

LAPORAN DATA SOURCE IDENTIFICATION – DATA MART FTIK

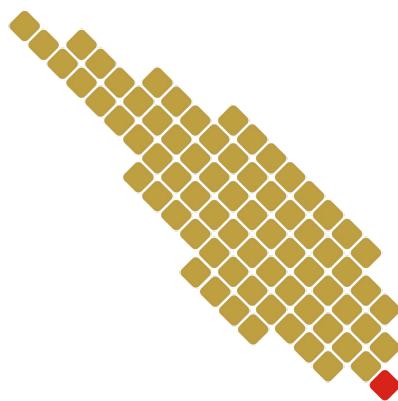
Disusun Oleh:

EIGI ARTAMEVIA (123450011)

SALSABILA PUTRI MAHARANI (123450070)

MUHAMMAD ZAKY ZAIDDAN (122440119)

HANIFAH INAYA SANI (123450123)



ITERA

PROGRAM STUDI SAINS DATA

FAKULTAS SAINS

INSTITUT TEKNOLOGI SUMATERA

2025

1. Identifikasi Sumber Data

Website FTIK tidak menyediakan database operasional atau API.

Karena itu, sumber data kita berasal dari **konten website** dan **dokumen publik** dalam bentuk:

(A). Website FTIK (HTML pages utama)

Jenis data yang bisa diambil:

- Berita fakultas
- Informasi kegiatan
- Agenda acara
- Informasi program studi
- Struktur organisasi

(B). Dokumen Publik (PDF)

Website FTIK juga menyediakan dokumen publik seperti:

- Laporan Tracer Study (PDF)
- Laporan kegiatan tertentu (jika ada)

Sumber data ini digunakan untuk:

- informasi unit, prodi, pejabat
- statistik umum yang dipublikasikan

(C). Manual Data Entry

tidak ada API atau database langsung, kita harus mengekstrak data dari website secara manual

- copy-paste judul berita
- mencatat tanggal acara
- menyalin nama program studi
- menyalin daftar pejabat FTIK

(D). Tidak ada OLTP dan tidak ada API

Website FTIK tidak menyediakan:

- database OLTP terbuka
- data mahasiswa
- prodi lengkap dalam bentuk tabel
- API kegiatan
- API berita

2. Data Source Analysis

(A) Halaman Berita

- **Struktur:**
 - a. Judul berita
 - b. Tanggal publikasi
 - c. Penulis (kadang ada)
 - d. Isi berita
 - e. Tag/kategori (kadang tidak konsisten)
- **Schema (hasil ekstraksi):**

Field	Tipe
title	string
publish_date	date
content	text
category	string
url	string

- **Kualitas:** Konsisten karena dipublikasikan resmi
- **Update:** bertambah 2–5 per bulan
- **Volume Awal:** ±300 artikel
- **Catatan:** Terkadang panjang artikel berbeda

(B) Halaman Agenda/Kegiatan

- **Struktur:**
 - a. nama acara
 - b. Tanggal
 - c. Lokasi
 - d. penyelenggara
- **Schema (hasil ekstraksi):**

Field	Tipe
event_name	string
event_date	date
event_time	string
location	string
organizer	string

- **Kualitas:** Bervariasi, tergantung input admin
- **Update:** mengikuti aktivitas prodi/fakultas
- **Volume Awal:** 10–50 per tahun
- **Catatan:** Kadang ada kegiatan tanpa detail lengkap

(C) Halaman Program Studi

- **Struktur:**
 - a. nama prodi
 - b. profil singkat
 - c. ketua prodi
- **Schema (hasil ekstraksi):**

Field	Tipe
prodi_name	string
kepala_program	string
deskripsi	text

- **Kualitas:** Sangat stabil
- **Update:** sangat stabil
- **Volume Awal:** ±10 prodi

- **Catatan:** Cocok untuk dimensi (Dim_Prodi)

(D) Struktur Organisasi

- **Struktur:**
 - a. nama pejabat
 - b. Jabatan
 - c. periode jabatan
- **Schema (hasil ekstraksi):**

Field	Tipe
name	string
position	string
unit	string
term_start	date
term_end	date

- **Kualitas:** Tinggi
- **Update:** berubah setiap 4 tahun
- **Volume Awal:** ±10 pejabat
- **Catatan:** Cocok untuk dimensi organisasi

3. Data Profiling (berdasarkan konten website FTIK)

(A).Distribusi Data

- Publikasi berita lebih banyak pada bulan menjelang akreditasi atau event besar.
- Kegiatan paling banyak dari Prodi Teknik Sipil dan PWK (berdasarkan berita).

(B).Null/Incomplete Values

- **Kategori berita:** banyak yang kosong
- **Lokasi agenda:** sebagian kosong
- **Penyelenggara acara:** kadang tidak muncul

(C).Outliers

- Kegiatan tanpa deskripsi
- Perubahan format tanggal karena masalah default WordPress

(D).Duplikasi

- Beberapa berita memiliki judul mirip, tetapi tidak duplikat identik
- Agenda kadang ditampilkan ulang di berita → perlu dedup

(E).Format Konsistensi

- Tanggal tidak seragam (format berbeda)
- Penulisan prodi kadang berbeda (“PWK”, “Planologi”)
- Nama pejabat tidak konsisten kapitalisasi