

Sistem Diagnosis Penyakit Jantung Menggunakan Metode *Modified K-Nearest Neighbor*

Kholif Beryl Gibran¹, Nurul Hidayat², Tibyani³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: 1kholifberylgibr@student.ub.ac.id, 2ntayadih@ub.ac.id, 3tibyani@ub.ac.id

Abstrak

Kardia termasuk kedalam unit yang sangat krusial untuk fisik manusia. Sebagai pemompa darah ke setiap anggota badan adalah fungsi utama dari kardia. Hilangnya fungsi jaringan kardia serta kelainan regulator kardia dan infeksi dapat disebabkan oleh kegagalan kardia terjadi ketika kardia tidak sanggup lagi untuk memenuhi tingkat kebutuhan nutrisi dan oksigen tubuh. Berdasar dari laporan Unit Kesehatan Dunia (WHO) di akibatkan oleh penyakit kardia sepertiga dari 58 juta manusia yang meninggal ditahun 2005 (Afriansyah 2009). Sebab dari kematian nomor 1 didunia untuk saat ini ialah penyakit kardia. Penyakit kardia menyebabkan sedikitnya 30% dari kematian yang ada di hamper seluruh dunia atau sekitar 17 setengah juta pada tahun 2005. Menurut Badan Kesehatan Dunia (WHO), sedangkan penyakit kardia coroner itu sendiri menyebabkan 60% kematian akibat penyakit kardia (WHO, 2007). Sedangkan 26,4 persen kematian dikarenakan oleh kardiovaskuler termasuk juga penyakit kardia coroner, hal itu didasarkan pada Sensus nasional yang dilakukan ditahun 2001 (Depkes RI, 2003). Menurut penjelasan persoalan yg telah di jabarkan dan juga menurut paparan dari riset - riset yang terdahulu , maka dari itu judul yg sesuai untuk riset ini ialah "Sistem Diagnosis Penyakit Jantung Menggunakan Metode Modified K-NN (MKNN)".

Kata kunci : Penyakit jantung, diagnosis, Modified K-NN

Abstract

Cardia is included in the unit which is very crucial for human physical. As a blood pump to each limb is the main function of the cardia. Loss of the function of the cardia tissue and abnormalities of the cardia regulator and infection can be caused by cardia failure occurs when the cardia is no longer able to meet the level of nutrition and oxygen needs of the body. Based on the report of the World Health Unit (WHO) caused by cardia disease one third of the 58 million people who died in 2005 (Afriansyah 2009). The cause of death number 1 in the world for now is cardia. Cardia disease causes at least 30% of deaths in almost the entire world or about 17 and a half million in 2005. According to the World Health Organization (WHO), while coronary cardia disease itself causes 60% of deaths due to cardia (WHO, 2007). While 26.4 percent of deaths are due to cardiovascular disease including coronary cardia, it is based on the National Census conducted in 2001 (MOH RI, 2003). According to the explanation of the problem that has been described and also according to the exposure of previous studies, therefore the appropriate title for this research is "Diagnosis System for Heart Disease Using the Modified K-NN (MKNN) Method".

Keywords : Heart Disease, Diagnosis, Modified K-NN

1. PENDAHULUAN

Kardia termasuk kedalam unit yang sangat krusial untuk fisik manusia. Sebagai pemompa darah ke setiap anggota badan adalah fungsi utama dari kardia. Hilangnya fungsi jaringan kardia serta kelainan regulator kardia dan infeksi dapat disebabkan oleh kegagalan kardia terjadi ketika kardia tidak sanggup lagi untuk memenuhi tingkat kebutuhan nutrisi dan oksigen tubuh. Berdasar dari laporan Unit Kesehatan Dunia (WHO) di akibatkan oleh penyakit kardia sepertiga dari 58 juta manusia yang meninggal ditahun 2005 (Afriansyah 2009).

Sebab dari kematian nomor 1 didunia untuk saat ini ialah penyakit kardia. Penyakit kardia menyebabkan sedikitnya 30% dari kematian yang ada di hampir seluruh dunia atau sekitar 17 setengah juta pada tahun 2005. Menurut Badan Kesehatan Dunia (WHO), sedangkan penyakit kardia coroner itu sendiri menyebabkan 60% kematian akibat penyakit kardia (WHO, 2007).

Tingginya kandungan dari kadar Kolesterol dalam darah merupakan suatu tanda penyakit kardia. Resiko hingga dua kali lebih besar kemungkinan untuk terkena kardia coroner dialami oleh individu dengan hiperkolesterol. Jika individu memiliki kolesterol di ambang normal maka dapat di katakan jika individu tersebut memiliki insiden penyakit diovaskular (stamler dan kaplan, 1983).

124-192 individu per 100 ribu peningkatan jumlah kemangkatan populasi yang ditemukan pada kalangan perempuan hal itu didasarkan pada kecendrungan meningkatnya penyakit peredaran sebagai sebab dari kematian ditahun 1995-2001 pada data SKRT. Prevalanso dari penyakit kardia coroner yang terjadi di Indonesia sebesar 7,2 persen dengan kcenderungn meningkatnya prevalansi dengan bertambahnya usia dan ditemukannya prevalansi yang lebih tinggi pada perempuan berdasarkan laporan Riskesdas ditahun 2007.

Klasifikasi merupakan bagian mesin pembelajaran yang termasuk proses penemuan suatu model atau function yang membeberkan atau memilah konsep atau data kelas, yang tujuannya ialah agar bisa memperhitungkan kelas dari sebuah objek dimana label dari objek tersebut belum diketahui. Metode - metode

didalam algoritma pengklasifikasian berdasar pada pembelajaran yang terbagi menjadi 2, yaitu metode learner eager dan learner lazy. Algoritma yang termasuk kedalam kelompok learner eager diantaranya ialah Artificial Network Neural (ANN), Support Machine Vector (SVM), Decision Tree dan Bayesan. Sementara itu Algoritma yang termasuk kedalam golongan learner lazy diantaranya ialah Rote Clasifier, K-Nearest Neighbor (K-NN), Fuzzy K-Nearest Neighbor (FK-NN) dan Regresi Linier (Prasetyo, 2014).

Modified K-NN memberikan pelabelan kelas untuk data yang diuji berdasar K data pelatihan yang sudah tervalidasi kemudian bobot dari setiap data pelatihan tersebut juga telah tervalidasi, bukan sekedar berdasar jarak paling dekat seperti yang dilakukan di K-NN. Modified K-NN juga memberikan peluang yg lebih banyak untuk data latih yang juga mempunyai validitas yg lumayan besar, Yang menjadikan pengklasifikasian tidak hanya sangat dipengaruhi oleh data yang tidak menentu atau punya validitas yang sedikit. Validitas dan bobot yang dipunyai oleh Modified K-NN sendiri bisa menanggapi kekurangan dari pengklasifikasian berdasar jarak yang paling dekat (Parvin, 2008). Terdapat 2 tahapan didalam Modified K-NN. Tahap kesatu ialah penghitungan validitas dari setiap data pelatihan. Pada bagian berikut nya ialah penghitungan bobot pada k data pelatihan yang terdekat dengan data pengujian.

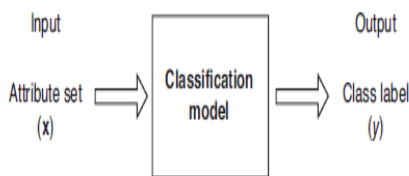
Menurut penjelasan persoalan yg telah di jabarkan dan juga menurut paparan dari riset - riset yang terdahulu , maka dari itu judul yg sesuai untuk riset ini ialah "Sistem Diagnosis Penyakit Kardia Menggunakan Metode *Modified K-NN* (MKNN)".

2. LANDASAN KEPUSTAKAAN

2.1 Klasifikasi

Pelatihan dan pengkasifikasian ialah proses dari klasifikasi. Tahap pembuatan pola disebut pelatihan sedangkan tahap penentuan golongan berdasar pola yang sudah dibuat disebut klasifikasi. Fungsionalitas Mining data ialah Salah satu fungsionalitas klasifikasi. Proses untuk klasifikasi bisa di lihat

pada gambar 2.7



Gambar 2.7 Proses Klasifikasi

Sumber: Tan et al, 2006

Dengan bantuan algoritma pengklasifikasian, perkelas yang dibangun dan dianalisa polanya dari set data pelatihan adalah tahap operasi pelatihan. Pada proses pelatihan sebenarnya perkelas dari data pelatihan sudah kita ketahui karena pola dari data pelatihan itulah yang akan digunakan pada proses pengklasifikasian dimana untuk proses pengklasifikasian kelas dari data pengujian memang belum diketahui.

2.2 MKNN

Algoritma K-NN pasti mempunyai kekurangan. Salah satu contohnya ialah kelas dari objek yang terbaru diputuskan berdasar pada voting dari mayoritas kelas pada K yang jaraknya terdekat. Untuk membenahi kinerja dari algoritma K-NN itu dalam menghitung pengklasifikasian, telah diterapkan beberapa modifikasi pada algoritma K-NN. Salah satu contoh dari modifikasi algoritma K-NN yang sudah diperkenalkan ialah algoritma Modified K-Nearest Neighbor (MKNN). Algoritma ini diperkenalkan oleh Parvin dkk yang tujuannya ialah untuk mengatasi permasalahan ditingkat keakuratan yang sedikit pada algoritma K-NN dan meningkatkan performa dari algoritma K-NN.

Modified K-NN memberikan pelabelan kelas untuk data yang diuji berdasar K data pelatihan yang sudah tervalidasi kemudian bobot dari setiap data pelatihan tersebut juga telah tervalidasi, bukan sekedar berdasar jarak paling dekat seperti yang dilakukan di K-NN. Modified K-NN juga memberi kesempatan yang lebih besar bagi data latih yang juga memiliki validitas yang cukup tinggi, sehingga pengklasifikasian tidak hanya terlalu terpengaruh di data yang tidak stabil atau punya validitas yang sedikit. Validitas dan bobot yang dipunyai oleh Modified K-NN sendiri bisa mengatasi kekurangan dari pengklasifikasian berdasar jarak terdekat (Parvin, 2008). Terdapat 2 tahapan didalam

Modified K-NN. Tahap kesatu ialah penghitungan validitas dari setiap data pelatihan. Tahap selanjutnya ialah penghitungan bobot pada K data pelatihan yang terdekat dengan data pengujian. Berikut adalah pseudo code dari algoritma MK-NN (Parvin, 2010).

```

keluaran_pelabelan := MKNN
(pelatihan_set, test_contoh)
dimulai
For i := 1 to ukuranlatih
Validity(i) := penghitungan validity of
i-th contoh;
akhiri for;
keluaranpelabelan:=Weighted_K-
NN(Validity,contohnya);
Return Output_label;
End
  
```

Urutan operasi Modified K-NN ialah sebagai berikut :

1. Mendapatkan tetangga paling dekat sebanyak jumlah K
2. Mendapatkan selisih antar data pelatihan dengan data pelatihan lain seperti pada rumus (2-1).

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

Keterangan:

- d = jarak
- i = variabel data
- x_i = sampel data
- y_i = data uji
- n = dimensi data

3. Menghitung validitas data latih.

Penghitungan validitas data latih ialah penghitungan yang dilakukan untuk menguji tingkat kevalidan setiap data pelatihan yang didasarkan pada pelabelan perkelas. Validitas data latih didefinisikan pada Rumus (2-2).

$$Validitas(x) = \frac{1}{k} \sum_{i=1}^k S(label(x), labelN_i(x)) \tag{2-2}$$

Keterangan:

- k = jumlah titik terdekat
- $label(x)$ = kelas x
- $label N_i(x)$ = label kelas titik terdekat x

Fungsi S yang terdapat di rumus (2-2) adalah Nilai mirip tidaknya antar titik dengan data tertentu dari tetangga terdekatnya. Fungsi S tersebut didefinisikan dengan rumus (2-3).

$$S(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

(2-3)

Keterangan:

a = kelas a pada data latih

b = kelas selain a pada data latih

a dan b adalah pelabel kelas dari suatu data pelatihan. Jika pelabel kelas a dan pelabel kelas b sama, maka S akan mempunyai nilai 1. Jika pelabel kelas a dan pelabel kelas b tidak sama, maka S akan mempunyai nilai 0.

4. Mendapatkan selisih antar data pelatihan dengan data uji seperti pada rumus (2-1).
5. *Weight voting*

Ciri pembeda dari Modified K-NN adalah *weighted voting* yang menggunakan tingkat kevalidan data dan tetangga yang dekat daripada hanya sekedar mengambil jumlah kelas atau golongan mayoritas. pemobot tiap tetangga terdekat di kalkulasi kan dengan rumus (2-4).

$$W_{(i)} = \frac{1}{d + \alpha}$$

(2-4)

Keterangan:

$W_{(i)}$ = bobot setiap tetangga

d = jarak Euclidean data uji dengan data latih

α = *smoothing regulator*, bernilai 0,5

Voting dari pembobotan ini yang kemudian akan ditotal perkelas di mana kelas yang total votenya tertinggi yang dipilih. Dari sini kemudian didapatkan rumus (2-5) yaitu rumus untuk melakukan *weight voting* pada MK-NN.

$$W_{(i)} = Validitas(i) \times \frac{1}{d + 0.5}$$

(2-5)

Keterangan:

$W_{(i)}$ = *weight voting*

$Validitas(i)$ = nilai validitas

d = jarak data uji dengan data latih

Weight voting menghasilkan dampak yg tinggi untuk sampel yg memiliki bobot validitas yg lebih besar dan juga lebih mendekati dari contoh hasil dari pengujian.

6. Menentukan kelas dari data uji.

Pengelompokkan dari data pengujian di tentukan dari nilai-nilai bobot voting yg paling besar sejumlah K yg telah di ditetapkan. Berdasarkan nilai *weight voting* yg di peroleh, kelas data dari nilai yang paling besarlah yang merupakan kelas data pengujian.

3. METODOLOGI

3.1 Tipe Penelitian

Di sini peneliti memakai riset bertipe non-implementatif analisis karena riset ini berfokus pada hasil algorithm pada persoalan yg di teliti.

3.2 Strategi Penelitian

Dalam riset ini peneliti memakai cara eksperimental, hal ini di pakai untuk melihat pengaruh dari variable tertentu yang terdapat dalam penelitian.

3.3 Partisipan Penelitian

Partisipan penelitian yang terlibat dalam penelitian adalah seorang dokter umum yang bekerja di Rumah Sakit Umum Daerah Jombang, yang berperan sebagai narasumber sekaligus pakar.

3.4 Lokasi Penelitian

Lokasi Penelitian terletak pada Rumah Sakit Umum Daerah Jombang untuk data sekunder, sedangkan lokasi pengolahan data terletak di Universitas Brawijaya laboratorium komputasi cerdas. Data yang diambil berupa gejala-gejala yang di alami penderita penyakit jantung.

3.5 Teknik Pengumpulan Data

Riset memakai data yang telah di dapat kan dari seorang pakar. Data yang di pakai yaitu gejala-gejala penyakit jantung yang di alami oleh pasien berdasarkan keterangan dari pakar.

3.6 Teknik Analisis Data

Disini data yang di pakai yaitu hasil dari data pengujian yang data itu di peroleh dari metode Modified K-NN. Terdapat 2 data yang di pakai yaitu data uji 1 dan data uji 2.

3.7 Implementasi Algoritma

Mula mula peneliti akan memmanualisasikan data agar penilti paham dengan algoritma yang akan di pakai dan juga akan mencoba menerapkan kedalam persoalan yang akan di selesaikan. Berikut nya menggambarkan flowchart agar bisa memperjelas alur dari Modified K-NN pada program.

4. PENGUJIAN DAN ANALISIS

Percobaan ini di laksanakan dengan cara menginputkan nilai K yg bermacam – macam yg nilai nya adalah 5, 10, 15, dan juga 20. Total dari semua data uji yg di pakai ada 172 data, lalu dari seluruh data tersebut terbagi lgi menjadi 2 kelpok yg memiliki 86 jenis data uji.

5. KESIMPULAN

Kesimpulan yg di peroleh menurut percobaan yg tlah di laksanakan dalam diagnosis penyakit kardia memakai MK-NN yaitu:

1. MK-NN bisa di terapkan dalam diagnosa penyakit pada kardia dengan tahap – tahap mengkalkulasikan jarak antar data latih, mengkalkulasikan nilai validitis data latih, mengkalkulasi jarak antar data latih dan data uji, lalu mengkalkulasi weight voting, hasil weight voting yg paling besar di ambil sebanyak K yg di pakai. Hasil weight voting yg diambil, nilai weight voting yg paling besar adalah jenis dari penyakit data uji.
2. Di percobaan ini diperoleh hasil untuk data uji ke 1 apabila nilai K nya besar maka accuracy nya akan semakin naik pula, kebalikan nya saat data uji ke 2 nilai accuracy nya tidak stabil naik dan turun, ketika di lihat hasil rata-rata accuracy untuk data uji ke 1 dan data uji ke 2 diperoleh hasil kalau semakin tinggi nilai K nya maka tingkat accuracy nya semakin tinggi juga.
3. Accuracy terbesar yg di peroleh setelah semua percobaan adalah 94,19% yg di

peroleh dari percobaan data ke 2 yg nilai K nya adalah 5, Accuracy terbesar yg di peroleh dri rata - rata percobaan data uji ke 1 dan data uji ke 2 adalah 93,02 yg di peroleh saat nilai K nya adalah 15.

4. Rata - rata terakhir dari accuracy adalah 91,425%

DAFTAR PUSTAKA

- Parvin H., Hoseinali., & Behrouz M. 2010. Modification on K-Nearest Neighbor Classification. *Global Journal of Computer Science and Technology* Vol.10 Issue 14 (Ver.1.0).
- Wafiyah, F., Hidayat, N., Perdana, R.S., 2017, Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK)*, 1(10), p.1210-1219
- Khotimah, H., 2015, Penentuan Status Gizi Balita menggunakan Metode Modified K-Nearest Neighbor (MK-NN) (Studi Kasus: Kecamatan Kertosono), Universitas Brawijaya, Malang.
- Putri, M.B.P., 2017. Diagnosis Penyakit pada Kucing Menggunakan Metode Modified K-Nearest Neighbor, Universitas Brawijaya, Malang.
- Keminfo., 2017. Faktor Penyebab Kecelakaan https://kominfo.go.id/content/detail/10368/rata-rata-tiga-orang-meninggal-setiap-jam-akibat-kecelakaan-jalan/0/artikel_gpr Diakses April 2018
- Han, J.W., Kamber, M. and Pei, J., 2012. *Data Mining Concepts and Techniques*. 3rd Edition, Morgan Kaufmann Publishers, Waltham.
- Zaiane., 1999. *Principles of Knowledge Discovery in Data* Chapter 1, University of Alberta Computing Science, Kanada.
- Tan, P.N., Steinbach, M., Karpatne, A., Kumar, V., 2006. *Introduction to Data Mining* 2nd Edition, Chapter 3.