

IMPLEMENTASI METODE K-NEAREST NEIGHBOR UNTUK MENGKLASIFIKASI JENIS PENYAKIT KATARAK

M. Safaat¹, A. Sahari², dan D. Lusiyanti³

^{1,2,3}Program Studi Matematika Jurusan Matematika FMIPA Universitas Tadulako

Jalan Soekarno – Hatta Km. 09 Tondo, Palu 94118, Indonesia

¹miratsafaat@gmail.com, ²agusman@untad.ac.id, ³desylusiyanti@untad.ac.id

ABSTRACT

The eyes is one of the five senses that are very important for humans that are used to see the beauty of nature and interact with the environment properly. If the eyes has a problems or diseases, it will be very severe. One of the disorders in the eye is cataract. Cataract if allowed, it will get worse for the sufferer. Therefore, the accuracy of determining the type and layout of early cataract is very important to prevent the more severe effects of cataract. One way to find out early on the type of cataract is by using the mathematical approach to data mining, namely the K-Nearest Neighbor (KNN) method. The concept of the KNN method is to find the nearest neighbor and choose the majority of the classes in the cluster. In this study, the system classified cataract types based on the symptoms experienced by cataract patients at Anutapura Palu Hospital whose research data was obtained from January 2018-March 2018 which amounted to 170 data. The results of this study indicate the accuracy of the KNN method for 170 data at 91.76%

Keywords : Cataract, Classification, K-Nearest Neighbor (KNN)

ABSTRAK

Mata merupakan salah satu panca indera yang sangat penting bagi manusia yang digunakan untuk melihat keindahan alam dan berinteraksi dengan lingkungan sekitar dengan baik. Jika mata mengalami gangguan atau penyakit mata, maka akan sangat berakibat parah. Salah satu gangguan yang ada pada mata adalah penyakit katarak. Penyakit katarak jika di biarkan maka akan semakin parah untuk penderitanya. Oleh karena itu, Ketepatan penentuan jenis dan tata letak katarak secara dini sangat penting untuk mencegah dampak keparahan katarak yang lebih parah. Salah satu cara mengetahui secara dini jenis penyakit katarak adalah dengan pendekatan matematika menggunakan data mining yaitu metode K-Nearest Neighbor (KNN). Konsep metode KNN adalah mencari tetangga terdekat dan memilih mayoritas kelas yang terdapat pada cluster. Pada penelitian ini, sistem mengklasifikasikan jenis katarak berdasarkan gejala yang dialami oleh pasien katarak di Rumah Sakit Anutapura Palu yang data penelitiannya didapatkan dari bulan Januari 2018-Maret 2018 yang berjumlah 170 data. Hasil penelitian ini menunjukkan akurasi metode KNN terhadap 170 data sebesar 91,76 %.

Kata Kunci : Katarak, Klasifikasi, K-Nearest Neighbor (KNN).

I. PENDAHULUAN

1.1. Latar Belakang

Mata merupakan salah satu panca indera yang sangat penting bagi manusia yang digunakan untuk melihat keindahan alam dan berinteraksi dengan lingkungan sekitar dengan baik. Jika mata mengalami gangguan atau penyakit mata, maka akan sangat berakibat fatal. Salah satu gangguan yang ada pada mata adalah penyakit katarak.

Secara etimologis Katarak berasal dari bahasa Yunani *Katarrhakies*, Inggris *Cataracta* yang berarti air terjun. Dalam bahasa Indonesia Bular dimana penglihatan seperti tertutup air terjun akibat lensa yang keruh. Menurut World Health Organisation (WHO) katarak adalah keruhan yang terjadi pada lensa mata, yang menghalangi sinar masuk ke dalam mata.

Secara umum katarak dapat dibagi dalam 6 jenis berdasarkan penyebabnya yaitu katarak Senilis, katarak kongenital, katarak traumatik, katarak komplikata, katarak toksik, dan katarak sekunder (Ilyas, 2003).

Ketepatan penentuan jenis dan tata letak katarak secara dini sangat penting untuk mencegah dampak keparahan katarak yang lebih parah. Prosedur utama diagnosis katarak (Gold Standard Prosedure) dilakukan menggunakan Computed Tomography (CT) scan dan Magnetic resonance Imaging (MRI). Penerapan gold standart prosedur tersebut tidak selalu dapat dilakukan karena tidak semua rumah sakit dan klinik mata memiliki fasilitas yang lengkap, alternatif diagnosis katarak dapat dilakukan melalui pemeriksaan secara fisik, pemeriksaan laboratorium, riwayat penyakit keluarga serta informasi lain yang terkait. Untuk itu diperlukan pengklasifikasian jenis penyakit katarak beserta gejala-gejalanya dengan memanfaatkan perkembangan teknologi informasi untuk menentukan hasil luaran yang benar (Rudi Haryanto dkk, 2016).

Penyakit katarak bisa dilakukan deteksi dini dengan menggunakan metode penyelesaian matematika, salah satunya dengan menggunakan pendekatan data mining. Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Metode K-Nearest Neighbor (K-NN) merupakan salah satu teknik dalam Knowledge Discovery in Database (KDD) dan data mining. Metode ini dapat digunakan untuk mengklasifikasi sedari dini jenis penyakit katarak.

K-Nearest Neighbor (K-NN) adalah suatu metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan data sampel. Diberikan suatu titik query, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik query. Nilai prediksi dari query akan ditentukan berdasarkan klasifikasi

tetangga. K-Nearest Neighbor dapat memberikan keputusan untuk mengklasifikasikan data dari data latih dan mendapatkan hasil yang baik jika menggunakan data dalam jumlah besar (Saxena et al., 2014).

Metode KNN memiliki beberapa kelebihan yaitu tangguh terhadap training data yang noisy dan efektif apabila data latihnya besar. Salah satu dasar inilah yang membuat penulis tertarik menggunakan metode KNN sebagai metode implementasi untuk mengklasifikasi jenis penyakit katarak.

Berdasarkan latar belakang diatas, maka penulis tertarik untuk mengimplementasikan metode K-Nearest Neighbor (K-NN) dalam mengklasifikasi jenis penyakit katarak yang dituangkan dalam sebuah tugas akhir yang berjudul "Implementasi Metode K-Nearest Neighbor (K-NN) Untuk Klasifikasi Jenis Penyakit Katarak".

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini yaitu bagaimana implementasi metode K-Nearest Neighbor (K-NN) dalam pengklasifikasian jenis penyakit katarak

II. METODE PENELITIAN

1. Memulai penelitian
2. Mengumpulkan studi literatur dengan mengumpulkan materi dari buku-buku artikel dan jurnal yang di temukan di media online.
3. Mengumpulkan data di tempat penelitian
4. Preprocessing data
5. Implementasi metode K-Nearest Neighbor (K-NN)
6. Menyimpulkan hasil penelitian
7. Hasil

III. HASIL DAN PEMBAHASAN

3.1. Hasil

3.1.2. Pengumpulan Data

Data diperoleh berupa data rekam medis penderita penyakit katarak yang 170 pasien. Adapun atribut yang digunakan adalah gejala pada mata yang berjumlah 11 gejala dan jenis katarak yang berjumlah 3 jenis katarak.

Tabel 1 : Atribut Data Penelitian

Gejala Pada Mata	Jenis Katarak
1. Penglihatan Kabur	1. Senilis
2. Penglihatan Silau	2. Kongenital
3. Penglihatan Ganda	3. Traumatik
4. Pupil mata berwarna putih	

5. Microftalmia	
6. Microcepaly	
7. Rw. Keluarga tidak bisa melihat	
8. Trauma bnturan keras pada mata	
9. Rw. Operasi	
10. Nyeri pada bagian mata	
11. Rw. Penyakit sistemik	

3.1.2. K-Fold Cross Validation

Dalam algoritma K-NN sebuah data baru diklasifikasikan berdasarkan jarak data baru tersebut dengan tingkat kemiripan data baru terdekat terhadap data pola. Jumlah data tetangga terdekat ditentukan dan dinyatakan dengan k. Penentuan nilai K terbaik dapat ditentukan dengan optimasi parameter, dalam penelitian ini k-Fold Cross Validation yang merupakan salah satu metode yang digunakan untuk mengetahui nilai K paling optimal dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak.

Menurut Fu (1994), K-Fold Cross Validation mengulang k-kali untuk membagi sebuah himpunan contoh secara acak menjadi k subset yang saling bebas, setiap ulangan disisakan satu subset untuk pengujian dan subset lainnya untuk pelatihan. Menurut Hastie et al (2008), dengan K=10 dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, sebab data training pada setiap fold cukup berbeda dengan data training yang asli. Menurut Hastie et al (2008), dengan K=10 dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, sebab data training pada setiap fold cukup berbeda dengan data training yang asli secara keseluruhan, 10-fold cross validation sama-sama direkomendasikan dan disepakati bersama.

Dari 170 data pasien yang di peroleh dari Polimata Rumah Sakit Anutapura maka dalam proses klasifikasi dengan menggunakan metode KNN,dapat diperoleh nilai K-optimal dengan menggunakan 10-fold Cross Validation, yang dimana dari 170 data tersebut akan dibagi menjadi 10 partisi, setelah dibagi menjadi 10 partisi dimana masing-masing partisi memiliki 17 data,yaitu 1 data testing dan 16 data training dengan menggunakan rumus euclid sebagai berikut :

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Keterangan :

X_{1i} = Data Training ke- i

X_{2i} = Data Uji/Testing ke- i

- i = Variabel Data
- d_i = Jarak Data ke- i
- p = Dimensi Data

Agar lebih mudah dalam proses perhitungan nilai euclid dengan data yang banyak, maka digunakan aplikasi microsoft excel untuk meminimalisasi kesalahan dalam perhitungan. Dari proses perhitungan di microsoft excel dengan metode euclid maka didapatkan hasil nilai K optimal yang terdapat pada tabel sebagai berikut :

Tabel 2 : Hasil pencarian Nilai K-Optimal dengan menggunakan Microsoft excel

Data	Partisi										Rata-rata	
	1	2	3	4	5	6	7	8	9	10		
K1	1	6	6	4	6	2	9	3	5	5	4,6	
K2	5	6	3	6	5	6	4	5	4	2	4,6	
K3	2	4	5	5	4	6	5	5	4	4	4,4	
K4	3	6	7	5	3	5	6	6	5	3	4,9	
K5	4	7	5	6	2	7	4	8	3	2	4,8	
K6	6	7	4	5	6	3	5	2	1	4	4,3	
K7	5	2	7	5	2	5	3	4	6	3	4,3	
K8	6	7	7	5	2	5	5	5	5	3	5	
K9	2	8	6	4	1	4	3	3	1	6	3,8	
K10	4	5	2	6	5	5	3	6	7	4	4,7	
K11	6	3	6	8	2	4	3	6	5	4	4,7	
K12	6	5	4	4	4	4	2	4	9	3	4,5	
K13	4	3	5	7	2	4	4	6	3	6	4,4	
K14	6	3	5	6	7	6	4	5	3	3	4,8	
K15	4	5	4	5	5	4	2	2	4	6	4,1	
K16	5	7	5	6	6	3	3	2	2	4	4,3	
											Max	5

3.2. Pembahasan

3.2.1. Implementasi Metode KNN

Setelah nilai K-optimal didapatkan maka langkah selanjutnya adalah implementasi metode KNN untuk mengklasifikasikan apakah pasien memiliki penyakit katarak.

Langkah yang dilakukan adalah Dicari nilai euclid berdasarkan data set :

Tabel 3 : Pencarian Nilai Euclid

No	Umur	L/P	1	2	3	4	5	6	7	8	9	10	11	Jenis Katarak	Hasil Euclid	Hasil Euclid
1	61	L	1	1	0	1	0	0	1	0	1	1	0	1	0	0
2	54	P	1	1	1	1	0	0	0	1	1	1	1	1	4	2
3	1	P	0	1	0	1	1	0	1	0	0	1	0	2	3	1,73205081
4	54	L	1	1	1	1	0	0	1	1	0	1	1	3	4	2
5	71	P	1	1	1	1	1	0	1	0	1	1	1	1	3	1,73205081
6	56	L	1	1	1	1	0	1	0	0	1	0	1	1	5	2,23606798

7	1	P	1	0	0	0	1	1	1	0	1	1	0	2	4	2
8	70	L	1	0	1	1	0	1	0	0	1	1	1	1	5	2,23606798
9	49	L	1	1	1	1	0	0	1	0	1	1	0	1	1	1
10	1	L	1	1	0	1	0	1	0	1	0	1	1	2	5	2,23606798
...
...
169	42	L	1	0	1	0	0	1	0	0	1	1	1	3	6	2,44948974
170	69	L	1	1	1	1	0	1	0	1	0	1	1	1	6	2,44948974

Setelah didapatkan nilai euclid dari dataset maka langkah selanjutnya adalah pemberian nilai ranking

Tabel 4 : Pencarian Nilai Ranking

No	Umur	L/P	1	2	3	4	5	6	7	8	9	10	11	Jenis Katarak	Hasil Euclid	Hasil Euclid	Ranking
1	61	L	1	1	0	1	0	0	1	0	1	1	0	1	0	0	9
2	54	P	1	1	1	1	0	0	0	1	1	1	1	1	4	2	6
3	1	P	0	1	0	1	1	0	1	0	0	1	0	2	3	1,73205081	7
4	54	L	1	1	1	1	0	0	1	1	0	1	1	3	4	2	6
5	71	P	1	1	1	1	1	0	1	0	1	1	1	1	3	1,73205081	7
6	56	L	1	1	1	1	0	1	0	0	1	0	1	1	5	2,23606798	5
7	1	P	1	0	0	0	1	1	1	0	1	1	0	2	4	2	6
8	70	L	1	0	1	1	0	1	0	0	1	1	1	1	5	2,23606798	5
9	49	L	1	1	1	1	0	0	1	0	1	1	0	1	1	1	9
10	1	L	1	1	0	1	0	1	0	1	0	1	1	2	5	2,23606798	5
...
...
169	42	L	1	0	1	0	0	1	0	0	1	1	1	3	6	2,44948974	4
170	69	L	1	1	1	1	0	1	0	1	0	1	1	1	6	2,44948974	4

Untuk menentukan hasil diagnosa maka kita melihat 8 nilai K optimal yaitu yang memiliki ranking 1 sampai 8, Maka didapatkan gejala Senilis adalah gejala yang memiliki jumlah paling banyak yaitu 76, maka dipastikan bahwa gejala katarak yang diderita pasien adalah katarak Senilis.

3.2.2. Uji Akurasi Metode KNN

Dari hasil pengujian akurasi didapatkan jumlah data yang benar diprediksi berjumlah 154 data dan data yang salah atau salah diprediksi berjumlah 16 data dengan rincian sebagai berikut.

1. Prediksi pasien senilis yang benar adalah sebesar 79 pasien
2. Prediksi pasien kongenital yang benar adalah sebesar 27 pasien
3. Prediksi pasien traumatik yang benar adalah sebesar 50 pasien
4. Pasien yang salah diprediksi adalah sebesar 14 pasien

$$\begin{aligned}
 \text{Akurasi data yang sesuai target} &= \frac{\text{Banyaknya data yang benar diprediksi}}{\text{Banyaknya data}} \times 100 \% \\
 &= \frac{156}{170} \times 100 \% \\
 &= 91,76 \%
 \end{aligned}$$

3.2.3 Aplikasi Menggunakan Gui Java

Dalam memudahkan dalam proses klasifikasi jenis katarak perlu adanya rancangan Graphic User Interface (GUI), dimana GUI ini berfungsi untuk perantara sistem dengan pengguna dan memudahkan pengguna untuk menggunakan program yang telah dibangun. Dalam rancangan GUI pada sistem klasifikasi jenis katarak ini dibuat menggunakan software NetBeans dengan bahasa Java. Berikut tampilan GUI yang dapat dilihat pada gambar berikut.



Gambar 1 : Tampilan Awal GUI

Tampilan awal pada aplikasi GUI Java untuk menentukan diagnosa jenis penyakit katarak dalam aplikasi ini memuat jenis gejala yang diderita pada mata, hasil implementasi metode KNN, serta logo Universitas Tadulako dan identitas peneliti sebagai tambahan desain. Untuk melihat diagnosa jenis katarak , klik pada pushbutton Diagnosa maka akan tampil gambar sebagai berikut :



Gambar 2. Tampilan Data Training

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan :

1. Dari hasil pengujian akurasi data yang berjumlah 170 data didapatkan jumlah data yang benar diprediksi berjumlah 156 data dan data yang salah diprediksi berjumlah 14 data dengan rincian sebagai berikut.
 - a. Prediksi pasien senilis yang benar adalah sebesar 79 pasien
 - b. Prediksi pasien kongenital yang benar adalah sebesar 27 pasien
 - c. Prediksi pasien traumatik yang benar adalah sebesar 50 pasien
 - d. Pasien yang salah diprediksi adalah sebesar 14 pasien
2. Akurasi yang diperoleh dari implementasi metode KNN dengan menguji 170 data adalah sebesar 91,76 %

DAFTAR PUSTAKA

- [1] Rudi. H, Achmad Basuki, Rini Nur Hasanah, *Klasifikasi penyakit mata katarak berdasarkan kelainan patologis dengan menggunakan algoritma Learning Vector Quantization*,. Universitas Brawijaya, 2016, Malang
- [2] Ilyas, S, *Katarak*, Universitas Indonesia, 2003, Jakarta
- [3] Fu L, *Neural Network In Computer Intelligence*, McGraw Hill, 1994, Singapura
- [4] Hastie Trevor, Tibshirani Robert, dan Jerome Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Springer, 2008, California
- [5] Saxena, K., Khan, Z., & Singh, S, *Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm*, 2014, International Journal of Computer Science Trends and Technology (IJCST),