



## *Application of Decision Tree for Health Status Classification, Compared to KNN and Naive Bayes*

### **Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan Naive Bayes**

**Arell S. Biyantoro<sup>1</sup>, Budi Prasetyo<sup>2</sup>**

<sup>12</sup>Information System, Faculty of Mathematics and Natural Sciences  
Universitas Negeri Semarang

E-Mail: <sup>1</sup>arelsaverro1234@students.unnes.ac.id, <sup>2</sup>bprasetyo@mail.unnes.ac.id

Makalah: Diterima 18 Februari 2024; Diperbaiki 1 Maret 2024; Disetujui 18 Maret 2024  
Corresponding Author: Budi Prasetyo

#### **Abstrak**

Fokus penelitian ini adalah pengujian algoritma *machine learning* untuk klasifikasi status kesehatan, dengan penekanan pada penggunaan Algoritma *Decision Tree*, *K-Nearest Neighbor (KNN)*, dan *Naive Bayes*. Metode penelitian ini menggunakan *dataset dummy* bertema kesehatan untuk melakukan tahapan perencanaan, pengumpulan data, preprocessing, instruksi, dan analisis hasil. Hasil penelitian menunjukkan bahwa algoritma *machine learning Decision Tree* memberikan akurasi tertinggi (95.45%) dalam memprediksi kesehatan berdasarkan variabel seperti usia (*Age*) dan intensitas olahraga (*ExerciseHours*). Intensitas olahraga lebih dari 2,5 jam per minggu dianggap sebagai faktor penting dalam menentukan kesehatan yang baik, sementara usia dan durasi olahraga tertentu mempengaruhi kategori kesehatan yang dihasilkan. Penelitian ini memberikan wawasan penting tentang penggunaan *machine learning* untuk membantu prediksi status kesehatan menggunakan variabel karakteristik individu dalam dataset.

**Keyword:** Decision Tree, Algoritma, Machine Learning, K-Nearest Neighbor, KNN, Naive Bayes, Kesehatan, Intensitas olahraga

#### **Abstract**

*This research focuses on testing machine learning algorithms for health status classification, with an emphasis on the use of Decision Tree, K-Nearest Neighbor (KNN), and Naive Bayes algorithms. This research method uses a health-themed dummy dataset to carry out the stages of planning, data collection, preprocessing, instruction, and analysis of results. The results show that the Decision Tree machine learning algorithm provides the highest accuracy (95.45%) in predicting health based on variables such as age (Age) and exercise intensity (ExerciseHours). Exercise intensity of more than 2.5 hours per week is considered an important factor in determining good health, while age and duration of certain exercises affect the resulting health category. This research provides important insights into the use of machine learning to aid health status prediction using individual characteristic variables in the dataset.*

**Keywords:** Decision Tree, Algorithm, Machine Learning, K-Nearest Neighbor, KNN, Naive Bayes, Health, Exercise intensity

### **1. PENDAHULUAN**

Kesehatan merupakan salah satu hal paling penting dan mahal di dunia, seseorang dapat melakukan aktivitas dengan lancar berkat adanya kesehatan di dalam dirinya tanpa terganggu[1]. Kesehatan tersebut sangat bisa dicapai dengan menerapkan olahraga yang teratur, olahraga tidak hanya memberikan kesehatan dan menjaganya, namun olahraga dengan dosis yang tepat juga dapat menjadi obat untuk berbagai penyakit, dan masih banyak lagi keuntungan fisik yang didapatkan[2]. Terdapat banyak bukti bahwa olahraga sangat berpengaruh pada kesehatan, bahkan olahraga yang dilakukan sebagai gaya hidup dinilai dapat memberikan rentan kesehatan yang lebih panjang dan menunda timbulnya 40 penyakit kronis[3].

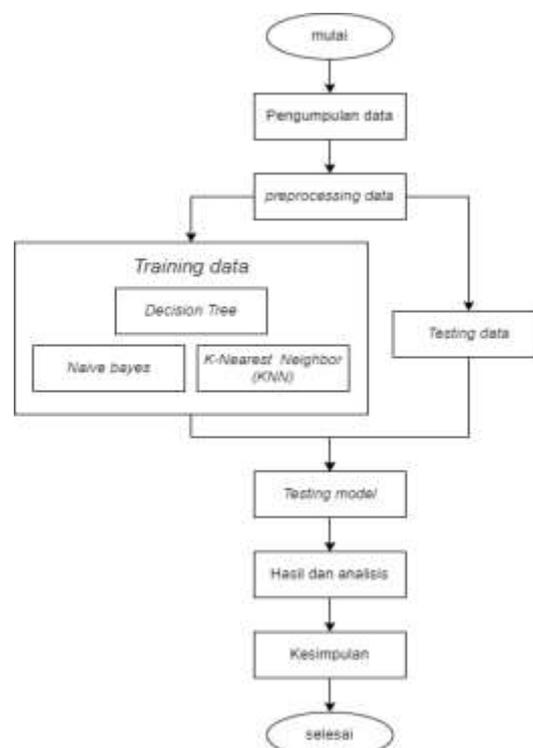
Tidak hanya intensitas dan takaran olahraga yang dapat mempengaruhi kesehatan, namun usia seseorang juga dianggap sangat berpengaruh pada penurunan kualitas hidup yang diakibatkan oleh kestabilan fisik dan psikis yang umumnya diakibatkan banyak hal yang telah seseorang tersebut lalui selama hidup, misalkan saja riwayat penyakit kronis yang dimiliki, masalah kehidupan sosial, dan masalah finansial yang sering dihadapi lalu mengakibatkannya stress[4]. Dalam teori kognitif beck mengenai depresi telah mengungkapkan bahwa perasaan tidak berdaya yang mempengaruhi perasaan emosional dapat diakibatkan karena pikiran negatif yang tidak rasional, penyebab stress, dan insiden[5]. Memasuki usia lanjut, manusia umumnya akan mengalami perubahan struktur otak yang nantinya juga akan mempengaruhi kinerja organ-organ lain pada orang tersebut[6].

Mengetahui tingkat kesehatan kelompok masyarakat dari setiap individunya pada suatu daerah sangat penting dan dapat dimanfaatkan karena erat hubungannya dengan lingkungan yang ditempati seperti kebersihan lingkungan, sanitasi, dan lain sebagainya[7]. Menentukan status kesehatan seseorang juga sangat penting dalam menentukan sumber daya manusia dan secara tidak langsung hal ini akan berpengaruh juga dengan[8]. Kesehatan individu pada suatu negara dibidang sangat berpengaruh pada negara, karena jika sumber daya manusia yang dimiliki kualitas kesehatan yang baik, maka negara tersebut akan berjalan semakin optimal[8].

Dengan kebutuhan-kebutuhan di atas, kita perlu melakukan klasifikasi untuk dapat dengan mudah memprediksi kesehatan seseorang berdasarkan karakteristik yang mirip dalam beberapa kelas[9]. Klasifikasi ini dapat dilakukan dengan berbagai cara atau metode. Salah satunya dengan menggunakan algoritma-algoritma yang populer digunakan seperti *Decision Tree*[9]. Selain *Decision Tree*, terdapat pula algoritma *K-Nearest Neighbor* (KNN), KNN merupakan algoritma yang sederhana tapi kuat dan efektif untuk mengklasifikasikan data. Kemudian terdapat algoritma *Naive bayes* yang menggunakan perhitungan probabilitas[10]. Penelitian ini akan membandingkan terlebih dahulu algoritma apa yang paling cocok untuk data yang digunakan kemudian algoritma dengan akurasi paling tinggi digunakan guna mendapatkan hasil paling optimal.

## 2. BAHAN DAN METODOLOGI PENELITIAN

Pada penelitian klasifikasi status kesehatan ini dilakukan dengan metode yang tersusun dalam beberapa tahap, diantaranya adalah perencanaan, pengumpulan data, preprocessing data, training data, hasil dan analisa, kesimpulan. Berikut ini bentuk jelas dari metodologi penelitian dalam bentuk flow.



Gambar 1. Metodologi Penelitian

## 2.1 Data Mining

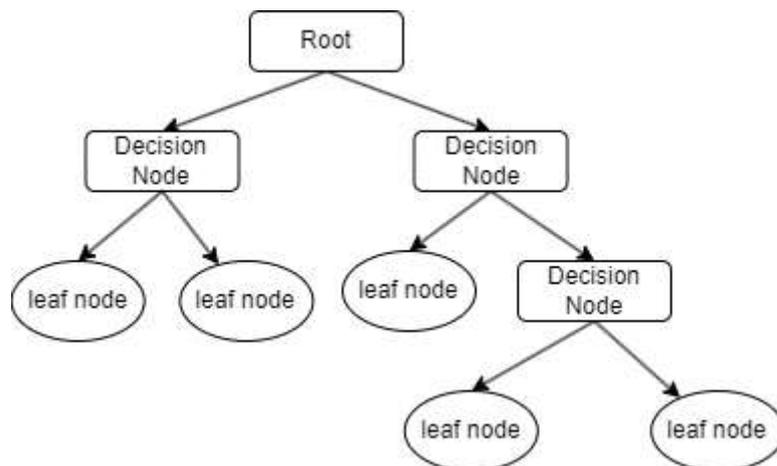
*Data mining* merupakan salah satu set teknik yang digunakan pada data yang sangat besar untuk mengeksplorasinya secara menyeluruh dan menampilkan berbagai relasi yang sangat kompleks[11]. Teknologi ini dapat digunakan untuk menemukan pola-pola yang sulit untuk diidentifikasi dan mencari informasi guna pemprediksian[11]. Pada *data mining*, pendekatan dilakukan dengan *discovery-based* dimana pencocokan pola dan penggunaan algoritma digunakan guna menentukan relasi kunci dalam data yang diolah[11]. Karena bertujuan untuk menampilkan penerapan *decision tree* pada pengolahan data dan membandingkannya dengan KNN and Naive Bayes untuk mendapat akurasi paling tinggi, Pada *data mining* kali ini penelitian menggunakan *data dummy* demi menghindari penggunaan data yang sensitif.

## 2.2 Klasifikasi

Klasifikasi merupakan salah satu metode dalam pengolahan data yang merupakan pengelompokan sekumpulan data sesuai dengan karakteristik atau ciri-ciri data itu sendiri satu sama lain[12]. Klasifikasi dilakukan dengan mengumpulkan sekumpulan data latih dengan beberapa kelas yang telah ditentukan[12]. Pengklasifikasian ini bisa dilakukan dengan dua cara, yaitu dengan cara yang manual atau dengan memanfaatkan teknologi yang ada[12]. Pada pemanfaatan teknologi di metode klasifikasi dapat dilakukan dengan menerapkan algoritma yang ditujukan untuk klasifikasi seperti *Decision Tree*, *Naive Bayes*, dan *K-Nearest Neighbor (KNN)*[13].

## 2.3 Decision Tree

*Decision Tree* merupakan salah satu algoritma pada pengolahan data metode klasifikasi[13]. Algoritma *machine learning* yang satu ini mempresentasikan strukturnya seperti struktur pohon untuk menemukan hasil keputusan[13]. Konsep dari *Decision Tree* ini adalah dengan menyajikan pernyataan-pernyataan bersyarat pada setiap langkah yang bercabang untuk pengambilan keputusan berdasarkan perhitungan dari data set itu sendiri[13]. Dalam metode Klasifikasi, algoritma *Decision Tree* merupakan yang paling banyak digunakan[12].



Gambar 2. Decision Tree

## 2.4 K-Nearest Neighbor (KNN)

Berbeda dengan model algoritma *Decision Tree*, *K-Nearest Neighbor (KNN)* merupakan algoritma yang melakukan klasifikasi data berdasarkan jarak paling dekat dengan objek target[13]. Konsep dari *K-Nearest Neighbor (KNN)* sangat sederhana, algoritma ini berjalan dengan mengklasifikasikan berdasarkan kesamaan ciri-ciri terhadap kelompok tertentu dari titik data terdekat atau tetangganya[13]. Dengan konsep seperti itu, *K-Nearest Neighbor (KNN)* akan memberikan hasil akhir yang kompetitif[13]. Untuk lebih jelasnya, berikut ini bentuk dari rumus persamaan pada algoritma *K-Nearest Neighbor (KNN)*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

## 2.5 Naive Bayes

Pengklasifikasian *Naive Bayes* adalah salah satu algoritma lain dari metode klasifikasi selain *Decision Tree* dan *K-Nearest Neighbor (KNN)*. *Naive Bayes* bekerja dengan melakukan pengklasifikasian yang

bersumber dari perhitungan probabilitas sederhana[13]. Algoritma *machine learning* yang satu ini dirancang untuk dapat difungsikan berdasarkan asumsi setiap variabel yang independen[13]. Secara umum, *Naive Bayes* biasanya digunakan untuk menemukan probabilitas lain dari data probabilitas tertentu yang telah dimiliki[13].

### 3. HASIL DAN ANALISIS

Hasil didapatkan setelah penelitian dilakukan dengan bantuan tool data mining yang bernama RapidMiner. RapidMiner adalah salah satu platform *data mining* yang tersedia gratis dan cukup populer[14]. Berikut ini hasil dan analisis pengolahan klasifikasi dengan algoritma yang telah dilakukan melalui RapidMiner.

#### 3.1 Pengumpulan Data

Penelitian ini telah menggunakan dataset dummy bertema kesehatan yang dihasilkan/dibuat melalui *Artificial Intelligence (AI) ChatGPT* (<https://chat.openai.com/>) dengan *prompt* sebagai berikut:

“Buat Dataset dalam bentuk csv dengan tema kesehatan dengan 3 kelas yaitu Good Fair Poor, 4 variable (Age, Gender, Weight, ExerciseHours, HealthStatus) dengan salah satunya adalah label untuk pengklasifikasian, panjang dataset 120 rows dan berikan atribut untuk nomor baris, Data mengandung mengandung missing value dan imbalance.”

Keterangan: Exercise Hours (jam olahraga) dihitung dalam satuan jam per minggu

dengan *prompt* di atas, *AI ChatGPT* menghasilkan data seperti dibawah ini:

Tabel 1. Data Awal

No_rows	Age	Gender	Weight	ExerciseHours	HealthStatus
1	45	Male	70	3	Good
2	50	Female	65		Good
3	35	Male	80	5	Good
4	28	Female	55	2	Good
5	60	Male		1	Fair
6	42	Female	68	4	Good
7	55	Male	75	6	Good
...	...	...	...	...	...

Pada tabel yang dihasilkan tersebut terdapat 5 (lima) atribut. Atribut “No\_rows” yang terletak pada kolom pertama merupakan penanda untuk berapa data yang telah tergenerate pada ChatGPT. Atribut tersebut tidak akan digunakan pada pemrosesan data karena tidak akan relevan. Kemudian atribut HealthStatus yang berisikan tiga kelas yaitu Good (Baik), Fair (Wajar/Normal), Poor (Tidak baik) menjadi hasil dari setiap data pada dataset, maka dari itu atribut tersebut layak diatur sebagai label.

#### 3.2 Preprocessing Data

Tahap selanjutnya setelah data didapatkan adalah melakukan *preprocessing data*. *Preprocessing data* adalah salah satu tahap yang dilakukan setelah dataset berhasil didapatkan, kemudian dibersihkan dari berbagai elemen yang tidak relevan untuk penelitian dan akhirnya mendapatkan dataset yang berkualitas[15]. Dalam penelitian ini, dataset akan melalui *preprocessing data* dengan beberapa. Pertama ada *Feature Selection*, dimana terdapat satu atribut bernama “No\_rows” yang tidak akan dimasukkan ke dalam pemrosesan data karena tidak relevan dan hanya berfungsi sebagai penomoran data di luar pemrosesan. Kemudian terdapat *Missing Data Treatment* yang dilakukan kepada beberapa atribut yang memiliki *missing value*.

Label: <b>HealthStatus</b>	Polynomial	0	Low: Poor (4)	High: Good (82)	Values: Good (82), Fair (34), ... [1 more]
Age	Integer	0	Min: 22	Max: 60	Average: 39.975
Gender	Binomial	0	Negative: Male	Positive: Female	Values: Female (60), Male (60)
Weight	Integer	5	Min: 50	Max: 90	Average: 67.816
ExerciseHours	Integer	10	Min: 1	Max: 6	Average: 3.336

Gambar 3. missing value

Dapat dilihat pada gambar diatas, RapidMiner menunjukkan bahwa dataset yang dimasukkan memiliki 6 *missing value* pada atribut “Weight” dan 10 *missing value* pada atribut “ExerciseHours”. Setelah melalui pertimbangan, putuskanlah untuk melakukan *treatment* yang berbeda pada dua (2) atribut tersebut. atribut “Weight” akan mendapatkan *treatment* berupa *Data Imputation* dan mengisi missing datanya dengan rata-rata (*average*) yang ditemukan dengan nilai 68. Sedangkan pada atribut “ExerciseHours” dilakukan *treatment Remove Missing Value* dan akan menghapus satu tupel pada *missing data* tersebut secara keseluruhan. Setelah dilakukannya *Missing Value Treatment*, data yang tersisa pada dataset adalah 110 data.

Setelah berhasil membersihkan dataset dari *missing value*, selanjutnya adalah memberikan label kepada salah satu atribut. Label sendiri sangat diperlukan pada metode klasifikasi karena *role* atribut ini digunakan sebagai *identifier* dan mengklasifikasi bagi objek atau data yang sedang olah[16]. Pada penelitian data set ini, yang berperan sebagai label adalah atribut “HealthStatus” dan berisikan tiga kelas yaitu Good, Fair, dan Poor. Berikut ini penampakan hasil setelah dataset dilakukan *Missing Value Treatment* dan pengaturan label pada atribut “HealthStatus”:

Row No. ↑	HealthStatus	Weight	Age	Gender	ExerciseHo...
1	Good	70	45	Male	3
2	Good	80	35	Male	5
3	Good	55	28	Female	2
4	Fair	68	60	Male	1
5	Good	68	42	Female	4
6	Good	75	55	Male	6
7	Good	72	48	Male	4
8	Fair	50	25	Female	2
9	Good	85	38	Male	5
10	Good	60	33	Female	3
11	Fair	52	29	Female	1
12	Good	77	40	Male	4
13	Good	63	36	Female	3
14	Good	73	58	Male	6

Gambar 4. data bersih

Hal terakhir yang dilakukan pada tahap *preprocessing* ini adalah melakukan *split data*. *split data* dilakukan untuk menguji akurasi atau performa dari algoritma yang digunakan. Kali ini, data akan di pisah dengan rasio 80% untuk *training model* dan 20% untuk *data testing*. Pembagian ini mengacu kepada *rule of thumb Pareto Principle* yang sering digunakan di bidang matematika, ekonomi, dan komputer, dengan rasio dataset menjadi 80:20[17]. Dengan begitu, data akan menjadi 88 untuk *training model* dan 22 untuk *data testing*.

### 3.3 Pengetesan Akurasi Algoritma

Pegimplementasian kepada algoritma *Decision Tree*, *K-Nearest Neighbor (KNN)*, dan *Naive Bayes* pertama-tama dilakukan untuk *testing* dan menentukan algoritma mana yang memiliki kelayakan dan akurasi paling tinggi pada dataset target. *Testing* dilakukan menggunakan hasil *split data* yang telah dihasilkan sebelumnya. 80% data yang merupakan *data training* akan dihubungkan/dimasukkan ke dalam algoritma untuk menjadi model atau sumber perhitungan. Kemudian algoritma yang akan diukur keakuratannya dihubungkan dengan operator *Apply Model* bersamaan dengan 20% data tersisa sebagai *data testing*.

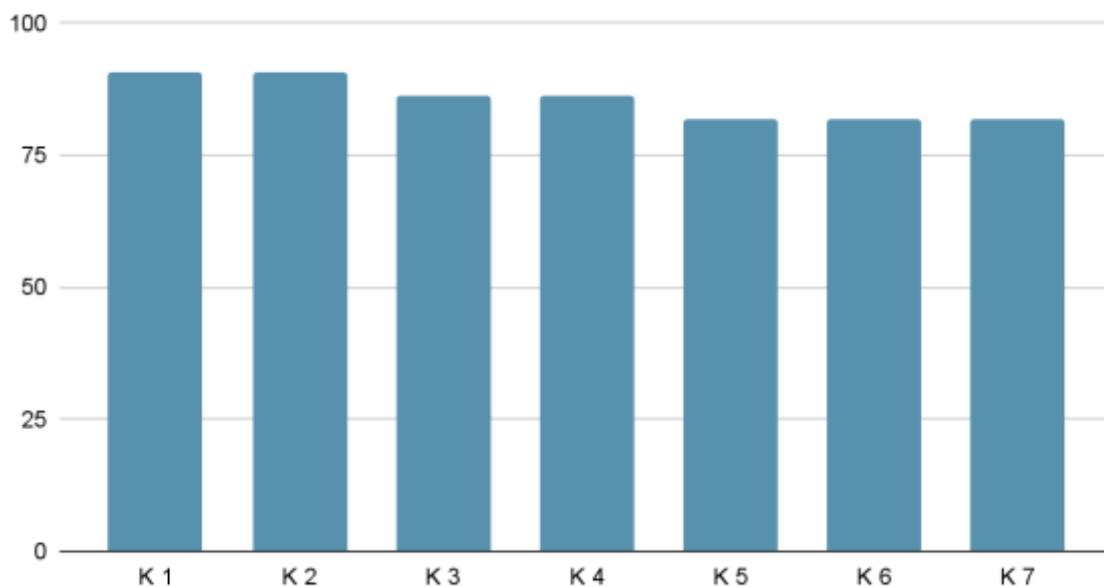
Pengujian pertama dilakukan kepada Algoritma *Decision Tree* dengan pengaturan *criteria gain\_ratio* dan *maximal depth* sebanyak 10 tingkat. Pada hasil pengujian ini didapatkan akurasi dari *Decision Tree* dengan 10 tingkatan mendapatkan akurasi sebesar 95.45% dari 22 *data testing*. Berikut tabel rincian hasil *testing Decision Tree*:

Tabel 2. akurasi decision tree

	true Good	true Fair	True Poor	lass Precision
Pred. Good	14	0	0	100%
Pred. Fair	0	6	0	100%
Pred. . Poor	0	1	1	50%
Class Recall	100%	85.71%	100%	

Selanjutnya, pengujian dilakukan kepada Algoritma *K-Nearest Neighbor (KNN)* dengan nilai K 1-7. Pengujian dilakukan dengan *split data* yang masih sama. Setelah dilakukan pengujian ternyata K dengan nilai satu (1) memiliki akurasi paling tinggi ketimbang nilai K lainnya. K dengan nilai 1 mendapatkan akurasi sebesar 90.91%. Untuk keakuratan setiap nilai K dan rincian hasil K 1 dapat dilihat pada grafik dan tabel dibawah.

#### Nilai K



Gambar 5. Grafik nilai k

Tabel 3. akurasi KNN K1

	true Good	true Fair	True Poor	lass Precision
Pred. Good	14	1	0	93.33%
Pred. Fair	0	6	1	85.71%
Pred. . Poor	0	0	0	0%
Class Recall	100%	85.71%	0%	

Terakhir, pengujian dilakukan pada Algoritma *Naive Bayes* menggunakan *split data* yang masih sama dengan dua algoritma sebelumnya. Setelah diuji menggunakan 22 data yang ada pada dataset, *Naive Bayes* mendapatkan akurasi sebesar 86.36%. Berikut rincian hasil pengujiannya:

Tabel 4. akurasi naive bayes

	true Good	true Fair	True Poor	lass Precision
Pred. Good	13	2	0	86.67%
Pred. Fair	1	5	0	83.33%
Pred. . Poor	0	0	1	100.00%
Class Recall	92.86%	71.43%	100%	

### 3.4 Implementasi Decision Tree

Setelah dilakukan pengujian, didapatkan bahwa Algoritma *Decision Tree* memiliki akurasi paling tinggi, yaitu sebesar 95.45%. Dengan begitu, pengimplementasian algoritma akan dilanjutkan menggunakan *Decision Tree* karena dianggap paling layak dan akurat. Implementasi algoritma ini juga masih dilakukan menggunakan tool yang sama, yaitu RapidMiner. Pada operator *Decision Tree* yang terdapat di RapidMiner dilakukan penyetalan beberapa parameter didalamnya, diantaranya adalah *maximal depth* pada angka 10, criteria pada pilihan “*gain\_ratio*”, *confidence* di angka 0.1, dan *minimal gain* pada angka 0.01.

Setelah dilakukan proses pada dataset dan algoritma ini, didapatkan bahwa, seseorang akan memiliki status kesehatan yang baik (Good) jika memiliki waktu olahraga lebih dari 2,5 jam/minggu atau berumur diantara lebih dari 27,5 tahun hingga 28,5 tahun. Orang yang berusia lebih dari 46 tahun, di antara lebih dari 28,5 tahun hingga 37,5 tahun, dan 27,5 tahun kebawah akan memiliki kesehatan yang aman (Fair) jika melakukan olahraga selama 2,5 jam/minggu. Sedangkan orang yang berusia diatas 37,5 tahun hingga 46 tahun akan beresiko memiliki kesehatan buruk (Poor) jika hanya melakukan olahraga 2,5 jam/minggu.



Gambar 6. hasil decision tree

Untuk lebih jelasnya, berikut ini bentuk deskriptif dari hasil *Decision Tree* yang diperoleh:

ExerciseHours > 2.500: Good {Good=54, Fair=0, Poor=0}

ExerciseHours ≤ 2.500

| Age > 37.500

| | Age > 46: Fair {Good=0, Fair=3, Poor=0}

| | Age ≤ 46: Poor {Good=0, Fair=0, Poor=3}

| Age ≤ 37.500

| | Age > 28.500: Fair {Good=0, Fair=17, Poor=0}

| | Age ≤ 28.500

| | | Age > 27.500: Good {Good=4, Fair=0, Poor=0}

| | | Age ≤ 27.500: Fair {Good=0, Fair=7, Poor=0}

### 4. KESIMPULAN

Penelitian ini ditujukan untuk memilih algoritma *machine learning* yang paling cocok digunakan untuk memprediksi kesehatan dilihat dari atribut intensitas olahraga, usia, umur, dan jenis kelamin. Isi dataset yang digunakan bukanlah data asli, melainkan *data dummy* yang dibuat menggunakan *Artificial Intelligence* yang bernama ChatGPT. Hal ini dilakukan untuk menghindari penggunaan data-data sensitif dan sebagai bahan

pembelajaran. Dari *prompt* yang diinputkan ke dalam ChatGPT, didapatkan 120 data yang mengandung *missing value* dan *imbalance*. 120 data tersebut diolah pada *preprocessing* dan dilakukan pula *split data* dan menghasilkan 88 *data training* dan 22 *data testing*.

Kemudian dilakukan pengujian menggunakan tiga (3) algoritma untuk metode klasifikasi yaitu *Decision Tree*, *K-Nearest Neighbor (KNN)*, dan *Naive Bayes*. Setiap Algoritma akan diuji kemudian ambil pengaturan yang mendapatkan persentase akurasi paling tinggi dan hasilnya akan dibandingkan lagi dengan algoritma lain yang diuji. Hasil pengujian didapatkan bawah algoritma *Decision Tree* mendapatkan akurasi paling tinggi dengan 95.45%, mengalahkan *KNN* (90.91%), dan *Naive Bayes* (86.36%). Dengan begitu Algoritma *machine learning* yang paling cocok dengan pengujian ini adalah *Decision Tree*.

Setelah ditemukan hasil akurasi paling tinggi, Implementasi algoritma *Decision Tree* dilakukan untuk mendapatkan prediksi dari model yang telah dibuat. Hasil dari implementasi tersebut adalah variabel atau atribut yang paling mempengaruhi dalam menentukan kesehatan seseorang dari pengujian tersebut adalah intensitas olahraga (*ExerciseHours*) dan umur (*Age*). Seseorang akan memiliki kesehatan yang bagus (*Good*) jika dirinya melakukan olahraga lebih dari 2,5 jam/minggu. Sebaliknya, seseorang akan memiliki kesehatan yang buruk (*Poor*) jika dirinya berusia 37,5 tahun hingga 46 tahun dan hanya melakukan olahraga maksimal 2,5 jam/minggu. Diluar dari kriteria tersebut, status kesehatan seseorang akan dinyatakan aman/normal (*Fair*).

## REFERENSI

- [1]. D. Kurnianto, P. Prodi, I. Keolahragaan, and P. Uny, "MENJAGA KESEHATAN DI USIA LANJUT."
- [2]. J. Vina, F. Sanchis-Gomar, V. Martinez-Bello, and M. Gomez-Cabrera, "Exercise acts as a drug; the pharmacological benefits of exercise," *British Journal of Pharmacology*, vol. 167, no. 1, pp. 1–12, Aug. 2012, doi: 10.1111/j.1476-5381.2012.01970.x.
- [3]. G. N. Ruegsegger and F. W. Booth, "Health Benefits of Exercise," *Cold Spring Harbor Perspectives in Medicine*, vol. 8, no. 7, p. a029694, May 2017, doi: 10.1101/cshperspect.a029694.
- [4]. F. N. Qonita, N. A. Salsabila, N. F. Anjani, and S. Rahman, "KESEHATAN PADA ORANG LANJUT USIA (Kesehatan Mental dan Kesehatan Fisik)," *PSIKOWIPA (Psikologi Wijaya Putra)*, vol. 2, no. 1, pp. 10–19, Jul. 2021, doi: 10.38156/psikowipa.v2i1.42.
- [5]. E. Şahin, N. Topkaya, C. Gençoğlu, and E. Ersanlı, "Prevalence and Correlates of Hopelessness among Turkish Elderly People Living with Family or in Nursing Homes," *Societies*, vol. 8, no. 2, p. 39, Jun. 2018, doi: 10.3390/soc8020039.
- [6]. I. Sampelan *et al.*, "HUBUNGAN DUKUNGAN KELUARGA DENGAN KEMANDIRIAN LANSIA DALAM PEMENUHAN AKTIVITAS SEHARI-HARI DI DESA BATU KECAMATAN LIKUPANG SELATAN KABUPATEN MINAHASA UTARA," 2015.
- [7]. R. H. D. P. K. dan I. Perilaku, "HUBUNGAN PERILAKU HIDUP SEHAT DENGAN STATUS KESEHATAN MASYARAKAT KELURAHAN UJUNG RELATIONSHIP BETWEEN HEALTHY BEHAVIOR AND HEALTH STATUS IN KELURAHAN UJUNG."
- [8]. Z. I. SKM., MKM., dr. N. S. Kes M., and W. M. Si, *Derajat Kesehatan Masyarakat Urban, Sub Urban, Dan Pesisir Kota Medan*. Merdeka Kreasi Group, 2022.
- [9]. D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 15–19, May 2023, doi: 10.57152/malcom.v3i1.591.
- [10]. A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naive Bayes, dan Decision Tree," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 1109–1119, Jun. 2022, doi: 10.35957/jatisi.v9i2.2048.
- [11]. I. Ernawati, "DATA MINING SEBAGAI SALAH SATU SOLUSI STRATEGI BISNIS," *Informatik : Jurnal Ilmu Komputer*, vol. 14, no. 1, p. 9, Apr. 2018, doi: 10.52958/iftk.v14i1.367.
- [12]. A. P. Wibawa, M. Guntur, A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, 2018.
- [13]. Eri Mardiani *et al.*, "Copyright @ Analisis Prediksi Pendapatan Penduduk dengan Metode K-Nearest Neighbor, Decision Tree, Naive Bayes, Ensemble Methods, dan Linear Regression," *INNOVATIVE: Journal Of Social Science Research*, Sep. 2023.
- [14]. P. Ristoski, C. Bizer, and H. Paulheim, "Mining the Web of Linked Data with Rapidminer," *SSRN Electronic Journal*, 2015, doi: 10.2139/ssrn.3199209.
- [15]. S. F. Pane and J. Ramdan, "Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter," *Jurnal Sistem Cerdas*, vol. 5, no. 1, pp. 12–20, May 2022, doi: 10.37396/jsc.v5i1.191.
- [16]. T. A. Kautsar, "Klasifikasi Multi Label Pada Jawaban Esai Menggunakan Algoritma Multi Label

Knearest Neighbor (Mlkn),” Aug. 15, 2023.

<https://repository.uinjkt.ac.id/dspace/handle/123456789/73665>

- [17]. R. Adinugroho, “Perbandingan Rasio Split data Training dan data Testing Menggunakan Metode Lstm dalam Memprediksi Harga Indeks Saham Asia,” Feb. 16, 2023.  
<https://repository.uinjkt.ac.id/dspace/handle/123456789/67314>