



Teknik Informatika
Institut Teknologi Sumatera

Metodologi Penelitian (IF25-41029)

Ragam Metodologi & Pengelolaan Open Dataset

Tim Dosen Metodologi Penelitian
Program Studi Teknik Informatika, ITERA
Semester Ganjil 2026/2027



Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Sub-CPMK Pertemuan 7

Setelah pertemuan ini, mahasiswa mampu:

1. **Memilih metodologi penelitian** yang tepat sesuai rumusan masalah (kuantitatif, kualitatif, eksperimental, atau rekayasa).
2. **Menerapkan strategi pencarian, validasi, kurasi, dan pengelolaan** open dataset secara etis.

CPMK1001 Sikap ilmiah dalam **merancang** solusi — metode dan data menentukan validitas jawaban RQ.

CPMK1004 Sikap ilmiah dalam **mendokumentasikan** — protokol data menjadi isi **Bab 3** proposal.

Yang Sudah Dimiliki

- **P5:** RQ, RO SMART, hipotesis.
- **P6:** pitch lima babak — dan pertanyaan penguji: *“Datanya dari mana? Kenapa metode ini?”*

Hari Ini: Membuktikan

- **Metode** apa yang cocok untuk jenis RQ Anda?
- **Data** apa yang menjadi bukti, dan dari mana?
- Bagaimana **mengelola** dan **memakai** data secara etis?

Babak **Solution** di pitch hanya meyakinkan jika metode dan sumber data Anda **dapat dipertanggungjawabkan**.

Dari Sistem ke Bukti

Pergeseran Pola Pikir

Pola Pikir Proyek

“TA saya adalah membangun aplikasi X.”

- Keberhasilan = aplikasi **jalan**.
- Bab 4 berisi tangkapan layar.
- Penguji bertanya: *“Lalu apa temuannya?”*

Pola Pikir Riset

“TA saya menjawab RQ dengan bukti dari sistem X.”

- Sistem adalah **instrumen**; data adalah **produknya**.
- Bab 4 berisi **pengukuran** dan **analisis**.
- Kesimpulan menjawab RQ.

Analogi: sistem adalah **pabrik**, data adalah **produk**. Penguji menilai mutu produk, bukan hanya kemegahan pabriknya.

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Kuantitatif

- Mengukur dengan **angka**.
- Menguji hipotesis / membandingkan.
- Contoh: akurasi model, waktu respons, skor SUS.

Kualitatif

- Memahami **makna** dan **alasan**.
- Menggali pengalaman dan konteks.
- Contoh: wawancara pengguna, observasi alur kerja.

Campuran

- Menggabungkan **keduanya**.
- Angka menjawab “seberapa”, kata menjawab “mengapa”.
- Contoh: uji usability + wawancara.

Pendekatan **bukan selera**. Ia ditentukan oleh **jenis RQ** yang sudah Anda rumuskan di P5.

Eksperimen

Memanipulasi variabel bebas dan mengukur dampaknya pada variabel terikat dalam kondisi terkendali.

Survei

Mengumpulkan data dari banyak responden melalui kuesioner atau wawancara terstruktur.

Studi Kasus

Menyelidiki satu atau beberapa kasus nyata secara mendalam dalam konteks aslinya.

Action Research

Peneliti terlibat langsung memperbaiki praktik di organisasi, lalu merefleksikan hasilnya.

Dawson (2009) menekankan: satu proyek dapat memakai **lebih dari satu** metode, asalkan setiap metode menjawab sub-RQ yang jelas.

Komponen Wajib

- **Variabel bebas:** yang diubah (mis. arsitektur model).
- **Variabel terikat:** yang diukur (mis. mAP, latensi).
- **Variabel kontrol:** yang dijaga tetap (data, epoch, perangkat).
- **Baseline:** pembandingan yang adil.
- **Pengulangan:** beberapa seed / run.

Studi Kasus Berjalan

- Bebas: YOLO-n **dengan** vs **tanpa** modul objek kecil.
- Terikat: mAP@0,5 objek kecil; latensi (ms).
- Kontrol: dataset, split, resolusi input, Jetson Orin Nano.
- Ulangan: 5 seed; uji statistik.



Kapan Cocok?

- Kontribusi utama berupa **artefak**: sistem, model, metode, framework.
- RQ berjenis **desain/rekayasa** (P5).
- Paling umum untuk TA Informatika.

Syarat Disebut Riset

- Artefak menjawab **masalah nyata** dengan celah di pustaka.
- Ada **evaluasi rigor** terhadap kriteria terukur.
- Ada **pengetahuan** yang dapat dibagikan (Hevner et al., 2004).

Enam aktivitas DSRM menurut Peffers et al. (2007).

Survei

- Menjangkau **banyak** responden dengan cepat.
- Instrumen baku lebih kuat: **SUS** (Brooke, 1996), TAM, UEQ.
- Perhatikan **ukuran sampel** dan **keterwakilan**.
- Uji coba kuesioner sebelum disebar.

Studi Kasus

- Menggali **satu konteks** secara mendalam (mis. satu UMKM, satu puskesmas).
- Data dari wawancara, observasi, dokumen, log.
- Kekuatan: kaya konteks; kelemahan: generalisasi terbatas.
- Tulis keterbatasan ini secara jujur (P5).

Memilih Metode Dari RQ

Peta Keputusan

Jenis RQ (P5)	Metode Utama	Data / Bukti	Contoh TA
Komparatif / kausal	Eksperimen	Metrik kinerja berulang	Membandingkan model deteksi
Desain / rekayasa	Design science + eksperimen	Metrik artefak vs kebutuhan	Merancang sistem antrean klinik
Relasional	Survei / analisis data sekunder	Data numerik banyak sampel	Akses LMS vs nilai akhir
Deskriptif	Survei, analisis log, studi kasus	Statistik deskriptif, pola	Pola serangan di log server
Eksploratif / “mengapa”	Studi kasus, wawancara	Transkrip, catatan observasi	Alasan UMKM menolak aplikasi kasir

Jika metode Anda tidak dapat dipetakan ke RQ mana pun, salah satunya perlu direvisi — biasanya **RQ-nya**.

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

	Kuantitatif	Kualitatif
Primer	Primer-Kuantitatif Hasil eksperimen sendiri, skor SUS, waktu tugas, log sistem yang Anda bangun.	Primer-Kualitatif Transkrip wawancara, catatan observasi, umpan balik terbuka.
Sekunder	Sekunder-Kuantitatif Open dataset (COCO, UCI), data BPS, data API publik.	Sekunder-Kualitatif Ulasan aplikasi, unggahan forum, dokumen kebijakan.

Primer = Anda kumpulkan sendiri untuk RQ ini. Sekunder = sudah dikumpulkan pihak lain untuk tujuan lain.

Data	Sumber	Bentuk
Waktu inferensi model diukur di laptop Anda	?	?
Komentar terbuka pengguna setelah uji coba aplikasi	?	?
Dataset ulasan produk dari Hugging Face	?	?
Rekaman wawancara dengan admin jaringan kampus	?	?
Data curah hujan harian dari portal data pemerintah	?	?
Citra UAV yang Anda ambil sendiri lalu dianotasi	?	?

Satu TA sering memakai **lebih dari satu kuadran** — mis. dataset publik untuk pelatihan + data primer lokal untuk pengujian.

Teknik	Kuadran	Cocok untuk	Upaya
Eksperimen & benchmark	Primer-kuant.	Membandingkan algoritma, model, konfigurasi	Sedang-tinggi
Survei & kuesioner	Primer-kuant./ kual.	Evaluasi UX, penerimaan teknologi	Sedang
Wawancara & observasi	Primer-kual.	Kebutuhan, alur kerja, alasan	Sedang
Log & telemetry sistem	Primer-kuant.	Perilaku pengguna nyata, kinerja jaringan	Rendah-sedang
Open dataset	Sekunder-kuant.	ML, data science, benchmark	Rendah
API & web scraping	Sekunder	Data media sosial, harga, ulasan	Sedang-tinggi

Rancangan

1. Tetapkan **metrik** sebelum menjalankan eksperimen.
2. Pilih **baseline** yang relevan dan terbaru (matriks SotA).
3. Kunci lingkungan: versi pustaka, perangkat keras, seed.
4. Ulangi dan laporkan **rerata $\pm$ simpangan baku**.

Catat Selalu

- Spesifikasi perangkat keras.
- File konfigurasi setiap run.
- Commit Git kode yang dipakai.
- Waktu eksekusi dan memori.

Eksperimen yang tidak dapat diulang oleh orang lain — atau oleh Anda sendiri bulan depan — **bukan bukti**.

Survei

- Instrumen baku > buatan sendiri.
- Skala Likert konsisten.
- Pilot ke 5 orang dulu.

Wawancara

- Pedoman semi-terstruktur.
- Rekam dengan izin.
- Transkrip lalu kodekan tema.

Observasi

- Amati tugas nyata pengguna.
- Catat waktu, kesalahan, kebingungan.
- Think-aloud bila perlu.

Bias yang Sering Terjadi

Responden hanya teman satu angkatan · pertanyaan menggiring (*"Apakah aplikasi ini sangat membantu?"*) · peneliti sendiri yang menjadi moderator sekaligus penilai.

API vs Web Scraping

Menambang Data Daring

API Resmi

- Data terstruktur (JSON).
- Ada aturan pakai dan kuota jelas.
- Lebih stabil dan **lebih etis**.
- Pilihan pertama jika tersedia.

Web Scraping

- Mengambil dari halaman HTML.
- Rapuh: struktur situs berubah.
- Wajib cek **robots.txt** dan **Terms of Service**.
- Beri jeda antar-permintaan; jangan membebani server.

Data yang dapat diakses publik **belum tentu boleh** dikumpulkan dan dipublikasikan ulang. Bedakan “terlihat” dengan “diizinkan”.

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Mengapa Open Dataset?

Kelebihan dan Jebakan

Kelebihan

- Hemat waktu pengumpulan.
- **Dapat dibandingkan** langsung dengan studi lain.
- Mendukung **reproduksibilitas**.
- Sering sudah beranotasi dan terdokumentasi.

Jebakan

- Tidak cocok dengan **konteks lokal** RQ Anda.
- Label salah, duplikat, atau bias.
- **Lisensi** melarang penggunaan tertentu.
- Sudah “jenuh”: hasil SotA sulit dilampaui.

Pola umum yang kuat: **open dataset** untuk pelatihan dan pembandingan + **data lokal kecil** untuk menguji generalisasi pada konteks Anda.

Repositori	Keunggulan	Perhatikan
Hugging Face Datasets	NLP, visi, audio; dapat dimuat langsung via pustaka <code>datasets</code>	Kualitas bervariasi; cek dataset card
Kaggle Datasets	Sangat banyak domain; ada notebook contoh	Banyak unggahan ulang tanpa sumber asli
UCI ML Repository	Benchmark klasik tabular, terdokumentasi	Ukuran relatif kecil
Zenodo	Data riset dengan DOI permanen	Pencarian kurang terkurasi
OpenML	Dataset + tugas + hasil eksperimen terstandar	Fokus tabular
Google Dataset Search	Mesin pencari lintas repositori	Hanya indeks, cek sumber asli
Portal Satu Data Indonesia (data.go.id), BPS	Data pemerintah dan statistik Indonesia	Format dan pembaruan beragam

Visi Komputer

- ImageNet, COCO
- VisDrone (citra drone)
- ChestX-ray8 (medis)

Bahasa & Teks

- IndoNLU (benchmark Indonesia)
- SmSA (sentimen)
- NusaX (bahasa daerah)

Sinyal & Sensor

- PhysioNet, MIT-BIH (EKG)
- UCI HAR (aktivitas)
- Sleep-EDF (tidur)

Cara tercepat menemukan dataset yang **diakui**: lihat kolom **Dataset** di matriks SotA P4. Dataset yang dipakai studi pembandingan adalah kandidat utama Anda.

Langkah

1. Mulai dari **paper pembanding** di matriks SotA.
2. Susun query: **tugas + modalitas + domain + “dataset”**.
3. Telusuri repositori dan Google Dataset Search.
4. Lacak ke **sumber asli** (paper/halaman resmi), bukan unggahan ulang.
5. Catat kandidat di **kartu evaluasi**.

Contoh Query

- "small object detection" UAV dataset
- rice pest detection drone images
- Indonesian sentiment dataset informal
- network intrusion dataset IoT

Kualitas Isi

- **Ukuran** cukup untuk latih dan uji?
- **Akurasi label** — cek sampel acak manual.
- **Keseimbangan kelas** dan nilai hilang.
- **Duplikat** atau hampir-duplikat.
- **Bias** sumber: lokasi, waktu, demografi.

Kelengkapan Dokumentasi

- Siapa membuat, untuk apa, dan bagaimana?
- Protokol anotasi dan kesepakatan antar-anotator.
- Split resmi latih/validasi/uji.
- Paper rujukan yang wajib disitasi.
- Kerangka: **Datasheets for Datasets** (Gebru et al., 2021).

Aturan emas: **baca README dan paper asli dataset** sebelum menulis satu baris kode.

Lisensi	Makna Singkat	Riset TA	Komersial
CC0 / Public Domain	Bebas dipakai tanpa syarat	✓	✓
CC BY 4.0	Bebas dengan atribusi	✓	✓
CC BY-NC	Atribusi, non-komersial	✓	–
CC BY-SA	Atribusi, turunan harus berlisensi sama	✓	🌕
Research-only / kustom	Sesuai perjanjian penyedia	🌕	–
Tanpa lisensi	Hak cipta tetap pada pemilik	🌕	–

Data kesehatan seperti MIMIC mensyaratkan **akses terkredensial** dan pelatihan etika sebelum diunduh. Rencanakan waktunya.

Isian	Contoh: VisDrone
Sumber asli	Paper + GitHub resmi penyusun
Modalitas & ukuran	Citra & video drone; ribuan citra beranotasi
Label	Kotak pembatas 10 kelas objek
Split resmi	Ada (latih / validasi / uji)
Lisensi	Cek halaman resmi sebelum dipakai
Relevansi RQ	🟡 Objek kecil ✓ hama –

Keputusan

- **Pakai sebagai pra-latih** untuk objek kecil dari udara.
- **Tidak cukup** untuk hama → perlu data lokal primer.
- Kebutuhan ini **menjustifikasi** kontribusi dataset di P5.



Prinsip Kurasi

- Folder `raw/` **tidak pernah diubah.**
- Setiap transformasi = **skrip**, bukan edit manual.
- Catat checksum dan versi sumber.

Struktur Folder

`data/raw/` — unduhan asli
`data/interim/` — hasil pembersihan
`data/processed/` — siap latih
`data/README.md` — data card

Sumber Kebocoran

- Citra dari **video/lokasi yang sama** ada di latih dan uji.
- Normalisasi dihitung dari **seluruh** data.
- Data **masa depan** dipakai memprediksi masa lalu.
- Tuning hyperparameter pada data uji.

Pencegahan

- **Group split**: per pasien, per lokasi, per sesi.
- **Temporal split** untuk data deret waktu.
- Hitung statistik pra-proses **hanya** dari data latih.
- Kunci data uji sampai evaluasi akhir.

Kapoor & Narayanan (2023) menemukan kebocoran data sebagai penyebab utama klaim ML yang **tidak dapat direproduksi** di berbagai bidang.

Mengelola Data Riset

Prinsip FAIR

Findable

Punya identitas tetap (DOI) dan metadata yang dapat dicari.

Accessible

Dapat diunduh melalui protokol terbuka yang jelas.

Interoperable

Format standar (CSV, COCO JSON, WAV), bukan format tertutup.

Reusable

Lisensi jelas + dokumentasi asal-usul data.

Praktik Harian

- Cadangan di dua lokasi (lokal + cloud kampus).
- Versi data dengan DVC atau catatan versi.

Merilis Dataset Anda

- Unggah ke Zenodo/Hugging Face dengan DOI.
- Sertakan data card, lisensi, dan cara sitasi.

Prinsip FAIR: Wilkinson et al. (2016).

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Persetujuan (Consent)

Jelaskan tujuan dengan bahasa sederhana; minta persetujuan eksplisit; beri hak mengundurkan diri.

Privasi & Anonimitas

Pakai ID anonim; hapus informasi yang dapat mengidentifikasi individu.

Minimalisasi

Kumpulkan **hanya** data yang benar-benar dibutuhkan untuk menjawab RQ.

Keamanan

Simpan terenkripsi; batasi akses; hapus data sensitif sesuai rencana.

Indonesia memiliki **UU No. 27 Tahun 2022 tentang Pelindungan Data Pribadi**. Data wajah, suara, lokasi, dan kesehatan termasuk data pribadi — termasuk dalam riset mahasiswa.

Situasi	Izin / Langkah
Melibatkan partisipan manusia (survei, uji pengguna)	Lembar persetujuan; konsultasi pembimbing terkait telaah etik
Data kesehatan atau anak-anak	Telaah komisi etik penelitian; kerja sama institusi
Data dari instansi / perusahaan	Surat izin resmi dan perjanjian penggunaan data
Menerbangkan drone, merekam di ruang publik	Izin lokasi / pemilik lahan; patuhi aturan penerbangan
Scraping situs / media sosial	Cek ToS, robots.txt; anonimkan akun pengguna

Mengurus izin bisa memakan **minggu hingga bulan**. Masukkan ke jadwal TA sejak proposal.

Protokol Pengumpulan Data

Isi Bab 3 Proposal

Delapan Elemen

1. Tujuan & sub-RQ yang dilayani.
2. Sumber / populasi & kriteria inklusi.
3. Instrumen (alat, kuesioner, sensor).
4. Prosedur langkah demi langkah.
5. Jumlah target & justifikasinya.
6. Kontrol kualitas (anotasi, pilot).
7. Penyimpanan & keamanan.
8. Risiko & rencana cadangan.

Uji Protokol

Protokol yang baik dapat dijalankan oleh **orang lain** dan menghasilkan data yang **setara**.

Lakukan **pilot kecil** (mis. 50 citra atau 5 responden) sebelum pengumpulan penuh.

Elemen	Rencana
Tujuan	Data uji lokal untuk Sub-RQ 1 (mAP objek kecil) dan Sub-RQ 3 (ablation).
Sumber	Sawah di 2 kabupaten mitra; ketinggian terbang 10–20 m; musim tanam berjalan.
Instrumen	Drone kamera 4K; GPS; pedoman anotasi 3 kelas hama.
Prosedur	Terbang terjadwal pagi; ekspor citra; anotasi oleh 2 anotator.
Jumlah	≥ 1.500 citra; dihitung dari kebutuhan uji per kelas.
Kualitas	Cohen's kappa antar-anotator $\geq 0,7$ pada 10% sampel; pilot 50 citra.
Etika & izin	Izin pemilik lahan; tidak merekam wajah/rumah; data mentah di server prodi.
Risiko	Cuaca $\rightarrow$ jadwal cadangan; data kurang $\rightarrow$ augmentasi + VisDrone pra-latih.

Pendahuluan

Ragam Metodologi

Jenis & Teknik Data

Open Dataset

Etika & Protokol

Studi Kasus & Penutup

Aspek	AI / Data Science	IMK / UX	Sistem / Jaringan
RQ	Model mana paling akurat untuk sentimen ulasan informal?	Apakah redesign menurunkan waktu tugas pengguna lansia?	Seberapa efektif QoS HTB menjaga throughput e-learning?
Metode	Eksperimen	Eksperimen pengguna + wawancara	Eksperimen testbed
Data	Open dataset + 500 ulasan lokal berannotasi	Waktu tugas, skor SUS, transkrip	Throughput, latensi, packet loss
Kuadran	Sekunder + primer kuant.	Primer kuant. + kual.	Primer kuant.
Etika	Anonimkan nama akun	Consent tertulis partisipan	Uji di lab, bukan jaringan produksi

Misi

1. Dari matriks SotA Anda, catat **dataset** yang dipakai studi pembandingan.
2. Telusuri minimal **dua repositori** (Hugging Face, Kaggle, UCI, Zenodo, OpenML).
3. Temukan **3 kandidat** dataset untuk RQ Anda.
4. Isi **kartu evaluasi** untuk setiap kandidat.
5. Putuskan: pakai, pakai sebagian, atau perlu data primer?

Bonus

Jika perlu data primer, tulis **draf protokol** dengan delapan elemen — ini akan langsung menjadi bahan Bab 3 proposal.

Intisari Pertemuan 7

1. **Metode** dipilih dari jenis RQ: eksperimen, design science, survei, studi kasus, atau kombinasinya.
2. Data dipetakan pada dua sumbu — **primer/sekunder** × **kuantitatif/kualitatif** — dan dikumpulkan dengan teknik yang sesuai.
3. **Open dataset** wajib divalidasi (kualitas, lisensi, kebocoran), dikurasi dengan skrip, dan dikelola menurut prinsip FAIR — semuanya dalam **protokol yang etis**.

Preview Pertemuan 8

UTS — evaluasi pemahaman metodologi, kajian pustaka, dan perumusan masalah (materi P1–P7)

- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. Dalam P. W. Jordan et al. (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE.
- Dawson, C. W. (2009). *Projects in Computing and Information Systems: A Student's Guide* (2nd ed.). Addison-Wesley.
- Geburu, T., et al. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*.
- Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

Terima Kasih

Tim Dosen Metodologi Penelitian
Program Studi Teknik Informatika, ITERA