

ANALISIS SENTIMEN TERHADAP LAYANAN SAMSAT DIGITAL NASIONAL (SIGNAL) MENGGUNAKAN METODE SVM

Slamet Kacung^{*1}, Caesare Pamungkas Putra Bagyana², Dwi Cahyono³
^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Dr. Soetomo
slamet@unitomo.ac.id

ABSTRAK

Peningkatan signifikan dalam penggunaan pelayanan publik secara digital telah terjadi dalam kurun waktu tiga tahun terakhir. Perubahan ini dipicu oleh dampak wabah *Covid-19* di Indonesia, yang mendorong sektor bisnis dan pelayanan untuk berinovasi guna terus memenuhi kebutuhan masyarakat. Dalam hal ini, Korps Lalu Lintas POLRI telah meluncurkan aplikasi bernama Samsat Digital Nasional (SIGNAL) untuk memudahkan pembayaran pajak kendaraan bermotor, menunjukkan komitmen dalam memperbaiki pelayanan publik melalui platform digital. Namun, dalam prakteknya, beberapa pengguna sering mengalami kendala seperti notifikasi kesalahan ketika melakukan pendaftaran seperti verifikasi wajah, kode otp dan verifikasi KTP, kemudian bug dalam penggunaan aplikasi seperti perizinan aplikasi dalam mengakses kamera pada beberapa versi android, bahkan kegagalan transaksi pembayaran saat menggunakan aplikasi SIGNAL. Menghadapi tantangan ini, sentiment analysis digunakan untuk mengevaluasi persepsi masyarakat. Selain itu, hasil prediksi dari tiga kernel yang terdapat dalam metode klasifikasi *Support Vector Machine* (SVM) dan kernel Linier mencapai 96.2%, RBF mencapai 94.12%, *polynomial* mencapai 85.5%. Namun hasil evaluasi menggunakan KFold linier mencapai 97.65%, KFold RBF mencapai 97.86%, dan KFold *polynomial* sebesar 69.36%.

Keyword : Sentiment analysis, Lexicon Based, TF-IDF, Support Vector Machine.

1. PENDAHULUAN

Perkembangan teknologi yang pesat membawa perubahan dan kemudahan bagi masyarakat dalam menjalankan aktivitas. Pada tahun 2021-2022 penduduk Indonesia yang terkoneksi oleh internet mencapai sekitar 210 juta jiwa dari total populasi 270 juta jiwa penduduk Indonesia tahun 2021, meningkatnya pengguna internet di dasari oleh berkembangnya sarana dan prasarana lingkungan sekitar terhadap era globalisasi industri 4.0 [1]. Seiring berkembangnya internet dan melonjaknya pengguna internet dari tahun ke tahun, dalam hal ini pemerintah Indonesia dalam bidang Teknologi Informasi melalui instansi Korps Lalu Lintas POLRI meluncurkan aplikasi SIGNAL (Samsat Digital Nasional) yang bertujuan untuk memudahkan masyarakat dalam pelayanan Samsat secara digital [2]. Namun sejak dipublikasikan, aplikasi tersebut muncul beberapa *error* dan *bug* yang mempengaruhi pelayanannya. Sehingga dalam penelitian ini perlu dilakukannya kajian terkait analisa sentimen terhadap komentar masyarakat terhadap aplikasi SIGNAL yang terdapat pada penyedia aplikasi *Play Store* dengan menerapkan machine learning klasifikasi menggunakan metode SVM dan pelabelan ulasan menggunakan metode pelabelan kamus *Lexicon* berbahasa Indonesia. Analisa sentimen merupakan proses untuk melihat suatu kecenderungan opini publik [3] terhadap topik apakah memiliki pandangan baik (positif), buruk (negatif), atau tidak berpihak (netral).

2. TINJAUAN PUSTAKA

Penelitian terkait analisis sentimen yang memanfaatkan penyedia aplikasi *Play Store* sebagai

datasetnya telah dilakukan oleh sejumlah peneliti dalam berbagai karya ilmiah. Penelitian [4] berfokus pada aplikasi Mobile JKN dan menerapkan parameter sentimen positif dan negatif. Evaluasi menggunakan metode *Maximum Entropy* (ME) menunjukkan akurasi sebesar 78.5%, presisi 78.49%, recall 81.53%, dan F-Measure 79.98%. Kombinasi metode ME dengan Metode Seleksi Fitur *Mutual Information* (MI) menghasilkan akurasi tertinggi sebesar 84.5% dan akurasi rata-rata 78.9%.

Penelitian lainnya dilakukan pada aplikasi Peduli Lindungi dan membandingkan metode *Naïve Bayes* (NB) dan *K-Nearest Neighbors* (KNN) [5]. Metode NB mencapai akurasi 70.48%, recall positif 68.57%, recall negatif 72.38%, presisi positif 71.29%, dan presisi negatif 69.72%. Sementara itu, metode KNN memiliki akurasi 73.33%, recall positif 71.49%, recall negatif 75.24%, presisi positif 74.25%, dan presisi negatif 72.48%. Hasil perbandingan menunjukkan keunggulan metode KNN dibandingkan dengan metode NB.

Penelitian berikutnya pada aplikasi mobile Samsat Mobile Jawa Barat (Sambara) dengan menerapkan algoritma *NB Classifier* dan seleksi fitur *Chi Square* (CS) [6]. Data latih dan data uji dibagi dengan rasio 80:20. Hasil penelitian menggunakan seleksi Fitur CS dengan nilai kritis 0.837 dan 244 fitur terseleksi mencapai akurasi tertinggi sebesar 94.4%, spesifisitas 93.5%, dan recall 95.3%.

Penelitian ini menggunakan metode *Support Vector Machine* (SVM) untuk mengevaluasi sejauh mana layanan aplikasi tersebut diterima oleh masyarakat umum sebagai pengguna, memberikan umpan balik kepada instansi terkait untuk

meningkatkan kualitas layanan. Penggunaan SVM dipilih karena algoritma ini dinilai memiliki akurasi yang baik dan tidak memiliki nilai *outlier* [7] di masa depan dan membandingkan performa dari 3 kernel SVM (*Support Vector Machine*) yaitu kernel Linear, Fungsi kernel linier adalah *dot product* antara dua vektor fitur dengan menggunakan persamaan 1 [8].

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i \langle x_i, x \rangle + b) \quad (1)$$

Radial Basis Function (RBF), RBF Kernel adalah salah satu kernel yang paling umum digunakan dalam SVM [9]. Ini memungkinkan SVM untuk menangani data yang tidak linier dengan sangat baik. RBF Kernel mengukur kesamaan berdasarkan jarak antara titik data dalam ruang fitur. Ini adalah salah satu kernel yang paling serbaguna dan efektif. Persamaan SVM kernel RBF dapat dilihat pada persamaan 2 [10].

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

Kernel Polynomial digunakan ketika data tidak dapat dipisahkan secara linier [11]. Ini mengubah data ke dalam dimensi yang lebih tinggi dengan polinomial dan mencoba memisahkan kelas menggunakan polinomial hyperplane. Persamaan SVM kernel Polynomial dapat dilihat pada persamaan 4 [10].

$$K(x_i, x_j) = ((x_i x_j) + c)^d \quad (3)$$

Dengan membandingkan performa pada tiga kernel tersebut dapat menyimpulkan kernel

manakah yang menghasilkan akurasi terbaik dan memberikan wawasan yang penting mengenai pendapat pengguna terhadap aplikasi SIGNAL.

3. METODE PENELITIAN

Metode SVM dalam penelitian ini digunakan untuk mengelompokkan data yang memiliki pola linier dan non linier [12] yang bertujuan untuk mengetahui sentimen pengguna aplikasi SIGNAL melalui ulasan yang diberikan di penyedia aplikasi *Play Store*. Metode penelitian yang digunakan diilustrasikan dalam Gambar 1, melalui beberapa tahapan, yaitu *scrapping* data ulasan, *pre-processing* dataset, *splitting* data untuk pelatihan dan pengujian, penerapan model machine learning SVM dan evaluasi kinerja SVM.



Gambar 1. Tahapan Penelitian (*Preprocessing*)

3.1. Scrapping Data Ulasan

Proses pengumpulan data dalam penelitian ini dilakukan dengan metode *scrapping* langsung terhadap aplikasi penyedia aplikasi Android. Data ulasan pengguna aplikasi SIGNAL diambil berdasarkan kategori ulasan yang paling relevan, terbaru, dan berbahasa Indonesia sebanyak 10000 data ulasan. Hasil dari ulasan terlihat pada Tabel 1.

Tabel 1. *Scrapping* ulasan SIGNAL di *Play Store*

User Name	Star	At	Content
User 1	3	1/10/2024 10:40	sudah bayar pajak kendaraan tapi belum verifikasi
User 2	5	1/9/2024 14:02	Sangat mudah digunakan dan proses gampang
User 3	5	6/10/2023 22:54	sangat mempermudah, bagi yang sedang merantau
User 4	2	1/2/2024 19:58	Susah daftar aplikasinya, verifikasi gagal terus
User 5	4	6/17/2023 17:49	Lumayan memudahkan namun proses nya lama
...	...	...	...
User 99	1	7/25/2023 10:07	email verifikasi kok ga dikirim kirim udah dicoba berulang kali tetep ga dikirim
User 100	2	1/27/2023 14:52	Tidak bisa melakukan pembayaran via Online Dana, LinkAja, Shopee semua terjadi kesalahan
User 101	1	1/22/2023 14:52	Buat device android 13 susah buat daftar nya. Selalu mentok di ijin pengaturan kamera
User 102	3	7/25/2023 21:52	proses dari daftar sampai dengan STNK ke tangan pengendara lumayan lama

3.2. Pre-processing Dataset

Tahap *pre-processing* adalah tahapan yang penting sebelum dilakukan pengujian dataset terhadap machine learning. Tujuan dari *pre-processing* dataset adalah untuk membersihkan, mengatur, dan mempersiapkan data mentah sehingga dapat diolah dengan lebih efektif oleh

machine learning. Tahapan *pre-processing* ini meliputi *Cleaning Data*, *Tokenizing* terlihat pada Tabel 2.

Tabel 2. Proses Cleaning Text dan Tokenize

Cleaning	Text Tokenize
sudah bayar pajak kendaraan tapi belum verifikasi	['sudah', 'bayar', 'pajak', 'kendaraan', 'tapi', 'belum', 'verifikasi']
sangat mudah digunakan dan proses gampang	['sangat', 'mudah', 'digunakan', 'dan', 'proses', 'gampang']
susah daftar aplikasinya verifikasi gagal terus	['susah', 'daftar', 'aplikasinya', 'verifikasi', 'gagal', 'terus']

Selanjutnya yaitu proses *Normalization*, *Stopword Removal* terlihat pada Tabel 3.

Tabel 3. Proses Normalisasi dan Stopwords

Text Normalized	Text Stopword
['sudah', 'bayar', 'pajak', 'kendaraan', 'tapi', 'belum', 'verifikasi']	['bayar', 'pajak', 'kendaraan', 'verifikasi']
['sangat', 'mudah', 'digunakan', 'dan', 'proses', 'gampang']	['mudah', 'proses', 'gampang']
['susah', 'daftar', 'aplikasinya', 'verifikasi', 'gagal', 'terus']	['susah', 'daftar', 'aplikasinya', 'verifikasi', 'gagal']

Setelah proses Normalisasi dan *Stopwords* selanjutnya *Stemming* terlihat pada Tabel 4.

Tabel 4. Proses Stemming

Text Stemming	Text Clean
['bayar', 'pajak', 'kendaraan', 'verifikasi']	bayar pajak kendaraan verifikasi
['mudah', 'proses', 'gampang']	mudah proses gampang
---	---
['susah', 'daftar', 'aplikasi', 'verifikasi', 'gagal']	susah daftar aplikasi verifikasi gagal

Setelah dilakukan *stemming*, proses selanjutnya adalah *labeling* sebelum proses SVM terlihat pada Tabel 5.

Tabel 5. Proses Pelabelan menggunakan kamus lexicon

Text Clean	Score	Sentiment
bayar pajak kendaraan verifikasi	0	neutral
mudah proses gampang	2	positive
---	---	---
susah daftar aplikasi verifikasi gagal	-2	negative

3.3. Visualisasi Hasil Pre-processing

Dataset yang telah diproses pada tahap *pre-processing* akan divisualisasikan menjadi grafik berdasarkan sentimennya untuk memberikan pemahaman yang lebih baik tentang pola sentimen yang ada dalam data tersebut dengan menggunakan teknik visualisasi seperti *Wordcloud*, *Diagram*

Batang dan *Diagram lingkaran (Pie chart)*. Sehingga, kita dapat melihat distribusi sentimen positif, negatif, dan netral

3.4. Split Data Latih & Uji

Dataset hasil *pre-processing* yang akan digunakan untuk melakukan klasifikasi SVM akan dibagi dengan rasio 70:30, di mana 70% datanya akan digunakan sebagai data latih, dan 30% sisanya akan digunakan sebagai data uji

3.5. Ekstraksi Fitur TF-IDF

Proses *TF-IDF (Term Frequency Inverse Document Frequency)* merupakan langkah kunci dalam pengolahan teks untuk pemodelan klasifikasi [13]. Dalam tahap ini kata-kata dalam dokumen diubah menjadi representasi numerik yang kemudian diproses oleh *machine learning*

3.6. Pengujian Data

Pengujian data dilakukan dengan menggunakan algoritma machine learning *Support Vector Machine Classifier* yang terdapat pada *library python scikit-learn* dengan membandingkan performa dari 3 kernel SVM yaitu kernel *linear*, *RBF* dan *polynomial* [9].

3.7. Evaluasi

Evaluasi hasil analisis penelitian dengan menggunakan *Confusion Matrix* melibatkan perbandingan prediksi model dengan nilai sebenarnya dalam bentuk tabel, yang memudahkan untuk mengidentifikasi kesalahan klasifikasi [14]. Dalam tabel ini, setiap elemen menunjukkan jumlah data yang diprediksi sesuai dengan kelas sebenarnya. Berikut ini adalah contoh dari confusion matrix yang terdapat pada Tabel 6.

Tabel 6. Confusion Matrix

Data Actual	Data Prediksi		
	Negative	Neutral	Positive
Negative	TN	NG	FN
Neutral	NN	TN	NP
Positive	FP	PN	TP

4. HASIL DAN PEMBAHASAN

4.1. Dataset Ulasan

Dataset ulasan pengguna aplikasi SIGNAL diambil berdasarkan kategori ulasan yang paling relevan, terbaru, dan berbahasa Indonesia sebanyak 10000 data ulasan dengan tujuan untuk menentukan algoritma terbaik dari ketiga kernel yaitu kernel linier, kernel RBF, dan kernel *polynomial* dengan menggunakan metode SVM. Sebagian Data ulasan hasil *pre-processing* terlihat pada Tabel 7.

Tabel 7. Dataset Ulasan Pengguna

userName	text_clean
User 1	stnk asli kirim alamat ya nya pilih pengiriman kantor pos ganggu
User 2	kode bayar lewat jam tentu kadaluarsa ya coba bayar mitra tokopedia
User 3	cepat praktis antri mana bayang
User 4	proses cetak kirim cepat muas mudah bayar pajak kendaraan
User 5	terimakasih layan bagus cuman kasih masuk proses disamsatnya bantu
User 6	aplikasi bantu tingkat performa nya
User 7	puas bantu jam kerja digunainnya
User 8	daftar sampai hasil kesah singkat kerja good job
User 9	aplikasi bantu taat pajak samsat mager antri loket samsat

Data ulasan pengguna pada Tabel 7. Diatas kemudian dilanjutkan proses pelabelan dataset sebagaimana terlihat pada Tabel 8.

Tabel 8. Pelabelan Dataset

text_clean	Skor	Label
stnk(0) asli(0) kirim(0) alamat(0) stnk(0) ya(0) nya(0) pilih(1) pengiriman(0) kantor(0) pos(0) ganggu(-1)	0	Neutral
kode(0) bayar(0) lewat(0) jam(0) tentu(0) kadaluarsa(-1) ya(0) coba(0) bayar(0) mitra(0) tokopedia(0)	-1	Negative
cepat(1) praktis(1) antri(0) mana(0) bayang(0)	2	Positive
proses(0) cetak(0) kirim(0) cepat(1) muas(0) mudah(1) bayar(0) pajak(0) kendaraan(0)	2	Positive
terimakasih(0) layan(0) bagus(1) cuman(0) kasih(1) masuk(0) proses(0) disamsatnya(0) bantu(1)	3	positive
aplikasi(0) bantu(1) tingkat(0) performa(0) nya(0)	1	positive
puas(1) bantu(1) jam kerja(1) digunainnya(0)	3	positive
daftar(0) sampai(0) hasil(0) kesah(0) singkat(0) kerja(1) good(0) job(0)	1	positive
aplikasi(0) bantu(1) taat(1) pajak(0) samsat(0) mager(0) antri(0) loket(0) samsat(0)	2	positive

4.2. Tahap Pengujian

Pengujian dilakukan dengan menggunakan metode klasifikasi *Support Vector Machine* [15] dengan membagi dataset 70% untuk data latih dan 30% untuk data uji yang melibatkan tiga kernel, yaitu kernel linier, kernel RBF, dan kernel *polynomial*.

Hasil evaluasi score setiap kernel terlihat pada tabel 9. Dan evaluasi dengan melibatkan KFold Crossvalidation terlihat pada tabel 10.

Tabel 9. Hasil Prediksi Kernel

Kernel	Accuracy	Precision	Recall	F1Score
Linier	96.19	95.13	94.74	94.92
RBF	94.12	91.52	92.67	91.92
Polynomial	85.5	77.67	84.32	80.22

Tabel 10. Hasil Prediksi dengan K Fold Cross Validation

Kernel	K1	K2	K3	K4	K5	Mean
Linier	97.1	97.5	98	97.9	97.7	97.65
RBF	97.3	97.6	98.0	98.2	98.2	97.86
Polynomial	94.2	95.1	97.2	98.0	97.3	96.36

5. KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa penggunaan metode *Support Vector Machine* (SVM) dengan kernel linear menghasilkan akurasi tertinggi dalam analisis sentimen terhadap layanan aplikasi SIGNAL (Samsat Digital Nasional) berdasarkan ulasan di Google Play Store pada skenario rasio 70:30. Tingkat akurasi yang berhasil dicapai adalah sebesar 96.19%, dengan konsistensi evaluasi yang tinggi menggunakan rata-rata KFold mencapai 97.65%.

Meskipun penggunaan kernel SVM dengan tipe RBF menghasilkan akurasi sedikit lebih rendah pada skenario yang sama, yaitu sebesar 94.12%, evaluasi rata-rata KFold menunjukkan kestabilan model dengan nilai sebesar 97.86%. Namun, penerapan kernel SVM dengan tipe *Polynomial* menghasilkan akurasi yang lebih rendah, yakni sebesar 85.5%, dengan nilai rata-rata KFold mencapai 96.36%. Hal ini mengindikasikan bahwa model mungkin kurang cocok untuk menangkap pola kompleks yang mungkin terdapat dalam dataset. Secara keseluruhan, penelitian ini merekomendasikan penggunaan metode SVM dengan kernel linear untuk mendapatkan *accuracy* terbaik dan RBF sebagai *K Fold Cross Validation* terbaik sebagai pilihan yang tepat dalam melakukan analisis sentimen terhadap layanan aplikasi SIGNAL.

DAFTAR PUSTAKA

- [1] M. Arif, "Profil Internet Indonesia 2022," *SRA Consult.*, no. June, pp. 1–104, 2022.
- [2] R. F. Meuthia, Ferdawati, Gustati, D. Handayani, and R. E. Sulastri, "Pengaruh Security, Privacy dan Familiarity terhadap Minat Penggunaan Aplikasi Samsat Digital Nasional (Signal) dengan E-Trust sebagai Variabel Intervening," *J. Akunt. Keuang. dan Bisnis*, vol. 15, no. 2, pp. 568–577, 2023.
- [3] R. A. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, and L. Efizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 211–218, 2023.
- [4] A. Filemon, H. Kaban, and N. Yudistira,

- “Analisis Sentimen Aplikasi E-Government berdasarkan Ulasan Pengguna menggunakan Metode Maximum Entropy dan Seleksi Fitur Mutual Information,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 1452–1458, 2021.
- [5] A. P. Wibowo, W. Darmawan, and N. Amalia, “Komparasi Metode Naïve Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi,” *IC-Tech*, vol. 17, no. 1, pp. 18–23, 2022.
- [6] 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.
- [7] H. Azis *et al.*, “Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah,” vol. 19, no. 3, pp. 286–294, 2020.
- [8] A. Z. Praghakusma and N. Charibaldi, “Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi),” *JSTIE (Jurnal Sarj. Tek. Inform.*, vol. 9, no. 2, p. 88, 2021.
- [9] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 2nd Edition (2019)*, vol. 44, no. 8, 2011.
- [10] R. Pambudi, “Deteksi Penggunaan Masker dengan Algoritma RBF Support Vector Machine,” *J. Mach. Learn. Comput. Intell.*, pp. 34–37, 2022.
- [11] C. J. C. BURGESS, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Scand. J. Med. Sci. Sport.*, vol. 28, no. 3, pp. 1056–1063, 2018.
- [12] J. Cervantes, X. Li, and W. Yu, “Splice site detection in DNA sequences using a fast classification algorithm,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, no. October, pp. 2683–2688, 2009.
- [13] S. Tedmori and A. Awajan, “Sentiment analysis main tasks and applications: A survey,” *J. Inf. Process. Syst.*, vol. 15, no. 3, pp. 500–519, 2019.
- [14] K. Saputra, “Perbandingan Kinerja Fungsi Kernel Algoritma Support Vector Machine Pada Klasifikasi Penyakit Padi,” *Ijccs*, vol. x, No.x, no. x, pp. 1–5, 2023.
- [15] M. A. Karyawan, A. Zainal, A. Saikhu, I. Teknologi, and S. Nopember, “KLASIFIKASI SINYAL EEG MENGGUNAKAN KOEFISIEN AUTOREGRESIF,” vol. 2, no. 1, pp. 1–13, 2011.