

KERANGKA ACUAN KERJA (KAK)
TUGAS BESAR MATA KULIAH KEAMANAN DATA

- Mata Kuliah : Keamanan Data
- Program Studi : S1 Data Science
- Fakultas : Informatika, Universitas Telkom
- Semester : Genap Tahun Akademik 2025/2026
- Dosen Pengampu : Ir Sidik Prabowo, S.T., M.T
- Versi : 1.0

Dokumen ini disusun sebagai panduan resmi untuk pelaksanaan Tugas Besar (*Project*) mata kuliah Keamanan Data. Tugas ini dirancang untuk mengintegrasikan konsep *Data Lifecycle Management* (DLM), privasi data sesuai UU PDP No. 27/2022, dan *best practices Data Science*. Project dikerjakan secara berkelompok (3-4 anggota, dengan komposisi kelompok yang telah ditentukan sebelumnya). Setiap kelompok bertanggung jawab penuh atas kelengkapan tugas, dengan koordinasi melalui LMS.

1. Latar Belakang

Dalam era Data Science, tantangan utama adalah menjaga keamanan dan privasi data selama siklus hidupnya (DLM), terutama pada tahap Generation dan Processing. *Synthetic data* (data buatan) dan *augmented data* (modifikasi data asli) menjadi solusi inovatif untuk melatih model *Machine Learning* (ML) tanpa risiko pelanggaran privasi, seperti *re-identification attack*. Project ini relevan dengan materi mata kuliah Keamanan Data yang membahas *encryption*, *anonymization*, dan *differential privacy*. Mahasiswa diharapkan menerapkan konsep ini pada kasus nyata, sambil mematuhi regulasi UU PDP (Pasal 16 tentang *pseudonimisasi*).

2. Tujuan

Mahasiswa mampu merancang dan mengimplementasikan pipeline generasi *synthetic/augmented data* untuk mendukung analisis Data Science yang menjamin pengamanan privasi.

- Menganalisis risiko privasi pada dataset asli menggunakan metrik evaluasi yang ada (seperti *K-Anonymity* dan *L-Diversity*).
- Mengembangkan pipeline generasi data *synthetic/augmented* yang mempertahankan utilitas data (mendekati distribusi asli) dengan tetap memenuhi standar privasi (e.g., $\epsilon < 1.0$ pada *differential privacy*).
- Mengevaluasi dampak data sintetis terhadap performa model ML dan mendiskusikan implikasi etis sesuai UU PDP.

3. Ruang Lingkup dan Batasan

- Tema Utama: Generated *Synthetic/Augmented* Data untuk Keamanan Privasi dalam Analisis Data Science.
 - o **Opsi 1:** *Synthetic Data* (e.g., menggunakan CTGAN atau SDV library untuk membuat dataset baru dari nol).
 - o **Opsi 2:** *Augmented Data* (e.g., modifikasi data asli via SMOTE atau *noise addition* dengan enkripsi).
- **Kasus Studi:** Dataset publik seperti "Telco Customer Churn" dari Kaggle (minimal 1.500 sampel, atribut sensitif: ID pelanggan, email). Hindari data real tanpa izin.
- **Teknologi:** Python (Pandas, Scikit-learn, TensorFlow/PyTorch untuk GANs). Tidak termasuk deployment cloud atau hardware khusus.
- **Batasan:** Fokus pada implementasi sederhana (tidak lebih dari 2.000 sampel output). Kelompok bertanggung jawab etis: data sintetis tidak boleh digunakan untuk tujuan komersial.

4. Metodologi

Menggunakan pendekatan iteratif: Analisis → Generasi → Evaluasi → Dokumentasi. (e.g; menggunakan Jupyter Notebook untuk prototipe).

1. Analisis Dataset Asli: Identifikasi sensitif data, hitung metrik privasi.
2. Generasi Data: Implementasikan pipeline (e.g., GANs untuk *synthetic*; *perturbation* untuk *augmented*).
3. Evaluasi: Train model ML (e.g., *Logistic Regression*), bandingkan metrik utility (F1-score) vs. *privacy*.
4. Diskusi: Bahas etika dan keterkaitan dengan DLM (tahap Storage & Processing).

5. Pembagian Tugas Kelompok

Kelompok terdiri dari 3-4 anggota (sudah ditentukan sebelumnya). Setiap anggota bertanggung jawab atas peran utama, dengan kolaborasi via GitHub atau Google Drive. Koordinator kelompok melaporkan progres mingguan ke dosen (langsung/via lms).

Tabel 1 Contoh pembagian peran tim Project

No.	Peran	Tanggung Jawab Utama	Anggota Kelompok (Contoh)
1	Koordinator	Pengelolaan jadwal, integrasi output, laporan progres	[Nama Anggota 1]
2	Analisis Data	Analisis risiko privasi & visualisasi dataset	[Nama Anggota 2]
3	Developer	Implementasi pipeline generasi data (coding utama)	[Nama Anggota 3]
4	Evaluator (opsional, jika 4 anggota)	Evaluasi model ML & diskusi etika	[Nama Anggota 4]

Catatan: Sesuaikan nama anggota dengan daftar kelompok resmi. Setiap anggota wajib berkontribusi minimal 25% (dievaluasi via commit log atau self-assessment).

6. Jadwal Pelaksanaan

Total durasi: 8 minggu, mulai Minggu 1 (10 November 2025). Presentasi di akhir Minggu 8. Progres dicek via milestone di LMS.

Minggu	Periode Tanggal	Kegiatan Utama	Output Milestone	Tanggung Jawab
1	10-16 Nov 2025	Formasi kelompok final, pemilihan tema & dataset, analisis awal dataset asli	Proposal singkat (1-2 halaman: tema, dataset, rencana)	Seluruh kelompok; submit ke LMS
2	17-23 Nov 2025	Analisis risiko privasi (hitung metrik K-Anonymity, dll.)	Laporan Analisis (Jupyter Notebook bagian 1)	Analisis Data
3	24-30 Nov 2025	Desain pipeline generasi data (pilih teknik: synthetic/augmented)	Desain Dokumen (diagram flowchart)	Developer & Koordinator
4	1-7 Des 2025	Implementasi generasi data (coding pipeline, generate sampel)	Dataset Sintetis/Augmented (CSV, minimal 1.500 baris)	Developer
5	8-14 Des 2025	Evaluasi awal: Train model ML pada data asli vs. sintetis	Hasil Evaluasi Sementara (tabel metrik, plot)	Evaluasi & Analisis Data

6	15-21 Des 2025	Refinement pipeline berdasarkan evaluasi, tambah diskusi etika (UU PDP)	Draft Laporan Lengkap (Jupyter Notebook full)	Seluruh kelompok
7	22-28 Des 2025	Finalisasi laporan, persiapan presentasi (demo coding)	Laporan Final (PDF 10-15 halaman) & Slide Presentasi	Koordinator & Seluruh
8	29 Des 2025 - 4 Jan 2026	Presentasi & Q&A (via Zoom/LMS, 10 menit/kelompok)	Demo Live & Feedback	Seluruh kelompok

Catatan: Keterlambatan >1 minggu kurangi nilai 10%.

7. Output

- Artefak Digital:
 - o Jupyter Notebook lengkap (kode + narasi, <50MB).
 - o Dataset output (CSV/JSON).
 - o Laporan PDF (struktur: Pendahuluan, Metode, Hasil, Kesimpulan, Referensi).
- Presentasi: Slide PowerPoint/Google Slides (10-15 slide) + demo 5 menit.
- Pengumpulan: Upload ke LMS sebelum deadline Minggu 8 (4 Jan 2026, pukul 23:59 WIB).

8. Kriteria Penilaian

Penilaian holistik oleh dosen (80%) + peer-review (20%). Total 100 poin.

Kriteria	Bobot (Poin)	Indikator Keberhasilan
Analisis Risiko Privasi	20	Akurasi metrik & identifikasi sensitif data (skor >80%)
Implementasi Pipeline	30	Kode bersih, pipeline berjalan (generate data valid), integrasi privasi
Evaluasi & Diskusi Etika	25	Perbandingan metrik utility-privacy solid; bahas pasal UU PDP relevan
Laporan & Presentasi	15	Kejelasan narasi, visualisasi bagus; demo lancar
Kolaborasi Kelompok	10	Kontribusi merata (via log), inovasi (e.g., tambah enkripsi)

9. Sumber Daya dan Dukungan

- **Tools** : Lab Komputer Telkom (Python 3.10+), library: Pandas, Scikit-learn, SDV/CTGAN (instal via Anaconda).
- **Referensi** :

- Buku: "*Privacy-Preserving Data Mining*" (Vaidya, 2006).
- Paper: "*Generating Synthetic Data with Differential Privacy*" (Abadi et al., 2016, arXiv).
- Dataset: [Kaggle.com/datasets](https://www.kaggle.com/datasets).
- Dukungan Dosen: Konsultasi mingguan ,tutorial GANs di Minggu 2. Hubungi via email [pakwowo@telkomuniversity.ac.id].

10. Penutup

Keberhasilan PROJECT bergantung pada disiplin dan kolaborasi antar anggota kelompok. Selamat bekerja, semoga proyek ini memperkaya pemahaman Keamanan Data untuk karir selanjutnya di industri Data Science.