Artikel Ilmiah.docx

by 1 1

Submission date: 26-Jan-2026 03:13AM (UTC-0500)
Submission ID: 2843262783
File name: Artikel_Ilmiah.docx (833.78K)
Word count: 4309
Character count: 26502

Sentiment Analysis of Public Reviews of the Brompit Application Using Naive Bayes and SVM Algorithms
[Analisis Sentimen Ulasan Masyarakat Aplikasi Brompit Menggunakan Algoritma Naive Bayes dan SVM]

Egha Arya Affandi [*,1)], Sumarno [2)], Uce Indahyanti [3)], Yulian Findawati [4)]

[1,2,3,4)] *Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia*
*Email Penulis Korespondensi: 221080200077@umsida.ac.id

Abstract. *BromPit application was developed to facilitate Honda motorcycle owners in booking service appointments, tracking vehicle status, and locating authorized service centers; however, user reviews on the Google Play Store indicate varied responses. This study focuses on sentiment analysis of BromPit application usage and compares the effectiveness of the Naive Bayes and Support Vector Machine methods in classifying textual data. The research procedure includes review data collection and labeling, preprocessing, training and testing data splitting, TF-IDF weighting, the application of SMOTE, and accuracy evaluation. The results show that positive reviews dominate with 7,157 entries, followed by 1,351 negative and 221 neutral reviews. Model testing indicates that Naive Bayes achieves an accuracy of 82.59%–83.05%, while SVM attains 81.90%–82.51%, indicating that Naive Bayes is more effective. These findings can be utilized to improve application features, services, and overall user experience.*

Keywords; *Sentiment, BromPit, Naive Bayes, SMOTE, SVM*

Abstrak. *Aplikasi BromPit dikembangkan untuk memudahkan pemilik sepeda motor Honda dalam melakukan pemesanan servis, pelacakan status kendaraan, dan pencarian bengkel resmi, namun ulasan pengguna pada Google Play Store menunjukkan beragam tanggapan. Studi ini berfokus pada analisis sentimen penggunaan aplikasi BromPit serta membandingkan efektivitas metode Naive Bayes dan Support Vector Machine dalam mengkategorikan data teks. Prosedur penelitian meliputi pengumpulan dan pelabelan data ulasan, preprocessing, pembagian data latih dan uji, pembobotan TF-IDF, penerapan SMOTE, serta evaluasi akurasi. Hasil penelitian menunjukkan bahwa ulasan positif mendominasi sebanyak 7.157 data, diikuti 1.351 negatif dan 221 netral. Pengujian model memperlihatkan akurasi Naive Bayes sebesar 82,59%–83,05% dan SVM 81,90%–82,51%, sehingga Naive Bayes dinilai lebih efektif. Temuan ini dapat dimanfaatkan untuk meningkatkan fitur, layanan, dan pengalaman pengguna aplikasi BromPit.*

Kata Kunci; *Sentimen, BromPit, Naive Bayes, SMOTE, SVM*

I. PENDAHULUAN

Tuntutan masyarakat memerlukan kebutuhan transportasi yang mampu mendukung aktivitas harian secara cepat dan efisien, mengingat padatnya rutinitas yang dijalani. Kebutuhan transportasi tidak hanya meliputi kendaraan yang digunakan untuk perjalanan masyarakat umum seperti bus kota, kereta api, dan taksi, serta alat transportasi pribadi termasuk mobil dan sepeda motor[1]. Sepeda motor menjadi pilihan utama bagi sebagian besar penduduk, karena dianggap lebih praktis dan hemat biaya. Kondisi tersebut terjadi karena keterbatasan sarana transportasi umum yang memadai bagi mobilitas masyaraka[2]. Kendaraan pribadi menjadi moda transportasi yang paling banyak digunakan oleh masyarakat dalam beraktivitas[3].

Merek Honda adalah sebagai salah satu produsen yang paling dipercaya dan diminati. Mengenai tingkat keunggulan produk mereka, Honda memiliki reputasi yang kuat untuk sepeda motor yang tidak menggunakan banyak bahan bakar, mesin yang dapat diandalkan dan suku cadang pengganti yang tahan lama[4]. Honda terus melakukan berbagai inovasi, salah satunya melalui aplikasi *BromPit*, yaitu aplikasi khusus yang diperuntukkan bagi pemilik sepeda motor Honda. Aplikasi ini dirancang untuk mempermudah pengguna dalam mengurus berbagai keperluan terkait kendaraannya, seperti pemesanan layanan servis di bengkel resmi AHASS, pelacakan status pemesanan kendaraan, pencarian lokasi bengkel dan diler terdekat, serta perawatan kendaraan. Menjaga sepeda motor dalam kondisi prima dan memastikannya bertahan lama memerlukan perhatian konsisten terhadap perbaikan, perawatan rutin, dan penggantian komponen yang aus tepat waktu[5].

Teknologi informasi menawarkan potensi untuk menyederhanakan tugas sehari-hari di banyak hal, termasuk pendidikan, ekonomi, interaksi sosial, budaya, dan banyak bidang lainnya[6]. Meskipun aplikasi *BromPit* memberikan kemudahan, tidak semua pengguna merasa puas. Berdasarkan ulasan di *Google Play Store*, terdapat tanggapan beragam, mulai dari pujian hingga keluhan. Evaluasi-evaluasi ini berfungsi lebih dari sekadar masukan, evaluasi-evaluasi ini juga dapat bertindak sebagai repositori data untuk memberikan penjelasan tambahan[7]. Beberapa keluhan mencakup tampilan aplikasi yang membingungkan, gangguan teknis, dan layanan yang belum optimal. Salah satu

keluhan yang sering muncul adalah jadwal pemesanan servis yang tidak sesuai, seperti jadwal yang tidak tersedia, kesalahan sistem, atau informasi bengkel yang tidak akurat.

Untuk memahami pandangan masyarakat secara menyeluruh, diperlukan suatu pendekatan yang dapat mengolah dan menganalisis tanggapan pengguna dalam bentuk teks. Salah satu pendekatan yang sesuai adalah analisis tren opini, yang mencoba mengidentifikasi kategori ulasan dari pengguna dikelompokkan ke dalam kategori positif, netral atau negatif. Dalam melakukan klasifikasi ini, diperlukan metode pengolahan data yang tepat yang nanti akan dipakai adalah algoritma *Naive Bayes*. Dibuktikan bahwa metode *Naive Bayes* memiliki kecepatan dan presisi yang cukup tinggi ketika digunakan pada kumpulan data yang besar[8]. *Support Vector Machine*, sebuah teknik pembelajaran mesin yang mengkategorikan data dengan menetapkan garis atau *hyperplane* yang memaksimalkan jarak antar kelas[9]. Kedua algoritma memiliki keunggulan berbeda, seperti *Naive Bayes* yang efisien secara komputasi dan SVM unggul untuk data kompleks.

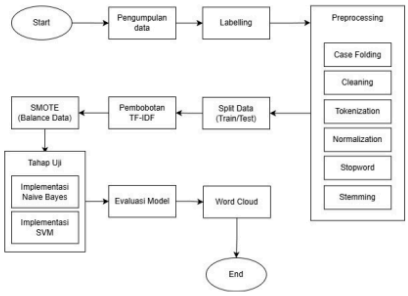
Penelitian untuk mengkategorikan pendapat pengguna *Twitter* ke dalam kelompok berdasarkan apakah sentimen negatif, positif dan netral. Prosesnya melibatkan pengambilan data, pelabelan, *preprocessing* teks, pembobotan pakai *TF-IDF*, dan pengukuran melalui *confusion matrix*. Temuannya menunjukkan bahwa pendekatan *Naive Bayes* memiliki kapasitas untuk mengkategorikan sentimen secara tepat sampai akurasi 79% pada Honda PCX, serta 72% pada Yamaha N-Max, sehingga efektif digunakan dalam analisis sentimen di media sosial[10].

Penelitian oleh Rifqi Yusril Muslikhin dan Ika Ratna Indra Astutik pada umpan balik pengguna untuk aplikasi BromPit dibagi ke dalam dua kategori utama yaitu positif dan negatif. Data penelitian yang digunakan terdiri atas 3.359 ulasan yang berasal dari platform *Google Play Store*. Teknik SVM dimanfaatkan untuk melakukan klasifikasi menggunakan kernel RBF, serta validasi dengan teknik *k-fold cross validation*. Hasil paling baik dicapai pada iterasi ke-8 memperoleh tingkat akurasi mencapai 87,6%, yang menunjukkan bahwa algoritma ini mampu menghasilkan performa yang tinggi dalam analisis sentimen pada ulasan aplikasi *BromPit*[11].

Hasil dari berbagai penelitian sebelumnya mengindikasikan bahwa penggunaan algoritma *Naive Bayes* memberikan hasil yang optimal dalam klasifikasi sentimen pengguna terhadap layanan dan produk, seperti terlihat pada studi terkait ulasan sepeda motor Honda dan Yamaha di *Twitter*, serta dalam penelitian mengkaji komentar yang disampaikan oleh pengguna aplikasi BromPit yang ada pada *Google Play Store* dengan menerapkan algoritma SVM. Penelitian tersebut menunjukkan akurasi tinggi. Namun belum mengevaluasi atau membandingkan efektivitas metode lain seperti *Naive bayes*, yang banyak dikenal karena kesederhanaan dan efisiensinya dalam klasifikasi data teks. Maka belum terdapat penelitian sehubungan dengan implementasi metode *Naive Bayes* yang dipakai untuk menilai sentimen publik terkait dengan aplikasi ini yaitu BromPit sebagai layanan digital resmi dari Honda untuk pemesanan servis kendaraan.

Tujuan dari studi ini adalah menganalisis kecenderungan opini publik terhadap layanan yang disediakan oleh aplikasi BromPit berdasarkan ulasan pengguna yang dikumpulkan dari *Google Play Store*, serta menilai tingkat akurasi algoritma *Naive Bayes* dan *Support Vector Machine* dalam melakukan klasifikasi sentimen umpan balik pengguna. Meliputi proses pengumpulan ulasan, pelabelan sentimen, preprocessing, representasi fitur menggunakan pembobotan *TF-IDF*, penanganan ketidakseimbangan kelas melalui metode SMOTE, pemisahan dataset ke dalam data pelatihan dan pengujian, serta pengukuran kinerja kedua model.

II. METODE



Gambar 1. Diagram tahapan penelitan

Delapan langkah diambil pada penelitian berikut, demi memperoleh hasil yang diinginkan, sebagaimana disajikan pada Gambar 1.

2.1 Pengumpulan data

Akuisisi data yang dilakukan melibatkan pengumpulan umpan balik pengguna khususnya dari aplikasi *BromPit*, yang dapat diakses melalui *platform* pada *Google Play Store* atau pada tautan https://play.google.com/store/apps/details?id=com.mpm.brompit, berjumlah 8.729 data. Data dikumpulkan dengan bantuan pustaka *Python* seperti *google-play-scraper*, dengan rentang waktu ulasan aplikasi pada Mei 2019 sampai Desember 2025. Atribut yang dipakai mencakup komentar, rating dan tanggal yang ditunjukkan pada Tabel 1 di bawah, serta atribut yang diproses yaitu komentar dan rating. Komentar berperan sebagai data utama yang diolah untuk menghasilkan fitur teks, sedangkan rating digunakan sejak tahap labeling sebagai label sentimen dan pada tahap evaluasi guna menilai performa algoritma *Naive Bayes* dan juga algoritma SVM.

Tabel 1. Contoh data ulasan aplikasi *BromPit*

Rating	Komentar	Tanggal
5	pelayanan sangat memuaskan	2025-10-30 03:30:56
3	tiga dlu ya msih prcobaan	2019-05-01 04:43:55
1	apalah makin kesini makin lemot	2025-10-30 03:07:54

2.2 *Labelling*

Tahap *labelling* merupakan langkah pelabelan data ke dalam kategori sentimen pada data ulasan yang telah dikumpulkan, sebagai dasar dalam proses klasifikasi. Dalam penelitian ini, proses *labelling* dilakukan secara otomatis berdasarkan nilai rating yang diungkapkan pengguna melalui aplikasi *BromPit* pada *platform Google Play Store*. Komentar yang mendapat skor 4 atau 5 dianggap memiliki nada emosi positif, sedangkan skor 1 atau 2 menunjukkan nada emosi negatif, dan skor 3 dipahami menunjukkan tidak adanya emosi yang kuat atau dianggap netral sebagaimana tercantum pada Tabel 2 di bawah.

Tabel 2. Data setelah pelabelan

Rating	Komentar	Label
5	pelayanan sangat memuaskan	positif
3	tiga dlu ya msih prcobaan	netral
1	apalah makin kesini makin lemot	negatif

2.3 *Preprocessing*

a. *Case Folding*

Langkah awal ini bertujuan mengganti seluruh huruf pada ulasan jadi kecil semua (*lowercase*). Tujuannya adalah menyamakan istilah yang memiliki arti sepadan, namun ditulis dengan kapitalisasi berbeda, bisa seperti kata "Bagus", "BAGUS", dan "bagus", agar dikenali sebagai satu kata yang serupa.

b. *Cleaning*

Setelah *case folding*, proses pembersihan data melibatkan penghapusan elemen-elemen yang tidak perlu. Ini termasuk penghapusan digit numerik, representasi simbolis, tanda baca, emotikon, alamat web, dan berbagai karakter unik yang tidak memiliki arti penting dalam konteks mengenali nada emosional dalam teks.

c. *Tokenization*

Kalimat-kalimat dibedah dalam prosedur ini, membaginya menjadi komponen-komponen kata individual yang disebut token, yang memungkinkan sistem untuk membedakan setiap kata sebagai entitas yang berbed. Contohnya, kalimat yakni "Aplikasi motor terbaiki" akan dipecah menjadi ["aplikasi", "motor", "terbaik"].

d. *Normalization*

Pada tahapan ini memodifikasi istilah atau ungkapan tidak baku menjadi versi baku, bahasa formal yang dapat dikenali. Misalnya, kata "gk", "ga", atau "nggak" akan dinormalisasi menjadi "tidak", sehingga meningkatkan akurasi dalam pemahaman teks oleh sistem.

e. *Stopword*

Stopword adalah kata-kata umum yang sering muncul dalam teks, namun sebenarnya tidak membantu dalam penilaian, yakni "yang", "di", "ke", dan lain-lain. Kata-kata ini akan dihapus agar hanya informasi yang relevan yang dianalisis lebih lanjut.

f. *Stemming*

Melalui proses ini, kata-kata ditransformasikan ke bentuk dasar. Misalnya, kata "membantu", "dibantu", atau "membantumu" akan di-*stemming* menjadi "bantu". Dengan *stemming*, kata-kata dengan makna serupa akan disederhanakan ke satu bentuk dasar, sehingga mengurangi redundansi fitur dalam proses analisis. Tabel 3 menunjukkan enam tahap *preprocessing* data teks yang dilakukan agar data siap diolah lebih lanjut.

Tabel 3. Data setelah *preprocessing*

Tahap	Komentar
Data Awal	Krn harus pake aplikasi ini yg Gambang trouble JD males servis ke dealer.
Case Folding	krn harus pake aplikasi ini yg gambang trouble jd males servis ke dealer.
Cleaning	krn harus pake aplikasi ini yg gambang trouble jd males servis ke dealer
Tokenization	[krn, harus, pake, aplikasi, ini, yg, gambang, trouble, jd, males, servis, ke, dealer]
Normalization	[karena, harus, pakai, aplikasi, ini, yang, gampang, bermasalah, jadi, malas, servis, ke, dealer]
Stopword Removal	[harus, pakai, aplikasi, gampang, bermasalah, malas, servis, dealer]
Stemming	[harus, pakai, aplikasi, gampang, masalah, malas, servis, dealer]

2.4 *Split data* (*train*/ *test*)

Tahap Fase selanjutnya melibatkan pembagian dataset menjadi sepasang segmen yang berbeda, yaitu set pelatihan (data latih) dan set pengujian (data uji). Bagian ini bertujuan dalam membangun model pengklasifikasian dengan memanfaatkan kumpulan data pelatihan untuk penilaian seberapa baik kinerja model tersebut dilakukan pada data uji yang belum digunakan pada tahap pelatihan, dengan tiga rasio pembagian data, yakni 70%–30%, 80%–20%, dan 90%–10% untuk data latih dan data uji.

2.5 Pembobotan *TF-IDF*

Setelah *preprocessing*, data teks diubah menjadi bentuk numerik agar bisa diklasifikasikan. Metode yang dipakai yakni *TF-IDF*, merupakan teknik yang memberi bobot pada kata menyesuaikan frekuensi dan signifikansi kata tersebut dalam konteks keseluruhan dokumen. *TF-IDF* terdapat tiga unsur utama:

Term Frequency (TF): Mengukur seberapa sering kata

$$TF(t,d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Keterangan :

$f_{t,d}$: frekuensi term (t) dalam dokumen (d)

$\sum_k f_{k,d}$: banyaknya seluruh kata di suatu dokumen (d)

Inverse Document Frequency (IDF): Ini menilai signifikansi kata dengan mempertimbangkan banyaknya suatu dokumen di mana kata tersebut itu muncul.

$$IDF(t) = \log\left(\frac{N}{n_t}\right) \quad (2)$$

N : jumlah dokumen secara keseluruhan dalam korpus

n_t : jumlah dokumen yang mencakup kata tertentu (t)

TF-IDF: Perhitungan pembobotan kata dilakukan dengan mengombinasikan nilai TF dan IDF melalui operasi perkalian.

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

Hasil kalkulasi TF-IDF ini adalah vektor numerik atau serangkaian angka yang menunjukkan seberapa penting kata-kata tertentu dalam satu dokumen jika dibandingkan dengan semua dokumen yang tersedia. Vektor inilah yang digunakan sebagai input pada tahapan klasifikasi memakai algoritma *Naive Bayes* dan SVM.

2.6 *SMOTE* (*balance data*)

Untuk mengatasi masalah distribusi data yang tidak merata di antara berbagai kategori sentimen, digunakan Teknik *SMOTE*. Teknik ini menambah data pada kelas yang memiliki proporsi data lebih kecil membuat sampel sintetis baru, bukan menyalin data yang sudah ada. Sampel tersebut dibentuk melalui perhitungan titik di antara data minoritas dan beberapa tetangga terdekatnya. Distribusi data menjadi lebih proporsional sehingga algoritma klasifikasi dapat mempelajari setiap kelas sentimen secara lebih optimal.

2.7 Implementasi *Naive Bayes*

Penentuan sentimen ulasan melalui penggunaan teknik *Naive Bayes*, yang bergantung pada informasi tekstual yang telah melalui beberapa tahapan pemrosesan dan diubah menjadi bentuk *vektor TF-IDF*.
Formulasi matematis dari *Teorema Bayes* tercantum di bawah ini:

$$P(C_i \mid x) = \frac{P(x \mid C_i) \cdot P(C_i)}{P(x)} \quad (4)$$

Keterangan:
$P(C_i \mid x)$: Probabilitas bersyarat dokumen x berada pada kelas C_i (*posterior probability*)
$P(x \mid C_i)$: Probabilitas bersyarat fitur x terjadi dalam kelas C_i (*likelihood*)
$P(C_i)$: Probabilitas prior kelas C_i (*prior*)
$P(x)$: Probabilitas keseluruhan fitur x di semua kelas (*evidence*)
Kelas probabilitas tertinggi akan dijadikan hasil klasifikasi akhir.

2.8 Implementasi SVM

Penentuan sentimen ulasan juga dilakukan memakai SVM dengan mengandalkan hasil pembobotan *TF-IDF*. SVM bekerja dengan mengidentifikasi garis pemisah terbaik yang memisahkan data ke dalam kelompok kelas yang berbeda dengan menciptakan celah jarak maksimum.
Fungsi keputusan SVM dinyatakan sebagai:

$$f(x) = w \cdot x + b \quad (5)$$

Keterangan:
w :vektor bobot yang menentukan posisi dan arah bidang pemisah
x :vektor fitur dari data input (hasil TF-IDF)
b :konstanta bias yang berfungsi menggeser *hyperplane*

Tujuan utama SVM adalah meminimalkan fungsi berikut agar diperoleh margin maksimum antar kelas:

$$\min_{w,b} \frac{1}{2}|w|^2 \text{ dengan syarat } y_i(w^T x_i + b) \geq 1 \quad (6)$$

Keterangan:
y_i : label kelas yang dimiliki data ke-i (positif, negatif, atau netral)
x_i : vektor fitur yang merepresentasikan data ke-i

2.9 Evaluasi model

Tujuan evaluasi model adalah guna mengukur tingkat kinerja model *Naive Bayes* dan SVM dapat memprediksi hasil dengan membandingkan prediksinya dengan label awal yang sebenarnya.

Accuracy adalah metrik yang mengukur prediksi akurat relatif terhadap total prediksi diukur dengan metrik ini. Metrik ini memberikan gambaran umum tentang efektivitas keseluruhan model pengklasifikasian.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision menggambarkan akurasi kinerja model dalam prediksi kelas positif, dengan kata lain, dari seluruh data yang diprediksi positif, berapa yang benar-benar positif.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall menggambarkan tingkat keberhasilan model dalam mengenali seluruh data diklasifikasikan sebagai positif. Artinya, dari keseluruhan data positif yang memang ada, berapa banyak yang berhasil dikenali model.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

F1-Score merupakan rerata harmonik dari *Precision* dan juga *Recall*, sehingga cocok dipakai ketika keduanya sama-sama penting, khususnya saat data tidak seimbang (misalnya: *fraud detection*).

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

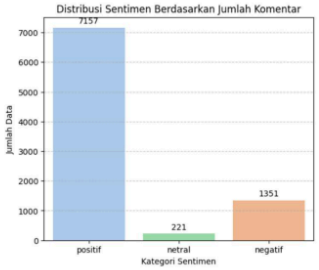
Keterangan :
TP = prediksi yang benar-benar sesuai dengan kelas positif.
TN = prediksi yang benar-benar sesuai dengan kelas negatif.
FP = kesalahan memprediksi negatif sebagai positif.
FN = kesalahan memprediksi positif sebagai negatif

2.10 *Word cloud*

Word Cloud adalah menyediakan gambaran grafis, menampilkan istilah yang paling sering muncul dalam kumpulan informasi tekstual. Kata-kata ditampilkan dalam berbagai ukuran, di mana semakin tinggi frekuensi sebuah kata, semakin besar pula ukuran tampilannya.

III. HASIL DAN PEMBAHASAN

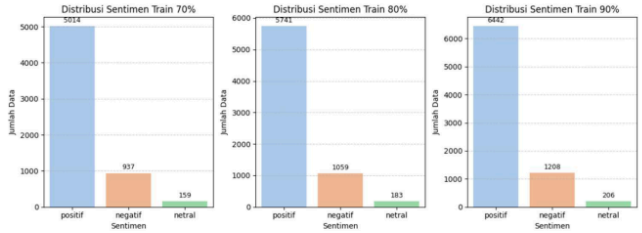
Klasifikasi sentimen terhadap ulasan pengguna pada aplikasi *BromPit* yang ditunjukkan pada Gambar 2, sebanyak 7.157 ulasan mencerminkan opini yang positif, sementara sejumlah besar ulasan, yaitu 1.351, menunjukkan opini yang negatif, dan 221 ulasan lainnya dikategorikan sebagai netral. Dominasi sentimen positif tingkat kepuasan pengguna yang tinggi terkait fungsionalitas aplikasi, sedangkan adanya ulasan negatif mengindikasikan bahwa masih terdapat pengguna yang mengalami kendala atau ketidakpuasan dalam penggunaan aplikasi.



Gambar 2. Diagram kategori sentimen

3.1 Skenario 1 (tanpa *SMOTE*)

Hasil pembagian data latih tanpa *SMOTE* yang tersaji pada Gambar 3, informasi yang menunjukkan bahwa di setiap perbandingan data, umpan balik positif jauh lebih banyak daripada umpan balik negatif dan netral. Pada *split* 70%, kelas positif berjumlah 5.014 data, sedangkan negatif hanya 937 dan netral 159. Sama pola data pada split 80% dan 90%, di mana kelas positif tetap mendominasi, kondisi tersebut menunjukkan bahwa data tidak seimbang.



Gambar 3. Distribusi sentimen (tanpa *SMOTE*)

Berdasarkan Tabel 4, 5, dan 6, model *Naive Bayes* menunjukkan akurasi yang tinggi pada kelas positif yakni *precision* 90–91%, *recall* 99%, dan *F1-score* 94–95% di semua rasio data. Kelas negatif memiliki performa sedang, yakni *precision* 86–89%, *recall* 56–62%, serta juga *F1-score* mencapai 68–73%. Namun, tidak ada prediksi yang berhasil untuk kelas netral, sehingga *precision*, *recall*, dan *F1-score* adalah 0%, mengartikan model tersebut kesulitan mengenali kelas minoritas ini akibat ketidakseimbangan data. Hal ini mengindikasikan perlunya penyeimbangan data.

Tabel 4. NB kelas positif (tanpa *SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	90%	99%	94%
80 : 20	91%	99%	95%
90 : 10	91%	99%	94%

Tabel 5. NB kelas netral (tanpa *SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	0%	0%	0%
80 : 20	0%	0%	0%
90 : 10	0%	0%	0%

Tabel 6. NB kelas negatif (tanpa *SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	86%	56%	68%
80 : 20	88%	61%	72%
90 : 10	89%	62%	73%

Berdasarkan Tabel 7, 8, dan 9, model SVM menunjukkan kinerja tinggi pada kelas positif yakni *precision* 94%, *recall* 97–98%, dan juga *F1-score* 96% di semua rasio data. Kelas negatif memiliki kinerja sedang, dengan *precision* 81–82%, *recall* 75–78%, dan *F1-score* 78–80%. Sama seperti pada *Naive Bayes*, model tidak berhasil mengenali kelas netral sehingga nilai *precision* 0% begitu juga nilai pada *recall* dan *F1-score*. Menekankan pentingnya penerapan metode penyeimbangan data, seperti *SMOTE*, agar semua kelas sentimen dapat dikenali.

Tabel 7. SVM kelas positif (tanpa *SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	***Percentage Recall***	***Percentage F1-Score***
70 : 30	94%	98%	96%
80 : 20	94%	97%	96%
90 : 10	94%	97%	96%

Tabel 8. SVM kelas netral (tanpa *SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	***Percentage Recall***	***Percentage F1-Score***
70 : 30	0%	0%	0%
80 : 20	0%	0%	0%
90 : 10	0%	0%	0%

Tabel 9. SVM kelas negatif (tanpa *SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	***Percentage Recall***	***Percentage F1-Score***
70 : 30	81%	75%	78%
80 : 20	81%	78%	80%
90 : 10	82%	76%	79%

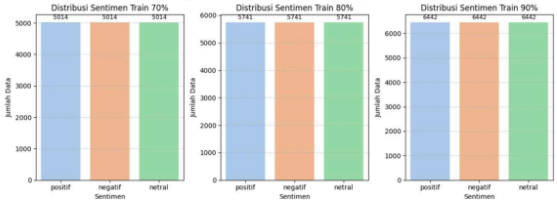
Sesuai data yang disajikan pada Tabel 10, akurasi SVM sedikit melebihi *Naive Bayes* pada tiap rasio pembagian data. Pada split 70:30, akurasi SVM mencapai 91,91% sementara *Naive Bayes* 89,84%. Pola serupa juga di rasio 80:20 dan 90:10, dengan SVM berada di kisaran 92,10–92,15% dan *Naive Bayes* 90,44–91,18%. Meskipun akurasi terlihat tinggi, model masih belum bisa mengenali kelas netral, sehingga nilai precision juga masih 0%, serupa dengan nilai *recall*, dan *F1-score* untuk kelas, yang menunjukkan bahwa performa model pada kelas minoritas belum optimal.

Tabel 10. Akurasi (tanpa *SMOTE*)

Rasio Split Data (%) (Train : Test)	*Accuracy*	
	Naive Bayes	*Support Vector Machine*
70 : 30	89,84%	91,91%
80 : 20	90,44%	92,15%
90 : 10	91,18%	92,10%

3.2 Skenario 2 (*SMOTE*)

Berdasarkan Gambar 4, penerapan *SMOTE* berhasil menyamakan rasio ulasan di setiap kelas sentimen. Pada *split* 70%, ketiga kelas memiliki jumlah yang serupa, yaitu 5.014, sehingga rasio antar kelas seimbang. Hal yang sama terlihat pada split 80% dan 90%, dengan masing-masing kelas memiliki jumlah identik 5.741 dan 6.442.



Gambar 4. Distribusi sentimen (*SMOTE*)

Berdasarkan Tabel 11,12, dan 13, penerapan *SMOTE* pada model *Naive Bayes* meningkatkan kemampuan model dalam mengenali kelas netral, dengan *precision* 4–8% dan *F1-score* 6–12%. Kelas positif mempertahankan performa tinggi yakni *precision* 97–98%, *recall* 84%, dan juga *F1-score* 90–91%, sedangkan kelas negatif menunjukkan peningkatan *recall* (82–86%) dengan *precision* 56–59%. Hal ini menunjukkan bahwa penyeimbangan data melalui *SMOTE* membantu model belajar dari semua kelas, terutama kelas minoritas.

Tabel 11. NB kelas positif (*SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	98%	84%	91%
80 : 20	98%	84%	90%
90 : 10	97%	84%	90%

Tabel 12. NB kelas netral (*SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	8%	21%	12%
80 : 20	7%	18%	10%
90 : 10	4%	13%	6%

Tabel 13. NB kelas negatif (*SMOTE*)

Rasio Split Data (%)	*Naive Bayes*		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	56%	82%	67%
80 : 20	57%	85%	68%
90 : 10	59%	86%	70%

Berdasarkan Tabel 14, 15, dan 16, penerapan *SMOTE* pada model SVM meningkatkan kemampuan model dalam memprediksi kelas netral, dengan *precision* 2–6% dan *F1-score* 4–9%. Kelas positif mempertahankan performa tinggi dengan *precision* 96–97%, *recall* 84–86%, dan *F1-score* 90–91%, sementara kelas negatif menunjukkan kinerja sedang hingga baik dengan *precision* 73–75%, *recall* 75–78%, dan *F1-score* 74–75%, menunjukkan bahwa *SMOTE* membantu menyeimbangkan pembelajaran model di seluruh kelas, sehingga prediksi menjadi lebih merata.

Tabel 14. SVM kelas positif (*SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	97%	86%	91%
80 : 20	97%	85%	90%
90 : 10	96%	84%	90%

Tabel 15. SVM kelas netral (*SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	6%	27%	9%
80 : 20	5%	26%	9%
90 : 10	2%	13%	4%

Tabel 16. SVM kelas negatif (*SMOTE*)

Rasio Split Data (%)	**SVM**		
	Percentage Precision	*Percentage Recall*	*Percentage F1-Score*
70 : 30	73%	75%	74%
80 : 20	73%	78%	75%
90 : 10	75%	76%	75%

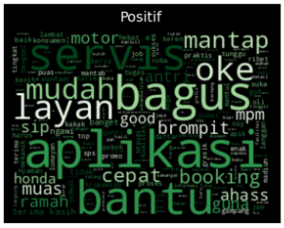
Berdasarkan Tabel 17, setelah penerapan *SMOTE*, akurasi kedua model sedikit menurun dibandingkan tanpa *SMOTE*, dengan *Naive Bayes* berada di kisaran 82,59–83,05% dan SVM 81,90–82,51%. Penurunan ini terjadi karena *SMOTE* menyeimbangkan data, sehingga model belajar dari semua kelas secara merata, termasuk kelas minoritas netral yang sebelumnya sulit dikenali oleh model. Meskipun akurasi keseluruhan sedikit menurun, model kini dapat memprediksi semua kelas sentimen dengan lebih optimal.

Tabel 17. Akurasi (*SMOTE*)

Rasio Split Data (%) (Train : Test)	*Accuracy*	
	Naive Bayes	*Support Vector Machine*
70 : 30	82,63%	82,51%
80 : 20	82,59%	82,36%
90 : 10	83,05%	81,90%

3.3 *Word cloud*

Pada kelas positif (Gambar 5), kata "bagus", "bantu", dan "mudah" mengindikasikan bahwa pengguna merasa aplikasi BromPit mempermudah proses servis Honda, memberikan layanan cepat, dan membantu mereka melakukan booking dengan nyaman.



Gambar 5. Word cloud positif

Pada kelas netral (Gambar 6), kata-kata seperti "baru", "error", dan "tolong" menandakan pengalaman yang bercampur, beberapa pengguna menganggap aplikasi bermanfaat, namun ada masalah teknis atau fitur yang belum optimal, sehingga respons mereka netral.

Gambar 6. Word cloud netral

Pada kelas negatif (Gambar 7), kata-kata seperti "lambat", "ribet", dan "harus" menunjukkan keluhan pengguna mengenai kesulitan penggunaan aplikasi, proses booking yang lama, dan fitur yang terasa membingungkan atau merepotkan.

Gambar 7. Word cloud negatif

IV. SIMPULAN

Penelitian ini menunjukkan bahwa ulasan pengguna aplikasi *BromPit* didominasi sentimen positif sebanyak 7.157 ulasan, sedangkan negatif 1.351 ulasan dan netral 221 ulasan. Temuan ini memperlihatkan bahwa sebagian besar pengguna puas saat menggunakan aplikasi. Setelah penerapan *SMOTE*, akurasi *Naive Bayes* berada di kisaran 82,59%–83,05% dan SVM 81,90%–82,51% untuk *split data* 70:30, 80:20, dan 90:10. *Naive Bayes* menjadi model terbaik, performa sedikit lebih tinggi dibanding SVM. *SMOTE* membantu model menangani kelas minoritas sehingga prediksi menjadi lebih merata. Kata-kata dominan dalam ulasan mencerminkan pengalaman pelanggan, seperti kata "bagus", "bantu", dan "mudah" menunjukkan bahwa pengguna menghargai kemudahan penggunaan aplikasi dan proses booking yang cepat. Kata "error", "tolong", dan "baru" menandakan masih adanya masalah teknis atau fitur yang perlu diperbaiki. Sementara kata "lambat", "ribet", dan "harus" menunjukkan kesulitan yang ditemui beberapa pengguna aplikasi. Temuan ini dapat digunakan guna memperbaiki fitur, meningkatkan layanan, dan memastikan pengalaman pelanggan lebih baik. Jadi berdasarkan temuan, dapat dikatakan bahwa mayoritas masyarakat memberikan sentimen positif terhadap aplikasi *BromPit* dan merasa puas dengan layanannya, serta *Naive Bayes* salah satu algoritma terbukti model yang paling optimal untuk menganalisis sentimen aplikasi *BromPit*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Sidoarjo atas dukungan fasilitas laboratorium komputer untuk menunjang pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada bagian administrasi kampus atas bantuan dan dukungan administratif selama proses penelitian berlangsung.

REFERENSI

[1] I. Ali and A. Rizki Rinaldi, "PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI PENJUALAN SEPEDA MOTOR TERLARIS," 2023.

[2] I. Setyo Hendriyanto and E. P. Saputro, "Value Jurnal Ilmiah Akuntansi Keuangan dan Bisnis PENGARUH PERSEPSI MEREK, HARGA, DAN KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN SEPEDA MOTOR DI TEGUH JAYA MOTOR PURWODADI," vol. 4, no. 1, 2023.

[3] Ardiansyah and Nur'aini, "JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0) IMPLEMENTASI ANALISIS SENTIMEN MASYARAKAT MENGENAI KENAIKAN HARGA BBM DENGAN METODE NAIVE BAYES," *Journal Of Information Systems And Informatics Engineering*, vol. 8, no. 1, pp. 1–9, 2024, doi: 10.35145/joisie.v8i1.3838.

[4] S. Syafrida and P. Putra, "Pengaruh Kualitas Pelayanan dan Kualitas Produk Terhadap Keputusan Pembelian Sepeda Motor Honda Vario," *Amsir Management Journal*, vol. 3, no. 2, pp. 79–92, Apr. 2023, doi: 10.56341/amj.v3i2.184.

[5] M. R. Effendi, F. T. Julfi, M. Narji, and D. Wanara, "Perancangan Aplikasi Berbasis Android Jadwal Service Sepeda Motor Pada Bengkel Ridho Motor," *Jurnal Teknologi Informatika dan Komputer*, vol. 7, no. 2, pp. 154–168, Sep. 2021, doi: 10.37012/jtik.v7i2.649.

[6] I. Indah Lestari, E. Purnawati, M. Akbar Setiawan, and D. Erla Mahmudah, "EDUKASI KETERAMPILAN TIK GUNA MENGURANGI TINGKAT GAGAP TEKNOLOGI (GAPTEK) MASYARAKAT PURWOKERTO," 2024.

[7] A. P. Nuriza and E. Novalia, "KLASIFIKASI DAN PREDIKSI ULASAN E-COMMERCE MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 9, no. 1, p. 207, Jul. 2025, doi: 10.35145/joisie.v9i1.4993.
[8] G. Darmawan, S. Alam, M. Imam Sulistyo, P. Studi Teknik Informatika, S. Tinggi Teknologi Wastukancana Purwakarta, and R. Artikel, "ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES INFO ARTIKEL ABSTRAK," vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123.
[9] E. Eskiyaturrofikoh and R. R. Suryono, "ANALISIS SENTIMEN APLIKASI X PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1408–1419, Aug. 2024, doi: 10.29100/jipi.v9i3.5392.
[10] R. Sayid Ali Al-Zaelani, Y. Raymond Ramadhan, and M. Andayani Komara, "ANALISIS SENTIMEN REVIEW PRODUK MOTOR HONDA PCX DAN YAMAHA N-MAX PADA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES," 2023.
[11] R. Y. Muslikhin and I. R. I. Astutik, "Analisis Sentimen Ulasan Aplikasi Brompit Di Google Play Store Menggunakan Algoritma Support Vector Machine," Feb. 2023.

Conflict of Interest Statement:
The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Artikel Ilmiah.docx

ORIGINALITY REPORT

19%	19%	18%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Muhammadiyah Sidoarjo Student Paper	15%
2	archive.umsida.ac.id Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Mohammad Bayu Anggara. "Comparison of Naïve Bayes and SVM Methods in Sentiment Analysis of User Reviews on the RSUD AL IHSAN Mobile Application", Competitive, 2025 Publication	<1%
5	Submitted to Universitas 17 Agustus 1945 Semarang Student Paper	<1%

Exclude quotes	On	Exclude matches	< 20 words
Exclude bibliography	On		