

## **ANALISA PERFORMA ALGORITMA *MACHINE LEARNING* DALAM PREDIKSI PENYAKIT LIVER**

Mahdiawan Nurkholifah <sup>1\*</sup>, Jasmarizal <sup>2</sup>, Yusran Umar <sup>3</sup>, Rahmadden <sup>4</sup>

<sup>1\*,2,3,4</sup> Program Studi Teknik Informatika, STMIK Amik Riau, Kota Pekanbaru, Provinsi Riau, Indonesia.

Email: 2210031802058@sar.ac.id <sup>1</sup>, 2210031802116@sar.ac.id <sup>2</sup>, 2210031802037@sar.ac.id <sup>3</sup>,  
rahmadden@sar.ac.id <sup>4</sup>

### **Histori Artikel:**

*Dikirim* 29 November 2022; *Diterima dalam bentuk revisi* 31 Desember 2022; *Diterima* 3 Januari 2023; *Diterbitkan* 10 Januari 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

### **Abstrak**

Saat ini dalam dunia kedokteran, penentuan penyakit peradangan hati menjadi hal yang tidak mudah dilakukan. Tetapi terdapat catatan rekam medis yang telah menyimpan gejala-gejala penyakit pasien dan diagnosis penyakit peradangan hati. Kelemahan-kelemahan yang dimiliki oleh metode manual, mendorong para peneliti untuk mengembangkan suatu metode yang tidak bergantung 100% pada manusia. Metode yang dikembangkan memanfaatkan komputer sebagai alat bantu untuk menganalisa data. Hal semacam ini tentu sangat berguna bagi para ahli kesehatan. Mereka dapat menggunakan catatan rekam medis yang telah ada sebagai bantuan untuk mengambil keputusan tentang diagnosis penyakit pasien. Pada penelitian ini melakukan analisa performa algoritma machine learning dengan membandingkan algoritma support vector machine, naïve bayes dan k-nearest neighbor. Penelitian ini bertujuan untuk mengetahui performa algoritma apa yang paling tinggi accuracy pada data penyakit liver. Dari hasil penelitian dengan menggunakan splinting data 80:20 dapat disimpulkan bahwa model algoritma Naïve Bayes memiliki performa lebih baik dari pada model algoritma lainnya ketika menggunakan teknik SMOTE dengan nilai accuracy 65,51%, sedangkan saat tidak menggunakan teknik SMOTE algoritma Support Vector Machine memiliki performa lebih baik dari pada model algoritma lainnya dengan nilai accuracy pada data tidak 72.41%.

**Kata Kunci :** Machine Learning; Support Vector Machine; Naïve Bayes; K-Nearest Neighbor.

### **Abstract**

Currently in the world of medicine, determining liver inflammation is something that is not easy to do. But there are medical records that have kept the patient's symptoms and diagnosis of liver inflammation. The weaknesses of the manual method encourage researchers to develop a method that does not depend 100% on humans. The developed method utilizes a computer as a tool to analyze data. This kind of thing is certainly very useful for health experts. They can use existing medical records as an aid in making decisions about the diagnosis of a patient's disease. In this study, we analyzed the performance of machine learning algorithms by comparing the support vector machine, naïve Bayes and k-nearest neighbor algorithms. This study aims to determine the performance of which algorithm has the highest accuracy in liver disease data. From the research results using splinting data 80:20 it can be concluded that the Naïve Bayes algorithm model has better performance than other algorithm models when using the SMOTE technique with an accuracy value of 65.51%, whereas when not using the SMOTE technique the Support Vector Machine algorithm has the highest performance. better than other algorithm models with an accuracy value on the data not 72.41%.

**Keyword:** Machine Learning; Support Vector Machine; Naïve Bayes; K-Nearest Neighbor.

## 1. Pendahuluan

Kesehatan merupakan salah satu hal terpenting dalam hidup manusia, hal ini menjadi dasar bahwasanya banyak ditemukannya temuan-temuan ilmiah baik berupa temua obat-obatan, alat kesehatan atau penemuan-penemuan baru di bidang kesehatan. Sejak dahulu banyak sekali penyakit-penyakit bermunculan entah itu datangnya dari virus, bakteri, parasite, sel kanker dan lain-lain. Salah satunya adalah organ tubuh hati (Noviriandini *dkk.*, 2019). Liver atau hati merupakan organ yang sangat penting dalam pengaturan homeostasis tubuh meliputi metabolisme, biotransformasi, sintesis, penyimpanan dan imunologi (Rhyzoma Grannata Rafsanjani *dkk.*, 2018). Hati adalah organ yang vital bagi manusia. Organ ini terletak di dalam rongga perut sebelah kanan, tepatnya di bawah diafragma. Terdapat beberapa fungsi kerja liver antara lain sebagai penawar dan penetralisir racun, mengatur sirkulasi hormon, mengatur komposisi darah yang mengandung lemak, gula, protein, dan zat lain (Pusporani *dkk.*, 2019).

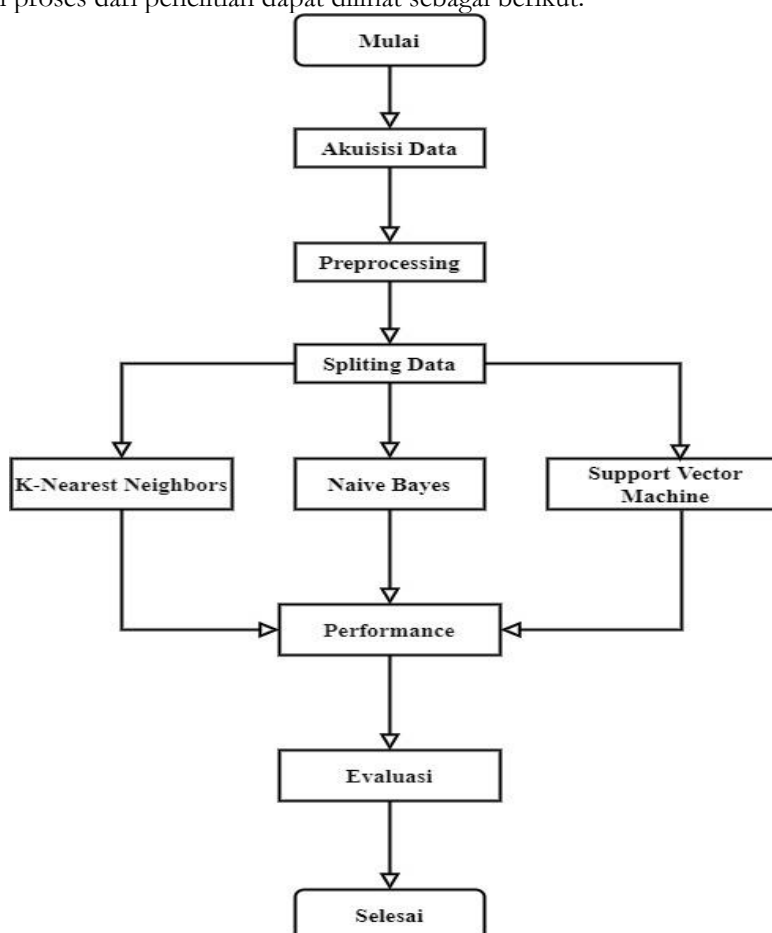
Saat ini dalam dunia kedokteran, penentuan penyakit peradangan hati menjadi hal yang tidak mudah dilakukan. Tetapi terdapat catatan rekam medis yang telah menyimpan gejala-gejala penyakit pasien dan diagnosis penyakit peradangan hati. Hal semacam ini tentu sangat berguna bagi para ahli kesehatan. Mereka dapat menggunakan catatan rekam medis yang telah ada sebagai bantuan untuk mengambil keputusan tentang diagnosis penyakit pasien (Handayani *dkk.*, 2019). Dalam mendiagnosis ada atau tidak penyakit liver dapat digunakan acuan dari hasil tes fungsi hati yang dilaksanakan di laboratorium. Tes tersebut antara lain yaitu transaminase serum, fosfatase alkali, total bilirubin, bilirubin terkonjugasi, total protein, albumin, serta rasio albumin dan globulin. Dari hasil tes tersebut dapat dilihat hasil tes yang signifikan sebagai ciri adanya gangguan fungsi hati dengan menggunakan algoritma pohon klasifikasi karena dapat memperoleh informasi mengenai data klasifikasi pasien penyakit liver (Handayani *dkk.*, 2019).

Kelemahan-kelemahan yang dimiliki oleh metode *manual*, mendorong para peneliti untuk mengembangkan suatu metode yang tidak bergantung 100% pada manusia. Metode yang dikembangkan memanfaatkan komputer sebagai alat bantu untuk menganalisa konten media sosial. Hal ini merupakan sebuah tantangan karena komputer hanya mengenali bahasa mesin, sedangkan konten yang beredar di dunia media sosial menggunakan bahasa yang dimengerti manusia. Permasalahan tersebut dapat diatasi dengan mengkuantifikasi data. Data yang telah terkuantifikasi selanjutnya dipisahkan, dikelompokkan, dengan pendekatan *Machine learning* (Laksana Utama, 2018). *Machine learning* merupakan metode berbasis komputer dimana komputer diberikan kemampuan untuk belajar dengan bantuan data tanpa harus terprogram terlebih dahulu (Laksana Utama, 2018). *Machine learning* telah berkembang, andal, dan alat pendukung di bidang medis dalam beberapa tahun terakhir (Akbar & Rahmaddeni, 2022). Maka tujuan dari penelitian ini adalah untuk melakukan perbandingan antara metode *Support Vector Machine* (SVM), *Naïve bayes*, *K-Nearest Neighbor*. Untuk mendapatkan Manakah dari metode tersebut yang paling akurat mendiagnosa penyakit peradangan hati. Penelitian ini diharapkan dapat membantu para tenaga kesehatan untuk mendiagnosa secara dini penyakit perdangan hati (Amrin & Omar Pahlevi, 2021).

Sebagai bahan acuan dan perbandingan, peneliti melakukan studi literatur dengan beberapa penelitian terdahulu yang terkait dengan tema metode-metode yang pernah digunakan untuk menyelesaikan prediksi penyakit peradangan hati (Amrin & Omar Pahlevi, 2021), diantaranya penelitian yang dilakukan oleh (Prabiantissa, 2021) dimana pada penelitian ini menggunakan algoritma *Support Vector Machine* dalam menentukan penyakit hati, dapat disimpulkan performa algoritma *Support Vector Machine* memiliki akurasi 82,36%. Selanjutnya penelitian yang dilakukan oleh (Fatchan *dkk.*, 2021) menggunakan algoritma *Naïve bayes* dalam menentukan penyakit liver, dapat disimpulkan performa algoritma *Naïve bayes* memiliki akurasi 84,00%. Selanjutnya penelitian yang dilakukan oleh (Ayudhitama & Pujianto, 2020) menggunakan algoritma *K-Nearest Neighbor* dalam menentukan penyakit liver, dapat disimpulkan performa algoritma *K-Nearest Neighbor* memiliki akurasi 66,03%.

## 2. Metode Penelitian

Penelitian ini dimulai dengan beberapa langkah terpisah, termasuk mengumpulkan data yang diperlukan untuk proses analisis. Mendistribusikan data latih dan data uji menggunakan rasio 70:30 (Akbar & Rahmaddeni, 2022). Selanjutnya merancang model algoritma *Support Vector Machine* (SVM), *Naïve bayes*, *K-Nearest Neighbor*. Hingga pada langkah terakhir yaitu, melatih dan menganalisa performa model-model algoritma *machine learning* yang diuji penelitian ini (Akbar & Rahmaddeni, 2022). Tahapan proses dari penelitian dapat dilihat sebagai berikut:



Gambar 1. Metode *Waterfall*

### 1) Akuisisi Data

Dataset pada penelitian ini adalah kumpulan data atau informasi pasien penyakit *liver* yang didapat dari situs *Kaggle Open Datasets* (kaggle.com) sebanyak 583 data. Kumpulan data ini memiliki 11 atribut, yaitu: *age*, *gender*, *total\_bilirubin*, *direct\_bilirubin*, *alkaline\_phosphotase*, *alanine\_aminotransferase*, *aspartate\_aminotransferase*, *total\_protiens*, *albumin*, *albumin\_and\_globulin\_ratio*, *outcome*. Deskripsi kumpulan data yang digunakan dapat dilihat pada tabel sebagai berikut:

Tabel 1. Atribut Kumpulan Data

No	Feature	Deskripsi
1.	Age(age)	Umur
2.	Gender(Gender)	Jenis Kelamin
3.	Total_bilirubin(TBI)	Jumlah Bilirubin
4.	Direct_bilirubin(DBI)	Bilirubin Langsung
5.	Alkaline_phosphotase(Aph)	Alkali Fosfatase

6.	Alanine_aminotransferase(AlAmt)	Alanin Aminotransferase
7.	Aspartate_aminotransferase(AsAmt)	Aspartat Aminotransferase
8.	Total_protiens(TP)	Jumlah Protein
9.	Albumin(Alb)	Albumin
10.	Albumin_and_globulin_ratio(AlGR)	Rasio Albumin dan Globulin
11.	Outcome(outcome)	Hasil

## 2) *Preprocessing*

Proses ini merupakan tahapan awal sebelum melakukan pengujian model algoritma, dimana dataset yang akan digunakan akan diolah menjadi data bersih yang siap untuk diuji. *Preprocessing* bertujuan untuk menghilangkan noise dan menyeragamkan bentuk data agar sesuai dan dapat diuji dengan model algoritma yang digunakan. Proses *preprocessing* mencakup dua tahapan yaitu *cleaning data* dan *transform data* (Akbar & Rahmaddeni, 2022). *Cleaning data* Proses *cleaning* dilakukan dengan menghilangkan noise, membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data (Ninditama *dkk.*, 2020). *Transform data* merupakan langkah terakhir *preprocessing* pada penelitian ini yang bertujuan untuk perubahan struktur fitur data yang digunakan agar menyesuaikan dengan kebutuhan model algoritma (Akbar & Rahmaddeni, 2022).

## 3) *Splinting Data*

Setelah tahapan *preprocessing*, tahap selanjutnya adalah mendistribusi kumpulan data kebentuk data latih dan data uji. Di tahapan *spliting data*, *library scikit-learn* yang telah disediakan di bahasa pemrograman *Python* digunakan. Data latih ialah informasi yang akan dalam melatih model, sedangkan data uji ialah informasi yang tidak diperlukan untuk melatih model dan digunakan sebagai data atau informasi untuk mengevaluasi akurat atau tidaknya dalam melatih model (Akbar *dkk.*, 2022).

## 4) *Support Vector Machine (SVM)*

*Support Vector Machine* merupakan sistem pembelajaran yang menggunakan hipotesis berupa fungsi-fungsi linear dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan dengan teori optimasi. Berdasarkan dengan karakteristiknya metode SVM dibagi menjadi dua yaitu linear dan non linear, SVM linear merupakan data yang dipisahkan secara linear yaitu memisahkan dua kelas pada hyperplane dengan soft margin. Sedangkan non linear yaitu merupakan fungsi dari kernel trick terhadap ruang yang berdimensi tinggi (Aldi Tangkelayuk & Evangs Mailoa, 2022).

## 5) *Naïve Bayes*

*Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan Teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Putri *dkk.*, 2021). Pada penelitian (Handoko & Neneng, 2021) Persamaan *Naïve Bayes* adalah:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Keterangan :

$P(C_i|X)$  = Probabilitas hipotesis  $C_i$  jika diberikan fakta atau record  $X$  (*Posterior probability*).

$P(X|C_i)$  = mencari nilai parameter yang memberi kemungkinan yang paling besar (*likelihood*).

$P(C_i)$  = Prior probability dari  $X$  (*Prior probability*)  $P(X)$  = Jumlah probability tuple yg muncul.

## 6) *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Prinsip kerja algoritma dari KNN yaitu menentukan dan mencari jarak terdekat dengan nilai *k* neighbor terdekat dalam data training dengan data yang akan diuji. Nilai *k* yang terbaik untuk algoritma ini tergantung berdasarkan nilai sebuah data, dimana biasanya nilai *k* yang tinggi mengurangi efek kesalahan atau noise pada proses klasifikasi, akan tetapi membuat sebuah batasan antar klasifikasi menjadi tidak maksimal (Aldi Tangkelayuk & Evangs Mailoa, 2022). Menurut (Ayudhitama & Pujianto, 2020) Kelebihan dari algoritma ini yaitu pelatihan dari data latihnya sangat cepat, sederhana, mudah dipelajari, tahan akan data pelatihan yang mengandung derau serta tetap efektif walaupun data latihnya besar. Alur pada algoritma *K-Nearest Neighbor*:

- a) Mencari nilai bobot dari setiap atribut pada kelas yang ada dan menghitung bobot dengan keanggotaan kelas yang ada pada data latih. Rumus mencari rata – rata bobot untuk setiap atribut yaitu :

$$w = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2)$$

Dimana penjabaran rumus diatas yaitu :

w, menunjukkan rata – rata bobot tiap atribut;

x<sub>n</sub>, menunjukkan data masukan ke-*n* tiap atribut;

n, menunjukkan jumlah data.

- a) Hitung jarak euclidean dengan rumus :

$$di = \sum_{i=1}^p w(x_{2i} - x_{1i})^2 \quad (3)$$

Dimana penjabaran rumus diatas yaitu :

w, menunjukkan nilai bobot dari setiap inputan;

x<sub>1</sub>, menunjukkan nilai data latih; 3)

x<sub>2</sub>, menunjukkan nilai data uji;

i, menunjukkan variabel data;

d, menunjukkan nilai jarak;

p, menunjukkan dimensi data.

- b) Untuk menentukan label dari data baru maka ambil sebanyak *k* tetangga terdekat dari label kelas tetangga sebelumnya.

## 7) *Evaluasi*

Untuk melihat, mengevaluasi, serta menganalisis kinerja model algoritma *machine learning* tim penulis menggunakan acuan *cross validation*. *Cross validation* atau dapat disebut estimasi rotasi yang merupakan suatu teknik konfirmasi model yang digunakan dalam menentukan dengan cara menerapkan hasil analisis statistik terhadap sejumlah besar data bebas. Teknik ini umumnya diaplikasikan untuk menghasilkan model prediktif dan memverifikasi ketepatan beberapa model saat digunakan. *K-fold cross validation* ialah satu-satunya metode estimasi rotasi yang mengubah data menjadi *k* bagian set data yang identik (Tempola *dkk.*, 2018). Persamaa berikut digunakan untuk mengukur kinerja prediksi tiap-tiap model algoritma yang digunakan.

$$accuracy = \frac{\sum \text{klasifikasi benar}}{\sum \text{data uji}} \times 100\% \quad (4)$$

### 3. Hasil dan Pembahasan

*Keagle open datasets* menjadi sumber untuk mendapatkan *dataset* yang digunakan dalam studi kasus ini. Dengan jumlah 583 data pasien penyakit *liver*. Berikut merupakan gambaran data dari sumber <https://www.kaggle.com/datasets/zgeakylz/heart-desease-data>

Tabel 2. Kumpulan *Dataset* Yang Digunakan

No	Age	Gende	TBI	DBI	Aph	AIAmt	AsAmt	TP	AIb	AIGR	Outcom
1.	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	1
2.	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
3.	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	1
4.	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	1
5.	72	Male	3.9	2.0	195	27	59	7.3	2.4	0.40	1
...	...	...	...	...	...	...	...	...	...	...	...
579	60	Male	0.5	0.1	500	20	34	5.9	1.6	0.37	2
580	40	Male	0.6	0.1	98	35	31	6.0	3.2	1.10	1
581	52	Male	0.8	0.2	245	48	49	6.4	3.2	1.00	1
582	31	Male	1.3	0.5	184	29	32	6.8	3.4	1.00	1
583	38	Male	1.0	0.3	216	21	24	7.3	4.4	1.50	2

Setelah data terkumpul, selanjutnya dilakukan proses *preprocessing* yang dilakukan agar ketika data digunakan dalam melatih model algoritma *machine learning*, data tersebut tidak terjadi kekosongan nilai maupun *transformasi* data. Berikut merupakan gambaran data proses *preprocessing*.

Tabel 3. Kumpulan Data Setelah *Preprocessing*

No	Age	Gende	TBI	DBI	Aph	AIAmt	AsAmt	TP	AIb	AIGR	Outcom
1.	65	0	0.7	0.1	187	16	18	6.8	3.3	0.90	0
2.	62	1	10.9	5.5	699	64	100	7.5	3.2	0.74	0
3.	62	1	7.3	4.1	490	60	68	7.0	3.3	0.89	0
4.	58	1	1.0	0.4	182	14	20	6.8	3.4	1.00	0
5.	72	1	3.9	2.0	195	27	59	7.3	2.4	0.40	0
...	...	...	...	...	...	...	...	...	...	...	...
579	60	1	0.5	0.1	500	20	34	5.9	1.6	0.37	1
580	40	1	0.6	0.1	98	35	31	6.0	3.2	1.10	0
581	52	1	0.8	0.2	245	48	49	6.4	3.2	1.00	0
582	31	1	1.3	0.5	184	29	32	6.8	3.4	1.00	0
583	38	1	1.0	0.3	216	21	24	7.3	4.4	1.50	1

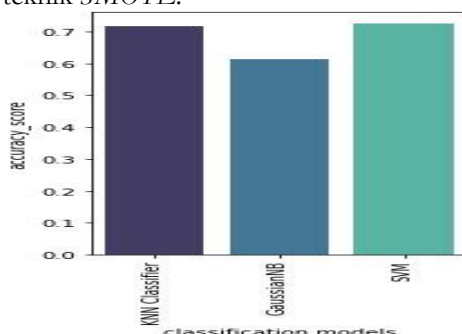
Proses *preprosesing* pertama dilakukan dengan mengubah nilai *outcome* 0 untuk data yang terdiagnosis penyakit liver dan 1 untuk data yang tidak terdiagnosis penyakit liver. Proses kedua mengkonversi nilai *gender* menjadi *categorical* dengan nilai 0 untuk *female* dan 1 untuk *male*. Proses ketiga menghapus data yang bernilai *null* pada *dataset*. Setelah *preprocessing* selesai proses berikutnya adalah mendistribusikan data latih dan data uji dengan rasio 80:20. Kemudian merancang model algoritma *Support Vector Machine* (SVM), *Naïve bayes*, *K-Nearest Neighbor*, sampai melakukan proses melatih dan menganalisa performa model-model algoritma *machine learning* yang diuji penelitian ini. Hasil analisa performa model-model algoritma *machine learning* yang diuji dapat dilihat pada tabel sebagai berikut:



Tabel 4. Hasil Klasifikasi *Before Smote*

Method	Accuracy	TPR	FPR	Precision	Recall
Naïve Bayes	61.20%	0.953	0.589	0.488	0.953
KNN	71.55%	0.768	0.524	0.869	0.768
SVM	72.41%	0.724	-	1.000	0.724

Pada tabel 4 terlihat bahwa klasifikasi SVM memiliki performa terbaik dengan *accuracy* 72.41%, akan tetapi pada nilai FPR (*False Positive Rate*) bernilai NaN atau kosong dibandingkan dengan model klasifikasi FPR Naïve Bayes dan KNN. Berikut merupakan gambar diagram perbandingan akurasi tanpa menggunakan teknik *SMOTE*.



Gambar 2. Perbandingan Akurasi Tanpa Menggunakan Teknik *SMOTE*

Karena terjadi *imbalance* pada *dataset* maka penelitian ini menerapkan teknik *SMOTE*. Pada proses *imbalance* data menggunakan teknik *SMOTE*, kumpulan data yang tidak seimbang akan diselesaikan secara *sampling*, terlihat pada tabel 4 sebelum menggunakan teknik *SMOTE* total data sebanyak 579 dan setelah menggunakan teknik *SMOTE* total data meningkat menjadi 682.

Tabel 5. Total Jumlah Sebelum Dan Sesudah *SMOTE*

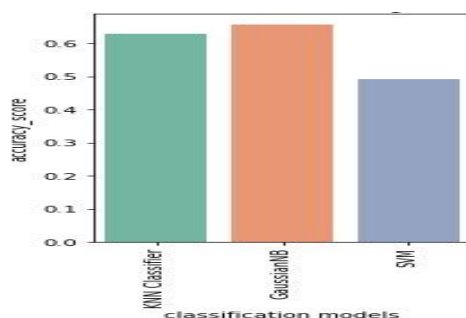
	Liver Patient	Not Liver Patient	Total
Sebelum <i>SMOTE</i>	414	165	579
Sesudah <i>SMOTE</i>	341	341	682

Pada Tabel 5 menunjukkan hasil klasifikasi sebelum menggunakan teknik *SMOTE*. Evaluasi kinerja terbaik akurasi dalam klasifikasi SVM. *Accuracy* dari *Naïve Bayes* 65,51% lebih tinggi dari KNN dan SVM. Peningkatan performa terjadi pada klasifikasi *Naïve Bayes* yang semula memiliki ketelitian masing-masing 61,20% ditunjukkan pada Tabel 4 meningkat menjadi 65,51% ditunjukkan pada Tabel 6.

Tabel 6. Hasil Klasifikasi Setelah menggunakan Teknik *SMOTE*

Method	Accuracy	TPR	FPR	Precision	Recall
Naïve Bayes	65.51%	0.867	0.479	0.534	0.867
KNN	62.93%	0.742	0.500	0.630	0.742
SVM	49.13%	0.850	0.583	0.233	0.850

Berikut merupakan gambar diagram perbandingan akurasi dengan menggunakan teknik *SMOTE*.



Gambar 3. Perbandingan Akurasi Dengan Menggunakan Teknik SMOTE

Berikut merupakan hasil perbandingan algoritma machine learning saat sebelum maupun sesudah menggunakan teknik SMOTE

Tabel 7. Hasil Perbandingan Model Algoritma Machine Learning

No	Model	Accuracy	
		Sebelum SMOTE	Sesudah SMOTE
1.	<i>K-Nearest Neighbor</i>	71.55%	62.93%
2.	<i>Naïve Bayes</i>	61.20%	65.51%
3.	<i>Support Vector Machine</i>	72.41%	49.13%

#### 4. Kesimpulan

Dari hasil pengujian dapat disimpulkan bahwa model algoritma *Naïve Bayes* memiliki performa lebih baik dari pada model algoritma lainnya ketika menggunakan teknik SMOTE dengan nilai *accuracy* 65,51%, sedangkan saat tidak menggunakan teknik SMOTE algoritma *Support Vector Machine* memiliki performa lebih baik dari pada model algoritma lainnya dengan nilai *accuracy* pada data tidak 72.41%.

#### 5. Ucapan Terima Kasih

Ucapan terimakasih kepada STMIK Amik Riau tempat kami menuntut ilmu dan rekan-rekan seperjuangan serta dosen pengampu mata kuliah *Machine Learning* di STMIK Amik Riau yang telah membimbing dalam penelitian ini.

#### 6. Daftar Pustaka

- Akbar, F., & Rahmaddeni. (2022). *Komparasi Algoritma Machine Learning Untuk Memprediksi Penyakit Alzheimer*. 8(2), 236–245.
- Akbar, F., Wira Saputra, H., Karel Maulaya, A., Fikri Hidayat, M., & Rahmaddeni. (2022). *Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke*. 2(October), 61–67.
- Aldi Tangkelayuk, & Evangs Mailoa. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>



- Amrin, & Omar Pahlevi. (2021). Data Mining Optimization Based on Particle Swarm Optimization For Diagnosis of Inflammatory Liver Disease. *Jite*, 5(1), 152–159. <http://ojs.uma.ac.id/index.php/jite>
- Ayudhitama, A. P., & Pujiyanto, U. (2020). Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer. *Jurnal Informatika Polinema*, 6(2), 1–9. <https://doi.org/10.33795/jip.v6i2.274>
- Fatchan, M., Tedi, N., Alfiyan, Kurniawan, & Widodo, E. (2021). Perbandingan Dalam Memprediksi Penyakit Liver Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor. *Jurnal Pelita Teknologi*, 16(1), 15–21.
- Handayani, P., Nurlelah, E., Raharjo, M., & Ramdani, P. M. (2019). Prediksi Penyakit Liver Dengan Menggunakan Metode Decision Tree dan Neural Network. *Computer Engineering, Science and System Journal*, 4(1), 55. <https://doi.org/10.24114/cess.v4i1.11528>
- Handoko, M. R., & Neneng. (2021). Sistem Pakar Diagnosa Penyakit Ispa Menggunakan Metode Naive Bayes Classifier Berbasis Web. *CSRID (Computer Science Research and Its Development Journal)*, 10(3), 127. <https://doi.org/10.22303/csr.id.10.3.2018.127-138>
- Laksana Utama, P. K. (2018). Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning. *Widya Duta: Jurnal Ilmiah Ilmu Agama dan Ilmu Sosial Budaya*, 13(1), 69. <https://doi.org/10.25078/wd.v13i1.436>
- Ninditama, I. P., Ninditama, I. P., Cholil, W., Akbar, M., & Antoni, D. (2020). Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang. *Jurnal TEKNO KOMPAK*, 15(2), 37–49. <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1156>
- Noviriandini, A., Handayani, P., & Syahriani. (2019). Prediksi Penyakit Liver Dengan Menggunakan Metode. *Prosiding TAU SNAR-TEK Seminar Nasional Rekayasa dan Teknologi*, November, 75–80.
- Prabiantissa, C. N. (2021). Klasifikasi pada Dataset Penyakit Hati Menggunakan Algoritma Support Vector Machine, K-NN, dan Naïve Bayes. *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, 219–224.
- Pusporani, E., Qomariyah, S., & Irhamah, I. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, 2(1), 25. <https://doi.org/10.12962/j27213862.v2i1.6810>
- Putri, H., Purnamasari, A. I., Dikananda, A. R., Nurdiawan, O., & Anwar, S. (2021). Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN. *Building of Informatics, Technology and Science (BITS)*, 3(3), 331–337. <https://doi.org/10.47065/bits.v3i3.1093>
- Rhyzoma Grannata Rafsanjani, Hidayat, N., & Dewi, R. K. (2018). Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes Dan Certainty Factor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4478–4482.
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), 577. <https://doi.org/10.25126/jtiik.201855983>