



## Klasifikasi Sel Tumor Payudara Menggunakan Algoritma *Support Vector Machine (SVM)*

Puspa Rismawati<sup>1</sup>, Adrianus Inu Natalisanto<sup>1,\*</sup>, Devina Rayzy Perwitasari Sutaji Putri<sup>1</sup>

1) Prodi Fisika, Fakultas MIPA, Universitas Mulawarman  
Jl. Gunung Kelua No. 4, Samarinda, Kalimantan Timur, Indonesia

\*E-mail korespondensi: [adrianus.inu@gmail.com](mailto:adrianus.inu@gmail.com)

### Article Info:

Received: 08-02-2023

Revised: 16-02-2023

Accepted: 05-06-2024

### Keywords:

Breast Cancer, Machine Learning, Support Vector Machine (SVM)

### Abstract

Cancer is a large group of diseases that can begin in almost any organ or tissue of the body when abnormal cells grow uncontrollably, beyond their normal limits to invade adjacent parts of the body and/or spread to other organs. There is a lot of information about breast cancer that can be accessed easily. Information about breast cancer can be processed with machine learning. Machine learning can discover new meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technology and statistical and mathematical techniques. The purpose of this research is to determine the value of the accuracy of the SVM model on training data and testing data; and to determine the precision value of the SVM model on training data and testing data. Wisconsin Breast Cancer (WBC) data available in the UCI Machine Learning Repository. The data have been processed using the Python programming language with a support vector machine (SVM) modeling algorithm. The results of this research indicate that the value of accuracy in training data was equal to 96,18%, the value of accuracy in testing data was equal to 97,07%, and the value of precision in the SVM model algorithm was obtained as large as 98,63% for training data and as large as 98,14% for data testing.



## PENDAHULUAN

Kanker payudara terbentuk ketika sel-sel di dalam payudara tumbuh secara abnormal dan tidak terkendali. Sel tersebut umumnya membentuk tumor [1]. Kanker payudara termasuk salah satu penyakit tidak menular yang cenderung terus meningkat setiap tahunnya. Data statistik *American Cancer Society* yang menunjukkan bahwa kanker payudara menempati urutan pertama dengan jumlah kasus terbanyak dari seluruh jenis kasus tumor ganas (kanker) di seluruh dunia [3]. Menurut WHO, ditinjau dari angka kejadian dan kematian per tahun 2020 kanker payudara menempati posisi pertama penyebab kematian di Indonesia.

Banyaknya jumlah kasus kanker payudara yang dijumpai saat ini, serta semakin mudahnya akses untuk mencari informasi-informasi terkait dengan kanker payudara tersebut dan hangatnya pembicaraan mengenai kecerdasan buatan atau yang disebut juga dengan *artificial intelligence (AI)*. AI nantinya dapat dijadikan pembelajaran untuk pengelompokan data kanker payudara berdasarkan beberapa parameter dengan mengkombinasikan beberapa teknik *machine learning (ML)*. Penelitian mengenai *machine learning* untuk kasus diagnosis maupun klasifikasi penyakit kanker payudara sebelumnya dilakukan oleh Prahartiwi & Dari (2021) serta

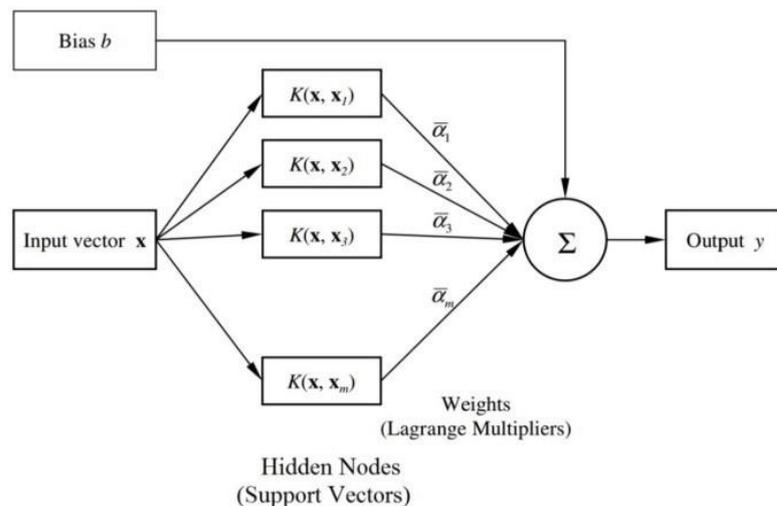
Chazar, C., & Erawan, B. (2020). Dari penelitian itu diperoleh nilai akurasi dari beberapa algoritma, yaitu *support vector machine*, *naive bayes* dan *decision tree*. Sementara penelitian ini bertujuan untuk mengetahui nilai akurasi dan presisi dari model SVM dengan pembagian data 50% untuk *training data* dan 50% untuk *testing data* serta memfokuskan pada satu algoritma saja. Data yang diolah ialah data *Wisconsin Breast Cancer (WBC)* yang tersedia di *UCI Machine Learning Repository*.

## TINJAUAN PUSTAKA

Tumor adalah benjolan tidak normal akibat pertumbuhan sel yang terjadi secara terus menerus. Tumor terbagi atas dua jenis, yaitu tumor jinak dan tumor ganas yang disebut dengan kanker [4]. Sel kanker dapat menyebar ke seluruh tubuh sehingga penyakit ini dapat mematikan. Penyakit kanker payudara atau *breast cancer* adalah penyakit yang ditandai dengan kelainan siklus sel khas yang menimbulkan kemampuan sel ini dapat menyebabkan pembelahan sel melebihi batas normal. Akibatnya, pembelahan sel yang terjadi menyerang jaringan biologis di dekatnya, dan dapat bermigrasi ke jaringan tubuh melalui sirkulasi darah [8].

SVM adalah pendekatan terbaru dari metode ML yang diterapkan di tingkat prediksi/prognosis kanker. Awalnya, SVM memetakan vektor *input* ke dalam ruang fitur dengan dimensi yang lebih tinggi dan mengidentifikasi *hyperplane* yang memisahkan titik data menjadi dua kelas. Jarak marjinal antara *hyperplane* keputusan dan contoh yang paling dekat dengan batas dimaksimalkan. Pengklasifikasian yang dihasilkan mencapai cukup banyak generalisasi dan karena itu dapat digunakan untuk klasifikasi sampel baru yang andal [5]. Persamaan model regresi SVM sebagai berikut [1] dan Gambar 1 merupakan diagram skema SVM [9].

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



**Gambar 1. Diagram Skema SVM**

Pengujian yang dilakukan adalah perhitungan persentase tingkat akurasi untuk memilih prognosis yang dianggap paling sesuai [2]. *Confusion matrix* merupakan tabel matriks yang terdiri dari dua kelas, yaitu kelas yang dianggap sebagai positif dan kelas yang dianggap sebagai negatif. *Matrix* ini berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada model klasifikasi.

*Confusion matrix* memberikan penilaian kinerja model klasifikasi berdasarkan jumlah objek yang diprediksi dengan benar dan salah [7].

**Tabel 1. Contoh Model Confusion Matrix [7]**

Classification		Predicted Class	
		Class = YES	Class = NO
Observed Class	Class = YES	A (true positive-TP)	B (false negative-FN)
	Class = NO	C (false positive-FP)	D (true negative-TN)

dengan keterangan:

*True Positive* (TP): Proporsi *positive* dalam *dataset* yang diklasifikasikan *positive*.

*True Negative* (TN): Proporsi *negative* dalam *dataset* yang diklasifikasikan *negative*.

*False Positive* (FP): Proporsi *positive* dalam *dataset* yang diklasifikasikan *negative*.

*False Negative* (FN): Proporsi *negative* dalam *dataset* yang diklasifikasikan *positive*

Setelah *testing data* dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung nilai akurasi, presisi, sensitivitas, spesififikasi serta nilai errornya seperti Tabel 2 [6].

**Tabel 2. Analisis data Confusion Matrix [6, 10]**

No	Ukuran	Rumus
1	Akurasi	$\frac{TP + TN}{TP + FN + TN + FP}$
2	Presisi	$\frac{TP + FP}{TP}$
3	Sensitivitas	$\frac{TP + FN}{TP}$
4	Spesififikasi	$\frac{FP + TN}{TN}$
5	Error	$\frac{FP + FN}{TP + FN + TN + FP}$

## METODE PENELITIAN

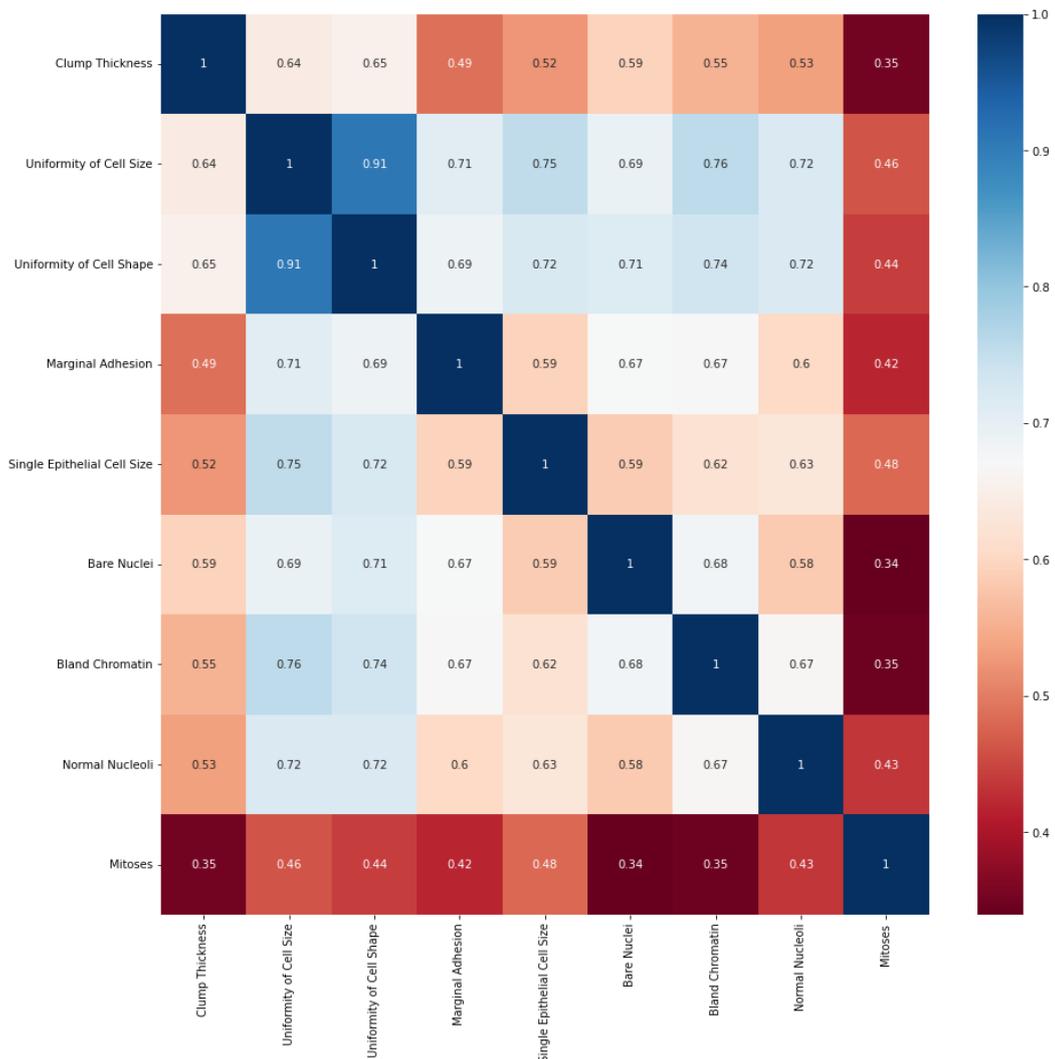
Penelitian ini menggunakan data sekunder yang diunduh pada *website database* kanker payudara dari *University of Wisconsin Hospitals*, Madison dari Dr. William H. Wolberg. Informasi mengenai *dataset Breast Cancer Wisconsin*. *Breast Cancer Wisconsin* memiliki 699 kasus (*benign*: 458 dan *malignant*: 241), 2 kelas (65,5% ganas dan 34.5% jinak), dan 11 parameter bernilai *integer* dan peralatan berupa laptop dan juga perangkat lunak *Jupyter Notebook App* digunakan untuk mengoperasikan ML dengan bahasa pemrograman Python dalam menentukan nilai akurasi dan presisi algoritma model SVM.

Dilakukan pengunduhan *dataset* kanker payudara pada laman *website database*, kemudian dilakukan penginputan data pada perangkat lunak *Jupyter Notebook App*. Dijalankan model algoritma SVM dengan membagi data menjadi dua bagian, yaitu *training data* dan *testing data*. Kemudian *training data* melatih pola atau hubungan antara parameter dan status sel kanker. Dilakukan hal yang sama terhadap *testing data*, jika hasil yang diperoleh dari *testing data* ini sesuai dengan yang diajarkan atau dilatih pada *training data*, maka akan dihasilkan *output*. Namun jika belum sesuai, maka model akan kembali mengulang proses pelatihan pada *training data*, dan dilakukan evaluasi menggunakan *confusion matrix*. Penelitian ini bertempat di

Laboratorium Fisika Komputasi dan Pemodelan, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda.

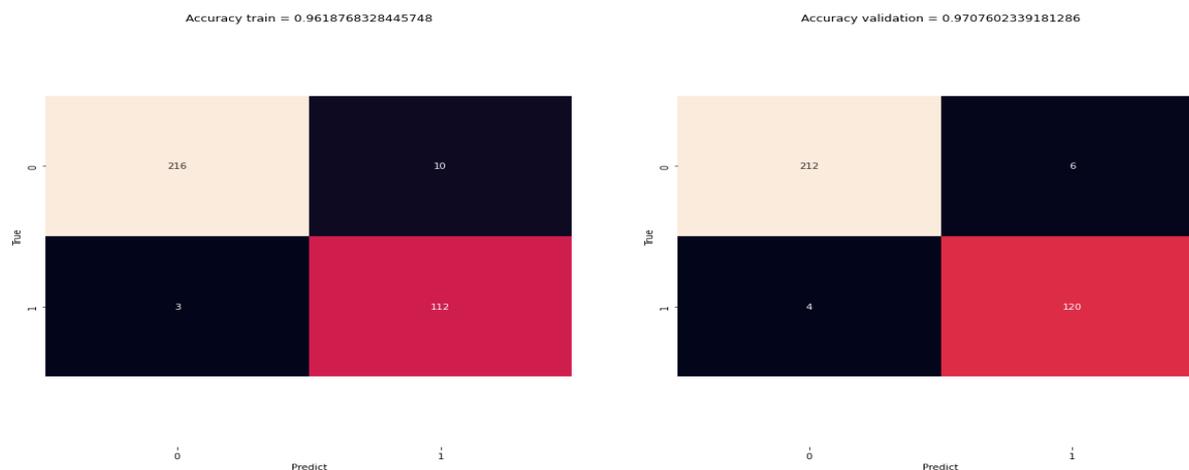
### HASIL DAN PEMBAHASAN

Pada Gambar 2 menunjukkan korelasi antara parameter pada *Dataset Breast Cancer Wisconsin* yang direpresentasikan berupa warna dan angka. Angka 1 dengan warna biru menunjukkan korelasi yang baik antara parameter pada *dataset*, sementara untuk korelasi yang buruk akan ditunjukkan dengan angka -1 dengan warna merah. Namun, dari hasil yang diperoleh korelasi yang terburuk pun nilainya tidak sampai -1. Korelasi antara parameter ini dikatakan baik ketika perubahan nilai parameter pada sumbu x berbanding lurus dengan perubahan nilai parameter pada sumbu y. Kemudian, korelasi antara parameter dikatakan buruk ketika perubahan nilai parameter pada sumbu x berbanding terbalik dengan perubahan nilai parameter pada sumbu y.



Gambar 2. Heatmap parameter *Dataset Breast Cancer Wisconsin*

Gambar 3 merupakan hasil dari evaluasi algoritma model SVM. *Matrix* bagian kiri ialah *matrix* hasil evaluasi *training data* dan *matrix* bagian kanan ialah hasil evaluasi *testing data*. Sumbu x adalah *class benign* dan *class malignant* untuk prediksi model SVM sementara sumbu y adalah *class benign* dan *class malignant* untuk informasi aktual. Berdasarkan penjelasan *confusion matrix* pada bab sebelumnya, baik pada *training data* maupun *testing data* warna hitam secara berturutan ialah FN dan FP sementara warna lainnya selain hitam ialah TP dan TN



**Gambar 3. Confusion Matrix model SVM**

Pada penjelasan Gambar 3 telah disebutkan mengenai keterangan dari *confusion matrix*, yaitu FN; FP; TP dan TN. FN adalah saat data pada prediksi model SVM terbaca *benign* yang diklasifikasikan *malignant*. FP adalah saat data pada prediksi model SVM terbaca *malignant* yang diklasifikasikan *benign*. TP adalah saat data pada prediksi model SVM terbaca *malignant* yang diklasifikasikan *malignant*. TN adalah saat data pada prediksi model SVM terbaca *benign* yang diklasifikasikan *benign*.

Setelah diketahui nilai dari *confusion matrix* baik untuk *training data* maupun *testing data* maka dapat dihitung nilai akurasi, nilai presisi, nilai sensitivitas, nilai spesififikasi, bahkan nilai error dari model SVM tersebut. Perhitungan dapat dilakukan menggunakan rumus yang tertera pada Tabel 2. Pada *training data* diperoleh nilai akurasi dalam persen sebesar 96,18%, nilai presisi dalam persen sebesar 98,63%, nilai sensitivitas dalam persen sebesar 95,57%, nilai spesififikasi dalam persen sebesar 97,39%, dan error sebesar 0,03%. Pada *testing data* diperoleh nilai akurasi dalam persen sebesar 97,07% dan nilai presisi dalam persen sebesar 98,14%, nilai sensitivitas dalam persen sebesar 97,24%, nilai spesififikasi dalam persen sebesar 96,77%, dan error sebesar 0,04%. Dapat dilihat dari nilai akurasi serta nilai presisi yang cukup tinggi untuk penggunaan 50% *training data* dan 50% *testing data* dari 699 *data record* yang ada. Maka, algoritma model SVM ini dapat dikatakan telah memiliki performa klasifikasi yang baik. Hal ini juga sejalan dengan tujuan dari SVM, yaitu menentukan *hyperplane* terbaik, yang bisa memberikan jarak terjauh (*maximum margin*) antara dua kelompok yang berbeda.

## KESIMPULAN

Setelah dilakukan proses evaluasi pada algoritma model SVM diperoleh nilai akurasi pada *training data*, yaitu sebesar 96,18% dan nilai akurasi pada *testing data*, yaitu sebesar 97,07%. Untuk nilai presisi pada algoritma model SVM diperoleh nilai sebesar 98,63% untuk *training data* serta sebesar 98,14% untuk *testing data*.

## UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan kepada peneliti sebelumnya dan juga pada sumber data, yaitu *website database* kanker payudara oleh Dr. William H. Wolberg. Data *Wisconsin Breast Cancer* (WBC) yang tersedia di *UCI Machine Learning Repository*.

## DAFTAR PUSTAKA

- [1] Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67-80.
- [2] Hammad, R., Kurniasih, J., Hasan, N. F., Dengen, C. N., & Kusri, K. (2019). Prototipe Machine Learning Untuk Prognosis Penyakit Demensia (The Prototype of Machine Learning for The Prognosis of Dementia). *JURNAL IPTEKKOM (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, 21(1), 17- 29.
- [3] Handayani, A., Jamal, A., & Septiandri, A. A. (2018). Evaluasi Tiga Jenis Algoritme Berbasis Pembelajaran Mesin untuk Klasifikasi Jenis Tumor Payudara. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*. <https://doi.org/10.22146/jnteti.v6i4.350>
- [4] Indrati, A., & Madenda, S. (2009). Ekstraksi Fitur Bentuk Tumor Payudara. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- [5] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*. Elsevier. <https://doi.org/10.1016/j.csbj.2014.11.005>
- [6] Mahendra, G. S., & Ernanda Aryanto, K. Y. (2019). SPK Penentuan Lokasi ATM Menggunakan Metode AHP dan SAW. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 49-56. <https://doi.org/10.25077/teknosi.v5i1.2019.49-56>
- [7] Maulana, M. S., Sabarudin, R., & Nugraha, W. (2019). Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 7(3), 202-206.
- [8] Nugraha, F. S., Shidiq, M. J., & Rahayu, S. (2019). ANALISIS ALGORITMA KLASIFIKASI NEURAL NETWORK UNTUK DIAGNOSIS PENYAKIT KANKER PAYUDARA. *Jurnal Pilar Nusa Mandiri*, 15(2), 149-156. <https://doi.org/10.33480/pilar.v15i2.601>
- [9] Seyam, M., Othman, F., & El-Shafie, A. (2017). Prediction of Stream Flow in Humid Tropical Rivers by Support Vector Machines. In *MATEC Web of Conferences* (Vol. 111). EDP Sciences. <https://doi.org/10.1051/mateconf/201711101007>
- [10] Umadevi, T. P., & Murugan, A. (2021). Enhanced handwritten document recognition using confusion matrix analysis. *Advances in Parallel Computing*, 39, 121- 126. <https://doi.org/10.3233/APC210131>