

Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4.5 Pada Klasifikasi Data Penyakit Infertilitas

Agung Prabowo^{*}, Sumita Wardani, Rico Wijaya Dewantoro, Wilfredo Wesly, Leonardo

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Prima Indonesia, Medan, Indonesia

Email: ^{1,*}agungprabowo@unprimdn.ac.id, ²sumitawardani@unprimdn.ac.id, ³rico@unprimdn.ac.id, ⁴wilfredowesly@gmail.com, ⁵kholeonardo100@gmail.com

Email Penulis Korespondensi: agungprabowo@unprimdn.ac.id

Abstrak—Pada dua dekade terakhir kesuburan pria mengalami penurunan. Penurunan tersebut disebabkan oleh faktor lingkungan, seperti kebiasaan gaya hidup yang dapat memengaruhi kualitas sperma pria. Teknologi kecerdasan buatan saat ini sedang berkembang sebagai metodologi untuk sistem pendukung keputusan kesehatan. Dalam proses untuk memprediksi penyakit infertilitas dapat dilakukan dengan menggunakan teknologi Machine Learning. Pada penelitian ini difokuskan melakukan perbandingan atau komparasi metode klasifikasi Random Forest dengan Decision Tree C4.5 untuk mengukur tingkat akurasi pada prediksi keberhasilan dari klasifikasi data penyakit infertilitas. Data untuk Fertility Dataset diperoleh dari UCI Machine Learning Repository dengan jumlah sebanyak 100 record data, 10 atribut dan 2 kelas atribut yaitu Normal dan Altered. Parameter yang digunakan yaitu usia, penyakit pada masa anak-anak, kecelakaan atau trauma, operasi bedah, konsumsi alkohol dan kebiasaan merokok. Kemudian evaluasi pengujian kedua metode tersebut yaitu dengan menggunakan 10fold Cross Validation. Berdasarkan hasil pengujian Random Forest dan Decision Tree C4.5, diperoleh hasil yaitu rata-rata akurasi dari Random Forest sebesar 87.20 % dan Decision Tree C4.5 dengan tingkat akurasi 85.90 %. Dari hasil yang diperoleh dapat disimpulkan bahwa Random Forest merupakan metode yang lebih unggul sebesar 1.3 % jika dibandingkan dengan Decision Tree C4.5 dalam memprediksi akurasi pada Fertility Dataset.

Kata Kunci: Klasifikasi; Infertilitas; Random Forest; Decision Tree C4.5

Abstract—Male fertility has declined over the past two decades. The decrease is due to environmental factors, such as lifestyle habits that can affect the quality of a man's sperm. Artificial intelligence technology is currently developing as a methodology for health decision support systems. In the process of predicting infertility can be done by applying Machine Learning technology. This study focuses on comparing the Random Forest classification method with Decision Tree C4.5 to see the level of accuracy in predicting the success of infertility data classification. Data for the Fertility Dataset was obtained from the UCI Machine Learning Repository with a total of 100 data records, 10 attributes and 2 attribute classes, namely Normal and Altered. The parameters used are age, childhood diseases, accidents or trauma, surgical operations, alcohol consumption and smoking habits. Then evaluate the testing of the two methods, namely by using 10fold Cross Validation. Based on the results of Random Forest and Decision Tree C4.5 testing, the average accuracy of Random Forest is 87.20% and Decision Tree C4.5 with an accuracy rate of 85.90%. From the results obtained, it can be concluded that Random Forest is a superior method by 1.3% when compared to Decision Tree C4.5 in predicting accuracy in the Fertility Dataset.

Keywords: Classification; Infertility; Random Forest; Decision Tree C4.5

1. PENDAHULUAN

Menurut Organisasi Kesehatan Dunia (WHO), tingkat kesuburan telah turun drastis. Infertilitas didefinisikan sebagai ketidakmampuan untuk hamil setelah satu tahun sering melakukan hubungan seksual tanpa menggunakan kontrasepsi. Infertilitas dapat digambarkan sebagai penyakit pada sistem reproduksi [1]. Angka infertilitas pasangan suami istri yang sulit memiliki anak di Indonesia sekitar 10%. Faktor laki-laki tetap menjadi penyebab utama ketidaksuburan [2]. Organisasi Kesehatan Dunia menemukan bahwa lebih dari 50% ketidaksuburan disebabkan oleh laki-laki, dengan air mani sebagai faktor terbesar [3]. Beberapa faktor dapat mempengaruhi kesuburan sperma, salah satunya gaya hidup dan lingkungan. Gaya hidup dan lingkungan tertentu dapat mempengaruhi konsentrasi sperma, motilitas sperma, dan morfologi sperma yang merupakan faktor-faktor yang mempengaruhi kualitas sperma [4].

Kriteria atau parameter yang digunakan untuk mengukur tingkat infertilitas didasarkan pada data sampel yang diambil dari UCI Machine Learning Repository berjudul "Fertility Dataset". Jurnal penelitian berjudul "Predicting Semen Quality Using Artificial Intelligence Methods" merupakan penelitian yang dilakukan dengan menggunakan data sampel yang sama [5]. Dalam proses memprediksi kualitas sperma, penelitian menggunakan faktor gaya hidup dan lingkungan sebagai acuan untuk menentukan kualitas sperma. Hasil klasifikasi atau keluaran yang diberikan adalah normal dan berubah.

Teknik pembelajaran mesin dapat digunakan untuk mengklasifikasikan data penyakit infertilitas untuk meningkatkan produktivitas dan konsistensi diagnostik. Banyak penelitian telah menggunakan teknik klasifikasi. Ini meningkatkan konsistensi hasil dan mendorong terciptanya taksonomi baru, yang memungkinkan teknologi berkembang [6].

Beberapa penelitian terdahulu yang menerapkan teknik untuk klasifikasi penyakit seperti pada penelitian dari [7] yaitu membandingkan Naïve Bayes dan Random Forest untuk klasifikasi penyakit liver dengan hasil yang diperoleh yaitu Random Forest lebih unggul dari perolehan akurasi dibandingkan dengan Naïve Bayes dengan akurasi sebesar 70.60%. Kemudian penelitian dari [8] yang melakukan komparasi K-Nearest Neighbor dan Random Forest pada klasifikasi pengobatan penyakit kutil dengan hasil yang diperoleh yaitu Random Forest lebih unggul dalam perolehan

akurasinya yaitu sebesar 85.50% untuk klasifikasi pengobatan penyakit kutil. Kemudian penelitian [9] yang meneliti tentang penerapan metode Decision Tree C4.5 untuk klasifikasi penyakit Diabetes Mellitus dengan hasil yang diperoleh yaitu tingkat akurasi klasifikasi penyakit Diabetes Mellitus yaitu sebesar 97.12%.

Kemudian beberapa penelitian terdahulu yang membahas mengenai klasifikasi penyakit infertilitas yaitu seperti pada penelitian oleh [10] meneliti tentang system pendukung keputusan dalam menentukan tingkat konsentrasi sperma dengan Fertility Dataset menggunakan metode MOORA dengan hasil yang diperoleh yaitu metode MOORA mampu melakukan pemeringkatan data konsentrasi tingkat kesuburan sperma. Kemudian penelitian oleh [11] meneliti tentang system pendukung keputusan dalam menentukan tingkat konsentrasi sperma dengan Fertility Dataset menggunakan ELECTRE dan MOORA dengan hasil yang diperoleh yaitu ELECTRE lebih unggul dari MOORA dari segi waktu eksekusi. Kemudian penelitian oleh [12] melakukan seleksi fitur pada Naïve Bayes dengan Forward Selection dalam klasifikasi penyakit infertilitas dengan hasil akurasi yang diperoleh yaitu 88 %. Kemudian penelitian dari [1] yang membahas mengenai optimasi metode Decision Tree C4.5 menggunakan Particle Swarm Optimization (PSO) untuk klasifikasi infertilitas dengan hasil yang diperoleh yaitu tingkat akurasi sebesar 93.33%.

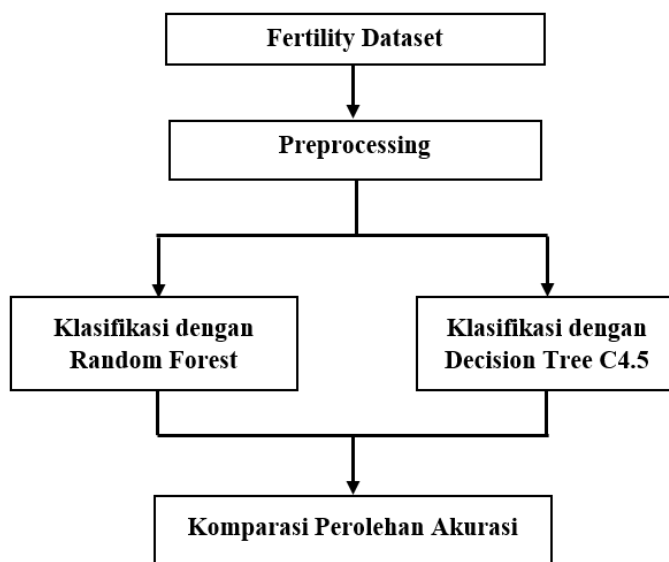
Random Forest dan Decision Tree C4.5 merupakan metode klasifikasi yang digunakan oleh penulis pada penelitian ini untuk melakukan klasifikasi pada Fertility Dataset. Penelitian ini dilakukan dengan tujuan untuk melakukan komparasi antara Random Forest dengan Decision Tree C4.5 untuk melihat metode yang lebih unggul ketepatan akurasi yang diperoleh dalam memprediksi akurasi dari keberhasilan metode klasifikasi penyakit infertilitas. Dataset yang digunakan dalam penelitian ini yaitu Fertility Dataset yang dipublikasikan melalui UCI Learning Machine Repository. Fertility Dataset ini terdiri atas 100 record dengan 10 atribut. Label dataset ini terdiri atas 2 class antara lain N (normal) dan O (altered). Dimana 88 record pada class N (normal) dan 12 class pada class O (altered).

Adapun penelitian ini bertujuan untuk melakukan komparasi antara Random Forest dan Decision Tree C4.5 untuk melihat kinerja yang paling unggul dari kedua metode tersebut dalam proses klasifikasi data penyakit infertilitas. Adapun pengukuran kinerja yang dilakukan yaitu berdasarkan nilai perolehan accuracy, precision dan recall dari masing-masing pengujian kedua metode tersebut

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan-tahapan yang dilakukan untuk memperoleh hasil pengujian pada penelitian ini dengan melakukan prosedur-prosedur seperti yang digambarkan pada Gambar 1 berikut:



Gambar 1. Tahapan Penelitian

Penjelasan dari tahapan penelitian pada Gambar 1 yang ditampilkan di atas sebagai berikut:

- Fertility Dataset yaitu data set yang akan diujikan pada penelitian ini yang diperoleh dari UCI Machine Learning Repository.
- Pre-Processing atau proses normalisasi data set untuk menghilangkan anomaly pada data set yang diujikan.
- Klasifikasi dengan Random Forest untuk melakukan klasifikasi pada data set dengan Random Forest.
- Klasifikasi dengan Decision Tree C4.5 untuk melakukan klasifikasi pada data set dengan Decision Tree C4.5.
- Komparasi Perolehan Akurasi yaitu untuk membahas mengenai perbandingan dari hasil pengujian kedua metode yang diusulkan pada penelitian ini kemudian memaparkan secara ringkas perbedaan hasil yang diperoleh dari kedua metode yang digunakan.

2.2 Fertility Dataset

Dataset yang digunakan dalam penelitian ini yaitu *Fertility Dataset* yang merupakan hasil penelitian dari David Gil, Jose Luis Girela, Joaquin De Juan, M. Jose Gomez-Torres, Magnus Johnsson pada tahun 2012. *Fertility* dipublikasikan melalui *UCI Learning Machine Repository*. Kemudian data set tersebut terdiri atas 100 records data dengan 10 atribut. Label dataset ini terdiri atas 2 class antara lain N (*Normal*) dan O (*Altered*). Dimana 88 record pada class N (*Normal*) dan 12 class pada class O (*Altered*).

2.3 Random Forest

Random Forest merupakan metode yang membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* [13]. Adapun tahapan penyelesaian dengan *Random Forest* sebagai berikut [14],[15]:

- Menentukan jumlah *tree* (k) yang dipilih dari total fitur m , di mana k kurang dari m .
- Kemudian sampel acak diambil sebanyak N pada dataset untuk setiap *tree*.
- Pada setiap *tree*, dilakukan pengambilan subset prediktor sebanyak secara acak. Dimana $m < p$, yang dimana p adalah jumlah variabel prediktor.
- Kemudian, proses pada langkah kedua dan ketiga diulangi sampai sebanyak k *tree*.
- Hasil prediksi diperoleh dari *vote* terbanyak dari hasil klasifikasi sebanyak *tree*.

2.4 Decision Tree C4.5

Decision Tree C4.5 digunakan untuk membentuk pohon keputusan dengan prediksi yang sangat kuat [16]. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan yang dapat dengan mudah dipahami [17]. Tahapan dalam Decision Tree C4.5, yaitu [18]:

- Menghitung nilai *Entropy* pada masing-masing atribut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

- Menghitung nilai *Information Gain* pada masing-masing atribut:

$$Info\ Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

- Menghitung nilai *Split Information* untuk masing-masing atribut:

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (3)$$

- Menghitung nilai *Gain Ratio* untuk masing-masing atribut:

$$Gain\ Ratio(A) = \frac{Info\ Gain(A)}{SplitInfo(A)} \quad (4)$$

- Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi akar (*splitting attribute*) dan atribut yang memiliki nilai *Gain Ratio* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*),
- Menghitung lagi nilai *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya,
- Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*). Mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa.

2.5 Confusion Matrix

Menghitung hasil kinerja klasifikasi dari masing-masing pengujian metode dengan *Confusion Matrix* untuk memperoleh hasil *Accuracy*, *Precision*, dan *Recall*. *Confusion Matrix* digunakan untuk menganalisis seberapa baik classifier mengenali data kelas yang berbeda [19]. Tabel *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1. Tabel *Confusion Matrix*

Actual Class	Assigned Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

True Positive dan *True Negative* adalah keadaan pada saat hasil prediksi sesuai dengan kondisi sebenarnya yang terjadi. *False Positive* dan *False Negative* adalah keadaan dimana hasil prediksi tidak sesuai dengan kondisi yang sebenarnya. Akurasi, presisi, dan *recall* dapat dihitung menggunakan rumus [20]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Precision = \frac{TP+TN}{TP+FP} \quad (7)$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan informasi data set yang digunakan, hasil dan pembahasan dari pengujian *Random Forest* dan *Naive Bayes* pada penelitian ini. Adapun pengujian dari metode yang digunakan yaitu menggunakan bantuan *tools* yang berguna dalam memudahkan dan mempersingkat prosedur pengujian dengan menggunakan *software Waikato Environment of Knowledge Analysis (WEKA)*.

3.1 Data dan Atribut Fertility Dataset

Pada Tabel 2 dan Tabel 3 berikut merupakan informasi atribut dan rincian data pada *Fertility Dataset* yang digunakan pada penelitian ini:

Tabel 2. Atribut *Fertility Dataset*

Atribut	Keterangan	Nilai Kategori
<i>Season</i> (X1)	Musim dimana analisis dilakukan 1)Musim dingin, 2)Musim semi, 3)Musim panas, 4)Musim gugur	(-1, -0,33, 0,33, 1)
<i>Age</i> (X2)	Umur pada saat analisis (18-36 tahun)	Skala [0-1]
<i>Childish Disease</i> (X3)	Penyakit anak (cacar air, campak, gondok, polio)	Ya, Tidak (0,1)
<i>Accident or Serious Trauma</i> (X4)	Kecelakaan atau trauma serius	Ya, Tidak (0,1)
<i>Surgical Intervention</i> (X5)	Mengalami pembedahan	Ya, Tidak (0,1)
<i>High Fever in the Last Year</i> (X6)	Demam tinggi pada 1 tahun terakhir Kurang dari tiga bulan yang lalu, lebih dari tiga bulan yang lalu,	(-1, 0, 1)
<i>Frequency of Alcohol Consumption</i> (X7)	Frekuensi konsumsi alkohol	Beberapa kali sehari, setiap hari, beberapa kali dalam seminggu, seminggu sekali, hampir tidak pernah, atau tidak pernah sama sekali (0, 0,2, 0,4, 0,6, 0,8, 1)
<i>Smoking Habbit</i> (X8)	Kebiasaan merokok	Tidak pernah, sesekali, setiap hari (-1, 0, 1)
<i>Number of Hours Spent Sitting Per Day</i> (X9)	Jumlah jam yang dihabiskan duduk per hari (Antara 1 hingga 16 jam)	Skala [0-1]
<i>Diagnosis</i>	Normal (N), Altered (O)	1,0

Tabel 3. Rincian Data *Fertility Dataset*

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	Class
1	-0.33	0.69	0	1	1	0	0.8	0	0.88	N
2	-0.33	0.94	1	0	1	0	0.8	1	0.31	O
3	-0.33	0.5	1	0	0	0	1	-1	0.5	N
4	-0.33	0.75	0	1	1	0	1	-1	0.38	N
5	-0.33	0.67	1	1	0	0	0.8	-1	0.5	O
6	-0.33	0.67	1	0	1	0	0.8	0	0.5	N
7	-0.33	0.67	0	0	0	-1	0.8	-1	0.44	N
8	-0.33	1	1	1	1	0	0.6	-1	0.38	N
9	1	0.64	0	0	1	0	0.8	-1	0.25	N
10	1	0.61	1	0	0	0	1	-1	0.25	N
11	1	0.67	1	1	0	-1	0.8	0	0.31	N
12	1	0.78	1	1	1	0	0.6	0	0.13	N
13	1	0.75	1	1	1	0	0.8	1	0.25	N
14	1	0.81	1	0	0	0	1	-1	0.38	N
15	1	0.94	1	1	1	0	0.2	-1	0.25	N
16	1	0.81	1	1	0	0	1	1	0.5	N
17	1	0.64	1	0	1	0	1	-1	0.38	N

18	1	0.69	1	0	1	0	0.8	-1	0.25	N
19	1	0.75	1	1	1	0	1	1	0.25	N
20	1	0.67	1	0	0	0	0.8	1	0.38	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	1	0.58	1	0	0	0	0.6	1	0.5	N
96	-1	0.67	1	0	0	0	1	-1	0.5	N
97	-1	0.61	1	0	0	0	0.8	0	0.5	N
98	-1	0.67	1	1	1	0	1	-1	0.31	N
99	-1	0.64	1	0	1	0	1	0	0.19	N
100	-1	0.69	0	1	1	0	0.6	-1	0.19	N

3.2 Tahapan Preprocessing Data

Tahapan preprocessing atau normalisasi data set dilakukan sebelum proses klasifikasi yang bertujuan untuk menghilangkan anomali dalam data set serta merupakan proses untuk menyeragamkan tipe data pada data set sehingga tidak ada lagi ketimpangan antara satu atribut dengan atribut lainnya di dalam data set tersebut. Sebelum melakukan tahapan klasifikasi pada data set yang diujikan, dilakukan proses normalisasi pada data set tersebut. Adapun teknik normalisasi yang digunakan pada penelitian ini menerapkan teknik normalisasi *min-max* dengan menormalisasikan data menjadi nilai minimal sama dengan 0 dan nilai maksimal sama dengan 1 dengan persamaan yang digunakan yaitu pada persamaan (8) berikut [7]:

$$Normalisasi\ Min\ Max = \frac{N - \min(n)}{\max(n) - \min(n)} \tag{8}$$

Adapun N merupakan data yang belum ternormalisasi. Kemudian $\min(n)$ adalah nilai minimum dari semua data dan $\max(n)$ adalah nilai maksimum dari semua data [21]. Adapun hasil normalisasi data yang diperoleh yaitu pada Tabel 4 berikut:

Tabel 4. Hasil Normalisasi *Immunotherapy Dataset*

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	Class
1	0,335	0,38	0	1	1	0,5	0,75	0,5	0,872	N
2	0,335	0,88	1	0	1	0,5	0,75	1	0,265	O
3	0,335	0	1	0	0	0,5	1	0	0,468	N
4	0,335	0,5	0	1	1	0,5	1	0	0,340	N
5	0,335	0,34	1	1	0	0,5	0,75	0	0,468	O
6	0,335	0,34	1	0	1	0,5	0,75	0,5	0,468	N
7	0,335	0,34	0	0	0	0	0,75	0	0,404	N
8	0,335	1	1	1	1	0,5	0,5	0	0,340	N
9	1	0,28	0	0	1	0,5	0,75	0	0,202	N
10	1	0,22	1	0	0	0,5	1	0	0,202	N
11	1	0,34	1	1	0	0	0,75	0,5	0,265	N
12	1	0,56	1	1	1	0,5	0,5	0,5	0,074	N
13	1	0,5	1	1	1	0,5	0,75	1	0,202	N
14	1	0,62	1	0	0	0,5	1	0	0,340	N
15	1	0,88	1	1	1	0,5	0	0	0,202	N
16	1	0,62	1	1	0	0,5	1	1	0,468	N
17	1	0,28	1	0	1	0,5	1	0	0,340	N
18	1	0,38	1	0	1	0,5	0,75	0	0,202	N
19	1	0,5	1	1	1	0,5	1	1	0,202	N
20	1	0,34	1	0	0	0,5	0,75	1	0,340	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	1	0,16	0,5681	1	0	0	0,5	0,5	1	N
96	0	0,34	0,4090	1	0	0	0,5	1	0	N
97	0	0,22	0,5909	1	0	0	0,5	0,75	0,5	N
98	0	0,34	0,9545	1	1	1	0,5	1	0	N
99	0	0,28	1	1	0	1	0,5	1	0,5	N
100	0	0,38	0,5227	0	1	1	0,5	0,5	0	N

3.3 Pengujian Klasifikasi dengan Random Forest

Setelah hasil dari tahapan normalisasi data diperoleh, langkah selanjutnya yaitu pengujian metode klasifikasi dan yang pertama dilakukan yaitu dengan klasifikasi *Fertility Dataset* dengan *Random Forest*. Pengujian *Random Forest* pada penelitian ini menggunakan pengujian berdasarkan nilai *10-fold cross validation* yaitu dengan nilai *fold* 1 sampai dengan 10 dan masing-masing diujikan untuk memperoleh jumlah data yang benar dan data yang salah dari hasil klasifikasi. Kemudian menghitung nilai *Accuracy*, *Precision* dan *Recall* berdasarkan *Confusion Matrix*. Untuk memudahkan proses klasifikasi, penulis dibantu dengan bantuan *Waikato Environment for Knowledge Analysis* (WEKA). Hasil klasifikasi yang diperoleh dengan pengujian *Random Forest* dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Klasifikasi dengan *Random Forest*

<i>Fold</i>	Jumlah Data Benar	Jumlah Data Salah	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)
1	91	19	91.00	80.90	91.00
2	89	11	89.00	80.90	89.00
3	88	12	88.00	84.80	88.00
4	86	14	86.00	82.60	86.00
5	89	11	89.00	86.90	89.00
6	86	14	86.00	82.60	86.00
7	86	14	86.00	82.60	86.00
8	86	14	86.00	82.60	86.00
9	86	14	86.00	82.60	86.00
10	85	14	85.00	82.00	85.00
	Rata-Rata		87.20	83.45	87.20

3.4 Pengujian Klasifikasi dengan *Decision Tree C4.5*

Kemudian dilakukan klasifikasi pada *Fertility Dataset* dengan *Decision Tree C4.5*. Pengujian dengan *Decision Tree C4.5* berdasarkan nilai *10-Fold Cross Validation* dari pengujian nilai *Fold* 1 sampai dengan 10. Adapun hasil klasifikasi dengan *Decision Tree C4.5* yaitu pada Tabel 6 berikut.

Tabel 6. Hasil Klasifikasi dengan *Decision Tree C4.5*

<i>Fold</i>	Jumlah Data Benar	Jumlah Data Salah	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)
1	88	12	88.00	81.80	88.00
2	88	12	88.00	81.80	88.00
3	85	15	85.00	77.10	85.00
4	85	15	85.00	77.10	85.00
5	84	16	84.00	81.40	84.00
6	87	13	87.00	82.00	87.00
7	85	15	85.00	80.20	85.00
8	86	14	86.00	77.20	86.00
9	86	14	86.00	77.20	86.00
10	85	15	85.00	77.10	85.00
	Rata-Rata		85.90	79.29	85.90

3.5 Komparasi Perolehan Akurasi

Setelah pengujian klasifikasi pada *Random Forest* dan *Decision Tree C4.5* selesai dilakukan, maka dapat dilihat hasil perbandingan dari hasil pengujian kedua metode tersebut dalam melakukan perolehan akurasi klasifikasi pada *Fertility Dataset*. Diketahui pada bagian sebelumnya bahwa hasil akurasi yang diperoleh dari pengujian pada kedua metode tersebut memperoleh hasil yang berbeda dan signifikan. Adapun komparasi hasil klasifikasi pada Tabel 7 berikut.

Tabel 7. Perbandingan Hasil Klasifikasi

Metode	Akurasi (%)
<i>Random Forest</i>	87.20
<i>Decision Tree C4.5</i>	85.90

4. KESIMPULAN

Pengujian pada kedua metode yang digunakan pada penelitian ini memperoleh hasil yang berbeda dengan perbedaan hasil akurasi yang signifikan pada proses klasifikasi *Fertility Dataset*. Berdasarkan hasil pengujian *Random Forest* dan *Decision Tree C4.5* dalam melakukan prediksi dari tingkat keberhasilan dan dapat diambil kesimpulan bahwa hasil pengujian menggunakan *Random Forest* memperoleh akurasi sebesar 87.20 %, kemudian *Decision Tree C4.5* memperoleh akurasi sebesar 85.90 %. Dari hasil pengujian yang telah dilakukan pada klasifikasi *Fertility Dataset*,

menunjukkan bahwa Random Forest memperoleh hasil yang lebih baik lebih dan lebih direkomendasikan jika dibandingkan dengan *Decision Tree C4.5*.

REFERENCES

- [1] T. W. Pratiwi and T. Arifin, "Optimasi Decision Tree Menggunakan Particle Swarm Optimization untuk Klasifikasi Kesuburan pada Pria," *Sist. J. Sist. Inf.*, vol. 10, no. 1, pp. 1–12, 2021, doi: 10.35957/jatiji.v7i3.361.
- [2] A. Amrulloh and A. P. Wibowo, "Implementasi Algoritma Decision Tree Untuk Mengklasifikasi Kondisi Kesuburan Pria," *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 1, no. 1, pp. 7–11, 2019, doi: 10.26905/jasiek.v1i1.3096.
- [3] U. Khaira, N. Syarif, and I. Hayati, "Prediksi Tingkat Fertilitas Pria Dengan Algoritma Pohon Keputusan Cart," *Progr. Stud. Sist. Informasi, Fak. Sains dan Teknol. Univ. Jambi*, vol. 5, no. 1, pp. 35–42, 2020.
- [4] A. H. Baksir, A. Fuad, F. Tempola, and Rosihan, "Prediksi Tingkat Kualitas Kesuburan Pria Dengan Jaringan Saraf Tiruan Backpropagation," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 2, pp. 107–112, 2020, doi: 10.33387/jiko.
- [5] D. Gil, J. L. Girela, J. De Juan, M. J. Gomez-Torres, and M. Johnsson, "Predicting seminal quality with artificial intelligence methods," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12564–12573, 2012, doi: 10.1016/j.eswa.2012.05.028.
- [6] S. J. Siregar, A. I. Lubis, and E. F. Ginting, "Penerapan Neural Network Dalam Klasifikasi Citra Permainan Batu Kertas Gunting dengan Probabilistic Neural Network," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 420–425, 2021, doi: 10.47065/bits.v3i3.1143.
- [7] A. I. Lubis, U. Erdiansyah, and R. Siregar, "Komparasi Akurasi pada Naive Bayes dan Random Forest dalam Klasifikasi Penyakit Liver," *J. Comput. Eng. Syst. Sci.*, vol. 7, no. 1, pp. 81–89, 2022.
- [8] U. Erdiansyah, A. Irmansyah Lubis, and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 208, 2022, doi: 10.30865/mib.v6i1.3373.
- [9] F. M. Hana and Universitas, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 4, no. 1, pp. 32–39, 2020.
- [10] A. Irmansyah Lubis, F. Setiawan, and L. Lusiyanti, "Penentuan Peringkat Konsentrasi Tingkat Kesuburan Sperma Menggunakan Metode MOORA," *Digit. Transform. Technol.*, vol. 1, no. 2, pp. 62–68, 2021, doi: 10.47709/digitech.v1i2.1116.
- [11] A. I. Lubis and F. Setiawan, "Komparasi Kinerja ELECTRE dan MOORA dalam Menentukan Konsentrasi Tingkat Kesuburan Sperma," vol. 13, no. 01, pp. 99–105, 2022, doi: 10.35970/infotekmesin.v13i1.1012.
- [12] A. I. Lubis and R. Chandra, "Forward Selection Attribute Reduction Technique for Optimizing Naïve Bayes Performance in Sperm Fertility Prediction," *Sinkron*, vol. 8, no. 1, pp. 275–285, 2023, doi: 10.33395/sinkron.v8i1.11967.
- [13] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [14] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 186, 2021, doi: 10.30865/mib.v5i1.2624.
- [15] A. B. Wibisono and A. Fahrurrozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.
- [16] N. Sunanto and G. Falah, "Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 208–216, 2022, doi: 10.36341/rabit.v7i2.2435.
- [17] R. Estian Pambudi, Sriyanto, and Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5," *Ijccs*, vol. x, No.x, no. x, pp. 1–5, 2022.
- [18] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [19] U. Erdiansyah, A. I. Lubis, and G. Syahputra, "Klasifikasi Penyakit Diabetic Retinopathy Menggunakan Multilayer Perceptron," *JAISE J. Artif. Intell. Softw. Eng.*, vol. 2, no. 1, pp. 1–6, 2022.
- [20] A. I. Lubis, S. Sibagariang, and N. Ardi, "Classification of Alzheimer Disease from MRI Image Using Combination Naïve Bayes and Invariant Moment," *Proc. 5th Int. Conf. Appl. Eng. ICAE 2022, 5 Oct. 2022, Batam, Indones.*, pp. 1–12, 2023, doi: 10.4108/eai.5-10-2022.2327750.
- [21] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *Media Inform. Budidarma*, vol. 5, no. 4, pp. 1580–1584, 2021, doi: 10.30865/mib.v5i4.3347.