

Analisis Sentimen Aplikasi Samsat Digital di Play Store Menggunakan Support Vector Machine

Nisrina Yulia Setyawati^{1)*}, Erni Widarti²⁾

¹⁾²⁾Sistem Informasi Kota Cerdas, Fakultas Teknik, Universitas Tunas Pembangunan Surakarta

E-mail: ^{1)*} nisyuliaa@gmail.com, ²⁾ erni.widarti@lecture.utp.ac.id

**Corresponding*

Abstrak

Teknologi memainkan peran penting dalam pengembangan layanan publik di era digital saat ini. Aplikasi Samsat Digital Nasional (SIGNAL) merupakan contoh nyata penerapan teknologi informasi dalam mempermudah proses pembayaran pajak kendaraan bermotor. Dengan memanfaatkan sistem digital, masyarakat dapat melakukan transaksi secara daring tanpa harus mengunjungi kantor Samsat secara fisik. Penelitian bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi SIGNAL di Play Store dengan menggunakan metode Support Vector Machine (SVM). Evaluasi dilakukan dengan membandingkan akurasi dari empat kernel SVM-linear, RBF, sigmoid, dan polynomial, pada tiga skenario. Hasil terbaik diperoleh pada skenario 1 dengan perbandingan data training 90% dan data testing 10%, menghasilkan akurasi sebesar 92,2%, precision 97,21%, recall 97,2%, dan f1-score 97,19%. Analisis juga menunjukkan bahwa sebagian besar ulasan pengguna bersifat positif, dengan 531 ulasan positif dan 219 ulasan negatif. Kata-kata yang sering muncul dalam ulasan positif meliputi "aplikasi", "mudah", "pajak", dan "bantu". Temuan ini menunjukkan bahwa aplikasi SIGNAL telah diterima dengan baik oleh pengguna dan dapat menjadi dasar bagi pengembang untuk terus meningkatkan layanan.

Kata Kunci— Analisis sentimen, Aplikasi SIGNAL, Support Vector Machine (SVM)

Abstract

Technology plays a crucial role in the development of public services in today's digital era. The National Digital Samsat application (SIGNAL) is a tangible example of the application of information technology aimed at facilitating the payment process for motor vehicle taxes. By utilizing digital systems, the public can conduct transactions online without the need to physically visit the Samsat office. This study aims to analyze user sentiment regarding the SIGNAL application on the Play Store using the Support Vector Machine (SVM) method. The evaluation was conducted by comparing the accuracy of four SVM kernels-linear, RBF, sigmoid, and polynomial, across three scenarios. The best results were obtained in scenario 1, with a training data ratio of 90% and testing data of 10%, yielding an accuracy of 92.2%, precision of 97.21%, recall of 97.2%, and an F1-score of 97.19%. The analysis also indicates that the majority of user reviews are positive, with 531 positive reviews and 219 negative reviews. The words that frequently appear in positive reviews include "application," "easy," "tax," and "help." These findings suggest that the SIGNAL application has been well received by users and can serve as a foundation for developers to continually improve the service.

Keywords— Sentiment Analysis, SIGNAL Application, Support Vector Machine (SVM)

1. Pendahuluan

Konsep pengembangan kota dengan melibatkan adanya teknologi atau biasa disebut dengan kota cerdas sudah banyak diterapkan di negara-negara maju di dunia. Indonesia juga mulai menerapkan konsep kota cerdas tersebut, seperti yang terdapat pada program Kementerian Komunikasi dan Informatika yaitu gerakan menuju 100 smart city. Gerakan tersebut berfokus untuk memaksimalkan teknologi dalam rangka

meningkatkan pelayanan masyarakat [1]. Meningkatnya jumlah penduduk kota yang mengakibatkan adanya perubahan perilaku dan kebiasaan menjadi alasan utama adanya program smart city ini karena berbagai pihak dituntut untuk membuat visi baru. Visi baru yang dimaksud adalah kota pintar dimana dapat didefinisikan sebagai suatu tatanan kota yang mampu memanfaatkan penggunaan teknologi dalam pemberian pelayanan publik sehingga dapat meningkatkan kualitas hidup masyarakat. Ketersediaan pelayanan publik berbasis teknologi atau dikenal dengan konsep E-Government ini menjadi tanggung jawab pemerintah sepenuhnya. Strategi implementasi smart city pada lingkup E-Government biasanya dimunculkan dalam bentuk aplikasi android [2].

Instansi pemerintahan yang mulai menerapkan e-gov dalam pelayanannya adalah samsat. Samsat atau bisa disebut dengan Sistem Administrasi Manunggal Satu Atap yang memberikan pelayanan dalam pembayaran pajak kendaraan bermotor (PKB), Sumbangan Wajib Dana Kecelakaan Lalu Lintas Jalan (SWDKLLJ), dan Penerima Negara Bukan Pajak (PNBP) Pengesahan Surat Tanda Nomor Kendaraan (STNK) [3]. Aplikasi yang diterapkan merupakan samsat digital nasional atau SIGNAL yang merupakan aplikasi pembayaran pajak kendaraan bermotor. Melalui aplikasi ini masyarakat dapat membayar pajak dari rumah tanpa harus datang ke kantor samsat. Masyarakat dapat menggunakan aplikasi ini dengan mengunduhnya melalui play store. Selayaknya aplikasi yang baru dikembangkan, pengembang memerlukan ulasan pengguna untuk memantau kualitas aplikasi sehingga dapat bersaing dengan aplikasi lainnya [4]. Ulasan aplikasi SIGNAL dapat ditulis oleh siapapun yang sudah mengunduhnya melalui play store serta juga dapat dilihat oleh siapa saja. Data ulasan sangat berpengaruh untuk calon pengguna karena berperan sebagai bahan pertimbangan penggunaan suatu aplikasi. Data ulasan dapat diolah dan berguna sebagai bahan perbaikan atau kritik dan saran dari pengguna untuk pengembang aplikasi. Namun data ulasan yang berjumlah sangat banyak menyebabkan kesulitan apabila diolah secara manual, padahal dari hasil olahan tersebut dapat dilihat kecenderungan pengguna apakah ke arah positif atau negatif [5]. Dengan melihat betapa pentingnya data ulasan maka dibutuhkan analisis sentimen untuk mengkategorikan topik dengan label positif, negatif, atau netral [6][7]. Salah satu metode yang dapat digunakan dalam analisis sentimen adalah Support Vector Machine [8][9].

Konsep Support Vector Machine (SVM) dikenalkan pada tahun 1992 yang merupakan sebuah konsep penggabungan antara teori komputasi seperti margin hyperplane, kernel, serta konsep pendukung lain. Salah satu machine learning ini digunakan digunakan untuk pengklasifikasian data dengan mencari garis pembatas. Jika dalam dua kelas pembatasnya disebut dengan garis, untuk tiga kelas disebut dengan plane, sedangkan untuk kelas lebih dari tiga disebut dengan hyperplane [10]. Pencarian garis pembatas dilakukan dengan menemukan nilai hyperplane menggunakan support vector dan nilai margin. Pada awalnya SVM bekerja secara linear classifier, namun selanjutnya diperluas pada permasalahan non-linear dengan memasukkan konsep kernel pada ruang berdimensi tinggi untuk memaksimalkan jarak antar kelas data [11]. Setiap kelas pada SVM akan disimbolkan dengan pattern tertentu, seperti pattern kotak untuk kelas negatif dan pattern lingkaran untuk kelas positif [12]. Metode SVM akan bekerja untuk menemukan hyperplane atau pembatas terbaik yang memisahkan antara kelas satu dengan kelas yang lainnya [13].

Penelitian ini bertujuan untuk menganalisis ulasan pengguna pada aplikasi SIGNAL pada PlayStore dengan menggunakan metode klasifikasi Support Vector Machine (SVM). Analisis dilakukan dengan membandingkan hasil akurasi antara kernel linear, kernel RBF, kernel sigmoid, dan kernel polynomial. Metode Support Vector Machine (SVM) dipilih karena memiliki hasil pengklasifikasian yang lebih baik dan akurat. Dengan diperolehnya informasi ini, maka nantinya juga dapat dijadikan sebagai acuan dalam menjaga popularitas, kualitas pelayanan, dan dalam rangka memperbaiki kekurangan serta evaluasi ke arah yang lebih baik.

2. Penelitian Terdahulu

Analisis sentimen dengan menggunakan metode Support Vector Machine (SVM) sudah pernah dilakukan sebelumnya, seperti penelitian [14] menggunakan metode support vector machine untuk melihat umpan balik dari penggunaan aplikasi novel online yaitu dreame dan wattpad. Data umpan balik akan dikategorisasi ke dalam label positif dan negatif dengan

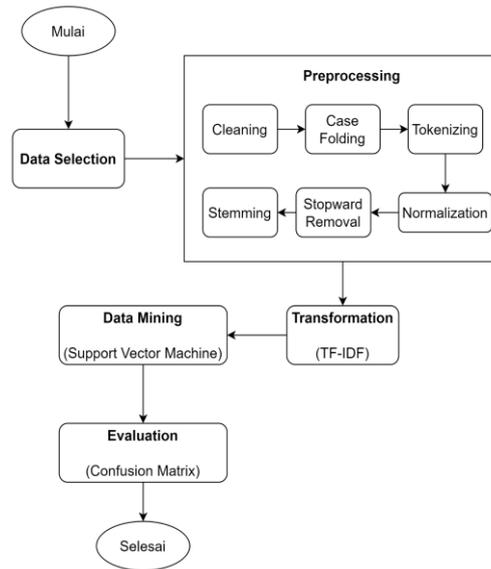
menggunakan tools vader lexicon di Python. Berdasarkan hasil analisis ulasan, diketahui bahwa aplikasi dreame lebih unggul daripada wattpad. Sedangkan untuk hasil tingkat akurasi wattpad sebesar 88,6 % dan dreame sebesar 87,45%. Penelitian [15] menganalisis sentimen ulasan aplikasi mola dengan menggunakan algoritma support vector machine. Pada penelitian ini menggunakan knowledge discovery in database (KDD) sebagai metodenya. Data yang diolah akan dibagi ke dalam 3 skenario kemudian dilakukan pembobotan. Hasil penelitian menunjukkan bahwa hasil terbaik pada skenario 1 dengan data training sebesar 90% dan data testing 10% serta memiliki tingkat akurasi sebesar 92,31%. Penelitian [16] melakukan analisis sentimen pada platform twitter mengenai sebuah maskapai penerbangan. Proses analisis dilakukan dengan menggunakan metode SVM yang pengujiannya dilakukan dengan instrumen confusion matrix dan diterapkan pada kernel linear, polynomial, RBF, dan sigmoid. Data yang digunakan dibagi menjadi tiga kelas yaitu positif, negatif, dan netral. Hasil penelitian menunjukkan nilai akurasi terbesar dimiliki oleh kernel RBF sebesar 84,73%.

Sedangkan penelitian [17] melakukan analisis sentimen pada platform Twitter dengan menggunakan metode SVM dan Naïve Bayes. Dataset yang digunakan merupakan kumpulan postingan twitter terhadap vaksin Covid-19. Penelitian ini dilakukan untuk membandingkan tingkat akurasi, presisi, recall, dan waktu training antara kedua metode yang digunakan. Hasil penelitian menunjukkan bahwa performa SVM berupa tingkat akurasi, presisi, dan recall lebih tinggi dibandingkan metode Naïve Bayes. Sedangkan bila dilihat dari segi waktu maka algoritma Naïve Bayes lebih efisien dibandingkan dengan metode SVM. Lalu, penelitian [18] melakukan analisis sentimen terhadap aplikasi ruang guru di Twitter dengan menggunakan metode Naïve Bayes, SVM, K-NN. Pada penelitian ini, metode yang digunakan adalah Cross-Industry Standard Process for Data Mining (CRISP-DM). Proses evaluasi dilakukan dengan membandingkan performa algoritma tanpa feature selection dan performa algoritma menggunakan feature selection. Hasil penelitian menunjukkan penggunaan feature selection menyebabkan meningkatnya performa dan akurasi algoritma. Dari beberapa algoritma yang digunakan, SVM dengan feature selection memiliki nilai akurasi dan performa yang paling tinggi.

Dari beberapa penelitian sebelumnya yang pernah dilakukan, diketahui bahwa keterkaitan dengan penelitian ini adalah analisis sentimen dengan menggunakan metode Support Vector Machine (SVM). Sedangkan perbedaannya adalah analisis sentimen dilakukan terhadap aplikasi pemerintahan yang mendukung keberhasilan smart city di Indonesia.

3. Metode Penelitian

Pada penelitian ini akan menggunakan metode KDD (Knowledge Discovery in Database). Metode ini sering digunakan dalam penelitian terkait data mining karena memiliki proses identifikasi data yang terorganisir sehingga data yang dihasilkan lebih mudah dipahami. Alur penelitian dengan menggunakan metode KDD dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

Alur penelitian pada gambar 1 dapat dijelaskan sebagai berikut.

a. Data Selection

Pada tahap ini akan dilakukan pengumpulan data dari Google Play Store. Kemudian juga akan dilakukan seleksi dan pelabelan data.

b. Pre-processing

Data mentah yang sudah dikumpulkan, selanjutnya akan diolah melalui tahapan preprocessing. Dimana tahap preprocessing memiliki 6 proses yaitu :

- 1) Cleaning, merupakan proses menghilangkan atribut kata yang tidak berpengaruh seperti tanda baca atau karakter kosong.
- 2) Case Folding, merupakan proses perubahan semua huruf yang ada menjadi huruf kecil
- 3) Tokenizing, merupakan proses yang memisahkan kalimat menjadi beberapa bagian kata
- 4) Normalization, merupakan proses perbaikan kesalahan seperti penggunaan ejaan agar setiap kata dapat memiliki makna yang setara
- 5) Stopword Removal, merupakan proses penghilangan kata yang dianggap tidak berpengaruh
- 6) Stemming, merupakan proses pengubahan kata-kata menjadi bentuk kata dasar.

c. Transformation

Pada tahap ini, data akan dibagi ke dalam 3 skenario, dengan skenario 1 (90% data latih dan 10% data uji), skenario 2 (80% data latih dan 20% data uji), skenario 3 (70% data latih dan 30% data uji). Lalu dilakukan pembobotan kata dengan TF-IDF.

d. Data Mining

Tahap data mining ini akan dilakukan klasifikasi sentimen ulasan pengguna dengan menggunakan empat kernel algoritma Support Vector Machine yaitu kernel linear, kernel RBF, kernel sigmoid, dan kernel polynomial.

e. Evaluation

Evaluasi akan dilakukan dengan menggunakan confusion matrix untuk mengetahui performansi dari setiap kernel. Hasil evaluasi akan berupa nilai *accuracy*, *precision*, *recall*, dan *f1-score*

4. Hasil dan Pembahasan

a. Data Selection

Data ulasan aplikasi SIGNAL dikumpulkan melalui google Play Store yang dilakukan dengan teknik scraping. Data ulasan yang berhasil dikumpulkan sebanyak 750 data, selanjutnya akan dilakukan pelabelan berdasarkan rating yang diberikan. Apabila rating < 3 akan masuk dalam label negatif sedangkan ulasan dengan rating lain maka akan masuk dalam label positif. Untuk label positif akan disimbolkan dengan angka 1 sedangkan label negatif disimbolkan oleh angka 0.

	content	score	sentiment
371	setiap mau melakukan pembayaran selalu error s...	2	0
294	Udah 2 kali download aplikasi ini,,,masih sama...	1	0
387	Setelah saya coba, saat ambil gambar tidak tau...	1	0
170	Sudah foto berapa kali tetap tidak bisa dengan...	1	0
213	Tidak bisa daftar,disuruh ke dukcapil.. Tapi s...	1	0

Gambar 2. Hasil Data Selection

b. Pre-processing

Tahap ini dilakukan pembersihan data terhadap data yang sudah dikumpulkan agar data tersebut dapat diolah. Contoh proses pembersihan yang dilakukan dapat dilihat pada tabel berikut.

Tabel 1 Hasil Pre-processing

Proses	Hasil
Data Ulasan	Seneng banget pakai apk SIGNAL, lebih mudah proses pembayaran pajaknya dan berkas Fisik diantar sampai ke tempatnya
Cleaning	Seneng banget pakai apk SIGNAL lebih mudah proses pembayaran pajaknya dan berkas Fisik diantar sampai ke tempatnya
Case Folding	seneng banget pakai apk signal lebih mudah proses pembayaran pajaknya dan berkas fisik diantar sampai ke tempatnya
Tokenizing	"seneng" "banget" "pakai" "apk" "signal" "lebih" "mudah" "proses" "pembayaran" "pajaknya" "dan" "berkas" "fisik" "diantar" "sampai" "ke" "tempatnya"
Normalization	"senang" "banget" "pakai" "aplikasi" "signal" "lebih" "mudah" "proses" "pembayaran" "pajaknya" "dan" "berkas" "fisik" "diantar" "sampai" "ke" "tempatnya"
Stopword Removal	"senang" "banget" "pakai" "aplikasi" "signal" "lebih" "mudah" "proses" "pembayaran" "pajaknya" "dan" "berkas" "fisik" "diantar" "sampai" "tempatnya"
Stemming	"senang" "banget" "pakai" "aplikasi" "signal" "lebih" "mudah" "proses" "bayar" "pajak" "dan" "berkas" "fisik" "antar" "sampai" "tempat"

c. Transformation

Pada tahap ini akan dilakukan pembagian data dan pembobotan TF-IDF. Pembagian data dilakukan dengan membagi menjadi dua kelompok yaitu data test dan data training. Kedua data tersebut nantinya akan dibagi lagi menjadi 3 skenario. Untuk hasil pembagian skenario dari data testing dan data training ditunjukkan pada tabel 2.

Tabel 2 Hasil Pembagian Data

Skenario Perbandingan	Data Training	Data Testing
90:10	675	75
80 : 20	600	150
70 : 30	525	225

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Gambar 2. Hasil Pembobotan TF-IDF

Langkah selanjutnya adalah melakukan pembobotan dengan metode TF-IDF. Pembobotan dilakukan terhadap setiap kata yang pada ulasan aplikasi SIGNAL. Gambar 3 merupakan contoh pembobotan TF-IDF pada data training 2.

d. Data Mining

Tahap ini merupakan tahap yang melakukan klasifikasi sentimen terhadap data ulasan aplikasi SIGNAL dengan membagi ke dalam 3 skenario. Proses klasifikasi dengan metode SVM ini akan dilakukan menggunakan empat kernel. Dimana hasil akurasi tiap kernel terhadap tiap skenario akan dipaparkan pada tabel 3.

Tabel 2 Hasil Klasifikasi SVM

Skenario	Kernel	Akurasi
90 : 10	<i>Linear</i>	88,79%
	RBF	93,46%
	<i>Polynomial</i>	97,2%
	<i>Sigmoid</i>	85,98%
80 : 20	<i>Linear</i>	88,26%
	RBF	91,55%
	<i>Polynomial</i>	93,58%
	<i>Sigmoid</i>	88,73%
70 : 10	<i>Linear</i>	91,22%
	RBF	92,48%
	<i>Polynomial</i>	94,04%

- Technol. ICCIT 2019*, no. December, pp. 1–5, 2019, doi: 10.1109/ICCIT48885.2019.9038348.
- [5] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepny Maulana Husain, “Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, doi: 10.29207/resti.v3i3.1118.
- [6] F. Zamachsari, G. Vangeran Saragih, and W. Gata, “Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection Algoritma Naive Bayes dan Support Vector Machine,” *Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 2, pp. 504–512, 2020.
- [7] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [8] 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.
- [9] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm,” *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [10] Styawati, N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.
- [11] E. Fitri, Y. Yuri, S. Rosyida, and W. Gata, “Sentiment Analysis of the Ruangguru Application Using Naive Bayes, Random Forest and Support Vector Machine Algorithms,” *J. Transform.*, vol. 18, no. 1, pp. 71–80, 2020.
- [12] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter,” *JOISM J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 97–108, 2021.
- [13] D. Haryadi and R. Mandala, “Prediksi Harga Minyak Kelapa Sawit Dalam Investasi Dengan Membandingkan Algoritma Naive Bayes, Support Vector Machine dan K-Nearest Neighbor,” *IT Soc.*, vol. 4, no. 1, 2019, doi: 10.33021/itfs.v4i1.1181.
- [14] S. I. Nurhafida and F. Sembiring, “Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM),” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 317–327, 2022.
- [15] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm,” *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.
- [16] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [17] F. Fitriana, E. Utami, and H. Al Fatta, “Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial

- Twitter Menggunakan Support Vector Machine dan Naive Bayes,” *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [18] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.