

Prediksi Kelulusan Mahasiswa Prodi Informatika dengan Algoritma Decision Tree (C4.5) dan Naive Bayes

Oleh :

Steven Gerrard

Ade Evianti, S.Kom., M.Kom

Hamzah Setiawan, S.Kom., M.Kom

Ika Ratna Indra Astutik, S.Kom., MT

Program Studi Informatika
Universitas Muhammadiyah Sidoarjo
2026



Latar Belakang



Persentase mahasiswa yang lulus tepat waktu adalah parameter utama penentu kualitas perguruan tinggi dan proses akreditasi

Realitanya, banyak mahasiswa menghadapi kendala untuk menuntaskan masa studinya dalam batas waktu ideal (maksimal 7 semester untuk S1)

PERMENDIRISTEKDIKTI NO. 44/2015

Dibutuhkan sebuah strategi berbasis data untuk memproyeksikan peluang kelulusan mahasiswa sejak dini

Rumusan Masalah & Tujuan

Rumusan masalah penelitian ini sebagai berikut :

1. Bagaimana membangun model prediksi untuk mengetahui ketepatan waktu kelulusan mahasiswa Universitas Muhammadiyah Sidoarjo?
2. Sejauh mana algoritma C4.5 dan Naive Bayes mampu memprediksi kelulusan mahasiswa tepat waktu berdasarkan atribut akademik dan non-akademik yang tersedia ?
3. Algoritma manakah yang memiliki performa prediksi yang lebih baik dalam menentukan ketepatan waktu kelulusan mahasiswa ?.

Tujuan dari penelitian ini adalah sebagai berikut :

1. Membangun model prediksi ketepatan waktu kelulusan mahasiswa Universitas Muhammadiyah Sidoarjo menggunakan algoritma C4.5 dan Naive bayes
2. Menganalisis kemampuan masing-masing algoritma dalam melakukan klasifikasi kelulusan tepat waktu berdasarkan data akademik dan atribut pendukung lainnya
3. Membandingkan hasil akurasi klasifikasi antara algoritma C4.5 dan Naive Bayes dalam memprediksi ketepatan waktu kelulusan mahasiswa

Manfaat Penelitian

- ✓ Memberikan Informasi awal untuk mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu sehingga dapat dilakukan intervensi dini seperti bimbingan akademik tambahan
- ✓ Menjadi alat bantu dalam evaluasi proses akademik mahasiswa dan sebagai dasar dalam pengamblan Keputusan strategis untuk peningkatan mutu Pendidikan
- ✓ Menjadi referensi dan dasar pengembangan penelitian lebih lanjut dalam bidang edukasi data mining, khususnya dalam penerapan algoritma klasifikasi untuk prediksi akademik

Perbandingan Penelitian

A. Wahyudi dkk
(2023)

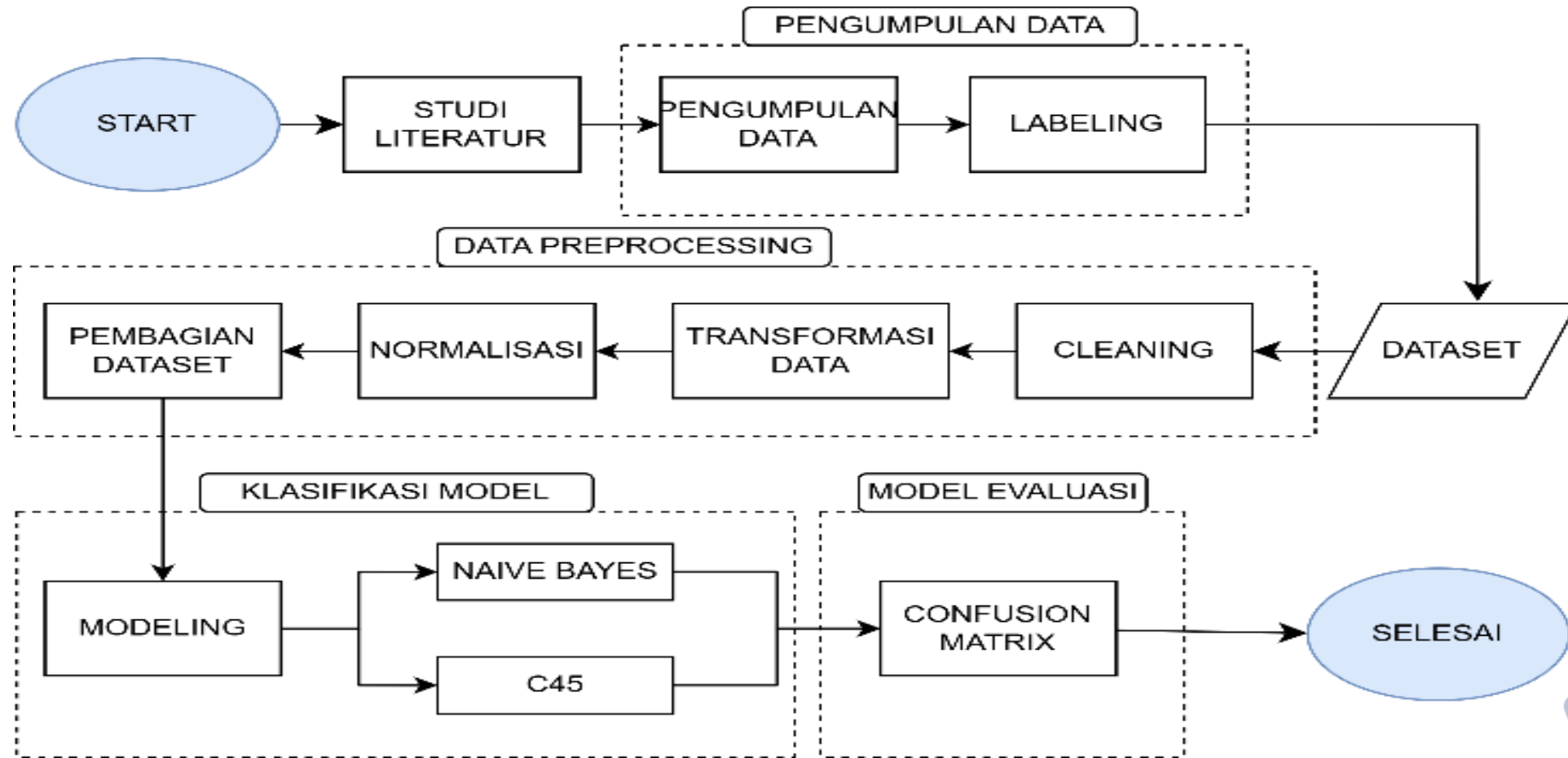
Kassymova dkk.
(2022)

Rovidatul dkk.
(2023)

Mursidil Arif, dkk.
(2023)

1. Penambahan Atribut Non-Akademik: Membuktikan bahwa nilai keaktifan non-akademik (SKEK) adalah penentu utama kelulusan, tidak sekadar mengandalkan nilai akademik awal seperti riset Wahyudi dan Kassymova
2. Lonjakan Akurasi: Berhasil menaikkan akurasi Naïve Bayes secara drastis dari 68% (pada riset Mursidil Arif di prodi yang sama) menjadi 94,13%.
3. Membongkar Overfitting C4.5: Mengkritisi hasil 100% C4.5 yang rawan overfitting, dan membuktikan bahwa Naïve Bayes justru memberikan prediksi yang lebih realistis (mendukung temuan Rovidatul)

Rancangan Penelitian



Dataset & Variabel Prediksi

Subjek

161 Entri rekam akademik mahasiswa
Program Studi Informatika Umsida



Faktor Non-Akademik

Jenis kelamin, status/skor kelulusan
PKMU, nilai BQ, nilai Ibadah, dan
akumulasi total poin SKEK



Faktor Akademik

Nilai Indeks Prestasi Semester (IPS 1
Hingga 6) dan IPK.



Data Prerocessing

```
...
      Nim      other
      self
0  22108020001 22108020001
1  22108020002 22108020002
2  22108020003 22108020003
3  22108020004 22108020004
4  22108020005 22108020005
..
...
157 221080200158 221080200158 ARRY WAHYU SETIAWAN
158 221080200159 221080200159 NaN
159 221080200160 221080200160 MUHAMMAD ANUGRAH RAMA DHANAWAN
160 221080200161 221080200161 NaN
161 221080200162 221080200162 NaN

      other
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
..
...
157      ARRY WAHYU SETIAWAN
158      NaN
159 MUHAMMAD ANUGRAH RAMA DHANAWAN
160      NaN
161      NaN
```

Cleaning

Membersihkan dataset dari isian kosong (*missing value*) dan format yang berantakan

Data Prerocessing

df_fix

	NO	Nim_x	Nama	Jenis Kelamin	IP51	IP52	IP53	IP54	IP55	IP56	IPK	NI
0	6.0	221080200006	ABDUL LATIF	L	3.47	3.58	3.89	3.92	3.81	3.27	3.69	
1	70.0	221080200070	ACH ISRO' MUJAHIDIN	L	3.67	3.8	3.72	3.06	3.36	3.6	3.52	
2	157.0	221080200157	ACHID SIROJ MUFADDHOL	L	3.52	3.48	3.7	1.97	2.56	0	2.74	
3	49.0	221080200049	ACHMAD RAIHAN RAMADHAN	L	3.58	3.5	3.59	3.56	3.51	3.33	3.51	
4	132.0	221080200132	ACHMAD SYAFIQ IZZUL KAHFI	L	3.13	2.12	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	
...
158	55.0	221080200055	YUDO DWINATA	L	3.40	3.6	3.65	3.65	2.68	3.6	3.44	
159	1.0	221080200001	YURIDA ZANI	P	3.45	3.8	3.63	3.8	3.91	3.72	3.72	
160	33.0	221080200033	ZAHRA NUR HABIBAH	Tidak ditemukan	0.00	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	
161	34.0	221080200034	ZAIDAN DAFA ROBANI HATUWE	L	3.50	3.72	3.74	3.57	3.41	3.67	3.59	
162	98.0	221080200098	ZULFIAN AHMAD ZULKARNAEN	Tidak ditemukan	0.00	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	Tidak ditemukan	

163 rows x 18 columns

Transformasi
Konversi Data guna menyelaraskan tatanan format data agar dapat diproses oleh model dengan lebih baik,

Data Prerocessing

Transformasi

```
df
```

	Nim	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPK	nilai_pkmu	status_pkmu	BQ	IBADAH	status_bq_ibadah	Total Point	status_final
0	221080200006	L	3.47	3.58	3.89	3.92	3.81	3.27	3.69	4.0	LULUS	325	4	LULUS	95	TIDAK LULUS TEPAT WAKTU
1	221080200070	L	3.67	3.80	3.72	3.06	3.36	3.60	3.52	4.0	LULUS	3	3	LULUS	130	LULUS TEPAT WAKTU
2	221080200157	L	3.52	3.48	3.70	1.97	2.56	0.00	2.74	4.0	LULUS	3	35	LULUS	60	TIDAK LULUS TEPAT WAKTU
3	221080200049	L	3.58	3.50	3.59	3.56	3.51	3.33	3.51	4.0	LULUS	3	3	LULUS	190	LULUS TEPAT WAKTU
4	221080200132	L	3.13	2.12	0.00	0.00	0.00	0.00	0.00	0.0	TIDAK LULUS	0	0	TIDAK LULUS	10	TIDAK LULUS TEPAT WAKTU
...
156	221080200055	L	3.40	3.60	3.65	3.65	2.68	3.60	3.44	4.0	LULUS	325	4	LULUS	175	LULUS TEPAT WAKTU
157	221080200001	P	3.45	3.80	3.63	3.80	3.91	3.72	3.72	4.0	LULUS	3	3	LULUS	250	LULUS TEPAT WAKTU
158	221080200033	P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	TIDAK LULUS	0	0	TIDAK LULUS	0	TIDAK LULUS TEPAT WAKTU
159	221080200034	L	3.50	3.72	3.74	3.57	3.41	3.67	3.59	3.7	LULUS	325	3	LULUS	165	LULUS TEPAT WAKTU
160	221080200098	L	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	TIDAK LULUS	0	0	TIDAK LULUS	0	TIDAK LULUS TEPAT WAKTU

161 rows x 16 columns

Data Prerocessing

Data Asli		Encoding
Jenis Kelamin	L	0
	P	1
Status_pkmu	LULUS	0
	TIDAK LULUS	1
Status_bq_ibadah	LULUS	0
	TIDAK LULUS	1
Status_final	LULUS TEPAT WAKTU	0
	TIDAK LULUS TEPAT WAKTU	1

Normalisasi

Menjalankan proses encoding, yaitu mengubah data kategorikal (seperti 'Laki-laki' menjadi '0') agar berbentuk numerik dan siap diproses algoritma

Hasil Pengujian

Cross Validation

Algoritma C4.5

FOLD	Akurasi
Fold 1	100.00%
Fold 2	96.88%
Fold 3	100.00%
Fold 4	93.75%
Fold 5	96.88%
Rata - rata	97.50%

Naive Bayes

FOLD	Akurasi
Fold 1	93.94%
Fold 2	96.88%
Fold 3	84.38%
Fold 4	93.75%
Fold 5	100.00%
Rata - rata	93.79%

Hasil Pengujian

Confusion Matrix 70:30

Algoritma C4.5

	True Terlambat	True Tepat
False Terlambat	29	0
False Tepat	0	20

Naive Bayes

	True Terlambat	True Tepat
False Terlambat	29	0
False Tepat	3	17

Analisa Dan Temuan Menarik

Algoritma	Data Training	Data Testing	Akurasi
C45	70	30	100%
	80	20	100%
	90	10	100%
Naïve Bayes	70	30	93.88%
	80	20	90.91%
	90	10	94.13%

Walaupun C4.5 unggul sempurna, angka 100% berpotensi menandakan adanya overfitting atau feature leakage dari fitur tunggal yang mendominasi prediksi. Performa Naïve Bayes dianggap lebih tahan terhadap dominasi satu fitur dan lebih merepresentasikan kondisi nyata di lapangan

Kesimpulan

Riset memvalidasi bahwa kedua metode machine learning ini andal, dengan C4.5 mencatat akurasi tertinggi, sedangkan Naïve Bayes menyuguhkan performa yang stabil dan realistis. Prediksi ini dapat ditunjuk sebagai fondasi pembuatan sistem informasi akademik kampus cerdas berbasis AI. Kemudian saran untuk penelitian selanjutnya untuk memperluas volume dataset ke lintas Angkatan dan menguji metode algoritma lainnya untuk daya generalisasi yang makin mantab.

Referensi

A. M. Limbong and M. Asbari, "Transformasi Standar Nasional dan Akreditasi Pendidikan Tinggi," *JOURNAL OF INFORMATION SYSTEMS AND MANAGEMENT*, vol. 03, no. 01, pp. 101–105, 2024, doi: <https://doi.org/10.4444/jisma.v3i1.905>

Tim Penulis UMSIDA, *Pedoman Akademik Universitas Muhammadiyah Sidoarjo Tahun 2023-2024*. Sidoarjo: UMSIDA Press, 2023.

E. F. Wati, L. Indriyani, E. Sunita, and A. D. Kuswanto, "Optimasi Machine Learning dalam Memprediksi Kelulusan Mahasiswa," *JURNAL REKAYASA PERANGKAT LUNAK*, vol. 6, no. 2, pp. 137–142, Nov. 2025, doi: <https://doi.org/10.31294/reputasi.v6i2.11067>.

J. Khatib Sulaiman, I. Attyyatlulatifah, M. Kamayani, U. D. Muhammadiyah HAMKA, K. Kunci Prediksi Kelulusan, and P. Algoritma, "Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa Teknik Informatika dengan Orange Data Mining," *Indonesian Journal of Computer Science*, vol. Vol. 13 No. 2, Apr. 2024, doi: <https://doi.org/10.33022/ijcs.v13i2.3796>.

D. Wardhani *et al.*, *Pengantar Data Mining PT. MIFANDI MANDIRI DIGITAL*. Deli Serdang: PT. Mifandi Mandiri Digital, 2023

N. Khasanah, D. Uki, E. Saputri, T. Hidayat, F. Aziz, and U. N. Mandiri, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 dengan RapidMiner: Studi Kasus Data Akademik Perguruan Tinggi XYZ," *Journal Computer Science*, vol. 4, no. 2, pp. 100–107, 2025, doi: <https://doi.org/10.31294/ijcs.v4i2.9647>

Z. Fatah and M. Hasanah, "Penerapan Decision Tree Pada Klasifikasi Kelulusan Mahasiswa," *Jurnal Mahasiswa Teknik Informatika*, vol. 4, pp. 225–230, Aug. 2025, doi: <https://doi.org/10.35473/jamastika.v4i2.4487>.

Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755

Dyah Ardyani Rizqi Azizah Adha, Aulia Noveesa Allanda, Diah Ayu Fatmasari, and Siska Narulita, "Performansi Algoritma C4.5 untuk Prediksi Kelulusan Mahasiswa," *Jurnal Cakrawala Informasi*, vol. 3, no. 2, pp. 9–17, Dec. 2023, doi: 10.54066/jci.v3i2.339.

A. Wahyudi, Kusri, and F. Wahyu, "PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DECISION TREE DAN NAÏVE BAYES," *Jurnal Permata Indonesia*, vol. 14, no. 2, pp. 132–138, Nov. 2023, doi: 10.59737/jpi.v14i2.276

K. R. Diska and K. Budayawan, "Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naive Bayes Classifier (Studi Kasus: Prodi Pendidikan Teknik Informatika)," *JURNAL PENDIDIKAN TAMBUSAI*, vol. 7 no 1, pp. 936–943, Feb. 2023, doi: <https://doi.org/10.31004/jptam.v7i1.5375>

G. K. Kassymova and O. Ndayizeye, "Prediction model of teacher candidate student graduation status: Decision Tree C4.5, Naive Bayes, and k-NN," *Jurnal Penelitian Hukum dan Pendidikan*, vol. 21, no. 2, pp. 1407–1418, 2022, doi: <https://doi.org/10.30863/ekspose.v21i2.3407>.

Imam Riadi, Rusydi Umar, and Rio Anggara, "Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode Naive Bayes," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 1, pp. 191–203, Jan. 2024, doi: 10.51454/decode.v4i1.308.

Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naive Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN)," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 109–119, Jun. 2024, doi: 10.52158/jacost.v5i1.489.

I. Penulis, N. Kholik Afandi UINSI Samarinda, and I. Artikel, "Literature Review is A Part of Research," *Sulawesi Tenggara Educational Journal*, vol. 1, no. 3, pp. 64–71, Dec. 2021, doi: <https://doi.org/10.54297/seju.v1i3.203>.

A. W. Anggraeni, A. S. Fitriani, and A. Eviyanti, "Penerapan Algoritma Support Vector Machine untuk Memprediksi Tingkat Partisipasi Pemilu terhadap Kualitas Pendidikan," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 21–27, Jun. 2024, doi: 10.29408/edumatic.v8i1.24838.

Z. Setiawan *et al.*, *BUKU AJAR DATA MINING*. Yogyakarta: PT. Sonpedia Publishing Indonesia, 2023. [Online]. Available: www.buku.sonpedia.com

R. Limia Budiarti and N. Kahar, "ANALISIS PREDIKSI KELULUSAN MAHASISWA FAKULTAS ILMU KOMPUTER UNIVERSITAS NURDIN HAMZAH MENGGUNAKAN METODE NAIVE BAYES," *JOURNAL OF INFORMATION TECHNOLOGY*, vol. Vol. 8 No.2, pp. 48–55, Dec. 2024, doi: <https://doi.org/10.53564/fortech.v8i2.1312>.

S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, vol. 27(4s), pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.

J. Allgaier and R. Pryss, "Cross-Validation Visualized: A Narrative Guide to Advanced Methods," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 2, pp. 1378–1388, Jun. 2024, doi: 10.3390/make6020065.

A. S. Fitriani, M. A. Rosid, C. Taurusta, and I. Fauzia, "Classification Using C4.5 Algorithm in Election Participation Prediction," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, pp. 1–9, Jul. 2020, doi: 10.1088/1757-899X/874/1/012016.

W. A. Firmansyach, U. Hayati, and Y. A. Wijaya, "ANALISA TERJADINYA OVERFITTING DAN UNDERFITTING PADA ALGORITMA NAIVE BAYES DAN DECISION TREE DENGAN TEKNIK CROSS VALIDATION," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, pp. 262–269, Feb. 2023, doi: <https://doi.org/10.36040/jati.v7i1.6329>.

M. M. Arif, H. Setiawan, A. S. Fitriani, F. Sains, and D. Teknologi, "Penggunaan Datamining Untuk Memprediksi Masa Studi Mahasiswa di Universitas Muhammadiyah Sidoarjo Dengan Algoritma Naive Bayes," vol. 4, no. 3, pp. 622–629, Jul. 2023, doi: <https://doi.org/10.30645/kesatria.v4i3.210.g209>.

Sekian,
Terimakasih

