

PREDIKSI TINGKAT KELULUSAN MAHASISWA S1 UNIVERSITAS AN NUUR DENGAN METODE DECISION TREE C4.5

Umar Haji Mussa'id¹, Agus Susilo Nugroho², Rahmawan Bagus Trianto³
Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas An Nuur Purwodadi,
Indonesia

ABSTRACT

One of the private universities in Purwodadi is An Nuur Purwodadi University. Many students have graduated from An Nuur Purwodadi University, but there are some students who did not graduate on time. This poses a problem and raises a significant question as to why these students did not graduate on time. A decision tree is a suitable data mining method for this research because it has the advantage of identifying and summarizing patterns in the data. The Decision Tree algorithm has an accuracy of 96.25%. The recall values for each class are 97.37% for the "Late" class and 95.00% for the "On-Time" class. Meanwhile, the precision values for each class are 94.87% for the "Late" class and 97.44% for the "On-Time" class

Keywords: *Student graduation1; decision tree2; university3; rapid miner4*

Correspondence:

Umar Haji Mussa'id
Universitas An Nuur Purwodadi, Email: Umarfattah00@gmail.com

PENDAHULUAN

Dalam proses akreditasi institusi suatu perguruan tinggi, gelar merupakan salah satu kriteria penilaiannya. Peringkat efisiensi pendidikan ideal perguruan tinggi menjadi salah satu faktor penilaian lulusan perguruan tinggi. Oleh karena itu kelulusan sangat membantu dalam menentukan akreditasi suatu universitas ketika mahasiswanya lulus tepat waktu. Di sisi lain, mahasiswa beruntung jika lulus tepat waktu karena bebas dari kewajiban biaya kuliah dan bisa langsung mendapatkan pekerjaan setelah selesai kuliah (Rohmawan 2018).

Salah satu perguruan tinggi swasta yang ada di Purwodadi adalah Universitas An Nuur Purwodadi. Sudah banyak mahasiswa yang lulus di Universitas An Nuur Purwodadi namun ada beberapa mahasiswa yang lulusnya tidak tepat waktu. Hal ini menjadikan adanya masalah yang timbul dan menjadi pertanyaan yang besar kenapa mahasiswa tersebut tidak lulus tepat waktu.

Decision tree adalah suatu metode data mining yang cocok dalam penelitian ini. karna decision tree memiliki kelebihan untuk mengidentifikasi dan merangkum pola-pola dalam data. Decisin tree juga sering di sebut sebagai alat untuk

menjelaskan data dan menghasilkan prediksi berdasarkan data tersebut.

Adapun tujuan dalam penelitian ini adalah untuk memprediksi tingkat kelulusan mahasiswa s1 keperawatan universitas an nuur dengan metode decision tree c4.5.

METODE PENELITIAN

Studi Literatur

Decision Tree adalah salah satu metode klasifikasi yang paling populer karena dapat dipengaruhi oleh manusia. Decision Tree adalah jenis model prediksi yang menggunakan struktur pohon atau struktur hierarki. Menurut konsep Decision Tree, data diubah menjadi pohon keputusan dan aturan keputusan. Informasi dalam Decision Tree sering kali dijelaskan menggunakan tabel dan catatan atribut. Saat membangun pohon, atribut menggambarkan parameter yang digunakan sebagai kriteria.

Algoritma C4.5 merupakan salah satu algoritma dalam penggalian data yang berfungsi untuk mengklasifikasikan suatu kelas. Cara kerja algoritma C4.5 dengan membentuk pohon keputusan untuk menghasilkan sebuah keputusan.

Pohon keputusan memiliki proses dengan membentuk data (tabel) menjadi bentuk model pohon, kemudian mengubah pohon menjadi sebuah rule, dan merubah rule menjadi lebih sederhana. adapun

rumus yang digunakan dalam penelitian sebagai berikut ini :

Entropy :

$$Entropy (S) = \sum_{i=1}^n p_i \log_2 p_i$$

Keterangan :

S = Himpunan Kasus

n = Jumlah partisi S

p_i = probabilitas yang didapat dari jumlah kelas dibagi total kasus

Gain :

$$Gains (S, A) = Entropy (S) - \sum_{i=1}^n p_i \log_2 p_i$$

* Entropy(si)

Keterangan :

S = Himpunan yang kasus

A = adalah Atribut

n = Jumlah dari atribut

|Si| = sebagai Jumlah partisi ke -i

|S| = sebagai jumlah kasus dalam S

Penelitian terkait dengan prediksi sudah banyak dilakukan oleh para peneliti sebelumnya, seperti prediksi produksi kopi, prediksi kelulusan dan prediksi resiko investasi saham. prediksi merupakan teknik kombinasi dari beberapa teknik data mining lainnya. Prediksi umumnya dipakai untuk menganalisis

peristiwa atau kejadian di masa lalu dalam urutan tertentu untuk memperkirakan peristiwa di masa depan.

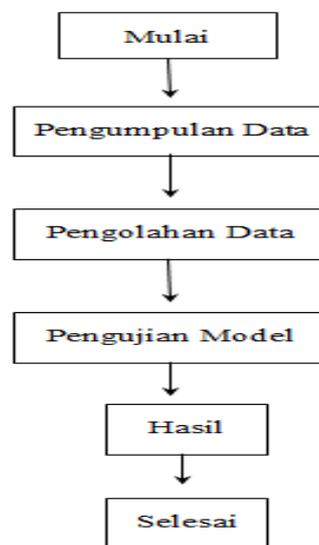
Database

Data yang digunakan dalam penelitian ini dikumpulkan pada tanggal 02 Juni sampai 21 Juli 2023 diperoleh dari Universitas An Nuur Purwodadi dengan menggunakan metode wawancara langsung dengan kaprodi S1 keperawatan Universitas An Nuur Purwodadi. Berdasarkan Hasil Wawancara yang dilakukan oleh peneliti mendapatkan informasi Database berisi tentang kelulusan mahasiswa An Nuur S1 Keperawatan yang berjumlah 111 mahasiswa. Hal tersebut menjadi sumber acuan utama penelitian database dan data mahasiswa.

Metodologi Penelitian

Metode penelitian ini dimulai dari studi literatur yang dilanjutkan dengan pengumpulan database. Pengumpulan database dilakukan pada tanggal 02 Juni sampai 21 Juli 2023 diperoleh dari Universitas An Nuur Purwodadi dengan menggunakan metode wawancara. Data yang telah dikumpulkan diolah terlebih dahulu di tahap pengolahan data. Proses Pengolahan data dilakukan untuk menentukan atribut-atribut yang mempunyai nilai signifikan tinggi. Proses Pengolahan data dilakukan

menggunakan aplikasi Rapidminer dengan tools select atribut. Setelah pengolahan data selesai dilakukan, maka kita akan lanjut ketahap selanjutnya yaitu pengujian model dari database tersebut dengan menggunakan algoritma decision Tree C4.5 kemudian dilihat performa model dengan mencari nilai Cross Validation. Semakin kecil nilai Cross Validation maka semakin baik pula performa model klasifikasi yang terbentuk. Metode penelitian ini dapat dilihat pada gambar 1 berikut:



Gambar 1 metode penelitian

Gambar 1 di atas menunjukkan alur metode penelitian yang dilakukan pada penelitian ini. Selanjutnya dilakukan proses penelitian dan pembahasan serta memunculkan hasil seperti yang dijelaskan pada bab selanjutnya.

HASIL DAN PEMBAHASAN

Sudah dijelaskan di bab sebelumnya bahwa pengumpulan data menggunakan teknik wawancara langsung dengan kaprodi S1 keperawatan Universitas An Nuur Purwodadi pada tanggal 02 juni sampai 21 juli 2023. Dalam proses wawancara peneliti mendapatkan informasi tentang Data kelulusan mahasiswa Universitas An Nuur Purwodadi sebanyak 111 Mahasiswa Universitas An Nuur Purwodadi. Adapun database yang diperoleh sebagai berikut.

NO	nim	nama mahasiswa	IP1	IP2	IP3	IP4	IP5	IP6	IP7	IP8	IPK	Label
			19	19,00	20	20	18	19	19	14	148	
1	0	RES 1	2,83	2,75	2,54	3,5	3,19	3	3,26	3,21	3,03	Telat
2	0	RES 2	3,29	3	3,55	3,35	3,64	3,71	3,87	3,71	3,51	Telat
3	0	RES 3	3,31	3,17	3,5	3,35	3,25	3,37	3,58	3,71	3,4	Tepat waktu
4	0	RES 4	3,29	3,14	3,75	3,6	3,72	3,82	3,87	3,86	3,62	Tepat waktu
5	0	RES 5	3,2	3,04	3,45	3,75	3,47	3,76	3,82	3,5	3,5	Tepat waktu
6	0	RES 6	2,93	2,92	3,2	3,35	3,06	3,21	3,71	3,57	3,23	Tepat waktu
7	0	RES 7	3,15	2,99	3,2	3,25	3,61	3,76	3,82	3,79	3,43	Telat
8	0	RES 8	3,13	3,02	3,55	3,7	3,47	3,82	3,87	3,86	3,54	Tepat waktu
9	0	RES 9	3,39	3,15	3,55	3,7	3,36	3,71	3,5	3,36	3,47	Tepat waktu
10	0	RES 10	3,36	3,16	3,7	3,25	3,33	3,42	3,82	3,93	3,48	Telat
100	0	RES 100	3,1	2,98	3,2	3,35	3,64	3,53	3,37	3,36	3,31	Tepat waktu
101	0	RES 101	3,29	3,1	3,6	3,45	3,19	3,68	3,61	3,5	3,43	Tepat waktu

Tabel 1 data mahasiswa annur

Tabel di atas memuat kode mahasiswa, program studi, organisasi, status pendidikan, indeks kinerja mahasiswa 1,2,3,4,5,6,7,8 dan indeks kinerja rata-rata. Untuk menghasilkan data

yang maksimal diperlukan pengolahan data.

Pengolahan Data

Data yang diperoleh dari akademik Universitas An Nuur Purwodadi akan dilakukan *preprocessing* terlebih dahulu menggunakan Rapid miner supaya dataset berkualitas dan sesuai dengan kebutuhan. Atribut-atribut yang akan digunakan nantinya adalah atribut-atribut yang mempunyai nilai signifikan tinggi. Dalam penelitian ini peneliti menggunakan nilai ip yang memiliki nilai di bawah 3.00 . Hasil preprocessing yang dilakukan dengan menggunakan Rapidminer menunjukkan bahwa atribut “Student Code”, “Program”, “Performance Index 5,6,7,8” dan “Average Performance Index” mempunyai signifikansi yang rendah, sehingga atribut yang digunakan berpengaruh terhadap kinerja. merupakan atribut indeks semester 1, 2, 3, 4. Organisasi, status pendidikan.

Nilai ip1 : 2.830

Nilai ip2 : 2.700

Nilai ip3 : 2.540

Nilai ip4 : 2.300

Nilai ip5 : 3.170

Nilai ip6 : 3.130

Nilai ip7 : 3.160

Nilai ip8 : 3.210

Dari nilai-nilai di atas kita dapat mengetahui mana data yang nantinya di buat untuk penelitian. Di atas sudah di jelaskan bahwa data yang pake adalah data yang memiliki nilai kurang dari 3.00 jadi data yang di pakai untuk penelitian nanti adalah ipk1 – ipk4 karna memiliki nilai kurang dari 3.00.

Hasil pengolahan data yang diperoleh adalah sebagai berikut :

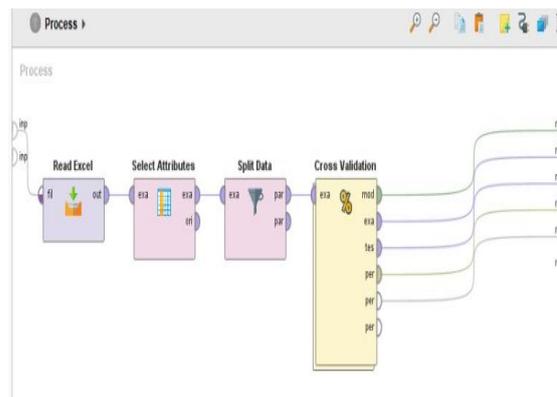
ORGANISASI	STATUS	IP1	IP2	IP3	IP4	LABEL
		19	19,00	20	20	
TIDAK	MANDIRI	2,8 3	2,75	2,54	3,5	Telat
TIDAK	MANDIRI	3,2 9	3	3,55	3,35	Telat
TIDAK	MANDIRI	3,3 1	3,17	3,5	3,35	Tepat waktu
TIDAK	MANDIRI	3,2 9	3,14	3,75	3,6	Tepat waktu
TIDAK	MANDIRI	3,2	3,04	3,45	3,75	Tepat waktu
AKTIF	MANDIRI	2,9 3	2,92	3,2	3,35	Tepat waktu
TIDAK	MANDIRI	3,1 5	2,99	3,2	3,25	Telat
TIDAK	MANDIRI	3,1 3	3,02	3,55	3,7	Tepat waktu
TIDAK	MANDIRI	3,3 9	3,15	3,55	3,7	Tepat waktu
TIDAK	MANDIRI	3,3 6	3,16	3,7	3,25	Telat
TIDAK	MANDIRI	3,2 2	2,96	3,3	3,1	Tepat waktu
TIDAK	MANDIRI	3,2 3	3,07	3,8	3,25	Tepat waktu

Tabel 2 hasil pengolahan data

Setelah pengolahan selesai dilakukan pengujian model menggunakan aplikasi *Rapid Miner* dengan menggunakan algoritma *Decision Tree*.

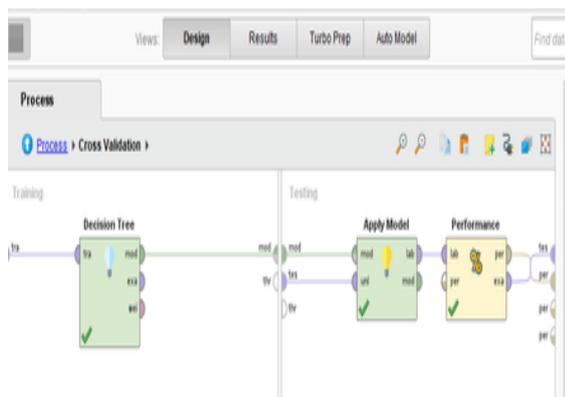
pengujian model menggunakan Algoritma *Decision Tree*

Dalam pengujian model peneliti menggunakan algoritma *Decision Tree* adapun permodelan yang di buat oleh peneliti sebagai berikut :



Gambar 2 permodelan rapid miner

Dari gambar di atas adalah permodelan yang sudah di buat . *Read Excell* digunakan untuk membaca database penelitian yang di gunakan.setelah itu masuk kedalam select atribut untuk menentukan atribut manakah yang memiliki nilai signifikan yang tinggi peneliti memilih menggunakan select atribut karna select atribut adalah tools yang paling mudah di gunakan dalam penelitian dan select atribut juga termasuk tools yang paling banyak di gunakan untuk melakukan penelitian. Setelah select atribut selesai lanjut ke proses split. *split* Data digunakan untuk membagi data menjadi data *training* dan data *testing*, data *training* sebanyak 30 % dan data *testing* sebanyak 70%. Dan *Cross validation* digunakan untuk pengujian model yang sudah di buat.



Gambar 3 permodelan algoritma

Gambar di atas adalah gambar permodelan algoritma *Decision Tree* untuk mengetahui Performa yang di dapat dari permodelan tersebut. Adapun hasil yang di dapat sebagai berikut :

Row No.	Label	prediction...	confidence...	confidence...	IP1	IP2	IP3	IP4	ORGANISASI	STATUS PE...
1	TELAT	TELAT	0,972	0,028	2,830	2,750	2,540	3,500	AKIF	MANDIRI
2	TELAT	TEPAT WAKTU	0	1	3,220	3,070	3,800	3,250	TIDAK	MANDIRI
3	TEPAT WAKTU	TEPAT WAKTU	0	1	3,180	3,030	3,750	3,600	TIDAK	MANDIRI
4	TEPAT WAKTU	TEPAT WAKTU	0	1	3,270	3,070	3,550	3,700	TIDAK	MANDIRI
5	TEPAT WAKTU	TEPAT WAKTU	0	1	3,440	3,020	3,450	3,200	TIDAK	MANDIRI
6	TEPAT WAKTU	TEPAT WAKTU	0	1	3,350	3,150	3,600	3,350	TIDAK	MANDIRI
7	TELAT	TELAT	0,972	0,028	3,080	2,980	3,200	3,200	TIDAK	MANDIRI
8	TEPAT WAKTU	TEPAT WAKTU	0	1	3,280	3,130	3,250	3,350	TIDAK	MANDIRI
9	TEPAT WAKTU	TEPAT WAKTU	0	1	3,270	3,050	3,750	3,600	TIDAK	MANDIRI
10	TELAT	TELAT	0,971	0,029	3,030	2,800	3,250	2,900	TIDAK	MANDIRI
11	TEPAT WAKTU	TEPAT WAKTU	0	1	3,420	3,320	3,850	3,750	TIDAK	MANDIRI
12	TELAT	TELAT	0,971	0,029	3,130	2,930	3,200	3,200	TIDAK	MANDIRI
13	TELAT	TELAT	0,971	0,029	3,200	2,870	3,600	3,350	TIDAK	MANDIRI
14	TELAT	TELAT	0,971	0,029	3,280	2,870	3,200	3,450	TIDAK	MANDIRI
15	TEPAT WAKTU	TEPAT WAKTU	0	1	3,390	3,250	3,650	3,650	TIDAK	MANDIRI

Gambar 4 Hasil *Cros validation*

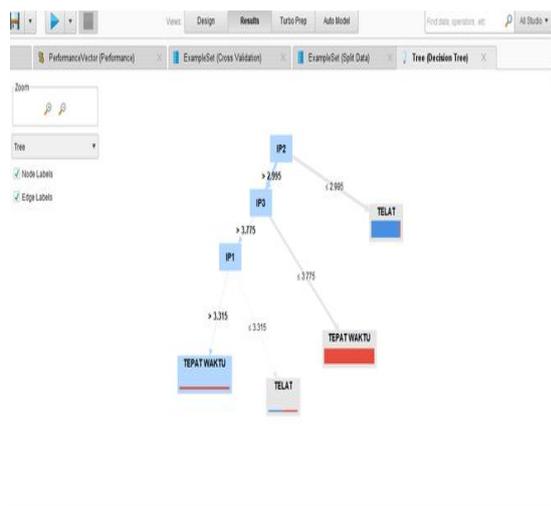
Gambar di atas adalah hasil dari *Cross Validation* dari gambar tersebut kita bisa melihat bahwa nomer satu dengan Label Telat meliki Prediction Telat nomer

dua dengan Label telat memiliki Prediksi Tepat waktu dan seterusnya

	Row TELAT	Row TEPAT WAKTU	class precision
Row TELAT	37	2	94,87%
Row TEPAT WAKTU	1	38	97,44%
class recall	97,37%	95,00%	

Gambar 5 Akurasi yang di dapat dari *Decision Tree*

Gambar di atas merupakan ilustrasi hasil kinerja yang dicapai oleh algoritma *Decision Tree*. Dari gambar tersebut terlihat bahwa algoritma *Decision Tree* mempunyai akurasi sebesar 96,25%. Nilai *recall* masing-masing kelas sebesar 97,37% untuk kelas Late dan 95,00% untuk kelas On-Time. Sedangkan nilai presisi tiap kelas sebesar 94,87% untuk kelas “Terlambat” dan 97,44% untuk kelas “Tepat Waktu”. Nilai akurasi tersebut disebabkan karena adanya beberapa data yang salah klasifikasi oleh sistem, yaitu 1 data berlabel “Terlambat” tergolong data “Tepat Waktu” dan 2 data berlabel “Tepat Waktu” tergolong data Terlambat.



Gambar 6 pohon keputusan

Dari gambar pohon keputusan tersebut kita bisa menyimpulkan bahwa :
Dari gambar pohon keputusan tersebut kita bisa menyimpulkan bahwa :

IP2 > 2.995 : Telat

IP3 > 3.775 : Tepat Waktu

IP1 > 3.315 : Tepat waktu {Telat = 0, Tepat waktu = 4}

IP1 ≤ 3.315 : Telat {Telat = 1, Tepat waktu = 1}

IP3 ≤ 3.775 : Tepat waktu {telat = 0, Tepat Waktu = 34}

IP2 ≤ 2,995 : Telat {Telat = 37, Tepat waktu = 1}

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian yang telah disampaikan pada bab sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Kinerja yang dicapai oleh algoritma Decision Tree mempunyai akurasi sebesar 90,25%. Nilai recall masing-masing kelas sebesar 97,37% untuk kelas tidak tepat waktu dan 95,00% untuk kelas tepat waktu. Sedangkan nilai presisi tiap kelas sebesar 94,87% untuk kelas “Terlambat” dan 97,44% untuk kelas “Tepat Waktu”.
2. IP di bawah 2,995 diperkirakan akan terlambat, sedangkan IP di atas 3,775 diperkirakan tepat waktu.

Saran

Adapun saran yang dapat dilakukan agar pengembangan sistem ini menjadi lebih baik adalah pada penelitian selanjutnya bisa menggunakan algoritma lain yang digunakan untuk penelitian ini.

DAFTAR PUSTAKA

- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640.
<https://doi.org/10.30865/mib.v5i2.2937>
- Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk

- Memprediksi Mahasiswa Lulus Tepat Waktu. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 265–270.
<https://doi.org/10.33480/jitk.v5i2.1214>
- Dengen, C. N., Kusriani, K., & Luthfi, E. T. (2020). Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Sisfotenika*, 10(1), 1.
<https://doi.org/10.30700/jst.v10i1.484>
- 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*, 1(1), 30–36.
<https://doi.org/10.35746/jtim.v1i1.11>
- Indah Lestari, Y., & Defit, S. (2021). *Jurnal Informatika Ekonomi Bisnis Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C.45*. 3, 148–154.
<https://doi.org/10.37034/infeb.v3i3.104>
- Mashlahah. (2013). *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*.
- Mayssara A. Abo Hassanin Supervised, A. (2014). Klasifikasi Predikat Kelulusan Mahasiswa dengan Metode CRISP-DM dan Naïve Bayes pada Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta. *Paper Knowledge . Toward a Media History of Documents*.
- Novita Sari, V., Yupianti, Y., & Maharani, D. (2018). Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan. *Jurteks*, 4(2), 133–140.
<https://doi.org/10.33330/jurteks.v4i2.53>
- Rakhman, A. (2017). Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization (PSO). *Smart Comp :Jurnalnya Orang Pintar Komputer*, 6(1), 193–197.
<http://ejournal.poltektegal.ac.id/index.php/smartcomp/article/view/466>
- Rohman, A., & Mujiyono, S. (2021). Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4.5. *Jurnal Prodi Teknik Informatika UNW "Multimatrix*, II(2), 1–5.
- Rohman, A., & Rufiyanto, A. (2019). Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5

- Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran. *Proceeding SINTAK 2019*, 134–139.
- Romadhona, A., Suprapedi, & Himawan, H. (2017). Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree. *Jurnal Teknologi Informasi*, 13(1), 69–83.
- Syahrin, A. (2013). Implementasi algoritma k-means untuk klasterisasi mahasiswa berdasarkan prediksi waktu kelulusan skripsi. *UPN "Veteran" Jatim*, 1–23.
- Wahyu, A. G. (2022). *Keunggulan Universitas An Nuur Purwodadi*. Universitas.123.Com.
<https://www.universitas123.com/news/keunggulan-universitas-an-nur-purwodadi>