



Implementasi Metode K-Nearest Neighbors (KNN) Untuk Klasifikasi Penyakit Jantung

Maulana Fansyuri^{1*}, Tyas Fachri Ramdhani², Rivan Saputra³, Deni Setiawan⁴

^{1,2,3,4}Pendidikan Teknik Informatika, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Banten

Email: : ^{1*}dosen02359@unpam.ac.id, ²Rtyasfachri@gmail.com, ³rivan7332@gmail.com,

⁴denisetiawannnn4@gmail.com

(* : coresponding author)

Abstrak – Penyakit jantung merupakan salah satu penyakit yang sangat berbahaya, penyakit ini juga bisa menyerang berbagai kalangan usia. Penyakit jantung meskipun tidak menular, akan menjadi serius bagi semua kalangan usia jika berbagai faktor seperti kurangnya aktivitas fisik, kebiasaan merokok, konsumsi alkohol dan pola hidup tidak sehat dilakukan. Salah satu yang menyebabkan penyakit ini timbul adalah penyumbatan aliran darah yang masuk ke dalam jantung. Penelitian ini memanfaatkan Teknik data mining untuk menggali informasi dari data besar dengan focus pada metode K-Nearest Neighbors (KNN). K-Nearest Neighbors adalah algoritma sederhana namun efektif untuk klasifikasi data yang bekerja dengan menentukan kedekatan objek baru. Metode ini termasuk dalam kategori supervised learning, yang menggunakan data berlabel untuk membangun model prediksi. Data yang digunakan dalam penelitian ini diambil dari Kaggle dengan judul “Dataset Penyakit Jantung” yang terdiri dari 1025 entri dan 13 atribut. Tujuan Penelitian ini adalah untuk memancarkan Tingkat akurasi metode K-NN serta mengembangkan aplikasinya dalam klasifikasi penyakit jantung.

Kata Kunci: Metode K-Nearest Neighbors; Jantung, Data Mining

Abstract – Heart disease is one of the most dangerous diseases, this disease can also attack various age groups. Heart disease, although not contagious, will be serious for all ages if various factors such as lack of physical activity, smoking habits, alcohol consumption and unhealthy lifestyles are carried out. One of the causes of this disease is the blockage of blood flow into the heart. This study utilizes data mining techniques to extract information from big data with a focus on the K-Nearest Neighbors (KNN) method. K-Nearest Neighbors is a simple but effective algorithm for data classification that works by determining the proximity of new objects. This method is included in the category of supervised learning, which uses labeled data to build a prediction model. The data used in this study was taken from Kaggle with the title "Heart Disease Dataset" consisting of 1025 entries and 13 attributes. The purpose of this study is to emit the level of accuracy of the K-NN method and develop its application in the classification of heart disease.

Keywords: K-Nearest Neighbors Method; Heart, Data Mining

1. PENDAHULUAN

Jantung merupakan organ tubuh yang sangat penting bagi manusia. Sebab, jantung berfungsi sebagai sistem peredaran darah pada tubuh serta alat untuk pernafasan bagi manusia. Penyakit jantung merupakan salah satu penyakit yang sangat berbahaya, walaupun penyakit ini tidak menular tetapi penyakit ini dapat menyerang berbagai kalangan usia baik tua maupun muda. Penyakit ini timbul karena berbagai macam faktor seperti kurang aktivitas sehat, kebiasaan merokok, kebiasaan mengkonsumsi minuman alkohol, dan hidup yang tidak sehat.

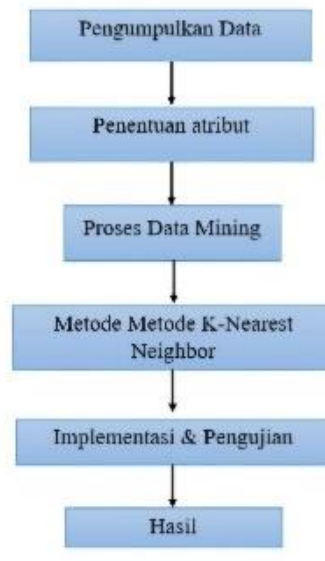
Data mining merupakan serangkaian proses agar dapat menemukan atau menggali informasi atau pengetahuan dari sebuah data yang besar. Ada banyak metode atau algoritma dalam proses data mining seperti K-Nearest Neighbors, K-Means clustering, dan lain-lain. Pada penelitian ini menggunakan salah satu metode data mining yaitu dengan K-Nearest Neighbors (KNN), algoritma ini sederhana namun sangat efektif untuk data yang besar. K-Nearest ini dapat digunakan untuk klasifikasi data dengan menentukan kedekatan objek baru atau tetangga terdekat. Metode ini merupakan bagian dari metode *supervised learning* yaitu salah satu metode dalam data mining yang melibatkan penggunaan data berlabel untuk membuat model supaya model dapat membuat prediksi atau klasifikasi pada data baru.

Untuk melakukan penelitian data yang di gunakan merupakan data yang bersumber dari browser kaggle dengan judul “Dataset Penyakit Jantung” dengan data berjumlah 1025 dan 13 atribut

pendukung. Penelitian ini bertujuan untuk mengetahui tingkat akurasi metode K-NN dan mengembangkan metode K-NN dalam klasifikasi penyakit jantung.

2. METODE

Penelitian ini dirancang untuk mengelompokkan sebuah data, metode penelitian ini terdiri dari beberapa tahapan berupa pengumpulan data, preprocessing, penerapan algoritma K-NN, tahap analisis, hasil evaluasi.



Gambar 1. Alur Metode Penelitian

2.1 Pengumpulan Data

Pada tahap ini dataset yang di kumpulkan dan dipakai merupakan data yag bersumber dari kaggle yang berjudul “Dataset Penyakit Jantung”. Dataset ini terdiri dari 1025 data, terdapat beberapa atribut pada dataset tersebut, yaitu ;

1. Sex : Jenis Kelamin (Direpresentasi menjadi 1 dan 0)
2. cp : nyeri dada (0:nyeri dada mengurangi suply darah ke jantung, 1:nyeri dada tidak berhubungan dengan jantung, 2:kejang eshopagus, 3: nyeri dada tidak menunjukkan tanda sakit)
3. Trestbps : Tekanan darah
4. Chol : Koresterol serum
5. Fbs: Fasting Blood Sugar
6. Restcg : Hasil Elektrokardiogram istirahat
7. Thalach : Denyut jantung maksimal pasien
8. Exang : Keadaan pasien ketika beraktivitas
9. Oldpeak : Depresi ST akibat latihan relative
10. Slope : Kemiringan segmen ST latihan puncak
11. Ca : Jumlah Pembuluh darah utama
12. Thal : Tes stress Thallium
13. Target : Memiliki Penyakit atau tidak (1=positive, 0=negativ) (atribut indikasi)

2.2. Metode K-Nearest Neighbors

Metode yang digunakan kali ini yaitu metode K-Nearest Neighbors (KNN) metode ini merupakan sebuah metode yang termasuk dalam supervised learning atau bisa diartikan sebagai metode yang penggunaannya membutuhkan data berlabel supaya model dapat membuat prediksi atau klasifikasi pada data baru. Kesamaan data dengan tabel menggunakan jarak Euclidian dengan Rumus:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

Keterangan :

D : Jarak

X : Dataset pelatihan

Y : Dataset testing

i : Dataset variabel

Tahapan-tahapan dalam metode K-Nearest Neighbors(KNN) adalah sebagai berikut:

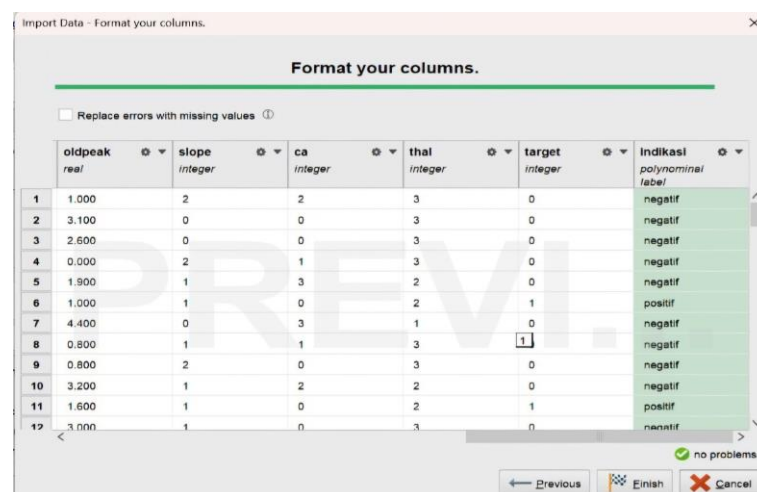
- Tentukan nilai k (Tetangga) terlebih dahulu
- Tentukan jarak diantara dataset testing dengan data pelatihan
- Urutkan jarak serta menentukan objek terdekat ke- k
- Menentukan kategori dari tetangga terdekat
- Menentukan kelas mayoritas dari K tetangga terdekat
- Prediksi atribut atau titik data uji diberi kelas mayoritas sebagai prediksi kelasnya

3. ANALISA DAN PEMBAHASAN

Penggunaan model klasifikasi dengan metode K-Nearest Neighbors (KNN) menggunakan software Rapidminer. Dataset yang dipakai yaitu dataset penyakit jantung yang berasal dari kaggle.com. Data ini terdiri dari 1025 data dan 13 atribut.

3.1 Transformasi Data Menggunakan RapidMiner

Transformasi data merupakan langkah awal yang penting dalam mempersiapkan data sebelum di proses algoritma K-Nearest Neighbors (KNN) pada RapidMiner. Berikut langkah proses transformasi data.

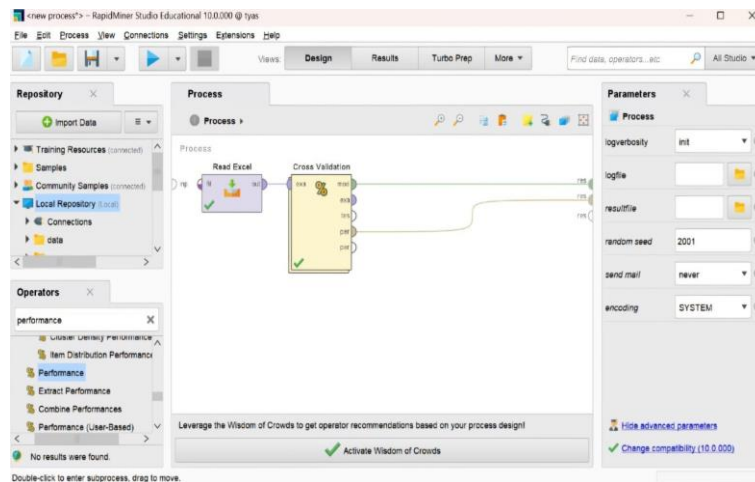


Gambar 2. Transformasi Dataset

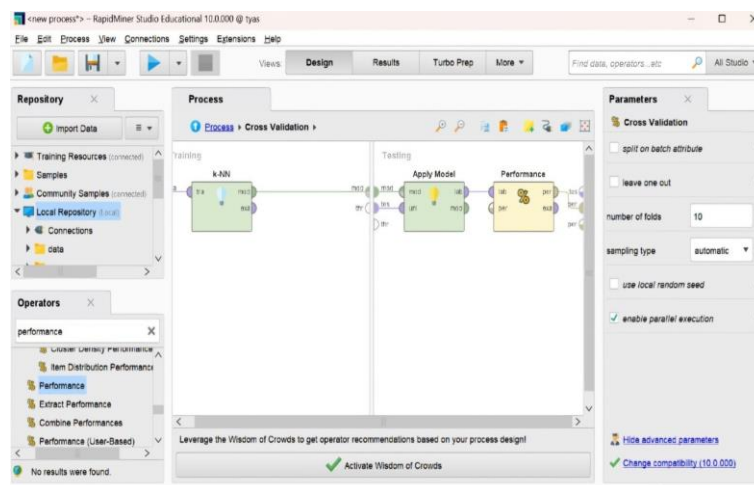
Gambar tersebut merupakan dataset yang mau di transformasikan dengan mengubah role dengan memasukan label, selanjutnya finish sampai data telah masuk ke dalam read excel.

3.2 Data Mining

Pada bagian ini penelitian dengan metode K-Nearest Neighbors dilanjutkan menggunakan tool yang ada pada RapidMiner untuk menganalisa bagaimana faktor-faktor penyakit jantung seperti *Chol* (Kolesterol serum), *ca* (Jumlah Pembuluh Darah Utama), *trestbps* (Tekanan Darah), dan lain-lain dapat mengindikasi penyakit jantung (target). Berikut adalah proses pada RapidMiner setelah mentranformasi data ke dalam RapidMiner.



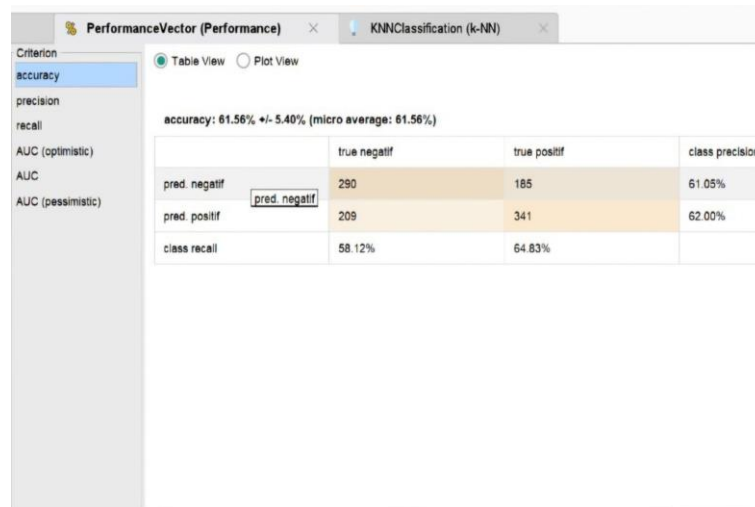
Gambar 3. Process



Gambar 4. Cross Validation Atau Metode K-Nearest Neighbors

Penerapan *Data Mining* dengan metode atau algoritma dari K-Nearest Neighbors dengan menggunakan aplikasi RapidMiner seperti yang terdapat pada Gambar 3 dan Gambar 4. Pada Gambar 3 Read Excel digunakan untuk membaca sebuah *dataset* yang disimpan dalam Excel dan pada Gambar 4 operator cross validation merupakan suatu metode *statistic* yang dapat digunakan untuk mengevaluasi kinerja metode yang digunakan, dalam Cross Validation ini dipisahkan menjadi dua subset, yakni data testing dan data testing(Romadloni et al., 2022). Performance tersebut memberikan sebuah daftar dari nilai kriteria kinerja.

3.3 Hasil Analisa



PerformanceVector (Performance)				KNNClassification (k-NN)			
☒ Table View ☐ Plot View							
accuracy: 61.56% +/- 5.40% (micro average: 61.56%)							
		true negatif	true positif	class precision			
pred. negatif	pred. negatif	290	185	61.05%			
pred. positif		209	341	62.00%			
class recall		58.12%	64.83%				

Gambar 5. Hasil Analisa

Berdasarkan Gambar 5 analisa dari hasil prediksi menggunakan metode K-Nearest Neighbors di dapatkan akurasi sebesar 61,56%. Dengan prediksi pasien yang memiliki penyakit jantung dan positive sebanyak 341 pasien dan pasien yang memiliki penyakit jantung namun negatif sebanyak 209 pasien, maka kelas tersebut mendapat presisi sebesar 62,00%. Sedangkan, prediksi pasien yang tidak memiliki penyakit jantung dan negatif sebesar 290 pasien dan prediksi pasien yang tidak memiliki penyakit jantung namun positif sebanyak 185 pasien dengan kelas presisi sebesar 61,05%. Sehingga didapatkan recall pasien yang memiliki penyakit jantung sebesar 64,83% dan pasien yang tidak memiliki penyakit jantung sebesar 58,12%.

4. KESIMPULAN

Dengan penerapan metode K-Nearest Neighbors (KNN) pada dataset penyakit jantung dengan total 1025 data sengan 13 kriteria dan menggunakan aplikasi RapidMiner dengan penggunaan operator berbagai macam seperti Read Excel, Cross Validation, K-Nearest Neighbors (KNN), Apply Model, dan Performance menghasilkan akurasi sebesar 61,56%, presisi sebesar 61,97%, dan recall sebesar 64,83%. Hasil ini menunjukkan bahwa model KNN yang dipakai belum optimal untuk digunakan klasifikasi penyakit jantung dan perlu di tingkatkan dan perlu dilakukan evaluasi serta pengembangan agar dapat meningkatkan kualitas secara keseluruhan.

REFERENCES

Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>