

Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode *K-Nearest Neighbor – Certainty Factor*

Elsa Nuramilus Shofia¹, Rekyan Regasari Mardi Putri², Achmad Arwan³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹echanuramilus@gmail.com, ²rekyan.rmp@ub.ac.id, ³arwan@ub.ac.id

Abstrak

Demam merupakan salah satu jenis gangguan kesehatan yang mengganggu produktivitas setiap orang bahkan dapat menyebabkan kematian dan masih menjadi permasalahan kesehatan Indonesia. Terdapat beberapa jenis demam yang perlu diwaspadai antara lain Demam Berdarah Dengue, Malaria dan Tifoid. Ketiga penyakit ini memiliki gejala yang mirip, sehingga banyak tenaga medis dan dokter internship yang seringkali melakukan kesalahan dalam mendiagnosis. Untuk itu diperlukan sebuah sistem yang mengadopsi pengetahuan pakar untuk mengatasi masalah tersebut yaitu sistem pakar. Metode yang digunakan untuk mendukung sistem pakar tersebut adalah metode *K-Nearest Neighbor – Certainty Factor* yang merupakan penggabungan 2 metode dimana hasil klasifikasi dari metode *K-Nearest Neighbor* akan diberi nilai kepastian oleh metode *Certainty Factor* sehingga menghasilkan suatu diagnosis penyakit. Pada penelitian ini data latih dan data uji yang digunakan berjumlah 143 data. Berdasarkan hasil pengujian variasi nilai *K* didapatkan akurasi sebesar 88.37%. Pada pengujian variasi data latih didapatkan akurasi sebesar 86.04%. Pada pengujian rasio data latih dan data uji didapatkan akurasi sebesar 95%. Pada pengujian variasi jumlah data uji didapatkan akurasi sebesar 90%. Pada pengujian variasi data uji didapatkan rata-rata akurasi sebesar 97.22%. Pada pengujian perbandingan metode, metode *k-nearest neighbor-certainty factor* menghasilkan akurasi sebesar 84.79%.

Kata kunci: sistem pakar, DBD, Malaria, Tifoid, *K-Nearest Neighbor*, *Certainty Factor*

Abstract

Fever is one of the health problems that disrupt everyone's productivity, even it can cause death and remain a health problem in Indonesia. There are several types of fever that needs to be wary, it includes dengue, malaria and typhoid. These three diseases have similar symptoms, so many medical personnel and doctors internship often make mistakes in diagnosing the disease. Therefore, an expert system is required to resolve the issue. The method used to support the expert system is K-Nearest Neighbor - Certainty Factor which is a merger of two methods in which the classification results of K-Nearest Neighbor to be rated certainty by a Certainty Factor method and resulting a diagnosis of the disease. In this study, the training data and test data used were 143 data. Based on test results obtained K value variation accuracy of 88.37%. On testing variations training data obtained an accuracy of 86.04%. In testing the ratio of training data and test data obtained an accuracy of 95%. In testing the variation of the number of test data obtained an accuracy of 90%. In testing the variety of test data obtained an average accuracy of 97.22%. In comparison testing method, the method k-nearest neighbor certainty factor gets an accuracy of 84.79%.

Keywords: expert system, dengue, Malaria, Typhoid, *K-Nearest Neighbor*, *Certainty Factor*

1. PENDAHULUAN

Kesehatan merupakan hal yang paling didambakan oleh setiap orang, karena dengan hidup sehat segala aktifitas dapat dilakukan dengan baik. Sehingga tidak heran jika seseorang melakukan berbagai cara untuk

menjaga kesehatannya. Namun demikian, bukan tidak mungkin seseorang tersebut dapat terserang suatu penyakit.

Demam merupakan salah satu jenis gangguan kesehatan yang mengganggu produktivitas setiap orang bahkan dapat menyebabkan kematian dan masih menjadi permasalahan kesehatan Indonesia. Terdapat 8

jenis demam yang perlu diwaspadai yaitu Demam Berdarah Dengue (DBD), *Tifoid*, Malaria, *Viral*, *Chicken Guinea*, Meningitis, Infeksi Saluran Kemih dan *HIV*.

Untuk mengetahui seseorang terserang penyakit demam jenis apa harus diketahui dari gejala yang dialami oleh orang tersebut. Padahal gejala pada penyakit Demam sangat mirip dan sulit dibedakan sehingga seringkali terjadi kesalahan dalam mendiagnosis karena keterbatasan pengalaman sehingga kecepatan mendiagnosis penyakit sangat terbatas dan terkadang kurang akurat. Masih banyak dokter internship dan tenaga medis yang sulit dalam mendiagnosis ketiga penyakit ini. Dokter internship adalah dokter yang baru menyelesaikan masa pendidikan profesi dan sedang melakukan program magang (Amir, 2007).

Untuk membantu tenaga medis dan dokter internship dalam mengetahui sejak dini jenis penyakit demam yang diderita, maka diperlukan suatu sistem yang memiliki kemampuan layaknya seorang dokter dalam mendiagnosis penyakit. Sistem tersebut adalah sistem pakar yang diadopsi oleh pengetahuan manusia ke dalam komputer agar dapat membantu menyelesaikan masalah yang biasanya dilakukan oleh pakar. Sistem pakar dapat melakukan penalaran sebagaimana seorang pakar salah satunya dengan menerapkan metode K-Nearest Neighbour - Certainty Factor.

Ada beberapa penelitian yang membahas tentang metode Certainty Factor dan K-Nearest Neighbour. Penelitian tersebut adalah Pengembangan Sistem Pakar Diagnosis Penyakit Tuberculosis dan Demam Berdarah Berbasis Web Menggunakan Metode Certainty Factor. Penelitian yang dilakukan oleh Yulianti Paula Bria dan Engelbertus Agung S. Takung ini bertujuan untuk membantu masyarakat dalam mendiagnosis penyakit TBC dan DBD sehingga dapat ditanggulangi lebih dini dan membantu dokter dalam pengambilan keputusan (Yulianti, 2015). Sedangkan penelitian yang menggunakan algoritma K-Nearest Neighbour adalah Model Diagnosis Tuberculosis Menggunakan K-Nearest Neighbour Berbasis Seleksi Atribut. Ratih Sari dan Purwanto melakukan perbandingan K-Nearest Neighbour dengan dua algoritma lain yaitu *Logistic Regression* dan *Decision Tree*. Input pada penelitian ini adalah karakteristik pasien, anamnesis, pemeriksaan tanda vital dan pemeriksaan radiologi sesuai dengan standar TB yang ditetapkan serta lama

batuk dan warna dahak. Hasil dari eksperimen menunjukkan bahwa model K-NN berbasis backward elimination memberikan kinerja yang terbaik jika dibandingkan dengan model-model lain. Model K-Nearest Neighbour berbasis backward elimination memberikan nilai akurasi sebesar 78.66% dan nilai AUC sebesar 0.806 yang mengindikasikan bahwa model adalah *good classification* (Ratih Sari, 2015).

2. STUDI LITERATURE

2.1 Sistem Pakar (*Expert System*)

Sistem Pakar adalah sistem yang berusaha mengadopsi pengetahuan para pakar dalam menyelesaikan permasalahan berbasis sistem komputer. Menurut Turban dan Aronson "sistem pakar merupakan sistem yang menggunakan pengetahuan manusia yang dimasukkan kedalam komputer untuk memecahkan permasalahan yang biasanya diselesaikan oleh pakar" (Turban, 2005). Dibawah ini merupakan komponen yang dibutuhkan dalam membangun sistem pakar, antara lain:

1. Memori Kerja (Working Memory)
2. Antar Muka Pengguna (User Interface)
3. Mekanisme Inferensi (Inference Machine)
4. Basis Pengetahuan (Knowledge Base)

2.2 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan sebuah algoritma yang melakukan klasifikasi terhadap objek yang memiliki jarak terdekat dengan objek tersebut. Algoritma *K-Nearest Neighbor* mengelompokkan data baru yang belum diketahui kelasnya dengan memilih data sejumlah k yang letaknya paling dekat dari data baru. *Class* paling banyak dari data terdekat sejumlah k akan dipilih sebagai *class* yang diprediksi untuk data baru. Pada umumnya nilai k menggunakan jumlah ganjil agar tidak terdapat jarak yang sama dalam proses klasifikasi. Jauh atau dekatnya tetangga dihitung menggunakan jarak Euclidean.

Perhitungan untuk mencari jarak menggunakan Euclidean ditunjukkan pada persamaan 2.1.

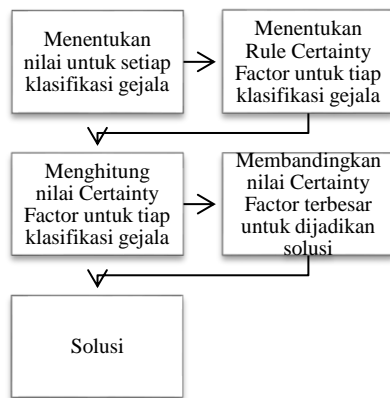
$$\begin{aligned} x_1 &= (x_{11}, x_{12}, \dots, x_{1n}) \\ x_2 &= (x_{21}, x_{22}, \dots, x_{2n}) \\ d(x_1, x_2) &= \sqrt{\sum_r^n (a_r(x_1) - a_r(x_2))^2} \dots \dots \end{aligned} \quad (2.1)$$

Dimana x_1 dan x_2 adalah dua *record* dengan n atribut. Persamaan menghitung jarak antara x_1 dan x_2 , bertujuan untuk menentukan perbedaan antara nilai-nilai atribut pada *record*

x_1 dan x_2 . Setelah diketahui jarak antar *record*, kemudian diambil sebanyak k tetangga terdekat untuk memprediksi label kelas dari record baru menggunakan label kelas tetangga. Diambil kelas mayoritas sebagai kelas target output data yang baru.

2.3 Certainty Factor

Metode *Certainty Factor* adalah teknik yang digunakan untuk mengatasi permasalahan ketidakpastian dalam pengambilan keputusan. *Certainty Factor* dapat terjadi dalam berbagai kondisi, misalnya terdapat beberapa *rule* yang berbeda dengan satu kondisi yang sama. Pada contoh kondisi ini harus dilakukan agregasi pada nilai *Certainty Factor* keseluruhan dari setiap kondisi yang ada. Adapun alur pada metode *Certainty Factor* adalah sebagai berikut:



Gambar 2.1 Diagram alur Metode *Certainty Factor*
 Sumber: (Fadrijn, 2015)

Alur pada metode *Certainty Factor* adalah sebagai berikut:

1. Menentukan nilai dari setiap gejala yang sudah ditentukan pakar dan yang diinputkan oleh user.
2. Menentukan *Rule Certainty Factor* untuk setiap gejala dari setiap penyakit. *Rule* ini berisi gabungan nilai dari pakar dan user.
3. Nilai dari *rule* diatas akan dihitung pada tiap penyakit.
4. Membandingkan nilai terbesar tiap penyakit untuk mendapatkan solusi.
5. Solusi yang telah didapatkan merupakan hasil akhir.

2.4 Demam

Demam adalah suatu keadaan saat suhu badan melebihi 37°C. Demam juga merupakan pertanda bahwa sel antibodi manusia sedang melawan virus atau bakteri. Terdapat 8 jenis demam yang perlu di waspadai antara lain Demam Berdarah Dengue, Tifoid, Malaria,

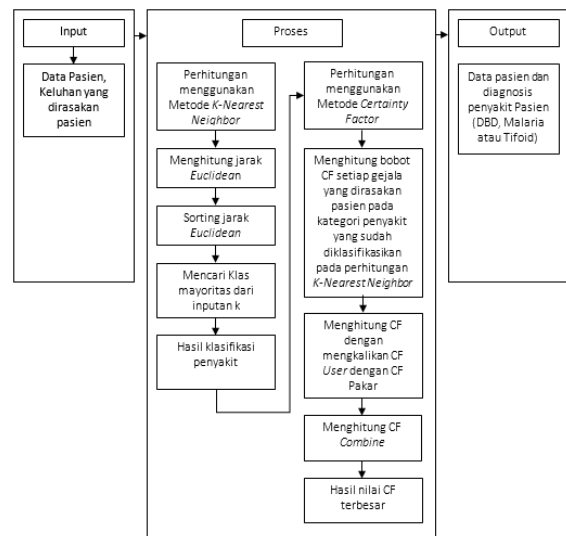
Chicken Guinea, Viral, Meningitis, Infeksi saluran kemih dan HIV. Dari 8 jenis demam tersebut 3 diantaranya memiliki gejala yang mirip yaitu Demam Berdarah Dengue, Malaria dan Tifoid.

3. METODOLOGI PENELITIAN

Metodologi mengenai tahapan-tahapan penelitian yang akan dilakukan. Berikut merupakan langkah-langkah yang akan dilakukan dalam pembuatan sistem “Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode *K-Nearest Neighbor – Certainty Factor*”

3.1 Model Perancangan Sistem

Model perancangan system merupakan tahapan-tahapan dari input awal kemudian proses pengolahan data hingga mendapatkan hasil *output* yang berupa hasil diagnosis penyakit. Model Perancangan system sistem pada diagnosis penyakit Demam: DBD, Malaria dan Tifoid menggunakan metode *K-Nearest Neighbor – Certainty Factor* akan ditunjukkan pada Gambar 3.3.



Gambar 3.1 Model Perancangan sistem

4. PERANCANGAN

Sistem yang dibuat merupakan perangkat lunak yang menggunakan metode *knn-certainty factor* untuk mendiagnosis penyakit. Dimana pada system ini akan diketahui apakah seorang pasien menderita penyakit dbd, malaria atau tifoid.

Data yang dibutuhkan pada penelitian ini adalah data pasien penyakit Demam: DBD, Malaria dan Tifoid yang akan dijadikan sebagai

data latih dan data uji, data gejala dan data nilai tingkatan gejala pada ketiga penyakit tersebut.

Tabel 4.1 Data Nilai Tingkatan Gejala Berdasarkan Jenis Penyakit

NO	VARIABEL	Tifoid	Malaria	DBD
		P1	P2	P3
I KELUHAN				
G1	Demam Intermittent (putus putus)	0.3	0.9	0.2
G2	Demam menggigil	0.4	0.9	0.4
G3	Demam terutama malam hari	0.8	0.2	0.6
G4	Demam lebih dari 1 minggu	0.8	0.7	0.1
G5	Sakit Kepala	0.9	0.7	0.8
G6	Sakit sakit tulang dan sendi	0.5	0.7	1
G7	Mual dan muntah	0.9	0.3	0.7
G8	Mencrret atau susah BAB (konstipasi)	0.7	0.2	0.2
G9	Nyeri Perut	0.5	0	0.9
II PEMERIKSAAN FISIK				
G10	Bintik merah (ptekie) pada kulit	0	0	1
G11	Lidah kotor (coated tongue)	0.9	0.4	0.2
G12	Bradikardi relative	0.8	0	0
G13	Pembesaran hati	0.5	0.5	0.3
G14	Pembesaran limpa	0.6	0.8	0.3
G15	Kulit lembab/ Keringat	0	0.9	0.2

Tabel 4.2 Data nilai tingkatan gejala

NO	VARIABEL	SKOR
G1	Demam Intermittent (putus putus)	
	Ya	0.8
	Tidak	0
G2	Demam menggigil	
	Berat	0.85
	Sedang	0.5
	Tidak ada	0
G3	Demam terutama malam hari	
	Ya	0.75
	Tidak	0
G4	Lama demam	
	> 7 hari	0.75
	4 - 7 hari	0.5
	1 - 3 hari	0.2
G5	Sakit Kepala	
	Berat	0.9
	Sedang	0.7
	Tidak ada	0
G6	Sakit sakit tulang dan sendi	
	Berat	0.9
	Sedang	0.75
	Tidak ada	0
G7	Mual dan muntah	
	Berat	0.85
	Sedang	0.6
	Tidak ada	0

G8	Mencrret atau susah BAB (konstipasi)	
	Ada	0.6
	Tidak ada	0.2
G9	Nyeri Perut	
	Berat	0.8
	Sedang	0.4
	Tidak ada	0
II PEMERIKSAAN FISIK		
G10	Bintik merah (ptekie) pada kulit	
	Berat	1
	Sedang	0.7
	Tidak ada	0
G11	Lidah kotor (coated tongue)	
	Berat	0.85
	Sedang	0.4
	Tidak ada	0
G12	Bradikardi relatif	
	Ya	0.75
	Tidak ada	0
G13	Pembesaran hati	
	Ya	0.5
	Tidak ada	0
G14	Pembesaran limpa	
	Ya	0.65
	Tidak ada	0
G15	Kulit lembab/ Keringat	
	Berat	0.85
	Sedang	0.4
	Tidak ada	0

Proses pengklasifikasian diagnosis penyakit menggunakan metode *knn – certainty factor* digambarkan dalam diagram alir seperti pada Gambar 4.1.



Gambar 4.1 Tahapan Metode KNN-Certainty Factor

Berikut penjelasan alur proses diatas:

1. Tenaga medis menginputkan data pasien dan keluhan yang dialami
Misalkan terdapat 10 data latih yang ditunjukkan pada Tabel 3.1 dan satu data uji yang ditunjukkan pada Tabel 3.2

Tabel 3.1 Data latih penyakit Demam

No Urut Pasien	Gejala															Penyakit	
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15		
10	0.8	0.85	0	0.5	0.7	0	0	0.2	0	0	0	0	0	0	0.4	MALARIA	
14	0	0.85	0	0.2	0.9	0	0	0.2	0	0	0	0	0	0	0.4	MALARIA	
18	0	0.85	0	0.5	0.9	0	0	0.2	0	0	0	0	0	0	0.4	MALARIA	
17	0	0	0	0.75	0.7	0.75	0.85	0.6	0	0	0.4	0.75	0.5	0.65	0	TYPHOID	
110	0	0	0	0.75	0.75	0.9	0	0	0.6	0.4	0.7	0.4	0.75	0	0.65	0	TYPHOID
84	0	0	0	0.75	0.9	0	0.85	0.6	0	0	0	0.75	0	0	0	0	TYPHOID
20	0	0.5	0.75	0.5	0.7	0.75	0.6	0.2	0.8	0.7	0.4	0	0.5	0.65	0	DBD	
38	0	0	0	0.2	0.9	0.75	0.6	0.2	0.8	0.7	0	0	0.5	0	0	DBD	
66	0	0	0	0.2	0.9	0.9	0.85	0.2	0	0.7	0.4	0	0	0	0	DBD	
109	0	0	0	0.2	0.7	0.9	0.85	0.2	0.4	0.7	0	0	0	0	0	DBD	

Tabel 3.2 Data uji penyakit Demam (Tifoid)

No Urut Pasien	Gejala															Penyakit
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	
25	0	0	0.75	0.5	0.9	0	0.85	0.6	0.4	0	0.85	0.75	0.5	0.65	0	TYPHOID

2. Melakukan normalisasi nilai data latih dan data uji.

Nilai min dan max dapat dilihat pada Tabel 3.3, dan untuk data latih ternormalisasi ditunjukkan pada Tabel 3.4, sedangkan untuk data uji ternormalisasi ditunjukkan pada Tabel 3.5

Tabel 3.3 Min-max normalization

Min	0	0	0	0.2	0.7	0	0	0.2	0	0	0	0	0	0	0	0
Max	0.8	0.85	0.75	0.75	0.9	0.9	0.85	0.6	0.8	1	0.85	0.75	0.5	0.65	0.85	

Tabel 3.4 Data latih ternormalisasi

No Urut Pasien	Gejala															Penyakit
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	
10	1	1	0	0.54	0	0	0	0	0	0	0	0	0	0	0.47	MALARIA
14	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0.47	MALARIA
18	0	1	0	0.54	1	0	0	0	0	0	0	0	0	0	0.47	MALARIA
17	0	0	0	1	0	0.83	1	1	0	0	0.47	1	1	1	0	TYPHOID
110	0	0	1	1	1	0	0	1	0.5	0.7	0.47	1	0	1	0	TYPHOID
84	0	0	0	1	1	0	1	1	0	0	0	1	0	0	0	TYPHOID
20	0	0.58	1	0.54	0	0.83	0.70	0	1	0.7	0.47	0	1	1	0	DBD
38	0	0	0	0	1	0.83	0.70	0	1	0.7	0	0	0	1	0	DBD
66	0	0	0	0	1	1	1	0	0	0.7	0.47	0	0	0	0	DBD

Tabel 3.5 Data uji ternormalisasi

No Urut Pasien	Gejala															Penyakit
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	
25	0	0	1	0.54	1	0	1	1	0.5	0	1	1	1	1	0	TYPHOID

3. Melakukan perhitungan klasifikasi *k-nearest neighbor* dengan mencari nilai jarak antara data uji terhadap data latih. Hasil perhitungan Euclidean yaitu jarak antara data uji terhadap data latih ditunjukkan pada Tabel 3.6

Tabel 3.6 Hasil perhitungan Euclidean

No Urut Pasien	Jarak	Penyakit
10	3.2359625	MALARIA
14	2.9612453	MALARIA
18	2.9105761	MALARIA
17	1.8523857	TYPHOID
110	1.7253662	TYPHOID
84	2.1110688	TYPHOID
20	2.2687546	DBD
38	2.6112201	DBD
66	2.7051428	DBD

Setelah didapatkan jarak data uji terhadap data latih, kemudian urutkan hasil

tersebut dari nilai terkecil ke terbesar, ditunjukkan pada tabel 3.7

Tabel 3.7 Hasil Euclidean dari yang terkecil

No Urut Pasien	Jarak	Penyakit
110	1.7253662	TYPHOID
17	1.8523857	TYPHOID
84	2.1110688	TYPHOID
20	2.2687546	THYPOID
38	2.6112201	DBD
66	2.7051428	DBD
18	2.9105761	MALARIA
14	2.9612453	MALARIA
10	3.2359625	MALARIA

Kemudian menentukan nilai *k* yang akan digunakan. Misal *k*=3, hasil ditunjukkan pada Tabel 3.8

Tabel 3.8 Hasil Euclidean dari yang terkecil berdasarkan *k*=3

No Urut Pasien	Jarak	Penyakit
110	1.7253662	TYPHOID
17	1.8523857	TYPHOID
84	2.1110688	TYPHOID

Setelah diketahui hasil klasifikasi dari perhitungan knn adalah penyakit tifoid, maka selanjutnya menghitung nilai kepastian dengan menggunakan *certainty factor*. Untuk mencari nilai CF adalah dengan mengkalikan data gejala pasien dengan data gejala pakar. Hasil perhitungan ini ditunjukkan pada Tabel 3.9.

Tabel 3.9 Hasil perhitungan nilai CF

CF	Tifoid
CF 1	0
CF 2	0
CF 3	0.8
CF 4	0.436364
CF 5	0.9
CF 6	0
CF 7	0.9
CF 8	0.7
CF 9	0.25
CF 10	0
CF 11	0.9
CF 12	0.8
CF 13	0.5
CF 14	0.6
CF 15	0

Setelah didapatkan nilai CF, kemudian menghitung *CF Combine* untuk mendapatkan nilai kepastian penyakit tersebut. Hasil

perhitungan *CF Combine* ditunjukkan pada Tabel 3.13.

Tabel 3.13 Hasil perhitungan *CF Combine*

CF Combine	Tifoid
Combine 1	0
Combine 2	0
Combine 3	0.8872727
Combine 4	0.9364628
Combine 5	0.9936463
Combine 6	0
Combine 7	0.9993646
Combine 8	0.9998094
Combine 9	0.999857
Combine 10	0
Combine 11	0.9999857
Combine 12	0.9999971
Combine 13	0.9999986
Combine 14	0.9999994
Combine 15	0
CF Combine	0.9999994

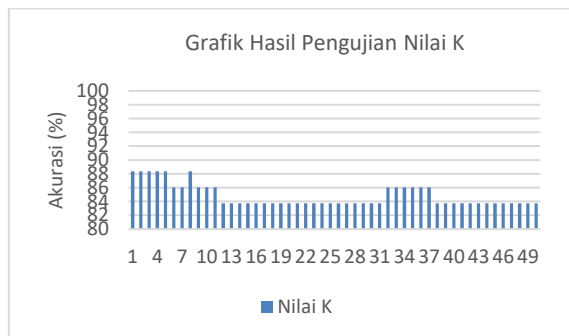
Berdasarkan perhitungan dan tabel diatas maka penyakit tifoid memiliki nilai kepastian sebesar 0.9999994.

5. PENGUJIAN DAN ANALISIS

5.1 Pengujian Variasi Nilai K

Pada pengujian variasi nilai k bertujuan untuk mengetahui apakah nilai k berpengaruh terhadap tingkat akurasi.

Pada analisis variasi nilai k digunakan 100 data latih, 43 data uji dan nilai k mulai dari k=1 sampai k=50. Hasil pengujian pengaruh nilai k ditunjukkan pada Gambar 5.1.



Gambar 5.1 Grafik Hasil Pengujian Variasi Nilai K

Pada grafik Gambar 5.1 diatas dapat diketahui bahwa dari nilai k=1 sampai k=5 menghasilkan akurasi sebesar 88.37%. Setelah itu akurasi mengalami naik turun. Pada nilai k=12 hingga k=31 menghasilkan akurasi 83.72%, pada nilai k=32 hingga k=37 menghasilkan akurasi 86.05%, pada nilai k=38 hingga k=50 menghasilkan akurasi 83.72%.

Pada percobaan tersebut nilai k yang semakin bertambah menghasilkan akurasi yang

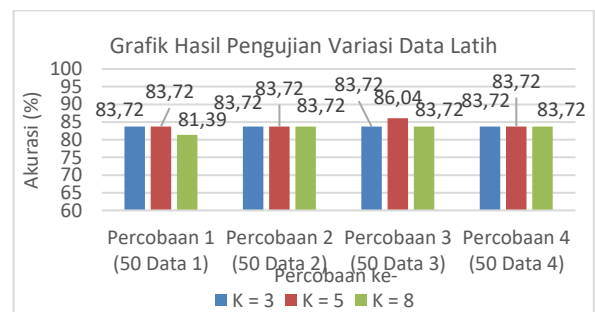
semakin rendah hingga mengalami naik turun akurasi, namun naik turun akurasi tidak begitu besar dan cenderung stabil. Nilai k yang bertambah akan membuat semakin banyak data tidak relevan terhadap data uji dan hal tersebut mengakibatkan tingkat kesalahan prediksi yang semakin besar dan menyebabkan tingkat akurasi turun.

Dari percobaan diatas dapat disimpulkan bahwa variasi nilai k berpengaruh pada akurasi. Untuk nilai k yang akan digunakan pada pengujian-pengujian adalah nilai k yang memiliki akurasi tertinggi yaitu nilai k=3, k=5 dan k=8 dengan akurasi sebesar 88.37%.

5.2 Pengujian Variasi Data Latih

Pengujian variasi data latih bertujuan untuk mengetahui apakah data latih yang berbeda-beda akan berpengaruh terhadap tingkat akurasi sistem. Pada pengujian ini jumlah data latih yang digunakan adalah sama namun dengan data yang berbeda.

Hasil pengujian pengaruh jumlah data latih ditampilkan dalam sebuah grafik pada Gambar 5.2.



Gambar 5.2 Grafik Hasil Pengujian Variasi Data Latih

Pada grafik Gambar 5.2 diatas dapat diketahui bahwa percobaan yang dilakukan dengan nilai k=3, k=5 dan k=8 pada percobaan ke-2 dan ke-4 memiliki kesamaan hasil akurasi yaitu 83.72%. Pada percobaan ke-3 dengan nilai k=5 memiliki hasil akurasi yang tertinggi di banding nilai k yang lain. Hal ini disebabkan karena meskipun data latih yang digunakan berjumlah sama, akan tetapi variasi data latihnya berbeda. Dimana pada percobaan ke-3 memiliki variasi data dengan jarak yang lebih berdekatan dengan data uji dibandingkan dengan percobaan ke-1, ke-2 dan ke-4 sehingga menghasilkan tingkat akurasi yang lebih tinggi.

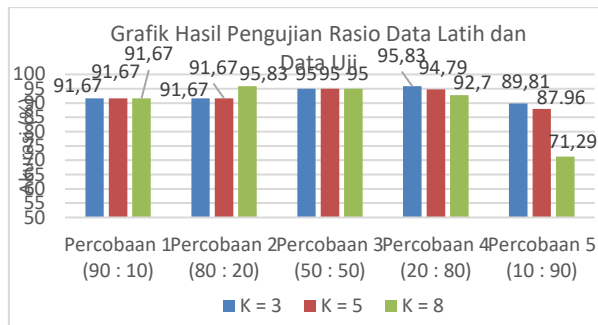
Dari hal tersebut dapat disimpulkan bahwa percobaan yang dilakukan dengan nilai k=3, k=5 dan k=8 memiliki akurasi yang naik turun

walaupun sebagian besar sama. Perbedaan variasi pada setiap percobaan membuat tingkat akurasi berbeda-beda dan mempengaruhi nilai jarak antara data latih dan data uji. Jika dalam suatu variasi data latih memiliki nilai jarak yang berdekatan dengan data uji akan membuat akurasi semakin tinggi dan begitu sebaliknya.

5.3 Pengujian Rasio Data Latih dan Data Uji

Pengujian rasio data latih dan data uji bertujuan untuk mengetahui apakah rasio data latih dan data uji yang berbeda-beda akan berpengaruh terhadap tingkat akurasi sistem.

Pengujian rasio data latih dan data uji ini diambil dari persentase perbandingan data latih dan data uji, yaitu 90:10, 80:20, 50:50, 20:80 dan 10:90. Berikut hasil grafik pengujian rasio data latih dan data uji ditunjukkan pada Gambar 5.3.



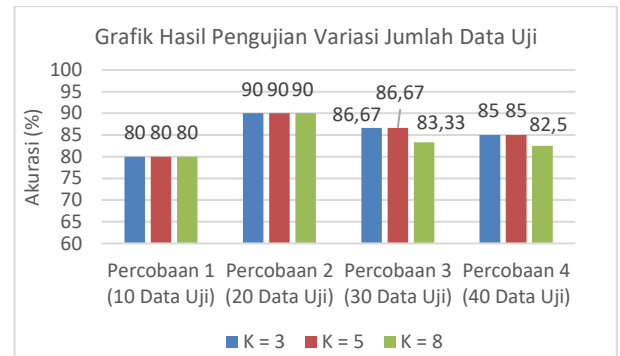
Gambar 5.3 Grafik Hasil Pengujian Rasio Data Latih dan Data Uji

Pada grafik Gambar 5.3 diatas dapat diketahui bahwa pada nilai k=8 percobaan ke-2 dan nilai k=3 pada percobaan ke-4 memiliki hasil akurasi tertinggi yaitu 95.83%, ini dikarenakan jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji pada setiap percobaan berpengaruh pada tingkat akurasi sistem. Seperti pada percobaan ke-2 jumlah kesalahan yang terjadi adalah 2 dan ke-3 jumlah kesalahan yang terjadi adalah 4, tetapi data uji yang digunakan pada percobaan ke-2 adalah 24 data dan pada percobaan ke-3 adalah 60 data, karena jumlah data uji pada percobaan ke-3 lebih banyak maka tingkat akurasi pada percobaan ke-3 lebih besar meskipun memiliki jumlah kesalahan lebih besar daripada percobaan ke-2. Jadi jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji berpengaruh besar pada tingkat akurasi.

5.4 Pengujian Variasi Jumlah Data Uji

Pengujian variasi jumlah data uji ini bertujuan untuk mengetahui apakah jumlah data uji yang berbeda-beda dapat mempengaruhi

tingkat akurasi sistem. Pada analisis variasi jumlah data uji ini digunakan nilai k dengan tingkat akurasi tertinggi yang didapatkan pada percobaan pertama yaitu k=3, k=5 dan k=8 dimana digunakan data uji dengan jumlah 10, 20, 30 dan 40 data dan data latih yang digunakan adalah data latih yang sama yaitu 100 data. Hasil pengujian variasi jumlah data uji ditampilkan dalam sebuah grafik pada Gambar 5.4.



Gambar 5.4 Grafik Hasil Pengujian Variasi Jumlah Data Uji

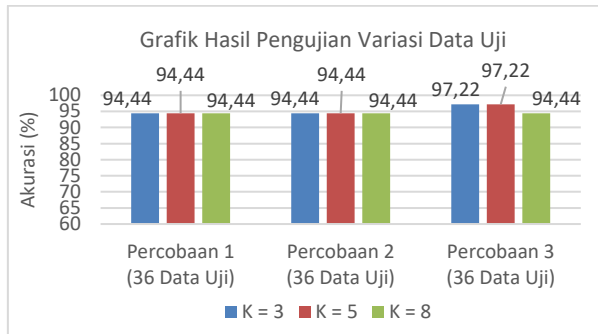
Pada grafik Gambar 5.4 diatas dapat diketahui bahwa percobaan yang dilakukan pada percobaan ke-1 dengan nilai k=3, k=5 dan k=8 memiliki nilai akurasi 80%. Sedangkan pada percobaan ke-2 dengan nilai k=3, k=5 dan k=8 memiliki nilai akurasi sebesar 90% yang merupakan nilai akurasi tertinggi. Pada percobaan ke-3 dan ke-4 tingkat akurasi mengalami naik turun tetapi tidak terlalu besar dan cenderung stabil.

Dari grafik diatas dapat disimpulkan bahwa yang dapat mempengaruhi tingkat akurasi adalah jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji pada setiap percobaan. Seperti pada percobaan ke-1 dan ke-2 jumlah kesalahan yang terjadi adalah 2, tetapi data uji yang digunakan pada percobaan ke-1 adalah 10 data dan pada percobaan ke-2 adalah 20 data, karena jumlah data uji pada percobaan ke-2 lebih banyak maka tingkat akurasi pada percobaan ke-2 lebih besar. Jadi dapat disimpulkan bahwa pada pengujian variasi jumlah data uji diengaruhi oleh 2 hal yaitu jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji.

5.5 Pengujian Variasi Data Uji

Pengujian variasi data uji ini bertujuan untuk mengetahui apakah variasi data uji yang berbeda-beda dapat mempengaruhi tingkat akurasi sistem. Pada analisis variasi data uji ini digunakan nilai k dan data latih dengan tingkat

akurasi tertinggi yang didapatkan pada percobaan pertama yaitu k=3, k=5 dan k=8 dimana digunakan variasi data uji berjumlah sama yaitu 36 data pada 3 kali percobaan dan data latih sebanyak 100 data latih. Hasil pengujian variasi data uji ditampilkan dalam sebuah grafik pada Gambar 5.5.

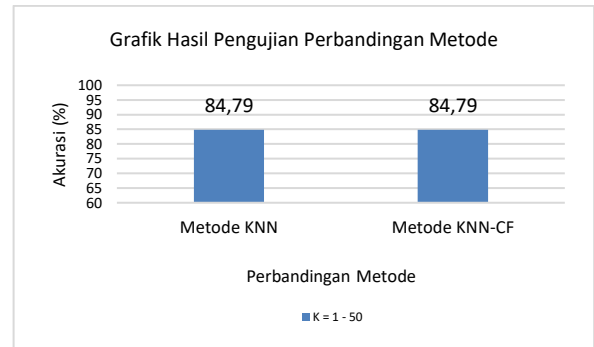


Gambar 5.5 Grafik Hasil Pengujian Variasi Data Uji

Pada grafik Gambar 5.5 diatas dapat diketahui bahwa percobaan yang dilakukan dengan nilai k=3, k=5 dan k=8 percobaan ke-1 dan ke-2 menghasilkan akurasi 94.44%. Akurasi tertinggi dihasilkan oleh percobaan ke-3 dengan nilai k=3 dan k=5 sebesar 97.22, hal ini disebabkan karena meskipun variasi data uji yang digunakan berjumlah sama, akan tetapi variasi data ujinya berbeda. Dimana pada percobaan ke-3 memiliki variasi data dengan jarak yang lebih berdekatan dengan data latih dibandingkan dengan percobaan lainnya sehingga menghasilkan tingkat akurasi yang lebih tinggi.

5.6 Pengujian Perbandingan Metode

Pada analisis perbandingan metode ini digunakan nilai akurasi dari dua metode yaitu metode *K-Nearest Neighbor – Certainty Factor* dan metode *K-Nearest Neighbor* dengan acuan perbandingan rata-rata akurasi yang dihasilkan untuk setiap k. Pengujian ini menggunakan k=1 hingga k=5, 100 data latih dan jumlah data uji yang tetap yaitu sebanyak 43 data. Hasil pengujian perbandingan metode ditampilkan dalam sebuah grafik pada Gambar 5.6.

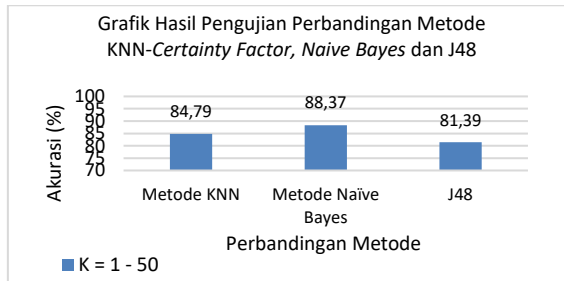


Gambar 5.6 Grafik hasil pengujian perbandingan metode

Dari grafik diatas dapat disimpulkan metode *K-Nearest Neighbor – Certainty Factor* memiliki tingkat akurasi yang sama dengan metode *K-Nearest Neighbor* yaitu sebesar 84.79%. Akurasi adalah menghitung hasil ketepatan diagnosis system dengan pakar dan hasil akurasi metode *K-Nearest Neighbor – Certainty Factor* mengikuti hasil klasifikasi *K-Nearest Neighbor*. Jadi dapat disimpulkan antara metode *K-Nearest Neighbor – Certainty Factor* dan *K-Nearest Neighbor* yang berperan dalam penentuan hasil akurasi adalah metode *K-Nearest Neighbor*, sedangkan metode *Certainty Factor* hanya menghitung nilai kepastian suatu penyakit berdasarkan hasil diagnosis yang dilakukan oleh metode *K-Nearest Neighbor*, apabila metode *K-Nearest Neighbor* tidak menghasilkan output maka metode *Certainty Factor* tidak dapat melakukan proses perhitungan nilai kepastian.

Untuk menguji kinerja metode *K-Nearest Neighbor – Certainty Factor* pada kasus Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid maka dilakukan pengujian perbandingan dengan 2 metode klasifikasi lainnya yaitu *naïve bayes* dan *J48* menggunakan Aplikasi Weka. *Naïve Bayes* adalah metode Bayesian learning yang paling cepat dan sederhana. Metode *Naive Bayes* merepresentasikan data gejala kedalam bentuk probability model. Probability model tersebut yang digunakan sebagai model untuk mengelompokkan suatu penyakit. Sedangkan metode *K-Nearest Neighbor* menyimpan sekelompok data training set sebagai acuan untuk mengelompokkan suatu penyakit. Metode *J48* merupakan salah satu kelas di paket classifier pada sistem weka yang mengimplementasikan algoritma *C45*. Algoritma *C4.5* adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Dalam pengujian ini data

latih yang digunakan berjumlah 100 data dan 43 data uji. Hasil pengujian perbandingan metode KNN-Certainty Factor, naïve bayes dan J48 ditampilkan dalam sebuah grafik pada Gambar 5.7.



Gambar 5.7 Grafik hasil pengujian perbandingan metode KNN-CF, Naive Bayes dan J48

Dari grafik diatas dapat disimpulkan bahwa metode *K-Nearest Neighbor – Certainty Factor* lebih baik daripada metode J48 yang menghasilkan akurasi 81.39%. Hal ini dikarenakan metode J48 dalam membuat cabang node tree hanya terbatas sehingga informasi yang didapatkan kurang optimal dan membuat akurasi menjadi turun. Tetapi metode *K-Nearest Neighbor – Certainty Factor* memiliki akurasi kurang dari *Naive Bayes* yang menghasilkan akurasi sebesar 88.37%. Hal ini disebabkan karena metode naïve bayes sudah mampu menentukan parameter yang diperlukan dalam proses pengklasifikasian hanya dengan jumlah data latih yang kecil atau sedikit sedangkan pada metode *K-Nearest Neighbor – Certainty Factor* diperlukan data latih yang besar sehingga hasil lebih efektif dan tangguh terhadap training data yang noisy.

6. PENUTUP

6.1 Kesimpulan

Berdasarkan hasil penelitian Sistem Pakar Diagnosis Penyakit Demam Menggunakan Metode *K-Nearest Neighbor – Certainty Factor*, dapat diambil kesimpulan sebagai berikut:

1. Sistem Pakar Diagnosis Penyakit Demam Menggunakan Metode *K-Nearest Neighbor – Certainty Factor* ini dibuat menggunakan Bahasa pemrograman PHP dan basis data MySQL sebagai tempat penyimpanan data masukan dan data *knowledge*. Sistem pakar ini mampu mengidentifikasi penyakit Demam dengan menginputkan 15 gejala dan kemudian dilakukan proses perhitungan menggunakan metode *K-Nearest Neighbor – Certainty*

Factor untuk mendapatkan hasil diagnosis berupa jenis penyakit sesuai dengan nilai kepastian terbesar.

2. Metode *K-Nearest Neighbor – Certainty Factor* menghasilkan tingkat akurasi yang tergolong baik dalam mengklasifikasikan dan memberi nilai kepastian jenis penyakit DBD, Malaria atau Tifoid dimana pada beberapa mekanisme pengujian menghasilkan tingkat akurasi yaitu:

- a. Didapatkan nilai akurasi maksimum sebesar 88.37% pada pengujian variasi nilai k, dengan $k=3$, $k=5$ dan $k=8$. Dimana nilai akurasi maksimum diperoleh dari nilai K yang kecil. Semakin banyak nilai K maka akan membuat semakin banyak data tidak relevan terhadap data uji dan hal tersebut mengakibatkan tingkat kesalahan prediksi yang semakin besar dan menyebabkan tingkat akurasi turun.
- b. Didapatkan nilai akurasi maksimum sebesar 86.04% pada pengujian variasi data latih, dimana perbedaan variasi data latih memberikan perbedaan tingkat akurasi. Hal ini disebabkan karena perbedaan tersebut akan mempengaruhi nilai jarak antara data latih dan data uji. Jika dalam suatu variasi memiliki nilai jarak yang berdekatan dengan data uji akan membuat akurasi semakin tinggi, begitupun sebaliknya.
- c. Didapatkan hasil akurasi maksimum sebesar 95.83% pada pengujian rasio data latih dan data uji pada percobaan 2 dengan nilai $k=8$ dan pada percobaan ke-4 dengan nilai $k=2$. Hal ini dikarenakan jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji pada setiap percobaan berpengaruh pada tingkat akurasi sistem. Jumlah data uji pada percobaan ke-3 lebih banyak maka tingkat akurasi pada percobaan ke-3 lebih besar meskipun memiliki jumlah kesalahan lebih besar daripada percobaan ke-2.
- d. Didapatkan hasil akurasi maksimum sebesar 90% pada pengujian variasi jumlah data uji dengan jumlah 20 data. Hal ini disebabkan karena jumlah kesalahan pada setiap percobaan dan banyaknya jumlah data uji pada setiap percobaan mempengaruhi tingkat akurasi yang dihasilkan.

- e. Didapatkan hasil akurasi maksimum sebesar 97.22% pada pengujian variasi data uji dimana perbedaan variasi data uji memberikan perbedaan tingkat akurasi. Hal ini disebabkan karena perbedaan variasi data uji akan mempengaruhi nilai jarak antara data latih dan data uji. Jika dalam suatu variasi memiliki nilai jarak yang berdekatan dengan data latih akan menghasilkan nilai akurasi semakin tinggi, dan sebaliknya.
 - f. Didapatkan hasil akurasi maksimum sebesar 84.79% pengujian perbandingan metode. Perbandingan nilai akurasi antara metode *K-Nearest Neighbor* – *Certainty Factor* dengan metode *K-Nearest Neighbor* menghasilkan akurasi yang sama, karena metode *Certainty Factor* hanya menghitung nilai kepastian suatu penyakit berdasarkan hasil diagnosis dari metode *K-Nearest Neighbor*.
3. Perbandingan nilai akurasi metode antara *K-Nearest Neighbor* – *Certainty Factor* dengan *K-Nearest Neighbor* menghasilkan akurasi yang sama yaitu 84.79%. Untuk perbandingan dengan metode yang lainnya, metode *K-Nearest Neighbor* – *Certainty Factor* lebih baik daripada metode J48 yang menghasilkan akurasi 81.39%. Tetapi metode *K-Nearest Neighbor* – *Certainty Factor* memiliki akurasi kurang dari *Naïve Bayes* yang menghasilkan akurasi sebesar 88.37%.

6.2 Saran

Beberapa saran untuk pengembangan lebih lanjut yang dapat diberikan oleh penulis antara lain:

1. Sistem ini hanya untuk mendiagnosis 3 jenis penyakit demam (DBD, Malaria dan Tifoid) sedangkan masih terdapat banyak jenis demam sehingga dapat ditambahkan jenis-jenis demam yang lain.
2. Pada penelitian selanjutnya perlu dilakukan penambahan data latih dan data uji, solusi dan cara pencegahan penyakit Demam sehingga informasi dan data pembelajaran sistem semakin banyak.
3. Sebaiknya pada penelitian selanjutnya metode sistem pakar yang dapat digunakan adalah dengan menggunakan metode *Fuzzy* – *K-Nearest Neighbor* atau *Modified* – *K-*

Nearest Neighbor dan penggabungan metode lainnya dengan *K-Nearest Neighbor*.

DAFTAR PUSTAKA

- Abdurrahman, Ginanjar. 2011. *Thesis Project: Penerapan Metode Tsukamoto (Logika Fuzzy) Dalam Sistem Pendukung Keputusan Untuk Menentukan Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Permintaan*. Yogyakarta: Universitas Negeri Yogyakarta.
- Anjas Sari, Nur. 2013. *Sistem Pakar Mendiagnosis Penyakit Demam Berdarah Menggunakan Metode Certainty Factor*. Medan: STMIK Budidarma Medan.
- Djunaedi, D. 2006. *Demam Berdarah Dengue (Dengue DBD) Epidemiologi, Immunopatologi, Patogenesis, Diagnosis dan Penatalaksanaannya*. Malang: UMM Press.
- Hidayati, N. 2001. *Thesis Project: Analisis Regresi Poisson Terhadap Faktor-faktor yang Mempengaruhi Penyakit Demam Typhoid di Provinsi Jawa Timur*. Surabaya: FMIPA ITS.
- Hidayat, Syaeful. 2010. *Aplikasi Untuk Mendeteksi Penyakit Pada Tanaman Tebu dan Cara Penanganannya Berbasis Web*. Bandung: Universitas Komputer Indonesia.
- Latumakulita, Luther. 2012. *Sistem Pakar Pendiagnosis Penyakit Anak Menggunakan Certainty Factor (CF)*. Manado: Jurnal Ilmiah Sains Vol. 12 No. 2.
- Paula, Yulianti. 2015. *Pengembangan Sistem Pakar Diagnosis Penyakit Tuberculosis dan Demam Berdarah Berbasis Web Menggunakan Metode Certainty Factor*. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi.