

DIAGNOSIS PENYAKIT HEPATITIS C BERDASARKAN FAKTOR-FAKTOR YANG MEMPENGARUHI MENGGUNAKAN METODE MODIFIED K-NEARST NEIGHBOR (MKNN)

Amelia Sukma Sekarsari¹, Hardian Oktavianto², Reni Umilasari³

Program Studi Teknik Informatika, Fakultas Teknik,

Universitas Muhammadiyah Jember.

e-mail: ameliasukma27@gmail.com

ABSTRAK

Hepatitis adalah kelainan hati berupa peradangan (sel) hati. Peradangan ini ditandai dengan meningkatnya kadar enzim hati. Peningkatan ini disebabkan adanya gangguan atau kerusakan membran hati. Ada dua faktor penyebabnya yaitu faktor infeksi dan faktor non infeksi. Faktor penyebab infeksi antara lain virus hepatitis dan bakteri. Berdasarkan banyaknya potensi penyebab munculnya hepatitis yang berkaitan tentang bakteri, virus dan infeksi obat. Jawaban dari setiap pertanyaan gejala dan potensi terburuk kemudian dijadikan sebagai masukan untuk mengklasifikasi Hepatitis. Maka dilakukanlah penarapan menggunakan metode MKNN untuk klasifikasi penyakit Hepatitis yang bertujuan untuk memudahkan ahli medis dalam memperoleh diagnosis sementara dengan mengklasifikasi gejala - gejala yang dialami oleh pasien dengan K data tetangga terdekat untuk mendapatkan nilai akurasi yang tinggi. MKNN merupakan algoritma yang dikembangkan dari algoritma KNN. Data yang digunakan yaitu Penyakit Hepatitis C di tahun 2019. Dari serangkaian pengujian dimulai dari penentuan Data Uji hingga mayoritas data dihasilkan K optimal yaitu K = 3 dengan presentase 90%, Akurasi dengan presentase 96%, Presisi dan Recall dengan presentase 83%.

Kata Kunci : Hepatitis C, Klasifikasi, Modified K- Nearst Neighbor

HEPATITIS C DISEASE DIAGNOSIS BASED ON FACTORS - FACTORS AFFECTING USING MODIFIED K-NEAREST NEIGHBOR (MKNN)

Amelia Sukma Sekarsari¹, Hardian Oktavian², Reni Umilasari³

Information Engineering Program, Faculty of Engineering,

Universitas Muhammadiyah Jember

e-mail: Ameliasukma27@gmail.com¹

ABSTRACT

Hepatitis is a liver disorder in the form of inflammation (cells) of the liver. This inflammation is characterized by increased levels of liver enzymes. This increase is due to interference or damage to the liver membrane. There are two factors, namely infectious factors and non-infectious factors. Factors that cause infection include viral hepatitis and bacteria. Based on the many potential causes of hepatitis related to bacteria, viruses and drug infections. The answers to each question of the symptoms and the worst potential are then used as input to classify Hepatitis. Then the implementation of the MKNN method for the classification of Hepatitis was carried out with the aim of making it easier for medical experts to obtain a provisional diagnosis by classifying the symptoms experienced by patients with K data from their closest neighbors to obtain high accuracy values. MKNN is an algorithm developed from the KNN algorithm. The data used is Hepatitis C in 2019. From a series of tests starting from determining the Test Data until the majority of the data is produced the optimal K, namely $K = 3$, shows accuracy with a percentage of 96%, Precision and Recall with a percentage 83%.

Keywords: Hepatitis C, Classification, Modified K- Nearest Neighbor.

1. PENDAHULUAN

Hepatitis adalah penyakit hati yang berupa peradangan (sel) hati. Peradangan ini ditandai dengan peningkatan kadar enzim hati. Peningkatan ini disebabkan gangguan atau kerusakan pada selaput hati. Ada dua faktor yaitu faktor infeksi dan faktor non infeksi. Berdasarkan banyak potensi penyebab hepatitis yang berhubungan dengan infeksi bakteri, virus dan obat. Kemudian jawaban dari setiap pertanyaan gejala dan pertanyaan mendasar yang paling serius digunakan sebagai masukan untuk klasifikasi hepatitis. Penerapan implementasi algoritma MKNN untuk klasifikasi penyakit Hepatitis bertujuan untuk memudahkan ahli medis dalam memperoleh diagnosis sementara dengan mengklasifikasi gejala-gejala yang dialami oleh pasien dengan K data tetangga terdekat untuk mendapatkan nilai akurasi yang tinggi. MKNN merupakan algoritma yang dikembangkan dari algoritma KNN, algoritma MKNN menambahkan proses baru untuk melakukan klasifikasi yaitu perhitungan nilai validitas untuk mempertimbangkan validitas antar data latih dan perhitungan weighted voting untuk menghitung bobot dari masing-masing terdekat. Penambahan 2 proses baru dalam MKNN diharapkan dapat memperbaiki setiap kesalahan pada proses KNN.

2. PENELITIAN TERKAIT

Penyakit Hepatitis

A. Data Mining

(Davies, 2004) Data mining merupakan suatu proses penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar .

B. Klasifikasi

(Han & Kamber 2006) Klasifikasi merupakan pemrosesan untuk menemukan sebuah model yang menjelaskan konsep atau kelas data, untuk kepentingan

tertentu, yang bias menggunakan pemodelan ini untuk memprediksi kelas objek yang labelnya tidak diketahui.

C. Modified K – Nearest Neighbor

Pada metode ini, terdapat dua langkah modifikasi yang menjadikannya berbeda dengan metode KNN diantaranya yaitu (Parvin, et al., 2009):

1. Proses Validitas Data latih

Pada algoritma Modified K-Nearest Neighbor dataset akan melalui tahap validasi terlebih dahulu/ validasi setiap titik dihitung sesuai dengan tetangganya dan dilakukan hanya satu kali (Parvin dkk, 2008).

Diantara H neighbor terdekat dari dataset x, nilai validasi (x) setiap dataset dilakukan dengan menghitung jumlah data yang memiliki label yang sama dengan label x. Adapun rumus untuk menghitung nilai validasi setiap dataset ialah (Parvin dkk, 2010): $validit(x) =$

$$\frac{1}{H} \sum_{i=1}^H S(Ni(x))(lbl(x), lbl(Ni(x))) \quad (2.2)$$

Keterangan :

H : label dari tetangga yang digunakan

$lbl(x)$: label dari dataset

$Ni(x)$: label dari tetangga terdekat dengan

$i = 1, 2, 3, \dots, H$.

S : Fungsi yang memperhitungkan kesamaan antara satu dataset dan dataset lainnya.

Berikut ialah definisi dari fungsi S : $(a.b) = \begin{cases} 1 & a = b \\ 0 & a \neq 0 \end{cases} \quad (2.3)$

Keterangan :

a = kelas a pada data latih

b = kelas lain selain a pada data latih

2. Proses Pembobotan Dataset (*Weight Voting*)

Hasil tahapan voting bobot digunakan untuk menentukan label kategori data yang akan diujikan. Jika ada beberapa kategori serupa dalam kumpulan data i , skor pemungutan suara berbobot dalam label kategori yang sama ditambahkan dan dibandingkan dengan label pemungutan suara berbobot dari kategori yang berbeda. Kemudian, label kategori dengan nilai voting tertimbang terbesar digunakan sebagai hasil akhir dari data label kategori yang akan diuji. Berikut proses perhitungan bobot akhir :

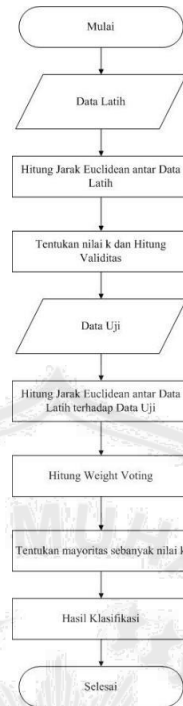
$$W(i) = \text{validity}(i) \times \frac{1}{de+a} \quad (2.6)$$

$w(i)$ dan $\text{validity}(i)$ merupakan bobot dan nilai validasi dari I tetangga terdekat antara dataset dan data testing. Teknik pembobotan diatas akan memberikan nilai yang tinggi untuk dataset yang memiliki nilai validasi besar serta kesamaan yang tinggi dengan data testing (Parvin dkk, 2010).

D. Metode Analisis Data

Metode analisis data yang digunakan adalah Metode Modified K – Nearest Neighbor yaitu teknik klasifikasi data Hepatitis C Virus for Egyptian Patients, dataset tahun 2019. Dalam proses klasifikasi Hepatitis C Virus for Egyptian Patients guna menentukan akurasi diagnosis pada metode Modified K- Nearest Neighbor.

E. Diagram Alur Penelitian



Gambar 3.1 Diagram Alur Penelitian

F. Pengumpulan Data

Pengumpulan data diperoleh dari Website UCI Machine learning Repository (<https://archive.ics.uci.edu/ml/index.php>) yaitu berupa dataset HCV (Hepatitis C Virus for Egyptian patients Data Set) tahun 2019. Data yang digunakan sebanyak 20 atributte.

3. HASIL DAN PEMBAHASAN

Pada bab ini membahas mengenai hasil yang diperoleh dari perhitungan yang telah dilakukan. Data yang telah didapat akan di klasifikasi menggunakan metode *Modified K-Nearst Neighbour* kemudian diolah untuk mendapatkan hasil akurasi yang tepat. Data yang digunakan adalah data kasus Hepatitis C di *Germany* periode 2019.

Tabel 4.1 Data Jumlah Virus Hepatitis C

C	Age	Gender	BMI	Fever	Nausea/Vomiting	Headache	Diarrhea	Fatigue & generalized bone ache	Jaundice	Epigastric pain	WBC	RBC	HGB
1	0,8275862	0	1	1	0	0	0	1	1	1	0,4867179	0,3597126	0,8
2	0,4827586	0	0,5384615	0	1	1	0	1	1	0	1	0,5099736	0
3	0,862069	0	0,8461538	1	1	1	1	0	0	0	0,1302964	0,6695088	0,4
4	0,5862069	1	0,8461538	0	1	0	1	0	1	0	0,3840834	0,8137982	0
5	0,9310345	0	0,7692308	0	0	1	0	1	1	1	0,0735456	0,657183	0,2
6	0,8965517	1	0	1	1	1	0	1	1	0	0,9653128	0,0549354	1
7	0,3448276	1	0,3076923	0	0	1	1	1	1	1	0,9472009	0,7744497	0,4
30	0,5517241	1	0,6153846	0	0	1	1	0	0	1	0,4768386	0,4904366	0,2

a. *Modified K – Nearst Neighbour* pada Jupyter

Data diolah menggunakan Jupyter yang di klasifikasi menggunakan metode *Modified K – Nearst Neighbour*. Output yang dapat dihasilkan dari eksekusi perintah Jupyter adalah hasil tingkat akurasi dan k optimal. Klasifikasi yang dihasilkan dari perintah metode *Modified K- Nearst Nighbour* pada Jupyter akan ditampilkan melalui hasil Run pada tools Jupyter, kemudian akan didapatkan hasil nilai akurasi dan klasifikasi k-optimal

b. Perhitungan Akurasi, Presisi dan Recall

Perhitungan akurasi, Presisi, dan Recall dilakukan menggunakan *Confusion Matrix*. Dimana pada *Confusion Matrix* dilakukan perbandingan antara data Real Class dan Klsifikasi yang didapat dari Mayoritas Data. Untuk perhitungan akurasi, presisi, dan recall sendiri digunakan persamaan dibawah ini :

$$\begin{aligned} \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\ &= \frac{9+10}{9+0+1+0} \times 100\% \\ &= \frac{14}{10} \times 100\% = 90\% \end{aligned}$$

$$\begin{aligned} \text{Presisi} &= \frac{TP+FP}{TP} \times 100\% \\ &= \frac{8+1}{8} \times 100\% = 83\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{TP+FN}{TP} \times 100\% \\ &= \frac{8+0}{8} = 83\% \end{aligned}$$

c. Penentuan Jumlah Nilai K – Optimal

Hasil perhitungan metode *Modified K – Nearst Neighbour* pada *Jupyter* ditampilkan berupa Data hasil nilai Akurasi dan K -Optimal pada penyakit Hepatitis C. Berikut ini adalah hasil metode *Modified K-Nearst Neighbour* pada penentuan nilai akurasi K -Optimal, hasil perhitungan nilai K menunjukkan bahwa pengujian menggunakan $K=3$ menunjukkan akurasi dengan presentase 90%. Pengujian menggunakan $K=5$ menunjukkan akurasi dengan nilai 80% presentase. Pengujian nilai $K=7$ menunjukkan akurasi dengan nilai 70% presentase.



4. KESIMPULAN

Berdasarkan penelitian untuk melakukan diagnosis penyakit hepatitis menggunakan metode *Modified K–Nearst Neighbor* maka dapat diambil kesimpulan sebagai berikut:

1. Implementasi metode *Modified K – Nearst Neighbor* untuk klasifikasi data Penyakit Hepatitis C bisa dilakukan secara akurat dengan nilai Akurasi 90%.
2. Hasil Presisi yang diperoleh dari metode *Modified K – Nearst Neighbor* tertinggi adalah 83%
3. Hasil Recall yang diperoleh dari metode *Modified K – Nearts Neighbor* penyakit Hepatitis C tertinggi adalah 83%.
4. K optimal pada metode *Modified K – Nearst Neighbor* ini ditunjukkan pada $K = 3$.

5. SARAN

Pada penelitian ini masih sangat jauh dari kata sempurna, disarankan pada penelitian selanjutnya agar:

1. Perhitungan manual klasifikasi *Modified K – Nearst Neighbor* di *Microsoft Excel* digunakan sebagai pembandingan hasil klasifikasi pada *Jupyter*. Nilai K yang dipilih lebih baik bisa disesuaikan dengan kebutuhan.
2. Pemilihan parameter jumlah tetangga yang digunakan dalam *Modified K – Nearst Neighbor* dapat disesuaikan dengan jenis data yang digunakan.

DAFTAR PUSTAKA

- Guarsa, Rinaldi. (2018). Sistem Pakar penentuan jenis penyakit hepatitis menggunakan metode logika fuzzy sugeno dan *certainy factor*. *Universitas Pendidikan Indonesia. Reposty.upi.edu*
- Turban, E, (2005). *Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1*. Andi: Yogyakarta
- Kesehatan, Kementrian, RI. (2014). *Pusat data dan Informasi*
- Han, & Kamber. (2006). *Data Mining: Concepts and Techniques*.
- Imanda, A. C., Furqon, M. T., & Hidayat, N. (2018). Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified KNearest Neighbor *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, Malang*.