna': 'bagaimana', 'spa': 'apa', 'dg': 'denga
    'x': 'nya', 'pdhal': 'padahal', 'tgl': 'tanggal', 'krng': 'kurang', 'lbh': 'lebih', 'hrs': 'harus', 'drpda': 'daripada', 'lbih': 'lebih',
    'klo': 'kalau', 'tpi': 'tapi', 'udh': 'sudah', 'moga': 'semoga'
}

def normalize(text):
    return ' '.join([normalisasi_kamus.get(word, word) for word in text.split()])

df['normalized'] = df['cleaned'].apply(normalize)
df[['cleaned', 'normalized']].head()

```

	cleaned	normalized
0	kenapa semua ebook yg udah didownload gak ke s...	kenapa semua ebook yang sudah di unduh tidak k...
1	kurang update buku yg aku cari semuanya gak ada	kurang update buku yang aku cari semuanya tida...
2	tolonglah kalau udah langganan premiumnya itu ...	tolonglah kalau sudah langganan premiumnya itu...
3	bagus	bagus
4	ok	oke

## Tokenizing

```
#tokenizing

from nltk.tokenize import word_tokenize

df['tokens'] = df['normalized'].apply(lambda x: word_tokenize(x))
df[['normalized', 'tokens']].head(10)
```

	normalized	tokens
0	kenapa semua ebook yang sudah di unduh tidak k...	[kenapa, semua, ebook, yang, sudah, di, unduh,...
1	kurang update buku yang aku cari semuanya tida...	[kurang, update, buku, yang, aku, cari, semuan...
2	tolonglah kalau sudah langganan premiumnya itu...	[tolonglah, kalau, sudah, langganan, premiumny...
3	bagus	[bagus]
4	oke	[oke]
5	halo saya susah masuk ke akun saya lagi padaha...	[halo, saya, susah, masuk, ke, akun, saya, lag...
6	saya mencoba mendaftar menggunakan email tapi ...	[saya, mencoba, mendaftar, menggunakan, email,...
7	tambahin pencarian dong buat yang sudah beli p...	[tambahin, pencarian, dong, buat, yang, sudah,...
8	kenapa belum ada novel tere liye lengkap	[kenapa, belum, ada, novel, tere, liye, lengkap]
9	min tolong bantu ya saya sudah beli paket buku...	[min, tolong, bantu, ya, saya, sudah, beli, pa...

## Stopword Removal

```
#Stopword Removal
# Buat factory stopwords
factory = StopWordRemoverFactory()
stop_words = set(factory.get_stop_words())

# Tambahkan stopwords buatan sendiri jika perlu
stopwords_tambahan = ['halo', 'nya', 'min', 'nih', 'sangat', 'lah']
stop_words.update(stopwords_tambahan)

# Hapus stopwords dari token
df['no_stopword'] = df['tokens'].apply(lambda x: [word for word in x if word not in stop_words])
df[['tokens', 'no_stopword']].head(10)
```

...	tokens	no_stopword
0	[kenapa, semua, ebook, yang, sudah, di, unduh,...	[semua, ebook, unduh, save, perpustakaan]
1	[kurang, update, buku, yang, aku, cari, semuan...	[kurang, update, buku, aku, cari, semuanya]
2	[tolonglah, kalau, sudah, langganan, premiumny...	[tolonglah, kalau, langganan, premiumnya, baca...
3	[bagus]	[bagus]
4	[oke]	[oke]
5	[halo, saya, susah, masuk, ke, akun, saya, lag...	[susah, masuk, akun, padahal, reset, password]
6	[saya, mencoba, mendaftar, menggunakan, email,...	[mencoba, mendaftar, menggunakan, email, katan...
7	[tambahin, pencarian, dong, buat, yang, sudah,...	[tambahin, pencarian, dong, buat, beli, premiu...
8	[kenapa, belum, ada, novel, tere, liye, lengkap]	[novel, tere, liye, lengkap]

## Stemming

```
#stemming / ubah kata ke bentuk dasar
stemmer = StemmerFactory().create_stemmer()
df['stemmed'] = df['no_stopword'].apply(lambda x: [stemmer.stem(word) for word in x])
df[['no_stopword', 'stemmed']].head(10)
```

...	no_stopword	stemmed
0	[semua, ebook, unduh, save, perpustakaan]	[semua, ebook, unduh, save, pustaka]
1	[kurang, update, buku, aku, cari, semuanya]	[kurang, update, buku, aku, cari, semua]
2	[tolonglah, kalau, langganan, premiumnya, baca...	[tolong, kalau, langgan, premium, baca, buku, ...
3	[bagus]	[bagus]
4	[oke]	[oke]
5	[susah, masuk, akun, padahal, reset, pasword]	[susah, masuk, akun, padahal, reset, pasword]
6	[mencoba, mendaftar, menggunakan, email, katan...	[coba, daftar, guna, email, kata, suruh, verif...
7	[tambahin, pencarian, dong, buat, beli, premiu...	[tambahin, cari, dong, buat, beli, premium, ta...
8	[novel, tere, liye, lengkap]	[novel, tere, liye, lengkap]
9	[bantu, beli, paket, buku, bulanan, terkendala...	[bantu, beli, paket, buku, bulan, kendala, buk...

## TF-IDF

```
#TF-IDF Vectorization
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['final_cleaned'])
y = df['label']
```

```
temp = pd.DataFrame({
    "Dokumen": doc_names[i],
    "Kata": feature_names[nz_idx],
    "Score pembobotan": row[nz_idx]
})

# urutkan dari skor terbesar
temp = temp.sort_values("Score pembobotan", ascending=False)

frames.append(temp)

tfidf_score_table = pd.concat(frames, ignore_index=True)

# rapikan tampilan angka (opsional)
tfidf_score_table["Score pembobotan"] = tfidf_score_table["Score pembobotan"].round(4)

tfidf_score_table.head(50)
```

	Dokumen	Kata	Score pembobotan
0	Dokumen 1 (negative)	makin	0.5088
1	Dokumen 1 (negative)	update	0.4611
2	Dokumen 1 (negative)	moga	0.4432
3	Dokumen 1 (negative)	ebook	0.4143
4	Dokumen 1 (negative)	banyak	0.4006

## TRAINING DAN TESTING

```
▶ # Split data ke 3 skenario
scenarios = {"70:30": 0.3, "80:20": 0.2, "90:10": 0.1}
results = {}

for name, test_size in scenarios.items():
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=test_size, stratify=y, random_state=42
    )

    model = SVC(kernel='linear')
    model.fit(X_train, y_train)

    y_pred = model.predict(X_test)
    acc = accuracy_score(y_test, y_pred)

    print(f"=== Skenario {name} (test_size={test_size}) ===")
    print(classification_report(y_test, y_pred, target_names=['negatif', 'positif']))
    print(f"Akurasi: {acc:.4f}")
    print()

    results[name] = {
        "test_size": test_size,
        "model": model,
        "y_train": y_train,
        "y_test": y_test,
        "y_pred": y_pred,
        "cm": confusion_matrix(y_test, y_pred),
        "accuracy": acc,
    }
```

```

... === Skenario 70:30 (test_size=0.3) ===
           precision    recall  f1-score   support

   negatif      0.80      0.93      0.86      171
   positif      0.85      0.63      0.73      111

   accuracy                    0.81      282
   macro avg      0.82      0.78      0.79      282
   weighted avg   0.82      0.81      0.81      282

Akurasi: 0.8121

```

```

=== Skenario 80:20 (test_size=0.2) ===
           precision    recall  f1-score   support

   negatif      0.78      0.93      0.85      114
   positif      0.85      0.59      0.70      74

   accuracy                    0.80      188
   macro avg      0.81      0.76      0.77      188
   weighted avg   0.81      0.80      0.79      188

Akurasi: 0.7979

```

```

=== Skenario 90:10 (test_size=0.1) ===
           precision    recall  f1-score   support

   negatif      0.78      0.89      0.84      57
   positif      0.79      0.62      0.70      37

   accuracy                    0.79      94
   macro avg      0.79      0.76      0.77      94
   weighted avg   0.79      0.79      0.78      94

Akurasi: 0.7872

```







Lampiran 2 Sampel *Dataset* Ulasan

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
Kenapa semua ebook yg udah didownload gak ke save di perpustakaan ya?	Positif	kenapa semua ebook yg udah didownload gak ke save di perpustakaan ya?	kenapa semua ebook yg udah didownload gak ke save di perpustakaan ya	kenapa semua ebook yang sudah diunduh tidak ke save di perpustakaan ya	['kenapa', 'semua', 'ebook', 'yang', 'sudah', 'di', 'unduh', 'tidak', 'ke', 'save', 'di', 'perpustakaan', 'ya']	['semua', 'ebook', 'unduh', 'save', 'perpustakaan']	['semua', 'ebook', 'unduh', 'save', 'pustaka']	semua ebook unduh save pustaka
kurang update. buku yg aku cari semuanya gak ada	Negatif	kurang update. buku yg aku cari semuanya gak ada	kurang update buku yg aku cari semuanya gak ada	kurang update buku yang aku cari semuanya tidak ada	['kurang', 'update', 'buku', 'yang', 'aku', 'cari', 'semuanya', 'tidak', 'ada']	['kurang', 'update', 'buku', 'aku', 'cari', 'semuanya']	['kurang', 'update', 'buku', 'aku', 'cari', 'semua']	kurang update buku aku cari semuanya
tolonglah, kalau udah langganan premiumnya itu bisa baca buku yang lebih banyak, banyak buku yang kucari ternyata harus beli satuan, terus ngapain	Negatif	tolonglah, kalau udah langganan premiumnya itu bisa baca buku yang lebih banyak, banyak buku yang kucari ternyata harus beli satuan, terus ngapain	tolonglah kalau udah langganan premiumnya itu bisa baca buku yang lebih banyak buku yang kucari ternyata harus beli satuan ngapain	tolonglah kalau sudah langganan premiumnya itu bisa baca buku yang lebih banyak buku yang kucari ternyata harus beli satuan ngapain	['tolonglah', 'kalau', 'sudah', 'langganan', 'premiumnya', 'itu', 'bisa', 'baca', 'buku', 'yang', 'lebih', 'banyak', 'buku', 'yang', 'kucari', 'ternyata', 'harus', 'beli', 'satuan', 'terus', 'ngapain']	['tolonglah', 'kalau', 'langganan', 'premiumnya', 'baca', 'buku', 'banyak', 'kucari', 'beli', 'satuan', 'ngapain']	['tolonglah', 'kalau', 'langganan', 'premiumnya', 'baca', 'buku', 'banyak', 'cari', 'nyata', 'beli', 'satu', 'terus', 'ngapain']	tolonglah kalau langganan premiumnya itu bisa baca buku yang lebih banyak buku yang kucari ternyata harus beli satuan ngapain

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
dah kalau ujungnya banyak yang harus beli satuan. dah gitu harga bulanan kaya beli paket kuota Telkom sel lagi		dah kalau ujungnya banyak yang harus beli satuan. dah gitu harga bulanan kaya beli paket kuota telkomsel lagi	dah kalau ujungnya banyak yang harus beli satuan. dah gitu harga bulanan kaya beli paket kuota telkomsel lagi	dah kalau ujungnya banyak yang harus beli satuan. dah gitu harga bulanan kaya beli paket kuota telkomsel lagi	'dah', 'kalau', 'ujungnya', 'banyak', 'yang', 'harus', 'beli', 'satuan', 'dah', 'gitu', 'harga', 'bulanan', 'kaya', 'beli', 'paket', 'kuota', 'telkomsel', 'lagi']	ya', 'banyak', 'beli', 'satuan', 'dah', 'gitu', 'harga', 'bulanan', 'kaya', 'beli', 'paket', 'kuota', 'telkom sel']	'kalau', 'ujung', 'banyak', 'beli', 'satu', 'dah', 'gitu', 'harga', 'bulan', 'kaya', 'beli', 'paket', 'kuota', 'telkom sel']	kalau ujung banyak beli satu dah gitu harga bulan kaya beli paket kuota telkomsel
bagus	Positif	bagus	bagus	bagus	['bagus']	['bagus']	['bagus']	bagus
ok	Positif	ok	ok	oke	['oke']	['oke']	['oke']	oke
halo ... saya susah masuk ke akun saya lagi padahal sudab reset password	Negatif	halo ... saya susah masuk ke akun saya lagi padahal sudab reset password	halo saya susah masuk ke akun saya lagi padahal sudab reset password	halo saya susah masuk ke akun saya lagi padahal sudah reset password	['halo', 'saya', 'susah', 'masuk', 'ke', 'akun', 'saya', 'lagi', 'padahal', 'sudah', 'reset', 'password']	['susah', 'masuk', 'akun', 'padahal', 'reset', 'password']	['susah', 'masuk', 'akun', 'padahal', 'reset', 'password']	susah masuk akun padahal reset password
saya mencoba mendaftar menggunakan email tapi kenapa bisa katanya disuruh	Negatif	saya mencoba mendaftar menggunakan email tapi kenapa bisa katanya disuruh	saya mencoba mendaftar menggunakan email tapi kenapa bisa katanya disuruh	saya mencoba mendaftar menggunakan email tapi kenapa bisa katanya	['saya', 'mencoba', 'mendaftar', 'menggunakan', 'email', 'tapi', 'kenapa', 'tidak', 'bisa', 'katanya', 'disuruh',	['mencoba', 'mendaftar', 'menggunakan', 'email', 'katanya', 'disuruh',	['coba', 'daftar', 'guna', 'email', 'kata', 'suruh', 'verifikasi', 'email', 'oke', 'hasil', 'verifikasi',	coba daftar guna email kata suruh verifikasi email oke hasil verifikasi

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
verifikasi email.ok e udh berhasil terverifikasi tapi kenapa ngga bisa masuk juga padahal sinyal oke dan selalu error bikin kesal perbaiki atau tidak sama sekali tergantung pada kemauan anda		verifikasi email.ok e udh berhasil terverifikasi tapi kenapa ngga bisa masuk juga padahal sinyal oke dan selalu error bikin kesal perbaiki atau tidak sama sekali tergantung pada kemauan anda	verifikasi emailok e udh berhasil terverifikasi tapi kenapa ngga bisa masuk juga padahal sinyal oke dan selalu error bikin kesal perbaiki atau tidak sama sekali tergantung pada kemauan anda	disuruh verifikasi email oke sudah berhasil terverifikasi tapi kenapa tidak bisa masuk juga padahal sinyal oke dan selalu error bikin kesal perbaiki atau tidak sama sekali tergantung pada kemauan anda	'verifikasi', 'email', 'oke', 'sudah', 'berhasil', 'terverifikasi', 'tapi', 'kenapa', 'tidak', 'bisa', 'masuk', 'juga', 'padahal', 'sinyal', 'oke', 'dan', 'selalu', 'error', 'bikin', 'kesal', 'perbaiki', 'atau', 'tidak', 'sama', 'sekali', 'tergantung', 'pada', 'kemauan', 'anda']	'email', 'oke', 'berhasil', 'terverifikasi', 'masuk', 'padahal', 'sinyal', 'oke', 'selalu', 'error', 'bikin', 'kesal', 'baik', 'sama', 'sekali', 'gantungan', 'mau']	'masuk', 'padahal', 'sinyal', 'oke', 'selalu', 'error', 'bikin', 'kesal', 'baik', 'sama', 'sekali', 'gantungan', 'mau']	masuk padahal sinyal oke selalu error bikin kesal baik sama sekali gantungan mau
tambahkan pencariannya dong buat yg udah beli premium atau gak tambahkan fitur filter biar gak usah scroll trs sampe bawah.	Positif	tambahkan pencariannya dong buat yg udah beli premium atau gak tambahkan fitur filter biar gak usah scroll trs sampe bawah.	tambahkan pencariannya dong buat yg udah beli premium atau gak tambahkan fitur filter biar gak usah scroll trs sampe bawah.	tambahkan pencariannya dong buat yang sudah beli premium atau tidak tambahkan fitur filter biar tidak usah scroll	['tambahkan', 'pencarian', 'dong', 'buat', 'yang', 'sudah', 'beli', 'premium', 'atau', 'tidak', 'tambahkan', 'fitur', 'filter', 'biar', 'tidak', 'usah', 'scroll',	['tambahkan', 'pencarian', 'dong', 'buat', 'beli', 'premium', 'tambahkan', 'fitur', 'filter', 'biar', 'usah', 'scroll', 'sampe',	['tambahkan', 'cari', 'dong', 'buat', 'beli', 'premium', 'tambahkan', 'fitur', 'filter', 'biar', 'usah', 'scroll', 'sampe',	tambahkan cari dong buat beli premium tambahkan fitur filter biar usah scroll terus sampe

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
mau istirahat dulu sebentar eh pas balik mau baca lagi masa iyah harus scroll trs		mau istirahat dulu sebentar eh pas balik mau baca lagi masa iyah harus scroll trs	mau istirahat dulu sebentar eh pas balik mau baca lagi masa iyah harus scroll trs	terus sampe bawah mau istirahat dulu sebentar eh pas balik mau baca lagi masa iyah harus scroll terus	'terus', 'sampe', 'bawah', 'mau', 'istirahat', 'dulu', 'seben tar', 'eh', 'pas', 'balik', 'mau', 'baca', 'lagi', 'ayah', 'harus', 'scroll', 'terus']	'bawah', 'mau', 'istirahat', 'dulu', 'seben tar', 'eh', 'pas', 'balik', 'mau', 'baca', 'masa', 'iyah', 'scroll', 'terus']	'bawah', 'mau', 'istirahat', 'dulu', 'seben tar', 'eh', 'pas', 'balik', 'mau', 'baca', 'masa', 'iyah', 'scroll', 'terus']	bawah mau istirahat dulu sebentar eh pas balik mau baca masa iyah scroll terus
kenapa belum ada novel tere liye lengkap ?	Negatif	kenapa belum ada novel tere liye lengkap ?	kenapa belum ada novel tere liye lengkap	kenapa belum ada novel tere liye lengkap	['kenapa', 'belum', 'ada', 'novel', 'tere', 'liye', 'lengkap']	['novel', 'tere', 'liye', 'lengkap']	['novel', 'tere', 'liye', 'lengkap']	novel tere liye lengkap
min, tolong bantu ya. saya sudah beli paket buku bulanan tapi terkendala karena bukunya tidak bisa diunduh . padahal jaringan internet saya stabil.	Positif	min, tolong bantu ya. saya sudah beli paket buku bulanan tapi terkendala karena bukunya tidak bisa diunduh . padahal jaringan internet saya stabil.	min tolong bantu ya saya sudah beli paket buku bulanan tapi terkendala karena bukunya tidak bisa diunduh padahal jaringan internet saya stabil kalau	min tolong bantu ya saya sudah beli paket buku bulanan tapi terkendala karena bukunya tidak bisa diunduh padahal jaringan internet saya stabil kalau	['min', 'tolong', 'bantu', 'ya', 'saya', 'sudah', 'beli', 'paket', 'buku', 'bulanan', 'tapi', 'terkendala', 'karena', 'bukunya', 'tidak', 'bisa', 'diunduh', 'padahal', 'jaringan', 'internet', 'saya', 'stabil', 'kalau',	['bantu', 'beli', 'paket', 'buku', 'bulanan', 'terkendala', 'buku', 'bulanan', 'diunduh', 'padahal', 'jaringan', 'internet', 'saya', 'stabil', 'g',	['bantu', 'beli', 'paket', 'buku', 'bulan', 'kendala', 'buku', 'unduh', 'padahal', 'jaringan', 'internet', 'stabil', 'kalau', 'trouble shooting', 'ratingnya',	bantu beli paket buku bulan kendala buku unduh padahal jaringan internet stabil kalau trouble shooting ratingnya

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
kalau sudah di troubles hooting nanti akan saya naikkan ratingnya. terima kasih. note, April 2025, sudah kembali normal.		kalau sudah di troubles hooting nanti akan saya naikkan ratingnya. terima kasih. note, april 2025, sudah kembali normal.	sudah di troubles hooting nanti akan saya naikkan ratingnya terima kasih note april sudah kembali normal	sudah di troubles hooting nanti akan saya naikkan ratingnya terima kasih note april sudah kembali normal	'sudah', 'di', 'troubleshooting', 'nanti', 'akan', 'saya', 'naikkan', 'ratingnya', 'terima', 'kasih', 'note', 'april', 'sudah', 'kembali', 'normal']	'naikkan', 'ratingnya', 'terima', 'kasih', 'note', 'april', 'normal']	'terima', 'kasih', 'note', 'april', 'normal']	terima kasih note april normal
aplikasinya sangat bagus   	Positif	aplikasinya sangat bagus   	aplikasinya sangat bagus	aplikasinya sangat bagus	['aplikasinya', 'sangat', 'bagus']	['aplikasinya', 'bagus']	['aplikasi', 'bagus']	aplikasi bagus
semoga makin banyak ebook yg update lagi	Positif	semoga makin banyak ebook yg update lagi	semoga makin banyak ebook yg update lagi	semoga makin banyak ebook yang update lagi	['semoga', 'makin', 'banyak', 'ebook', 'yang', 'update', 'lagi']	['semoga', 'makin', 'banyak', 'ebook', 'update']	['moga', 'makin', 'banyak', 'ebook', 'update']	moga makin banyak ebook update
perkembangan yang luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di	Positif	perkembangan yang luar biasa, readernya susah semakin baik. meski ada beberapa yang perlu di	perkembangan yang luar biasa readernya susah semakin baik meski ada beberapa yang perlu di	perkembangan yang luar biasa readernya susah semakin baik meski ada beberapa yang perlu di	['perkembangannya', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'ada', 'beberapa', 'yang', 'perlu', 'di']	['perkembangan', 'luar', 'biasa', 'readernya', 'susah', 'semakin', 'baik', 'meski', 'ada', 'beberapa', 'yang', 'perlu']	['kembangan', 'luar', 'biasa', 'readernya', 'susah', 'makin', 'baik', 'meski', 'beberapa', 'perlu']	kembangkan yang luar biasa readernya susah makin baik meski beberapa

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
tingkatkan tapi gramedia digital sekarang sudah sangat oke		tingkatkan tapi gramedia digital sekarang sudah sangat oke	tingkatkan tapi gramedia digital sekarang sudah sangat oke	tingkatkan tapi gramedia digital sekarang sudah sangat oke	'tingkatkan', 'tapi', 'gramedia', 'digital', 'sekarang', 'sudah', 'sangat', 'oke']	a', 'perlu', 'tingkatkan', 'gramedia', 'digital', 'sekarang', 'oke']	'tingkat', 'gramedia', 'digital', 'sekarang', 'oke']	perlu tingkat gramedia digital sekarang oke
saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook, sangat kecewa!	Negatif	saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook, sangat kecewa!	saya baru update aplikasi dan semua bacaan yang pernah sy download hilang gak bersisa padahal ada ratusan ebook, sangat kecewa	saya baru update aplikasi dan semua bacaan yang pernah saya download hilang tidak bersisa padahal ada ratusan ebook sangat kecewa	['saya', 'baru', 'update', 'aplikasi', 'dan', 'semua', 'bacaan', 'yang', 'pernah', 'saya', 'download', 'hilang', 'tidak', 'bersisa', 'padahal', 'ada', 'ratusan', 'ebook', 'sangat', 'kecewa']	['baru', 'update', 'aplikasi', 'semua', 'bacaan', 'pernah', 'download', 'hilang', 'bersisa', 'padahal', 'ratusan', 'ebook', 'kecewa']	['baru', 'update', 'aplikasi', 'semua', 'baca', 'pernah', 'download', 'hilang', 'sisa', 'padahal', 'ratus', 'ebook', 'kecewa']	baru update aplikasi semua baca pernah download hilang sisa padahal ratusan ebook kecewa
lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua??	Negatif	lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua??	lohh setelah diupdate koleksi yg sudah saya unduh kok hilang semua	lohh setelah diupdate koleksi yang sudah saya unduh kok hilang semua	['lohh', 'setelah', 'diupdate', 'koleksi', 'yang', 'sudah', 'saya', 'unduh', 'kok', 'hilang', 'semua']	['lohh', 'diupdate', 'koleksi', 'unduh', 'kok', 'hilang', 'semua']	['lohh', 'diupdate', 'koleksi', 'unduh', 'kok', 'hilang', 'semua']	lohh diupdate koleksi unduh kok hilang semua
Senang sekali bisa baca banyak	Positif	senang sekali bisa baca banyak	senang sekali bisa baca banyak	senang sekali bisa baca banyak	['senang', 'sekali', 'bisa', 'baca', 'banyak',	['senang', 'sekali', 'baca', 'banyak'	['senang', 'sekali', 'baca', 'banya	senang sekali baca banyak







content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
buku dalam satu aplikasi.		buku dalam satu aplikasi.	buku dalam satu aplikasi	buku dalam satu aplikasi	'buku', 'dalam', 'satu', 'aplikasi']	, 'buku', 'satu', 'aplikasi']	k', 'buku', 'satu', 'aplikasi']	k buku satu aplikasi
gimana indonesia jadi negara maju, orang yg minat baca juga harus bayar dulu buat baca bukunya, udah mah literasi rendah dipersulit Juga	Negatif	gimana indonesia jadi negara maju, orang yg minat baca juga harus bayar dulu buat baca bukunya, udah mah literasi rendah dipersulit juga	gimana indonesia jadi negara maju orang yg minat baca juga harus bayar dulu buat baca bukunya udah mah literasi rendah dipersulit juga	gimana indonesia jadi negara maju orang yang minat baca juga harus bayar dulu buat baca bukunya udah mah literasi rendah dipersulit juga	['gimana', 'indonesia', 'jadi', 'negara', 'maju', 'orang', 'yang', 'minat', 'baca', 'juga', 'harus', 'bayar', 'dulu', 'buat', 'baca', 'bukunya', 'udah', 'mah', 'literasi', 'rendah', 'dipersulit', 'juga']	['gimana', 'indonesia', 'jadi', 'negara', 'maju', 'orang', 'minat', 'baca', 'bayar', 'dulu', 'buat', 'baca', 'bukunya', 'udah', 'mah', 'literasi', 'rendah', 'dipersulit']	['gimana', 'indonesia', 'jadi', 'negara', 'maju', 'orang', 'minat', 'baca', 'bayar', 'dulu', 'buat', 'baca', 'buku', 'udah', 'mah', 'literasi', 'rendah', 'sulit']	gimana indonesia jadi negara maju orang minat baca bayar dulu buat baca bukunya udah mah literasi rendah dipersulit
Tidak puas sama sekali, parah. Sulit sekali aksesnya padahal baru mulai beli 3 buku sudah mengalami kesulitan akses.	Negatif	tidak puas sama sekali, parah. sulit sekali aksesnya padahal baru mulai beli 3 buku sudah mengalami kesulitan akses.	tidak puas sama sekali parah sulit sekali aksesnya padahal baru mulai beli buku sudah mengalami kesulitan akses	tidak puas sama sekali parah sulit sekali aksesnya padahal baru mulai beli buku sudah mengalami kesulitan akses	['tidak', 'puas', 'sama', 'sekali', 'parah', 'sulit', 'sekali', 'akses', 'nya', 'padahal', 'baru', 'mulai', 'beli', 'buku', 'sudah', 'mengalami', 'kesulitan', 'akses',	['puas', 'sama', 'sekali', 'parah', 'sulit', 'sekali', 'akses', 'padahal', 'baru', 'mulai', 'beli', 'buku', 'mengalami', 'kesulitan', 'akses', 'akhir', 'buang',	['puas', 'sama', 'sekali', 'parah', 'sulit', 'sekali', 'akses', 'padahal', 'baru', 'mulai', 'beli', 'buku', 'alami', 'sulit', 'akses', 'akhir', 'buang', 'duit',	puas sama sekali parah sulit sekali akses padahal baru mulai beli buku alami sulit akses akhir buang duit

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
Dan akhirnya buang2 duit saja ini. Akses parah, buang2 duit dan waktu saja. Padahal baru coba mulai beli 3 buku. Sudah kapok.		dan akhirnya buang2 duit saja ini. akses parah, buang2 duit dan waktu saja. padahal baru coba mulai beli 3 buku. sudah kapok.	dan akhirnya buang duit saja ini akses parah buang duit dan waktu saja padahal baru coba mulai beli buku sudah kapok	dan akhirnya buang duit saja ini akses parah buang duit dan waktu saja padahal baru coba mulai beli buku sudah kapok	'dan', 'akhir', 'nya', 'buang', 'duit', 'saja', 'ini', 'akses', 'parah', 'buang', 'duit', 'dan', 'waktu', 'saja', 'padahal', 'baru', 'coba', 'mulai', 'beli', 'buku', 'sudah', 'kapok']	'duit', 'akses', 'parah', 'buang', 'duit', 'waktu', 'padahal', 'baru', 'coba', 'mulai', 'beli', 'buku', 'kapok']	'akses', 'parah', 'buang', 'duit', 'waktu', 'padahal', 'baru', 'coba', 'mulai', 'beli', 'buku', 'kapok']	akses parah buang duit waktu padahal baru coba mulai beli buku kapok
enak kagumi saya	Positif	enak kagumi saya	enak kagumi saya	enak kagumi saya	['enak', 'kagumi', 'saya']	['enak', 'kagumi']	['enak', 'kagum']	enak kagum
sangat bagus untuk membaca jadi lebih mudah dan praktis	Positif	sangat bagus untuk membaca jadi lebih mudah dan praktis	sangat bagus untuk membaca jadi lebih mudah dan praktis	sangat bagus untuk membaca jadi lebih mudah dan praktis	['sangat', 'bagus', 'untuk', 'membaca', 'jadi', 'lebih', 'mudah', 'dan', 'praktis']	['bagus', 'membaca', 'jadi', 'lebih', 'mudah', 'praktis']	['bagus', 'baca', 'jadi', 'lebih', 'mudah', 'praktis']	bagus baca jadi lebih mudah praktis
saya pakai paket setahun, tapi tunggu promo dari gramedia di akhir tahun bisa dapat	Positif	saya pakai paket setahun, tapi tunggu promo dari gramedia di akhir tahun bisa dapat	saya pakai paket setahun tapi tunggu promo dari gramedia di akhir tahun bisa dapat	saya pakai paket setahun tapi tunggu promo dari gramedia di akhir tahun bisa dapat	['saya', 'pakai', 'paket', 'setahun', 'tapi', 'tunggu', 'promo', 'dari', 'gramedia', 'di', 'akhir', 'tahun', 'bisa', 'dapat', 'harga',	['pakai', 'paket', 'setahun', 'tunggu', 'promo', 'gramedia', 'akhir', 'tahun', 'harga', 'semoga	['pakai', 'paket', 'tahun', 'tunggu', 'promo', 'gramedia', 'akhir', 'tahun', 'harga', 'an',	pakai paket tahun tunggu promo gramedia akhir tahun harga an moga

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
harga 700 an ... semoga ada terus promonya		harga 700 an ... semoga ada terus promonya	harga an semoga ada terus promonya	harga an semoga ada terus promonya	'an', 'semoga', 'ada', 'terus', 'promonya']	', 'terus', 'promonya']	'moga', 'terus', 'promonya']	terus promonya
mantap	Positif	mantap	mantap	mantap	['mantap']	['mantap']	['mantap']	mantap
akun gak bisa masuk, trus minta diverifikasi tp gak ada notif di gmail, tanya CS tp no respon, padahal dulu sdh beli banyak buku	Negatif	akun gak bisa masuk, trus minta diverifikasi tp gak ada notif di gmail, tanya cs tp no respon, padahal dulu sdh beli banyak buku	akun gak bisa masuk trus minta diverifikasi tp gak ada notif di gmail tanya cs tp no respon padahal dulu sdh beli banyak buku	akun tidak bisa masuk trus minta diverifikasi tapi tidak ada notif di gmail tanya cs tapi no respon padahal dulu sudah beli banyak buku	['akun', 'tidak', 'bisa', 'masuk', 'trus', 'minta', 'diverifikasi', 'tapi', 'tidak', 'ada', 'notif', 'di', 'gmail', 'tanya', 'cs', 'tapi', 'no', 'respon', 'padahal', 'dulu', 'sudah', 'beli', 'banyak', 'buku']	['akun', 'masuk', 'trus', 'minta', 'diverifikasi', 'notif', 'gmail', 'tanya', 'cs', 'no', 'respon', 'padahal', 'dulu', 'beli', 'banyak', 'buku']	['akun', 'masuk', 'trus', 'minta', 'verifikasi', 'notif', 'gmail', 'tanya', 'cs', 'no', 'respon', 'padahal', 'dulu', 'beli', 'banyak', 'buku']	akun masuk trus minta verifikasi notif gmail tanya cs no respon padahal dulu beli banyak buku
enak baca bukunya	Positif	enak baca bukunya	enak baca bukunya	enak baca bukunya	['enak', 'baca', 'bukunya']	['enak', 'baca', 'bukunya']	['enak', 'baca', 'bukunya']	enak baca buku
kurang memuaskan Krn hak pembeli dibatasi sehingga TDK bisa mencerdaskan	Negatif	kurang memuaskan krn hak pembeli dibatasi sehingga tdk bisa mencerdaskan	kurang memuaskan krn hak pembeli dibatasi sehingga tdk bisa mencerdaskan	kurang memuaskan karena hak pembeli dibatasi sehingga tidak bisa mencerdaskan	['kurang', 'memuaskan', 'karena', 'hak', 'pembeli', 'dibatasi', 'sehingga', 'tidak', 'bisa', 'mencerdaskan']	['kurang', 'memuaskan', 'hak', 'pembeli', 'dibatasi', 'sehingga', 'mencerdaskan']	['kurang', 'muas', 'hak', 'beli', 'batas', 'cerdas', 'bangsa', 'sejati']	kurang memuaskan hak beli batas cerdas bangsa sejati

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
bangsa yg sejatinya ketika sudah membeli buku adalah milik konsumen sepenuhnya TDK perlu hrs baca lewat media anda...k apok saya beli online		bangsa yg sejatinya ketika sudah membeli buku adalah milik konsumen sepenuhnya tdk perlu hrs baca lewat media anda...k apok saya beli online	bangsa yg sejatinya ketika sudah membeli buku adalah milik konsumen sepenuhnya tdk perlu hrs baca lewat media andakap ok saya beli online	askan bangsa yang sejatinya ketika sudah membeli buku adalah milik konsumen sepenuhnya tidak perlu hrs baca lewat media andakap ok saya beli online	skan', 'bangsa', 'yang', 'sejatinya', 'ketika', 'sudah', 'membeli', 'buku', 'adalah', 'milik', 'konsumen', 'sepenuhnya', 'sepenuhnya', 'tidak', 'perlu', 'harus', 'baca', 'lewat', 'media', 'andakapok', 'saya', 'beli', 'online']	'bangsa', ', 'sejatinya', 'membeli', 'buku', 'milik', 'konsumen', 'sepenuhnya', 'perlu', 'baca', 'lewat', 'media', 'andakapok', 'beli', 'online']	'beli', 'buku', 'milik', 'konsumen', 'sepenuh', 'perlu', 'baca', 'lewat', 'media', ', 'andakapok', 'beli', 'online']	beli buku milik konsumen sepenuhnya perlu baca lewat media andakapok beli online
easily read ..without must visit at the store	Positif	easily read ..without must visit at the store	easily read without must visit at the store	easily read without must visit at the store	['easily', 'read', 'without', 'must', 'visit', 'at', 'the', 'store']	['easily', 'read', 'without', 'must', 'visit', 'at', 'the', 'store']	['easily', 'read', 'without', 'must', 'visit', 'at', 'the', 'store']	easily read without must visit at the store
aplikasi yang sangat membantu untuk menemukan buku kesukaan Trims	Positif	aplikasi yang sangat membantu untuk menemukan buku kesukaan trims	aplikasi yang sangat membantu untuk menemukan buku kesukaan trims	aplikasi yang sangat membantu untuk menemukan buku kesukaan trims	['aplikasi', 'yang', 'sangat', 'membantu', 'untuk', 'menemukan', 'buku', 'kesukaan', 'trims']	['aplikasi', 'membantu', 'menemukan', 'buku', 'kesukaan', 'trims']	['aplikasi', 'bantu', 'temu', 'buku', 'suka', 'trims']	aplikasi bantu temu buku suka trims
good	Positif	good	good	good	['good']	['good']	['good']	good
sering logout	Negatif	sering logout	sering logout	sering logout	['sering', 'logout']	['sering', ', 'sering']	['sering', ', 'sering']	sering

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
otomatis , risih banget apalagi kalo lagi asik2nya baca		otomatis , risih banget apalagi kalo lagi asik2nya baca	otomatis risih banget apalagi kalo lagi asiknya baca	otomatis risih banget apalagi kalo lagi asiknya baca	'otomatis', 'risih', 'banget', 'apalagi', 'kalo', 'lagi', 'asiknya', 'baca']	'logout', 'otomatis', 'risih', 'banget', 'kalo', 'asiknya', 'baca']	'logout', 'otomatis', 'risih', 'banget', 'kalo', 'asiknya', 'baca']	logout otomatis risih banget kalo asiknya baca
mantap aplikasi	Positif	mantap aplikasi	mantap aplikasi	mantap aplikasi	['mantap', 'aplikasi']	['manta p', 'aplikasi']	['manta p', 'aplikasi']	mantap aplikasi
Mohon diperbaiki. Akun sering logout jadi lumayan mengganggu sekali	Negatif	mohon diperbaiki. akun sering logout jadi lumayan mengganggu sekali	mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali	mohon diperbaiki akun sering logout jadi lumayan mengganggu sekali	['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']	['mohon', 'diperbaiki', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'mengganggu', 'sekali']	['mohon', 'baik', 'akun', 'sering', 'logout', 'jadi', 'lumayan', 'ganggu', 'sekali']	mohon baik akun sering logout jadi lumayan ganggu sekali
cuma bisa dipakai dua device dan sering ke logout sendiri	Negatif	cuma bisa dipakai dua device dan sering ke logout sendiri	cuma bisa dipakai dua device dan sering ke logout sendiri	cuma bisa dipakai dua device dan sering ke logout sendiri	['cuma', 'bisa', 'dipakai', 'dua', 'device', 'dan', 'sering', 'ke', 'log', 'out', 'sendiri']	['cuma', 'dipakai', 'device', 'sering', 'log', 'out', 'sendiri']	['cuma', 'pakai', 'device', 'sering', 'log', 'out', 'sendiri']	cuma pakai device sering logout sendiri
Tambahkan fitur kutipan dan highlight tulisan untuk yg format pdf	Positif	tambahkan fitur kutipan dan highlight tulisan untuk yg format pdf	tambahkan fitur kutipan dan highlight tulisan untuk yg format pdf	tambahkan fitur kutipan dan highlight tulisan untuk yg format pdf	['tambahkan', 'fitur', 'kutipan', 'dan', 'highlight', 'tulisan', 'untuk', 'yang', 'format', 'pdf']	['tambahkan', 'fitur', 'kutipan', 'highlight', 'tulisan', 'format', 'pdf']	['tambahkan', 'fitur', 'kutipan', 'highlight', 'tulisan', 'format', 'pdf']	tambahkan fitur kutipan highlight tulisan format pdf

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
so insightful and meaningful for me as a college student   	Positif	so insightful and meaningful for me as a college student   	so insightful and meaningful for me as a college student	so insightful and meaningful for me as a college student	['so', 'insightful', 'and', 'meaningful', 'for', 'me', 'as', 'a', 'college', 'student']	['so', 'insightful', 'and', 'meaningful', 'for', 'me', 'as', 'a', 'college', 'student']	['so', 'insightful', 'and', 'meaningful', 'for', 'me', 'as', 'a', 'college', 'student']	so insightful and meaningful for me as a college student
Ribed, gk bisa baca lewat pc. Harus unduh di hp dl. Padahal di hp tulisannya kecil2, susah bacanya . Nyesel banget udah beli buku disini.	Negatif	ribed, gk bisa baca lewat pc. harus unduh di hp dl. padahal di hp tulisannya kecil2, susah bacanya . nyesel banget udah beli buku disini.	ribed gk bisa baca lewat pc harus unduh di hp dl padahal di hp tulisannya kecil susah bacanya nyesel banget udah beli buku disini	ribet tidak bisa baca lewat pc harus unduh di hp dulu padahal di hp tulisannya kecil susah bacanya nyesel banget sudah beli buku disini	['ribet', 'tidak', 'bisa', 'baca', 'lewat', 'pc', 'harus', 'unduh', 'di', 'hp', 'padahal', 'di', 'hp', 'tulisan', 'ya', 'kecil', 'susah', 'bacanya', 'nyesel', 'banget', 'sudah', 'beli', 'buku', 'disini']	['ribet', 'baca', 'lewat', 'pc', 'unduh', 'hp', 'dulu', 'padahal', 'hp', 'tulisan', 'kecil', 'susah', 'bacanya', 'nyesel', 'banget', 'beli', 'buku', 'disini']	['ribet', 'baca', 'lewat', 'pc', 'unduh', 'hp', 'dulu', 'padahal', 'hp', 'tulisan', 'kecil', 'susah', 'baca', 'nyesel', 'banget', 'beli', 'buku', 'sini']	ribet baca lewat pc unduh hp dulu padahal hp tulis kecil susah baca nyesel banget beli buku sini
Tidak ada fitur untuk highlight.	Negatif	tidak ada fitur untuk highlight.	tidak ada fitur untuk highlight	tidak ada fitur untuk highlight	['tidak', 'ada', 'fitur', 'untuk', 'highlight']	['fitur', 'highlight']	['fitur', 'highlight']	fitur highlight
Opsi memory zoom tidak ada ketika mode	Negatif	opsi memory zoom tidak ada ketika mode	opsi memory zoom tidak ada ketika mode	opsi memory zoom tidak ada ketika mode	['opsi', 'memory', 'zoom', 'tidak', 'ada', 'ketika', 'mode']	['opsi', 'memory', 'zoom', 'mode', 'mema']	['opsi', 'memory', 'zoom', 'mode', 'baca', 'sehing']	opsi memory zoom mode baca sehin

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
membaca, sehingga sering setting zoom layar untuk kenyamanan membaca		membaca, sehingga sering setting zoom layar untuk kenyamanan membaca	membaca sehingga sering setting zoom layar untuk kenyamanan membaca	membaca sehingga sering setting zoom layar untuk kenyamanan membaca	'membaca', 'sehingga', 'sering', 'setting', 'zoom', 'layar', 'untuk', 'kenyamanan', 'membaca']	'sehingga', 'sering', 'setting', 'zoom', 'layar', 'kenyamanan', 'membaca']	g', 'sering', 'setting', 'zoom', 'layar', 'nyaman', 'baca']	gg sering setting zoom layar nyaman baca
perlu diviralin sih ada aplikasi ini	Positif	perlu diviralin sih ada aplikasi ini	perlu diviralin sih ada aplikasi ini	perlu diviralin sih ada aplikasi ini	['perlu', 'diviralin', 'sih', 'ada', 'aplikasi', 'ini']	['perlu', 'diviralin', 'sih', 'aplikasi']	['perlu', 'diviralin', 'sih', 'aplikasi']	perlu diviralin sih aplikasi
Woiiii yang bener ajaaa di aplikasii . Baru langgan an kemarin sudah disuruh update aplikasi. Setelah update buku2 yang sudah didownload hilang semuaaaaa. Gilaa masak harus gitu terus. Lebih	Negatif	woiiii yang bener ajaaa di aplikasii . baru langgan an kemarin sudah disuruh update aplikasi. setelah update buku2 yang sudah didownload hilang semuaaaaa. gilaa masak harus gitu terus. lebih	woiiii yang bener ajaaa di aplikasii baru langgan an kemarin sudah disuruh update aplikasi setelah update buku yang sudah didownload hilang semuaaaaa gilaa masak harus gitu terus lebih	woiiii yang bener saja di aplikasii baru langgan an kemarin sudah disuruh update aplikasi setelah update buku yang sudah diunduh semua gila masak gitu terus enak pas	['woiiii', 'yang', 'bener', 'saja', 'di', 'aplikasii', 'baru', 'langganan', 'kemarin', 'sudah', 'disuruh', 'update', 'aplikasi', 'setelah', 'update', 'buku', 'yang', 'sudah', 'di', 'unduh', 'ilang', 'semua', 'gila', 'masak', 'harus', 'gitu', 'terus', 'lebih', 'enak',	['woiiii', 'bener', 'aplikasi', 'baru', 'langgan', 'kemarin', 'disuruh', 'update', 'aplikasi', 'update', 'buku', 'ilang', 'semua', 'gila', 'masak', 'terus', 'lebih', 'enak', 'pas', 'diupdate']	['woiiii', 'bener', 'aplikasi', 'baru', 'langgan', 'kemarin', 'suruh', 'update', 'aplikasi', 'update', 'buku', 'unduh', 'ilang', 'semua', 'gila', 'masak', 'harus', 'gitu', 'terus', 'lebih', 'enak', 'pas',	woiiii bener aplikasi baru langgan kemarin suruh update aplikasi update buku unduh ilang semua gila masak harus gitu terus lebih enak pas

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
enak pas sebelum diupdate .		enak pas sebelum diupdate .	enak pas sebelum diupdate	sebelum diupdate	'pas', 'sebelum', 'diupdate']		'diupdate']	diupdate
aplikasi sangat mengecewakan, sudah beli paketnya tapi malah gak kebuka	Negatif	aplikasi sangat mengecewakan, sudah beli paketnya tapi malah gak kebuka	aplikasi sangat mengecewakan sudah beli paketnya tapi malah gak kebuka	aplikasi sangat mengecewakan sudah beli paketnya tapi malah tidak kebuka	['aplikasi', 'sangat', 'mengecewakan', 'sudah', 'beli', 'paketnya', 'tapi', 'malah', 'tidak', 'kebuka']	['aplikasi', 'mengecewakan', 'beli', 'paketnya', 'malah', 'kebuka']	['aplikasi', 'kecewakan', 'beli', 'paket', 'malah', 'buka']	aplikasi kecewa beli paket malah buka
terimakasih	Positif	terimakasih	terimakasih	terimakasih	['terimakasih']	['terimakasih']	['terimakasih']	terimakasih
Format tampilan kurang nyaman untuk baca di hp atau ereader krn kecil.	Negatif	format tampilan kurang nyaman untuk baca di hp atau ereader krn kecil.	format tampilan kurang nyaman untuk baca di hp atau ereader krn kecil	format tampilan kurang nyaman untuk baca di hp atau ereader karena kecil	['format', 'tampilan', 'kurang', 'nyaman', 'untuk', 'baca', 'di', 'hp', 'atau', 'ereader', 'karena', 'kecil']	['format', 'tampilan', 'kurang', 'nyaman', 'baca', 'hp', 'ereader', 'kecil']	['format', 'tampilan', 'kurang', 'nyaman', 'baca', 'hp', 'ereader', 'kecil']	format tampilan kurang nyaman baca hp ereader kecil
Hai, itu ada beberapa ebook error. Saat aku mau beli "janji kematian Lucifer" situ tertulis (error-buku tidak	Negatif	hai, itu ada beberapa ebook error. saat aku mau beli "janji kematian lucifer" situ tertulis (error-buku tidak	hai itu ada beberapa ebook error saat aku mau beli janji kematian lucifer situ tertulis error buku tidak ditemuk	hai itu ada beberapa ebook error saat aku mau beli janji kematian lucifer situ tertulis error buku tidak ditemuk	['hai', 'itu', 'ada', 'beberapa', 'ebook', 'error', 'saat', 'aku', 'mau', 'beli', 'janji', 'kematian', 'lucifer', 'situ', 'tertulis', 'error', 'buku']	['hai', 'beberapa', 'ebook', 'error', 'aku', 'mau', 'beli', 'janji', 'kematian', 'lucifer', 'situ', 'tertulis', 'error', 'buku']	['hai', 'beberapa', 'ebook', 'error', 'aku', 'mau', 'beli', 'janji', 'mati', 'lucifer', 'situ', 'tertulis', 'error', 'buku']	hai beberapa ebook error aku mau beli janji kematian lucifer situ tertulis error buku

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
ditemukan) padahal aku bisa baca previewnya.		ditemukan) padahal aku bisa baca previewnya.	an padahal aku bisa baca previewnya	an padahal aku bisa baca previewnya	'tidak', 'ditemukan', 'padahal', 'aku', 'bisa', 'baca', 'preview', 'nya']	'ditemukan', 'padahal', 'aku', 'baca', 'preview']	'temu', 'padahal', 'aku', 'baca', 'preview']	temu padahal aku bisa preview
gremedia semakin berkuwalitas	Positif	gremedia semakin berkuwalitas	gremedia semakin berkuwalitas	gremedia semakin berkuwalitas	['gremedia', 'semakin', 'berkuwalitas']	['gremedia', 'semakin', 'berkuwalitas']	['gremedia', 'makin', 'berkuwalitas']	gremedia makin berkuwalitas
Di 2025 ini tolong ditambahkan fitur untuk menandai teks (kayak stabilo atau garis bawah). Sayangnya sekali baca buku tapi nggak bisa tandai teks bacaan yang penting. Semoga inovasinua seperti google	Positif	di 2025 ini tolong ditambahkan fitur untuk menandai teks (kayak stabilo atau garis bawah). sayang sekali baca buku tapi nggak bisa tandai teks bacaan yang penting. semoga inovasinua seperti google	di ini tolong ditambahkan fitur untuk menandai teks kayak stabilo atau garis bawah sayang sekali baca buku tapi nggak bisa tandai teks bacaan yang penting semoga inovasinua seperti google	di ini tolong ditambahkan fitur untuk menandai teks kayak stabilo atau garis bawah sayang sekali baca buku tapi nggak bisa tandai teks bacaan yang penting semoga inovasinua seperti google	['di', 'ini', 'tolong', 'ditambahkan', 'fitur', 'untuk', 'menandai', 'teks', 'kayak', 'stabilo', 'atau', 'garis', 'bawah', 'sayang', 'sekali', 'baca', 'buku', 'tapi', 'tidak', 'bisa', 'tandai', 'teks', 'bacaan', 'yang', 'penting', 'semoga', 'inovasinua', 'seperti', 'google',	['ditambahkan', 'fitur', 'menandai', 'teks', 'kayak', 'stabilo', 'garis', 'bawah', 'sayang', 'sekali', 'baca', 'buku', 'tandai', 'teks', 'bacaan', 'yang', 'penting', 'semoga', 'inovasinua', 'playbook']	['tambah', 'fitur', 'tanda', 'teks', 'kayak', 'stabilo', 'garis', 'bawah', 'sayang', 'sekali', 'baca', 'buku', 'tanda', 'teks', 'baca', 'penting', 'moga', 'inovasinua', 'google', 'playbook']	tambah fitur tanda teks kayak stabilo garis bawah sayang sekali baca buku tanda teks bacaan yang penting moga inovasinua seperti google

content	label	lower	cleaned	normalized	tokens	no_stop_word	stemmed	final_cleaned
playbook		playbook	playbook	playbook	'playbook']			
Alhamdulillah bisa baca majalah bobo	Positif	alhamdulillah bisa baca majalah bobo	alhamdulillah bisa baca majalah bobo	alhamdulillah bisa baca majalah bobo	['alhamdulillah', 'bisa', 'baca', 'majalah', 'bobo']	['alhamdulillah', 'baca', 'majalah', 'bobo']	['alhamdulillah', 'baca', 'majalah', 'bobo']	alhamdulillah bisa baca majalah bobo
	Positif				[]	[]	[]	
Aplikasinya bagus tapi kok gak ada mode gelap ya. Sayang banget, gak kaya Google Play Book.	Negatif	aplikasinya bagus tapi kok gak ada mode gelap ya. sayang banget, gak kaya google play book.	aplikasinya bagus tapi kok gak ada mode gelap ya sayang banget gak kaya google play book	aplikasinya bagus tapi kok tidak ada mode gelap ya sayang banget tidak kaya google play book	['aplikasinya', 'bagus', 'tapi', 'kok', 'tidak', 'ada', 'mode', 'gelap', 'ya', 'sayang', 'banget', 'gak', 'kaya', 'google', 'play', 'book']	['aplikasinya', 'bagus', 'kok', 'mode', 'gelap', 'sayang', 'banget', 'kaya', 'google', 'play', 'book']	['aplikasi', 'bagus', 'kok', 'mode', 'gelap', 'sayang', 'banget', 'kaya', 'google', 'play', 'book']	aplikasinya bagus kok mode gelap sayang banget kaya google play book
baikkkkk dulu	Positif	baikkkkk dulu	baikkkkk dulu	baik dulu	['baik', 'dulu']	['baik', 'dulu']	['baik', 'dulu']	baik dulu
fleksibel dan worth it buat kantong mahasiswa  selalu repurchase	Positif	fleksibel dan worth it buat kantong mahasiswa  selalu repurchase	fleksibel dan worth it buat kantong mahasiswa selalu repurchase	fleksibel dan worth it buat kantong mahasiswa selalu repurchase	['fleksibel', 'dan', 'worth', 'it', 'buat', 'kantong', 'mahasiswa', 'selalu', 'repurchase']	['fleksibel', 'worth', 'it', 'buat', 'kantong', 'mahasiswa', 'selalu', 'repurchase']	['fleksibel', 'worth', 'it', 'buat', 'kantong', 'mahasiswa', 'selalu', 'repurchase']	fleksibel worth it buat kantong mahasiswa selalu repurchase

## Lampiran 3 Hasil Terjemah


### Bahasa Inggris

#### ABSTRACT

Sasikirana, Priaji Dewa, 2026. **The Analysis of User Satisfaction on the Gramedia Digital Application Using the Support Vector Machine (SVM) method**. Thesis. Library and Information Science, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang Advisors: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Ach. Nizam Rifqi, M.A

Keywords: *sentiment analysis, user reading satisfaction, Gramedia Digital, Support Vector Machine*

Gramedia Digital is a digital reading service platform in response to technological advancements. This research aims to analyze user satisfaction with the Gramedia Digital application by conducting sentiment analysis on user reviews from the Google Play Store using the Support Vector Machine (SVM) method. User reviews reveal a discrepancy between expectations and actual experiences, including technical issues such as automatic logout, limited features, and application instability. The research used a dataset of 1,032 reviews between January 2020 and May 2025. It applied a text mining approach consisting of several stages: data collection, expert-based labeling, preprocessing (including case folding, cleaning, normalization, tokenization, stopword removal, and stemming), TF-IDF weighting, and classification using SVM. For model evaluation, the researcher employed a confusion matrix with accuracy, precision, recall, and F1-score metrics across three data-splitting scenarios (70:30, 80:20, and 90:10). The results identify the 70:30 scenario as the best-performing configuration, achieving an accuracy of 81%, indicating strong alignment between predicted and actual labels. The model also attains a precision of 82%, demonstrating high correctness in positive predictions, a recall of 78%, reflecting its ability to identify positive instances, and an F1-score of 79%, indicating a balanced performance between precision and recall. The research result shows that the SVM method performs sentiment analysis effectively. Based on the sentiment analysis results, user reading satisfaction remains relatively low, as indicated by a higher proportion of negative reviews compared to positive ones, along with frequent complaints regarding feature limitations, application stability, and access issues after payment. These findings indicate that sentiment analysis serves as a valuable tool to understand user perceptions and reading satisfaction, while also providing recommendations for Gramedia Digital to improve the quality of its application services.

 <p>Penerjemah Rizka Yanuari NIPBP 197801242023212005</p>	<p>Tanggal 4-05-2026</p>
--	------------------------------

## Bahasa Arab

## مستخلص البحث

ساسيكيرانا، براجي ديفا، ٢٠٢٦. تحليل رضا مستخدمي تطبيق غراميديا الرقمي باستخدام طريقة آلة للنتجه الداعمة (SVM). البحث الجامعي. قسم المكتبات وعلوم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فخرس حسن ريزا محفوظ، الماجستير؛ المشرف الثاني: أحمد نظام رقمي، الماجستير.

الكلمات الرئيسية: تحليل مشاعر، رضا مستخدمين لقراءة، غراميديا رقمي، آلة منتهج داعمة.

غراميديا الرقمي هو منصة خدمة قراءة رقمية ظهرت استجابةً للتقدم التكنولوجي. هدف هذا البحث إلى تحليل رضا المستخدمين عن تطبيق غراميديا الرقمي من خلال تحليل المشاعر للتعليقات على منجر جوجل بلاي باستخدام طريقة آلة للنتجه الداعمة (SVM). تُظهر تعليقات المستخدمين وجود عدم توافق بين التوقعات والتجربة، مثل الصعوبات التقنية المتمثلة في تسجيل الخروج التلقائي، محدودية الميزات، وكذلك اضطرابات استقرار التطبيق. تم استخدام ١٠٣٢ تعليقًا خلال الفترة من يناير ٢٠٢٠ حتى مايو ٢٠٢٥. استخدم هذا البحث منهج التنقيب النصي من خلال مراحل جمع البيانات، ووضع العلامات من قبل خبراء اللغة، والمعالجة المسبقة (تحويل الحالات، التنظيف، التجزئة، إزالة كلمات الوقف، والجزر)، وتوزين TF-IDF، وكذلك التصنيف باستخدام SVM. تم تقييم النموذج باستخدام مصفوفة الالتباس مع مقاييس الدقة والثبات والاسترجاع ونسبة ف١ من خلال ثلاثة سيناريوهات لتقسيم البيانات (٧٠:٣٠، ٨٠:٢٠، و٩٠:١٠). أفضل سيناريو هو ٧٠:٣٠ مع دقة تبلغ ٨١٪ مما يشير إلى التوافق بين نتائج التنبؤ والتسميات الفعلية، والثبات يبلغ ٨٢٪ مما يشير إلى صحة التنبؤات الإيجابية، والاسترجاع يبلغ ٧٨٪ مما يشير إلى قدرة النموذج على التعرف على البيانات الإيجابية، ونسبة ف١ تبلغ ٧٩٪ مما يعكس التوازن بين الثبات والاسترجاع. أظهرت نتائج هذا البحث أن طريقة SVM قادرة على إجراء تحليل المشاعر بشكل جيد إلى حد ما. بناءً على تحليل المشاعر، أشار رضا المستخدمين للقراءة أن أغلب المستخدمين يشعرون بعدم الرضا، واتضح ذلك من خلال وجود مراجعات سلبية أكثر من المراجعات الإيجابية بالإضافة إلى الشكاوى المتعلقة بمحدودية الميزات، واستقرار التطبيق، ومشاكل الوصول بعد الدفع. تشير هذه النتائج إلى أن تحليل المشاعر يمكن استخدامه كأداة لفهم تصورات ورضا المستخدمين للقراءة، وكذلك لتقديم توصيات لغراميديا الرقمي في تحسين جودة خدمات التطبيق.

<p>Penerjemah</p>  <p>M. Mubasysir Munnir, MA NIPPPK: 198605132023211019</p>		<p>Tanggal 04-05-2026</p>
---	--	-------------------------------

## Lampiran 4 Turnitin

### 21% Overall Similarity




The combined total of all matches, including overlapping sources, for each database.

#### Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

---

#### Top Sources

- 16%  Internet sources
  - 12%  Publications
  - 17%  Submitted works (Student Papers)
-