

Praktikum 1 IF3170 Intelligensi Artifisial

Open University Learning Analytics

Dipersiapkan Oleh Tim Asisten Lab AI '22
Versi: 1.0 28/10/2025



Deadline 1: Rabu, 28 Oktober 2025 09.00 WIB
Deadline 2: Rabu, 28 Oktober 2025 21:00 WIB

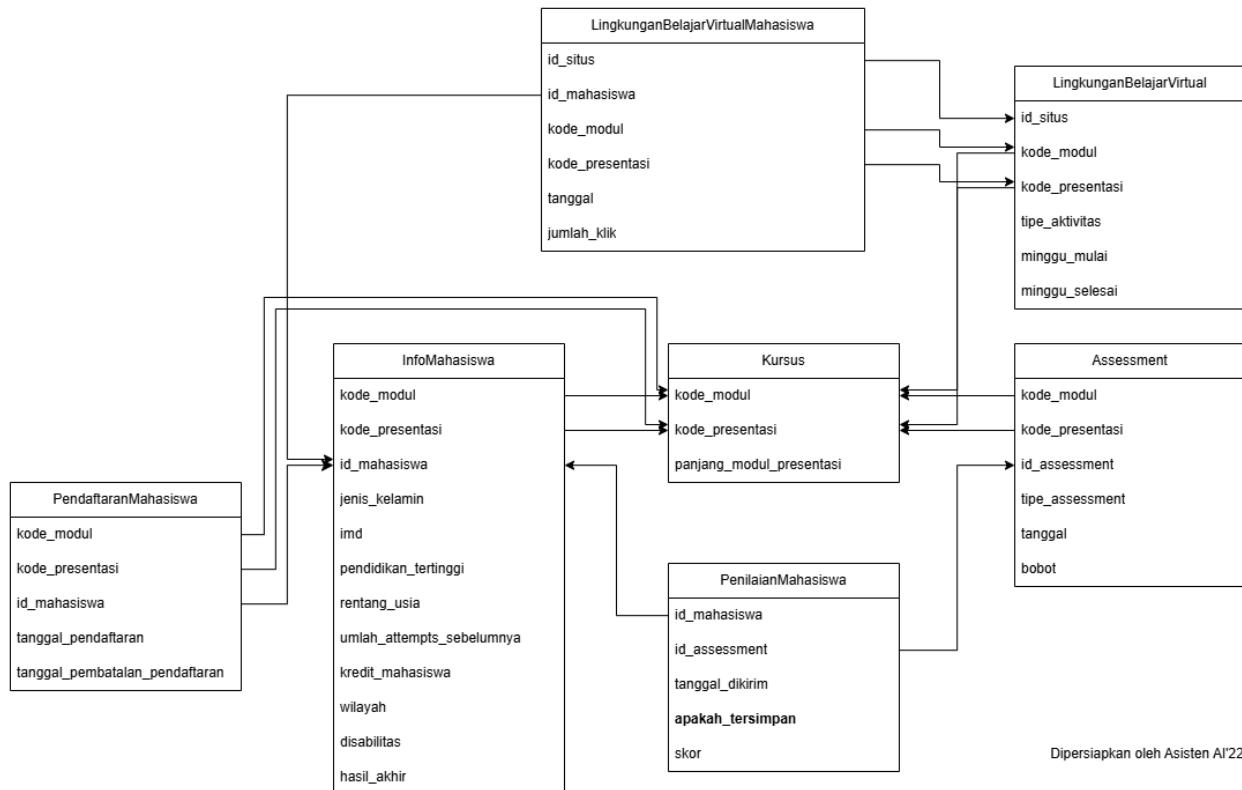
Tujuan

Praktikum 1 pada kuliah IF3170 Intelligensi Buatan bertujuan untuk memberikan pengalaman langsung kepada peserta kuliah dalam menerapkan fase-fase awal krusial dari framework CRISP-DM: Data Understanding dan Data Preparation. Praktikum ini berfokus pada proses investigasi, pembersihan, dan rekayasa data dari sumber multi-file yang kompleks untuk menghasilkan satu dataset final yang bersih, terstruktur, dan siap untuk dianalisis.

Spesifikasi

Sebuah universitas online ingin memahami perilaku belajar dan faktor-faktor yang mempengaruhi keberhasilan mahasiswa mereka. Untuk melakukan analisis di masa depan, mereka membutuhkan dataset terpadu yang menggabungkan semua data dari sistem terpisah.

Pada praktikum ini, Anda **tidak diminta untuk membuat atau melatih model machine learning**. Tugas Anda adalah bertindak sebagai Data Scientist yang bertanggung jawab untuk mengkonsolidasikan data mentah dari 7 file .csv terpisah. Tugas Anda adalah melakukan seluruh tahapan Data Preparation untuk menghasilkan satu master dataframe final.



Berikut daftar keterangan dari setiap fitur.

Nama Fitur	Tipe Data	Keterangan
kode_modul	VARCHAR(45)	Kode identifikasi modul.
kode_presentasi	VARCHAR(45)	Kode identifikasi presentasi. Terdiri dari tahun dan “B” untuk presentasi yang dimulai pada Februari dan “J” untuk presentasi yang dimulai pada Oktober.
id_assessment	INT	Nomor identifikasi penilaian.
tipe_assessment	VARCHAR(45)	Jenis penilaian. Terdapat tiga jenis penilaian: <ul style="list-style-type: none"> - Penilaian yang Diberi Nilai oleh Dosen (TMA). - Penilaian yang Diberi Nilai oleh Komputer (CMA).

		- Ujian Akhir (Exam).
tanggal	INT	Informasi tentang tanggal pengumpulan akhir penilaian, dihitung sebagai jumlah hari sejak awal presentasi modul. Tanggal awal presentasi memiliki nomor 0.
bobot	INT	Bobot penilaian dalam %.
id_situs	INT	Nomor identifikasi situs.
tipe_aktivitas	VARCHAR(45)	Aktivitas yang terkait dengan modul.
minggu_mulai	INT	Minggu dimana penggunaan modul dimulai.
minggu_selesai	INT	Minggu dimana penggunaan modul selesai.
id_mahasiswa	VARCHAR(45)	Nomor identifikasi unik untuk mahasiswa.
jenis_kelamin	VARCHAR(3)	Jenis kelamin mahasiswa.
wilayah	VARCHAR(45)	Wilayah geografis tempat mahasiswa tinggal saat mengikuti presentasi modul.
pendidikan_tertinggi	VARCHAR(45)	Tingkat pendidikan tertinggi mahasiswa saat masuk ke presentasi modul.
imd	VARCHAR(16)	Band Indeks Kemiskinan Multidimensi dari tempat tinggal mahasiswa selama presentasi modul.
rentang_usia	VARCHAR(16)	Rentang usia mahasiswa.
jumlah_attempts_sebelumnya	INT	Jumlah banyaknya percobaan yang dilakukan mahasiswa di modul tertentu.
kredit_mahasiswa	INT	Jumlah total kredit untuk modul yang sedang dipelajari oleh mahasiswa.
disabilitas	VARCHAR(3)	Apakah mahasiswa telah menyatakan memiliki disabilitas.
hasil_akhir	VARCHAR(45)	Hasil akhir mahasiswa dalam modul presentasi.
tanggal	INT	Tanggal akses siswa ke materi, diukur sebagai jumlah hari sejak dimulainya presentasi modul.

jumlah_klik	INT	Jumlah interaksi mahasiswa dengan materi pada hari tersebut.
tanggal_pendaftaran	INT	Tanggal pendaftaran mahasiswa pada presentasi modul, ini adalah jumlah hari yang dihitung terhadap awal presentasi modul (misalnya, nilai negatif -30 berarti mahasiswa mendaftar ke presentasi modul 30 hari sebelum dimulainya).
tanggal_pembatalan_pendaftaran	INT	tanggal pembatalan pendaftaran mahasiswa dari presentasi modul, jumlah hari yang dihitung relatif terhadap awal presentasi modul. Mahasiswa yang telah menyelesaikan kursus memiliki kolom ini kosong.
tanggal_dikirim	INT	Tanggal pengiriman tugas oleh mahasiswa, dihitung sebagai jumlah hari sejak dimulainya presentasi modul.
apakah_tersimpan	INT	Status yang menunjukkan bahwa hasil penilaian telah dipindahkan dari presentasi sebelumnya.
skor	FLOAT	Nilai siswa dalam penilaian ini. Rentang nilainya adalah 0 hingga 100. Nilai di bawah 40 dianggap sebagai Gagal. Nilai-nilai tersebut berada dalam rentang 0 hingga 100.

Dataset bisa diakses melalui link [berikut](#).

Berikut merupakan beberapa hal yang harus dilakukan oleh praktikan:

EDA (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) adalah langkah krusial dalam proses analisis data yang melibatkan pemeriksaan dan visualisasi dataset untuk mengungkap pola, tren, anomali, dan *insight*. Ini merupakan langkah awal sebelum menerapkan teknik statistik dan machine learning yang lebih lanjut. EDA membantu Anda memahami data secara mendalam, memungkinkan Anda membuat keputusan yang lebih terarah dan dapat merumuskan hipotesis untuk analisis lebih lanjut.

Pada tahap ini, Anda diminta untuk melakukan EDA dengan membuat visualisasi serta penjelasan terkait dataset yang dimiliki. Gunakan visualisasi yang tepat dilengkapi dengan penjelasan untuk setiap pertanyaan atau pernyataan yang ingin disampaikan.

Data Preprocessing

Langkah ini adalah hal pertama yang dilakukan setelah seorang Data Scientist memiliki pemahaman umum tentang data. Data mentah jarang siap untuk training, sehingga perlu dilakukan langkah-langkah untuk membersihkan dan memformat data agar dapat diinterpretasikan oleh model machine learning.

Dengan melakukan data preprocessing, Anda memastikan bahwa dataset siap untuk training model, yang akan menghasilkan hasil machine learning yang lebih akurat dan andal. Langkah-langkah ini sangat penting untuk mentransformasi data mentah menjadi format yang dapat dipelajari secara efektif oleh algoritma machine learning dan digunakan untuk membuat prediksi.

Pada tahap ini, Anda diminta untuk melakukan proses-proses yang sekiranya tepat digunakan untuk dataset yang diberikan, dengan penjelasan rinci untuk setiap poin:

1. Data Integration

Ini adalah proses menggabungkan data dari berbagai sumber terpisah. Anda diminta untuk merancang dan mengimplementasikan strategi untuk menggabungkan 7 file dataset OULAD. Anda harus mengidentifikasi primary key dan foreign key yang relevan (lihat diagram skema data) untuk melakukan join/merge pada dataframe Anda.

2. Data Cleaning

Ini adalah proses mendeteksi dan menangani data yang tidak akurat atau tidak lengkap. Anda diminta untuk melakukan pembersihan pada dataframe terintegrasi, termasuk menangani missing values (misalnya pada kolom score atau data demografi) dengan metode yang dapat dijustifikasi (imputasi, penghapusan), serta menangani anomali data.

3. Data Transformation

Ini adalah proses mengubah data dari format mentah menjadi format yang lebih sesuai untuk analisis, seringkali melalui *Feature Engineering*. Ini adalah bagian paling menantang. Anda wajib melakukan agregasi pada data transaksional (seperti studentVle.csv dan studentAssessment.csv) untuk membuat fitur-fitur baru (misal: total_clicks, average_score). Anda juga diminta melakukan *encoding* data kategorikal dan scaling fitur numerik.

4. Data Reduction/Compression

Ini adalah proses mengurangi kompleksitas (jumlah fitur) data. Setelah melakukan feature engineering yang ekstensif, Anda diminta menerapkan teknik reduksi data.

Terapkan PCA (Principal Component Analysis) pada fitur-fitur numerik dan lakukan Feature Selection (misal: analisis korelasi) untuk mengidentifikasi fitur-fitur yang paling informatif.

Asisten telah menyiapkan *notebook template* untuk Anda gunakan dengan tautan sebagai [berikut](#).

Untuk PPT penjelasan dapat diakses melalui link [berikut](#).

Kelompok

Pembagian kelompok ditentukan sendiri oleh mahasiswa dengan mengisi [sheets kelompok](#) berikut ini dengan 1 kelompok terdiri dari dengan **maksimal anggota sebanyak 2 orang**.

QnA

Pertanyaan dapat ditanyakan pada [link QnA](#) berikut. Pastikan pertanyaan yang ditanyakan tidak berulang.

Aturan

Terdapat beberapa hal yang harus diperhatikan dalam penggeraan tugas ini, yakni:

1. Jika terdapat hal yang tidak dimengerti, silahkan ajukan pertanyaan kepada asisten melalui [link QnA](#) yang telah diberikan di atas. Pertanyaan yang diajukan secara personal ke asisten **tidak akan dijawab** untuk menghindari perbedaan informasi yang didapatkan oleh peserta kuliah.
2. Dilarang melakukan **plagiarisme, menggunakan AI dalam bentuk apapun secara tidak bertanggungjawab, dan melakukan kerjasama antar kelompok**. Pelanggaran pada poin ini akan menyebabkan pemberian **nilai E** pada setiap anggota kelompok.
3. Tidak ada batasan untuk *library* yang boleh digunakan.

Deliverables

- **Deadline 1 (Rabu, 29 Oktober 2025 09:00 WIB)**
 - Link ke notebook penggeraan dengan format penamaan file NIM1_NIM2_Kelas_Deadline1.ipynb
(Contoh: 13523998_13523999_K1_Deadline1.ipynb)
- **Deadline 2 (Rabu, 29 Oktober 2025 21:00 WIB)**

- Link ke notebook penggeraan dengan format penamaan file NIM1_NIM2_Kelas_Deadline2.ipynb
(Contoh: 13523998_13523999_K1_Deadline2.ipynb). Jika terdapat revisi atau tambahan dari notebook pertama, silakan tambah markdown cell di notebook penggeraan yang berisi daftar perubahan yang dilakukan.
- PowerPoint yang **minimal** berisi penjelasan terkait beberapa hal berikut:
 - EDA
 - Data Cleaning
 - Data Transformation
 - Data Reduction/Compression
 - Data Integration
- PowerPoint dikumpulkan dalam bentuk file .pdf dengan format penamaan file NIM1_NIM2_Kelas_PPT.pdf (Contoh: 13523998_13523999_K1_PPT.pdf).
- Notebook penggeraan yang dikumpulkan harus memiliki akses **Editor** untuk **Anyone with the link**.
- Pengumpulan dilakukan melalui **edunex**.
- Jika terdapat masalah pada edunex, silakan gunakan tautan [berikut](#).
- Tugas yang terlambat dikumpulkan tidak akan diterima.
- Pengumpulan dilakukan oleh NIM terkecil.

Referensi

- <https://www.geeksforgeeks.org/machine-learning/data-integration-in-data-mining/>
- <https://www.geeksforgeeks.org/data-analysis/data-cleaning-in-data-mining/>
- <https://www.geeksforgeeks.org/machine-learning/data-transformation-in-machine-learning/>
- <https://www.geeksforgeeks.org/dbms/data-reduction-in-data-mining/>