

PRAKTIKUM
PENERAPAN METODOLOGI PENGEMBANGAN DATA MINING
CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING
(CRISP-DM)

Studi Kasus: Kelulusan Mahasiswa



Dosen Pengampu : Fitri Ayuning Tyas, M.Kom.

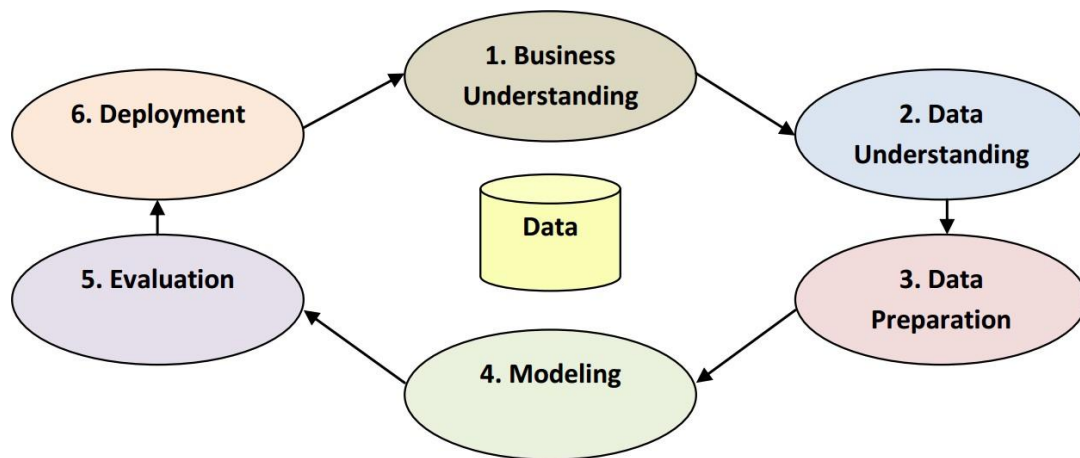
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS, TEKNOLOGI, DAN KESEHATAN
UNIVERSITAS MUHAMMADIYAH BREBES

A. Tujuan Praktikum

Mahasiswa mampu melakukan analisis permasalahan *data science* menggunakan tahapan CRISP-DM.

B. Pendahuluan

CRISP-DM adalah salah satu standard proses yang mampu mendukung penggunaan *data mining* untuk menyelesaikan masalah bisnis. Metode ini dirancang untuk membimbing organisasi melalui tahap-tahap yang terlibat dalam proyek *data mining*, mulai dari *business understanding* hingga implementasi solusi. CRISP-DM terdiri dari 6 tahapan, yaitu:



1. ***Business Understanding***: Memahami pengetahuan bisnis, objek bisnis, dan tujuan pemodelan untuk tujuan bisnis.
2. ***Data Understanding***: Menentukan jenis data yang diperlukan, sumber data, dan data yang harus diabaikan. Selama tahap ini, tim proyek mengumpulkan data yang diperlukan, mengeksplorasi karakteristiknya, dan memahami kualitasnya.
3. ***Data Preparation***: Data yang telah dikumpulkan diolah dan disiapkan untuk analisis. Hal ini mencakup pembersihan data, penggabungan data dari berbagai sumber, dan transformasi data.
4. ***Modeling***: Di tahap ini, model statistik atau model machine learning dikembangkan dan diuji. Berbagai teknik model dapat digunakan tergantung pada tujuan proyek.
5. ***Evaluation***: Mengevaluasi model yang dihasilkan pada tahap pemodelan untuk mengetahui kualitas dan efektivitasnya sebelum menerapkannya di lapangan serta menentukan apakah model tersebut benar-benar mencapai tujuan bisnis.
6. ***Deployment***: Menerapkan model yang dihasilkan untuk membantu dalam pengambilan keputusan.

C. Studi Kasus

STUDI KASUS KELULUSAN MAHASISWA DI UNIVERSITAS SUKA BELAJAR

1. *Business Understanding*

a. *Problem*

Budi adalah Rektor di Universitas Suka Belajar. Universitas Suka Belajar memiliki masalah besar karena rasio kelulusan mahasiswa tiap angkatan sangat rendah. Budi ingin memahami dan membuat pola dari profil mahasiswa yang bisa lulus tepat waktu dan yang tidak lulus tepat waktu.

Dengan pola tersebut, Budi bisa melakukan konseling, terapi, dan memberi peringatan dini kepada mahasiswa kemungkinan tidak lulus tepat waktu untuk memperbaiki diri, sehingga akhirnya bisa lulus tepat waktu.

b. *Objective*

Menemukan pola dari mahasiswa yang lulus tepat waktu dan tidak.

2. *Data Understanding*

Budi mengumpulkan data dari sistem informasi akademik di universitasnya yang memuat data profil mahasiswa dan indeks prestasi semester mahasiswa, dengan atribut seperti di bawah.

Atribut	Deskripsi
Nama	Nama Mahasiswa
Jenis Kelamin	Laki-laki atau Perempuan
Status Mahasiswa	Mahasiswa atau Bekerja
Umur	Umur mahasiswa
Status Nikah	Menikah atau Belum Menikah
IPS (Indeks Prestasi Semester) Semester 1-8	1,00 – 4,00

IPK (Indeks Prestasi Kumulatif)	1,00 – 4,00
---------------------------------	-------------

Status Kelulusan	Terlambat atau Tepat
------------------	----------------------

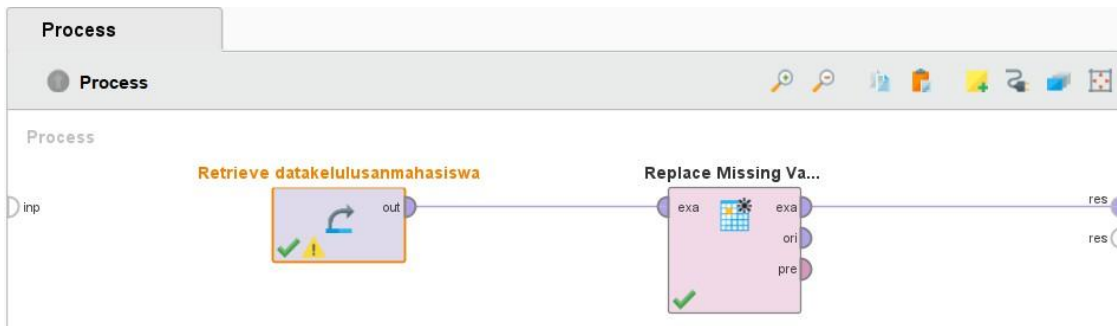
3. *Data Preparation*

Dataset yang digunakan adalah datakelulusanmahasiswa.xls. Pastikan bahwa data ini telah bersih dari *noise* dan dapat digunakan untuk analisis data.

Row No.	STATUS KE...	NAMA	JENIS KELA...	STATUS MA...	UMUR	STATUS NI...	IPS 1	IPS 2	IPS 3
1	TERLAMBAT	ANIK WIDAYA...	PEREMPUAN	BEKERJA	28	BELUM MENI...	2.760	2.800	3.200
2	TERLAMBAT	DWI HESTYN...	PEREMPUAN	MAHASISWA	32	BELUM MENI...	3	3.300	3.140
3	TERLAMBAT	MURYA ARIE...	PEREMPUAN	BEKERJA	29	BELUM MENI...	3.500	3.300	3.700
4	TERLAMBAT	NANIK SUSA...	PEREMPUAN	MAHASISWA	27	BELUM MENI...	3.170	3.410	3.610
5	TERLAMBAT	RIFKA ISTIQF...	PEREMPUAN	BEKERJA	29	BELUM MENI...	2.900	2.890	3.300
6	TERLAMBAT	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MENI...	2.950	2.820	3.090
7	TEPAT	FARIKHATUN ...	PEREMPUAN	MAHASISWA	26	BELUM MENI...	2.760	3.140	2.600
8	TEPAT	FIFI SUNALISA	PEREMPUAN	MAHASISWA	27	BELUM MENI...	2.620	2.890	2.320
9	TERLAMBAT	HENDRIK MU...	PEREMPUAN	BEKERJA	25	MENIKAH	3.600	3.540	3.520
10	TERLAMBAT	IMAM AGUNG ...	PEREMPUAN	BEKERJA	28	BELUM MENI...	2.710	2.550	1.770
11	TERLAMBAT	IMAM SANTO...	PEREMPUAN	BEKERJA	27	BELUM MENI...	3.140	3.460	3.400
12	TERLAMBAT	IRFAN EKO ...	PEREMPUAN	BEKERJA	32	BELUM MENI...	2.670	2.300	1.570
13	TERLAMBAT	IWAN HAMBALI	PEREMPUAN	BEKERJA	26	BELUM MENI...	2.570	2.820	2.200
14	TERLAMBAT	M SYAIFULLAH	PEREMPUAN	BEKERJA	31	BELUM MENI...	2.710	3	2.650
15	TERLAMBAT	DIANA LAILY F...	PEREMPUAN	MAHASISWA	26	BELUM MENI...	3.240	3.380	3.440
16	TERLAMBAT	DONNY PRA...	LAKI - LAKI	MAHASISWA	27	BELUM MENI...	2.860	2.860	2.450
17	TERLAMBAT	EDI JATMIKO	PEREMPUAN	MAHASISWA	27	BELUM MENI...	2.710	3.270	2.540

Terdapat 379 data mahasiswa dengan 15 atribut. Ternyata, terdapat *missing value* sejumlah 10 data di atribut IPK dan IPS 8. *Missing Value* dipecahkan dengan menambahkan data dengan nilai rata-rata. Hasilnya adalah data bersih tanpa *missing value*.

Name	Type	Missing	Statistics	Filter (15 / 15 attributes)	Search for Attributes
IPS 1	Real	0	0.330	3.790	2.854
IPS 2	Real	0	Min 0.500	Max 3.960	Average 2.812
IPS 3	Real	0	Min 0.160	Max 3.960	Average 2.901
IPS 4	Real	0	Min 0	Max 3.910	Average 2.782
IPS 5	Real	0	Min 0.200	Max 3.880	Average 2.782
IPS 6	Real	0	Min 0	Max 4	Average 2.877
IPS 7	Real	0	Min 0	Max 3.910	Average 2.531
IPS 8	Real	7	Min 0	Max 4	Average 1.272
IPK	Real	3	Min 0.870	Max 3.850	Average 2.933

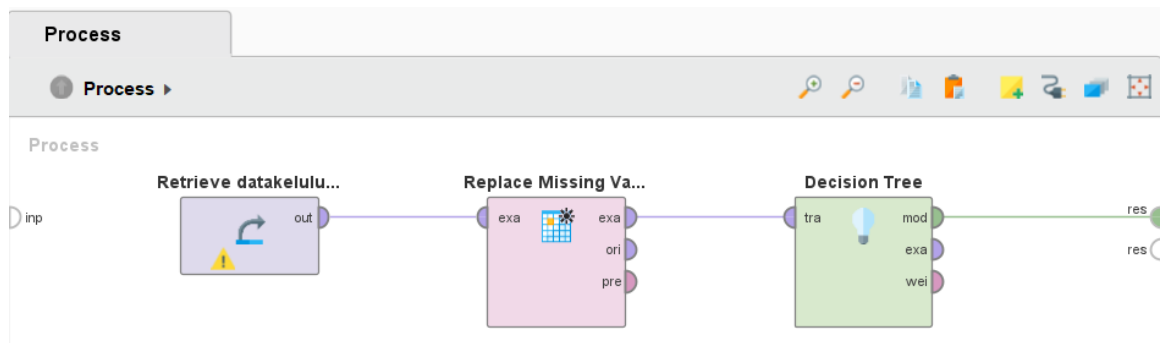


Operator *Replace Missing Value* dapat digunakan untuk membersihkan data dari *noise*.

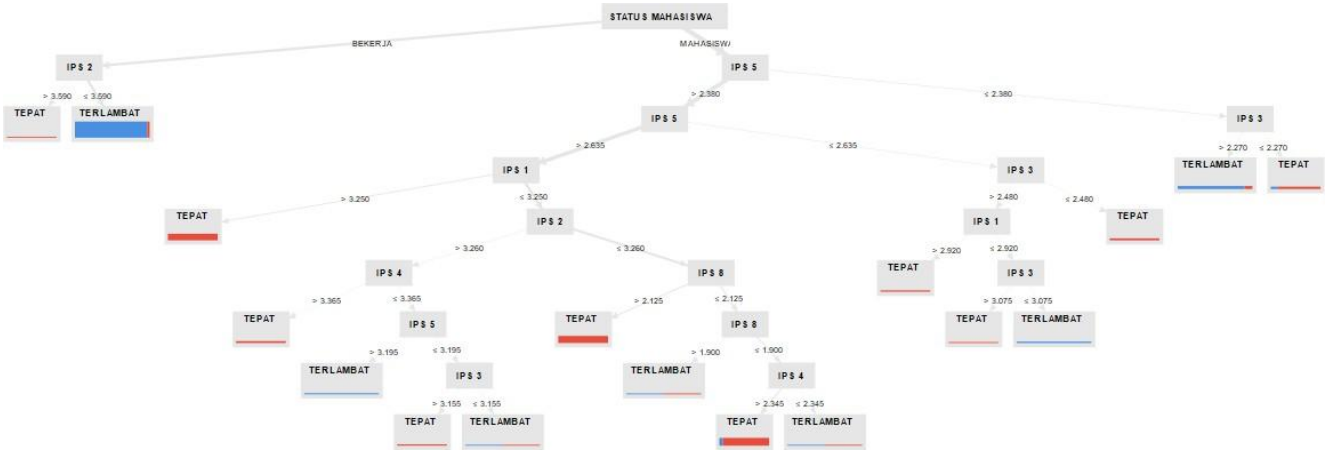
Name	Type	Missing	Statistics	Filter (15 / 15 attributes):	Search for Attributes
IPS 1	Real	0	0.330	3.790	2.854
IPS 2	Real	0	Min 0.500	Max 3.960	Average 2.812
IPS 3	Real	0	Min 0.160	Max 3.960	Average 2.901
IPS 4	Real	0	Min 0	Max 3.910	Average 2.782
IPS 5	Real	0	Min 0.200	Max 3.880	Average 2.782
IPS 6	Real	0	Min 0	Max 4	Average 2.877
IPS 7	Real	0	Min 0	Max 3.910	Average 2.531
IPS 8	Real	0	Min 0	Max 4	Average 1.272
IPK	Real	0	Min 0.870	Max 3.850	Average 2.933

4. *Modeling*

Modelkan dataset dengan *Decision Tree*. Pola yang dihasilkan bisa berbentuk *tree* atau *if-then*.



Hasil pola dari data berupa *decision tree* (pohon keputusan) berikut.



5. Evaluation

Hasil pola dari data juga bisa berupa peraturan *if-then*.

```

Tree (Decision Tree) ExampleSet (//Local Repository/data/datakelulusanmahasiswa)

Tree
STATUS MAHASISWA = BEKERJA
| IPS 2 > 3.590: TEPAT {TERLAMBAT=0, TEPAT=2}
| IPS 2 ≤ 3.590: TERLAMBAT {TERLAMBAT=127, TEPAT=4}
STATUS MAHASISWA = MAHASISWA
| IPS 5 > 2.380
| | IPS 5 > 2.635
| | | IPS 1 > 3.250: TEPAT {TERLAMBAT=0, TEPAT=50}
| | | IPS 1 ≤ 3.250
| | | | IPS 2 > 3.260
| | | | | IPS 4 > 3.365: TEPAT {TERLAMBAT=0, TEPAT=10}
| | | | | IPS 4 ≤ 3.365
| | | | | | IPS 5 > 3.195: TERLAMBAT {TERLAMBAT=4, TEPAT=0}
| | | | | | IPS 5 ≤ 3.195
| | | | | | | IPS 3 > 3.155: TEPAT {TERLAMBAT=0, TEPAT=5}
| | | | | | | IPS 3 ≤ 3.155: TERLAMBAT {TERLAMBAT=1, TEPAT=1}
| | | | | IPS 2 ≤ 3.260
| | | | | | IPS 8 > 2.125: TEPAT {TERLAMBAT=0, TEPAT=54}
| | | | | | IPS 8 ≤ 2.125
| | | | | | | IPS 8 > 1.900: TERLAMBAT {TERLAMBAT=1, TEPAT=1}
| | | | | | | IPS 8 ≤ 1.900
| | | | | | | | IPS 4 > 2.345: TEPAT {TERLAMBAT=4, TEPAT=56}
| | | | | | | | IPS 4 ≤ 2.345: TERLAMBAT {TERLAMBAT=1, TEPAT=1}
| | | | IPS 5 ≤ 2.635
| | | | | IPS 3 > 2.480
| | | | | | IPS 1 > 2.920: TEPAT {TERLAMBAT=0, TEPAT=5}
| | | | | | IPS 1 ≤ 2.920
| | | | | | | IPS 3 > 3.075: TEPAT {TERLAMBAT=0, TEPAT=2}
| | | | | | | IPS 3 ≤ 3.075: TERLAMBAT {TERLAMBAT=6, TEPAT=0}
| | | | | IPS 3 ≤ 2.480: TEPAT {TERLAMBAT=0, TEPAT=11}
| | IPS 5 ≤ 2.380
| | | IPS 3 > 2.270: TERLAMBAT {TERLAMBAT=17, TEPAT=2}
| | | IPS 3 ≤ 2.270: TEPAT {TERLAMBAT=2, TEPAT=12}
    
```

Atribut atau faktor yang paling berpengaruh adalah Status Mahasiswa, IPS 2, IPS 5, dan IPS 1. Sementara itu, atribut atau faktor yang tidak berpengaruh adalah Nama, Jenis Kelamin, Umur, IPS 6, IPS 7, dan IPS 8.

6. ***Deployment***

Setelah menemukan pola dari mahasiswa yang lulus tepat waktu dan tidak, Budi dapat memberlakukan beberapa kebijakan di bawah ini:

- a. Budi membuat program peningkatan disiplin dan pendampingan ke mahasiswa di semester awal (1-2) dan semester 5, karena faktor yang paling menentukan kelulusan mahasiswa ada di dua semester itu.
- b. Budi membuat peraturan melarang mahasiswa bekerja paruh waktu di semester awal perkuliahan, karena beresiko tinggi di kelulusan tepat waktu.
- c. Budi membuat program kerja paruh waktu di dalam kampus, sehingga banyak pekerjaan kampus yang bisa intens ditangani, sambil mendidik mahasiswa supaya memiliki pengalaman kerja. Dan yang paling penting mahasiswa tidak meninggalkan kuliah karena pekerjaan.
- d. Budi memasukkan pola dan model yang terbentuk ke dalam sistem informasi akademik, secara berkala di-*update* setiap semester. Sistem informasi akademik dibuat cerdas, sehingga bisa mengirimkan email analisis pola kelulusan secara otomatis ke mahasiswa sesuai profilnya.