

LEMBAR KERJA MAHASISWA (LKM)

LK.8 Perancangan Project Data Science

Nama	: Amanda Febriyanti
Tanggal	: minggu/ 07 Desember 2025
Kelas	: 5A
Judul Project	: Clustering Kecamatan Berdasarkan Komposisi Usia Siswa SD di Kabupaten Enrekan

A. Instruksi

Peserta diminta untuk merancang sebuah proyek Data Science yang berfokus pada permasalahan di bidang pendidikan. Rancangan proyek ini harus disusun secara sistematis berdasarkan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) yang mencakup enam tahapan utama, yaitu:

1. Business Understanding (Pemahaman Bisnis)
2. Data Understanding (Pemahaman Data)
3. Data Preparation (Persiapan Data)
4. Modeling (Pemodelan)
5. Evaluation (Evaluasi)
6. Deployment (Penerapan)

Pada setiap tahapan, peserta diharapkan dapat:

1. Menjelaskan tujuan dan fokus kegiatan pada tahap tersebut.
2. Menguraikan langkah-langkah yang dilakukan serta teknik atau metode yang digunakan.
3. Menjelaskan jenis dan sumber data yang diperlukan.
4. Menunjukkan hasil atau keluaran yang diharapkan dari tiap tahap.

Gunakan contoh kasus nyata atau permasalahan aktual di dunia pendidikan, seperti: Prediksi prestasi belajar siswa, Analisis tingkat kehadiran, Deteksi dini siswa berisiko tidak lulus, atau Rekomendasi pembelajaran adaptif berbasis data.

Hasil akhir dari tugas ini berupa dokumen rancangan proyek Data Science lengkap yang menggambarkan alur proses dari awal hingga implementasi model, serta menunjukkan bagaimana solusi berbasis data dapat memberikan manfaat nyata bagi peningkatan mutu pendidikan.

B. Format Perancangan

Tahapan CRISP-DM	Instruksi untuk Peserta	Rancangan Implementasi
1. Business Understanding (Pemahaman Bisnis)	<div><div>1. Pilih konteks pendidikan (contoh: sekolah, universitas, pelatihan).</div><div>2. Identifikasi permasalahan yang dapat diselesaikan dengan data science.</div><div>3. Rumuskan tujuan bisnis (contoh: meningkatkan prestasi siswa, menurunkan tingkat ketidakhadiran).</div></div>	Konteks proyek berada pada pendidikan tingkat sekolah dasar di Kabupaten Enrekang. Permasalahan utama adalah variasi komposisi usia siswa per kecamatan, di mana terdapat siswa terlalu muda (<7 tahun) dan terlalu tua (>12 tahun) untuk jenjang SD. Tujuan bisnis dari proyek ini adalah mengelompokkan kecamatan berdasarkan profil usia siswa untuk membantu pemerintah mengidentifikasi daerah yang memerlukan penyesuaian kebijakan, pemerataan siswa, dan evaluasi usia masuk sekolah.
2. Data Understanding (Pemahaman Data)	<div><div>1. Jelaskan sumber data (contoh: data nilai siswa, absensi, data keluarga).</div><div>2. Sebutkan jenis data (numerik, kategorikal, teks, waktu).</div></div>	Sumber data berasal dari dataset “Jumlah Siswa Menurut Kelompok Umur SD Kabupaten Enrekang Tahun 2024/2025”. Data berjenis numerik (jumlah dan persentase usia siswa) serta kategorikal (nama kecamatan, status sekolah). Fitur yang digunakan meliputi: jumlah siswa <7 tahun, 7–12 tahun, >12 tahun, persentasenya masing-masing, serta total siswa per kecamatan. Karena metode yang digunakan adalah clustering (unsupervised learning), tidak ada variabel target.

	3. Deskripsikan fitur dan target yang akan digunakan.	
3. Data Preparation (Persiapan Data)	<ol style="list-style-type: none"> 1. Tuliskan langkah pembersihan data: hapus duplikat, tangani nilai kosong, dan outlier. 2. Transformasi data: normalisasi, encoding data kategorikal. 	Dilakukan pembersihan data dengan menghapus baris header tambahan, menangani nilai kosong, menghilangkan data duplikat, dan memeriksa outlier pada kolom jumlah siswa. Data numerik lalu dinormalisasi menggunakan Min-Max Scaling agar skala antar fitur seimbang. Kolom kategorikal yang tidak diperlukan dalam pemodelan diabaikan. Setelah transformasi, dataset siap diproses dengan algoritma clustering.
4. Modeling (Pemodelan)	<ol style="list-style-type: none"> 1. Pilih algoritma yang sesuai (contoh: Decision Tree, Random Forest, Logistic Regression). 2. Jelaskan alasan pemilihan algoritma. 	Algoritma yang dipilih adalah K-Means Clustering , karena sesuai untuk data numerik dan dapat mengelompokkan kecamatan berdasarkan pola usia siswa. Jumlah cluster ditentukan melalui Elbow Method atau Silhouette Score. Model menghasilkan kelompok seperti: kecamatan dengan dominasi usia ideal, kecamatan dengan banyak siswa terlalu muda, atau kecamatan dengan banyak siswa terlalu tua.
5. Evaluation (Evaluasi)	Pilih metode evaluasi yang akan digunakan misalkan menggunakan cross-validation atau confusion matrix.	Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur kualitas cluster. Selain itu, digunakan visualisasi seperti grafik perbandingan kelompok

		usia dan diagram radar untuk melihat karakteristik tiap cluster. Evaluasi memastikan bahwa cluster yang terbentuk memiliki perbedaan yang jelas dan bermakna.
6. Deployment (Penerapan / Implementasi)	Buat rancangan deploymentnya / tampilan interface nya	<p>Hasil clustering dipresentasikan dalam dashboard. Interface menampilkan:</p> <ul style="list-style-type: none"> • Menu pilihan Kecamatan • Grafik distribusi usia siswa • Label cluster tiap kecamatan • Insight atau rekomendasi kebijakan <p>Dashboard dapat dibuat dengan Streamlit atau ditampilkan sebagai prototype UI yang memperlihatkan tampilan halaman analisis cluster dan peta sebaran kecamatan.</p>