

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Data dari CNN Indonesia menyebutkan bahwa Indonesia membutuhkan 15 juta talenta digital hingga tahun 2030, sementara hingga saat ini baru tersedia sekitar 12 juta talenta digital [1]. Hal ini mendorong berbagai institusi, termasuk Telkom Indonesia, untuk mencetak SDM unggul di bidang teknologi. Sejak 2021, Telkom meluncurkan program Digistar guna menyiapkan talenta digital menghadapi industri 4.0. Namun, berdasarkan wawancara dengan divisi HC Com, program ini belum memiliki sistem pendataan yang terstruktur, sehingga menyulitkan identifikasi dan pengembangan peserta secara sistematis.

Pada 2023, HC Com Telkom Indonesia mengembangkan Candidate Relationship Management System (CRMS) sebagai repositori data peserta Digistar, pemantau aktivitas, dan alat pendukung strategi funneling talenta. Sistem ini mengikuti tren global digitalisasi manajemen SDM. Susanto & Hamzali (2024) menyebut sistem digital seperti Applicant Tracking System dapat meningkatkan efisiensi rekrutmen dan akurasi pemetaan kandidat [2]. Sebelum diintegrasikan ke dalam CRMS, data harus melalui proses pembersihan dan standarisasi guna memastikan validitas dan konsistensinya. Sebelum data diimplementasikan ke dalam sistem langkah pembersihan dan standarisasi merupakan langkah yang krusial. Menurut studi dari Canidium, data yang bersih telah melalui tahap “data cleansing dan transformation” untuk memastikan konsistensi format, akurasi, dan kelengkapan sebelum integrasi sistem [3].

Permasalahan utama yang dihadapi adalah proses pembersihan data yang masih dilakukan secara manual dengan memeriksa data satu per satu, sehingga memakan waktu dan rentan terhadap *human error*. Data yang tidak konsisten dan tidak bersih dapat menimbulkan duplikasi data ketika format email tidak seragam, atau data tidak terdeteksi sistem karena tidak sesuai format standar. Tantangan spesifik yang sering muncul adalah inkonsistensi penulisan nama jurusan dan universitas yang tidak sesuai dengan referensi di tabel master, seperti variasi "Teknik Informatika" dan "Informatika", atau "Telkom University" yang kadang ditulis sebagai "Universitas Telkom". Kondisi ini membuat proses pembersihan dan standarisasi data menjadi tidak efisien.

Tantangan lain yang dihadapi adalah proses pencocokan data feedback dengan data registrasi yang juga dilakukan secara manual. Data feedback yang hanya memuat informasi terbatas seperti nama lengkap dan email perlu dicocokkan dengan data registrasi untuk mendapatkan informasi lengkap peserta. Proses pencocokan

manual ini membutuhkan waktu yang lama dan berpotensi menimbulkan kesalahan dalam identifikasi peserta, sehingga menghambat efektivitas pengelolaan data dalam CRMS.

Untuk mengatasi permasalahan tersebut, dibuatlah sistem otomatisasi untuk pembersihan dan standarisasi data menggunakan teknik *fuzzy matching*. Teknik ini memungkinkan sistem untuk mencocokkan entri yang serupa meskipun terdapat perbedaan penulisan atau kesalahan ketik [4]. Selain itu, sistem otomatisasi juga dilengkapi dengan fitur pencocokan file registrasi dan feedback secara otomatis untuk mengatasi proses manual yang memakan waktu lama dalam mengidentifikasi dan menggabungkan data peserta. Otomatisasi ini diharapkan dapat mempercepat proses pembersihan data, mengurangi potensi *human error*, serta memastikan konsistensi dalam penulisan nama jurusan, universitas, dan data lainnya di seluruh dataset. Dengan demikian, proses pembersihan dan standarisasi data akan menjadi lebih efisien, akurat, dan dapat mendukung optimalisasi fungsi CRMS dalam mengelola talenta digital program Digistar.

## 1.2 Rumusan Masalah dan Solusi

Berdasarkan latar belakang yang telah diuraikan maka dapat diidentifikasi masalah yang ada sebagai berikut:

1. Bagaimana upaya untuk mengurangi *human error* serta mempercepat proses pembersihan dan standarisasi data, termasuk pencocokan data feedback dengan data registrasi yang selama ini masih dilakukan secara manual?
2. Bagaimana cara mengatasi ketidakteraturan dan inkonsistensi dalam penulisan nama jurusan dan universitas peserta?

Solusi dari permasalahan diatas adalah sebagai berikut:

1. Membangun sistem otomatisasi pengolahan data yang terdiri dari dua fitur utama, yaitu:
  - a. Fitur pencocokan data feedback dan data registrasi, yang akan mempercepat proses pelengkapan data secara otomatis.
  - b. Fitur pembersihan dan standarisasi data, yang akan mengurangi proses manual dalam membersihkan data duplikat, mencocokkan nama jurusan dan universitas dengan referensi pada tabel master, serta menyelaraskan format data agar konsisten dan sesuai standar.
2. Menggunakan teknik *fuzzy matching* untuk mencocokkan variasi penulisan nama jurusan dan universitas berdasarkan tabel referensi secara otomatis.

### 1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, tujuan yang ingin dicapai adalah sebagai berikut:

1. Mengurangi *human error* dan mempercepat proses pembersihan, standarisasi, serta pencocokan data dengan mengotomatiskan langkah-langkah yang sebelumnya dilakukan secara manual.
2. Meningkatkan konsistensi data dengan menyelaraskan penulisan nama jurusan dan universitas secara otomatis dengan menerapkan *fuzzy matching*.

### 1.4 Batasan Masalah

Untuk membatasi ruang lingkup pembahasan, berikut adalah batasan masalah yang diterapkan:

1. Data yang ditampilkan dalam laporan ini adalah data dummy untuk menjaga kepatuhan terhadap Undang-Undang Perlindungan Data Pribadi (UU PDP).
2. Penyajian data hasil pembersihan dilakukan menggunakan Streamlit.
3. Sistem otomasi yang dikembangkan berfokus pada mempercepat proses pembersihan dan standarisasi data sebelum integrasi ke dalam *Candidate Relationship Management System* (CRMS), tanpa pengembangan fitur analisis lanjutan.

### 1.5 Penjadwalan Kerja

Untuk merealisasikan solusi yang telah dirancang, dibutuhkan perencanaan waktu yang terstruktur agar setiap tahapan dapat diselesaikan secara tepat waktu dan efisien. Jadwal ini mencakup seluruh proses mulai dari analisis kebutuhan hingga penyusunan laporan akhir. Berikut merupakan jadwal pelaksanaan kerja:

Tabel 1. 1 Pelaksanaan Kerja

No	Deskripsi	Januari				Februari				Maret				April				Mei				Juni		
		2025				2025				2025				2025				2025				2025		
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
1	Analisis Kebutuhan & Spesifikasi																							

No	Deskripsi	Januari				Februari				Maret				April				Mei				Juni		
		2025				2025				2025				2025				2025				2025		
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
2	Desain Sistem (Flowchart, Use case)																							
3	Pengembangan Fitur 1: Pencocokan Feedback-Registrasi																							
4	Pengembangan Fitur 2: Pembersihan & Standarisasi Data																							
5	Pengujian Fungsional Fitur																							
6	Pengembangan Antarmuka Pengguna (Streamlit)																							
7	Pengujian Sistem secara keseluruhan																							
8	Perbaikan dan Optimasi Berdasarkan Hasil Pengujian																							
9	Finalisasi Dokumentasi dan Laporan Akhir																							