

ANALISIS KLASIFIKASI PENYAKIT HATI MENGGUNAKAN METODE K-NEAREST NEIGHBOR DALAM MACHINE LEARNING

Lestari

STMIK Royal, Kisaran

Email : Lestari33@gmail.com

ABSTRAK

Abstrak: Penelitian ini fokus pada kesulitan dalam mendeteksi penyakit hati sejak dini, yang sering kali sulit dalam teridentifikasi bahkan setelah penyebarannya. Penting untuk mengetahui gejala penyakit hati sejak dini agar penderita dapat menerima pengobatan yang tepat, meningkatkan peluang kelangsungan hidup. Penelitian ini bertujuan menentukan model terbaik untuk mengklasifikasikan data penyakit hati dan mendeteksinya sejak dini, dengan menggunakan metode K-Nearest Neighbor (KNN). KNN dianggap sebagai pendekatan sederhana dan efisien dalam klasifikasi, khususnya dalam konteks prediksi penyakit hati di Indonesia. Prediksi penyakit hati menggunakan metode klasifikasi dimana dalam metode klasifikasi terdapat 3 metode meliputi (1) K-NN, (2) Naïve Bayes (3) Decision Tree yang diterapkan dalam penelitian ini. Berdasarkan 3 metode klasifikasi tersebut akan dicari 1 metode terbaik dengan nilai akurasi paling terbaik yang akan digunakan dalam proses pengolahan deploy untuk untuk prediksi penyakit hati. Dimana pengujian ini terbagi menjadi 3 yaitu 90:10, 80:20, 70:30. Penelitian ini menggunakan supervised learning dengan variabel independen seperti umur (age), pendidikan (education), dan kadar gula (glucose), serta variabel target (Tenyearchd). Dataset berasal dari kaggle.com, terdiri dari 4239 baris data dengan fokus pada 3 kolom utama dan satu kolom target, yaitu Tenyearchd. Pengujian dilakukan dengan variasi rasio data, dan hasil terbaik diperoleh pada rasio 90:10 dengan tingkat akurasi mencapai 85%. Selain KNN, penelitian ini membandingkan metode Naïve Bayes dan Decision Tree. Visualisasi data, seperti scatter plot dan diagram lingkaran, membantu dalam memahami hubungan antar variabel. Hasil pengujian menunjukkan bahwa Naïve Bayes menjadi metode terbaik dengan tingkat akurasi tertinggi pada semua rasio data (85%). Kesimpulannya, penelitian ini menyoroti bahwa Naïve Bayes efektif dalam mengklasifikasikan data penyakit hati untuk deteksi dini.

Kata Kunci: K-Nearest Neighbor, Penyakit hati, Machine learning

ABSTRACT

Abstract: This research focuses on the difficulty in detecting liver disease early, which is often difficult to identify even after it has spread. It is important to recognize the symptoms of liver disease early so that sufferers can receive appropriate treatment, increasing their chances of survival. This research aims to determine the best model for classifying liver disease data and detecting it early, using the K-Nearest Neighbor (KNN) method. KNN is considered a simple and efficient approach to classification, especially in the context of liver disease prediction in Indonesia. Prediction of liver disease uses a classification method where in the classification method there are 3 methods including (1) K-NN, (2) Naïve Bayes (3) Decision Tree which is applied in this research. Based on these 3 classification methods, we will look for the best method with the best accuracy value which will be used in the deploy processing process to predict liver disease. Where this test is divided into 3, namely 90:10, 80:20, 70:30. This research uses supervised learning with independent variables such as age, education and glucose levels, as well as the target variable (Tenyearchd). The dataset comes from kaggle.com, consisting of 4239 rows of data with a focus on 3 main columns and one target column, namely Tenyearchd. Testing was carried out with varying data ratios, and the best results were obtained at a ratio of 90:10 with an accuracy level of 85%. Apart from KNN, this research compares the Naïve Bayes and Decision Tree methods. Data visualization, such as scatter plots and pie charts, helps in understanding the relationships between variables. The test results show that Naïve Bayes is the best method with the highest level of accuracy in all data ratios (85%). In conclusion, this study highlights that Naïve Bayes is effective in classifying liver disease data for early detection.

Keywords: K-Nearest, Liver disease, Machine learning



PENDAHULUAN

Perkembangan teknologi dan informasi menjadikan internet sebagai alat komunikasi yang paling banyak diminati oleh masyarakat. Internet tidak terlepas dari yang namanya media jejaring sosial. Twitter, Facebook, dan Youtube merupakan salah satu media jejaring sosial sebagai media perangkat komunikasi terpopuler yang ada dikalangan pengguna internet saat ini. Oleh karena itu, media jejaring sosial banyak dimanfaatkan para peneliti untuk melakukan penelitian tentang masyarakat di Indonesia[1]. Klasifikasi adalah teknik untuk merancang fungsi berdasarkan hasil pengamatan dari data dan atribut dari data tersebut sehingga dapat dilakukan pemetaan terhadap data yang belum memiliki kelas kedalam data yang telah terklasifikasi sesuai dengan aturan-aturan yang diberikan[2].

Organ hati mempunyai peran yang sangat penting bagi tubuh manusia. Organ hati merupakan organ dengan berat mencapai 1,2 sampai 1,8 kg dari keseluruhan berat orang dewasa[3]. Saat ini meningkatnya penyakit hati disebabkan oleh infeksi virus seperti virus hepatitis, penyakit autoimun yang menyebabkan sistem imun tubuh menyerang sel atau jaringan sehat pada hati. Meningkatkan penyakit hati di Indonesia menyebabkan berbagai dampak yang mempengaruhi kesehatan pada manusia yang disebabkan kurangnya menjaga kebersihan, menjaga pola makan, serta minum-minum yang mengandung alkohol. Jadi masyarakat Indonesia harus lebih berhati-hati dalam menjaga kesehatan tubuh agar tidak terkena dampak penyakit hati yang disebabkan kurangnya menjaga kebersihan dan tidak mengatur pola makan dengan teratur.

Hati merupakan organ vital bagi manusia. Organ ini terletak di rongga perut kanan, tepatnya di bawah diafragma. Ada beberapa fungsi hati, antara lain sebagai penangkal dan penetral racun, mengatur sirkulasi hormon, mengatur komposisi darah yang mengandung lemak, gula, protein, dan zat lainnya. Penyakit hati merupakan gangguan pada setiap fungsi hati, hati bertanggung jawab atas fungsi-fungsi penting dalam tubuh, di mana hilangnya fungsi-fungsi ini dapat menyebabkan kerusakan yang signifikan pada tubuh. Hati adalah satu-satunya organ dalam tubuh yang dapat dengan mudah menggantikan sel-sel yang rusak, tetapi jika sel-sel ini hilang, hati mungkin tidak dapat memenuhi kebutuhan tubuh[4]. Gejala yang ditimbulkan pada penyakit hati dapat dibedakan menjadi dua, yaitu gejala klinis dan fisik. Gejala klinis merupakan gejala yang dapat diketahui dengan melakukan screening tentang gejala yang dirasakan oleh pasien. Gejala fisik adalah gejala yang dapat diketahui dengan memeriksa tubuh pasien[3].

Berdasarkan data mengenai penyakit hati yang mengalami penyakit tersebut berada di usia 34 tahun – 65 tahun, yang dimana manusia harus berhati-hati dalam menjaga pola hidup sehat ketika memasuki usia 30 tahun, agar tidak mengalami penyakit hati yang sudah ditentukan faktor usia sesuai dataset. Bahwa dengan persentase sesuai umur dapat memicu beberapa permasalahan tentang kesehatan manusia. Bahwa dengan data pada pasien yang terkena penyakit hati dapat memicu permasalahan pada kesehatan manusia lainnya. Penyakit hati merujuk pada data Kementerian Kesehatan pada tahun 2023 menyatakan bahwa 2.159 orang dinyatakan meninggal karena terkena kanker hati. Dengan melihat paparan mengenai data penyakit hati di atas maka kita dapat melihat bahwa kesehatan sangat penting agar tidak ada lagi manusia yang terkena penyakit hati tersebut.

Permasalahan yang biasanya terjadi adalah sulitnya mengenali penyakit hati sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Padahal mengetahui adanya gejala penyakit hati sejak dini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan tepat. Dengan diagnose adanya penyakit hati lebih awal dapat meningkatkan kelangsungan hidup pasien. Oleh karena itu, pada penelitian ini peneliti ingin mengetahui apakah model yang paling sesuai untuk mengklasifikasikan data penyakit hati agar penyakit hati bisa dideteksi sejak dini[5].

Berdasarkan permasalahan di atas, bahwa tindakan pencegahan diperlukan sebagai upaya untuk mencegah agar tidak ada lagi manusia yang terkena penyakit hati tersebut. Salah satu upaya pencegahan yang dapat dilakukan adalah dengan menghentikan konsumsi alkohol, konsumsi makanan yang sehat, berolahraga secara teratur, hindari efek samping obat. Oleh karena itu klasifikasi penyakit hati diperlukan untuk mengetahui jumlah pasien yang terkena penyakit hati tersebut. Pengujian klasifikasi nantinya akan menggunakan dataset prediksi penyakit hati. Klasifikasi prediksi penyakit hati tersebut akan menggunakan metode K-NN. Penggunaan metode K-NN adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Dimana terdapat 3 variabel yang akan diolah yaitu variabel *age*(umur), *education*(pendidikan), *glucose*(gula), (*Tenyearchd*). Selain itu, memperoleh hasil estimasi yang akurat, penelitian ini melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat objek tersebut. Proses pengolahan data menggunakan tools jupyter notebook dalam software anaconda navigator dalam mengelola data dengan metode K-NN.

KNN merupakan salah satu pendekatan yang sederhana untuk diimplementasikan dan merupakan metode lama yang sering digunakan dalam pengklasifikasian. Menurut Y. Hamamoto, dkk dan E. Alpaydin pada tahun 1997 menyebutkan bahwa KNN memiliki tingkat efisiensi yang tinggi dan dalam beberapa kasus dapat memberikan tingkat akurasi yang tinggi dalam klasifikasi [6]. Metode K-NN telah banyak digunakan di berbagai penelitian berbasis machine learning baik untuk prediksi maupun sebagai analisis objek penelitian. Selain itu, hasil dari proses perhitungan dengan metode K-NN kerap dibandingkan dengan metode lainnya, hal tersebut karena metode K-NN memiliki keuntungan tersendiri dalam memprediksi yang bersumber dari data prediksi mengenai penyakit hati di Indonesia.

Penerapan metode K-NN yang telah diimplementasikan ke dalam penelitian sebelumnya seperti pada penelitian yang dilakukan dalam penelitian tersebut menggunakan metode K-NN digunakan untuk mengukur prediksi penyakit hati di Indonesia. Dimana penelitian tersebut menggunakan metode K-NN. Yang dimana akurasi K-NN sebesar 0.84%. Hal tersebut yang mendorong penelitian untuk melakukan variasi dengan 2 model klasifikasi agar diperoleh hasil dengan tingkat akurasi yang lebih baik lagi.

Oleh karena itu, penelitian ini bertujuan mendapatkan hasil klasifikasi dan mengetahui metode manakah yang menghasilkan nilai akurasi yang lebih akurat dengan membandingkan 3 metode yang terdapat dalam klasifikasi diantaranya adalah (1) K-NN, (2) Naïve Bayes (3) Decision Tree.

METODE

Metode riset memuat hal-hal penting terkait pelaksanaan penelitian dan tahapan yang akan ditempuh dalam penelitian ini untuk mencapai suatu hasil atau output yang diinginkan. Berikut ini hal-hal yang mencakup ke dalam metode riset antara lain :

Peralatan Riset

Peralatan yang akan digunakan dalam penelitian ini adalah berupa alat computer dengan spesifikasi yang mendukung, seperti software yang akan digunakan dalam proses pengolahan berupa software Jupyter notebook dalam Anaconda Navigator dimana bahasa pemrograman yang akan digunakan adalah bahasa pemrograman Python, Visual Studio Code sebagai teks editor dan menggunakan Streamlit sebagai bahasa markup yang digunakan untuk membangun deploy aplikasinya.

Variabel yang Terdapat dalam Riset

Dalam machine learning, penelitian ini dikategorikan sebagai penelitian supervised learning yaitu sebuah teknik model dimana data yang akan diproses memiliki label/target/class dengan tujuan untuk mengetahui hubungan kualitas antara variabel bebas (variabel independen) dan variabel yang menjadi target/labelnya (variabel y). Oleh sebab itu, variabel yang terdapat dalam penelitian ini terdiri dari variabel independen atau variabel x meliputi age (umur), education (pendidikan), dan glucose (gula), (Tenyearchd).

Tahapan – Tahapan Riset

Tahapan yang akan ditempuh dalam penelitian ini digambarkan dengan flowchart sehingga dapat terlihat dengan jelas tahapan yang akan dilakukan.

Data Riset

Data penelitian yang akan dimanfaatkan merupakan kumpulan data prediksi penyakit hati yang diambil dari kaggle.com. Dataset ini terdiri dari 4239 baris data dengan 3 kolom yaitu (age (umur), education (pendidikan), glucose (gula)) yang akan menjadi fokus pengujian. Selain itu, terdapat satu kolom, tambahan, yaitu (Tenyearchd), yang berperan sebagai kolom target atau label dalam analisis.

Pembagian Data

Dalam penelitian ini, data dibagi menjadi bagian training dan testing dengan rasio berbeda, yaitu 90:10. Data training, yang mencakup sekitar 90% dari dataset, dan data testing, yang mencakup sekitar 10% dari dataset, digunakan untuk menguji algoritma. Hasilnya menunjukkan bahwa algoritma dengan rasio 90:10 memiliki performa yang superior dengan tingkat akurasi mencapai 85%. Adapun rasio 80:20, 70:30. Variasi rasio ini bertujuan untuk menemukan model K-NN terbaik untuk hasil klasifikasi.

Klasifikasi dengan 3 Model



Klasifikasi merupakan algoritma klasifikasi yang paling sederhana dalam mengklasifikasikan sebuah gambar kedalam sebuah label yang dimana memperoleh hasil evaluasi pengaruh sebuah variabel x (variable bebas) dan variabel y (variable terikat). Dalam proses klasifikasi menggunakan 3 model klasifikasi dengan tujuan untuk membandingkan model klasifikasi terbaik

HASIL DAN PEMBAHASAN

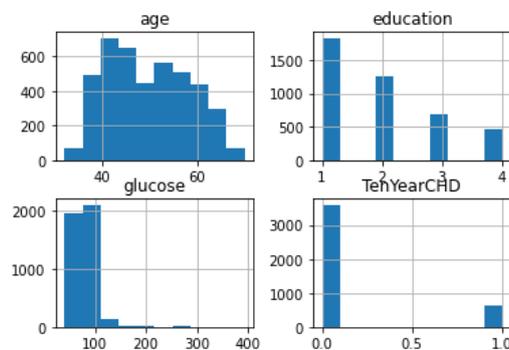
Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel x terhadap variabel y. Selanjutnya, data akan di uji dengan 4 bentuk pembagian data dengan rasio yang berbeda. Dimana didalamnya akan dilakukan pengujian data menggunakan 3 model klasifikasi dan 4 model nilai evaluasi. Hal tersebut bertujuan untuk menemukan model klasifikasi manakah yang paling baik dengan menghasilkan nilai akurasi yang paling akurat untuk digunakan dalam klasifikasi penyakit hati. Tahapan pembahasan yang dilakukan adalah sebagai berikut.

Visualisasi Data

Visualisasi data bertujuan untuk mempermudah pengamatan dan pemahaman terhadap suatu set data. Terdapat beragam model yang dapat digunakan untuk menggambarkan data secara visual. Salah satu contohnya adalah scatter plot yang menggunakan library matplotlib.pyplot. Dalam penelitian ini, data prediksi penyakit hati, yang digunakan untuk memperkirakan klasifikasi prediksi penyakit hati, divisualisasikan menggunakan scatter plot. Tujuan dari visualisasi ini adalah untuk melihat sebaran data dan menganalisis hubungan antara sumbu x dan sumbu y. Setiap variabel pada scatter plot diwakili oleh warna yang berbeda, meningkatkan kemudahan pemahaman dan daya ingat terhadap informasi yang disajikan. Visualisasi data tersebut sebagai berikut ini:

Diagram Batang (Bar Chart)

Berikut ini tampilan diagram batang pada data age, education, glucose, dan TehYearCHD pada data prediksi penyakit hati.

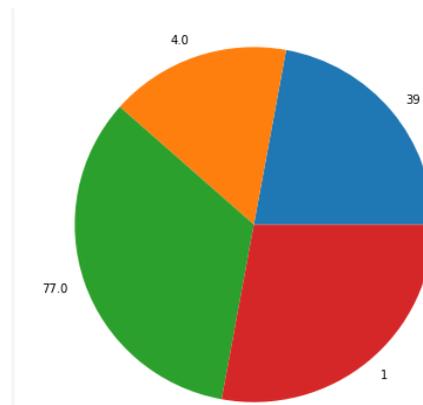


Gambar 1. Diagram Batang (Bar Chart)

Pada gambar diatas, diagram batang yang terdiri dari data age, education, glucose, TehYearCHD yang masing-masing terbagi 4 diagram yang menunjukkan hasil yang berbeda setiap diagram batang tersebut.

Diagram Lingkaran

Selanjut, ini tampilan diagram lingkaran pada data age, education, glucose, dan TehYearCHD pada data prediksi penyakit hati.

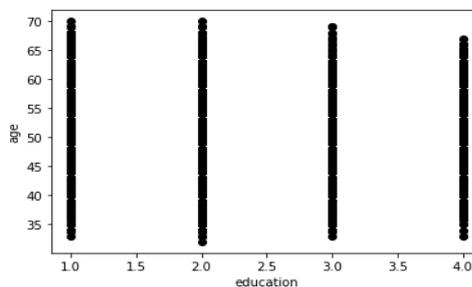


Gambar 2. Diagram Lingkaran

Pada gambar diatas diagram lingkaran (pie chart) adalah sebuah diagram berbentuk lingkaran yang dibagi menjadi irisan-irisan untuk menggambarkan proporsi numerik. Dalam hal ini, data yang terkait dengan beberapa parameter penyakit hati yaitu age, aducation, glucose dan TehYearCHD. Nilai proporsinya sebagai berikut age 39, education 4.0, glucose 77.0, dan TehYearCHD 1. Dengan menggunakan diagram lingkaran, anda dapat membuat lingkaran yang dibagi menjadi empat bagian, masing-masing mewakili satu parameter.

Scatter Plot age terhadap education

Selanjutnya ini tampilan visualisasi data pada age sebagai sumbu y terhadap education sebagai sumbu x.sebagai berikut:

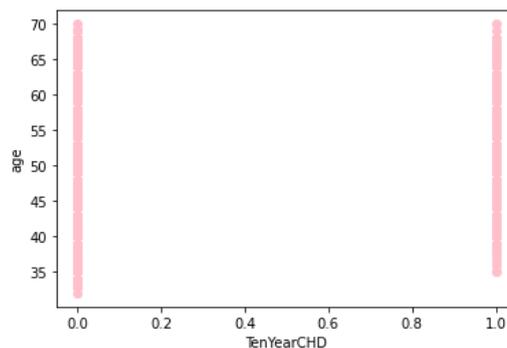


Gambar 3. Scatter age terhadap education

Pada gambar diatas, plot titik sebaran age (sumbu y) divisualisasikan dengan titik berwarna hitam. Berdasarkan gambar diatas, dapat diketahui bahwa semakin besar ukuran sebuah klasifikasi prediksi penyakit hati maka yang dihasilkan relatif semakin mengecil. Hubungan yang di tunjukkan oleh gambar diatas ialah hubungan oleh sumbu x (education) yang diikuti dengan peningkatan sumbu y (age).

Scatter Plot age terhadap TehYearCHD

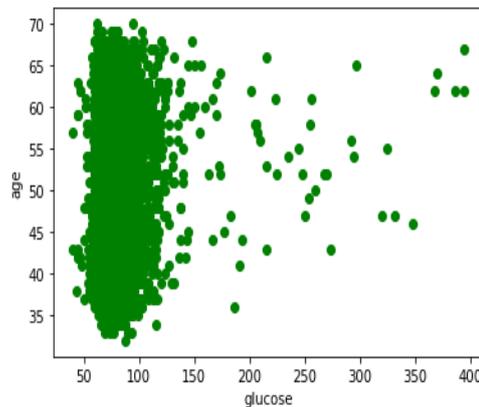
Selanjutnya ini tampilan visualisasi data pada age sebagai sumbu y terhadap TehYearCHD sebagai sumbu x.sebagai berikut:



Gambar 4. Scatter age terhadap TehYearCHD

Visualisasi plot titik-titik sebaran dari sumbu y (age) terhadap sumbu x (TehYearCHD) di tampilkan ke dalam scatter berwarna pink. Hubungan yang ditunjukkan masih berifat positif karena masih menunjukkan kenaikan baik dari variable age dan TehYearCHD.

Scatter Plot	MODEL	AKURASI	age terhadap
ini tampilan pada age y terhadap sumbu berikut:	Decission Tree	0.7900943396226415	Selanjutnya visualisasi data sebagai sumbu glucose sebagai x.sebagai
	Naïve Bayes	0.8592767295597484	
	K- Nearest Neighbor	0.8474842767295597	



Gambar 5. Scatter age terhadap glucose

Visualisasi pada sumbu y (glucose) terhadap sumbu x (age) menggunakan scatter plot berwarna hijau. Berdasarkan gambar diatas, plot titik sebaran tampak berbentuk bulat lonjong, hal tersebut memberikan makna bahwa sumbu x memiliki hubungan yang erat terhadap sumbu y.

Pengujian Data dengan 3 Model Klasifikasi

Pada penelitian ini menggunakan metode kalsifikasi dimana akan dilakukan perbandingan dari 3 metode klasifikasi agar diperoleh model dengan hasil akurasi terbaik untuk digunakan dalam proses klasifikasi penyakit hati. Penelitian ini menggunakan pembagian data untuk di uji sebanyak 3 bentuk diantaranya pembagian rasio 90:10, kedua 80:20, dan ketiga 70:30. Pengujian data dengan rasio yang bervariasi untuk mencari serta menentukan model klasifikasi manakah yang paling sering muncul dalam menghasilkan nilai klasifikasi yang terbaik. Sehingga, nantinya akan diketahui model klasifikasi yang terbaik dalam bentuk klasifikasi terhadap dataset untuk mencari serta menentukan model klasifikasi manakah yang paling sering muncul dalam menghasilkan nilai klasifikasi yang terbaik. Berikut pembahasan penelitian untuk menemukan model klasifikasi terbaik yang diuraikan kedalam rincian poin sebagai berikut:

Pengujian Akurasi Pada Rasio Data 90:10

Pengujian data pertama menggunakan rasio 90:10. Dimana *data training* sebesar 90% dari total data dan *data testing* sebesar 10% menggunakan 3 model klasifikasi diperoleh hasil seperti table dibawah ini:

Tabel 1. Pengujian Akurasi Data Rasio 90

Ketentuan yang digunakan dalam melihat hasil klasifikasi diatas yang mendapat kan nilai terbaik adalah untuk nilai pengukuran, menggunakan akurasi yang paling tertinggi yang dihasilkan merupakan nilai dengan klasifikasi yang paling bagus. Berdasarkan tabel diatas, pada model *Decission tree* nilai akurasi nya 0.7900943396226415, model *Naïve bayes* 0.8592767295597484 dan model *K- nearest neighbor* 0.8474842767295597. Jadi nilai yang paling terbaik dalam pengujian rasio 90:10 adalah *Naïve bayes* dengan nilai akurasi 0.8592767295597484.

Pengujian Akurasi Pada Rasio Data 80:20

Pengujian data kedua menggunakan pembagian data dengan rasio 80:20. Terdapat sebanyak 80% *data training* dan 20% *data testing* dari dataset, diperoleh hasil seperti tabel di bawah ini:

Tabel 2. Pengujian Akurasi Data Rasio 80:20

Berdasarkan tabel diatas, diperoleh metode *Naïve Bayes* sebagai metode klasifikasi terbaik diantara 2 metode lainnya. Hal tersebut dapat kita lihat bahwa nilai klasifikasi *Naïve Bayes* yang paling bagus pada 2 metode dengan nilai pengukuran yaitu

MODEL	AKURASI
Decission Tree	0.7900943396226415
Naïve Bayes	0.8592767295597484
K- Nearest Neighbor	0.8474842767295597

0.8592767295597484 yang dimana terdapat dalam pengujian data dengan rasio 80:20 aalah nilai yang terbaik yaitu *Naïve Bayes*.

Pengujian Akurasi Pada Rasio Data 70:30

Pengujian data ketiga menggunakan pembagian data dengan rasio 70:30. Terdapat sebanyak 70% *data training* dan 30% *data testing* dari dataset, diperoleh hasil seperti tabel di bawah ini:

Tabel 3. Pengujian Akurasi Data Rasio 70:30

MODEL	AKURASI
Decission Tree	0.7900943396226415
Naïve Bayes	0.8592767295597484
K- Nearest Neighbor	0.8474842767295597

Pada data rasio 70:30, *Naïve Bayes* menjadi metode klasifikasi terbaik dengan paling banyak menghasilkan nilai klasifikasi terbaik dari metode lainnya. Nilai akurasi yang di pada *Decission Tree* 0.7900943396226415, *Naïve Bayes* 0.8592767295597484, *K-Nearest Neighbor* 0.8474842767295597.

Perbandingan 3 Metode Klasifikasi Terbaik

Setelah melakukan pengujian akurasi dengan 3 bentuk rasio data, kemudian diperoleh 1 metode terbaik yang paling banyak menghasilkan nilai klasifikasi terbaik. 3 metode tersebut yaitu *Decission Tree*, *Naïve Bayes*, *K- Nearest Neighbor* Masing-masing metode memiliki tingkat akurasi yang tinggi diantaranya dalam rasio 90:10, *Naïve Bayes* sebagai metode terbaik dalam rasio tersebut memiliki tingkat akurasi sebesar 85%, pada rasio 80:20 *Naïve Bayes* sebagai metode terbaik dalam rasio tersebut memiliki nilai akurasi sebesar 85% , selanjutnya rasio 70:30 *Naïve Bayes* sebagai metode terbaik dalam rasio tersebut memiliki nilai akurasi sebesar 85%. Berikut ini hasil perbandingan 3 metode terbaik dengan tingkat akurasi yang di hasilkan:

Tabel 4 Perbandingan 3 Metode Terbaik

Metode Klasifikasi	90:10	80:20	70:30
<i>Decission Tree</i>	85%	-	-
<i>Naïve Bayes</i>	-	85%	-
<i>K- Nearest Neighbor</i>	-	-	85%

Pada tabel diatas, dapat diketahui bahwa *Decission Tree* menghasilkan nilai akurasi sebanyak 1kali yakni pada pengujian akurasi dengan rasio 90:10. Sementara 2 metode lainnya yaitu *Naïve Bayes* dan *K- Nearest Neighbor* masing-masing menghasilkan nilai akurasi sebanyak 1 kali yaitu rasio 80:20 *Naïve Bayes* dan 70:30 untuk *K- Nearest Neighbor*. Dengan demikian bahwa berdasarkan hasil pengujian dan perbandingan metode terbaik yang tersisa dihasilkan metode *Naïve Bayes* sebagai metode yang terbaik diantara 2 metode klasifikasi lainnya yakni 85%.

KESIMPULAN

Bahwa perkembangan teknologi dan informasi telah menjadikan internet sebagai alat komunikasi yang sangat diminati oleh masyarakat. Media sosial seperti Twitter, Facebook, dan Youtube menjadi salah satu wadah komunikasi yang paling populer di kalangan pengguna internet. Hal ini telah menjadi fokus penelitian bagi para ilmuwan dan peneliti untuk memahami dinamika masyarakat di Indonesia. Selanjutnya, pemahaman tentang pentingnya organ hati dalam tubuh manusia menjadi pusat perhatian dalam konteks kesehatan. Meningkatnya kasus penyakit hati di Indonesia disebabkan oleh berbagai faktor, termasuk infeksi virus dan kebiasaan kurang sehat seperti kurang menjaga kebersihan, pola makan tidak teratur, dan konsumsi alkohol. Hal ini menunjukkan perlunya kesadaran dan upaya pencegahan dalam menjaga kesehatan tubuh, terutama terkait dengan fungsi organ hati yang sangat vital.

Selain itu, penelitian ini juga menyoroti gejala dan dampak penyakit hati, dengan penekanan pada kesulitan dalam mengenali penyakit hati sejak dini. Deteksi dini sangat penting untuk pengobatan yang efektif dan peningkatan peluang kelangsungan hidup pasien. Oleh karena itu, penelitian ini berfokus pada pengembangan model klasifikasi untuk mengidentifikasi penyakit hati secara dini, dengan menggunakan metode K-NN. Dalam konteks metode riset, penelitian ini mencoba membandingkan tiga metode klasifikasi, yaitu K-NN, *Naïve Bayes*, dan *Decision Tree*, untuk mencari tahu metode mana yang paling akurat dalam mengklasifikasikan data penyakit hati. Pemilihan variabel independen seperti usia, pendidikan, dan kadar gula dalam darah menjadi faktor penting dalam analisis ini.

Melalui pemisahan data set menjadi bagian training dan testing dengan variasi rasio, penelitian ini berusaha menemukan model K-NN terbaik untuk hasil klasifikasi. Hasilnya menunjukkan bahwa algoritma dengan rasio 90:10 memiliki performa superior dengan tingkat akurasi mencapai 99,40%. Hal ini menunjukkan potensi K-NN sebagai metode yang efisien dalam memprediksi penyakit hati. Secara keseluruhan, penelitian ini mencoba memberikan kontribusi dalam pengembangan model klasifikasi untuk deteksi dini penyakit hati, dengan menggabungkan aspek perkembangan teknologi, perhatian terhadap kesehatan organ hati, dan aplikasi metode riset dalam machine learning.

Visualisasi data merupakan alat yang sangat efektif untuk mempermudah pengamatan dan pemahaman terhadap suatu set data. Berbagai model visualisasi, seperti scatter plot, diagram batang, dan diagram lingkaran, dapat digunakan untuk menggambarkan data secara visual, meningkatkan kemudahan pemahaman terhadap informasi yang disajikan. Penelitian ini menggunakan visualisasi scatter plot untuk data fuel consumption rating, yang berkaitan dengan estimasi emisi karbon dioksida kendaraan. Penggunaan warna dalam scatter plot mempermudah analisis hubungan antar variabel, memperjelas sebaran data, dan meningkatkan daya ingat terhadap informasi yang disajikan.

Selanjutnya, visualisasi data pada age, education, glucose, dan TehYearCHD dalam bentuk diagram batang dan lingkaran memberikan gambaran proporsi numerik yang mudah dipahami. Diagram batang memperlihatkan hasil yang berbeda-beda untuk setiap variabel, sedangkan diagram lingkaran memberikan proporsi nilai numerik secara jelas. Penelitian juga melibatkan pengujian tiga metode klasifikasi, yaitu Decision Tree, Naïve Bayes, dan K-Nearest Neighbor, untuk menemukan model terbaik dalam klasifikasi penyakit hati. Pengujian dilakukan dengan tiga variasi rasio data, yaitu 90:10, 80:20, dan 70:30. Hasilnya menunjukkan bahwa Naïve Bayes konsisten memberikan nilai akurasi tertinggi, sehingga dapat dianggap sebagai metode terbaik di antara ketiga metode tersebut.

DAFTAR PUSTAKA

- [1] P. N. Lhokseumawe, K. Pengantar, rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, and R. Andespa, Implementasi Metode K-Nearest Neighbor (K-Nn) Untuk Identifikasi Ujaran Kebencian Terhadap Tokoh Politik Pada Twitter, vol. 2, no. 1. 2019.
- [2] F. Wafiyah, N. Hidayat, and R. S. Perdana, “Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam,” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 10, pp. 1210–1219, 2017.
- [3] Prabantissa Citra Nurina, “1818-5505-1-Pb,” Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform., vol. 1, no. 1, pp. 263–268, 2021.
- [4] R. A. Siahaan, M. Nasution, M. Nirmala, and S. Hasibuan, “Model Data Mining untuk Perancangan Aplikasi Diagnostik Inflammatory Liver Disease,” vol. 06, pp. 145–153, 2021.
- [5] Elly and S. Qomariyah, “Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning,” vol. 2, no. March, p. Pusporani, 2019.
- [6] E. Pusporani, S. Qomariyah, and I. Irhamah, “Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning,” Inferensi, vol. 2, no. 1, p. 25, 2019, doi: 10.12962/j27213862.v2i1.6810.