

Metode *Support Vector Machine* Untuk Klasifikasi Data Penyakit Hati Yang *Imbalance*

Support Vector Machine Method for Classification Imbalance Liver Disease Data

Nur Rahmah^{a,1,*}, Purnawansyah^{a,2}, dan Fitriyani Umar^{a,3}

^a Program Studi Informatika Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231
¹ nurrahma2002@gmail.com; ² purnawansyah@umi.ac.id; ³ fitriyani.umar@umi.ac.id;
*corresponding author

| Informasi Artikel | ABSTRAK |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Diserahkan : 2 Maret 2024 Diterima : 29 April 2024 Direvisi : 30 April 2024 Diterbitkan : 30 April 2024</p> <p>Kata Kunci: Support Vector Machine Penyakit Hati Data Tidak Seimbang</p> | <p>Penelitian ini bertujuan untuk melakukan klasifikasi penyakit hati menggunakan metode <i>Support Vector Machine</i> (SVM) dan <i>Synthetic Minority Oversampling Technique</i> (SMOTE). Data yang digunakan berupa data sekunder yang diperoleh dari situs <i>Kaggle</i> dengan jumlah data sebanyak 582 sampel. Data tersebut terdiri dari 10 fitur yang digunakan sebagai variabel masukan SVM. Proses klasifikasi dilakukan dengan membagi data menjadi data training 70% dan data testing 30%. Hasil dari penelitian yang telah dilakukan ialah dengan menggunakan metode <i>support vector machine</i> mampu melakukan klasifikasi data penyakit hati dengan hasil klasifikasi yang menunjukkan nilai 0 dan 1. Dimana nilai 0 menandakan bahwa pasien tersebut tidak mengidap penyakit hati dan nilai 1 menandakan bahwa pasien tersebut mengidap penyakit hati. Berdasarkan proses klasifikasi data penyakit hati yang telah dilakukan memperoleh nilai akurasi performansi yaitu 67,06%, dan berdasarkan proses visualisasi data yang telah dilakukan dalam proses pengklasifikasian data tersebut ditemukan ketidakseimbangan data penyakit hati. Ketidakseimbangan data yang peroleh dilakukan <i>oversampling</i> menggunakan metode SMOTE untuk menyeimbangkan data. Penelitian telah melakukan proses penyeimbangan data penyakit hati sehingga tenaga medis lebih terbantuan dalam mendeteksi penyakit hati yang diderita oleh pasien.</p> |
| <p>Keywords: <i>Support Vector Machine</i> <i>Liver Disease</i> <i>Imbalance Data</i></p> | <p>ABSTRACT</p> <p><i>This research aims to classify liver disease using Support Vector Machine (SVM) and Synthetic Minority Oversampling Technique (SMOTE). The data used is secondary data obtained from the Kaggle website with a total of 582 samples. With a total of 582 samples. The data consists of 10 features which are used as SVM input variables. The classification process is carried out by dividing the data into 70% training data and 30% testing data. The result of research that has been done is to use the support vector machine method is able to classify liver disease data with classification results showing values 0 and 1. Where the value 0 indicates that the patient does not have liver disease and the value 1 indicates that the patient does not have liver disease. Where the value 0 indicates that the patient does not have liver disease and the value 1 indicates that the patient has liver disease. The patient has liver disease. Based on the data classification process data classification process that has been carried out obtains a performance accuracy value of 67.06%, and based on the data visualization process that has been carried out in the data classification process. The data classification process, there is an imbalance of liver disease data. Liver disease data. The imbalance of data obtained is oversampling using the SMOTE method to balance the data. Research has done the process of balancing liver disease data so that the medis are more assisted in detecting liver disease in detecting liver disease suffered by patients.</i></p> |

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. Pendahuluan

Pada bidang kesehatan, perkembangan ilmu kedokteran mengalami kemajuan yang sangat pesat yang ditandai dengan ditemukannya penyakit baru yang belum teridentifikasi sebelumnya. Salah satu penyakit yang saat ini yaitu penyakit pada organ hati. Salah satunya adalah penyakit hepatitis [1]. Hati merupakan organ vital manusia yang memiliki fungsi kompleks dan beragam seperti menawarkan dan menetralkan zat-zat racun yang tidak bisa diserap oleh usus, menyaring darah yang datang dari usus melalui vena porta, kemudian menyimpan dan mengubah bahan makanan dari vena porta untuk selanjutnya bahan makanan tersebut dikirim ke dalam darah sesuai dengan kebutuhan [2]. Berdasarkan data dari WHO, penyakit yang memiliki kasus paling banyak menyerang hati manusia adalah hepatitis dan sirosis. Penyakit hati yang sudah akut akan mempengaruhi fungsi-fungsi hati, tetapi penyakit hati tersebut dapat diketahui gejala klinis maupun fisik yang timbul pada pasien. Gejala klinis dapat diketahui dari apa yang dirasakan oleh pasien, sedangkan gejala fisik dapat diketahui dari keadaan tubuh pasien. Gejala penyakit hati ada banyak dan kompleks, serta penyakit hati memiliki kemiripan gejala dengan beberapa penyakit [3].

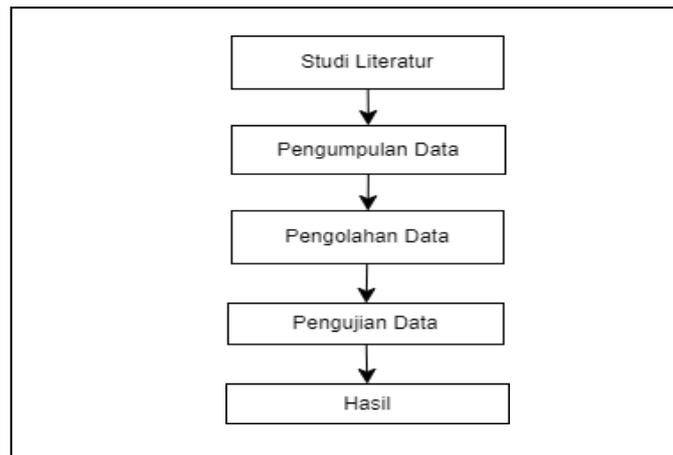
Hati merupakan salah satu organ yang cukup besar dan penting pada tubuh kita. Bagian yang penting pada hati ini terdiri dari hepatosit, yang merupakan sel epitel dengan konfigurasi yang unik. Pada dasarnya hati ini merupakan kelenjar eksorin, oleh karena mensekresi cairan empedu yang dialirkan ke dalam duodenum. Selain itu juga merupakan kelenjar endokrin dan penyaring darah [4]. SVM adalah teknik *supervised learning* yang mempunyai tingkat akurasi dan kualitas yang baik sehingga membuatnya menjadi sangat diminati diantara algoritma yang lainnya. Akan tetapi, untuk implementasinya diperlukan tahap pelatihan *sequential training* dan harus melalui proses pengujian. Kelebihan metode SVM adalah kemampuannya mengidentifikasi *hyperplane* yang terpisah sehingga bisa memaksimalkan margin dari kelas yang berbeda [5]. SVM atau Support Vector Machine merupakan salah satu metode terbaik yang dapat digunakan dalam masalah klasifikasi pola. Metode SMOTE merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan cara membuat *instance* baru dari *minority class* dengan pembentukan *convex* kombinasi dari *instances* yang saling berdekatan [6]. Penelitian sebelumnya yang dilakukan oleh Nurdin, dkk pada tahun 2022 mengenai klasifikasi data daftar ulang calon mahasiswa Universitas Islam Negeri Sunan Ampel Surabaya. Bertujuan untuk membandingkan hasil klasifikasi performa metode SMOTE dan ADASYN dalam menangani imbalanced data pada data daftar ulang calon mahasiswa Universitas Islam Negeri Sunan Ampel Surabaya. Hasil menunjukkan bahwa metode SMOTE dapat menyeimbangkan data secara seimbang dibandingkan dengan ADASYN. Dari hasil pengujian metode SMOTE lebih cocok digunakan daripada metode ADASYN karena nilai *ROCAUC* SMOTE lebih tinggi dari ADASYN [7]. Penelitian yang dilakukan oleh Setiawati, dkk pada tahun 2019 yaitu membandingkan metode klasifikasi *Support Vector Machine* (SVM) dengan *Naïve Bayes Classifier* (NBC) untuk mengolah ILPD dataset. Perbandingan 2 metode tersebut adalah akurasi klasifikasi dan waktu eksekusi dalam melakukan klasifikasi penyakit lever. Hasil penelitian tersebut menunjukkan bahwa dalam hal akurasi klasifikasi Metode SVM menghasilkan nilai akurasi lebih baik dari pada NBC, sedangkan dalam hal waktu pemrosesan NBC lebih cepat dari pada SVM [8]. Pada penelitian yang dilakukan oleh Nuriana pada tahun 2021 bertujuan untuk membandingkan metode yang tepat pada dataset penyakit hati. Dataset tersebut memiliki atribut berbeda-beda sehingga penentuan klasifikasi dapat dilakukan secara akurat. klasifikasi tersebut yaitu Naïve Bayes, KNN, dan SVM. Performa dari ketiga metode tersebut dibandingkan untuk mendapatkan metode yang terbaik untuk dataset hati, dengan cara menentukan nilai performanya yang berupa nilai *accuracy*, *precision*, *recall*, dan *F-measure*. Hasil penelitian menunjukkan bahwa dari ketiga algoritma, SVM memiliki rata-rata performa paling baik dengan akurasi sebesar 82,36% [9].

Berdasarkan pendahuluan di atas maka dalam penelitian ini membahas mengenai prediksi penyakit hati dengan menggunakan metode *Support Vector Machine* (SVM) dan SMOTE digunakan untuk mengatasi permasalahan ketidakseimbangan kelas. Penelitian ini diharapkan dapat mengatasi masalah pendiagnosis penyakit hati sejak dini dan juga dapat memberikan hasil implementasi prediksi berupa nilai akurasi dan performansi. Penelitian ini mengambil judul “Klasifikasi Penyakit Hati Menggunakan Metode *Support Vector Machine* (SVM) dan SMOTE”. Secara umum dalam proses klasifikasi menggunakan data *training* dan data *testing*.

II. Metode

A. Tahapan Penelitian

Tahapan penelitian ini berisikan tahapan-tahapan penelitian yang akan dilakukan, tahapan penelitian ini berdasarkan tahapan metode penelitian yang tercantum dibawah ini :



Gambar 1. Tahap penelitian

Penjelasan :

1. Studi Literatur

Pada tahap ini, merupakan salah satu teknik yang digunakan untuk mencari ide atau sumber referensi dalam penelitian sama seperti penelitian pada umumnya, yaitu dengan cara mengumpulkan dan mempelajari literatur - literatur yang saling berkaitan ataupun saling mendukung dalam metode SVM.

2. Pengumpulan Data

Untuk pengumpulan data pada penelitian ini menggunakan data sekunder, yang dimana data sekunder adalah data yang telah dibuat atau dikumpulkan oleh orang lain yang dapat digunakan untuk tujuan penelitian yang diperoleh dengan cara riset kepustakaan, membaca buku atau jurnal yang berkaitan dengan masalah klasifikasi data.

3. Pengolahan Data

Pengolahan data adalah proses pengolahan data mentah menjadi informasi yang berguna dan mudah diterima. Data mentah biasanya berupa angka atau catatan yang tidak memiliki arti bagi pengguna, sehingga membutuhkan proses pengolahan untuk mengubahnya menjadi informasi berguna menggunakan teknik dan metode SVM.

4. Pengujian Data

Pada fase ini akan dilakukan pengujian terhadap sistem yang telah dibuat pada tahap klasifikasi. Pengujian ini dilakukan untuk mengukur keberhasilan sistem dengan mengukur tingkat akurasi. Proses pengujian dilakukan dengan cara pengujian akurasi.

5. Hasil

Hasil dapat diperoleh Ketika seluruh tahapan telah diselesaikan terlebih dahulu mulai tahap studi literatur, pengumpulan data, pengolahan data, pengujian data, hasil pengujian hingga klasifikasi metode SVM.

B. Tinjauan Pustaka

1. Klasifikasi

Klasifikasi merupakan suatu proses yang bertujuan untuk menentukan suatu objek kedalam suatu kelas atau kategori yang sudah ditentukan sebelumnya. Klasifikasi adalah proses dari pembangunan terhadap suatu model yang mengklasifikasikan suatu objek sesuai dengan atribut – atributnya. Klasifikasi data ataupun dokumen juga dapat dimulai dari membangun aturan klasifikasi tertentu yang menggunakan data training yang sering disebut sebagai tahapan pembelajaran dan pengujian digunakan sebagai data testing. Beberapa tugas dari klasifikasi yang melibatkan proses pembangunan terhadap model yang dibentuk untuk melakukan prediksi target atau variabel dari data set yang sudah jelas, ataupun variabel independent [10].

2. Penyakit Hati

Hati merupakan organ dalam tubuh manusia yang memiliki peran sangat penting. Peran hati adalah sebagai penetral racun yang masuk ke dalam tubuh melalui makanan ataupun dari lingkungan. Tanpa adanya hati manusia tidak akan mampu bertahan hidup. Kerusakan pada organ hati mampu membuat kemampuan tubuh manusia terganggu pada saat memecah sel darah merah dari racun- racun yang terkandung di dalamnya. Akibatnya racun-racun tersebut akan menetap pada tubuh kita [11].

Kerusakan hati yang parah dapat dikenali dengan perubahan warna bola mata dan kulit menjadi kuning dan juga membuat air seni atau kencing menjadi gelap. Hati yang telah rusak akan berdampak pada kemampuan tubuh dalam memecah protein. Faktor penyebab kerusakan organ hati manusia ialah terlalu banyak mengkonsumsi obat antibiotik, penyalahgunaan narkoba, kebiasaan merokok dan minum minuman beralkohol, serangan virus hepatitis, dan terlalu banyak makan dan minum yang mengandung kolesterol tinggi [12].

3. *Support Vektor Machine*

Support Vector Machine dikenalkan pertama kali oleh Vapnik tahun 1992 sebagai salah satu metode *learning machine* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*. Metode ini menggunakan hipotesis berupa fungsi – fungsi linier dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan.

4. *Synthetic Minority Oversampling Technique* (SMOTE)

Ketidakeimbangan data terjadi jika jumlah objek suatu kelas data lebih banyak dibandingkan dengan kelas lain. Kelas data yang objeknya lebih banyak disebut kelas mayor sedangkan lainnya disebut kelas minor. Pengaruh penggunaan data tidak seimbang untuk membuat model sangat besar pada hasil model yang diperoleh. Pengolahan algoritma yang tidak menghiraukan ketidakseimbangan data akan cenderung diliputi oleh kelas mayor dan mengacuhkan kelas minor.

Metode SMOTE diusulkan oleh sebagai salah satu solusi dalam menangani data tidak seimbang dengan prinsip yang berbeda dengan metode *over sampling* yang telah diusulkan sebelumnya. Apabila metode *over sampling* berprinsip memperbanyak pengamatan secara acak, metode SMOTE menambah jumlah data kelas minor agar setara dengan kelas mayor dengan cara membangkitkan data buatan [16].

5. *Confusion Matrix*

Confusion matrix memberikan keputusan yang diperoleh dalam training dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi [8].

Confusion Matrix

| <i>Predicted Values</i> | <i>Actual Service</i> | |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | 1 | 0 |
| 1 | <p style="text-align: center;"><i>True Positive^a</i></p> <p>Contoh : Seorang dokter memberitahukan kepada seorang pasien wanita bahwa saat ini ia sedang hamil</p> | <p style="text-align: center;"><i>False Negative^b</i></p> <p>Contoh : Seorang dokter memberitahukan kepada seorang pasien pria bahwa saat ini ia sedang hamil (<i>Type 1 Error</i>)</p> |
| 0 | <p style="text-align: center;"><i>False Positive^c</i></p> <p>Contoh : Seorang dokter memberitahukan kepada seorang pasien wanita bahwa saat ini ia tidak hamil (<i>Type 2 Error</i>)</p> | <p style="text-align: center;"><i>True Negative^d</i></p> <p>Contoh : Seorang dokter memberitahukan kepada seorang pasien pria bahwa saat ini ia tidak hamil</p> |
| ^a | True Positive (TP) : Proporsi positif dalam data set yang diklasifikasikan positif | |
| ^b | False Negative (FN) : Proporsi positif dalam data set yang diklasifikasikan negatif | |
| ^c | False Positive (FP) : Proporsi negatif dalam data set yang diklasifikasikan positif | |
| ^d | True Negative (TN) : Proporsi positif dalam data set yang diklasifikasikan negatif | |

Berikut adalah persamaan medol *confusion matrix*

- 1) Nilai *accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan (1) :

$$Accuracy = \frac{P+N}{P+N+P+N} \quad (1)$$

- 2) *Precision* adalah rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif, yang dihitung menggunakan persamaan (2) :

$$Precision = \frac{P}{P+P} \quad (2)$$

- 3) *Recall* (Sensitifitas) adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data yang bernilai positif, yang dihitung dengan menggunakan persamaan (3) :

$$Recall = \frac{P}{P+N} \quad (3)$$

III. Hasil dan Pembahasan

A. Hasil Penelitian

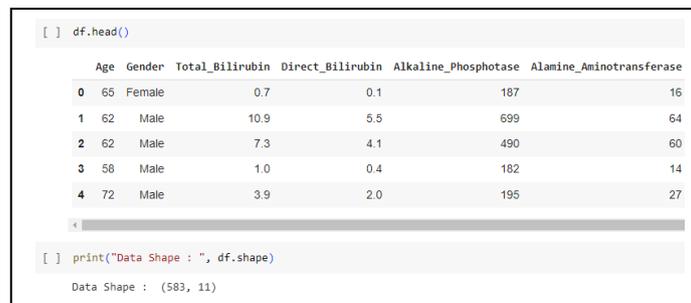
Sebelum melakukan proses pengklasifikasian data menggunakan metode *Support Vector Machine* (SVM), hal yang paling utama dilakukan pada pengolahan data penyakit hati ialah melakukan pengecekan keseimbangan hasil data penyakit hati. Apabila ditemukan hasil data penyakit hati memiliki rasio yang tidak seimbang maka peneliti melakukan proses penyeimbangan data menggunakan metode SMOTE kemudian melakukan proses pengklasifikasian data penyakit hati menggunakan metode SVM. Adapun tahapan pengecekan keseimbangan data dapat dilihat pada langkah-langkah berikut ini.

1. Penginputan Dataset

Data pada penelitian ini diperoleh dari situs *website* www.kaggle.com secara gratis. Proses pengambilan data dari *website* umumnya melibatkan penggunaan *Application Programming Interface* (API) atau dapat dilakukan dengan melakukan pengunduhan langsung pada situs web. Kemudian data yang telah diperoleh diinputkan kedalam *google colab*.

2. Understanding data

Setelah melakukan penginputan dataset, langkah selanjutnya ialah melakukan *understanding* data untuk membantu pemilihan kernel yang sesuai dengan karakteristik data yang ingin diolah. Pada penelitian ini kernel yang digunakan ialah kernel linear karena kernel linear merupakan jenis kernel yang cocok dengan jenis data linear, yang dimana data liner mengacu pada kasus yang memiliki batasan keputusan yang hanya terdiri atas dua kelas. Adapun proses *understanding* data dapat dilihat pada gambar dibawah ini.



```
[ ] df.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase |
|---|-----|--------|-----------------|------------------|----------------------|--------------------------|
| 0 | 65 | Female | 0.7 | 0.1 | 187 | 16 |
| 1 | 62 | Male | 10.9 | 5.5 | 699 | 64 |
| 2 | 62 | Male | 7.3 | 4.1 | 490 | 60 |
| 3 | 58 | Male | 1.0 | 0.4 | 182 | 14 |
| 4 | 72 | Male | 3.9 | 2.0 | 195 | 27 |

```
[ ] print("Data Shape : ", df.shape)
```

Data Shape : (583, 11)

Gambar 2. Heading dan data shape

Pada gambar diatas menjelaskan proses *heading* dan *shape* data yang bertujuan untuk menampilkan lima data teratas serta menampilkan jumlah baris dan kolom pada data yang ingin diklasifikasikan. Pada proses diatas peneliti mampu menentukan kernel yang sesuai untuk diterapkan pada proses pengklasifikasian dimana dataset penelitian hanya memiliki dua batasan Keputusan yaitu penderita penyakit hati atau bukan penderita penyakit hati.



```
df.Dataset.value_counts()
```

| | |
|---|-----|
| 0 | 416 |
| 1 | 167 |

Name: Dataset, dtype: int64

Gambar 3. Rasio data penyakit hati

Pada gambar 3 menunjukkan penggalan proses *understanding* data yang menampilkan informasi-informasi terkait *value* dari hasil Keputusan yang dimiliki oleh data tersebut. Dimana ditampilkan bahwa dari keseluruhan data yang dimiliki, terdapat 416 hasil yang dinyatakan sehat (0) dan 167 hasil yang dinyatakan sebagai penderita penyakit hati (1). Dari hasil rasio data penyakit hati diatas terdapat rasio data yang memiliki selisih yang cukup tinggi sehingga membuat hasil keputusan dari data penyakit hati menjadi tidak seimbang. Oleh karenanya diperlukan penyeimbangan data menggunakan metode SMOTE sehingga hasil klasifikasi data menggunakan SVM akan lebih optimal.

B. Pembahasan

Pada penelitian ini mengangkat suatu kasus penyeimbangan data penyakit hati menggunakan metode SMOTE dan pengklasifikasian data penyakit hati menggunakan metode SVM agar dapat membantu tenaga medis dalam mengetahui kondisi pasien dengan cepat dan lebih spesifik.

Tahapan pertama yang harus dilakukan oleh peneliti ialah melakukan *preprocessing* data dengan tujuan untuk memastikan model yang digunakan dapat bekerja dengan baik dan memberikan hasil yang optimal. Adapun bagian proses *preprocessing* data ialah *cleaning* data, numerisasi data, dan standarisasi data.

```
nan_data = df[df.isna().any(axis=1)]
nan_data
```

| ilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotransferase |
|----------|------------------|----------------------|--------------------------|----------------------------|
| 0.9 | 0.3 | 189 | 23 | 33 |
| 0.8 | 0.2 | 230 | 24 | 46 |
| 0.6 | 0.2 | 180 | 12 | 15 |
| 1.3 | 0.6 | 106 | 25 | 54 |

```
[ ] df = df.dropna()
```

Gambar 4. Memfilter data not a number (NaN)

```
df.drop_duplicates(keep="first")
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase |
|---|-----|--------|-----------------|------------------|----------------------|
| 0 | 65 | Female | 0.7 | 0.1 | 187 |
| 1 | 62 | Male | 10.9 | 5.5 | 699 |
| 2 | 62 | Male | 7.3 | 4.1 | 490 |
| 3 | 58 | Male | 1.0 | 0.4 | 182 |
| 4 | 72 | Male | 3.9 | 2.0 | 195 |

Gambar 5. Menghapus duplikasi baris

Pada gambar diatas menjelaskan proses *cleaning* data, yang bertujuan untuk menghilangkan *missing value* dan data duplikat agar proses pengklasifikasian data dapat berjalan dengan optimal.

```
[ ] encode = LabelEncoder()
df['Age'] = encode.fit_transform(df['Age'].values)
df['Gender'] = encode.fit_transform(df['Gender'].values)
df['Total_Bilirubin'] = encode.fit_transform(df['Total_Bilirubin'].values)
df['Direct_Bilirubin'] = encode.fit_transform(df['Direct_Bilirubin'].values)
df['Alkaline_Phosphotase'] = encode.fit_transform(df['Alkaline_Phosphotase'].values)
df['Alamine_Aminotransferase'] = encode.fit_transform(df['Alamine_Aminotransferase'].values)
df['Aspartate_Aminotransferase'] = encode.fit_transform(df['Aspartate_Aminotransferase'].values)
df['Total_Proteins'] = encode.fit_transform(df['Total_Proteins'].values)
df['Albumin'] = encode.fit_transform(df['Albumin'].values)
df['Albumin_and_Globulin_Ratio'] = encode.fit_transform(df['Albumin_and_Globulin_Ratio'].values)
df['Dataset'] = encode.fit_transform(df['Dataset'].values)
```

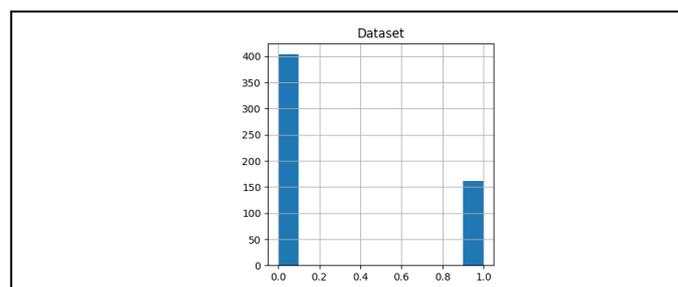
Gambar 6. Numerisasi data

Tahapan selanjutnya setelah proses *cleaning* data ialah numerisasi data, yang merupakan suatu proses mengubah data dalam bentuk non-numerik menjadi representasi numerik sehingga dapat digunakan oleh algoritma *machine learning*.

```
scaler = StandardScaler()
scaler.fit(x)
x = scaler.transform(x)
```

Gambar 7. Standarisasi data

Gambar 7 diatas menjelaskan proses standarisasi data, yang merupakan proses perubahan distribusi nilai fitur dalam dataset sehingga memiliki rata-rata nol dan deviasi standar satu. Tahapan selanjutnya ialah melakukan visualisasi data terhadap dataset agar peneliti memahami karakteristik data, memastikan pemilihan fitur yang tepat dan membuat Keputusan yang lebih baik dalam pemrosesan data dan pemilihan model yang akan digunakan.



Gambar 8. Hasil visualisasi data

Pada gambar diagram diatas dapat dilihat kategori dan frekuensi dari Dataset yang mewakili hasil pengelompokan kategori terhadap para pasien penderita penyakit hati. Dimana dapat dilihat dari hasil visualisasi data di atas bahwa terdapat 2 kategori, yaitu: sehat (0) dan penderita penyakit hati (1) dengan frekuensi sebanyak 416 orang dinyatakan sehat dan 167 orang lainnya dinyatakan sebagai penderita penyakit hati. Selanjutnya masuk pada tahapan *splitting* data, adapun prosesnya dapat dilihat pada gambar 10.

```

] x_train, x_test, y_train, y_test = train_test_split(x,y, test_size= 0.3, random_state=42
] print(x.shape, x_train.shape, x_test.shape)
(566, 10) (396, 10) (170, 10)

```

Gambar 9. *Splitting* data

Pada gambar diatas merupakan proses pembagian data, dimana pembentukan model klasifikasi dilakukan dengan membagi dataset menjadi dua bagian yaitu data training (70%) dan data testing (30%).

Berdasarkan hasil pengecekan data maka ditemukan data yang tidak seimbang pada data penyakit hati yang telah diinput pada google colab, sehingga dibutuhkan proses penyeimbangan data menggunakan metode SMOTE. Hal ini dikarenakan penggunaan data dengan rasio yang tidak seimbang dalam pembangunan sebuah model akan sangat mempengaruhi hasil dari model yang dibangun. Algoritma klasifikasi yang mengabaikan ketidakseimbangan data akan cenderung mengabaikan sampel minoritas dan hanya berfokus pada sampel mayoritas. Dataset dalam penelitian ini memiliki perbedaan rasio yang cukup besar antar sampel mayoritas dan minoritas sehingga dapat memengaruhi hasil dari model yang dibangun.

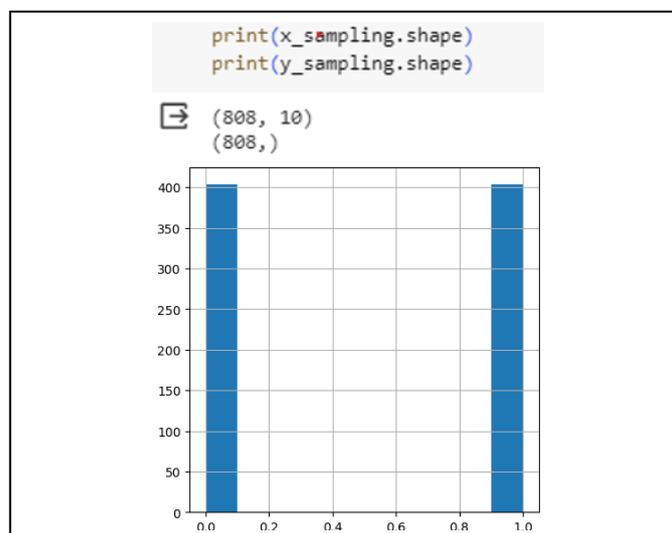
```

▶ sm = SMOTE(random_state=42)
x_sampling, y_sampling = sm.fit_resample(x,y)

```

Gambar 10. Metode SMOTE

Pada gambar 10 di atas menampilkan proses penerapan metode SMOTE (*Synthetic Minority Oversampling Technique*) dalam melakukan penyeimbangan hasil keputusan data penyakit hati yang merupakan pengembangan dari metode *over sampling* dengan memperbanyak sampel baru yang berasal dari kelas minoritas untuk membuat proporsi data menjadi seimbang dengan sampel mayoritas yang mendominasi frekuensi dataset. Setelah melalui berbagai proses penyeimbangan data maka diperoleh hasil seperti pada gambar 11.

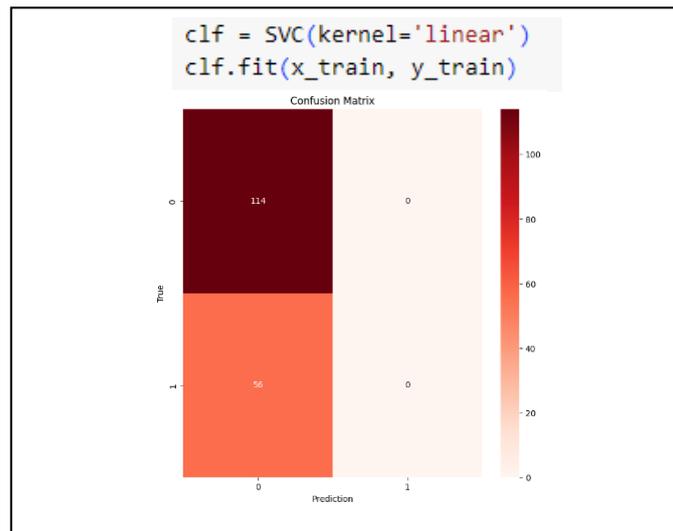


Gambar 11. Hasil penyeimbangan data

Pada gambar 11 menunjukkan hasil penyeimbangan data penyakit hati, dimana data penyakit hati yang sebelumnya mengalami ketidakseimbangan data. Setelah melalui proses penyeimbangan data menggunakan metode SMOTE maka hasil akhirnya menunjukkan bahwa hasil keputusan dari data penyakit hati telah mengalami proses penambahan rasio data yang menjadikan data penyakit hati telah seimbang.

Setelah data penyakit hati telah seimbang maka proses yang dilakukan oleh peneliti ialah melakukan proses pengklasifikasian data menggunakan metode SVM. Tahapan pertama yang harus dilakukan ialah membangun model pengklasifikasian menggunakan metode *support vector machine* dengan pemilihan

linear kernel serta *library support vector machine* pada *sklearn* untuk memudahkan proses pengklasifikasian data penyakit hati.

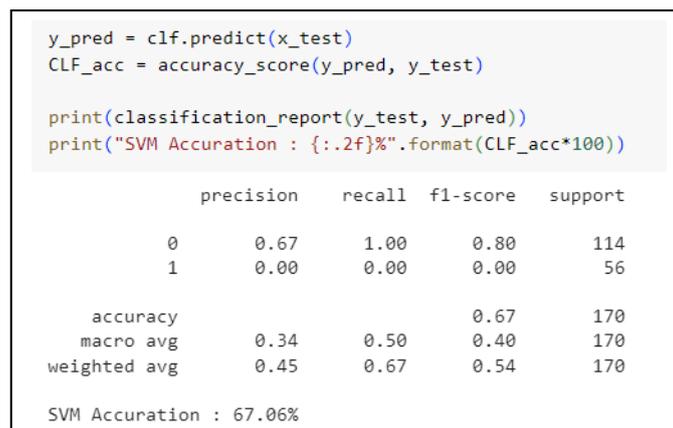


Gambar 12. Model klasifikasi dan *confusion matrix*

Keterangan :

1. *True Positives* (TP), merupakan kondisi dimana data yang sebenarnya dan data prediksi memiliki hasil yang menyatakan bahwa hasil yang diperoleh sama-sama positif.
TP = 114
2. *False Positives* (FP) merupakan kondisi dimana data prediksi menyatakan hasil positif namun data asli akan menampilkan hasil negatif.
FP = 56
3. *False Negatives* (FN) merupakan kondisi dimana data prediksi menyatakan hasil negatif namun data asli akan menampilkan hasil positif.
FN = 0
4. *True Negatives* (TN) merupakan kondisi dimana data yang sebenarnya dan data prediksi memiliki hasil yang menyatakan bahwa hasil yang diperoleh sama-sama negatif.
TN = 0

Berdasarkan hasil pemodelan klasifikasi maka hasil pengklasifikasian data penyakit hati menggunakan metode support vector machine maka diperoleh hasil classification report sebagai berikut :



Gambar 13. *Classification report*

Pada gambar 13 menjelaskan terkait hasil klasifikasi data penyakit hati menggunakan metode SVM yang telah melalui proses penyeimbangan data menggunakan metode SMOTE, dengan nilai akurasi 67,06%. Dimana 0 merupakan kode untuk nilai data negatif dengan nilai *precision* 0,67, *recall* 1,00, *f1-score* 0,80, dan *support* 114, serta 1 merupakan kode untuk nilai data positif dengan nilai *precision* 0,00, *recall* 0,00, *f1-score* 0,00, dan *support* 56. Setelah memperoleh hasil *classification report* maka tahapan akhir yang harus dilakukan ialah *testing data*. *Testing* yang bertujuan untuk menguji model *machine learning* yang telah dibangun.

```

new_data = {'Age': [20],
            'Gender': [0],
            'Total_Bilirubin': [37],
            'Direct_Bilirubin': [45],
            'Alkaline_Phosphatase': [160],
            'Alamine_Aminotransferase': [25],
            'Aspartate_Aminotransferase': [83],
            'Total_Proteins': [30],
            'Albumin': [18],
            'Albumin_and_Globulin_Ratio': [6]}

new_data = pd.DataFrame(new_data)
new_data

Age  Gender  Total_Bilirubin  Direct_Bilirubin  Alkaline_Phosphatase  Alamine_Aminotransferase
0    20      0              37                45                160                    25

[ ] scaled_new_data = scaler.transform(new_data)
    y_pred_new = clf.predict(scaled_new_data)

print("The new data prediction : ", y_pred_new)

The new data prediction : [0]

```

Gambar 14. Testing data

Berdasarkan gambar diatas, menampilkan hasil testing yang telah dilakukan sehingga dapat disimpulkan bahwa pasien dengan kriteria sesuai dengan inputan pada Gambar 14 dinyatakan tidak menderita penyakit hati (0).

IV. Kesimpulan dan saran

Berdasarkan hasil penelitian yang telah dilakukan maka diperoleh kesimpulan sebagai bahwa Metode *support vector machine* mampu melakukan klasifikasi data penyakit hati dengan hasil klasifikasi yang menunjukkan nilai 0 dan 1, di mana nilai 0 menandakan bahwa pasien tersebut tidak mengidap penyakit hati dan nilai 1 menandakan bahwa pasien tersebut mengidap penyakit hati. Berdasarkan proses visualisasi data yang telah dilakukan dalam proses pengklasifikasian data tersebut ditemukan ketidakseimbangan data penyakit hati. Penerapan metode SMOTE pada penelitian ini mampu meningkatkan kinerja SVM dalam mengklasifikasi data penyakit hati yang tidak seimbang menjadi seimbang. Hasil dari proses pengklasifikasian data penyakit hati menggunakan metode SVM menghasilkan nilai akurasi performansi yaitu 67,06%. Berdasarkan hasil kesimpulan yang telah diperoleh, maka penulis memberikan saran yang dapat dijadikan masukan oleh peneliti selanjutnya. Diharapkan agar peneliti selanjutnya menggunakan jumlah data yang lebih banyak sehingga proses pengklasifikasian lebih efisien serta proses penyeimbangan data dapat dilakukan dengan mudah. Diperlukan pengujian dengan menggunakan metode pengklasifikasian lain seperti Naïve Bayes atau KNN. Diharapkan hasil penelitian ini dapat bermanfaat bagi perkembangan sistem pendukung keputusan diagnosis penyakit hati berbasis machine learning.

Daftar Pustaka

- [1] R. Munawarah, O. Soesanto, and M. R. Faisal, "Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis," *Kumpul. J. Ilmu Komput.*, vol. 04, no. 01, pp. 103–113, 2016. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] L. Octaria Sitanggang and N. Bahtiar, "Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode Support Vector Machine (SVM) Polynomial (Studi Kasus : Data Pasien Hati India)," *J. Masy. Inform.*, vol. 10, no. 1, pp. 20–27, 2019, doi: 10.14710/jmasif.10.1.31490.
- [3] A. I. Falatehan, N. Hidayat, and K. C. Brata, "Sistem Pakar Diagnosis Penyakit Hati Menggunakan Metode Fuzzy Tsukamoto Berbasis Android," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 8, pp. 2373–2381, 2018.
- [4] W. Erawati, "Prediksi Penyakit Hati dengan Menggunakan Model Algoritma Neural Network," *J. Techno Nusa Mandiri*, vol. XII, no. 2, pp. 157–166, 2015.
- [5] F. Bei and S. Sudin, "Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (SVM)," *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [7] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *J. Tek. Inform. dan Sist. Inf.*, vol. 8,

- no. 1, pp. 24–32, 2022, doi: 10.28932/jutisi.v8i1.4004.
- [8] I. Setiawati, A. P. Wibowo, and A. Hermawan, “Implementasi Decision Tree Untuk Mendiagnosis Penyakit Lever,” *J. Inf. Syst. Manag.*, vol. 1, no. 1, pp. 13–17, 2019.
- [9] Prabiantissa Citra Nurina, “Klasifikasi pada Dataset Penyakit Hati Menggunakan Algoritma Support Vector Machine, K-NN, dan Naive Bayes,” *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, vol. 1, no. 1, pp. 263–268, 2021.
- [10] A. S. Nugroho, A. B. Witarto, and D. Handoko, “Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika,” *Kuliah Umum Ilmu Komputer.Com*, vol. 2, no. 1, pp. 842–847, 2020, doi: 10.1109/CCDC.2011.5968300.
- [11] N. D. Prayoga, N. Hidayat, and R. K. Dewi, “Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2666–2671, 2018.
- [12] P. A. Putri and H. Mustafidah, “Sistem Pakar untuk Mendiagnosa Penyakit Hati Menggunakan Metode Forward Chaining (Expert System for Diagnosing Liver Disease Using Forward Chaining),” *Juita*, vol. 1, no. 4, pp. 2086–9398, 2011.
- [13] M. I. Parapat, M. T. Furqon, and Sutrisno, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3165–3166, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [14] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [15] N. W. S. Agustini, D. Priadi, and R. V. Atika, “Klasifikasi Penyakit Gigi dan Mulut Menggunakan Metode Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 2, pp. 802–810, 2018, doi: 10.15578/jpbkp.v17i1.781.
- [16] R. Azmatul Barro, I. D. Sulvianti, and M. Afendi, “Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu,” *Xplore J. Stat.*, vol. 1, no. 1, pp. 1–6, 2013.