



# **PREDIKSI DATA KELULUSAN MAHASISWA**

---

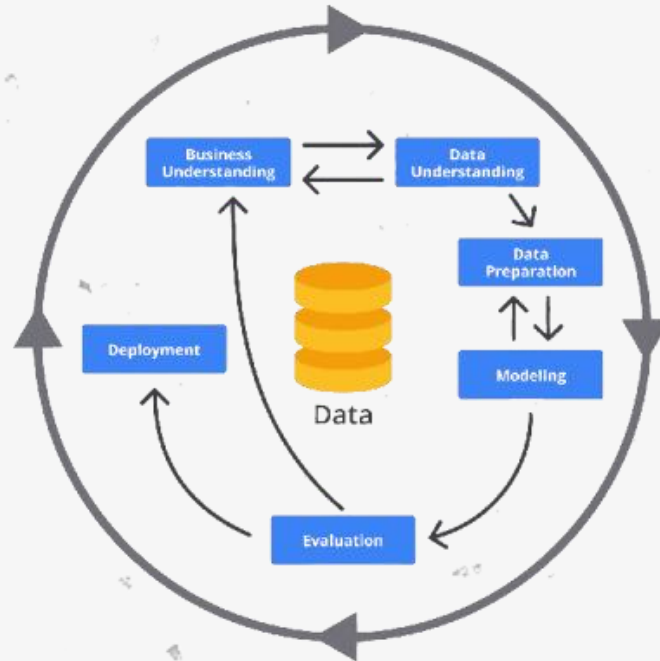
DISYA NURUL ARIZA – DATA SCIENTIST

# Pendahuluan

- Pendidikan tinggi memiliki peran penting dalam mencetak sumber daya manusia yang berkualitas.
- Salah satu indikator keberhasilan perguruan tinggi adalah tingkat kelulusan mahasiswanya.
- Data kelulusan mahasiswa memberikan gambaran mengenai pencapaian akademik, efektivitas kurikulum, serta kualitas proses pembelajaran.
- Analisis terhadap data kelulusan ini penting untuk mengetahui tren, faktor pendukung, dan kendala yang dihadapi mahasiswa dalam menyelesaikan studi tepat waktu.
- Dengan adanya data tersebut, perguruan tinggi dapat melakukan evaluasi dan perbaikan untuk meningkatkan kualitas pendidikan dan layanan akademik.

# Metodologi

Proses analisis data mengikuti kaidah CRISP-DM (Cross-Industry Standard Process for Data Mining).



# Business Understanding I

Mendefinisikan Masalah

01

## Target User

Pihak Kampus

03

## Kenapa Penting

Kelulusan tepat waktu penting untuk akreditasi kampus, efisiensi biaya mahasiswa, serta evaluasi kualitas pendidikan tinggi secara nasional.

02

## Masalah

Masih banyak mahasiswa yang belum lulus tepat waktu. Kampus juga belum memiliki sistem analisis untuk mendeteksi sejak dini mahasiswa yang berisiko terlambat.

04

## Kondisi saat ini

Data kelulusan tersedia, tetapi baru digunakan sebatas laporan deskriptif. Belum ada analisis prediktif untuk memahami pola keterlambatan kelulusan.

# Business Understanding 2

## Mendefinisikan KPI (Key Performance Indicators)

**Ekspektasi Manfaat:** Meningkatkan kelulusan tepat waktu

**Metric:** Persentase kelulusan tepat waktu

**Target:**  $\geq 80\%$

Menekan lama studi mahasiswa

**Metric:** Rata-rata lama studi

**Target:**  $\leq 4$  tahun (S1)

Menjaga kualitas akademik lulusan

**Metric:** Rata-rata IPK

**Target:**  $\geq 3,25$

Mengurangi mahasiswa tidak lulus/dropout

**Metric:** Dropout rate

**Target:**  $\leq 5\%$

## Mendefinisikan faktor (features)

- **Akademik:** IPK, jumlah SKS per semester, nilai mata kuliah inti.
- **Demografis:** usia, jenis kelamin, status pernikahan dan status mahasiswa
- **Target (Label):** status kelulusan (tepat waktu / tidak lulus).

# DATA UNDERSTANDING I

## Sumber Data:

berasal dari Sistem Informasi Akademik  
(SIAKAD)

## Validasi Data:

- memastikan kelengkapan (tidak ada data kosong)
- konsistensi format (IPK 0–4)
- keakuratan (sesuai catatan resmi)
- deteksi outlier (misalnya IPK > 4 atau lama studi < 6 semester).

# DATA UNDERSTANDING 2

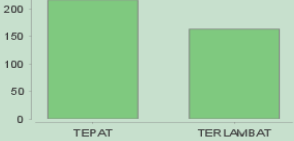
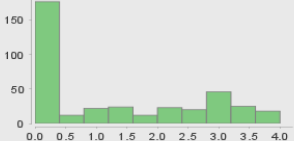
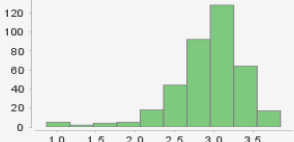
TAMPILAN DATA – 379 data dengan atribut 14



Row No.	STATUS KE...	NAMA	JENIS KEL...	STATUS MA...	UMUR	STATUS NI...	IPS 1	IPS 2	IPS 3	IPS 4	IPS
1	TERLAMBAT	ANIK WIDAY...	PEREMPUAN	BEKERJA	28	BELUM MEN...	2.760	2.800	3.200	3.170	2.98
2	TERLAMBAT	DWI HESTY...	PEREMPUAN	MAHASISWA	32	BELUM MEN...	3	3.300	3.140	3.140	2.84
3	TERLAMBAT	MURYA ARIE...	PEREMPUAN	BEKERJA	29	BELUM MEN...	3.500	3.300	3.700	3.290	3.53
4	TERLAMBAT	NANIK SUSA...	PEREMPUAN	MAHASISWA	27	BELUM MEN...	3.170	3.410	3.610	3.360	3.48
5	TERLAMBAT	RIFKA ISTIQ...	PEREMPUAN	BEKERJA	29	BELUM MEN...	2.900	2.890	3.300	2.850	2.98
6	TERLAMBAT	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MEN...	2.950	2.820	3.090	3.100	2.78
7	TEPAT	FARIKHATU...	PEREMPUAN	MAHASISWA	26	BELUM MEN...	2.760	3.140	2.600	2.950	3.23
8	TEPAT	FIFI SUNALISA	PEREMPUAN	MAHASISWA	27	BELUM MEN...	2.620	2.890	2.320	2.500	2.50
9	TERLAMBAT	HENDRIK M...	PEREMPUAN	BEKERJA	25	MENIKAH	3.600	3.540	3.520	3.390	3.52
10	TERLAMBAT	IMAM AGUN...	PEREMPUAN	BEKERJA	28	BELUM MEN...	2.710	2.550	1.770	2.110	1.93
11	TERLAMBAT	IMAM SANTO...	PEREMPUAN	BEKERJA	27	BELUM MEN...	3.140	3.460	3.400	3.430	3.27

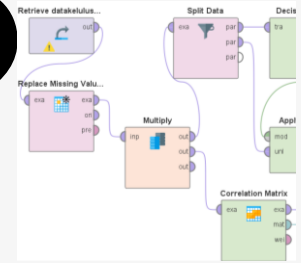
# DATA UNDERSTANDING (3)

Statistika Deskriptif Exploratory Data Analysis (EDA), Tidak Ada missing value

Name	Type	Missing	Statistics	Filter (15 / 15 attributes): <input type="text" value="Search for Attributes"/>
<div>how the data in a table</div> <div>Label</div> <div>STATUS KELULUSAN</div>	Binominal	0	 <div>Negative TERLAMBAT</div> <div>Positive TEPAT</div> <div>Values TEPAT (216)</div> <div><a href="#">Details...</a></div>	
<div>IPS 8</div>	Real	0	 <div>Min 0</div> <div>Max 4</div> <div>Average 1.272</div> <div>Deviation 1.388</div> <div><a href="#">Open visualizations</a></div>	
<div>IPK</div>	Real	0	 <div>Min 0.870</div> <div>Max 3.850</div> <div>Average 2.933</div> <div>Deviation 0.458</div> <div><a href="#">Open visualizations</a></div>	



# DATA UNDERSTANDING (4)



Memeriksa hubungan antar variable  
menambahkan operator correlation matrix

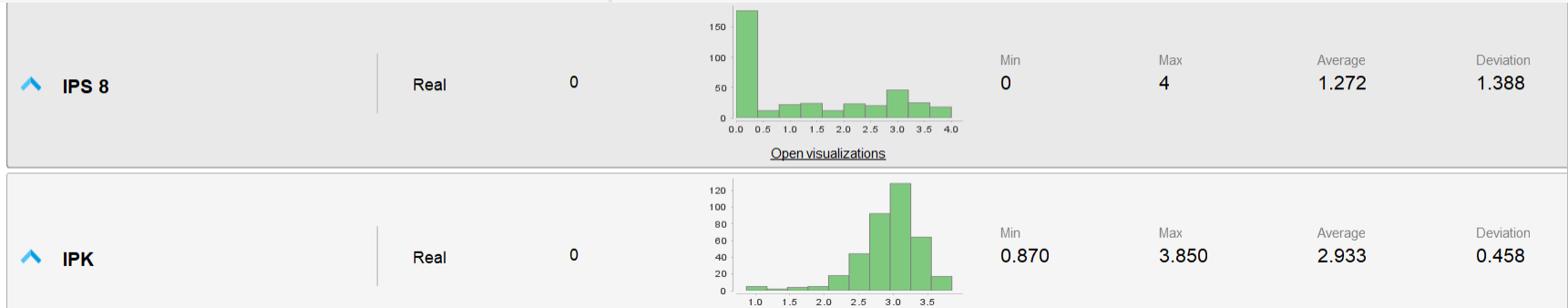
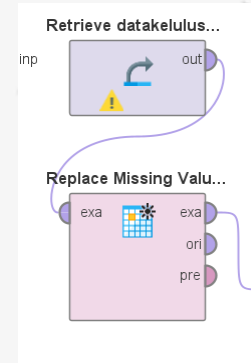
Attribut...	NAMA	JENIS ...	STATU...	UMUR	STATU...	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK
UMUR	?	0.015	-0.204	1	0.148	-0.060	-0.064	-0.094	-0.100	-0.115	-0.027	-0.018	-0.144	-0.068
STATUS...	?	0.002	-0.123	0.148	1	0.051	0.136	0.003	0.060	0.054	0.074	0.089	-0.016	0.061
IPS 1	?	-0.200	0.228	-0.060	0.051	1	0.543	0.620	0.549	0.562	0.455	0.265	-0.128	0.551
IPS 2	?	-0.191	0.234	-0.064	0.136	0.543	1	0.593	0.704	0.692	0.587	0.464	0.083	0.661
IPS 3	?	-0.178	0.168	-0.094	0.003	0.620	0.593	1	0.698	0.679	0.578	0.313	-0.099	0.688
IPS 4	?	-0.142	0.303	-0.100	0.060	0.549	0.704	0.698	1	0.799	0.721	0.513	0.077	0.741
IPS 5	?	-0.153	0.323	-0.115	0.054	0.562	0.692	0.679	0.799	1	0.675	0.468	0.043	0.743
IPS 6	?	-0.105	0.166	-0.027	0.074	0.455	0.587	0.578	0.721	0.675	1	0.585	0.046	0.683
IPS 7	?	-0.083	0.047	-0.018	0.089	0.265	0.464	0.313	0.513	0.468	0.585	1	0.107	0.539
IPS 8	?	0.146	0.109	-0.144	-0.016	-0.128	0.083	-0.099	0.077	0.043	0.046	0.107	1	0.081
IPK	?	-0.194	0.158	-0.068	0.061	0.551	0.661	0.688	0.741	0.743	0.683	0.539	0.081	1

# DATA PREPARATION I

- Merubah data nominal (kategori) menjadi numerik
- Mengatasi missing data
- Mengatasi outliers
- Melakukan penyamaan
- skala data (feature scaling)

# DATA PREPARATION 2

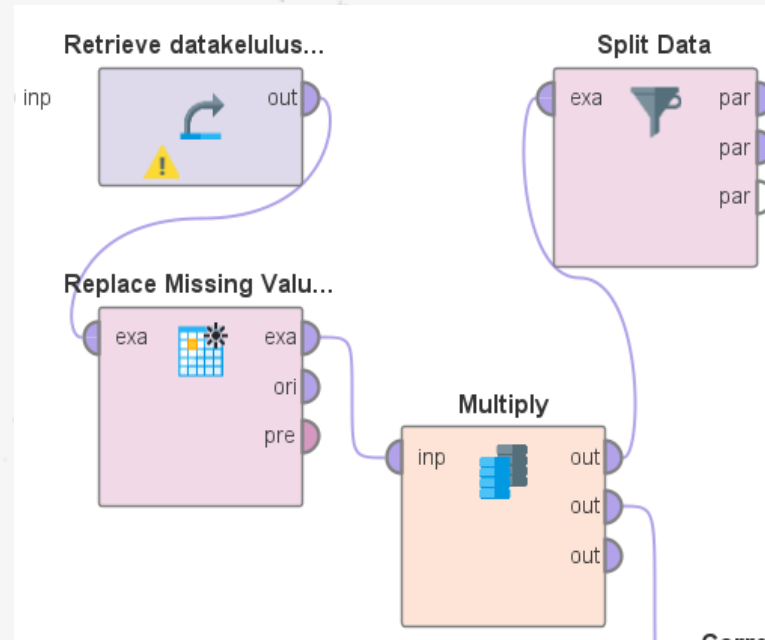
Mengatasi missing data dengan menambahkan  
operator replance missing value



# DATA PREPARATION 3

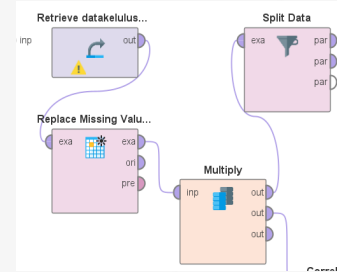
Menambahkan operator multiply, agar datanya

bisa bercabang dan bisa disambungkan.



# DATA PREPARATION 4

Menambahkan operator split data untuk membagi data menjadi data training dan data testing



Open in Turbo Prep Auto Model Interactive Analysis Filter (303 / 303 examples): all

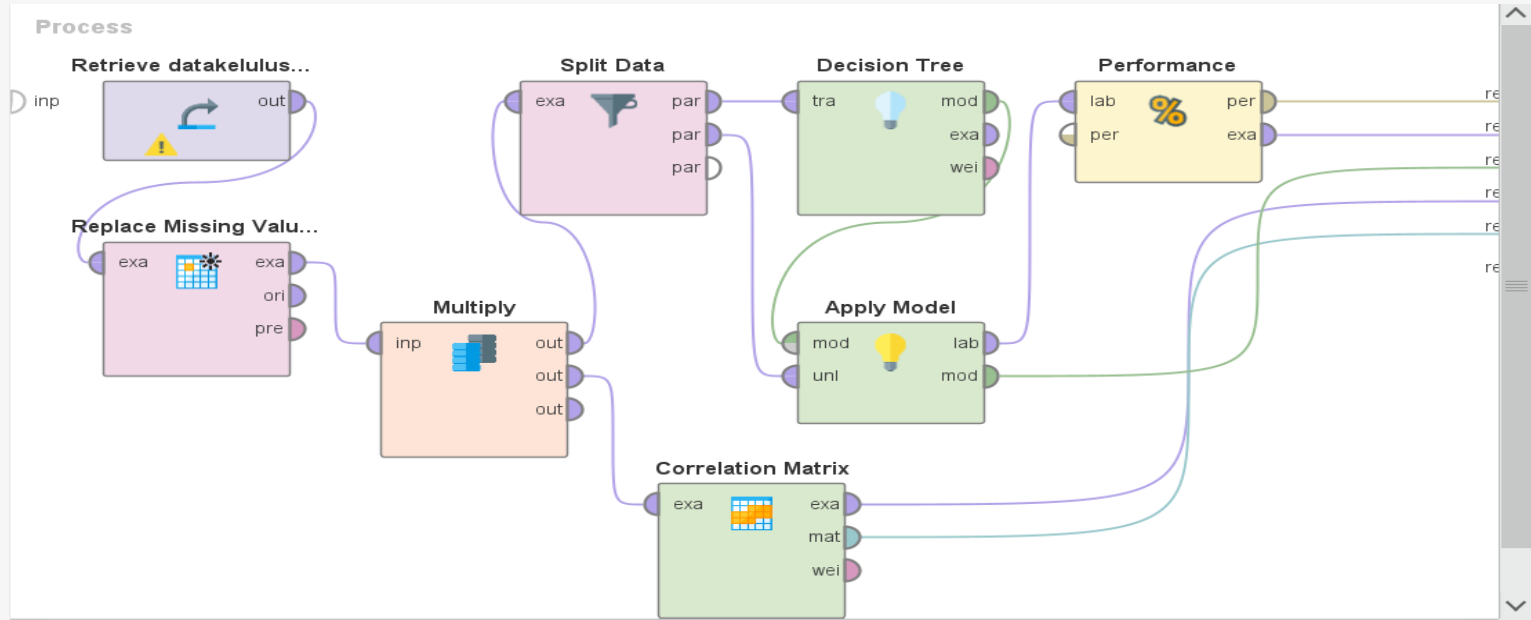
Row No.	STATUS KE...	NAMA	JENIS KEL...	STATUS MA...	UMUR	STATUS NI...	IPS 1	IPS 2	IPS 3	IPS 4	IPS
1	TERLAMBAT	ANIK WIDAY...	PEREMPUAN	BEKERJA	28	BELUM MEN...	2.760	2.800	3.200	3.170	2.98
2	TERLAMBAT	DWI HESTY...	PEREMPUAN	MAHASISWA	32	BELUM MEN...	3	3.300	3.140	3.140	2.84
3	TERLAMBAT	MURYA ARIE...	PEREMPUAN	BEKERJA	29	BELUM MEN...	3.500	3.300	3.700	3.290	3.53
4	TERLAMBAT	RIFKA ISTIQ...	PEREMPUAN	BEKERJA	29	BELUM MEN...	2.900	2.890	3.300	2.850	2.98
5	TERLAMBAT	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MEN...	2.950	2.820	3.090	3.100	2.78

Open in Turbo Prep Auto Model Interactive Analysis Filter (76 / 76 examples): all

Row No.	STATUS KE...	NAMA	JENIS KEL...	STATUS MA...	UMUR	STATUS NI...	IPS 1	IPS 2	IPS 3	IPS 4	IPS
1	TERLAMBAT	NANIK SUSA...	PEREMPUAN	MAHASISWA	27	BELUM MEN...	3.170	3.410	3.610	3.360	3.48
2	TERLAMBAT	M SYAIFULL...	PEREMPUAN	BEKERJA	31	BELUM MEN...	2.710	3	2.650	2.270	2.13
3	TERLAMBAT	DIANA LAILY ...	PEREMPUAN	MAHASISWA	26	BELUM MEN...	3.240	3.380	3.440	3.300	3.56
4	TERLAMBAT	ANDI SUNAR...	LAKI - LAKI	BEKERJA	28	BELUM MEN...	3	3.330	2.880	2.800	3.36
5	TERLAMBAT	MURDIANTO...	LAKI - LAKI	MAHASISWA	32	BELUM MEN...	2.450	2.450	2.140	1.950	1.82

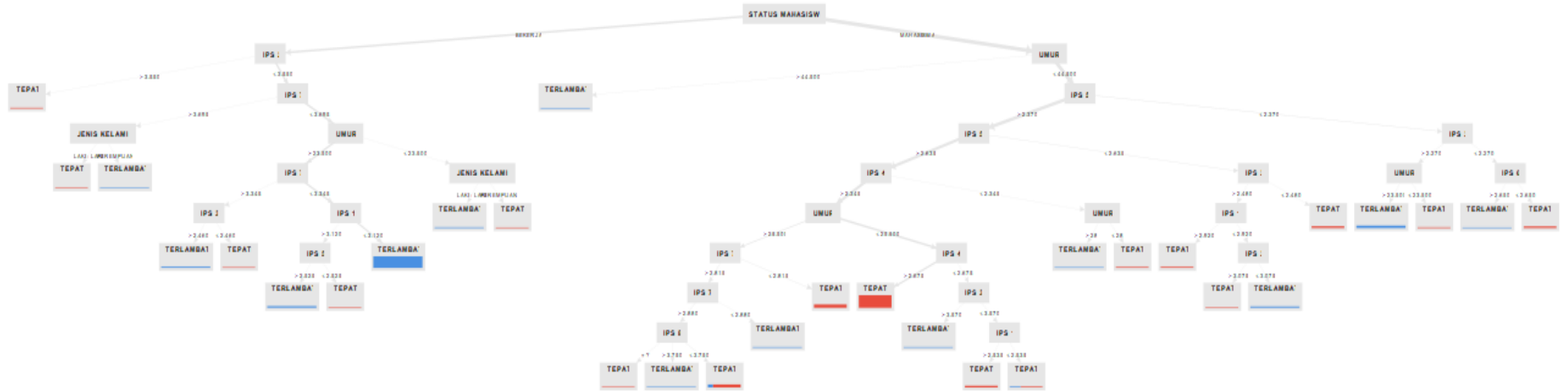
# MODELLING I

Algoritma yang digunakan Decision Tree (pohon keputusan). Data training masuk ke **Decision Tree**, lalu model yang dihasilkan diterapkan ke data uji menggunakan **Apply Model**.



# MODELLING 2

## Visualisasi dari Decision Tree



# EVALUASI I

Menambahkan operator Apply Model untuk melatih data testing

Row No.	STATUS KE...	prediction(S...	confidence(...	confidence(...	NAMA	JENIS KEL...	STATUS MA...	UMUR	STATUS NI...	IPS 1	IPS
1	TERLAMBAT	TEPAT	0.158	0.842	DWI HESTY...	PEREMPUAN	MAHASISWA	32	BELUM MEN...	3	3.30
2	TERLAMBAT	TERLAMBAT	1	0	MURYA ARIE...	PEREMPUAN	BEKERJA	29	BELUM MEN...	3.500	3.30
3	TERLAMBAT	TERLAMBAT	1	0	RIFKA ISTIQ...	PEREMPUAN	BEKERJA	29	BELUM MEN...	2.900	2.89
4	TERLAMBAT	TERLAMBAT	1	0	HENDRIK M...	PEREMPUAN	BEKERJA	25	MENIKAH	3.600	3.54
5	TERLAMBAT	TERLAMBAT	1	0	IMAM AGUN...	PEREMPUAN	BEKERJA	28	BELUM MEN...	2.710	2.55
6	TERLAMBAT	TERLAMBAT	1	0	IWAN HAMBALI	PEREMPUAN	BEKERJA	26	BELUM MEN...	2.570	2.82
7	TERLAMBAT	TEPAT	0	1	DONNY PRA...	LAKI - LAKI	MAHASISWA	27	BELUM MEN...	2.860	2.86
8	TERLAMBAT	TEPAT	0	1	AHMAD IBROZI	LAKI - LAKI	MAHASISWA	27	BELUM MEN...	2.670	2.68
9	TERLAMBAT	TERLAMBAT	1	0	ANDI SUNAR...	LAKI - LAKI	BEKERJA	28	BELUM MEN...	3	3.33
10	TERLAMBAT	TERLAMBAT	1	0	ANNY AZIZAH	LAKI - LAKI	BEKERJA	26	BELUM MEN...	3.120	3.23



# EVALUASI 2

- Performance (Classification) digunakan untuk menghitung metrik performa.
- Metrik utama: Akurasi dengan 90.79% yang artinya data dan pengolahan baik atau akurat

Criterion	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
accuracy	accuracy: 90.79%		
	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	28	2	93.33%
pred. TEPAT	5	41	89.13%
class recall	84.85%	95.35%	

# INSIGHT

- Mahasiswa dengan IPK  $\geq 3.25$  dan jumlah SKS yang stabil tiap semester memiliki kemungkinan lulus tepat waktu yang jauh lebih tinggi.
- IPK yang rendah sering kali berkorelasi dengan keterlambatan kelulusan akibat pengulangan mata kuliah atau penundaan skripsi.
- Mahasiswa yang mengambil  $< 18$  SKS per semester cenderung menambah masa studi hingga lebih dari 8 semester.
- Sebaliknya, mahasiswa yang konsisten mengambil SKS penuh setiap semester lebih cepat menyelesaikan studi.
- Mahasiswa yang bekerja sambil kuliah atau berusia di atas 25 tahun memiliki kemungkinan lebih tinggi untuk tidak lulus tepat waktu.
- Hal ini berkaitan dengan beban waktu dan prioritas belajar yang terbagi.
- Mahasiswa dengan nilai tinggi di mata kuliah inti (core subjects) menunjukkan tingkat ketepatan kelulusan tertinggi.
- Variasi besar antar-semester (fluktuasi IPK) menjadi sinyal risiko keterlambatan.

# REKOMENDASI BISNIS / KEBIJAKAN KAMPUS

1. Implementasi Sistem *Early Warning* untuk Mahasiswa Berisiko: Gunakan model prediksi ini untuk mendeteksi sejak dini mahasiswa yang berpotensi tidak lulus tepat waktu (berdasarkan IPK, SKS, atau status kerja), Mahasiswa dengan nilai risiko tinggi dapat diberi pendampingan akademik intensif, misalnya melalui konseling akademik atau bimbingan studi personal.
2. Optimalisasi Pengambilan SKS dan Rencana Studi: Buat sistem rekomendasi SKS yang menyeimbangkan beban studi dengan performa akademik, agar mahasiswa dapat merencanakan masa studi secara realistis dan optimal.
3. Penguatan Mata Kuliah Inti (Core Subjects): Lakukan evaluasi terhadap mata kuliah inti yang paling sering menjadi hambatan kelulusan, dan sediakan kelas remedial atau tutor sebaya (peer tutoring).
4. Program Pembinaan untuk Mahasiswa Bekerja: Sediakan kebijakan fleksibel seperti kelas malam atau hybrid learning agar mahasiswa yang bekerja tetap dapat menyelesaikan studi tepat waktu.
5. Monitoring Berkelanjutan oleh Dosen Wali: Dosen wali dapat menggunakan hasil analisis prediktif untuk melacak performa mahasiswa tiap semester dan memberi intervensi lebih awal.

# IMPLIKASI BISNIS / MANFAAT UNTUK KAMPUS

- Meningkatkan Akreditasi Program Studi melalui peningkatan persentase kelulusan tepat waktu.
- Efisiensi biaya operasional dengan menekan jumlah mahasiswa yang memperpanjang studi.
- Peningkatan reputasi institusi karena menunjukkan penerapan *data-driven decision making* dalam pengelolaan akademik.
- Dukungan kebijakan pendidikan adaptif, yang disesuaikan berdasarkan data performa mahasiswa.

**TERIMAKASIH**