

Penerapan Algoritma K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa

Rizki Alexander Manullang¹, Fricles Ariwisanto Sianturi²

^{1,2} Program Studi Teknik Informatika, STMIK Pelita Nusantara Jl. St. Iskandar Muda No. 1 Medan, Indonesia
¹ rizkimanullang03@gmail.com; ² sianturifricles@gmail.com;

ABSTRAK

Kata Kunci:
Data Mining
K-Nearest Neighbor
Prediksi Kelulusan

STMIK Pelita Nusantara Medan merupakan salah satu lembaga perguruan tinggi komputer Swasta yang berada di Kota Medan Provinsi Sumatera Utara yang memiliki visi menjadi perguruan tinggi unggul dan pusat pendidikan komputer yang menghasilkan lulusan berdaya saing nasional pada tahun 2024. Dalam mewujudkan visi STMIK Pelita Nusantara pada tahun 2024, terdapat beberapa masalah yang dialami oleh pihak perguruan tinggi, yang menyebabkan visi tersebut tidak dapat terpenuhi dengan baik, salah satunya adalah tidak seimbangnya ratio jumlah mahasiswa yang mendaftar dengan jumlah mahasiswa yang lulus dalam setiap tahunnya. seperti yang terjadi pada data kelulusan mahasiswa prodi Teknik Informatika pada tahun 2018, terdapat sejumlah 212 orang mahasiswa yang mendaftar pada tahun 2014 ke perguruan tinggi STMIK Pelita Nusantara. Dan yang berhasil menyelesaikan jenjang pendidikan akademik S1 (Strata Satu) pada tahun 2018 tersebut hanya 120 orang, dapat disimpulkan bahwa terjadi penurunan jumlah kelulusan mahasiswa yang kurang seimbang. Teknik data mining banyak digunakan untuk mengatasi banyak permasalahan. Salah satunya dengan kemampuan yang dimiliki teknik data *mining* yaitu klasifikasi, klasifikasi merupakan suatu teknik dalam data *mining* untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel, mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot disebut dengan Algoritma *K-Nearest Neighbor*. Berdasarkan hasil uji coba dengan *Algoritma K-Nearest Neighbor* diharapkan dapat membantu STMIK Pelita Nusantara dalam memprediksi kelulusan mahasiswa dengan efektif.

Keywords:
Data Mining
K-Nearest Neighbor
Graduation Prediction

ABSTRACT

STMIK Pelita Nusantara Medan is one of the private computer tertiary institutions located in Medan City, North Sumatra Province which has a vision of becoming a superior university and computer education center that produces graduates with national competitiveness in 2024. In realizing the vision of STMIK Pelita Nusantara in 2024, there are several problems experienced by the university, which causes the vision to not be fulfilled properly, one of which is the unbalanced ratio of the number of students who register with the number of students who graduate each year. As happened in the graduation data of Informatics Engineering Study Program students in 2018, there were a total of 212 students who registered in 2014 to STMIK Pelita Nusantara College. And those who managed to complete the undergraduate academic education level (Strata One) in 2018 were only 120 people, it can be concluded that there was an unbalanced decrease in the number of student graduations. Data mining techniques are widely used to overcome many problems, one of which is the ability of data mining techniques, namely classification, classification is a technique in data mining to group data based on data attachment to sample data, look for cases by calculating the proximity between new cases and old cases. based on weight matching is called the K-Nearest Neighbor Algorithm. Based on

the results of trials with the K-Nearest Algorithm, it is hoped that it can help STMIK Pelita Nusantara in predicting student graduation effectively.

I. Pendahuluan

Seiring dengan berkembangnya teknologi informasi pada saat ini yang semakin berkembang pesat, kebutuhan akan informasi menjadi salah satu hal yang sangat penting dan semakin dibutuhkan dalam berbagai sektor, sebagai contoh adalah dalam sektor pendidikan. Dalam sistem lembaga perguruan tinggi, unsur terpenting dalam penyelenggaraan perguruan tinggi adalah mahasiswa [2], STMIK Pelita Nusantara Medan merupakan salah satu lembaga sektor pendidikan perguruan tinggi komputer swasta yang berada di kota Medan Provinsi Sumatera Utara yang memiliki visi menjadi perguruan tinggi unggul dan pusat pendidikan komputer yang menghasilkan lulusan berdaya saing nasional pada tahun 2024. Dalam mewujudkan visi STMIK Pelita Nusantara pada tahun 2024, terdapat beberapa masalah yang dialami oleh pihak perguruan tinggi, yang menyebabkan visi tersebut tidak dapat terpenuhi dengan baik, salah satunya adalah tidak seimbang ratio jumlah mahasiswa yang mendaftar dengan jumlah mahasiswa yang lulus dalam setiap tahunnya. Seperti yang terjadi pada data kelulusan mahasiswa prodi Teknik Informatika pada tahun 2018, terdapat sejumlah 212 orang mahasiswa yang mendaftar pada tahun 2014 ke perguruan tinggi STMIK Pelita Nusantara, yang berhasil menyelesaikan jenjang pendidikan akademik S1 (Strata Satu) pada tahun 2018 tersebut hanya 120 orang, dari jumlah tersebut dapat kita simpulkan bahwa terjadi penurunan jumlah kelulusan mahasiswa yang kurang seimbang sekitar 40% yang tidak lulus, yang tentunya dapat mempengaruhi mimpi STMIK Pelita Nusantara dalam mewujudkan visi pada tahun 2024.

Berdasarkan latar belakang masalah diatas maka identifikasi masalah dalam penelitian adalah ditemukannya kesulitan dalam meminimalisir jumlah mahasiswa yang tidak lulus dalam setiap tahunnya. Dan batasan masalah dari penelitian ini adalah, atribut yang digunakan dalam penelitian prediksi kelulusan ini adalah IP semester 1-6, data kelulusan yang akan di prediksi merupakan data tahun 2020 dan data yang digunakan merupakan data tahun 2018 program studi Teknik Informatika STMIK Pelita Nusantara. Dari masalah yang teridentifikasi tersebut, dengan demikian penulis membuat rumusan. Bagaimana merancang Aplikasi Algoritma *K-Nearest Neighbour* dalam memprediksi kelulusan mahasiswa dan data apa yang digunakan untuk melakukan klasifikasi prediksi kelulusan mahasiswa. Dengan data mining kita dapat melakukan pengklasifikasian, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan data dalam jumlah yang besar [3]. Data *Mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [4], Data *mining* adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis [6]. Tujuan dari penelitian yang dilakukan adalah untuk mengetahui hasil analisa dari prediksi kelulusan mahasiswa menggunakan teknik klasifikasi data mining dan menerapkan perhitungan metode Algoritma *K-Nearest Neighbor* dengan cara memanfaatkan data sebelumnya dengan menguji data tersebut dengan algoritma KNN dan merancang suatu aplikasi berbasis *web* yang didalamnya menerapkan metode KNN untuk memprediksi kelulusan mahasiswa. Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [5]. *K-Nearest Neighbor* (KNN) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample [1]. *Nearest Neighbor* adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [7].

II. Metode

2.1 Tahapan Penelitian

1. Pengumpulan Data

Pada tahap ini merupakan aktifitas mencari atau mengumpulkan data untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian seperti:

- a. Literatur
Pada tahap studi literatur penelitian ini mengumpulkan dasar-dasar teori dan informasi dari buku dan jurnal yang berkaitan dengan kebutuhan penelitian terkait dengan pokok bahasan masalah. Dalam studi literatur dilakukan mengenai studi kasus yang berkaitan dengan implementasi data mining menggunakan algoritma *K-Nearest Neighbor*.
- b. Observasi
Observasi adalah aktifitas terhadap suatu proses atau objek dengan cara mengamati secara langsung untuk mendapatkan informasi yang jelas. Tujuan observasi adalah untuk mendeskripsikan kegiatan yang terjadi serta orang-orang yang terlibat di dalamnya. Dengan adanya metode observasi ini peneliti gunakan untuk mengumpulkan data secara langsung, adapun data yang dikumpulkan yaitu data mahasiswa program studi Teknik Informatika dan data kelulusan tahun 2018-2020.
- c. Wawancara
Adalah proses komunikasi lisan yang bertujuan menggali informasi tertentu dengan cara terstruktur dan dilakukan dua orang atau lebih. Tujuan dari wawancara adalah untuk memperoleh informasi dan memperluas data yang didapat dari pihak-pihak lain seperti Kaprodi Teknik Informatika STMIK Pelita Nusantara, dimana pewawancara memberikan pertanyaan-pertanyaan untuk dijawab oleh pihak yang diwawancarai. Hasil dari wawancara tersebut akan dijadikan acuan untuk proses penelitian lebih lanjut.

2.2 Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik *clustering*, yaitu mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga terdekat. Pertama sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (*neighbor*). Lalu, untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan titik pada data testing, maka digunakan rumus *Euclidean* dengan persamaan:

$$d(a, b) = \sum_{i=0}^n (X_i - Y_i)^2 \quad (1)$$

Keterangan:

- d (a,b) : jarak euclidian
x : data 1
y : data 2
i : fitur ke –
n : jumlah fitur. (Dzikrulloh, Indriati, & Setiawan, 2017).

Kedekatan antara dua kasus dapat dihitung dengan mencari nilai *similarity* sebagai berikut:

$$\frac{\sum_{i=1}^n (T_i, S_i) * W_i}{W_i} \quad (2)$$

Keterangan:

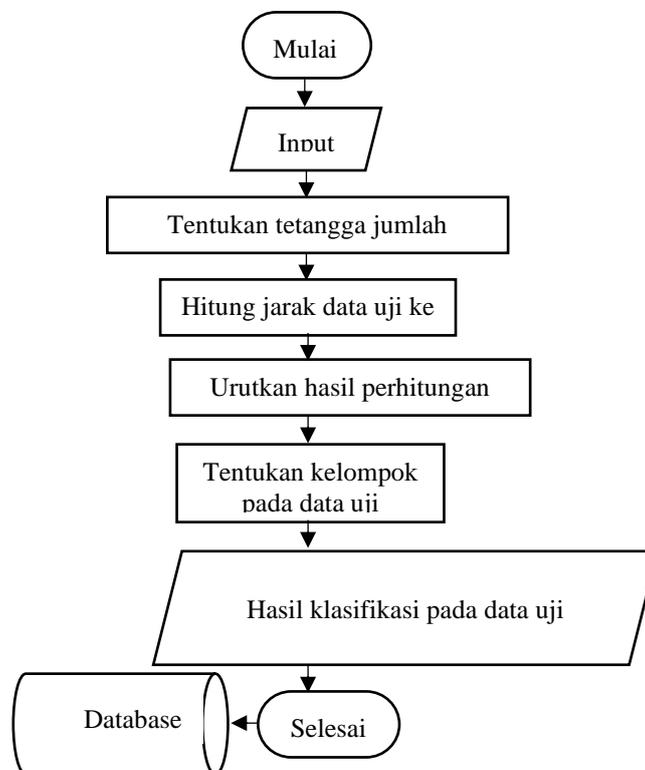
- T : kasus baru
S : kasus yang ada dalam penyimpanan
N : jumlah atribut dalam setiap kasus
I : atribut individu antara 1 s.d n
Wi : bobot yang diberikan pada atribut ke-i

Kemiripan (*similarity*) ini dinyatakan dengan 1 (mirip) dan 0 (tidak mirip), secara matematika dapat ditulis

$$s = \begin{cases} 1 & \text{jika } x = y \\ 0 & \text{jika } x \neq y \end{cases} \quad (3)$$

(Muliono, Lubis, Khairina, 2019).

2.3 Proses Algoritma K-Nearest Neighbor (KNN).



Gambar 1. Proses Metode K-Nearest Neighbor.

Sumber : [1]

Langkah-langkah *K-Nearest Neighbor*:

1. Menentukan parameter K (jumlah tetangga paling dekat), Parameter K pada testing ditentukan berdasarkan nilai K optimum pada saat training.
2. Menghitung kuadrat jarak *euclid* (*euclidean distance*) masing-masing objek terhadap data sampel yang diberikan

3. Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclidian* terkecil
4. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
5. Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi. [2].

III. Hasil dan Pembahasan

3.1 Tahapan Penelitian

Pada tahap pengolahan data, data diperoleh dengan cara observasi dan wawancara secara langsung ke Kampus STMIK Pelita Nusantara. Data yang digunakan adalah data kelulusan mahasiswa Program Studi Teknik Informatika tahun 2018 dan data kelulusan mahasiswa Program Studi Teknik Informatika tahun 2020. Pengambilan data disesuaikan dengan kebutuhan pada penelitian

Berikut merupakan sampel data training dan sampel data testing kelulusan mahasiswa tahun 2018 dan tahun 2020 yang sudah di selection pada table berikut ini.

Tabel 1. Data Sampel (Data *Training*).

No	Nama	Nim	Indeks Prestasi						Keterangan
			S1	S2	S3	S4	S5	S6	
1	Ahmad Sholihin Hulu	140121xxx	2.45	2.67	3.00	3.15	3.25	3.75	Lulus
2	Andrew W Hutapea	140121xxx	2.79	2.89	3.00	3.10	3.33	3.56	Lulus
3	Mustofa J Batubara	140121xxx	0.56	2.00	0	0	0	0	Tidak Lulus
4	Susi Y Sitorus	140121xxx	0	0	1.23	0	0.78	0	Tidak Lulus
5	Windy Manullang	140121xxx	0.46	1.23	0	0	0	0	Tidak Lulus

Tabel 2. Data Sampel (Data *Testing*).

No	Nama	Nim	Indeks Prestasi						Keterangan
			S1	S2	S3	S4	S5	S6	
1	Sarah A Sihombing	160121xxx	3.20	3.12	3.23	3.01	3.56	3.79	?
2	Aditya Utami	160121xxx	2.68	2.79	3.22	3.00	3.78	3.78	?
3	Kordia Br Turnip	160121xxx	1.53	2.67	2.89	3.20	3.45	4.00	?
4	Rista	160121xxx	2.56	2.79	3.22	3.45	3.67	3.89	?
5	Sry Wahyu Ningsih	160121xxx	2.79	3.45	3.20	3.12	3.24	3.78	?

3.2 Analisa Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample. Diberikan suatu titik *query*, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik *query*. Nilai prediksi dari query akan ditentukan berdasarkan klasifikasi tetangga. Sebelum melakukan perhitungan dengan metode K-Nearest Neighbor, terlebih dahulu harus menentukan data latih dan data uji. Kemudian akan dilakukan proses perhitungan untuk mencari jarak menggunakan Euclidean. Setelah itu, akan dilakukan tahapan perhitungan dengan metode KNN. Terdapat beberapa tahapan yang digunakan dalam Algoritma *K-Nearest Neighbor*, yaitu:

a. Langkah 1

Menentukan parameter K (jumlah tetangga paling dekat), misalnya nilai parameter K=3. Dalam hal penentuan K dibutuhkan data training dan data testing.

1. Data Training

Data training adalah: data yang digunakan pada saat pemanfaatan algoritma guna mencari model yang pas.

Berikut data training dapat dilihat pada tabel 3. dibawah ini.

No	Nama	Nim	Indeks Prestasi						Keterangan
			S1	S2	S3	S4	S5	S6	
1	Ahmad Sholihin Hulu	140121xxx	2.45	2.67	3.00	3.15	3.25	3.75	Lulus
2	Andrew W Hutapea	140121xxx	2.79	2.89	3.00	3.10	3.33	3.56	Lulus
3	Mustofa J Batubara	140121xxx	0.56	2.00	0	0	0	0	Tidak Lulus
4	Susi Y Sitorus	140121xxx	0	0	1.23	0	0.78	0	Tidak Lulus
5	Windy Manullang	140121xxx	0.46	1.23	0	0	0	0	Tidak Lulus

2. Data Testing

Data testing adalah: data yang akan di uji ke data training

Berikut data testing dapat dilihat pada tabel 4. dibawah ini:

b. Langkah 2

No	Nama	Nim	Indeks Prestasi						Keterangan
			S1	S2	S3	S4	S5	S6	
1	Sarah A Sihombing	160121xxx	3.20	3.12	3.23	3.01	3.56	3.79	?
2	Aditya Utami	160121xxx	2.68	2.79	3.22	3.00	3.78	3.78	?
3	Kordia Br Turnip	160121xxx	1.53	2.67	2.89	3.20	3.45	4.00	?
4	Rista	160121xxx	2.56	2.79	3.22	3.45	3.67	3.89	?
5	Sry Wahyu Ningsih	160121xxx	2.79	3.45	3.20	3.12	3.24	3.78	?

Menghitung jarak antara data uji dengan data training dengan Euclidean Distance.

Jarak data uji 1 dengan data training

$$\begin{aligned}
 &= \sqrt{(2.45 - 3.20)^2 + (2.67 - 3.12)^2 + (3.00 - 3.23)^2 + (3.15 - 3.01)^2 + (3.25 - 3.56)^2 + (3.75 - 3.79)^2} \\
 &= \sqrt{0.5625 + 0.2025 + 0.0529 + 0.0196 + 0.0961 + 0.0016} \\
 &= \sqrt{0.9352} \\
 &= 0.967057
 \end{aligned}$$

Jarak data uji 2 dengan data training

$$\begin{aligned}
 &= \sqrt{(2.79 - 2.68)^2 + (2.67 - 2.79)^2 + (3.00 - 3.22)^2 + (3.10 - 3.00)^2 + (3.33 - 3.78)^2 + (3.56 - 3.78)^2} \\
 &= \sqrt{0.0121 + 0.0144 + 0.0484 + 0.01 + 0.2025 + 0.0484} \\
 &= \sqrt{0.3358} \\
 &= 0.579482
 \end{aligned}$$

Jarak data uji 3 dengan data training

$$\begin{aligned}
 &= \sqrt{(0.56 - 1.53)^2 + (2.00 - 2.67)^2 + (0 - 2.89)^2 + (0 - 3.20)^2 + (0 - 3.45)^2 + (0 - 4.00)^2} \\
 &= \sqrt{0.9409 + 0.4489 + 8.3521 + 10.24 + 11.9025 + 16} \\
 &= \sqrt{47.8844} \\
 &= 6.919855
 \end{aligned}$$

Jarak data uji 4 dengan data training

$$\begin{aligned}
 &= \sqrt{(0 - 2.56)^2 + (0 - 2.79)^2 + (1.23 - 3.22)^2 + (0 - 3.45)^2 + (0.78 - 3.67)^2 + (0 - 3.89)^2} \\
 &= \sqrt{6.5536 + 7.7841 + 3.9601 + 11.9025 + 8.3521 + 15.1321} \\
 &= \sqrt{53.6845} \\
 &= 7.326970
 \end{aligned}$$

Jarak data uji 5 dengan data training

$$\begin{aligned}
 &= \sqrt{(0.46 - 2.79)^2 + (1.23 - 3.45)^2 + (0 - 3.2)^2 + (0 - 3.12)^2 + (0 - 3.24)^2 + (0 - 3.78)^2} \\
 &= \sqrt{5.4289 + 4.9284 + 10.24 + 9.7344 + 10.4976 + 14.2884} \\
 &= \sqrt{55.1177} \\
 &= 7.424129
 \end{aligned}$$

Tabel 5. Euclidean distance data testing.

No	Nama	Nim	Indeks Prestasi						Keterangan	Euclidean Distance
			S1	S2	S3	S4	S5	S6		
1	Ahmad Sholihin Hulu	140121xxx	2.45	2.67	3.00	3.15	3.25	3.75	Lulus	0.967057
2	Andrew W Hutapea	140121xxx	2.79	2.89	3.00	3.10	3.33	3.56	Lulus	0.579482
3	Mustofa J Batubara	140121xxx	0.56	2.00	0	0	0	0	Tidak Lulus	6.919855
4	Susi Y Sitorus	140121xxx	0	0	1.23	0	0.78	0	Tidak Lulus	7.326970
5	Windy Manullang	140121xxx	0.46	1.23	0	0	0	0	Tidak Lulus	7.424129

Nilai *Euclidean Distance* diatas diperoleh dari hasil perhitungan data uji 1 sampai ke data uji ke-215 berdasarkan langkah 2 diatas.

b. Langkah 3

Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclidean* terkecil.

Tabel 6. Pengurutan Peringkat Jarak.

No	Nama	Nim	Indeks Prestasi						Keterangan	Euclidean Distance	Peringkat Jarak Terkecil
			S1	S2	S3	S4	S5	S6			
1	Ahmad Sholihin Hulu	140121xxx	2.45	2.67	3.00	3.15	3.25	3.75	Lulus	0.967057	69
2	Andrew W Hutapea	140121xxx	2.79	2.89	3.00	3.10	3.33	3.56	Lulus	0.579482	32
3	Mustofa J Batubara	140121xxx	0.56	2.00	0	0	0	0	Tidak Lulus	6.919855	155
4	Susi Y Sitorus	140121xxx	0	0	1.23	0	0.78	0	Tidak Lulus	7.326970	158
5	Windy Manullang	140121xxx	0.46	1.23	0	0	0	0	Tidak Lulus	7.424129	176

Data yang digunakan itu berdasarkan nilai dari tabel 4.6 dimana dilakukan pengurutan nilai *distance* mulai dari angka terkecil hingga ke angka terbesar.

c. Langkah 4

Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)

Setelah mendapatkan peringkat pada jarak masing-masing data, maka pada langkah selanjutnya mengumpulkan kategori Y (*klasifikasi nearest neighbor*). Pada langkah 1, parameter (k) yang kita tentukan adalah 3. Maka berdasarkan perhitungan KNN, dapat hasil klasifikasi seperti berikut ini:

Tabel 7. Hasil Klasifikasi KNN.

No	Nama	Nim	Keterangan	Euclidean Distance	Peringkat Jarak Terkecil	Hasil Klasifikasi
1	Ahmad Sholihin Hulu	140121xxx	Lulus	0.967057	69	Lulus
2	Andrew W Hutapea	140121xxx	Lulus	0.579482	32	Lulus
3	Mustofa J Batubara	140121xxx	Tidak Lulus	6.919855	155	Lulus
4	Susi Y Sitorus	140121xxx	Tidak Lulus	7.326970	158	Lulus
5	Windy Manullang	140121xxx	Tidak Lulus	7.424129	176	Lulus

Pada tabel 7 hasil klasifikasi KNN, data no 1 memperoleh hasil klasifikasi lulus, hasil tersebut diperoleh dari hasil nilai *distance* (0.967057) yang diurutkan ke semua jumlah *distance* yang sudah dihitung berdasarkan nilai K=3 atau tetangga terdekat yang masuk ke peringkat sejumlah K.

d. Langkah 5

Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksikan kategori objek.

Setelah mengumpulkan kategori Y dengan K=3, maka hasil pengujian data testing 1 ke sejumlah data training sebanyak 210 data, memperoleh hasil klasifikasi bahwa data testing 1 berada dalam klasifikasi LULUS. Hasil kategori mayoritas dapat dilihat pada tabel dibawah ini:

Tabel 8. Hasil Kategori Mayoritas

No	Nama	Keterangan	Euclidean Distance	Peringkat Jarak Terkecil	Klasifikasi
1	Ahmad Sholihin Hulu	Lulus	0.414487	1	Lulus
2	Andrew W Hutapea	Lulus	0.439431	2	Lulus
3	Mustofa J Batubara	Tidak Lulus	0.468401	3	Lulus

Pada tabel 4.9 menunjukkan bahwa atribut klasifikasi LULUS yang paling banyak muncul. Dengan begitu maka dapat disimpulkan bahwa suatu data dengan Nama = Sarah S Sihombing. Nim = 140121xxx. S1=3.20. S2=3.12. S3=3.23. S4=3.01. S5=3.56. S6=3.79 Maka hasil prediksinya adalah LULUS. Untuk keterangan hasil klasifikasinya bisa dilihat pada tabel 4.10 baris 6 dibawah ini:

Tabel 9. Hasil Semua Klasifikasi Mayoritas

No	Nama	Nim	Indeks Prestasi						Keterangan
			S1	S2	S3	S4	S5	S6	
1	Ahmad Sholihin Hulu	140121xxx	2.45	2.67	3.00	3.15	3.25	3.75	Lulus
2	Andrew W Hutapea	140121xxx	2.79	2.89	3.00	3.10	3.33	3.56	Lulus
3	Mustofa J Batubara	140121xxx	0.56	2.00	0	0	0	0	Tidak Lulus
4	Susi Y Sitorus	140121xxx	0	0	1.23	0	0.78	0	Tidak Lulus
5	Windy Manullang	140121xxx	0.46	1.23	0	0	0	0	Tidak Lulus
6	Sarah A Sihombing	140121xxx	3.20	3.12	3.23	3.01	3.56	3.79	Lulus

Berdasarkan dari hasil pengujian prediksi kelulusan menggunakan *Algoritma K-Nearest Neighbor*, jumlah data testing sebanyak 215 data yang di uji terhadap data training sebanyak 210 data, diperoleh hasil klasifikasi LULUS sebanyak 148 orang, dan klasifikasi TIDAK LULUS sebanyak 67 orang. Hasil klasifikasi dapat dilihat pada tabel dibawah ini:

Tabel 10. Hasil Pengujian Keseluruhan Data testing sebanyak 215 Data

No	Nama	Nim	Indeks Prestasi						Sebelum Diuji	Setelah Diuji
			S1	S2	S3	S4	S5	S6		
1	Sarah A Sihombing	160121xxx	3.20	3.12	3.23	3.01	3.56	3.79	?	Lulus
2	Aditya Utami	160121xxx	2.68	2.79	3.22	3.00	3.78	3.78	?	Lulus
3	Kordia Br Turnip	160121xxx	1.53	2.67	2.89	3.20	3.45	4.00	?	Lulus
4	Rista	160121xxx	2.56	2.79	3.22	3.45	3.67	3.89	?	Lulus
5	Sry Wahyu Ningsih	160121xxx	2.79	3.45	3.20	3.12	3.24	3.78	?	Lulus

IV. Kesimpulan

Berdasarkan penelitian yang dilakukan mengenai Penerapan Algoritma K-Nearest Neighbour untuk Memprediksi Kelulusan Mahasiswa, maka dapat diambil kesimpulan sebagai berikut:

1. Dengan menguji data sebanyak 215 data dengan menerapkan metode KNN, didapatkan hasil perhitungan jumlah mahasiswa yang lulus untuk Program Studi Teknik Informatika sebanyak 148 mahasiswa, dan jumlah yang tidak lulus sebanyak 67 mahasiswa.

2. Dari hasil penelitian di atas tentang Penerapan Algoritma *K-Nearest Neighbor* untuk memprediksi kelulusan mahasiswa, data yang digunakan sebagai data training merupakan data kelulusan mahasiswa tahun 2018, dan data yang digunakan sebagai data testing merupakan data kelulusan mahasiswa tahun 2020 jurusan Teknik Informatika dengan atribut yang digunakan adalah: IP semester 1-6.
3. Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa metode KNN dapat digunakan untuk memprediksi kelulusan mahasiswa di STMIK Pelita Nusantara.

Daftar Pustaka

- [1] Eska, J. (2016). Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksieran. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*.
- [2] Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM: Jurnal Teknologi Informasi Dan Multimedia*. <https://doi.org/10.35746/jtim.v1i1.11>
- [3] Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- [4] Mulyati, S., Husein, S. M., & Ramdhan, R. (2020). RANCANG BANGUN APLIKASI DATA MINING PREDIKSI KELULUSAN UJIAN NASIONAL MENGGUNAKAN ALGORITMA (KNN) K-NEAREST NEIGHBOR DENGAN METODE EUCLIDEAN DISTANCE PADA SMPN 2 PAGEDANGAN. *JIKA (Jurnal Informatika)*. <https://doi.org/10.31000/jika.v4i1.2288>
- [5] Rohman, A. (2015). MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK PREDIKSI KELULUSAN MAHASISWA. *Neo Teknika*. <https://doi.org/10.37760/neoteknika.v1i1.350>
- [6] Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. In *Jurnal Teknik Komputer AMIK BSI (JTK)*.
- [7] Mustafa, M. S., & Simpen, I. W. (2015). Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar). *Creative Information Technology Journal*. <https://doi.org/10.24076/citec.2014v1i4.27>