

ANALISIS DAN PERBANDINGAN PERFORMA KINERJA METODE *DECISION TREE* DAN KNN DALAM MEMPREDIKSI RISIKO PENYAKIT KARDIOVASKULAR

Rafael Evaldo Setianto

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: ¹rafael.535220086@stu.untar.ac.id

ABSTRAK

Penyakit kardiovaskular atau biasa disebut CVD merupakan penyebab utama kematian global saat ini. Prediksi risiko CVD menjadi penting dalam upaya pencegahan dan penanganan dini. Metode *machine learning* seperti *Decision Tree* dan *K-Nearest Neighbors* (KNN) telah digunakan untuk memprediksi risiko penyakit CVD. Studi ini mengevaluasi performa kinerja kedua metode *machine learning* dalam memprediksi risiko CVD berdasarkan dataset yang mencakup variable penting seperti usia, jenis kelamin, tekanan darah, dan kadar kolesterol. Hasil analisis ini menunjukkan bahwa performa *Decision Tree* dan KNN bervariasi tergantung pada dataset dan parameter yang digunakan. *Decision Tree* cenderung memberikan hasil yang lebih akurat dalam beberapa skenario, sementara KNN unggul dalam situasi lain. Namun demikian, kedua metode memiliki kelebihan dan kekurangan masing-masing, yang perlu dipertimbangkan dalam konteks aplikasi klinis. Penemuan ini memberikan wawasan penting bagi praktisi Kesehatan dalam memilih metode terbaik untuk memprediksi risiko penyakit CVD. Penggunaan metode *machine learning* dapat meningkatkan diagnosis dini dan intervensi yang efektif dalam menangani CVD. Kesimpulannya, pemahaman yang mendalam tentang performa kinerja *Decision Tree* dan KNN dapat membantu meningkatkan penanganan risiko penyakit CVD secara lebih efektif dan efisien.

Kata kunci—Analisis perbandingan, Performa Kinerja Metode, *Decision Tree*, KNN, Prediksi risiko penyakit kardiovaskular

ABSTRACT

Cardiovascular disease (CVD) is the leading cause of death globally today. CVD risk prediction is important for prevention and early treatment. Machine learning methods such as Decision Tree and K-Nearest Neighbors (KNN) have been used to predict CVD risk. This study evaluates the performance of both machine learning methods in predicting CVD risk based on a dataset that includes important variables such as age, gender, blood pressure, and cholesterol levels. The results of this analysis show that the performance of Decision Tree and KNN varies depending on the dataset and parameters used. Decision Tree tends to provide more accurate results in some scenarios, while KNN excels in other situations. However, both methods have their own advantages and disadvantages, which need to be considered in the context of clinical applications. These findings provide important insights for healthcare practitioners in choosing the best method to predict CVD disease risk. The use of machine learning methods can improve early diagnosis and effective intervention in treating CVD. In conclusion, a better understanding of the performance of Decision Tree and KNN can help improve the management of CVD risk more effectively and efficiently.

Keywords—Comparative analysis, Method Performance, *Decision Tree*, KNN, Cardiovascular Disease Risk Prediction

1. PENDAHULUAN

Penyakit kardiovaskular atau biasa disebut CVD telah menjadi salah satu penyebab utama kematian di seluruh dunia, dengan dampak yang signifikan pada Kesehatan Masyarakat dan biaya perawatan Kesehatan. Dengan perkembangan teknologi dan metode analisis data di masa yang sekarang ini, upaya untuk memprediksi risiko CVD ini telah menjadi fokus penelitian yang penting dalam upaya pencegahan dan penanganan penyakit ini.

Dalam rangka meningkatkan prediksi risiko CVD, berbagai metode pembelajaran mesin telah diusulkan dua diantaranya yang paling umum digunakan adalah metode *Decision Tree* dan *K-Nearest Neighbors* (KNN). Kedua metode ini menawarkan pendekatan yang kuat dan fleksibel untuk pemodelan data yang kompleks, termasuk dalam konteks memprediksi risiko kesehatan.

Tujuan dari penelitian ini adalah untuk melakukan analisis perbandingan performa kinerja antara metode *Decision Tree* dan KNN dalam memprediksi risiko penyakit kardiovaskular. Melalui perbandingan ini, diharapkan dapat diketahui metode mana yang lebih akurat dan lebih efektif dalam memprediksi risiko individu yang terkena penyakit kardiovaskular. Dengan pemahaman yang lebih baik tentang kelebihan dan kekurangan masing masing metode, diharapkan dapat meningkatkan upaya pencegahan, diagnosis dan pengelolaan secara efektif untuk penyakit kardiovaskular.

2. METODE PENELITIAN

Algoritma yang akan saya gunakan adalah *Decision Tree* dan *K-Nearest Neighbors* (KNN). *Decision Tree* merupakan algoritma Machine Learning yang menghasilkan model prediktif dalam bentuk pohon struktur Keputusan. Pohon Keputusan ini menggunakan aturan-aturan yang diturunkan dari fitur-fitur data untuk memprediksi nilai. Penerapan Algoritma *Decision Tree* akan memecah dataset menjadi subset-subset yang lebih kecil berdasarkan aturan-aturan yang ditentukan, dengan tujuan memaksimalkan kehomogenan setiap subset dalam nilai target. Pemilihan fitur terbaik untuk pemecahan dilakukan berdasarkan kriteria seperti entropi, Gini impurity, atau informasi gain.

Kelebihan algoritma *Decision Tree* adalah Mudah dipahami dan diinterpretasi, dapat mengatasi data kategorikal dan numerik, serta tidak memerlukan asumsi distribusi data tertentu. Begitu pula algoritma *Decision Tree* juga memiliki kelemahan yaitu rentan terhadap overfitting jika tidak diatur dengan baik, dan kecenderungan untuk tidak memperhatikan interaksi antar fitur.

K-Nearest Neighbors (KNN) merupakan algoritma Machine learning yang berbasis instansi, yang bertujuan mengklasifikasikan instansi baru berdasarkan mayoritas label dari k-instansi terdekat dalam fitur. Penerapan Algoritma KNN ini menggunakan jarak seperti jarak Euclidean untuk menentukan k-instansi terdekat dari instansi yang ingin diprediksi. Jumlah tetangga terdekat yang akan dipertimbangkan atau yang biasa disebut nilai k dipilih sebelumnya. Kemudian, label mayoritas dari tetangga-tetangga ini akan diambil sebagai prediksi untuk instansi yang diberikan.

Kelebihan algoritma *K-Nearest Neighbors* (KNN) adalah sederhana dan mudah diimplementasikan, tidak memerlukan proses pelatihan, dapat menangani data multikelas dan data non linier. Ada juga kelemahan dari algoritma *K-Nearest Neighbors* (KNN) yaitu sensitif terhadap pemilihan nilai parameter k, diperlukannya perhitungan jarak yang mahal, dan tidak cocok untuk data dengan dimensi yang sangat tinggi.

2.1 Metode Evaluasi

Dalam penelitian kali ini, metode evaluasi yang akan saya gunakan untuk membandingkan performa kinerja antara metode *Decision Tree* dan KNN dalam memprediksi risiko penyakit kardiovaskular. Dibawah ini penjelasan tentang metode evaluasi yang saya gunakan:

1. Pembagian Dataset

Dataset akan dibagi menjadi dua yaitu data latih dan data uji. Data latih ini akan digunakan untuk melatih model, sedangkan data uji akan digunakan untuk menguji kinerja model yang dilatih.

2. Validasi silang (Cross-Validation):

Metode validasi silang ini akan saya gunakan untuk menghindari overfitting dan memperoleh estimasi kinerja model yang lebih akurat. Sebuah skema validasi silang yakni validasi silang K-Fold akan digunakan untuk membagi data menjadi subset-subset yang saling tidak tumpang tindih, Dimana setiap subset digunakan secara bergantian sebagai set data latih dan data uji.

3. Matriks Evaluasi:

Berbagai matrik evaluasi saya gunakan untuk mengukur kinerja model dalam memprediksi risiko penyakit kardiovaskular. Matriks yang saya gunakan adalah:

- Akurasi (*Accuracy*): Tingkat kebenaran prediksi dari semua prediksi yang dibuat oleh model. Ini menggambarkan seberapa tepat model dalam memprediksi kelas data secara keseluruhan.
- Presisi (*Precision*): Ketepatan dalam mengidentifikasi prediksi positif dibandingkan dengan total prediksi positif yang dilakukan oleh model. Presisi memberikan gambaran tentang seberapa tepat model dalam mengklasifikasikan sampel sebagai kelas positif.
- Recall (*Sensitivity*): Kemampuan model untuk mengidentifikasi semua instansi positif yang sebenarnya. Recall mengukur seberapa baik model dalam menemukan semua instansi positif yang ada dalam data.
- F1-Score: Nilai gabungan yang mencerminkan keseimbangan antara presisi dan recall. F1-Score memberikan gambaran holistic tentang kinerja model dengan mempertimbangkan baik presisi maupun recall. Semakin tinggi nilai F1-Scorenya, semakin baik juga keseimbangan antara recall dan presisi model.

4. Perbandingan Performa:

Setelah model Decision Tree dan KNN dilatih dan dievaluasi menggunakan matriks evaluasi diatas dan hasilnya dibandingkan, perbandingan dapat dilakukan berdasarkan hasil akurasi, presisi, recall, F1-Score, atau metrik evaluasi yang relevan. Analisis perbandingan akan menyoroti kelebihan dan kelemahan masing masing metode dalam konteks memprediksi risiko penyakit kardiovaskular.

2.1.1 Metode Penelitian

Metode penelitian kali ini menggunakan metode klasifikasi. Metode penelitian dengan menggunakan metode klasifikasi adalah pendekatan yang digunakan untuk mengklasifikasikan atau mengelompokkan data dalam kategori atau kelas tertentu berdasarkan fitur-fitur yang diberikan. Tujuan dari metode ini adalah untuk memprediksi atau mengklasifikasikan berdasarkan pola atau karakteristik tertentu yang ditemukan dalam data.

Prosesnya dimulai dengan mengumpulkan data yang relevan lalu kemudian disiapkan dan diproses untuk dianalisis. Langkah selanjutnya adalah membagi data menjadi dua subset, yaitu data latih dan data uji. Data latih digunakan untuk melatih model klasifikasi, sedangkan data uji digunakan untuk pengujian kinerja model data yang telah dilatih.

Ada berbagai metode klasifikasi yang digunakan, seperti *Decision Tree*, *K-Nearest Neighbors* (KNN), *Support Vector Machines* (SVM), dan lainnya. Setiap metode klasifikasi pasti memiliki kelebihan dan kelemahan di masing-masing metodenya, dan pemilihan metode yang tepat tergantung pada karakteristik data dan tujuan dari penelitian.

Selama proses pelatihan, model klasifikasi akan mempelajari pola atau aturan dari data latih untuk pengklasifikasian data secara benar. Setelah proses pelatihan selesai, model akan diuji menggunakan data uji untuk mengukur kinerjanya. Hasil evaluasi kinerja model biasanya menggunakan matriks seperti akurasi (*Accuracy*), presisi (*Precision*), recall (*Sensitivity*), dan F1-Score yang telah dijelaskan sebelumnya.

Metode klasifikasi dapat diterapkan dalam berbagai bidang, termasuk analisis risiko Kesehatan, deteksi fraud keuangan, klasifikasi gambar, dan lain-lain. Dengan menggunakan metode klasifikasi ini, peneliti dapat menghasilkan prediksi atau klasifikasi yang sangat akurat berdasarkan data yang dianalisis.

2.2. Rumus Matematika

Rumus perhitungan dilakukan dengan mencari information gain yang paling baik ditunjukkan pada rumus 1.

$$Gain(S, A) = Entropy(S) = \sum_{v \in values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad 1$$

Rumus perhiyungan jarak Euclidean ditunjukkan pada rumus 2

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad 2$$

3. HASIL DAN PEMBAHASAN

Dataset yang saya gunakan adalah dataset “*Cardiovascular_Disease_Dataset.csv*”. Dataset ini merupakan Kumpulan data yang menyediakan informasi mengenai berbagai atribut dengan risiko penyakit kardiovaskular pada sejumlah pasien. Dataset ini memiliki jumlah sample sebanyak 1001 sample dan data ini bersumber dari Kaggle, dan tipe data pada dataset ini bersifat numerik. Berikut adalah penjelasan mengenai setiap atribut yang terdapat dalam dataset ini:

Patientid: Atribut ini merupakan identifikasi unik untuk setiap pasien yang ada dalam dataset. Setiap nilai dalam kolom ini mempresentasikan nomor identifikasi pasien.

1. Age: Atribut ini mencatat usia pasien dalam tahun
2. Gender: Atribut ini mencatat jenis kelamin pasien, yang dapat berupa laki laki atau Perempuan.
3. Chestpain: Atribut ini mencatat tipe nyeri dada yang dirasakan oleh pasien, yang mungkin dibagi menjadi beberapa kategori seperti “Angina tipikal”, “Angina atipikal”, “Nyeri non-anginal”, atau “Asimtomatik”.
4. RestingBP: Atribut ini mencatat tekanan darah istirahat pasien, diukur dalam milimeter air raksa (mmHg).
5. Serumcholesterol: Atribut ini mencatat kadar kolesterol serum pasien, mungkin diukur dalam milligram per desiliter (mg/dL).
6. Fastingbloodsugar: Atribut ini mungkin mencatat kadar glukosa darah pasien setelah puasa, yang mungkin diungkapkan sebagai “Kurang dari 120 mg/dL” atau “Lebih dari 120 mg/dL”.
7. Restingelectro: Atribut ini mencatat hasil elektrodiagram istirahat pasien, mungkin dalam bentuk “Normal” atau “Tidak normal”.
8. Maxheartrate: Atribut ini mencatat detak jantung maksimum yang dicapai oleh pasien selama tes, diukur dalam denyut per menit (bpm).
9. Exercieangia: Atribut ini mencatat apakah pasien mengalami angina yang diinduksi oleh latihan fisik, yang mungkin didukung sebagai “Ya” atau “Tidak”.
10. Oldpeak: Atribut ini mungkin mengukur depresi segmen ST dari elektrokardiogram, yang merupakan perubahan relative terhadap posisi diam.
11. Slope: Atribut ini mungkin mencatat kemiringan segmen ST puncak latihan fisik pasien, yang dapat berupa “Naik”, “Datar”, atau “Turun”.
12. Noofmajorvessels: Atribut ini mencatat jumlah pembuluh darah besar yang berkurang dalam angiografi pasien.
13. Target: Atribut ini merupakan variable target yang menunjukkan apakah pasien didiagnosis mengalami penyakit kardiovaskular atau tidak. Nilai dalam atribut ini mungkin berupa “Positif” (menderita penyakit kardiovaskular) atau “Negatif” (tidak menderita penyakit kardiovaskular).

Hasil dari eksperimen saya dengan menggunakan metode Decision Tree dan KNN adalah:

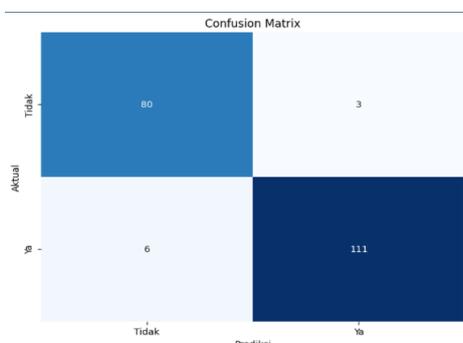
1. Metode K-Nearest Neighbors (KNN):

- Memiliki akurasi model: 0.955.
- Tabel classification report hasil prediksi risiko penyakit kardiovaskular ditampilkan pada tabel (1).

Tabel 1 Menampilkan classification report metode KNN.

Precision	0.95
Recall	0.96
F1-Score	0.95
Akurasi	0.95
Rata-rata	0.95

- Confusion Matrixnya: $[[80,3] [6, 111]]$.
- Gambar 1 Grafik *confusion matrix* hasil prediksi risiko penyakit kardiovaskular metode KNN.



Gambar 1 Grafik *confusion matrix* metode KNN.

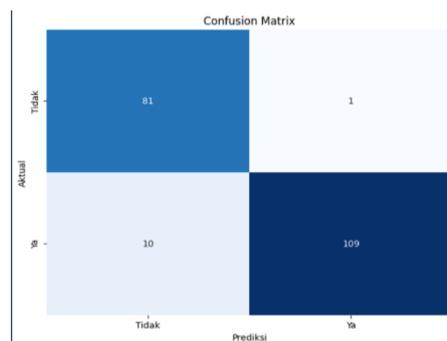
2. Metode *Decision Tree*:

- Memiliki akurasi model: 0.945.
- Tabel *classification report* hasil prediksi risiko penyakit kardiovaskular ditampilkan pada tabel (2).

Tabel 2. Menampilkan *classification report* metode *Decision Tree*

Precision	0.95
Recall	0.95
F1-Score	0.95
Akurasi	0.95
Rata-rata	0.95

- Confusion Matrixnya: $[[81,1] [10, 109]]$.
- Gambar 2 menampilkan Grafik *confusion matrix* hasil prediksi risiko penyakit kardiovaskular metode *Decision Tree*.



Gambar 2 Grafik *confusion matrix* metode *Decision Tree*.

Pembagian data menjadi data latih dan data uji adalah salah satu praktik umum dalam *Machine Learning*. Pembagian ini memiliki tujuan untuk mengevaluasi kinerja model pada data yang tidak pernah dilihat sebelumnya, sehingga dapat memberikan perkiraan prediksi yang lebih baik tentang seberapa baik model tersebut akan berperforma pada dunia nyata. Pembagian umum yang digunakan adalah data latih 80% dan data uji 20%. Berikut penjelasan lebih lanjut mengenai mengapa pembagian ini umumnya digunakan:

1. Data latih (80%):

- Data latih ini digunakan untuk melatih model mesin. Model menggunakan data ini untuk mempelajari pola atau hubungan antara fitur-fitur input (*variabel independent*) dengan label target (*variabel dependen*). Semakin banyak data yang digunakan untuk pelatihan, semakin baik model dapat memahami hubungan antara fitur-fitur tersebut dan membuat prediksi yang akurat.
- Dalam machine learning, jumlah data latih yang cukup besar penting untuk menghindari overfitting. Overfitting terjadi ketika model terlalu menghafal data latih dan tidak dapat menggeneralisasi dengan baik ke data yang baru. Dengan menggunakan 80% dari data sebagai data latih, kita dapat memberikan model cukup informasi untuk belajar tanpa terlalu banyak “menghafal”.

2. Data uji (20%):

- Data uji digunakan untuk menguji kinerja model setelah melatihnya dengan data latih. Data ini seharusnya tidak pernah dilihat oleh model selama proses pelatihan. Penggunaan data uji memungkinkan kita untuk mengukur seberapa baik model yang dilatih dapat melakukan prediksi pada data baru yang tidak dapat digunakan dalam proses pelatihan.
- Dengan menggunakan 20% dari data sebagai data uji, kita dapat memiliki sejumlah data yang cukup besar untuk mengevaluasi kinerja model dengan baik. Jumlah data yang cukup besar di data uji memungkinkan kita untuk mendapatkan perkiraan yang lebih stabil tentang seberapa baik model akan berperforma pada data baru.

Pembagian data latih sebesar 80% dan data uji sebesar 20% adalah pilihan yang umum karena keseimbangan yang baik antara melatih model dengan cukup banyak data dan menguji model pada data yang cukup untuk mengevaluasi kinerjanya. Namun, dalam beberapa kasus tertentu, pembagian yang berbeda mungkin lebih sesuai tergantung pada sifat dan ukuran dari dataset yang digunakan.

Skenario Eksperimen yang dilakukan melibatkan beberapa Langkah yang terstruktur untuk mengevaluasi dan membandingkan kinerja dua metode klasifikasi yaitu metode Decision Tree dan metode KNN, dalam memprediksi risiko penyakit kardiovaskular. Berikut adalah scenario yang saya lakukan:

1. Pengumpulan Data:

Langkah pertama adalah mengumpulkan dataset yang berisi informasi klinis tentang individu seperti usia, jenis kelamin, tekanan darah, tipe nyeri dada, dan faktor risiko lainnya yang dapat berhubungan dengan penyakit kardiovaskular. Dataset ini harus mencakup label target yang menunjukkan apakah individu menderita penyakit kardiovaskular atau tidak.

2. Pembersihan dan pra-pemrosesan data:

Setelah dataset dikumpulkan, Langkah selanjutnya adalah membersihkan dan mempersiapkan data sebelum digunakan. Proses ini mencakup penanganan nilai yang hilang, pengkodean variabel kategori, penskalaan fitur numerik, dan pemisahan data latih dan data uji. Dengan melakukan ini, kita memastikan bahwa data siap digunakan untuk melatih dan menguji model prediksi dengan akurat.

3. Pembagian data:

Pembagian data biasanya dibagi menjadi dua yaitu data latih dan data uji. Pada umumnya, pembagian ini dilakukan dengan proporsi 80% data latih dan 20% data uji. Data latih digunakan

untuk melatih model, sedangkan data uji digunakan untuk menguji atau mengevaluasi kinerja model yang telah dilatih.

4. Pelatihan model:

Dua metode klasifikasi yaitu Decision Tree dan KNN, dilatih menggunakan data latih. Selama pelatihan, model belajar hubungan antara fitur-fitur input (*variabel independent*) dengan label target (*variabel dependen*) menggunakan algoritma yang sesuai dengan masing-masing metode.

5. Evaluasi model:

Setelah model dilatih, kinerja keduanya akan dievaluasi menggunakan data uji, matriks evaluasi yang umumnya digunakan adalah akurasi, yang digunakan untuk mengukur seberapa baik model dapat memprediksi kelas yang benar dari data uji.

6. Analisis hasil:

Hasil dari kedua metode klasifikasi dibandingkan untuk menentukan model mana yang memiliki kinerja yang lebih baik dalam memprediksi risiko penyakit kardiovaskular. Analisis ini dapat melibatkan perbandingan akurasi serta pertimbangan lain seperti kompleksitas model, kecocokan fitur, dan sebagainya.

7. Interpretasi dan kesimpulan:

Berdasarkan hasil evaluasi dan dianalisis, kesimpulan diambil tentang kinerja relative antara Decision Tree dan KNN dalam memprediksi risiko penyakit kardiovaskular. Hasil eksperimen ini dapat digunakan untuk memberikan wawasan yang berharga bagi pemangku kepentingan dalam pengembangan model prediksi penyakit kardiovaskular yang lebih akurat dan efektif di masa yang akan datang.

Hasil analisis dan perbandingan performa kinerja metode *Decision Tree* dan KNN dalam memprediksi risiko penyakit kardiovaskular adalah:

1. Akurasi model:

- Akurasi merupakan metrik utama dalam mengevaluasi kinerja model klasifikasi. Dari eksperimen ini, akurasi dari kedua model yaitu Decision Tree, diukur menggunakan data uji yang sebelumnya tidak pernah dilihat oleh model.
- Model dengan akurasi yang lebih tinggi dianggap memiliki kinerja yang lebih baik dalam memprediksi risiko penyakit kardiovaskular.

2. Kinerja model Decision Tree:

Hasil analisis menunjukkan bahwa model Decision Tree memiliki akurasi sebesar 0.945. Hal ini menunjukkan bahwa model Decision Tree dapat memprediksi risiko penyakit kardiovaskular dengan akurasi 0.95 pada dataset yang digunakan.

3. Kinerja model KNN:

Sementara itu, model KNN memiliki akurasi sebesar 0.955. Dengan demikian, model KNN dapat memprediksi risiko penyakit kardiovaskular dengan akurasi sekitar 0.95 pada dataset yang sama

4. Perbandingan performa:

Dari hasil analisis dapat dilihat bahwa model KNN memberikan akurasi yang lebih tinggi dalam memprediksi risiko penyakit kardiovaskular dibandingkan dengan model Decision Tree.

Oleh karena itu, model KNN dapat dianggap sebagai pilihan yang lebih baik untuk memprediksi risiko penyakit kardiovaskular berdasarkan dataset yang digunakan. Analisis hasil dari penelitian ini adalah memberikan saya wawasan yang berharga tentang kinerja relative dari dua metode klasifikasi tersebut dalam konteks memprediksi risiko penyakit kardiovaskular. Berikut adalah analisis yang saya peroleh dari penelitian:

1. Penelitian ini memberikan saya pemahaman yang luar biasa tentang kinerja metode Decision Tree dan KNN dalam memprediksi risiko penyakit kardiovaskular.

2. Dengan menggunakan dataset yang tersedia, saya dapat mengevaluasi akurasi dari kedua metode yang saya pilih sehingga dapat membandingkannya untuk menentukan model mana yang lebih baik dalam memprediksi risiko penyakit kardiovaskular.

Kendala yang saya hadapi selama penulisan artikel ilmiah ini adalah:

1. Salah satu kendala yang mungkin dihadapi adalah kurangnya data atau kualitas data yang buruk. Jika dataset memiliki nilai yang hilang atau tidak terdistribusi dengan baik, ini dapat memengaruhi kinerja model.
2. Selain itu, pemilihan parameter yang optimal untuk kedua metode (misalnya, kedalaman pohon untuk Decision Tree atau jumlah tetangga untuk KNN) juga dapat menjadi kendala. Memilih parameter yang salah dapat menghasilkan model yang tidak optimal.

Kelebihannya adalah:

1. Metode Decision Tree memiliki kelebihan dalam hal interpretasi karena strukturnya dapat divisualisasikan secara intuitif. Ini memungkinkan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi prediksi risiko penyakit kardiovaskular.
2. KNN, di sisi lain, tidak membuat asumsi tentang distribusi data, sehingga dapat berkinerja dengan baik dengan data yang kompleks atau tidak terstruktur.

Kekurangannya adalah:

1. Decision Tree rentan terhadap overfitting jika tidak diatur dengan benar. Jika terlalu kompleks, model dapat “menghafal” data latih dan gagal menggeneralisasi dengan baik pada data uji.
2. KNN mungkin tidak efisien untuk dataset yang sangat besar, karena perlu menghitung jarak antara setiap titik data uji dengan setiap titik data latih.

Dengan memahami kendala, kelebihan dan kekurangan dari masing-masing metode, penelitian ini dapat memberikan saya panduan yang berharga dalam memilih metode yang paling sesuai untuk memprediksi risiko penyakit kardiovaskular berdasarkan kebutuhan dan karakteristik dataset yang ada.

Perbandingan performa kinerja algoritma saya ditampilkan pada tabel (3)

Tabel 3 Perbandingan performa kinerja algoritma

Metode	Akurasi	Kelebihan	Kekurangan
Decision Tree	Tinggi	Interpretasi mudah	Rentan terhadap overfitting
		Mampu menangani korelasi nonlinear	Sensitif terhadap perubahan kecil dalam data
		Visualisasi struktur model yang intuitif	
KNN	Tinggi	Tidak membuat asumsi tentang distribusi data	Membutuhkan komputasi yang tinggi
		Efektif untuk data kompleks atau tidak terstruktur	Sensitif terhadap fitur yang tidak relevan
		Tidak perlu melatih model	

KESIMPULAN

1. Kedua metode, baik metode Decision Tree maupun metode KNN, menunjukkan akurasi yang tinggi dalam memprediksi risiko penyakit kardiovaskular berdasarkan dataset yang digunakan.

2. Decision Tree memiliki keunggulan dalam interpretasi dan visualisasi struktur model yang intuitif, sementara KNN tidak membuat asumsi tentang distribusi data dan efektif untuk data kompleks atau tidak terstruktur.
3. Namun Decision tree rentan terhadap overfitting dan sensitive terhadap perubahan kecil dalam data, sementara KNN membutuhkan komputasi yang tinggi dan sensitive terhadap fitur yang tidak relevan.

Saran

1. Dalam kasus Dimana interpretasi dan pemahaman faktor-faktor yang mempengaruhi prediksi penting, penggunaan Decision Tree dapat Diutamakan.
2. Namun, jika data kompleks atau data tidak terstruktur, atau jika tidak ada asumsi tentang distribusi data yang dapat dibuat, KNN bisa menjadi pilihan terbaik.
3. Penting untuk memperhatikan penyetulan parameter dan pengoptimalan model untuk mengurangi risiko overfitting pada Decision Tree dan meningkatkan efisiensi komputasi pada KNN.
4. Selain itu, perlu dilakukan penelitian yang berkelanjutan dengan dataset yang lebih besar dan beragam untuk menguji kinerja kedua metode dalam berbagai konteks klinis.

Dengan mempertimbangkan kesimpulan ini, pemilihan metode yang tepat akan memungkinkan pengembangan model prediksi penyakit kardiovaskular yang lebih akurat dan efektif, yang dapat memberikan manfaat yang signifikan dalam praktik klinis perawatan Kesehatan.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih sebesar-besarnya kepada semua pembaca yang telah menghabiskan waktu untuk membaca artikel ilmiah ini. Penelitian ini tidak akan menjadi kenyataan tanpa kontribusi dari berbagai pihak. Saya juga ingin mengungkapkan terima kasih kepada diri saya sendiri sebagai penulis artikel ilmiah ini yang telah memberikan upaya maksimal dalam setiap tahap penelitian ini. Tidak lupa, saya ingin mengucapkan terima kasih yang luar biasa besar kepada partisipan yang telah ikut andil dengan data mereka dalam penelitian ini. Tak kalah pentingnya, saya ingin menyampaikan terima kasih kepada dosen, keluarga atas dukungan moral dan semangat yang diberikan. Terakhir, namun tidak kalah pentingnya, saya ingin menyampaikan penghargaan yang tulus kepada para pembaca atas waktu dan perhatian mereka dalam membaca artikel ilmiah ini. Semoga artikel ini memberikan wawasan yang berharga dan kontribusi yang bermanfaat dalam pemahaman dan penelitian yang lebih lanjut tentang prediksi risiko penyakit kardiovaskular.

DAFTAR PUSTAKA

- [1] B. Charbuty and A. M. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [2] R. Latha et al., "Stock Movement Prediction using KNN Machine Learning Algorithm," 2022 International Conference on Computer Communication and Informatics (ICCCI), Jan. 2022, doi: 10.1109/iccci54379.2022.9740781.
- [3] C. Krittanawong et al., "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Scientific Reports*, vol. 10, no. 1, Sep. 2020, doi: 10.1038/s41598-020-72685-1.
- [4] P. Ghosh et al., "Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, Jan. 2021, doi: 10.1109/access.2021.3053759.
- [5] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of Cardiovascular Diseases based on Machine Learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, May 2021, doi: 10.52810/tiot.2021.100035.
- [6] P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, "Artificial intelligence, machine learning, and

- cardiovascular disease*,” *Clinical Medicine Insights. Cardiology*, vol. 14, p. 117954682092740, Jan. 2020, doi: 10.1177/1179546820927404.
- [7] V. Bhamare, S. R. Chikhale, N. S. Sawakare, A. Y. Kurkunde, and M. S. Autade, “*Heart disease prediction using machine learning algorithms*,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 4, pp. 559–564, Apr. 2024, doi: 10.22214/ijraset.2024.59580.
- [8] Y. Ying et al., “*An enhanced machine learning approach for effective prediction of IgA nephropathy patients with severe proteinuria based on clinical data*,” *Computers in Biology and Medicine*, vol. 173, p. 108341, May 2024, doi: 10.1016/j.compbimed.2024.108341.
- [9] A. Lakshmanarao, T. V. S. Krishna, T. S. R. Kiran, V. M. K. Chinta, S. Ushanag, and N. Supriya, “*Heart disease prediction using ML through enhanced feature engineering with association and correlation analysis*,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, p. 1122, May 2024, doi: 10.11591/ijeecs.v34.I.pp1122-1130.
- [10] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, “*A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis*,” *PloS One*, vol. 16, no. 2, p. e0245909, Feb. 2021, doi: 10.1371/journal.pone.0245909.
- [11] M. Ali, I. S. Amiri, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, “*Hear disease prediction using supervised machine learning algorithms: Performance analysis and comparison*,” *Computers in Biology and Medicine*, vol. 136, p. 104672, Sep. 2021, doi: 10.1016/j.compbimed.2021.104672.
- [12] X. Lang, D. Wu, and W. Mao, “*Comparison of supervised machine learning methods to predict ship propulsion power at sea*,” *Ocean Engineering*, vol. 245, p. 110387, Feb. 2022, doi: 10.1016/j.oceaneng.2021.110387.
- [13] R. O. Alabi et al., “*Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer*,” *International Journal of Medical Informatics*, vol. 136, p. 104068, Apr. 2020, doi: 10.1016/j.ijmedinf.2019.104068.
- [14] P. Dinesh and P. Kalyan Sundaram, “*Medical image prediction for diagnosis of breast cancer disease Comparing the machine learning algorithms: SVM, KNN, Logistic Regression, Random Forest, and Decision Tree to measure accuracy*,” *ECS Transactions*, vol. 107, no. 1, pp. 12681–12691, Apr. 2022, doi: 10.1149/10701.12681ecst.
- [15] G. S. R. Thummala and R. Baskar, “*Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy*,” *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)*, Jul. 2022, doi: 10.1109/icses55317.2022.9914044.
- [16] H. Rajaguru and S. R. S. Chakravarthy, “*Analysis of Decision Tree and K-Nearest neighbor algorithm in the classification of breast cancer*,” *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 12, pp. 3777–3781, Dec. 2019, doi: 10.31557/apjcp.2019.20.12.3777.
- [17] M. Bansal, A. Goyal, and A. Choudhary, “*A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Shor Term Memory algorithms in machine learning*,” *Decision Analytics Journal*, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.
- [18] W. Ren et al., “*A deep learning-based method for the prediction of temporal lobe injury in patients with nasopharyngeal carcinoma*,” *Physica Medica*, vol. 121, p. 103362, May 2024, doi: 10.1016/j.ejmp.2024.103362.
- [19] K. Meixner et al., “*Variability modeling of products, processes, and resources in cyber-physical production systems engineering*,” *Journal of Systems and Software/the Journal of Systems and Software*, vol. 211, p. 112007, May 2024, doi: 10.1016/j.jss.2024.112007.
- [20] Z. Gong and D. Huo, “*Tool condition monitoring in micro milling of brittle materials*,” *Precision Engineering*, vol. 87, pp. 11–22, May 2024, doi: 10.1016/j.precisioneng.2024.01.010.