

BAB IV

HASIL DAN PEMBAHASAN

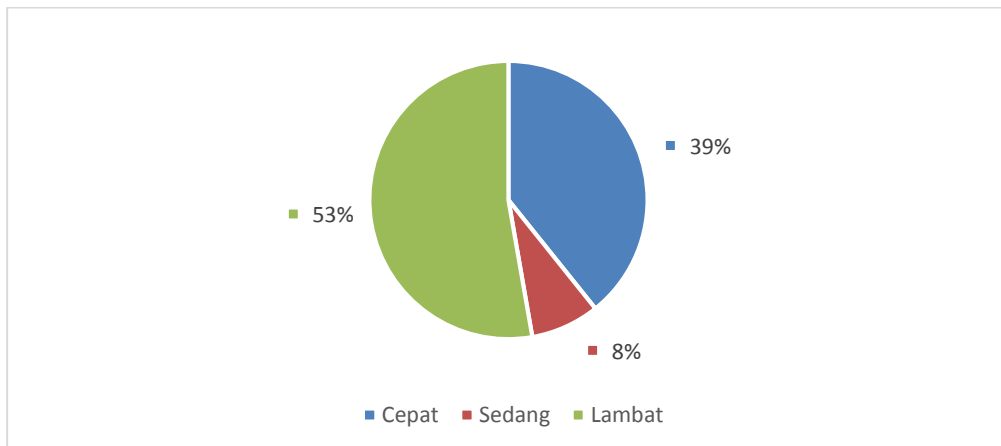
4.1. Data

Sebelum melakukan proses klasifikasi data, terlebih dahulu dilakukan pengumpulan data. Data yang akan diolah dalam pembahasan ini adalah data siswa di SMK Islam Al-Barokah Poncowarno. Data siswa yang digunakan adalah data siswa Alumni. Atribut-atribut data yang digunakan adalah Jurusan, Nilai rata-rata, Prestasi Akademik/Non-Akademik dan Ekstrakurikuler. Data diperoleh dari SMK Islam Al-Barokah Poncowarno yaitu data alumni 5 tahun terakhir (Tahun 2017 - 2021).

4.2. Deskripsi Data

Pada Proses Tahap Pertama dalam perbandingan metode Naïve Bayes, K-NN adalah menyiapkan data training. Data training yang diambil ada 400 data siswa alumni 5 tahun terakhir (Tahun 2017 - 2021) yang berasal dari data arsip sekolah, setelah dilakukan analisis ternyata data *tracer study* yang ada belum tersinkron dengan data dapodik, data yang di dapat dari arsip sekolah sangat sederhana. Berikut merupakan deskripsi data dari alumni SMK Islam Al-Barokah Poncowarno selama 5 tahun terakhir (Tahun 2017 - 2021).

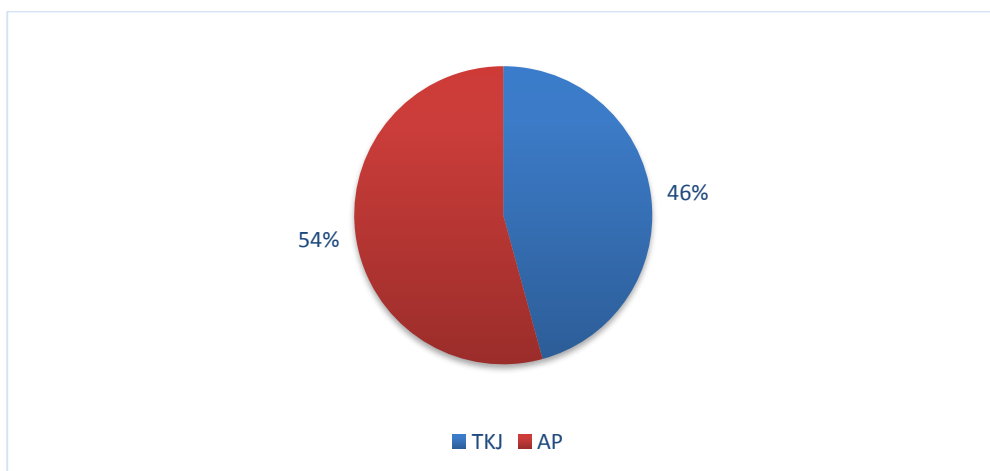
1. Persentase data alumni SMK Al-Barokah Poncowarno berdasarkan status lama mendapatkan pekerjaan dengan kelas dibawah 12 bulan dan di atas 12 bulan dapat dilihat pada gambar 4.1.



Gambar 4.1 Persentase Status Lama Mendapatkan Pekerjaan.

Pada gambar 4.1 diatas menunjukkan persentase status lama mendapatkan pekerjaan dari alumni SMK Islam Al-Barokah Poncowarno dari data 5 tahun terakhir, dimana total data adalah 400 data alumni. Diketahui 39 % atau 157 alumni dengan lama mendapatkan pekerjaan kurang dari 12 bulan atau cepat , sedangkan 8 % atau 32 alumni dengan lama mendapatkan pekerjaan antara 13 bulan sampai dengan 18 bulan atau sedang, dan 211 alumni lainnya atau 53 % alumni dengan waktu tunggu mendapat pekerjaan lebih dari 18 bulan atau lambat.

2. Persentase data alumni SMK Islam Al-Barokah Poncowarno berdasarkan jurusan yaitu TKJ dan AP dapat dilihat pada gambar 4.2.



Gambar 4.2 Persentase Alumni Per-Jurusan.

Pada gambar 4.2 diatas menunjukkan persentase jurusan dari alumni SMK Islam Al-Barokah Poncowarno dari data 5 tahun terakhir dimana total data adalah 400 data alumni. Diketahui 54 % atau 217 alumni dengan jurusan AP, sedangkan 46% lainnya atau 183 alumni jurusan TKJ.

3. Persentase data alumni SMK Islam Al-Barokah Poncowarno berdasarkan Jenis Kelamin dapat dilihat pada gambar 4.3.



Gambar 4.3 Persentasi Alumni Berdasarkan Jenis Kelamin.

Pada gambar 4.3 diatas menunjukkan persentase berdasarkan Jenis Kelamin. dari alumni SMK Islam Al-Barokah Poncowarno dari data 5 tahun terakhir dimana total data adalah 400 data alumni. Diketahui 59 % atau 237 alumni dengan Jenis Kelamin Perempuan, sedangkan 41% lainnya atau 133 alumni dengan Jenis Kelamin laki-laki.

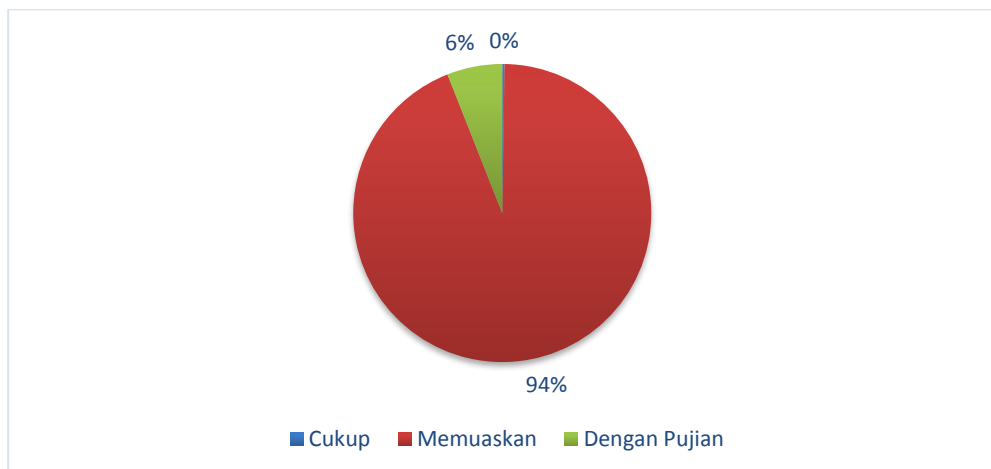
4. Persentase data alumni SMK Islam Al-Barokah Poncowarno berdasarkan Keikutsertaan Kegiatan Ekstrakurikuler dapat dilihat pada gambar 4.4.



Gambar 4.4 Persentasi Alumni Berdasarkan Keikutsertaan Ekstrakurikuler.

Pada gambar 4.4 diatas menunjukkan persentase berdasarkan Keikutsertaan Ekstrakurikuler dari alumni SMK Islam Al-Barokah Poncowarno dari data 5 tahun terakhir dimana total data adalah 400 data alumni. Diketahui 89 % atau 357 alumni mengikuti kegiatan ekskul, sedangkan 11% lainnya atau 43 alumni tidak mengikuti ekstrakurikuler.

5. Persentase data alumni SMK Islam Al-Barokah Poncowarno berdasarkan Rata-Rata Nilai dapat dilihat pada gambar 4.5.



Gambar 4.5 Persentasi Berdasarkan Rata – Rata Nilai

Pada gambar 4.5 diatas menunjukkan persentase berdasarkan Rata-rata Nilai dari alumni SMK Islam Al-Barokah Poncowarno dari data 5 tahun terakhir dimana total data adalah 400 data alumni. Diketahui 94 % atau 376 alumni dengan rata-rata nilai memuaskan, sedangkan 6% lainnya atau 34 alumni dengan rata-rata nilai dengan pujian.

4.3. Pembagian data *training* dan data *testing*

Pada penelitian ini, menggunakan teknik split data (hold-out validation) karena beberapa alasan tertentu. Pertama, dataset yang digunakan mungkin memiliki ukuran yang cukup besar, sehingga melakukan cross-validation dengan jumlah lipatan yang tinggi akan memakan banyak waktu dan sumber daya komputasi. Dengan teknik split data, kita hanya perlu membagi dataset menjadi dua bagian, yaitu dataset pelatihan dan dataset pengujian, yang lebih efisien dalam hal waktu dan pengolahan data.

Kedua, dalam beberapa kasus, model yang digunakan dapat relatif stabil dan tidak terlalu sensitif terhadap variasi data. Oleh karena itu, evaluasi model dengan teknik split data sudah cukup dapat memberikan hasil yang konsisten dan memadai untuk tujuan penelitian.

Meskipun cross-validation dapat memberikan evaluasi yang lebih komprehensif dan meminimalkan risiko overfitting, untuk penelitian ini, teknik split data dipilih karena alasan efisiensi komputasi dan stabilitas model. Dengan demikian, teknik split data telah memberikan evaluasi yang memadai untuk perbandingan antara algoritma K-Nearest Neighbor (K-NN) dan Naïve Bayes dalam klasifikasi waktu tunggu alumni dalam memperoleh pekerjaan pada studi kasus SMK Islam Al-Barokah Poncowarno.

Proses pertama sebelum melakukan analisis klasifikasi data *tracer* alumni SMK Islam Al-Barokah Poncowarno dengan menggunakan metode *naive* Bayes dan K-NN, data *tracer* alumni SMK Islam Al-Barokah Poncowarno dibagi menjadi data training dan data testing dengan proporsi pembagian 80:20 Dalam pembelajaran mesin dan pemodelan statistik, semakin besar ukuran dataset pelatihan, semakin baik model dapat memahami pola dan variabilitas dalam data.

Dengan memberikan 80% data untuk pelatihan, model memiliki lebih banyak contoh data untuk belajar dan menyesuaikan diri dengan berbagai situasi. Pembagian 80%-20% juga memberikan keseimbangan antara kualitas pelatihan model dan keefisienan waktu dan komputasi yang dibutuhkan untuk melatih model. Jika kita membagi terlalu banyak data untuk pengujian, waktu dan sumber daya komputasi yang diperlukan untuk melatih model mungkin meningkat secara signifikan.

Adapun perhitungan pembagian data training dan data testing sebagai berikut:

$$\begin{aligned}\text{Jumlah data training} &= \text{Proporsi data training} \times N \\ &= 80\% \times 400 \\ &= 320 \\ \text{Jumlah Data} &= N - \text{Jumlah data training} \\ \text{Testing} &= 400 - 320 \\ &= 80\end{aligned}$$

Berdasarkan perhitungan diperoleh hasil bahwa 320 data pertama hasil randomisasi data untuk masing-masing variabel penelitian berfungsi sebagai data training yang dapat dilihat pada Lampiran. Data training ini akan digunakan dalam proses perhitungan pada metode naive Bayes serta perhitungan jarak pada metode K-NN, sedangkan 80 data terakhir untuk masing-masing variabel penelitian berfungsi sebagai data testing yang nantinya digunakan untuk melakukan proses klasifikasi dan pengukuran tingkat akurasi.

4.4. Analisis Lama masa Tunggu alumni mendapatkan pekerjaan Dengan Naïve Bayes.

Pada proses klasifikasi metode naive Bayes terdapat tiga alur yaitu menghitung probabilitas prior, menghitung probabilitas setiap variabel bebas pada setiap kelompok dan menghitung perkalian probabilitas prior dan probabilitas variabel bebas pada setiap kelompok. Adapun data yang

digunakan untuk perhitungan klasifikasi adalah menggunakan data training 80% sebanyak 320 data, sedangkan data testing 20% sebanyak 80 data. Adapun tahapan-tahapan klasifikasi metode *naive* Bayes dalam klasifikasi lama masa tunggu alumni mendapatkan pekerjaan r sebagai berikut:

1. Menghitung Probabilitas *prior* (awal) setiap kelas.

Tahapan klasifikasi metode *naive* Bayes yang pertama yaitu menghitung probabilitas *prior* pada ketiga kelompok menggunakan data *training*. Tabel 4.1 Menyajikan data Hasil nilai probabilitas *prior* setiap kelas yang didapat.

Tabel 4.1. Probabilitas Setiap Kelas

	Cepat	Sedang	Lama
$P(\text{cepat, sedang, lama})$	0.416	0.078	0.506

Pada tabel 4.1 merupakan hasil probabilitas setiap kelas / kelompok, ini dapat kita hitung dengan persamaan (2.4) seperti pada Probabilitas cepat , dapat kita hitung sebagai berikut:

$P(Y_{\text{cepat}}) = 0,416 * 320 = 133,12$ Alumni dengan probabilitas cepat pada data training yang berjumlah 320 data , begitu juga dengan kelas-kelas lainnya dapat kita hitung dengan persamaan (2.4).

2. Menghitung setiap probabilitas variabel bebas pada setiap kelas.

Perhitungan nilai variabel bebas pada kedua kelompok kelas dilakukan pada data *testing* berdasarkan data *training*. Pada data testing diketahui alumni memiliki atribut variabel (X_1) Jurusan, (X_2) Jenis Kelamin, (X_3) Ekstrakurikuler, (X_4) Rata-rata nilai .Untuk menghitung probabilitas setiap variabel bebas pada ke tiga kelas menggunakan persamaan (2.6) maka hasil yang didapatkan dapat dilihat pada tabel 4.2.

Tabel 4.2. Probabilitas Setiap Variabel Bebas Terhadap Kelas.

Attribute	Parameter	Cepat	Lama	Sedang
Ekskul	value=IKUT	0.842	0.981	0.840
Ekskul	value=TIDAK IKUT	0.158	0.019	0.160
Jenis Kelamin	value=P	0.594	0.617	0.720
Jenis Kelamin	value=L	0.406	0.383	0.280
Jurusan	value=TKJ	0.887	0.130	0.600
Jurusan	value=AP	0.113	0.870	0.400
Nilai	value=Memuaskan	1.000	0.920	0.800
Nilai	value=Dengan pujian	0.000	0.080	0.160
Nilai	value=Cukup	0.000	0.000	0.040

Pada tabel 4.2 menampilkan hasil perhitungan probabilitas setiap variabel bebas terhadap kelas cepat, sedang dan lama. Pada tabel tersebut dapat kita ketahui probabilitas paling besar dari setiap atribut dan parameternya, probabilitas ini akan mempengaruhi terhadap hasil klasifikasi metode *naïve* bayes, semakin besar probabilitas, maka atribut dan parameter tersebut akan sangat berpengaruh, sebagai contoh atribut nilai dengan parameter memuaskan dengan probabilitas 1,000 pada kelas cepat, maka ini dapat diindikasikan bahwa setiap alumni yang memiliki nilai memuaskan dia akan semakin besar peluang untuk mendapatkan pekerjaan lebih cepat.

3. Menghitung perkalian probabilitas prior dan probabilitas setiap variabel bebas pada setiap kelompok (posterior).

Tahap ketiga adalah menghitung perkalian probabilitas *prior* dan probabilitas *posterior* pada data *testing*. Pada data testing pertama diketahui alumni memiliki Jurusan (X_1) TKJ, Jenis Kelamin (X_2) Laki-laki, Ekstrakurikuler (X_3) Ikut, Rata-rata nilai (X_4) Memuaskan. Pada data testing kedua diketahui alumni memiliki Jurusan (X_1) AP, Jenis Kelamin (X_2) Laki-laki, Ekstrakurikuler (X_3) Ikut, Rata-rata nilai (X_4) Memuaskan. Adapun perhitungan perkalian probabilitas *prior* dan probabilitas *posterior* dapat dihitung menggunakan persamaan (2.7), dan hasil perhitungan dapat dilihat pada tabel 4.3.

Tabel 4.3. Contoh Probabilitas *Posterior* Data Testing Pertama.

Jurusan	Confidence (Cepat)	Confidence (Lama)	Confidence (Sedang)	Keterangan	Klasifikasi
TKJ	0.800	0.144	0.056	Cepat	Cepat
AP	0.092	0.874	0.034	Lama	Lama

Pada data testing pertama alumni memiliki atribut nilai dengan parameter memuaskan dengan probabilitas pada kelas cepat adalah 1, lalu alumni memiliki atribut jurusan dengan TKJ dengan probabilitas pada kelas cepat adalah 0,887, berdasarkan data yang dapat kita lihat pada tabel 4.3 terhadap probabilitas setiap atribut dan parameter terhadap kelas, maka untuk data testing alumni yang pertama nilai confidence cepat lebih tinggi dari confidence kelas lainnya, dihasilkan confidence 0.800 lebih tinggi dari confidence kelas lainnya dan hasil klasifikasinya pun cepat.

Pada data testing kedua alumni memiliki atribut nilai dengan parameter memuaskan dengan probabilitas pada kelas cepat adalah 1, lalu alumni memiliki atribut jurusan dengan AP dengan probabilitas pada kelas cepat adalah 0,092, berdasarkan data yang dapat kita lihat pada tabel 4.2 terhadap probabilitas setiap atribut dan parameter terhadap kelas, maka untuk data testing alumni yang pertama nilai confidence lama lebih tinggi dari confidence kelas lainnya, dihasilkan confidence 0.874 lebih tinggi dari confidence kelas lainnya dan hasil klasifikasinya pun lama.

4. Mengevaluasi Hasil Klasifikasi

Untuk mengevaluasi hasil klasifikasi pada metode naive Bayes dilakukan pengukuran tingkat akurasi dengan menghitung peluang kesalahan klasifikasi menggunakan precision recall dan *confusion matrix* pada persamaan (2.9) – (2.14). Adapun hasil akurasi dari perhitungan dapat dilihat pada tabel 4.4.

Tabel 4.4. *Convusion Matrix* Metode Naïve Bayes.

	true Cepat	true Lama	true Sedang	class precision
pred. Cepat	18	1	2	85.71%
pred. Lama	4	48	1	90.57%
pred. Sedang	2	0	4	66.67%
class recall	75.00%	97.96%	57.14%	

Hasil akurasi yang didapat kan adalah $\frac{18+48+4}{80} = 87,50\%$.

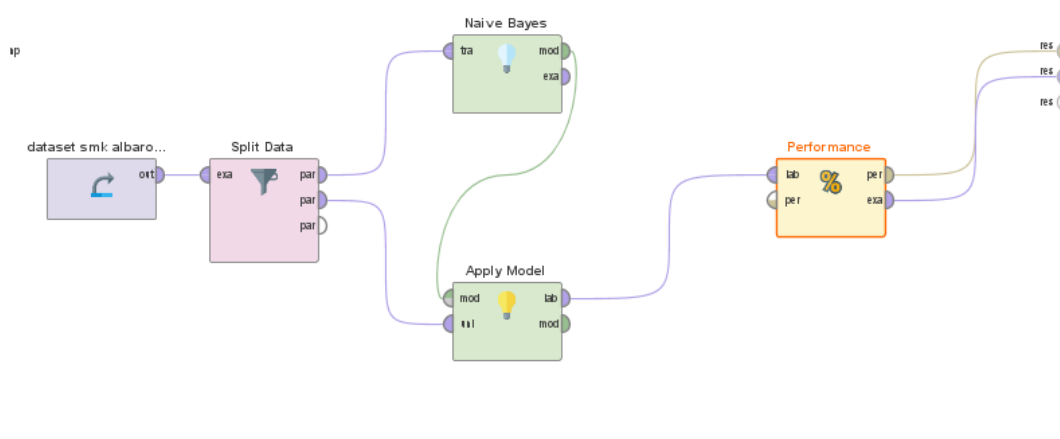
Hasil akurasi yang didapat adalah 87,50% dimana *naïve* bayes berhasil diklasifikasi kelas cepat dengan 18 Klasifikasi benar dengan hanya 3 Klasifikasi salah, dan Klasifikasi kelas Lama 48 benar dengan hanya 5 Klasifikasi salah, sedangkan pada kelas sedang 4 kali benar dengan 2 kali salah. Dari total 80 data training pada ujicoba *naïve* bayes hanya memiliki kesalahan pada 10 Klasifikasi.

4.4.1. Naïve Bayes dengan Rapid Miner

Proses ini merupakan implementasi dari proses pembuatan model klasifikasi pada pengklasifikasian data. Pengolahan data training dilakukan sehingga terbentuk sebuah model klasifikasi. Pada proses ini digunakan aplikasi Rapid Miner sebagai alat bantu untuk membuat proses data mining. Berikut adalah langkah-langkah algoritma Naïve bayes menggunakan Rapid Miner.

1. Import Terlebih dahulu data set/data training.
2. Setelah itu dilakukan Perubahan Role terhadap target/label.
3. Kemudian dilakukan Filter terhadap data yang missing sehingga Proses tidak terjadi Error.
4. Kemudian dilakukan Proses Naïve Bayes.

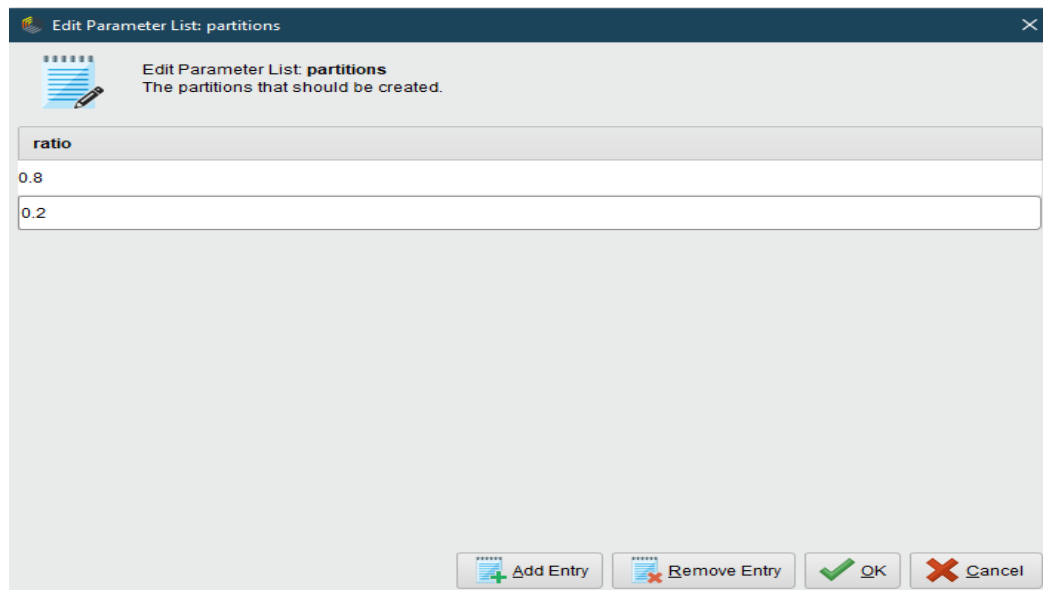
Berikut Adalah Rancangan Skema Naïve bayes Menggunakan Rapid Miner dapat dilihat pada gambar 4.6.



Gambar 4.6 Hasil Skema Naïve Bayes dengan Rapid Miner

Terdapat split data yang digunakan untuk membedakan antara data *testing* dengan data *training* dengan presentase 80% data training dan 20% data testing. Untuk *Performance* dari metode dapat dianalisis menggunakan *performance*.

Data testing yang akan dilakukan merupakan data alumni 5 tahun terakhir diambil sampel 20% dari total data, dan data training diambil 80% dari total data, berikut pengaturan pada aplikasi Rapid Miner dapat dilihat pada gambar 4.7.



Gambar 4.7 Data *Training* dan Data *Testing*

Berikut ini adalah hasil dari Naïve Bayes menggunakan Rapid Miner dengan nilai *confidence* masing-masing tertera pada gambar 4.8.

Row No.	KETERANGAN	prediction(K...	confidence(...	confidence(...	confidence(...	Jurusan	JK	PENDIDIKAN...	EKSKUL	RATA-RATA	KETERA...
1	LAMA	LAMA	0	1.000	0.000	TKJ	L	SMP	IKUT	81	TUNTA...
2	SEDANG	SEDANG	0.000	0.042	0.958	AP	L	SMP	IKUT	82	TUNTA...
3	LAMA	LAMA	0	1.000	0.000	AP	L	SMA	IKUT	83	TUNTA...
4	LAMA	LAMA	0.000	0.973	0.027	TKJ	L	SMA	IKUT	82	TUNTA...
5	CEPAT	CEPAT	1.000	0.000	0.000	AP	L	SMA	IKUT	81	TUNTA...
6	LAMA	LAMA	0	1.000	0.000	AP	P	SMA	IKUT	80	TUNTA...
7	LAMA	LAMA	0	1.000	0.000	TKJ	P	SMP	IKUT	85	TUNTA...
8	LAMA	LAMA	0.000	0.963	0.037	AP	P	SMP	IKUT	83	TUNTA...
9	LAMA	LAMA	0	1.000	0.000	AP	P	SMA	IKUT	84	TUNTA...
10	LAMA	LAMA	0	1.000	0.000	AP	P	SMP	IKUT	88	TUNTA...
11	LAMA	LAMA	0	1.000	0.000	AP	L	S1	IKUT	80	TUNTA...
12	LAMA	LAMA	0.000	1.000	0.000	TKJ	L	SMA	IKUT	82	TUNTA...
13	SEDANG	SEDANG	0.008	0.003	0.989	AP	L	SLTP	IKUT	79	TUNTA...
14	LAMA	LAMA	0	1.000	0.000	AP	P	SLTP	TIDAK IKUT	81	TUNTA...

Gambar 4.8. Hasil Percobaan metode Naïve Bayes dengan Rapid Miner

Sedangkan hasil akurasi dapat dilihat pada gambar 4.9.

accuracy: 87.50%

	true Cepat	true Lama	true Sedang	class precision
pred. Cepat	18	1	2	85.71%
pred. Lama	4	48	1	90.57%
pred. Sedang	2	0	4	66.67%
class recall	75.00%	97.96%	57.14%	

Gambar 4.9. Hasil Akurasi Percobaan metode Naïve Bayes dengan Rapid Miner

Dari hasil diatas didapatkan total accuracy dari percobaan analisis data alumni 5 tahun terakhir dalam kasus masa tunggu alumni untuk mendapatkan pekerjaan yaitu 95 % dimana *precision* dan *recall* masing-masing prediksi sedang, lama, sedang dapat dilihat pada gambar 4.9. Dalam perhitungan ini dapat diketahui bahwa metode Naïve Bayes sangat mungkin digunakan untuk analisis prediksi factor-faktor yang dapat mempengaruhi lama waktu alumni dalam mendapatkan pekerjaan.

4.5. Analisis Lama masa Tunggu alumni mendapatkan pekerjaan Dengan K-NN.

Pada proses klasifikasi metode K-NN terdapat tiga alur yaitu menentukan nilai parameter K, menghitung jarak Euclid antara data training dan data testing, kemudian menentukan ranking dari hasil perhitungan jarak. Data yang digunakan untuk perhitungan klasifikasi adalah menggunakan data training 80% sebanyak 320 data, sedangkan data testing 20% sebanyak 80 data. Adapun tahapan-tahapan dalam proses pengklasifikasian dengan metode K-NN sebagai berikut:

1. Menentukan nilai Parameter K.

Klasifikasi Waktu tunggu alumni dalam mendapatkan pekerjaan menggunakan metode K-NN dilakukan dengan menentukan nilai parameter K terlebih dahulu. Adapun nilai K yang digunakan yaitu K=1, K=3, K=5, K=7 dan K=9. Menghitung Jarak Euclid

Pada tahap selanjutnya dilakukan perhitungan jarak *Euclidean* pada objek antara data testing dan data training berdasarkan Persamaan (2.8). Perhitungan jarak pada penelitian ini menggunakan angka yang berasal dari kodingan kategori pada variabel bebas. Adapun perhitungan jarak Euclid antara data testing pertama dan data training pertama hingga data training ke-320 sebagai berikut:

$$\begin{aligned}d(X_1, X_1) &= \sqrt{(x_{11} - x_{11})^2 + (x_{12} - x_{12})^2 + (x_{13} - x_{13})^2 + (x_{14} - x_{14})^2} \\ &= \sqrt{(1 - 1)^2 + (2 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} \\ &= 1\end{aligned}$$

$$\begin{aligned}d(X_2, X_1) &= \sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + (x_{23} - x_{13})^2 + (x_{24} - x_{14})^2} \\ &= \sqrt{(2 - 1)^2 + (2 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} \\ &= 1.414214\end{aligned}$$

$$. = 1.414214$$

.

..

.

$$d(X_{320}, X_1)$$

$$= \sqrt{(x_{3201} - x_{11})^2 + (x_{3202} - x_{12})^2 + (x_{3203} - x_{13})^2 + (x_{3204} - x_{14})^2}$$

$$= \sqrt{(2 - 1)^2 + (1 - 2)^2 + (1 - 1)^2 + (2 - 2)^2}$$

$$= 1$$

2. Menentukan perhitungan K dari hasil perhitungan jarak.

Setelah jarak euclid didapatkan kemudian dilakukan perbandingan jarak dengan tetangga atau K yang ditentukan misal k=1, dimana jarak yang diambil adalah jarak yang paling kecil dari jumlah tetanga tersebut. Berikut hasil perhitungan jarak antara data testing pertama dengan data training dapat dilihat pada tabel 4.5.

Tabel 4.5. Hasil Perhitungan Jarak dan K=1

data training ke -	data testing ke -	jarak (d)	Hasil K=1
1	1	1	0
2	1	1.414214	0
3	1	1	0
4	1	1.732051	0
5	1	1.732051	0
6	1	1	0
7	1	0	1
8	1	1	0
9	1	1	0
10	1	1.414214	0
11	1	1	0
12	1	1	0
13	1	1	0
.	.	.	.
..
.	.	.	.
320	1	1	0

Pada tabel 4.5 diatas dilakukan perbandingan jarak terhadap semua data training dimana jumlah K yang ditentukan adalah 1, dimana setiap jarak *euclidean* yang dihasilkan kemudian dibandingkan dengan data data jarak *euclidean* dari data training 1 – 320, jika jarak tersebut lebih kecil maka dilakukan pelabelan waktu tunggu berdasarkan label kelas data testing cepat, sedang atau lama.

Sebagai contoh pada data training ke 7 dimana jarak terhadap data testing pertama adalah 0 maka 0 akan dibandingkan dengan data jarak 1 sampai 320 dan jika 0 merupakan terkecil maka akan dilakukan pelabelan waktu sesuai dengan data waktu di data training ke 7 yaitu 1 cepat. Dari hasil data testing pertama maka akan dilakukan pengelompokan terhadap label data yang paling sering muncul.

3. Evaluasi Hasil Klasifikasi

Pada evaluasi hasil Klasifikasi dari uji coba dilakukan dengan menggunakan nilai K =1, sampai K=13. Berikut adalah hasil akurasi dari setiap percobaan menggunakan nilai K yang berbeda. Berikut hasil percobaan dari mulai K=1 sampai dengan K=13 dapat dilihat pada tabel 4.6.

Tabel 4.6. Hasil Percobaan Perhitungan Menggunakan Nilai K.

Nilai K	Akurasi
1	72.50%
3	57.50%
5	57.50%
7	78.75%
9	83.75%
11	93.75%
13	84.75%

Nilai K pada algoritma KNN mendefinisikan berapa banyak tetangga yang akan diperiksa untuk menentukan klasifikasi pada data tersebut. Maka perbedaan Nilai K pada setiap pengujian mempengaruhi terhadap akurasi. Pada penelitian ini akan digunakan hasil nilai akurasi yang paling besar dari percobaan menggunakan nilai

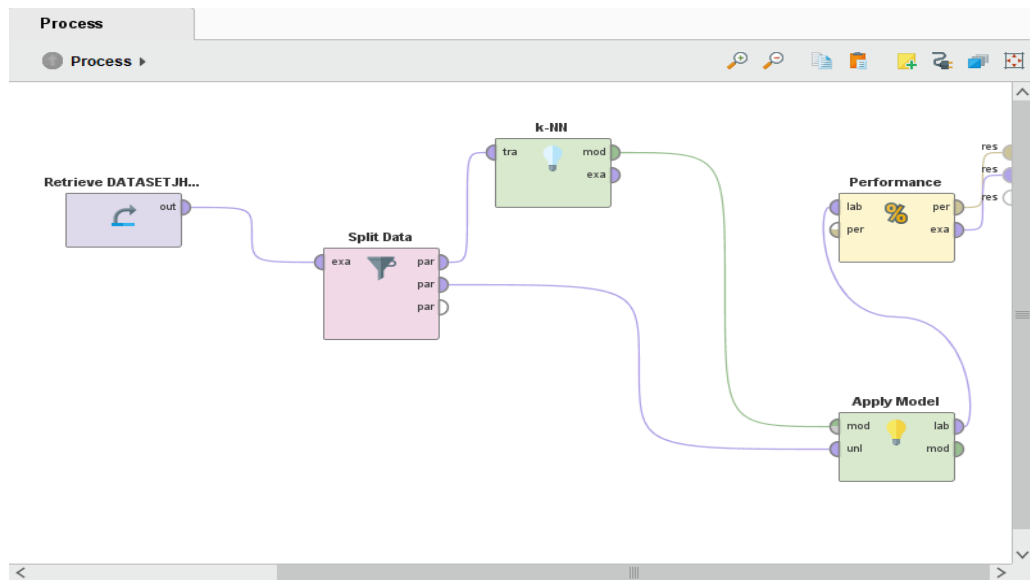
K yang berbeda. Setelah di lakukan uji coba pada nilai K= 1 sampai dengan K=11 didapatkan adanya peningkatan akurasi. Namun, pada K=13 terjadi adanya penurunan tingkat akurasi. Sehingga akurasi terbaik terdapat pada K=11 yaitu sebesar 93.75% .

4.5.1. KNN dengan Rapid Miner

Proses ini merupakan implementasi dari proses pembuatan model klasifikasi pada pengklasifikasian data. Pengolahan data training dilakukan sehingga terbentuk sebuah model klasifikasi. Pada proses ini dibagi menjadi dua yaitu pembentukan *Naïve Bayes* Pada proses ini digunakan aplikasi *Rapid Miner* sebagai alat bantu untuk membuat proses data mining. Berikut adalah langkah-langkah algoritma *Naïve Bayes* menggunakan *Rapid Miner*.

1. Import Terlebih dahulu data set/data training.
2. Setelah itu dilakukan Perubahan Role terhadap target/label.
3. Kemudian dilakukan Split datz untuk membedakan data training dan data testing.
4. Kemudian dilakukan Proses KNN.
5. Setelah itu masukan *Performance* untuk mengetahui tingkat akurasi metode

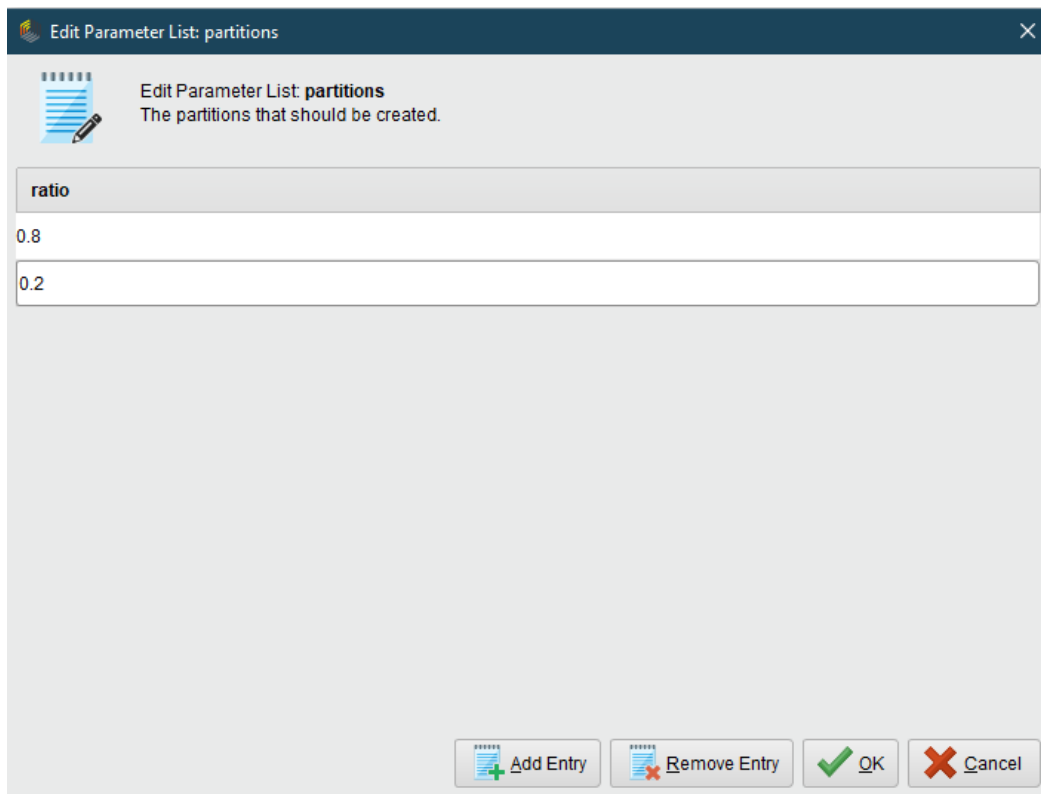
Berikut adalah Rancangan Full Skema KNN menggunakan Rapid Miner dapat dilihat pada gambar 4.10.



Gambar 4.10. Hasil Skema Naïve Bayes Dengan Rapid Miner

Terdapat split data yang digunakan untuk membedakan antara data *testing* dengan data *training* dengan presentase 80% data training dan 20% data testing. Untuk *Performance* dari metode akan dianalisis menggunakan *performance*.

Data testing yang akan dilakukan merupakan data alumni 5 tahun terakhir diambil sampel 20% dari total data, dan data training diambil 80% dari total data, berikut pengaturanya pada aplikasi Rapid Miner dapat dilihat pada gambar 4.11.



Gambar 4.11 Data *Training* Dan Data *Testing*

Berikut ini adalah hasil dari K-NN menggunakan Rapid Miner dengan nilai $K=11$, *confidence* masing-masing tertera pada gambar dibawah.

Hasil pada percobaan ini dihasilkan seperti gambar 4.12 di bawah ini.

Row No.	KETERANGAN	prediction(K...	confidence(...	confidence(...	confidence(...	Jurusan	JK	EKSKUL	Nilai
1	Cepat	Cepat	0.545	0.455	0	TKJ	L	IKUT	Memuaskan
2	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan
3	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan
4	Cepat	Cepat	0.545	0.455	0	TKJ	L	IKUT	Memuaskan
5	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan
6	Lama	Lama	0.182	0.636	0.182	AP	P	IKUT	Memuaskan
7	Sedang	Sedang	0.262	0.262	0.475	TKJ	P	IKUT	Dengan pujian
8	Lama	Lama	0.182	0.636	0.182	AP	P	IKUT	Memuaskan
9	Lama	Lama	0.182	0.636	0.182	AP	P	IKUT	Memuaskan
10	Lama	Lama	0.171	0.743	0.086	AP	P	IKUT	Dengan pujian
11	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan
12	Cepat	Cepat	0.545	0.455	0	TKJ	L	IKUT	Memuaskan
13	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan
14	Cepat	Cepat	0.750	0.100	0.150	AP	P	TIDAK IKUT	Memuaskan
15	Cepat	Cepat	0.545	0.455	0	TKJ	L	IKUT	Memuaskan
16	Cepat	Cepat	0.750	0.100	0.150	AP	P	TIDAK IKUT	Memuaskan
17	Lama	Lama	0.182	0.636	0.182	AP	P	IKUT	Memuaskan
18	Cepat	Cepat	0.650	0.250	0.100	TKJ	L	TIDAK IKUT	Memuaskan
19	Lama	Lama	0.091	0.636	0.273	AP	L	IKUT	Memuaskan

ExampleSet (80 examples, 5 special attributes, 4 regular attributes)

Gambar 4.12. Hasil Percobaan Metode KNN Dengan Rapid Miner

Hasil akurasi menggunakan K=11 dapat dilihat pada gambar 4.13 di bawah ini.

accuracy: 93.75%

	true Cepat	true Lama	true Sedang	class precision
pred. Cepat	24	2	3	82.76%
pred. Lama	0	47	0	100.00%
pred. Sedang	0	0	4	100.00%
class recall	100.00%	95.92%	57.14%	

Gambar 4.13. Hasil Akurasi Percobaan Metode KNN Dengan RapidMiner

Dari hasil gambar 4.13 diatas didapatkan total accuracy dari percobaan analisis data alumni 5 tahun terakhir dalam kasus masa tunggu alumni untuk mendapatkan pekerjaan yaitu 93,75 % dimana precission dan recall masing-masing prediksi sedang, lama, sedang dapat dilihat pada gambar 4.13. Pada penelitian ini, metode KNN sangat mungkin digunakan untuk analisis prediksi faktor-faktor apa saja yang dapat mempengaruhi lama waktu alumni dalam mendapatkan pekerjaan.

4.6. Perbandingan tingkat akurasi metode Naïve bayes dan K-NN

Pengukuran tingkat akurasi baik pada metode naive Bayes maupun K-NN dilakukan dengan menghitung peluang kesalahan klasifikasi. Pada proses klasifikasi diharapkan melakukan klasifikasi pada semua obyek dengan benar, sehingga semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan. Berikut adalah hasil akurasi dari masing-masing metode dapat dilihat pada tabel 4.7.

Tabel 4.7. Hasil Akurasi Metode Naïve bayes dan K-NN

Metode	Akurasi
Naïve Bayes	87.50%
K-NN	93.75%

Pada Tabel 4.7 dapat diketahui bahwa pada metode Naïve Bayes menunjukkan hasil akurasi sebesar 87.50% dan pada metode K-NN menunjukkan hasil akurasi sebesar 93.75%. Hal ini menunjukkan bahwa metode K-NN bekerja lebih baik dibandingkan dengan metode K-NN untuk klasifikasi waktu tunggu alumni dalam mendapatkan pekerjaan dengan nilai akurasi yang lebih tinggi.