

# Implementasi Model Klasifikasi Jenis Kanker Payudara Menggunakan Algoritma SVM dan *Logistic Regression* berbasis Web

<sup>1</sup>Nunung Nurjanah, <sup>2</sup>Arphilia Nur Rani, <sup>3</sup>Hanny Hikmayanti Handayani, <sup>4</sup>Anis Fitri Nur Masruriyah,  
<sup>1,2,3,4</sup>Universitas Buana Perjuangan Karawang  
Kota Karawang, Indonesia

<sup>1</sup>if20.nunungnurjanah@mhs.ubpkarawang.ac.id, <sup>2</sup>if20.arphiliarani@mhs.ubpkarawang.ac.id,  
<sup>3</sup>hanny.hikmayanti@ubpkarawang.ac.id, <sup>4</sup>anis.masruriyah@ubpkarawang.ac.id

\*Penulis Korespondensi

Diajukan : 31/07/2023

Diterima : 06/09/2023

Dipublikasi : 02/10/2023

## ABSTRAK

Menurut Organisasi Kesehatan Dunia (WHO), ada 7 juta pasien kanker payudara, 5 juta di antaranya meninggal setiap tahun. Berdasarkan data Globocan untuk 2018 menunjukkan tingkat kematian rata-rata 17 per 100.000 orang dan insiden 2,1 per 100.000 orang yang menyerang perempuan di Indonesia. Hal itu, menyebabkan kanker payudara ini merusak genetic pada DNA dari sel epitel payudara menjalar ke ductus. Tujuan penelitian ini untuk mengklasifikasi jenis kanker (jinak atau ganas) yang diderita. Perbedaan penelitian sebelumnya dengan penelitian ini adalah metode pengujian algoritma yang dipilih. Pada penelitian ini menggunakan algoritma SVM dan Logistic Regression dengan SMOTE. Beberapa tahapan yang digunakan pada penelitian ini dimulai dengan pengumpulan data, kemudian *pre-processing*. Selanjutnya implementasi, evaluasi dan *deployment* pada sistem. Adapun metode K-Fold Cross Validation digunakan pada penelitian ini untuk melakukan partisi pada data. Sedangkan evaluasi model menggunakan *confusion matrix*. Berdasarkan tujuan penelitian, *deployment* dilakukan menggunakan *flask* untuk melakukan mengimplementasikan model pada sistem. Adapun metode pengembangan sistem yang digunakan pada penelitian ini yaitu RAD dengan beberapa tahapan. Tahapan dimulai dengan analisis kebutuhan, prototype dan implementasi. Berdasarkan hasil dari penelitian ini menunjukkan bahwa *accuracy* yang didapat sebesar 1.0, *precision* 1.0 dan *recall* 1.0. Selain itu, *accuracy* yang didapatkan pada sistem yaitu 90%. Maka dari itu, diharapkan berdasarkan hasil penelitian ini dapat membantu tenaga medis untuk mengklasifikasikan jenis kanker payudara, guna melakukan pengobatan secara cepat dan tepat pada penderita penyakit kanker payudara.

**Kata kunci:** Kanker payudara, K-Fold Cross Validation, *Logistik Regresi*, SVM

## I. PENDAHULUAN

Kanker payudara dikatakan sebagai jenis kanker yang umum terjadi pada wanita dan tingkat kematiannya kedua setelah jenis kanker lainnya (Azis et al., 2019). Kanker juga menjadi salah satu tantangan besar bagi manusia karena dapat mempengaruhi berbagai organ tubuh. Kanker payudara ini dapat menyerang semua usia (Nurelasari, 2018). Menurut Organisasi Kesehatan Dunia (WHO), ada 7 juta pasien kanker payudara, 5 juta di antaranya meninggal setiap tahun. Berdasarkan data Globocan 2018 tingkat kematian rata-rata 17 per 100.000 jiwa dan insiden sebanyak 2,1 per 100.000 jiwa menyerang wanita di Indonesia (KUSUMA et al., 2022). Kanker payudara ini menyebabkan kerusakan *genetic* pada DNA dari sel epitel payudara menjalar

ke *ductus*. Kanker payudara ini dapat dicegah dengan melakukan pemeriksaan payudara diri sendiri (SADARI) (Choirunnasih, 2019).

Penyakit kanker payudara ini telah terbukti dapat diprediksi dan diklasifikasi dengan menggunakan pengetahuan data mining (Nurnawati, 2022)(Solikin, 2021)(Dikshit et al., 2022)(Random et al., 2021)(Harafani, 2020)(Ramadhan & Kurniawati, 2020). Data mining digunakan untuk menemukan pola, korelasi dan tren untuk memilah-milah data dengan jumlah yang besar(Normah et al., 2022). Sebelumnya, Athalla dkk melakukan penelitian dengan metode K-Nearest Neighbor untuk mengklasifikasikan penyakit kanker payudara. Penelitian ini menggunakan perhitungan jarak menurut Minkowski berdasarkan jarak terdekat dengan objek. Hasil dari penelitian menggunakan metode K Nearest Neighbor diperoleh akurasi sebesar 93%(Khadijah & Kusumaningrum, 2019). Secara umum kanker payudara ini dibagi menjadi dua, yaitu jinak dan ganas(Random et al., 2021). Kanker jinak ini biasanya ditandai dengan benjolan kecil, bulat dan lembut. Sedangkan kanker ganas biasa ditandai dengan bentuk payudara yang tidak simetris, kasar, dan terasa nyeri (Athalla et al., 2018). Pendeteksian secara dini kanker payudara membantu dalam mengurangi resiko kematian yang tinggi. Pemerintah Indonesia sendiri bahkan membuat program mengenai deteksi dini mengenai kanker payudara (Azis et al., 2019). Berdasarkan penelitian sebelumnya, maka penelitian ini perlu dilakukan untuk mengklasifikasi jenis kanker (jinak atau ganas) yang diderita. Hal ini dapat membantu untuk melakukan pengobatan secara cepat dan tepat pada penderita penyakit kanker payudara.

## II. STUDI LITERATUR

### SVM

Support Vector Machine (SVM) adalah metode untuk memecahkan masalah klasifikasi dan regresi. Namun, SVM lebih dikenal dan lebih banyak digunakan dalam konteks klasifikasi (Sonjaya et al., 2022). SVM juga dapat menemukan *hyperplane* (pemisah) terbaik untuk memisahkan menjadi dua kelas dan memaksimalkan jarak antara kedua kelas (Pratama et al., 2018). SVM dapat dibagi menjadi dua kategori, yaitu SVM *linear* dan SVM *non-linear*. SVM *linear* digunakan ketika data dapat dipisahkan secara *linear* dapat dipisahkan menggunakan *hyperplane* dengan margin yang lembut (*soft margin*). Sementara itu, SVM *non-linear* melibatkan penggunaan fungsi kernel untuk mengubah ruang fitur menjadi ruang yang memiliki dimensi yang lebih tinggi (Rahman Isnain et al., 2021). Adapun rumus SVM dapat dilihat pada Tabel 1 sebagai berikut:

Tabel 1. Rumus Support Vector Machine (SVM)

SVM	Jenis Kernel	Definisi Rumus
Linear	Linear	$K(x, y) = x, y$
Non Linear	<i>Polynomial</i>	$K(x, y) = (x, y + 1)^p$
	RBF	$K(x, y) = e^{- x.y /2\sigma^2}$
	<i>Sigmoid</i>	$(x, y) = \tanh(Kx.y - \delta)$

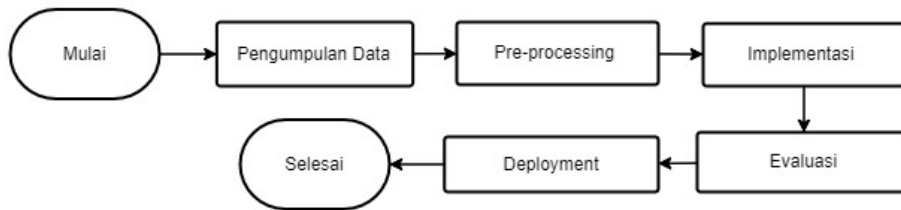
Sumber: (Sonjaya et al., 2022)

### Logistic Regression

Algoritma *Logistic Regression* digunakan untuk melakukan klasifikasi dengan menghitung suatu kejadian yang akan terjadi atau tidak. Proses klasifikasi dapat terjadi jika variabel dependen adalah variabel dikotomi. Dimana variabel dikotomi direpresentasikan sebagai angka 1 dan 0(Ismafillah et al., 2023). *Logistic Regression* akan mengidentifikasi hubungan antara variabel dependen (Y) sebagai kejadian yang akan terjadi atau tidak, dengan variabel independen yaitu *categorical* atau *continuous* (Gunawan et al., 2020).

### III. METODE

Dalam penelitian ini, metode yang digunakan yaitu Support Machine Learning dan Logistic Regression. Diagram alir penelitian dapat dilihat pada Gambar 1



Gambar 1. Diagram alir penelitian

#### Pengumpulan Data

Dataset yang digunakan diperoleh dari [Breast cancer prediction | Kaggle](#) dengan format .csv dengan jumlah 569 data dengan 32 atribut. Pada dataset ini terdapat atribut diagnosis yang menunjukkan kanker jinak (benign) atau ganas (malignant). Selain itu, dataset ini terbagi menjadi tiga fitur yaitu Mean (rata-rata), Standar deviasi (se), dan terburuk (worst). Fitur tersebut terbagi untuk setiap kolom dengan berbagai ukuran yang terdiri dari *radius*, *teksture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*.

#### Pre-Processing

Adapun beberapa teknik yang dilakukan pada tahapan ini dijelaskan sebagai berikut;

- Mengubah Tipe Data**  
Perubahan tipe data yang dilakukan adalah mengubah tipe data 'object' menjadi 'category' untuk memudahkan pembacaan data.
- Seleksi Fitur**  
Pada tahap ini dilakukannya seleksi fitur dengan menggunakan Teknik correlation based fitur. Seleksi fitur ini dapat menemukan subset fitur asli dengan melakukan pendekatan yang berbeda berdasarkan informasi (Dikshit et al., 2022). Seleksi fitur ini juga digunakan untuk melihat keterhubungan antara satu atribut dengan satu atau dua atribut lain yang ada pada dataset.
- Cleaning data**  
Pembersihan data dilakukan pada penelitian ini guna untuk menyelesaikan data dari *noisy*, *outlier*, *missing value* maupun duplikasi data, sehingga data siap digunakan pada tahap selanjutnya (Nikmatun & Waspada, 2019).
- Normalisasi Data**  
Tahapan normalisasi data ini digunakan untuk membantu menghilangkan perbedaan skala antar atribut. Adapun metode normalisasi data yang digunakan yaitu metode *Min-max scaling*.
- Transformasi Data**  
Transformasi data yang dilakukan pada penelitian ini untuk mengubah suatu data kategorik menjadi data numerik, sehingga data akan mudah untuk proses memodelkan data. Adapun Teknik yang digunakan yaitu *label encoding*.

#### Implementasi

Sebelum implementasi model dilakukan ada tahap ini, penelitian ini menerapkan Teknik SMOTE. *Synthetic Minority Oversampling Technique* (SMOTE) adalah sebuah algoritma preprocessing yang umum digunakan untuk menangani masalah ketidakseimbangan data. Selanjutnya dilakukan implementasi model dengan menggunakan algoritma SVM dan *Logistic Regression*.

**Evaluasi**

Pada penelitian ini, evaluasi yang digunakan yaitu *Confusion matrix*. *Confusion matrix* merupakan suatu table untuk menyatakan jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021). *Confusion matrix* dilakukan untuk mengetahui tingkat performa suatu model dengan berdasarkan presisi dan akurasi (Sonjaya et al., 2022). Adapun rumus confusion matrix yang digunakan yaitu ditunjukkan pada Tabel 2.

Tabel 2. Rumus Evaluasi

Aktual		
Prediksi	A	B
A	TP (True Positive)	FP (False Positive)
B	FN (False Negative)	TN (True Negative)

Sumber: (Sonjaya et al., 2022)

Sedangkan data yang diuji dilakukan partisi dengan menggunakan Teknik K-Fold Cross Validation. K-Fold Cross Validation adalah metode pengujian *Cross Validation* yang digunakan untuk mengevaluasi kinerja suatu Algoritma dengan membagi sampel data menjadi beberapa kelompok secara acak dan membagi data tersebut menjadi K kelompok fold. Setiap kelompok fold akan digunakan secara bergantian sebagai data uji, sedangkan kelompok fold lainnya akan digunakan sebagai data latih (Cahyanti et al., 2020).

**Deployment**

Model pada penelitian kemudian akan diterapkan pada sistem berbasis web dengan menggunakan *flask*. *Flask* adalah sebuah *web framework* yang digunakan dalam bahasa pemrograman Python (Ningtyas & Setiyawati, 2021). Selanjutnya, sistem yang dibuat menggunakan perancangan model pengembangan *Rapid Application Development (RAD)*. Adapun tahapan pada RAD adalah analisis kebutuhan, *prototype* dan implementasi (Normawati & Prayogi, 2021). Kemudian pengujian sistem yang digunakan adalah *blackbox testing*. Pengujian *black box* adalah sebuah metode perancangan data uji yang berdasarkan pada spesifikasi perangkat lunak. Data uji tersebut dieksekusi pada perangkat lunak dan hasil keluaran dari perangkat lunak diperiksa untuk memastikan apakah sesuai dengan yang diharapkan (Fahrezi et al., 2022). Setelah itu, dilakukan perhitungan akurasi secara manual pada sistem, guna mengetahui performa implementasi model pada sistem tersebut.

**IV. HASIL DAN PEMBAHASAN**

**Pra-Pemrosesan**

Pada tahap *pre-processing* telah dilakukannya perubahan tipe data guna untuk membantu kecocokan data dalam dataset. Perubahan tipe data ditunjukkan pada gambar dibawah ini:

```
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   id                    569 non-null   int64
1   diagnosis              569 non-null   object
2   radius_mean            569 non-null   float64
3   texture_mean          569 non-null   float64
..   ..                    ..
```

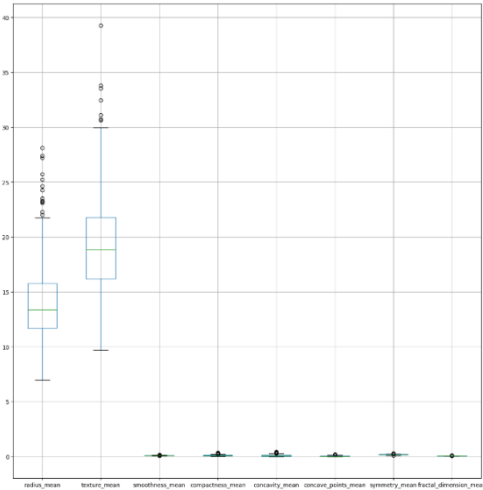
Gambar 2. *Before* Ubah Tipe Data

```
[ ]
df['diagnosis'] = df['diagnosis'].astype('category')
df.dtypes
id                int64
diagnosis         category
radius_mean       float64
texture_mean      float64
..               ..
```

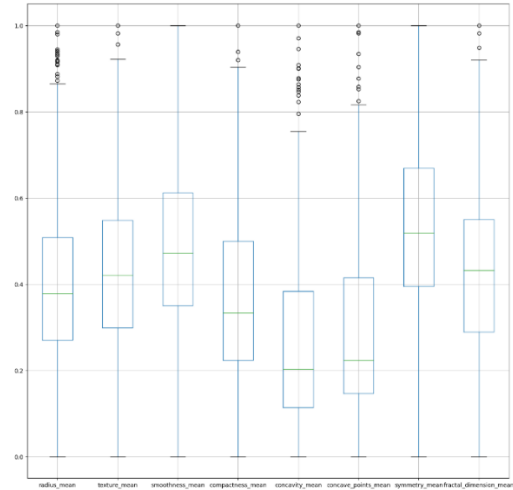
Gambar 3. *After* Ubah Tipe Data

Sumber: Hasil Pengolahan Data

Kemudian, dilakukannya seleksi fitur untuk mengidentifikasi atribut-atribut yang memiliki pengaruh besar pada data. (Tiana & Wahyuni, 2020). Selanjutnya, Data dibersihkan agar tidak adanya noise atau outlier. Adapun *cleaning data* dapat ditunjukkan pada gambar berikut:



Gambar 4. Before Cleaning Data



Gambar 5. After Cleaning Data

Sumber: Hasil Pengolahan Data

Setelah dilakukan *cleaning* data, maka data yang berjenis numerik dilakukan normalisasi agar mempermudah dalam proses pensklaan data. Sedangkan data yang kategorik dilakukan transformasi data guna untuk mempermudah dalam pengujian (Sonjaya et al., 2022). Hasil normalisasi data dapat ditunjukkan pada gambar berikut:

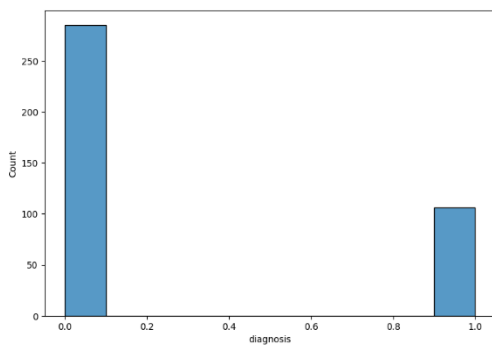
diagnosis	radius_mean	texture_mean	smoothness_mean	compactness_mean	diagnosis	radius_mean	texture_mean	smoothness_mean	compactness_mean	
1	M	20.57	17.77	0.08474	0.07864	M	0.939274	0.400995	0.351796	0.331766
2	M	19.69	21.25	0.10960	0.15990	M	0.872476	0.574129	0.745213	0.786698
4	M	20.29	14.34	0.10030	0.13280	M	0.918020	0.230348	0.598038	0.634979
6	M	18.25	19.98	0.09463	0.10900	M	0.763170	0.510945	0.508308	0.501736
7	M	13.71	20.83	0.11890	0.16450	M	0.418552	0.553234	0.892388	0.812451

Gambar 6. Before Normalisasi Data

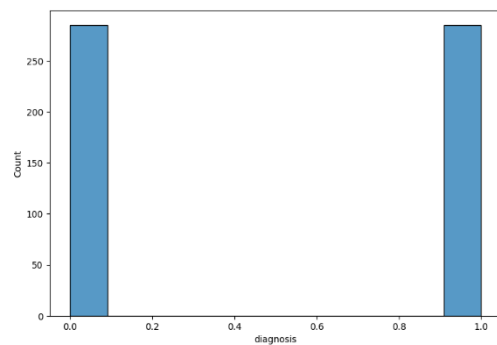
Gambar 7. After Normalisasi Data

Sumber: Hasil Pengolahan Data

Kemudian, berdasarkan data kanker payudara yang diperoleh dari sumber Kaggle menunjukkan adanya ketidak seimbangan, maka perlu diterapkannya metode oversampling untuk mencapai keseimbangan data. Demikian, setelah penerapan metode oversampling, data akan menjadi seimbang. Teknik yang digunakan yaitu SMOTE. Adapun hasil penerapan metode oversampling ditunjukkan pada gambar dibawah ini:



Gambar 8. Before SMOTE



Gambar 9. After SMOTE

Sumber: Hasil Pengolahan Data

### Modeling

Pada tahap modelling ini dilakukan klasifikasi data. Selanjutnya, data yang telah dilakukan pre-processing dipartisi terlebih dahulu menggunakan Teknik K-Fold Cross Validation.

Kemudian setelah dipartisi, dilakukan modeling menggunakan Algoritma SVM jenis *kernel polynomial* dan *Logistic Regression* dengan Teknik SMOTE. SMOTE ini merupakan suatu Teknik *oversampling*. Selanjutnya, akan dilakukan evaluasi performa algoritma SVM dengan menggunakan SMOTE atau Teknik *oversampling*.

### Evaluasi dan Perbandingan

Metode *k-fold cross-validation* digunakan untuk mengurangi bias yang mungkin muncul pada data acak (Sonjaya et al., 2022). Dalam penelitian ini, digunakan metode *10-fold cross-validation* yang melibatkan pembagian dataset menjadi 10-fold yang berbeda dengan ukuran yang sama. Setiap fold digunakan sebagai set pengujian. Sedangkan fold yang lain digunakan sebagai set pelatihan, dan proses ini diulang sebanyak 10 kali untuk menguji model secara *komprehensif* (Qadrini, 2023). Kemudian, dilakukan klasifikasi dengan algoritma SVM dan *Logistic Regression* dengan Teknik SMOTE dan tanpa SMOTE. Adapun hasil klasifikasi didapatkan nilai *accuracy*, *precision* dan *recall* dengan SMOTE dapat ditunjukkan pada Tabel 3

Tabel 3. Hasil Evaluasi Model dengan SMOTE

K-Fold	SVM			Logistic Regression		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	0.96	0.96	0.96	0.92	0.96	0.90
2	0.98	1.0	0.96	0.96	1.0	0.92
3	1.0	1.0	1.0	0.94	0.92	0.92
4	1.0	1.0	1.0	0.98	1.0	0.97
5	0.98	0.95	1.0	0.92	0.90	0.90
6	0.91	0.88	0.91	0.91	0.85	0.95
7	0.92	0.92	0.92	0.96	0.96	0.96
8	0.94	0.93	0.96	0.98	0.96	1.0
9	0.98	1.0	0.96	0.96	1.0	0.93
10	1.0	1.0	1.0	0.91	1.0	0.84
<b>Rata-rata</b>	<b>0.97</b>	<b>0.96</b>	<b>0.97</b>	<b>0.94</b>	<b>0.96</b>	<b>0.93</b>

Sumber: Hasil Pengolahan Data

Berdasarkan Tabel 3 diperoleh bahwa *accuracy*, *precision* dan *recall* tertinggi terdapat pada K-Fold 10 menggunakan metode SVM. Adapun hasil evaluasi yang didapat yaitu *accuracy 1.0*, *precision 1.0* dan *recall 1.0*. Selanjutnya, dilakukan klasifikasi dengan metode SVM dan Logistic Regression tanpa menggunakan Teknik SMOTE. Adapun hasil evaluasi model tanpa SMOTE ditunjukkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model Tanpa SMOTE

K-Fold	SVM			LR		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	0.97	1.0	0.91	0.95	1.0	0.83
2	0.94	0.90	0.90	0.92	0.9	0.81
3	0.92	0.81	0.9	0.94	1.0	0.8
4	1.0	1.0	1.0	0.97	1.0	0.83
5	0.92	0.66	0.8	1.0	1.0	1.0
6.	0.94	0.8	1.0	0.94	1.0	0.75
7.	0.94	0.90	0.90	0.97	1.0	0.90
8.	0.92	0.93	0.87	0.89	1.0	0.75
9.	0.97	1.0	0.93	0.94	1.0	0.87
10.	0.94	0.90	0.90	0.97	1.0	0.90
<b>Rata-rata</b>	<b>0.95</b>	<b>0.89</b>	<b>0.91</b>	<b>0.95</b>	<b>0.99</b>	<b>0.84</b>

Sumber: Hasil Pengolahan Data

Berdasarkan Tabel 4 menunjukkan accuracy, precision, recall tertinggi terdapat pada K-Fold 9. Adapun hasil *accuracy* yaitu 0.97, *precision* 1.0 dan *recall* 0.93. Bersumber dari penelitian sebelumnya (Sonjaya et al., 2022) (Widyaningsih et al., 2021) menyebutkan bahwa K-Fold10 merupakan pilihan terbaik untuk mendapatkan estimasi yang akurat. Selanjutnya dilakukan perbandingan hasil evaluasi model ditunjukkan guna menentukan kinerja model untuk diimplementasikan ke sistem. Adapun hasil perbandingan model ditunjukkan pada Tabel 5.

Tabel 5. Perbandingan Model

Metode	Accuracy	Precision	Recall
SVM	0.94	0.90	0.90
LR	<b>0.97</b>	<b>1.0</b>	<b>0.90</b>
SVM + SMOTE	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>
LR + SMOTE	0.91	1.0	0.84

Sumber: Hasil Pengolahan Data

Adapun hasil perbandingan evaluasi model dengan SMOTE pada penelitian ini tertinggi adalah *accuracy* 1.0, *precision* 1.0 dan *recall* 1.0 dengan metode SVM. Sedangkan hasil evaluasi model tanpa SMOTE tertinggi adalah *Accuracy* 0.97, *precision* 1.0 dan *recall* 0.90 dengan metode LR. Maka hasil perbandingan evaluasi model dengan menggunakan SMOTE cenderung lebih tinggi dibandingkan model tanpa SMOTE. Adapun dilihat dari hasil perbandingan model menunjukkan adanya peningkatannya yang signifikan pada SVM dengan SMOTE dibandingkan SVM tanpa SMOTE. Selanjutnya, hasil dari model diimplementasikan pada sistem dengan menggunakan *flask*.

## Deployment

### a. Identifikasi Pengguna

Adapun identifikasi pengguna dilakukan untuk mengidentifikasi kebutuhan pengguna pada sistem.

Tabel 6. Identifikasi Pengguna

No	Pengguna	Keterangan
1.	Tenaga Medis	Tenaga Medis merupakan pengguna yang dapat melakukan klasifikasi

Berdasarkan Tabel 6 identifikasi pengguna yaitu Tenaga medis. Tenaga medis dapat melakukan fungsi klasifikasi pada sistem.

#### b. Analisis Kebutuhan Fungsional dan Non Fungsional

Adapun Analisis kebutuhan pengguna dibagi menjadi 2 bagian, yaitu kebutuhan fungsional dan non-fungsional.

##### 1. Kebutuhan Fungsional

Tabel 7. Kebutuhan Fungsional

No	Deskripsi Kebutuhan	Aksi
1.	Sistem dapat menampilkan hasil klasifikasi	Klasifikasi

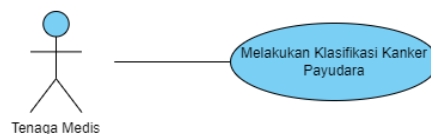
Berdasarkan Tabel 7 menunjukkan bahwa aksi dari sistem terdiri dari aksi klasifikasi saja. Sistem ini akan melakukan proses klasifikasi berdasarkan data yang diberikan. Output dari hasil klasifikasi akan menampilkan jenis diagnosis kanker payudara, yaitu jinak (*benign*) atau ganas (*malignan*).

##### 2. Kebutuhan Non Fungsional

1. sistem yang disajikan oleh beberapa *software web browser*.
2. Sistem dapat mudah digunakan.

#### c. Usecase

Adapun usecase dari sistem klasifikasi kanker payudara dapat dilihat pada Gambar 12.



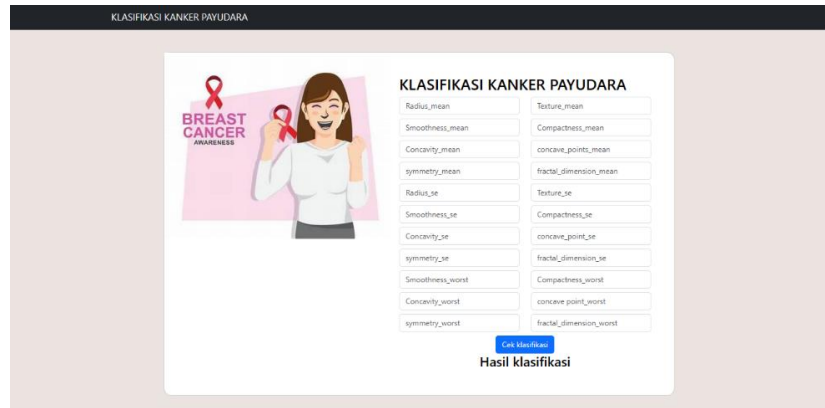
Gambar 10. Usecase Diagram klasifikasi

Berdasarkan Gambar 10 menunjukkan bahwa aktor dari pengguna sistem ini adalah petugas medis. Petugas medis ini dapat melakukan proses klasifikasi kanker payudara dengan menginputkan data pasien ke sistem.

#### d. Implementasi Antar Muka

Adapun tampilan sistem dapat dilihat pada Gambar 11.





Gambar 11. Tampilan sistem

Berdasarkan Gambar 11 menunjukkan tampilan sistem klasifikasi kanker payudara. Sistem ini mempunyai satu fungsi yaitu cek klasifikasi. Sistem ini akan berjalan ketika petugas medis menginput data pasien berdasarkan kolom yang ditampilkan. Setelah diinputkan sistem akan memproses data pasien dan menampilkan hasil klasifikasi.

**Pengujian**

Pengujian sistem yang dilakukan menggunakan *blackbox testing* dan uji validasi. Pengujian blackbox testing dilakukan pada fungsi sistem. Adapun fungsi sistem ini terdiri dari kelas uji cek klasifikasi dengan nilai input benar dan salah. Hasil pengujian *blackbox testing* dapat dilihat pada Tabel 8.

Tabel 8. Pengujian *Blackbox*

No	Kelas Uji	Nilai Input	Skenario Pengujian	Hasil yang diharapkan	Status
1.	Cek Klasifikasi	Benar	Mengisi semua Form	Dapat Menampilkan hasil	Valid
		Salah	Tidak mengisi semua	Menampilkan error	Valid

Sumber: Hasil Pengolahan Data

Berdasarkan Pengujian sistem pada Tabel 8 menunjukkan bahwa fungsi sistem dari cek klasifikasi berfungsi dengan baik. Adapun skenario pengujian sistem ini terdiri dengan mengisi semua form input dan tidak mengisi semua form input. Mengisi semua form input akan menampilkan hasil klasifikasi kanker payudara. Sedangkan tidak mengisi semua form input akan menampilkan pesan error. Adapun hasil uji validasi dapat dilihat pada Gambar 12.

No	Kelas Uji	Hasil yang diharapkan	Hasil sistem	Status
1.	Kelas uji 1	Kanker Ganas	Kanker Ganas	Valid
2.	Kelas uji 2	Kanker Ganas	Kanker Ganas	Valid
3.	Kelas uji 3	Kanker Ganas	Kanker Ganas	Valid
4.	Kelas uji 4	Kanker Ganas	Kanker Ganas	Valid
5.	Kelas uji 5	Kanker Ganas	Kanker Ganas	Valid
6.	Kelas uji 6	Kanker Jinak	Kanker Jinak	Valid
7.	Kelas uji 7	Kanker Jinak	Kanker Jinak	Valid
8.	Kelas uji 8	Kanker Jinak	Kanker Jinak	Valid
9.	Kelas uji 9	Kanker Jinak	Kanker Jinak	Valid
10.	Kelas uji 10	Kanker Jinak	Kanker Ganas	Invalid

Gambar 12. Uji Validasi  
 Sumber: Hasil Pengolahan Data

Adapun perhitungan akurasi secara manual sebagai berikut:

Jumlah data sesuai	: 9
Jumlah tidak sesuai	: 1
Total data testing	: 10

$$\text{Accuracy} = 9/10 \times 100 = 90\%$$

Berdasarkan Gambar 12, hasil uji validasi dengan perhitungan manual menunjukkan bahwa accuracy sebesar 90%. Hal ini menunjukkan bahwa hasil uji validasi cenderung memiliki hasil yang baik.

## V. KESIMPULAN

Penelitian ini menggunakan data kanker payudara yang diperoleh dari *kaggle*. Adapun beberapa tahapan yang digunakan pada *preprocessing* ini adalah mengubah tipe data, seleksi fitur, *cleaning* data, normalisasi data, transformasi data. Selanjutnya dilakukan Teknik SMOTE untuk menangani data yang tidak seimbang. Kemudian dilakukannya implementasi pada model dengan algoritma SVM dan *Logistic Regression* menggunakan K-Fold Cross Validation. Hasil perbandingan evaluasi model yang memiliki nilai terbaik pada penelitian ini yaitu metode SVM dengan SMOTE. Adapun hasil *accuracy* sebesar 1.0, *precision* 1.0 dan *recall* 1.0. Selanjutnya, *deployment* dilakukan menggunakan metode RAD. Adapun pengujian sistem menggunakan *blackbox testing* menunjukkan fungsi sistem berjalan dengan baik. Selanjutnya, dilakukan uji validasi sistem terhadap sistem mendapatkan *accuracy* 90%. Maka dari itu, diharapkan berdasarkan hasil penelitian ini dapat membantu tenaga medis untuk mengklasifikasikan jenis kanker payudara, guna melakukan pengobatan secara cepat dan tepat pada penderita penyakit kanker payudara..

## VI. REFERENSI

- Atthalla, I. N., Jovandy, A., & Habibie, H. (2018). Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor. *Prosiding Annual Research Seminar*, 4(1), 148–151.
- Azis, A. I. S., Irma Surya Kumala Idris, Budy Santoso, & Yasin Aril Mustofa. (2019). Pendekatan Machine Learning yang Efisien untuk Prediksi Kanker Payudara. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 458–469. <https://doi.org/10.29207/resti.v3i3.1347>
- Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39–43. <https://doi.org/10.33096/ijodas.v1i2.13>
- Choirunnasih. (2019). Jurnal ilmiah. *Jurnal Ilmiah*, 10(2), 1–94. <https://doi.org/10.33087/jiubj.v23i1.3348>
- Dikshit, P., Dey, B., Shukla, A., Singh, A., Chadha, T., & Sehgal, V. K. (2022). Prediction of Breast Cancer using Machine Learning Techniques. *ACM International Conference Proceeding Series, April 2019*, 382–387. <https://doi.org/10.1145/3549206.3549274>
- Fahrezi, A., Salam, F. N., Ibrahim, G. M., & Rahman, R. (2022). Pengujian Black Box Testing pada Aplikasi Inventori Barang Berbasis Web di PT . AINO Indonesia. *Jurnal Ilmu Komputer Dan Pendidikan*, 1(1), 1–5.
- Gunawan, M. I., Sugiarto, D., & Mardianto, I. (2020). Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic

- Regression. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(3), 280. <https://doi.org/10.26418/jp.v6i3.40718>
- Harafani, H.-. (2020). Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara. *Jurnal Infortech*, 1(2), 131–139. <https://doi.org/10.31294/infortech.v1i2.7398>
- Ismafillah, D., Rohana, T., & Cahyana, Y. (2023). Implementasi Model Support Vector Machine dan Logistic Regression Untuk Memprediksi Penyakit Stroke. *JURIKOM (Jurnal Riset Komputer)*, 10(1), 2407–389. <https://doi.org/10.30865/jurikom.v10i1.5478>
- Khadijah, K., & Kusumaningrum, R. (2019). Ensemble Classifier untuk Klasifikasi Kanker Payudara. *It Journal Research and Development*, 4(1), 61–71. [https://doi.org/10.25299/itjrd.2019.vol4\(1\).3540](https://doi.org/10.25299/itjrd.2019.vol4(1).3540)
- 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>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Ningtyas, D. F., & Setiyawati, N. (2021). Implementasi Flask Framework pada Pembangunan Aplikasi Purchasing Approval Request. *Jurnal Janitra Informatika Dan Sistem Informasi*, 1(1), 19–34. <https://doi.org/10.25008/janitra.v1i1.120>
- Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- Nurelasari, E. (2018). Penerapan Metode Metode Neural Network Berbasis Algoritma Genetika Untuk Prediksi Penyakit Kanker Payudara. *Speed – Sentra Penelitian Engineering Dan Edukasi*, 10(3), 1–11.
- Nurnawati, E. K. (2022). Penerapan Algoritma Decision Tree Untuk Memprediksi Kanker Payudara menggunakan Data Mining dan Machine Learning. *Jurnal Dinamika Informatika*, 11(2), 103–112.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(4), 1704–1708. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1351>
- Qadrini, L. (2023). *Undersampling dan K-Fold Random Forest Untuk Klasifikasi Kelas Tidak Seimbang*. 4(4), 1967–1974. <https://doi.org/10.47065/bits.v4i4.3141>
- Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, 2(1), 31–37. <https://t.co/NfhmfMjtXw>

- 
- Ramadhan, I., & Kurniawati, K. (2020). Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naive Bayes. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 21. <https://doi.org/10.30865/jurikom.v7i1.1755>
- Random, D., Untuk, F., Kanker, P., Widodo, A. M., Anwar, N., Irawan, B., Meria, L., Wisnujati, A., Komputer, F. I., Unggul, U., Unggul, E., Barat, J., Ekonomi, F., Unggul, U. E., & Barat, J. (2021). *Komparasi Performansi Algoritma Pengklasifikasi KNN, Bagging*. 367–372.
- Solikin, I. (2021). Teknik Data Mining untuk Prediksi Kanker Payudara yang Efisien. *FIDELITY : Jurnal Teknik Elektro*, 3(3), 63–67. <https://doi.org/10.52005/fidelity.v3i3.100>
- Sonjaya, C. B., Fitri, A., Masruriyah, N., & Sulistya, D. (2022). *The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease*. 5(2), 166–175.
- Tiana, E., & Wahyuni, S. (2020). Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(2), 130. <https://doi.org/10.30865/json.v1i2.1766>
- Widyaningsih, Y., Arum, G. P., & Prawira, K. (2021). Aplikasi K-Fold Cross Validation Dalam Penentuan Model Regresi Binomial Negatif Terbaik. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 15(2), 315–322. <https://doi.org/10.30598/barekengvol15iss2pp315-322>