Jurnal Sains dan Edukasi Sains
Vol. 9, No. 1, Februari 2026: 1–8
https://doi.org/10.24246/juses.v9i1p1-8

Komparasi Support Vector Machine dengan Logistic Regression terhadap klasifikasi pada data asma

Tiara Utary Grace Moelyono[1], Adi Setiawan[2,*], Rachel Wulan Nirmalasari Wijaya[1]

[1]Program Studi Matematika, Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah 50711

[2]Program Studi Magister Sains Data, Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah 50711

*email korespondensi: adi.setiawan@uksw.edu

Diterima: 27 November 2025; **Direvisi**: 23 Desember 2025; **Disetujui**: 8 Januari 2026; **Diterbitkan**: 9 Maret 2026

ABSTRAK

Single Nucleotide Polymorphisms (SNP) atau Polimorfisme Nukleotida Tunggal merupakan salah satu jenis variasi genetik yang paling umum pada genom manusia. Salah satu contoh data SNP adalah data asma yang diperoleh dari paket program R yaitu SNPassoc. Penyakit asma diantaranya dipengaruhi oleh kondisi genetik manusia. Untuk melihat pengaruh SNP pada genetik manusia dapat dilakukan klasifikasi yang merupakan sebuah proses untuk memisahkan kelas data satu dengan yang lainnya. Penelitian ini bertujuan membandingkan kinerja (*performance*) metode Support Vector Machine (SVM) dan Logistic Regression (LR) untuk klasifikasi terkena penyakit asma atau tidak pada data asma. Empat statistik digunakan guna melihat tingkat kebaikan metode atau kinerja dari metode yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, dan F1-*Score*. Pada proporsi data uji 20%, metode SVM lebih unggul dengan akurasi, presisi, recall, dan F1-Score sebesar 78,76%, 78,85%, 100%, 88,12% dibandingkan metode LR dengan hasil 77,09%, 79,34%, 96,09%, 78,14%. Demikian juga untuk proporsi data uji yang lain, memberikan hasil yang analog. Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa pada kasus ini metode SVM cenderung lebih baik dari pada metode LR.

Kata-kata kunci: *Akurasi; Klasifikasi; Logistic regression; SNP; Support Vector Machine*

PENDAHULUAN

Asma adalah penyakit inflamasi kronik yang mengakibatkan terjadinya pembengkakan serta penyempitan pada saluran jalan nafas sehingga menyebabkan peningkatan hiperaktivitas serta hiperresponsive jalan nafas. Asma dipicu beberapa hal seperti virus, alergen dan aktivitas fisik sehingga menyebabkan bersin, kesulitan bernapas, sesak napas dan batuk yang dapat berbeda seiring waktu dan intensitasnya (Permatasari & Yanti, 2020). Asma ditandai dengan ditandai dengan mengi, sesak nafas, dada terasa berat, dan batuk terutama hal itu terjadi pada malam hari, hiperinflamasi, serta hiperventilasi pernafasan (Arismendi et al., 2024).

Single Nucleotide Polymorphisms (SNP) atau polimorfisme nukleotida tunggal merupakan salah satu jenis variasi genetik yang paling umum pada genom manusia. SNP pada gen yang berperan dalam perbaikan DNA, regulasi siklus sel, metabolisme, dan sistem imun diketahui berhubungan dengan kerentanan genetik terhadap berbagai penyakit (Zhang et al., 2023). Oleh karena itu, SNP banyak digunakan dalam studi asosiasi genetik sebagai representasi variasi genetik yang berkaitan dengan kerentanan terhadap penyakit (Xia et al., 2022). Perkembangan teknologi genotyping SNP berbasis amplicon berkapasitas tinggi memungkinkan diperolehnya data SNP dalam jumlah besar dengan tingkat akurasi yang tinggi, sehingga mendukung pemanfaatannya dalam analisis komputasional modern.

Dalam konteks analisis data genetik tersebut, metode klasifikasi dalam supervised learning digunakan untuk membangun model prediktif berdasarkan karakteristik data, dengan variabel keluaran berupa kategori, seperti kondisi sakit atau tidak sakit (Zhu et al., 2024). Metode klasifikasi dapat diterapkan pada data SNP untuk memprediksi status penyakit seseorang, dengan algoritma yang umum digunakan antara lain Random Forest, Logistic Regression (LR), Decision Tree, K-Nearest Neighbor, Support Vector Machine (SVM), dan Artificial Neural Network.

Penelitian sebelumnya mengenai perbandingan klasifikasi dengan metode SVM dan LR telah dilakukan dalam mengklasifikan data echocardiogram (e.g., lihat Firdausanti et al. (2022). Hasil penelitian tersebut menunjukkan bahwa metode SVM lebih baik daripada LR, ditunjukkan dengan nilai akurasi yang lebih tinggi.

Penelitian ini bertujuan membandingkan kinerja metode SVM dan LR. Kedua metode akan digunakan dalam klasifikasi pada data SNP. Hasil klasifikasi akan dievaluasi menggunakan Confusion Matrix sehingga dapat terlihat metode mana yang lebih baik dalam mengklasifikasikan case control penyakit Asma berdasar data genetik SNP.

EKSPERIMEN

Klasifikasi merupakan proses analisis data yang memisahkan kelas data satu dengan lainnya untuk menentukan kategori suatu objek. Proses klasifikasi dilakukan melalui dua tahap, yaitu *training* dan *testing*. Pada tahap *training*, data sampel dimasukkan ke dalam tabel berisi atribut, jumlah data yang telah terklasifikasi sesuai target, serta nilai *entropy* dan *gain*. Model klasifikasi dibangun menggunakan algoritma SVM dan LR. Selanjutnya, tahap *testing* dilakukan dengan memasukkan data uji yang memiliki atribut sama seperti data *training*, kemudian dibandingkan dengan aturan hasil pelatihan untuk menentukan kelas targetnya.

Metode SVM bekerja dengan mengoptimalkan batas pemisah atau margin antara dua kelas (Ferdowsi et al., 2024). Klasifikasi dilakukan dengan membentuk *hyperplane* pada ruang berdimensi tinggi yang secara optimal memisahkan data menjadi dua kelompok. Untuk data dengan m data pelatihan $x_1, y_1, \dots, x_m, y_m$ dengan $x_i \in R^N$ dan $y_m \in \{-1,1\}$, SVM mencari *weight* w dan *bias* b yang memenuhi persamaan (Cortes & Vapnik, 1995)

$$x^T x + b = 0. \tag{1}$$

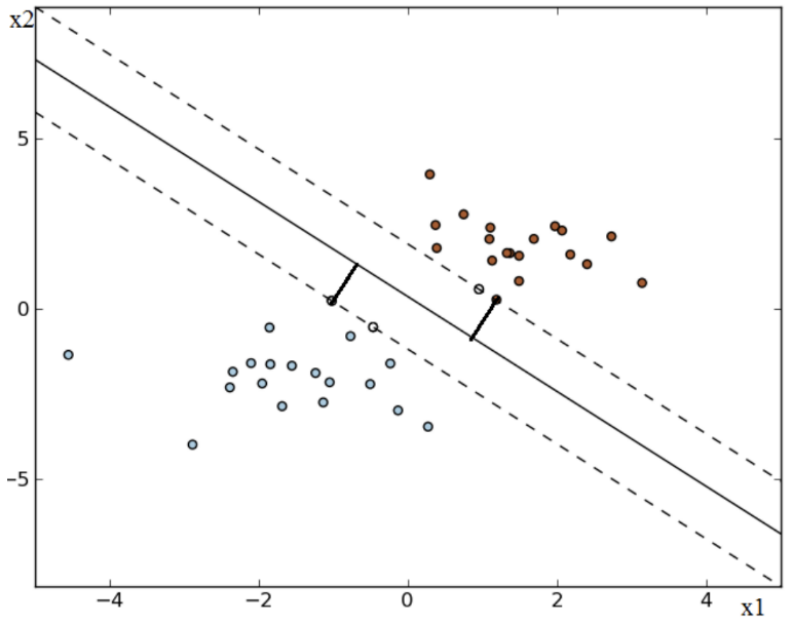
dengan w^T merupakan vektor parameter bobot dan b adalah bias. *Hyperlane* optimal dihitung menggunakan:

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i \left(x_i^T \cdot x\right) + b. \tag{2}$$

Nilai bias b diperoleh menggunakan salah satu support vector yang terletak tepat pada margin optimal. Dengan memanfaatkan kondisi margin tersebut, nilai bias dirumuskan sebagai:

$$b = y_i - w^T x_i. \tag{3}$$

Sebagai ilustrasi **Gambar 1**, dua kelompok data berwarna merah dan hijau dipisahkan menggunakan *hyper-plane*. Titik terdekat dari *hyperplane* disebut *support vector*, sedangkan garis putus-putus menunjukkan *margin* atau jarak terdekat titik dari *hyperplane. Support vector* berperan penting dalam menentukan posisi *hyperplane*, karena perubahan titik ini dapat mengubah batas klasifikasi.



Gambar 1. *Support Vector Machine*

Pada beberapa kondisi, *kernel linear* tidak mampu membagi data secara optimal (Ilham & Gumilang, 2023). Untuk itu digunakan kernel *non-linear* dengan kernel trick yang memetakan data ke *feature space* berdimensi lebih tinggi sehingga pemisahan secara linear dapat dilakukan. Jenis *kernel* yang umum digunakan ditunjukkan pada **Tabel 1**.

Tabel 1. Kernel pada metode klasifikasi SVM

Kernel	Persamaan
Linear	$K(x_i, x) = x_i^T x$
Polinomial	$K(x_i, x) = (\gamma x_i^T x)^p, \gamma > 0$
RBF (Radial Basis Function)	$K(x_i, x) = \text{eks}(-\gamma \lvert x_i - x \rvert^2), \gamma > 0$
Sigmoid	$K(x_i, x) = \tanh(\gamma x_i^T x + r)$

Algoritma LR dipergunakan untuk menjelaskan hubungan variabel dependen dan independen dalam bentuk kategori (Pramakrisna et al., 2022). Berbeda dari regresi linear, LR menghasilkan output kategorikal seperti 0 atau 1, benar-salah, atau ya-tidak. Model logistic regression dapat memiliki berbagai bentuk, termasuk logistik regression biner, one-vs-rest, atau multinomial logistic regression (Erlin et al., 2022). Model yang digunakan dalam penelitian ini adalah *binary logistic regression*, karena data keluaran berupa dua kategori, yaitu 0 (tidak menderita asma) dan 1 (menderita asma). Persamaannya dinyatakan sebagai (Yang et al., 2022):

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \quad (4)$$

dengan β_0 merupakan konstanta (*intercept*) sedangkan β_1, β_2, … ,β_n adalah koefisien yang menunjukkan besarnya pengaruh masing-masing variabel independen terhadap hasil prediksi. Variabel *p* menyatakan probabilitas bahwa suatu observasi termasuk ke dalam kelas positif dengan nilai $0 \leq p \leq 1$.

Evaluasi model dilakukan menggunakan *confusion matrix*, yang terdiri dari empat elemen yaitu *True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)* sebagaimana pada **Tabel 2** (Indrajaya et al., 2022). Dari *confusion matrix* tersebut, dapat dihitung ukuran kinerja model seperti akurasi, presisi, *recall*, dan *F1-score* (**Tabel 3**) (Setiawan, 2022).

Tabel 2. *Confusion matrix*

		Prediksi	
		Kelas 1	**Kelas 2**
Aktual	Kelas 1	TAPI	FN
	Kelas 2	FP	TN

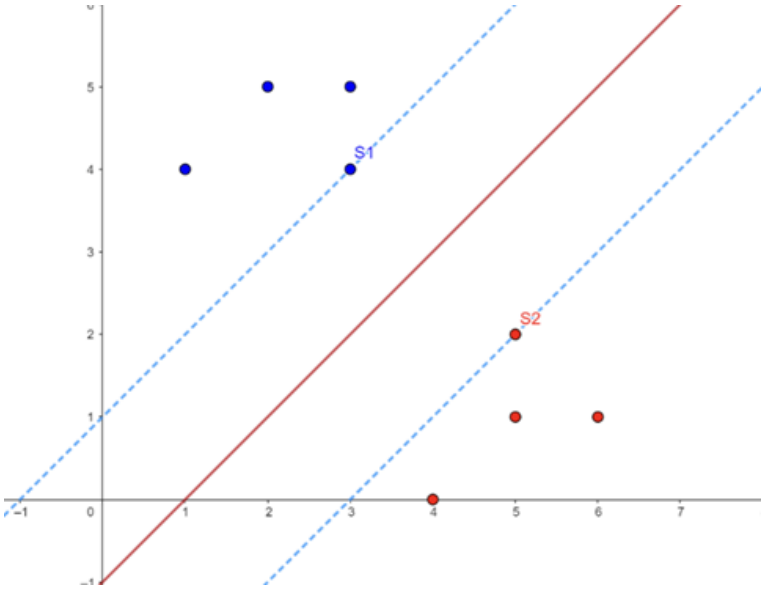
Tabel 3. Ukuran kebaikan model

Nama	**Persamaan**
Akurasi	$\frac{TP + TN}{TP + TN + FP + FN}$
Presisi	$\frac{TP}{TP + FP}$
Sensitivitas (*recall*)	$\frac{TP}{TP + FN}$
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

Sebagai ilustrasi, digunakan data sederhana pada **Tabel 4** yang diklasifikasikan menggunakan SVM dan LR dengan target warna. Dengan metode SVM diperoleh *hyperplane* terbaik $-0.5x_1 + 0.5x_2 + 0.5 = 0$ dengan $w = (-0.5, 0.5)$ dan $b = 0.5$. *Support vector* S1 dan S2 serta *margin* digambarkan pada **Gambar 3**. Hasil perhitungan menunjukkan akurasi, presisi, *recall*, dan *F1-score* sebesar 100%, menandakan pemisahan sempurna. Model LR dengan persamaan $-34.509 + 15.037x_1 - 8.517x_2$ juga menghasilkan metrik evaluasi 100% yang menunjukkan klasifikasi yang sangat akurat.

Tabel 4. Contoh data sederhana

x_1	x_2	**Warna**
1	4	Biru
2	5	Biru
3	5	Biru
3	4	Biru
6	1	Merah
4	0	Merah
5	2	Merah
5	1	Merah



Gambar 2. Klasifikasi data sederhana dengan metode SVM

Data utama yang digunakan dalam penelitian ini adalah data "Asthma" dari paket R "SNPassoc". Data ini memiliki 57 kolom dan 1578 baris (**Gambar 3**), yang setelah pembersihan *missing values* menjadi 1076 baris (**Gambar 4**). Data dikodekan menggunakan *machine learning encoding* dan dibagi menjadi data *training* serta *testing* untuk menghindari *overfitting* (Zhao et al., 2024). Skoring dilakukan pada kolom *country* dan *genetic code* sebagaimana pada **Tabel 5** dan **Tabel 6**, sedangkan skoring untuk *Gender* bernilai biner (0 untuk pria dan 1 untuk wanita).

	country	gender	age	bmi	smoke	casecontrol	rs4490198	rs4849332	rs1367179
1	Germany	Males	42.80630	20.14797	1	0	GG	TT	GC
2	Germany	Males	50.22861	24.69136	0	0	GG	GT	GC
3	Germany	Males	46.68857	27.73230	0	0	GG	TT	GC
4	Germany	Females	47.86311	33.33187	0	0	AG	GT	GG
5	Germany	Females	48.44079	25.23634	0	1	AG	GG	GG
6	Germany	Females	36.90075	29.37240	1	0	AG	GT	GG
7	Germany	Females	51.01437	41.09978	0	0	GG	TT	GC
8	Germany	Males	41.04586	24.93372	0	0	AG	TT	GG
9	Germany	Females	43.32101	33.30499	0	0	AG	GT	GG

Gambar 3. Data asma awal

	country	gender	age	bmi	smoke	casecontrol	X1	X2	X3	X4	X5	X6	X7
1	0	0	42.80630	20.14797	1	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0	0	50.22861	24.69136	0	0	0.0	0.5	0.0	0.0	0.5	0.5	0.5
3	0	0	46.68857	27.73230	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0	1	47.86311	33.33187	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
5	0	1	48.44079	25.23634	0	1	0.5	1.0	0.5	0.5	1.0	1.0	1.0
6	0	1	36.90075	29.37240	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
7	0	1	51.01437	41.09978	0	0	0.0	0.0	0.0	0.0	0.5	0.5	0.5
8	0	0	41.04586	24.93372	0	0	0.5	0.0	0.5	0.5	0.5	0.5	0.5
9	0	1	43.32101	33.30499	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Gambar 4. Data asma setelah dilakukan konversi

Tabel 5. Skoring kolom negara

Konversi	Nama negara	Jumlah data	Konversi	Nama negara	Jumlah data
1	Germany	131	6	Sweden	215
2	Switzerland	86	7	Australia	19
3	Belgium	12	8	Estonia	5
4	UK	106	9	France	152
5	Spain	277	10	Norway	73

Tabel 6. Skoring kolom kode genetik

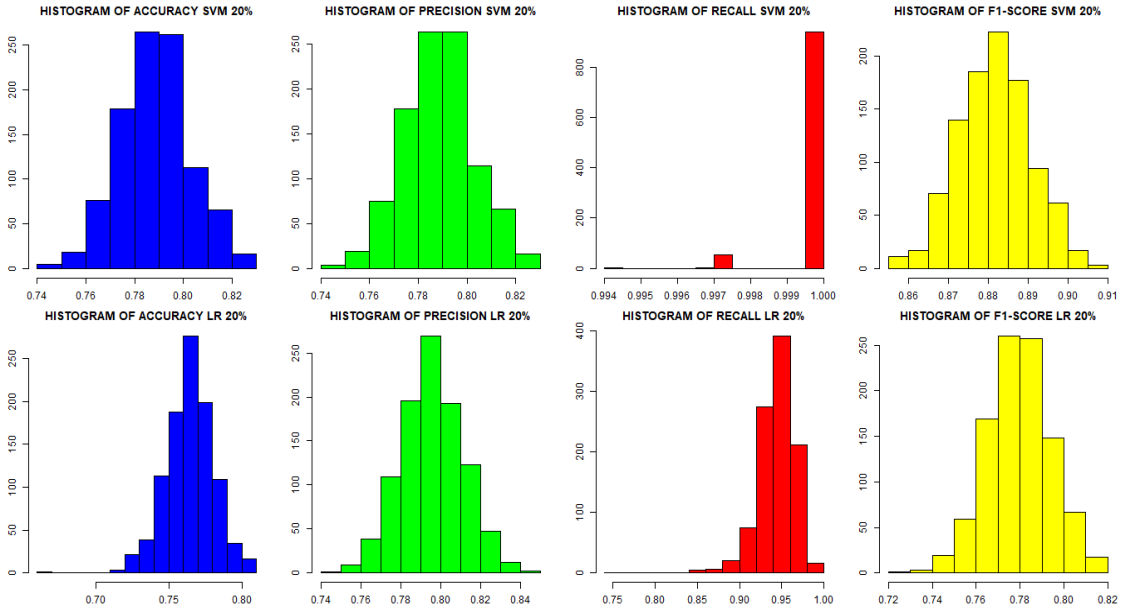
AA	AC	CC	CC	CG	GG
0	0,5	1	0	0,5	1
AA	AG	GG	CC	CT	TT
0	0,5	1	0	0,5	1
AA	AT	TT	GG	GT	TT
0	0,5	1	0	0,5	1

HASIL DAN BAHASAN

Percobaan dilakukan dengan mengulang pemilihan data training dan data testing secara random sebanyak bilangan besar *B* kali dan dalam hal ini digunakan *B* = 1000 kali. Perbandingan nilai akurasi, presisi, *recall*, dan *F1-score* untuk proporsi uji data 40% ditunjukkan pada **Gambar 5**. Berdasarkan histogram, nilai akurasi, presisi, dan *F1-score* kedua metode menunjukkan distribusi yang relatif normal, sedangkan nilai *recall* pada metode SVM cenderung menceng ke kanan. Secara lebih terperinci hasil tersebut juga dinyatakan pada **Tabel 7**.

Tabel 7. Tabel statistik deskriptif hasil pengulangan 1000 kali untuk tiap-tiap metode yaitu SVM dan LR pada data uji 40%.

Statistik	SVM				LR			
	Akurasi	Presisi	*Recall*	F1-*Score*	Akurasi	Presisi	*Recall*	F1-*Score*
Min.	0,74	0,74	0,99	0,85	0,7090	0,7530	0,8601	0.7365
1st Qu.	0,77	0,77	1,00	0,87	0,7529	0,7850	0,9304	0,7691
Median	0,788	0,78	1,00	0,88	0,7650	0,7966	0,9448	0,7800
Mean	0,789	0,78	0,99	0,88	0,7638	0,7961	0,9424	0,7795
3rd Qu.	0,79	0,79	1,00	0,88	0,7748	0,8059	0,9560	0,7900
Max.	0,82	0,82	1,00	0,90	0,8023	0,8371	0,9824	0,8174



Gambar 5. Histogram data uji 40%

Pengujian dilakukan menggunakan SVM dan *Logistic Regression* (LR), berdasarkan pembagian proporsi data *training* dan *testing*. Hasil pengujian median setiap metrik evaluasi untuk metode SVM dan LR disajikan pada **Tabel 8** dan **Tabel 9**. Berdasarkan **Tabel 8**, metode SVM menunjukkan nilai akurasi tertinggi sebesar 0,7884 pada proporsi uji data 10%, sedangkan presisi tertinggi sebesar 0,7898 diperoleh pada proporsi 40%. Nilai *recall* tertinggi yaitu 1,0000 muncul pada proporsi 40%, 30%, dan 20%, sementara *F1-score* tertinggi sebesar 0,8826 juga diperoleh pada proporsi 40%.

Untuk metode *Logistic Regression* (**Tabel 9**), akurasi tertinggi sebesar 0,7683 dicapai pada proporsi uji data 30%, presisi tertinggi sebesar 0,7966 pada proporsi 40%, *recall* tertinggi sebesar 0,9667 pada proporsi 10%, dan *F1-score* tertinggi sebesar 0,7839 pada proporsi 20%.

Tabel 8. Hasil median uji data dengan metode SVM

Proporsi uji data	Akurasi	Presisi	*Recall*	F1-*Score*
40%	0,7897	0,7898	1,0000	0,8826
30%	0,7882	0,7882	1,0000	0,8816
20%	0,7876	0,7885	1,0000	0,8812
10%	0,7884	0,7890	0,9990	0,8812

Tabel 9. Hasil median uji data dengan metode LR

Proporsi uji data	Akurasi	Presisi	*Recall*	F1-*Score*
40%	0,7650	0,7966	0,9448	0,7801
30%	0,7683	0,7953	0,9547	0,7809
20%	0,7709	0,7934	0,9609	0,7814
10%	0,7768	0,7945	0,9667	0,7839

Hasil pengujian pada **Tabel 8** dan **Tabel 9** menunjukkan bahwa metode *Support Vector Machine* menghasilkan nilai akurasi, presisi, *recall*, dan *F1-score* yang lebih tinggi dibandingkan *Logistic Regression*. Performa SVM yang lebih baik dapat dijelaskan melalui kemampuannya dalam memodelkan hubungan non-linear menggunakan *kernel trick*, sehingga mampu membentuk *hyperplane* optimal dalam ruang berdimensi tinggi.

Sebaliknya, regresi logistik cenderung terpengaruh oleh masalah *multikolinearitas*, yaitu kondisi ketika variabel prediktor saling berkorelasi tinggi sehingga menyebabkan ketidakstabilan pada koefisien (Arora et al., 2018). Kondisi ini mempengaruhi performa klasifikasi karena regresi logistik mengasumsikan hubungan linear antara variabel independen dan *log odds*. Ketidaksesuaian hubungan tersebut dapat menghasilkan kesalahan klasifikasi.

Lebih jauh SVM mengambil pendekatan geometris dalam klasifikasi, yang membuatnya lebih efisien pada ruang fitur besar dan lebih tangguh terhadap data non-linear dibandingkan pendekatan probabilistik pada regresi logistik. Hasil penelitian ini kemudian dibandingkan dengan penelitian-penelitian sebelumnya seperti yang ditunjukkan pada **Tabel 10**. Berdasarkan beberapa literatur, kinerja SVM dan LR bervariasi tergantung pada jenis data yang digunakan. Pada penelitian Sitompul et al. (2023) dan Osei-Nkwantabisa & Ntumy (2024), akurasi LR sedikit lebih tinggi daripada SVM, namun tidak berbeda signifikan. Sebaliknya, penelitian oleh Gusti (2023) dan Firdausanti et al. (2022) menunjukkan bahwa SVM memiliki akurasi lebih tinggi. Hal ini menunjukkan bahwa keunggulan kedua metode bersifat kontekstual dan sangat bergantung pada karakteristik data.

Tabel 10. Penelitian sebelumnya mengenai perbandingan SVM dan LR

Penulis	*Data* yang digunakan	Akurasi SVM	Akurasi LR
Sitompul et al. (2023)	Ulasan Aplikasi Grab	85,5%	85,6%
Osei-Nkwantabisa & Ntumy (2024)	Penyakit jantung	81%	85%
Gusti (2023)	Tweet Bencana Alam	80,41%	70,74%
Firdausanti et al. (2022)	*Echocardiogram Dataset*	91,54%	88%
Baihaqi et al. (2023)	*Sentiment Tweet*	88,8%	90,5%

KESIMPULAN

Pada proporsi uji data 20%, hasil perbandingan memperjelas bahwa metode *Support Vector Machine* menghasilkan kinerja (akurasi, presisi, *recall* dan *F1 Score*) yang lebih tinggi dibandingkan metode *Logistic Regression* dengan median akurasi 78,76%, presisi 78,85%, *recall* 100%, dan *F1-Score* 88,12%. Hasil yang analog juga berlaku untuk proporsi uji data yang lain yaitu 10%, 30% dan 40%. Penelitian selanjutnya dapat dilakukan dengan memberi skoring yang berbeda untuk kode genetik, atau hanya menggunakan kode genetik saja tanpa fitur lainnya, maupun memisahkan *case control* yang positif dan negatif. Demikian juga proporsi data uji juga bisa digunakan yang lain seperti 5%, 15%, 25%, 35% dan 45%. Hasil penelitian dapat dibandingkan untuk melihat adanya perubahan maupun peningkatan dalam akurasi kedua metode tersebut.

DAFTAR PUSTAKA

Arismendi, E., Ribo, P., García, A., Torrego, A., Bobolea, I., Casas-Saucedo, R., Palomino, R., Picado, C., Munoz-Cano, R., & Valero, A. (2024). Asthma control according to GINA 2023: Does changing the criteria improve asthma control? *Journal of Clinical Medicine*, *13*(22), 6646. https://doi.org/10.3390/jcm13226646

Arora, A., Lin, J.-J., Gasperian, A., Maldjian, J., Stein, J., Kahana, M., & Lega, B. (2018). Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings. *Journal of Neural Engineering*, *15*(6). https://doi.org/10.1088/1741-2552/aae131

Baihaqi, K. A., Setyawan, I., Manongga, D., Purnomo, H. D., Hendry, H., Fauzi, A., & Hananto, A. (2023). A comparison Support Vector Machine, Logistic Regression and Naive Bayes for classification sentimen analisys user mobile app. *International Journal of Artificial Intelligence Research*, *7*(1), 64–71. https://doi.org/10.29099/ijair.v7i1.962

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

Erlin, Marlim, Y. N., Junadhi, Suryati, L., & Agustina, N. (2022). Deteksi dini penyakit diabetes menggunakan Machine Learning dengan algoritma Logistic Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *11*(2), 88–96. https://doi.org/10.22146/jnteti.v11i2.3586

Ferdowsi, M., Kwan, B. H., Tan, M. P., Saedon, N. I., Subramaniam, S., Hashim, N. F. I. A., Nasir, S. S. M. Z., Chee, K. H., & Goh, C. H. (2024). Classification of vasovagal syncope from physiological signals on tilt table testing. *BioMedical Engineering OnLine 2024 23:1*, *23*(1), 37. https://doi.org/10.1186/s12938-024-01229-9

Firdausanti, N. A., Ningrum, R. A., & Qomariyah, S. (2022). Comparisons of Logistic Regression and Support Vector Machines in classification of echocardiogram dataset. *Inferensi*, *5*(2), 85–90. https://doi.org/10.12962/j27213862.v5i2.14121

Gusti, K. W. (2023). Perbandingan metode Support Vector Machine dan Logistic Regression untuk klasifikasi bencana alam. *Informatik : Jurnal Ilmu Komputer*, *19*(2), 134–140. https://doi.org/10.52958/iftk.v19i2.6355

Ilham, M., & Gumilang, F. (2023). Sentiment analysis kualitas pelayanan maskapai penerbangan. *Jurnal Siliwangi Seri Sains dan Teknologi*, *9*(2), 38–43.

Osei-Nkwantabisa, A. S., & Ntumy, R. (2024). *Classification and prediction of heart diseases using machine learning algorithms*. https://doi.org/10.48550/arXiv.2409.03697

Permatasari, D., & Yanti, B. (2020). Perbedaan diagnosis asma, penyakit paru obstruktif kronik dan asthma-COPD overlap syndrome. *Jurnal Kedokteran Syiah Kuala*, *20*(3). https://doi.org/10.24815/jks.v20i3.18640

Pramakrisna, F. D., Adhinata, F. D., Annisa, N., & Tanjung, F. (2022). Aplikasi klasifikasi SMS berbasis web menggunakan algoritma Logistic Regression. *Teknika*, *11*(2), 90–97. https://doi.org/10.34148/teknika.v11i2.466

Setiawan, A. (2022). Perbandingan penggunaan jarak Manhattan, jarak Euclid, dan jarak Minkowski dalam klasifikasi menggunakan metode KNN pada data Iris. *Jurnal Sains Dan Edukasi Sains*, *5*(1), 28–37. https://doi.org/10.24246/juses.v5i1p28-37

Sitompul, K. P. J., Pratama, A. R., & Baihaqi, K. A. (2023). Komparasi algoritma Naive Bayes, Support Vector Machine, dan Logistic Regression pada analisis sentimen pengguna aplikasi transportasi online. *Kumpulan Jurnal Ilmu Kompute*, *10*(1), 27–38.

Xia, Z., Wang, Y., Liu, F., Shu, H., & Huang, P. (2022). Association between TNF-α-308, +489, −238 polymorphism, and COPD susceptibility: An updated meta-analysis and trial sequential analysis. *Frontiers in Genetics*, *12*, 772032. https://doi.org/10.3389/fgene.2021.772032

Yang, Z., Wang, J., & Zhu, Y. (2022). Few-Shot classification with contrastive learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 13680 LNCS*. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.48550/arXiv.2209.08224

Zhang, L., Pozsgai, E., Song, Y., Macharia, J., Alfatafta, H., Zheng, J., Li, Z., Liu, H., & Kiss, I. (2023). The relationship between single nucleotide polymorphisms and skin cancer susceptibility: A systematic review and network meta-analysis. *Frontiers in Oncology*, *13*, 1094309. https://doi.org/10.3389/fonc.2023.1094309

Zhao, J., Wan, R., & Miao, D. (2024). Conflict analysis triggered by three-way decision and Pythagorean fuzzy rough set. *International Journal of Computational Intelligence Systems 2024 17:1*, *17*(1), 17. https://doi.org/10.1007/s44196-023-00378-4

Zhu, X., Li, J., Liu, Y., Ma, C., & Wang, W. (2024). A survey on model compression for large language models. *Transactions of the Association for Computational Linguistics*, *12*, 1556–1577. https://doi.org/10.1162/tacl_a_00704