

# RANCANG BANGUN APLIKASI PREDIKSI KANKER PAYUDARA DENGAN PENDEKATAN *MACHINE LEARNING*

Nur Tri Ramadhanti Adiningrum<sup>1</sup>, Resa Rianti<sup>2</sup>, Cahyo Prianto<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Logistik dan Bisnis Internasional; Bandung, Jawa Barat, Indonesia

*Riwayat artikel:*

*Received: 18 Juli 2023*

*Accepted: 10 Agustus 2023*

*Published: 11 September 2023*

**Keywords:**

Kanker payudara;

Prediksi;

*Machine Learning*;

Support Vector Machine;

*Web Base*.

**Correspondent Email:**

[nurtrira06@gmail.com](mailto:nurtrira06@gmail.com)

**Abstrak.** Kanker payudara merupakan penyakit terbanyak yang ditemukan pada wanita dan angka kematiannya menempati posisi kedua kasus kanker yang dapat mempengaruhi lebih dari 2,1 juta orang pada tahun 2020. Berdasarkan kasus tersebut dapat diketahui kanker payudara adalah kanker penyebab utama kematian wanita. Namun kematian ini dapat dikurangi dengan melakukan deteksi dini terhadap sel kanker. Oleh karena itu, pencegahan kanker berperan penting dalam proses pengobatan dan membantu meningkatkan angka pemulihan hidup. Prediksi kanker dapat membantu pasien untuk berkonsultasi dengan dokter lebih cepat. Sehingga, prediksi kanker yang tepat sangat penting untuk memperbarui perawatan pasien kanker payudara. 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.

**Abstract.** Breast cancer is the most common disease found in women and the death rate ranks second in cancer cases which can affect more than 2.1 million people in 2015. Based on these cases, it can be seen that breast cancer is the leading cause of death for women. However, cancer prevention plays an important role in the treatment process and helps increase the survival rate. Cancer prediction can help patients to consult doctors more quickly. Thus, precise cancer prediction is very important to update the care of breast cancer patients. Machine Learning techniques can be used to predict breast cancer because they can capture high-level interactions between data that may produce better predictions to differentiate between benign and malignant. Therefore, this study used an effective classification approach Support Vector Machine (SVM). SVM is a model used to predict breast cancer. To simplify the prediction process, breast cancer prediction results are implemented in the form of a web base with the Django framework so that doctors can make decisions quickly. The modeling results show that the prediction of breast cancer using SVM obtains an accuracy of 98.24%. So that the model made is very good so that the application will work to predict well.

## 1. PENDAHULUAN

Salah satu penyakit paling berbahaya di dunia adalah kanker[1]. Sel penyebab kanker mungkin memiliki perubahan atau mutasi genetik[2]. Suatu kanker yang paling umum pada wanita yaitu kanker payudara. Kanker payudara masih memiliki tingkat kematian tertinggi kedua dari semua jenis kanker[3]. Sebuah studi menggunakan *database* dari GLOBOCAN, CDC, dan gudang kesehatan World Health Organization (WHO) menilik bahwa kanker payudara adalah penyakit mematikan yang merenggut ribuan nyawa setiap tahun[4]. Pada tahun 2020, sebuah studi yang dilakukan oleh Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa lebih dari 2.000.000 kasus baru dan lebih dari 600.000 kematian dilaporkan karena kanker payudara dalam satu tahun[4].

Berdasar kasus-kasus tersebut dapat diketahui bahwa kanker payudara adalah jenis kanker yang menjadi penyebab utama kematian wanita. Namun, kematian ini dapat dikurangi dengan deteksi dini sel kanker[5]. Diagnosis dini dan prediksi kanker payudara dapat meningkatkan kemungkinan bertahan hidup karena dapat membantu melakukan perawatan yang tepat waktu untuk pasien[4]. Prediksi kanker yang tepat waktu dapat membantu pasien untuk berkonsultasi dengan dokter tepat waktu[4]. Setiap kemajuan dalam prediksi dan diagnosis kanker merupakan investasi yang signifikan dalam hidup yang panjang dan sehat. Oleh karena itu, menjaga aspek pengobatan dan pemulihan pasien dalam hal kualitas hidup menjadi sangat penting[6].

Salah satu teknik yang paling sering digunakan dalam studi data medis adalah klasifikasi[7]. Hingga akhirnya, teknik *Machine Learning* otomatis menawarkan potensi untuk meningkatkan akurasi diagnosis kanker payudara sekaligus membuat prosesnya lebih cepat dan efisien[8].

Berbagai penelitian telah dilakukan terkait dengan kanker payudara, yang merupakan jenis kanker yang paling sering menyerang wanita dan memiliki salah satu dari dua tingkat kematian tertinggi dari jenis kanker lainnya. Sebuah studi yang dilakukan oleh Azminuddin DKK mengevaluasi keefektifan pendekatan pembelajaran mesin yang digunakan untuk memprediksi kanker pankreas. Pada penelitian tersebut diatas, tujuan dari makalah ini adalah

untuk meningkatkan produktivitas metode *Machine Learning* melalui proses pre-processing untuk memprediksi kanker payudara[3].

Dalam penelitian lain, Anusha dkk mengevaluasi penerapan *Machine Learning* dalam aplikasi medis seperti pendeteksian jenis sel kanker. Sel-sel kanker diklasifikasikan sebagai jinak (B) atau ganas (M). Dalam proyek ini, dilakukan analisis menggunakan Support Vector Machine (SVM) pada dataset Kanker Payudara Wisconsin. Dari metodologi yang digunakan, teknik SVM diidentifikasi sebagai teknik yang kuat untuk analisis prediktif. Dan berdasarkan temuan yang didapat, disimpulkan bahwa SVM menggunakan kernel Gaussian adalah teknik yang paling cocok untuk prediksi kekambuhan atau non-kekambuhan kanker payudara[5].

Berdasarkan paparan sebelumnya, diperlukan suatu kebutuhan berupa teknik komputasi untuk membantu penerapan analisis dalam diagnosis [9]. Hal ini karena kemajuan teknologi membuat pendekatan ini jauh lebih efektif seiring berjalannya waktu[10]. Selain itu, pada tahap awal penyakit, indikator penyakit dapat lebih mudah diidentifikasi melalui teknologi daripada pemeriksaan manual[11]. Prediksi kanker payudara menggunakan *Machine Learning* berperan penting dalam diagnosa awal penyakit sehingga dapat meminimalkan dan keterlambatan deteksi sel kanker dapat dihindari [12]. Hingga saat ini, berbagai model *Machine Learning* (ML) telah menjadi penggunaan utama dalam mendiagnosis [13].

Pendekatan *Machine Learning* untuk deteksi dini kanker payudara oleh Muawia dalam penelitian ini menyajikan metode yang efektif berdasarkan Support Vector Machine (SVM) dengan metode pemilihan fitur yang tepat yang hanya mempertimbangkan fitur yang memiliki pengaruh tinggi dan mengabaikan yang lainnya. Pemilihan SVM diambil setelah melakukan percobaan dengan tujuh pengklasifikasi yang populer di bidang diagnosis kanker payudara, hasil eksperimen menunjukkan bahwa pendekatan classifier berbasis SVM lebih unggul dibandingkan metode lainnya. Hasil eksperimen mencerminkan bahwa pendekatan classifier berbasis SVM mengungguli classifier lainnya dengan memperoleh akurasi tertinggi, mencapai 97,4% [14].

Sebuah penelitian lain yang dilakukan oleh dalam membuat aplikasi Prediksi Penjualan Kopi 372 berbasis *web*. *Framework* yang digunakan dalam pembangunan aplikasi ini adalah Django dengan *database* MySQL. Prediksi penjualan dilakukan dengan menerapkan algoritma regresi linier berganda menggunakan data transaksi sebanyak 372 kopi, cuaca harian Kota Bandung, hari libur nasional, dan hari libur bersama tahun 2018-2019. Berdasarkan hasil evaluasi model prediksi menunjukkan bahwa Gajua Kopi memiliki model prediksi terbaik, dengan MAE = 104.259.000, RSME = 114.408.000, dan R2 = 0,57581 [15].

Pada penelitian ini, dalam melakukan prediksi dini kanker payudara, metode yang diterapkan ialah menggunakan *Machine Learning* dengan algoritma Support Vector Machine[6]. Support Vector Machine dapat digunakan untuk melakukan prediksi dan diagnosis kanker payudara, dan menjadi topik utama penelitian dan telah terbukti sebagai teknik yang kuat dari berbagai jenis algoritma *Machine Learning*[14]. Salah satu keuntungan menggunakan model *Machine Learning* dibandingkan model statistik adalah jumlah fleksibilitas dalam menangkap interaksi tingkat tinggi antara data, yang mungkin menghasilkan prediksi yang lebih baik[4].

Disamping itu, penggunaan *website* dapat membantu pengguna dalam melakukan pekerjaan karena mudah digunakan dan mampu berproses secara cepat. Django adalah kerangka kerja *web* yang dibangun di atas bahasa pemrograman Python yang dibuat dengan tujuan untuk membuat aplikasi *web* yang aman, tangguh, dan menarik. Django, yang didirikan oleh Django Software Foundation, terus menerima pemutakhiran, menjadikannya kerangka kerja *web* pilihan bagi banyak pengembang aplikasi *web* baru[16].

Tujuan utama dalam penelitian ini adalah untuk memprediksi dan mendiagnosis kanker payudara menggunakan *Machine Learning* dengan model yang memiliki akurasi paling tinggi berdasarkan penelitian sebelumnya[6]. Support Vector Machine (SVM) adalah model *Machine Learning* yang digunakan untuk memprediksi perkembangan kanker payudara karena memiliki kinerja yang jauh lebih baik daripada model pembelajaran mesin lainnya [6]. Hasil prediksi kanker payudara dapat

digunakan oleh pengguna seperti dokter dalam menentukan keputusan dengan cepat. Selain itu, model akan dibuat dalam bentuk *website* agar prediksi dapat mudah digunakan. Visualisasi hasil prediksi tersebut akan ditampilkan berbasis *web* dengan *framework* Django agar lebih mudah dipahami dan untuk digunakan oleh tenaga medis.

## 2. TINJAUAN PUSTAKA

### 2.1. Kanker Payudara

Kanker payudara adalah bentuk kanker yang dimulai secara internal di jaringan payudara dan sering menyebar ke bagian tubuh lainnya. Ini adalah istilah umum untuk tumor yang dapat berkembang di jaringan payudara dari segala jenis sel[17]. Kanker payudara berkembang ketika beberapa sel payudara mulai berperilaku berbeda dari biasanya. Sel-sel ini terus berkembang biak dan membelah lebih cepat daripada sel sehat, mengakibatkan kekacauan atau massa. Dari payudara wanita, sel dapat berpindah ke kelenjar getah bening atau bagian lain dari tubuhnya.[18].

### 2.2. Machine Learning

*Machine Learning* adalah studi tentang membuat komputer belajar dan meningkatkan dengan cara yang meniru atau mengungguli kemampuan belajar manusia[19]. Teknik pembelajaran mesin otomatis dapat menyederhanakan proses ini dengan memilih model secara otomatis, mengoptimalkan *hyperparameter*, dan mengurangi kebutuhan intervensi manusia. Pendekatan ini juga dapat mengurangi beban kerja ahli radiologi dan ahli patologi serta menghemat waktu dan tenaga dalam diagnosis data pencitraan medis[8].

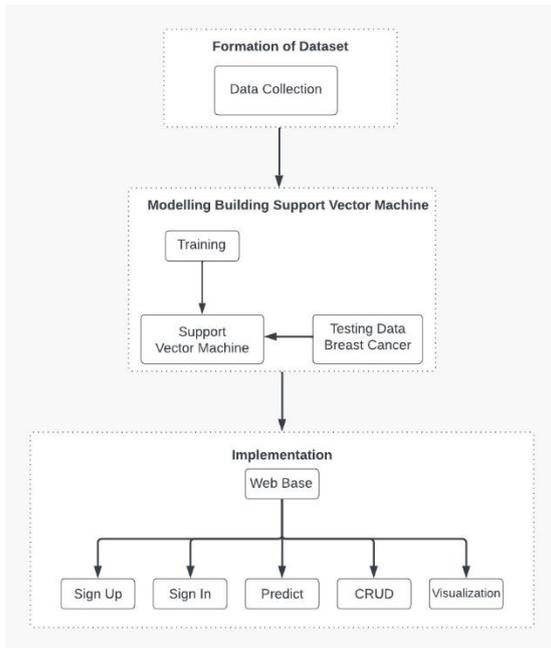
### 2.3. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma yang digunakan untuk menemukan *hyperplane* yang dengan jelas membagi titik data dari kelas yang berbeda[20].

### 2.4. Framework Django

Kerangka web Django memiliki Python sebagai bahasa pemrograman utamanya. Django adalah kerangka kerja *web* yang dirancang untuk membuat aplikasi daring yang dinamis, aman, dan kaya fitur [21].

## 3. METODE PENELITIAN



Gambar 1. Metodologi Penelitian

Tujuan dalam proses *data mining* adalah melaksanakan pengetahuan dari pengumpulan data. Eksplorasi, pembuatan model, dan konstruksi adalah tiga elemen utama dari latihan pengetahuan ini.[22].

Berdasarkan Gambar 1[23], metodologi penelitian dikelompokkan menjadi tiga tahap utama. Pada bagian Formation of Dataset hal yang dilakukan adalah mengumpulkan data dan melakukan data understanding, pada proses ini data dieksplorasi dan dilakukan pre-processing data. Pada tahap Modelling Building Support Vector Machine, data yang telah diolah kemudian ditraining dan dilakukan pembangunan model dan prediksi dengan SVM. Pada tahap Implementation, model prediksi dikembangkan dengan cara diterapkan dalam bentuk *web* dengan *framework* Django.

### 3.1. Formation Dataset (Kumpulan Formasi Dataset)

#### 3.1.1. Data Collection (Pengumpulan Data)

Data digunakan dalam penelitian ini untuk menganalisis dan mengevaluasi setiap metode klasifikasi yang diterapkan. Informasi yang dimasukkan dalam penelitian ini adalah data kanker payudara. yang didapatkan melalui situs bernama Kaggle pada tahun 2022 yang dapat dikunjungi melalui link yang dicantumkan sebagai berikut:

<https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>.

#### 3.1.2. Eksplorasi Data

Pada tahap eksplorasi dilakukan preprocessing data, transformasi data, pemilihan atribut, dll [22]. Data yang digunakan berlabel dan memiliki 32 kolom dan 569 record. Dataset ini berisi 357 kasus kanker payudara jinak dan 212 kasus kanker payudara ganas. Terdiri dari 32 kolom, dengan kolom pertama menunjukkan nomor ID, kolom kedua menunjukkan hasil diagnosis (jinak atau ganas), diikuti dengan rata-rata, standar deviasi dan rata-rata pengukuran terburuk dari sepuluh fitur. Fitur-fitur ini diperoleh dari penggunaan matriks korelasi dengan nilai tinggi antar atribut. Atribut yang dipilih sebanyak 21 atribut. Berikut adalah atribut dan maknanya [24] yang telah dipilih berdasarkan nilai korelasi atribut yang tinggi:

Tabel 1. Atribut Dataset Dan Artinya

Atribut	Arti
id	ID unik
diagnosis	Diagnosis masalah (M = <i>malignant</i> /ganas, B = <i>benign</i> /jinak)
Radius_mean	Jari-jari lobus/kelenjar (rata-rata jarak dari pusat ke titik-titik di perimeter) Perimeter adalah garis keliling inti sel yang diukur sebagai jumlah dari jarak.
Perimeter_mean	Perimeter Luar Lobus (ukuran rata-rata tumor inti)
Area_mean	Rata-rata Luas Lobus
Compactness_mean	Rata-rata dari rasio volume dan luas permukaan tumor (rata-rata dari $\text{perimeter}^2 / \text{area} - 1.0$ )
Concavity_mean	Rata-rata keparahan bagian cekung dari kontur. Sebuah cekungan di bagian bawah toraks (dada) yang disebabkan oleh perpindahan tulang rawan xiphoid ke belakang.
Concave_ponts_mean	Rata-rata untuk jumlah bagian cekung dari kontur
Radius_se	Jari-jari lobus/kelenjar (Standard Error dari pusat ke titik-titik di perimeter) Perimeter adalah garis keliling inti sel yang diukur sebagai jumlah dari jarak.
Texture_se	Standard Error Tekstur Permukaan (standar deviasi nilai <i>grayscale</i> )
Perimeter_se	Standard Error Perimeter Luar Lobus (ukuran Standard Error tumor inti)

Area_se	Standard Error Luas Lobus
Compactness_se	Standard Error dari rasio volume dan luas permukaan tumor (Standard Error dari $\text{perimeter}^2 / \text{area} - 1.0$ )
Concavity_se	Standard Error keparahan bagian cekung dari kontur. Sebuah cekungan di bagian bawah toraks (dada) yang disebabkan oleh perpindahan tulang rawan xiphoid ke belakang.
Concave points_se	Standard Error untuk jumlah bagian cekung dari kontur
Fractal_dimension_se	Standard Error dimensi fraktal (Standard Error "coastline approximation" - 1) Analisis fraktal dari gambar spesimen jaringan payudara memberikan deskripsi numerik pola pertumbuhan tumor sebagai angka kontinu antara 1 dan 2. Angka ini, dimensi fraktal, adalah ukuran yang objektif dan dapat direproduksi dari kompleksitas arsitektur jaringan spesimen biopsi.
Radius_worst	Keparahan jari-jari lobus/kelenjar (keparahan jarak dari pusat ke titik-titik di perimeter) Perimeter adalah garis keliling inti sel yang diukur sebagai jumlah dari jarak.
Texture_worst	Keparahan Tekstur Permukaan (standar deviasi nilai <i>grayscale</i> )
Perimeter_worst	Perimeter Luar Lobus (ukuran keparahan tumor inti)
Area_worst	Keparahan Luas Lobus
Compactness_worst	Keparahan dari rasio volume dan luas permukaan tumor (keparahan dari $\text{perimeter}^2 / \text{area} - 1.0$ )
Concavity_worst	Keparahan bagian cekung dari kontur. Sebuah cekungan di bagian bawah toraks (dada) yang disebabkan oleh perpindahan tulang rawan xiphoid ke belakang.
Concave points_worst	Keparahan untuk jumlah bagian cekung dari kontur

### 3.2. Model Building SVM (Pemodelan SVM)

#### 3.2.1. Training Data (Pelatihan Data)

Penelitian ini dimulai dengan memisahkan data pelatihan dan pengujian menggunakan fungsi "train\_test\_split" yang diimport dari library python sci-kit learn dalam seksi pemilihan model. Sebagian besar data training sebanyak 80% dan testing sebanyak 20% atau 455 data digunakan sebagai data training dan

sisanya sebanyak 114 data digunakan untuk data testing. Data pengujian kemudian digunakan dalam proses pemodelan tanpa label data untuk memastikan bahwa model dapat berfungsi dengan jenis data yang akan diuji. Kemudian dilakukan proses fitting antara  $x_{\text{train}}$  dan  $y_{\text{train}}$ .

#### 3.2.2. Modelling SVM (Memodelkan dengan SVM)

Pada proses pemilihan metode dan model yang tepat dilakukan dalam pengembangan dan penilaian model agar kinerja prediksi yang dihasilkan tinggi[22]. Langkah selanjutnya setelah melatih dan membuat klasifikasi model adalah menguji klasifikasi model terhadap data untuk melihat seberapa akuratnya. Pengujian menggunakan data hasil pengelasan dengan produk yang berbeda[25].

SVM adalah algoritma *Machine Learning* yang digunakan untuk klasifikasi dan regresi dengan pengawasan[26]. Metode SVM mengonversi data pelatihan awal ke dimensi yang lebih tinggi menggunakan tangkapan nonlinier. Data dari dua kelas selalu dapat dipisahkan oleh *hyperplane* dalam situasi ini karena dimensi baru akan menemukan *hyperplane* untuk dibagi secara linear dan dengan kekacauan nonlinear yang diperlukan ke dimensi yang lebih tinggi. SVM menentukan ini menggunakan margin dan vektor dukungan. Dalam metode ini harus diusahakan untuk memilih fungsi *classifier* atau separator terbaik yang dapat membedakan antara dua kelas yang berbeda. Untuk memisahkan dua jenis objek yang berbeda, teknik ini mencari melalui fungsi pemisah (*hyperplanes*) dalam jumlah tak terbatas. *Hyperplane* terbaik adalah yang berada di tengah dua kelompok objek yang termasuk dalam kelas yang berbeda. Ini dapat dinyatakan dalam masalah SVM untuk optimasi klasifikasi linier, seperti yang ditunjukkan di bawah ini[25].

$$\min_{\vec{w}} \tau(\vec{w}) = \frac{1}{2} \|\vec{w}\|^2 \dots \dots \dots (1)$$

$$\frac{1}{2} \|\vec{w}\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \dots \dots \dots (2)$$

Dengan syarat :

$$y_i(x_i \cdot w + b) \geq 1, i = 1,2,3, \dots, n \dots \dots (3)$$

$$y_i(x_1 \cdot w_1 + x_1 \cdot w_1 + b) \geq 1 \dots \dots \dots (4)$$

Dimana  $x_i$  adalah data input  $y_i$  adalah keluaran dari data  $x_i$ ,  $w$ ,  $b$  adalah parameter-parameter yang kita cari nilainya. Dalam rumus di atas, ingin meminimalkan fungsi tujuan (obyektif function)  $\frac{1}{2} \|\vec{w}\|^2$  atau memaksimalkan kuantitas  $\|\vec{w}\|^2$  dengan memperhatikan pembatas  $y_i(x_i \cdot w + b) \geq 1$ . Bila output data  $y_i = +1$ , maka pembatas menjadi  $(x \cdot w + b) \geq 1$ .

### 3.2.3. Testing Data (Tes Data)

Setelah pemodelan dibangun dan dilatih, maka model *Machine Learning* di uji menggunakan data testing. Hal ini nantinya akan menentukan apakah model dapat bekerja dengan baik atau tidaknya.

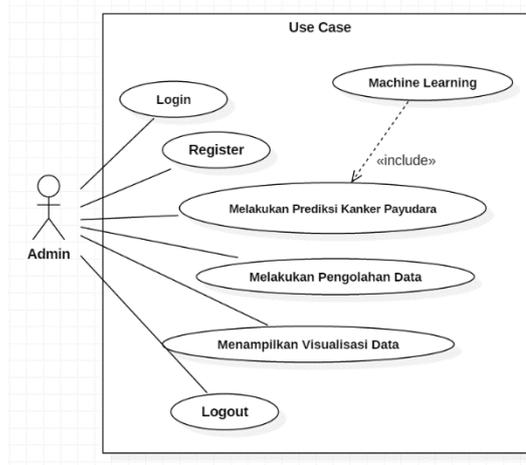
### 3.2.4. Evaluasi

Tingkat kinerja pola algoritma yang dikembangkan diperiksa selama tahap ini. [27]. Tahap evaluasi bertujuan untuk mengevaluasi kualitas model yang dibangun. Pendekatan yang digunakan pada titik ini menggunakan alat uji matriks kebingungan dan akurasi untuk menganalisis keluaran model dan menentukan seberapa sering prediksi data yang dibuat akurat jika dibandingkan dengan data aktual. Pengujian Precision, Recall, dan F1-SCORE, serta Accuracy, dapat digunakan untuk menguji metode klasifikasi [28]. Perbandingan antara data yang dikategorikan benar dengan semua data menghasilkan nilai akurasi. Nilai akurasi adalah rasio jumlah data kategori positif yang dikategorikan benar terhadap semua data kategori positif. Penarikan mengungkapkan proporsi data kategori positif yang diklasifikasikan dengan benar oleh sistem [29]. Perbandingan presisi dan gain rata-rata tertimbang disediakan oleh F1-SCORE [28]. Berdasarkan hal tersebut evaluasi model yang digunakan adalah presisi, Recall, F1-SCORE dan Akurasi.

### 3.3. Implementasi Aplikasi

Penerapan *website* dari hasil pemodelan ditentukan dengan menerapkan model pada data yang akan dievaluasi dan menjadi pondasi sistem data mining [22]. Use case diagram adalah jenis diagram UML (Unified Modeling Language) yang paling umum digunakan untuk mengilustrasikan bagaimana suatu sistem berinteraksi dengan aktor-aktor yang ada dalam sistem tersebut. Diagram ini dapat berfungsi sebagai ilustrasi yang berguna untuk

menunjukkan dengan jelas konteks suatu sistem sehingga komponen-komponen sistem dapat terlihat [30]. Diagram ini memfokuskan pada fungsi sistem dan cara aktor terlibat dengan sistem tersebut. Use case diagram pada Figure 3, mengartikan bahwa aplikasi dapat diakses oleh admin dan yang dapat melakukan *login*, registrasi, prediksi kanker payudara, CRUD data, menampilkan visualisasi data, dan *logout*.



Gambar 2. Use Case Diagram

## 4. HASIL DAN PEMBAHASAN

### 4.1. Hasil Evaluasi Model Machine Learning

Dalam penelitian ini, setelah melewati proses pemodelan *Machine Learning* menggunakan beberapa model. Model yang digunakan adalah Support Vector Machine (SVM) untuk prediksi kanker payudara. Berdasarkan Figure 2, pengujian nilai presisi untuk label Tumor Ganas 1.00 sedangkan untuk label Tumor Jinak 0.97, dan nilai Recall untuk label Tumor Ganas adalah 0.95 sedangkan untuk label Tumor Jinak 1.00, dan hasil F1-Score memiliki nilai 0.98 untuk label Tumor Ganas 0.99 untuk label Tumor Jinak dan Confusion Matrix memiliki nilai akurasi sebesar 0.98% [28].

	precision	recall	f1-score	support
Tumor Ganas (Terdeteksi Kanker)	1.00	0.95	0.98	42
Tumor Jinak (Tidak Terdeteksi Kanker)	0.97	1.00	0.99	72
accuracy			0.98	114
macro avg	0.99	0.98	0.98	114
weighted avg	0.98	0.98	0.98	114



Gambar 3. Hasil Klasifikasi dan Confusion Matrix

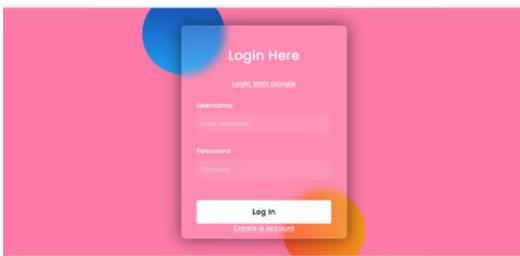
accuracy score: 98.24561403508771 %

Gambar 4. Skor Akurasi

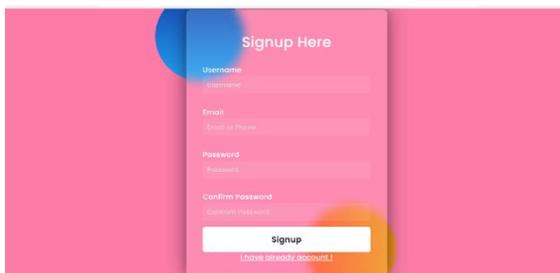
Berdasarkan Figure 3, hasil pemodelan menunjukkan bahwa pendekatan classifier berbasis SVM memperoleh akurasi yaitu 98,24%.

4.2. Hasil Implementasi Aplikasi

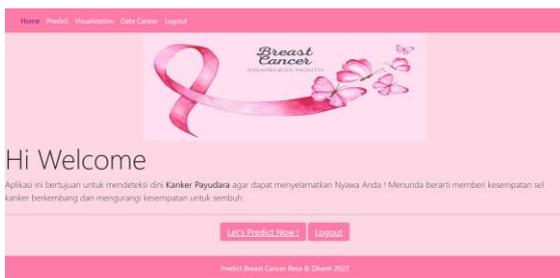
Pada bagian implementasi website prediksi dengan framework Django, hasil UI dapat dilihat pada gambar berikut.



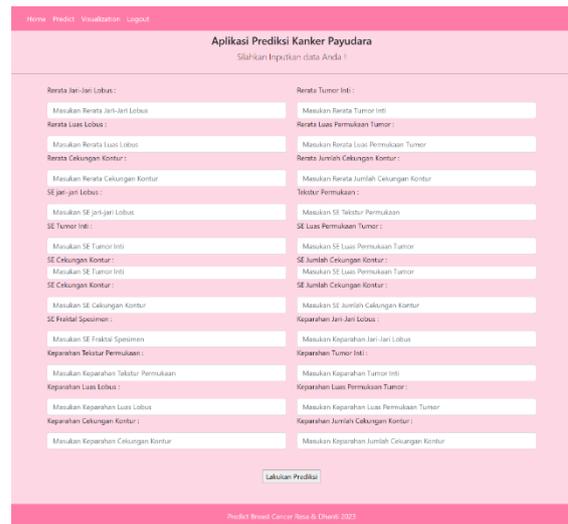
Gambar 5. Halaman Login Aplikasi



Gambar 6. Halaman Registrasi Aplikasi



Gambar 7. Halaman Depan Aplikasi



Gambar 8. Halaman Prediksi Kanker



Gambar 9. Halaman Hasil Prediksi Aplikasi

SE Luas Permukaan	SE Cakungan Kontur	SE Jumlah Cakungan Kontur	SE Fraktal Spektren	Keparahan Jari2 Lobus	Keparahan Tabur Permukaan	Keparahan Luas Tumor Inti	Keparahan Luas Permukaan Tumor	Keparahan Luas Cakungan Kontur	Keparahan Jumlah Cakungan Kontur	Action	
489	450	400	439	425	41	497	516	516	520	483	Edit Delete
135	177	336	309	419	199	465	511	210	278	401	Edit Delete
455	368	465	407	399	261	455	451	464	438	468	Edit

Gambar 10. Halaman Data Kanker Payudara



Gambar 11. Halaman Tambah Data Kanker Payudara Aplikasi

Gambar 12. Halaman Edit Data Kanker Payudara



Gambar 13. Halaman Visualisasi Aplikasi

## 5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa poin yaitu:

- Model prediksi yang dirancang menggunakan algoritma Support Vector Machine yang dibuat berhasil melakukan prediksi kanker payudara dengan baik, sehingga dapat disimpulkan bahwa model yang dibangun memiliki performa yang baik untuk memprediksi kanker payudara dengan menggunakan sembilan belas variabel independen.
- Berdasarkan model yang dibuat didapatkan nilai akurasi sebesar 98.24%. Akurasi tersebut merupakan nilai akurasi yang baik, sehingga dapat dikatakan model *Machine Learning* dapat berperforma baik untuk memprediksi kanker payudara.
- Berdasarkan uji validitas, nilai akurasi 0,9824 menunjukkan bahwa Diagnosis dipengaruhi oleh faktor independen 98.24%. Nilai sisa dari akurasi tersebut adalah 0,176 yang artinya Diagnosis dipengaruhi oleh faktor lain yang tidak diketahui sebesar 1.76%.
- Visualisasi data dari hasil model prediksi kanker payudara dapat digunakan menjadi bentuk aplikasi berbasis *web* dengan menggunakan *framework* Django. Dengan aplikasi tersebut, admin dapat melakukan

prediksi kanker payudara dengan mudah dan dengan cepat.

- Penelitian ini masih menggunakan dataset dari Kaggle. Maka, sumber data yang digunakan perlu dikombinasikan dengan data yang berasal dari Rumah Sakit untuk melakukan testing model prediksi.

## UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada pihak-pihak yang telah mendukung kelancaran penelitian ini dari proses penulisan hingga proses penerbitan, yaitu orang tua kami, rekan-rekan kami, penerbit, dan masih banyak lagi yang tidak dapat kami sebutkan satu persatu. Selain itu kami juga mengucapkan rasa terima kasih kepada pihak-pihak dari Jurnal Informatika dan Teknik Elektro Terapan yang telah melakukan membaca dan mengevaluasi penulisan kami dan memberikan kami kesempatan untuk mempublikasikan hasil karya ilmiah kami. Semoga karya ini dapat menjadi referensi yang bermanfaat bagi para pembaca serta dapat bermanfaat bagi masyarakat dalam arti luas.

## DAFTAR PUSTAKA

- L. Ayu Marcelina and A. Sri Yuliningtyas, "Penerapan Manajemen Stress sebagai Terapi Komplementer bagi Penyintas Kanker di Komunitas Kanker Indonesia," *Jurnal Bakti Masyarakat Indonesia*, vol. 4, no. 3, 2021, doi: <https://doi.org/10.24912/jbmi.v4i3.13467>.
- R. Y. Robot, M. F. Durry, and C. F. Kairupan, "Morfologi, Patogenesis, dan Imunoterapi Kanker Paru Tipe Adenokarsinoma," *Medical Scope Journal*, vol. 3, no. 1, p. 74, 2021, doi: 10.35790/msj.3.1.2021.33544.
- A. I. S. Azis, I. Surya Kumala Idris, B. Santoso, and Y. Aril Mustofa, "Pendekatan Machine Learning yang Efisien untuk Prediksi Kanker Payudara," *Jurnal Rekayas a Sistem dan T eknol ogi Informasi*, vol. 3, no. 3, pp. 458–469, 2019, Accessed: Oct. 22, 2022. [Online]. Available: <http://jurnal.iaii.or.id/index.php/RESTI/article/view/1347/180>
- S. Raj Gupta, "Prediction Time Of Breast Cancer Tumor Recurrence Using Machine Learning," *Cancer Treat Res Commun*, vol.

- 32, pp. 2–9, 2022, doi: <https://doi.org/10.1016/j.ctarc.2022.100602>.
- [5] A. Bharat, N. Pooja, and R. A. Reddy, “Using Machine Learning algorithms for breast cancer risk prediction and diagnosis,” *IEEE Third International Conference on Circuits, Control, Communication and Computing*, 2018, doi: <https://doi.org/10.1109/CIMCA.2018.8739696>.
- [6] M. A. Naji, S. el Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, “Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis,” *Procedia Comput Sci*, vol. 191, pp. 487–492, 2021, doi: <https://doi.org/10.1016/j.procs.2021.07.062>.
- [7] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 4, pp. 89–95, 2021, doi: <https://doi.org/10.33365/jtsi.v2i4.1373>.
- [8] A. E. E. Rashed, A. M. Elmorsy, and A. E. M. Atwa, “Comparative evaluation of automated machine learning techniques for breast cancer diagnosis,” *Biomed Signal Process Control*, vol. 86, 2023, doi: <https://doi.org/10.1016/j.bspc.2023.105016>.
- [9] J. A. Carter, C. S. Long, B. P. Smith, T. L. Smith, and G. L. Donati, “Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes,” *Expert Systems With Applications*, vol. 115, pp. 245–255, 2019, doi: <https://doi.org/10.1016/j.eswa.2018.08.002>.
- [10] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, “An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators,” *Healthcare Analytics*, vol. 2, Nov. 2022, doi: <https://doi.org/10.1016/j.health.2022.100118>.
- [11] J. Chaki, S. T. Ganesh, S. K. Cidham, and S. A. Theertan, “Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3204–3225, Jun. 2022, doi: <https://doi.org/10.1016/j.jksuci.2020.06.013>.
- [12] Romsida, “Transformasi Peran Akuntan dalam Era Revolusi Industri 4.0 dan Tantangan Era Society 5.0,” *Inovbiz : Jurnal Inovasi Bisnis*, vol. 7, pp. 206–212, 2019, Accessed: Jul. 18, 2023. [Online]. Available: [www.ejournal.polbeng.ac.id/index.php/IBP](http://www.ejournal.polbeng.ac.id/index.php/IBP)
- [13] A. Nicolucci *et al.*, “Prediction of complications of type 2 Diabetes: A Machine learning approach,” *Diabetes Res Clin Pract*, vol. 190, 2022, doi: <https://doi.org/10.1016/j.diabres.2022.110013>.
- [14] Muawia A. Elsadig, “A Machine Learning Approach For Breast Cancer Early Detection,” *J Theor Appl Inf Technol*, vol. 99, no. 5, pp. 1044–1053, 2021, Accessed: Jan. 17, 2023. [Online]. Available: <http://www.jatit.org/volumes/Vol99No5/4V099No5.pdf>
- [15] M. Aditia Farhan, “Pengembangan Aplikasi Prediksi Penjualan di 372 Kopi Menggunakan Algoritma Multiple Linear Regression,” Bandung, 2021.
- [16] D. Saputra and R. Fathoni Aji, “Analisis Perbandingan Performa Web Service Rest Menggunakan Framework Laravel, Django Dan Ruby On Rails Untuk Akses Data Dengan Aplikasi Mobile (Studi Kasus: Portal E-Kampus STT Indonesia Tanjungpinang),” *Bangkit Indonesia*, vol. 2, no. 2, pp. 17–22, 2018, doi: [10.52771/bangkitindonesia.v7i2.90](https://doi.org/10.52771/bangkitindonesia.v7i2.90).
- [17] G. Verras, L. Tchabashvili, D.-D. Chlorogiannis, F. Mulita, and M. Argentou, “Updated Clinical Evidence on the Role of Adipokines and Breast Cancer: A Review,” *MDPI*, vol. 15, no. 5, 2023, doi: <https://doi.org/10.3390/cancers15051572>.
- [18] Md. M. Hassan *et al.*, “A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction,” *Decision Analytics Journal*, vol. 7, 2023, doi: <https://doi.org/10.1016/j.dajour.2023.100245>.
- [19] M. M. Teye, “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions,” *MDPI*, vol. 12, no. 5, Apr. 2023, doi: <https://doi.org/10.3390/computers12050091>.
- [20] M. Cakir, M. Yilmaz, M. A. Oral, H. O. Kazanci, and Okan Oral, “Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture,”

- Journal of King Saud University-Science*, vol. 35, no. 6, Aug. 2023, doi: <https://doi.org/10.1016/j.jksus.2023.102754>
- [21] D. Saputra and R. F. Aji, "Analisis Perbandingan Performa Web Service Rest Menggunakan Framework Laravel, Django Dan Ruby On Rails Untuk Akses Data Dengan Aplikasi Mobile (Studi Kasus: Portal E-Kampus STT Indonesia Tanjungpinang)," *Bangkit Indonesia*, vol. 2, no. 7, pp. 17–22, 2018.
- [22] M. Ali Machmudi, "Uji Pengaruh Karakteristik Dataset Pada Performa Algoritma Klasifikasi," *Journal of Computer, information system, & technology management*, vol. 1, no. 2, pp. 7–11, 2018, Accessed: Jan. 21, 2023. [Online]. Available: <http://e-journal.unipma.ac.id/index.php/RESEARCH/article/view/2449/pdf>
- [23] Pratiwi. Andini, E. Budhiarti Nababan, and Amalia, "Detection of the Use of Mask to Prevent the Spread of COVID-19 Using SVM, Haar Cascade Classifier, and Robot Arm," *Journal of Computing and Applied Informatics (JoCAI)*, vol. 6, no. 2, pp. 125–137, 2022, Accessed: Jan. 20, 2023. [Online]. Available: <https://talenta.usu.ac.id/JoCAI/article/view/9289/5173>
- [24] M. H. Memon and Z. Wang, "Breast Cancer Detection in the IOT Health Environment Using Modified Recursive Feature Selection," *Hindawi: Wireless Communications and Mobile Computing*, vol. 2019, no. 5176705, pp. 1–19, 2019, doi: 10.1155/2019/5176705.
- [25] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan SMAW (Shield Metal Arc Welding)," *Jurnal Ilmiah Edutic*, vol. 5, no. 1, pp. 17–25, 2018.
- [26] Samsudiney, "Penjelasan Sederhana tentang Apa Itu SVM?," 2019. <https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02> (accessed Jan. 11, 2023).
- [27] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing (JAIC)*, vol. 5, no. 2, p. 103, 2021, Accessed: Jul. 18, 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [28] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb ," *Jurnal Pendidikan Informatika (EDUMATIC)*, vol. 5, no. 1, pp. 11–20, 2021, doi: 10.29408/edumatic.v5i1.3125.
- [29] F. Satria, Zamhariri, and M. Apun Syaripudi, "Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah Dan Ilmu Komunikasi UIN Raden Intan Lampung," *Jurnal Ilmiah MATRIK*, vol. 2, no. 1, pp. 28–35, 2020, Accessed: Jan. 20, 2023. [Online]. Available: <https://journal.binadarma.ac.id/index.php/jurnal/matrik/article/download/836/474/>
- [30] Tri A. Kurniawan, "Pemodelan Use Case (UML): Evaluasi Terhadap Beberapa Kesalahan Dalam Praktik," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 5, no. 1, pp. 77–86, 2018, doi: 10.25126/jtiik.201851610.