ANALISIS KINERJA BEBERAPA FUNGSI *KERNEL* PADA KLASIFIKASI *SUPPORT VECTOR MACHINE* TERHADAP DATA PENDERITA PENYAKIT *LIVER*

(SKRIPSI)

Oleh

CITRA MARIA MAGDALENA NAIBAHO
2017031073

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

ABSTRACT

PERFORMANCE ANALYSIS OF SEVERAL KERNEL FUNCTIONS IN SUPPORT VECTOR MACHINE CLASSIFICATION OF DATA ON LIVER DISEASE PATIENTS

BY

CITRA MARIA MAGDALENA NAIBAHO

Support Vector Machine (SVM) is one of the machine learning methods used for classification by dividing data into two different classes. The principle of the SVM method is to find the best separating function called hyperplane. If the data cannot be separated linearly then the data is nonlinear data. One method to overcome this is to use a kernel function. The purpose of this research is to apply the SVM method to determine the performance of the best kernel function based on the highest accuracy value for the classification of liver disease patients. Based on the analysis results, the best kernel function for this case is the Radial Basis Function (RBF) kernel function with a cost value = 12 and gamma = 2.5 on a split dataset of 90% training and 10% testing, with an accuracy value of 84.63%. By using the kernel function and the split dataset proportion, the parameters **w** and b are obtained, as follows:
w_{Age} = -5,1951, w_{TB} = -7,3708, w_{DB} = -9,1687, w_{AP} = -9,2968, w_{AA} = -10,6287, $w_{Asp\ A}$ = -10,3948, w_{TP} = 0,1057, $w_{Albumin}$ = 9,3402, $w_{AG\ R}$ = 2,5066 , b = 0,8868.

Keywords: Support Vector Machine; Radial Basis Function Kernel; Classification; Liver Disease

ABSTRAK

ANALISIS KINERJA BEBERAPA FUNGSI *KERNEL* PADA KLASIFIKASI *SUPPORT VECTOR MACHINE* TERHADAP DATA PENDERITA PENYAKIT *LIVER*

OLEH

CITRA MARIA MAGDALENA NAIBAHO

Support Vector Machine (SVM) adalah salah satu metode *machine learning* yang digunakan untuk pengklasifikasikan dengan membagi data menjadi dua kelas yang berbeda. Prinsip metode SVM adalah mencari fungsi pemisah disebut *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 *liver*. Berdasarkan hasil analisis diperoleh, bahwa fungsi *kernel* terbaik untuk kasus ini adalah fungsi *kernel Radial Basis Function* (RBF) dengan nilai *cost* = 12 dan *gamma* = 2,5 pada *split* dataset 90% *training* dan 10% *testin*g, dengan nilai akurasi sebesar 84,63%. Dengan menggunakan fungsi *kernel* dan proporsi *split* dataset tersebut diperoleh parameter **w** dan b, yakni sebagai berikut:
w_{Age} = -5,1951, w_{TB} = -7,3708, w_{DB} = -9,1687, w_{AP} = -9,2968, w_{AA} = -10,6287, $w_{Asp\ A}$ = -10,3948, w_{TP} = 0,1057, $w_{Albumin}$ = 9,3402, $w_{AG\ R}$ = 2,5066 , b = 0,8868.

Kata Kunci : *Support Vector Machine*; *Kernel Radial Basis Function*; Klasifikasi; Penyakit *liver*

ANALISIS KINERJA BEBERAPA FUNGSI *KERNEL* PADA KLASIFIKASI *SUPPORT VECTOR MACHINE* TERHADAP DATA PENDERITA PENYAKIT *LIVER*

Oleh

CITRA MARIA MAGDALENA NAIBAHO
2017031073

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

Judul Skripsi	:	**ANALISIS KINERJA BEBERAPA FUNGSI *KERNEL* PADA KLASIFIKASI *SUPPORT VECTOR MACHINE* TERHADAP DATA PENDERITA PENYAKIT *LIVER***
Nama Mahasiswa	:	**Citra Maria Magdalena Naibaho**
Nomor Pokok Mahasiswa	:	**2017031073**
Jurusan	:	**Matematika**
Fakultas	:	**Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI,

1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.	**Dra. Dorrah Azis, M.Si.**
NIP. 19740726 200003 2 001	NIP. 19610128 198811 2 001

Mengetahui,

Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.** .....

   Sekretaris : **Dra. Dorrah Azis, M.Si.** .....

   Penguji

   Bukan Pembimbing : **Drs. Nusyirwan, M.Si.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711101 200501 1 002

Tanggal Lulus Ujian Skripsi : **28 Maret 2024**

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama	: Citra Maria Magdalena Naibaho
Nomor Pokok Mahasiswa	: 2017031073
Jurusan	: Matematika
Judul Skripsi	: **Analisis Kinerja Beberapa Fungsi *Kernel* Pada Klasifikasi *Support Vector Machine* Terhadap Data Penderita Penyakit *Liver***

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah katya penulisan ilmiah Universitas Lampung. Baik gagasan, data maupun pembahasannya adalah benar karya saya sendiri yang saya susun dengan mengikuti norma dan etika yang berlaku dan saya memastikan bahwa tingkat similaritas tidak lebih dari 25%. Jika dikemudian hari terbukti bahwa pernyataan saya ini tidak benar, saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 28 Maret 2024

Penulis

Citra Maria Magdalena Naibaho

NPM. 2017031073

RIWAYAT HIDUP

Penulis dilahirkan pada tanggal 04 Maret 2001 di Balige, anak pertama dari empat bersaudara pasangan Bapak Lindon Naibaho dan Ibu Mariani Simanjuntak. Penulis memulai pendidikan di TK ST. Lusia Siborongborong pada tahun 2006-2007, pendidikan tingkat dasar di SD ST. Lusia Siborongborong pada tahun 2007-2013. Pada Tahun 2013, penulis melanjutkan ke pendidikan tingkat menengah pertama di SMP ST.Lusia Siborongborong sampai dengan tahun 2016, dan kemudian melanjutkan pendidikan tingkat menengah atas di SMA Negeri 1 Siborongborong dan diselesaikan pada tahun 2019.

Pada tahun 2020 melalui jalur SBMPTN, penulis diterima dan terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Selama menjadi mahasiswa, penulis aktif dalam organisasi Himpunan Mahasiswa Jurusan Matematika (HIMATIKA)sebagai anggota Kaderisasi & Kepemimpinan, anggota Dinas ADKESMA Badan Eksekutif Mahasiswa (BEM FMIPA) dan menjadi anggota Doa & Pemerhati (DP) di organisasi Persekutuan Oikumene Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam (POMMIPA), Mentoring Agama Katolik organisasi KMK Unila pada tahun 2020-2022.

Pada Januari tahun 2023 penulis melaksanakan Praktik Kerja Lapangan (PKL) di PT. Bank Rakyat Indonesia Kantor Cabang Tanjung Karang. Pada Tahun yang sama di bulan Juli penulis melaksanakan Kuliah Kerja Nyata (KKN) di desa Bina Karya Jaya, Kecamatan Putra Rumbia, Lampung Tengah selama 40 hari.

KATA INSPIRASI

“Apapun juga yang kamu perbuat, perbuatlah dengan segenap hatimu seseperti untuk Tuhan dan bukan untuk manusia”
(Kolose 3:23)

“Lakukanlah segala pekerjaanmu dalam kasih”
(1 Korintus 16:14)

“Tak semua usaha itu dipermudah, tapi semua yang berusaha pasti berubah”
(2 Tawariks 15:7)

“Sebuah rencana yang hebat dapat gagal hanya karena kurangnya kesabaran”
(Konfusius)

“Jika kita tau bahwa penyesalan itu datang terlambat maka jangan berikan peluang untuk penyesalan itu hadir dalam hidup mu”
(Penulis)

PERSEMBAHAN

Segala puji dan syukur kepada Tuhan Yesus Kristus yang telah memberikan rahmat, pertolongan dan anugerah-Nya melaui orang-orang yang membimbing dan mendukung dengan berbagai cara sehingga penulis dapat menulis dan menyelesaikan skripsi ini. Dengan penuh ketulusan, penulis mempersembahkan karya ini untuk :

Kedua Orang Tua Tercinta:

Ibu dan Ayah yang telah memberikan dukungan moril, maupun materi serta do'a yang tiada henti untuk kesuksesan saya dan keberhasilan dalam penulisan skripsi ini. Kasih sayang serta pengorbanan yang tak tergantikan sehingga penulis selalu kuat dalam melewati segala halangan dan rintangan.

Bapak, Mamak terimalah karya ini sebagai salah satu kado kecil untuk membalas semua pengorbanan kalian yang tanpa lelah berjuang mempertaruhkan nyawa hingga segalanya demi hidupku.

Bapak Ibu Dosen Pembimbing dan Penguji:

Dosen pembimbing dan penguji yang sangat berjasa dan tidak lelah memberikan arahan serta masukan sehingga penulis dapat menyelesaikan skripsi ini.

Kepada adik-adik yang kusayangi:

Terimakasih kepada adik-adikku yang selalu memberikan semangat dan dukungan dalam menyelesaikan skripsi ini.

SANWACANA

Puji syukur penulis ucapkan kehadirat Tuhan Yang Maha Esa yang telah memberikan berkah dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **"Analisis Kinerja Beberapa Fungsi *Kernel* pada Klasifikasi *Support Vector Machine* Terhadap Data Penderita Penyakit *Liver*"**.

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

1. Ibu Dr. Khoirin Nisa, S.Si, M.Si., selaku dosen Pembimbing I yang telah banyak memberikan bimbingan, kritik, saran, dan *support* dalam penyelesaian skripsi ini.
2. Ibu Dra. Dorrah Aziz, M.Si. selaku dosen Pembimbing II atas bantuan dan bimbingan kepada penulis selama proses penyusunan skripsi ini berlangsung.
3. Bapak Drs. Nusyirwan, M.Si., selaku Dosen Pembahas yang telah memberikan kritik dan saran yang membangun selama proses penyusunan skripsi.
4. Ibu Widiarti, S.Si., M.Si., selaku dosen pembimbing akademik.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku ketua Jurusan Matematika.
6. Bapak Dr. Eng. Heri Satria, S.Si.,M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen, staff, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Kedua orang tuaku, Bapak Lindon Naibaho dan Ibu Mariani Simanjutak yang telah memberikan seluruh tenaga, fikiran, semangat, dukungan, nasehat, doa yang tiada hentinya dan menjadi sumber semangat kepada penulis.
9. Abang dan adik-adikku, Santo Paulusan Naibaho, Santa Lidia Naibaho, Santi

Maria Naibaho, serta seluruh keluarga besar yang telah memberikan semangat, keceriaan, dukungan serta doa nya untuk penulis.

10. Teman seperjuangan, Arinda Lunetta Azhwar, Siska Nabila Azahra, Niken Nadia, Lathifatul Hana, Ainun Sila, Maya Puspitasari, Irma Lumbantoruan, Yolanda Nababan, Abdul Kholiq, Daniel Simatupang yang selalu berbagi ilmu, pengalaman, kelauh kesah dan sabar menemani penulis dalam penyusunan skripsi.
11. Sahabat ku, Deckris Nababan, Karima, Stevanus Dio, Windy Wati yang telah memberi semangat, dukungan dan doa kepada penulis untuk menyelesaikan skripsi.
12. Teman-teman Matematika Angkatan 2020 atas kebersamaan dan persaudaraannya selama ini.
13. Rekan-Rekan KKN Unila Periode II Tahun 2023 Desa Bina Karya Jaya, Kecamatan Putra Rumbia untuk kebersamaan selama 40 hari.
14. Keluarga Besar MENTORING KAG, POMMIPA, BEM FMIPA, dan Bidang Kaderisasi Kepemimpinan HIMATIKA FMIPA Universitas Lampung.sebagai wadah bagi penulis untuk mengembangkan minat dan bakat.
15. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.
16. Almamater tercinta Universitas Lampung.

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, 28 Maret 2024
Penulis

Citra Maria Magdalena Naibaho
NPM. 2017031073

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Klasifikasi adalah salah satu teknik yang umum digunakan dalam bidang data *mining* dan *machine learning*. Klasifikasi merupakan sebuah proses untuk menemukan sebuah model yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui (Tan, *et al*., 2006). Klasifikasi bertujuan untuk menemukan suatu aturan keputusan yang dapat memprediksi kategori data *testing* yang berasal dari distribusi yang serupa dengan data *training*. Salah satu metode dalam klasifikasi ialah *Support Vector Machine* (SVM). *Support Vector Machine* (SVM) adalah salah satu metode *machine learning* untuk menemukan fungsi pengklasifikasi yang dapat membagi data menjadi dua kelas yang berbeda (Vapnik, 1995). Prinsip SVM, ialah upaya dalam mencari *hyperplane* terbaik yang berfungsi sebagai pemisah antara dua *class* atau multi *class* pada *input space* (Nugraha, *et al*., 2019).

Dalam proses pengembangan model klasifikasi menggunakan SVM, fungsi *kernel* memainkan peran penting karena membantu dalam pemetaan dataset ke ruang dimensi yang lebih tinggi untuk mendapatkan interpretasi yang lebih baik pada model klasifikasi (Nanda, *et al*., 2018). Fungsi *kernel* yang berbeda, seperti linear, polinomial, *radial basis function* (RBF), dan sigmoid, memiliki karakteristik yang berbeda dalam transformasi data. Memilih fungsi *kernel* yang sesuai untuk tugas klasifikasi tertentu sangat penting, tetapi seringkali merupakan tantangan dalam praktiknya. Sehingga perlu dilakukan perbandingan beberapa fungsi *kernel* yang

digunakan dalam algoritma SVM untuk mencapai hasil kinerja terbaik dari setiap fungsi *kernel*.

Berbagai penelitian sebelumnya yang mengkaji mengenai nilai akurasi terbaik berdasarkan pengujian besarnya persentase dataset dan perbandingan fungsi *kernel*, yakni Praghakusma & Charibaldi (2021), Nachev & Teodosiev (2015) dan Ginanjar & Feta (2019). Penelitian Praghakusma & Charibaldi (2021), diperoleh dari penelitian bahwa *kernel* linear memiliki akurasi tertinggi dibandingkan dengan *kernel* polinomial dan *kernel* sigmoid, penelitian Nachev & Teodosiev (2015), diperoleh hasil bahwa fungsi *kernel* RBF memiliki akurasi tertinggi dibandingkan dengan *kernel* lainnya. Sedangkan, pada penelitian Ginanjar & Feta (2019), Berdasarkan hasil kinerja yang dilakukan dengan dataset kedelai, keduanya dapat bekerja dengan baik pada masalah klasifikasi. Namun, dari kedua fungsi *kernel*, *Radial Basis Function* (RBF) mengklasifikasikan lebih baik dari fungsi *kernel* lainnya.

Berdasarkan penjabaran di atas, maka dalam penelitian ini akan dikaji tentang kinerja fungsi *kerne*l pada metode *Support Vector Machine* (SVM) terhadap klasifikasi penderita penyakit *liver*.

1.2 Tujuan Penelitian

Adapun tujuan penelitian ini adalah menerapkan metode SVM untuk mengetahui kinerja fungsi *kernel* terbaik berdasarkan nilai akurasi tertinggi terhadap data klasifikasi penderita penyakit *liver*.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dapat mengetahui performa fungsi *kernel* terbaik yang menghasilkan akurasi tertinggi pada *Support Vector Machine* (SVM) pada data penderita penyakit *liver*;
2. Penelitian ini diharapkan dapat menjadi referensi atau landasan bagi penelitian yang akan dilakukan di masa mendatang.

II. TINJAUAN PUSTAKA

2.1 Data *Mining*

Data *mining* adalah proses eksplorasi, analisis, dan filtrasi data yang besar guna mengungkap pola, tren, dan keterhubungan baru dalam data (Sumiran, 2018). Berdasarkan definisi lain, data *mining* juga merujuk pada ekstraksi dan identifikasi pengetahuan yang berguna yang berguna dari data besar yang menggunakan matematika, statistika, *artificial intelligence* (AI) *dan machine learning* untuk keperluan di masa depan (Hermawati, 2013). *Data mining* sering juga disebut sebagai *Knowledge Discovery in Database* (KDD).

Berikut pengelompokan data mining (Buulolo, 2020) :

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

2.2 Tahapan Data *Mining*

Data *mining* memiliki tahapan-tahapan dalam pemrosesannya. Berikut merupakan tahapan-tahapan data *mining*:

1. Data *selection*, yaitu kumpulan database operasional yang dipilih atau diseleksi berdasarkan kebutuhan atau kepentingan sebelum melakukan proses data mining, kemudian disimpan dalam sebuah berkas atau tempat penyimpanan yang berbeda dengan *database* operasional sebelumnya agar mempermudah penggunaan data selanjutnya.
2. *Preprocessing*, yaitu membersihkan data dari isi yang tidak sempurna dari database seperti data yang hilang atau tidak valid, baik karena kesalahan pengetikan atau atribut yang tidak relevan agar tidak mengurangi nilai mutu atau akurasi yang akan dihasilkan dari data tersebut. Dalam tahap ini performasi akan berpengaruh sebab data yang dihasilkan dapat berkurang jumlah dan kompleksitasnya.
3. *Transformation*, merupakan metode yang digunakan untuk mengubah format atau struktur data sebelum memulai proses data *mining*. Kualitas dari tahap ini sangat penting karena karateristik beberapa metode yang digunakan dalam proses *data mining* tergantung pada tahap *transformation*. Dalam tahap ini, salah satu proses yang dilakukan adalah *scaling* data untuk memastikan data numerik memiliki rentang yang seragam.

   Terdapat dua cara yang digunakan dalam *scaling* data, yaitu:

   a. *Min Max Normalization* merupakan metode *scaling* data dengan melakukan transformasi linear terhadap data asli.

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

   dengan:

   x = nilai yang di amati

   x_{min} = nilai x minimum

   x_{max} = nnilai x maximum

b. *Z-Score Normalization* (*Standard Scaler*) adalah teknik di mana nilai-nilai atribut akan dinormalisasi berdasarkan *mean* dan standar deviasi.

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

dengan:

x = nilai yang di amati

$\overline{x}$ = rata rata nilai (mean)

s = standar deviasi

4. *Interpretation* (*Evaluation*) merupakan tahapan dalam mengartikan pola-pola informasi yang dihasilkan dari pemrosesan data ke dalam bentuk yang dapat dimengerti oleh pihak yang berkepentingan. Selain itu, tahap ini juga bertujuan untuk memeriksa kesesuaian pola informasi dengan fakta atau hipotesis yang telah ada sebelumnya.

2.3 *Machine Learning*

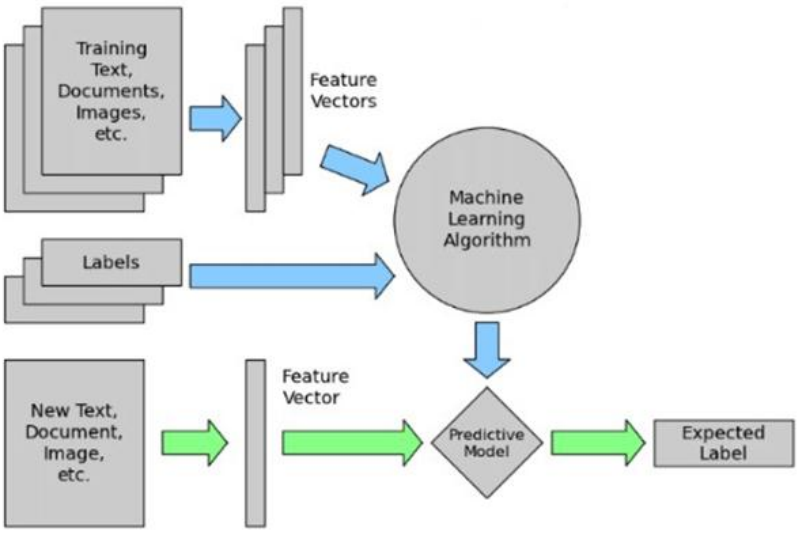
Machine learning merupakan salah satu cabang gabungan antara pengaplikasian ilmu komputer dan algoritma matematika melalui learning yang bersumber dari data, sehingga diperoleh prediksi di waktu kedepan (Goldberg & Holland, 1988). Proses *learning* ini melibatkan dua tahap utama, yaitu (*training*) dan pengujian (*testing*) (Huang, et.al., 2006). Pada *training*, model pembelajaran dibuat dengan menganalisis sampel data *training* menggunakan algoritma *learning* (Dhage & Raina, 2016). Sedangkan, pada *testing* model yang telah dilatih digunakan untuk membuat prediksi dengan memanfaatkan mesin eksekusi.

Machine learning dipisah menjadi tiga, yakni *supervised learning*, *unsupervised learning*, *reinforcement learning* (Somvanshi & Chavan, 2016). Menurut Kotsiantis (2007), jika contoh berlabel yang diketahui (sesuai *output* benar) maka *learning* disebut dengan *supervised*. Hal tersebut berbeda dengan *unsupervised learning*, yakni contoh tidak berlabel. Sedangkan, *reinforcement learning* mempunyai ide bahwa harus mengatasi tujuan tanpa adanya notifikasi dari

komputer secara jelas jika tujuan tersebut telah tercapai (Das & Nene, 2017). *Supervised learning* bertujuan untuk memprediksi hasil berdasarkan *input*, sedangkan tujuan *unsupervised learning* adalah menjelaskan hubungan (*associations*), serta pola diantara data-data input.

2.4 *Supervised Learning*

Supervised learning dibagi menjadi dua jenis masalah, yakni klasifikasi dan regresi. Dalam masalah klasifikasi, variabel terikat memiliki bentuk kategori, seperti merah dan biru atau penyakit dan tidak ada penyakit. Sementara, dalam masalah regresi, variabel terikat berupa nilai kontinu, seperti dollar atau berat (Brownlee, 2016).



Gambar 1. Model *Supervised Learning*

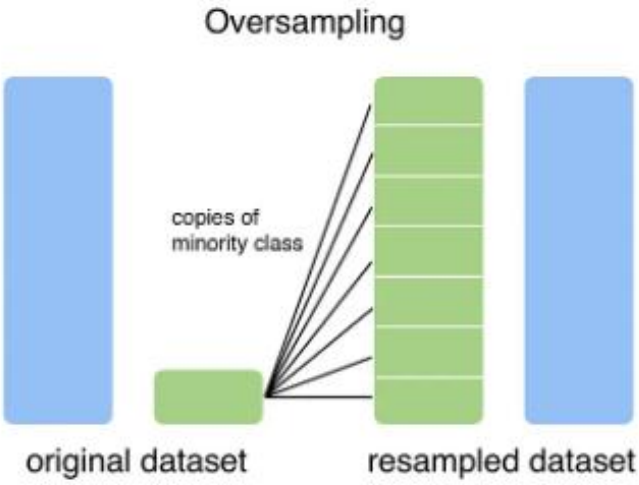
Menurut Nasteski (2017), Gambar 1 algoritma membuat perbedaan antara data yang diamati *X*, yaitu data *training*, banyak kasus data terstruktur diberikan ke model selama proses *training*. Pada proses ini, algoritma *supervised learning* membangun model prediksi. Setelah *training*, model yang memprediksi paling banyak kemungkinan label untuk *set* sampel baru *X* di kumpulan *testing*. Langkah

selanjutnya adalah diklasifikasikan tergantung pada sifat target y, yakni sebagai klasifikasi atau regresi.

2.5 *Random Oversampling*

Ketidakseimbangan data terjadi dimana suatu kelompok kelas memiliki jumlah data yang jauh berbeda dibandingkan dengan kelas lainnya. Kelas yang memiliki jumlah data lebih banyak disebut dengan *majority class* dan kelas yang mempunyai jumlah data lebih sedikit disebut dengan *minority class* (Barro dkk., 2013). Untuk mengatasi ketidakseimbangan data, terdapat beberapa metode yang dapat digunakan. Salah satunya dengan menggunakan *resampling*. Pendekatan *resampling* terdiri menjadi 3 yaitu, *Random Oversampling* (ROS), *Random Undersampling* (RUS), dan *Hibrida* yang menggabungkan kedua pendekatan *sampling* (Jian dkk., 2016).

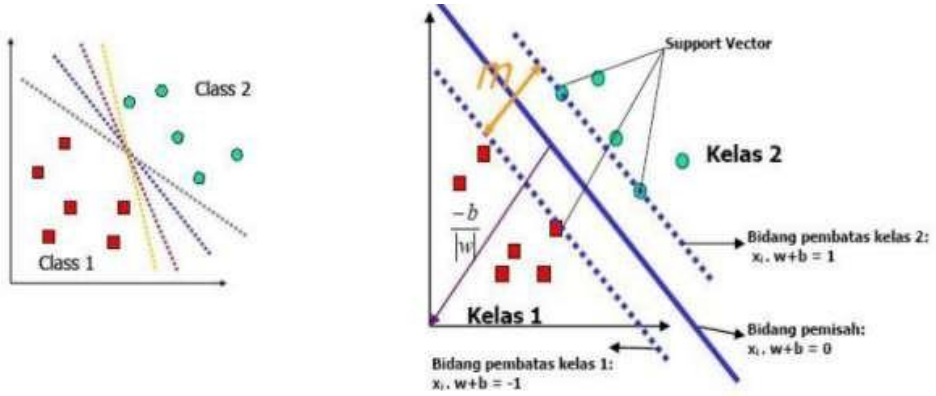
Random Oversampling adalah teknik dimana data dari kelas minoritas ditambahkan ke dalam data *training* secara acak. Proses penambahan ini diulang hingga jumlah data kelas minoritas setara dengan jumlah kelas mayoritas. *Random Oversampling* bertujuan untuk meningkatkan ukuran kelas minoritas dengan mensintesis sampel baru atau dataset *training* dengan menduplikasi secara acak sampel kelas minoritas (Yu dkk., 2017).



Gambar 2. Proses *Random Oversampling*

2.6 *Support Vector Machine*

Support Vector Machine (SVM) merupakan salah satu *supervised learning* yang pertama kali diperkenalkan oleh Vapnik, dkk., pada tahun 1992. Algoritma SVM memiliki tujuan untuk menemukan *hyperplane* yang paling optimal dengan *margin* maksimum yang bertindak untuk memisahkan dua kelas yang berbeda (Goh & Lee, 2019). *Margin* sendiri merupakan jarak antara *hyperplane* dengan data terdekat dari masing-masing kelas.



Gambar 3. *Support Vector Machine* menemukan *hyperplane* terbaik

Pada Gambar 3 menunjukkan bahwa terdapat dua pola yang merupakan anggota dari dua buah kelas, yaitu +1 dan -1. Pola yang tergabung dalam kelas -1 disimbolkan dengan kotak berwarna merah, sedangkan kelas +1 disimbolkan dengan lingkaran berwarna hijau. Pada Gambar 3 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* (Adinegoro dkk., 2015). Tujuan utama dari klasifikasi adalah mencari *hyperplane* permisah antara kedua kelas.

Misalkan data yang tersedia direpresentasikan dalam bentuk vektor :

$$\vec{d} \coloneqq \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\},$$

Dengan $x_i \in R$ dan $y_i \in \{-1,1\}$
Diasumsikan data tersebut terpisah secara sempurna ke dalam dua kelas yaitu -1 dan 1 oleh *hyperplane*, yang didefiniskan (Zaki, 2014):

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

dengan :

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

b = jarak dari *hyperplane* ke titik pusat

Sehingga menurut Cortes dan Vapnik (1995), diperoleh persamaan :

$$[(\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}}) + b] \geq +1 \; untuk \; y_i = +1 \qquad (2.4)$$

$$[(\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}}) + b] \leq -1 \; untuk \; y_i = -1 \qquad (2.5)$$

dengan :

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

y_i = label kelas dari x_i

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

$$y_i(\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}} + b) \geq 1, i = 1,2,3 \dots, N \qquad (2.6)$$

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 $\frac{1}{\|\vec{w}\|}$.

$$d = d_1 + d_2 = \frac{1}{\|\vec{w}\|}(|\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{1}} + b||\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{2}} + b|) = \frac{2}{\|\vec{w}\|} \quad (2.7)$$

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

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

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

$$L = \frac{1}{2}\|\boldsymbol{w}\|^{\boldsymbol{2}} - \sum_{i=1}^{I} a_i[y_i(\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}} + b) - 1]$$

$$L = \tfrac{1}{2}\|\boldsymbol{w}\|^{\boldsymbol{2}} - \textstyle\sum_{i=1}^{I} a_i y_i(\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}} + b) - \sum_{i=1}^{I} a_i \quad (2.9)$$

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):

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

$$\textstyle\sum_{i=1}^{I} a_i y_i = 0 \quad (2.10)$$

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

$$\vec{\boldsymbol{w}} - \sum_{i=1}^{I} a_i y_i \vec{\boldsymbol{x}}_{\boldsymbol{i}} = 0$$

$$\vec{\boldsymbol{w}} = \textstyle\sum_{i=1}^{I} a_i y_i \vec{\boldsymbol{x}}_{\boldsymbol{i}} \quad (2.11)$$

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

$$L = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{I} a_i y_i (\boldsymbol{w^T}.\vec{\boldsymbol{x}}_{\boldsymbol{i}} + b) - \sum_{i=1}^{I} a_i$$

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

$$L = \frac{1}{2}\left(\sum_{i=1}^{I} a_i y_i \vec{\boldsymbol{x}}_{\boldsymbol{i}}.\sum_{j=1}^{I} a_j y_j \vec{\boldsymbol{x}}_{\boldsymbol{j}}.\right) - \left(\sum_{i=1}^{I} a_i y_i \vec{\boldsymbol{x}}_{\boldsymbol{i}}.\sum_{i=1}^{I} a_j y_j \vec{\boldsymbol{x}}_{\boldsymbol{j}} + 0 - \sum_{i=1}^{I} a_i\right)$$

$$L = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{I} a_i a_j y_i y_j \overrightarrow{\boldsymbol{x}_{\boldsymbol{i}}}\overrightarrow{\boldsymbol{x}_{\boldsymbol{j}}} - \left(\sum_{i=1}^{I}\sum_{j=1}^{I} a_i a_j y_i y_j \overrightarrow{\boldsymbol{x}_{\boldsymbol{i}}}\overrightarrow{\boldsymbol{x}_{\boldsymbol{j}}} - \sum_{i=1}^{I} a_i\right)$$

$$L = \sum_{i=1}^{I} a_i - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{I} a_i a_j y_i y_j \overrightarrow{\boldsymbol{x}_{\boldsymbol{i}}}\overrightarrow{\boldsymbol{x}_{\boldsymbol{j}}} \qquad (2.12)$$

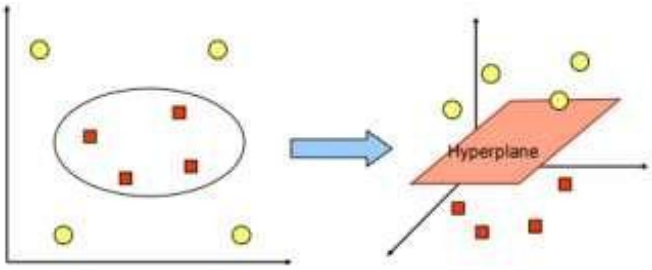
Dimana $a_i \geq 0, \sum_{i=1}^{I} a_i y_i = 0$

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

$$\overrightarrow{\boldsymbol{w_t}} = \sum_{i=1}^{I} a_i y_i K\left(\overrightarrow{\boldsymbol{x_i}}, \overrightarrow{\boldsymbol{x_j}}\right), b = -\frac{1}{2}(\boldsymbol{w^T x^+} + \boldsymbol{w^T x^-}) \qquad (2.13)$$

Setelah proses telah dilakukan, maka diperoleh $a_i > 0$ yang disebut dengan *support vector* dan sisanya memiliki nilai $a_i = 0$. Fungsi keputusan yang dihasilkan hanya bergantung pada nilai nilai *support vector*. Di dunia nyata, data jarang terpisah secara linear, sehingga untuk mengatasi permasalahan non-linear, SVM menggunakan fungsi *kernel*. Konsep dasar dari *kernel* adalah mentransformasikan data ke dalam dimensi ruang fitur (*feature space*) yang lebih tinggi. Dengan menggunakan metode kernel, SVM dapat menyelesaikan kasus non-linear dengan mengubah data menjadi linear (Hamel, 2009). Adapun metode *kernel* dirumuskan dengan :

$$K\left(\overrightarrow{\boldsymbol{x_i}}, \overrightarrow{\boldsymbol{x_j}},\right) = \emptyset(\overrightarrow{\boldsymbol{x_i}},).\emptyset\left(\overrightarrow{\boldsymbol{x_j}},\right) \qquad (2.14)$$



Gambar 4. Transformasi Data dalam *Feature Space*

Pada Gambar 4 menunjukkan di sebelah kiri terlihat *input space* X memuat beberapa data kelas kuning dan merah tidak dapat dipisahkan secara linear oleh *hyperplane*. Di sebelah kanan, gambar mengambarkan pemetaan data ke dalam ruang dengan dimensi lebih tinggi (dimensi tiga) sehingga dua kelas dapat dipisahkan secara linear oleh sebuah *hyperplane*. Berikut notasi matematika dari yang mewakili pemetaan tersebut :

$$\emptyset: R^d \rightarrow R^q, d < q \tag{2.15}$$

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

$$f\left(\emptyset(\overrightarrow{\boldsymbol{x_i}})\right) = sign(\boldsymbol{w^T}.\emptyset(\overrightarrow{\boldsymbol{x_i}}) + b)$$

$$f\left(\emptyset(\overrightarrow{\boldsymbol{x_i}})\right) = sign\left(\sum_{i=1}^{n} a_i y_i \emptyset(\overrightarrow{\boldsymbol{x_i}}).\emptyset\left(\overrightarrow{\boldsymbol{x_j}}\right) + b\right)$$

$$f\left(\emptyset(\overrightarrow{\boldsymbol{x_i}})\right) = \left(\sum_{i=1}^{n} a_i y_i K\left(\overrightarrow{\boldsymbol{x_i}}, \overrightarrow{\boldsymbol{x_j}}\right) + b\right) \tag{2.16}$$

dengan :

$\boldsymbol{x_i}$ = data *input* x baris ke-i

$\boldsymbol{x_j}$ = data *input* x kolom ke-j

y_i = kelas *output* baris ke-i

b = bias

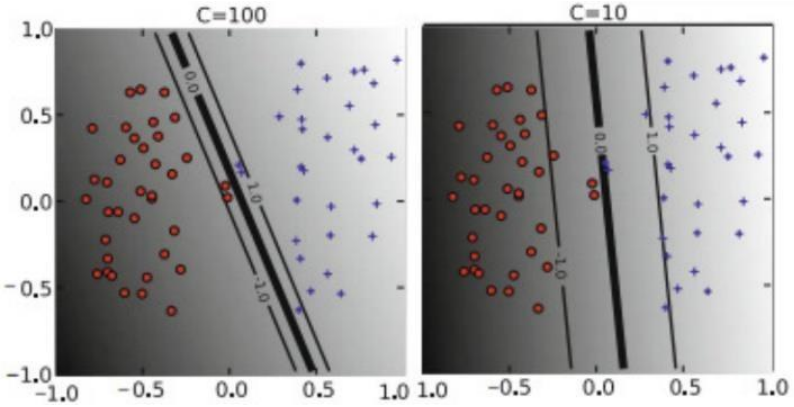
a_i = *support vector*

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

Beberapa fungsi *kernel* yang umumnya digunakan dalam SVM sebagai berikut (Han dkk., 2012) :

a. *Kernel* Linear

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \vec{x}_j \tag{2.17}$$



Gambar 5. *Kernel* Linear

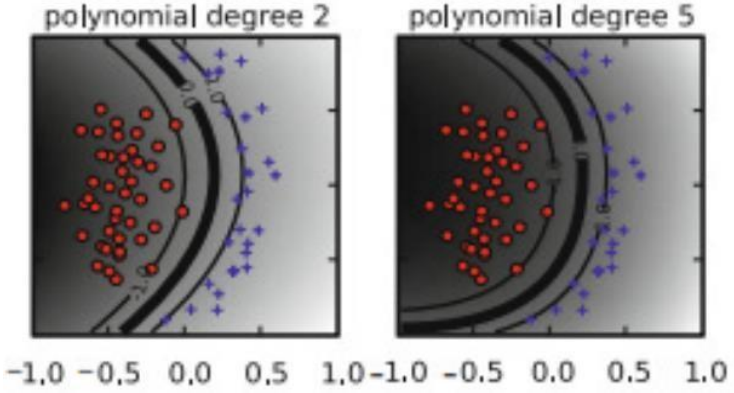
Pada Gambar 5 jika nilai *cost* yang rendah (10) maka nilai *margin error* yang rendah, memperlebar nilai *margin* dan mengabaikan *ignore point* yang dekat dengan *decision boundary*, dan juga sebaliknya jika nilai *cost* yang tinggi (100) maka *margin error* besar dan mempersempit *margin*.

b. *Kernel* Polynomial

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i . \vec{x}_j + 1)^d \tag{2.18}$$

dimana:

d : Derajat polinomial

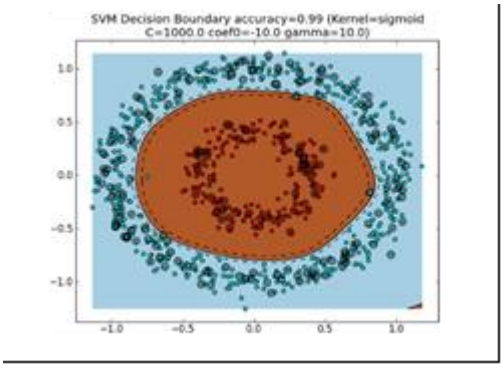


Gambar 6. *Kernel* Polinomial

Derajat atau *degree* pada *kernel* polinomial dapat mengendalikan fleksibilitas dari hasil klasifikasi, semakin tinggi nilai derajat pada *kernel* polinomial memungkinkan *decision boundary* yang lebih fleksibel (Ben-Hur & Weston, 2010). Menurut Kowalczyk (2017), menggunakan derajat polinomial yang terlalu tinggi dapat mengakibatkan *overfitting* atau model yang digunakan terlalu fokus pada data *training*, sehingga pengujian dengan data *testing* yang berbeda akan menyebabkan penurunan akurasi.

c. *Kernel* Sigmoid

$$K(\vec{x}_i, \vec{x}_j) = tanh(y\vec{x}_i.\vec{x}_j - \delta) \tag{2.19}$$



Gambar 7. *Kernel* Sigmoid

Menurut (Al-Mejibli, dkk., 2020) penggunaan *gamma* yang terlalu tinggi pada *kernel* sigmoid cenderung menurunkan tingkat akurasi pada suatu klasifikasi tetapi tetap bergantung pada jumlah fitur yang digunakan, semakin banyak jumah fitur maka *gamma* yang digunakan cenderung kecil, dan sebaliknya.

d. *Kernel Radial Basis Function* (RBF)

$$K(\vec{x}_i, \vec{x}_j) = e^{-\left(\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}\right)}, \sigma > 0 \tag{2.20}$$

Kernel RBF mempunyai dua parameter, yaitu *Gamma* dan *Cost*. Parameter *cost* sebagai parameter yang digunakan untuk menghindari misklasifikasi pada setiap sampel dalam *training* dataset. Parameter *Gamma dapat* menentukan seberapa jauh pengaruh dari satu sampel *training* dataset

dengan nilai rendah berarti “jauh”, dan nilai tinggi berarti “dekat”. Jika nilai *gamma* rendah, maka titik yang berada jauh dari garis pemisah yang logis dalam perhitungan untuk garis pemisah. Sedangkan, ketika *gamma* tinggi berarti titik-titik berada di sekitar garis yang logis akan dipertimbangkan dalam perhitungan (Meyer & Wien, 2015).

Penggunaan *kernel* dapat dibedakan sesuai dengan data yang digunakan. *kernel* linear digunakan pada saat data yang akan diklasifikasikan dapat dipisahkan oleh *hyperplane* berbentuk garis. Dalam artian lain, *kernel* linear digunakan pada data berdimensi dua. Sebaliknya, *kernel* non-linear digunakan pada data yang dipisahkan oleh *hyperplane* berbentuk bidang di ruang berdimensi tinggi (Puspitasari, dkk., 2018).

2.7 Evaluasi Model

Pengujian atau evaluasi model bertujuan untuk menilai seberapa baik kinerja dari model pada tahapan pembelajaran menggunakan *kernel* SVM. Ada berbagai metode untuk mengukur kinerja model, salah satunya adalah menggunakan *confusion matrix*. *Confusion matrix* menggambarkan jumlah data *testing* yang diklasifikasikan dengan benar dan salah(Indriani, 2014).

Tabel 1. *Confusion Matrix*

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

a. *True Postive* (TP) adalah data diprediksi positif dan data sebenarnya adalah positif.
b. *True Negative* (TN) adalah data diprediksi negatif dan data sebenarnya adalah negatif.

c. *False Positive* (FP) adalah data diprediksi positif dan data sebenarnya adalah negatif.
d. *False Negative* (FN) adalah data diprediksi negatif dan data sebenarnya adalah positif.

Berdasarkan hasil dari *confusion matrix*, performa model dapat diukur melalui beberapa metrik, seperti *accuracy*, *precision, recall* (*sensitivity/true positive rate*), dan $f1-score$ (Saputro dan Sari, 2019). Masing-masing perhitungannya didefinisikan sebagai berikut:

a. *Accuracy*, adalah proporsi dari data yang terklasifikasikan dengan benar terhadap keseluruhan data. Pengukuran ini digunakan untuk mengevaluasi tingkat kebenaran klasifikasi (Hamel, 2009). Semakin tinggi nilai akurasi, semakin baik pula kualitas klasifikasi yang dihasilkan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.21)$$

Akurasi klasifikasi tidak mencukupi sebagai ukuran kriteria standar, terutama dalam kasus ketidakseimbangan antar kelas. Ini karena akurasi cenderung memberikan hasil yang baik hanya untuk kelas mayoritas, sedangkan prediksi untuk kelas minoritas menjadi buruk.

b. *Precision*, yaitu melihat seberapa sering model memprediksi positif dan secara aktual prediksi itu benar dengan perumusan sebagai berikut:

$$Precision = \frac{TP}{FP+TP} \quad (2.22)$$

c. *Recall*, adalah seberapa sering model memprediksi positif pada data yang memiliki klasifikasi aktual yang positif dengan perumusan sebagai berikut:

$$Recall = \frac{TP}{FN+TP} \quad (2.23)$$

d. *F1-score,* yaitu merupakan hubungan antara data berlabel positif dari hasil klasifikasi yang menunjukan keseimbangan antara *precision* dan *recall*

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (2.24)$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2023/2022 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 *liver* yang diambil dari website https://www.kaggle.com/ datasets/uciml/indian-liver-patient-records. Jumlah data yang digunakan pada penelitian ini yaitu sebanyak 583 data dan terdapat 10 variabel yaitu, variabel *age*, *total bilirubin* variabel (TB), *direct bilirubin* variabel (DB), *alkaline phosphotase* variabel (AP), *alamine aminotransferase* variabel (AA), *aspartate aminotransferase* variabel (Asp A), *total protiens* variabel (TP), variabel *albumin*, *Albumin and Globulin* (AG) dan diagnosis variabel y.

Dimana:

Age : Tahun ; *Total Bilirubin* (TB) : mg/dL ; *Direct Bilirubin* (DB) : mg/dL ; *Alkaline Phosphotase* (AP) : IU/L ; *Alamine Aminotransferase* (AA) : IU/L ; *aspartate aminotransferase* (Asp A) : IU/L ; *total protiens* : g/dL ; *albumin* : g/dL ; *Albumin and Globulin* : *Ratio*.

3.3 Metode Penelitian

Langkah-langkah yang dilakukan pada penelitian ini sebagai berikut :

1. Melakukan visualisasi data untuk menggambarkan dan mendeskripsikan data yang digunakan. Pada tahap ini ditampilkan *bar chart* untuk menunjukkan perbandingan jumlah antara penderita penyakit *liver* dan tidak penderita penyakit *liver*. Selain itu dilakukannya analisis eksplorasi data dengan statistik deskriptif.
2. Melakukan *preprocessing* data, yaitu :
    a. Melakukan *cleaning data* untuk mendeteksi keberadaan data hilang atau data duplikat.
    b. *Scaling Data*, melakukan transformasi data menggunakan *standard scaler*.
    c. *Handling Data Categorical*¸ dengan memberikan label pada data yang berbentuk kategorik dengan menggunakan *one hot encoding.*
3. *Handling Imbalance Data*

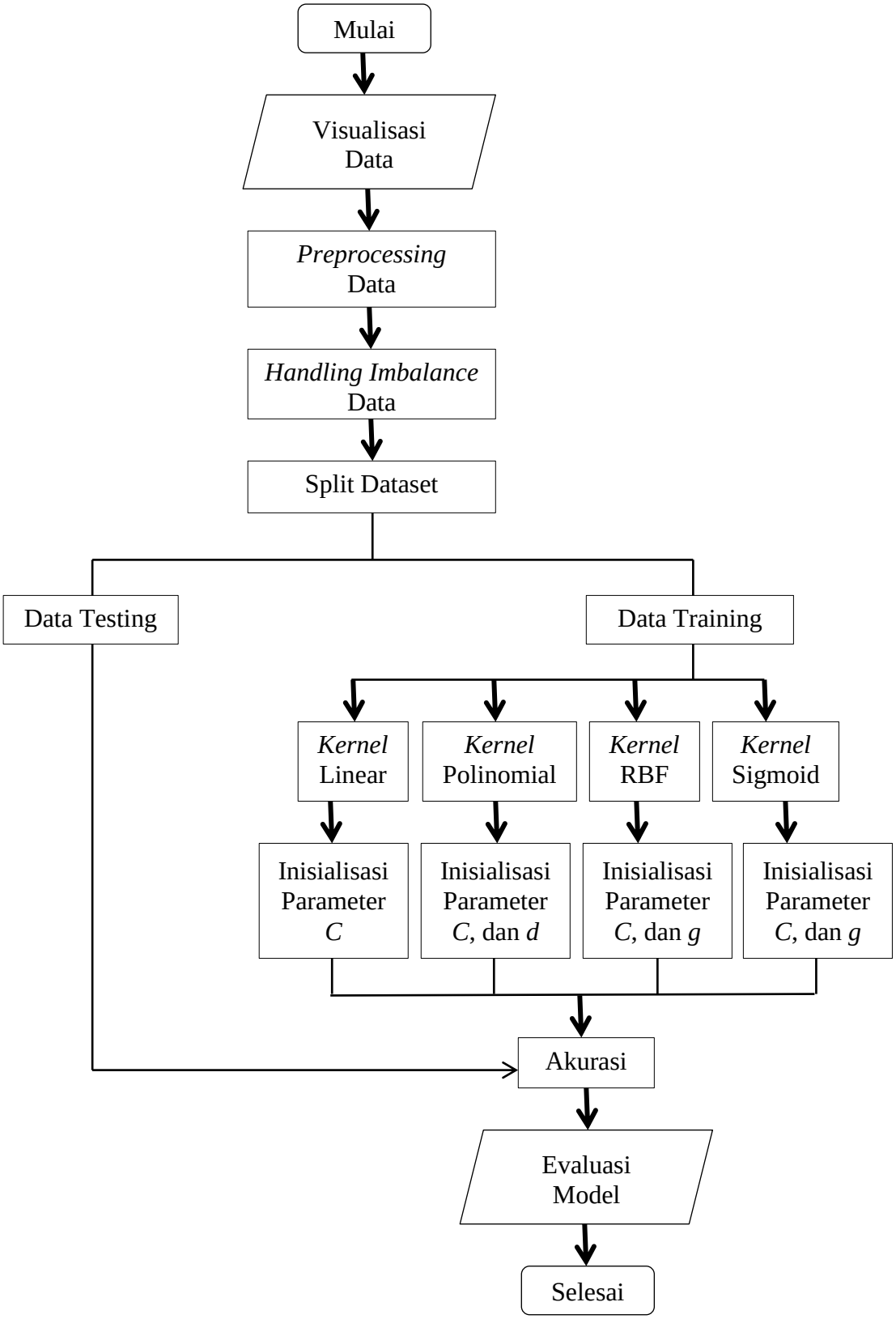
    Mengatasi *imbalance* data dengan menggunakan *Random Oversampling* (ROS). *Random Oversampling* bertujuan untuk meningkatkan ukuran kelas minoritas dengan mensintesis sampel baru atau dataset *training* dengan menduplikasi secara acak sampel kelas minoritas.
4. Melakukan *splitting data* menjadi 2 (*training* dan *testing*) dengan 4 skema yaitu, 60% data *training* dan 40% data *testing*, skema 70% data *training* dan 30% data *testing*, skema 80% data *training* dan 20% data *testing*, skema 90% data *training* dan 10% data *testing* yang diambil secara acak dari dataset penelitian.
5. Melakukan proses klasifikasi SVM dengan fungsi *kernel* Linear, Polynomial, Sigmoid, dan *Radial Basis Function* RBF.
    a. Proses Klasifikasi SVM dengan fungsi *kernel* Linear
        - Menetapkan fungsi yang akan digunakan untuk klasifikasi SVM dengan *kernel* Linear.

- Menentukan parameter yang relevan untuk fungsi *kernel* yang dipilih.
- Membangun model SVM dengan menggunakan fungsi *kernel* Linear.
- Menghasilkan *confusion matrix* dan menghitung kinerja klasifikasi berdasarkan ukuran akurasi, *precision, recall*, dan *F1-score*.

b. Proses Klasifikasi SVM dengan fungsi *kernel* Polynomial
   - Menetapkan fungsi yang akan digunakan untuk klasifikasi SVM dengan *kernel* Polynomial.
   - Menentukan parameter yang relevan untuk fungsi *kernel* yang dipilih.
   - Membangun model SVM dengan menggunakan fungsi *kernel* Polynomial.
   - Membentuk *confusion matrix* dan menghitung performa klasifikasi berdasarkan ukuran akurasi, *precision, recall*, dan *F1-score.*

c. Proses Klasifikasi SVM dengan fungsi *kernel* Sigmoid
   - Menetapkan fungsi yang akan digunakan untuk klasifikasi SVM dengan *kernel* Sigmoid.
   - Menentukan parameter ya.ng relevan untuk fungsi *kernel* yang dipilih.
   - Membangun model SVM dengan menggunakan fungsi *kernel* Sigmoid.
   - Membentuk *confusion matrix* dan menghitung performa klasifikasi berdasarkan ukuran akurasi, *precision, recall*, dan *F1-score*.

d. Proses Klasifikasi SVM dengan fungsi *kernel Radial Basis Function* RBF
   - Menetapkan fungsi yang akan digunakan untuk klasifikasi SVM dengan *kernel* RBF.
   - Menentukan parameter ya.ng relevan untuk fungsi *kernel* yang dipilih.
   - Membangun model SVM dengan menggunakan fungsi *kernel* RBF.
   - Membentuk *confusion matrix* dan menghitung performa klasifikasi berdasarkan ukuran akurasi, *precision, recall*, dan *F1-score*.



Gambar 8. Diagram Alir untuk Tahapan Analisis Data Penyakit *Liver*.

V. KESIMPULAN

Setelah melakukan proses *machine learning* dengan menggunakan *Random Oversampling* (ROS) yang digunakan pada metode *Support Vector Machine* (SVM) untuk menyeimbangkan data pada klasifikasi penderita *liver*. Diperoleh hasil *kernel* Linear memiliki tingkat akurasi terbaik pada *split* dataset 90% dengan parameter *cost* = 5, *kernel* Polinomial memiliki tingkat akurasi terbaik pada *split* dataset 90% dengan parameter *cost* = 12 dan *degree* = 3, *kernel* RBF memiliki tingkat akurasi terbaik pada *split* dataset 90% dengan paraemter *cost* = 12 dan *gamma* = 2,5, dan *kernel* Sigmoid memiliki tingkat akurasi terbaik pada *split* dataset 70% dengan parameter *cost* = 2 *gamma* = 0,1. Berdasarkan hasil yang diperoleh dari keempat fungsi *kernel*, *kernel* RBF memiliki tingkat akurasi tertinggi dibandingkan dengan *kernel* lainnya, dengan parameter terbaik yaitu c*ost* = 12, *degree* = 3, dan *gamma* = 2,5. Pada hasil klasifikasi SVM dengan skema data 90% data *training* dan 10% data *testing* dengan menggunakan *Kernel* RBF diperoleh nilai akurasi sebesar 84,63%. Dengan menggunakan fungsi *kernel* RBF dan proporsi *split* dataset tersebut dapat diperoleh parameter **w** dan b, yakni sebagai berikut:

w_{Age} = -5,1951, w_{TB} = -7,3708, w_{DB} = -9,1687, w_{AP} = -9,2968, w_{AA} = -10,6287, $w_{Asp\ A}$ = -10,3948, w_{TP} = 0,1057, $w_{Albumin}$ = 9,3402, $w_{AG\ R}$ = 2,5066 , b = 0,8868.

DAFTAR PUSTAKA

Adinegoro, A., Atmaja, R.D., & Purnamasari, R. 2015. Deteksi Tumor Otak dengan Ektrasi Ciri dan Feature Selection Mengunakan Linear. *Proceeding of Engineering*. **2**(2): 2532.

Al-Mejibli, I.S., Alwan, J.K., & Abd, D.H. (2020). The effect of gamma value on support vector machine performance with different kernels. *International Journal of Electrical and Computer Engineering*. **10**(5):5497–5506.

Barro, R.A., Sulvianti, I.D., & Afendi, F.M. 2013. Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. *Journal of Statistics*. **1**(1).

Ben-Hur, A., & Weston, J. 2010. A User's Guide to Support Vector Machines. *Methods in Molecular Biology* (Clifton, N. J.). **609**:223-239.

Buulolo, E. 2020. *Data Mining Untuk Perguruan Tinggi*. Deepublish Publisher.

Brownlee, J. 2016. *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. Melbourne. Australia: Jason Brownlee.

Cortes. C, & Vapnik, V. 1995. Support Vector Network, Machine Learning. *Kluwer Academic Publisher*. 20(3):273-297.

Das, S., & Nene, M.J. 2017. A Survey on Types of Machine Learning Techniques in Intrusion Prevention Systems, hlm. 2296-2299. Proceeding of International Conference on Wireless Communications.

Dhage, S.N. & Raina, C.K. 2016. A review on Machine Learning Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*. **4**(3):395-399.

Ginanjar, A.R., & Feta, N.R. 2019. Komparasi Fungsi Kernel Metode Support Vector Machine Untuk Pemodelan Klasifikasi Terhadap Penyakit Tanaman Kedelai. *BRITech*. **1**(1):33-39.

Goh, R.Y., & Lee, L.S. 2019. Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches. *Hindawi: Advances in Operations Research.* **2019**:1-31.

Goldberg, D.E., & Holland, J.H. 1988. Genetic algorithms and machine learning. *Machine Learning*. **3**(2):95-99.

Gullo, F. 2015. From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. *Physics Procedia*. **62**:18–22.

Gorunescu, F. 2011. *Data Mining : Concept, Model and Techniques*. Springer. Berlin.

Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts and Techniques* 3 rd Ed. Morgan Kaufmann. USA.

Hamel, L. 2009. Knowledge Discovery with Support Vector Machines. *In Knowledge Discovery with Support Vector Machines*.

Hermawati, F.A. 2013. *Data Mining*. Edisi ke-1. Andi Offset. Yogyakarta.

He, H., Zhang, W., & Zhang S. 2018. A Novel Ensemble Method for Credit Scoring: Adaption of Different Imbalance Ratios. *Expert Systems with Applications*. **98**:105–117.

Huang, G.B., Zhu, Q.Y., & Siew, C.K. 2006. Extreme learning machine: theory and applications. *Neurocomputing*. **70**(1–3):489-501.

Indriani, A. 2014. Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier, hlm. 5-9. Prosiding Seminar Nasional Aplikasi Teknologi Informasi.

Jian, C., Gao, J., & Ao, Y. 2016. A new sampling method for classifying imbalanced data based on support vector machine ensemble. *Journal of Neurocomputing*. **193**: 115–122.

Kotsiantis, S.B. 2007. Supervised Machine Learning: A Review of Classification Techniques. *Informatica*. **31**(3):249-268.

Kowalczyk, A. 2017. *Support vector machines succinctly*. Syncfusion Inc. USA.

Meyer, D., & Wien, F.T. 2015. Support vector machines. *The Interface to libsvm in package e1071*. **28**:20.

Nachev, A., & Teodosiev, T. 2015. Using Support Vector Machines for Direct Marketing Models. *IJEAT*. **4**(4):183-190.

Nanda, M.A., Seminar, K.B., Nandika, D., & Maddu, A. 2018. A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection. *Information*. **9**(1):1-14.

Naufal, A.R., Wahono, R.S., & Syukur, A. 2015. Penerapan Bootstrapping untuk Ketidakseimbangan Kelas dan Weighted Information Gain untuk Feature Selection pada Algoritma Support Vector Machine untuk Prediksi Loyalitas Pelanggan. *Journal of Intelligent Systems*. **1**(2):98-108.

Nasteski, V. 2017. An overview of the supervised machine learning methods. *Horizons.B.* **4**(12):1-11.

Nugraha, Y.R., Wibawa, A.P., & Zaeni, I.A.E. 2019. Particle Swarm Optimization-Support Vector Machine (PSO-SVM) Algorithm for Journal Rank Classification, hlm. 69-73. Proceedings - 2019 2nd International Conference of Computer and Informatics Engineering: Artificial Intelligence Roles in Industrial Revolution 4.0, IC2IE 2019.

Praghakusma, A.Z., & Charibaldi, N. 2021. Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi). *Jurnal Sarjana Teknik Informatika*. **9**(2):33-42.

Puspitasari, A.M., Ratnawati, D.E., & Widodo, A.W. 2018. Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **2**(2):802-810.

Saputro, I.W. & Sari, B.W. 2019. Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*. **6**(1):1-11.

Sir, Y.A. & Soepranoto, A.H.H. 2022. Pendekatan Resampling Data untuk Menangani Masalah Ketidakseimbangan Kelas. *J-ICON*. **10**(1): 31-38.

Shwartz, S., & David, S.B. 2014. *Understanding Machine Learning From Theory to Algorithm.* Cambridge University Press. New York.

Somvanshi, M., & Chavan, P. 2016. A Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine. 2016 *International Conference on Computing Communication Control and Automation* (ICCUBEA). 1-7.

Sumiran, K. 2018. An Overview of Data Mining Techniques and Their Application in Industrial Engineering. *Asian Journal of Applied Science and Technology*. **2**(2):947-953.

Tan, P., Steinbach, M., & Kumar, V. 2006. *Introduction to Data Mining*. KMedia. Yogyakarta.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag. New York.

Yu, D., Hu, J., Tang, Z., & Shen, H. 2017. Neurocomputing Improving proteinATP binding residues prediction by boosting SVMs with random undersampling. *Journal of Neurocomputing*. **104**:180–190.

Zaki, M.J. 2014. *Data Mining and Analysis : Fundamental Concepts and Algorithms*. Cambridge University Press 32. New York.