

PERBANDINGAN ALGORITME K-NEAREST NEIGHBOR DAN NAÏVE BAYES DALAM KLASIFIKASI KELULUSAN MAHASISWA

Muhammad Fahmi As Shiddiqi, Henny Dwi Bhakti

Teknik Informatika, Universitas Muhammadiyah Gresik

Jl. Sumatera No.101 Gn. Malang, Randuagung, Kec. Kebomas, Kabupaten Gresik, Jawa Timur 61121

fahmiasdq90@gmail.com

ABSTRAK

Kelulusan merupakan tujuan akhir bagi setiap orang dalam rangkaian proses menyelesaikan studi dalam memperoleh gelar, penghargaan ataupun yang lainnya. Pada Program Studi Teknik Informatika Universitas Muhammadiyah Gresik terdapat banyak mahasiswa yang lulus tidak tepat waktu (terlambat) dan mahasiswa yang *Drop Out* (DO), hal tersebut merupakan masalah penting bagi Program Studi, dikarenakan jumlah total dari mahasiswa yang mendaftar tidak sebanding dengan mahasiswa lulus yang berdampak pada penilaian Akreditasi dimana akan berakibat pada menurunnya calon mahasiswa baru di Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. Sangat penting bagi Program Studi mengetahui klasifikasi kelulusan mahasiswa. Dalam penelitian ini akan dilakukan klasifikasi kelulusan mahasiswa menggunakan algoritme *K-Nearest Neighbor* dan *Naïve Bayes*. Hasil pengujian evaluasi performa menggunakan *Confusion Matrix* didapatkan nilai rata-rata *precision* dan *recall* dari kedua algoritme. Algoritme *K-Nearest Neighbor* mendapatkan rata-rata nilai *precision*, dan *recall* lebih tinggi sebesar 85,1% dan 92,9%. Sedangkan untuk algoritme *Naïve Bayes* mendapatkan nilai rata-rata *precision* dan *recall* lebih rendah sebesar 73% dan 73,5%. Sementara hasil pengujian menggunakan *K-Fold Cross Validation* mendapatkan hasil bahwa algoritme *K-Nearest Neighbor* dinilai lebih baik dari algoritme *Naïve Bayes* dengan melihat nilai hasil dari *accuracy*, *precision*, dan *recall* algoritme *K-Nearest Neighbor* dengan nilai *accuracy* sebesar 84,2%, *precision* sebesar 92,3% dan *recall* sebesar 82,6%.

Kata kunci : Klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes*, *Confusion Matrix*, *K-Fold Cross Validation*.

1. PENDAHULUAN

Kelulusan merupakan tujuan akhir bagi setiap orang dalam rangkaian proses menyelesaikan studi untuk memperoleh nilai, penghargaan, gelar ataupun yang lainnya. Bagi mahasiswa kelulusan merupakan suatu capaian yang luar biasa terutama bila dapat menyelesaikan studinya tepat waktu. Dalam perguruan tinggi mahasiswa merupakan asset yang sangat penting bagi institusi pendidikan oleh karena itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya [1].

Umumnya mahasiswa mampu menyelesaikan studi mereka secara tepat waktu, ada juga yang terlambat menyelesaikan studi (lebih dari 8 semester), namun tidak sedikit pula mahasiswa yang mampu menyelesaikan studinya lebih cepat selama 7 semester. Banyaknya mahasiswa yang masa studinya tepat waktu akan berdampak baik terhadap penilaian akreditasi program studi.

Pada Program Studi Teknik Informatika Universitas Muhammadiyah Gresik sendiri terdapat banyak mahasiswa yang lulus tidak tepat waktu (terlambat) dan mahasiswa yang *Drop Out* (DO) dikarenakan melebihi batas masa studi. Hal tersebut merupakan suatu masalah penting bagi Program Studi, dikarenakan jumlah total dari mahasiswa yang mendaftar tidak sebanding dengan jumlah total mahasiswa yang lulus (menyelesaikan studi) yang nantinya akan berdampak pada penilaian Akreditasi. Sangat penting bagi program studi mengetahui klasifikasi kelulusan mahasiswa menggunakan data mining. Informasi tersebut dapat menjadi dasar dalam

membuat rencana strategis dalam rangkaian pembelajaran supaya meningkatkan jumlah mahasiswa dengan masa studi lulus tepat waktu.

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan kelas dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan. Dalam klasifikasi terdapat beberapa algoritme antara lain, *K-Nearest Neighbor*, *Naïve Bayes*, *Neural Network*, *Decision Tree*, dan *Support Vector Machine* (SVM). Algoritme *K-Nearest Neighbor* merupakan salah satu teknik klasifikasi data yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot. Algoritme *K-Nearest Neighbor* merupakan satu metode yang termasuk kedalam *supervised learning*, dimana hasil dari klasifikasi ditentukan oleh kelas yang paling banyak muncul (mayoritas) [1].

Salah satu kelebihan dari metode KNN adalah tangguh terhadap *training data* dan efektif apabila *training data*-nya dalam jumlah yang besar [2].

Pada penelitian sebelumnya yang dilakukan oleh Aninda Zulaifa A. dan Yogiek Indra K. (2019) yang berjudul “Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritme *K-Nearest Neighbor*” mendapat kesimpulan, bahwa penerapan Algoritme *K-Nearest Neighbor* dirasa sesuai dilihat dari nilai pengujian data testing sebanyak 12 kali percobaan. Penelitian lainnya yang dilakukan oleh Yulia Rizki Amalia (2018) yang berjudul “Penerapan

Data Mining Untuk Prediksi Penjualan Terlaris Menggunakan Metode *K-Nearest Neighbor*” mendapatkan kesimpulan, bahwa dengan menerapkan model prediksi menggunakan metode *K-Nearest Neighbor* mampu memberikan nilai akurasi klasifikasi yang tinggi. Pandu Yuli Santoso dan Dewi Kusumaningsih (2018) dengan judul penelitian “Algoritme *K-Nearest Neighbor* Untuk Memprediksi Kelulusan Ujian Nasional Berbasis Desktop Pada SMAN 12 Tangerang” mendapatkan kesimpulan, nilai akurasi rata-rata sistem yang didapat setelah dilakukan 5 kali testing dalam melakukan prediksi sebesar 88,5% dengan nilai tertinggi pada $K = 7$.

Penelitian lainnya yang dilakukan oleh Eka Wahyu Sholeha, Selviana Yunita, dkk (2022) dengan judul penelitian “Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan *K-Nearest Neighbor*” mendapatkan kesimpulan, bahwa didapatkan akurasi tertinggi ketika seluruh data menggunakan huruf kecil untuk kedua Algoritme dengan akurasi 52,35%. Penelitian lainnya yang dilakukan oleh Raemon Syaljumari, dkk (2021) dengan judul penelitian “Akurasi Klasifikasi Pengguna Hotspot WiFi Dengan Menggunakan Metode *K-Nearest Neighbor*” mendapatkan kesimpulan, bahwa berdasarkan tabel presentase terbesar adalah $K = 3$ dengan presentase akurasi 95% dan *error* sebesar 5%.

Penelitian lainnya yang dilakukan oleh Tias Mugi Rahayu, Besse Arnawisuda Ningsi, Isnurani, dan Irvana Arofah (2021) dengan judul penelitian “Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Dengan Metode *Naïve Bayes*” mendapat kesimpulan, dari total data yang berjumlah 2478 data mahasiswa, diketahui 61,9% lulus tepat waktu berjumlah 697 mahasiswa dan 38,1% lulus tidak tepat waktu berjumlah 429 mahasiswa. Dari perhitungan menggunakan Algoritme *Naïve Bayes* bahwa 225 data testing yang diuji menunjukkan 156 data terklasifikasi secara benar. Hasil klasifikasi ketepatan kelulusan mahasiswa menggunakan Algoritme *Naïve Bayes* diperoleh tingkat akurasi sebesar 69,33%.

Dalam penelitian ini bertujuan untuk menganalisis perbandingan metode Algoritme *K-Nearest Neighbor* dan *Naïve Bayes* dalam melakukan klasifikasi waktu penyelesaian studi mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. Secara waktu penyelesaian studi, bagaimana klasifikasi waktu penyelesaian studi dengan waktu 7 semester (lebih cepat), 8 semester (tepat waktu), dan lebih dari 8 semester (terlambat), klasifikasi waktu penyelesaian studi akan menggunakan 3 kelas tersebut. Adapun parameter yang digunakan dalam penelitian ini adalah Indeks Prestasi Semester (IPS) yang terdiri dari IPS 1, IPS 2, IPS 3, IPS 4, IPS 5, dan IPS 6, fitur Indeks Prestasi Kumulatif (IPK), Maka Kuliah Pilihan (MKP), Kuliah Kerja Nyata (KKN), dan Tugas Khusus (KP). Metode yang digunakan dalam melakukan penelitian ini adalah metode *K-Nearest Neighbor* (KNN) dan *Naïve Bayes*.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan analisa yang dilakukan secara *automatic* (otomatis) pada data yang berjumlah besar dan kompleks yang bertujuan untuk mendapatkan nilai kecenderungan atau pola yang keberadaannya tidak disadari. *Data mining* merupakan proses menemukan sesuatu yang bermakna oleh suatu korelasi baru, pola dan juga tren yang terdapat dengan cara memilah-milah data yang berukuran besar, dimana data tersebut disimpan dalam *repository*, menggunakan teknologi sosialisasi pola serta statistik dan teknik matematika [3].

[4] *Data mining* adalah proses yang memakai teknik statistik, teknik matematika, kecerdasan protesis, *machine learning* dalam melakukan ekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait oleh database yang besar.

2.2. Klasifikasi

Klasifikasi merupakan proses penemuan model membedakan kelas data, atau dengan cara mengklasifikasi data kedalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya [1].

Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (*target*) f yang memetakan tiap himpunan label kelas yang telah terdefinisi sebelumnya [5].

Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui. Metode klasifikasi yang umum digunakan diantaranya adalah *Decision Tree*, *K-Nearest Neighbor*, *Naïve bayes*, *Neural Network*, *C4.5*, dan *Support Vector Machine* [6].

2.3. Algoritme *K-Nearest Neighbor*

Algoritme *K-Nearest Neighbor* merupakan metode klasifikasi berdasarkan tetangga terdekat dengan konsep sederhana, kuat pada data non-linier, dan dapat digunakan dalam kasus multi-kelas. [7]

Algoritme *K-Nearest Neighbor* merupakan metode klasifikasi terhadap sekumpulan data berdasarkan mayoritas, yang bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan kategori yang sama dari sampel data training [8].

K-Nearest Neighbor termasuk dalam *supervised learning*, yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Hasil dari klasifikasi diambil dari kelas yang paling banyak muncul, yang menjadi kelas hasil klasifikasi [9].

2.4. Algoritme *Naïve Bayes*

Algoritme *Naïve Bayes* merupakan salah satu algoritme yang masuk dalam *supervised learning*. Algoritme *Naïve Bayes* merupakan metode pengklasifikasian berdasarkan nilai probabilitas setiap atribut yang terdapat di dalam data set, klasifikasi *naïve bayes* didasarkan pada teorema bayes (*bayes theorem*). Algoritma *Naïve Bayes* diperkenalkan oleh

ilmuwan inggris Thomas Bayes. Algoritma *Naive Bayes* mengadopsi ilmu statistika yaitu menggunakan teori kemungkinan (probabilitas) untuk menyelesaikan sebuah kasus supervised learning, artinya dalam himpunan data terdapat label, class atau target sebagai acuan [10].

Keunggulan dari algoritme ini adalah tingkat pemrosesan yang cepat dan tingkat akurasi yang tinggi mesikipun dalam jumlah data set yang besar.

2.5. Penelitian Terkait

Berdasarkan penelitian yang dilakukan oleh Aninda Zulaifa Abidin dan Yogiek Indra Kurniawan dengan judul Aplikasi Klasifikasi Penerimaan Kartu Indonesia Sehat Menggunakan Algoritme *K-Nearest Neighbor* membahas mengenai penerapan Algoritme *K-Nearest Neighbor*. Dalam sistem Aplikasi Klasifikasi Penerimaan Kartu Indonesia Sehat dirasa sesuai, dilihat dari nilai pengujian data sebanyak 12 kali menghasilkan rata-rata nilai *accuracy* 97,66%, *precision* 98,5%, dan *recall* 96,5%.

Penelitian lain yang dilakukan oleh Aria Pratama, Farid Ali Ma'ruf, Iin, Ade Rizki Rinaldi, dan Faturrhohman yang berjudul Klasifikasi Penerimaan Beasiswa Menggunakan Algoritme *K-Nearest Neighbor*, pada penerapan Algoritme *K-Nearest Neighbor* dalam melakukan klasifikasi penerimaan beasiswa, didapatkan nilai *accuracy* 78,45% nilai *precision* 25%, dan nilai *recall* 9,52%.

Penelitian lain ditulis oleh Ari Rudiyan, Akhmad Erik Dzulkifli, dan Khabib Munazar yang berjudul Klasifikasi Kebaran Hutan Menggunakan Algoritme *K-Nearest Neighbor*, membahas tentang penerapan Algoritme *K-Nearest Neighbor* untuk klasifikasi kebaran hutan dilakukan pengujian dengan data testing 30% dan data training 70% dari 14.201 data. Mendapatkan nilai *accuracy* tertinggi sebesar 92% dengan nilai K=18, dan diperoleh nilai yang sama pada nilai K=28.

Penelitian serupa yang dilakukan oleh Tias Mugi Rahayu, Besse Arnawisuda Ningsi, Isnurani, dan Irvana Arofah yang berjudul Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Dengan Metode *Naive Bayes*, membahas mengenai penerapan Algoritme *Naive Bayes* dalam melakukan klasifikasi ketepatan waktu kelulusan mahasiswa, dari total data sebanyak 2478 data mahasiswa, diketahui 61,9% mahasiswa lulus tepat waktu berjumlah 697 mahasiswa dan 38,1% lulus tidak tepat waktu berjumlah 429 mahasiswa. Dari perhitungan menggunakan Algoritme *Naive Bayes* bahwa 225 data testing yang diuji menunjukkan 156 data terklasifikasi secara benar. Hasil klasifikasi ketepatan kelulusan mahasiswa menggunakan Algoritme *Naive Bayes* diperoleh nilai *accuracy* 69,33%.

Pada penelitian lain yang dilakukan oleh Sri Hartati dan Haris Anom SAN dengan judul Algoritme *Naive Bayes* untuk Prediksi Kelulusan Mahasiswa, membahas mengenai penerapan Algoritme *Naive Bayes* dalam melakukan prediksi kelulusan mahasiswa

dari total dataset kelulusan mahasiswa sejumlah 321 mendapatkan nilai akurasi hasil prediksi sebesar 80%, dengan nilai *precision* sebesar 88% dan nilai *recall* sebesar 88%.

3. METODE PENELITIAN

3.1. Representasi Data

Dalam mengetahui klasifikasi kelulusan mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik dibutuhkan beberapa variabel prediktor untuk mendapatkan klasifikasi terhadap tingkat kelulusan mahasiswa yang menjadi variabel respons. Penentuan variabel tersebut didasari untuk melihat kemampuan mahasiswa dalam proses akademik mahasiswa tiap semesternya dalam rangkaian pembelajaran sebagai acuan mahasiswa dapat menyelesaikan studinya secara lebih cepat, tepat waktu atau terlambat, semua data dari variabel tersebut dihitung menggunakan Algoritme *K-Nearest Neighbor* dan *Naive Bayes*.

Tabel 1. Atribut Data

Atribut	Tipe Atribut	Keterangan
Indes Prestasi Semester (IPS 1 – 6)	Numerik	Indeks Prestasi Semester didapatkan dari nilai indeks prestasi semester 1 sampai semester 6.
Indeks Prestasi Kumulatif (IPK)	Numerik	Indeks Prestasi Kumulatif didapatkan dari nilai indeks prestasi mahasiswa di semester 6.
Mata Kuliah Pilihan (MKP)	Numerik	Mahasiswa menyelesaikan 6 MKP pada semester 6 dengan 1 = ya dan 0 = tidak.
Kuliah Kerja Nyata (KKN)	Numerik	Mahasiswa menyelesaikan KKN pada semester 6 dengan 1 = ya dan 0 = tidak.
Tugas Khusus (KP)	Numerik	Mahasiswa menyelesaikan KP pada semester 6 dengan 1 = ya dan 0 = tidak.

Data variabel yang diperoleh dalam penelitian ini berupa data akademik mahasiswa yang berkaitan dengan atribut Indeks Prestasi Semester (IPS) yang terdiri dari IPS 1, IPS 2, IPS 3, IPS 4, IPS 5, dan IPS 6, fitur Indeks Prestasi Kumulatif (IPK), Mata Kuliah Pilihan (MKP), Kuliah Kerja Nyata (KKN), dan Tugas Khusus (KP). Data variabel respons kelulusan mahasiswa diklasifikasikan menjadi 3 label, yaitu lebih cepat, tepat waktu, dan terlambat.

3.2. Preprocessing Data

Data latih dan data uji terlebih dahulu dilakukan proses *preprocessing* data untuk menemukan data yang tidak sesuai dengan kriteria dengan cara mengeliminasi data tersebut, dimana terdapat data dari mahasiswa yang lulus selama ≤ 7 Semester melalui jurnal/artikel tanpa melalui proses sidang Seminar Proposal (Sempro), menurut peraturan terbaru pada

tahun 2022 di Universitas Muhammadiyah Gresik dimana semua proses kelulusan mahasiswa melalui jurnal/artikel diharuskan melalui proses sidang Seminar Proposal terlebih dahulu. Pada proses *preprocessing* ini ditemukan beberapa data mahasiswa

yang lulus tanpa melalui proses Seminar Proposal sebanyak 7 data pada data training dan 4 data pada data testing.

3.3. Perhitungan Algoritme K-Nearest Neighbor

3.3.1. Persiapan Data

Tabel 2. Data Training Hasil *Preprocessing*

No	IPS						IPK	MKP	KKN	KP	Kelas
	1	2	3	4	5	6					
1	3,57	3,79	3,69	3,02	3,65	3,76	3,57	1	0	0	3
2	3,33	3,34	3,29	3,4	3,37	3,67	3,4	0	0	0	3
3	3,57	3,37	3,08	2,89	1,86	3,15	2,98	0	0	0	3
4	3,43	3,16	3,56	3,04	3,39	3,67	3,38	1	0	0	3
5	2,21	3,13	3,44	3,33	3,27	3,74	3,36	0	0	0	3
6	3,81	3,79	3,35	3,52	3,31	3,57	3,54	0	1	0	3
7	3,4	3,58	2,6	3,17	2,56	3,07	3,15	0	0	0	3
8	3,24	3,08	3,25	2,77	2,87	3,61	3,13	1	0	0	3
9	3,33	3,55	3,38	3,23	3,33	3,67	3,41	1	0	0	3
10	3,07	2,76	2,74	3,24	2,18	1,74	2,69	1	0	0	3
...
91	3,67	3,39	3,48	3,56	3,76	3,74	3,53	1	0	0	2

Tabel 3. Data Testing Hasil *Preprocessing*

No	IPS						IPK	MKP	KKN	KP	Kelas
	1	2	3	4	5	6					
1	3,36	3,53	3,44	3,56	3,3	1,44	3,1	0	0	0	3
2	3,86	3,57	2,63	3,55	3,21	2,56	3,28	0	0	0	3
3	3,29	3,5	3,46	3,75	3,37	2,19	3,24	0	0	0	3
4	3,32	3,47	3,67	3,9	2,96	2,71	3,36	0	0	0	3
5	2,93	3,07	2,75	2,48	2,53	1,19	2,58	0	0	0	3
6	2,93	3,37	2,88	3,7	2,93	2,59	3,05	0	0	0	3
7	3,71	3,7	3,19	3,04	2,96	2,61	3,33	0	0	0	3
8	3,69	3,69	3,6	3,6	2,52	3,5	3,43	1	0	0	3
9	3,21	2,98	3,16	3,23	2,91	3,5	3,15	0	0	0	3
10	3,33	3,27	3,46	3,4	2,59	3,42	3,24	1	0	0	3
...
72	3,57	3,53	3,31	3,25	3,7	3,61	3,49	1	0	1	2

3.3.2. Menentukan Nilai K

Dalam penelitian ini nilai K yang digunakan adalah K=3.

3.3.3. Menghitung Jarak Euclidean

$$d_{x_1d_{y_1}} = \sqrt{(3,57 - 3,36)^2 + (3,79 - 3,53)^2 + (3,69 - 3,44)^2 + (3,02 - 3,56)^2 + (3,65 - 3,3)^2 + (3,76 - 1,44)^2 + (3,57 - 3,1)^2 + (1 - 0)^2 + (0 - 0)^2 + (0 - 0)^2}$$

$$= 2,6817$$

Tabel 4. Hasil Perhitungan Jarak *Euclidean*

No	Jarak	Kelas	Urutan Nilai Terkecil
81	1,9125	2	1
19	1,9294	3	2
10	1,9447	3	3
72	1,9802	3	4
7	2,0171	3	5
44	2,1104	3	6
87	2,1486	2	7

No	Jarak	Kelas	Urutan Nilai Terkecil
46	2,2291	3	8
17	2,2359	3	9
...
68	6,5298	2	91

Pada Tabel 4 dilakukan pengurutan pengurutan peringkat data berdasarkan nilai jarak *Euclidean*. Proses selanjutnya dilakukan pengelompokan data yang akan diklasifikasikan berdasarkan nilai K.

Tabel 5. Hasil Klasifikasi KNN K = 3

No	Jarak	Kelas	Urutan Nilai Terkecil	Klasifikasi
81	1,9125	2	1	Ya
19	1,9294	3	2	Ya
10	1,9447	3	3	Ya

Dari proses perhitungan sampai pengelompokan klasifikasi pada Tabel 5 mendapatkan

hasil bahwa nilai dari data testing 1 hasil klasifikasi K = 3 masuk kedalam kelas 3.

3.4. Perhitungan Algoritme Naïve Bayes

3.4.1. Menghitung nilai probabilitas awal tiap kelas

$$P(C_i) = \frac{\sum C_i}{n}$$

Keterangan:
 $P(C_i)$: Probabilitas label kelas C_i
 $\sum C_i$: Jumlah data dengan label kelas C_i
 n : Jumlah total data latih

Label Kelas 1 :

$$P(1S) = \frac{\sum 1}{n} = \frac{4}{91} = 0,044$$

Label Kelas 2 :

$$P(2S) = \frac{\sum 2}{n} = \frac{37}{91} = 0,407$$

Label Kelas 3 :

$$P(3S) = \frac{\sum 3}{n} = \frac{50}{91} = 0,549$$

3.4.2. Menghitung probabilitas awal tiap fitur

Tabel 6. Probabilitas Awal Fitur IPS 1

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,32	0,24
Baik	0	0,62	0,60
Cukup	0,25	0,03	0,08
Kurang	0	0	0,04
Sangat Kurang	0	0,03	0,04

Tabel 7. Probabilitas Awal Fitur IPS 2

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,22	0,18
Baik	0	0,68	0,58
Cukup	0,25	0,08	0,14
Kurang	0	0	0,08
Sangat Kurang	0	0,03	0,02

Tabel 8. Probabilitas Awal Fitur IPS 3

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,54	0,18
Baik	0,25	0,43	0,42
Cukup	0	0,03	0,30
Kurang	0	0	0,08
Sangat Kurang	0	0	0,02

Tabel 9. Probabilitas Awal Fitur IPS 4

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,49	0,04
Baik	0,25	0,51	0,46
Cukup	0	0	0,38
Kurang	0	0	0,08
Sangat Kurang	0	0	0,04

Tabel 10. Probabilitas Awal Fitur IPS 5

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,86	0,18
Baik	0,25	0,11	0,38
Cukup	0	0,03	0,18
Kurang	0	0	0,16
Sangat Kurang	0	0	0,10

Tabel 11. Probabilitas Awal Fitur IPS 6

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,75	0,84	0,58
Baik	0,25	0,14	0,26
Cukup	0	0,03	0,04
Kurang	0	0	0,04
Sangat Kurang	0	0	0,08

Tabel 12. Probabilitas Awal Fitur IPK

$x_k C_i$	Kelas		
IPK	1	2	3
Sangat Baik	0,75	0,51	0,14
Baik	0,25	0,46	0,56
Cukup	0	0,03	0,28
Kurang	0	0	0
Sangat Kurang	0	0	0,02

Tabel 13. Probabilitas Awal Fitur MKP

$x_k C_i$	Kelas		
MKP	1	2	3
Sangat Baik	0,75	0,51	0,14
Kurang	0,25	0,46	0,56

Tabel 14. Probabilitas Awal Fitur KKN

$x_k C_i$	Kelas		
KKN	1	2	3
Sangat Baik	0	0,35	0,04
Kurang	0	0,65	0,96

Tabel 15. Probabilitas Awal Fitur KP

$x_k C_i$	Kelas		
KKN	1	2	3
Sangat Baik	0,75	0,27	0
Kurang	0,25	0,73	1

Tabel 16. Tabel Data Training Ke-1

No	IPS						IPK	MKP	KKN	KKN
	1	2	3	4	5	6				
1	S	S	S	B	S	S	S	SB	K	K
	B	B	B		B	B	B			

$$P(1|X) = 0,04 \times 0,75 \times 0,75 \times 0,75 \times 0,25 \times 0,75 \times 0,75 \times 0,75 \times 1 \times 0 \times 0,25$$

$$P(1|X) = 0$$

$$P(2|X) = 0,41 \times 0,32 \times 0,22 \times 0,54 \times 0,51 \times 0,86 \times 0,84 \times 0,51 \times 0,92 \times 0,65 \times 0,73$$

$$P(2|X) = 0,0013$$

$$P(3|X) = 0,55 \times 0,24 \times 0,18 \times 0,18 \times 0,46 \times 0,18 \times 0,58 \times 0,14 \times 0,54 \times 0,96 \times 1$$

$$P(3|X) = 0$$

Dari perhitungan diatas, hasil terbesar diperoleh pada $P(2|X)$. Maka untuk data training ke-1 diklasifikasikan pada kelas-2.

Tabel 17. Hasil Klasifikasi Data Training

No.	1	2	3	Kelas
1	0	0,0013	0	2
2	0	0,0001	0,0020	3
3	0	0	0,00004	3
4	0	0,0009	0,0010	3
5	0	0	0,0001	3
6	0	0,00001	0	2
7	0	0	0,0001	3
8	0	0	0,0009	2
9	0	0,0002	0,0007	2
...
91	0	0,0030	0,0013	2

3.4.3. Menghitung probabilitas akhir tiap kelas

$$P(C_i) = \frac{\sum C_i}{n}$$

Keterangan:
 $P(C_i)$: Probabilitas label kelas C_i
 $\sum C_i$: Jumlah data dengan label kelas C_i
 n : Jumlah total data latih

Label Kelas 1 :

$$P(1S) = \frac{\sum 1}{n} = \frac{8}{91} = 0,088$$

Label Kelas 2 :

$$P(2S) = \frac{\sum 2}{n} = \frac{43}{91} = 0,473$$

Label Kelas 3 :

$$P(3S) = \frac{\sum 3}{n} = \frac{40}{91} = 0,440$$

3.4.4. Menghitung probabilitas akhir tiap fitur

Tabel 18. Probabilitas Akhir Fitur IPS 1

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,35	0,13
Baik	0	0,63	0,65
Cukup	0,13	0,02	0,10
Kurang	0	0	0,05
Sangat Kurang	0	0	0,08

Tabel 19. Probabilitas Akhir Fitur IPS 2

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,21	0,10
Baik	0	0,72	0,58
Cukup	0,13	0,07	0,18
Kurang	0	0	0,10
Sangat Kurang	0	0	0,05

Tabel 20. Probabilitas Akhir Fitur IPS 3

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,51	0,08
Baik	0,13	0,47	0,43

Cukup	0	0,02	0,38
Kurang	0	0	0,10
Sangat Kurang	0	0	0,03

Tabel 21. Probabilitas Akhir Fitur IPS 4

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,37	0
Baik	0,13	0,63	0,38
Cukup	0	0	0,48
Kurang	0	0	0,10
Sangat Kurang	0	0	0,05

Tabel 22. Probabilitas Akhir Fitur IPS 5

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,81	0,05
Baik	0,13	0,16	0,40
Cukup	0	0	0,23
Kurang	0	0,02	0,20
Sangat Kurang	0	0	0,13

Tabel 23. Probabilitas Akhir Fitur IPS 6

$x_k C_i$	Kelas		
IPS	1	2	3
Sangat Baik	0,88	0,86	0,48
Baik	0,13	0,12	0,33
Cukup	0	0,02	0,05
Kurang	0	0	0,05
Sangat Kurang	0	0	0,10

Tabel 24. Probabilitas Akhir Fitur IPK

$x_k C_i$	Kelas		
IPK	1	2	3
Sangat Baik	0,88	0,51	0
Baik	0,13	0,49	0,60
Cukup	0	0	0,38
Kurang	0	0	0
Sangat Kurang	0	0	0,03

Tabel 25. Probabilitas Akhir Fitur MKP

$x_k C_i$	Kelas		
MKP	1	2	3
Sangat Baik	1	0,93	0,43
Kurang	0	0,07	0,58

Tabel 26. Probabilitas Akhir Fitur KKN

$x_k C_i$	Kelas		
KKN	1	2	3
Sangat Baik	1	0,23	0,03
Kurang	0	0,77	0,98

Tabel 27. Probabilitas Akhir Fitur KP

$x_k C_i$	Kelas		
KP	1	2	3
Sangat Baik	0,38	0,23	0
Kurang	0,63	0,77	1

Tabel 28. Hasil Pengujian Data Testing Algoritme Naïve Bayes

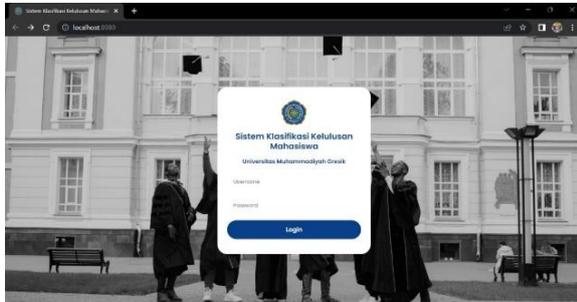
No	1	2	3	Kelas		Akurasi
				Prediksi	Asal	
1	0	0	0,00000001	3	3	Benar
2	0	0,00000002	0	2	3	Salah
3	0	0	0,0001	3	3	Benar
4	0	0,0000004	0	2	3	Salah
5	0	0	0,00001	3	3	Benar
6	0	0,00000001	0	2	3	Salah
7	0	0	0,000003	3	3	Benar
8	0	0,00004	0	2	3	Salah
9	0	0	0,0003	3	3	Benar
10	0	0	0,0462	3	3	Benar
...
72	0	0,0006	0	2	2	Benar

Dari proses pengujian kedua algoritme diatas, mendapatkan nilai akurasi dalam melakukan klasifikasi kelulusan mahasiswa dari masing-masing algoritme. Algoritme *K-Nearest Neighbor* dalam melakukan klasifikasi kelulusan mahasiswa mendapat nilai akurasi sebesar 90,27%, sedangkan pada algoritme *Naïve Bayes* dalam melakukan klasifikasi kelulusan mahasiswa mendapat nilai akurasi sebesar 91,66%.

4. HASIL DAN PEMBAHASAN

4.1. Halaman Login

Pada halaman *login*, *user* terlebih dahulu memasukkan *username* dan *password* yang sesuai pada *database* pada sistem. Jika *username* dan *password* tidak sesuai maka *user* tidak dapat masuk untuk mengakses sistem.



Gambar 1. Halaman Login

4.2. Halaman Pilih Metode

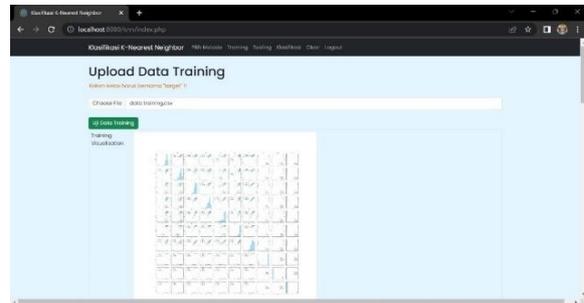
Setelah halaman *login* sistem, selanjutnya *user* dialihkan menuju halaman pilih metode, pada halaman ini *user* diharuskan memilih salah satu metode untuk perhitungan klasifikasi yang akan digunakan.



Gambar 2. Halaman Pilih Metode

4.3. Halaman Data Training

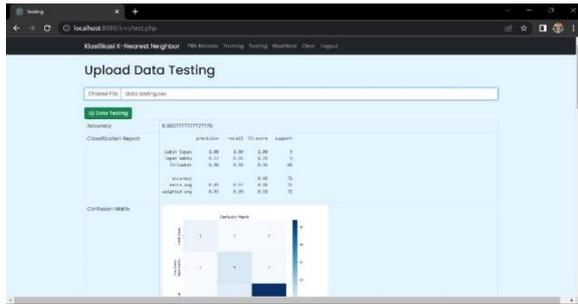
Pada halaman data training, *user* diharuskan meng-upload file data training berupa file excel dengan format *.csv*, setelah di upload data otomatis akan tersimpan pada library dan ditampilkan pada halaman data training.



Gambar 3. Halaman Data Training

4.4. Halaman Data Testing

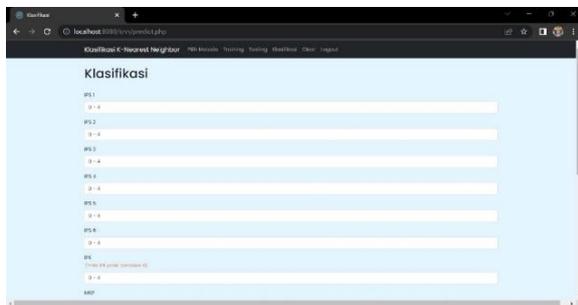
Pada halaman data testing ini serupa dengan halaman data training, *user* diharuskan meng-upload file data testing berupa file excel dengan format *.csv*, setelah di upload data otomatis akan tersimpan pada library dan sistem akan menampilkan data testing serta nilai pengujian dari *Confusion Matrix*.



Gambar 4. Halaman Data Testing

4.5. Halaman Klasifikasi

Pada halaman ini, *user* dapat melakukan pengujian data tunggal baru dengan memasukkan nilai pada kolom yang tersedia, kemudian *user* perlu menekan tombol “Klasifikasi” pada bagian bawah untuk mendapatkan hasil klasifikasi.



Gambar 5. Halaman Klasifikasi

4.6. Pengujian Confusion Matrix Algoritme K-Nearest Neighbor

Tabel 29. Hasil Pengujian Algoritme K-Nearest Neighbor

Kelas	Precision	Recall	Accuracy
1	100%	100%	90,27%
2	57,1%	88,8%	
3	98,1%	90%	
Rata-rata	85,1%	92,9%	

Dari Tabel 29 di atas di ketahui akurasi dari algoritme *K-Nearest Neighbor* sebesar 90,27%. Kemudian *precision* dan *recall* dari kelas 1 “Lebih Cepat” adalah 100%, dan 100%. Artinya model dari *K-Nearest Neighbor* ini memiliki nilai *precision* kelas 1 “Lebih Cepat” sebesar 100% yaitu semua prediksi pada kelas 1 “Lebih Cepat” semuanya benar dari keseluruhan data. Kemudian, nilai *precision* dan *recall* dari kelas 2 “Tepat Waktu” algoritme *K-Nearest Neighbor* memiliki nilai *precision* sebesar 57,1% yaitu semua prediksi kelas 2 “Tepat Waktu” yang benar sebesar 57,1% yang benar benar kelas 2 “Tepat Waktu”, dan memiliki nilai *recall* dari kelas 2 “Tepat Waktu” sebesar 88,8% dari keseluruhan data yang benar. Kemudian, nilai *precision* dan *recall* kelas 3 “Terlambat” dari algoritme *K-Nearest Neighbor* memiliki nilai *precision* kelas 3 “Terlambat” sebesar 98,1% yaitu semua prediksi kelas 3 “Terlambat” yang benar sebesar 98,1% yang benar benar kelas 3

“Terlambat”, dan memiliki nilai *recall* dari kelas 3 “Terlambat” sebesar 90% dari keseluruhan data yang benar.

4.7. Pengujian Confusion Matrix Algoritme Naive Bayes

Tabel 30. Hasil Pengujian Algoritme Naive Bayes

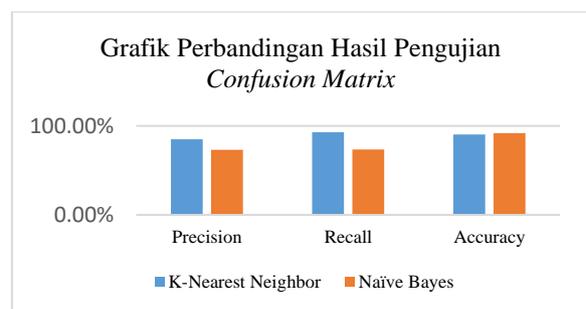
Kelas	Precision	Recall	Accuracy
1	50%	66,7%	91,66%
2	71,4%	55,6%	
3	96,7%	98,3%	
Rata-rata	73%	73,5%	

Dari Tabel 30 di atas di ketahui *accuracy* dari algoritme *Naive Bayes* sebesar 91,6%. Nilai *precision* dan *recall* dari kelas 1 “Lebih Cepat” adalah 50% dan 66,7%. Artinya model dari *Naive Bayes* ini memiliki nilai *precision* kelas 1 “Lebih Cepat” sebesar 50% yaitu semua prediksi yang benar pada kelas 1 “Lebih Cepat” hanya 50% yang benar benar kelas 1 “Lebih Cepat”, dan memiliki nilai *recall* dari kelas 1 “Lebih Cepat” sebesar 66,7% dari keseluruhan data yang benar. Kemudian, nilai *precision* dan *recall* dari kelas 2 “Tepat Waktu” adalah 71,4% dan 55,6%. Artinya model dari *Naive Bayes* ini juga memiliki nilai *precision* kelas 2 “Tepat Waktu” sebesar 71,4% yaitu semua prediksi kelas 2 “Tepat Waktu” yang benar sebesar 71,4% yang benar benar kelas 2 “Tepat Waktu”, dan memiliki nilai *recall* dari kelas 2 “Tepat Waktu” sebesar 55,5% dari keseluruhan data yang benar. Kemudian, nilai *precision* dan *recall* dari kelas 3 “Terlambat” adalah 97,6% dan 83,3%. Artinya model dari *Naive Bayes* ini juga memiliki nilai *precision* kelas 3 “Terlambat” sebesar 96,7% yaitu semua prediksi kelas 3 “Terlambat” yang benar sebesar 96,7% yang benar benar kelas 3 “Terlambat”, dan memiliki nilai *recall* dari kelas 3 “Terlambat” sebesar 83,3% dari keseluruhan data yang benar.

4.8. Perbandingan Hasil Pengujian Confusion Matrix

Tabel 31. Perbandingan Hasil Pengujian Confusion Matrix

Precision		Recall		Accuracy	
KNN	NB	KNN	NB	KNN	NB
85,1%	73%	92,9%	73,5%	90,27%	91,66%



Gambar 6. Grafik Perbandingan Confusion Matrix

Dari Tabel 31 dan Gambar 4 dapat dilihat bahwa perbandingan hasil pengujian *Confusion Matrix* mendapat hasil rata-rata bahwa kedua algoritme dinilai baik dalam melakukan proses klasifikasi dengan mendapatkan nilai *accuracy* diatas 90%, namun algoritme *Naïve Bayes* dinilai lebih baik dari algoritme *K-Nearest Neighbor* dengan nilai *accuracy* sebesar 91,6%. Algoritme *K-Nearest Neighbor* mendapatkan rata-rata nilai *precision*, dan *recall* lebih tinggi dari algoritme *Naïve Bayes* yaitu sebesar 85,1% dan 92,9%. Sedangkan untuk algoritme *Naïve Bayes* mendapatkan rata-rata nilai *precision* dan *recall* lebih rendah yaitu sebesar 73% dan 73,5%.

4.9. Pengujian K-Fold Cross Validation

Dalam pengujian ini digunakan sebanyak 3 *Fold Cross Validation* dimana dari total 163 dataset masing – masing dibagi menjadi 40 data testing dengan cara pengambilan 40 data di awal, 40 data di tengah, dan 40 data pada akhir dataset dengan perbandingan 123 data training dan 40 data testing pada setiap iterasi untuk kedua algoritme.

Tabel 32. Hasil Pengujian *K-Fold Cross Validation* Algoritme *K-Nearest Neighbor*

K-Fold ke	Precision	Recall	Accuracy
1	95%	80%	92,5%
2	87%	73%	65%
3	95%	95%	95%

Dari Tabel 32 diatas dapat dilihat bahwa nilai rata-rata tertinggi dari pengujian *K-Fold Cross Validation* didapatkan pada *K-Fold* ke-3 dengan nilai *accuracy* sebesar 95% dan nilai rata – rata *precision* dan *recall* sebesar 95%.

Tabel 33. Hasil Pengujian *K-Fold Cross Validation* Algoritme *Naïve Bayes*

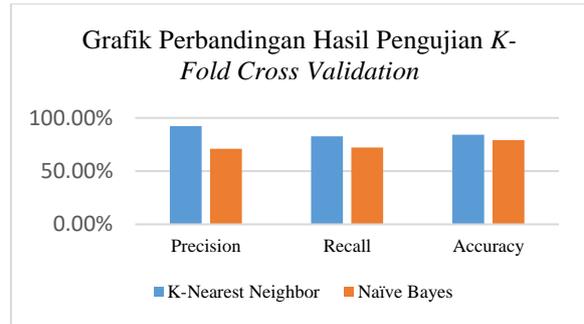
K-Fold ke	Precision	Recall	Accuracy
1	50%	47%	67,5%
2	77%	85%	80%
3	86%	85%	90%

Dari Tabel 33 diatas dapat dilihat bahwa nilai rata-rata tertinggi dari pengujian *K-Fold Cross Validation* didapatkan pada *K-Fold* ke-3 dengan nilai *accuracy* sebesar 90% dan nilai rata – rata *precision* dan *recall* sebesar 86% dan 85%.

4.10. Perbandingan Hasil Pengujian K-Fold Cross Validation

Tabel 34. Perbandingan Hasil Pengujian *K-Fold Cross Validation*

Precision		Recall		Accuracy	
KNN	NB	KNN	NB	KNN	NB
92,3%	71%	82,6%	72,3%	84,2%	79,2%



Gambar 7. Grafik Perbandingan *K-Fold Cross Validation*

Melihat pada Tabel 34 dan Gambar 7 perbandingan hasil pengujian *K-Fold Cross Validation* mendapat hasil rata-rata bahwa algoritme *K-Nearest Neighbor* dinilai lebih baik dari algoritme *Naïve Bayes* dengan melihat nilai hasil dari *accuracy*, *precision*, dan *recall* algoritme *K-Nearest Neighbor* mendapatkan hasil yang lebih baik dengan nilai *accuracy* sebesar 84,2% nilai *precision* sebesar 92,3% dan nilai *recall* sebesar 82,6%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dari implementasi dan pengujian sistem yang telah dilakukan mendapat beberapa kesimpulan antara lain: Proses klasifikasi kelulusan mahasiswa Universitas Muhammadiyah Gresik dapat dikatakan berhasil, dilihat dari nilai *accuracy* yang didapat dari kedua algoritme cukup tinggi. Algoritme *Naïve Bayes* dinilai lebih baik daripada algoritme *K-Nearest Neighbor* dilihat dari nilai *accuracy* yang didapat sebesar 91,6%, meskipun nilai *accuracy* dari kedua algoritme tergolong baik dengan mendapat nilai *accuracy* diatas 90%.

Hasil pengujian evaluasi performa menggunakan *Confusion Matrix* didapatkan nilai rata-rata *precision* dan *recall* dari kedua algoritme. Algoritme *K-Nearest Neighbor* mendapatkan rata-rata nilai *precision*, dan *recall* lebih tinggi dari algoritme *Naïve Bayes* yaitu sebesar 85,1% dan 92,9%. Sedangkan untuk algoritme *Naïve Bayes* mendapatkan rata-rata nilai *precision* dan *recall* lebih rendah yaitu sebesar 73% dan 73,5%. Sementara hasil pengujian menggunakan *K-Fold Cross Validation* mendapatkan hasil bahwa algoritme *K-Nearest Neighbor* dinilai lebih baik dari algoritme *Naïve Bayes* dengan melihat nilai hasil dari *accuracy*, *precision*, dan *recall* algoritme *K-Nearest Neighbor* mendapatkan hasil yang lebih baik dengan nilai *accuracy* sebesar 84,2% nilai *precision* sebesar 92,3% dan nilai *recall* sebesar 82,6%.

Berdasarkan hasil penelitian dan pengujian yang dilakukan, saran untuk penelitian selanjutnya yakni menambahkan atribut data baru yaitu pada semester berapa mahasiswa tersebut mengambil tugas akhir (skripsi) dengan mengubah range pengambilan semua nilai yang semula diambil pada semester 6 diubah dengan pengambilan semua nilai pada semester 7. Atau dapat ditambahkan dengan melakukan pembobotan pada data seperti *Weighted Naïve Bayes*

atau modifikasi metode seperti *Modified K-Nearest Neighbor* (M-KNN), bisa juga dengan menggunakan metode klasifikasi lain seperti *Support Vector Machine* (SVM), *Decision Tree*, dan *Neural Network* supaya mendapatkan hasil penelitian yang lebih baik.

DAFTAR PUSTAKA

- [1] Y. Saputra dan Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Techno.COM*, pp. Vol.17, No.4, 2018
- [2] Y. R. Amalia, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Study Kasus PT. Bintang Multi Sarana Palembang)," *Skripsi Fakultas Sains dan Teknologi*, 2018
- [3] D. T. Larose, *Data Mining Methods and Models*, New Jersey: John Willey & Sons, Inc, 2006
- [4] E. Turban, *Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1*, Yogyakarta: Andi, 2005
- [5] F. A. Hermawati, *Data Mining*, Yogyakarta: Andi, 2013
- [6] S. Diansyah, "Klasifikasi Tingkat Kepuasan Pengguna Menggunakan Metode K-Nearest Neighbor (KNN)," *Jurnal Sistim Informasi dan Teknologi*, pp. Vol.4, No.1, 2022
- [7] E. Prasetyo, R. Purbaningtyas dan R. D. Adityo, "Cosine K-Nearest Neighbor in Milkfish Eye Classification," *International Journal of Intelligent Engineering & Systems*, pp. Vol.13, No.3, 2019
- [8] P. Putra, A. M. H. Pardede dan S. Syahputra, "Analisis Metode K-Nearest Neighbor Dalam Klasifikasi Data Iris Bunga," *Jurnal Teknik Informatika Kaputama (JTIK)*, pp. Vol.6, No. 1, 2022
- [9] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Berlin: Heidelberg:Springer, 2011
- [10] B. A. N. I. I. A. Tias Mugi Rahayu, "Klasifikasi Ketepatan Kelulusan Mahasiswa Dengan Metode Naive Bayes," *ISSN Vol.15 No.10*, 2021