G-Tech : Jurnal Teknologi Terapan
Vol. 8, No. 4, Oktober 2024, hlmn. 2411-2419
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Komparasi Algoritma *Machine Learning* Untuk Menganalisis Sentimen Ulasan Pada Aplikasi Digital Korlantas Polri

Siti Delimasari[1]✉, Kusrini[2]

[1, 2] Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 01-08-2024
Direvisi : 15-08-2024
Diterima : 25-08-2024

ABSTRAK

Aplikasi Digital Korlantas Polri merupakan aplikasi *mobile* yang memberikan kemudahan bagi masyarakat dalam memperpanjang SIM. Analisis sentimen terhadap ulasan pengguna membantu Korlantas Polri mengidentifikasi persepsi publik terhadap layanan yang diberikan. Penelitian ini bertujuan mengevaluasi lima algoritma *machine learning* mana yang paling baik performanya dari *Support Vector Machine (SVM), Naive Bayes, Random Forest, K-Nearest Neighbors (KNN),* dan *Logistic Regression* dalam proses analisis sentimen. Evaluasi dilakukan dengan mengukur akurasi, presisi, *recall* dan F1 *measure.* Terdapat 10.000 ulasan diberi label dengan validasi ahli bahasa, diproses ulang, diberi bobot kata setelah data dikumpulkan. Teknik *over-sampling minoritas sintetis (SMOTE)* diterapkan sebelum pembagian data untuk pelatihan dan pengujian. Hasil evaluasi menunjukkan *Random Forest* dan SVM melakukannya dengan paling baik. *Random Forest* memiliki akurasi 90,77%, *recall* 90,77%, dan nilai F1 tertinggi yaitu 90,79%. SVM memiliki presisi tertinggi dengan 91,14% di antara algoritma lainnya, yang menunjukkan potensi besar kedua algoritma ini menganalisis sentimen ulasan aplikasi digital Korlantas Polri.

Kata Kunci:

Analisis Sentimen; Aplikasi Digital Korlantas Polri; *Machine Learning*

Keywords **:**

Sentiment Analysis, Digital Korlantas Polri Application, Machine Learning

ABSTRACT

Korlantas Polri Digital Application is one of the mobile applications that provides ease for the public in extending the driving license. Sentiment analysis of user reviews can help korlantas polri identify public perception of the given service. The study aims to evaluate which of the five machine learning algorithms performed best from Support Vector Machine (SVM), Naive Bayes, Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression in sentiment analysis. The evaluation was done by measuring accuracy, precision, recall and F1 measure. There were 10,000 reviews labelled with linguistic validation, re-processed, and word weighted after data was collected. Synthetic minority over-sampling techniques (SMOTE) are applied before data splitting for training and testing. The evaluation shows that Random Forest and SVM do the best. Random Forest has an accuracy of 90.77%, recall 90.77%, and its highest F1 rating is 90.79%. SVM has the highest precision with 91.14% among other algorithms, which shows the great potential of both of these algorithms in the analysis of sentiment reviews of digital applications Korlantas Polri.

Corresponding Author **:**
Siti Delimasari
Teknik Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia
Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, DIY
Email: sari@swu.ac.id

PENDAHULUAN

Aplikasi *mobile* semakin populer dalam kehidupan sehari-hari di era digital, termasuk layanan pemerintah seperti perpanjangan SIM. *Platform* resmi untuk memperpanjang SIM adalah aplikasi digital yang disediakan oleh Korlantas Polri (Kaburuan & Setiawan, 2023). Aplikasi tersebut saat ini telah diunduh oleh banyak pengguna di *Google Play Store*, dengan *rating* 3,6 dan 103 ribu ulasan positif, negatif, dan netral berdasarkan peringkat bintang atau *rating* yang diberikan pengguna aplikasi. *rating* bintang memberikan gambaran umum tentang kepuasan pengguna (Bahtiar et al., 2023).

Dalam konteks aplikasi digital Korlantas Polri, analisis sentimen dapat membantu mengidentifikasi persepsi publik terhadap layanan yang diberikan melalui ulasan pengguna. (Jardim & Mora, 2021).Ulasan Pengguna dievaluasi menggunakan beberapa algoritma *machine learning* untuk mengetahui performa terbaik dari hasil akurasi, presisi, *recall* serta nilai f1.

Penelitian sebelumnya oleh (Yu, 2024) menggunakan *Random Forest, Logistic Regression, Long Short-Term Memory (LSTM) networks* dan *Convolutional Neural Networks (CNN)* untuk memprediksi sentimen positif atau negatif dalam ulasan. Tujuan dari penelitian Yu untuk mengevaluasi efektivitas berbagai algoritma *machine learning* dalam menganalisis sentimen pelanggan melalui ulasan Amazon. Hasil penelitian menunjukkan bahwa *Logistic Regression* menunjukkan kinerja terbaik dengan akurasi tertinggi, diikuti oleh *Random Forest* dan CNN. LSTM, meskipun diharuskan untuk menangani data urutan dengan baik, menunjukkan kinerja yang sangat rendah. Penelitian ini memberikan pemahaman penting tentang pilihan algoritma yang tepat untuk tugas klasifikasi sentimen dan menunjukkan bahwa kompleksitas algoritma tidak selalu sejalan dengan kinerja yang lebih baik.

Temuan ini juga menekankan pentingnya optimisasi dan pemahaman mendalam tentang data dan tugas yang dihadapi dalam analisis sentimen. Optimasi pada satu algoritma memang dapat meningkatkan performa spesifik dari algoritma tersebut, namun arah penelitian kali ini adalah untuk memahami perbandingan kinerja alami dari berbagai algoritma yaitu *Support Vector Machine (SVM), Random Forest, Naive Bayes, K-Nearest Neighbors (KNN),* dan *Logistic Regression* dalam konteks analisis sentimen.

Dengan membandingkan lima algoritma ini tanpa optimasi berlebihan, diharapkan dapat mengidentifikasi kekuatan dan kelemahan intrinsik dari masing-masing metode, serta memberikan rekomendasi yang lebih umum dan aplikatif untuk berbagai jenis data ulasan di masa depan. Optimasi bisa menjadi langkah lanjutan setelah pemahaman mendasar tentang performa setiap algoritma diperoleh. Studi-studi sebelumnya menunjukkan bahwa komparasi berbagai algoritma seringkali memberikan wawasan yang lebih komprehensif dibandingkan fokus pada optimasi satu algoritma saja. Hasil komparasi dalam penelitian ini diharapkan dapat menemukan algoritma terbaik yang sesuai untuk analisis sentimen pada dataset ulasan aplikasi digital Korlantas Polri dengan mengetahui nilai akurasi, presisi, *recall* dan *f1 measure* pada masing-masing algoritma yang dibandingkan.

METODE PENELITIAN

Metode kuantitatif digunakan dalam penelitian ini dengan tujuan untuk membandingkan performa berbagai algoritma *machine learning* dalam analisis sentimen ulasan pengguna aplikasi digital Korlantas Polri. Jenis penelitian komparatif dengan membandingkan kinerja lima algoritma *machine learning* yaitu *Support Vector Machine (SVM), Random Forest, Naive Bayes, K-Nearest Neighbors (KNN),* dan *Logistic Regression*.

Langkah pertama yang dilakukan melalui pengumpulan data. Metode pengumpulan data dalam penelitian ini dilakukan melalui proses *web scraping*. Sebanyak 10.000 ulasan dipilih secara acak untuk dianalisis. Dalam penelitian ini menggunakan dataset ulasan dari aplikasi *Google Play Store* untuk menganalisis sentimen pada ulasan aplikasi digital Korlantas Polri. Alasan utama penggunaan dataset dari satu sumber ini adalah karena ulasan di *Google Play Store* dilengkapi

dengan pelabelan berbasis peringkat, yang memudahkan proses klasifikasi sentimen. Penelitian sebelumnya menunjukkan bahwa ulasan dengan pelabelan peringkat memiliki kualitas data yang lebih terstruktur dan konsisten, yang sangat membantu dalam membangun model machine learning yang akurat (Sasikala P et al., 2020).

Langkah kedua adalah pemberian label positif, negatif atau netral pada data ulasan pengguna menggunakan pelabelan berbasis peringkat pada *rating* aplikasi (Alhaqq & Ruldeviyani, 2022).

Pada langkah ketiga yaitu *pre-processing* data melibatkan *data cleansing*, *case folding* yang akan menghindari duplikasi data, *tokenizing* yang merupakan proses penguraian hasil dari *case folding* menjadi token-token atau kata-kata dengan dibantu *library* nltk, Selanjutnya proses formalisasi yang mengubah kata menjadi baku sesuai dengan kamus besar bahasa Indonesia. *Stopword removal* yaitu kata-kata yang tidak penting nantinya akan diseleksi dan dihapus, dan proses yang terakhir adalah stemming untuk mengurangi kata-kata ke bentuk dasar, lebih tepatnya membantu mempermudah dalam pemrosesan teks dan meningkatkan kinerja algoritma. Pada intinya dalam skenario ini, data teks akan diubah menjadi representasi numerik yang bisa diproses oleh algoritma *machine learning*.

Setelah *pre-processing* data kemudian dilakukan Teknik SMOTE dimana dengan melakukan SMOTE akan menciptakan sampel-sampel tambahan dari kelas minoritas, sehingga membantu meningkatkan kinerja klasifikasi pada data yang memiliki ketidakseimbangan kelas (Blagus & Lusa, 2013)

Selanjutnya, pilih model pembelajaran mesin menggunakan *Super Vector Machine, Naïve Bayes, K-nearest neighbor, Random Forest,* dan *Logistic Regression*. Model *machine learning* belajar dari data latih dan mengoptimalkan parameter selama pelatihan. Data latih dan uji adalah dua bagian dari data yang digunakan untuk melatih model *machine learning*. Data uji digunakan untuk menguji hasil pembelajaran model. Untuk meningkatkan kualitas *machine learning*, penelitian ini membagi 20% data uji dan 80% data latih. Semakin banyak data yang digunakan untuk melatih model, semakin besar kemungkinan bahwa model dapat belajar dari variasi yang lebih besar dari data yang saat ini digunakan, dan itu akan berakibat, jika model diuji dengan data yang belum pernah dilihat sebelumnya, diharapkan dapat memberikan hasil yang lebih akurat.

Pengujian performa model dengan menggunakan data uji membantu memahami sejauh mana model dapat memprediksi dengan benar dan mengetahui skor akurasi, presisi, *recall*, dan f1 *measure* untuk model *machine learning*. Setelah pengujian model, algoritma terbaik untuk menganalisis sentimen ulasan pengguna juga akan diketahui.

HASIL DAN PEMBAHASAN

Dataset

Penggunaan dataset terdiri dari 10.000 ulasan. Penggunaanan ulasan *Google Play Store* juga memungkinkan untuk dapat menghindari masalah heterogenitas dan variasi dalam format serta kualitas data yang sering terjadi ketika menggunakan dataset dari berbagai sumber. *Scraping* data dilakukan dengan mengumpulkan data ulasan pengguna dengan bantuan dari *library google-play-scraper* dan IDE *Google Collab*.

Tabel 1 menunjukkan data yang diambil dari *web scrapping* pada aplikasi tersebut berupa nama pemberi ulasan, nilai bintang yang diberikan, waktu ulasan dikirim dan isi *review* ulasan. Ulasan-ulasan ini mencakup berbagai peringkat bintang yang diberikan oleh pengguna, yang kemudian dilabeli menjadi sentimen positif, negatif, atau netral berdasarkan peringkat tersebut yang selanjutnya divalidasi oleh ahli bahasa sebelum proses *pre-processing* data. Berikut tampilan dataset hasil *scrapping* dilihat pada Tabel 1.

Tabel 1. Dataset Hasil Scrapping

Username	Score	At	Content
Andi Wibowo	3	2023-12-12 02:52:49	Untuk server website eRikkes tolong diperbaiki...
Resma Putro	4	2023-12-12 01:58:21	Semua menunya berjalan baik dan sangat memba...
Haidar Annihrir	2	2023-12-12 04:38:26	Sudah menyelesaikan semua tes, tinggal pengaju...
Nizar D Rahmat	5	2023-12-27 13:04:53	Aplikasi bagus. Hanya saja, ada beberapa HP ya...
Lukman Nur	5	2023-12-17 14:33:09	Aplikasinya sangat membantu, memudahkan utk...

Pre-processing data mencakup beberapa prosedur, seperti membersihkan data untuk menghilangkan duplikat dan karakter yang tidak relevan, setelah itu teks dilipat menjadi huruf kecil untuk konsistensi. Tokenisasi memecah teks ulasan menjadi kata-kata unik dengan menggunakan *library* NLTK. Kata-kata menjadi bentuk baku yang dapat digunakan dalam Kamus Besar Bahasa Indonesia setelah diformalisasi. Kata-kata yang tidak berguna dihapus dengan *stopword removal*. *Stemming* mengubah bentuk aslinya dari kata-kata untuk mempermudah pemrosesan teks.

Untuk mengatasi ketidakseimbangan kelas dalam data, metode *Synthetic Minority Over-sampling Technique* (SMOTE) digunakan. SMOTE meningkatkan kinerja klasifikasi pada data yang tidak seimbang.

Pemodelan

Penelitian ini menggunakan *Super Vector Machine (SVM)*, *Naiva Bayes, K-Nearest Neighbors (KNN), Random Forest*, dan *Logistic Regression* sebagai teknik pemodelan.

1. *Super Vector Machine (SVM).*
   SVM adalah algoritma klasifikasi dua kelas yang biasanya digunakan untuk membuat SVM multi-kelas. SVM multi-kelas menggunakan klasifikasi satu lawan satu untuk memisahkan data uji dengan margin yang signifikan (Vankdothu & Hameed, 2022).
   SVM pada dasarnya digunakan untuk klasifikasi linear, tetapi juga dapat digunakan untuk masalah non-linear. Untuk mengkategorikan data yang dapat dipisahkan secara linear, biasanya menggunakan SVM linear. Misalkan ada data xi = {x1,x2,...,xn} ε R n dan yi ε {-1,+1}. Langkah pertama yang harus dilakukan yaitu menemukan *hyperplane* sebagai pemisah antara dua kelas dengan fungsi linear. Proses ini dapat digambarkan secara matematis sebagai berikut.:(Soebroto, 2019)

$$f(x) = w.x_i + b \quad \textbf{(1)}$$

Dimana w adalah bobot *vector support* atau *vector* tegak lurus dengan *hyperplane*, yang dapat digambarkan sebagai berikut:

$$\sum_{i=1}^{n} a_i y_i x_i \quad \textbf{(2)}$$

Di mana x_i adalah data ke-I, y_i adalah kelas data ke-I, a_i adalah nilai α data ke-I, dan b adalah nilai ambang.

$$b = -\frac{1}{2}(w.x^{+} + w.x^{-}) \quad \textbf{(3)}$$

Data pada kelas positif atau negatif dapat diklasifikasikan dengan fungsi keputusan klasifikasi tanda (f(x)). Data pendukung kelas positif x+ dan negatif x- memiliki nilai alfa tertinggi.

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x, x_i) + b \quad \textbf{(4)}$$

Hyperplane terbaik terletak di tengah antara dua set objek dari dua k, di mana m adalah jumlah vector pendukung, αi adalah nilai bobot tiap titik data, dan K adalah fungsi kernel. Berdasarkan nilai *sign*(f(x)) = 1 dan sign(f(x)) =-1, data diklasifikasikan sebagai kelas positif dan negatif.

2. *Naïve Bayes*
Penentuan kelas mana sekumpulan fitur (sifat) akan termasuk, algoritma klasifikasi *Naive Bayes* menggunakan probabilitas. *Naive Bayes* kami menggunakan buku NLTK. (Motz et al., 2022)
Persamaan teorema Bayes seperti berikut:

$$\boldsymbol{P(H|X) = \frac{P(H|X).PX}{P(X)}} \quad \textbf{(5)}$$

Untuk data dari kelas yang belum diketahui, X adalah keterangan, dan H adalah hipotesis data dari kelas tertentu. P (H|X) adalah kemungkinan hipotesis H berdasarkan kondisi X, kemungkinan hipotesis H prior, kemungkinan hipotesis H berdasarkan kondisi pada hipotesis H P, dan kemungkinan X adalah kemungkinan X.

3. *K-Nearest Neighbors (KNN)*
Algoritma *K-Nearest Neighbor* adalah suatu metode algoritma klasifikasi yang bekerja berdasarkan tingkat kemiripan yang dihitung berdasarkan jarak (*distance*) terdekat dari data pembelajarannya (data latih dan data uji) (Boiculese et al., 2013). Metode yang paling banyak digunakan peneliti untuk menghitung jarak pada algoritma K-NN adalah metode Euclidean distance (Liu & Zhang, 2012). Ketepatan metode KNN berubah seiring dengan jumlah tetangga dan tingkat informasi yang digunakan untuk klasifikasi. Sementara itu, mereka menunjukkan rincian tentang variasi nilai maksimum dan minimum akurasi dengan ukuran set klasifikasi dan jumlah tetangga (Maleki et al., 2021).
Metode *Euclidean distance* secara umum digunakan untuk data numerik. Untuk rumus perhitungan *Euclidean distance*:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad \textbf{(6)}$$

Dimana d(x, y) jarak kedekatan antara data x ke data y, xi data testing (data uji) ke i , yi data *training* (data latih) ke I, n jumlah atribut 1 sampai n.

4. *Random Forest*
Random Forest adalah algoritma *machine learning* yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil. Leo Breiman dan Adele Cutler menciptakan algoritma *random forest*. Algoritma ini berasal dari ide pembelajaran kelompok, yaitu proses menggabungkan berbagai pengklasifikasi untuk meningkatkan kinerja model dan memecahkan masalah yang kompleks. (Trivusi, 2021).

Random Forest adalah klasifikator yang terdiri dari kumpulan klasifikasi berstruktur pohon {h(x, k), k = 1,...} di mana {k} adalah vektor acak independen yang didistribusikan secara identik dan setiap pohon mengeluarkan suara unit untuk kelas yang paling populer pada input x.

5. *Logistic Regression*
Logistic Regression adalah algoritma statistik yang digunakan untuk memprediksi probabilitas suatu variabel dependen kategorikal berdasarkan satu atau lebih variabel independen. *Logistic Regression* memperkirakan probabilitas suatu hasil hanya dengan menggunakan dua nilai (Cam et al., 2024)
Logistic Regression dimulai dengan persamaan berikut:

$$\text{Probability of outcome}(\hat{Y}_i) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_iX_i}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_iX_i}} \quad \textbf{(7)}$$

Dalam regresi logistik, Y ^_i adalah probabilitas yang diperkirakan untuk berada dalam satu hasil biner kategori (i) terhadap yang lain, dan bukan hasil berkelanjutan.

Dalam regresi logistik, $e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_iX_i}$ adalah persamaan regresi linear untuk variabel independen yang dinyatakan dalam skala logit, bukan dalam format linear asli.

Lima algoritma *machine learning* yaitu *Super Vector Machine (SVM), Naiva Bayes, K-Nearest Neighbors (KNN), Random Forest,* dan *Logistic Regression* dilatih menggunakan data latih yang telah di-*oversampling* dengan SMOTE. Parameter *default* digunakan tanpa optimasi tambahan untuk mengevaluasi performa alami dari setiap algoritma. Model dievaluasi menggunakan data uji dengan metrik utama sebagai berikut:

Tabel 2. Definisi Metrik Evaluasi Model Machine Learning

Metrik Evaluasi	Keterangan
Akurasi	Persentase prediksi yang benar dari keseluruhan prediksi
Presisi	Proporsi prediksi positif yang benar dari total prediksi positif
Recall	Proporsi kejadian positif yang benar-benar teridentifikasi dari total kejadian positif
F1 Score	Harmonic mean dari presisi dan recall

Tabel 2 mengandung definisi metrik evaluasi utama yang digunakan untuk mengevaluasi seberapa baik model *machine learning* berfungsi dalam analisis sentimen. *Recall* menunjukkan persentase prediksi yang benar dari semua prediksi yang dibuat oleh model, yang menunjukkan seberapa baik model mengklasifikasikan data. Presisi menunjukkan persentase dari semua prediksi positif yang benar-benar positif, yang menunjukkan ketepatan model dalam mengidentifikasi kelas positif. Akurasi menunjukkan persentase dari semua prediksi positif yang benar-benar positif. Terakhir, skor F1 adalah rata-rata harmonis dari presisi dan *recall*, ini mengimbangi kedua metrik dan membantu ketika distribusi kelas tidak seimbang. Metrik ini memberikan gambaran lengkap tentang seberapa efektif model menganalisis sentimen ulasan jika digabungkan.

Evaluasi

Setelah melalui tahapan pengumpulan data, *preprocessing*, pembobotan kata, penggunaan teknik SMOTE, kemudian pembagian data, dan pelatihan model, Tabel 3 dan Gambar 1 menunjukkan hasil dari evaluasi performa lima algoritma *machine learning* yaitu *Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), Random Forest,* dan *Logistic Regression.*

Random forest menunjukkan akurasi tertinggi 90,76%, diikuti oleh *Support Vector Machine* 90,30% dan *Logistic Regression* 83,58%. Sedangkan pada nilai presisi *Support Vector Machine* memberikan nilai tertinggi 91,13%, baru kemudian *Random Forest* 90,92% dan *Logistic Regression* 83,87%. Random forest kembali unggul dengan nilai *recall* tertinggi 90,76%, *Support Vector Machine*

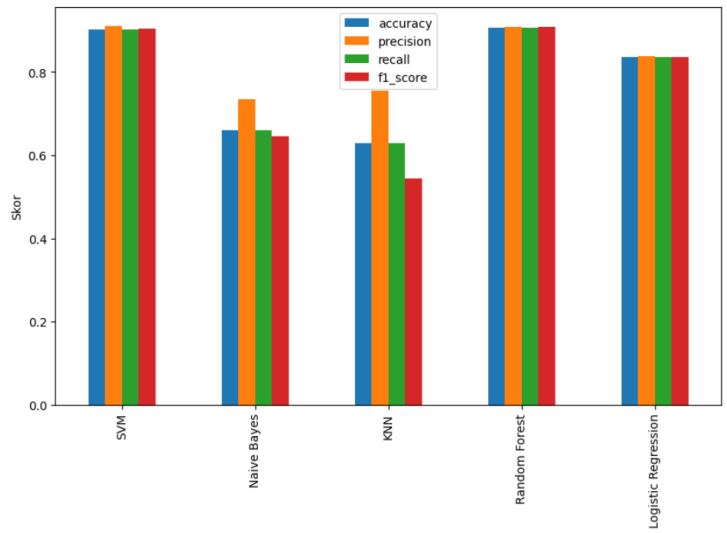
90,30% kemudian *logistic regression* 83,62%. *Random Forest* memiliki F1 *Score* tertinggi 90,78%, disusul *Supprot Vector Machine* 90,40% dan *Logistic Regression* 83,62%.

Hasil menunjukkan bahwa *Random Forest* adalah yang terbaik untuk menilai sentimen ulasan pengguna aplikasi Digital Korlantas Polri. Selanjutnya adalah *Logistic Regression* dan *Support Vector Machine*. Tiga algoritma lainnya mengalahkan algoritma Naive Bayes dan K-Nearest Neighbors.

Penelitian ini mendukung temuan sebelumnya (Yu, 2024) yang menunjukkan bahwa *Logistic Regression* dan *Random Forest* adalah algoritma yang efektif untuk analisis sentimen. Meskipun dalam penelitian ini Random Forest menunjukkan performa terbaik, disusul oleh *Support Vector Machine* kemudian *Logistic Regression*.

Tabel 3 Performa Algoritma Machine Learning

Algoritma	Akurasi (%)	Presisi (%)	Recall (%)	F1 Score (%)
Support Vector Machine (SVM)	90,30	91,13	90,30	90,40
Naive Bayes	66,02	73,54	66,02	64,66
K-Nearest Neighbors (KNN)	62,89	75,58	62,89	54,38
Random Forest	90,76	90,92	90,76	90,78
Logistic Regresssion	83,58	83,87	83,58	83,62



Gambar 1. Hasil Evaluasi Model Machine Learning

KESIMPULAN DAN SARAN

Kesimpulan

Hasil perbandingannya bisa dilihat bahwa algoritma *Random Forest* adalah algoritma terbaik dalam analisis sentimen ulasan aplikasi Digital Korlantas Polri, karena kinerjanya yang kompetitif. *Support Vector Machine* dan *Logistic Regression* dapat dipertimbangkan untuk digunakan dalam aplikasi serupa di masa depan.

Saran

Rekomendasi untuk penelitian lanjutan adalah melakukan optimasi parameter untuk algoritma terbaik guna meningkatkan kinerja lebih lanjut. Menggunakan dataset yang lebih besar dan lebih bervariasi untuk menguji generalisasi model. Penelitian selanjutnya dapat lebih mengeksplorasi penggunaan algoritma deep learning seperti LSTM dengan optimasi yang tepat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Prof. Dr. Kusrini, M.Kom., yang telah memberikan bimbingan, dukungan, dan pengetahuan yang sangat berharga selama proses pembuatan jurnal ini. Ucapan terima kasih juga kepada keluarga yang terus mendoakan dan kepada semua orang yang telah membantu dan mendukung penulis dalam menyusun dan menyelesaikan jurnal ini.

REFERENSI

Alhaqq, R. I., & Ruldeviyani, Y. (2022). Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store. https://www.researchgate.net/publication/367216412

Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. Journal of Information Systems and Informatics, 5(3), 915–927. https://doi.org/10.51519/journalisi.v5i3.539

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data (Vol. 14). http://www.biomedcentral.com/1471-2105/14/106

Boiculese, V. L., Dimitriu, G., & Moscalu, M. (2013). Improving recall of k-nearest neighbor algorithm for classes of uneven size. 2013 E-Health and Bioengineering Conference, EHB 2013. https://doi.org/10.1109/EHB.2013.6707403

Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. Heliyon, 10(1). https://doi.org/10.1016/j.heliyon.2023.e23784

Jardim, S., & Mora, C. (2021). Customer reviews sentiment-based analysis and clustering for market-oriented tourism services and products development or positioning. Procedia Computer Science, 196, 199–206. https://doi.org/10.1016/j.procs.2021.12.006

Kaburuan, E. R., & Setiawan, N. R. (2023). Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 12(1), 105–116. https://doi.org/10.32736/sisfokom.v12i1.1614

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining Text Data (Vol. 9781461432234, pp. 415–463). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_13

Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164. https://doi.org/10.1016/j.eswa.2020.113981

Motz, A., Ranta, E., Calderon, A. S., Adam, Q., Alzhouri, F., & Ebrahimi, D. (2022). Live Sentiment Analysis Using Multiple Machine Learning and Text Processing Algorithms. Procedia Computer Science, 203, 165–172. https://doi.org/10.1016/j.procs.2022.07.023

Sasikala P, Mary, L., & Sheela, I. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS.

Soebroto, A. A. (2019). Buku Ajar AI, Machine Learning & Deep Learning. https://www.researchgate.net/publication/348003841

Trivusi. (2021, March 20). Penjelasan Lengkap Algoritma Support Vector Machine (SVM). Trivusi.

Vankdothu, R., & Hameed, M. A. (2022). Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning. Measurement: Sensors, 24. https://doi.org/10.1016/j.measen.2022.100440

Yu, B. (2024). Comparative Analysis of Machine Learning Algorithms for Sentiment Classification in Amazon Reviews. In Business, Economics and Management EMFRM (Vol. 2023).