

ANALISIS PENGGUNAAN ALGORITMA KLASIFIKASI DALAM PREDIKSI KELULUSAN MENGGUNAKAN ORANGE DATA MINING

¹⁾Dinda Safitri, ²⁾Shofa Shofiah Hilabi, ³⁾Fitria Nurapriani

^{1,2,3)}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

^{1,2,3)}Jl. H.S.Ronggo Waluyo, Karawang – Jawa Barat - Indonesia

E-mail : si19.dindasafitri@mhs.ubpkarawang.ac.id, shofa.hilabi@ubpkarawang.ac.id

ABSTRAK

Kelulusan secara tepat waktu dalam Perguruan Tinggi merupakan harapan dari mahasiswa. Salah satu syarat untuk lulus dalam studinya, mahasiswa harus menempuh tahapan akhir yaitu menyelesaikan tugas akhir atau skripsi. Tetapi waktu kelulusan tidak selalu dapat memprediksi kapan mahasiswa akan lulus. Banyak faktor yang menyebabkan kelulusan mahasiswa seperti IPK, SKS, status pekerjaan dan lain sebagainya. Melihat hal tersebut penting adanya sebuah metode yang dapat memprediksi kelulusan mahasiswa, akan tetapi pada beberapa Perguruan Tinggi belum memiliki metode tersendiri untuk dapat memperkirakan kelulusan mahasiswanya apakah mahasiswa tersebut dapat lulus tepat waktu atau tidak. Untuk mengatasi hal tersebut, maka diperlukan sebuah model untuk dapat memprediksi kelulusan mahasiswa. Pada penelitian kali ini dilakukan analisis terhadap 3 metode yaitu Naive Bayes, K-NN dan Neural Network. Tujuan daripada penelitian ini yaitu untuk mengetahui metode mana yang lebih tepat digunakan dalam memprediksi kelulusan. Pada penelitian ini juga dilakukan perbandingan antara ketiga metode tersebut, dan didapatkan metode terbaik yaitu metode K-NN dengan nilai akurasi 89%.

Kata Kunci: Kelulusan, Klasifikasi, Naive Bayes, K-NN, Neural Network

ABSTRACT

Timely graduation in Higher Education is the expectation of students. One of the requirements to graduate in their studies, students must take the final stage, namely completing the final project or thesis. But the time of graduation is not always able to predict when college students will graduate. Many factors cause student graduation such as GPA, credits, employment status and so on. Seeing this, it is important to have a method that can predict student graduation, but some universities do not have their own method to be able to estimate student graduation whether the student can graduate on time or not. To overcome this, a model is needed to be able to predict student graduation. In this study, an analysis of 3 methods was carried out, namely Naive Bayes, K-NN and Neural Network. The purpose of this study is to find out which method is more appropriate to use in predicting graduation. In this study, a comparison was also made between the three methods, and the best method was obtained, namely the K-NN method with an accuracy value of 89%.

Keyword: *Graduated, Clasification, Naive Bayes, K-NN, Neural Network*

PENDAHULUAN

Berdasarkan penelitian terdahulu, kelulusan dalam Perguruan Tinggi merupakan suatu masa ketika mahasiswa berharap untuk lulus secara tepat waktu, begitu juga dengan setiap kampus yang menginginkan mahasiswanya untuk lulus tepat waktu [1]. Pada penelitian tahun 2021 yang dilakukan oleh Virtusena, Johar dan Andang mengatakan bahwa gelar sarjana di Universitas seringkali dapat dijangkau empat atau delapan semester, akan tetapi mahasiswa tidak selalu dapat

merampungkan studi mereka dalam kerangka waktu yang dipertimbangkan [2].

Pada penelitian ditahun 2022 oleh Zeniarja mengatakan bahwa akreditasi program Universitas dipengaruhi oleh persentase yang lebih besar dari siswa yang lulus tepat waktu [3]. Untuk memperoleh kebijaksanaan hasil prediksi dan memastikan bahwa semua siswa lulus tepat waktu, beberapa faktor yang mempengaruhi jumlah lulusan perguruan tinggi perlu diperiksa. Beberapa penelitian telah melakukan berbagai metode dalam memprediksi seperti pada penelitian yang

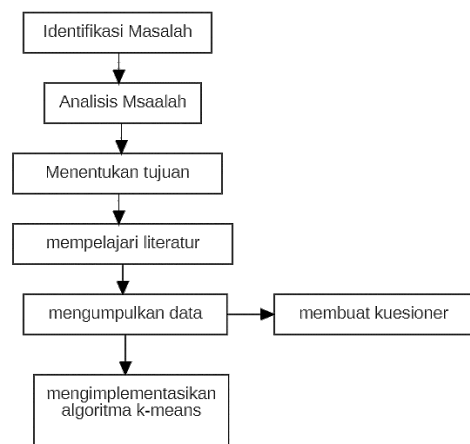
dilakukan Khaerunnisa telah membuktikan tingkat akurasi yang baik sebesar 96,55% dalam memprediksi kelulusan menggunakan algoritma K-Means dengan mengklaster berdasarkan data akademik mahasiswa [4]. Beberapa metode lain yang sering digunakan dalam prediksi yaitu metode klasifikasi dengan algoritma c4.5. Dengan menggunakan nilai akurasi, algoritma C4.5 dapat memprediksi hasil, dan dengan menggunakan metode ini, pohon keputusan akan dibangun untuk menentukan kelulusan siswa [5]. Kemudian pada penelitian lain nya yang menggunakan metode algoritma Naive Bayes menghasilkan tingkat akurasi 88,16% [6].

Akan tetapi pada beberapa Universitas belum memiliki metode dalam memprediksi kelulusan mahasiswa nya, seperti pada Universitas Buana Perjuangan Karawang khususnya pada Program Studi Sistem Informasi dalam memperkirakan kelulusan nya masih melihat berdasarkan IPK mahasiswa, sehingga belum bisa diketahui apakah mahasiswa tersebut bisa lulus tepat atau terlambat. Untuk penting adanya sebuah metode dalam memprediksi kelulusan ini agar dapat diidentifikasi apakah seorang mahasiswa lulus tepat waktu atau tidak. Pada penelitian kali ini digunakan beberapa algoritma dalam metode klasifikasi yaitu, algoritma naive bayes, K-NN, Neural Network menggunakan Orange Data Mining sebagai tools dari metode tersebut dan menggunakan data tingkat stress mahasiswa menghadapi skripsi sebagai salah satu faktor yang mempengaruhi kelulusan. Dengan tujuan mengetahui tingkat akurasi dari masing-masing algoritma agar dapat dilihat metode mana yang lebih akurat atau tepat. Dan juga untuk mengetahui pola kelulusan mahasiswa.

METODE

Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini yaitu dengan studi pustaka dan kuesioner, yang dimana dalam hal ini kuesioner tersebut dibuat dalam bentuk google formulir dan disebarluaskan kepada mahasiswa Sistem Informasi khususnya pada angkatan 2017-2018. Adapun alur dari tahapan penelitian ini yaitu seperti pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Analisis Data

Pada sesi ini, data siswa yang digunakan untuk teknik analisis data kuantitatif merupakan data mahasiswa Sistem Informasi Universitas Buana Perjuangan Karawang angkatan 2017-2018, berupa IPK, SKS, status pekerjaan dan tingkat stres mahasiswa. Adapun metode dalam analisis data ini yaitu analisis deskriptif dan teknik klasifikasi menggunakan beberapa algoritma.

Naive Bayes

Metode yang dikategorikan secara statistik yang disebut Naive Bayes dapat memperkirakan probabilitas keanggotaan kelas [7]. Untuk mengkategorikan himpunan data pelatihan tinggi sering digunakan Naive Bayes. Teorema Bayes sebagai berikut :

$$P(C | A) = \frac{P(A | C) P(C)}{P(A)} \quad (1)$$

K-Nearest Neighbor

K-NN atau K- Nearest Neighbor, adalah algoritma yang menemukan kasus dengan memperkirakan seberapa dekat kasus baru dengan kasus lama [8]. Menemukan jarak terpendek antara data yang perlu dievaluasi dengan terdekatnya adalah ide mendasar K-NN. Data pengujian dan nilai jarak data pelatihan diurutkan dimulai dengan nilai terendah[9]. Berikut rumus dari K-NN :

$$d_{\text{Euclidian}}(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Tahapan :

1. Menentukan parameter k (jumlah paling terdekat)
2. Menghitung kuadrat jarak Euclidean
3. Mengurutkan hasil pada tahap 2 secara ascending
4. Klasifikasi nearest neighbor berdasarkan nilai k
5. Kategori objek dapat diprediksi memiliki mayoritas terbesar menggunakan kategori terdekat

Neural Network

Jaringan saraf adalah representasi algoritmik dari cara neuron berfungsi. Setiap neuron di otak terhubung ke yang lain dan masing-masing dari mereka mentransmikan informasi [10]. Beberapa tahapan dalam Neural Network yaitu :

1. Input Layer, yaitu nilai jumlah data yang dimasukkan untuk pembelajaran. Nilainya tidak lebih dari jumlah variabel atau jumlah data.
2. Hidden Layer, yaitu neuron dalam jaringan. Jika jumlahnya Data yang tidak mencukupi menyebabkan underfitting, yang membuat jaringan kurang mampu mengenali sinyal dan pola dalam kumpulan data. Ketika jumlahnya berlebihan, overfitting terjadi, yang berarti tidak ada cukup data set pelatihan untuk melatih

semua neuron. Masing-masing nilai terbaik digunakan untuk menentukan tingkat pembelajaran, momentum, siklus pelatihan, dan lapisan input.

3. Output Layer, yaitu berapa banyak jumlah output dari proses pembelajaran algoritma neural network.

Analisis Kebutuhan

Pada tahap ini, algoritma klasifikasi tersebut digunakan untuk menganalisis data yang akan dikelompokkan. Informasi yang diperlukan disajikan sebagai parameter, yang dijadikan sebagai acuan dalam memprediksi kelulusan yaitu IPK, status pekerjaan, tingkat stres mahasiswa, dan kategori kelulusan. Proses klasifikasi disini menggunakan Orange data mining Tools. Berikut merupakan kategori tingkat stress mahasiswa :

Tabel 1. Kategori Tingkat Stress

Norma Kategorisasi	Kategori	Inisial
$X > \mu + 1,5\alpha$	Sangat Tinggi	1
$\mu + 0,5\alpha < X \leq \mu + 1,5\alpha$	Tinggi	2
$\mu + 0,5\alpha < X \leq \mu + 0,5\alpha$	Sedang	3
$\mu - 1,5\alpha < X \leq \mu + 0,5\alpha$	Rendah	4
$X \leq \mu - 1,5\alpha$	Sangat Rendah	5

Perhitungan pada kategori tingkat stress ini didapat melalui kuesioner dengan google formulir. Yang mana untuk menjabarkannya menggunakan norma kategorisasi.

Keterangan :

X = total skor

μ = mean teoritis

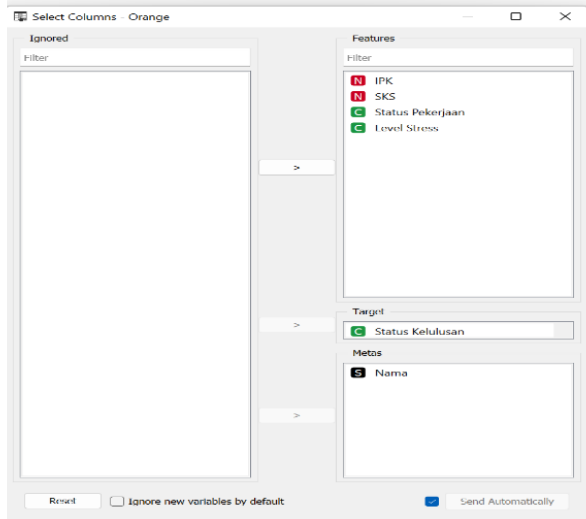
α = standar deviasi

HASIL

Data Selection Process

Sebelum melakukan proses klasifikasi, dilakukan proses pemilihan data terlebih dahulu atau disebut juga dengan selection data. Berdasarkan Gambar 2 ialah proses pemilihan data. Atribute yang akan digunakan yaitu

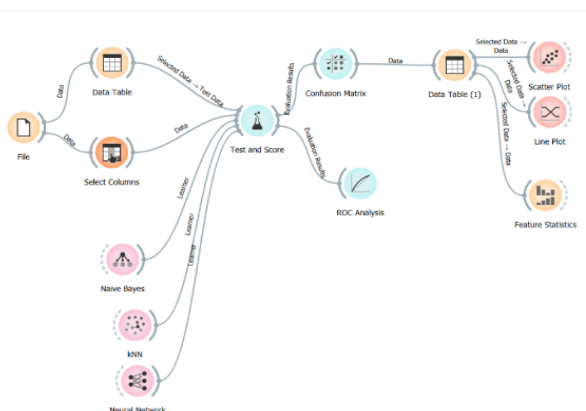
Nama, IPK, SKS, status pekerjaan, kategori stress dan status kelulusan sebagai attribute target. Pada saat selection process ini dataset mahasiswa Sistem Informasi tidak ada nilai yang hilang sehingga hanya dilakukan pemilihan data saja.



Gambar 2. Selection Data Process

Data Mining Process

Data mining digunakan untuk menguji model klasifikasi pada Orange Tools untuk mengetahui perbandingan metode agar dapat memilih metode terbaik dengan akurasi yang tepat dalam memprediksi kelulusan mahasiswa Sistem Informasi Universitas Buana Perjuangan Karawang. Seperti yang terlihat



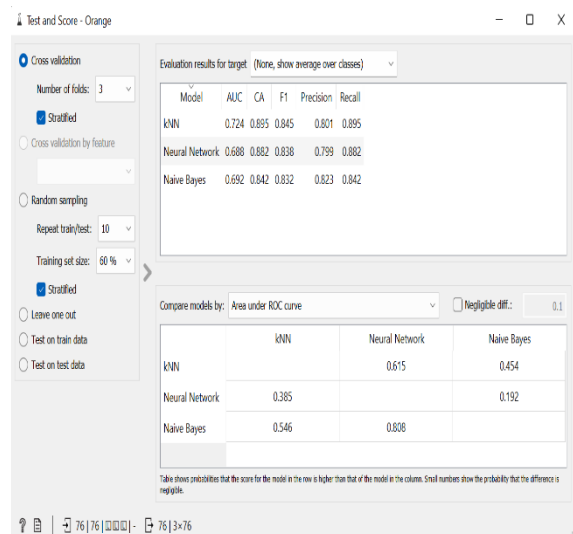
Gambar 3. Data Mining dengan Orange Tool

pada Gambar 3.

Hasil Simulasi 3 Model Klasifikasi

Hasil nilai tes diperoleh seperti Gambar 4. dengan menjalankan simulasi model klasifikasi menggunakan sekumpulan atribut dengan 1 atribut sebagai target dan 2 atribut numerik, yaitu sks dan Indeks Prestasi Kumulatif (IPK) mahasiswa.

Berdasarkan 76 record data mahasiswa Sistem Informasi yang telah diuji, maka diperoleh hasil dari menghitung presisi, recall, dan akurasi untuk setiap model, seperti yang ditunjukkan pada Gambar 4. Gambar 4 juga membandingkan tiga model AUC, diketahui bahwa model dengan nilai AUC tertinggi adalah pada metode K-NN yaitu sebesar 0.724. AUC digunakan untuk mengukur kinerja diskriminatif dengan memperkirakan probabilitas output yang diseleksi secara acak. Semakin besar AUC maka semakin baik hasil klasifikasi yang digunakan.



Gambar 4. Hasil Test and Score

Analisis Hasil Evaluasi dengan Confusion Matrix

Confusion Matrix merupakan Indikator kinerja untuk masalah klasifikasi machine learning. Confusion Matrix tabel yang berisi empat kemungkinan kombinasi angka prediksi dan aktual. Gambar 5. menunjukkan temuan evaluasi untuk model K-NN, Gambar 6. menunjukkan hasil Confusion Matrix untuk model Naive Bayes, dan Gambar 7. menunjukkan nilai Confusion Matrix untuk

model Neural Network.

Gambar 5. menunjukkan bahwa nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) masing-masing adalah 68, 0, 0, 0, dan 8. Nilai akurasi, presisi, dan penarikan metode K-NN adalah sebagai berikut:

$$\text{Accuracy} = \frac{68+0}{68+0+8+0} \times 100\% = 89\%$$

$$\text{Precision} = \frac{68}{68+8} \times 100\% = 89\%$$

$$\text{Recall} = \frac{68}{68+0} \times 100\% = 100\%$$

		Predicted		Σ
		Tepat	terlambat	
Actual	Tepat	68	0	68
	terlambat	8	0	8
Σ		76	0	76

Gambar 5. Hasil K-NN

Gambar 6. menunjukkan bahwa nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) masing-masing adalah 63, 1, 5, dan 7. Nilai-nilai metode Naive Bayes untuk Akurasi, Presisi, dan Recall adalah sebagai berikut:

$$\text{Accuracy} = \frac{63+1}{63+7+5+1} \times 100\% = 84\%$$

$$\text{Precision} = \frac{63}{63+7} \times 100\% = 90\%$$

$$\text{Recall} = \frac{63}{63+5} \times 100\% = 92\%$$

		Predicted		Σ
		Tepat	terlambat	
Actual	Tepat	63	5	68
	terlambat	7	1	8
Σ		70	6	76

Gambar 6. Hasil Naive Bayes

Gambar 7. menunjukkan bahwa nilai True Positive (TP), True Negative (TN), False

Positive (FP), dan False Negative (FN) masing-masing adalah 67, 0, 1, dan 8. Nilai metode Neural Network untuk Akurasi, Presisi, dan Recall adalah sebagai berikut:

$$\text{Accuracy} = \frac{67+0}{67+0+1+8} \times 100\% = 88\%$$

$$\text{Precision} = \frac{67}{67+8} \times 100\% = 89\%$$

$$\text{Recall} = \frac{67}{67+1} \times 100\% = 98\%$$

		Predicted		Σ
		Tepat	terlambat	
Actual	Tepat	67	1	68
	terlambat	8	0	8
Σ		75	1	76

Gambar 7. Hasil Neural Network

Analisis Perbandingan

Berdasarkan temuan evaluasi Confusion Matrix maka didapatkan nilai perbandingan Accuracy, Precision dan Recall dari ketiga metode diatas, seperti pada Tabel 2.

Tabel 2. Hasil Perbandingan

Metode	Accuracy	Precision	Recall
K-NN	89%	89%	100%
Naive Bayes	84%	90%	92%
Neural Network	88%	89%	98%

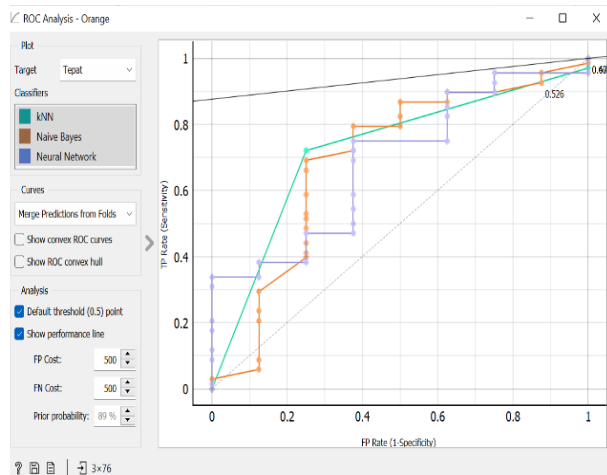
Berdasarkan tabel 2 dapat dilihat bahwa kinerja dari metode K-Nearest Neighbor lebih baik dari model Naive Bayes dan Neural Network. Akurasi klasifikasi tersebut dipengaruhi oleh banyaknya data yang diuji dalam proses simulasi yang dilakukan.

Analisis dengan ROC Curve

Cara mudah untuk membandingkan nilai akurasi setiap model klasifikasi secara grafis adalah dengan menggunakan model kurva ROC. Gambar 8. dan 9. menunjukkan

hasil grafis ROC.

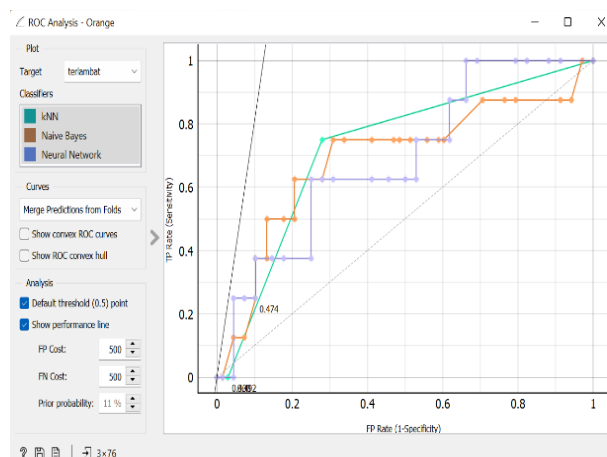
Pada Gambar 8. menunjukkan hasil analisis dengan ROC pada kelulusan mahasiswa Sistem Informasi Universitas Buana Perjuangan Karawang yang termasuk kategori lulus Tepat Waktu pada masing-masing model



Gambar 8. Kategori Tepat Waktu pada ROC Curve

yaitu, K-NN sebesar 0.600, Naive Bayes sebesar 0.526, dan Neural Network sebesar 0.472. K-NN dan Neural Network yang kurvanya mendekati titik 0.1 memiliki nilai akurasi terbaik untuk studi kasus pada model ini.

Pada Gambar 9. menunjukkan bahwa hasil analisis dengan ROC pada kelulusan mahasiswa Sistem Informasi Universitas Buana Perjuangan Karawang yang termasuk kategori lulus Tepat Waktu pada masing-masing model yaitu, K-NN sebesar 0.400, Naive Bayes



Gambar 9. Kategori Terlambat pada ROC Curve

sebesar 0.474, dan Neural Network sebesar 0.492. Maka untuk studi kasus pada model ini nilai akurasi paling baik yaitu K-NN dan Neural Network, karena mendekati titik 0.1 pada kurvanya.

KESIMPULAN

Pada penelitian ini untuk memprediksi kelulusan mahasiswa digunakan 3 metode yaitu Naive Bayes, K-NN dan Neural Network. Berdasarkan hasil olah ketiga metode tersebut didapatkan satu metode yang lebih baik yaitu metode algoritma K-NN (K-Nearest Neighbor) dengan jumlah mahasiswa yang lulus tepat waktu sebanyak 68 orang dalam kategori level stress 2. Dan juga algoritma tersebut menghasilkan nilai akurasi lebih tinggi dibandingkan 2 metode lainnya yaitu sebesar 89% dengan nilai presisi sebesar 89% dan recall 100%. Sedangkan pada metode Naive Bayes nilai akurasi sebesar 84% dengan nilai presisi sebesar 90% dan recall 99%. Kemudian pada metode Neural Network nilai akurasi sebesar 88% dengan nilai presisi 89% dan recall 98%.

Kemudian berdasarkan hasil olah level stress pada penelitian ini, membuktikan bahwa banyaknya mahasiswa yang lulus tepat waktu berada pada stress level 2. Hal ini menunjukkan bahwa tingkat stres tidak terlalu menjadi hambatan bagi mahasiswa untuk lulus tepat waktu.

Untuk itu pada penelitian selanjutnya disarankan melakukan perbandingan antar beberapa algoritma seperti perbandingan antara algoritma clustering dan classification. Agar dapat diketahui lebih dalam akurasi penggunaan dari kedua algoritma tersebut.

DAFTAR PUSTAKA

- [1] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.
- [2] V. Virtusena, A. Johar, and A. Wijanarko, "Mahasiswa Fakultas Teknik Unib Menggunakan Algoritme K-

- Means,” *J. Rekursif*, vol. 9, no. 2, pp. 206–225, 2021.
- [3] J. Zeniarja and A. Salam, “Perancangan Sistem Prediksi Kelulusan Mahasiswa Universitas Dian Nuswantoro Menggunakan Unified Modelling Language,” vol. 3, no. 1, pp. 25–36, 2021.
- [4] A. Khaerunnisa, “Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma K-Means Clustering,” *J. Ris. Mat.*, pp. 67–76, 2022.
- [5] N. Azwanti, “Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu),” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018, doi: 10.24176/simet.v9i1.1627.
- [6] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, “Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes,” *Technologia*, vol. 13, no. 3, pp. 207–214, 2022.
- [7] Amril, Sutan, Yana, and Bayu, “Accounting Information System,” pp. 17–30, 2019.
- [8] N. A. Sinaga, B. H. Hayadi, and Z. Situmorang, “Perbandingan Akurasi Algoritma Naïve Bayes, K-Nn Dan Svm Dalam Memprediksi Penerimaan Pegawai,” *J. Tek. Inf. dan Komput.*, vol. 5, no. 1, p. 27, 2022, doi: 10.37600/tekinkom.v5i1.446.
- [9] E. Purwaningsih and E. Nurelasari, “Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa,” *Syntax J. Inform.*, vol. 10, no. 01, pp. 46–56, 2021, doi: 10.35706/syji.v10i01.5173.
- [10] R. Ridwan, H. Lubis, and P. Kustanto, “Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 286, 2020, doi: 10.30865/mib.v4i2.2035.