

PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE TERHADAP ANALISIS SENTIMEN PADA ULASAN APLIKASI LINKAJA

Elvin Alan Pratama^{1*}, Aji Primajaya², Iqbal Maulana³

^{1,2,3} Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang;
Jl. H.S. Ronggowaluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

Keywords:

Sentiment Analysis,
Knowledge Discovery in
Database, LinkAja, Root
Cause Analysis, Support
Vector Machine

Correspondent Email:

xxxxxxxxxx@xxxx.xxx

Abstrak. Kehadiran dompet digital berperan dalam meningkatnya volume transaksi uang elektronik di Indonesia. LinkAja merupakan salah satu dompet digital dengan rating dan pengguna terendah di Indonesia dibandingkan dompet digital lainnya. Penelitian ini bertujuan untuk menerapkan model klasifikasi Support Vector Machine (SVM) dalam analisis sentimen ulasan pengguna LinkAja serta menginterpretasikan hasilnya sebagai rekomendasi peningkatan layanan. Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD) dalam text mining. Proses dalam penelitian ini melibatkan beberapa tahapan, yaitu data selection, data preprocessing, data transformation, data mining, evaluation dan knowledge. Dataset diperoleh dengan teknik scraping dari Google Plays Store menggunakan Python, didapatkan 2,000 ulasan dari Januari hingga November 2024, dengan 944 ulasan positif dan 1,056 ulasan negatif. Model terbaik diperoleh pada skenario pembagian data 70:30 dengan performa unggul, yaitu accuracy 88%, precision 90%, recall 83%, dan F1-score 87%. Hasil word cloud menunjukkan kata-kata yang sering muncul pada tiap kelas sentimen. Pada sentimen positif, kata-kata seperti "aplikasi" dan "saldo" mendominasi, hal ini menunjukkan kemudahan penggunaan dan pengiriman saldo. Sementara itu pada sentimen negatif didominasi oleh "aplikasi" dan "pakai", hal ini mengindikasikan bahwa pengguna terkendala saat menggunakan aplikasi tersebut. Melalui teknik Root Cause Analysis (RCA) mengidentifikasi enam faktor yang memengaruhi kepuasan pengguna, diantaranya dukungan pelanggan yang kurang, crash pada perangkat rendah, kendala pada koneksi, UI/UX yang membingungkan, dan minimnya pemantauan bug. Hasil dari pencarian akar masalah ini kemudian divisualisasikan dalam Diagram Ishikawa.



JITET is licensed under
a Creative Commons
Attribution-NonCommercial
4.0 International License

Abstract. The presence of digital wallets has contributed to the increasing volume of electronic money transactions in Indonesia. LinkAja is one of the digital wallets with the lowest rating and user base in the country compared to other digital wallets. This study aims to apply the Support Vector Machine (SVM) classification model in analyzing user sentiment towards LinkAja reviews and interpreting the results as recommendations for service improvements. The research method follows the Knowledge Discovery in Database (KDD) framework in text mining. The process involves several stages, including data selection, preprocessing, transformation, data mining, evaluation, and knowledge extraction. The dataset was obtained through web scraping from Google Play Store using Python, collecting 2,000 reviews from January to November 2024, consisting of 944 positive reviews and 1,056 negative reviews. The best model was achieved with a 70:30 data split scenario, yielding superior performance with an accuracy of 88%, precision of 90%, recall of 83%, and an F1-score of 87%. The word cloud results

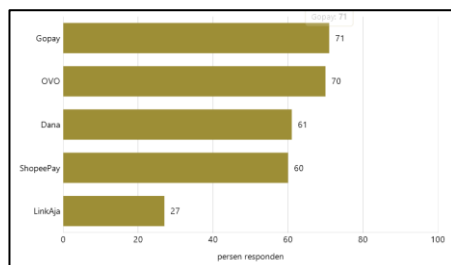
highlight frequently occurring words in each sentiment category. In positive sentiment, words such as "application" and "balance" dominate, indicating ease of use and smooth balance transfers. Meanwhile, negative sentiment is dominated by "application" and "use," suggesting that users face difficulties while using the app. Through the Root Cause Analysis (RCA) technique, six key factors affecting user satisfaction were identified, including poor customer support, crashes on low-end devices, connectivity issues, confusing UI/UX, and inadequate bug monitoring. The findings from this root cause analysis are then visualized in an Ishikawa Diagram.

1. PENDAHULUAN

Perkembangan pengguna *smartphone* di tanah air terus mengalami peningkatan. Pada tahun 2023 sendiri perkembangan yang dicapai adalah sebesar 83,6 dan akan diprediksi akan terus tumbuh hingga 89,2% di tahun 2025 mendatang [1]. Saat ini, *smartphone* telah menjadi perangkat utama untuk berkomunikasi sekaligus mengakses berbagai layanan digital [2]. Semua layanan tersebut tersedia dalam bentuk aplikasi yang dapat diunduh dan dipasang melalui *platform* distribusi aplikasi digital, salah satunya adalah Google Play Store.

Sejalan dengan peningkatan pengguna *smartphone*, diikuti juga oleh transaksi uang elektronik yang mengalami hal sama. Volume transaksi uang elektronik di Indonesia terus mengalami peningkatan signifikan terhitung sejak awal 2019, dengan volume transaksi tertinggi tercatat pada Februari 2020, yang mencapai total 2.372.349.430 transaksi, sedangkan nilai transaksi tertinggi terjadi pada Desember 2023, mencapai Rp184 triliun [3].

Salah satu metode pengelolaan dan pembayaran uang berbasis elektronik ialah dompet digital. Di Indonesia sendiri, terdapat berbagai macam aplikasi dompet digital seperti Gopay, OVO, Dana, Shopeepay, dan Link Aja.



Gambar 1.1 Dompet Digital di Indonesia Berdasarkan Banyaknya Pengguna (Responden)

Merujuk pada data dari Insight Asia mengenai penggunaan dompet digital di Indonesia, dapat diketahui bahwa LinkAja

memiliki pengguna paling sedikit dibanding dompet digital lainnya dengan persentase sebesar 27%.

Sejak pertama kali diluncurkan pada tahun 2019, aplikasi LinkAja telah diunduh oleh 10 juta pengguna dengan total *rating* mencapai 3,6 dan memperoleh ulasan sebanyak 781,282 di Google Play Store. *Rating* tersebut tergolong rendah bila dibandingkan dengan *rating* yang didapat oleh aplikasi dompet digital pesaingnya. Selain *rating* pengguna, ulasan juga berperan penting untuk memahami respons pengguna dalam menilai kualitas dari sebuah aplikasi. Ulasan-ulasan tersebut bisa dijadikan informasi yang bermanfaat bagi pengembang. Untuk mendapatkan pemahaman mengenai penilaian suatu produk, dapat dilakukan melalui analisis sentimen [4].

Analisis sentimen merupakan cabang dari *Natural Language Processing* (NLP) yang berfokus pada pembangunan sistem untuk mengenali dan mengambil opini dari teks[5]. Sentimen adalah perasaan, sikap, evaluasi, atau emosi mendasar yang terkait dengan suatu opini. Orientasi pada sentimen dapat berupa positif, negatif, atau netral [6]. Maka sentimen-sentimen ini dapat digunakan untuk mengkaji emosi dan ulasan masyarakat melalui sentimen positif dan negatif.

Salah satu algoritma yang sering digunakan adalah *Support Vector Machine* (SVM) karena mampu bekerja dengan akurasi 85-90% dalam mengklasifikasikan sentimen positif dan negatif [4]. *Support Vector Machine* menggunakan sebuah *hyperplane* untuk memisahkan dua kelas data dengan *margin* selebar mungkin[7]. Algoritma *Support Vector Machine* telah banyak diterapkan dan terbukti efektif untuk analisis sentimen dan tugas *text mining* [8].

Pada penelitian sebelumnya yang dilakukan oleh [9], menunjukkan bahwa kombinasi lexicon InSet dengan *Support Vector Machine* menghasilkan akurasi 87,83%. Nilai ini lebih

unggul 0,17% jika dibandingkan dengan kombinasi lexicon VADER dengan *Support Vector Machine* dengan nilai akurasi 87,66%.

Sementara itu, penelitian yang dilakukan oleh [10] menggunakan algoritma K-NN menghasilkan nilai akurasi sebesar 76% jika k bernilai 3. Sementara untuk k bernilai 4 menghasilkan nilai akurasi sebesar 76%. Dalam penelitian tersebut juga dilakukan analisis masalah berdasarkan sentimen-sentimen yang diperoleh menggunakan teknik *Root Cause Analysis* (RCA).

Penelitian ini akan mengklasifikasikan polaritas sentimen ulasan aplikasi LinkAja menjadi positif dan negatif menggunakan lexicon InSet. Proses klasifikasi akan dilakukan dengan algoritma *Support Vector Machine* yang dioptimalkan menggunakan hyperparameter tuning untuk mencari parameter terbaik. Hasil klasifikasi selanjutnya akan dievaluasi menggunakan Confusion Matrix, Kurva ROC, dan AUC. Hasil evaluasi ini akan digunakan untuk mengidentifikasi akar penyebab rendahnya rating aplikasi melalui *Root Cause Analysis* (RCA), yang akan divisualisasikan dalam bentuk diagram Ishikawa.

2. TINJAUAN PUSTAKA

2.1 Google Play Store

Google Play Store merupakan sebuah *platform* yang memfasilitasi pengguna untuk mengunduh dan menggunakan beragam aplikasi dari pihak ketiga. Jutaan pengguna membagikan informasi pribadi ke Google maupun ke layanan pihak ketiga demi dapat mengakses aplikasi di ponsel dan tablet Android [11].

2.2 LinkAja

LinkAja merupakan produk aplikasi terkemuka yang dikembangkan oleh PT Fintek Karya Nusantara (Finarya) dan telah mendapatkan registrasi resmi dari Bank Indonesia sejak 21 Februari 2019. LinkAja memberikan kemudahan dalam melakukan pembayaran dan pembelian untuk berbagai kebutuhan di berbagai lokasi, seperti di *platform* E-commerce, layanan ojek *online*, *merchant*, dan berbagai macam platform lainnya [12].

2.3 Analisis Sentimen

Analisis sentimen merupakan bidang penelitian yang bertujuan untuk mengevaluasi opini dan perasaan individu terhadap suatu entitas, seperti produk, layanan, individu, atau organisasi, yang diekspresikan dalam bentuk teks. Proses ekstraksi pendapat ini melibatkan penggunaan algoritma agar teks yang dianalisis dapat dipahami dengan baik. [13].

2.4 Knowledge Discovery In Database

Knowledge Discovery in Database (KDD) merupakan proses untuk mengeksplorasi dan menganalisis kumpulan data guna memperoleh informasi dan pengetahuan yang bermanfaat. Informasi ini dapat menjadi dasar untuk pengambilan keputusan [14]. Proses KDD secara umum terdiri dari *data selection*, *data preprocessing*, *data transformation*, *data mining*, *evaluation*, dan *knowledge*. [15].

2.5 Algoritma Support Vector Machine

Support Vector Machine adalah sebuah algoritma linear yang dapat digunakan untuk berbagai tujuan seperti regresi, klasifikasi estimasi *density*, deteksi kebaruan, dan lain-lain. Pada kasus klasifikasi dua kelas, *Support Vector Machine* bekerja dengan mencari sebuah *hyperplane* yakni garis atau bidang pemisah antara dua kelas data dengan *margin* sebesar mungkin [16].

2.6 Kamus Lexicon

Kamus *lexicon* adalah sebuah kamus yang berfungsi untuk memberikan nilai atau bobot pada setiap kata dalam sebuah kalimat. Nilai-nilai tersebut kemudian digabungkan untuk menghitung sentimen keseluruhan dari kalimat tersebut berdasarkan hasil pembobotan yang telah dilakukan [17].

2.7 Root Cause Analysis

Causal Analysis atau biasa disebut sebagai *Root Cause Analysis* (RCA), adalah sebuah teknik yang digunakan untuk mencari atau mengidentifikasi penyebab utama atau akar dari suatu masalah. Tujuan *Root Cause Analysis* (RCA) adalah menganalisis dan memahami alasan-alasan di balik terjadinya sebuah masalah tersebut sejak awal, sehingga penyebab utama dari masalah tersebut dapat diidentifikasi dan bukan hanya sekedar gejala-gejalanya [18].

2.8 Word Cloud

Word Cloud atau biasa juga disebut *tag cloud*, adalah teknik visualisasi data yang menekankan data teks penting berdasarkan jumlah kemunculannya. Pada *Word Cloud*, semakin sering kata muncul, maka semakin besar ukurannya yang ditampilkan yang lebih besar, sebaliknya kata-kata dengan frekuensi rendah akan muncul dengan ukuran yang lebih kecil [19].

2.9 Diagram Ishikawa

Diagram Ishikawa atau biasa disebut diagram *fishbone*, pertama kali diperkenalkan oleh Kaoru Ishikawa untuk digunakan dalam kontrol kualitas dan rekayasa industri. Diagram ini sangat berguna dalam berbagai situasi yang memerlukan pemahaman mendalam tentang hubungan sebab-akibat [20].

3. METODE PENELITIAN

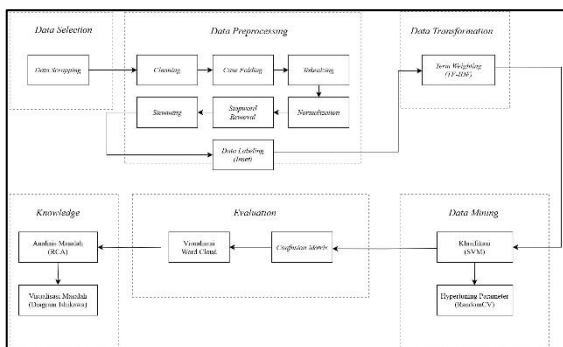
Dalam penelitian ini, metode yang digunakan adalah *Knowledge Discovery in Database (KDD)* dalam *text mining*. Proses dalam penelitian ini melibatkan beberapa tahapan, yakni sebagai berikut:

3.1. Data Selection

Langkah pertama adalah mengumpulkan data ulasan aplikasi LinkAja dari Google Play Store menggunakan Google Colab dan bahasa pemrograman Python, dengan rentang waktu Januari-November 2024. Ulasan yang tidak relevan, seperti simbol, angka, atau abjad acak, akan disaring karena tidak mengandung sentimen atau informasi untuk klasifikasi.

3.2. Data Preprocessing

Setelah data ulasan terkumpul, akan dilakukan penghilangan *noise* pada data, semisal tautan, karakter khusus, atau *emoticon* yang tidak relevan. Berikut adalah tahapan *preprocessing*:



Gambar 2.1 Rancangan Penelitian

a) Data Cleaning

Cleaning bertujuan untuk menghilangkan data yang tidak relevan atau mengganggu dalam penelitian, seperti angka (1234567890), tanda baca (.,,;!?), symbol-simbol (<@#&%*/~) dan juga emoji.

b) Case Folding

Case folding bertujuan untuk untuk menyeragamkan semua data per karakter menjadi huruf kecil atau *lowercase*. Sebagai contoh kata “Mudah” akan diubah menjadi “mudah”, kemudian “Sulit” menjadi “sulit”.

c) Tokenizing

Tokenizing membagi data teks menjadi segmen-segmen kecil yang disebut dengan *token*. *Token* ini berbentuk karakter individual, kata, atau pun frasa, tergantung pada kebutuhan analisis. Contohnya adalah kalimat ‘aplikasi ini bagus banget’ menjadi [aplikasi, ini, bagus, banget].

d) Normalization

Dalam proses *normalization* dilakukan perubahan pada kata-kata yang sebelumnya tidak baku menjadi kata baku. Tujuannya untuk menyetarakan kata agar sama. Contohnya adalah kata ‘ngga’ menjadi ‘tidak’, ‘tf’ menjadi ‘transfer’.

e) Stopword Removal

Pada tahap eliminasi *stopword*, kata-kata yang dianggap tidak relevan atau tidak berarti dalam sebuah teks akan di hapus. Contohnya seperti kata “ke”, “itu”, “di”, “yang”, dan sebagainya.

f) Stemming

Dalam proses *stemming* semua data per kata yang mempunyai imbuhan harus dihapus sehingga menghasilkan kata sesuai kata bakunya. Contohnya seperti pada kata ‘namanya’ menjadi ‘nama’ dan ‘aplikasinya’ menjadi ‘aplikasi’.

g) Labeling

Selanjutnya, data akan melalui proses pelabelan menggunakan kamus InSet (*Indonesian Sentiment Lexicon*). Hal ini bertujuan untuk menentukan polaritas sentimen, apakah sentimen-sentimen tersebut memiliki polaritas positif atau negatif. Untuk sentimen yang bersifat netral tidak akan dilanjutkan ke dalam proses *data transformation*.

3.3. Data Transformation

Setelah tahap pembersihan data, langkah selanjutnya adalah pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk melakukan seleksi fitur dalam menentukan sejauh mana pentingnya suatu kata dalam ulasan berdasarkan frekuensinya.

3.4. Data Mining

Sebelum klasifikasi, data dibagi menjadi data latih dan data uji. Analisis sentimen kemudian dilakukan dengan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan polaritas ulasan menjadi sentimen positif atau negatif. Proses klasifikasi mencakup *hyperparameter tuning* menggunakan *Randomized Search Cross Validation* untuk menentukan kernel dan parameter terbaik.

3.5. Evaluation

Setelah analisis sentimen, evaluasi dilakukan menggunakan *Confusion Matrix* untuk mengukur *accuracy*, *precision*, *recall*, dan *F1-score* dengan algoritma *Support Vector Machine*. Kinerja klasifikasi juga dievaluasi melalui ROC dan AUC. Hasil klasifikasi disajikan dalam bentuk *Word cloud* untuk visualisasi kata-kata positif dan negatif yang sering muncul.

3.6. Knowledge

Selanjutnya, analisis masalah dilakukan dengan menggunakan *Root Cause Analysis* (RCA) untuk menemukan akar utama dari masalah tersebut. Analisis ini diambil berdasarkan sentimen-sentimen negatif yang telah diklasifikasikan. Hasil dari pencarian akar masalah tersebut kemudian akan divisualisasikan melalui diagram Ishikawa.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Proses pengambilan data dilakukan dengan menerapkan Teknik *Scrapping Data* menggunakan bantuan tool Library Scraper pada bahasa pemrograman Python terhadap ulasan pengguna aplikasi LinkAja dari Google Play Store. Data yang terkumpul sebanyak 2,000 yang terdiri dari berbagai jenis ulasan dalam Bahasa Indonesia. Pada *dataset* ini terdiri dari beberapa atribut, yang meliputi

Review ID, Username, Rating, Review Text,

Review ID	Username	Rating	Review Text	Date
0	62760029-5064-4028-a408-a1e690ce8e89	5	Muattapp	2024-12-02 10:06:15
1	4b58b3-8305-43f5-95c5-b0b11d5e43a	1	Saya sangat kecewa terhadap pihak link aja. Sa	2024-12-02 09:49:07
2	edc3be92-89e0-4895-b00a-838e3c5956c	1	LIVE CHAT TIDAK BERGUNA HHHH GANTI NOMOR T	2024-12-02 09:45:42
3	626add8b-9419-4958-b011-b0d7653271	1	Belum Daftar sudah dikirim kode verifikasi kee	2024-12-02 05:43:04
4	b4442770-2ac9-4365-8387-3a8004458d2	1	aplikasi sampahn	2024-12-02 05:40:20
...	...	...	...	...
1995	4e1d1cbe-7444-49d0-8744-ba030baad519	1	Kop dana sy tdk masuk	2024-07-03 05:08:37
1996	0b53c3de-dea4-4ab8-879b-39371f8c288	1	Aplikasi nya Sangat lambat proses tsul service	2024-07-03 04:50:12
1997	7d92cc3c-3d1a-4b22-8b9f-c1eb8b8d824e	2	Aplikasi gmn jln	2024-07-03 04:03:36
1998	221ab263-a802-4198-ba67-079089a12008	3	Sangat baik	2024-07-03 03:20:38
1999	7a0c3895-8a4c-4028-808a-17e09564c2e9	1	Gila susah bgt upgrade service mana uda masuk	2024-07-02 23:37:41

Gambar 4.1 Data Ulasan Pengguna Hasil Scrapping

dan Date.

4.2. Preprocessing

Dataset yang telah didapatkan masih sangat tidak terstruktur dan belum optimal untuk proses *data mining*. Oleh karena itu, dataset diproses terlebih dahulu pada tahap *preprocessing*.

1. Cleaning

Proses *cleaning* dilakukan dengan menghapus URL, angka, tanda baca, dan simbol-simbol yang tidak relevan untuk klasifikasi.

Tabel 4.1 Dataset Sebelum dan Sesudah Cleaning

No	Sebelum	Sesudah
1	Kecewa transfer ke mybca dari linkaja.di awal seblm trnsfer 297500,biaya transfer25000.saya coba transfer berhasil.ternyata yang masuk uang transferan ke mybca yg tiba jadi mngkin yg lain juga 260000.mngkin yg lain juga mengalami seperti saya???	Kecewa transfer ke mybca dari linkajadi awal seblm trnsfer biaya transfersaya coba transfer berhasilternyata yang masuk uang transferan ke mybca yg tiba jadi mngkin yg lain juga mengalami seperti saya
2	Makin gajelas ni aplikasi udah ga layak pakai Uninstall dahlah!!!	Makin gajelas ni aplikasi udah ga layak pakai Uninstall dahlah

2. Case Folding

Case folding bertujuan untuk menyamakan format teks dengan mengubah seluruh huruf menjadi huruf kecil, untuk mempermudah analisis selanjutnya.

Tabel 4.2 Dataset Sebelum dan Sesudah Case Folding

No	Sebelum	Sesudah
----	---------	---------

1.	Sejauh ini aplikasi nya aman ga ada kendala pokoknya bagus	sejauh ini aplikasi nya aman ga ada kendala pokoknya bagus
2.	Makin gajelas ni aplikasi udah ga layak pakai Uninstall dahlah	makin gajelas ni aplikasi udah ga layak pakai uninstall dahlah

3. Tokenizing

Tokenizing bertujuan untuk mengubah data ulasan yang semula berupa kalimat menjadi potongan kata per kata. Proses ini mempermudah tahap transformasi berikutnya dengan memproses kata-kata secara individual, bukan keseluruhan kalimat.

Tabel 4.3 Dataset Sebelum dan Sesudah *Tokenizing*

No	Sebelum	Sesudah
1.	belum daftar sudah dikirim kode verifikasi keemail entah email siapa yg dikirimkan hadew	belum,daftar,sudah,di kirim,kode,verifikasi, keemail,entah,email,si apa,yang,dikirimkan,h adew
2.	makin gajelas ni aplikasi udah ga layak pakai uninstall dahlah	makin,ga,jelas,nih,apli kasi,udah,ga,layak,pak ai,unihnstall,dahlah

4. Normalization

Mengubah kata-kata tidak baku dalam ulasan menjadi bentuk kata baku. Pada tahap ini, digunakan dokumen khusus yang berisi daftar kata baku dan tidak baku, yang berfungsi untuk mengganti kata tidak baku pada data ulasan dengan kata yang sesuai.

Tabel 4.4 Dataset Sebelum dan Sesudah *Normalization*

No	Sebelum	Sesudah
1.	ini, apk, apaan, dari, kemarin, transaksi, beli, token, listrik, gk, masuk, di, minta, kode, token, gk, dikirim, tp, saldo, berkurang	ini,apk,apaan, dari,kemarin, transaksi, beli,token,listrik, tidak,masuk,di, meminta,kode, token,tidak,dikirim, tapi,saldo,berkurang
2.	jelek, banget, la, aplikasi, ini, live, chat, ga, berfungsi, saldo, udh, ketarik, tapi,	jelek,banget,lah, aplikasi,ini,live, chat,tidak, berfungsi,saldo, sudah,tertarik,

token, listrik, ngga, dikirim, kirim, aduan, pengajuan, melalui, live, chatt, apaa, ngga, bisa, diketik, aneh, bgtt	tapi,token,listrik, tidak,dikirim, kirim,aduan, pengajuan,melalui, live,chat,apa, tidak,bisa,diketik, aneh,banget
---	---

5. Stopword Removal

Stopword removal bertujuan untuk membuang kata-kata umum yang tidak relevan seperti kata penghubung dan kata ganti, untuk menyisakan kata-kata bermakna yang mendukung analisis sentimen.

Tabel 4.5 Dataset Sebelum dan Sesudah *Stopword Removal*

No	Sebelum	Sesudah
1	kenapa,proses,upgrade,premium,lama,sekali,saya,sudah,menunggu,lebih,dari,jam,tidak,ada,konfirmasi	proses, upgrade, premium, jam, konfirmasi
2	jangan,pakai,apk,ini,saya,transaksi,qr,di,link,nya,sukses,tapi,ke,penerima,tidak,ada,hadeh	pakai, apk, transaksi, qr, link, nya, sukses, penerima, hadeh

6. Stemming

Stemming yaitu proses penghilangan imbuhan yang ada pada kata, baik di awal, akhir, maupun kombinasi keduanya sehingga kata tersebut menjadi bentuk dasarnya. Tujuan dari tahap ini adalah untuk mengurangi frekuensi kata turunan agar fokus tetap pada bentuk dasar kata. Pada tahap ini, digunakan *library* Sastrawi pada Python.

Tabel 4.6 Dataset Sebelum dan Sesudah *Stemming*

No	Sebelum	Sesudah
1.	Xmotong, saldo, uang, orang, no, token, listrik, dikasih, laporan, cs, hasilnya, dikirim, no, token, listrikny, kecewa, berat	motong, saldo, uang, orang, no, token, listrik, kasih, laporan, cs, hasil, kirim, no, token, listrik, kecewa, berat
2.	profesional, menanggapi, keluhan, pelanggan, gagal, bayar, status, sukses, merekomen	profesional, tanggap, keluhan, pelanggan, gagal, bayar, status, sukses, rekomendasi, pakai, ewallet, kecewa

	dasikan, memakai, ewallet, kecewa	
--	-----------------------------------	--

7. Data Labelling

Pada tahap pelabelan, data diberi bobot sentimen menggunakan kamus InSet *lexicon* dengan menghitung skor dari jumlah kata positif dan negatif. Hasilnya, dari 2,000 data, sebanyak 867 diklasifikasikan sebagai sentimen positif dan 1,133 sebagai negatif.

	stemming	label	weight
0	mantap, kali	positive	5 + 0 = 5
1	kecewa, link, isi, saldo, tanggal, kematin, ku...	positive	0 + 0 + -2 + 2 + 0 + 0 + -3 + 0 + 0 + 3 + 3 + ...
2	live, chat, guna, ganti, nomor, aplikasilii, m...	negative	3 + 3 + 1 + -3 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + ...
3	daftar, kirim, kode, verifikasi, email, email,...	positive	-1 + 2 + 0 + 0 + 0 + 0 + 2 + 0 = 3
4	aplikasi, sampah	negative	0 + 0 = 0
5	bell, pulsa, bayar, anjir	negative	-2 + 0 + -1 + -1 = -4
6	nih, aplikasi, layak, pakai, unihstaili, dah	negative	0 + 0 + -3 + 0 + 0 + 0 = -3
7	email, nonaktif, digantilive, chat, respon, ya...	positive	0 + 0 + 0 + 3 + 2 + 0 + 3 + 0 = 8
8	bagus	negative	-2 = -2
9	baru, rutin, deteksi, aplikasi, curiga	negative	0 + 0 + 0 + 0 + -1 = -1

Gambar 4.2 Hasil Labelling InSet

4.3. Data Transformation

Pada tahap ini, dilakukan pembobotan kata menggunakan metode TF-IDF, yang menghitung frekuensi kemunculan kata (TF) dan seberapa signifikan kata tersebut dalam keseluruhan dokumen (IDF). Kombinasi TF dan IDF menghasilkan bobot yang merepresentasikan pentingnya masing-masing kata.

	apk	banget	bikin	cipta	customer	game	kapok	layak
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.594424	0.000000	0.259898	0.381979	0.000000	0.381979	0.000000	0.000000
2	0.000000	0.222534	0.000000	0.000000	0.21647	0.000000	0.376174	0.238661

Gambar 4.3 Hasil Pembobotan TF-IDF

4.4. Data Mining

Pada Pada tahapan ini akan dilakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi LinkAja menggunakan *Support Vector Machine* (SVM) dan menggunakan *Random Search Cross-Validation* dengan skenario sebagai berikut :

Tabel 4.7 Skenario Pembagian Split Data

Skenario	Pembagian Data	Data Latih	Data Uji
Skenario 1	90:10	1800	200
Skenario 2	80:20	1600	400
Skenario 3	70:30	1400	600
Skenario 4	60:40	1200	800
TOTAL		2000	

Model diuji dengan berbagai skenario pembagian data, yaitu rasio 90:10, rasio 80:20, rasio 70:30, dan rasio 60:40. Pengujian skenario ini akan dilakukan dua kali, untuk skenario pertama akan dilakukan tanpa menggunakan *Random Search CV*, sedangkan untuk skenario kedua akan menggunakan *Random Search CV*.

Tabel 4.8 Hasil Akurasi Data Mining

Skenario	Kernel	Tanpa Rando mCV	Rando mCV	Peningkatan
90:10	Linear	0.87	0.87	↑0.00
	RBF	0.84	0.87	↑0.03
	Polynomial	0.80	0.82	↑0.02
	Sigmoid	0.89	0.88	↓0.01
80:20	Linear	0.89	0.89	↑0.00
	RBF	0.86	0.88	↑0.02
	Polynomial	0.77	0.86	↑0.09
	Sigmoid	0.87	0.88	↑0.01
70:30	Linear	0.88	0.88	↑0.00
	RBF	0.85	0.88	↑0.03
	Polynomial	0.76	0.87	↑0.11
	Sigmoid	0.87	0.87	↑0.00
60:40	Linear	0.86	0.86	↑0.00
	RBF	0.85	0.87	↑0.02
	Polynomial	0.75	0.86	↑0.11
	Sigmoid	0.85	0.85	↑0.00

Perbandingan kinerja model SVM dengan empat kernel yakni *linear*, RBF, *polynomial*, dan *sigmoid* pada berbagai skenario 90:10 hingga 60:40 menunjukkan bahwa *Random Search CV* dapat meningkatkan akurasi model. Pada skenario 90:10, kernel RBF dan *polynomial* masing-masing meningkat sebesar 0,03% dan 0,02%. Pada skenario 80:20, kernel RBF meningkat 0,02% , *polynomial* meningkat 0,09% dan *sigmoid* meningkat 0,01%. Pada 70:30, kernel RBF meningkat 0,03% dan *polynomial* meningkat 0,11% dan. Pada 60:40, kernel RBF meningkat 0,02% dan *polynomial* meningkat 0,11%.

4.5. Evaluation

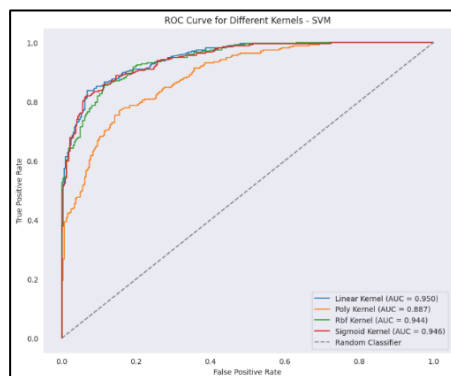
Setelah model selesai dibuat, setiap skenario diuji untuk mengevaluasi performa dengan menghitung *accuracy*, *precision*, *recall*, dan *F1-score*. Perbandingan menunjukkan hasil terbaik pada skenario 70:30, yang diukur dengan nilai AUC dan grafik ROC.

Tabel 4.9 Hasil AUC

Skenario	Kernel	AUC	Rata-Rata
90:10	Linear	0.95	0.9300
	RBF	0.94	
	Polynomial	0.88	

	<i>Sigmoid</i>	0.95	
80:20	<i>Linear</i>	0.95	0.9275
	RBF	0.94	
	<i>Polynomial</i>	0.87	
	<i>Sigmoid</i>	0.95	
70:30	<i>Linear</i>	0.95	0.9325
	RBF	0.94	
	<i>Polynomial</i>	0.89	
	<i>Sigmoid</i>	0.95	
60:40	Linear	0.95	0.9250
	RBF	0.87	
	<i>Polynomial</i>	0.94	
	<i>Sigmoid</i>	0.94	

Hasil menunjukkan bahwa kernel *linear* dan *sigmoid* pada skenario 70:30 memberikan performa terbaik dengan rata-rata AUC sebesar 0,9325. Di mana kernel *linear* mencapai nilai AUC 0,950 dan kernel *sigmoid* 0,944. Oleh karena itu, model dengan kernel *linear* dipilih untuk tahap *knowledge*, sebagaimana ditampilkan pada kurva ROC berikut.



Gambar 4.4 Grafik ROC

Berikut adalah rincian *confusion matrix* untuk kernel linear pada skenario 70:30, di mana model dilatih menggunakan *Random Search* CV menghasilkan nilai *accuracy* 88%, *precision* 90%, *recall* 83%, serta *F1-score* 87%.

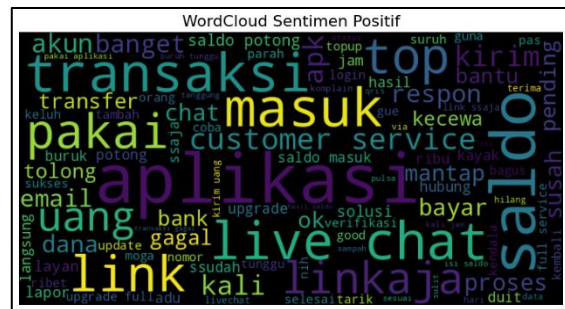
Tabel 4.10 Evaluasi *Confusion Matrix*

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	88%	90%	83%	87%
RBF	88%	89%	85%	87%
Polynomial	87%	88	83%	85%
Sigmoid	87%	90	81%	86%

4.6. Knowledge

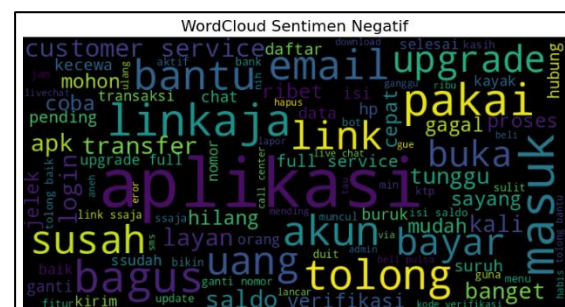
Visualisasi *word cloud* menampilkan kata-kata yang paling sering muncul dalam analisis sentimen pengguna, seperti "saldo", "transaksi", "masuk", "pakai", dan "chat". Kata

"saldo" mencerminkan kemudahan akses dan pengiriman dana, sedangkan "transaksi" mengindikasikan kelancaran proses pembayaran. Kata "masuk" mengacu pada kemudahan login, sementara "pakai" menunjukkan bahwa aplikasi dianggap praktis. Adapun kata "chat" menandakan bahwa fitur layanan pelanggan berjalan dengan baik.



Gambar 4.5 Hasil *Word Cloud* Positif

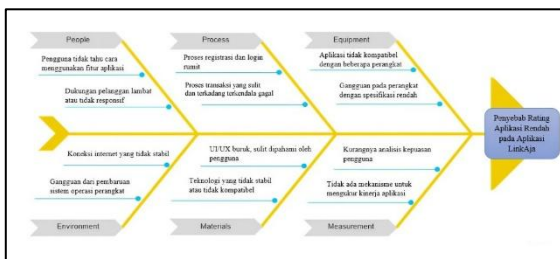
Word cloud sentimen negatif memperlihatkan kata-kata dominan seperti "aplikasi", "pakai", "upgrade", "susah", "tolong", "email", dan "uang", yang muncul dengan ukuran lebih besar menandakan frekuensi tinggi dalam ulasan negatif. Kata "aplikasi" mengindikasikan masalah teknis atau kinerja sistem yang tidak optimal, sementara "pakai" dan "susah" menunjukkan pengalaman pengguna yang tidak intuitif. Keluhan terhadap "upgrade" mencerminkan frustrasi pasca-pembaruan, dan "email" berkaitan dengan kendala verifikasi akun. Kata "uang" menyoroti masalah terkait transaksi, sedangkan "tolong" mengisyaratkan permintaan bantuan yang tidak ditangani dengan baik oleh layanan pelanggan.



Gambar 4.6 Hasil *Word Cloud* Negatif

Analisis penyebab rendahnya rating aplikasi LinkAja dilakukan melalui *Root Cause Analysis*. Hasil tersebut mencakup enam faktor utama: manusia, proses, perangkat, lingkungan, bahan, dan pengukuran. Faktor manusia

mencakup kurangnya tutorial dan lambatnya layanan pelanggan, sehingga perlu panduan interaktif dan peningkatan responsivitas. Pada perangkat, sering terjadi *crash* di ponsel berspesifikasi rendah, maka dibutuhkan optimasi dan peningkatan kompatibilitas. Lingkungan mencakup koneksi tidak stabil dan masalah *update* OS, sehingga solusi yang diusulkan adalah penyimpanan data *offline* dan notifikasi koneksi terputus. Dari sisi bahan, UI/UX yang membingungkan dan *bug* pada validasi saldo perlu diatasi dengan desain ulang antarmuka dan perbaikan sistem validasi. Terakhir, kurangnya pemantauan *bug* dan *feedback* diselesaikan melalui sistem pelaporan otomatis dan peningkatan respons terhadap masukan pengguna.



Gambar 4.7 Diagram Ishikawa

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, diantaranya:

- Penelitian ini berhasil membuktikan bahwa algoritma *Support Vector Machine* (SVM) dengan kernel linear, yang dioptimalkan menggunakan menerapkan *tuning hyperparameter* menggunakan *Random Search Cross-Validation* (CV), efektif dalam mengklasifikasikan sentimen dan mengoptimasi proses klasifikasi ulasan aplikasi LinkAja. Model terbaik diperoleh pada skenario pembagian data 70:30 menggunakan kernel linear dengan performa unggul, yaitu *accuracy* 89%, *precision* 90%, *recall* 83%, *F1-score* 87%.
- Hasil analisis sentimen dan evaluasi performa model berhasil digunakan untuk mengidentifikasi akar masalah rendahnya rating aplikasi LinkAja. Analisis menggunakan diagram Ishikawa menunjukkan bahwa faktor utama yang memengaruhi *rating* rendah meliputi

masalah teknis, kurangnya respons terhadap keluhan pengguna, dan keterbatasan mekanisme pengukuran kepuasan pengguna.

- Untuk penelitian selanjutnya, diharapkan *dataset* yang digunakan agar lebih besar untuk meningkatkan nilai keakurasian data, menggunakan *dataset* yang diperoleh dari *platform* atau media sosial lain, serta dapat menerapkan fitur seleksi dengan *information gain* untuk meningkatkan nilai akurasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Y. Faizal, O. K. Zaini, and Hasrul, "Analisis Faktor-Faktor Yang Menentukan Perilaku Konsumen Smartphone Xiaomi (Studi Pada Mahasiswa/I Fakultas Ekonomi Dan Bisnis Universitas Pakuan Tahun 2022)," *NAMARA: Jurnal Management Pratama*, vol. 1, no. 1, pp. 1–17, 2022.
- [2] B. J. Blažič and A. J. Blažič, "Overcoming the digital divide with a modern approach to learning digital skills for the elderly adults," *Educ Inf Technol (Dordr)*, vol. 25, no. 1, pp. 259–279, 2020, doi: 10.1007/s10639-019-09961-9.
- [3] Y. Sari and E. Winarni, "The Impact of Cryptocurrency and Electronic Money Use On The Circulating Money Supply and Monetary Stability in Indonesia Cryptocurrency in Indonesia," *Nomico Journal*, vol. 1, no. 8, pp. 78–93, 2024.
- [4] Harnelia and A. R. Saputra, "ANALISIS SENTIMEN REVIEW SKINCARE SKINTIFIC DENGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4095.
- [5] M. Isnain, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput Sci*, vol. 227, pp. 168–175, 2023, doi: 10.1016/j.procs.2023.10.514.
- [6] X. Zhang, *Encyclopedia of Machine Learning and Data Mining*, vol. 3, no. 1959. 2017. doi: 10.1007/978-1-4899-7687-1.
- [7] X.-D. Zhang, "Support Vector Machines (SVM)," *Gesture*, vol. 23, no. 6, pp. 349–361,

- 2001, doi: 10.1016/B978-0-12-804494-0.00015-2.
- [8] M. Isnani, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput Sci*, vol. 227, pp. 168–175, 2023, doi: 10.1016/j.procs.2023.10.514.
- [9] M. F. N. Fathoni, E. Y. Puspaningrum, and A. N. Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem : Jurnal Informatika dan Sains Teknologi*, vol. 2, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v2i3.112.
- [10] N. P. Wahyuningtyas, D. E. Ratnawati, and N. Y. Setiawan, "Root Cause Analysis (RCA) berbasis Sentimen menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Pengunjung Kolam Renang Brawijaya)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 5, pp. 2515–2520, 2023.
- [11] N. Viennot, E. Garcia, and J. Nieh, "A measurement study of Google Play," *Performance Evaluation Review*, vol. 42, no. 1, pp. 221–234, 2014, doi: 10.1145/2591971.2592003.
- [12] Fintek Karya Nusantara, "Tentang LinkAja," <https://www.linkaja.id/tentang>.
- [13] N. Gupta and R. Agrawal, *Application and techniques of opinion mining*. INC, 2020. doi: 10.1016/B978-0-12-818699-2.00001-9.
- [14] W. Astuti, R. Kurniawan, and Y. A. Wijaya, "Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3314–3319, 2024.
- [15] T. P. Utama and M. S. Haibuan, "Penerapan Algoritma Naïve Bayes Dan Forward Selection Untuk Prediksi Penyakit Stroke," *Jurnal Teknoinfo*, vol. 17, no. 2, pp. 351–357, 2023.
- [16] X. Zhang, *Encyclopedia of Machine Learning and Data Mining*, vol. 3, no. 1959. 2017. doi: 10.1007/978-1-4899-7687-1.
- [17] M. F. N. Fathoni, E. Y. Puspaningrum, and A. N. Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem : Jurnal Informatika dan Sains Teknologi*, vol. 2, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v2i3.112.
- [18] G. Damele, G. Bazzana, F. Andreis, F. Aquilio, S. Arnoldi, and E. Pessi, "Process improvement through Root Cause Analysis," *Achieving Quality in Software*, pp. 35–47, 1996, doi: 10.1007/978-0-387-34869-8_5.
- [19] K. Padmanandam, S. P. V. D. S. Bheri, L. H. Vegesna, and K. Sruthi, "A Speech Recognized Dynamic Word Cloud Visualization for Text Summarization," *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, pp. 609–613, 2021.
- [20] M. E. Davey and J. A. Morell, "Method for Using Rubric Ratings on Fishbone Diagrams to Compare Case Studies," *J Multidiscip Eval*, vol. 16, no. 34, pp. 56–64, 2020.