BAB III

ANALISA DAN PERANCANGAN

3.1. Arsitektur Sistem

Arsitektur sistem adalah representasi abstrak dari suatu sistem, yang menggambarkan komponen-komponennya, interaksi antar komponen tersebut, dan bagaimana mereka bekerja bersama untuk mencapai tujuan sistem secara keseluruhan. arsitektur sistem adalah fondasi yang penting dalam penelitian. Tanpa arsitektur yang baik, penelitian dapat menjadi tidak terarah, tidak efisien, dan sulit untuk direplikasi. Oleh karena itu, investasi dalam perencanaan dan perancangan arsitektur sistem yang matang adalah langkah yang krusial untuk keberhasilan penelitian.

Dalam penelitian ini, tahapan pertama dimulai dengan pengumpulan data. Dalam pengumpulan data disini diambli dari data kuisioner yang telah dibagikan sebelumnya kepada Mahasiswa Fakultas Sains dan Teknologi Universitas Labuhanbatu. Data ini sudah diseleksi berdasarkan relevansi dengan tujuan penelitian yang akan dilakukan. Sebelum data diproses, dilakukan pembersihan data dan normalisasi untuk data yang lebih optimal.

Kemudian data yang telah didapat tadi di proses dengan pengelompokan data kedalam sejumlah *cluster* (K). Setelah pengelompokan data berdasarkan cluster selesai, kemudian ditentukan jumlah *cluster* yang optimal.

Setelah itu, kemudian masuk kedalam pelabelan data. Pada pelabelan data, data yang telah di kelompokkan dilabeli setiap karakteristik data didalam *cluster*-

nya. Data yang telah dilabeli kemudian dijadikan target dalam klasifikasi *K-Nearest Neighbor*.

Pada tahapan *K-Nearest Neighbor,* data dibagi menjadi dua, yaitu data latih *(training)* dan data uji *(testing)*. Dimana Data *training* digunakan untuk melatih model dan Data *testing* digunakan untuk mengukur performa model yang dihasilkan.

Selanjutnya, sistem ini melakukan Evaluasi Model, yang bertujuan untuk menguji kinerja model klasifikasi. Evaluasi dilakukan dengan menggunakan *K-Nearest Neighbor*dan menghitung metrik evaluasi yang relevan untuk perhitungan akurasi dan presisi. Hasil dari evaluasi ini di interpretasikan untuk memberikan gambaran yang jelas mengenai keunggulan dan kelemahan dari setiap metode yang digunakan.

Dengan arsitektur yang sistematis ini, proses analisis data dapat dilakukan dengan lebih optimal, dan hasil yang diperoleh dapat dijadikan dasar dalam pengambilan keputusan serta rekomendasi yang berkaitan dengan analisa pola asuh orang tua terhadap kesejahteraan emosional remaja pada mahasiswa Fakultas Sains dan Teknologi Universitas Labuhanbatu.

3.1.1. Perbandingan antara Metode *K-Nearest Neighbor* dengan *K-Means Clustering*

Untuk perbandingan antara metode *K-Nearest Neighbor* dan Metode *K-Means Clustering* akan disajikan dalam bentuk tabel berikut:

Tabel 3.1 Perbandingan Antara Dua Metode

Aspek	Metode K-Nearest Neighbor	Metode K-Means Clustering
Dasar	Algoritma supervised learning yang digunakan untuk klasifikasi dan regresi dengan dasar pemikiran sangat intuitif dan sederhana.	Algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kemiripan karakteristiknya untuk memaksimalkan kemiripan data dalam satu cluster dan meminimalkan kemiripan data antar cluster.
Tugas	Klasifikasi dan Regresi	Clustering
Data yang dibutuhkan	Data Berlabel	Data Tidak Berlabel
Cara Kerja	Berdasarkan jarak tetangga terdekat	Berdasarkan pusat cluster
Kelebihan	Sederhana dan fleksibel	Efisien dan mudah di implementasikan
Kekurangan	Membutuhkan banyak memori, sensitif terhadap data yang tidak relevan.	Sensitif terhadap inisialisasi, sulit menentukan jumlah cluster.

3.1.2. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan di Universitas Labuhanbatu pada Mahasiswa Fakultas Sains dan Teknologi yang berlokasi di Jl. SM. Raja Aek Tapa No.126 A KM 3.5, Bakaran Batu, Kecamatan Rantau Selatan, Kabupaten Labuhanbatu, Sumatera Utara dengan kode pos 21418. Lokasi ini dipilih karena peneliti ingin menganalisa pola asuh orang tua terhadap kesejahteraan emosional remaja pada tingkat mahasiswa. Penelitian dijadwalkan berlangsung pada bulan Desember 2024, dengan fokus utama pada pengaruh pola asuh orang tua untuk mengetahui kesejahteraan emosional remaja.

3.1.3. Populasi dan Sampel

Populasi dalam penelitian ini adalah remaja yang ada di Labuhanbatu. Dari populasi tersebut, sampel yang dipilih untuk penelitian ini adalah mahasiswa Universitas Labuhanbatu pada Fakultas Sains dan Teknologi. yang berlokasi di Jl. SM. Raja Aek Tapa No.126 A KM 3.5, Bakaran Batu, Kecamatan Rantau Selatan. Pemilihan Fakultas Sains dan Teknologi sebagai sampel didasarkan pada pertimbangan bahwa fakultas ini merupakan salah satu fakultas dengan keberagaman yang banyak, sehingga relevan untuk menganalisis kesejahteraan emosional remaja.

3.1.4. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui observasi dan kuesioner. Observasi digunakan untuk mengamati langsung emosional mahasiswa Fakultas Sains dan Teknologi di Universitas Labuhanbatu, termasuk cara konsentrasi belajar, tingkah laku diluar ruang kelas, dan tutur bahasa

mahasiswa. Sementara itu, kuesioner digunakan untuk mengumpulkan data dari mahasiswa terhadap sifat dan perasaan mahasiswa terhadap orangtuanya.. Kuesioner ini berisi pertanyaan-pertanyaan yang dirancang untuk menggali faktor-faktor seperti pengaruh mahasiswa terhadap pengambilan keputusan dan pengaruh perasaan mahasiswa.

3.2. Desain Aktivitas Sistem

Pengolahan data pada penelitian ini dilakukan menggunakan rumus dari *K-Nearest Neighbor* dan Metode *K-Means Clustering*. Kedua metode ini diterapkan untuk mengklasifikasikan data sesuai dengan pengaruh pola asuh orang tua terhadap kesejahteraan emosional remaja. Untuk prosesnya yaitu sebagai berikut.

3.2.1. Langkah-Langkah Penerapan *K-Nearest Neighbor* dan Metode *K-Means Clustering*

Pada penerapan metode *Naive Bayes* dan Neural Network terdapat beberapa tahapan yang harus dilakukan yaitu sebagai berikut.

1.2.1.1 Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mengumpulkan atau mengukur informasi dari berbagai sumber untuk mendapatkan gambaran yang lengkap dan akurat tentang suatu topik atau masalah. Data yang terkumpul ini nantinya akan dianalisis dan diinterpretasikan untuk menghasilkan kesimpulan atau jawaban atas pertanyaan penelitian.

Penelitian ini dilakukan dengan membagikan kuisioner kepada 55 mahasiswa fakultas Sains dan Teknologi Universitas Labuhanbatu.

Tabel 3.2 Kuisioner

No	Pertanyaan
BAGIAN I	
	Nama
	Usia
	Jenis Kelamin
	Prodi
	Status Hubungan Dengan Orangtua
BAGIAN II	
1	Bertutur dengan saya dengan nada yang hangat dan bersahabat
2	Tidak membantu saya sebanyak mana yang saya perlukan
3	Membiarkan saya untuk melakukan hal yang saya suka
4	Nampak seperti bersikap dingin terhadap saya
5	Memahami masalah dan kebimbangan saya
6	Menyayangi saya
7	Mendukung saya membuat keputusan sendiri
8	Tidak mau saya bersikap dewasa
9	Mengawasi apa yang saya lakukan
10	Mencampuri urusan pribadi saya
11	Senang apabila berbicara dengan saya
12	Sering senyum kepada saya
13	Cenderung memanjakan saya
14	Kelihatan tidak memahami keperluan dan kehendak saya
15	Membiarkan saya membuat keputusan sendiri
16	Membuat saya merasa diinginkan
17	Membuat saya merasa lebih baik ketika saya gelisah
18	Banyak berbicara dengan saya
19	Mencoba untuk membuat saya mandiri (tidak bergantung dengan mereka)

20	Merasa saya bisa menjaga diri sendiri ketika ketika saya tidak berada di sekitar mereka
21	Memberikan batasan pada saya
22	Tidak membiarkan saya pergi sesering yang saya inginkan
23	Melindungi saya
24	Memuji saya
25	Memperhatikan cara saya berpakaian

Pertanyaan ini terdiri dari 25 kuisioner yang dibagi menjadi 2 kategori yaitu 1-12 tentang Kasih Sayang dan Perhatian dan 13-25 tentang Proteksi Berlebihan/Kontrol.

Dalam penelitian ini data dikumpulkan dengan 5 kategori kuisioner sebagai berikut :

Tabel 3.3 Keterangan Kuisioner

No	Pilihan	Keterangan
1	1	Tidak Pernah
2	2	Jarang
3	3	Kadang-Kadang
4	4	Sering

Semakin kecil skala yang di ambil maka semakin kecil kemungkinan pengaruh pola asuh terhadap mental anak. Dalam pembagian kuisioner ini didapatkan data mentah yang memiliki variasi dalam pemilihan jawaban.

Tabel 3.4 Hasil Kuisioner

	Nama	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	18	20	21	22	23	24	25
2	R1	3	1	3	2	3	4	3	1	2	2	3	3	3	2	4	3	3	3	3	3	4	3	4	2	1
3	R2	3	1	3	2	3	4	2	3	4	3	2	3	3	3	2	3	4	4	4	3	4	4	4	2	3
4	R3	4	1	4	1	4	4	3	1	3	2	3	4	4	2	4	4	4	3	3	3	1	3	4	3	2
5	R4	4	1	2	1	3	4	2	3	3	2	3	4	3	1	3	3	4	4	4	1	2	2	4	3	4
6	R5	4	1	1	1	4	4	2	1	4	3	4	4	4	1	3	4	4	3	3	3	4	3	4	2	1
7	R6	3	1	1	1	3	4	3	1	4	3	4	4	4	2	1	3	4	4	4	4	3	2	4	4	4
8	R7	4	1	3	1	4	4	2	1	2	1	4	4	4	1	2	4	4	4	2	2	3	2	4	4	3
9	R8	3	3	4	2	3	3	4	1	1	1	2	2	1	3	4	2	3	2	4	4	1	1	4	2	1
10	R9	3	2	4	1	4	4	4	1	2	2	3	3	1	2	4	4	4	4	4	4	1	3	4	3	1
11	R10	3	2	4	2	3	4	4	1	1	1	2	2	1	3	4	2	2	3	4	4	1	1	3	3	1
12	R11	3	3	4	3	2	3	4	1	1	2	3	2	1	2	3	2	2	2	4	4	2	1	3	2	1
13	R12	4	2	4	1	3	4	4	1	2	2	3	3	1	2	4	4	4	2	4	4	2	1	4	3	1
14	R13	4	3	4	1	4	4	4	1	3	2	2	3	1	3	4	3	4	2	4	4	2	2	4	3	1
15	R14	2	2	4	2	2	4	4	1	1	1	2	2	1	1	4	2	4	2	4	4	2	1	4	2	1
16	R15	4	3	4	1	3	4	4	1	3	3	3	3	2	3	3	4	3	4	3	3	3	4	4	3	2
17	R16	3	2	4	2	3	4	3	2	2	3	3	3	2	1	4	3	3	2	3	4	2	2	4	3	2
18	R17	3	2	2	2	2	4	3	3	4	3	2	2	1	3	4	3	4	3	4	3	4	4	4	3	3
19	R18	3	3	4	1	3	4	4	2	3	3	2	2	3	3	3	2	2	3	4	4	2	1	4	2	1
20	R19	4	1	4	2	4	4	4	2	3	2	4	4	3	1	3	4	4	4	4	3	4	4	4	3	4
21	R20	3	2	4	2	3	3	4	1	2	2	2	2	1	4	3	2	2	2	4	4	2	1	2	1	1
22	R21	3	2	4	2	3	4	4	1	3	2	3	3	2	3	4	2	3	3	4	4	2	3	4	2	2
23	R22	4	1	4	1	3	4	2	3	2	1	2	2	3	3	2	3	3	2	4	4	2	2	3	2	1
24	R23	4	3	4	2	2	4	4	1	1	2	3	3	1	1	4	2	4	3	4	4	3	3	4	3	2
25	R24	4	1	4	2	4	4	4	1	2	2	3	3	2	2	4	4	4	3	4	4	3	3	4	3	3
26	R25	3	1	4	2	3	3	4	1	2	2	3	2	2	2	3	3	3	3	4	4	3	3	4	2	2
27	.....																									
28	.....																									
29	.....																									
30	R95	3	2	4	2	3	3	4	1	2	3	3	2	2	4	4	3	4	4	3	3	1	1	2	2	2
31	R96	3	2	4	2	2	4	4	2	2	1	3	3	2	3	4	4	4	3	4	4	2	1	4	2	2
32	R97	4	1	4	1	4	3	2	1	4	2	4	2	1	3	4	3	4	4	4	4	3	4	3	2	1
33	R98	2	3	4	1	3	3	4	1	3	3	3	2	1	2	4	3	4	3	4	4	1	1	3	1	1
34	R99	3	4	4	1	4	4	4	1	1	1	4	2	1	2	4	4	4	2	4	4	1	1	2	2	1
35	R100	3	1	2	3	2	4	2	1	4	4	3	2	1	4	2	3	4	4	4	4	4	3	4	2	4
36	R101	4	1	4	1	4	4	4	1	3	1	4	4	1	1	4	4	4	4	4	4	1	1	4	3	1

1. Pelabelan Data

Pelabelan data adalah proses pemberian label atau kategori pada data. Dalam konteks supervised learning (pembelajaran terbimbing) seperti K-NN, pelabelan data sangat penting karena algoritma ini belajar dari data yang sudah memiliki label. Label ini digunakan sebagai "jawaban" yang benar untuk data pelatihan, sehingga model dapat belajar untuk memprediksi label data baru yang belum berlabel.

Pada data berikut dibagi menjadi dua bagian yaitu data training dan data testing. Data training terdiri dari 101 data yang diperoleh melalui kuisioner

sedangkan data testing diambil sebagai sampel sebanyak 15 data. Data testing yang digunakan dalam penelitian ini sebanyak 15 data, yang merupakan data sampel penelitian. Data ini akan diolah dan dianalisis menggunakan model yang telah dibangun dengan bantuan *data training*. Proses ini bertujuan untuk mengukur performa model klasifikasi dalam mengidentifikasi pola dan menghasilkan prediksi yang akurat sesuai dengan data yang diujikan.

Tabel 3.5 Data Testing

No	Nama	Kasih Sayang	Proteksi	Total	Tingkat Emosional
1	R1	30	38	68	Normal
2	R2	33	43	76	Pengaruh
3	R3	34	40	74	Pengaruh
4	R4	32	38	70	Pengaruh
5	R5	33	39	72	Pengaruh
6	R6	32	43	75	Pengaruh
7	R7	31	39	70	Pengaruh
8	R8	29	32	61	Normal
9	R9	33	39	72	Pengaruh
10	R10	29	32	61	Normal
11	R11	31	29	60	Normal
12	R12	33	36	69	Normal
13	R13	35	37	72	Pengaruh
14	R14	27	32	59	Normal
15	R15	36	41	77	Pengaruh

Data training yang digunakan dalam penelitian ini berjumlah 101 data, yang berfungsi sebagai dasar untuk melatih model klasifikasi. Data ini membantu proses pengolahan data dengan mengenalkan pola dan karakteristik tertentu kepada model,

sehingga model dapat belajar untuk membuat prediksi yang akurat pada tahap pengujian. Dengan *data training* ini, diharapkan model dapat memahami hubungan antara variabel-variabel yang ada untuk mendukung tujuan penelitian.

Tahap pengolahan data dikumpulkan dari 15 atribut menjadi 3 atribut berdasarkan pilihan kuisioner data mentah yang masuk dan ditampilkan seperti berikut:

Tabel 3.6 Atribut

1	No	Nama	Kasih Sayang	Proteksi	Total	Tingkat Emosional
2	1	R1	30	38	68	Normal
3	2	R2	33	43	76	Pengaruh
4	3	R3	34	40	74	Pengaruh
5	4	R4	32	38	70	Pengaruh
6	5	R5	33	39	72	Pengaruh
7	6	R6	32	43	75	Pengaruh
8	7	R7	31	39	70	Pengaruh
9	8	R8	29	32	61	Normal
10	9	R9	33	39	72	Pengaruh
11	10	R10	29	32	61	Normal
12	11	R11	31	29	60	Normal
13	12	R12	33	36	69	Normal
14	13	R13	35	37	72	Pengaruh
15	14	R14	27	32	59	Normal
16	15	R15	36	41	77	Pengaruh
17	16	R16	34	35	69	Normal
18	17	R17	32	43	75	Pengaruh
19	18	R18	34	34	68	Normal
20	19	R19	38	45	83	Pengaruh
21	20	R20	30	29	59	Normal
22	21	R21	34	38	72	Pengaruh
23	22	R22	29	34	63	Normal
24	23	R23	33	38	71	Pengaruh
25	24	R24	34	43	77	Pengaruh
26	25	R25	30	38	68	Normal
27	26	R26	35	38	73	Pengaruh
28	.....					
29	.....					
30	100	R100	31	43	74	Pengaruh
31	101	R101	35	36	71	Pengaruh

2. Data Processing

Data Processing (Pemrosesan Data) merupakan sebuah proses pemilahan data dari data mentah menjadi data yang lebih relevan dalam penelitian ini. Proses ini mengubah data mentah, yang mungkin tidak terorganisir dan sulit diinterpretasikan, menjadi data yang terstruktur, relevan, dan mudah dipahami. Data processing melibatkan berbagai tahapan, dan tujuannya adalah untuk meningkatkan kualitas data dan membuatnya lebih berguna untuk pengambilan keputusan, analisis, atau tujuan lainnya. Dalam penelitian ini data di proses melalui 2 metode yaitu metode K-Nearest Neighbor dan metode K-Means Clustering.

1) Metode K-Nearest Neighbor

Tujuan utama dari pengumpulan data adalah untuk mendapatkan informasi yang relevan dan andal untuk menjawab pertanyaan penelitian atau mencapai tujuan penelitian. Proses ini melibatkan identifikasi, pengelolaan, dan pemilihan data yang relevan untuk mencapai tujuan penelitian. K-Nearest Neighbor mengklasifikasikan data baru berdasarkan "kedekatannya" dengan data lain dalam data pelatihan. "Kedekatan" ini diukur menggunakan metrik jarak. Algoritma ini tidak membuat asumsi tentang bentuk distribusi data. Dalam penelitian ini, data yang digunakan terdiri dari dua jenis dataset, yaitu *data training* dan *data testing*. *Data training* digunakan untuk melatih model klasifikasi agar dapat mengenali pola dan karakteristik yang ada pada data, sedangkan *data testing* berfungsi untuk menguji akurasi dan performa model yang telah dilatih.

Tabel 3.7 Prediksi

No	Nama	Kasih Sayang	Proteksi	Total	Tingkat Emosional
1	R1	30	38	68	Normal
2	R2	33	43	76	Pengaruh
3	R3	34	40	74	Pengaruh
4	R4	32	38	70	Pengaruh
5	R5	33	39	72	Pengaruh
6	R6	32	43	75	Pengaruh
7	R7	31	39	70	Pengaruh
8	R8	29	32	61	Normal
9	R9	33	39	72	Pengaruh
10	R10	29	32	61	Normal
11	R11	31	29	60	Normal
12	R12	33	36	69	Normal
13	R13	35	37	72	Pengaruh
14	R14	27	32	59	Normal
15	R15	36	41	77	Pengaruh
16	R16	31	40	71	?

Menentukan nilai K yang tepat dalam algoritma K-Nearest Neighbor (K-NN) adalah langkah penting untuk mendapatkan model yang akurat. Nilai K ini merepresentasikan jumlah tetangga terdekat yang akan dipertimbangkan dalam proses klasifikasi atau regresi. Dalam kasus ini, penentuan nilai K diambil dari pertimbangan domain masalah. Jadi, nilai K pada penelitian ini adalah 3.

Dalam metode K-Nearest Neighbor (K-NN), perhitungan jarak antar data merupakan langkah krusial. Jarak ini menentukan seberapa mirip atau dekat suatu data dengan data lainnya. Jarak Euclidean adalah jarak "garis lurus" antara dua titik dalam ruang berdimensi n. Jarak Euclidean memberikan ukuran seberapa "dekat" atau "mirip" dua titik dalam ruang berdimensi n. Semakin kecil jarak Euclidean

antara dua titik, semakin mirip kedua titik tersebut. Sebaliknya, semakin besar jarak Euclidean antara dua titik, semakin tidak mirip kedua titik tersebut.

Secara matematis, rumus jarak Euclidean antara dua titik x dan y dalam ruang berdimensi n dinyatakan sebagai berikut:

$$d(x, y) = \sqrt{((x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

Setelah menghitung jarak pada data, maka data dicari 3 tetangga terdekat berdasarkan kluster yang kita tentukan sehingga didapat nilai 69, 70, 72. Nilai nilai ini kemudian dilakukan prediksi dimana kemungkinan terbesar data baru yang muncul.

Tabel 3.8 Euclidean

No	Nama	Kasih Sayang	Proteksi	Total	Tingkat Emosional
7	R7	31	39	70	Pengaruh
9	R9	33	39	72	Pengaruh
12	R12	33	36	69	Normal
16	R16	31	40	71	?

Prediksi dari 3 tetangga terdekat nilai yang diperoleh adalah 2 terpengaruh dan 1 nilai dalam kategori normal. Sehingga prediksi nilai baru yang muncul adalah terpengaruh.

2) Metode *K-Means Clustering*

Metode *K-Means Clustering* merupakan algoritma *unsupervised learning* yang populer digunakan untuk mengelompokkan data ke dalam sejumlah *cluster* (kelompok) berdasarkan kemiripan karakteristiknya. Tujuannya adalah untuk meminimalkan variasi (atau inersia) dalam setiap

cluster dan memaksimalkan variasi antar *cluster*. Dalam penelitian ini di kelompokkan data testing menjadi sebagai berikut :

Tabel 3.9 Klustering

No	Nama	Kasih Sayang	Proteksi	Total	Tingkat Emosional
1	R14	27	32	59	Normal
2	R11	31	29	60	Normal
3	R8	29	32	61	Normal
4	R10	29	32	61	Normal
5	R1	30	38	68	Normal
6	R12	33	36	69	Normal
7	R4	32	38	70	Pengaruh
8	R7	31	39	70	Pengaruh
9	R9	33	39	72	Pengaruh
10	R13	35	37	72	Pengaruh
11	R5	33	39	72	Pengaruh
12	R3	34	40	74	Pengaruh
13	R6	32	43	75	Pengaruh
14	R2	33	43	76	Pengaruh
15	R15	36	41	77	Pengaruh

Data testing yang digunakan dalam penelitian ini sebanyak 15 data, yang merupakan data sampel penelitian. Data ini akan diolah dan dianalisis menggunakan klasifikasi berdasarkan frekuensi pernyataan dari kuisioner. Proses ini bertujuan untuk mengukur performa model klasifikasi dalam mengidentifikasi pola dan menghasilkan prediksi yang akurat sesuai dengan data yang diujikan.

K-Means bekerja dengan cara iteratif, yaitu dengan berulang kali menetapkan titik-titik data ke cluster terdekat dan memperbarui pusat cluster (centroid) hingga konvergensi.

Langkah-Langkah Perhitungan K-Means

1. Menentukan Nilai K:

   Pilih jumlah cluster (K) yang diinginkan. Nilai K ini harus ditentukan sebelum memulai algoritma. Paka cluster ini data dibagi menjadi 2 sehingga K=2

2. Inisialisasi Centroid:

   Pilih K titik data acak dari dataset sebagai pusat cluster awal (centroid). Titik centroid ini diambil adalah (27,32) (33,39)

   Menetapkan Data ke Cluster Terdekat:

3. Hitung jarak antara setiap titik data dengan semua centroid.

   Tetapkan setiap titik data ke cluster dengan centroid terdekat. Jarak dapat dihitung menggunakan berbagai metode, seperti jarak Euclidean.

   $d(x, y) = \sqrt{((x_1 - y_1)^2 + (x_2 - y_2)^2}$

4. Memperbarui Centroid:

   Hitung rata-rata fitur dari semua titik data dalam setiap cluster.

Gunakan rata-rata ini sebagai centroid baru untuk cluster tersebut. Rata-rata fitur dari K1 adalah (31,29) sehingga centroid baru untuk cluster 1 adalah (31,29). Dan untuk Rata-rata fitur dari K2 adalah (35,37) sehingga centroid baru untuk cluster 1 adalah (35,37).

5. Iterasi:

   Ulangi langkah 3 hingga 4 kali hingga centroid tidak lagi berubah secara signifikan atau mencapai batas iterasi yang ditentukan.

3. Klasifikasi Data K-Nearest Neighbor

Pengujian metode dalam penelitian ini dilakukan menggunakan aplikasi Rapidminner. RapidMiner adalah perangkat lunak yang menyediakan lingkungan terpadu untuk data science. Berikut ini adalah beberapa fungsi utama RapidMiner:

a) Persiapan Data (Data Preparation): RapidMiner memiliki fitur yang lengkap untuk membersihkan, mengubah, dan menyiapkan data sebelum dianalisis. Ini termasuk penanganan nilai yang hilang, konversi tipe data, dan transformasi data.

b) Pemodelan Prediktif (Predictive Modeling): RapidMiner menyediakan berbagai algoritma machine learning untuk membangun model prediktif. Ini mencakup algoritma klasifikasi, regresi, pengelompokan, dan asosiasi.

c) Evaluasi Model (Model Evaluation): RapidMiner memungkinkan pengguna untuk mengevaluasi kinerja model machine learning menggunakan berbagai metrik evaluasi. Ini membantu dalam memilih model terbaik untuk masalah yang diberikan.

d) Visualisasi Data (Data Visualization): RapidMiner menawarkan berbagai alat visualisasi untuk membantu pengguna memahami data dan hasil analisis. Ini mencakup grafik, diagram, dan visualisasi interaktif lainnya.

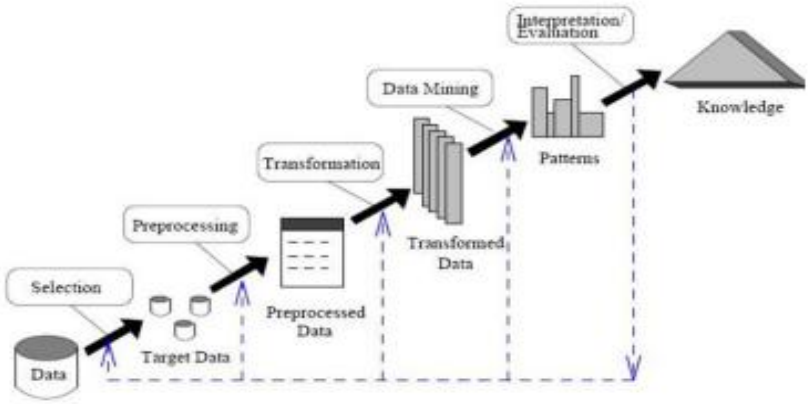
e) Automasi Proses (Process Automation): RapidMiner memungkinkan pengguna untuk mengotomatiskan seluruh alur kerja data science, dari persiapan data hingga evaluasi model. Ini membantu meningkatkan efisiensi dan produktivitas.

f) Integrasi (Integration): RapidMiner dapat diintegrasikan dengan berbagai sumber data, termasuk database, file, dan layanan web. Ini memudahkan pengguna untuk mengakses dan mengolah data dari berbagai sumber.

g) Ekstensibilitas (Extensibility): RapidMiner memiliki arsitektur yang terbuka dan dapat diperluas.

2. Evaluasi Model

Dalam mengevaluasi model, penelitian ini juga menggunakan aplikasi Rapidminner. RapidMiner menawarkan berbagai alat visualisasi untuk membantu pengguna memahami data dan hasil analisis. Ini mencakup grafik, diagram, dan visualisasi interaktif lainnya.



Gambar 3.2 Alur Proses Evaluasi Model

Sumber : Aas Nurjannah, Ahmad Rifai.

3. Analisis Hasil

Analisis hasil adalah proses penting dalam setiap metode ilmiah atau penelitian. Tujuannya adalah untuk memahami dan menginterpretasikan data yang telah dikumpulkan dan diolah, sehingga menghasilkan kesimpulan yang bermakna dan relevan.

Analisis hasil digunakan untuk menguji hipotesis yang telah dirumuskan sebelumnya. Apakah data yang diperoleh mendukung atau menolak hipotesis tersebut. Berdasarkan analisis hasil, peneliti dapat menarik kesimpulan yang valid dan relevan mengenai fenomena yang diteliti.