

DIAGNOSA PENYAKIT PARKINSON DENGAN ALGORITMA *K-NEAREST NEIGHTBOR* DAN *DECISION TREE C4.5*

DIAGNOSIS OF PARKINSON'S DISEASE USING K-NEAREST NEIGHTBOUR AND DECISION TREE C4.5 ALGORITHMS

¹Anita Desiani*, ²Narti, ³Irmeilyana, ⁴Indri Ramayanti, ⁵Muhammad Arhami

^{1,2,3}Prodi Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya, Indralaya, Indonesia

⁴Prodi Departemen Parasitologi, Fakultas Kedokteran, Universitas Muhammadiyah Palembang, Palembang, Indonesia

⁵Program Studi Teknik Informatika, Politeknik Negeri Lhokseumawe, Aceh, Indonesia
*e-mail: anita_desiani@unsri.ac.id

Abstrak

Parkinson adalah suatu penyakit dimana neurologis mempengaruhi neuron dopaminergik, yang dibuktikan dengan kematian sel-sel otak yang ada, hilangnya pigmentasi substantia nigra, adanya inklusi sitoplasma, dan penurunan kadar dopamin di substantia nigra pars compacta dan corpus striatum. Penyakit parkinson dapat didiagnosa dengan melakukan pengklasifikasian untuk mengukur tingkat akurasi. Tujuan dari penelitian ini adalah untuk melakukan diagnosa penyakit Parkinson dengan dua algoritma yang berbeda, yaitu algoritma *K-Nearest Neighbor* (KNN) dan algoritma C4.5 dengan metode pelatihan *Percentage split* dan validasi *K-fold cross* yang nantinya kan dibandingkan satu sama lain. Dari penelitian ini, nilai presisi yang dimiliki penderita Parkinson's disease algoritma C4.5 split persentasenya adalah 96%. Begitu juga untuk nilai recall yang dimiliki oleh penderita penyakit Parkinson yaitu sebesar 93%. Nilai akurasi algoritma *K-Nearest Neighbor* (KNN) adalah 82% untuk metode pelatihan pada *percentage split* dan 76,8% dengan metode validasi *K-fold cross* dan 89% untuk algoritma C4.5 dengan metode pelatihan pada *Percentage split* dan 81% dengan metode validasi *K-fold cross*.

Kata kunci: *C4.5, K- Fold Cross Validation, K-Nearest Neighbor, Parkinson, Percentage Split*

Abstract

Parkinson's is a disease in which neurological damage affects dopaminergic neurons, as evidenced by the death of existing brain cells, loss of substantia nigra pigmentation, presence of cytoplasmic inclusions, and decreased levels of dopamine in the substantia nigra pars compacta and corpus striatum. Parkinson's disease can be diagnosed by classifying it to measure the level of accuracy. The purpose of this research is to diagnose Parkinson's disease with two different algorithms, namely the K-Nearest Neighbor (KNN) algorithm and the C4.5 algorithm with the split percentage training method and K-fold cross validation which will later be compared with each other. From this study, the percentage of precision that is owned by patients with Parkinson's disease algorithm C4.5 split is 96%. Likewise for the recall value that is owned by people with Parkinson's disease, which is equal to 93%. The accuracy value of the K-Nearest Neighbor (KNN) algorithm is 82% for the training method on the split percentage and 76.8% with the K-fold cross

validation method and 89% for the C4.5 algorithm with the training method on the split percentage and 81% with the K-Nearest Neighbor (KNN) algorithm. K-fold cross validation.

Keywords: C4.5, K- Fold Cross Validation, K-Nearest Neighbor, Parkinson, Percentage Split

1 PENDAHULUAN

Parkinson adalah suatu penyakit pada sistem saraf pusat yang mengenai neuron dopaminergic [1]. Penyakit parkinson ditandai oleh degenerasi sel-sel saraf dalam otak, hilangnya pigmentasi di substansia nigra, adanya inklusi sitoplasmik, serta penurunan dopamin di substansia nigra pars kompakta dan korpus striatum [2]. Penyakit parkinson biasanya muncul pada usia 40-70, rata-rata diatas usia 60 tahun yang didominasi oleh laki-laki [3]. Kriteria diagnosis klinis penyakit parkinson dapat ditegakkan jika terdapat dua dari tiga tanda kardinal komplikasi motorik yaitu rigiditas, tremor, dan bradykinesia [4]. Penyakit parkinson dapat didiagnosa dengan melakukan pengklasifikasian untuk mengukur tingkat akurasi. Untuk mengukur tingkat akurasi pada dataset dapat dilakukan dengan teknik klasifikasi data *mining* [5]. Data mining merupakan proses untuk menemukan hubungan maupun pola baru yang sebelumnya belum diketahui untuk membangun sebuah model prediksi [6]. Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari jumlah kelas yang tersedia [5]. Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) ke satu jumlah label kelas yang tersedia [5]. Pada penelitian ini akan melakukan deteksi penyakit parkinson dengan menggunakan algoritma klasifikasi data mining *K-Nearest Neighbor* (KNN) dan C4.5. Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut [7]. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi yang bersifat prediktif digunakan untuk mengklasifikasikan data yang memiliki atribut berupa numerik ataupun kategorial [6]. Algoritma C4.5 disebut juga sebagai algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya [8].

Algoritma KNN menjadi algoritma dengan pelatihan yang super cepat, sederhana, dan mudah dipelajari [9]. Algoritma KNN telah banyak digunakan pada banyak masalah klasifikasi sehingga menjadi salah satu algoritma yang banyak diteliti [10]. Kekurangan dari algoritma KNN terletak pada nilai k, komputasi kompleks, keterbatasan memori, dan mudah tertipu dengan atribut yang tidak relevan [9]. Kelebihan algoritma C4.5 dapat menangani atribut *continue* dan diskrit beserta pengolahan *training* data dengan *missing value* untuk memperbaiki informasi pada gain menggunakan gain ratio [11]. Selain itu, algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dan dapat menangani atribut bertipe diskret dan numeric [8]. Dalam mengkonstruksi pohon, algoritma C4.5 membaca seluruh sampel data *training* dari storage dan memuatnya ke memori [12]. Hal ini yang menjadi kelemahan dari algoritma C4.5 dalam kategori skalabilitas yang berarti algoritma ini hanya dapat digunakan jika data *training*nya dapat disimpan secara keseluruhan dan pada waktu yang bersamaan pada memori [12].

Penelitian mengenai diagnosis penyakit parkinson telah dilakukan berdasarkan kombinasi algoritma data mining dan seleksi fitur dengan menerapkan algoritma J48 yang menghasilkan nilai akurasi terbaik yaitu sebesar 96.923% dengan ROC 0.982 dan running time 0.10 detik [13]. Penelitian lainnya yang telah dilakukan untuk diagnosis penyakit parkinson adalah algoritma klasifikasi dengan performa terendah dimiliki oleh algoritma Naive Bayes dengan nilai akurasi sebesar 74% [14]. Pada penelitian ini akan melakukan

deteksi penyakit Parkinson dengan dua algoritma yang berbeda, yaitu algoritma klasifikasi data mining *K-Nearest Neighbor* (KNN) dan C4.5.

2 TINJAUAN PUSTAKA

Tinjauan penyakit Parkinson salah satu penyakit degenerasi otak yang paling banyak diderita setelah Alzheimer. Penelitian ini menerapkan algoritma klasifikasi data mining *K-Nearest Neighbor* dan Naïve Bayes. Penelitian ini menghasilkan nilai akurasi mencapai 96% untuk algoritma *K-Nearest Neighbor* dan 74% untuk algoritma Naïve Bayes, yang menunjukkan tingginya tingkat keberhasilan yang diraih dalam melakukan diagnosa terhadap penyakit Parkinson [14].

Penelitian ini dilakukan sebagai salah satu solusi untuk melakukan diagnosa terhadap penyakit. Algoritma *K-Nearest Neighbor* diterapkan untuk mempermudah proses diagnosa. Sementara Naïve Bayes yang memiliki rekam jejak dengan hasil yang lebih akurat untuk data dengan jumlah yang banyak.

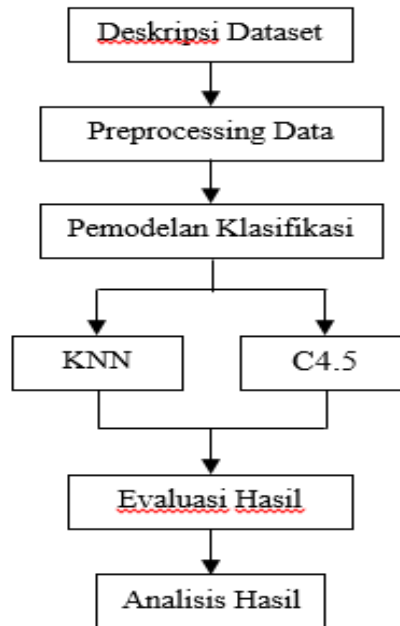
Penelitian ini fokus pada penyakit Parkinson. Penerapan algoritma *K-Nearest Neighbor* memperoleh hasil yang menunjukkan bahwa algoritma ini dapat diterapkan dengan baik. Namun, berbeda dengan *K-Nearest Neighbor*, Naïve Bayes memperoleh hasil yang masih diperlukan evaluasi kembali agar dapat melakukan diagnosa yang lebih akurat.

Berdasarkan review terhadap beberapa penelitian yang telah dilakukan tidak terdapat penelitian yang menerapkan algoritma C4.5 untuk diagnosa penyakit Parkinson. Algoritma C4.5 dimungkinkan dapat membantu hasil diagnosa agar lebih akurat. Penerapan algoritma *K-Nearest Neighbor* dan C4.5 ini memberikan uji banding yang penting dalam menentukan langkah yang dapat dipilih untuk melakukan diagnosa terhadap penyakit Parkinson.

Dengan demikian, dengan melihat beberapa penelitian yang telah dilakukan oleh para peneliti terdahulu, dapat ditarik beberapa kesimpulan terkait dengan keberhasilan evaluasi kinerja algoritma klasifikasi data mining *K-Nearest Neighbor* dan C4.5. Penerapan algoritma *K-Nearest Neighbor* dan C4.5 ini menjadi salah satu factor keberhasilan dalam melakukan diagnosa penyakit Parkinson. Dengan demikian, beberapa penelitian yang telah dilakukan baik dengan nilai akurasi yang lebih rendah maupun jenis algoritma yang berbeda menunjukkan bahwa evaluasi serta perbaikan penelitian perlu dilakukan secara berkala untuk meningkatkan kinerja algoritma yang dapat digunakan.

3 METODE PENELITIAN

Penelitian untuk mendiagnosa penyakit Parkinson ini dilakukan dengan menggunakan metode *K-Nearest Neighbor* (KNN) dan C4.5 yang dilakukan dengan beberapa langkah yang telah ditentukan, yaitu :



Gambar 1. Tahapan Penelitian

Gambar 1. Menggambarkan tahapan penelitian yang dilakukan dengan deskripsi dataset, *Preprocessing* data, pemodelan klasifikasi, evaluasi hasil, dan analisis hasil. Dataset yang diuji pada penelitian ini diambil dari situs *Kaggle* ([https://www.kaggle.com/datasets /debasisdotcom/parkinson-diseasedetection](https://www.kaggle.com/datasets/debasisdotcom/parkinson-diseasedetection)), yang berjudul *Parkinson Disease Detection* yang dapat dilihat pada Tabel 1.

Tabel 1. Dataset sebagai Data Input Algoritma

No	Atribut	Keterangan Atribut
1	name	Nama
2	MDVP:Fo(Hz)	88.33-260.11
3	MDVP:Fhi(Hz)	102.14-592.03
4	Ketebalan Substrat (h)	1.6 mm
5	Konduktifitas Bahan	5.8×10^7 S/m
6	MDVP:Jitter(Abs)	0
7	MDVP:RAP	0-0.02
8	MDVP:PPQ	0-0.02
9	Jitter:DDP	0-0.06
10	MDVP:Shimmer	0.01-0.12
11	MDVP:Shimmer(dB)	0.09-1.3
12	Shimmer:APQ3	0-0.06
13	Shimmer:APQ5	0.01-0.08
14	MDVP:APQ	0.01-0.14
15	Shimmer:DDA	0.01-0.17
16	NHR	0-0.31
17	NHR	8.44-33.05
18	RPDE	0.26-0.69
19	DFA	0.57-0.83
20	spread1	-7.96-2.43
21	spread1	0.01-0.45
22	PPE	0.04-0.53
23	status	0 : NO 1 : YES
24	D2	1.42-3.67

Pada Tabel 1. Terdapat 24 atribut yang terdiri dari 23 atribut mengenai ciri serta 1 atribut status. Dataset ini terdiri dari 195 sample yang terdiri dari 48 sample data no dan 147 sample data yes. *Preprocessing* adalah suatu langkah yang dilakukan untuk mentransformasikan data yang ada sesuai dengan format data yang bisa diproses [15]. *Preprocessing* yang dilakukan membuang atribut yang tidak terlalu berpengaruh yaitu atribut name yang bertipe objek. Sehingga, atribut pada dataset ini yang digunakan menjadi 23 atribut. Pada penelitian ini, diambil 80% dari data untuk menjadi data *training* set kemudian sisa 20% menjadi data test untuk *percentage split*. Pada *K-fold cross validation* nilai k-fold yang dipilih pada penelitian ini yaitu 10 sehingga 10 kali *training* secara bergantian.

Preprocessing K-Nearest Neighbor adalah suatu metode yang dilakukan untuk pengambilan keputusan dengan pembelajaran terawasi dimana akibat dari data masukan yang baru diklasifikasi sesuai dengan data yang terdekat pada data nilai

Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain [16]:

1. Menentukan parameter *K*.
2. Menghitung kuadrat jarak Euclid (*query instance*) masing-masing objek terhadap data sampel yang diberikan menggunakan rumus (1)

$$Apq = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Pada Keterangan

A : Jarak Kedekatan

p : Data *Training*

q : Data *Testing*

i : Atribut Individu

n : Jumlah Atribut Individu

3. Urutkan semua objek tersebut ke berdasarkan kelompok dengan jarak Euclid terkecil.

4. Mengumpulkan kategori *Y* (Klasifikasi *Nearest Neighbor*).

5. Prediksi nilai query instance yang telah didapat sesuai dengan nilai *K-Nearest Neighbor* yang paling banyak.

Pemilihan atribut sebagai akar, didasarkan dari nilai gain yang paling tinggi sesuai dengan atribut yang dimiliki yang dimiliki [17]. Untuk mengetahui nilai gain dapat dihitung dengan menggunakan rumus rumus (2). seperti yang tertera [17]:

$$Gain(P, Q) = Entropy(P) - \sum_{i=0}^n \frac{|P_i|}{|P|} * Entropy(P_i) \quad (2)$$

Keterangan:

P : Himpunan Kejadian

Q : Atribut

n : Total Partisi pada Q

|Si| : Total Kejadian pada Partisi Ke-i

|S| : Total Kejadian pada Partisi P

Hal yang harus dilakukan sebelum mencari gain ialah mencari nilai Entropi terlebih dahulu. Nilai entropi bermanfaat dalam penentuan tingkat informatif dari masukan atribut untuk menghasilkan atribut yang baru [17]. Rumus dasar pada rumus (3). [17]:

$$Entropy(P) = \sum_{i=0}^n -p_i * \log_2 p_i \quad (3)$$

Keterangan:

P : Himpunan kejadian

n : Total partisi P

Pi : Proporsi dari Si terhadap P

Confusion matrix merupakan metode yang dimanfaatkan untuk menentukan tingkat akurasi pada data mining yang membagi klasifikasi prediksi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* merujuk pada *tuple positive* yaitu berlabel benar yang diberikan oleh classifier, *True negative* dengan

tuple negative berlabel dengan benar yang diberikan oleh *classifier* [9]. *False Positive* dengan *tuple negative* yang berlabel salah yang diberikan oleh *classifier*, serta *false negative* dengan *tuple positive* yang tidak diberikan label dengan benar oleh *classifier* [9]. Evaluasi *confusion matrix* adalah hasil evaluasi yang didapat pada tabel *matrix* yang dapat dilihat pada Tabel 2.

Tabel 2. *Confusion Matrix*

Kelas		Nilai Aktual	
		<i>Positive</i>	<i>Negatif</i>
Nilai Prediksi	<i>Positive</i>	<i>True positive</i>	<i>False Negative</i>
	<i>Negatif</i>	<i>Flase positive</i>	<i>True Negative</i>

Tabel 2. Menunjukkan nilai *actual* dan prediksi dari masing-masing kasus *positive* dan *negative* yang diklasifikasi baik benar maupun salah. Evaluasi penelitian ini dengan menggunakan *confusion matrix* untuk mendapatkan tingkat akurasi, presisi, dan *recall*. Nilai akurasi pada klasifikasi merupakan nilai persentase ketepatan *record* data yang diklasifikasi dengan benar yang dilakukan sesudah uji coba untuk hasil klasifikasi [18]. Presisi merupakan perbandingan uji yang diperkirakan bernilai positif dan dikira bernial positif benar pula untuk data yang aslinya. *Recall* merupakan perbandingan dari kasus positif yang sebelumnya telah diperkirakan bernilai positif benar [19]. Untuk melihat kinerja ketepatan algoritma KNN dan C4.5 dalam mendiagnosa penyakit parkinson maka akan diukur akurasi dengan menggunakan rumus (4), presisi dengan menggunakan rumus (5), dan *recall* dengan menggunakan rumus (6) [20].

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{Presisi} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

Pada tahapan ini, hasil dari penelitian yang diperoleh saat evaluasi kinerja pada KNN dan C4.5 untuk hasil diagnosa kedua kinerja diagnosa parkinson akan dibandingkan untuk melihat keakuratan yang lebih baik dalam mendiagnosa penyakit parkinson.

4 HASIL DAN PEMBAHASAN

1. Algoritma K-Nearest Neightbhor (KNN)

Pada algoritma KNN kita akan menggunakan data *testing* sebesar 20% dan data *training* sebesar 80% menggunakan pengujian *percentage split*. Pada *K-fold cross validation* nilai *k-fold* yang ditentukan yaitu 10. Pada algoritma KNN ini, terdapat 10 nilai *k* yang diambil adalah 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Dari percobaan yang dilakukan akan diperoleh nilai *precision*, *recall*, dan *accuracy* dari setiap nilai *k* yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pemodelan KNN

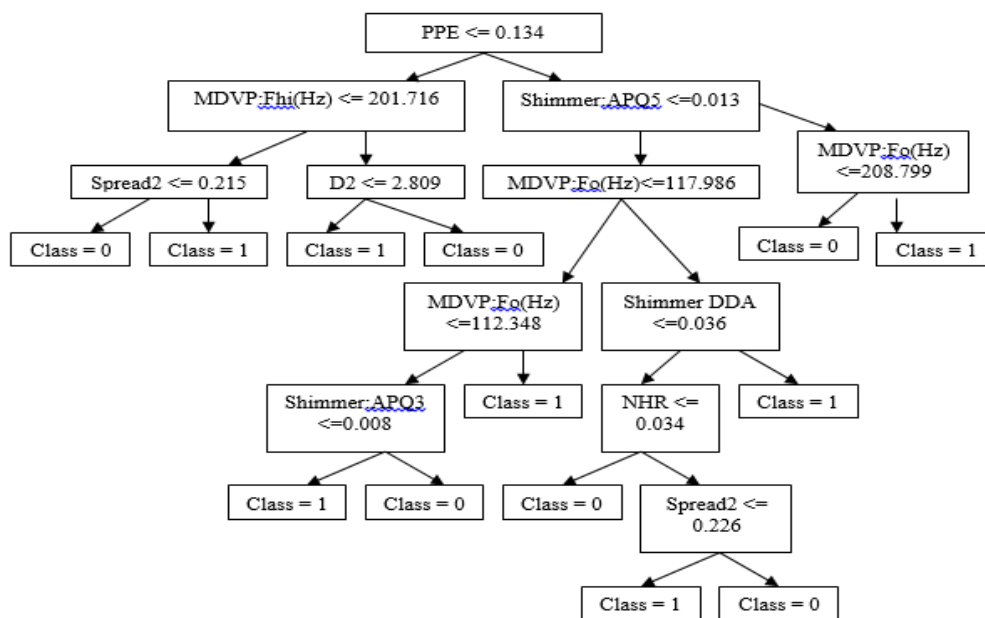
Nilai K	Target Kelas	<i>Persentase Split</i>			<i>K-fold cross validation</i>		
		Presisi	<i>Recall</i>	Akurasi	Presisi	<i>Recall</i>	Akurasi
1	0	86%	50%	82%	52%	31%	75.8%
	1	81%	96%		80%	90%	
2	0	64%	75%	79%	46%	65%	72%
	1	88%	81%		87%	75%	
3	0	86%	50%	82%	46%	38%	73.7%
	1	81%	96%		81%	86%	

4	0	78%	58%	82%	44%	44%	72%
	1	83%	93%				
5	0	80%	33%	76.9%	53%	42%	76.2%
	1	76%	96%				
6	0	67%	50%	76.9%	49%	46%	74.7%
	1	80%	89%				
7	0	75%	25%	74%	53%	33%	76.3%
	1	74%	96%				
8	0	75%	50%	79.4%	48%	33%	74.8%
	1	81%	93%				
9	0	75%	50%	79.4%	56%	29%	77%
	1	81%	93%				
10	0	78%	58%	79%	52%	31%	79.4%
	1	83%	93%				

Hasil pemodelan pada Tabel 3. Menunjukkan *training* yang dilakukan pada algoritma KNN diperoleh bahwa nilai akurasi yang tertinggi terdapat pada metode *training percentage split* yaitu sebesar 82% sedangkan untuk *K-fold cross validation* hanya sebesar 76.8%.

2. Algoritma C4.5

Pada perhitungan menggunakan algoritma C4.5 kita harus mencari nilai Gain tertinggi yang dimiliki oleh data dari penyakit parkinson. Dari hasil perhitungan Gain, diperoleh pohon keputusan dengan Gain tertinggi pada atribut PPE. Untuk nilai PPE yang tidak memenuhi syarat maka harus melihat nilai dari atribut lainnya untuk mendiagnosa penyakit parkinson seperti pada gambar 2.



Gambar 2. Pohon Keputusan C4.5

Gambar 2. Menunjukkan pohon keputusan dari algoritma C4.5 yang memperoleh aturan linguistik untuk mendiagnosa penyakit parkinson sebagai berikut : **IF PPE <= 0.134 AND MDVP:Fhi(Hz) <= 201.716 AND D2 <= 2.809 THEN** Diagnosa Penyakit Parkinson = 0 (no) **OR** Diagnosa Penyakit Parkinson = 1 (yes)

IF PPE <= 0.134 AND MDVP:Fhi(Hz) <= 201.716 AND Spread2 <= 0.215 THEN Diagnosa Penyakit Parkinson = 0 (no) **OR** Diagnosa Penyakit Parkinson = 1 (yes)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 208.799 THEN Diagnosa Penyakit Parkinson = 0 (no) **OR** Diagnosa Penyakit Parkinson = 1 (yes)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 117.986 AND Shimmer:DDA <= 0.036 THEN Diagnosa Penyakit Parkinson = 1 (yes)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 117.986 AND Shimmer:DDA <= 0.036 AND NHR <= 0.034 AND Spread2 <= 0.226 THEN Diagnosa Penyakit Parkinson = 0 (no) **OR** Diagnosa Penyakit Parkinson = 1 (yes)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 117.986 AND Shimmer:DDA <= 0.036 AND NHR <= 0.034 THEN Diagnosa Penyakit Parkinson = 0 (no)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 117.986 AND Shimmer:Fo(Hz) <= 112.348 THEN Diagnosa Penyakit Parkinson = 1 (yes)
IF PPE <= 0.134 AND Shimmer:APQ5 <= 0.013 AND MDVP:Fo(Hz) <= 117.986 AND Shimmer:Fo(Hz) <= 112.348 AND Shimmer:APQ3 <= 0.008 THEN Diagnosa Penyakit Parkinson = 0 (no) **OR** Diagnosa Penyakit Parkinson = 1 (yes)

Setelah memperoleh aturan pohon keputusan dari penyakit Parkinson dengan algoritma C4.5, selanjutnya akan dicari perhitungan menggunakan *Percentage split* dengan 80% *training* set dan test set 20%, maka diperoleh

[[9, 1][3, 26]]

Dengan presisi dan recall yang dapat dilihat pada Tabel 4.

Tabel 4. Percentage Split

Kelas	presisi	recall
0	0.75	0.90
1	0.96	0.90

Tabel 4. menunjukkan bahwa pada algoritma C4.5 hasil precision dengan *percentage split* pada label no (0) diperoleh 0,75 sedangkan pada label yes (1) diperoleh 0,96. Demikian juga nilai recall untuk label no (0) diperoleh 0,90 sedangkan pada label yes(1) diperoleh 0,90. Nilai akurasi dari uji algoritma C4.5 dengan *Percentage split* sebesar 89%. Hasil yang diperoleh menunjukkan bahwa algoritma C4.5 baik untuk mendiagnosa penyakit Parkinson.

Selanjutnya akan dilakukan perhitungan menggunakan metode *training K-fold cross validation* dengan nilai K-Fold sebesar 10, maka diperoleh :

[[17, 21][8 110]]

Dengan presisi dan recall yang dapat dilihat pada Tabel 5.

Tabel 5. Percentage Split

Kelas	presisi	recall
0	0.68	0.45
1	0.84	0.93

Tabel 5. menunjukkan hasil perhitungan metode *training* dari algoritma C4.5 ini diperoleh hasil precision dengan *K-fold cross validation* pada label no (0) diperoleh 0,68 sedangkan pada label yes (1) diperoleh 0,84. Demikian juga nilai recall untuk label no (0) diperoleh 0,45 sedangkan pada label yes(1) diperoleh 0,93. Nilai akurasi untuk algoritma C4.5 dengan *K-fold cross validation* sebesar 81%. Hasil penelitian ini menunjukkan bahwa algoritma C4.5 cukup baik untuk mendiagnosa penyakit parkinson. Dari kedua metode *training* yang dilakukan pada algoritma C4.5 diperoleh bahwa nilai akurasi yang tertinggi

terdapat pada metode *training Percentage split* yaitu sebesar 89% sedangkan untuk *K-fold cross validation* sebesar 81%.

3. Perbandingan Hasil Kedua Algoritma

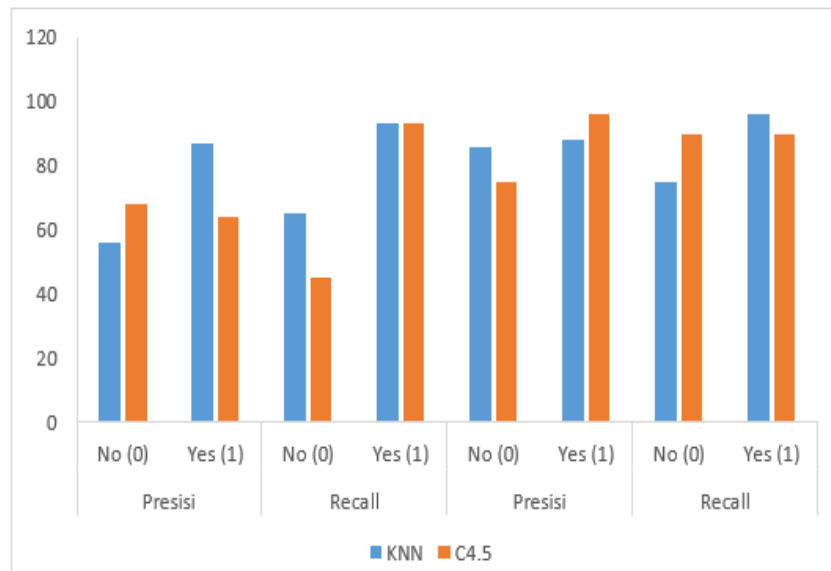
Dari hasil penelitian untuk diagnosa penyakit Parkinson dengan algoritma KNN dan C4.5 ini dapat dilihat bahwa keduanya memiliki nilai yang cukup baik serta tidak berbeda jauh baik dari nilai presisi, recall, maupun accuracy yang dapat dilihat pada Tabel 6.

Tabel 6. Nilai *Precision*, *Recall*, dan *Accuracy* Algoritma KNN dan C4.5

Algoritma	Metode Training	Kelas	Presisi	Recall	Akurasi
<i>K-Nearest Neighbor</i> (KNN)	<i>Persentase Split</i>	No (0)	86%	75%	82%
		Yes (1)	88%	96%	
<i>K-Nearest Neighbor</i> (KNN)	<i>K-fold cross validation</i>	No (0)	56%	65%	76.8%
		Yes (1)	87%	93%	
C4.5	<i>Persentase Split</i>	No (0)	75%	90%	89%
		Yes (1)	96%	90%	
C4.5	<i>K-fold cross validation</i>	No (0)	68%	45%	81%
		Yes (1)	84%	93%	

Pada Tabel 6. Dapat dilihat bahwa nilai accuracy pada algoritma C4.5 lebih baik dibandingkan nilai accuracy pada algoritma KNN baik dengan metode *training percentage split* maupun *K-fold cross validation*. Pada metode *training presentase split*, nilai accuracy pada algoritma KNN sebesar 82%, sedangkan accuracy pada algoritma C4.5 sebesar 89%. Demikian pula dengan metode *training K-fold cross validation*, nilai accuracy pada algoritma KNN sebesar 76.8%, sedangkan accuracy pada algoritma C4.5 sebesar 81%. Untuk nilai presisi, algoritma KNN dengan *percentagesplit* cukup baik yaitu sebesar 88%. Namun, algoritma C4.5 memiliki nilai presisi yang lebih baik dengan metode *training* yang sama yaitu *Percentage split* sebesar 96%.

Nilai recall C4.5 sebesar 90% untuk metode *training percentagesplit* dan 45% untuk metode *training K-fold cross validation* pada label no(0) serta 93% untuk metode *training K-fold cross validation* pada label yes (1). Nilai recall dari algoritma KNN pada metode *training Percentage split* yaitu sebesar 75% pada label no(0) dan pada metode *training K-fold cross validation* yaitu sebesar 65% pada label no(0). Nilai recall dari algoritma KNN sangat baik pada metode *training percentagesplit* yaitu sebesar 96% pada label yes(1) dan pada metode *training K-fold cross validation* yaitu sebesar 93% pada label yes(1). Hasil precision dan recall pada algoritma KNN dan C4.5 dapat dilihat pada gambar 3.



Gambar 3. Nilai *Precision* dan *Recall* dari hasil klasifikasi menggunakan metode *percentage split* dan *K-fold cross validation*

Dari Gambar 3. dapat dilihat bahwa nilai *Precision* dan *Recall* yang diperoleh dengan algoritma KNN dan C4.5 baik dari metode *training percentage split* maupun *K-fold cross validation* tidak jauh berbeda. Kedua algoritma tersebut cukup baik untuk mendiagnosa penyakit parkinson.



Gambar 4. Rata-rata Keseluruhan Nilai *Precision* dan *Recall* dengan metode *percentage split* dan *K-fold cross validation*

Dari Gambar 4. dapat dilihat bahwa nilai rata-rata *Precision* dan *Recall* yang diperoleh dengan algoritma KNN dan C4.5 baik dari metode *training percentage split* maupun *K-fold cross validation* sudah cukup konsisten. Kedua algoritma tersebut cukup baik untuk mendiagnosa penyakit parkinson.

5 KESIMPULAN

Berdasarkan pengaplikasian KNN dan C4.5 untuk mendiagnosa penyakit parkinson yang dilakukan dalam penelitian ini, diperoleh bahwa nilai *precision* dan *recall* dari label *yes* (1) lebih besar dibandingkan nilai *precision* dan *recall* dari label *no* (0). Untuk nilai tertinggi dari *precision* terdapat pada label *yes* (1) dengan algoritma C4.5 *percentage split*. Begitu juga

untuk nilai recall, label yes (1) memiliki nilai yang lebih tinggi pada KNN *percentage split* dibandingkan nilai recall pada label no (0). Algoritma KNN baik pada *percentage split* maupun *K-fold cross validation* memberikan nilai akurasi yang sangat baik. Namun, algoritma C4.5 memberikan akurasi yang lebih tinggi lagi baik untuk *percentage split* maupun untuk *K-fold cross validation* dibandingkan nilai akurasi pada algoritma KNN.

DAFTAR PUSTAKA

- [1] D. A. H. Pitra, "Disfungsi Mitokondria pada Penyakit Parkinson," *Baiturrahmah Med. J.*, vol. 1, no. 2, pp. 77–82, 2021, [Online]. Available: <https://jurnal.unbrah.ac.id/index.php/brmj/article/download/1038/pdf>
- [2] K. Ngurah and D. Putra, "Program Vo2Max Dengan Resiko Jatuh Pada Pasien," vol. 05, no. 01, pp. 1–7, 2023.
- [3] O. Tambun, S. Marisdina, and E. Bahar, "Faktor-Faktor Yang Mempengaruhi Kualitas Hidup Penderita Penyakit Parkinson Dengan Menggunakan," *Artik. Penelit. Neurona*, vol. 38, no. 4, p. 270, 2021.
- [4] H. P. Catur Ari Setianto, Eko Arisetijono, Dessika Rahmawati, Zamroni Afif, Rodhiyan Rakhmatiar, Neila Raisa, *Melawan Parkinson: Diagnosis dan Tata Laksana Holistik Penyakit Parkinson*. Universitas Brawijaya Press, 2023.
- [5] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [6] M. Y. Rizki, E. Sartika, Y. Pratama, S. Rasika, A. P. Windarto, and A. D. A. N. Pembahasan, "Implementasi C45 Dalam Memprediksi Index Prestasi Mahasiswa / i Menurut Kebiasaan Belajar," *Semin. Nas. Teknol. Komput. Sains*, pp. 297–303, 2020.
- [7] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [8] J. P. Gultom and A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin," *Kumpul. Artik. Karya Ilm. Fak. Ilmu Komput.*, vol. 02, no. 01, pp. 11–19, 2020.
- [9] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.
- [10] R. T. Prasetyo, "SELEKSI FITUR DAN OPTIMASI PARAMETER k-NN BERBASIS ALGORITMA GENETIKA PADA DATASET MEDIS," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 213–221, 2020, doi: 10.51977/jti.v2i2.319.
- [11] Yadi, "Implementation Algorithm C4.5 Classification Of Prospective Scholarship Recipients," *J. SimanteC*, vol. 11, no. 1, pp. 27–32, 2022.
- [12] Nurfitriyani, Islamiyah, and A. P. A. Masa, "Penerapan Klasifikasi Algoritma C4.5 Dan Algoritma C5.0 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Website Sistem Informasi Terpadu Layanan Program Studi (SIPLO)," *J. Media Inform. Budidarma*, vol. 7, pp. 1877–1887, 2023, doi: 10.30865/mib.v7i4.6433.
- [13] S. Agustiani, A. Mustopa, A. Saryoko, W. Gata, and S. K. Wildah, "Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 153–160, 2020, doi: 10.31294/p.v22i2.8174.
- [14] W. Z. Aprilita, R. Akbar, R. Cahyadi Prayogi, and Rahmaddeni, "Komparasi Algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam Klasifikasi Penyakit Parkinson," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 188–193, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [15] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5,

- Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [16] A. M. Widodo, N. Anwar, B. Irawan, A. Wisnujati, and L. Meria, “Komparasi Performansi Algoritma Pengklasifikasi KNN, Bagging Dan Random Forest Untuk Prediksi Kanker Payudara,” *Proceeding KONIK (Konferensi Nas. Ilmu Komputer)*, vol. 5, pp. 367–372, 2021, [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=FOwZ8hUAAAAJ&pagesize=100&citation_for_view=FOwZ8hUAAAAJ:fFSKOagxvKUC
- [17] Y. Y. Nur and A. Hermawan, “Implementasi Algoritma C.45 Untuk Menentukan Jurusan Calon Siswa Sekolah Menengah Kejuruan (Smk),” *J. Simantec*, vol. 11, no. 2, pp. 175–184, 2023, doi: 10.21107/simantec.v11i2.20384.
- [18] M. N. Akbar, “Klasifikasi Kanker Menggunakan Algoritma Nnge , Random Forest , Dan Random Commitee,” *J. Inform. dan Teknol.*, vol. 5, pp. 289–298, 2020, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/instek/index>
- [19] S. P. Backar, P. Purnawansyah, H. Darwis, and W. Astuti, “Hybrid Fourier Descriptor Naïve Bayes dan CNN pada Klasifikasi Daun Herbal,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 2, pp. 126–133, 2023, doi: 10.30591/jpit.v8i2.5186.
- [20] M. R. Santoso and P. Musa, “Rekomendasi Kesehatan Janin Dengan Penerapan Algoritma C5.0 Menggunakan Classifying Cardiocography Dataset,” *J. Simantec*, vol. 9, no. 2, pp. 65–76, 2021, doi: 10.21107/simantec.v9i2.10730.