



JADWAL BELAJAR HARIAN

Berikut adalah rencana studi intensif selama 15 hari untuk mempelajari dasar-dasar **data mining** dan mempersiapkan diri untuk lomba. Setiap hari akan difokuskan pada topik tertentu, dengan alokasi waktu rata-rata 4-6 jam per hari.

Hari 1-3: Pemahaman Dasar Data Mining

1. **Tujuan:** Memahami konsep dasar, proses, dan aplikasi data mining.
 - Pelajari konsep dasar:
 - Apa itu data mining?
 - Proses CRISP-DM (Cross Industry Standard Process for Data Mining).
 - Jenis-jenis data mining (klasifikasi, klustering, asosiasi, prediksi, dll).
 - Baca buku: "Data Mining: Concepts and Techniques" (Bab 1-2).
 - Tonton video pengantar di YouTube (misalnya dari channel StatQuest atau Kaggle Learn).
 - Diskusikan dalam tim untuk memastikan semua memahami dasar teori.
 2. **Output:**
 - Mind map atau catatan tentang proses dan jenis data mining.
-

Hari 4-6: Pemahaman Data dan Pengolahan Awal

1. **Tujuan:** Memahami teknik eksplorasi data dan pre-processing.
 - Pelajari dasar-dasar **Python** atau **R** untuk data analysis.
 - Eksplorasi pustaka seperti:
 - **Python:** Pandas, NumPy, Matplotlib, Seaborn.
 - **R:** dplyr, ggplot2.

- Pelajari teknik:
 - Pembersihan data (mengatasi missing values, outliers).
 - Transformasi data (normalisasi, encoding).
 - Visualisasi data untuk eksplorasi.

2. **Output:**

- Notebook (Jupyter atau RStudio) dengan eksplorasi data sederhana.
-

Hari 7-9: Teknik Data Mining

1. **Tujuan:** Memahami algoritma utama dalam data mining.

- **Klasifikasi:**
 - Algoritma: Decision Tree, Random Forest, K-Nearest Neighbors (KNN).
- **Klustering:**
 - Algoritma: K-Means, Hierarchical Clustering.
- **Asosiasi:**
 - Algoritma: Apriori, FP-Growth.
- Gunakan dataset dari Kaggle atau UCI Machine Learning Repository untuk eksperimen.

2. **Praktik:**

- Implementasi sederhana menggunakan **scikit-learn** di Python.
- Tinjau hasil dengan metrik evaluasi (accuracy, precision, recall, dll).

3. **Output:**

- Notebook yang berisi implementasi algoritma dan analisis hasil.
-

Hari 10-12: Studi Kasus & Penguatan

1. **Tujuan:** Menerapkan teknik pada studi kasus data mining.

- Pilih 1-2 studi kasus yang sesuai dengan tema lomba.
- Lakukan analisis end-to-end:
 - Eksplorasi data.
 - Pre-processing.

- Pemilihan dan penerapan algoritma.
- Evaluasi hasil.

2. **Output:**

- Studi kasus yang siap dipresentasikan.
-

Hari 13-14: Simulasi Lomba

1. **Tujuan:** Mempraktikkan kemampuan dalam kondisi lomba.

- Simulasikan penyelesaian masalah dalam waktu terbatas.
- Fokus pada pembagian tugas tim:
 - Pemrosesan data.
 - Pemodelan.
 - Dokumentasi/presentasi.
- Evaluasi hasil dan catat perbaikan yang diperlukan.

2. **Output:**

- Latihan soal yang selesai dalam batas waktu.
-

Hari 15: Penyempurnaan dan Persiapan Lomba

1. **Tujuan:** Menyempurnakan strategi dan mempersiapkan mental.

- Review semua materi dan studi kasus.
- Siapkan slide presentasi yang profesional.
- Latihan presentasi di depan rekan/mentor untuk mendapatkan masukan.

2. **Output:**

- Tim siap secara teknis dan mental untuk lomba.
-

Sumber Belajar Tambahan

1. **Platform Online:**

- [Kaggle Learn](#)

- [DataCamp](#)
- [Coursera](#)

2. **Buku:**

- "Data Science for Beginners" oleh Andrew Park.
- "Introduction to Data Mining" oleh Tan, Steinbach, dan Kumar.

3. **Tools:**

- Jupyter Notebook atau Google Colab.
 - Python (Anaconda distribution).
-

Kalau dipecah per-hari

Hari 1: Pengantar Data Mining

Waktu:

- **09:00-10:30:** Belajar konsep dasar data mining (Apa itu data mining, proses CRISP-DM).
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Baca materi dari buku "Data Mining: Concepts and Techniques" (Bab 1-2).
 - **14:00-15:30:** Diskusi tim dan buat mind map tentang data mining.
-

Hari 2: Jenis Data Mining dan Proses CRISP-DM

Waktu:

- **09:00-10:30:** Pelajari jenis-jenis data mining (klasifikasi, klustering, asosiasi, prediksi).
- **10:30-11:00:** Istirahat.
- **11:00-12:30:** Tonton video penjelasan tentang jenis data mining (StatQuest, YouTube).

- **14:00-15:30:** Diskusi tim untuk merumuskan aplikasi data mining di kehidupan nyata.
-

Hari 3: Eksplorasi Tools dan Dataset

Waktu:

- **09:00-10:30:** Instal dan eksplorasi tools (Python, Jupyter Notebook, Pandas, Matplotlib).
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Cari dan eksplorasi dataset sederhana (misalnya dari Kaggle).
 - **14:00-15:30:** Diskusi tim tentang temuan awal dari dataset.
-

Hari 4: Eksplorasi Data dan Preprocessing

Waktu:

- **09:00-10:30:** Pelajari eksplorasi data (menggunakan Pandas, visualisasi dengan Seaborn).
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Pelajari teknik pembersihan data (missing values, outliers).
 - **14:00-15:30:** Diskusi tim sambil mencoba preprocessing pada dataset yang ditemukan.
-

Hari 5: Transformasi dan Visualisasi Data

Waktu:

- **09:00-10:30:** Pelajari teknik transformasi data (normalisasi, encoding).
- **10:30-11:00:** Istirahat.
- **11:00-12:30:** Latihan membuat visualisasi data (Seaborn, Matplotlib).
- **14:00-15:30:** Diskusi tim untuk menganalisis hasil visualisasi.

Hari 6: Evaluasi Data yang Sudah Diproses

Waktu:

- **09:00-10:30**: Evaluasi hasil preprocessing dan eksplorasi.
 - **10:30-11:00**: Istirahat.
 - **11:00-12:30**: Siapkan data untuk digunakan dalam algoritma data mining.
-

Hari 7: Pengenalan Klasifikasi

Waktu:

- **09:00-10:30**: Pelajari algoritma Decision Tree dan Random Forest.
 - **10:30-11:00**: Istirahat.
 - **11:00-12:30**: Implementasi algoritma menggunakan **scikit-learn**.
 - **14:00-15:30**: Diskusi hasil evaluasi algoritma.
-

Hari 8: Klustering

Waktu:

- **09:00-10:30**: Pelajari algoritma K-Means dan Hierarchical Clustering.
 - **10:30-11:00**: Istirahat.
 - **11:00-12:30**: Implementasi algoritma klustering pada dataset.
 - **14:00-15:30**: Diskusi hasil evaluasi.
-

Hari 9: Asosiasi dan Penerapan Lain

Waktu:

- **09:00-10:30**: Pelajari algoritma Apriori dan FP-Growth.

- **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Implementasi asosiasi menggunakan Python.
 - **14:00-15:30:** Analisis hasil asosiasi dengan tim.
-

Hari 10: Studi Kasus 1

Waktu:

- **09:00-10:30:** Pilih studi kasus pertama.
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Lakukan eksplorasi dan preprocessing data.
 - **14:00-15:30:** Terapkan algoritma dan evaluasi hasil.
-

Hari 11: Studi Kasus 2

Waktu:

- **09:00-10:30:** Pilih studi kasus kedua.
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Eksplorasi dan preprocessing data.
 - **14:00-15:30:** Terapkan algoritma dan evaluasi hasil.
-

Hari 12: Penyempurnaan Studi Kasus

Waktu:

- **09:00-10:30:** Review hasil dari kedua studi kasus.
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Diskusi tim untuk menyusun laporan dan presentasi awal.
-

Hari 13: Simulasi Lomba 1

Waktu:

- **09:00-11:30:** Simulasi penyelesaian soal dalam waktu terbatas.
 - **11:30-12:00:** Istirahat.
 - **12:00-13:30:** Evaluasi hasil simulasi dan diskusikan perbaikan.
-

Hari 14: Simulasi Lomba 2

Waktu:

- **09:00-11:30:** Simulasi soal lain dengan pembagian tugas yang lebih efektif.
 - **11:30-12:00:** Istirahat.
 - **12:00-13:30:** Diskusi hasil dan penyempurnaan.
-

Hari 15: Finalisasi dan Latihan Presentasi

Waktu:

- **09:00-10:30:** Finalisasi laporan dan slide presentasi.
 - **10:30-11:00:** Istirahat.
 - **11:00-12:30:** Latihan presentasi dengan tim.
 - **14:00-15:30:** Review akhir dan persiapan mental untuk lomba.
-

catatan materi per hari berdasarkan jadwal yang telah disusun:

Hari 1: Pengantar Data Mining

Materi:

- **Data Mining** adalah proses menemukan pola atau informasi yang berguna dari kumpulan data besar.
 - **CRISP-DM (Cross Industry Standard Process for Data Mining):**
 - **Business Understanding:** Menentukan tujuan bisnis.
 - **Data Understanding:** Pengumpulan dan eksplorasi data.
 - **Data Preparation:** Pembersihan dan transformasi data.
 - **Modeling:** Menerapkan algoritma data mining.
 - **Evaluation:** Menilai model.
 - **Deployment:** Penerapan model pada skala penuh.
 - **Jenis-jenis Data Mining:**
 - **Klasifikasi:** Menentukan kategori data berdasarkan label yang ada (misal: spam vs tidak spam).
 - **Klusterisasi:** Mengelompokkan data berdasarkan kemiripan (misal: segmentasi pelanggan).
 - **Asosiasi:** Menemukan pola hubungan antar item dalam dataset (misal: analisis keranjang belanja).
 - **Prediksi:** Meramalkan nilai berdasarkan data yang ada.
-

Hari 2: Jenis Data Mining dan Proses CRISP-DM

Materi:

- **Klasifikasi:**
 - Tujuannya adalah untuk memprediksi kategori data.
 - **Algoritma populer:** Decision Tree, KNN, Random Forest.
 - **Klusterisasi:**
 - Tujuannya adalah untuk mengelompokkan data berdasarkan kesamaan.
 - **Algoritma populer:** K-Means, Hierarchical Clustering.
 - **Asosiasi:**
 - Menemukan hubungan antar item (misal: "Jika seseorang membeli A, maka dia cenderung membeli B").
 - **Algoritma populer:** Apriori, FP-Growth.
-

Hari 3: Eksplorasi Tools dan Dataset

Materi:

- **Python:**
 - **Pandas:** Digunakan untuk manipulasi data dan analisis data tabular.
 - **NumPy:** Digunakan untuk komputasi numerik.
 - **Matplotlib/Seaborn:** Digunakan untuk visualisasi data.
 - **Google Colab** atau **Jupyter Notebook:**
 - Platform untuk menulis dan menjalankan kode Python secara interaktif.
 - **Dataset:**
 - Cari dataset dari sumber seperti **Kaggle** atau **UCI Machine Learning Repository**.
 - Pastikan untuk memilih dataset yang sesuai dengan tujuan lomba.
-

Hari 4: Eksplorasi Data dan Preprocessing

Materi:

- **Exploratory Data Analysis (EDA):**
 - Analisis statistik dasar (mean, median, mode).
 - Visualisasi distribusi data menggunakan histogram, boxplot, dan scatter plot.
 - **Preprocessing:**
 - Mengatasi **missing values**: Mengisi nilai yang hilang dengan rata-rata, median, atau teknik lain.
 - Mengatasi **outliers**: Identifikasi dan penanganan data yang menyimpang.
 - **Normalisasi** dan **Standardisasi**: Proses untuk menyamakan skala fitur.
-

Hari 5: Transformasi dan Visualisasi Data

Materi:

- **Transformasi Data:**
 - **One-Hot Encoding:** Teknik untuk mengubah data kategori menjadi bentuk numerik.
 - **Min-Max Scaling** dan **Standardization:** Metode untuk menskalakan data agar sesuai dengan rentang tertentu.
 - **Visualisasi Data:**
 - **Matplotlib:** Visualisasi dasar seperti grafik garis dan pie chart.
 - **Seaborn:** Visualisasi data yang lebih kompleks, termasuk heatmaps, pairplots, dll.
-

Hari 6: Evaluasi Data yang Sudah Diproses

Materi:

- **Evaluasi Preprocessing:**
 - Periksa hasil preprocessing dengan memeriksa distribusi data dan visualisasi.
 - Pastikan tidak ada masalah besar dengan data yang telah dibersihkan dan ditransformasi.
-

Hari 7: Pengenalan Klasifikasi

Materi:

- **Klasifikasi:** Proses untuk memprediksi label kategori berdasarkan fitur.
 - **Decision Tree:** Model yang membagi data berdasarkan fitur-fitur terbaik untuk klasifikasi.
 - **Random Forest:** Ensemble method yang menggunakan banyak decision trees untuk meningkatkan akurasi.
 - **K-Nearest Neighbors (KNN):** Algoritma yang mengklasifikasikan data

berdasarkan kedekatannya dengan data lain.

Hari 8: Klustering

Materi:

- **K-Means**: Algoritma klustering yang membagi data ke dalam jumlah kluster yang sudah ditentukan.
 - Langkah-langkah: Inisialisasi kluster, iterasi berdasarkan jarak, update pusat kluster.
 - **Hierarchical Clustering**: Membangun dendrogram yang menunjukkan pengelompokan hierarkis data.
-

Hari 9: Asosiasi dan Penerapan Lain

Materi:

- **Apriori Algorithm**: Digunakan untuk menemukan aturan asosiasi dalam dataset besar, terutama dalam analisis keranjang belanja.
 - Menentukan itemset yang sering muncul dan aturan asosiasi.
 - **FP-Growth**: Alternatif yang lebih efisien daripada Apriori untuk menemukan pola asosiasi.
-

Hari 10: Studi Kasus 1

Materi:

- Terapkan **EDA**, **preprocessing**, dan **algoritma klasifikasi** atau **klustering** pada dataset pilihan.
 - Evaluasi model menggunakan **confusion matrix**, **accuracy**, **precision**, **recall**.
-

Hari 11: Studi Kasus 2

Materi:

- Pilih dataset lain dan terapkan teknik yang sama dari hari sebelumnya.
 - Bandingkan hasil dari dua studi kasus untuk mencari pola atau kesalahan yang terjadi.
-

Hari 12: Penyempurnaan Studi Kasus

Materi:

- **Fine-tuning Model:** Mengubah parameter model untuk meningkatkan performa (misalnya menggunakan Grid Search untuk Decision Tree atau Random Forest).
 - **Cross-Validation:** Teknik untuk mengevaluasi model menggunakan pembagian data menjadi beberapa subset.
-

Hari 13: Simulasi Lomba 1

Materi:

- Terapkan semua teknik yang telah dipelajari dalam skenario lomba simulasi.
 - Fokus pada alokasi tugas dan manajemen waktu yang efektif selama simulasi.
-

Hari 14: Simulasi Lomba 2

Materi:

- Ulangi simulasi lomba dengan perubahan dataset atau masalah baru.
 - Latihan membagi tugas antar anggota tim dan evaluasi kinerja.
-

Hari 15: Finalisasi dan Latihan Presentasi

Materi:

- Penyusunan **laporan** dan **slide presentasi**:
 - Ikuti struktur yang jelas: Pendahuluan, Data yang Digunakan, Proses, Hasil, Kesimpulan.
 - Latihan **presentasi**: Fokus pada cara menyampaikan hasil dengan jelas dan efektif.
-