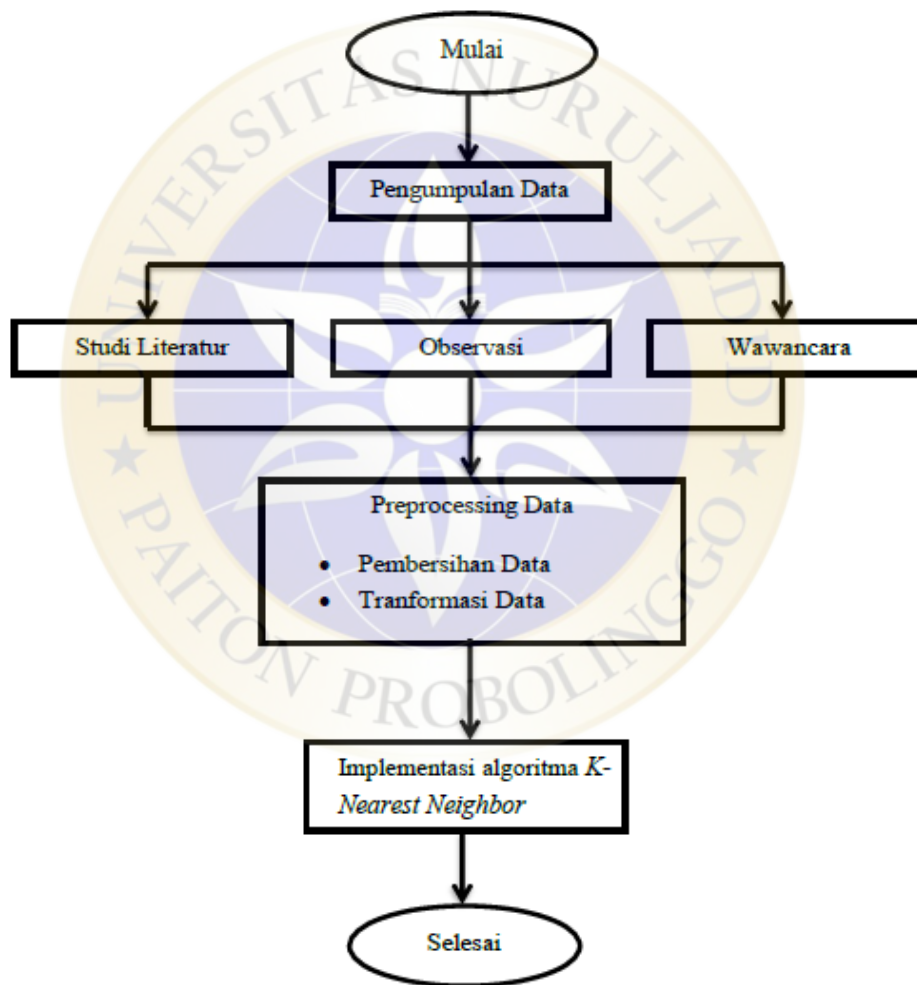


BAB III METODE PENELITIAN

3.1. Kerangka Penelitian

Pada bab ini akan membahas tentang kerangka penelitian yang akan dilakukan dalam menyelesaikan masalah yang dibahas. Adapun kerangka penelitian ini dapat dilihat pada gambar dibawah ini:



Gambar 3.1. Kerangka Penelitian

Dari gambar kerangka penelitian diatas dapat dilihat bahwa secara garis besar proses utama dari penelitian ini ada *study literatur*, pengumpulan data *preprocessing data*, implementasi, hasil dan kesimpulan. Adapun pembahasan lebih rincinya adalah sebagai berikut:

3.2.Pengumpulan Data

Pengumpulan data disini adalah tahap kedua setelah studi literatur yang mana menggunakan observasi dan wawancara.

3.2.1.Studi Literatur

Tahap ini adalah tahap awal peneliti untuk mencari referensi, materi, atau teori yang ingin digunakan. Tahap ini dilakukan untuk membantu peneliti untuk memahami teori dan konsep yang sedang dilakukan, dengan mencari tentang landasan teori dan peneliti sebelumnya yang berkaitan dengan perkembangan belajar secara umum dan metode *K-Nearest Neighbor*. Materi atau teori yang dipaparkan bersumber dari jurnal dan beberapa buku literatur, maupun dari *website* resmi.

3.2.2. Observasi

Observasi adalah cara atau teknik untuk mengumpulkan data dengan melakukan pengamatan dan menganalisa objek yang akan diteliti dilapangan secara langsung untuk mendapatkan data yang valid yang dibutuhkan oleh peneliti.

3.2.3. Wawancara

Wawancara adalah suatu komunikasi semacam percakapan untuk memperoleh suatu informasi. Dilakukan dengan tanya jawab sambil bertatap muka antara penanya dan narasumber.

3.3. Preprocessing Data

Preprocessing data adalah langkah pertama dalam data mining sebagai proses mengubah data mentah ke dalam data yang lebih mudah dipahami. Karena kebanyakan data mentah tidak memiliki format yang teratur dan tidak mudah dipahami. Tahap ini penting karena sebagian besar proses data mining terletak pada tahap *preprocessing* ini dan cukup banyak menghabiskan waktu. *Preprocessing* ini disebut juga preparasi data atau proses sterilisasi data agar pada saat proses data dilakukan memiliki keakuratan pola data.

NO	NAMA	KOTA ASAL	TANGGAL LAHIR	STATUS	LEMBAGA	ANGKATAN	TAJWID	BTQ	FA	FIQIH	AQIDAH	UJIAN AKHIR	KETERANGAN
1	Umrotul Hasanah	Probolinggo	14 Oktober 1999	Lama	UNUJA	2018	87	85	80	77	82	82,2	Lulus
2	Indah Indrawati	Probolinggo	01 Mei 1997	Baru	UNUJA	2018	63	70	54	64	65	63,2	Tidak
3	Tartila	Sumenep	02 Mei 2000	Baru	UNUJA	2018	95	74	85	63	91	81,6	Lulus
4	Nur Mardatillah	Denpasar	05 September 1999	Baru	UNUJA	2018	85	70	81	77	80	78,6	Lulus
5	Nur Ayu Febriana	Situbondo	04 Februari 2000	Baru	UNUJA	2018	71	75	64	53	76	67,8	Tidak
6	Aliffida Ussi Adalia	Probolinggo	19 November 1999	Baru	UNUJA	2018	70	70	68	70	68	69,2	Tidak
7	Siti Hamidah	Situbondo	18 Desember 1999	Baru	UNUJA	2018	83	95	85	70	80	82,6	Lulus
8	Siti Fatimah	Probolinggo	12 November 1999	Baru	UNUJA	2018	79	63	63	70	54	65,8	Tidak
9	Ummi Rifqotul Rofiqoh	Singaraja	21 Maret 2000	Baru	UNUJA	2018	81	86	75	73	84	79,8	Lulus
10	Zakiyatus Safiroh	Jember	14 Oktober 2000	Baru	UNUJA	2018	86	76	85	70	75	78,4	Lulus
11	Nadzirotul Jamilah	Sumenep	19 Juli 2001	Baru	UNUJA	2018	86	85	76	83	76	81,2	Lulus
12	Nur Hilla Lina Kamila	Situbondo	28 Maret 1997	Baru	UNUJA	2018	71	69	74	70	73	71,4	Lulus
13	Isyti Rozanah	Sumenep	11 September 2001	Lama	UNUJA	2018	76	73	91	85	72	79,4	Lulus
14	Mahbubatul Hafifi	Bondowoso	14 Januari 2000	Lama	UNUJA	2018	85	74	82	86	72	79,8	Lulus
15	Hanifatul Rodiyah	Situbondo	24 Mei 1999	Lama	UNUJA	2018	80	80	73	95	75	80,6	Lulus
16	Ika Maziyyatus Sholihah	Probolinggo	25 November 1998	Lama	UNUJA	2018	75	75	81	72	70	74,6	Lulus
17	Nur Khoififah	Bondowoso	23 Januari 1999	Baru	UNUJA	2018	63	69	61	71	68	66,4	Tidak
18	Wildatul Aluf	Situbondo	22 November 2000	Lama	UNUJA	2018	80	85	73	74	71	76,6	Lulus
19	Yeni Cahyani	Usuku	30 Juli 1999	Baru	UNUJA	2018	54	65	66	40	59	56,8	Tidak
20	Siti Qomariah	Lumajang	23 Oktober 1999	Baru	UNUJA	2018	74	90	75	75	70	76,8	Lulus
21	Kholifatun Nisa'	Jember	27 Januari 2001	Lama	UNUJA	2018	82	87	90	84	81	84,8	Lulus
22	Kamalayah	Gresik	10 Agustus 2000	Lama	UNUJA	2018	78	76	72	70	81	75,4	Lulus

Gambar 3.2 Data Awal

3.3.1. Pembersihan Data

Dalam proses ini ada beberapa langkah yang harus dilakukan terlebih dahulu, dimulai dengan pembersihan dan pengecekan data pada data yang sudah didapatkan sehingga dapat dibedakan jika ada data yang kurang lengkap atau ada data yang salah. Selanjutnya dilakukan transformasi data dari data yang telah diperoleh, namun sebelum proses transformasi data harus ditentukan atribut apa yang akan dianalisis.

Karena pada proses ini tidak semua atribut dipilih sehingga perlu adanya penyeleksian atribut mana saja yang akan digunakan pada penelitian ini. Pada Gambar 3.2 sebelumnya terdapat 12 atribut yang terdiri dari Nama, Kota asal, Tanggal Lahir, Status, Lembaga, Angkatan, Tajwid, BTQ, FA, Fiqih, Aqidah, Nilai Akhir dan Keterangan. Akan tetapi setelah dilakukan penyeleksian hanya 9 atribut saja yang dipilih pada penelitian ini, yaitu Nama, Status, BTQ, FA, Fiqih, Aqidah dan Keterangan.

NAMA	STATUS	TAJWID	BTQ	FA	FIQIH	AQIDAH	Nilai Akhir	KETERANGAN
Umrotul Hasanah	Lama	87	85	80	77	82	82,2	Lulus
Indah Indrawati	Baru	63	70	54	64	65	63,2	Tidak
Tartila	Baru	95	74	85	63	91	81,6	Lulus
Nur Mardatillah	Baru	85	70	81	77	80	78,6	Lulus
Nur Ayu Febriana	Baru	71	75	64	53	76	67,8	Tidak
Aliffida Ussi Adalia	Baru	70	70	68	70	68	69,2	Tidak
Siti Hamidah	Baru	83	95	85	70	80	82,6	Lulus
Siti Fatimah	Baru	79	63	63	70	54	65,8	Tidak
Ummi Rifqotur Rofiqoh	Baru	81	86	75	73	84	79,8	Lulus
Zakiyatus Safiroh	Baru	86	76	85	70	75	78,4	Lulus
Nadzrotul Jamilah	Baru	86	85	76	83	76	81,2	Lulus
Nur Hilla Lina Kamila	Baru	71	69	74	70	73	71,4	Lulus
Isyti Rozanah	Lama	76	73	91	85	72	79,4	Lulus
Mahbubatul Hafifi	Lama	85	74	82	86	72	79,8	Lulus
Hanifatur Rodiyah	Lama	80	80	73	95	75	80,6	Lulus
Ika Maziyatus Sholihah	Lama	75	75	81	72	70	74,6	Lulus
Nur Khofifah	Baru	63	69	61	71	68	66,4	Tidak
Wildatul Aluf	Lama	80	85	73	74	71	76,6	Lulus
Yeni Cahyani	Baru	54	65	66	40	59	56,8	Tidak
Siti Qomariah	Baru	74	90	75	75	70	76,8	Lulus
Kholifatun Nisa'	Lama	82	87	90	84	81	84,8	Lulus

Gambar 3.3. Hasil seleksi atribut

3.3.2. Tranformasi Data

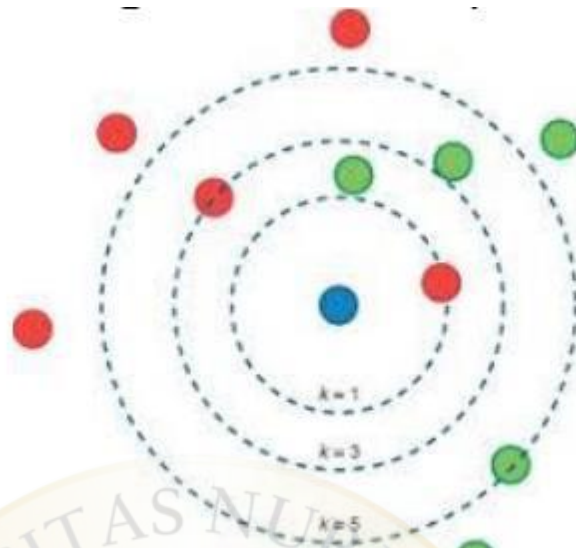
Tahap ini adalah tahap dimana data yang diperoleh diubah atau dikonversi menjadi *numeric*. Data yang akan dikonversi disini hanya di Keterangan, 1= Lulus dan 0= Tidak. Sedangkan untuk atribut yang lain masih tetap.

NAMA	STATUS	TAJWID	BTQ	FA	FIQIH	AQIDAH	Nilai Akhir	Keterangan	
Umrotul Hasanah	Lama	87	85	80	77	82	82,2	1	
Indah Indrawati	Baru	63	70	54	64	65	63,2	0	Lulus = 1
Tartila	Baru	95	74	85	63	91	81,6	1	Tidak = 0
Nur Mardatillah	Baru	85	70	81	77	80	78,6	1	
Nur Ayu Febriana	Baru	71	75	64	53	76	67,8	0	
Aliffida Ussi Adalia	Baru	70	70	68	70	68	69,2	0	
Siti Hamidah	Baru	83	95	85	70	80	82,6	1	
Siti Fatimah	Baru	79	63	63	70	54	65,8	0	
Ummi Rifqotur Rofiqoh	Baru	81	86	75	73	84	79,8	1	
Zakiyatus Safiroh	Baru	86	76	85	70	75	78,4	1	
Nadzירותul Jamilah	Baru	86	85	76	83	76	81,2	1	
Nur Hilla Lina Kamila	Baru	71	69	74	70	73	71,4	1	
Isyti Rozanah	Lama	76	73	91	85	72	79,4	1	
Mahbubatul Hafifi	Lama	85	74	82	86	72	79,8	1	
Hanifatur Rodiyah	Lama	80	80	73	95	75	80,6	1	
Ika Maziyyatus Sholihah	Lama	75	75	81	72	70	74,6	1	
Nur Khofifah	Baru	63	69	61	71	68	66,4	0	
Wildatul Aluf	Lama	80	85	73	74	71	76,6	1	
Yeni Cahyani	Baru	54	65	66	40	59	56,8	0	
Siti Qomariah	Baru	74	90	75	75	70	76,8	1	
Kholifatun Nisa'	Lama	82	87	90	84	81	84,8	1	

gambar 3.4. Hasil Tranformasi Data

3.4. Implementasi Algoritma *K-Nearest Neighbor*

Pada tahap ini data dianalisis dan variabel dikelompokkan satu sama lain. Kemudian, setelah data dianalisis, model diterapkan sesuai dengan jenis datanya. pembagian data menjadi data latih dan data uji juga diperlukan untuk pemodelan. Algoritma *K-Nearest Neighbor* (K-NN) adalah metode yang digunakan untuk melakukan klasifikasi terhadap objek baru berdasarkan k tetangga terdekatnya. K-NN adalah algoritma pembelajaran *supervised learning*, di mana hasil query baru diklasifikasikan menurut mayoritas kategori di K-NN. Maka kelas yang paling banyak muncul adalah kelas hasil klasifikasi. *Nearest Neighbor* adalah metode pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, dimana klasifikasi algoritma ini menggunakan voting terbanyak dari obyek k dan menggunakan ketetanggan sebagai nilai prediksi. (Mustafa & Simpen, n.d.)



Gambar 3.5. Ilustrasi *K-Nearest Neighbor*

Langkah-langkah *K-Nearest Neighbor* sebagai berikut:

1. Menentukan parameter k (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak *Euclidean* antara data uji baru yang akan dievaluasi
3. Mengurutkan jarak *Euclidean* dari yang terkecil
4. Mengumpulkan kategori y berdasarkan nilai k
5. Mencari jumlah kelas yang sering muncul sebagai hasil prediksi. (Nurul et al., 2021)

3.4.1. Persiapan Data

Pada tahap ini yang harus kita lakukan adalah dengan menyiapkan data latih terlebih dahulu. Data Latih adalah data yang diambil dari waktu sebelumnya yang sudah diketahui kelas atau labelnya. Pembagian dataset ini memiliki perbandingan 75% : 25%. Dengan perbandingan tersebut maka jumlah data training adalah 173 data sedangkan data testing adalah 58 data. Hasil pembagian data training tersebut ditunjukkan pada Tabel 3.1 sedangkan data testing ditunjukkan pada Tabel 3.2. hasil pembagian data training dan data testing ini selanjutnya akan digunakan dalam proses klasifikasi dengan menggunakan *K-Nearest Neighbor*.

Tabel 3.1. Sampel Data *Training*

NO	X1	X2	X3	X4	X5	X6	X7
1	76	65	75	90	70	75,2	1
2	76	82	80	83	78	79,8	1
3	70	63	70	65	65	66,6	0
4	76	75	81	76	70	75,6	1
5	78	68	71	77	70	72,8	1
.
.
.
169	82	80	80	70	60	74,4	1
170	77	63	87	72	65	72,8	1
171	67	80	85	85	70	77,4	1
172	83	87	79	93	85	85,4	1
173	87	80	84	82	65	79,6	1

Tabel 3.2. Samapel Data *Testing*

NO	X1	X2	X3	X4	X5	X6	X7
1	65	85	77	73	82	76,4	?
2	85	95	85	70	70	80,4	?
3	86	76	72	85	85	76,6	?
4	65	89	85	85	85	78,6	?
5	85	90	67	71	71	77,6	?
.
.
.
54	84	87	75	88	73	81,4	?
55	75	78	83	79	88	80,6	?
56	70	76	80	74	74	76,6	?
57	63	69	61	71	71	66,4	?
58	81	85	76	85	85	81,4	?

Data latih yang saya gunakan diatas memiliki 6 atribut data yang bertipe numerik. Dan untuk kelasnya menggunakan angka 1 = Lulus dan angka 0 = Tidak.

3.4.2. Normalisasi Data

Tahap selanjutnya adalah membuat tabel untuk perhitungan normalisasinya. Pada penelitian ini menggunakan rumus *min-max* untuk normalisasinya.

$$\text{Normalisasi} = \frac{X - \text{data}_{\min}}{\text{data}_{\max} - \text{data}_{\min}}$$

Dimana :

X = Nilai pertama pada kolom

data_{\max} = Nilai maksimum pada kolom

data_{\min} = Nilai minimum pada kolom

Lakukan perhitungan tersebut pada masing-masing semua atribut.

Tabel 3.3. Sampel Hasil Normalisasi Data

NO	X1	X2	X3	X4	X5	X6	X7
1	0,714285714	0,111111111	1,266666667	2,571428571	0,833333333	1,197183099	1
2	0,714285714	2	1,6	0	1,055555556	1,521126761	1
3	0,285714286	-0,111111111	0,933333333	1,571428571	0,694444444	0,591549296	0
4	0,714285714	1,222222222	1,666666667	1,714285714	0,833333333	1,225352113	1
5	0,857142857	0,444444444	1	3,571428571	0,833333333	1,028169014	1
6	1,714285714	2,444444444	1,6	1,285714286	1,111111111	1,901408451	1
7	1,357142857	1	1,933333333	-1,714285714	0,944444444	1,352112676	1
8	0,357142857	1,222222222	0,533333333	4,428571429	1	0,676056338	0
9	1,357142857	1,777777778	2,4	-2,142857143	1,333333333	2,112676056	1
10	-0,071428571	1,777777778	2,266666667	0	1,472222222	1,225352113	1
11	-0,071428571	1,222222222	2,466666667	1,571428571	0,555555556	0,943661972	1
12	0,071428571	0,666666667	0,133333333	2,857142857	1,111111111	0,845070423	1
13	0,642857143	2,666666667	1,933333333	2,285714286	1,25	1,788732394	1
14	0,785714286	2,111111111	1,933333333	0,714285714	0,916666667	1,521126761	1
15	1	0,888888889	0,8	1,857142857	0,5	0,802816901	0
16	0,285714286	2,333333333	1,6	-0,571428571	1,333333333	1,549295775	1
17	-0,071428571	1,222222222	0,466666667	3	0,611111111	0,492957746	1
18	1,357142857	1,111111111	1,733333333	1,714285714	0,888888889	1,521126761	1
19	1,357142857	0,666666667	1,666666667	2,857142857	1,111111111	1,436619718	1
20	0,642857143	1,777777778	1	-2,571428571	1,055555556	1,38028169	1

3.4.3. Menentukan Parameter k

Sebenarnya tidak ada cara khusus dalam menentukan nilai terbaik untuk menentukan nilai k terbaik dalam algoritma *K-Nearest Neighbor*. Nilai k yang sangat kecil, misalnya $k = 2$ atau $k = 3$ akan memberikan hasil yang kurang akurat. Sedangkan nilai k terlalu besar dapat memberikan hasil yang lebih akurat dan batas yang lebih bagus. Oleh karena itu nilai k yang digunakan pada kasus ini adalah $k = 9$. Nilai $k = 9$ merupakan nilai ganjil yang dipilih untuk menghindari jumlah kelas yang sama pada saat perhitungan kelas k .

3.4.4. Menghitung Euclidean Distance atau Jarak

Setelah menentukan nilai $k = 9$. Langkah selanjutnya adalah menghitung jarak antara data *training* dan data *testing* menggunakan rumus *Euclidean* menggunakan *Software Microsoft Excel*.

Perhitungan data *training* pertama terhadap data *testing* pertama

$$\begin{aligned}d(1,174) &= \sqrt{\sum_{i=1}^n (X_i - X_{174})^2} \\d(1,174) &= \\&= \sqrt{(X_{2baris\ 1} - X_{2baris\ 174})^2 + (X_{3baris\ 1} - X_{3baris\ 174})^2 \\&\quad + (X_{4baris\ 1} - X_{4baris\ 174})^2 \\&\quad + (X_{5baris\ 1} - X_{5baris\ 174})^2 \\&\quad + (X_{6baris\ 1} - X_{6baris\ 174})^2} \\d(1,174) &= \\&= \sqrt{(0,714285714 - -0,071428571)^2 + (0,111111111 - \\&\quad 2,777777778)^2 + (1,266666667 - 1,4)^2 + (2,571428571 + \\&\quad 0,714285714)^2 + (0,833333333 - 1,166666667)^2 + \\&\quad (1,197183099 - 1,281690141)^2} \\d(1,174) &= 3,023347229\end{aligned}$$

Perhitungan data *training* kedua terhadap data *testing* pertama

$$\begin{aligned}
d(2,174) &= \sqrt{(X2_{\text{baris } 2} - X2_{\text{baris } 174})^2 + (X3_{\text{baris } 2} - X3_{\text{baris } 174})^2} \\
&\quad + (X4_{\text{baris } 2} - X4_{\text{baris } 174})^2 + (X5_{\text{baris } 2} - X5_{\text{baris } 174})^2 \\
&\quad + (X6_{\text{baris } 2} - X6_{\text{baris } 174})^2 \\
d(2,174) &= \sqrt{(0,714285714 - 0,071428571)^2 + (2 - 2,777777778)^2 + (1,6 - 1,4)^2} \\
&\quad + (0 - 0,714285714)^2 + (1,055555556 - 1,166666667)^2 \\
&\quad + (1,521126761 - 1,281690141)^2 \\
d(2,174) &= 1,161179446
\end{aligned}$$

Perhitungan seperti contoh diatas dilakukan terhadap data *training* pertama sampai dengan terakhir (1-173) dan data *testing* pertama samapi terakhir (174- 231), sehingga akan diperoleh hasil seperti pada Tabel 3.4 dibawah ini :

Tabel 3.4. Hasil Pengurutan *Distance*

NO	Distance	Rangking	Kelas
1	3,023347229	129	1
2	1,161179446	10	1
3	2,784711175	112	0
4	1,742788293	42	1
5	3,595663431	155	1
6	1,988665131	61	1
7	3,171015576	139	1
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
167	2,035960789	65	1
168	1,244683438	12	1
169	1,515934876	27	1
170	3,468940831	150	1
171	3,394920234	146	1
172	2,251143639	80	1
173	1,858294481	50	1

3.4.5. Menentukan Kelas Mayoritas

Dengan mengurutkan jarak k terdekat berdasarkan nilai jarak minimum. Sehingga dari seluruh data pengurutan nilai jarak pada Tabel 3.4, akan diambil 9 data teratas, sebagaimana pada tabel 3.5. Hasil prediksi perkembangan belajar dapat dilihat dengan melihat perbandingan kelas yang ada. Jika jumlah data dengan kelas Lulus lebih banyak dibandingkan dengan kelas Tidak Lulus maka hasil prediksinya adalah Lulus. Dan jika jumlah data dengan kelas Lulus lebih sedikit dari kelas Tidak Lulus maka hasil prediksinya adalah Tidak Lulus.

Tabel 3.5. Hasil Prediksi Data Uji Satu dengan $k = 9$

NO	Euclidian Distance	Rangking	Hasil
86	0,197328378	1	1
36	0,535945718	2	1
100	0,724952106	3	1
143	1,008581148	4	1
149	1,046494387	5	1
156	1,080015196	6	1
14	1,090115101	7	1
132	1,145004166	8	1
114	1,15457396	9	1

Tabel 3.6 Hasil Penentuan Kelas

Ket	Hasil 1	Hasil 0
K = 1	1	0
K = 3	3	0
K = 5	5	0
K = 7	7	0
K = 9	9	0
K = 11	11	0
K = 13	13	0

Dari hasil penentuan kelas mayoritas pada Tabel 3.6 data testing pertama dengan jumlah kelas berupa $0 = 0$. Dan jumlah kelas berupa $1 = 9$. Maka dapat dipastikan bahwa data testing pertama masuk dalam kategori 1 yaitu Lulus.