

BAB IV HASIL DAN PEMBAHASAN

4.1 *Data Selection* (Seleksi Data)

Pada penelitian ini data yang digunakan adalah dataset berupa record yang terdiri dari data siswa SMKN 1 Penawartama 2018-2021. Data terdiri atas NISN, Nama, Nilai, Status Ekonomi, Kedisiplinan, Presensi, Prestasi

Row No.	Prestasi	No	Nis	Status Ekon...	Kedisiplinan	Presensi	Nilai
1	tidak ada	1	181444023	mampu	kurang baik	baik sekali	baik sekali
2	ada	2	181444028	mampu	baik	baik sekali	baik sekali
3	tidak ada	3	181444035	mampu	kurang baik	baik sekali	baik sekali
4	tidak ada	4	181444037	mampu	kurang baik	baik sekali	baik sekali
5	tidak ada	5	181444040	mampu	kurang baik	baik sekali	baik sekali
6	tidak ada	6	181444043	mampu	kurang baik	baik sekali	baik sekali
7	tidak ada	7	181444048	mampu	kurang baik	baik sekali	baik sekali
8	tidak ada	8	181444050	mampu	kurang baik	baik sekali	baik sekali
9	tidak ada	9	181444066	mampu	kurang baik	baik sekali	baik sekali
10	tidak ada	10	181444069	mampu	kurang baik	baik sekali	baik sekali
11	tidak ada	11	181444084	mampu	kurang baik	baik sekali	baik sekali
12	tidak ada	12	18144088	mampu	kurang baik	baik sekali	baik sekali
13	tidak ada	13	181444089	mampu	kurang baik	baik sekali	baik sekali
14	ada	14	181444116	?	kurang baik	?	?
15	?	15	181444118	?	kurang baik	?	baik sekali

Gambar 4.1 *Data Selection*

4.2 *Pre-Processing/Cleaning* (Pemilihan Data)

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Dalam penelitian ini, proses *cleaning* yang dilakukan

adalah menghilangkan kolom nilai, karena dalam penelitian ini yang diprediksi adalah nilai (prestasi).

Gambar 4.2 Data Cleaning

Row No.	Prestasi	No	Nis	Status Ekon...	Kedisiplinan	Presensi	Nilai
1	tidak ada	1	181444023	mampu	kurang baik	baik sekali	baik sekali
2	ada	2	181444028	mampu	baik	baik sekali	baik sekali
3	tidak ada	3	181444035	mampu	kurang baik	baik sekali	baik sekali
4	tidak ada	4	181444037	mampu	kurang baik	baik sekali	baik sekali
5	tidak ada	5	181444040	mampu	kurang baik	baik sekali	baik sekali
6	tidak ada	6	181444043	mampu	kurang baik	baik sekali	baik sekali
7	tidak ada	7	181444048	mampu	kurang baik	baik sekali	baik sekali
8	tidak ada	8	181444050	mampu	kurang baik	baik sekali	baik sekali
9	tidak ada	9	181444066	mampu	kurang baik	baik sekali	baik sekali
10	tidak ada	10	181444069	mampu	kurang baik	baik sekali	baik sekali
11	tidak ada	11	181444084	mampu	kurang baik	baik sekali	baik sekali
12	tidak ada	12	18144088	mampu	kurang baik	baik sekali	baik sekali
13	tidak ada	13	181444089	mampu	kurang baik	baik sekali	baik sekali

4.3 Transformation (Transformasi Data)

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Tahap ini mengubah data ke dalam format yang sesuai untuk kebutuhan data mining.

1. Transformasi Status Ekonomi

Transformasi ini digunakan untuk menentukan range Status Ekonomi, dengan kategori seperti pada tabel 4.1.

Tabel 4.1 Transformasi Status Ekonomi

Status Ekonomi	Range
Kurang	< 1.500.000
Cukup	1.500.000-2.500.000
Mampu	>2.500.000

2. . Transformasi Kedisiplinan

Transformasi ini digunakan untuk menentukan range Kedisiplinan, dengan kategori seperti pada table 4.2.

Tabel 4.2 Transformasi Kedisiplinan

Kedisiplinan	Range
Kurang Disiplin	Tidak mematuhi tata tertib>70%
Cukup Disiplin	Kadang-kadang mematuhi tata tertib70%
Disiplin	Selalu mematuhi tata tertib70%

3. . Transformasi Presensi

Transformasi ini digunakan untuk menentukan range Presensi, dengan kategoriseperti pada tabel 4.3.

Tabel 4.3 Transformasi Presensi

Presensi	Range
Tidak Rajin	Kehadiran < 70%
Rajin	Kehadiran > 70%

4.4 Data Mining (Data Mining)

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tujuan dari

tahap ini adalah untuk menganalisis *database* sesuai algoritma yang digunakan.
Pada tabel data training berikut hanya ditampilkan 606 data

Tabel 4.4 Data Training

No	Nis	Status Ekonomi	Kedisiplinan	Presensi	Nilai	Prestasi
1	181444023	mampu	kurang baik	baik sekali	baik sekali	tidak ada
2	181444028	mampu	baik	baik sekali	baik sekali	ada
3	181444035	mampu	kurang baik	baik sekali	baik sekali	tidak ada
4	181444037	mampu	kurang baik	baik sekali	baik sekali	tidak ada
5	181444040	mampu	kurang baik	baik sekali	baik sekali	tidak ada
6	181444043	mampu	kurang baik	baik sekali	baik sekali	tidak ada
7	181444048	mampu	kurang baik	baik sekali	baik sekali	tidak ada
8	181444050	mampu	kurang baik	baik sekali	baik sekali	tidak ada
9	181444066	mampu	kurang baik	baik sekali	baik sekali	tidak ada
10	181444069	mampu	kurang baik	baik sekali	baik sekali	tidak ada
11	181444084	mampu	kurang baik	baik sekali	baik sekali	tidak ada
12	181444088	mampu	kurang baik	baik sekali	baik sekali	tidak ada
13	181444089	mampu	kurang baik	baik sekali	baik sekali	tidak ada
14	181444116		kurang baik	baik sekali	baik sekali	ada
15	181444118		kurang baik	baik sekali	baik sekali	ada

Tree

No > 604.500: ada {tidak ada=1, ada=2}

No ≤ 604.500

| No > 2.500

| | No > 21.500

| | | No > 539.500

| | | | No > 545.500

| | | | | Kedisiplinan = baik: tidak ada {tidak ada=4, ada=0}

| | | | | Kedisiplinan = baik sekali: ada {tidak ada=1, ada=4}

| | | | | Kedisiplinan = kurang baik

| | | | | | Nis > 46002753.500: tidak ada {tidak ada=42, ada=2}

| | | | | | Nis ≤ 46002753.500

| | | | | | | No > 567: ada {tidak ada=0, ada=2}

| | | | | | | No ≤ 567: tidak ada {tidak ada=4, ada=0}

| | | | | No ≤ 545.500: ada {tidak ada=0, ada=6}

| | | | No ≤ 539.500

| | | | | No > 74.500

| | | | | | No > 507.500

| | | | | | | No > 511.500: tidak ada {tidak ada=28, ada=0}

| | | | | | | No ≤ 511.500

| | | | | | | | No > 509.500: ada {tidak ada=0, ada=2}

| | | | | | | | No ≤ 509.500: tidak ada {tidak ada=1, ada=1}

| | | | | | No ≤ 507.500

| | | | | | | Kedisiplinan = baik

| | | | | | | | No > 172.500: ada {tidak ada=0, ada=4}

| | | | | | | | No ≤ 172.500: tidak ada {tidak ada=2, ada=0}

| | | | | | | | Kedisiplinan = baik sekali: tidak ada {tidak ada=2, ada=0}

| | | | | | | | Kedisiplinan = kurang baik

| | | | | | | No > 312.500: tidak ada {tidak ada=194, ada=0}
 | | | | | | | No ≤ 312.500
 | | | | | | | No > 310.500: ada {tidak ada=0, ada=2}
 | | | | | | | No ≤ 310.500: tidak ada {tidak ada=225, ada=4}
 | | | | No ≤ 74.500
 | | | | | No > 70.500: ada {tidak ada=0, ada=4}
 | | | | | No ≤ 70.500
 | | | | | No > 33.500
 | | | | | | | No > 36.500: tidak ada {tidak ada=32, ada=2}
 | | | | | | | No ≤ 36.500: ada {tidak ada=0, ada=3}
 | | | | | | | No ≤ 33.500: tidak ada {tidak ada=12, ada=0}
 | | No ≤ 21.500
 | | | Kedisiplinan = baik sekali: ada {tidak ada=0, ada=5}
 | | | Kedisiplinan = kurang baik: tidak ada {tidak ada=14, ada=0}
 | No ≤ 2.500: tidak ada {tidak ada=1, ada=1}

Mencari entropy dan gain untuk menentukan node root awal, diketahui dari proses diatas bahwa dengan menggunakan tools rapidminer node root awal adalah age. Disini kita akan melakukan pengukuran nilai entropy dan gain dengan cara menghitung manual. Proses ini adalah untuk pembuktian apakah tools rapidminer berjan dengan baik

Hasil Akurasi Decision Tree C4.5 dan Split Validation, pada pembagian data sampling 70% dan data testing 98.64% maka didapatkan akurasi sebesar 98,02%

accuracy: 98.02%

	true tidak ada	true ada	class precision
pred. tidak ada	561	10	98.25%
pred. ada	2	34	94.44%
class recall	99.64%	77.27%	

Gambar 4. 5 Nilai Akurasi Split Validation

Menghitung akurasi:

Rumus:

$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + FP + TP)}$$

$$\text{Akurasi} = \frac{(561+34)}{(561+34+10+2)} = \frac{595}{607} = 98,02\%$$

Berikut adalah hasil precision yaitu mendapatkan 94,44% seperti gambar dibawah ini

precision: 94,44% (positive class: ada)

	true tidak ada	true ada	class precision
pred. tidak ada	561	10	98,25%
pred. ada	2	34	94,44%
class recall	99,64%	77,27%	

Gambar 4. 6 Nilai Precision Dengan Split Validation

Menghitung Precision, Rumus

$$precision = \frac{TP}{(TP+FP)}$$

$$p(1) = \frac{561}{(561+10)} = \frac{561}{571} = 98,25\% \quad ; p(0) = \frac{34}{(34+2)} = \frac{34}{36} = 94,44\%$$

recall: 77,27% (positive class: ada)

	true tidak ada	true ada	class precision
pred. tidak ada	561	10	98,25%
pred. ada	2	34	94,44%
class recall	99,64%	77,27%	

Gambar 4. 7 Nilai Confusion Matrix Dengan Split Validation

Menghitung Recall, Rumus : $Recall = \frac{TP}{(TP+FN)}$

$$R(1) = \frac{561}{(561+2)} = \frac{561}{569} = 99,54\% \quad R(0) = \frac{34}{(34+10)} = \frac{34}{44} = 77,27\%$$



Gambar 4.8 Nilai AUC dengan *Decision Tree*

PerformanceVector

PerformanceVector:
accuracy: 98.02%

ConfusionMatrix:

True:	tidak	ada	ada
tidak	ada:	561	10
ada:	2	34	

precision: 94.44% (positive class: ada)

ConfusionMatrix:

True:	tidak	ada	ada
tidak	ada:	561	10
ada:	2	34	

recall: 77.27% (positive class: ada)

ConfusionMatrix:

True:	tidak	ada	ada
tidak	ada:	561	10
ada:	2	34	

AUC (optimistic): 0.984 (positive class: ada)

AUC: 0.963 (positive class: ada)

AUC (pessimistic): 0.941 (positive class: ada)

Tabel 4. 6 Hasil dari Decision Tree dan Split Validation

No	Algoritma C45	Hasil Nilai
1	Accuracy	98,02%
2	Precision	94,44%
3	Recall / Confusion Matrix	77.72%
4	AUC	0.96