

LEMBAR KERJA MAHASISWA (LKM)

LK.8 Perancangan Project Data Science

Nama	: NESSA DENANTA SARI
Tanggal	: 06 -12 - 2025
Kelas	: 5AI - A
Judul Project : Analisis Gender dan Prediksi Risiko Putus Sekolah Siswa SMA/MA Kota Makassar dengan Algoritma Machine Learning Menggunakan Data 2021–2024	

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)	<ol style="list-style-type: none">1. Pilih konteks pendidikan (contoh: sekolah, universitas, pelatihan).2. Identifikasi permasalahan yang dapat diselesaikan dengan data science.3. Rumuskan tujuan bisnis (contoh: meningkatkan prestasi siswa, menurunkan	<ol style="list-style-type: none">1. Konteks: Pendidikan Menengah Atas/Madrasah Aliyah (SMA/MA) di tingkat Kota Makassar.2. Permasalahan: Tingginya angka putus sekolah yang tidak terdeteksi secara dini, mengakibatkan hilangnya potensi sumber daya manusia

		<p>dan pemborosan anggaran pendidikan</p> <p>3. Tujuan Bisnis (Tujuan Proyek):</p> <ol style="list-style-type: none"> Membangun model Klasifikasi untuk memprediksi siswa yang memiliki Risiko Tinggi untuk putus sekolah. Menyediakan Sistem Peringatan Dini (Early Warning System) untuk Dinas Pendidikan dan pihak sekolah. Mengidentifikasi faktor gender dan wilayah yang paling berkontribusi terhadap risiko putus sekolah.
--	--	---

2. Data Understanding (Pemahaman Data)	<ol style="list-style-type: none"> Jelaskan sumber data (contoh: data nilai siswa, absensi, data keluarga). Sebutkan jenis data (numerik, kategorikal, teks, waktu). Deskripsikan fitur dan target yang akan digunakan. 	<ol style="list-style-type: none"> Sumber: Data Statistik Siswa Putus Sekolah per Kecamatan. Fitur: Jumlah Putus Sekolah (L/P) per Kelas, Kecamatan, Tahun. Target: Risiko Putus Sekolah (Biner: Tinggi / Rendah).
3. Data Preparation (Persiapan Data)	<ol style="list-style-type: none"> Tuliskan langkah pembersihan data: hapus duplikat, tangani nilai kosong, dan outlier. Transformasi data: normalisasi, encoding data kategorikal. 	<ol style="list-style-type: none"> Pembersihan: Imputasi <i>missing values</i> dengan nol (0), hapus duplikat. Transformasi: 1. <i>Feature Engineering</i> (menciptakan variabel Target Biner). 2. One-Hot Encoding untuk fitur Kecamatan. 3. Standard Scaling pada fitur numerik.
4. Modeling (Pemodelan)	<ol style="list-style-type: none"> Pilih algoritma yang sesuai (contoh: Decision Tree, Random Forest, Logistic Regression). Jelaskan alasan pemilihan algoritma. 	<ol style="list-style-type: none"> Algoritma: Random Forest Classifier. Alasan: Efektif untuk klasifikasi, tahan <i>overfitting</i>, dan mampu menyediakan Feature Importance untuk interpretasi faktor risiko.

5. Evaluation (Evaluasi)	Pilih metode evaluasi yang akan digunakan misalkan menggunakan cross-validation atau confusion matrix.	<ol style="list-style-type: none"> 1. Menggunakan Confusion Matrix dan Cross-Validation. 2. Metrik Prioritas: Recall pada kelas Risiko Tinggi (untuk meminimalisir <i>False Negatives</i>), serta Accuracy dan F1-Score.
6. Deployment (Penerapan / Implementasi)	Buat rancangan deploymentnya tampilan interface nya	Platform: Gradio. Interface: Input data Kecamatan dan statistik putus sekolah, Output berupa prediksi status "RISIKO TINGGI" atau "RISIKO RENDAH" dengan persentase probabilitas.