

SENTIMENT ANALYSIS OF THE SAMBARA APPLICATION USING THE SUPPORT VECTOR MACHINE ALGORITHM

Thoriq Janati Firdaus^{*1}, Jamaludin Indra², Santi Arum Puspita Lestari³, Hanny Hikmayanti⁴

^{1,2,3,4}Information Engineering, Faculty of Computer Science, Universitas Buana Perjuangan Karawang, Indonesia
Email: ¹if20.thoriqfirdaus@mhs.ubpkarawang.ac.id, ²jamaludin.indra@ubpkarawang.ac.id,
³santi.arum@ubpkarawang.ac.id, ⁴hanny.hikmayanti@ubpkarawang.ac.id

(Article received: August 17, 2024; Revision: August 25, 2024; published: September 3, 2024)

Abstract

Rapid technological developments have opened up new opportunities for public services by utilizing digital application innovations. One example is the West Java Samsat Mobile (SAMBARA) designed by the West Java Provincial Revenue Agency (BAPENDA). The SAMBARA application is expected to accelerate annual vehicle tax payment obligations, but several reviews on the Playstore show user dissatisfaction with SAMBARA's performance. This study aims to conduct a sentiment analysis of SAMBARA application reviews using the Support Vector Machine algorithm. SAMBARA user review data on Google Playstore was collected using the python programming language google play scraper library on google colab resulting in 1620 data on January 2, 2024. The data pre-processing stage involves various steps such as data cleaning, lowercase conversion, tokenization, stemming, stop words removal, normalization, and the use of the TF-IDF method. The data is then labeled positive and negative, positive for reviews with scores of 4 and 5 and negative labels for reviews with scores of 1 to 3. The Support Vector Machine (SVM) algorithm is used for classification, a well-known method for accurate classification. Model evaluation was conducted using a confusion matrix to calculate the precision, recall, and F1-Score values. The evaluation results provide an overview of the performance of the classification algorithm in grouping user reviews into positive and negative categories. The evaluation results show that the SVM algorithm provides quite good performance with an accuracy value of 88.75%, precision 87.51%, recall 81.25%, and F1-Score 83.71% which can be the basis for improving the quality of service of the SAMBARA application. Because the Sambara application has a negative sentiment of 73.4%, it can be concluded that it still gets a bad rating in terms of use.

Keywords: sambara, sentiment analysis, Support Vector Machine

ANALISIS SENTIMEN APLIKASI SAMBARA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Abstrak

Perkembangan teknologi yang pesat telah membuka peluang baru bagi pelayanan publik dengan memanfaatkan inovasi aplikasi digital. Salah satu contohnya adalah Samsat Mobile Jawa Barat (SAMBARA) yang dirancang oleh Badan Pendapatan Daerah Provinsi Jawa Barat (BAPENDA). Aplikasi SAMBARA ini diharapkan dapat mempercepat kewajiban pembayaran pajak kendaraan tahunan, namun beberapa ulasan di Playstore menunjukkan ketidakpuasan pengguna terhadap kinerja SAMBARA. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan aplikasi SAMBARA dengan menggunakan algoritma *Support Vector Machine*. Data ulasan pengguna SAMBARA di Google Playstore dikumpulkan menggunakan library google play scraper bahasa pemrograman python di google colab resulting in 1620 data pada tanggal 2 Januari 2024. Tahap pre-processing data melibatkan berbagai langkah seperti pembersihan data, pengonversian huruf kecil, tokenisasi, stemming, stop words removal, normalisasi, dan penggunaan metode TF-IDF. Data kemudian diberi label positif dan negatif, positif untuk ulasan dengan skor 4 dan 5 serta label negatif untuk ulasan dengan skor 1 sampai 3. Algoritma *Support Vector Machine* (SVM) digunakan untuk klasifikasi, sebuah metode terkenal untuk klasifikasi yang akurat. Evaluasi model dilakukan menggunakan confusion matrix untuk menghitung nilai precision, recall, dan F1-Score. Hasil evaluasi memberikan gambaran tentang performa algoritma klasifikasi dalam mengelompokkan ulasan pengguna ke dalam kategori positif dan negatif. Hasil evaluasi menunjukkan bahwa algoritma SVM memberikan performa yang cukup baik dengan nilai akurasi sebesar 88.75%, precision 87.51%, recall 81.25%, dan F1-Score 83.71% yang dapat menjadi dasar untuk meningkatkan kualitas pelayanan aplikasi SAMBARA. Karena aplikasi Sambara memiliki sentimen negatif sebesar 73,4%, maka dapat disimpulkan masih mendapatkan rating buruk dari segi penggunaan.

Kata kunci: analisis sentimen, sambara, *Support Vector Machine*

1. PENDAHULUAN

Kemajuan teknologi yang semakin cepat membawa kemudahan bagi masyarakat dalam berbagai aspek kehidupan. Perkembangan teknologi ini menghasilkan data dalam jumlah besar, yang dapat memberikan informasi yang bermanfaat jika diolah dan dimanfaatkan dengan baik[1]. Inovasi dalam pelayanan publik telah menjadi kebijakan pemerintah sebagai *respons* terhadap kemajuan teknologi dan informasi[2]. Salah satu bentuk pelayanan publik yang disediakan oleh pemerintah adalah Samsat Mobile Jawa Barat (SAMBARA), yang mengembangkan aplikasi SAMBARA guna meningkatkan kualitas pelayanan. Perkembangan teknologi komputer di bidang pelayanan ini mengarahkan kita menuju sistem identitas digital, yang kini mulai kita rasakan manfaatnya. Sistem ini, yang sebelumnya berbasis manual, secara bertahap beralih menjadi digital.

SAMBARA adalah inovasi elektronik yang diluncurkan oleh Badan Pendapatan Daerah Provinsi Jawa Barat (BAPENDA), yang berfungsi untuk memeriksa pajak kendaraan di wilayah Jawa Barat. Melalui aplikasi ini, pengguna dapat secara *online* melihat informasi mengenai pajak kendaraan bermotor, baik untuk roda dua maupun mobil, di provinsi Jawa Barat[3].

Aplikasi SAMBARA, yang dirilis pada 2 Januari 2018, telah diunduh oleh lebih dari 1 juta pengguna. Aplikasi ini menawarkan berbagai layanan, seperti pembayaran pajak kendaraan secara *online*, informasi tentang lokasi layanan Samsat di sekitar Anda, jadwal Samsat Keliling dan Samsat Pakai, serta banyak lagi[4]. Namun, di antara banyaknya pengguna yang telah mengunduh aplikasi ini, ada beberapa yang merasa kurang puas dengan kinerja SAMBARA. Hal ini terlihat dari ulasan pengguna di *Playstore*. Ulasan tersebut memberikan masukan setelah mereka menggunakan aplikasi, dan rating yang diberikan bisa menjadi referensi untuk menganalisis serta memahami kebutuhan dan keluhan pengguna. Analisis sentimen juga bisa dimanfaatkan untuk meningkatkan performa aplikasi ini[5].

Beberapa ulasan di *Playstore* mengindikasikan ketidakpuasan pengguna terhadap kinerja aplikasi SAMBARA. Oleh karena itu, analisis terhadap ulasan, rating, dan sentimen pengguna menjadi penting untuk memahami kebutuhan serta keluhan mereka, dengan tujuan meningkatkan kualitas layanan aplikasi. Analisis sentimen adalah metode yang digunakan untuk menilai emosi pengguna. Klasifikasi sentimen, yang mencakup kategori negatif atau positif, bertujuan untuk memprediksi polaritas sentimen berdasarkan data ulasan pengguna[6]. Analisis sentimen juga dikenal sebagai kecerdasan buatan emosional dan berguna dalam

pemrosesan bahasa alami, analisis teks, serta komputasi linguistik[7]. Ini membantu mengidentifikasi, mengekstrak, mengukur, dan mempelajari informasi secara terstruktur, serta meningkatkan akurasi, presisi, *recall*, dan *F1-Score* dalam hasil algoritma klasifikasi sentimen. Akurasi menunjukkan seberapa baik model dapat mengklasifikasikan data dengan benar, presisi mengukur keakuratan antara data yang diharapkan dengan hasil prediksi model, *recall* mencerminkan proporsi data yang diklasifikasikan dengan benar, dan *F1-Score* adalah rata-rata harmonis dari presisi dan *recall*[8]. Terdapat beberapa penelitian yang telah dilakukan terkait analisis sentimen pada topik tertentu dengan memanfaatkan data ulasan dari *Google Playstore*.

Salah satu penelitian di bidang ini adalah Analisis Sentimen Ulasan Aplikasi Mobile Jamsostek dengan Menggunakan Strategi *Support Vector Machine*, nilai akurasi pada penelitian tersebut mendapatkan 96%, dengan rata-rata untuk nilai presisi adalah 92%, rata-rata nilai *recall* adalah 96%, dan rata-rata nilai *f1-score* adalah 94%[9].

Penelitian lainnya tentang Analisis Sentimen Aplikasi Zenius menggunakan Strategi Calculated Relpase, review ini menimbulkan ekspektasi yang besar dalam melakukan sistem pengelompokan pada survey aplikasi Zenius di *Google Playstore*, pada kelas 0 nilai akurasi, review, dan skor *f-1* yang diberikan sebesar 90%, sedangkan untuk kelas 1 nilai akurasi dan skor *f-1* yang diberikan sebesar 90%, review *esteem* sebesar 91%, dan *precision esteem* sebesar 90% yang menunjukkan bahwa model tersebut sangat bagus untuk mengurutkan teks.[10].

Penelitian tentang Urutan Opini Audit Aplikasi WhatsApp di *Playstore* Menggunakan Strategi K-Nearest Neighbor. Dengan nilai *f1* sebesar 89,87% untuk kelas positif, nilai akurasi sebesar 84%, nilai presisi sebesar 87,65%, dan nilai *recall* sebesar 92,21 persen, metode K-NN mencapai kinerja model terbaik pada pembagian data 90:10 penelitian ini. Sedangkan pada kelas negatif, nilai akurasi sebesar 68,42%, review sebesar 56,52%, dan nilai *f1* sebesar 61,90% pada $k = 14$ dan $Edge = 20$. [11].

Pada penelitian menggunakan metode Random Forest untuk Analisis Sentimen Ulasan Aplikasi Dana, berdasarkan perbandingan hasil data pelatihan dan pengujian serta analisis. 80%:20%, nilai *precision* 84%, *recall* 84%, *F1-Score* 84%, dan nilai *accuracy* 84% dengan kedalaman pohon 65 dan jumlah pohon 400[12].

Pada Memanfaatkan Fitur Chi-Square K-NN yang Disempurnakan dan Metode BM25 untuk Analisis Sentimen Aplikasi TikTok, 5 pengujian tingkat fitur digunakan dalam pengujian ini, dan hasil terbaik diperoleh dengan tingkat fitur 50% dan

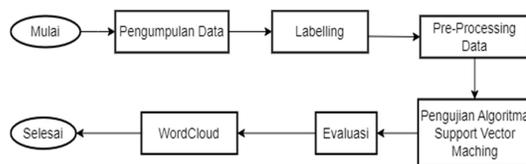
$k = 20$ sehingga, hasil terbaik diperoleh dengan nilai *precision* sebesar 70,03%, *recall* sebesar 67,22%, *accuracy* sebesar 83,33%, dan *f-measure* sebesar 66,26%. Dapat disimpulkan bahwa penambahan fitur seleksi dapat meningkatkan hasil dalam hal *recall*, *f-measure*, *precision*, dan *accuracy*[13].

Analisis sentiment ulasan SAMBARA telah dilakukan sebelumnya menggunakan algoritma Naïve Bayes dan Fitue Chi-Square, Analisis menunjukan hasil kinerja metode Naive Bayes dan pemilihan fitur Chi-square dalam mengklasifikasikan review aplikasi Sambara menggunakan data uji sebanyak 179 data review mempunyai nilai kritis sebesar 0,837 atau spesifisitas sebesar 94,4%, pada beberapa kasus tercapai nilai akurasi tertinggi menggunakan pemilihan fitur chi-square. Klasifikasi tanpa chi-square mencapai akurasi 92%, spesifisitas 94,3%, dan *recall* 94%[14].

Tujuan dari penelitian ini adalah untuk melihat apa yang orang katakan tentang aplikasi SAMBARA di Google Playstore. Perhitungan yang digunakan dalam ulasan ini adalah *Support Vector Machine* untuk menerapkan strategi karakterisasi[15]. Penelitian ini fokus pada analisis sentimen terhadap aplikasi SAMBARA di Google Playstore, dengan dua jenis sentimen yang diidentifikasi berdasarkan nilai ulasan. Sentimen positif diberi skor 4 hingga 5, sementara sentimen negatif diberi skor 1 hingga 3. Metode *Support Vector Machine* diharapkan dapat memberikan hasil akurasi terbaik dalam klasifikasi data tersebut.

2. METODE PENELITIAN

Ada beberapa tahapan untuk mendapatkan hasil terbaik dalam menganalisa sentimen dan melihat tingkat akurasinya. Berikut tahapan yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

Pada Gambar 1 dapat dilihat Tahapan penelitian dimulai dari pengumpulan data ulasan aplikasi. Kedua labelling data untuk positif dan negatif. Ketiga, dataset yang sudah di peroleh akan di preprocessing untuk dilakukan penerapan algoritma. Keempat, implementasi algoritma menggunakan *Support Vector Machine*. Kelima, Tahapan evaluasi penelitian menggunakan *confusion matrix*. Tahap terakhir visualisasi menggunakan *Wordcloud*.

A. Pengumpulan Data

Data dikumpulkan melalui dua metode, yaitu studi literatur dan pengambilan data (*crawling*).

Pengumpulan ulasan pengguna aplikasi SAMBARA di *Google Playstore* dilakukan menggunakan *library google-play-scraper* dalam bahasa pemrograman Python melalui *Google Colaboratory*. Dari proses ini, diperoleh 1620 data yang mencakup informasi seperti nama pengguna, teks ulasan, dan skor ulasan. Data yang dikumpulkan adalah ulasan terbaru di *Google Playstore* pada tanggal 2 Januari 2024, dengan rentang periode pengambilan data dari 25 Maret 2023 hingga 2 Januari 2024.

B. Labelling

Setelah tahap pengumpulan data, langkah berikutnya adalah memberikan label pada setiap ulasan. Proses ini melibatkan dua kategori label: positif dan negatif. Ulasan dengan skor 4 hingga 5 dikategorikan sebagai positif, sementara ulasan dengan skor 1 hingga 3 dikategorikan sebagai negatif[16].

C. Preprocessing

Preprocessing Data merupakan tahap yang mengubah data agar sesuai dengan format yang diinginkan dan dapat diproses. Ada beragam macam proses pre-processing tetapi tergantung kebutuhan, secara umum diantaranya *Cleaning*, *Case Folding*, *Tokenizing*, *Stemming*, *Stop Words*, *Normalize*, *TF-IDF*[17].

1. *Case Folding* : tahapan di mana huruf besar diganti dengan huruf kecil. Proses ini dilakukan dengan tujuan untuk memudahkan pencarian dan mendapatkan format standar teks[18].
2. *Cleaning* : Proses pembersihan data untuk menghilangkan kata yang tidak diperlukan. Sesuatu seperti ini: Misalnya, hapus tanda "@" di nama pengguna atau "#" di hashtag[19].
3. *Tokenizing* : tahap untuk membagi kalimat menjadi bagian-bagian kata tersendiri. Angka, tanda baca, spasi, karakter individual dalam kalimat, dan karakter khusus (tab, jeda baris, *link*, *backslace*, emoji, *mention*, *hashtags*, URL) juga dihilangkan pada tahap ini[20].
4. *Stemming* : Digunakan untuk menghilangkan kata yang mengandung imbuhan. Semua kata yang mempunyai imbuhan diubah menjadi kata dasar pada saat proses *stemming*[21].
5. *Normalize* : Normalisasi kata merupakan proses perubahan kata-kata yang tidak baku seperti singkatan dan bahasa gaul menjadi baku, kemudian menggantinya dengan kata yang ada di kamus KBBI[22].
6. *Stopword*: Proses menghilangkan kata-kata yang kurang penting atau tidak berhubungan dari kosakata yang berkaitan dengan kasus yang diteliti, biasanya preposisi dan subjek[23].
7. *TF-IDF* memiliki tujuan untuk menghitung bobot setiap kata yang digunakan sebagai elemen, dengan semakin banyaknya dokumen yang diproses, jumlah fiturnya pun bertambah. Proses ini terdiri dari dua bagian, yaitu TF

(Term Frequency) dan IDF (Inverse Document Frequency). TF mengacu pada kuantitas kejadian setiap kata dalam laporan; semakin sering kata muncul, semakin tinggi nilai TF-nya. Sebaliknya, IDF adalah ukuran frekuensi kemunculan kata dalam dokumen. Nilai IDF berbanding terbalik satu sama lain, jadi kata yang jarang muncul dalam dokumen akan memiliki nilai IDF yang lebih tinggi. Mengalikan nilai TF dan IDF menghasilkan hasil pembobotan kata TF-IDF, yang memberikan bobot lebih rendah untuk kata yang sering digunakan dalam banyak dokumen dan bobot lebih tinggi untuk kata yang jarang digunakan. di seluruh dokumen dalam koleksi[24].

D. Penerapan Algoritma *Support Vector Machine*
 Proses ini merupakan proses pengolahan data yang dilakukan pada saat preprocessing dan kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM). *Support Vector Machine* (SVM) adalah teknik pembelajaran yang sangat populer dengan berbagai properti yang diinginkan. SVM memiliki dasar teori yang kuat dan seringkali menghasilkan klasifikasi yang lebih akurat dibandingkan banyak algoritma lain di berbagai aplikasi. Banyak penelitian telah menunjukkan bahwa SVM adalah salah satu metode klasifikasi teks yang paling akurat, dan SVM juga sering digunakan dalam klasifikasi sentimen[25].

$$Support\ Vector\ Machine = K(x_i, x) = x^t x \quad (1)$$

E. Evaluasi
 Data yang diterapkan pada algoritma tersebut kemudian dievaluasi, dengan tujuan untuk mengevaluasi model dari algoritma klasifikasi yang diterapkan. Evaluasi ini menggunakan confusion matrix yang menghitung nilai presisi dan perolehan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2x \frac{precision \times recall}{precision + recall} \quad (4)$$

Keterangan :

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

F. Wordcloud

Setelah tahap evaluasi selanjutnya akan dilakukan tahap visualisasi menggunakan Wordcloud. Tahap ini akan menampilkan gambar

pada seluruh dokumen sentimen ulasan positif dan negatif pada aplikasi sambara.

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Tahap pengumpulan data menggunakan perpustakaan Google_Play_Scrapper dan data review terbaru dari Google Playstore untuk mengumpulkan ulasan pengguna untuk aplikasi SAMBARA yang berjumlah 1620 data dan hanya 3 data yang akan di tampilkan. Contoh data yang dikumpulkan menggunakan perpustakaan Google_Play_Scrapper adalah sebagai berikut pada tabel 1.

Tabel 1 Pengumpulan Data

Username	Score	Content
Septian hadiwibowo	1	Loading data to server mulu,, gak pernah di maintenance kah? Jadi tidak berguna sama sekali ini aplikasi
Anton yunus	5	Sangat membantu untuk cek.kendaraan dan pajak
Sandi kusdiantoro	3	Tidak berfungsi dengan baik. Sungguh mengecewakan.

B. Labelling

Proses label pada setiap kata dalam data berdasarkan skor review data teks. Nilai label berisi dua kategori, positif dan negatif, dengan nilai 0 dan 1. Label 0 adalah sentimen negatif dan 1 adalah sentimen positif. Tabel berikut menunjukkan pelabelan data berdasarkan rating, dan dapat di lihat pada tabel 2.

Tabel 2 Labelling

Content	Score	Label
Loading data to server mulu,, gak pernah di maintenance kah? Jadi tidak berguna sama sekali ini aplikasi	1	0
Sangat membantu untuk cek.kendaraan dan pajak	5	1
Tidak berfungsi dengan baik. Sungguh mengecewakan.	3	0

C. Preprocessing

1. Case Folding, langkah untuk menguba semua huruf menjadi huruf kecil(lowercase). Contoh Case Folding sebagai berikut dalam tabel 3.

Tabel 3 Case Folding

Content	Case Folding
Loading data to server mulu,, gak pernah di maintenance kah? Jadi tidak berguna sama sekali ini aplikasi	loading data to server mulu,, gak pernah di maintenance kah? jadi tidak berguna sama sekali ini aplikasi
Sangat membantu untuk.cek.kendaraan dan pajak	sangat membantu untuk.cek.kendaraan dan pajak
Tidak berfungsi dengan baik. Sungguh mengecewakan.	tidak berfungsi dengan baik.sungguh mengecewakan.

- Cleaning merupakan proses penghapusan karakter non abjad. Contoh cleaning sebagai berikut dalam tabel 4.

Tabel 4 Cleaning

Case Folding	Cleaning
loading data to server mulu,, gak pernah di maintenance kah? jadi tidak berguna sama sekali ini aplikasi	loading data to server mulu gak pernah di maintenance kah jadi tidak berguna sama sekali ini aplikasi
sangat membantu untuk.cek.kendaraan dan pajak	sangat membantu untuk cek kendaraan dan pajak
tidak berfungsi dengan baik.sungguh mengecewakan.	tidak berfungsi dengan baik sungguh mengecewakan

- Tokenizing, proses memisahkan teks per teks menjadi string atau kata berdasarkan setiap kata yang di dapat dari ulasan masyarakat. Contoh tokenizing sebagai berikut dalam tabel 5.

Tabel 5 Tokenizing

Cleaning	Tokenizing
loading data to server mulu gak pernah di maintenance kah jadi tidak berguna sama sekali ini aplikasi	['loading', 'data', 'to', 'server', 'mulu', 'gak', 'pernah', 'di', 'maintenance', 'kah', 'jadi', 'tidak', 'berguna', 'sama', 'sekali', 'ini', 'aplikasi']
sangat membantu untuk cek kendaraan dan pajak	['sangat', 'membantu', 'untuk', 'cek', 'kendaraan', 'dan', 'pajak']
tidak berfungsi dengan baik sungguh mengecewakan	['tidak', 'berfungsi', 'dengan', 'baik', 'sungguh', 'mengecewakan']

- Stemming, proses penghapusan kata imbuhan pada awalan dan akhiran. Stemming di lakukan untuk menjadikan kata lebih sederhana yang menjadikanya seperti kata dasar. Contoh stemming sebagai berikut dalam tabel 6.

Tabel 6 Stemming

Tokenizing	Stemming
['loading', 'data', 'to', 'server', 'mulu', 'gak', 'pernah', 'di', 'maintenance', 'kah', 'jadi', 'tidak', 'berguna', 'sama', 'sekali', 'ini', 'aplikasi']	loading data to server mulu gak pernah di maintenance kah jadi tidak guna sama sekali ini aplikasi
['sangat', 'membantu', 'untuk', 'cek', 'kendaraan', 'dan', 'pajak']	sangat bantu untuk cek kendara dan pajak
['tidak', 'berfungsi', 'dengan', 'baik', 'sungguh', 'mengecewakan']	tidak fungsi dengan baik sungguh kecewa

- Normalize merupakan langkah untuk menggantikan kata ejaan yang kurang tepat dan kata ejaan dengan singkatan yang akan di koreksi berdasarkan kata yang tertulis di KBBI, contoh normalize sebagai berikut dalam tabel 7.

Tabel 7 Normalize

Stemming	Normalize
loading data to server mulu gak pernah di maintenance kah jadi tidak guna sama sekali ini aplikasi	['loading', 'data', 'to', 'server', 'mulu', 'gak', 'pernah', 'di', 'maintenance', 'kah', 'jadi', 'tidak', 'guna', 'sama', 'sekali', 'ini', 'aplikasi']
sangat bantu untuk cek kendara dan pajak	['sangat', 'bantu', 'untuk', 'cek', 'kendara', 'dan', 'pajak']
tidak fungsi dengan baik sungguh kecewa	['tidak', 'fungsi', 'dengan', 'baik', 'sungguh', 'kecewa']

- Stopword adalah kata-kata yang sering muncul tetapi tidak dianggap memiliki makna signifikan. Contoh stopwords dapat dilihat dalam tabel 8.

Tabel 8 Stopword

Normalize	Stopword
['loading', 'data', 'to', 'server', 'mulu', 'gak', 'pernah', 'di', 'maintenance', 'kah', 'jadi', 'tidak', 'guna', 'sama', 'sekali', 'ini', 'aplikasi']	loading data to server mulu gak maintenance kah aplikasi

['sangat', 'bantu', 'untuk', 'cekk', 'kendara', 'dan', 'pajak']	bantu cek kendara pajak
['tidak', 'fungsi', 'dengan', 'baik', 'sungguh', 'kecewa']	fungsi sungguh kecewa

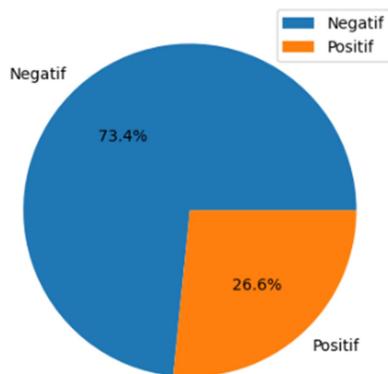
7. TF-IDF Tahap pembobotan kata ini bertujuan untuk menghitung nilai setiap kata dalam teks. Kualitas-kualitas ini kemudian akan digunakan sebagai elemen dalam tabel 9.

Tabel 9 TF-IDF

Term	TF	IDF	TFIDF
App	2	5.1637540135 94269	0.6821407186988 224
Stnk	3	4.0861951341 23991	0.7252843407049 465
daftar	1	4.9919037566 67609	0.3297195032222 685
masuk	1	4.3254248231 89826	0.7669698255979 689

D. Klasifikasi Algoritma Support Vector Machine

Pada tahap ini, digunakan 1620 data yang telah melalui proses prapemrosesan dan diklasifikasikan ke dalam dua label sentimen: positif dan negatif. Pengujian dilakukan dengan menggunakan seluruh data, di mana 80% dari data tersebut diacak sebagai data latih, dan 20% sisanya diacak sebagai data uji. Pengujian pertama dilakukan untuk mengukur performa berdasarkan kernel yang digunakan. Pada Gambar 2, menunjukkan hasil sentimen terhadap aplikasi sambara.



Gambar 2 Hasil Sentimen

Pada Gambar 2, menjelaskan bahwa kelas yang mempunyai nilai sentimen positif sebesar 26,6% dan sentimen negatif sebesar 73,4%. Kemudian, data yang digunakan untuk mencari nilai akurasi menggunakan algoritma adalah data testing. Berikut pada tabel 10, menunjukkan hasil performa algoritma *Support Vector Machine*. Berikut Tabel 10 menunjukkan hasil performa algoritma *Support Vector Machine*

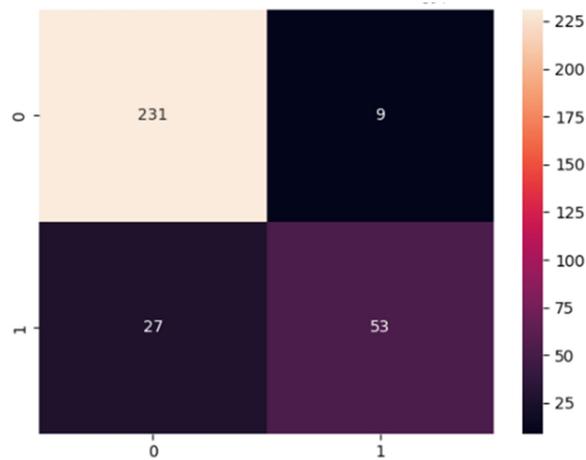
Tabel 10 Hasil Performa Algoritma Support Vector Machine

No	Algoritma	Akurasi	Pesisi	Recall	F1-Score
1	Support Vector Machine	88,75%	87,51	81,25	83,71

Berdasarkan pada tabel 10, algoritma *Support Vector Machine* mendapat nilai akurasi 88.75%, presisi 87.51%, Recall 81.25%, dan F1-Score 83.71%. Dapat disimpulkan bahwa nilai performa akurasi dari model algoritma *Support Vector Machine* cukup baik dalam klasifikasi pada aplikasi sambara.

E. Evaluasi

Pada tahap Evaluasi ini menggunakan Confussion Matrix untuk mengetahui bagaimana klasifikasi algoritma bekerja pada jumlah data label yang di prediksi. Evaluasi ini juga melakukan perhitungan ulang pada hasil akurasi yang di peroleh setiap algoritma. Visualisasi *Confussion Matrix* dapat dilihat pada Gambar 3.



Gambar 3 Confussion Matrix

Vector Machine (SVM) menunjukkan bahwa SVM merupakan kalkulasi yang benar-benar ideal untuk menganalisis sentimen yang berhubungan dengan survei aplikasi SAMBARA, dengan tingkat ketepatan sebesar 88,75%, akurasi sebesar 87,51%, dan review sebesar 81,25%. Dengan menambah jumlah data dan mempertimbangkan metode pengolahan data yang lebih baik untuk meningkatkan kinerja algoritma, data ini dapat dijadikan sebagai rekomendasi untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] T. A. U. Azmi, L. Hakim, D. C. R. Novitasari, and W. D. U. D. Utami, "Application Random Forest Method for Sentiment Analysis in Jamsostek Mobile Review," *Telematika*, vol. 20, no. 1, p. 117, 2023, doi: 10.31315/telematika.v20i1.8868.
- [2] M. F. D. Samudera and V. I. Pertiwi, "Inovasi Pelayanan Publik Melalui Jamsostek Mobile (JMO) (Studi Kasus di BPJS Ketenagakerjaan Cabang Rungkut Kota Surabaya)," *JPAP J. Penelit. Adm. Publik*, vol. 8, no. 1, pp. 152–172, 2022, doi: 10.30996/jpap.v8i1.6617.
- [3] A. Ramdani, "Analisis Deskriptif Terhadap Inovasi Layanan Aplikasi SAMBARA (Samsat Mobile Jawa Barat)," *J. Acad. Praja*, vol. 3, no. 01, pp. 37–43, 2020, doi: 10.36859/jap.v3i01.141.
- [4] M. F. Mulyaddin, M. R. Mahendra, and M. Dewi, "Analisis Sentimen Komentar Pengguna Aplikasi SAMBARA di Google Playstore," vol. 39, pp. 85–91, 2024.
- [5] H. Faisal, A. Febriandirza, and F. N. Hasan, "Analisis Sentimen Terkait Ulasan Pada Aplikasi PLN Mobile Menggunakan Metode Support Vector Machine," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 5, no. 1, pp. 303–312, 2024.
- [6] Y. Khoiruddin, A. Fauzi, and A. M. Siregar, "Analisis Sentimen Gojek Indonesia Pada Twitter Menggunakan Algoritme Naive Bayes Dan Support Vector Machine," *J. Ilm. Komput.*, vol. 19, pp. 391–400, 2023.
- [7] I. Zulfahmi, J. Williem Iskandar, P. V Medan, S. Tuan, and D. Serdang, "Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decision Tree," *J. Penelit. Rumpun Ilmu Tek.*, vol. 3, no. 1, pp. 11–21, 2024, [Online]. Available: <https://doi.org/10.55606/juprit.v3i1.3096>
- [8] Y. Romadhoni and K. F. H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 118–124, 2022, doi: 10.30591/jpit.v7i2.3191.
- [9] V. Fitriyana, Lutfi Hakim, Dian Candra Rini Novitasari, and Ahmad Hanif Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *J. Buana Inform.*, vol. 14, no. 01, pp. 40–49, 2023, doi: 10.24002/jbi.v14i01.6909.
- [10] I. M. Juniandika, I. Bagus, and M. Mahendra, "Analisis Sentimen Aplikasi Zenius Menggunakan Metode Logistic Regression," vol. 1, pp. 1171–1178, 2023.
- [11] M. Riski, M. Fikry, and Yusra, "Klasifikasi Sentimen Ulasan Aplikasi WhatsApp di Play Store Menggunakan Metode K-Nearest Neighbor," *Media Online*, vol. 4, no. 1, pp. 438–444, 2023, doi: 10.30865/klik.v4i1.1050.
- [12] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," ... *Tek. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] A. Purnamawati, M. N. Winarto, and M. Mailasari, "Analisis Sentimen Aplikasi TikTok menggunakan Metode BM25 dan Improved K-NN Fitur Chi-Square," *J. Komtika (Komputasi dan Inform.)*, vol. 7, no. 1, pp. 97–105, 2023, doi: 10.31603/komtika.v7i1.8938.
- [14] M. Irfan, D. Widiyanto, and J. Jayanta, "Analisis Sentimen Pada Ulasan Aplikasi Samsat Mobile Jawa Barat (Sambara) Menggunakan algoritma Naive bayes Classifier dengan seleksi fitur Chi Square," ... *Mhs. Bid. Ilmu ...*, pp. 337–346, 2022, [Online]. Available: [https://conference.upnvj.ac.id/index.php/senamika/article/view/2179%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/do](https://conference.upnvj.ac.id/index.php/senamika/article/view/2179%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/download/2179/1748)
- [15] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [16] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [17] S. Shevira, I. Made, A. D. Suarjaya, and P. Wira Buana, "Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia," *JITTER-Jurnal Ilm. Tek. dan Komput.*, vol. 3, no. 2, 2022.

- [18] T. D. Ramadhan, D. Wahiddin, and E. E. Awal, "Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes," *Sci. Student J. Information, Technol. Sci.*, vol. IV, no. 1, pp. 82–87, 2023, [Online]. Available: www.tripadvisor.com
- [19] M. Wijaya, "Penerapan Algoritma Naive Bayes Dan KNN Dalam Menganalisis Sentimen Aplikasi Tiktok Shop Seller Center Berdasarkan Review Google Playstore," vol. V, pp. 115–122, 2024.
- [20] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," vol. 17, no. 1, pp. 25–38, 2022.
- [21] Y. Alfiansah, A. Siregar, and ..., "Penerapan Algoritma K-Nearest Neighbors untuk Analisis Sentimen pada Buletin APTIKOM," ... *Student J. ...*, vol. III, pp. 125–132, 2022, [Online]. Available: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/431%0Ahttps://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/431/345>
- [22] J. A. Pangetsu, Y. Cahyana, and H. Y. Novita, "Penerapan Algoritma Support Vector Machine Untuk," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 1, p. 40, 2024, [Online]. Available: <https://t.co/dKngKqLqXI>
- [23] A. Syihabudin, A. Ratna Juwita, and A. Rizki Pratama, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Terhadap Produk Motor Matic Honda Beat dan Scoopy," *Sci. Student J. Information, Technol. Sci.*, vol. IV, no. 1, pp. 95–101, 2023.
- [24] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer Dike : Jurnal Ilmu Multidisiplin," vol. 1, pp. 37–43, 2023.
- [25] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjee.v3i2.10885.

