

Efektifitas Hasil Analisis Sentimen Aplikasi SIGNAL Berbasis Lexicon-Based dan Random Forest

Nayra Zanetti Windy Rahmantya^{a1}, I Gusti Ngurah Anom Cahyadi Putra^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹nayra.zanetti05@gmail.com
²anom.cp@unud.ac.id

Abstract

SIGNAL (Samsat Digital Nasional) is a digital innovation developed by the Indonesian National Police to simplify vehicle tax payments, STNK validation, and other administrative services online. As the number of users grows, various user opinions are reflected in the form of reviews on the Google Play Store. The research adopts a lexicon-based approach by extracting positive and negative keywords directly from the dataset to classify sentiments in user-generated reviews. A sentiment label is assigned based on the frequency and dominance of positive or negative terms within each review. To evaluate the effectiveness of this lexicon-based classification, the Random Forest machine learning algorithm is employed as a benchmark. These findings indicate that the lexicon-based approach, when built from domain-specific vocabulary, can effectively classify sentiment with minimal computational resources while maintaining competitive performance. This research contributes to the development of lightweight sentiment analysis systems and highlights the potential of hybrid methods for enhancing accuracy.

Keywords: sentiment analysis, lexicon-based, random forest, SIGNAL application, user reviews, text classification.

1. Pendahuluan

Perkembangan teknologi digital telah mendorong instansi pemerintahan untuk menyediakan layanan publik berbasis aplikasi guna meningkatkan efisiensi dan aksesibilitas bagi masyarakat. Salah satu bentuk transformasi digital tersebut adalah kehadiran aplikasi SIGNAL (Samsat Digital Nasional), yang dikembangkan oleh Korlantas Polri untuk mempermudah masyarakat dalam melakukan pengesahan STNK, pembayaran pajak kendaraan bermotor, serta layanan administrasi kendaraan lainnya secara daring.

Seiring meningkatnya jumlah pengguna, aplikasi SIGNAL menerima berbagai ulasan dari masyarakat yang mencerminkan pengalaman, persepsi, serta tingkat kepuasan terhadap layanan yang diberikan. Ulasan tersebut, yang umumnya dipublikasikan melalui Google Play Store, mengandung informasi berharga yang dapat dianalisis untuk mengetahui bagaimana tanggapan publik terhadap kinerja aplikasi ini. Namun, ulasan dalam bentuk teks cenderung bersifat subjektif dan tidak terstruktur, sehingga diperlukan pendekatan sistematis untuk mengolah dan memahami sentimen yang terkandung di dalamnya.

Salah satu pendekatan yang dapat digunakan untuk menganalisis sentimen dari data teks adalah analisis sentimen berbasis kamus (lexicon-based). Metode ini tidak memerlukan pelatihan model seperti pada machine learning, sehingga lebih ringan secara komputasi dan mudah diterapkan. Dengan menggunakan daftar kata positif dan negatif, setiap ulasan dapat diklasifikasikan ke dalam kategori sentimen positif, negatif atau netral berdasarkan kata-kata yang terkandung di dalamnya. Meskipun pendekatan lexicon-based cukup efektif, untuk mengukur akurasi, tetap diperlukan perbandingan dan pengujian dengan algoritma Random Forest. Oleh karena itu, dalam penelitian ini juga diterapkan algoritma Random Forest sebagai

metode perbandingan dan pengujian. Random Forest merupakan algoritma klasifikasi yang cukup populer dan kuat karena menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat dengan melakukan tahap pengujian dan perbandingan akurasi.

Penelitian ini bertujuan untuk membandingkan performa pendekatan lexicon-based dengan metode Random Forest dalam menganalisis sentimen terhadap ulasan pengguna aplikasi SIGNAL. Dengan membandingkan kedua pendekatan, diharapkan dapat diketahui masing-masing akurasi dari metode serta memberikan gambaran tentang pendekatan yang paling sesuai. Selain itu, penelitian ini juga memberikan kontribusi terhadap pengembangan sistem analisis sentimen ringan dan efisien yang dapat digunakan untuk memantau kepuasan pengguna.

2. Metode Penelitian

2.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan *lexicon-based* untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi SIGNAL. Pendekatan ini bersifat *rule-based*, yaitu dengan mencocokkan kata-kata dalam ulasan terhadap daftar kata positif dan negatif yang telah ditentukan sebelumnya. Metode ini dipilih karena sederhana, efisien dan tidak memerlukan proses pelatihan model.

Selain itu juga digunakan algoritma *Random Forest* untuk melakukan pengujian dari pendekatan yang dilakukan oleh *lexicon-based*. Algoritma ini hanya berfungsi sebagai penguji dan perbandingan hasil akurasi antara menggunakan pendekatan *lexicon-based* atau *random forest*.

2.2 Sumber Data

Data diperoleh dari ulasan pengguna aplikasi SIGNAL di Google Play Store, yang diambil melalui *platform* dataset yaitu *Kaggle*. Jumlah data yang dianalisis sebanyak 1418 ulasan yang diambil dari salah satu dataset *Kaggle* yang tersedia dan terakhir diperbarui pada 10 bulan yang lalu. Kemudian data disimpan dalam format CSV dengan tiga atribut utama:

- username (nama pengguna)
- at (tanggal memberikan ulasan)
- content (isi ulasan)
- score (penilaian bintang 1–5)

Tabel 1. Contoh dataset hasil ulasan aplikasi SIGNAL

username	score	at	content
Pengguna Google	5	2024-09-03 06:54:42	Mantapp, bisa bayar pajak online, surat diantar, jossss semoga makin sukses
Pengguna Google	5	2024-09-03 03:26:45	Saran utk login menggunakan NIK saja, karena klo login pake no hp kadang suka lupa hp yg mana yg di input. Kadang no hp juga suka ganti krn ke blokir atau hilang hp nya dll
Pengguna Google	5	2024-09-03 06:35:15	Mantap GBU

2.3 Preprocessing Data Teks

Untuk mendapatkan hasil yang lebih akurat, data teks ulasan terlebih dahulu dibersihkan melalui tahapan *preprocessing*, yaitu :

- **Cleaning:** Menghapus karakter-karakter yang tidak relevan dalam sistem (seperti emoji, tanda baca ataupun simbol-simbol khusus).
- **Normalization:** Membuat seluruh kalimat menjadi seragam (seperti mengubah semua ke dalam huruf kecil)
- **Tokenization:** Memisahkan kalimat menjadi kata per kata.
- **Stopword:** Menghapus kata-kata umum dan tidak memiliki makna penting jika dihilangkan.
- **Stemming:** Mengubah kata menjadi bentuk dasar.

Tabel 2. Proses *preprocessing*

ulasan	cleaning	normalization	tokenization	stopward	stemming
Mantapp,bisa	Mantapp bisa	mantapp bisa	"mantapp",	"bisa", "bayar",	"bisa", "bayar",
bayar pajak	bayar pajak	bayar pajak	"bisa", "bayar",	"pajak",	"pajak",
online, surat	online surat	online surat	"pajak",	"online",	"online",
diantar, jossss	diantar jossss	diantar jossss	"online",	"surat",	"surat", "antar",
semoga makin	semoga makin	semoga makin	"surat",	"diantar",	"sukses"
sukses	sukses	sukses	"diantar",	"sukses"	
			"jossss",		
			"semoga",		
			"makin",		
			"sukses"		

2.4 Klasifikasi Kata Sentimen

Daftar kata sentimen disusun secara manual berdasarkan *wordcloud* yang didapatkan dari hasil dataset ulasan pengguna aplikasi SIGNAL. Setiap ulasan dihitung jumlah kemunculan kata positif dan negatifnya. Sentimen ditentukan berdasarkan perbandingan jumlah kata dari masing-masing kategori ditentukan berdasarkan ambang batas (*threshold*):

- Skor Total > +1 → Positif
- Skor Total < -1 → Negatif
- -1 ≤ Skor Total ≤ +1 → Netral

Dan diperoleh hasil klasifikasi berdasarkan kata positif dan negative yang digambarkan melalui *wordcloud* dan dibagi menjadi 2 tipe klasifikasi yaitu :

Hasil pengujian pada data uji menunjukkan bahwa metode ini menghasilkan sentimen analisis dengan tingkat akurasi sebesar 58,18%.

Sudah 1 bulan stnk saya tidak dikirim	negatif	netral
Pelayanannya bagus	positif	positif
Cepat untuk perpanjangan STNK ga perlu datang ke samsat 1	positif	positif
Keren	positif	positif
Total data: 1418		
Prediksi benar: 825		
Akurasi: 58.18%		

Gambar 3. Hasil dari pengujian *Lexicon-Based*

b. Hasil Pengujian *Random Forest*

Sebagai pembandingan, dilakukan pengujian menggunakan algoritma *Random Forest*. Data teks diubah menjadi fitur numerik menggunakan teknik TF-IDF, kemudian dilatih dengan model *Random Forest*. Model ini diuji pada data yang sama dengan *lexicon-based*.

Hasil pengujian menunjukkan bahwa *Random Forest* mampu mengklasifikasikan ulasan secara benar, menghasilkan akurasi sebesar 88,32%.

	precision	recall	f1-score	support
Negatif	0.84	0.36	0.50	45
Positif	0.89	0.99	0.93	229
accuracy			0.88	274
macro avg	0.86	0.67	0.72	274
weighted avg	0.88	0.88	0.86	274
Akurasi Total: 88.32%				

Gambar 4 . Hasil dari pengujian *Random Forest*

3. Hasil dan Diskusi

Setelah dilakukan proses klasifikasi dan pengujian terhadap ulasan pengguna aplikasi SIGNAL menggunakan pendekatan *lexicon-based* dan Algoritma *Random Forest*, diperoleh hasil yang berbeda antara keduanya. Analisis dilakukan dengan total dataset 1418 baris dan didapatkan hasil perbandingan keduanya sebagai berikut.

3.1. Hasil Pendekatan *Lexicon-Based*

Tabel 3. Hasil dari pendekatan *Lexicon-Based*

Review	Label Asli	Hasil Prediksi
Jelek apk nya gk bs daftar	negatif	negatif
Mantab semoga semakin lancar.	positif	positif
GABISA DAFTAR	negatif	netral
Gak bisa di pakai error terus padahal sinyal bagus	negatif	netral
Ga bisa diganain	negatif	netral
Mau registrasi kok susah banget	netral	negatif

Pada pendekatan *lexicon-based*, klasifikasi sentimen dilakukan dengan menghitung skor berdasarkan jumlah kata positif dan negatif yang ditemukan dalam setiap ulasan. Kata-kata tersebut berasal dari kamus sentimen yang dibentuk berdasarkan frekuensi kata pada masing-masing label sentiment dan disimpan pada kamus kata pada program analisis yang terbagi menjadi 2 klasifikasi kata yaitu “positif” dan “negatif”, jika diluar dari klasifikasi kata tersebut maka program akan menghasilkan hasil “netral”. Pendekatan ini menghasilkan akurasi yang cukup buruk yaitu 58,18%.

3.1. Hasil Pengujian *Random Forest*

Tabel 4. Hasil dari pendekatan *Random Forest*

precision	recall	F1-score
0.84	0.36	0.50
0.89	0.99	0.93

Pada algoritma *Random Forest* menunjukkan akurasi yang jauh lebih tinggi. Model ini berhasil mengklasifikasikan sentiment ulasan pengguna aplikasi SIGNAL jauh lebih akurat sehingga dapat menghasilkan tingkat akurasi sebesar 88,32%. Metode ini menunjukkan keunggulan dalam memahami pola yang lebih kompleks dalam data yang jumlahnya besar.

4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa pendekatan *lexicon-based* dan menggunakan algoritma *random forest* yang dilakukan dengan dataset yang sama dan perhitungan klasifikasi. Dapat disimpulkan bahwa pendekatan *lexicon-based* efektif dalam mengklasifikasikan sentimen apabila didukung oleh kamus sentimen yang sesuai dengan domain data. Menurut hasil pengujian, jika hanya menggunakan pendekatan *lexicon-based* hanya memiliki akurasi dibawah 60% sedangkan jika menggunakan algoritma *random forest* mendapatkan hasil akurasi yang lebih tinggi yaitu 88%. Hal ini menunjukkan bahwa pendekatan machine learning dapat memberikan peningkatan performa akurasi. Penelitian ini bertujuan untuk mengetahui persepsi pengguna terhadap aplikasi SIGNAL (Samsat Digital Nasional) melalui pendekatan analisis sentimen berbasis *lexicon*.

Sentimen negatif umumnya berkaitan dengan kendala teknis seperti kesulitan login, error sistem, dan keterlambatan proses. Sebaliknya, sentimen positif banyak ditemukan pada ulasan yang menyoroti kemudahan penggunaan, kecepatan layanan, dan tampilan antarmuka yang nyaman.

Pendekatan *lexicon-based* terbukti cukup efektif dan efisien dalam mengklasifikasikan opini pengguna tanpa memerlukan pelatihan model atau dataset. Namun memiliki hasil akurasi yang sangat rendah jika dibandingkan dengan penggunaan algoritma *random forest*. Pendekatan ini sangat sesuai untuk analisis awal atau eksploratif terhadap opini publik secara cepat.

Daftar Pustaka

- [1] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, “Analisis Sentimen Menggunakan Metode Lexicon-Based Dan Support Vector Machine Pada Presiden Dan Wakil Presiden Indonesia Periode 2024–2029,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5604.
- [2] S. A. Nugraha, “Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara,” 2025.
- [3] Ahmad Fatihin, “Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine (Svm) Dan Pendekatan Lexicon Based,” 2022.
- [4] N. Y. Setyawati and E. Widarti, “Analisis Sentimen Aplikasi Samsat Digital di Play Store

- Menggunakan Support Vector Machine,” vol. 3, no. 2, 2024.
- [5] R. Chandra and E. M. Sipayung, “Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Menggunakan Algoritma Naive Bayes Classifier,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 3, pp. 156–164, Jan. 2025, doi: 10.25077/teknosi.v10i3.2024.156-164.
- [6] R. Chandra and E. M. Sipayung, “Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Menggunakan Algoritma Naive Bayes Classifier,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 3, pp. 156–164, Jan. 2025, doi: 10.25077/teknosi.v10i3.2024.156-164.

Halaman ini sengaja dibiarkan kosong