

ANALISIS SENTIMEN APLIKASI LINKAJA DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN RANDOM FOREST

Kaeren¹, Andrianingsih²

Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika
Universitas Nasional

Jl. Sawo Manila, Pejaten Ps. Minggu Jakarta, Indonesia

kaerenkaeren2021@student.unas.ac.id, andrianingsih@civitas.unas.ac.id

Abstrak

Dompot digital telah menjadi bagian penting dalam transaksi keuangan modern, dengan LinkAja sebagai salah satu penyedia layanan di Indonesia. Meskipun termasuk lima besar dompet digital populer, LinkAja memiliki jumlah ulasan tinggi di *Google Play Store*, tetapi proporsi pengguna dan rata-rata ratingnya lebih rendah dibandingkan dengan pesaingnya. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi LinkAja dengan mengklasifikasikan opini mereka ke dalam kategori positif dan negatif menggunakan algoritma Naïve Bayes dan Random Forest. Hasil penelitian menunjukkan bahwa mayoritas ulasan pengguna bersentimen negatif, dengan permasalahan utama meliputi kesulitan *login*, kendala transaksi, proses *upgrade* lambat, serta layanan pelanggan yang kurang responsif. Hal ini menyebabkan *rating* aplikasi LinkAja didominasi oleh skala terendah, yaitu 1. Banyaknya pengalaman kurang menyenangkan mendorong pengguna untuk aktif memberikan ulasan, yang sebagian besar berisi keluhan terhadap layanan aplikasi. Dalam proses klasifikasi, Random Forest menunjukkan performa lebih unggul dengan akurasi sebesar 82%, presisi 84%, *recall* 81%, dan *f1-score* 82%. Sedangkan Naïve Bayes memiliki nilai akurasi sebesar 79%, presisi 83%, *recall* 78%, dan *f1-score* 79%. Random Forest juga lebih efektif dalam mengidentifikasi sentimen negatif dibandingkan Naïve Bayes. Analisis lebih lanjut mengungkapkan bahwa fitur aplikasi dan kemudahan penggunaan menjadi faktor utama dalam ulasan positif, meskipun jumlahnya lebih sedikit dibandingkan ulasan negatif.

Kata Kunci : Analisis sentimen, Dompot digital, LinkAja, Naïve Bayes, Random Forest

Abstract

Digital wallets have become an important part of modern financial transactions, with LinkAja as one of the service providers in Indonesia. Even though it is among the top five popular digital wallets, LinkAja has a high number of reviews on the *Google Play Store*, but the proportion of users and average rating are lower compared to its competitors. This research aims to analyze the sentiment of user reviews of the LinkAja application by classifying their opinions into positive and negative categories using the Naïve Bayes and Random Forest algorithms. The research results show that the majority of user reviews are negative, with the main problems including difficulty logging in, transaction problems, slow upgrade processes, and less responsive customer service. This causes the LinkAja application rating to be dominated by the lowest scale, namely 1. The many unpleasant experiences encourage users to actively provide reviews, most of which contain complaints about the application service. In the classification process, Random Forest shows superior performance with an accuracy of 82%, precision of 84%, recall of 81%, and *f1-score* of 82%. Meanwhile, Naïve Bayes has an accuracy value of 79%, precision of 83%, recall of 78%, and *f1-score* of 79%. Random Forest is also more effective in identifying negative sentiment than Naïve Bayes. Further analysis revealed that app features and ease of use were the main factors in positive reviews, although they were fewer in number than negative reviews.

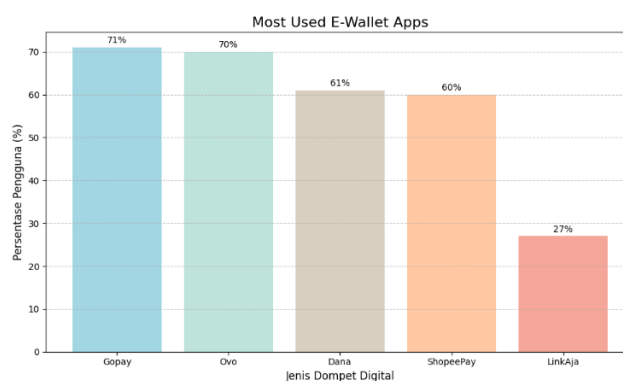
Keywords: Sentiment analysis, Digital wallet, LinkAja, Naïve Bayes, Random Forest

PENDAHULUAN

Dalam era digital yang terus berkembang, teknologi keuangan atau *financial technology (fintech)* menjadi pilar penting dalam mendorong inovasi ekonomi serta meningkatkan efisiensi, kelancaran, dan keandalan sistem pembayaran (Permana, 2021). Salah satu layanan *fintech* yang mendapatkan perhatian luas dan mengalami peningkatan dalam penggunaannya adalah dompet digital (*e-wallet*) (Yanti & Isnaeni, 2022). Dompot digital banyak digunakan masyarakat Indonesia karena

memberikan kenyamanan dan keamanan dalam berbagai transaksi harian serta mampu membantu mereka dalam manajemen keuangan (Putu et al., 2024). Salah satu aplikasi dompet digital yang populer di Indonesia adalah LinkAja.

LinkAja merupakan layanan penyedia jasa pembayaran yang dikembangkan oleh PT Fintek Karya Nusantara (Finarya) dan bekerja sama dengan aplikasi MyPertamina sebagai penyedia layanan untuk transaksi *e-payment* pada aplikasi tersebut (Raffi et al., 2023). Diantara lima jenis dompet digital populer lainnya seperti Dana, Gopay, Ovo, dan ShopeePay, aplikasi LinkAja memiliki rata-rata rating yang paling rendah serta jumlah proporsi pengguna yang paling sedikit diantara keempatnya. Informasi proporsi jumlah pengguna didasarkan pada hasil laporan survei mengenai dompet digital yang dilakukan oleh Insight Asia pada bulan September 2022 dalam risetnya yang berjudul “*Consistency That Leads: 2023 E-Wallet Industry Outlook*” melibatkan 1.300 responden yang dilakukan di beberapa daerah di Indonesia seperti Jabodetabek (Dhanesworo, 2022). Proporsi pengguna untuk masing-masing pengguna dompet digital pada hasil survei tersebut divisualisasikan pada *bar chart* dibawah ini:



Gambar 1. Persentase pengguna e-wallet

Berdasarkan persentase pengguna *e-wallet* hasil survei oleh Insight Asia, menunjukkan bahwa dari kelima jenis dompet digital yang terdiri dari Gopay, Ovo, Dana, ShopeePay, dan LinkAja, diketahui bahwa aplikasi LinkAja menempati posisi paling terakhir dengan jumlah proporsi pengguna yang paling rendah hanya mencapai 27%. Selain informasi proporsi pengguna oleh hasil survei Insight Asia, informasi proporsi rating untuk kelima jenis dompet digital berdasarkan data dari *Google Play Store*, menunjukkan bahwa aplikasi LinkAja memperoleh rata-rata rating yang paling rendah namun jumlah ulasannya lebih banyak dibandingkan dengan aplikasi yang memiliki jumlah pengguna dan rata-rata rating yang lebih tinggi. Hal ini mengindikasikan bahwa pengguna memiliki keterlibatan tinggi untuk menyampaikan pengalaman mereka, baik dalam bentuk pujian maupun keluhan. Kondisi ini memunculkan peluang untuk menggali lebih dalam pola sentimen pengguna dan faktor-faktor yang memengaruhi persepsi mereka terhadap aplikasi tersebut sehingga dapat dijadikan rekomendasi dan masukan bagi aplikasi LinkAja untuk memperbaiki layanannya sehingga dapat meningkatkan jumlah proporsi pengguna aplikasi tersebut.

Tujuan penelitian ini adalah menganalisis sentimen ulasan pengguna aplikasi LinkAja di *Google Play Store* menggunakan algoritma Naïve Bayes dan Random Forest. Analisis ini bertujuan memahami persepsi pengguna serta memberikan rekomendasi untuk meningkatkan kualitas layanan aplikasi. Naïve Bayes dipilih karena kemampuannya mengklasifikasikan teks dengan cepat dan stabil pada data sampel yang besar maupun kecil (Wandani, 2021). Sementara Random Forest dipilih karena algoritma ini unggul dalam menangani fitur kompleks serta mengurangi *overfitting* dan mampu menghasilkan akurasi yang bagus (Indrayanto et al., 2023). Dengan membandingkan kedua algoritma ini, penelitian juga diharapkan dapat mengidentifikasi metode yang lebih akurat dalam analisis sentimen ulasan aplikasi LinkAja.

PENELITIAN RELEVAN

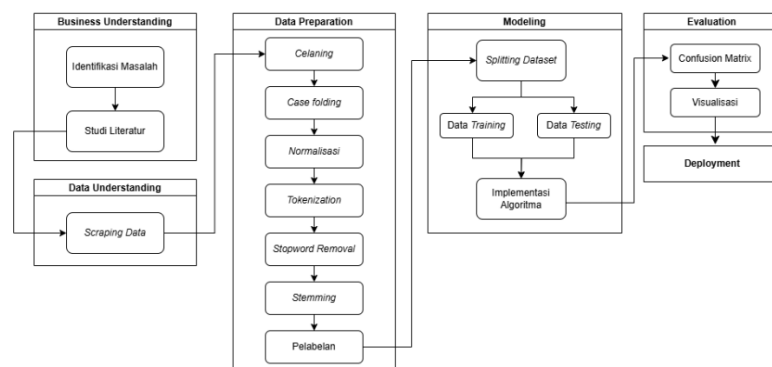
Penelitian sebelumnya yang dilakukan oleh Sitepu et al. (Sitepu et al., 2024) dalam penelitiannya yang berjudul “*Analisis Sentimen Terhadap Aplikasi LinkAja Menggunakan Metode Naïve Bayes Classifier*” menganalisis ulasan LinkAja di *Google Play Store* dengan 100 ulasan dan menghasilkan akurasi yang relatif rendah, yaitu 75% menggunakan Naïve Bayes. Namun, jumlah data yang digunakan masih terbatas, dan analisis hanya berfokus pada hasil akurasi algoritma. Penelitian lain oleh Turjaman & Budi (Turjaman & Budi, 2022) menggunakan algoritma SVM dalam menganalisis sentimen ulasan aplikasi LinkAja di Twitter dengan hasil akurasi sebesar 80%. Namun, opini di Twitter yang digunakan sebagai dataset cenderung bersifat umum dan kurang relevan dalam konteks aplikasi. Sementara itu, penelitian oleh Aryanti et al. (Aryanti et al., 2023) pada aplikasi Primaku menggunakan Random Forest menghasilkan nilai akurasi sebesar 88%, menunjukkan performa klasifikasi algoritma yang baik.

Oleh karena itu, penelitian ini akan menganalisis sentimen ulasan aplikasi LinkAja di *Google Play Store* dengan pendekatan yang lebih komprehensif. Berbeda dengan penelitian sebelumnya yang hanya berfokus pada akurasi algoritma dan menggunakan jumlah data terbatas, penelitian ini mengombinasikan dua algoritma yaitu Naïve Bayes dan Random Forest, dengan jumlah data yang lebih besar dan lebih relevan. Selain itu, penelitian ini tidak hanya menghasilkan akurasi yang lebih tinggi dibandingkan penelitian sebelumnya, tetapi juga memberikan analisis lebih mendalam terkait isu utama yang sering dihadapi pengguna. Hasil penelitian ini diharapkan dapat memberikan rekomendasi yang lebih bermanfaat bagi pengembang aplikasi dalam meningkatkan layanannya.

METODE PENELITIAN

Desain Penelitian

Penelitian ini dilakukan secara sistematis dengan menerapkan metode CRISP-DM untuk menggambarkan keseluruhan proses analisis. Menurut Singgalen, salah satu kerangka kerja yang efektif untuk diterapkan dalam proses data mining guna mendukung optimalisasi manajemen organisasi atau perusahaan adalah *Cross-Industry Standard Process for Data Mining (CRISP-DM)* (Singgalen, 2023). Setiap tahapan disusun secara terstruktur agar hasil yang diperoleh sesuai dengan tujuan penelitian. Adapun alur penelitian dapat dilihat pada ilustrasi berikut:



Gambar 2. Desain Penelitian

Desain penelitian dengan menggunakan metode CRISP-DM memiliki tahapan yang terdiri dari:

1. *Business Understanding*

Tahap *Business Understanding* peneliti berusaha memahami tujuan penelitian yang akan dilakukan. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi LinkAja berdasarkan ulasan di *Google Play Store* guna mengetahui persepsi mereka terhadap aplikasi tersebut. Selain itu, penelitian ini juga bertujuan untuk menentukan algoritma yang lebih efektif dalam analisis sentimen dengan membandingkan akurasi yang dihasilkan oleh algoritma Naive Bayes dan Random Forest.

2. **Data Understanding**

Pada tahap *data understanding*, peneliti mengidentifikasi serta menganalisis struktur dan karakteristik data yang digunakan dalam penelitian ini. Data ulasan dikumpulkan melalui teknik *scraping* dari *Google Play Store* sebagai sumber utama. Dataset yang dihasilkan mencakup berbagai atribut penting yang mendukung proses penelitian. Tahap *data understanding* juga mencakup visualisasi frekuensi kata yang menggambarkan kecenderungan kata-kata yang sering muncul di ulasan pengguna pada dataset sebelum dilakukan *preparation data*

3. **Data Preparation**

Pada tahap *Data Preparation*, peneliti akan melakukan preparasi dataset yang mencakup beberapa proses, seperti pembersihan data menggunakan teknik *preprocessing* serta pelabelan sentimen berbasis lexicon. Tahap *preprocessing* meliputi beberapa tahap antara lain: *Cleaning*, *Case Folding*, *Normalisasi*, *Tokenization*, *Stopword Removal*, dan *Stemming*.

4. **Modeling**

Pada tahap ini, peneliti berfokus pada pembangunan model menggunakan algoritma Naïve Bayes dan Random Forest. Sebelum proses klasifikasi, dataset yang telah melalui tahap *preprocessing* akan dibagi menjadi data latih (80%) dan data uji (20%). Setelah pembagian data, klasifikasi akan dilakukan menggunakan kedua algoritma tersebut untuk menganalisis performanya.

5. **Evaluation**

Pada tahap *evaluation*, akan dilakukan evaluasi dengan mengukur kinerja model yang telah dibangun pada tahap *modeling* menggunakan *confusion matrix* untuk mengetahui nilai akurasi, presisi, *recall*, dan *F1-Score* dari algoritma. Selain mengevaluasi kinerja algoritma, pada tahap *evaluation* juga mencakup visualisasi *wordcloud* untuk setiap kategori sentimen yaitu positif dan negatif, guna mengetahui kecenderungan ulasan pengguna berdasarkan kata-kata yang sering mereka ungkapkan di kolom komentar.

6. **Deployment**

Pada tahap ini, akan dijelaskan hasil penelitian dalam bentuk laporan dan mencakup sebuah *dashboard* interaktif yang mengimplementasikan hasil analisis sentiment menggunakan *framework streamlit*.

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas penerapan dan hasil dari setiap tahapan penelitian yang dilakukan dengan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Penjelasan lebih lanjut mengenai setiap tahap akan diuraikan dalam penjelasan berikut ini:

1. **Business Understanding**

Berdasarkan survei Insight Asia dan Populix, LinkAja termasuk dalam lima dompet digital populer di Indonesia, tetapi memiliki proporsi pengguna dan *rating* terendah di *Google Play Store*, meskipun jumlah ulasannya lebih banyak dibanding dompet digital lain. Hal ini menimbulkan pertanyaan mengenai kecenderungan sentimen pengguna, apakah lebih dominan positif atau negatif. Oleh karena itu, penelitian ini bertujuan menganalisis ulasan pengguna untuk mengidentifikasi permasalahan aplikasi menggunakan teknik analisis sentimen. Selain itu, penelitian ini juga bertujuan membandingkan akurasi algoritma Naïve Bayes dan Random Forest dalam klasifikasi sentimen. Hasilnya diharapkan dapat membantu meningkatkan kualitas layanan LinkAja menjadi lebih baik.

2. **Data Understanding**

Pada tahap ini akan dilakukan pemahaman dari dataset penelitian. Dataset ulasan dalam penelitian ini dikumpulkan menggunakan teknik *scraping* dengan memanfaatkan *library*

google-play-scraper di *google colab*. Proses *scraping* dilakukan dengan mengambil 2000 ulasan terbaru dari aplikasi *LinkAja* dengan ID 'com.telkom.mwallet' di *Google Play Store*.

```
[ ] #Import Library
from google_play_scraper import app

import pandas as pd
import numpy as np
import datetime

from google_play_scraper import reviews, Sort

app_id = 'com.telkom.mwallet'

def get_reviews(app_id, lang='id', count=2000, sort=Sort.NEWEST, filter_score_with=None, filter_device_with=None, continuation_token=None):
    try:
        result, continuation_token = reviews(
            app_id,
            lang=lang,
            country='id',
            sort=sort,
            count=count,
            filter_score_with=filter_score_with,
            filter_device_with=filter_device_with,
            continuation_token=continuation_token
        )
    except:
```

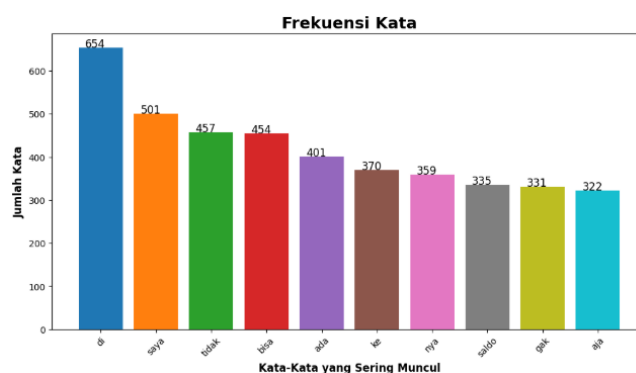
Gambar 3. Proses Scraping Data

Gambar 3 merupakan *script* yang digunakan untuk proses *scraping* data ulasan dari *google play store*. Data ulasan diambil sebanyak 2000 data terbaru berbahasa Indonesia. Contoh sampel data ulasan hasil *scraping* ditampilkan dalam sebuah dataframe, kemudian disimpan dalam format file csv. Gambar 4 dibawah ini merupakan contoh sampel data ulasan hasil *scraping*:

	Review ID	Username	Rating	Review Text	Date
0	da4d2137-ed0b-4f1d-b2d7-1fda2a724a8b	STEVEN DC	1	Customer servicenya lama merespon keluhan pela...	2024-11-04 05:42:26
1	cddb3244-e520-4ce2-b82e-ab05440a6d2a	Ahmad anis	1	Aplikasi gk jelas pusban chat robot .malez apk...	2024-11-04 05:24:17
2	da730ac9-c372-4b73-b6c3-2552e68f180a	Ali Machrus	3	Masukannya saat masukkan pin pembayaran harus...	2024-11-04 04:31:04
3	130c03ca-9395-4cd1-ae09-351f52b9e7af	Adi Vanda	1	Aplikasi sampah	2024-11-04 02:14:02
4	56b82bc7-38ec-4148-9f1d-4496cd838054	Zaenal Abidin	1	Sebaiknya tersedia fitur hapus Akun dan ganti ...	2024-11-04 02:10:55
5	8e151c1c-79b9-4e75-9c49-dd034f5c1d66	Putri Sulistian	1	Baru kali ini kecewa setelah memakai aplikasi ...	2024-11-04 00:39:51
6	7f0471b2-757a-49f3-b98a-170a19e065aa	Gerakan Milenial Pataba	5	Mntap	2024-11-03 23:48:20
7	465b5c39-195b-40fe-8dbc-0a0642752ad5	Pekanbaru Riau	5	Klo ganti nmr bisa gk	2024-11-03 15:49:50
8	eada6d20-9eae-4d6b-ab6d-dfb41f657939	Agi Hanafi	5	sangat bagus	2024-11-03 14:08:48
9	51c08787-67c0-4b90-bf6c-167736b9e57c	Fariz Hasbullah	1	Top up emoney 100k, ga bs diupdate alias ilang...	2024-11-03 13:09:08
10	38cec8a3-6824-4bfb-92ee-24be9f141403	Indah Sartika	1	App yg bermasalah udah motong trx tapi hasilnya...	2024-11-03 09:44:51

Gambar 4. Dataset Hasil Scraping

Gambar 4 mengenai sampel dataset hasil *scraping*, menunjukkan bahwa dataset terdiri dari beberapa atribut yang meliputi *Review ID*, *Username*, *Rating*, *Review Text* dan *Date*. Masing-masing atribut memiliki tipe data dan *value* yang berbeda.



Gambar 5. Frekuensi Kata

Gambar 5 menampilkan *bar chart* yang menunjukkan frekuensi kata dalam ulasan pengguna sebelum tahap *data preparation*. Kata dengan kemunculan tertinggi, seperti “di”, “saya”, dan “tidak”, sebagian besar merupakan *stopwords* yang kurang informatif dalam analisis. Namun, kata “saldo” dan “gak” yang cukup sering muncul mengindikasikan adanya

permasalahan, kemungkinan terkait kendala transaksi. Hal ini menunjukkan perlunya pembersihan data agar analisis lebih fokus pada kata-kata yang benar-benar mencerminkan pengalaman dan keluhan pengguna.

3. Data Preparation

Sebelum melakukan *preparation* data, peneliti terlebih dahulu memfilter dataset dengan hanya menggunakan dua atribut yaitu *Rating* dan *Text Review*. Atribut lainnya tidak digunakan karena tidak memberikan informasi langsung yang relevan mengenai sentimen pengguna (Nurzaman et al., 2024). Dataset juga akan dibersihkan dari data duplikat, sehingga dari 2000 data ulasan, tersisa 1859 data untuk proses analisis selanjutnya. Berikut adalah teknik *preprocessing* dan pelabelan yang dilakukan pada tahap *preparation*:

a. Cleaning

Pada tahap *cleaning*, dataset akan dibersihkan dari elemen yang tidak relevan seperti menghapus emoji, simbol, *mention*, dan atribut lain yang dapat mengganggu proses analisis.

Tabel 1. Tahap *Cleaning*

Data Ulasan	Hasil <i>Cleaning</i>
BUKAN LINKAJA SYARIAH BIASA. Klo ganti nmr bisa gk	BUKAN LINKAJA SYARIAH BIASA Klo ganti nmr bisa gk

b. Case Folding

Pada tahap ini, seluruh teks dalam data ulasan akan dikonversi ke huruf kecil (*lowercase*) guna menjaga konsistensi dan mempermudah proses analisis.

Tabel 2. Tahap *Case Folding*

Data Ulasan	Hasil <i>Case Folding</i>
BUKAN LINKAJA SYARIAH BIASA Klo ganti nmr bisa gk	bukan linkaja syariah biasa klo ganti nmr bisa gk

c. Normalisasi

Pada tahap ini, kata-kata tidak baku, singkatan, typo, dan slang dalam teks akan dikonversi ke bentuk baku untuk memastikan keselarasan dan konsistensi..

Tabel 3. Tahap Normalisasi

Data Ulasan	Hasil Normalisasi
bukan linkaja syariah biasa klo ganti nmr bisa gk	bukan linkaja syariah biasa kalau mengganti nomor bisa tidak

d. Tokenization

Tahap ini mencakup pemisahan pada teks menjadi unit-unit kecil yang disebut dengan token. Pemisahan ini dilakukan sesuai dengan penulisan spasi pada teks agar proses perhitungan frekuensi kata dapat dilakukan dengan mudah.

Tabel 4. Tahap *Tokenization*

Data Ulasan	Hasil <i>Tokenization</i>
bukan linkaja syariah biasa kalau mengganti nomor bisa tidak	[bukan, linkaja, syariah, biasa] [kalau, mengganti, nomor, bisa, tidak]

e. Stopword Removal

Pada tahap ini, kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dalam teks akan dihapus.

Tabel 5. Tahap *Stopword Removal*

Data Ulasan	Hasil <i>Stopword Removal</i>
[bukan, linkaja, syariah, biasa]	[linkaja, syariah]
[kalau, mengganti, nomor, bisa, tidak]	[mengganti, nomor, tidak]

f. *Stemming*

Pada tahap ini, akan dilakukan sebuah proses mengubah kata berimbuhan menjadi bentuk kata dasar. Misalnya “berlari” menjadi “lari”, “membaca” menjadi “baca” (Wiratama Putra et al., 2022)

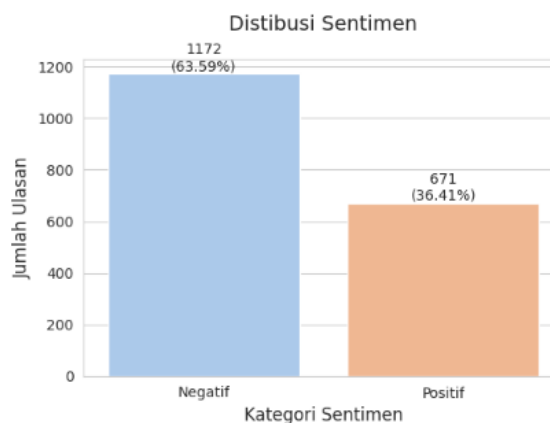
Tabel 6. Tahap *Stemming*

Data Ulasan	Hasil <i>Stemming</i>
[linkaja, syariah]	linkaja syariah
[mengganti, nomor, tidak]	[ganti, nomor, tidak]

Selain *preprocessing*, tahap *preparation* juga melibatkan pelabelan dataset. Kolom sentimen dikategorikan ke dalam dua label, yaitu positif dan negatif. Pelabelan ini dilakukan dengan menggunakan kamus *lexicon* yang diperoleh dari *Kaggle*. Sebelum diberi label, dataset hasil *preprocessing* akan dihapus dari data duplikat, sehingga data ulasan yang tersisa pada tahap ini berjumlah 1843. Hasil proses pelabelan akan disimpan dalam file dengan format csv.

tokenize	stopword removal	stemming_data	sentimen
['customer', 'servicenya', 'lama', 'merespon', '...]	['customer', 'servicenya', 'merespon', 'keluha...]	customer servicenya merespon keluh langgan	Positif
['aplikasi', 'tidak', 'jelas', 'pusban', 'chat...]	['aplikasi', 'tidak', 'pusban', 'chat', 'robot...]	aplikasi tidak pusban chat robot malas aplikas...	Negatif
['masukannya', 'saaf', 'masukkan', 'pin', 'pem...]	['masukannya', 'masukkan', 'pin', 'pembayaran'...]	masuk masuk pin bayar pakai sidik jari tulis p...	Negatif
['aplikasi', 'sampah']	['aplikasi', 'sampah']	aplikasi sampah	Negatif
['sebaiknya', 'tersedia', 'fitur', 'hapus', 'a...]	['tersedia', 'fitur', 'hapus', 'akun', 'mengga...]	sedia fitur hapus akun ganti nomor telepon bia...	Negatif
['baru', 'kali', 'ini', 'kecewa', 'setelah', '...]	['kali', 'kecewa', 'memakai', 'aplikasi', 'ber...]	kali kecewa pakai aplikasi bertahunnya transf...	Negatif
['mantap']	['mantap']	mantap	Positif

Gambar 6. Sampel *Labeling Dataset*



Gambar 7. Distribusi Sentimen

Hasil dari proses pelabelan pada dataset yang divisualisasikan pada Gambar 7 menunjukkan bahwa dari total 1843 data, distribusi untuk sentimen negatif dan positif adalah 1172 ulasan berlabel Negatif dan 671 ulasan berlabel Positif. Tingginya persentase sentimen negatif mencerminkan ketidakpuasan pengguna, yang mungkin disebabkan oleh masalah teknis, layanan pelanggan, atau fitur aplikasi. Temuan ini dapat menjadi acuan bagi pengembang untuk mengidentifikasi permasalahan dan meningkatkan kualitas aplikasi.

4. Modeling

Pada tahap *Modeling*, penelitian ini membangun model klasifikasi menggunakan data hasil *preparation*. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji untuk mengoptimalkan performa model. Pembangunan model ditunjukkan pada Gambar 10 dibawah

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

nb = MultinomialNB()
rf = RandomForestClassifier(random_state=42)

from sklearn.pipeline import Pipeline

model1 = Pipeline([('vectorizer', cvec)
                    ,('classifier', nb)])
model2 = Pipeline([('vectorizer', cvec)
                    ,('classifier', rf)])
```

Gambar 10. Modeling Algoritma

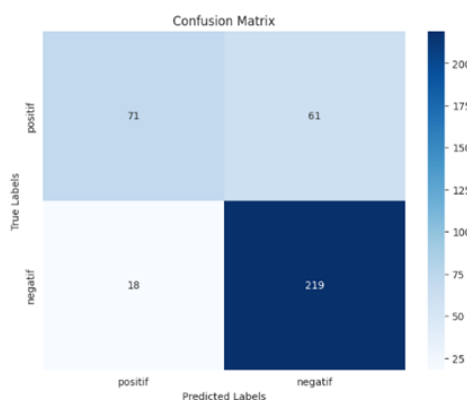
5. Evaluation

Pada tahap *Evaluation*, model yang telah dikembangkan di tahap *modeling* akan dievaluasi untuk mengukur kemampuannya dalam mengklasifikasikan ulasan pengguna. Penilaian dilakukan dengan menghitung metrik akurasi, presisi, *recall*, dan *F1-Score* guna membandingkan kinerja masing-masing algoritma.

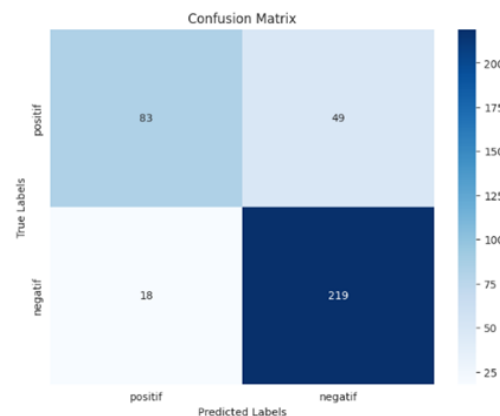
Tabel 7. Perbandingan Kalsifikasi Algoritma

Parameter	Naïve Bayes	Random Forest
Akurasi	0.785908	0.818428
Presisi	0.830909	0.843233
Recall	0.785908	0.818428
F1-Score	0.797833	0.824934

Berdasarkan tabel perbandingan algoritma, diketahui bahwa nilai presentase akurasi pada algoritma Random Forest lebih tinggi dibandingkan dengan Naive Bayes. Namun, kedua algoritma memiliki nilai presentase presisi yang sama-sama baik ditunjukkan dengan persentase mencapai 80% lebih. Selain itu, kinerja model algoritma pada penelitian ini juga diukur menggunakan *confusion matrix* untuk mengetahui performa prediksi model algoritma secara lebih rinci.



Gambar 11. Confusin Matrix Naive Bayes



Gambar 12. Confusion Matrix Random Forest

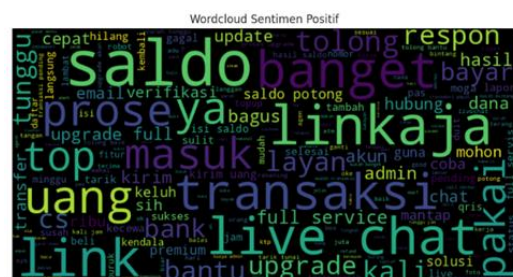
Berdasarkan *Confusion Matrix* dan tabel perbandingan klasifikasi, Random Forest menunjukkan performa lebih baik secara keseluruhan dibandingkan Naïve Bayes. Random

Forest berhasil memprediksi lebih banyak ulasan positif dengan jumlah kesalahan pada ulasan positif yang lebih sedikit. Keduanya memiliki jumlah kesalahan pada ulasan negatif (*False Positive*) yang sama. Hal ini menunjukkan bahwa Random Forest lebih unggul dalam mengenali ulasan positif sekaligus mempertahankan akurasi pada ulasan negatif.

Pada tahap *evaluation*, terdapat juga visualisasi *wordcloud* mengenai ulasan positif dan negatif yang menggambarkan kata-kata ulasan yang sering dibahas oleh pengguna. Berikut adalah *wordcloud* untuk masing-masing sentimen:



Gambar 13. *Wordcloud* Sentimen Negatif

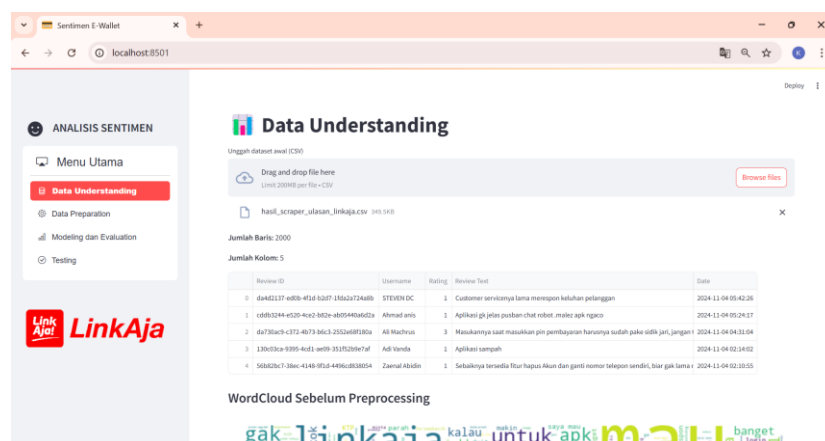


Gambar 14. *Wordcloud* Sentimen Positif

Permasalahan utama pada aplikasi LinkAja yang ditunjukkan oleh *wordcloud* sentimen negatif meliputi kesulitan akses akun (misalnya login), kendala transaksi (seperti gagal transfer atau pengelolaan saldo), serta kekecewaan terhadap fitur dan layanan, termasuk proses *upgrade* yang lambat. Pengguna juga mengeluhkan sistem yang dianggap ribet, buruk, dan lama, serta kurangnya layanan pelanggan yang memadai. Sedangkan *Wordcloud* sentimen positif menunjukkan pengguna mengapresiasi fitur aplikasi, kelancaran transaksi, kemudahan penggunaan, dan kecepatan layanan, meskipun beberapa aspek seperti saldo dan *live chat* juga muncul dalam sentimen negatif.

6. Deployment

Pada tahap *deployment*, hasil analisis sentimen aplikasi LinkAja yang telah dilakukan menggunakan algoritma Naïve Bayes dan Random Forest akan disajikan dalam bentuk visualisasi interaktif. *Dashboard* berbasis *streamlit* ini mencakup seluruh tahapan analisis untuk mempermudah interpretasi data. Berikut adalah tampilan *dashboard* tersebut:



Gambar 13. Dashboard Penelitian

Gambar 13 merupakan visualisasi dari tampilan *dashboard* yang terdiri dari empat menu utama: *Data Understanding*, *Data Preparation*, *Modeling & Evaluation*, dan *Testing*. Menu *Data Understanding* memungkinkan pengguna mengunggah dataset dan

menampilkan *wordcloud* dari ulasan pengguna. *Data Preparation* menampilkan dataset setelah *preprocessing*, termasuk *wordcloud* terbaru, daftar kata dengan frekuensi tertinggi, dan distribusi sentimen. *Modeling & Evaluation* mencakup pembagian data latih-uji serta evaluasi performa algoritma Naïve Bayes dan Random Forest menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix*. Terakhir, menu *Testing* memungkinkan pengguna menguji model dengan memasukkan teks untuk diprediksi sentimennya.

SIMPULAN

Pada penelitian ini, analisis sentimen terhadap ulasan pengguna aplikasi LinkAja dilakukan dengan menggunakan algoritma Naïve Bayes dan Random Forest. Hasil penelitian menunjukkan bahwa mayoritas ulasan pengguna terhadap aplikasi LinkAja bersentimen negatif, dengan keluhan utama terkait kesulitan *login*, kendala transaksi, proses *upgrade* yang lambat, sistem yang rumit, dan layanan pelanggan yang kurang responsif. Akibatnya, aplikasi LinkAja memperoleh banyak ulasan, tetapi didominasi oleh komentar negatif, yang berkontribusi pada rendahnya *rating* pengguna dibandingkan dengan dompet digital lainnya.

Dari hasil klasifikasi sentimen, algoritma Random Forest memiliki performa lebih baik dibandingkan Naïve Bayes, dengan akurasi 82% berbanding 79%, terutama dalam mengidentifikasi sentimen negatif. Untuk meningkatkan kepuasan pengguna, pengembang perlu memperbaiki keandalan sistem login, transaksi, dan layanan pelanggan. Penelitian selanjutnya dapat memperluas cakupan data, menggunakan metode pelabelan yang lebih bervariasi, serta mengeksplorasi algoritma lain seperti *Support Vector Machine* (SVM) atau model berbasis *deep learning* untuk meningkatkan akurasi klasifikasi

DAFTAR PUSTAKA

- Aryanti, R., Misriati, T., & Sagiyo, A. (2023). Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data. *Journal of Computer System and Informatics (JoSYC)*, 5(1), 218–227. <https://doi.org/10.47065/josyc.v5i1.4562>
- Dhanesworo, S. (2022). *Dompot Digital Makin Diminati*. Katadata.Co.Id. <https://katadata.co.id/infografik/639fe20e7f7e1/dompot-digital-makin-diminati>
- Indrayanto, C. G., Ratnawati, D. E., & Rahayudi, B. (2023). Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(3), 1131–1139. <http://j-ptiik.ub.ac.id>
- Nurzaman, N., Suarna, N., & Prihartono, W. (2024). Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 967–974. <https://doi.org/10.36040/jati.v8i1.8708>
- Permana, R. I. (2021). Analisis Faktor Penggunaan Dompet Digital di Kalangan Mahasiswa Perguruan Tinggi Surabaya. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 312–322. <https://doi.org/10.35957/jatisi.v8i1.584>
- Putu, N., Krismajayanti, A., Nurmallasari, M. R., Prawitasari, P. P., Dewiningrat, A. I., Ayu, I., Megawati, P., Kusnita, K. L., Diah, P., & Sanjiwani, A. (2024). Tren Revolusioner: Bagaimana E-Wallet Mengubah Konsumen di Era Modern? *Journal of Islamic Business Management Studies*, 5(1), 41–51.
- Raffi, M., Suharto, A., & Maulana, I. (2023). Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma Naïve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 6(1), 450–462. <https://doi.org/10.31539/intecom.v6i1.6117>
- Singgale, Y. A. (2023). Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM. *Building of Informatics, Technology and Science (BITS)*, 4(4), 1870–1882. <https://doi.org/10.47065/bits.v4i4.3199>
- Sitepu, M., Yohanna, M., & Manurung, S. V. B. H. (2024). *METODE NAÏVE BAYES CLASSIFIER*. 8(1), 44–50.
- Turjaman, R. M., & Budi, I. (2022). Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter). *Jurnal Darma Agung*, 30(2), 266. <https://doi.org/10.46930/ojsuda.v30i2.1672>
- Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 651–665.
- Wiratama Putra, T., Triayudi, A., & Andrianingsih, A. (2022). Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(1), 20–26. <https://doi.org/10.35870/jtik.v6i1.368>
- Yanti, L. R., & Isaeni, N. (2022). Analisis Faktor-Faktor Penggunaan Dompet Digital (E-Wallet) sebagai Alat Transaksi di Tinjau dari Perspektif Ekonomi Islam. *Journal of Islamic Economic and Finance*, 3(3), 157–167.