

## Perbandingan Performa Algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* untuk Analisis Sentimen Pengguna Aplikasi Retail di Android

Akhmad Ghiffary Budianto<sup>1)</sup>, Rusilawati<sup>2)</sup>, Arief Trisno Eko Suryo<sup>3)</sup>, Gunawan Rudi Cahyono<sup>4)</sup>, Andry Fajar Zulkarnain<sup>5)</sup>, Martunus<sup>6)</sup>

<sup>1)2)3)4)5)6)</sup> Program Studi Rekayasa Elektro, Fakultas Teknik, Universitas Lambung Mangkurat  
Jl. Achmad Yani Km. 33,5 Banjarbaru, Kalimantan Selatan 70714, Indonesia

<sup>1)</sup> ghiffaryb04@gmail.com

<sup>2)</sup> habsyi.sila@gmail.com

<sup>3)</sup> arief.suryo@ulm.ac.id

<sup>4)</sup> gunawan.cahyono@ulm.ac.id

<sup>5)</sup> andry.zulkarnain@ulm.ac.id

<sup>6)</sup> martunusunus@gmail.com

### Abstrak

Platform Google Play Store menyediakan fitur pemberian masukan dan penilaian dari pelanggan terhadap aplikasi yang tersedia. Aplikasi retail di android dapat menjadikan hal tersebut sebagai salah satu metrik utama untuk memantau kinerja. Analisis sentimen merupakan salah satu bagian dari NLP dan *text mining* yang digunakan untuk melihat sentimen dari opini dalam ulasan pengguna di aplikasi tersebut. Penelitian ini bertujuan untuk membandingkan performa algoritma SVM dan *logistic regression* dalam klasifikasi sentimen ulasan pengguna aplikasi tersebut. Dari 4250 data ulasan pengguna, 3400 data digunakan sebagai *data training* dan 850 data digunakan sebagai *data testing*. Hasilnya didapatkan klasifikasi dengan SVM dan *logistic regression* sama-sama memiliki tingkat akurasi 0,87, namun namun *Logistic Regression* menunjukkan performa yang lebih baik dalam mengidentifikasi sentimen positif dengan nilai F1 score 0.89 dibandingkan SVM (0.78). *Logistic Regression* juga memiliki misklasifikasi yang lebih sedikit (110) dan nilai *True Positive* (TP) dan *True Negative* (TN) yang lebih tinggi dibandingkan SVM. Secara keseluruhan, *Logistic Regression* menunjukkan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi retail dibandingkan SVM.

**Kata kunci:** *Natural Language Processing, Support Vector Machine, Logistic Regression, Sentimen, Klasifikasi Teks*

### Abstract

The Google Play Store platform provides a feature for customers to provide feedback and ratings for available applications. Retail applications on Android can use this as one of the main metrics to monitor performance. Sentiment analysis is a part of NLP and text mining that is used to see the sentiment of opinions in user reviews of the application. This study aims to compare the performance of SVM and logistic regression algorithms in classifying the sentiment of user reviews of the application. From 4250 user review data, 3400 data were used as training data and 850 data were used as testing data. The results obtained are that the classification with SVM and logistic regression both have an accuracy level of 0.87, but Logistic Regression shows better performance in identifying positive sentiment with an F1 score of 0.89 compared to SVM (0.78). Logistic Regression also has fewer misclassifications (110) and higher True Positive (TP) and True Negative (TN) values compared to SVM. Overall, Logistic Regression shows better performance in classifying the sentiment of user reviews of retail applications compared to SVM.

**Keywords:** *Natural Language Processing, Support Vector Machine, Logistic Regression, Sentiment, Text Classification*

## 1. PENDAHULUAN

Penggunaan smartphone bagi masyarakat sudah menjadi hal yang umum di era internet dan teknologi. Smartphone dapat melakukan berbagai aktivitas dengan menggunakan aplikasi yang tersedia pada *mobile app stores* sesuai *operating system* yang digunakan. Salah satu *mobile app stores* yaitu *Google Play Store*. Platform *Google Play Store* memungkinkan pengguna untuk mengekspresikan pendapat mereka tentang pengalaman menggunakan suatu aplikasi melalui ulasan dan rating bintang yang dapat dilihat oleh calon pengguna aplikasi lainnya [1]. Masukan dan penilaian pelanggan selalu menjadi salah satu metrik utama yang dapat digunakan untuk meninjau kinerja dan memberikan rekomendasi yang sesuai untuk meningkatkan fungsionalitas yang disediakan oleh aplikasi [2].

*Text mining* menjadi salah satu bagian dari *Big Data Analytics*. Klasifikasi teks menggunakan pengklasifikasian *machine learning* dapat dilakukan dengan dua cara yaitu dengan *supervised* dan *unsupervised learning*. *Supervised learning* dilakukan dengan data yang memiliki label sedangkan *unsupervised learning* dilakukan dengan data tanpa label [3]. Setelah pemberian label pada dataset, pelatihan data dilakukan untuk mengambil keluaran yang layak untuk mendapatkan informasi yang dibutuhkan [4]. *Text mining* dengan *Bag-of-Words* juga digunakan dalam menganalisis perihal politik dan pemilu kepala daerah di media sosial [5].

Analisis sentimen merupakan penggalian opini yang digunakan untuk mengenali emosi konten di suatu web. Hal ini biasanya digunakan untuk mendapatkan suara pengguna dari produk, layanan, film, berita, isu tertentu dan sebagainya. Tujuan analisis sentimen yaitu untuk menentukan sentiment seseorang mengenai subjek/objek tertentu [6]. Penelitian terdahulu mengenai analisis sentimen umumnya digunakan untuk mengetahui ulasan dan komentar pengguna micro blogging sites (Facebook, twitter, dan sebagainya) [7], opini tentang perkembangan teknologi dibidang kecerdasan buatan [8], ulasan pengguna di *platform Google Play Stores* dan *Apple Stores* [9]. Untuk algoritma yang digunakan pada analisis sentiment, sering terdapat perbandingan antara SVM dan *naïve bayes* [6], SVM dan *sentiwordnet* [9], *naïve bayes* dan *logistic regression* [3]. Namun belum ada yang membandingkan performa antara algoritma SVM dan *logistic regression* pada analisis sentiment yang menggunakan model Tervektorisasi.

Penelitian ini bertujuan untuk membandingkan performa dari algoritma SVM dan *logistic regression* dari dataset ulasan pengguna aplikasi retail di *Google Play Stores* pada kasus analisis sentimen. Dataset ulasan pengguna tersebut dibuat dalam model tervektorisasi. Performa *machine learning* untuk klasifikasi sentimen dapat dilihat pada akurasi, *precision*, *recall*, *f-1 score* dan *confussion matrix*.

## 2. TINJAUAN PUSTAKA

### 2.1 *Natural Language Processing (NLP) dan Text Classification*

*Natural Language Processing (NLP)* bertujuan untuk memahami teks misalnya dalam menentukan struktur kalimat ataupun sentimen kalimatnya. Sedangkan tujuan dari text mining yaitu untuk memperoleh informasi yang terstruktur dari teks. Hubungan NLP dan text mining yaitu text mining menggunakan konsep dari NLP. Konsep NLP dalam text mining yang umum digunakan antara lain [10]:

- 1) Token yaitu unit dasar teks yang memiliki makna, seperti kata, angka, atau simbol. Memisahkan teks menjadi token, atau proses tokenisasi, merupakan langkah penting dalam penambangan teks, karena banyak algoritma tidak dapat memproses elemen yang lebih kecil dari token.
- 2) *Part-of-Speech (POS)* yaitu kategori leksikal dari setiap token sebagai contoh kata benda sebagai contoh pada kata benda, kata sifat atau tanda baca. Penandaan POS yaitu proses pengklasifikasian token secara otomatis, memberikan informasi semantik tambahan pada teks
- 3) Lemma dan Stem yaitu bentuk dasar dari sebuah kata. Lemma mewakili bentuk kanonik dari kata, sedangkan Stem merupakan fragmen kata yang tidak berubah. Contoh lemma dari kata "menginduksi" adalah "induksi", sedangkan stemnya adalah "induc-".

- 4) Pemisahan Kalimat: Proses identifikasi batas kalimat dalam teks. Metode pemisahan kalimat harus mempertimbangkan perbedaan antara titik di akhir kalimat dan titik di akhir akronim atau singkatan. Memisahkan teks menjadi kalimat penting untuk mengekstrak ide-ide unik, meskipun konteks keseluruhan dokumen juga perlu dipertimbangkan.
- 5) Entitas: Segmen teks yang relevan dengan domain tertentu, dan dapat terdiri dari satu atau lebih token.

*Text mining* adalah bagian dari *data mining*. Proses penggalian informasi dan pengetahuan yang menarik dalam teks yang tidak terstruktur disebut dengan *Intelligent Text Analysis*, *Text Data Mining* atau *Knowledge-Discovery in Texts* (KDT) [11]. *Text classification* merupakan tugas pada NLP dalam mengklasifikasikan atau mengkategorikan teks dan menghitung bobot setiap kata, yaitu banyaknya kemunculan setiap kata [12]. Vektorisasi teks atau penyematan kata melibatkan representasi atau pemetaan kata atau dokumen suatu korpus ke vektor numerik bilangan atau bilangan real. Karena sebagian besar algoritma pada machine learning menangani masukan numerik, ini merupakan langkah penting dalam NLP dan analisis sentimen berbasis machine learning [13].

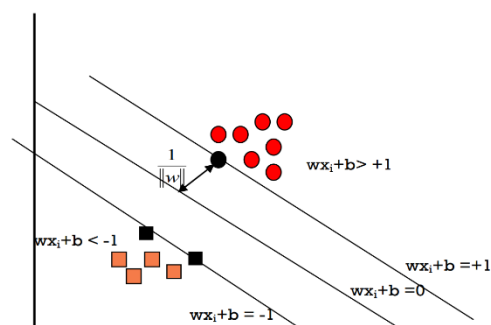
## 2.2 Analisis Sentimen

Analisis sentiment merupakan salah satu bagian topik bahasan dalam NLP. Analisis sentiment berfokus pada analisis opini, sentimen, perilaku atau pandangan masyarakat yang tertuang dalam bentuk teks. Oleh karena itu, analisis sentiment sering disebut sebagai penggalian opini [8]. Adapun tujuan dari sentimen analisis antara lain: memahami opini publik, analisis kondisi sekarang untuk meningkatkan reputasi, mengembangkan produk yang lebih baik lagi dan menentukan strategi pemasaran yang cocok sesuai dengan sentimen opini publik. Analisis sentimen terbagi menjadi dua kelompok berdasarkan sumber datanya [14] yaitu :

- 1) Analisis sentimen kasar (*Coarse-grained sentiment analysis*) merupakan proses pengkategorian keseluruhan isi dokumen ke dalam bentuk positif atau negatif. Pendekatan ini cocok untuk mendapatkan gambaran umum tentang sentiment yang mendominasi dalam suatu teks yang Panjang.
- 2) Analisis sentimen halus (*Fine-grained sentiment analysis*) yaitu sentiment pada Tingkat kalimat yang membagi klasifikasi sentiment ke dalam hal yang lebih spesifik seperti sangat positif, positif, netral, negatif dan sangat negatif. Hal ini membuat analisis yang lebih detail dan akurat dapat dilakukan terhadap opini yang tertuang dalam setiap kalimat.

## 2.3 Support Vector Machine (SVM)

SVM merupakan metode klasifikasi data linier atau nonlinier dengan menggunakan data nonlinier yang dipetakan untuk mengubah training data ke dimensi yang lebih tinggi. Metode ini mencari *hyperplane* dengan memaksimalkan margin atau jarak antar kelas [9]. Bentuk *hyperplane* dan margin maksimal yang terbentuk dapat dilihat pada Gambar 1. Mengingat kelas dalam klasifikasi ini menggunakan konsep *One Against One* (OAO) sehingga metode ini cocok untuk mengklasifikasikan kasus kategori biner (1 atau 0).



Gambar 1. Grafik ilustrasi *hyperplane* dan margin yang terbentuk pada SVM [15]

## 2.4 Logistic Regression

*Logistic Regression* merupakan strategi penilaian untuk memecah kumpulan data dimana setidaknya ada satu faktor otonom yang menentukan hasilnya. Hasilnya diukur dengan variabel dikotomis (yang hanya terdapat dua kemungkinan hasil). Dalam *Logistic Regression*, variabel dependen bersifat berpasangan atau dikotomis yaitu hanya membuat informasi biner untuk kode 1 (benar, positif, dan sebagainya) atau kode 0 (salah, negatif dan sebagainya) [16]. Regresi logistik banyak digunakan dalam aplikasi *machine learning*. Model ini menangkap vektor variabel dan mengevaluasi koefisien atau bobot untuk setiap variabel masukan dan kemudian memprediksi kelas pada teks yang dinyatakan sebagai vektor kata [3]. Secara matematis fungsi regresi logistik memperkirakan fungsi linier berganda yang didefinisikan sebagai berikut:

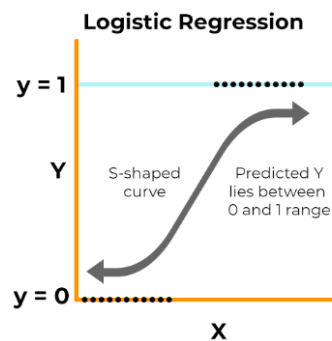
$$\text{logit}(S) = b_0 + b_1M_1 + b_2M_2 + b_3M_3 + \dots + b_kM_k \quad (1)$$

Keterangan:

$S$  = probabilitas kehadiran fitur yang diamati

$M_1, M_2, \dots, M_k$  = nilai prediktor

$B_0, B_1, \dots, B_k$  = intersep pada model



Gambar 2. Grafik ilustrasi klasifikasi dengan *Logistic Regression*

## 2.5 Pengukuran Performa pada Klasifikasi

*Confusion matrix* merupakan salah satu metode dalam melakukan evaluasi performa pada *machine learning* di kasus klasifikasi. Gambar 3 menunjukkan ilustrasi dari pengukuran performa pada klasifikasi biner. *Precision* merupakan jumlah ulasan positif yang sebenarnya dari jumlah total ulasan yang diberikan secara positif. *Recall* adalah jumlah ulasan positif yang benar dari ulasan positif yang sebenarnya. *F-measure* digunakan untuk menghitung metode presisi dan *recall* yang terbobot. *Accuracy* merupakan ukuran kinerja yang paling penting yaitu ukuran rasio observasi yang diprediksi terhadap jumlah total yang diobservasi. Akurasi yang memiliki nilai tinggi menunjukkan model *machine learning* yang baik [17].

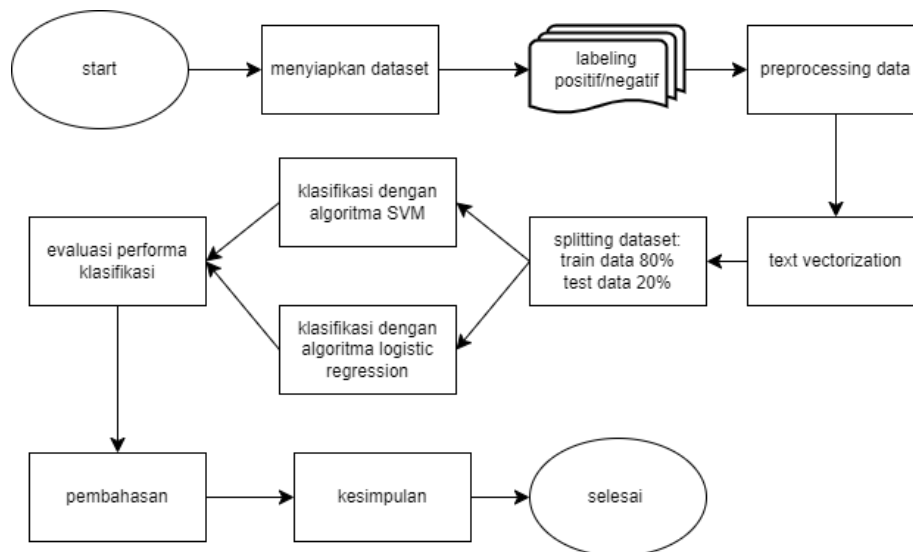
		Actual Value (Obtained by experiment)		
		Positives	Negatives	
Predicted Value (Predicted by the test)	Positives	<b>True Positive</b> (Correct)	<b>False Positive</b> (Incorrect)	← Precision (Positive Predicted Value)
	Negatives	<b>False Negative</b> (Incorrect)	<b>True Negative</b> (Correct)	
		↑ Recall (Sensitivity)	↑ Specificity	← NPV (Negative Predicted Value)
				ACC, F-measure, MMC

Gambar 3. *Confusion matrix* pada klasifikasi [18]

### 3. METODE PENELITIAN

Metode dalam penelitian ini dapat dilihat pada Gambar 4. Proses komputasi menggunakan bahasa python dengan menggunakan *google colab*. Adapun *library* python yang digunakan antara lain: *pandas*, *numpy*, *scikit-learn*, dan *matplotlib*.

Dataset yang disiapkan berisikan data *username*, *score rating*, *at* (tanggal dan waktu pemberian ulasan), konten (berisikan ulasan pengguna dalam bahasa indonesia), sentimen (label positif / negatif) dan *content* (ulasan pengguna dalam bahasa inggris). Dataset yang digunakan sebanyak 4250 baris dan 6 kolom seperti yang terlihat pada gambar 5. Selanjutnya melakukan *preprocessing data* berupa penghapusan simbol, *emoticon* dan angka pada teks ulasan. Selain itu melakukan perubahan singkatan dalam bahasa inggris seperti “*i’ve*” menjadi “*i have*”.



Gambar 4. Diagram alir penelitian

Setelah melakukan *preprocessing data*, maka dilanjutkan dengan melakukan *text vectorization*. Dalam *text vectorization*, tokenisasi adalah langkah pertama untuk mendapatkan sentimen penting yang memuat kata-kata dari sebuah teks ulasan. Dengan menggunakan token ini, penandaan *Part-of-Speech* (*POS Tagging*) dilakukan untuk mendapatkan kata-kata yang mengandung sentimen tertentu. Kata-kata yang mengandung sentimen tertentu bisa jadi termasuk dalam kata sifat yang mendeskripsikan kata benda, kata kerja dan kata keterangan yang ada pada teks ulasan pengguna [19].

Data yang telah selesai dilakukan tokenisasi dan *POS tagging*, selanjutnya dilakukan pemisahan antara data training dan testing dengan rasio 80%:20%. Sehingga data training berjumlah 3400 data dan data testing berjumlah 850 data.

Machine learning untuk klasifikasi dilakukan dengan 2 algoritma sebagai bahan perbandingan performa klasifikasi yaitu algoritma SVM dan *logistic regression*. Pemilihan kedua algoritma ini sebagai bahan perbandingan performa klasifikasi karena kedua algoritma ini cocok dalam kasus klasifikasi biner (1/0 atau positif/negatif). Algoritma SVM disini menggunakan *kernel linear* dalam klasifikasi. Setelah kedua algoritma tersebut dijalankan dan mendapat *training data* dan *testing data* terbaik. Maka selanjutnya dilakukan proses evaluasi dengan membandingkan *accuracy*, *precision*, *recall*, *f-1 score* dan *confusion matrix* yang terbentuk. Proses analisis dan penarikan kesimpulan dilakukan terhadap perbandingan performa kedua algoritma tersebut dalam melakukan klasifikasi sentimen yang terbaik.

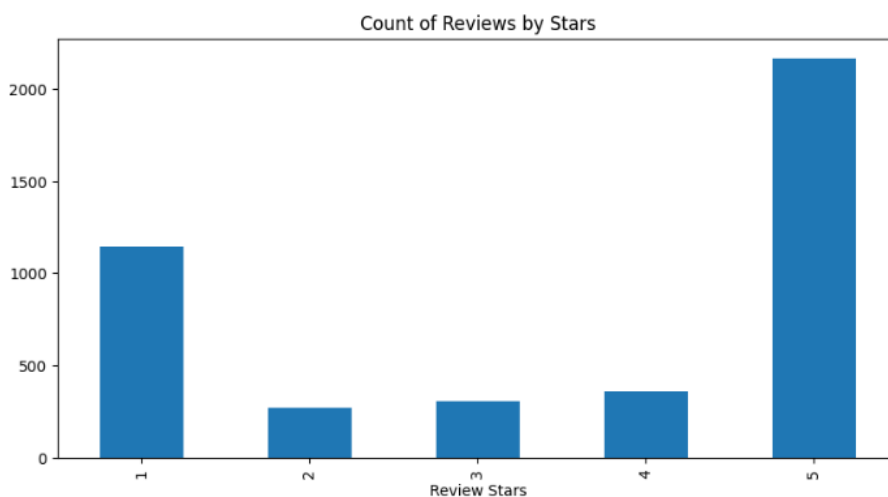
	userName	score	at	konten	sentimen	content
0	indrawan avianto	3	4/30/2024 2:30	Fitur kuitansi elektroniknya sulit dicari	negative	The electronic receipt feature is difficult to...
1	Gusviandy (Gusvi)	1	4/29/2024 14:14	Malas sudah sebenarnya pakai apps dan layanan ...	negative	I'm actually lazy about using these apps and s...
2	YamamotoChristine	3	4/29/2024 10:34	Aplikasi ini sangat membantu sebenarnya, tapi ...	negative	This application is actually very helpful, but...
3	yasmin julianti	1	4/28/2024 14:07	Pesanan dikirim dalam 4 jam, padahal pilih wak...	negative	The order was sent within 4 hours, even though...
4	Fajar Andilla	4	4/28/2024 12:25	Tolong sediakan tombol cancel sebelum pesanan ...	negative	Please provide a cancel button before the orde...
...	...	...	...	...	...	...
4245	Muhammad Imron Praza Mulyadana	1	3/5/2020 10:19	Di dekat lokasi rumah saya banyak indomaret, b...	negative	Near the location of my house there are many I...
4246	Leo Nababan	5	3/5/2020 9:36		Mantap	Excellent
4247	ivonni kusumawati	4	3/5/2020 7:22		good	Good
4248	Komang Safitri	5	3/5/2020 7:22	setelah di update, kenapa jadi ga bisa bayar v...	positive	After being updated, why can't you pay via vir...
4249	Lunnley Nescense	5	3/5/2020 2:14	Selama menggunakan klikindomaret nggak pernah ...	positive	During using Klikindomaret there was never a s...

4250 rows x 6 columns

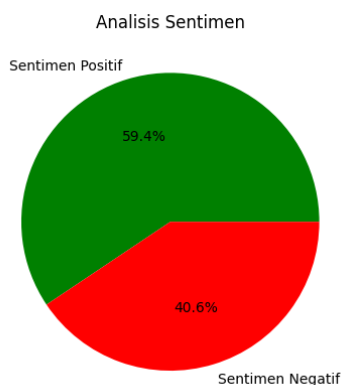
Gambar 5. Dataset ulasan pengguna sebanyak 4250 baris data

#### 4. PEMBAHASAN

Pada tahap awal, statistik deskriptif dari dataset ulasan pengguna aplikasi retail pada android digunakan untuk memberikan gambaran dari variasi penilaian pengguna secara rating bintang dengan sentimen dari ulasan pengguna. Gambar 6 menunjukkan plotting dari data statistik deskriptif dataset yang digunakan. Terlihat lebih dari 2000 pengguna memberikan rating bintang 5 (penilaian sangat baik) terhadap penggunaan aplikasi. Hal ini sesuai dengan nada sentimen positif yang berjumlah 59.4% dari total 4250 data ulasan. Sedangkan sisanya 40.6% memiliki nada sentimen negatif pada penggunaan aplikasi tersebut.



Gambar 6. Bar chart dari rating ulasan pengguna



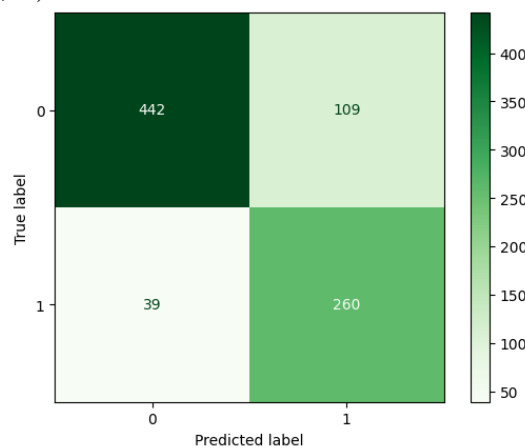
Gambar 7. Pie chart dari sentimen ulasan pengguna

Setelah dilakukan proses *data training* dan *data testing* pada algoritma SVM dan *logistic regression*, hasil pada data testing dapat dilihat pada tabel 1 berikut:

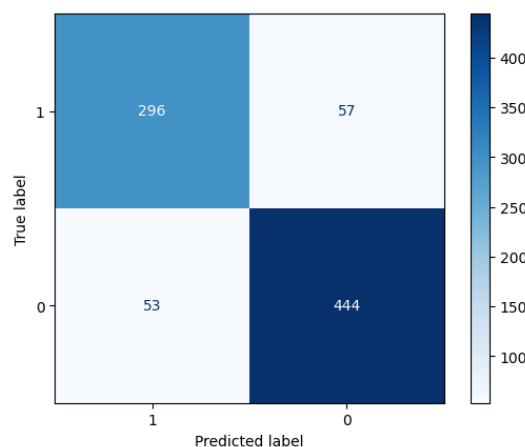
Tabel 1. Perbandingan performa algoritma SVM dan *Logistic regression* pada *data testing*

No	Algoritma Klasifikasi	Label	Precision	Recall	f-1 Score
1	SVM (accuracy = 0.87)	Positif	0.70	0.87	0.78
		Negatif	0.91	0.80	0.86
2	Logistic Regression (accuracy = 0.87)	Positif	0.89	0.89	0.89
		Negatif	0.85	0.84	0.84

Berdasarkan hasil klasifikasi pada Tabel 1, kedua algoritma klasifikasi (SVM dan *logistic regression*) sama-sama memiliki performa yang baik karena memiliki nilai akurasi yang sama yaitu 0,87. Namun *Logistic Regression* memiliki nilai *precision* dan *recall* yang lebih tinggi untuk label sentimen positif, sehingga hal ini berarti algoritma *logistic regression* menunjukkan performa yang lebih dalam mengidentifikasi data teks ulasan pengguna yang sebenarnya memiliki label sentimen positif dibandingkan algoritma SVM. Hal ini dapat dilihat dari nilai *f-1 score logistic regression* (0,89) lebih baik daripada SVM (0,78). Sedangkan dalam identifikasi label negatif sebenarnya kedua algoritma tersebut tidak jauh berbeda untuk nilai *f-1 score SVM* (0,86) dan *logistic regression* (0,84).



Gambar 8. *Confussion matrix* dari SVM



Gambar 9. *Confussion matrix* dari *logistic regression*

Gambar 8 dan 9 menunjukkan *confussion matrix* untuk performa dalam klasifikasi pada kedua algoritma. Algoritma SVM memiliki nilai *True Positive (TP)* sebanyak 260 dan *True Negative*

(TN) sebanyak 442 serta total misklasifikasi sebanyak 148. Sedangkan algoritma *logistic regression* memiliki nilai *True Positive* sebanyak 296 dan *True Negative* sebanyak 444 serta total misklasifikasi sebanyak 110. Dari jumlah *data testing* yang digunakan sebanyak 850 data. Hal ini menunjukkan algoritma *logistic regression* lebih baik dalam melakukan klasifikasi sentimen dibandingkan dengan algoritma SVM. Hal ini terlihat dari jumlah misklasifikasi yang *logistic regression* yang lebih sedikit daripada algoritma SVM. Selain itu *logistic regression* juga memiliki nilai TP dan TN yang lebih baik.

## 5. KESIMPULAN

Analisis sentimen dilakukan terhadap ulasan pengguna aplikasi retail di *platform Android*. Statistik deskriptif menunjukkan bahwa mayoritas pengguna memberikan rating bintang 5 (penilaian sangat baik) dan sentimen positif pada ulasan mereka. Algoritma klasifikasi SVM dan *Logistic Regression* digunakan untuk mengklasifikasikan sentimen ulasan. Kedua algoritma memiliki akurasi yang sama (0.87), namun *Logistic Regression* menunjukkan performa yang lebih baik dalam mengidentifikasi sentimen positif dengan nilai F1 score 0.89 dibandingkan SVM (0.78). *Confusion matrix* menunjukkan bahwa *Logistic Regression* memiliki misklasifikasi yang lebih sedikit (110) dibandingkan SVM (148). Hal ini menunjukkan bahwa *Logistic Regression* lebih baik dalam mengklasifikasikan sentimen ulasan pengguna. Selain itu, *Logistic Regression* memiliki nilai *True Positive* (TP) dan *True Negative* (TN) yang lebih tinggi dibandingkan SVM.

Walaupun hasil akurasi pada klasifikasi sentimen dengan algoritma SVM dan *logistic regression* menghasilkan nilai yang sama (0,87). Namun, *Logistic Regression* secara keseluruhan menunjukkan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi retail dibandingkan SVM. Hal ini dapat dilihat dari *F-1 score* yang lebih tinggi, jumlah misklasifikasi yang lebih sedikit, dan TP/TN yang lebih tinggi.

## DAFTAR PUSTAKA

- [1] S. Hassan, C. Tantithamthavorn, C.-P. Bezemer, and A. E. Hassan, "Studying the dialogue between users and developers of free apps in the google play store," *Empir. Softw. Eng.*, vol. 23, pp. 1275–1312, 2018.
- [2] S. Venkatakrisnan, A. Kaushik, and J. K. Verma, "Sentiment analysis on google play store data using deep learning," *Appl. Mach. Learn.*, pp. 15–30, 2020.
- [3] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve Bayes and logistic regression," in *2017 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, 2017, pp. 1–5.
- [4] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [5] W. T. H. Putri and R. Hendrowati, "Penggalian Teks Dengan Model Bag of Words Terhadap Data Twitter," *J. Muara Sains, Teknol. Kedokteran, dan Ilmu Kesehatan*, vol. 2, no. 1, pp. 129–138, 2018.
- [6] S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, IEEE, 2016, pp. 106–111.
- [7] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment analysis using SVM: A systematic literature review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, 2018.
- [8] N. A. R. Putri, "Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," *J. Sains dan Inform.*, vol. 9, no. 2, pp. 136–145, 2023.
- [9] M. Fikri and R. Sarno, "A comparative study of sentiment analysis using SVM and SentiWordNet," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, pp. 902–909, 2019.
- [10] A. Lamurias and F. M. Couto, "Text Mining for Bioinformatics Using Biomedical Literature." 2019.



- 
- [11] K. S. Srujan, S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish, and H. M. Keerthi Kumar, "Classification of amazon book reviews based on sentiment analysis," in *Information Systems Design and Intelligent Applications*, Springer, 2018, pp. 401–411.
  - [12] W. A. Qader, M. M. Ameen, and B. I. Ahmed, "An overview of bag of words; importance, implementation, applications, and challenges," in *2019 international engineering conference (IEC)*, IEEE, 2019, pp. 200–204.
  - [13] H. D. Abubakar, M. Umar, and M. A. Bakale, "Sentiment classification: Review of text vectorization methods: Bag of words, Tf-Idf, Word2vec and Doc2vec," *SLU J. Sci. Technol.*, vol. 4, no. 1 & 2, pp. 27–33, 2022.
  - [14] R. L. Atimi and E. E. Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains Dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022.
  - [15] B. Santosa and A. Umam, *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark*. Penebar Media Pustaka.
  - [16] K. Bhargava and R. Katarya, "An improved lexicon using logistic regression for sentiment analysis," in *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, IEEE, 2017, pp. 332–337.
  - [17] M. Wankhade, A. C. S. Rao, S. Dara, and B. Kaushik, "A sentiment analysis of food review using logistic regression," *Int J Sci Res Comput Sci Eng Inform Technol*, pp. 2–17, 2017.
  - [18] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of artificial intelligence in stock market prediction based on machine learning," *arXiv Prepr. arXiv2107.01031*, 2021.
  - [19] A. M. Rajeswari, M. Mahalakshmi, R. Nithyashree, and G. Nalini, "Sentiment analysis for predicting customer reviews using a hybrid approach," in *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, IEEE, 2020, pp. 200–205.

#### **Biodata Penulis**

**Akhmad Ghiffary Budianto**, merupakan dosen Program Studi Sarjana Rekayasa Elektro di Universitas Lambung Mangkurat. Penulis mendapatkan gelar Sarjana Teknik Industri di Universitas Muhammadiyah Malang pada 2016 dan Magister Teknik Industri di Institut Teknologi Sepuluh Nopember Surabaya pada 2019. Penulis memiliki ketertarikan bidang riset di riset operasi, simulasi komputer, pengambilan keputusan, optimasi dan *big data analytics*.

**Rusilawati**, penulis adalah dosen Program Studi Sarjana Rekayasa Elektro di Universitas Lambung Mangkurat. Penulis memperoleh gelar Sarjana pada bidang Teknik Elektro di Institut Teknologi Nasional Malang pada 1997, gelar Magister di tahun 2010 dan Doktor di tahun 2018 pada bidang Teknik Elektro di Insitut Teknologi Sepuluh Nopember Surabaya. Selanjutnya penulis juga mengambil pendidikan profesi keinsinyuran pada tahun 2022 di Universitas Lambung Mangkurat. Penulis memiliki ketertarikan riset di bidang kestabilan sistem tenaga listrik.

**Arief Trisno Eko Suryo**, penulis merupakan dosen Program Studi Sarjana Rekayasa Elektro di Universitas Lambung Mangkurat. Penulis memperoleh gelar Sarjana pada bidang Teknik Elektro di Universitas Muhammadiyah Malang pada 2014 dan Magister Teknik Elektro di Universitas Brawijaya pada 2020. Penulis memiliki bidang riset di sistem tenaga listrik, sistem kontrol dan kelistrikan.

**Gunawan Rudi Cahyono**, penulis adalah dosen Program Studi Sarjana Rekayasa Elektro di Universitas Lambung Mangkurat. Penulis memperoleh gelar Sarjana pada bidang Teknik Elektro di Institut Teknologi Nasional Malang pada 1998 dan gelar Magister Teknik Elektro di Institut Teknologi Sepuluh Nopember Surabaya pada 2009. Penulis memiliki ketertarikan bidang riset pada artifical intelligence dan sistem kontrol kelistrikan.

**Andry Fajar Zulkarnain**, penulis adalah dosen Program Studi Sarjana Rekayasa Elektro di Universitas Lambung Mangkurat. Penulis memperoleh gelar Sarjana Sains Terapan pada bidang Teknik Komputer Jaringan dan Media Digital di Institut Teknologi Bandung pada 2014 dan gelar Magister Teknik Elektro di Institut Teknologi Bandung pada 2017. Penulis memiliki ketertarikan bidang riset mengenai arsitektur komputer, *Internet of Things (IoT)*, *real-time and embedded system*.

**Martunus**, penulis merupakan mahasiswa pada Program Studi Sarjana Rekayasa Elektro Universitas Lambung Mangkurat. Saat ini penulis aktif dalam keorganisasian himpunan mahasiswa Rekayasa Elektro ULM sebagai ketua himpunan.