

## BAB IV

### HASIL DAN PEMBAHASAN

Pada bab ini penulis akan membahas mengenai perhitungan data yang digunakan. Data tersebut akan dihitung menggunakan algoritma Adaboost, Bagging dan PSO Pada Algoritma Knearest Neighbor Untuk Prediksi Penyakit Hati.

#### 4.1 Persiapan Data

Data yang digunakan dalam penelitian ini adalah dataset dari Kaggle yaitu data penyakit hati , Jumlah record data 615 data terdiri dari 12 atribut dan 1 Label (Blood Donor, suspect Blood Donor, Hepatitis, Fibrosis, Cirrhosis). Atribut yang digunakan Age, Sex, ALB (Albumin) , ALP (Alkaline Phosphatase), ALT (Alanine Aminotransferase), AST (Aspartate Aminotransferase), BIL (Bilirubin) , CHE (Cholinesterase), CHOL (Cholesterol), CREA (Creatinine), GGT (Gamma-Glutamyl Transferase), PROT (Protein). Kumpulan data ini digunakan untuk memprediksi apakah seorang pasien kemungkinan besar menderita penyakit hati. Data penyakit hati terdapat pada gambar 4.1 dibawah ini:

| No | Category  | Age | Sex | ALB  | ALP  | ALT  | AST  | BIL  | CHE   | CHOL | CREA | GGT  | PROT |
|----|-----------|-----|-----|------|------|------|------|------|-------|------|------|------|------|
| 1  | Blood Don | 32  | m   | 38.5 | 52.5 | 7.7  | 22.1 | 7.5  | 6.93  | 3.23 | 106  | 12.1 | 69   |
| 2  | Blood Don | 32  | m   | 38.5 | 70.3 | 18   | 24.7 | 3.9  | 11.17 | 4.8  | 74   | 15.6 | 76.5 |
| 3  | Blood Don | 32  | m   | 46.9 | 74.7 | 36.2 | 52.6 | 6.1  | 8.84  | 5.2  | 86   | 33.2 | 79.3 |
| 4  | Blood Don | 32  | m   | 43.2 | 52   | 30.6 | 22.6 | 18.9 | 7.33  | 4.74 | 80   | 33.8 | 75.7 |
| 5  | Blood Don | 32  | m   | 39.2 | 74.1 | 32.6 | 24.8 | 9.6  | 9.15  | 4.32 | 76   | 29.9 | 68.7 |
| 6  | Blood Don | 32  | m   | 41.6 | 43.3 | 18.5 | 19.7 | 12.3 | 9.92  | 6.05 | 111  | 91   | 74   |
| 7  | Blood Don | 32  | m   | 46.3 | 41.3 | 17.5 | 17.8 | 8.5  | 7.01  | 4.79 | 70   | 16.9 | 74.5 |
| 8  | Blood Don | 32  | m   | 42.2 | 41.9 | 35.8 | 31.1 | 16.1 | 5.82  | 4.6  | 109  | 21.5 | 67.1 |
| 9  | Blood Don | 32  | m   | 50.9 | 65.5 | 23.2 | 21.2 | 6.9  | 8.69  | 4.1  | 83   | 13.7 | 71.3 |
| 10 | Blood Don | 32  | m   | 42.4 | 86.3 | 20.3 | 20   | 35.2 | 5.46  | 4.45 | 81   | 15.9 | 69.9 |
| 11 | Blood Don | 32  | m   | 44.3 | 52.3 | 21.7 | 22.4 | 17.2 | 4.15  | 3.57 | 78   | 24.1 | 75.4 |
| 12 | Blood Don | 33  | m   | 46.4 | 68.2 | 10.3 | 20   | 5.7  | 7.36  | 4.3  | 79   | 18.7 | 68.6 |
| 13 | Blood Don | 33  | m   | 36.3 | 78.6 | 23.6 | 22   | 7    | 8.56  | 5.38 | 78   | 19.4 | 68.7 |
| 14 | Blood Don | 33  | m   | 39   | 51.7 | 15.9 | 24   | 6.8  | 6.46  | 3.38 | 65   | 7    | 70.4 |
| 15 | Blood Don | 33  | m   | 38.7 | 39.8 | 22.5 | 23   | 4.1  | 4.63  | 4.97 | 63   | 15.2 | 71.9 |
| 16 | Blood Don | 33  | m   | 41.8 | 65   | 33.1 | 38   | 6.6  | 8.83  | 4.43 | 71   | 24   | 72.7 |
| 17 | Blood Don | 33  | m   | 40.9 | 73   | 17.2 | 22.9 | 10   | 6.98  | 5.22 | 90   | 14.7 | 72.4 |
| 18 | Blood Don | 33  | m   | 45.2 | 88.3 | 32.4 | 31.2 | 10.1 | 9.78  | 5.91 | 102  | 48.5 | 76.5 |
| 19 | Blood Don | 33  | m   | 36.6 | 57.1 | 38.9 | 40.3 | 24.9 | 9.52  | 5.5  | 112  | 27.6 | 69.3 |
| 20 | Blood Don | 33  | m   | 42   | 63.1 | 32.6 | 34.9 | 11.2 | 7.01  | 4.05 | 105  | 19.1 | 68.1 |
| 21 | Blood Don | 33  | m   | 44.3 | 49.8 | 32.1 | 21.6 | 13.1 | 7.44  | 5.59 | 103  | 30.2 | 74   |
| 22 | Blood Don | 33  | m   | 46.7 | 88.3 | 23.4 | 23.9 | 7.8  | 9.42  | 4.62 | 78   | 29.5 | 74.3 |

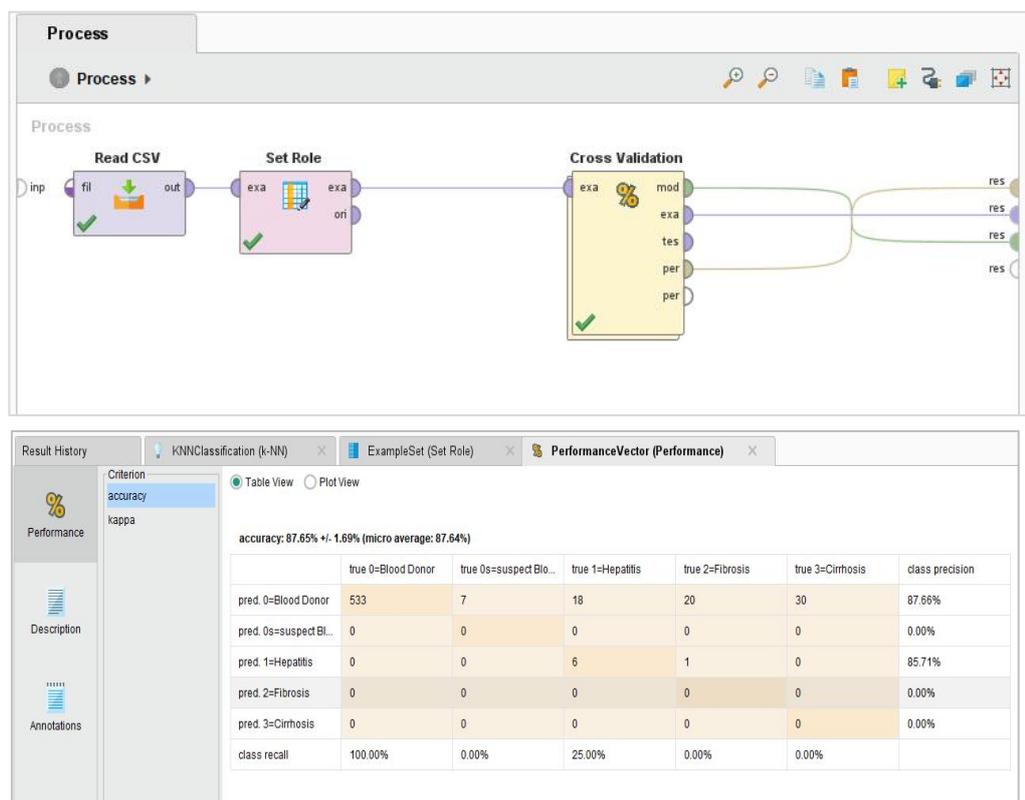
Gambar 4.1 Data Prediksi Penyakit Hati

## 4.2 Pemodelan

Pemilihan dan penerapan teknik pemodelan yang sesuai merupakan langkah penting dalam tahap analisis data. Hal ini dilakukan untuk memastikan bahwa metode yang digunakan dapat memberikan hasil prediksi yang akurat dan relevan dengan tujuan penelitian. Pemodelan pada penelitian ini menggunakan teknik data mining Prediksi.

### 4.2.1 Penelitian Menggunakan Algoritma Knearest Neighbor

Hasil Penerapan data pada Rapidminer untuk Prediksi Penyakit Hati menggunakan algoritma K-NN ditunjukkan pada gambar 4.2 dibawah ini :



Gambar 4.2 Confusion Matrix Hasil Prediksi Penyakit Hati Menggunakan Algoritma K-NN pada Rapidminer

Gambar 4.2 merupakan confusion matrix yang digunakan untuk mengevaluasi performa model klasifikasi yang berbasis algoritma *K-Nearest Neighbors* (K-NN). Confusion matrix adalah tabel yang memberikan gambaran visual mengenai hasil klasifikasi model dengan membandingkan antara nilai prediksi dan nilai sebenarnya dari suatu dataset.

Pada confusion matrix, terdapat empat komponen utama:

- a. True Positive (TP) – Jumlah sampel yang diklasifikasikan dengan benar sebagai kelas positif.
- b. True Negative (TN) – Jumlah sampel yang diklasifikasikan dengan benar sebagai kelas negatif.
- c. False Positive (FP) – Jumlah sampel yang salah diklasifikasikan sebagai kelas positif padahal seharusnya negatif (*Type I error*).
- d. False Negative (FN) – Jumlah sampel yang salah diklasifikasikan sebagai kelas negatif padahal seharusnya positif (*Type II error*).

Dalam eksperimen ini, model K-NN dievaluasi menggunakan confusion matrix, dan hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 87,65%. Nilai akurasi 87,65% mengindikasikan bahwa model K-NN memiliki tingkat keberhasilan yang cukup tinggi dalam mengklasifikasikan data dengan benar, meskipun masih terdapat kesalahan klasifikasi (false positive dan false negative). Namun, akurasi saja tidak cukup untuk menilai kualitas model secara keseluruhan, sehingga perlu

mempertimbangkan metrik lain seperti presisi, recall, dan F1-score agar evaluasi model lebih komprehensif.

#### 4.2.2 Penelitian Menggunakan Algoritma Knearest Neighbor dan PSO

Hasil Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar 4.3 di bawah ini:

Table View Plot View

accuracy: 98.21% +/- 1.20% (micro average: 98.21%)

|                        | true 0=Blood Donor | true 0s=suspect Blo... | true 1=Hepatitis | true 2=Fibrosis | true 3=Cirrhosis | class precision |
|------------------------|--------------------|------------------------|------------------|-----------------|------------------|-----------------|
| pred. 0=Blood Donor    | 533                | 3                      | 0                | 0               | 0                | 99.44%          |
| pred. 0s=suspect BL... | 0                  | 3                      | 0                | 0               | 0                | 100.00%         |
| pred. 1=Hepatitis      | 0                  | 1                      | 21               | 1               | 0                | 91.30%          |
| pred. 2=Fibrosis       | 0                  | 0                      | 2                | 19              | 2                | 82.61%          |
| pred. 3=Cirrhosis      | 0                  | 0                      | 1                | 1               | 28               | 93.33%          |
| class recall           | 100.00%            | 42.86%                 | 87.50%           | 90.48%          | 93.33%           |                 |

Gambar 4.3 Confusion Matrix PSO Dan K-NN

Gambar 4.3 adalah confusion matrix yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan confusion matrix, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. Confusion matrix tersebut berisi hasil dari eksperimen yang dilakukan dengan menggunakan teknik PSO (Particle Swarm Optimization) dan K-NN. Dalam confusion matrix kita dapat melihat PSO dan K-NN hasil akurasi yang dihasilkan adalah 98,21 %.

Perhitungan Akurasi :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

### Total Prediksi Benar (Diagonal Confusion Matrix)

Berdasarkan diagonal utama confusion matrix (jumlah klasifikasi yang benar untuk masing-masing kelas):

- Blood Donor (533)
- Suspected Blood Donor (3)
- Hepatitis (21)
- Fibrosis (19)
- Cirrhosis (28)

Jumlah total klasifikasi yang benar:

$$533+3+21+19+28=604$$

Total Data yang Diuji

Total jumlah sampel dalam confusion matrix adalah jumlah semua elemen dalam tabel:

$$533+3+0+0+0+0+3+0+0+0+0+1+21+1+0+0+2+19+2+0+1+1+28=615$$

Menghitung Akurasi

$$Akurasi = \frac{604}{615} \times 100\%$$

$$Akurasi = 98.21\%$$

Model PSO dan K-NN menghasilkan akurasi 98,21%, menunjukkan performa yang sangat baik dalam klasifikasi dataset ini.

### 4.2.3 Penelitian Menggunakan Algoritma Knearest Neighbor dan Adaboost

Hasil Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar 4.4 di bawah ini:

Table View Plot View

accuracy: 98.70% +/- 1.03% (micro average: 98.70%)

|                        | true 0=Blood Donor | true 0s=suspect Blo... | true 1=Hepatitis | true 2=Fibrosis | true 3=Cirrhosis | class precision |
|------------------------|--------------------|------------------------|------------------|-----------------|------------------|-----------------|
| pred. 0=Blood Donor    | 533                | 3                      | 0                | 0               | 0                | 99.44%          |
| pred. 0s=suspect Bl... | 0                  | 3                      | 0                | 0               | 0                | 100.00%         |
| pred. 1=Hepatitis      | 0                  | 1                      | 22               | 0               | 0                | 95.65%          |
| pred. 2=Fibrosis       | 0                  | 0                      | 1                | 21              | 2                | 87.50%          |
| pred. 3=Cirrhosis      | 0                  | 0                      | 1                | 0               | 28               | 96.55%          |
| class recall           | 100.00%            | 42.86%                 | 91.67%           | 100.00%         | 93.33%           |                 |

Gambar 4.4 Confusion Matrix Adaboost Dan K-NN

Gambar 4.4 adalah confusion matrix yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan confusion matrix, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. Confusion matrix tersebut berisi hasil dari eksperimen yang dilakukan dengan menggunakan teknik Adaboost dan K-NN. Dalam confusion matrix kita dapat melihat Adaboost dan K-NN hasil akurasi yang dihasilkan adalah 98,70 %.

#### 4.2.4 Penelitian Menggunakan Algoritma Knearest Neighbor dan Bagging

Hasil Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar 4.5 di bawah ini:

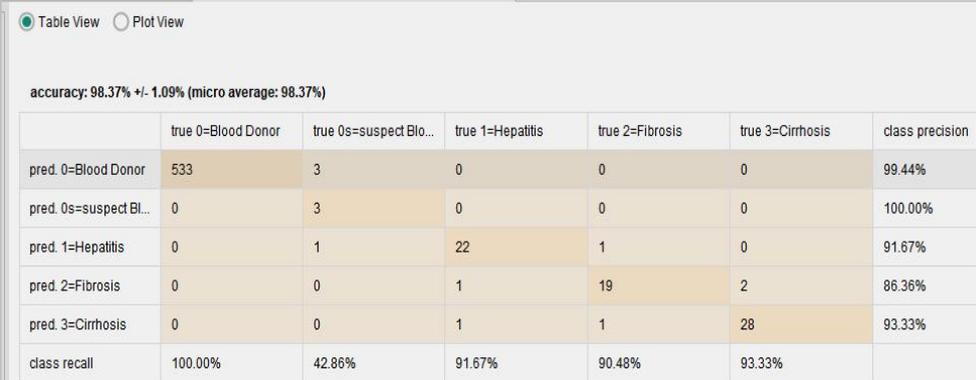


Table View Plot View

accuracy: 98.37% +/- 1.09% (micro average: 98.37%)

|                        | true 0=Blood Donor | true 0s=suspect Blo... | true 1=Hepatitis | true 2=Fibrosis | true 3=Cirrhosis | class precision |
|------------------------|--------------------|------------------------|------------------|-----------------|------------------|-----------------|
| pred. 0=Blood Donor    | 533                | 3                      | 0                | 0               | 0                | 99.44%          |
| pred. 0s=suspect Bl... | 0                  | 3                      | 0                | 0               | 0                | 100.00%         |
| pred. 1=Hepatitis      | 0                  | 1                      | 22               | 1               | 0                | 91.67%          |
| pred. 2=Fibrosis       | 0                  | 0                      | 1                | 19              | 2                | 86.36%          |
| pred. 3=Cirrhosis      | 0                  | 0                      | 1                | 1               | 28               | 93.33%          |
| class recall           | 100.00%            | 42.86%                 | 91.67%           | 90.48%          | 93.33%           |                 |

Gambar 4.5 Confusion Matrix Bagging Dan K-NN

Gambar 4.5 adalah confusion matrix yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan confusion matrix, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. Confusion matrix tersebut berisi hasil dari eksperimen yang dilakukan dengan menggunakan teknik Bagging dan K-NN. Dalam confusion matrix kita dapat melihat Bagging dan K-NN hasil akurasi yang dihasilkan adalah 98,37 %.

#### 4.2.5 Perbandingan Menggunakan Optimasi Dan Tidak

##### Menggunakan Optimasi

Perbandingan hasil eksperimen yang tidak menggunakan Optimasi dan eksperimen yang menggunakan Optimasi ditunjukkan pada tabel dibawah ini:

Tabel 4.1 Perbandingan Menggunakan Optimasi Dan Tidak Menggunakan Optimasi

| <b>Penelitian</b>               | <b>Tingkat Akurasi</b> |
|---------------------------------|------------------------|
| K-Nearest Neighbor              | 87,65%                 |
| K-Nearest Neighbor dan PSO      | 98,21%                 |
| K-Nearest Neighbor dan Adaboost | 98,70%                 |
| K-Nearest Neighbor dan Bagging  | 98,37%                 |

Hasil penelitian menunjukkan adanya peningkatan akurasi yang signifikan pada model prediksi penyakit hati setelah menerapkan metode optimasi dan teknik penggabungan dibandingkan dengan model dasar K-Nearest Neighbor (KNN). Model KNN tanpa optimasi menghasilkan akurasi sebesar 87,65%. Dengan menggunakan metode Particle Swarm Optimization (PSO) untuk mengoptimalkan parameter, akurasi meningkat menjadi 98,21%, yang berarti terdapat peningkatan sebesar 10,56%. Selanjutnya, penerapan teknik Adaboost memberikan

akurasi tertinggi, yaitu 98,70%, dengan peningkatan sebesar 11,05% dibandingkan model dasar. Sementara itu, teknik Bagging juga menunjukkan peningkatan yang signifikan dengan akurasi sebesar 98,37%, yaitu peningkatan 10,72% dibandingkan KNN tanpa optimasi.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa semua metode yang diterapkan berhasil meningkatkan performa model dalam memprediksi penyakit hati. Dari ketiga pendekatan tersebut, Adaboost memberikan akurasi terbaik, menjadikannya metode yang paling unggul dalam penelitian ini yaitu 98,70%, dengan peningkatan sebesar 11,05%.