

# Analisis Kinerja Algoritma Support Vector Machine dengan Pendekatan Oversampling pada Prediksi Kanker Payudara

Ahmad Rifa'i<sup>1</sup>, Marta Ardiyanto<sup>2</sup>

<sup>12</sup>Universitas Duta Bangsa Surakarta

<sup>12</sup>Surakarta, Indonesia

Email: <sup>1</sup>[ahmad\\_rifai@udb.ac.id](mailto:ahmad_rifai@udb.ac.id), <sup>2</sup>[marta.ardiyanto@udb.ac.id](mailto:marta.ardiyanto@udb.ac.id)

## Abstract

*Breast cancer is one of the leading causes of death among women worldwide, and early detection remains a major challenge. The imbalance between the amount of data in the healthy and cancer-positive classes reduces the ability of classification models to recognize minority cases. This study aims to improve the accuracy and balance of breast cancer classification model performance using the Support Vector Machine (SVM) algorithm with the application of oversampling techniques. The data used is a breast cancer diagnosis dataset with an imbalanced class proportion. The research stages include data pre-processing, oversampling application on training data, SVM model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The test results show that oversampling increases the model's accuracy from 95.91% to 98.83%, recall from 89.06% to 96.88%, and F1-score from 94.21% to 98.41%, while precision remains high at 100%. These findings confirm that the combination of SVM and oversampling is effective in addressing data imbalance and improving the generalization capabilities of machine learning-based breast cancer detection systems.*  
**Keywords:** Breast cancer, Support Vector Machine, Oversampling, Classification, Machine learning

## Abstraksi

*Kanker payudara merupakan salah satu penyebab utama kematian pada perempuan di dunia, dan ketepatan deteksi dini masih menjadi tantangan utama. Ketidakseimbangan jumlah data antara kelas sehat dan positif kanker menurunkan kemampuan model klasifikasi dalam mengenali kasus minoritas. Penelitian ini bertujuan meningkatkan akurasi dan keseimbangan performa model klasifikasi kanker payudara menggunakan algoritma Support Vector Machine (SVM) dengan penerapan teknik oversampling. Data yang digunakan merupakan dataset diagnosis kanker payudara dengan proporsi kelas tidak seimbang. Tahapan penelitian meliputi pra-pemrosesan data, penerapan oversampling pada data pelatihan, pelatihan model SVM, serta evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa oversampling meningkatkan akurasi model dari 95,91% menjadi 98,83%, recall dari 89,06% menjadi 96,88%, dan F1-score dari 94,21% menjadi 98,41%, sedangkan presisi tetap tinggi pada 100%. Temuan ini menegaskan bahwa kombinasi SVM dan oversampling efektif dalam mengatasi ketidakseimbangan data serta meningkatkan kemampuan generalisasi sistem deteksi kanker payudara berbasis pembelajaran mesin.*

**Kata Kunci:** Kanker payudara, Support Vector Machine, Oversampling, Klasifikasi, Machine learning

## **1. PENDAHULUAN**

Kanker payudara merupakan salah satu penyakit paling mematikan pada wanita, terjadi ketika sel-sel di jaringan payudara tumbuh tidak terkendali dan membentuk tumor ganas yang dapat menyebar ke bagian tubuh lain. Secara global, kanker ini menyumbang sekitar 25% dari seluruh kasus kanker dan 12% dari kasus baru pada wanita, menjadikannya jenis kanker paling umum pada perempuan [1][2]. Di Amerika Serikat saja, pada tahun 2024 diperkirakan terdapat lebih dari 310.720 kasus baru kanker payudara invasif [3]. Fakta tersebut menegaskan bahwa kanker payudara masih menjadi masalah kesehatan global yang serius dan memerlukan sistem deteksi dini yang lebih akurat [4].

Permasalahan utama dalam deteksi dini kanker payudara terletak pada ketepatan klasifikasi antara tumor jinak dan ganas. Kesalahan dalam diagnosis dapat menyebabkan keterlambatan penanganan medis yang berdampak fatal bagi pasien. Tantangan ini diperburuk oleh kondisi ketidakseimbangan kelas pada data medis, di mana jumlah data pasien dengan tumor jinak jauh lebih banyak dibandingkan dengan tumor ganas. Ketidakseimbangan ini menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan gagal mengenali kasus minoritas secara akurat.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan penerapan algoritma pembelajaran mesin (machine learning) guna meningkatkan akurasi deteksi kanker payudara. Salah satu algoritma yang terbukti efektif dalam klasifikasi data medis adalah Support Vector Machine (SVM), yang memiliki kemampuan generalisasi tinggi serta efisien dalam menangani data berdimensi besar [5]. Beberapa studi terkini juga menunjukkan bahwa SVM memberikan performa yang konsisten dengan tingkat kesalahan rendah pada berbagai dataset medis, termasuk deteksi kanker payudara [6]. Namun demikian, performa SVM dapat menurun ketika diterapkan pada data yang tidak seimbang.

Untuk mengatasi hal tersebut, penelitian ini menerapkan teknik Random Oversampling guna menyeimbangkan distribusi data sebelum dilakukan proses klasifikasi. Teknik ini bekerja dengan menggandakan secara acak sampel dari kelas minoritas agar proporsi data menjadi seimbang [7]. Melalui pendekatan ini, diharapkan model SVM dapat meningkatkan sensitivitas (recall) terhadap kasus kanker tanpa mengorbankan akurasi keseluruhan. Proses evaluasi model dilakukan menggunakan confusion matrix, yang memberikan gambaran menyeluruh mengenai kinerja klasifikasi dengan mengukur tingkat akurasi, presisi, dan recall, serta menilai kestabilan dan tingkat kesalahan klasifikasi (false positive dan false negative) [8].

Gap pada penelitian ini terletak pada belum adanya analisis mendalam terkait pengaruh penerapan teknik oversampling terhadap kinerja algoritma SVM dalam mendeteksi kanker payudara. Oleh karena itu, penelitian ini memberikan kontribusi dengan mengombinasikan oversampling dan SVM untuk menghasilkan model klasifikasi yang lebih akurat dan adil terhadap setiap kelas data.

Tujuan dari penelitian ini adalah untuk menganalisis pengaruh penerapan oversampling terhadap kinerja algoritma SVM pada dataset Breast Cancer. Evaluasi

dilakukan dengan membandingkan hasil akurasi, presisi, dan recall model sebelum dan sesudah penerapan oversampling. Hasil penelitian diharapkan mampu menunjukkan sejauh mana teknik oversampling dapat meningkatkan performa model SVM dalam mendeteksi kanker payudara secara lebih efektif, serta memberikan kontribusi bagi pengembangan sistem deteksi dini pada data medis dengan karakteristik tidak seimbang

## **2. TINJAUAN PUSTAKA**

### **2.1. Machine Learning**

Berbagai penelitian telah menggunakan pendekatan pembelajaran mesin (machine learning) untuk membantu diagnosis kanker payudara. Secara umum, algoritma seperti Support Vector Machine (SVM), Random Forest, Naïve Bayes, dan K-Nearest Neighbor (K-NN) sering digunakan untuk memprediksi kategori tumor ganas atau jinak. Namun, tantangan utama dalam penerapan model tersebut adalah ketidakseimbangan data (class imbalance) dan ukuran dataset yang besar, yang dapat mempengaruhi akurasi dan generalisasi model prediktif. Penelitian terdahulu menunjukkan bahwa performa model sangat dipengaruhi oleh proses pemilihan fitur, metode validasi, serta jenis algoritma yang digunakan [9].

### **2.2. Prediksi**

Prediksi kanker payudara menggunakan machine learning bertujuan mengklasifikasikan apakah suatu kasus bersifat ganas atau jinak secara lebih cepat dan akurat. Berbagai algoritma telah digunakan, seperti SVM, Random Forest, Naïve Bayes, dan K-NN. Penelitian [9] menunjukkan bahwa dengan pemilihan fitur yang tepat, SVM dan Random Forest mampu mencapai akurasi tinggi, sedangkan penelitian [10] menegaskan keunggulan SVM dengan akurasi tertinggi dibanding algoritma lain. Temuan ini menjadi dasar pemilihan SVM dalam penelitian ini karena efektif untuk prediksi pada data medis yang kompleks. Namun, masih terbatas penelitian yang mengombinasikan SVM dengan teknik penyeimbangan data dan optimasi parameter, sehingga penelitian ini berupaya mengisi gap tersebut dengan mengembangkan model prediksi berbasis SVM yang lebih stabil dan akurat.

### **2.3. Imbalanced Data**

Ketidakseimbangan data (imbalanced dataset) pada kasus kanker payudara menyebabkan model cenderung lebih akurat pada kelas mayoritas (jinak) dan kurang sensitif terhadap kelas minoritas (ganas). Untuk mengatasinya, penelitian ini menggunakan Random Oversampling, yaitu teknik yang menggandakan secara acak sampel pada kelas minoritas sehingga distribusi kedua kelas menjadi seimbang [7]. Dengan demikian, model SVM dapat belajar secara lebih proporsional dan meningkatkan kemampuan dalam mendeteksi kasus ganas.

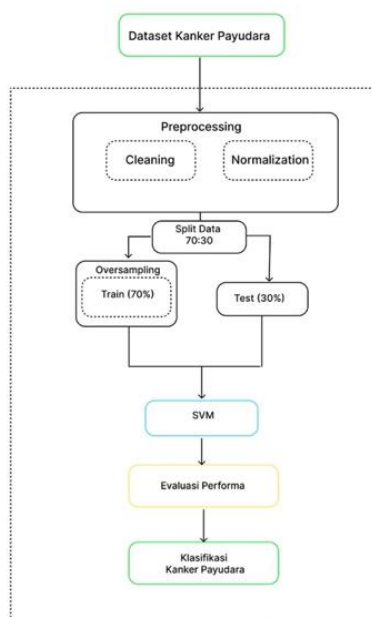
## **2.4. Support Vector Machine (SVM)**

Penelitian terdahulu menunjukkan bahwa SVM sebagai model machine learning telah digunakan untuk diagnosis kanker payudara dengan kombinasi fitur dan metode berbeda. Tantangan utama yang sering dihadapi adalah ukuran dataset yang besar serta ketidakseimbangan antara kelas ganas dan jinak. Studi [9] menemukan bahwa analisis fitur membantu pemilihan algoritma terbaik, di mana SVM dan Random Forest mencapai akurasi 96,5%. Penelitian [10] juga membandingkan beberapa algoritma—SVM, Random Forest, Naïve Bayes, dan K-NN—dan menunjukkan bahwa SVM memberikan akurasi tertinggi 97,9%, sehingga dianggap paling efektif untuk prediksi kanker payudara. Penelitian lain [11] menunjukkan K-NN unggul dengan akurasi 97,51%, sedangkan [12] menemukan Decision Tree memiliki akurasi tertinggi 96,89%. Sementara itu, studi [13] menunjukkan model DNN memiliki akurasi 92%, dengan Random Forest memberikan hasil terbaik pada kurva ROC. Secara umum, penelitian-penelitian tersebut menegaskan potensi tinggi algoritma pembelajaran mesin, terutama SVM, dalam diagnosis kanker payudara meskipun masih dihadapkan pada tantangan ketidakseimbangan data dan optimasi parameter model.

Berdasarkan analisis penelitian terdahulu, belum banyak studi yang secara khusus mengintegrasikan penanganan imbalanced dataset melalui teknik resampling dengan optimasi parameter SVM dalam satu kerangka model prediksi kanker payudara. Sebagian besar penelitian hanya menitikberatkan pada perbandingan akurasi antar-algoritma tanpa mengkaji secara mendalam pengaruh ketidakseimbangan kelas, efisiensi komputasi, serta kestabilan performa model pada data medis yang sesungguhnya. Oleh karena itu, masih diperlukan penelitian yang menghadirkan model SVM teroptimasi dengan strategi penyeimbangan data untuk menghasilkan sistem diagnosis yang lebih akurat, stabil, dan aplikatif secara klinis.

## **3. METODE PENELITIAN**

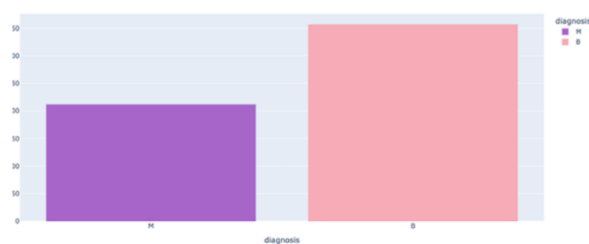
Metode penelitian yang digunakan dalam pengembangan model prediksi kanker payudara berbasis algoritma Support Vector Machine (SVM). Setiap tahap dirancang untuk memastikan data yang digunakan berkualitas baik, model terlatih secara optimal, dan hasil klasifikasi memiliki tingkat akurasi yang tinggi. Berikut penjelasan tiap proses pada alur penelitian yang digunakan dalam pengembangan model klasifikasi kanker payudara. Metode yang diusulkan dapat dilihat pada gambar 1.



Gambar 1. Metode yang diusulkan

### 3.1. Pengumpulan Data

Dalam proses pengumpulan data, para peneliti memperoleh dataset untuk studi ini dari sumber data publik yang tersedia di situs UCI machine learning [1], Kumpulan data kanker payudara ini terdiri dari total 569 data dengan 32 atribut. Kumpulan data ini mencakup 357 kasus kanker payudara jinak, di mana tumor bersifat tidak agresif, dan 212 kasus ganas, di mana tumor bersifat agresif dan dapat menyebar. Contoh kumpulan data ini ditampilkan pada Gambar 2. Dataset Kanker Payudara.



Gambar 2. Dataset Kanker Payudara

### 3.2. Preprocessing

Pada tahap preprocessing, data yang sudah diimpor akan dilakukan Analisa dan pengolahan dengan melakukan penghapusan fitur yang tidak diperlukan dan pembersihan data (cleaning data) agar data tidak terjadi duplikat pada saat penelitian. Kemudian dataset dari kanker payudara akan dilakukan dengan pengolahan yang berbeda

dengan menggunakan Over Sampling, nantinya akan dilakukan eksperimen dan akan dilakukan perbandingan.

### **3.3. Split Data**

Pembagian data dilakukan setelah tahap Random Oversampling pada data latih. Random Oversampling adalah metode penyeimbangan kelas dengan cara menggandakan secara acak sampel dari kelas minoritas (ganas) hingga proporsinya setara dengan kelas mayoritas (jinak). Teknik ini dipilih karena dataset kanker payudara umumnya imbalanced, sehingga tanpa penyeimbangan model cenderung bias ke kelas mayoritas dan lemah dalam mendeteksi kasus ganas. Setelah data latih seimbang, dataset dibagi menjadi dua subset: 70% sebagai data pelatihan (hasil oversampling) dan 30% sebagai data pengujian. Skema ini memastikan model belajar dari data yang representatif dan seimbang, sementara evaluasi dilakukan pada data asli agar performa model mencerminkan kondisi nyata.

### **3.4. Modelling**

Model klasifikasi pada data pelatihan dibangun menggunakan Support Vector Machine (SVM). SVM bekerja dengan mencari hyperplane pemisah terbaik antara kelas jinak dan ganas dengan margin maksimum, sehingga mampu menghasilkan klasifikasi yang akurat. Jika data tidak terpisah secara linear, SVM dapat memakai fungsi kernel (mis. RBF) untuk memetakan data ke ruang lebih tinggi agar pemisahan lebih optimal. SVM dipilih karena stabil pada data berdimensi tinggi, memiliki generalisasi baik, dan efektif untuk klasifikasi medis.

Selain itu, ditampilkan pula konfigurasi parameter yang digunakan dalam model Support Vector Machine (SVM) sebagai model dengan kinerja terbaik pada penelitian ini, Sebagai berikut:

Tabel 1. Parameter Model Support Vector Machine (SVM)

Parameter	Nilai
Kernel	rbf
C	01.00
gamma	scale
random_state	42

Pada Tabel 1, Model Support Vector Machine (SVM) dengan parameter kernel='rbf' menunjukkan bahwa model menggunakan Radial Basis Function, yang efektif untuk memisahkan data non-linear. Nilai C=1.0 berfungsi sebagai parameter regularisasi untuk mengontrol keseimbangan antara meminimalkan kesalahan pelatihan dan kompleksitas model, semakin besar nilainya, semakin sedikit kesalahan yang diizinkan tetapi berisiko overfitting. Parameter gamma='scale' secara otomatis menyesuaikan pengaruh setiap titik data terhadap batas keputusan. Sementara itu, random\_state=42 digunakan untuk menjaga hasil pelatihan agar reproducible atau konsisten saat kode dijalankan ulang.

### **3.5. Evaluasi Performa**

Akurasi model klasifikasi yang dihasilkan oleh algoritma pembelajaran mesin dievaluasi dan dibandingkan dengan confusion matrix. Tujuannya adalah mengidentifikasi algoritma yang menghasilkan akurasi atau mengoptimalkan akurasi dibandingkan dengan metode sebelumnya. Langkah ini melibatkan penyetelan model, penyesuaian hiperparameter, dan pemilihan pendekatan yang paling efektif berdasarkan metrik evaluasi. Hasil penelitian menghasilkan prediksi prediksi kanker payudara yang efisien dan akurat.

Accuracy merupakan metrik yang digunakan untuk mengevaluasi seberapa baik model memprediksi output yang benar berdasarkan data input, dengan rumus perhitungan seperti pada Persamaan (1):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Precision merupakan metrik evaluasi yang mengukur proporsi prediksi positif yang benar dari seluruh prediksi positif yang dibuat oleh model, dengan rumus perhitungan seperti pada Persamaan (2):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan metrik evaluasi yang menilai proporsi data positif aktual yang berhasil diidentifikasi dengan benar oleh model dibandingkan total data positif aktual dalam dataset, dengan rumus perhitungan seperti pada Persamaan (3):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1-Score adalah metrik evaluasi berupa rata-rata harmonik precision dan recall yang memberikan ukuran performa seimbang, terutama pada dataset tidak seimbang, dengan rumus perhitungan seperti pada Persamaan (4):

$$\text{F1 - Score} = 2 \times \frac{TP}{2TP + FP + FN} \quad (4)$$

Setelah evaluasi model dilakukan, kinerja model dapat diamati dengan jelas. Evaluasi ini menentukan apakah model mampu mengidentifikasi kanker payudara secara akurat atau tidak.

## **4. HASIL DAN PEMBAHASAN**

### **4.1. Hasil Penelitian**

Bagian ini menyajikan hasil penelitian yang diperoleh dari proses implementasi dan pengujian model klasifikasi kanker payudara menggunakan algoritma pembelajaran

mesin. Hasil yang ditampilkan mencakup performa masing-masing model berdasarkan nilai akurasi, presisi, recall dan F1-Score.

Berikut disajikan hasil implementasi model Support Vector Machine (SVM) yang telah dilakukan pada data kanker payudara, baik dengan penerapan teknik oversampling maupun tanpa oversampling, untuk melihat pengaruh keseimbangan data terhadap performa klasifikasi.

Tabel 2. Hasil Evaluasi Performa Model SVM Sebelum dan Sesudah Oversampling

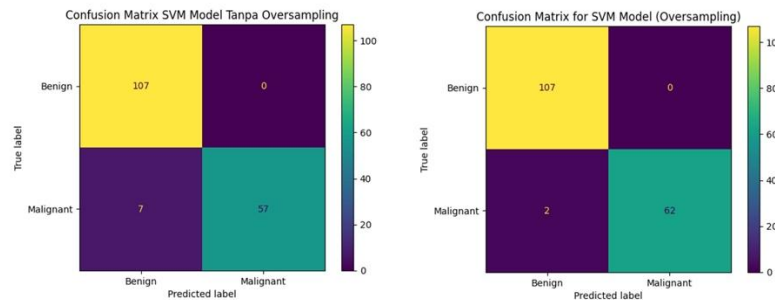
Evaluasi	SVM	SVM + Oversampling
Akurasi Training	98.49	98.4
Akurasi Testing	95.91	98.83
Presisi	100	100
Recall	89.06	96.88
F1-Score	94.21	98.41

Berdasarkan hasil pengujian yang ditampilkan pada Tabel 2, Terlihat bahwa oversampling mengubah perilaku belajar SVM secara nyata. Sebelum data diseimbangkan, model berlatih pada distribusi yang didominasi kelas jinak sehingga cenderung memilih prediksi mayoritas karena itu paling aman, akibatnya banyak kasus ganas luput terdeteksi dan recall berada pada tingkat yang lebih rendah. Setelah oversampling, jumlah contoh ganas pada data latih meningkat sehingga SVM memperoleh representasi pola kelas minoritas yang lebih kaya. Dampaknya, batas keputusan SVM menjadi lebih adil dan tidak lagi berat ke kelas mayoritas, sehingga kemampuan menangkap kasus ganas meningkat drastis, hal ini terlihat dari recall yang naik dari 89,06% menjadi 96,88%.

Kenaikan akurasi pada data uji dari 95,91% menjadi 98,83%, menunjukkan bahwa model kini lebih mampu mengenali kedua kelas secara seimbang pada data baru, bukan hanya unggul di kelas mayoritas. Sementara itu, penurunan kecil akurasi pada data latih dari 98,49% menjadi 98,40%, justru masuk akal karena model tidak lagi terlalu menyesuaikan diri dengan dominasi kelas jinak di data pelatihan, sehingga indikasi overfitting berkurang dan model menjadi lebih robust. Presisi yang tetap 100% menandakan bahwa peningkatan sensitivitas terhadap kasus ganas tidak diikuti peningkatan false positive, sehingga F1-score ikut terdorong naik dari 94,21% menjadi 98,41%. Secara keseluruhan, oversampling efektif mengurangi bias akibat imbalanced dataset dan membuat performa SVM lebih stabil serta relevan untuk deteksi kanker payudara.

Berikut disajikan Hasil dari Confusion Matrix model Support Vector Machine (SVM) sebelum dan sesudah penerapan teknik oversampling:

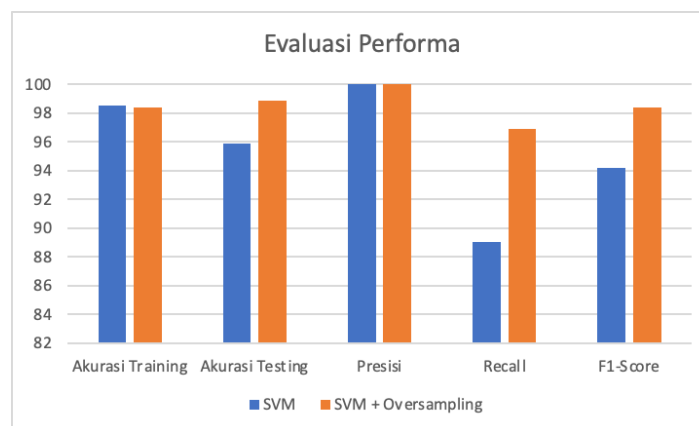




Gambar 3. Confusion Matrix Model SVM Sebelum dan Sesudah Oversampling

Pada Gambar 3, confusion matrix menunjukkan perbedaan signifikan antara model SVM tanpa oversampling dan dengan oversampling, khususnya pada nilai False Negative (FN). Sebelum oversampling, terdapat 7 kasus kanker ganas yang salah diklasifikasikan sebagai jinak, sedangkan setelah oversampling jumlah FN menurun menjadi hanya 2. Penurunan ini sangat penting dalam konteks deteksi kanker, karena FN berarti pasien yang sebenarnya mengidap kanker justru teridentifikasi sehat, yang bisa berakibat fatal secara klinis.

Berikut disajikan grafik dan evaluasi performa model Support Vector Machine (SVM) sebelum dan sesudah penerapan teknik oversampling, yang bertujuan untuk menunjukkan pengaruh penyeimbangan data terhadap peningkatan akurasi dan kemampuan klasifikasi model.



Gambar 4. Evaluasi Performa Model SVM Sebelum dan Sesudah Oversampling

Pada Gambar 4, Grafik tersebut menunjukkan hasil evaluasi performa model SVM sebelum dan sesudah oversampling. Terlihat bahwa SVM dengan oversampling mengalami peningkatan performa pada sebagian besar metrik, terutama akurasi testing, recall, dan F1-score. Akurasi training dan presisi relatif stabil di angka tinggi sekitar 98% - 100%, namun recall meningkat signifikan setelah oversampling, menunjukkan bahwa model menjadi lebih baik dalam mengenali kelas minoritas. Secara keseluruhan, teknik oversampling berhasil membuat model lebih seimbang dan meningkatkan generalisasi tanpa mengorbankan akurasi.

## 4.2. Pembahasan

Hasil evaluasi performa menunjukkan bahwa penerapan teknik Oversampling membantu SVM karena pada data imbalanced hyperplane SVM terdorong ke arah kelas mayoritas sehingga kelas minoritas kurang terwakili dan recall rendah; dengan menambah sampel minoritas, distribusi jadi lebih seimbang, margin pemisah lebih representatif, dan recall serta F1-Score naik tanpa mengorbankan akurasi berarti. Ini krusial untuk data medis yang sering imbalanced dan menuntut sensitivitas tinggi agar kasus sakit tidak terlewat, sementara SVM sendiri cocok di medis karena kuat pada data kecil berdimensi tinggi, mampu memodelkan batas non-linear lewat kernel, dan relatif robust serta lebih mudah ditelusuri lewat support vectors. Bukti penguat dari literatur juga menunjukkan SVM banyak dipakai di kesehatan dan performanya meningkat ketika ketidakseimbangan kelas ditangani dengan teknik oversampling/penyeimbangan, terutama untuk menaikkan sensitivitas deteksi pasien [14].

## 5. KESIMPULAN

Berdasarkan hasil penelitian, penerapan oversampling pada model SVM terbukti meningkatkan performa klasifikasi kanker payudara secara signifikan. Setelah oversampling diterapkan, akurasi meningkat dari 95,91% menjadi 98,83%, recall dari 89,06% menjadi 96,88%, dan F1-score dari 94,21% menjadi 98,41%, dengan presisi tetap stabil pada 100%. Peningkatan recall dan F1-score menegaskan bahwa penyeimbangan data memperbaiki kemampuan model dalam mengenali kelas minoritas tanpa menurunkan kinerja keseluruhan. Dengan demikian, kontribusi utama penelitian ini adalah menunjukkan bahwa oversampling merupakan langkah efektif untuk mengatasi ketidakseimbangan kelas pada klasifikasi kanker payudara berbasis SVM dan menghasilkan model yang lebih sensitif serta konsisten pada data uji.

## DAFTAR PUSTAKA

- [1] K. Varshini, R. K. Sethuramamoorthy, V. Kumar, S. A. Shree, and S. Deivarani, "Breast cancer prediction using machine learning techniques," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6 Special Issue, pp. 2026–2032, 2020, doi: 10.5120/ijca2022922490.
- [2] M. A. Rahman *et al.*, "Advancements in Breast Cancer Detection: A Review of Global Trends, Risk Factors, Imaging Modalities, Machine Learning, and Deep Learning Approaches," *BioMedInformatics*, vol. 5, no. 3, p. 46, Aug. 2025, doi: 10.3390/biomedinformatics5030046.
- [3] A. N. Giaquinto *et al.*, "Breast cancer statistics 2024," *CA. Cancer J. Clin.*, vol. 74, no. 6, pp. 477–495, Nov. 2024, doi: 10.3322/caac.21863.
- [4] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020, doi: 10.1109/ACCESS.2020.3016715.
- [5] A. Bilal, A. Imran, T. I. Baig, X. Liu, E. Abouel Nasr, and H. Long, "Breast cancer diagnosis using support vector machine optimized by improved quantum inspired

- grey wolf optimization,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–25, 2024, doi: 10.1038/s41598-024-61322-w.
- [6] R. Shafique *et al.*, “Breast Cancer Prediction Using Fine Needle Aspiration Features and Upsampling with Supervised Machine Learning,” *Cancers (Basel)*, vol. 15, no. 3, p. 681, Jan. 2023, doi: 10.3390/cancers15030681.
- [7] T. Yulian and E. R. Susanto, “Sistemasi: Jurnal Sistem Informasi Analisis Perbandingan Teknik Oversampling dan SMOTEENN pada Algoritma Machine Learning untuk Prediksi Penyakit Kanker Payudara Comparative Analysis of Oversampling and SMOTEENN Techniques in Machine Learning Algorithms f,” vol. 14, pp. 1318–1331, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>.
- [8] E. S. Septiany, H. H. Handayani, T. Al Mudzakir, and A. F. N. Masruriyah, “Optimasi Metode Support Vector Machine Menggunakan Seleksi Fitur Recursive Feature Elimination dan Forward Selection untuk Klasifikasi Kanker Payudara,” *TIN Terap. Inform. Nusantara*, vol. 5, no. 2, pp. 144–154, 2024, doi: 10.47065/tin.v5i2.5324.
- [9] S. Ara, A. Das, and A. Dey, “Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms,” *2021 Int. Conf. Artif. Intell. ICAI 2021*, pp. 97–101, 2021, doi: 10.1109/ICAI52203.2021.9445249.
- [10] Y. Khourdifi and M. Bahaj, “Applying best machine learning algorithms for breast cancer prediction and classification,” *2018 Int. Conf. Electron. Control. Optim. Comput. Sci. ICECOCS 2018*, pp. 1–5, 2019, doi: 10.1109/ICECOCS.2018.8610632.
- [11] Y. Tewari, E. Ujjwal, and L. Kumar, “Breast Cancer Classification Using Machine Learning,” *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022*, pp. 168–171, 2022, doi: 10.1109/ICACITE53722.2022.9823932.
- [12] F. S. Fatima, A. Jaiswal, and N. Sachdeva, “Lung Cancer Detection Using Machine Learning Techniques,” *Crit. Rev. Biomed. Eng.*, vol. 50, no. 6, pp. 45–58, 2022, doi: 10.1615/critrevbiomedeng.v50.i6.40.
- [13] F. Teixeira, J. L. Z. Montenegro, C. A. Da Costa, and R. Da Rosa Righi, “An analysis of machine learning classifiers in breast cancer diagnosis,” *Proc. - 2019 45th Lat. Am. Comput. Conf. CLEI 2019*, 2019, doi: 10.1109/CLEI47609.2019.235094.
- [14] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, “An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review,” *Information*, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235.