

Komparasi Metode SVM dan Logistic Regression untuk Klasifikasi Hipotesa Penyakit Kanker Paru Paru Berdasarkan Gejala Awal

Shafara Rahmaeda¹, Rastri Prathivi²

Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia
E-mail: shafararhm@gmail.com¹, vivi@usm.ac.id²

Abstract

Lung cancer is the uncontrolled growth of cancer cells in lung tissue that occurs due to various carcinogenic substances. Throughout Indonesia, this disease is still the leading cause of death from cancer. The main risk factors include smoking habits, exposure to cigarette smoke, chest pain. Namely, classification is one way of early detection that can reduce the death rate of lung cancer. Various classification techniques have been proposed in various fields such as machine learning and expert systems. In machine learning, there are two methods used in classification, namely SVM and Logistic Regression. The advantage of SVM is to divide data into hyperplanes so that the data space is divided into two classes. SVM theory begins by collecting data that can be separated by a straight line using a hyperplane, then grouped by class. While Logistic Regression is used to describe the relationship between categorical response variables and covariates. Specifically, there is a direct relationship between the independent variable and the logarithm of the probability of an event occurring. This study aims to compare which is the best using the SVM algorithm and the Logistic Regression Algorithm in the classification of lung cancer. The lung cancer disease dataset has a total of 309 data where the data is separated into two parts, namely 70% training data consisting of 216 data, while 30% test data consists of 93 data. The performance used in predicting the model is Accuracy, Precision, Recall, and F1-Score. From the research conducted, the Accuracy value of the Logistic Regression Algorithm was 97.85%. In this case, the Logistic Regression algorithm has better performance in classifying lung cancer than the SVM algorithm.

Keywords: Lung cancer, early detection, classification, machine learning

Abstrak

Kanker paru-paru adalah perkembangan sel kanker yang tidak terkendali di jaringan paru-paru yang terjadi akibat berbagai zat karsinogenik. Di seluruh Indonesia, penyakit ini masih menjadi penyebab utama kematian akibat kanker. Faktor risiko utama meliputi kebiasaan merokok, paparan asap rokok, nyeri dada. Yakni klasifikasi salah satu cara deteksi dini yang dapat mengurangi Tingkat kematian kanker paru-paru. Beragam teknik klasifikasi telah diajukan dalam berbagai bidang seperti pembelajaran mesin dan sistem pakar. Dalam ilmu pembelajaran mesin, terdapat dua metode yang digunakan dalam klasifikasi, yaitu Support Vector Machine (SVM) dan Logistic Regression. Keunggulan dari SVM untuk membagi data ke dalam hyperplane sehingga ruang data terbagi menjadi dua kelas. Teori SVM dimulai dengan mengumpulkan data yang dapat dipisahkan dengan garis lurus menggunakan hyperplane, lalu dikelompokkan berdasarkan kelas. Sedangkan Logistic Regression digunakan untuk menggambarkan hubungan antara variabel respons kategori dengan kovariat. Secara spesifik terdapat hubungan langsung antara variabel bebas dengan logaritma kemungkinan terjadinya suatu peristiwa. Penelitian ini bertujuan untuk membandingkan mana yang paling baik dengan menggunakan algoritma SVM dan Algoritma Logistic Regression pada klasifikasi penyakit kanker paru-paru. Dataset

penyakit kanker paru-paru memiliki jumlah data sebanyak 309 data dimana Data dipisahkan menjadi dua bagian, yaitu 70% data latih yang terdiri dari 216 data, Sementara 30% data uji terdiri dari 93 data. Performa yang digunakan dalam memprediksi model adalah Accuracy, Precision, Recall, dan F1-Score. Dari penelitian yang dilakukan memperoleh nilai Accuracy pada Algoritma Logistic Regression sebesar 97.85%. Dalam hal ini algoritma Logistic Regression memiliki performa lebih baik dalam melakukan klasifikasi penyakit kanker paru-paru dibandingkan algoritma SVM.

Kata Kunci: Kanker paru-paru, deteksi dini, klasifikasi, pembelajaran mesin

1. Pendahuluan

Kanker merupakan penyakit yang sering disebut terjadi ketika sekelompok sel-sel yang ganas tumbuh tidak normal dan tanpa terkendali, melewati batas sehat, lalu menyebar ke organ-organ tubuh yang sehat melalui pembuluh darah atau pembuluh getah bening untuk membentuk massa sel-sel yang baru. Namun, perlu diingat bahwa kanker tidak dapat menular [1]. Di Seluruh dunia kanker paru – paru masih menjadi penyebab utama kematian akibat kanker. Menurut informasi terbaru dari Global Cancer Observatory (GLOBOCAN), pada tahun 2020 dilaporkan sekitar 2,2 juta kasus kanker paru-paru baru dan 1,8 juta kematian karena penyakit tersebut. Di Indonesia, kanker paru-paru berada di posisi kedua di Asia Tenggara dan posisi ke-23 di Asia. Jumlah kasus baru mencapai 141,1 juta dengan jumlah kematian mencapai 85,1 juta [2] [3] [4]. Paru-paru adalah organ penting di dalam tubuh manusia yang bertugas menjadi bagian dari sistem pernapasan untuk memperlancar aliran udara bersama Trakea. Meskipun begitu, organ ini rentan terkena penyakit yang bisa memengaruhi kerjanya. Beberapa penyakit yang sering menyerang paru-paru, tidak peduli usia atau jenis kelamin, antara lain Asma, PPOK, Pneumonia, Tuberkulosis, dan kanker paru [5]. Sebagian besar kasus kanker paru dikaitkan dengan kebiasaan merokok, dimana sekitar 90% terjadi pada pria dan sekitar 70% pada wanita yang terpapar asap rokok [6].

Menurut Kementerian Kesehatan Republik Indonesia (2014), paparan asap rokok adalah faktor utama yang paling umum menyebabkan kanker paru-paru, baik dari perokok aktif maupun pasif [7]. Terlebih lagi, kenaikan risiko terkena kanker paru-paru juga disebabkan oleh polusi udara dan paparan lingkungan kerja. Seperti nitrogen, senyawa hidrokarbon, Tar, nikotin, benzopiren, fenol dan cadmium mengandung zat yang tidak sehat di dalam asap rokok. Tar salah satu komponen utama yang berbahaya di dalam asap rokok jugamenyebabkan penyakit serius, seperti kanker, penyakit jantung, bronchitis serta gangguan kehamilan [8]. Ada faktor lain yang memiliki pengaruh besar dalam penyebab terjadinya kanker paru-paru, seperti kebiasaan minum alkohol, faktor genetik, dan obesitas [9][10]. Sebuah penelitian mengungkapkan bahwa penyintas kanker paru-paru yang kurang bergerak tidak hanya mengalami gejala batuk yang berkepanjangan, tetapi juga sering merasa kelelahan, nyeri, dan kesulitan bernapas yang lebih intens dibandingkan dengan mereka yang aktif secara fisik [11]. Berdasarkan sejumlah fakta yang telah diidentifikasi, tujuan penelitian ini adalah untuk memprediksi peluang seseorang menerima diagnosis kanker paru-paru, dengan mempertimbangkan informasi terkait dengan lingkungan (seperti tingkat polusi udara), gaya hidup (seperti kebiasaan merokok, paparan asap rokok, dan konsumsi alkohol), serta kondisi kesehatan individu (seperti tingkat alergi, risiko genetik, gejala seperti nyeri dada, batuk, sesak napas, penebalan kuku, dan tingkat mendengkur).

Pentingnya deteksi awal melalui klasifikasi kanker untuk memungkinkan tenaga kesehatan memberikan intervensi medis dengan cepat. Menyadari penyakit sejak awal sangat penting untuk mengenali potensi atau masalah kesehatan sebelum berubah menjadi kondisi yang lebih serius. Pendekatan ini sangatlah penting dalam meningkatkan efektivitas pengobatan serta memberikan peluang kesembuhan yang lebih tinggi. Dengan

mendiagnosis penyakit lebih awal, prosedur medis dapat dilakukan dengan lebih cepat, seringkali lebih efektif, dan kurang invasif dibandingkan dengan penanganan saat penyakit telah berkembang lebih lanjut. Sebagai contoh, dalam situasi kanker, deteksi yang lebih awal dapat dicapai dengan menjalani pemeriksaan rutin dan mengenali gejala-gejala yang muncul pada tahap awal [12][13]. Strategi ini telah terbukti efektif dalam meningkatkan tingkat kesuksesan pengobatan serta mengurangi angka kematian akibat kanker (Cancer Research UK). Menurut penelitian yang dilakukan National Cancer Institute pada tahun 2019, menemukan bahwa deteksi kanker paru-paru pada tahap awal dapat meningkatkan peluang hidup lima tahun dari sekitar 18% menjadi lebih dari 55%. Kanker paru-paru adalah kondisi yang serius yang sering kali disebabkan oleh pertumbuhan tumor ganas di epitel bronkus, sehingga sulit untuk dideteksi secara dini sebelum menjadi parah. Sebagian besar, sekitar 85% kasus kanker paru-paru diidentifikasi pada stadium lanjut. Karena itu, sangatlah Penting untuk mendeteksi dan mengklasifikasikan nodul pada paru-paru dengan akurat guna memastikan diagnosis yang tepat [14].

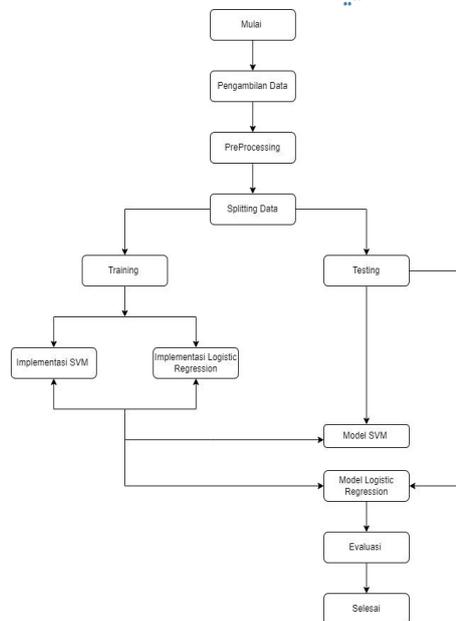
Klasifikasi merupakan metode analisis data yang membantu dalam memperkirakan label kelas mana yang cocok untuk dimasukkan ke dalam sampel. Beragam teknik klasifikasi telah diajukan dalam berbagai bidang seperti pembelajaran mesin dan sistem pakar. Agar terhindar dari efek samping yang berbahaya, bahkan risiko kematian, kanker paru-paru harus segera diobati dan dicegah setelah diagnosis ditegakkan. Sangatlah penting untuk mengenali serta mengatasi persoalan sejak pada tahap awal [15]. Pembelajaran mesin merupakan subkategori dari kecerdasan buatan yang menekankan pemanfaatan algoritma dan metode khusus dalam memprediksi, mengenali pola, dan mengklasifikasikan data. Support Vector Machine dan Logistic Regression dua metode tersebut sering dipakai dalam klasifikasi. Salah satu keunggulan dari Support Vector Machine (SVM) adalah popularitasnya, selain dari kemampuannya dalam klasifikasi. SVM tidak tergantung pada jumlah atribut dan mampu menangani masalah yang berkaitan dengan dimensi dengan baik. Dalam hal komputasi, Support Vector Machine (SVM) dapat melakukan pelatihan dengan cepat, dan teknik pembelajarannya dapat mengatasi tantangan yang berkaitan dengan keraguan. Regresi Logistik digunakan untuk menjelaskan hubungan antara variabel respons kategori dengan kovariat. Secara spesifik, terdapat hubungan langsung antara variabel bebas dengan logaritma kemungkinan terjadinya suatu peristiwa.

Hasil penelitian yang dilakukan oleh Nor Kumalasari Caecar Pratiwi, Nur Ibrahim, dan Sofia Saidah menunjukkan nilai akurasi, presisi, serta recall berturut-turut sebesar 86%, 85%, dan 97,2% [9]. Penelitian lain yang telah dilakukan oleh Dina Mariana dan Hafiz Irsyad menunjukkan bahwa nilai presisi mencapai 62,33%, sementara recall mencapai 78,50%, dengan tingkat akurasi sebesar 65,63% [16]. Hasil penelitian lain dari Muhammad Iqbal Yunan Helmi, Dian Anggraeni, dan Alfian Futuhul Hadi menunjukkan bahwa metode Support Vector Machine (SVM) yang diuji menggunakan kernel linear melalui 5-fold cross validation dan juga 10-fold cross validation, berhasil mencapai tingkat ketepatan klasifikasi sebesar 90%. Adapun, dalam konteks yang berbeda, metode Naïve Bayes juga dilakukan dengan 5-fold cross validation dan 10-fold cross validation, dan berhasil mencapai tingkat ketepatan klasifikasi sebesar 75% [17]. Studi lain yang dilakukan oleh Priscilia Lovita memperlihatkan tingkat akurasi sebesar 87,7% [4].

2. Metodologi Penelitian

Data dalam penelitian ini mencakup informasi tentang gejala awal pada pasien yang telah didiagnosis menderita kanker paru-paru, juga pasien yang tidak mengidap kanker paru-paru. Baik metode SVM maupun Logistic Regression mengaplikasikan data latih dan data uji, dengan kinerjanya dinilai melalui confusion matrix. Pendekatan ini sejalan dengan metodologi yang diterapkan dalam studi terbaru [18], yang membandingkan

berbagai algoritma machine learning dalam deteksi kanker paru-paru. Temuan dari penelitian ini menunjukkan perbandingan keefektifan antara SVM dan Logistic Regression dalam memprediksi kanker paru-paru dari gejala awal. Sehingga mampu meningkatkan kemampuan untuk mendeteksi kanker paru-paru lebih awal. Gambar 1 langkah langkah yang dilakukan pada penelitian ini.



Gambar 1. Alur Penelitian

2.1. Pengambilan Data

Dataset yang dipergunakan dalam penelitian ini terdiri 309 data dengan 16 kolom yang berasal dari dataset publik yang ada di Kaggle. Atribut data set ini meliputi berbagai informasi seperti usia, jenis kelamin, kebiasaan merokok, warna jari, tingkat kecemasan, tekanan teman sebaya, riwayat penyakit jangka panjang, tingkat kelelahan, reaksi alergi, mendengkur, asupan alkohol, batuk, napas tersengal, kesulitan menelan, nyeri dada, serta hasil diagnosis kanker paru-paru.

2.2. PreProcessing Data

PreProcessing data yaitu tahap mengolah data mentah agar siap untuk proses selanjutnya. Proses ini mencakup menghilangkan data yang tidak bermanfaat atau kosong serta merapikan data dengan mengaplikasikan teknik perawatan dan transformasi data. Proses ini berhasil mengatur data yang akan diproses dengan lebih terstruktur, sehingga memudahkan dalam melakukan pemodelan [19].

2.3. Splitting Data

Setelah proses preprocessing selesai, data dipisahkan menjadi data latih dan data uji. Data dipisahkan menjadi dua bagian, yaitu 70% data latih yang terdiri dari 216 data untuk melatih model atau instruksi. Sementara 30% sisanya digunakan untuk mengukur keberhasilan atau kegagalan penelitian, yang biasa disebut data uji, terdiri dari 93 data. Saat ini, data pelatihan sedang digunakan untuk menjalankan penelitian yang relevan [20].

2.4. Model SVM dan Logistic Regression

Algoritma SVM digunakan untuk tujuan klasifikasi maupun regresi. Cara kerjanya mengacu pada konsep Structural Risk Minimization (SRM), dengan tujuan untuk membagi data ke dalam hyperplane sehingga ruang data terbagi menjadi dua kelas. Teori

SVM dimulai dengan mengumpulkan data yang dapat dipisahkan dengan garis lurus menggunakan hyperplane, lalu dikelompokkan berdasarkan kelas. [12] Berikut rumus prediksi untuk klasifikasi menggunakan SVM :

$$f(X) = W \cdot X + b \tag{1}$$

Rumus dari klasifikasi linear SVM

Keterangan:

Rumus dari klasifikasi linear SVM

Keterangan:

W = vector bobot

X = vector fitur dari sampel yang akan diprediksi

b = bias atau intersep

$$f(X) = \sum_{i=1}^N a_i y_i K(x_i, x) + b \tag{2}$$

Rumus untuk klasifikasi non-linear SVM

Keterangan:

N = jumlah total sampel dalam dataset pelatihan.

a_i = koefisien Lagrange yang diperoleh dari pelatihan.

y_i = label kelas dari sampel pelatihan i .

$K(x_i, x)$ = fungsi kernel yang mengukur kesamaan antara sampel pelatihan dan sampel yang akan diprediksi.

Logistic Regression adalah suatu metode regresi yang diterapkan secara khusus untuk memproyeksikan koneksi antara variabel dependen dua nilai dengan berbagai indikator yang bisa berupa data kontinu, kategorikal, atau campuran. Metode ini tidak memerlukan asumsi tentang distribusi multivariat yang normal atau keseimbangan matriks kovarians variabel, sehingga dapat digunakan untuk data dengan skala yang beragam.[15]. Berikut rumus dari Logistic Regression:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = B_0 + B_1 X \tag{3}$$

Keterangan :

\ln = Logaritma Natural

\hat{p} = Probabilitas Logistik, $B_0 + B_1 X$

2.5. Evaluasi

Tabel Confusion Matrix memberikan informasi tentang kesalahan yang terjadi dalam model dengan membandingkan prediksi yang tepat serta yang tidak tepat untuk setiap kelas. Dalam confusion matrix, terdapat beberapa evaluasi yang sering dimanfaatkan seperti presisi, akurasi, recall, dan skor F1 yang populer digunakan. Akurasi menggambarkan sejauh mana keakuratan prediksi dibandingkan dengan total data, mencerminkan tingkat kesesuaian antara hasil prediksi dan nilai aktual. Untuk menilai kinerja model klasifikasi dalam memprediksi penyakit kanker paru-paru, kita menggunakan confusion matrix. Tabel 2 menggambarkan klasifikasi dua kelas berdasarkan [21].

Tabel 1. Rumus Confusion Matrix

Prediction		1	0
	1	TP	FP
	0	FN	TN

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{6}$$

$$\text{F1 - Score} = 2x \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Keterangan:

TP (True Positive) = Data positif diklasifikasikan secara benar.

TN (True Negative) = Data negatif diklasifikasikan secara benar.

FP (False Positive) = Data negatif diklasifikasikan menjadi positif.

FN (False Negative) = Data positif diklasifikasikan menjadi negatif.

3. Hasil dan Pembahasan

3.1. Proses Pengumpulan Data

Pada Tabel 1 terdapat rincian jumlah data yang digunakan dalam penelitian ini dan dalam proses pengolahan data. Data tersebut mencakup identifikasi jenis kelamin pasien (1 untuk laki-laki dan 2 untuk perempuan), rentang usia pasien 38 hingga 87 tahun, keberadaan kebiasaan merokok pasien (Ya=2 atau Tidak=1), penanda fisik berupa jari yang menguning (Ya=2 atau Tidak=1), tingkat kecemasan pasien (Ya=2 atau Tidak=1), penilaian terhadap tekanan dari teman sebaya atau sosial terhadap pasien (Ya=2 atau Tidak=1), keberadaan penyakit kronis pada pasien (Ya=2 atau Tidak=1), kecenderungan kelelahan pasien (Ya=2 atau Tidak=1), keberadaan alergi pada pasien (Ya=2 atau Tidak=1), keberadaan asma pada pasien (Ya=2 atau Tidak=1), kebiasaan mengonsumsi alkohol pasien (Ya=2 atau Tidak=1), seringnya pasien mengalami batuk (Ya=2 atau Tidak=1), kesulitan bernapas pada pasien (Ya=2 atau Tidak=1), kesulitan menelan makanan atau cairan pada pasien (Ya=2 atau Tidak=1), adanya nyeri pada area dada pasien (Ya=2 atau Tidak=1), serta diagnosa kanker paru-paru pada pasien (Ya=2 atau Tidak=1).

Tabel 2. Atribut Kumpulan Data

No	Jenis Kelamin	Usia	Merokok	Jari Kuning	Kecemasan	Tekanan Teman Sebaya	...	Kanker
0	1	69	1	2	2	1	...	2
1	1	74	2	1	1	1	...	2
2	2	59	1	1	1	2	...	1
3	1	63	2	2	2	1	...	1
4	2	63	1	2	1	1	...	1
...
304	2	56	1	1	1	2	...	2
305	1	70	2	1	1	1	...	2
306	1	58	2	1	1	1	...	2
307	1	67	2	1	2	1	...	2
308	1	62	1	1	1	2	...	2

3.2. PreProcessing Data

Agar analisis tetap akurat, langkah ini mencakup penghapusan atau penggantian nilai yang kosong agar data tidak terpengaruh. Selain itu, perbaikan dilakukan pada data yang bermasalah seperti duplikasi atau inkonsistensi agar memastikan keakuratan dan konsistensi data yang digunakan. Agar semua fitur dalam dataset memiliki skala yang seragam. Membersihkan data adalah tindakan yang krusial untuk meningkatkan kualitas data serta memastikan bahwa analisis dan model yang dibuat berdasarkan data tersebut tepat.

Tabel 3. Correlation Matrix

Variabel 1	Variabel 2	Korelasi
Kesulitan Menelan	Jari Kuning	0.35
Tekanan	Jari Kuning	0.37
Sesak Nafas	Kelelahan	0.44
Kecemasan	Kesulitan Menelan	0.49
Kecemasan	Jari Kuning	0.57

Di Tabel 3 terlihat bahwa korelasi antara 0.35 hingga 0.37 dapat dikategorikan sebagai korelasi lemah di mana terdapat hubungan antar variabel namun tidak begitu signifikan. Korelasi dengan nilai antara 0.44 hingga 0.57 dapat dikategorikan sebagai korelasi moderat, menandakan hubungan yang cukup signifikan antar variabel meskipun tidak terlalu kuat.

3.3. Evaluasi

Berdasarkan analisis akurasi perbandingan antara algoritma SVM dan Regresi Logistik, Confusion Matrix mengalami kekacauan seiring dengan proses klasifikasi yang terlibat dalam menampilkan apakah data diprediksi dengan benar atau salah. Salah satu langkah yang bisa diambil untuk membentuk confusion matrix ialah dengan menghitung jumlah hasil perhitungan yang tepat dan yang tidak tepat. Setelah mendapatkan nilai matriks, kita dapat menghitung akurasi. Berikut merupakan hasilnya.

Tabel 4. Confusion Matrix SVM

Prediction		1	0
	1	4	3
	0	1	85

Di Tabel 4 memperlihatkan confusion matrix untuk algoritma SVM. Hasil prediksi model ditunjukkan dalam bentuk kelas 1 dan 0, dipisahkan dalam kolom masing-masing. Terdapat 85 data yang benar diprediksi sebagai kelas 0 True Negatif (TN), 4 data diprediksi sebagai kelas 1 True Positive (TP), 3 data salah diprediksi sebagai kelas 1 False Positif (FP), dan 1 data salah diprediksi sebagai kelas 0 False Negatif (FN).

Tabel 5. Confusion matrix Logistic Regression

Prediction		1	0
	1	5	2
	0	0	86

Di Tabel 5, confusion matrix dari algoritma logistic regression menunjukkan perbedaan. Terdapat 86 data yang benar diprediksi sebagai kelas 0 True Negative (TN), 5 data yang benar diprediksi sebagai kelas 1 True Positive (TP), 2 data yang salah diprediksi sebagai kelas 1 False Positive (FP), dan tidak ada data yang salah diprediksi sebagai kelas 0 False Negative (FN). Dari hasil Confusion Matrix, peneliti telah melakukan uji akurasi untuk kedua algoritma tersebut. Perbandingan hasil dari setiap algoritma akan dilakukan untuk mengevaluasi tingkat ketepatan. dapat melihat hasil uji akurasi model dalam tabel berikut.

Tabel 6. Hasil Perbandingan Tingkat Performa

Method	Accuracy(%)	Precision(%)	Recall(%)	F1Score(%)
SVM	95.70%	80%	57.14%	66.67%
Logistic Regression	97.85%	100%	71.42%	83.33%

Dari data yang tersaji di Tabel 6, dapat diamati bahwa Logistic Regression menunjukkan tingkat akurasi yang lebih tinggi daripada SVM. Logistic Regression mencapai akurasi sebesar 97.85%, sedangkan SVM hanya mencapai 95.70% dalam pengaturan dataset yang terdiri dari 309 data.

4. Kesimpulan

Hasil pengujian dan perbandingan dalam penelitian ini menunjukkan bahwa kedua metode, yaitu SVM dan Regresi Logistik, sama-sama menunjukkan performa yang baik dalam memprediksi kanker paru-paru dari gejala awal, dengan tingkat akurasi rata-rata mencapai 96.775%. Perbedaan dalam tingkat akurasi antara metode SVM dan Regresi Logistik adalah sebesar 2.15%, dengan akurasi Regresi Logistik lebih tinggi daripada metode SVM.

Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih yang sangat mendalam kepada semua pihak yang telah bekerja keras dalam proses terciptanya jurnal penelitian ini. Terutama kepada Ibu Rastri Prathivi, yang merupakan dosen pembimbing penulis, serta kepada semua anggota keluarga besar FTIK Prodi Teknik Informatika Universitas Semarang, tercinta Bapak dan Ibu, sahabat-sahabat, penulis juga ingin menyampaikan kesadaran bahwa keberhasilan penelitian ini tidak akan terjadi tanpa bantuan, arahan, dan masukan yang telah diberikan, yang mungkin tidak semuanya dapat disebutkan satu per satu. Semoga semua kebaikan dan pertolongan yang telah diberikan kepada penulis akan mendapat berkah berlipat dari Allah SWT. Dalam rangka penyempurnaan penelitian ini di masa yang akan datang, saya mengharapkan kritik dan saran yang membangun dari berbagai pihak.

Daftar Pustaka

- [1] R. D. Marzuq, S. A. Wicaksono, And N. Y. Setiawan, "Prediksi Kanker Paru-Paru Menggunakan Algoritme Random Forest Decision Tree," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 7, No. 7, Pp. 3448–3456, 2023.
- [2] F. Islami *Et Al.*, "Annual Report To The Nation On The Status Of Cancer, Part 1: National Cancer Statistics," *J. Natl. Cancer Inst.*, Vol. 113, No. 12, Pp. 1648–1669, 2021, Doi: 10.1093/Jnci/Djab131.
- [3] D. Septhya *Et Al.*, "Implementasi Algoritma Decision Tree Dan Support Vector Machine Untuk Klasifikasi Penyakit Kanker Paru," *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 3, No. 1, Pp. 15–19, 2023, Doi: 10.57152/Malcom.V3i1.591.
- [4] P. L. Paelongan And I. Palupi, "Lung Cancer Prediction Model Using Logistic Linear Regression With Imbalanced Dataset," *Ind. J. Comput.*, Vol. 7, No. 2, Pp. 1–14, 2022, Doi: 10.34818/Indojc.2022.7.2.616.
- [5] U. Khultsum, F. Sarasati, And G. Taufik, "Penerapan Metode Mobile-Net Untuk Klasifikasi Citra Penyakit Kanker Paru-Paru," *Jurikom (Jurnal Ris. Komputer)*, Vol. 9, No. 5, P. 1366, 2022, Doi: 10.30865/Jurikom.V9i5.4918.
- [6] R. . Putri, "Implemantasi Metode Convolutional Neural Network Dan Ekstraksi Glcm Pada Klasifikasi Kanker Paru," *Semin. Nas. Rekayasa 2022*, Vol. 23, No. 4,

- Pp. 70–79, 2022.
- [7] S. Firyal Nabila, D. Setiawan Hendyca Putra, S. Farlinda, E. Tri Ardianto, J. Kesehatan, And P. Negeri Jember, “J-Remi: Jurnal Rekam Medik Dan Informasi Kesehatan Analisis Faktor Risiko Pada Penyakit Karsinoma Paru (C34) Pasien Rawat Inap Di Rumah Sakit Baladhika Husada Jember,” Vol. 2, No. 2, Pp. 244–254, 2021.
- [8] S. Akbar, M. S. Qisam, And G. A. Yasmin, “Identifikasi Kanker Paru-Paru Menggunakan Metode Ekuialisasi Histogram Dan Lbp,” *J. Electron. Instrum.*, Vol. 1, No. 1, Pp. 1–6, 2023.
- [9] N. K. C. Pratiwi, N. Ibrahim, And S. Saidah, “Prediksi Kanker Paru Menggunakan Grid Search Untuk Optimasi Hyperparameter Pada Algoritma Mlp Dan Logistic Regression,” *Elkomika J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, Vol. 12, No. 3, P. 556, 2024, Doi: 10.26760/Elkomika.V12i3.556.
- [10] M. Annan, M. Mustofa, And H. N. Wahiid, “Penggunaan Algoritma Knn Dalam Deteksi Awal Kanker Paru-Paru Menggunakan Data Medis,” Vol. 8, Pp. 485–493, 2024.
- [11] A. Desiani *Et Al.*, “Perbandingan Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Support Vector Machine Dan K-Nearest Neighbor,” *J. Process.*, Vol. 18, No. 1, Pp. 54–62, 2023, Doi: 10.33998/Processor.2023.18.1.700.
- [12] M. Tiara *Et Al.*, “Pemanfaatan Algoritma Adasyn Dan Support Vector Machine Dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru,” Vol. 8, No. 5, Pp. 8773–8778, 2024.
- [13] D. Benaya, “Implementasi Random Forest Dalam Klasifikasi Kanker Paru-Paru,” *Jointer J. Informatics Eng.*, Vol. 5, No. 01, Pp. 27–31, 2024, Doi: 10.53682/Jointer.V5i01.331.
- [14] Jatnika Fahmi Idris, Rafid Ramadhani, And Muhammad Malik Mutoffar, “Klasifikasi Penyakit Kanker Paru Menggunakan Perbandingan Algoritma Machine Learning,” *J. Media Akad.*, Vol. 2, No. 2, 2024, Doi: 10.62281/V2i2.145.
- [15] N. R. Muntari And K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” *J. Ilmu Komput. Dan Teknol.*, Vol. 3, No. 1, Pp. 1–6, 2022, Doi: 10.35960/Ikomti.V3i1.766.
- [16] H. Irsyad And D. Mariana, “Klasifikasi Pneumonia Pada Chest X-Ray Paru-Paru Dengan Ekstraksi Fitur Local Binary Pattern Menggunakan Support Vector Machine,” *J. Ilm. Betrik*, Vol. 12, No. 1, Pp. 54–62, 2021, Doi: 10.36050/Betrik.V12i1.294.
- [17] M. I. Y. Helmi, D. Anggraeni, And A. F. Hadi, “Diagnosis Penderita Penyakit Kanker Paru Menggunakan Support Vector Machine Dan Naïve Bayes,” *Stat. J. Theor. Stat. Its Appl.*, Vol. 21, No. 1, Pp. 1–4, 2021, Doi: 10.29313/Jstat.V21i1.7566.
- [18] A. Asuntha And A. Srinivasan, “Deep Learning For Lung Cancer Detection And Classification,” *Multimed. Tools Appl.*, Vol. 79, No. 11, Pp. 7731–7762, 2020, Doi: 10.1007/S11042-019-08394-3.
- [19] I. A. Ricky, I. F. Hanif, F. N. Hasan, E. S. Sinduningrum, Z. Halim, And N. Nunik, “Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression,” *J. Linguist. Komputasional*, Vol. 5, No. 2, P. 77, 2022, [Online]. Available: <https://T.Co/23c4krbjp>
- [20] Ginanjar Abdurrahman And H. Oktavianto, “Klasifikasi Teks Mining Untuk Deteksi Kanker Menggunakan Support Vector Machine,” *Justify J. Sist. Inf. Ibrahimy*, Vol. 3, No. 1, Pp. 16–20, 2024, Doi: 10.35316/Justify.V3i1.5028.
- [21] A. S. Rahayu, A. Fauzi, And R. Rahmat, “Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (Svm) Pada Analisis Sentimen Spotify,” *J. Sist. Komput. Dan Inform.*, Vol. 4, No. 2, P. 349, 2022, Doi: 10.30865/Json.V4i2.5398.