BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Perhitungan *K-Nearest Neightbors*

Dalam penelitian ini akan dilakukan beberapa tahapan penelitian, sebagai berikut:

1. Dataset

   Data yang akan digunakan dalam analisis adalah data dari tahun 2021 dan 2022, yang diperoleh dari STMIK Surya Intan Kotabumi. Data tersebut mencakup 95 mahasiswa Teknik informatika dan Sistem informasi dan berisi atribut-atribut seperti Indeks Prestasi Semester (IPS), Waktu Studi, Absensi, dan Jenis Kelamin [7].

Tabel 4. 1 Sampel dataset Mahasiswa

No	Nama	IPS 1	IPS 2	IPS 3	IPS 4	Waktu studi	Absensi	Jenis Kelamin	Output
1	Ridwan	3,7	3,7	3,7	3,8	4	80	Laki-laki	Standar
2	Lilia	3,6	3,7	3,7	3,7	4	78	Perempuan	Standar
3	Wawan	3,8	3,8	3,8	3,9	3	95	Laki-laki	Cepat
4	Rizki	3,7	3,7	3,8	3,7	4	80	Laki-laki	Standar
5	Saskia	3,7	3,7	3,8	3,8	4	85	Perempuan	Standar
6	Fatur	3,2	3,2	3,3	3,5	4	68	Laki-laki	Standar
7	Deri	3,3	3,4	3,4	3,3	4	70	Laki-laki	Standar
8	Mawar	3,5	3,6	3,5	3,7	4	75	Perempuan	Standar
9	Bunga	3,7	3,8	3,8	3,8	4	88	Perempuan	Standar

10	Kiki	3,9	3,9	3,8	3,9	3	95	Laki-laki	Cepat

2. *Pre-processing*

   Pemilihan atribut dilakukan menggunakan fitur "*Select Attribute*" dan "*Replace Missing*" untuk menghindari adanya data kosong dan *noise*. Pada tahap ini, data dibersihkan dari nilai yang hilang dan *noise* agar tidak menimbulkan masalah saat diolah. Setelah melalui proses *pre-processing*, diperoleh 665 catatan data yang siap digunakan.

Tabel 4. 2 Attribute Indeks Prestasi Semester (IPS)

Nama	IPS 1	IPS 2	IPS 3	IPS 4
Ridwan	3,7	3,7	3,7	3,8
Lilia	3,6	3,7	3,7	3,7
Wawan	3,8	3,8	3,8	3,9
Rizki	3,7	3,7	3,8	3,7
Saskia	3,7	3,7	3,8	3,8
Fatur	3,2	3,2	3,3	3,5
Deri	3,3	3,4	3,4	3,3
Mawar	3,5	3,6	3,5	3,7
Bunga	3,7	3,8	3,8	3,8
Kiki	3,9	3,9	3,8	3,9

Tabel 4. 3 Attribute Waktu Studi

Nama	Waktu Studi
Ridwan	4 Semester
Lilia	4 Semester
Wawan	3 Semester
Rizki	4 Semester
Saskia	4 Semester
Fatur	4 Semester
Deri	4 Semester
Mawar	4 Semester
Bunga	4 Semester
Kiki	3 Semester

Tabel 4. 4 Attribute Absensi

Nama	Absensi
Ridwan	80
Lilia	78
Wawan	95
Rizki	80
Saskia	85
Fatur	68
Deri	70
Mawar	75
Bunga	88
Kiki	95

Tabel 4. 5 Attribute Jenis Kelamin

Nama	Jenis Kelamin
Ridwan	Laki-laki
Lilia	Perempuan
Wawan	Laki-laki
Rizki	Laki-laki
Saskia	Perempuan
Fatur	Laki-laki
Deri	Laki-laki
Mawar	Perempuan
Bunga	Perempuan
Kiki	Laki-laki

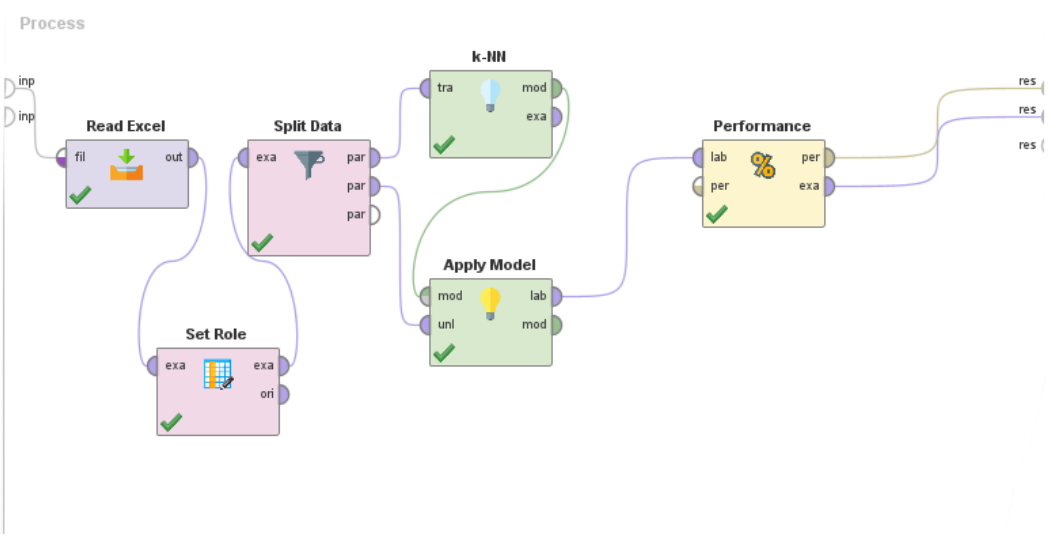
3. Pengujian Metode dengan *10-fold Cross Validation*

   Data diuji menggunakan metode *k-Nearest Neighbor*. Dataset dibagi menjadi dua bagian: (1) data *training* dan (2) data *testing*, menggunakan *10-fold cross validation*. Kemudian, data *training* diklasifikasikan dengan *k-Nearest Neighbor* untuk menghasilkan akurasi dalam masa studi mahasiswa. Pengujian menggunakan teknik *validasi* bertujuan untuk mendapatkan nilai akurasi. Dataset diuji dengan metode yang diusulkan menggunakan aplikasi RapidMiner 10.1.

4. *Result*

   Metode *K-Nearest Neighbor* (K-NN) diterapkan dalam penelitian ini untuk menentukan nilai akurasi kinerja siswa berdasarkan tingkat

kelulusan. Pengolahan dan pengujian data pada metode K-NN dilakukan menggunakan RapidMiner sebagai perangkat lunak implementasi. Penelitian ini melibatkan pengujian dengan metode K-NN, di mana percobaan dilakukan sebanyak 10 kali dengan nilai k dari 1 hingga 10, menggunakan dataset masa studi mahasiswa yang terdiri dari 665 catatan.



Gambar 4. 1 Pengolahan Data dengan Rapidminer

Tabel 4. 6 Perbandingan K-Nearest Neightbor

Nilai k	K-Terdekat
1	100%
2	100%
3	80%
4	80%
5	80%
6	80%
7	80%
8	80%
9	80%
10	80%
Rata-rata	94%

Hasil percobaan menunjukkan bahwa menggunakan metode *K-Nearest Neighbor* (K-NN), diperoleh akurasi rata-rata sebesar 94%. Hal ini menunjukkan bahwa K-NN dapat meningkatkan akurasi klasifikasi kinerja mahasiswa dengan baik. Selain itu, perlu dicatat bahwa tingkat akurasi dapat dipengaruhi oleh berbagai faktor, termasuk jumlah dan kualitas data yang digunakan, serta pengaturan parameter seperti jumlah tetangga terdekat (k) dalam KNN atau klastering data dalam metode lainnya.

4.2 Perhitungan Naïve Bayes

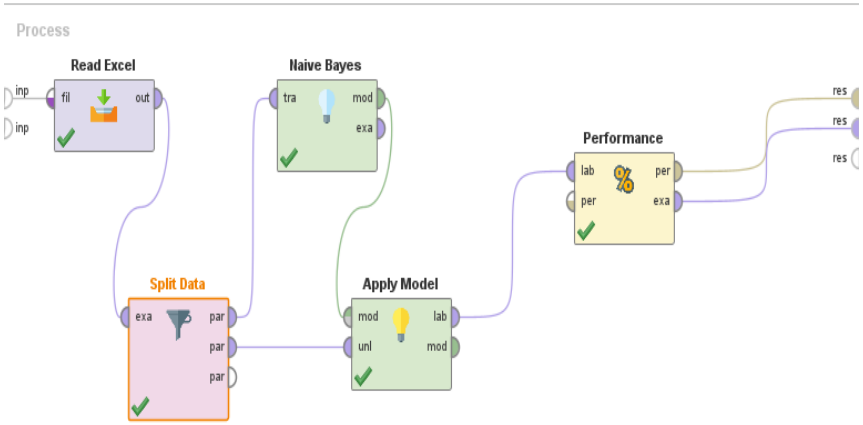
Dengan menggunakan data mahasiswa, kita dapat menentukan tingkat ketepatan waktu mahasiswa yang masih dalam masa studi untuk lulus tepat waktu. Berdasarkan hasil pengumpulan data, beberapa sampel data dari beberapa mahasiswa telah berhasil diperoleh [8].

Tabel 4. 7 Sample data Mahasiswa

No	Nama	IPS 1	IPS 2	IPS 3	IPS 4	Waktu studi	Absensi	Jenis Kelamin	Output
1	Ridwan	3,7	3,7	3,7	3,8	4	80	Laki-laki	Standar
2	Lilia	3,6	3,7	3,7	3,7	4	78	Perempuan	Standar
3	Wawan	3,8	3,8	3,8	3,9	3	95	Laki-laki	Cepat
4	Rizki	3,7	3,7	3,8	3,7	4	80	Laki-laki	Standar
5	Saskia	3,7	3,7	3,8	3,8	4	85	Perempuan	Standar
6	Fatur	3,2	3,2	3,3	3,5	4	68	Laki-laki	Standar
7	Deri	3,3	3,4	3,4	3,3	4	70	Laki-laki	Standar
8	Mawar	3,5	3,6	3,5	3,7	4	75	Perempuan	Standar
9	Bunga	3,7	3,8	3,8	3,8	4	88	Perempuan	Standar
10	Kiki	3,9	3,9	3,8	3,9	3	95	Laki-laki	Cepat

1. Pengujian dengan Rapidminer

   Berikut adalah hasil uji data dengan bantuan alat RapidMiner dengan menggunakan data yang sama, yang dapat dilihat pada Gambar 4.2



Gambar 4. 2 Pengujian Naive Bayes dengan Rapidminer

2. Hasil pengujian Naïve Bayes dengan Rapidminer

   Berikut ini adalah hasil uji coba dengan menggunakan alat RapidMiner. Hasil ini merupakan hasil uji data yang dapat dilihat pada Gambar 4.3.

Table View Plot View

accuracy: 100.00%

	true Standar	true Cepat	class precision
pred. Standar	38	0	100.00%
pred. Cepat	0	9	100.00%
class recall	100.00%	100.00%	

Gambar 4. 3 Hasil Perhitungan Naive Bayes

3. Hasil Uji Validasi

Berikut ini adalah hasil dari uji coba penerapan model klasifikasi Naïve Bayes untuk prediksi mahasiswa tepat waktu dengan menggunakan alat RapidMiner. Uji coba dilakukan dengan 95 record data dan melakukan 5 kali uji coba menggunakan confusion matrix. Hasilnya dapat dilihat pada Tabel 4.8.

Tabel 4. 8 Hasil Uji Validasi Naive Bayes

Training	Testing	Pembagian Data	Hasil Presentase
8	2	90 : 10	100%
15	4	80 : 20	100%
23	6	70 : 30	100%
30	8	60 : 40	100%
39	9	50 : 50	100%

Berdasarkan hasil pengujian di atas, didapatkan akurasi yang stabil dengan contoh menggunakan 50% data training dan 50% data testing dari total 95 record data. Pengujian ini menghasilkan akurasi 100% terhadap penerapan model klasifikasi Naïve Bayes untuk prediksi mahasiswa tepat waktu. Berikut ini adalah perhitungan manual terhadap hasil pengujian validasi.

a. Hasil *recall*

Prediksi Standar : $\frac{38}{38+0} = \frac{38}{38} = 1\% \ (100\%)$

Prediksi Cepat : $\frac{9}{0+9} = \frac{9}{9} = 1\% \ (100\%)$

b. Hasil Precision

Prediksi Standar : $\frac{38}{0+38} = \frac{38}{38} = 1 = 100\%$

Prediksi Cepat : $\frac{9}{0+9} = \frac{9}{9} = 1\% \ (100\%)$

c. Hasil Akurasi

$$\frac{38+9}{38+0+0+9}$$

Accuracy :

Accuracy : $\frac{47}{47}$

Accuracy : 1 (100%)

Hasil pengujian menunjukkan tingkat akurasi sebesar 100% menggunakan algoritma Naïve Bayes, dengan pembagian data menjadi 50% untuk pelatihan dan 50% untuk pengujian, dari total 95 rekaman data yang tersedia.

4.3 Perbandingan K-Nearest Nightbor dan Naïve Bayes

Perbandingan antara algoritma K-NN dan Naive Bayes berdasarkan akurasi yang ada :

Tabel 4. 9 Perbandingan KNN dan Naive Bayes

Metode	Akurasi	Kelebihan	Kekurangan
Naïve bayes	100%	1. Efisien dalam waktu komputasi karena mengasumsikan independensi fitur 2. Cocok untuk dataset dengan jumlah fitur besar	1. Sensitif terhadap fitur yang sebenarnya tidak independen. 2. Tidak cocok untuk dataset di mana hubungan antar fitur sangat kompleks.
K-Nearest Nightbor	94%	1. Sederhana untuk dipahami dan diimplementasikan. 2. Cocok untuk dataset dengan persebaran data tidak teratur.	1. Memerlukan penyimpanan data yang besar karena klasifikasi membutuhkan data training lengkap. 2. Sensitif terhadap skala fitur dan pemilihan parameter K yang tepat

Dalam kasus ini, Naive Bayes memiliki akurasi yang lebih tinggi (100%) dibandingkan KNN (94%), namun keduanya memiliki kelebihan dan keterbatasan masing-masing tergantung pada sifat dataset yang digunakan.