Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Analisis Sentimen Ulasan Pengguna Aplikasi M-Pajak di *Google Play Store* Menggunakan Algoritma *Naïve Bayes*

DOI: http://dx.doi.org/10.35889/jutisi.v15i1.3402

Nindita Nashwa Azahra[1*], Arochman[2], Bambang Ismanto[3]
Teknik Informatika, Institut Widya Pratama, Pekalongan, Indonesia
*e-mail *Corresponding Author:* ninditanashwa@gmail.com

Abstract

Digital-based public services require service quality evaluation to ensure user satisfaction, one way of doing this is through sentiment analysis of M-tax app reviews on the Google Play Store. The aim of this study is to test the tendency of user sentiment toward the M-Tax application by applying the naïve bayes algorithm. The dataset utilized in this research consists of 5,263 user reviews, which were processed through several preprocessing steps, including case folding, text cleaning, tokenization, stopword removal, and stemming. The analyzed variable is user review text classified into positive and negative sentiment categories. The naïve bayes algorithm was employed to perform sentiment classification. To assess the model performance, a confusion matrix was employed with an 80:20 split between train and testt data, along with k-fold cross-valiidation using k = 5. The findings show that negative sentiment dominates the reviews at 88.34%, while positive sentiment accounts for 11.66%. The model attained an accuracy of 92%, indicating that the Naïve Bayes algorithm performs efecctively in classifying user sentiment and can serve as a basis for evaluating improvements in digital taxation services.

Keywords: *Sentiment analysis; Naive Bayes; Application reviews; TF-IDF; M-Pajak*

Abstrak

Pelayanan publik berbasis digital menuntut adanya evaluasi kualitas layanan untuk memastikan kepuasan pengguna, salah satunya melalui analis sentiment ulasan aplikasi M-Pajak di *Google Play Store*. Adapun tujuan dari penelitian ini yaitu untuk mengkaji kecenderungan sentimen pada ulasan pengguna aplikasi M-Pajak di *Google Play Store* dengan menerapkan algoritma *Naïve Bayes*. Data yang dianalisis berjumlah 5.263 ulasan pengguna, yang diproses dengan tahapan *prerocessing* yaitu *case folding, cleaning text, tokeniation, stopword, dan stemming*. Variabel yang dianalisis berupa teks ulasan pengguna dengan kelas sentimen positif dan negatif. Klasifikasi dilakukan dengan menggunakan algoritma *Naïve Bayes*. Untuk menilai performa model, digunakan *confusiont matrix* sebagai alat evaluasi dengan pembagian data latih dan data sebesar 80:20 serta teknik *k-fold cross-validation* dengan nilai k = 5. Penelitian ini menunjukkan hasil bahwa sentimen negatif mendominasi sebesar 88,34%, sedangkan sentimen positif sebesar 11,66%. Model menghasilkan akurasi evaluasi sebesar 92%, sehingga algoritma *Naïve Bayes* dinilai tepat dan efisien untuk menganalisis sentimen pada ulasan aplikasi M-Pajak. dan dapat digunakan sebagai dasar evaluasi peningkatan layanan digital perpajakan.

Kata kunci: *Analisis sentimen; Naive Bayes; Ulasan aplikasi; TF-IDF; M-Pajak*

1. Pendahuluan

Pemanfaatan teknologi informasi pada sektor publik telah menjadi kebutuhan untuk meningkatkan efisiensi, transparasi, serta akuntabilitas dalam penyelenggaraan pelayanan publik [1]. Pelayanan publik yang efektif seharusnya tidak dipersulit dengan birokrasi yang rumit, karena kondisi tersebut dapat menimbulkan keengganan masyarakat dalam mengurus

layanan [2]. Keberhasilan implementasi layanan digital tidak hanya dipengaruhi oleh ketersediaan teknologi, tetapi juga oleh tingkat penerimaan dan kepuasan pengguna. Oleh karena itu, evaluasi terhadap layanan digital pemerintah menjadi hal yang penting agar dapat memenuhi kebutuhan dan masyarakat. Salah satu bentuk layanan digital pemerintah di bidang perpajakan adalah aplikasi M-Pajak yang dikembangkan oleh Direktorat Jenderal Pajak. Aplikasi ini dikembangkan untuk memudahkan masyarakat dalam mencari informasi dan melakukan administrasi perpajakan melalui perangkat seluler. Namun, berdasarkan ulasan pengguna di *Google Play Store* yang jumlahnya mencapai ribuan, aplikasi M-Pajak masih menerima tanggapan yang beragam. Ulasan yang diberikan pengguna pada *platform Google Play Store* memiliki peran penting dalam memengaruhi keputusan calon pengguna lain untuk mengunduh atau menggunakan suatu aplikasi [3]. Sejumlah pengguna menyampaikan keluhan terkait kendala teknis, seperti error sistem, proses login yang lambat, dan kurangnya responsivitas aplikasi, sementara pengguna lainnya memberikan penilaian positif terhadap kemudahan fitur yang tersedia. Variasi ulasan tersebut menunjukkan adanya perbedaan persepsi pengguna terhadap kualitas layanan digital M-Pajak.

Ulasan pengguna pada *platform* digital merupakan bahan informasi yang penting untuk memahami pengalaman dan persepsi masyarakat terhadap suatu layanan. Namun, jumlah ulasan yang besar dan berbentuk teks menyebabkan analisis secara manual menjadi tidak efisien dan berpotensi subjektif. Oleh karena itu diperlukan pendekatan berbasis komputasi melalui teknik *text mining*, khususnya analisis sentimen. Analisis sentimen diterapkan untuk mengidentifikasi opini pengguna ke kategori positif, negatif, atau netral, sehingga dapat dijadikan dasar dalam pengambilan keputusan strategis terkait pengembangan fitur dan peningkatan kepuasan pengguna [4]. Selain itu, analisis sentimen telah banyak dimanfaatkan di berbagai sektor, termasuk bisnis dan sosial, karena pendpat pengguna memiliki pengaruh yang signifikan terhadap pola perilaku serta kebijakan yang diambil oleh suatu organisasi [5]. Analisis sentimen dapat diterapkan melalui beragam teknik klasifikasi, salah satunya dengan memanfaatkan algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* meruoakan metode klasifikasi berbasis probabilitas yang sering digunakan dalam analisis teks karena memiliki struktur yang sederhana, proses komputasi yang efisien, serta kemampuan dalam mengolah data berukuran besar. Meskipun mengenakan anggapan independensi antar fitur, algoritma ini terbukti mampu memberikan hasil yang reliabel dalam berbagai penelitian analisis sentimen [6].

Berdasarkan permasalahan tersebut, tujuan dari penelitian ini yaitu untuk mengkaji kecenderungan sentmen pada ulasan pengguna aplikasi M-Pajak yang terdapat di *Google Play Store* dengan menerapkan algoritma *Naïve bayes*. Penelitian inii diharapkan dapat memberikan pemahaman mengenai persepsi pengguna terhadap aplikasi M-Pajak, sekaligus menunjukkan proporsi dominasi sentimeen positif dan negatif yang muncul dalam ulasan pengguna.

2. Tinjauan pustaka

Berbagai studi sebelumnya yang sesuai dengan topik penelitian ini ialah penelitian yang dilaksanakan oleh Susanto et al. yang mengkaji analisis sentimen ulasan pengguna aplikasi pajak online kendaraan (newsakpole) dengan menerapkan algoritma *naïve bayes*. Dataset yang dimanfaatkan terdiri atas 2.414 ulasan positif dan 2.395 ulasan negatif. Proses evaluasi model dilaksanakan dengan metode 10-fold *cross-validation*, dan hasilnya mengindikasikan bahwa algoritma yang diterapkan mampu mencapai nilai akurasi sebesar 88%. Temuan tersebut menyatakan bahwa algoritma *naïve bayes* mempunyai kinerja yang memadai dalam mengklasifikasikan sentimen pengguna terhadap layanan perpajakan berbasis digital [7].

Penelitian yang dilaksanakan oleh Larasati et al. yang membahas analisis sentimen terhadap ulasan pengguna aplikasi dompet digital Dana dengan menggunakan metode random forest. Data dikumpulkan menggunakan teknik *web scraping*, kemudian dilakukan *preprocessing* teks sebelum proses klasifikasi. Metode *Random Forest* diterapkan dalam penelitian ini untuk mengategorikan sentimen ulasan menjadi tiga kelompok, yaitu sentimen positif, negatif, dan netral. Variabel yang dianalisis berupa teks ulasan pengguna dan label sentimen. Evaluasi kinerja model dilaksanakan melalui metrik akurasi, presisi, *recall,* dan f-measure dengan membagi data latiih dan data uji sebanyakk 80:20. Temuan tersebut menunjukan bahwa metode *random forest* bisa memberikan performa yang cukp baik dengan akurasi 84% dalam mengklasifikasikan sentimen ulasan pengguna [8].

Penelitian yang dilaksanakan oleh Rahayu et al. yang membahas tentang Penerapan analisis sentimen juga dilakukan pada aplikasi teknologi finansial, salah satunya pada aplikasi

FLIP, Penelitian ini bertujuan untuk mengetahui kesesuaian antara ulasan pengguna dan rating tinggi yang diperoleh aplikasi tersebut. Data berupa teks ulasan diproses menggunakan tahapan text mining, selanjutnya diklasifikasikan dengan algoritma *K-Nearest Neighboor* (KNN). Variabel yang dianalisis meliputi teks ulasan pengguna dan kelas sentimen. Evaluasi performa dilakukan dengan membagi data latih dan data uji sebesar 80:20. Hasil penelitian ini mengindikasikan bahwa mayoritas ulasan pengguna bersentimen positif, dengan akurasi klasifikasi sebesar 76,68%, serta nilai presisi dan recall yang tergolong tinggi [9].

Kajian mengenai analisis sentimen pada aplikasi layanan perpajakan sebelumnya telah diterapkan pada aplikasi M-Pajak. Salah satu penelitian yang membahas topik tersebut dilakukan oleh Abdullah Faisol dengan menerapkan beberapa algoritma klasifikasi, yaitu *Support Vector Machine (SVM), Naive Bayes*, dan *K-Nearest Neighborrs* (KNN). Data yang dianalisis berupa teks ulasan pengguna yang diklasifikasikan ke dua kelas sentimen, yakni positif dan negatif. Proses evaluasi kinerja model dilakukan dengan menerapkan teknik validasi silang 10-fold menggunakan parameter dengan parameter pengukuran berupa akurasi, presisi, dan recall. Hasil dari penelitian ini mengindikasikan bahwa algoritma SVM memberikan performa paling optimal dibandingkan algoritma lainnya, sehingga dinilai efisien dalam klasifikasi sentimen ulasan pengguna aplikasi M-Pajak [10].

Sejumlah penelitian sebelumnya menunjukkan bahwa analisiis sentimen telah banyak diterapkan pada ulasan aplikasi layanan keuangan dan layanan publik dengan menerapkan berbagai metode klasifikasi, seperti *K-Nearest Neighbor, Support Vector Machinne, Random Forest*, serta pendekatan perbandingan beberapa algoritma. Kemudian pada penelitian yang dilakukan oleh Abdullah faisol juga membahas mengenai aplikasi M-Pajak, namun penelitian tersebut berfokus pada perbandingan kinerja beberapa algoritma klasifikasi untuk menentukan metode terbaik berdasarkan metrik evaluasi tertentu.

Berbeda dengan penelitian-penelitian tersebut, penelitian ini secara khusus berfokus pada analisis sentimen terhadap ulasan pengguna aplikasi M-Pajak dengan menerapkan satu algoritma yaitu naive bayes, tanpa melakukan perbandingan dengan metode lain. Pendekatan ini dipilih untuk memperoleh gambaran sentimen pengguna yang lebih terfokus dan konsisten terhadap layanan perpajakan digital yang disediakan. Dengan demikian, hasil analisis yang diperoleh diharapkan mampu merepresentasikan kondisi persepsi pengguna secara lebih akurat, sehingga hasil penelitian ini dapat memberikan pemetaan persepsi pengguna yang lebih jelas dan aplikatif sebagai bahan evaluasi peningkatan kualitas layanan.

3. Metodologi

Penelitian ini akan melalui 5 tahapan, yaitu pengumpulan data, pemberian label, *preprocessing* (*case folding*, *cleaning text*, *tokenizing*, *stopword*, dan *stemming*) untuk menyiapkan teks agar dapat diolah secara optimal, pembobotan, dan pengujian.

3.1 Pengumpulan data

Data yang di gunakan pada penelitiian ini yaitu ulasan pengguna terhadap aplikasi M-Pajak yang terdapatt di *Google Play Store*. Pengumpulan data dilaksanakan dengan metode *web scraping*, yaitu metode pengambilan dan ekstraksi informasi dari halaman web secara terstruktur menggunakan bantuan perangkat lunak [11].

Berdasarkan inforrmasi yang terdapat pada *Google Play Store*, aplikasi M-Pajak mempunyai ulasan sebanyak 10,3 ribu ulasan. Ulasan tersebut kemudian akan dilakukan *scarping* guna memperoleh data valid. Data yang dianggap valid pada penelitian ini adalah ulasan dari pengguna yang mencantumkan komentar secara utuh serta memberikan rating. Sementara itu, ulasan tanpa komentar tidak dimasukkan sebagai data dalam penelitian.

3.2 Pemberian label

Tahap selanjutnya adalah pelabelan data, yang dilakukan dengan meninjau isi komentar pengguna aplikasi. Pelabelan dilakukan untuk mengkategorikan teks sebagai sentimen positif maupun negatif [12].

Proses pelabelan sentimen dilakukan dengan mengacu pada penilaian yang diberikan pengguna pada *Google Play Store*. Ulasan yang memiliki nilai rating 4 dan 5 diklasifikasikan sebagai sentimen positif, sedangkan nilai rating 1 dan 2 dimasukkan ke kategori sentimen negatif. Sementara itu, ulasan dengan rating 3 tidak dipakai karena menunjukan kecenderungan sentimen netral.

3.3 *Preprocessing*

Pada tahap ini, data ulasan dipersiapkan melalui *preprocessing* data agar lebih konsisten dan siap dianalisis. *Preprocessing* yang dilakukan pada data yang diperoleh dari web berperan penting untuk meningkatkan kualitas hasil analisis sentimen, karena dapat memperbaiki akurasi dan meminimalkan kesalahan interpretasi polaritas. *Preprocessing* data dilakukan melalui beberapa tahapan [13]. Setiap tahapan dijelaskan secara rinci sebagai berikut:

1) *Case folding*

Pada proses *case folding*, keseluruhan karakter pada teks diubah ke dalam format huruf kecil untuk menghindari perbedaan makna akibat variasi penulisan huruf.

2) *Cleaning text*

Tahap *cleaning text* dilakukan untuk membersihkan karakter/simbol yang tidak sesuai dalam teks, seperti karakter non-alfabet, simbol, angka, serta tautan URL, sehingga teks menjadi lebih bersih dan siap diproses lebih lanjut.

3) *Tokenization*

Tokenization proses pemecahan suatu kalimat atau teks menjadi unit kata tunggal.

4) *Stopword*

Stopword yaitu mengeliminasi kata umum yang tidak mempunyai nilai informatif, seperti yang, dengan, itu.

5) *Stemming*

Stemming yaitu mengembalikan kalimat kebentuk dasarnya.

3.4 Pembobotan

Proses pembobotan dilaksanakan dengan menerapkan teknik *Term Frequency–Inverse Document Frequency* (TF-IDF). Setiap term diberikan bobot sesuai frekueensi kemunculannya di dokumen dan kontribusinya terhadap keseluruhan dataset, sehingga kata-kata yang kurang informatif dapat diminimalkan dan fitur yang lebih representatif terhadap sentimen dapat dipertahankan [14]. Perhitungan TF-IDF dapat dilihat pada rumus berikut:

$$TF(t,d) = \frac{f(t,d)}{\sum f(t,d)}$$ / (TF) = (1)

Inverse document frequency (IDF) = $$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$ (2)

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t)$$ (3)

Keterangan:

t: kata
d: dokumen ulasan
f(t, d): frekuensi kemunculan katta t pada dokumen d
N: jumlah keseluruhan dokumen
df(t): jumlah dokumen yang memuat kata t

3.5 Pengujian

Setelah melewati proses diatas, penelitian ini dilanjutkan ke tahap pengujian untuk memperoleh hasil klasifikasi ulasan pengguna aplikasi. Tahap pengujian mencakup dua proses utama, yaitu yaitu pengklasifikasian sentimen dengan algoritma *naïve bayes* dan evaluasi performa model yang digunakan dengan *confusion matrix*. Pada proses klasifikasi, digunakan algoritma *Naïve Bayes* karena metode ini dinilai efektif dan sesuai untuk menangani data berbasis teks [15]. Perhitungan *naïve bayes* dapat dilihat pada rumus berikut:

$$P(C_k \mid x) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i \mid C_k)}{P(x)}$$ (4)

Keterangan :

Ck : kelas sentimen ke-k (positif atau negatif)
x : dokumen ulasan
xi : fitur ke-i dalam dokumen
P(Ck): probabilitas awal (prior) kelas Ck

P(xi|Ck) : probabilitas kemunculan fitur xi pada kelas Ck

Pada penelitian ini, kinerja model dievaluasi menggunkan *confusion matrix* sebagai alat untuk mengukur performa klasifikasi. *Confusion matrix* menampilkan perbandingan antara data sebenarnya dan prediksi model, sehingga dilakukan perhitungan metrik evaluasi seperti akurasi, presisi, dan recall secara kuantitatif [16]. Perhitungan *confusion matriix* dapat dilihat pada rumus berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

Keterangan:
TP: jumlah data positif yang diprediksi positif
TN: jumlah data negatif yang diprediksi negatif
FP: jumlah data negatif yang diprediksi positif
FN: jumlah data positif yang diprediiksi negatif

4. Hasil dan Pembahasan

4.1 Pengumpulan data

Pengumpulan data dilaksanakan melalui teknik *web scraping* dengan memakai bahasa pemrograman python untuk mengekstraksi ulasan dari *platform Google Play Store*. Sebanyak 5.427 ulasan berhasil dikumpulkan. Data yang didapat mencakup komentar pengguna yang tidak kosong dan disertai penilaian aplikasi dalam rentang bintang 1–5. hasil pengumpulan data kemudian disimpan dalam bentuk file berformat .csv agar dapat diolah pada tahap berikutnya.

```
# === 1. Install library ===
!pip install google-play-scraper pandas

# === 2. Import library ===
from google_play_scraper import reviews, Sort
import pandas as pd
```

Gambar 1. Pustaka python yang digunakan untuk tahapan web scarping

Tabel 1. Dataset hasil scarping data

No.	Ulasan	Rating
1.	Aplikasi gak membantu banget. mau daftar, mau lapor...	1
2.	Menguras pulsa pendaftaran gagal terus muter muter	1
3.	Mantap	5
4.	kok prah si kode otp buat hp g muncul trus pas terakhi...	1
5.	buruk banget sih apk burik klo ga ngebantu ga usah …	1
…	…	…
5423.	Mudah banget sekarang	5
5424.	Waw keren	5
5425.	Mantab, ini yang resmi ya?	5
5426.	Buat billing jadi lebih mudah ðŸ˜□ semoga fitur fitur …	5
5427.	Cakepp	5

4.2 Pemberian label

Data yang diperoleh dari hasil *scarping* kemudian akan di lakukan proses pelabelan positif dan negatif, ulasan pengguna yang akan digunakan yaitu ulasan dengan nilai rating 4 dan 5 yang diklasifikasikan sebagai sentimen positif, sedangkan nilai rating 1 dan 2 dimasukkan

ke kategori sentimen negatif. Sementara itu, ulasan dengan rating 3 tidak digunakan karena cenderung mengarah pada sentimen netral. Maka dari 5427 data diperoleh 5263 data hasil dari pelabelan yang akan digunakan ke proses selanjutnya.

Tabel 2. Hasil dari pelabelan data

No.	Label	Jumlah
1.	Positif	4158
2.	Negatif	1105

4.3 Preprocessing

Tahap berikutnya yaitu *preprocessing*, meliputi proses case folding, tokenization, stoppword removal, dan stemming. Pada tahap ini, peneliti memanfaatkan pustaka python yaitu Sastrawi, untuk melakukan pembersihan dan standarisasi teks.

```
#  Install library yang dibutuhkan ===
!pip install Sastrawi pandas

#  Import library ===
import pandas as pd
import re
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
```

Gambar 2. Pustaka python yang digunakan dalam proses preprocessing

Tabel 3. Dataset yang telah melalui tahap preprocessing

No.	Ulasan	Sentimen	Rating
1.	aplikasi gak bantu banget mau daftar mau lapor pajak ...	Negatif	1
2.	uras pulsa daftar gagal terus muter muter	Negatif	1
3.	mantap	Positif	5
4.	kok prah si kode otp buat hp g muncul trus pas terakhi...	Negatif	1
5.	buruk banget sih apk burik klo ga ngebantu ga usah …	Negatif	1
…	…	…	…
5259.	mudah banget sekarang	Positif	5
5260.	waw keren	Positif	5
5262.	mantab yang resmi ya	Positif	5
5262.	Buat billing jadi lebih mudah moga fitur fitur lain sege …	Positif	5
5263.	cakep	Positif	5

4.4 Pembobotan

Setelah melewati tahap preprocessing akan dilanjutkan ke tahap pembobotan, Pada tahap ini, peneliti memanfaatkan pustaka python yaitu *TfidfVectorize*r dan *CountVectorizer*.

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
```

Gambar 3. Pustaka python yang digunakan dalam proses pembobotan

Tabel 4. Penggalan data hasil dari pembobotan

Kata	Idf_weights	Kata	Idf_weights
mpajak	7.749345	sesuai	4.976756
sabar	7.238519	otp	4.393610
uninstall	7.056198	bagus	4.133933
telat	6.902047	bantu	3.909893
rumit	6.768516	sulit	3.668987
kacau	6.496582	login	3.359020
lancar	5.957586	pajak	2.918368
lambat	5.649284	mau	2.836690
website	5.159078	aplikasi	2.059549

Dari total 5263 ulasan, dilakukan proses *preprocessing* yang mencakup pembersihan teks, normalisasi, dan penghapusan karakter yang tidak relevan. Pada tahap ini ditemukan bahwa 143 record memiliki nilai kosong (NaN) pada kolom hasil pembersihan teks (*cleaned*). *Record* tersebut tidak dapat diproses oleh TF-IDF karena tidak mengandung informasi tekstual yang valid, sehingga harus dihapus sebelum proses analisis sentimen dilakukan. Dengan demikian, jumlah data yang digunakan untuk proses klasifikasi adalah 5120 ulasan.

4.5 Pengujian

4.5.1 Klasifikasi menggunakan algoritma naïve bayes

Tahapan akhir dalam penelitian ini yaitu tahap pengujian. Pada tahap ini, peneliti memanfaatkan pustaka python *sklearnnaive bayes* yang memanfaatkan kelas MultinomialNB berfungsi untuk melakukan proses klasifikasi.

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
```

Gambar 4. Pustaka python yang digunakan dalam proses klasifikasi

```
=== JUMLAH ULASAN ===
predicted_sentiment
Negatif    4523
Positif     597
Name: count, dtype: int64

=== PERSENTASE ULASAN ===
predicted_sentiment
Negatif    88.34
Positif    11.66
Name: count, dtype: float64
```

Gambar 5. Hasil klasifikasi sentimen

Berdasarkan hasil klasifikasi sentimen terhadap seluruh ulasan pengguna, diperoleh dua kategori utama, yaitu sentimen positif dan sentimen negatif. Dari total 5120 ulasan yang berhasil diproses, sebanyak 4523 ulasan (88,34%) termasuk dalam kategori sentimen negatif, sedangkan 597 ulasan (11,66%) dikategorikan sebagai sentimen positif.

Tabel 5. Hasil klasifikasi

No.	Ulasan	Sentimen
1.	aplikasi gak bantu banget mau daftar mau lapor pajak ...	Negatif
2.	uras pulsa daftar gagal terus muter muter	Negatif
3.	mantap	Positif
4.	kok prah si kode otp buat hp g muncul trus pas terakhi...	Negatif

No.	Ulasan	Sentimen
5.	buruk banget sih apk burik klo ga ngebantu ga usah …	Negatif
…	…	…
5116.	mudah banget sekarang	Positif
5117.	waw keren	Positif
5118.	mantab yang resmi ya	Positif
5119.	Buat billing jadi lebih mudah moga fitur fitur lain sege …	Positif
5120.	cakep	Positif

4.5.2 Evaluasi model menggunakan confusion matrix

Pada tahap ini, peneliti memanfaatkan pustaka python yaitu sklearn.model selection berfungsi untuk proses pembagian data dan validasi model, kemudian proses klasifikasi dilakukan dengan memanfaatkan kelas MultinomialNB yang tersedia dalam pustaka sklearn.naive_bayes yang berfungsi untuk proses klasifikasi, sklearn.metrics yang berfungsi untuk evaluasi performa model, dan imblearn.over_sampling berfungsi untuk menangani ketidakseimbangan kelas (*imbalanced data*) dengan cara membentuk data sintetis untuk kelas minoritas.

```
import pandas as pd
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_score
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import classification_report, confusion_matrix
from imblearn.over_sampling import SMOTE
from sklearn.pipeline import Pipeline
import numpy as np
```

Gambar 6. Pustaka python dalam proses evaluasi model

```
              precision    recall  f1-score   support

     Negatif       0.94      0.96      0.95       830
     Positif       0.82      0.75      0.79       194

    accuracy                           0.92      1024
   macro avg       0.88      0.86      0.87      1024
weighted avg       0.92      0.92      0.92      1024

[[799  31]
 [ 48 146]]
F1 (macro) CV: mean=0.8769 std=0.0128
```

Gambar 7. Hasil evaluasi model

Dataset pada penelitian ini terdiri atas 5.120 data ulasan. Setelah dilakukan proses preprocessing, seluruh data yang valid tetap dapat digunakan untuk tahap pelatihan model. Pada tahap evaluasi, Dataset terbagi menjadi data latih dan data uji dengan menerapkan metode *train-test split* dengan perbandingan 80:20. Oleh karena itu, jumlah data uji pada *classification report* adalah sekitar 1.024 data, yang merupakan 20% dari keseluruhan dataset. Selain evaluasi menggunakan data uji, validasi model diterapkan melalui skema *k-fold cross-validation* dengan nilai k = 5. Teknik ini dilakukan untuk memastikan kestabilan kinerja model pada berbagai subset data dan meminimalkan risiko overfitting.

Berdasarkan hasil pengujian performa model klasifikasi, diperoleh nilai akurasi sebesar 92% yang menunjukkan bahwa model dapat mengklasifikasikan ulasan pengguna aplikasi M-Pajak dengan tingkat ketepatan yang tinggi. Nilai akurasi ini mengindikasikan bahwa pendekatan probabilistik yang sederhana seperti *Naïve Bayes* tetap efisen dalam menangani data teks ulasan layanan publik yang memiliki karakteristik bahasa informal dan beragam.

Jika ditinjau berdasarkan masing-masing kelas, model memperlihatkan kinerja yang sangat baik pada kelas sentimen negatif dengan tingkat nilai precision 94%, tingkat nilai *recall* 96%, dan *f1-score* sebesar 95%. Tingginya nilai r*ecall* pada kelas negatif menunjukkan bahwa

sebagian besar ulasan negatif berhasil teridentifikasi dengan benar oleh model. Hal ini dapat disebabkan oleh karakteristik ulasan negatif yang umumnya lebih eksplisit dalam menyampaikan keluhan, kritik, atau ketidakpuasan terhadap layanan, sehingga pola kata yang muncul lebih konsisten dan mudah dipelajari oleh model. Temuan ini sejalan dengan beberapa penelitian terdahulu pada analisis sentimen aplikasi layanan publik yang menyatakan bahwa sentimen negatif cenderung lebih mudah dikenali dibandingkan sentimen positif.

Sebaliknya, performa model pada kelas sentimen positif menunjukkan nilai yang lebih rendah, dengan tingkat nilai *precision* 82%, tingkat nilai r*ecall* 75%, dan *f1-score* sebesar 79%. Perbedaan performa ini menunjukkan adanya pengaruh ketidakseimbangan jumlah data, di mana ulasan negatif mendominasi dataset dibandingkan ulasan positif. Ketidakseimbangan data menyebabkan model lebih banyak mempelajari pola sentimen negatif, sehingga kemampuan dalam mengenali sentimen positif menjadi relatif lebih rendah. Fenomena ini juga ditemukan pada beberapa penelitian sebelumnya yang menggunakan dataset ulasan aplikasi keuangan dengan proposi kelas yg tidak seimbang.

Meskipun demikian, nilai macro F1-score sebesar 87% yang memperlihatkan bahwa model tetap memiliki performa yang cukup baik dalam mengklasifikasikan kedua kelas tanpa memperhatikan perbedaan proporsi jumlah data. Disisi lain, nilai *weighted F1-score* sebesar 92% menunjukan bahwa secara keseluruhan performa model tetap tinggi ketika mempertimbangkan distribusi data aktual.

4.6 Pembahasan

Berdasarkan pemaparan penelitian terdahulu, dapat diketahui bahwa penelitian ini memiliki keterkaitan langsung dengan studi yang dilakukan oleh Abdullah Faisol yang sama-sama mengkaji analisis sentimen pada aplikasi M-Pajak [10]. Perbedaan utama terletak pada pendekatan yang digunakan, di mana penelitian ini secara khusus memfokuskan pada penggunaan satu algoritma, yaitu *Naïve Bayes*, dengan tujuan mengevaluasi sejauh mana algoritma sederhana berbasis probabilistik mampu memberikan performa optimal pada data ulasan layanan publik. Hasil dari penelitian tersebut menunjukkan bahwa algoritma SVM memperoleh akurasi tertinggi sebesar 86,41%. Meskipun penelitian sebelumnya menyatakan algoritma SVM memiliki performa yang lebih unggul, algoritma *Naïve Bayes* dalam penelitian ini tetap mampu mencapai tingkat akurasi yang tinggi sebesar 92%. Temuan tersebut mengindikasikan bahwa *Naïve Bayes* masih relevan dan efektif dalam menganalisis sentimen ulasan aplikasii layanan publik, khususnya pada data yang mengandung bahasa informal seperti ulasan pengguna *Google Play Store*.

Kemudian penelitian yang mengkaji mengenai analisis sentimen pada ulasan aplikasi M-pajak juga pernah dilakukan dengan menggunakan metode *Knowledge Discoverry in Database* (KDD) dengan algoritma XGBoost untuk klasifikasinya [17]. Melalui metode KDD, algoritma XGBoost mampu menghasilkan tingkat akurasi yang cukup tinggi dalam mengklasifikasikan sentimen ulasan pengguna apliikasi M-Pajak di *Google Play Store*. Hal ini mengidentifikasikan bahwa pemanfaatan algoritma berbasis ensemble dapat memberikan performa yang bagus dalam proses klasifikasi teks. Jika dibandingkan dengan penelitian tersebut, pendekatan yang digunakan dalam penelitian ini relatif lebih sederhana karena tidak menggunakan kerangka KDD secara menyeluruh, melainkan berfokus pada tahapan *text preprocessing*, pembagian data menggunakan *cross validation*, serta proses klasifikasi menggunakan algoritma *Naïve Bayes*. Meskipun menggunakan pendekatan yang lebih sederhana, model yang dihasilkan dalam penelitian ini tetap mampu mencapai tingkat akurasi sebesar 92%. Hasil tersebut menunjukkan bahwa algoritma dengan kompleksitas yang lebih rendah tetap mampu menghasilkan kinerja yang baik dalam melakukan klasifikasi sentimen, khususnya pada data ulasan pengguna yang bersifat informal seperti pada *Google Play Store*.

Penelitian lain yang relevan dilakukan oleh Wahyudi Ariannor dkk yang membahas analisiis sentimen pengguna pada aplikasi layanan pajak kendaraan bermotor (info pajak kalsel) menggunakan algoritma Naïve Bayes. Hasil penelitian tersebut menunjukkan bahwa metode *Naïve Bayes* mampu digunakan secara efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi layanan perpajakan dengan nilai akurasi sebesar 92% dengan pembagian data 90:10 [18]. Temuan tersebut menunjukkan bahwa *Naïve Bayes* merupakan metode yang cukup efektif untuk melakukan klasifikasi sentimen pada data ulasan aplikasi berbasis teks. Hasil tersebut sejalan dengan penelitian ini yang juga memanfaatkan algoritma *Naïve Bayes* dan memperoleh tingkat akurasi yang lebih tinggi yaitu sebesar 92%, sehingga

semakin memperkuat bahwa metode ini masih relevan digunakan dalam analisis sentmen ulasan aplikasi pada platform *Google Play Store.*

Dengan demikian, penelitian ini memberikan kontribusi akademik berupa penguatan bukti empiris bahwa algoritma dengan kompleksitas yang lebih rendah tetap mampu menghasilkan performa yang kompetitif. Selain itu, penelitian ini juga memberikan implikasi praktis bagi pengembang sistem analisis sentimen untuk mempertimbangkan penggunaan metode yang lebih sederhana namun efisien secara komputasi dalam pengolahan data teks berskala besar.

5. Simpulan

Hasil dari analisis sentimen terhadap ulasan pengguna aplikasi M-Pajak menunjukan mayoritas ulasan bersifat negatif, yaitu sebesar 88,34%, sedangkan ulasan positif hanya mencapai 11,66%. Tingginya proposi sentimen negatif ini menunjukkan bahwa sebagian besar pengguna masih mengalami kendala atau ketidakpuasan dalam penggunaan aplikasi, sehingga temuan ini menjadi indikasi penting bagi pengembang untuk melakukan evaluasi serta peningkatan kualitas layanan.

Untuk memastikan keakuratan hasil, digunakan *confusion matrix* sebagai alat evaluasi dengan pembagian dataset menjdi data latih dan data uji sebesar 80:20 serta teknik *k-fold cross-validation* (k = 5). Model yang dibangun menunjukkan performa yang kuat dengan akurasi sebesar 92%. Pada kelas Negatif, model menghasilkan tingkat nilai precision 94%, tingkat nilai *recall* 96%, dan f1-score sebesar 95%, menandakan kemampuan yang sangat baik dalam mengenali ulasan negatif yang jumlahnya dominan. Sementara itu, performa pada kelas Positif lebih rendah (*precision* 82%, *recall* 75%, *f1-score* 79%) akibat ketidakseimbangan jumlah data antar kelas. *Macro F1-score* tercatat sebesar 87% dan weighted *F1-score* mencapai 92% semakin menunjukan bahwa model bekerja stabil pada berbagai skenario evaluasi.

Secara keseluruhan, temuan ini memperlihatkan bahwa naïve bayes merupakan metode yang tepat dan efiektif untuk melakukan analisis sentimen pada ulasan aplikasi pemerintah seperti M-Pajak. Selain memetakan persepsi pengguna secara objektif, model ini juga menjadi bahan masukan yang bernilai untuk pengembang dalam upaya meningkatkan kualitas layanan dan pengalaman pengguna di masa mendatang.

Daftar Referensi

[1] Rifdan, Haerul, H. Sakawati, and M. N. Yamin, “Analisis penerapan e-government dalam meningkatkan kualitas pelayanan publik di kecamatan tallo kota makassar,” *J. Gov. Polit.*, vol. 4, pp. 49–61, 2024.

[2] W. Ariannor, D. R. Safitri, and R. Rahmi, “Analisis Usability Sistem Permohonan Legalisir Menggunakan Metode System Usability Scale,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, pp. 415-426, 2023, doi: 10.35889/jutisi.v12i2.1122.

[3] S. A. Saputra, D. Rosiyadi, W. Gata, and S. M. Husain, “Google Play E-Wallet Sentiment Analysis Using Naive Bayes Algorithm Based on Particle Swarm Optimization,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019.

[4] Z. Rahman, P. Sakinah, Y. Hendra, B. Satria, F. Maulana, and A. Q. Ayun, “Sentiment Analysis of Gojek App Reviews on Google Play Store with Natural Language Processing Using Naive Bayes Algorithm,” *Jatilima J. Multimed. Dan Teknol. Inf.*, vol. 06, no. 03, pp. 60–69, 2024, [Online]. Available: https://journal.cattleyadf.org/index.php/jatilima/index

[5] B. Liu, *Sentiment Analysis and Opinion Mining*. Toronto: Springer Nature, 2022.

[6] S.N. Abdullah, "Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Naïve Bayes. Progresif: Jurnal Ilmiah Komputer, vol. 19, n0. 2, pp. 858-866, 2023. doi:http://dx.doi.org/10.35889/progresif.v19i2.1348

[7] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Sattriedi Wahyu Binabar, “Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 234–241, 2022, doi: 10.37859/coscitech.v3i3.4343.

[8] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” vol. 6, no. 9, pp. 4305–4313, 2022.

[9] S. Rahayu, Y. Mz, J. E. Bororing, and R. Hadiyat, “Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP,” Edumatic : Jurnal Pendidikan Informatika, vol. 6, no. 1, pp. 98–106, 2022,

doi: 10.29408/edumatic.v6i1.5433.
[10] A. Faisol, “Analisis Sentimen M-Pajak Menggunakan Algoritma Support Vector Machine, Naive Bayes dan KNN,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 7, no. 5, pp. 1231–1239, 2024, doi: 10.32672/jnkti.v7i5.8076.
[11] S. Satriajati, S. B. Panuntun, and S. Pramana, “Implementasi Web Scraping Dalam Pengumpulan Berita Kriminal Pada Masa Pandemi Covid-19,” *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 300–308, 2021, doi: 10.34123/semnasoffstat.v2020i1.578.
[12] A. Alkabkabi and M. Taileb, “Ensemble Learning Sentiment Classification for Un-labeled Arabic Text,” *Commun. Comput. Inf. Sci.*, vol. 1097 CCIS, no. January, pp. 203–210, 2019, doi: 10.1007/978-3-030-36365-9_17.
[13] W. Ariannor, S. M. A. B. Alshalwi, and B. Susarianto, “Sentiment Analysis of Netizens on Constitutional Court Rulings in the 2024 Presidential Election,” *IJIE (Indonesian J. Informatics Educ.*, vol. 8, no. 2, p. 90-102, 2024, doi: 10.20961/ijie.v8i2.94614.
[14] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, “Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes,” *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.
[15] Rina Noviana and Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
[16] F. M. Amin, “Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial,” *J. Eng. Res.*, vol. 6, no. 5, pp. 1-12, 2022.
[17] J. Sistem, “Analisis Sentimen Ulasan Aplikasi M-Pajak pada Google Play Store,” Jurnal Sistem Informasi dan Aplikasi, vol. 3, no. 1, pp. 11–25, 2025.
[18] W. Ariannor, E. A. Kusuma, F. Fadilah, and M. Arsyad, “Analyzing User Sentiments in Motor Vehicle Tax Applications Using the Naïve Bayes Algorithm,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, pp. 91-101, 2024, doi: 10.35889/progresif.v20i1.1694.