BAB I
PENDAHULUAN

1.1 Latar Belakang

Data kependudukan merupakan elemen krusial dalam perencanaan dan evaluasi kebijakan di berbagai sektor pemerintahan. Di Dinas Kependudukan dan Pencatatan Sipil (Dukcapil) Kabupaten Madiun, penyediaan data agregat kependudukan yang akurat bagi berbagai instansi dan pihak yang berkepentingan merupakan salah satu layanan yang dilakukan secara terus-menerus. Sebelum adanya integrasi sistem yang memadai, setiap kali terdapat permintaan, petugas Dukcapil harus mengatur, memilah, dan merekapitulasi data agregat kependudukan tersebut secara manual dari puluhan hingga ratusan *spreadsheet* sebelum dapat didistribusikan. Proses rekapitulasi manual ini memakan waktu yang tidak sedikit serta sangat rentan terhadap kesalahan (*human error*), sehingga menghambat kecepatan dan efisiensi pelayanan administrasi kependudukan.

Untuk mengatasi permasalahan tersebut dan mempermudah petugas dalam mendistribusikan data, dikembangkanlah Sistem Rekap Data Kependudukan (SERDADU). Sistem ini bertujuan untuk mengotomatisasi proses pengelolaan data agregat kependudukan, menggantikan alur kerja manual dengan proses ekstraksi dan transformasi otomatis ke dalam basis data terpusat. Dengan hadirnya SERDADU, diharapkan penyediaan dan distribusi data agregat kependudukan kepada instansi atau masyarakat yang membutuhkan dapat dilakukan dengan jauh lebih cepat, akurat, dan efisien, sehingga langsung menjawab kendala waktu dan beban kerja manual yang sebelumnya dialami oleh petugas Dukcapil.

Penelitian pengembangan model visualisasi data kependudukan di Kecamatan Medan Amplas menunjukkan bahwa *dashboard* interaktif, peta tematik, dan grafik distribusi mampu menyajikan informasi demografis secara lebih efisien dan intuitif untuk mendukung pengambilan keputusan [1].

Penerapan visualisasi data dan informasi kependudukan berbasis web juga memperlihatkan bahwa data penduduk dapat dipantau melalui tampilan visual yang lebih mudah dibaca, sehingga informasi perkembangan kependudukan lebih mudah dipahami oleh pengguna [2].

Pada objek penelitian ini, inventarisasi awal terhadap folder dataset AGREGAT_202501 menunjukkan terdapat 123 file Excel agregat dengan 33 kelompok varian nama file, serta memungkinkan untuk mengolah data dari semester sebelumnya dan semester selanjutnya dengan jumlah file hingga ribuan. Di sisi lain, konfigurasi import aktif pada Sistem Rekap Data Kependudukan (SERDADU) memetakan 13 kategori utama, yaitu pendidikan, pekerjaan, agama, jenis kelamin, status perkawinan, kepala keluarga, wajib KTP, kepemilikan KK, KIA, akta kelahiran, akta kawin cerai, kelompok umur, dan umur tunggal.

Jumlah dan variasi tersebut menimbulkan persoalan nyata pada proses migrasi data. Beberapa file menggunakan struktur *header* IDEM, KODE, dan WILAYAH, sedangkan file lain hanya menggunakan KODE dan WILAYAH. Format kode wilayah juga tidak seragam karena sebagian ditulis bertitik dan sebagian lain tanpa titik. Perbedaan ini berisiko menimbulkan kesalahan identifikasi kategori data, kesalahan pemetaan kolom, inkonsistensi wilayah, dan kegagalan pemuatan data ke basis data target apabila tidak ditangani dengan mekanisme transformasi yang terkontrol.

Untuk menangani kondisi tersebut, SERDADU menerapkan proses *Extract*, *Transform*, *Load* dengan pendekatan *rule-based* dan *string matching*. Pendekatan *rule-based* digunakan untuk memilih aturan pemetaan kategori dan atribut target, sedangkan *string matching* digunakan untuk mengenali variasi nama file, nama sheet, *header* kolom, alias, serta identitas wilayah. Berdasarkan kondisi tersebut, penelitian ini difokuskan pada analisis implementasi ETL pada SERDADU untuk melihat bagaimana data agregat yang heterogen diarahkan menjadi data yang lebih terstruktur.

1.2 Rumusan Masalah

1. Bagaimana implementasi proses *Extract*, *Transform*, *Load* pada migrasi data agregat kependudukan dengan pendekatan *rule-based* dan *string matching* pada sistem SERDADU?
2. Bagaimana kesesuaian implementasi tersebut dalam mengenali variasi struktur data sumber dan memetakan hasil transformasi ke basis data target yang sesuai?

1.3 Batasan Masalah

1. Penelitian difokuskan pada analisis implementasi ETL pada sistem SERDADU dengan ruang lingkup 13 kategori import utama dan dataset file Excel agregat yang relevan sebagai sumber migrasi data.
2. Analisis dibatasi pada proses import melalui halaman admin, layanan import utama, konfigurasi pemetaan data, pendekatan *rule-based*, *string matching*, serta pemuatan hasil transformasi ke basis data target; penelitian ini tidak membahas machine learning, OCR, jalur import melalui terminal, atau uji performa sistem.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis implementasi proses *Extract*, *Transform*, *Load* pada migrasi data agregat kependudukan dengan pendekatan *rule-based* dan *string matching* pada sistem SERDADU.
2. Mengevaluasi kesesuaian implementasi tersebut dalam mengenali variasi struktur data sumber dan memetakan hasil transformasi ke basis data target yang sesuai.

1.5 Manfaat Penelitian

1. Penelitian ini diharapkan memberikan manfaat berupa gambaran yang lebih terstruktur mengenai implementasi ETL pada SERDADU, menunjukkan titik-titik kritis pada proses import data agregat

kependudukan, serta menjadi acuan awal untuk evaluasi dan perbaikan mekanisme migrasi data pada sistem yang memiliki karakteristik serupa.

2. Dengan demikian, hasil penelitian tidak hanya menjelaskan alur proses yang sedang berjalan, tetapi juga membantu memperjelas bagian mana dari implementasi import yang sudah sesuai dan bagian mana yang masih berpotensi disempurnakan.