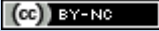


# Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor

Hasran <sup>a,1</sup>

<sup>a</sup> Universitas Muslim Indonesia, Jln. UripSumoharjo Km.5 , Makassar 90231, Indonesia

<sup>1</sup> jayaannur7@gmail.com

INFORMASI ARTIKEL	ABSTRAK
Diterima : 15 Januari 2020 Direvisi : 19 Februari 2020 Diterbitkan : 31 Maret 2020	Secara global, penyebab kematian nomor satu setiap tahunnya adalah penyakit kardiovaskuler. Penyakit kardiovaskuler adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah(Kemenkes RI, 2014). K-Nearest Neighbor (KNN) adalah metode yang mencari kelompok objek dalam data training yang paling mirip dengan objek pada data baru atau data testing (Lestari, 2014). Penelitian ini mencakup pengukuran performa (akurasi, presisi, recall dan f-measure) metode KNN dengan nilai K 3 hingga 9 pada objek 1000 data pasien penyakit jantung yang diperoleh dari pusat dataset UCI Machine Learning Repository. Hasil dari pengukuran performa diperoleh nilai K terbaik adalah 6 dimana nilai akurasi 85%, presisi 78%, recall 93% dan f-measure sebesar 85%.
<b>Kata Kunci:</b> K-Nearest Neighbor Akurasi Presisi recall f-measure	

## I. Pendahuluan

Jantung merupakan organ manusia yang berperan dalam sistem peredaran darah. Penyakit jantung adalah sebuah kondisi dimana jantung tidak dapat melaksanakan tugasnya dengan baik. Secara global, penyebab kematian nomor satu setiap tahunnya adalah penyakit kardiovaskuler. Penyakit kardiovaskuler adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah.

Data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data[1]. Data mining akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data. Hasil keluaran dari data mining ini dapat dijadikan untuk memperbaiki pengambilan keputusan, Salah satu metode yang sering digunakan dalam mengolah data mining adalah metode K-Nearest Neighbor (KNN). KNN dilakukan dengan mencari kelompok objek dalam data training yang paling dekat atau mirip dengan objek pada data baru atau data testing. Dalam ilmu kesehatan menggunakan metode KNN juga dapat digunakan untuk mencari solusi dari pasien baru berdasarkan kedekatan kasus pasien lama [2].

Penelitian ini mencoba untuk menghitung performa metode K-Nearest Neighbor (KNN) pada data penyakit kardiovaskuler. Performa yang dihitung terkait akurasi, presisi, recall dan f-measure

## II. Metode

### A. Data Mining

Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat.

Data Mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering dan asosiasi. Proses dalam tahap data mining terdiri dari tiga langkah Utama, yaitu data Preparation Pada langkah ini, data dipilih, dibersihkan, dan dilakukan preprocessed mengikuti pedoman dan knowledge dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh. Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai. Namun semakin besar data yang diolah maka semakin besar pula waktu prosesnya[3][4].

### B. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. kNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. diperlukan suatu sistem klasifikasi sebagai

sebuah sistem yang mampu mencari informasi. Contoh kasus, misal diinginkan untuk mencari solusi terhadap masalah seorang pasien baru dengan menggunakan solusi dari pasien lama. Perhitungan jarak ketetanggaan menggunakan algoritma eucliden seperti yang ditunjukkan pada persamaan 1.

$$euc = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2}$$

Dimana  $a = a_1, a_2, \dots, a_n$ , dan  $b = b_1, b_2, \dots, b_n$  mewakili  $n$  nilai atribut dari dua record. Untuk atribut dengan nilai kategori..

### Akurasi

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual [5]. rumus akurasi dipaparkan pada persamaan 2.

### Presisi

Presisi didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih[6]. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut. rumus presisi ditunjukkan pada persamaan 3.

### Recall

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Rumus Recall diuraikan pada persamaan 4.

### F-Measure

Measure adalah harmonic mean antara nilai presisi dan recall, F-measure juga kadang disebut dengan nama F1-Score. Rumus F-Measure dijabarkan pada persamaan 5.

$$AKURASI = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$PRESISI = \frac{TP}{(TP+FP)}$$

$$RECALL = \frac{TP}{(TP+FN)}$$

$$F\text{-Measure} = 2 \cdot \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})}$$

Keterangan Variabel:

- TP : True Positive
- TN : True Negative
- FP : False Positive
- FN : False Negative.

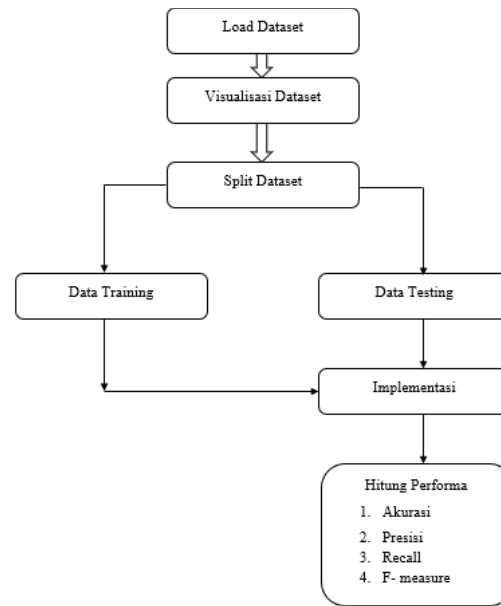
### Rancangan / Model

Rancangan atau model pada penelitian ini seiring dengan alur proses metode KNN dimana dimulai dari mengumpulkan dataset, visualisasi dataset, split atau membagi dataset, implementasi metode KNN hingga menghitung performa metode KNN pada pada dataset tersebut. Flowchart model ditunjukkan pada Gambar berikut.

### Sampel dan Data

Data yang diolah pada penelitian ini memiliki beberapa atribut age (umur), gender (jenis kelamin), height (tinggi badan), weight (berat badan), ap\_hi (tekanan darah sistolik), ap\_lo (tekanan darah diastolik), kolesterol (kolesterol), gluc (gula darah), smoke (status perokok), olco (status peminum), active (status fisik). kesuluruhan nilai bersifat bilangan desimal, sedangkan nilai gender dan active dilabelkan kedalam nilai biner.

Data diperoleh dari UCI Machine Learning Repository, yang mana data tersebut diunggah pada tahun 1988 dan terus diupdate hingga tahun 2018. data di kelola oleh Hungarian Institute of Cardiology. Budapest oleh Andras Janosi, M.D. , University Hospital, Zurich, Switzerland oleh William Steinbrunn, M.D., University Hospital, Basel, Switzerland oleh Matthias Pfisterer, M.D. dan V.A. Medical Center, Long Beach dan Cleveland Clinic Foundation oleh Robert Detrano, M.D., Ph.D.



Gambar 1. Flowchart Rancangan Penelitian

### Teknik Analisis

Teknik analisis dilakukan dengan cara menghitung performa metode KNN dimana performa yang diukur yaitu akurasi, presisi, recall dan f-measure menggunakan persamaan 2, 3 dan 4. Keseluruhan tahap penelitian ini menggunakan scikit-learn library sebagai machine learning tools.

Tahap awal yang dilakukan adalah mengumpulkan data, data yang diperoleh sebanyak 1000 data pasien, data yang kami gunakan pada penelitian ini dapat di unduh pada <http://bit.ly/datasetcardio>. Tahap selanjutnya adalah melakukan split data dimana 90% digunakan sebagai data training dan 10% sebagai data testing.

Tahap selanjutnya adalah menerapkan metode KNN menggunakan data testing dan data training yang telah dipersiapkan sebelumnya. Pada tahap terakhir dilakukan perhitungan performa dari seluruh data testing dengan berbagai simulasi ketetangaan pada metode KNN. Tabel 1 menunjukkan beberapa perintah yang digunakan menggunakan scikit-learn library.

Tabel 1 Source code implementasi metode KNN

Ket	Source Code
Load	<code>dataset = pd.read_csv('cardio.csv')</code>
Split	<code>x = dataset.iloc[:, 1:12]</code> <code>y = dataset.iloc[:, 12]</code> <code>x_train, x_test, y_train, y_test = train_test_split(x,y, random_state = 0, test_size=0.1)</code>
Train	<code>classifier = KNeighborsClassifier(n_neighbors=3,p=2,metric='euclidean')</code> <code>classifier.fit (x_train, y_train)</code>
Test	<code>y_pred = classifier.predict(x_test)</code>
Result	<code>cm = confusion_matrix(y_test, y_pred)</code> <code>cm</code> <code>accuracy_score(y_test, y_pred)</code> <code>precision_score(y_test, y_pred, average = None)</code> <code>recall_score(y_test, y_pred, average = None)</code> <code>f1_score(y_test, y_pred, average=None )</code>

### III. Hasil dan Pembahasan

Kesimpulan pengujian dilakukan setelah proses metode KNN, Pengujian performa pada penelitian ini disimulasikan pada berbagai nilai ketetanggan metode KNN dengan nilai K=3 hingga nilai K=9 untuk melihat nilai performa terbaik. Hasil uji performa metode KNN di tunjukkan pada Table 2.

Tabel 2 Hasil uji performa metode KNN

Nilai K	Akurasi	Presisi	Recall	F <sub>1</sub> -score
3	0,81	0,76	0,87	0,81
4	0,79	0,72	0,89	0,80
5	0,83	0,79	0,87	0,83
6	0,85	0,78	0,93	0,85
7	0,83	0,80	0,85	0,82
8	0,79	0,72	0,89	0,8
9	0,82	0,78	0,85	0,82

Tabel 3 menunjukkan perbandingan beberapa penelitian sebelumnya dengan objek yang sama

Tabel 3 Perbandingan hasil penelitian

Penelitian	Metode	akurasi
Paper ini	KNN	85%
(Chugh dkk, 2017)	GBT, DT	80%, 83%
[8]	KNN	70%
[9]	Chart, ID3, DT	83%, 72%, 82%

### IV. Kesimpulan

Berdasarkan hasil penelitian ini maka penulis dapat menarik beberapa kesimpulan, yaitu dengan simulasi nilai K=3 hingga K=9 maka diketahui bahwa nilai K 6 memiliki nilai akurasi paling baik yaitu sebesar 85% dan nilai presisinya 78% recall 93% dan F-measure 85%. Pada penelitian selanjutnya dapat mengimplementasikan penelitian ketahap selanjutnya berupa decision support system serta dapat menghitng usability aplikasi[10]

#### Ucapan Terima Kasih

Penulis mengucapkan terima kasih pada Fakultas Ilmu Komputer Universitas Muslim Indonesia atas dukungan penggunaan sumberdaya Lab. Riset untuk pengelolaan data penelitian ini

#### Daftar Pustaka

- [1] M. M. Baharuddin, T. Hasanuddin, and H. Azis, "Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269–274, 2019.
- [2] A. Fitria and H. Azis, "Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 102–106, 2018.
- [3] A. A. Karim, H. Azis, and Y. Salim, "Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana 1," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 84–87, 2018.
- [4] L. Nurhayati and H. Azis, "Perancangan Sistem Pendukung Keputusan Untuk Proses Kenaikan Jabatan Struktural Pada Biro Kepegawaian," *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 6–7, 2016.

- 
- [5] H. Azis, R. D. Mallongi, D. Lantara, and Y. Salim, "Comparison of Floyd-Warshall Algorithm and Greedy Algorithm in Determining the Shortest Route," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIconCIT 2018*, pp. 294–298, 2018.
  - [6] N. Fadhillah, Huzain Azis, and D. Lantara, "Validasi Pencarian Kata Kunci Menggunakan Algoritma Levenshtein Distance Berdasarkan Metode Approximate String Matching," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 3–7, 2018.
  - [7] S. Chugh, K. Arivu Selvan, and R. K. Nadesh, "Prediction of heart disease using apache spark analysing decision trees and gradient boosting algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 263, no. 4, pp. 0–10, 2017.
  - [8] M. Lestari, "Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung," *Fakt. Exacta*, vol. 7, no. September 2010, pp. 366–371, 2014.
  - [9] V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," *Caribb. J. Sci. Technol.*, vol. 1, no. December, pp. 208–217, 2013.
  - [10] Rosmasari *et al.*, "Usability Study of Student Academic Portal from a User's Perspective," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIconCIT 2018*, pp. 108–113, 2018.