

# PERBANDINGAN KINERJA SVM DAN REGRESI LOGISTIK UNTUK PREDIKSI KANKER PAYUDARA

Galih Rakasiwi<sup>1\*</sup>, Fuad Nur Hasan<sup>2</sup>

Program Studi Informatika, Fakultas Teknik dan Informatika

Universitas Bina Sarana Informatika

1,2 Universitas Bina Sarana Informatika Jakarta

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus

Ibukota Jakarta 10450, Indonesia

[galihfascalkenan@gmail.com](mailto:galihfascalkenan@gmail.com)<sup>1\*</sup> [fuad.fnu@bsi.ac.id](mailto:fuad.fnu@bsi.ac.id)<sup>2</sup>

## Abstrak

Kanker payudara adalah jenis kanker umum pada wanita, dan deteksi dini penting untuk meningkatkan kesembuhan. Penelitian ini membandingkan dua algoritma pembelajaran mesin, Support Vector Machine (SVM) dan Regresi Logistik, dalam mendeteksi kanker payudara menggunakan dataset dari UCI. Data dibagi 80:20 untuk pelatihan dan pengujian, lalu diskalakan. Model SVM dan Regresi Logistik dilatih, dievaluasi menggunakan akurasi, presisi, recall, dan confusion matrix, serta diuji dengan K-Fold Cross Validation 10 fold. Hasil menunjukkan SVM mencapai akurasi 0.9561, presisi 0.9714, recall 0.9577, sementara Regresi Logistik mencapai akurasi 0.9737, presisi 0.9722, recall 0.9859. Namun, Cross Validation menunjukkan rata-rata akurasi SVM (0.9577) lebih tinggi dari Regresi Logistik (0.9490), menandakan performa SVM lebih konsisten. Kesimpulannya, kedua model berkinerja baik, dengan SVM sedikit unggul dalam konsistensi. Kedua model menunjukkan kinerja sangat baik dan dapat menjadi alat efektif dalam deteksi dini kanker payudara.

**Kata Kunci :** Regresi logistik, Support Vector Machine, akurasi, scikit-learn, K fold Cross Validation, Kanker Payudara

## Abstract

*Journal of Research and Application of Informatics Students (JRAMI) is a publication media for journals or scientific articles resulting from the Research and Final Assignments of college students in the Informatics Study Program, Faculty of Engineering and Computer Science. Publications are published online every 3 months, February, May, August and November. JRAMI writing templates must be fulfilled by writers who will send research journals to be published in JRAMI. All manuscripts are typed with Single space in A4, One Columns, top / bottom / right margins = 2.5 cm and Left = 3 cm, and maximal 8-10 pages. The font used is Times New Roman, Justify. The title is composed of no more than 12 syllables, 14pt, bold. Author's name, institution, address and email address of the size of 11pt. The abstract consists of 100 - 200 syllables, 10pt. The contents of the article (Introduction s. Bibliography) is written with Font 11 pt. The contents of the abstract are a summary of the background of the problem, the formulation of the problem, the purpose of the study, the methodology used and the results of the study.*

**Keyword :** Regresi logistik, Support Vector Machine, akurasi, scikit-learn, K fold Cross Validation, Kanker Payudara

## PENDAHULUAN

Kanker adalah penyakit yang bisa mempengaruhi hampir seluruh organ dalam tubuh manusia. Hasilnya, ia mampu membuka jaringan biologis yang berdekatan dan terhubung ke jaringan lain melalui darah aliran. (Kusuma et al., 2022) Kanker payudara adalah jenis kanker yang paling umum terjadi pada wanita di Indonesia, kanker payudara dapat terjadi karena sejumlah sel tumbuh dan berkembang tidak terkendali atau tidak normal. (Ravly Andryan et al., 2022) Ada dua jenis kanker, yaitu ganas dan jinak. Kanker ganas cenderung menyebar dan dapat merusak jaringan di sekitarnya, sedangkan kanker jinak tidak menyebar dan tidak mengganggu sel-sel di sekitarnya. Meskipun tidak ada gejala spesifik yang bisa digunakan untuk mendiagnosis kanker payudara, kanker ini dapat diprediksi dengan mengamati tanda-tanda yang muncul pada payudara. (Febrian, 2020) Dari kasus-kasus tersebut, dapat ditarik kesimpulan bahwa kanker payudara merupakan salah satu penyebab utama kematian pada

wanita. Kanker ini terjadi karena sel-sel dalam jaringan payudara berkembang secara tidak terkendali, membentuk benjolan di dalam payudara. (Judu, 2023)

*International Agency for Research on Cancer (IARC)*, badan kesehatan dunia, memproyeksikan bahwa lebih dari 400.000 wanita meninggal setiap tahun karena kanker payudara. Banyak orang masih beranggapan bahwa tumor sama dengan kanker, padahal sebenarnya tumor tidak selalu bersifat kanker. (Chazar & Widhiaputra, 2020) Seiring berjalannya waktu dan berkembangnya teknologi, sistem dengan bantuan computer dapat membantu para ahli di bidang medis. Dan dengan bantuan sistem computer ini dapat membantu para ahli patologi dalam melakukan pemeriksaan dengan lebih konsisten dan efisien. (Erwandi, 2020)

Berdasarkan paparan diatas, dalam penelitian ini peneliti melakukan akurasi, *presisi*, *recall*, prediksi kanker payudara menggunakan *support vector machine (SVM)* dan Regresi Logistik, hasil dari analisis akan lebih diketahui Tingkat akurasi SVM dan regresi Logistik, dimana hasilnya adalah lebih akurat menggunakan metode Regresi Logistik. Selain itu untuk model SVM dan Regresi Logistik di uji lagi dengan metode *K Fold Cross Validation*. Dimana hasilnya terdapat perbedaan dengan evaluasi akurasi. Hasilnya model SVM lebih baik dalam memprediksi kanker payudara, berbeda karna jika menggunakan *K Fold Cross Validation* terdapat 10x pengulangan, dan di ambil rata ratanya di setiap model. Dari penelitian tersebut menunjukan bahwa SVM lebih baik dalam memprediksi kanker payudara.

## **PENELITIAN RELEVAN**

Pada penelitian ini, metode Support Vector Machine diterapkan pada Breast Cancer Coimbra Data Set. Penerapan Backward Elimination bertujuan untuk mengoptimalkan performa suatu model dengan sistem kerja pemilihan mundur dan memilih atribut yang paling relevan pada proses klasifikasi. Hasil penelitian klasifikasi pasien kanker payudara menggunakan metode Support Vector Machine menghasilkan nilai akurasi sebesar 65,22% dan nilai AUC sebesar 0,700 yang termasuk ke dalam kategori Fair Classification. Sedangkan hasil penelitian klasifikasi pasien kanker payudara menggunakan metode Support Vector Machine dengan Backward Elimination menghasilkan nilai akurasi sebesar 95,65% dan nilai AUC sebesar 1,000 yang termasuk ke dalam kategori Excellent Classification. (Resmiati & Arifin, 2021)

Teknik Machine Learning dapat dilakukan untuk memprediksi kanker payudara karena dapat menangkap interaksi tingkat tinggi antar data yang mungkin menghasilkan prediksi yang lebih baik untuk membedakan antara jinak dan ganas. Oleh karena itu, penelitian ini digunakan pendekatan klasifikasi yang efektif Support Vector Machine (SVM). SVM adalah model yang digunakan untuk melakukan prediksi kanker payudara. Untuk mempermudah proses prediksi, hasil prediksi kanker payudara diterapkan dalam bentuk web base dengan framework Django untuk dapat digunakan oleh dokter dalam menentukan keputusan dengan cepat. Hasil modeling menunjukkan bahwa prediksi kanker payudara menggunakan SVM memperoleh akurasi sebesar 98,24%. Sehingga model yang dibuat sangat baik sehingga aplikasi akan bekerja untuk memprediksi dengan baik. (Tri et al., 2023)

Kanker payudara merupakan penyebab kematian kedua tertinggi pada perempuan di seluruh dunia. Untuk mengatasi masalah yang terus berkembang ini, penelitian ini dirancang sebuah model yang dapat memprediksi kanker payudara dengan memanfaatkan dataset kemudian diolah menggunakan metode Prediksi Regresi Logistik. Metode ini tepat untuk memprediksi data yang digunakan karena kemampuannya dalam menangani variabel dependen yang bersifat kategorikal serta memberikan output dalam bentuk probabilitas. Penelitian ini menggunakan dataset sebanyak 306 sample dengan 4 atribut. Data yang digunakan Langkah penelitian meliputi pengumpulan data, preprocessing, pemodelan dengan regresi logistic dan evaluasi hasil menggunakan matrix seperti confusion matrix, MAE, MSE, dan R-Square. Hasil menunjukkan akurasi prediksi sebesar 86%, dengan nilai MSE 0,137 dan R-Square 0,309. Penelitian ini menunjukkan efektivitas regresi logistic dalam memprediksi kelangsungan hidup pasien pasca operasi kanker payudara. Meskipun demikian dengan menerapkan algoritma yang berbeda

penelitian ini dapat memilih set atribut terbaik yang signifikan guna meningkatkan nilai akursi prediksi pada pasien pasca operasi kanker payudara. (Sujana et al., 2024)

## METODE PENELITIAN

Masing masing langkah penelitian pada gambar diatas dapat diuraikan sebagai berikut:

### 1. *Dataset*

Dataset pada penelitian ini menggunakan data sintetis dari *scikit-learn*, data yang dibuat simulasi dan algoritma komputasi berdasarkan teknologi kecerdasan buatan *generative*.

### 2. *Preprocessing data*

Preprocessing data pada kode ini dimulai dengan memuat dataset kanker payudara menggunakan `sklearn.datasets.load_breast_cancer()`, di mana fitur-fitur (X) digunakan untuk memprediksi apakah tumor ganas (1) atau jinak (0). Data kemudian dibagi menjadi data latih dan uji menggunakan `train_test_split`, dengan 80% untuk pelatihan dan 20% untuk pengujian. Fitur-fitur diskalakan menggunakan `StandardScaler` agar setiap fitur memiliki mean 0 dan standar deviasi 1, sehingga mencegah perbedaan skala antar fitur memengaruhi model. Selanjutnya, metode K-Fold Cross Validation (10 kali) digunakan untuk mengevaluasi model, memastikan performa yang lebih akurat dan representatif. Preprocessing ini penting untuk memastikan data siap dan model dapat bekerja optimal.

### 3. *Pembagian data*

Metode yang digunakan untuk membagi data adalah "*split-test train split*", di mana data dibagi menjadi dua data yaitu, data uji dan data latih. Penelitian ini, pembagian data dilakukan dengan cara perbandingan 80% untuk data latih, yaitu 455 data, dan 20% untuk data uji, yaitu 114 data.

### 4. *Pelatihan Model*

Dalam penelitian ini, ada dua model yang digunakan untuk memprediksi kanker payudara yaitu SVM dan Regresi Logistik. Kedua model ini dipilih karena efektivitasnya dalam memprediksi kanker payudara. Data dibagi menjadi data *training* dan data *testing*, dan model dilatih menggunakan data *testing* dengan bantuan tools *Spyder*. Kedua model tersebut digunakan untuk memprediksi kanker payudara ganas (*malignant*) dan kanker payudara jinak (*benign*).

### 5. *Evaluasi Model*

Dalam penelitian ini, model SVM (*Support Vector Machine*) dan Regresi Logistik digunakan untuk melakukan klasifikasi pada dataset kanker payudara. Metode yang dipakai untuk memprediksi dan mengevaluasi model dilakukan dengan menghitung *akurasi*, *presisi*, *recall*, prediksi, dan confusion matrix untuk masing-masing model. Hasil evaluasi ini, yang dilakukan menggunakan tools *Spyder*, membantu dalam menentukan metode yang lebih baik dalam memprediksi kanker payudara dan memberikan informasi yang penting dalam pengambilan keputusan klinis.

### 6. *Visualisasi*

Visualisasi data merupakan langkah penting dalam penelitian ini untuk memberikan gambaran yang jelas tentang pola, distribusi, dan hubungan antar variabel dalam dataset kanker payudara yang digunakan.

### 7. *Validasi model dengan K-fold*

*Cross-validasi* adalah teknik evaluasi model yang digunakan untuk mengukur seberapa baik hasil analisis statistik dapat diterapkan pada data yang tidak digunakan dalam pelatihan model tersebut. "Dengan validasi silang, kita dapat mengoptimalkan kompleksitas model agar kesalahan klasifikasi atau aproksimasi pada kumpulan data uji yang tidak diketahui dapat diminimalkan." (Santoso et al., 2021)

## HASIL DAN PEMBAHASAN

### *Pembahasan Algoritma*

penelitian ini menggunakan beberapa matrik yaitu akurasi, presisi dan recall, untuk memprediksi kanker payudara yang dihitung dengan pemrosesan data:

*confusion matrix* memiliki 4 sel utama yaitu:

- True Positive (TP)*: jumlah data positif yang benar benar diprediksi positif
- False Positive (FP)*: jumlah data negatif yang salah diprediksi positif

- c. *True Negative* (TN): jumlah data negatif yang benar benar diprediksi negatif
- d. *False Negative* (FN): jumlah data positif yang salah diprediksi negatif

Disini peneliti akan mengidentifikasi nilai nilai *confusion matrix* model SVM sebagai berikut:

- a. *True Positive* (TP): memiliki nilai sebesar: 68
- b. *False Positive* (FP): memiliki nilai sebesar: 2
- c. *True Negative* (TN): memiliki nilai sebesar: 41
- d. *False Negative* (FN): memiliki nilai sebesar: 3

Disini peneliti akan mengidentifikasi nilai nilai *confusion matrix* model Regresi logistik sebagai berikut:

- a. *True Positive* (TP): memiliki nilai sebesar: 70
- b. *False Positive* (FP): memiliki nilai sebesar: 2
- c. *True Negative* (TN): memiliki nilai sebesar: 41
- d. *False Negative* (FN): memiliki nilai sebesar: 1

Berikut ini adalah perhitungan dari beberapa matrix:

- a. Akurasi

Berikut adalah perhitungan matrix akurasi model SVM:

$$Akurasi = \frac{68 + 41}{68 + 2 + 41 + 3}$$

$$Akurasi = \frac{109}{114}$$

Hasilnya:

$$Akurasi = 0,9561$$

Perhitungan matrix akurasi model regresi logistik:

$$Akurasi = \frac{70 + 41}{70 + 2 + 41 + 1}$$

$$Akurasi = \frac{111}{114}$$

Hasilnya:

$$Akurasi = 0,9736$$

- b. *Precision*

Berikut adalah perhitungan *matrix precision* model SVM:

$$Precision = \frac{68}{68 + 2}$$

$$Precision = \frac{68}{70}$$

Hasilnya:

$$Precision = 0,9714$$

perhitungan *matrix precision* model regresi logistik:

$$Precision = \frac{70}{70 + 2}$$

$$Precision = \frac{70}{72}$$

Hasilnya

$$Precision = 0,9722$$

- c. *Recall*

Berikut adalah perhitungan *matrix recall* model SVM:

$$Recall = \frac{68}{68 + 3}$$

$$Recall = \frac{68}{71}$$

Hasilnya:

$$Recall = 0,9577$$

Perhitungan *matrix recall* model regresi logistik:

$$Recall = \frac{70}{70 + 1}$$

$$Recall = \frac{70}{71}$$

$$Recall = 0,9859$$

Hasilnya:

**Validasi model dengan K-Fold**

Dalam penelitian ini memakai *K fold Cross Validation* yang digunakan untuk menguji suatu model. Setiap model menggunakan 10 *fold* untuk mengetahui hasil akurasi rata rata *K fold cross validation* peneliti menggunakan tools *spyder*. Berikut table yang berisi 10 *fold* setiap model:

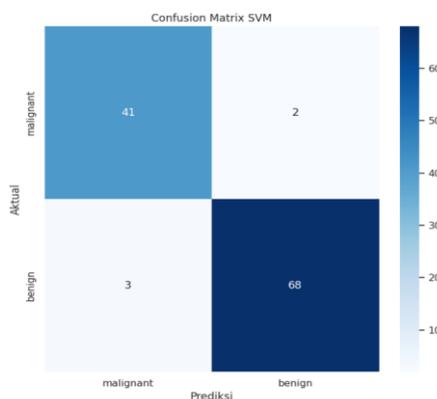
Tabel 1. 10 Fold Validation

Regresi Logistik	SVM
0.98245614	0.98245614
0.92982456	0.96491228
0.98245614	0.98245614
0.94736842	0.96491228
0.9122807	0.96491228
0.94736842	0.94736842
0.98245614	0.98245614
0.96491228	0.94736842
0.92982456	0.94736842
0.91071429	0.89285714
Akurasi rata rata : 0.9489	Akurasi rata rata : 0.9577

Sumber: (Penelitian, 2024)

**Visualisasi**

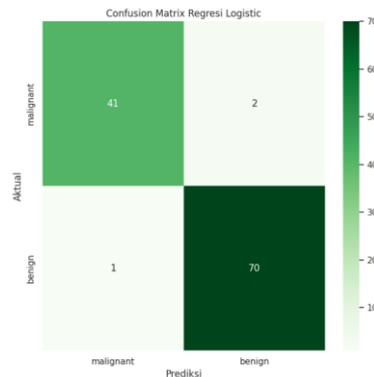
Berikut adalah gambar confusion matrix-nya yang dihasilkan dari pemrosesan data menggunakan tools *Spyder*:



Gambar 5. Confusion matrix model SVM

Sumber: (Penelitian, 2024)

Heatmap pertama menunjukkan confusion matrix untuk model SVM. Di sini, kita dapat melihat jumlah prediksi yang benar dan salah untuk masing-masing kelas (positif dan negatif). Warna biru yang lebih gelap menunjukkan angka yang lebih besar.



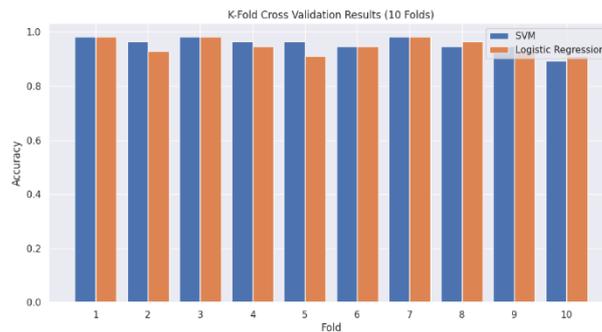
Gambar 6 Confusion matrix model regresi logistik

Sumber: (Penelitian, 2024)

Heatmap kedua menunjukkan Confusion matriks untuk model **Regresi Logistik**. Seperti halnya SVM, ini juga menunjukkan kategori yang sama, namun perbedaannya terletak pada hasil akurasi dari Regresi Logistik.

**Visualisasi K Fold Cross Validation**

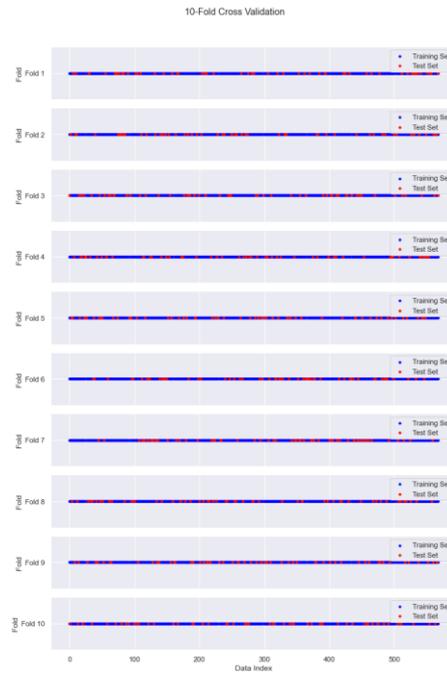
Dalam penelitian ini, terdapat diagram batang dan scatter plot. Diagram batang ini digunakan untuk membandingkan akurasi masing-masing fold antara model SVM dan Logistic Regression, memudahkan identifikasi model mana yang lebih konsisten. Pemrosesan data ini dilakukan menggunakan tools Spyder.



Sumber: (Penelitian, 2024)

Gambar 7. Diagram batang K fold Validation

*Scatter plot* untuk Menunjukkan distribusi data latih dan uji pada setiap *fold*, memastikan bahwa pembagian data dilakukan secara merata dan acak untuk setiap *fold*. Pemrosesan data ini menggunakan tools *spyder*



Sumber: (Penelitian, 2024)

Gambar 8.Scatter plot K fold cross validation

## SIMPULAN

Kanker payudara adalah jenis kanker umum di antara wanita, dan deteksi dini sangat penting untuk meningkatkan peluang kesembuhan dan mengurangi angka kematian. Penggunaan pembelajaran mesin dalam deteksi kanker payudara menunjukkan potensi signifikan dalam membantu profesional medis. Penelitian ini bertujuan untuk memberikan informasi berharga kepada profesional medis, peneliti, dan pengembang sistem kesehatan terkait dengan kinerja dua algoritma pembelajaran mesin, yaitu Support Vector Machine (SVM) dan Regresi Logistik, dalam mendeteksi kanker payudara. Dataset kanker payudara dari UCI Machine Learning Repository dibagi menjadi data latih dan uji dengan rasio 80:20 dan diskalakan menggunakan StandardScaler. Dua model, SVM dengan kernel linear dan Regresi Logistik, dilatih dan diuji. Evaluasi menggunakan metrik akurasi, presisi, recall, dan confusion matrix. K-Fold Cross Validation dengan 10 fold digunakan untuk estimasi kinerja yang lebih akurat.

Berdasarkan hasil penelitian ini, model SVM mencapai akurasi 0.9561, presisi 0.9714, dan recall 0.9577, sementara model Regresi Logistik mencapai akurasi 0.9737, presisi 0.9722, dan recall 0.9859, yang menunjukkan performa lebih baik. Namun, hasil 10-fold Cross Validation mengindikasikan bahwa akurasi rata-rata SVM adalah 0.9577, sedangkan Regresi Logistik adalah 0.9490, menunjukkan performa SVM yang lebih konsisten. Berdasarkan hasil evaluasi dan analisis, model Support Vector Machine (SVM) menunjukkan sedikit keunggulan dalam hal konsistensi dibandingkan dengan Regresi Logistik dalam mendeteksi kanker payudara. Kedua model ini telah menunjukkan kinerja yang sangat baik dan dapat dianggap sebagai alat yang efektif dalam mendukung deteksi dini kanker payudara, yang pada akhirnya dapat memberikan kontribusi signifikan dalam pengambilan keputusan klinis dan meningkatkan peluang kesembuhan pasien.

**DAFTAR PUSTAKA**

- Chazar, C., & Widhiaputra, B. E. (2020). *INFORMASI (Jurnal Informatika dan Sistem Informasi) Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine*.
- Erwandi, R. (2020). *Klasifikasi Kanker Payudara Menggunakan Residual Neural Network*. <https://doi.org/10.21108/indojc.2020.5.1.373>
- Febrian, F. (2020). Perbandingan Teknik Klasifikasi Neural Network, Support Vector Machine, dan Naive Bayes dalam Mendeteksi Kanker Payudara. *BINA INSANI ICT JOURNAL*, 7(1), 53–62.
- Judu, H. (2023). *KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN METODE RANDOM FOREST*.
- KUSUMA, J., HAYADI, B. H., WANAYUMINI, W., & ROSNELLY, R. (2022). Komparasi Metode Multi Layer Perceptron (MLP) dan Support Vector Machine (SVM) untuk Klasifikasi Kanker Payudara. *MIND Journal*, 7(1), 51–60. <https://doi.org/10.26760/mindjournal.v7i1.51-60>
- Ravly Andryan, M., Fajri, M., Nina Sulistyowati, dan, Informatika, T., Singaperbangsa Karawang Jl Ronggo Waluyo, U. H., Timur, T., & Barat, J. (2022). KOMPARASI KINERJA ALGORITMA XGBOOST DAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA PENYAKIT KANKER PAYUDARA. *Jurnal Informatika Dan Komputer*, 6(1), 1–5.
- Resmiati, R., & Arifin, T. (2021). *SISTEMASI: Jurnal Sistem Informasi Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination*. <http://sistemasi.ftik.unisi.ac.id>
- Santoso, J. T., Kom, S., Kom, M., Buatan, K., & Syaraf Buatan, J. (2021). *Kecerdasan Buatan & Jaringan Syaraf Buatan*.
- Sujana, S., Ratna Juwita, A., Faisal, S., Karawang, P., Jl HSRonggo Waluyo, K., Timur, T., & Barat, J. (2024). Penerapan Metode Regresi Logistik Untuk Memprediksi Peristiwa Biner Pasien Pasca Operasi Kanker Payudara. *Journal of Information System Research (JOSH)*, 5(4), 1095–1101. <https://doi.org/10.47065/josh.v5i4.5639>
- Tri, N., Adiningrum, R., Rianti, R., & Prianto, C. (2023). RANCANG BANGUN APLIKASI PREDIKSI KANKER PAYUDARA DENGAN PENDEKATAN MACHINE LEARNING. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3), 2830–7062. <https://doi.org/10.23960/jitet.v11i3%20s1.3351>

---

**BIOGRAFI PENULIS**

<p>Galih Rakasiwi (15200154)</p>	<p>Nama : Galih Rakasiwi Nim : 15200154 Prodi : Informatika Fakultas : Teknik dan Informatika Institusi : Universitas Bina Sarana Informatika Email : galihfascalkenan@gmail.com</p>
<p>Fuad Nur Hasan S.Kom, M.Kom sebagai dosen pembimbing</p>	<p>NIDN : 0309129201 Nama : Fuad Nur Hasan, S.Kom., M.Kom Fakultas : Teknik dan informatika Program studi : Informatika (S1) Institusi : Universitas Bina Sarana Informatika Email : fuad.fnu@bsi.ac.id</p>

BUKTI SUBMIT



11 dari 1.036 < > ✎

[JRAMI] Submission Acknowledgement Kotak Masuk x



**Alusyanti Primawati** <pengembangjurnalunindra@gmail.com>  
kepada saya ▾

1 Okt 2024, 20.47 ☆ 😊 ↶ ⋮

[Terjemahkan ke Indonesia](#)

Galih Rakasiwi:

Thank you for submitting the manuscript, "PERBANDINGAN KINERJA SVM DAN REGRESI LOGISTIK UNTUK PREDIKSI KANKER PAYUDARA" to Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI). With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Manuscript URL:  
<https://jim.unindra.ac.id/index.php/jrami/author/submission/13219>  
Username: galihrakasiwi

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Alusyanti Primawati  
Pengembang Jurnal Unindra (JRAMI)