IMPLEMENTASI *SUPPORT VECTOR MACHINE* DENGAN *RANDOM OVERSAMPLING* UNTUK MENGATASI DATA TAK SEIMBANG PADA KLASIFIKASI PENDERITA PENYAKIT *CARDIOVASCULAR*

(Skripsi)

Oleh

CANTIKA MERITA
NPM 2117031112

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

ABSTRACT

IMPLEMENTATION OF SUPPORT VECTOR MACHINE WITH RANDOM OVERSAMPLING TO OVERCOME IMBALANCE DATA IN CLASSIFYING PATIENTS WITH CARDIOVASCULAR DISEASE

BY

CANTIKA MERITA

Support Vector Machine (SVM) is one of the machine learning methods used for classification by dividing data into two different classes. The working principle of SVM is to find the best separating function (hyperplane). If the data cannot be separated linearly, then the data is nonlinear. One method to overcome this is by using a kernel function. The objective of this study is to apply the SVM method to determine the best kernel function based on the highest accuracy value for classifying patients with cardiovascular disease. However, the dataset used in this study has an imbalance data problem. Therefore, Random Oversampling (ROS) was used to address this issue. The results of the study indicate that the best kernel function for balanced data is the Radial Basis Function (RBF) function, with a gamma parameter of 0,1 and a cost of 1 in a 90% training and 10% testing data scheme, yielding an accuracy value of 73,86%.

Keywords : Cardiovascular; Kernel Radial Basis Function; Imbalance Data; Random Oversampling; Support Vector Machine.

ABSTRAK

IMPLEMENTASI *SUPPORT VECTOR MACHINE* DENGAN *RANDOM OVERSAMPLING* UNTUK MENGATASI DATA TAK SEIMBANG PADA KLASIFIKASI PENDERITA PENYAKIT *CARDIOVASCULAR*

OLEH

CANTIKA MERITA

Support Vector Machine (SVM) adalah salah satu metode *machine learning* yang digunakan untuk pengklasifikasikan dengan membagi data menjadi dua kelas yang berbeda. Prinsip kerja SVM adalah mencari fungsi pemisah (*hyperplane*) yang terbaik. Apabila data tidak dapat dipisahkan secara linear maka data tersebut merupakan data nonlinear. Salah satu metode untuk mengatasi hal tersebut adalah dengan menggunakan fungsi *kernel*. Tujuan penelitian ini adalah menerapkan metode SVM untuk mengetahui kinerja fungsi *kernel* terbaik berdasarkan nilai akurasi tertinggi terhadap klasifikasi penderita penyakit *cardiovascular*. Namun, pada dataset yang digunakan dalam penelitian memiliki masalah ketidakseimbangan data (*imbalance data*). Oleh karena itu, digunakan *Random Oversampling* (ROS) untuk mengatasi masalah tersebut. Hasil penelitian menunjukkan bahwa fungsi *kernel* terbaik pada data yang seimbang adalah fungsi *Radial Basis Function* (RBF), dengan parameter *gamma* 0,1 dan *cost* 1 pada skema data *training* 90% dan *testing* 10% didapat nilai akurasi sebesar 73,86%.

Kata Kunci : *Cardiovascular*; *Kernel Radial Basis Function*; Ketidakseimbangan Data; *Random Oversampling*; *Support Vector Machine*.

IMPLEMENTASI *SUPPORT VECTOR MACHINE* DENGAN *RANDOM OVERSAMPLING* UNTUK MENGATASI DATA TAK SEIMBANG PADA KLASIFIKASI PENDERITA PENYAKIT *CARDIOVASCULAR*

CANTIKA MERITA

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

Judul Skripsi	:	**IMPLEMENTASI *SUPPORT VECTOR MACHINE* DENGAN *RANDOM OVERSAMPLING* UNTUK MENGATASI DATA TAK SEIMBANG PADA KLASIFIKASI PENDERITA PENYAKIT *CARDIOVASCULAR***
Nama Mahasiswa	:	Cantika Merita
Nomor Pokok Mahasiswa	:	**2117031112**
Program Studi	:	**Matematika**
Fakultas	:	**Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Subian Saidi, S.Si., M.Si.
NIP 198008212008121001

Misgiyati, S.Pd., M.Si.
NIP 198509282023212032

2. Ketua Jurusan Matematika

Dr.Aang Nuryaman, S.Si., M.Si.
NIP 197403162005011001

MENGESAHKAN

1. Tim penguji

   Ketua : **Dr. Subian Saidi, S.Si., M.Si.** .....

   Sekretaris : **Misgiyati, S.Pd., M.Si.** .....

   Penguji
   Bukan Pembimbing : **Dr. Khoirin Nisa, S.Si., M.Si.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP 197110012005011002

Tanggal Lulus Ujian Skripsi: **26 Mei 2025**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:.

Nama	:	**Cantika Merita**
Nomor Pokok Mahasiswa	:	**2117031112**
Jurusan	:	**Matematika**
Judul Skripsi	:	**Implementasi *Support Vector Machine* dengan *Random Oversampling* untuk Mengatasi Data Tak Seimbang pada Klasifikasi Penderita Penyakit *Cardiovascular***

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 4 Juni 2025

Penulis,

Cantika Merita

RIWAYAT HIDUP

Penulis bernama Cantika Merita lahir di Kota Liwa, Kabupaten Lampung Barat pada tanggal 29 Maret 2003. Penulis merupakan anak pertama dari empat bersaudara dari pasangan Burnawan dan Lena Asmara.

Penulis mengawali pendidikan di Taman Kanak-kanak (TK) Dharma Wanita pada tahun 2009-2010. Kemudian menempuh pendidikan Sekolah Dasar (SD) di SDN 2 Hanakau tahun 2009-2015. Melanjutkan ke Sekolah Menengah Pertama (SMP) di SMPN 2 Liwa dan lulus pada tahun 2018. Kemudian penulis melanjutkan pendidikan Sekolah Menengah Atas (SMA) di SMAN 2 Liwa dan lulus pada tahun 2021. Pada tahun 2021, penulis berhasil lulus Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) dan diterima di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

Selama menjadi mahasiswa, penulis aktif dalam organisasi Badan Eksekutif Mahasiswa (BEM) FMIPA dan organisasi Rohani Islam (Rohis) FMIPA Unila. Selain itu, penulis juga aktif dalam berbagai kegiatan yang mendukung pengembangan diri, termasuk mengikuti seminar dan workshop yang berkaitan dengan ilmu pengetahuan dan teknologi. Pada tahun 2024, penulis melakukan Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Kabupaten Lampung Barat dan Kuliah Kerja Nyata (KKN) di Desa Labuhan Ratu, Kecamatan Lampung Timur, Lampung. Dengan latar belakang pendidikan dan pengalaman tersebut penulis berharap penelitian ini dapat memberikan kontribusi yang berarti, khususnya di bidang matematika terutama dalam ranah statistika.

KATA INSPIRASI

"Dan janganlah engkau berjalan di bumi ini dengan sombong, karena sesungguhnya engkau tidak akan dapat menembus bumi dan tidak akan mampu menjulang setinggi gunung."
(QS. Al-Isra: 37)

"Dan barang siapa yang bertakwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya."
(Q.S At-Talaq: 4)

"*The greatest glory in living lies not in never falling, but in rising every time we fall*."
(Nelson Mandela)

PERSEMBAHAN

Dengan mengucapkan Alhamdulillah dan penuh rasa syukur kepada Allah SWT atas segala nikmat dan petunjuk-Nya, saya dapat menyelesaikan skripsi ini dengan baik dan sesuai dengan waktu yang ditentukan. Dengan hati yang penuh rasa syukur dan kebahagiaan, saya ingin menyampaikan terima kasih saya kepada:

Bapak dan Ibu Tercinta

Terima kasih yang sebesar-besarnya kepada orang tuaku atas segala pengorbanan, motivasi, doa, ridho, dan dukungannya yang tiada henti selama ini. Terima kasih telah mengajarkan pelajaran hidup yang sangat berharga, yang membuatku memahami makna sejati dari perjalanan hidup, sehingga kelak dapat menjadi pribadi yang bermanfaat bagi banyak orang.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sangat berjasa, selalu membimbing, memberikan arahan, dan juga ilmu yang sangat bermanfaat saat proses pembuatan skripsi ini

Seluruh Keluarga dan Sahabat-sahabatku

Terimakasih kepada semua keluarga dan orang-orang baik yang telah memberikan pengalaman, dorongan, inspirasi, dan doa-doanya, serta yang senantiasa mendukung dalam situasi apa pun.

Almamater Tercinta

Universitas Lampung

SANWACANA

Alhamdulillah, puji dan syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan tepat waktu dengan judul " Implementasi *Support Vector Machine* dengan *Random Oversampling* untuk Mengatasi Data Tak Seimbang pada Klasifikasi Penderita Penyakit *Cardiovascular*".

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan dengan baik tanpa adanya arahan, bimbingan, serta kritik dan saran dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Bapak Dr. Subian Saidi, S.Si., M.Si. selaku Pembimbing 1 yang telah dengan penuh perhatian memberikan arahan, bimbingan, serta dukungan moral dan motivasi yang sangat membantu penulis dalam menyelesaikan skripsi ini.
2. Ibu Misgiyati, S.Pd., M.Si. selaku Pembimbing II yang telah memberikan arahan, bimbingan, serta dukungan yang sangat berharga, sehing- ga penulis dapat menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si. selaku Penguji yang telah memberikan kritik, saran, dan evaluasi yang sangat membantu penulis untuk memperbaiki skripsi ini.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D. selaku dosen pembimbing akademik.
6. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Cinta pertama dan panutanku, Bapak Burnawan dan Ibu Lena Asmara. Terima kasih atas segala pengorbanan dan tulus kasih yang di berikan. Bapak dan ibu memang tidak sempat merasakan pendidikan bangku perkuliahan,

namun mereka mampu memberikan yang terbaik, tak kenal lelah mendoakan serta memberikan perhatian dan dukungan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana. Semoga bapak dan ibu sehat selalu, panjang umur dan senantiasa bahagia.

8. Adik-adik tercinta Chandra, Naufal, dan Nayla dan Seluruh keluargaku atas segala dukungan, semangat, dan doa yang tiada henti.
9. Sahabat satu bimbingan, Dea Ayu Setiawati atas segala dukungan, semangat, dan kerja sama yang telah terjalin selama proses menyelesaikan skripsi.
10. Lisa Andriyani, Dinda Meilani, Aulia adilah, Yulina Putri, dan Rani Tias yang telah menyemangati dan menemani penulis selama proses perkuliahan.
11. Cantika Merita, ya diri saya sendiri. Apresiasi sebesar-besarnya karena telah bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terima kasih karena sudah bertahan dan terus berusaha untuk tidak menyerah, serta senantiasa menikmati setiap prosesnya yang bisa dibilang tidak mudah.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat banyak kekurangan baik dalam penyajian maupun penulisan. Oleh sebab itu, saran dan kritikan yang membangun senantiasa penulis harapkan demi menyempurnakan skripsi ini.

Bandar Lampung, 4 Juni 2025
Penulis,

Cantika Merita

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Klasifikasi merupakan salah satu teknik yang sering digunakan dalam data *mining* dan *machine learning*. Klasifikasi adalah proses untuk menemukan sebuah model yang dapat menjelaskan dan membedakan konsep atau kategori data, dengan tujuan memprediksi kategori suatu objek yang belum diketahui kategorinya (Susanti, *et al*., 2023). Klasifikasi bertujuan untuk menemukan aturan keputusan yang mampu memprediksi kategori data uji yang berasal dari distribusi yang mirip dengan data *training*. Salah satu metode yang digunakan dalam klasifikasi adalah *Support Vector Machine* (SVM).

SVM adalah metode pembelajaran *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space* (Cindy, *et al*., 2024). *Hyperplane* merupakan pemisah terbaik yang ditentukan dengan mengukur margin *hyperplane* dan menemukan titik maksimumnya. Margin adalah jarak antara *hyperplane* dengan data terdekat dari dua kelas yang berbeda, yang dikenal sebagai *support vector* (Umar, *et al*., 2020). Permasalahan yang sering kali ditemui pada saat melakukan klasifikasi adalah ketidakseimbangan pada data (*imbalance* data). Ketidakseimbangan dalam data ini dapat mengurangi prediksi pengambilan keputusan (Fitriani & Febrianto, 2021).

Imbalance data adalah kondisi ketika terjadi perbedaan antara jumlah data yang signifikan pada kelas mayoritas dan kelas minoritas. Perbedaan rasio data membuat *classifier* mengambil keputusan yang keliru pada saat klasifikasi yang cenderung lebih memilih kelas mayoritas dan mengabaikan kelas minoritas (Sir & Soepranoto, 2022). Kelas yang memiliki jumlah data lebih banyak disebut dengan kelas mayoritas dan kelas yang memiliki jumlah data lebih sedikit disebut dengan kelas minoritas (Akbar & Hayaty, 2020). Ketidakseimbangan pada data dapat memberikan dampak yang buruk pada hasil klasifikasinya, yaitu kelas minoritas sering kali salah diklasifikasikan sebagai kelas mayoritas (Siringoringo & Jaya, 2018). Dalam melakukan klasifikasi kita perlu melakukan *hyperparameter tuning* dengan *grid search* untuk membantu dalam menentukan parameter terbaik untuk memperoleh akurasi yang tinggi.

Hyperparameter tuning adalah proses mencari nilai parameter optimal atau parameter terbaik. *Hyperparameter tuning* berfungsi untuk membantu model mencari nilai parameter yang tepat pada dataset agar mendapatkan hasil kinerja yang maksimal (Andini, *et al.*, 2022). Salah satu metode *hyperparameter tuning* yang paling sederhana adalah *grid search*, cara kerja *grid search* yaitu dengan melakukan pencarian secara menyeluruh pada semua kombinasi *hyperparameter* yang ditentukan pada *grid konfigurasi* (Dana, *et al.*, 2024). Kelebihan *grid search* adalah kemampuannya untuk dijalankan secara paralel, sehingga setiap percobaan berjalan secara mandiri tanpa dipengaruhi oleh urutan waktu (Nurcahyo & Sasongko, 2023).

Dalam pengembangan model klasifikasi menggunakan SVM, fungsi *kernel* memiliki peran penting karena memungkinkan pemetaan dataset ke ruang dimensi yang lebih tinggi, sehingga membantu dalam memperoleh interpretasi yang lebih baik terhadap model klasifikasi (Nugroho, 2024). Menurut (Wahyuni & Kusumodestoni, 2024) algoritma SVM memiliki konsep *kernel* yang membedakan antara SVM linear dan SVM non linear. SVM linear bekerja pada data yang dapat dipisahkan secara linier, menggunakan *kernel* linier untuk membangun *hyperplane* yang efektif dalam memisahkan dua kelas di ruang fitur

berdimensi tinggi, sehingga memungkinkan klasifikasi yang akurat. Sementara itu SVM non linear digunakan untuk data yang tidak dapat dipisahkan secara linier, dengan menerapkan pendekatan *kernel* untuk memudahkan dalam pemetaan data. Beberapa *kernel* yang umum digunakan dalam SVM meliputi *kernel* polinomial, *radial basis function* (RBF), sigmoid, dan *kernel* linier.

Penyakit *Cardiovascular* (CVD), atau yang dikenal sebagai penyakit jantung, merupakan salah satu kondisi kronis yang menjadi penyebab utama kematian di seluruh dunia. Pada tahun 2017, penyakit *cardiovascular* telah menyebabkan lebih dari 17 juta kematian, hilangnya 330 juta tahun kehidupan, dan 35,6 juta tahun hidup dengan kecacatan di seluruh dunia (Roth, *et al.,* 2018). Prediksi penyakit *cardivascular* telah menjadi fokus penting dalam penelitian medis mengingat dampak signifikan yang ditimbulkannya. Upaya untuk memprediksi penyakit ini secara dini bertujuan mengembangkan model yang efektif dalam menilai risiko individu terhadap penyakit *cardiovascular*. Dengan demikian, langkah ini memungkinkan evaluasi awal serta perencanaan tindakan pencegahan yang tepat untuk mengurangi dampak penyakit *cardiovascular.* Permasalahan tersebut dapat diatasi dengan proses data *mining* yaitu klasifikasi.

Berbagai penelitian terdahulu mengenai penggunaan SVM pada kasus *imbalance* data sudah banyak dilakukan, salah satunya penelitian Purnajaya, *et al*., (2024) melakukan penelitian mengenai penggunaan SVM untuk identifikasi *subtype* kanker tiroid dengan pendekatan *oversampling* dan *under sampling* dalam meningkatkan akurasi, sensitivitas, dan spesifisitas. Hasilnya menunjukkan bahwa kinerja klasifikasi yang tinggi dengan data sampel yang memiliki sensitivitas yang unggul dengan perbedaan akurasi data tidak seimbang sebesar 85% dan akurasi data seimbang sebesar 87%. Penelitian lainnya dilakukan Mutmainah (2021), dengan membandingkan *Random Oversampling* (ROS) dan *Random Undersampling* (RUS) pada kemungkinan penyakit stroke. Penanganan *imbalance* data dilakukan antar *class* 1 (stroke) dan *class* 0 (tidak stroke) dengan distribusi antar data yang sama. Hasil yang didapatkan pada penggunaan teknik

ROS mendapat performa yang lebih tinggi yaitu 95% dari pada teknik RUS yang mendapat performa 76%.

Penelitian lainnya dilakukan oleh Cindy, *et al*., (2024), melakukan klasifikasi kasus *monkeypox* dengan pendekatan *oversampling* dan *undersampling* untuk mengatasi ketidakseimbangan kelas dengan metode *support vector machine*. Peneliti melakukan evaluasi melalui *confusion matrix* menilai akurasi, sensitivitas, spesifisitas, dan AUC. Di mana akurasi rata-rata mencapai 67,1% untuk data yang tidak seimbang dan 36,5% untuk data yang seimbang dan spesifisitas untuk data tidak seimbang lebih rendah 0,2 dari pada data seimbang 0,4. Namun, nilai AUC tetap sedikit lebih tinggi untuk data tidak seimbang 0,6 dari pada data seimbang 0,4 yang menunjukkan kemampuan SVM untuk membedakan antara kelas meskipun adanya kelas tidak seimbang. Hasil penelitian menunjukkan akurasi yang lebih tinggi dalam mendiagnosis *monkeypox* menggunakan SVM, meskipun terdapat ketidakseimbangan kelas.

Berdasarkan penjabaran di atas, maka dalam penelitian ini akan dikaji tentang implementasi *Support Vector Machine* dengan *Random Oversampling* untuk mengatasi ketidakseimbangan pada klasifikasi penderita penyakit *cardiovascular*.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain:

1. Mengatasi ketidakseimbangan data pada penderita penyakit c*ardiovascular* dengan menggunakan ROS.
2. Menerapkan metode SVM untuk mengetahui kinerja fungsi *kernel* terbaik berdasarkan nilai akurasi tertinggi terhadap data klasifikasi penderita penyakit c*ardiovascular.*

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memberikan pemahaman bagaimana cara mengatasi ketidakseimbangan data dengan ROS dan melakukan klasifikasi dengan SVM.
2. Penelitian ini diharapkan dapat menjadi referensi atau landasan bagi penelitian yang akan dilakukan pada masa mendatang.

II. TINJAUAN PUSTAKA

2.1 Data *Mining*

Data *mining* adalah proses eksplorasi, analisis, dan filtrasi data yang besar guna memperoleh hubungan baru yang mempunyai arti, pola, dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika (Mardi, 2017). Konsep dasar data *mining* adalah menentukan informasi tersembunyi dalam sebuah basis data dan merupakan bagian dari *Knowledge Discovery in Database* (KDD) untuk menemukan informasi dan pola yang berguna dalam data (Tarigan, *et al*., 2024).

Berikut pengelompokan data *mining* (Ginting & Simanjuntak, 2021) :

1. Deskripsi (*Description*);
2. Estimasi (*Estimation*);
3. Prediksi (*Prediction*);
4. Klasifikasi (*Classification*);
5. Pengelompokan (*Clustering*); dan
6. Asosiasi (*Association*).

2.2 Tahapan Data *Mining*

Data *mining* memiliki tahapan-tahapan utama , yaitu sebagai berikut (Urva, *et al.*, 2023) :

1. Data *Integration*, dilakukan untuk membuat sekumpulan data target yang berasal dari data asli, dengan memilih variabel dan sampel data tersebut.
2. Data *selection*, merupakan kumpulan database operasional yang dipilih dan diseleksi berdasarkan kebutuhan sebelum dilakukannya proses data *mining*, kemudian data hasil seleksi akan disimpan dalam suatu penyimpanan yang berbeda dengan database operasional sebelumnya .
3. *Preprocessing* data, yaitu penyeleksian data untuk membersihkan data yang tidak sempurna seperti data hilang, data duplikat dan memperbaiki kesalahan dalam penginputan data yang tidak relevan agar tidak mengurangi nilai akurasi dari data tersebut.
4. Transformasi data, yaitu metode yang digunakan untuk mengubah format atau struktur dari data sebelum dilakukannya proses data *mining*. Dalam data *mining* terdapat beberapa metode yang dapat digunakan tergantung karakteristik dari datanya. Salah satu proses yang dilakukan dalam *transformasi* data adalah *scaling* data , yaitu proses penyeragaman data numerik agar memiliki rentang nilai yang sama atau seragam.

   Scaling data dapat dilakukan dengan dua cara , yaitu :

   a. *Min Max Normalization* , yaitu metode *scaling* data dengan melakukan transformasi linier pada data asli.

   $$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.1)$$

   dengan:

   x = nilai yang diamati,

   x_{min} = nilai x minimum,

   x_{max} = nilai x maximum.

b. *Z-score Normalization* (*Standard Scaler),* yaitu teknik yang dilakukan untuk menormalisasikan nilai-nilai atribut berdasarkan *mean* dan standar deviasi.

$$x_{standard} = \frac{x-\bar{x}}{s} \quad (2.2)$$

dengan:

x = nilai yang diamati

$\bar{x}$ = rata-rata nilai (*mean*)

s = standar deviasi

5. Data *mining*, yaitu proses pengambilan pola yang menarik dengan menggunakan metode data *mining* tertentu (misalnya, *classification, clustering, regression*) yang dapat menghasilkan representatif yang tepat dari hasil *output.*
6. *Interpretasi/evaluation* data, interpretasi bertujuan untuk menafsirkan hasil yang diperoleh dengan memvisualisasikan pola model dari data yang ada.

2.3 *Machine Learning*

Machine Learning merupakan bidang teknologi yang sedang banyak digunakan pada masa sekarang untuk membuat algoritma dengan data yang berukuran besar (*big data*) (Hasanah & Nugraha, 2023) . Pada tahun 2020, (Roihan, *et al*., 2020) mendefinisikan *machine learning* sebagai salah satu aplikasi komputer dan algoritma matematika yang diperoleh dengan mempelajari data yang menghasilkan sebuah prediksi di masa mendatang. Proses *learning* memiliki dua tahapan, yaitu latihan (*training*) yaitu tahap proses pembelajaran terhadap suatu data yang telah diketahui kategori dan pengujian (*testing*) yaitu tahapan evaluasi terhadap kinerja model dari hasil pelatihan (Solihin, *et al*., 2022). *Machine learning* terbagi menjadi empat, yaitu *Supervised Learning, Unsupervised Learning, Semi-supervised,* dan *Reinforcement Learning* (Farwati, *et al.*, 2023).

1. *Supervised learning* adalah metode klasifikasi dengan menggunakan teknik pengumpulan data yang sepenuhnya diberikan label untuk kelas yang tidak diketahui.
2. U*nsupervised learning* merupakan metode klasifikasi yang di dalam pengumpulan datanya tidak diberikan label dan tidak bisa mengidentifikasi contoh pada kelas yang ada.
3. *Semi supervised* merupakan gabungan dari s*upervised learning* dan u*nsupervised learning* , di mana sebagian dari dataset memiliki label dan sebagian tidak memiliki label untuk pelatihan.
4. *Reinforcement learning* adalah metode klasifikasi yang bekerja pada lingkungan yang dinamis dengan tujuan yang harus terselesaikan tanpa ada informasi eksplisit dari komputer bahwa tujuan sudah tercapai.

2.4 Klasifikasi

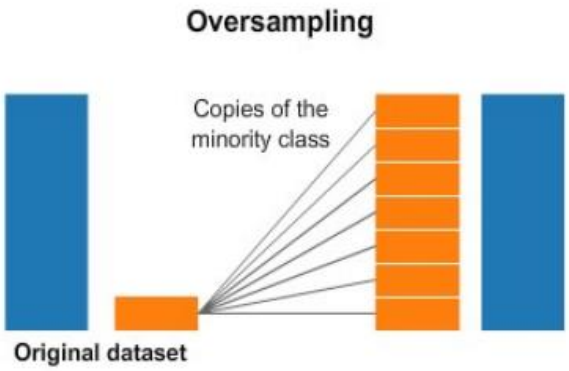
Klasifikasi merupakan proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas pada data, yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya belum diketahui (Yunita, 2017). Klasifikasi data terdiri dari dua langkah, yang pertama adalah *learning* (*fase training*), di mana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, di mana data *testing* digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (Putri, *et al*., 2013). Menurut Leidiyana (2013), proses klasifikasi didasarkan pada empat komponen, sebagai berikut :

a. Kelas, yaitu variabel dependen berupa kategorik yang merepresentasikan (label) yang terdapat pada objek.
b. Prediktor, yaitu variabel independen yang direpresentasikan oleh karakteristik (atribut) data.

c. *Training* dataset, merupakan satu set data yang berisi nilai dari kelas dan *predictor* yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor.*
d. *Testing* dataset, yaitu data baru yang akan diklasifikasikan oleh model yang telah dibuat untuk menghitung tingkat akurasi dari klasifikasi.

2.5 *Imbalance* Data

Imbalance data adalah kondisi di mana terjadi perbedaan antara jumlah data yang signifikan pada kelas mayoritas dan kelas minoritas. Kelas yang memiliki jumlah data lebih banyak disebut dengan kelas mayoritas dan kelas yang memiliki jumlah data lebih sedikit disebut dengan kelas minoritas (Akbar & Hayaty, 2020). Salah satu metode yang digunakan untuk mengatasi *imbalance* data adalah *random oversampling* (ROS). ROS adalah proses penambahan data pada *training* dari kelas minoritas secara acak, yang dilakukan secara berulang sampai jumlah data dari kelas minoritas sama dengan kelas mayoritas (Prasetya, 2022). Tujuan dari ROS yaitu menciptakan keseimbangan pada jumlah sampel pada kelas mayoritas dan kelas minoritas dengan menambahkan duplikat dari beberapa data yang ada pada kelas tersebut
(Hasbi & Sasongko, 2024).



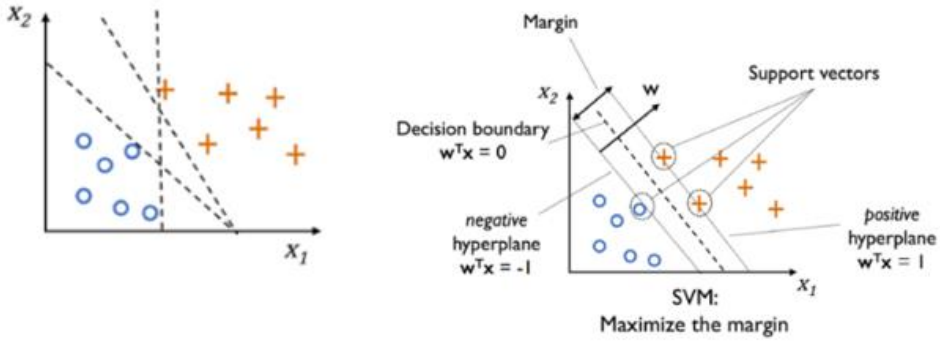
Gambar 1. Proses *Random Oversampling*

2.6 *Hyperparameter Tuning Grid Search*

Hyperparameter tuning berfungsi untuk membantu model mencari nilai parameter yang tepat pada dataset agar meningkatkan akurasi secara maksimal (Amalia, *et al*., 2022). Salah satu metode *hyperparameter tuning* yang paling sederhana adalah *grid search* , cara kerja *grid search* yaitu dengan melakukan pencarian secara menyeluruh ke semua kombinasi *hyperparameter* yang ditentukan pada *grid konfigurasi* (Dana, *et al*., 2024). Menentukan kombinasi dari model dan *hyperparameter* dilakukan dengan uji coba satu persatu terhadap kombinasi model dan melakukan validasi pada setiap kombinasi dengan menggunakan *grid search* (Fatmawati & Rifai, 2023). *Grid search* merupakan metode *hyperparameter tuning* yang digunakan untuk mengoptimalkan nilai akurasi. Keunggulan metode ini adalah parameter kandidat dihasilkan secara berurutan atau sistematis sehingga setiap kumpulan data akan dievaluasi menggunakan parameter kandidat yang sama (Andini, *et al.*, 2022). Kelebihan *grid search* adalah kemampuannya untuk dijalankan secara paralel, di mana setiap percobaan berjalan secara mandiri tanpa dipengaruhi oleh urutan waktu (Nurcahyo & Sasongko, 2023).

2.7 *Support Vector Machine* (SVM)

Support Vector Machine merupakan salah satu teknik dalam *supervised learning* yang digunakan dalam masalah regresi dan klasifikasi, seperti *Support Vector Regression* dan *Support Vector Classification* (Oktafiani & Rianto, 2023). SVM digunakan untuk menemukan *hyperplane* optimal dengan memberikan jarak atau pemisah antara dua kelas dan margin maksimum (Akbar, *et al*., 2024). Margin merupakan jarak antara titik data terdekat dengan *hyperplane.*



Gambar 2. Menentukan *hyperplane* terbaik dengan SVM

Pada Gambar 2 menunjukkan bahwa terdapat dua pola yang merupakan anggota dari dua buah kelas, yaitu +1 dan -1. Pola yang tergabung dalam kelas -1 disimbolkan dengan lingkaran berwarna biru, sedangkan kelas +1 disimbolkan dengan tanda tambah berwarna oranye. Pada Gambar 2 terlihat berbagai alternatif bidang pemisah yang dapat memisahkan semua dataset sesuai dengan kelasnya. Gambar sebelah kiri merupakan alternatif bidang pemisah sesuai kelasnya, sedangkan gambar sebelah kanan merupakan bidang pemisah terbaik optimal (*hyperplane*) dengan jarak terbesar. Adapun data yang terletak pada bidang pembatas disebut dengan *support vector* (Rangkuti, 2023). Tujuan utama dari klasifikasi adalah mencari *hyperplane* pemisah antara kedua kelas.

Misalnya data yang ada direpresentasikan dalam bentuk vektor berikut :

$$\vec{\boldsymbol{d}} = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\},$$

dengan $x_i \in R \; dan \; y_i \in \{-1,1\}$.

Diasumsikan data terpisah secara sempurna oleh *hyperplane* ke dalam dua kelas yaitu -1 dan 1, yang didefinisikan sebagai berikut (Zaki & Meira, 2014) :

$$\boldsymbol{w}^T.\vec{\boldsymbol{x}} + b = 0 \qquad (2.3)$$

dengan :

$\boldsymbol{w}$ = vektor normal pada *hyperplane*

b = jarak dari *hyperplane* ke titik pusat

Menurut Cortes & Vapnik (1995) , didapatkan persamaan sebagai berikut :

$$[(\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_i}}) + b] \geq +1 \; untuk \; y_i = +1 \quad (2.4)$$

$$[(\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_i}}) + b] \leq -1 \; untuk \; y_i = -1 \quad (2.5)$$

dengan :

$\boldsymbol{x_i}$ = himpunan data *training* , $i = 1,2, \ldots \; , n$,

y_i = label kelas dari x_i.

Persamaan (2.4) dan (2.5) dapat disederhanakan menjadi :

$$y_i(\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_i}} + b) \geq 1, i = 1,2,3, \ldots \; , n \quad (2.6).$$

Pemaksimalan jarak terdekat antara *hyperplane* dengan *pattern* dilakukan untuk menghitung margin maksimum antar kelas. Margin didefinisikan sebagai $d = d_1 + d_2$, sehingga margin akan memiliki nilai maksimum jika $d_1 = d_2$. Margin maksimum dapat didapatkan dengan memaksimalkan jarak antara *hyperplane* dengan titik terdekatnya yaitu $\frac{1}{\|\vec{w}\|}$.

$$d = d_1 + d_2 = \frac{1}{\|\vec{w}\|}\left(|\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_1}}) + b||\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_2}} + b|\right) = \frac{2}{\|\vec{w}\|} \quad (2.7).$$

Berdasarkan persamaan di atas, maka untuk mencari margin maksimal sama dengan meminimumkan nilai $\|w\|^2$, secara matematis dinyatakan sebagai berikut:

$$min\frac{1}{2}\|\vec{w}\|^2 \quad (2.8).$$

Optimasi dapat dilakukan dengan menggunakan *Lagrange multiplier* sebagai berikut :

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{l} a_i \left[y_i(\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_i}}) + b\right] - 1$$

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{l} a_i \, y_i(\boldsymbol{w^T}.\overrightarrow{\boldsymbol{x_i}} + b) - \sum_{i=1}^{l} a_i \quad (2.9)$$

a_i merupakan *lagrange multiplier* dengan nilai nol atau positif ($a_i \geq 0$).

Optimasi dilakukan dengan meminimalkan L terhadap **w** dan b sebagai berikut (Hamel, 2009):

$$\frac{\partial L}{\partial b} = 0$$

$$\sum_{i=1}^{l} a_i\, y_i = 0 \tag{2.10}$$

$$\frac{\partial L}{\partial w} = 0$$

$$\vec{\boldsymbol{w}} \sum_{i=1}^{l} a_i\, y_i \vec{\boldsymbol{x_i}} = 0$$

$$\vec{\boldsymbol{w}} = \sum_{i=1}^{l} a_i\, y_i \vec{x_i} \tag{2.11}.$$

Selain itu, optimasi dapat dilakukan dengan memaksimalkan L terhadap a_i dengan substitusi persamaan (2.10) dan (2.11) ke dalam persamaan (2.9) sebagai berikut :

$$L = \frac{1}{2}\|\vec{\boldsymbol{w}}\|^2 - \sum_{i=1}^{l} a_i\, y_i(\boldsymbol{w^T}\, \vec{\boldsymbol{x_i}} + b) - \sum_{i=1}^{l} a_i$$

$$L = \frac{1}{2}(\boldsymbol{w^T}.\vec{\boldsymbol{w}}) - \left(\sum_{i=1}^{l} a_i\, y_i \boldsymbol{w^T}\, \vec{\boldsymbol{x_i}} + \sum_{i=1}^{l} a_i\, y_i b - \sum_{i=1}^{l} a_i\right)$$

$$L = \frac{1}{2}\left(\sum_{i=1}^{l} a_i\, y_i \vec{\boldsymbol{x_i}} \,.\, \sum_{j=1}^{l} a_j\, y_j \vec{\boldsymbol{x_j}}\right) - \left(\left(\sum_{i=1}^{l} a_i\, y_i \vec{\boldsymbol{x_i}}.\sum_{j=1}^{l} a_j\, y_j\, \vec{\boldsymbol{x_j}}\right) + 0 - \sum_{i=1}^{l} a_i\right)$$

$$L = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i\, a_j y_i y_j \vec{\boldsymbol{x_i}}\, \vec{\boldsymbol{x_j}} - \left(\sum_{i=1}^{l}\sum_{j=1}^{l} a_i\, a_j y_i y_j \vec{\boldsymbol{x_i}}\, \vec{\boldsymbol{x_j}} - \sum_{i=1}^{l} a_i\right)$$

$$L = \sum_{j=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i\, a_j y_i y_j \vec{\boldsymbol{x_i}}\, \vec{\boldsymbol{x_j}} \tag{2.12}$$

dimana $a_i \geq 0, \sum_{i=1}^{l} a_i y_i = 0$.

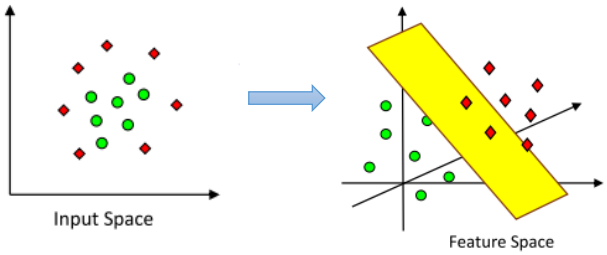
Nilai a_i akan diperoleh dengan penyelesaian persamaan (2.12) yang digunakan untuk mencari *primal* variabel dengan rumus :

$$\vec{\boldsymbol{w_l}} = \sum_{i=1}^{l} a_i\, y_i K\left(\vec{\boldsymbol{x_i}}, \vec{\boldsymbol{x_j}}\right), b = -\frac{1}{2}(\boldsymbol{w^T x^+} + \boldsymbol{w^T x^-}) \tag{2.13}$$

Setelah proses telah dilakukan, maka diperoleh $a_i > 0$ yang disebut dengan *support vector* dan sisanya memiliki nilai $a_i = 0$. Keputusan yang dihasilkan oleh SVM hanya bergantung pada nilai dari *support vector*. Umumnya data yang sepenuhnya terpisah secara linear sangat jarang ditemui di dunia nyata, sehingga untuk menangani masalah non linear SVM dapat menggunakan fungsi *kernel*. Fungsi *kernel* bekerja dengan mentransformasikan data ke ruang fitur berdimensi lebih tinggi (*feature space*).

Penyelesaian kasus non linier dapat diatasi dengan SVM yang telah dikembangkan yaitu dengan menggunakan *kernel trick* yang dapat mengubah data menjadi linier (Hamel, 2009). Adapun *kernel trick* dirumuskan dengan :

$$K(\overrightarrow{x_i}, \overrightarrow{x_j}) = \phi(\overrightarrow{x_i}).\phi(\overrightarrow{x_j}) \quad (2.14).$$



Gambar 3. Transformasi Data dalam *Feature Space*

Pada Gambar 3 menunjukkan data berdimensi dua tidak dapat dipisahkan secara linear oleh *hyperplane*. Pada Gambar 3 mengilustrasikan pemetaan data ke dalam ruang dengan dimensi lebih tinggi (dimensi tiga) sehingga dua kelas dapat dipisahkan secara linear oleh *hyperplane*. Berikut notasi matematika dari *mapping* tersebut :

$$\phi; R^d \rightarrow R^q, d < q \quad (2.15).$$

Umumnya, transformasi ϕ tidak diketahui sehingga diganti dengan fungsi *kernel* $K = (\boldsymbol{x_i}, \boldsymbol{x_j})$. Hasil klasifikasi dapat diperoleh dari persamaan :

$$f(\phi(\overrightarrow{x_i})) = sign(\boldsymbol{w^T}.\phi(\overrightarrow{x_i}) + b)$$
$$f(\phi(\overrightarrow{x_i})) = sign\left(\sum_{i=1}^{n} a_i y_i\, \phi(\overrightarrow{x_i}).\phi(\overrightarrow{x_j}) + b\right)$$
$$f(\phi(\overrightarrow{x_i})) = sign\left(\sum_{i=1}^{n} a_i y_i\, K(\overrightarrow{x_i}, \overrightarrow{x_j}) + b\right) \quad (2.16)$$

dengan :

x_i = data *input x* baris ke-i

x_j = data *input x* kolom ke-j

y_i = kelas *output* baris ke-i

b = bias

a_i = *support vector*

$sign = notasi\ (+\ atau\ -), jika\ f(\phi(x)) > 0$ maka data dimasukkan ke kelas +1, sedangkan jika $f(\phi(x)) < 0$ maka data dimasukkan ke kelas -1.

Beberapa fungsi *kernel* yang umumnya digunakan dalam SVM sebagai berikut (Han & Kamber, 2006) :

a. *Kernel* Linier

$$K(\vec{x_i}, \vec{x_j}) = \vec{x_i} . \vec{x_j} \quad (2.17)$$

b. *Kernel* Polynomial

$$K(\vec{x_i}, \vec{x_j}) = (\vec{x_i} . \vec{x_j} + 1)^d \quad (2.18)$$

c. *Kernel* Sigmoid

$$K(\vec{x_i}, \vec{x_j}) = \tanh\ (\gamma \vec{x_i} . \vec{x_j} - \delta) \quad (2.19)$$

d. *Kernel Radial Basis Function* (RBF)

$$K(\vec{x_i}, \vec{x_j}) = e^{-\frac{\|\vec{x_i} \cdot \vec{x_j}\|^2}{2\sigma^2}}, \sigma > 0 \quad (2.20)$$

2.8 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui seberapa baik kinerja dari model pada pengolahan data dengan menggunakan *kernel* SVM. Indikator penilaian dalam melakukan klasifikasi ada banyak, salah satunya adalah *confusion matrix*. C*onfusion matrix* digunakan untuk menentukan persentase nilai dari *accuracy, precision,* dan *recall* (Atmanegara & Handayani, 2024).

Tabel 1. *Confusion matrix*

Kelas Asli	Kelas Prediksi	
	Prediksi Positif	Prediksi Negatif
Aktual Positif	*True Positive* (TP)	*False Positip* (FP)
Aktual Negatif	*False Negatif* (FN)	*True negatif* (TN)

1. *True Positive* (TP) , adalah data yang diprediksi positif dan data sebenarnya adalah positif .
2. *True Negative* (TN), adalah data yang diprediksi negatif dan data sebenarnya adalah negatif.
3. *False Positive* (FP) , adalah data yang diprediksi positif dan data sebenarnya adalah negatif.
4. *False Negative* (FN), adalah data yang diprediksi negatif dan data sebenarnya adalah positif.

Berdasarkan tabel *confusion matrix* di atas, dapat dilakukan perhitungan untuk mengukur performa dari model *accuracy, precision*, *recall* dan *F1-score.* Berikut adalah perhitungan dari masing-masing model (Saputro & Sari, 2020) :

a. *Accuracy,* adalah total keseluruhan seberapa sering model benar mengklasifikasi.

$$Accuracy = \frac{TP + TN}{Total}$$

b. *Precision,* yaitu ketika model memprediksi positif dan seberapa sering prediksi itu benar.

$$Precision = \frac{TP}{FP + TP}$$

c. *Recall* (*Sensitivity/True Positive Rate*), yaitu ketika kelas aktualnya positif dan seberapa sering model memprediksi positif.

$$Recall = \frac{TP}{FN + TP}$$

d. *F1-score*, yaitu rata-rata harmonik dari *Precision* dan *Recall* .

$$F1\text{-}score = 2 \times \frac{precision \times recall}{precision + recall}$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2024/2025 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu pengetahuan Alam Universitas lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini merupakan data sekunder mengenai penderita penyakit *cardiovascular* yang diambil dari website https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset. Jumlah data yang digunakan pada penelitian ini sebanyak 68.783 data yang terdiri dari 12 variabel yaitu variabel *age, gender, height, weight, ap-high, ap-low, cholesterol, glucose, smoke, alcohol, physical activity,* dan diagnosis *cardiovascular*. Sebagian contoh data dapat dilihat pada lampiran.

3.3 Metode Penelitian

Langkah-langkah analisis data yang dilakukan dalam penelitian ini sebagai berikut :

1. Melakukan visualisasi data dengan *bar chart* untuk menggambarkan dan mendeskripsikan perbandingan data antara jumlah penderita penyakit *cardiovascular* dan tidak penderita *cardiovascular*. Selain itu dilakukannya analisis eksplorasi data untuk mengetahui deskripsi dari masing-masing variabel yang digunakan dalam penelitian.
2. *Preprocessing* data
   a. C*leansing* data, melakukan deteksi data hilang.
   b. *Scaling* data, melakukan transformasi data dengan menggunakan *standard scaler.*
3. *Handling imbalance* data
   Melakukan *handling imbalance* data dengan menggunakan *random over sampling*. Tujuan dari *random oversampling* adalah untuk meningkatkan ukuran kelas minoritas dengan *mensistensi* sampel baru atau data *training* dengan menduplikat secara acak sampel dari kelas minoritas.
4. Melakukan *splitting* data
   Membagi data menjadi 2 bagian yaitu data *training* dan data *testing* dengan menggunakan 3 skema yaitu, 70% data *training* dan 30% data *testing*, 80% data *training* dan 20% data *testing*, 90% data *training* dan 10% data *testing* yang diambil secara acak dari dataset penelitian.
5. Melakukan klasifikasi SVM dengan menerapkan *hyperparameter tuning grid search* untuk mendapatkan parameter terbaik dalam melakukan prediksi menggunakan fungsi *kernel* linear, polinomial, sigmoid dan *radial basis function* (RBF).
6. Evaluasi terhadap model
   Model yang sudah dibangun selanjutnya akan diuji untuk mengetahui seberapa baik performa dari model yang dihasilkan dengan *confusion matrix.*

V. KESIMPULAN

Setelah melakukan klasifikasi dengan menggunakan metode *Support Vector Machine* (SVM) pada data penderita penyakit *cardiovascular*, dapat diambil kesimpulan sebagai berikut :

1. Data penderita penyakit *cardiovascular* setelah dilakukan *balancing* dengan menggunakan ROS, berhasil menjadi seimbang (*balance*) yaitu masing-masing sebanyak 34.742 data untuk kategori "c*ardiovascular*" dan "tidak *cardiovascular*".
2. Analisis SVM menggunakan data yang telah dilakukan *balancing* dengan ROS menghasilkan model terbaik yaitu model kernel RBF dengan parameter *gamma* sebesar 0,1 dan *cost* sebesar 1 pada data *training* 90% dan *testing* 10%, didapat nilai *accuracy* sebesar 73,86%, *precision* sebesar 78,86%, *recall* sebesar 71,96%, dan *f1-score* sebesar 75,25%.

DAFTAR PUSTAKA

Atmanegara, R.C., & Handayani, W. 2024. Customer Churn Analysis Using Machine Learning to Improve Customer Retention on Vissie Net. *International Journal of Scientific Research and Management* . **12**(09):7379–7387.

Akbar, K., & Hayaty, M. 2020. Data Balancing untuk Mengatasi Imbalance Dataset pada Prediksi Produksi Padi. *Information Technology Journal of* UMUS. **2**(2):1-14.

Akbar, M., Utami, G.R., Aulia, B.P., & Kurniawan, R. 2024. Penerapan Machine Learning dalam Pengklasifikasian Indeks Saham SRI-Kehati. *Prosiding Seminar Nasional Sains Data.* **4**(1):64-76.

Amalia, Radhi, M., Sitompul, D.R.H., Sinurat, S.H., & Indra, E. 2022. Prediksi Harga Mobil Menggunakan Algoritma Regresi dengan Hyperparameter Tuning. *Jurnal Sistem Informasi dan Ilmu Komputer Prima* **4**(2):28–32.

Andini, E., Faisal, M.R., Herteno, R., Nugroho, R.A., Abadi, F., & Muliadi. 2022. Peningkatan Kinerja Prediksi Cacat Software dengan Hyperparameter Tuning pada Algoritma Klasifikasi Deep Forest. *Jurnal Mnemonic* **5**(2):119–27.

Cindy, Sabatini, T., & Itan, V. 2024. Implementasi Support Vector Machine untuk Klasifikasi Kasus Monkeypox: Pendekatan Oversampling dan Undersampling untuk Mengatasi Ketidakseimbangan Kelas. *Journal of Digital Ecosystem for Natural Sustainability* **4**(1):38-43.

Cortes, & Vapnik. 1995. Support Vector Networks. *Kluwer Academic Publisher.* **20**(3):273–97.

Dana, A.R., Kristananda, R.V., Wibowo, M.B.S., & rasetya, D.A. 2024. Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke. *Seminar Nasional Informatika Bela Negara.* **4**:66-75.

Farwati, M., Salsabila, I.T., Navira, K.R., & Sutabri, T. 2023. Analisis Pengaruh Teknologi Artificial Intelligence (Ai) dalam Kehidupan Sehari-hari. *Jurnal Sistem Informasi & Manajemen* **11**(01):39-45.

Fatmawati, & Rifai, N.A.K. 2023. Klasifikasi Penyakit Diabetes Retinopati Menggunakan Support Vector Machine dengan Algoritma Grid Search Cross-validation. *Jurnal Riset Statistika* **3**(1):79–86.

Fitriani, M.A., & Febrianto, D.C. 2021. Data Mining for Potential Customer Segmentation in the Marketing Bank Dataset. *Jurnal Informatika* **9**(1):25-32.

Hamel, L. 2009. *Knowledge Discovery with Support Vector Machines*. Boken-New Jersey. Canada.

Han, J., & Kamber, M. 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufmann. Amsterdam .

Hasanah, F.Rahmah.U., & Nugraha, M.K.R. 2023. Predictive Analysis Pendidikan Menggunakan Machine Learning di Sumatera Barat. *JOSTECH Journal of Science and Technology* **3**(1):79–88.

Hasbi, Sasongko, T.B. 2024. Optimasi Performa Random Forest dengan Random Oversampling dan SMOTE pada Dataset Diabetes. *Jurnal Media Informatika Budidarma* **8**(3):1756-1767.

Leidiyana, H. 2013. Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer* **1**(1):65-76

Mardi, Y. 2017. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika* **2**(2):213–19.

Mutmainah, S., 2021. Penanganan Imbalance Data pada Klasifikasi Kemungkinan Penyakit Stroke. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, **1**(1):10-16.

Nugroho, G.N.C. 2024. Klasifikasi Gender Berdasarkan Sidik Jari Menggunakan Principal Component Analysis dan Support Vector Machine. *Journal of Information Engineering and Educational Technology* **8**(1):45–53.

Oktafiani, R., & Rianto, R. 2023. Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata. *Jurnal Nasional Teknologi dan Sistem Informasi* **9**(2):113–121.

Prasetya, J. 2022. Penerapan Klasifikasi Naive Bayes dengan Algoritma Random Oversampling dan Random Undersampling pada Data Tidak Seimbang Cervical Cancer Risk Factors. *Jurnal Matematika* **2**(2):11–22.

Purnajaya, A.R., Darmawan, J., Yamin, V., & Charles. 2024. Pendekatan dengan Oversampling dan Undersampling untuk Meningkatkan Akurasi Diagnostik Kanker Tiroid. *Jurnal Pustaka Data (Pusat Akses Kajian Database, Analisa Teknologi, dan Arsitektur Komputer)* **4**(1):33–39.

Putri, Y.R., Mukhlash, I., & Hidayat, N. 2013. Prediksi Pola Kecelakaan Kerja pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 dan C5.0. *Jurnal Sains dan Seni Pomits* **2**(1):1-6.

Rangkuti, F.A. 2023. Implementasi Support Vector Machine untuk Mendeteksi Kalimat Pelecehan Seksual pada Media Sosial Facebook. *Jurnal Teknologi Informasi* **4**(1):275–84.

Roihan, A., Sunarya, P.A., & Rafika, A.S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *Jurnal Indonesian Journal on Computer and Information Technology* **5**(1):75-82.

Roth, G.A., Abate, D., Abate, K.H., Obay, S.M., Abbafati, C., Abbasi, N., *et al*., 2018. Global, Regional, and National Age-sex-specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* **392**(10159):1736-1788

Saputro, I.W., & Sari, B.W. 2020. Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal* **6**(1):1-11.

Sir, Y.A., & Soepranoto, A.H.H. 2022. Pendekatan Resampling Data untuk Menangani Masalah Ketidakseimbangan Kelas. *Jurnal Komputer dan Informatika* **10**(1):31–38.

Siringoringo, R., & Jaya, I.K. 2018. Ensemble Learning dengan metode Smote Bagging pada Klasifikasi Data Tidak Seimbang. *Information System Development* **3**(2).

Solihin, A., Mulyana, D.I., & Yel, M.B. 2022. Klasifikasi Alat Musik Tradisional Papua Menggunakan Metode Transfer Learning dan Data Augmentasi. *Jurnal Sistem Komputer dan Kecerdasan Buatan* **5**(2):36-44.

Susanti, Y., Choyyin, M.G., Priyatna, A., & Lestari, S. 2023. Perbandingan Penerapan Algoritma Decision Tree C.45 dan Naïve Bayes dalam Analisis Kelulusan Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan. *Jurnal Sistem Informasi dan Manajemen Basis Data* **6**(2):117–23.

Tanty, Ginting, B.S., & Simanjuntak, M. 2021. Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus : Puskesmas Bahorok). *Jurnal Ilmu Komputer dan Informatika* **5**(2).

Tarigan, P.M.S., Hardnata, J.T., Qurniawan, H., Safii, M., & Winajaya, R. 2022. Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap). *Jurnal Sistem Informasi Ibrahimy* **3**(1):55–65.

Toha, A., Purwono, P., & Gata, W. 2022. Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter Grid Search CV. *Buletin Ilmiah Sarjana Teknik Elektro* **4**(1):12–21.

Umar, R., Riadi, I., & Faroek, D.A. 2020. Komparasi Image Matching Menggunakan Metode K-Nearest Neighbor (KNN) dan Metode Support Vector Machine (SVM). *Journal of Applied Informatics and Computing* **4**(2):124-131.

Wahyuni, S.D., & Kusumodestoni, R.H. 2024. Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting. *Journal Bulletin of Information Technology* **5**(2):56-64.

Yunita, D., 2017. Perbandingan Algoritma K-Nearest Neighbor dan Decision Tree untuk Penentuan Risiko Kredit Kepemilikan Mobil. *Jurnal Informatika Universitas Pamulang* **2**(2):103-107.

Zaki, M.J., & Meira Jr, W. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.