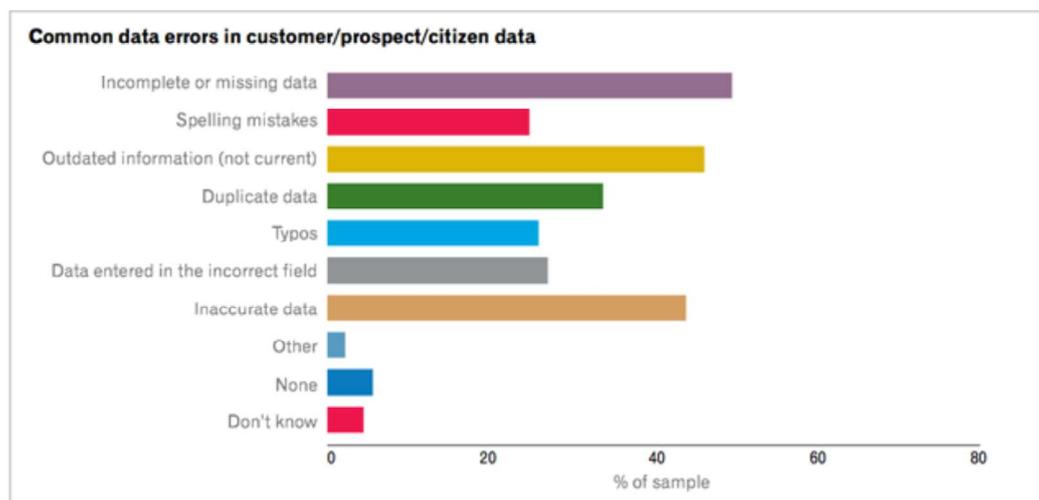


BAB I PENDAHULUAN

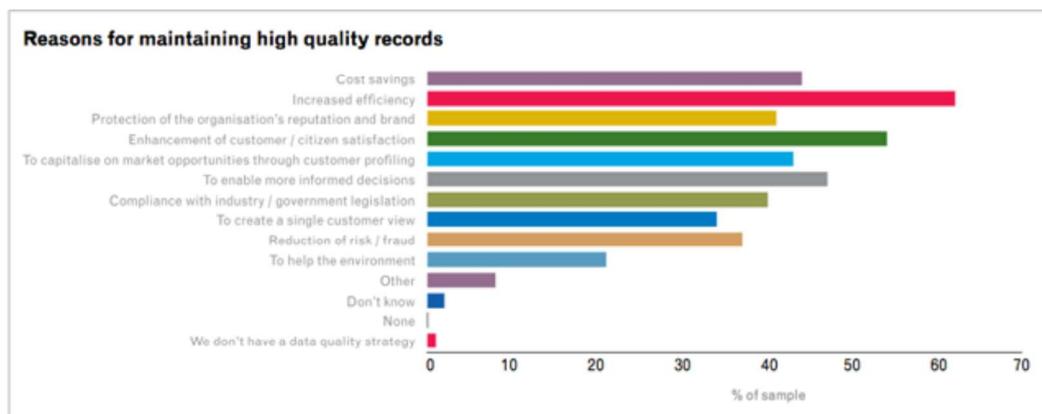
I.1 Latar Belakang

Data merupakan suatu aset bagi perusahaan yang bernilai sangat tinggi dan tidak akan terkonsumsi ketika digunakan, tidak seperti aset finansial atau aset fisik. Perusahaan selalu perlu mengelola datanya, agar dapat memenuhi kebutuhan bisnis. Untuk memastikan data dapat memenuhi kebutuhan bisnis, perusahaan harus bekerja sama dengan pengguna data agar dapat menentukan kebutuhan tersebut, termasuk karakteristik yang membuat data berkualitas tinggi. Data berkualitas tinggi merupakan data yang tersedia, lengkap, relevan, tepat waktu, akurat, konsisten, bermakna dan dapat dipahami. Data yang berkualitas tinggi dapat memberikan keuntungan diantaranya: meningkatkan pengalaman pengguna, produktivitas yang lebih tinggi, berkurangnya resiko, kemampuan untuk bertindak berdasarkan peluang, dan meningkatkan pendapatan (DAMA International, 2017). Namun kenyataannya, banyak data yang belum memiliki kualitas tinggi, sehingga kebutuhan bisnis pun menjadi terhambat.



Gambar I 1 Kesalahan umum pada data (Kramer, 2015)

Gambar I 1 merupakan hasil survey dari lembaga yang menemukan bahwa sebagian besar perusahaan, lebih dari 90 persennya melaporkan bahwa terdapat kesalahan dalam data kontak mereka (Kramer, 2015). Kesalahan-kesalahan ini menyebabkan turunnya kualitas data. Beberapa kesalahan yang memiliki persentase terbesar diantaranya *incomplete or missing data*, *outdated information*, dan *inaccurate data*. Data yang tidak lengkap, informasi yang terlalu lama, maupun data yang tidak akurat dapat menyebabkan kerugian pada perusahaan. Lembaga tersebut memperkirakan rata-rata perusahaan kehilangan 12 persen pendapatannya karena data yang tidak akurat yang menyebabkan penurunan produktivitas, sumber daya yang terbuang, dan peluang yang hilang untuk pemasaran lintas saluran karena terdapat celah dalam catatan data kontak (Kramer, 2015).



Gambar I 2 Alasan pentingnya memelihara kualitas data (Kramer, 2015)

Berdasarkan gambar I 2 merupakan hasil penjelasan bisnis mengapa memelihara kualitas data itu penting. Salah satu persentase terbesarnya merupakan *increased efficiency*. Maka, kualitas data yang baik dapat meningkatkan efisiensi perusahaan (Kramer, 2015).

Maka dari itu, *Data Quality Management* (DQM) dibutuhkan sebagai pemeliharaan kualitas data, agar data pada perusahaan memiliki kualitas yang baik. *Data Cleansing* merupakan salah satu metode yang dapat dilakukan untuk meningkatkan kualitas data. Data cleansing berfokus pada identifikasi data yang

melibatkan serangkaian aktivitas pada bagian tertentu yang berguna untuk membersihkan ketidaksesuaian dalam basis data (Noor, 2019).

Lalu, bagaimana caranya menentukan apakah data tersebut memiliki kualitas yang baik atau tidak? *Data Quality Rules* adalah salah satu caranya. *Data Quality Rules* memberikan aturan tentang data yang tidak baik, dan memberikan juga solusi untuk memperbaiki data tersebut (Fan & Geerts, 2012). *Data Quality Rules* di setiap perusahaan berbeda-beda karena menyesuaikan kondisi dan kebutuhan bisnis mereka. Dengan adanya *Data Quality Rules*, penggunaan *Data Cleansing* menjadi lebih terarah sesuai dengan peraturan dari setiap perusahaan.

Pada aplikasi *Data Quality Management*, terdapat berbagai paket atau modul-modul *cleansing* yang sudah ada dan sudah pernah dibuat sebelumnya, berdasarkan *rules* yang telah ditentukan. Dengan adanya modul yang sudah pernah dibuat, pengerjaan *Data Cleansing* pun bisa lebih cepat dengan memilah Langkah-langkah yang ada dan menyesuainya dengan *rules* yang akan digunakan. Namun, dari sisi aplikasi, penetapan rules pada modul *cleansing* membuat aplikasi menjadi tidak dinamis. Modul hanya akan bisa dijalankan untuk suatu *case* tertentu apabila menyesuaikan dengan rule. Agar modul dapat dibuat menjadi lebih dinamis dan dapat digunakan secara umum, perlu dilakukan penyesuaian pada modul-modul tersebut. Dilain sisi, beberapa modul yang harusnya bersambung seperti *Cleansing Pattern* dengan *Profiling Pattern*, saat ini belum tersambung. Maka dari itu perlu diperlukan juga integrasi modul untuk mempermudah pengguna dalam penggunaan aplikasi *Data Quality Management*.

Aplikasi *Data Quality Management* pun sebelumnya masih berjalan secara lokal pada penelitian sebelumnya, yang mana masih menggunakan *platform windows* untuk menjalankannya, dan masih menggunakan konfigurasi masing-masing peneliti-peneliti sebelumnya. Dengan kondisi tersebut, aplikasi tidak akan bisa dijalankan pada server tanpa dilakukan penyesuaian. Pada umumnya, server yang digunakan untuk memasang aplikasi web merupakan server berbasis *open source* yang salah satunya merupakan Linux. Maka dari itu, penyesuaian pun dibutuhkan agar aplikasi *Data Quality Management* dapat dijalankan pada *platform open*

source. Penyesuaian yang dibutuhkan disinipun berupa penyesuaian konfigurasi, serta penyesuaian *syntax* agar dapat dijalankan pada *platform open source*.

Pada penelitian kali ini, peneliti berfokus pada integrasi *Data Cleansing*, yaitu mengintegrasikan Modul Profiling Pattern dengan modul Cleansing Pattern agar menjadi satu alur, mengintegrasikan modul Cleansing yang baru, serta menyesuaikan aplikasi *Data Quality Management* agar dapat dijalankan pada *open source platform*. Selain itu, peneliti juga berfokus pada penyesuaian modul baik modul eksisting maupun modul baru agar dapat digunakan secara umum tanpa harus berdasarkan suatu *rule*. Tujuan pengintegrasian modul Profiling Pattern dengan Cleansing Pattern yaitu untuk memotong alur proses yang peneliti anggap sia-sia dan menyulitkan pengguna. Dengan itu, siapapun yang menggunakan aplikasi ini dapat membersihkan data mereka sesuai dengan format dan keinginan mereka agar dapat membantu dalam peningkatan pengambilan keputusan selanjutnya.

I.2 Perumusan Masalah

Berdasarkan uraian masalah yang telah dipaparkan pada latar belakang, maka dapat diperoleh permasalahan yang akan dikaji dalam penelitian ini antara lain:

1. Bagaimana mengintegrasikan serta mempersingkat alur dari modul eksisting pada aplikasi *Data Quality Management* agar lebih mudah digunakan?
2. Bagaimana mengintegrasikan serta memasang modul *cleansing* baru pada aplikasi *Data Quality Management* menggunakan *open source tools*?
3. Bagaimana penyesuaian agar aplikasi *Data Quality Management* dapat berjalan pada platform Linux?

I.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian yang dikaji ini antara lain sebagai berikut:

1. Mengintegrasikan serta mempersingkat alur dari modul eksisting pada aplikasi *Data Quality Management* agar lebih mudah digunakan.
2. Mengintegrasikan serta memasang modul *cleansing* baru pada aplikasi *Data Quality Management* menggunakan *open source tools*.
3. Menyesuaikan aplikasi *Data Quality Management* agar dapat berjalan pada platform Linux.

I.4 Batasan Penelitian

Adapun batasan dalam melakukan penilitan ini adalah sebagai berikut:

1. Penelitian ini dilakukan hanya pada bidang *data cleansing* dan *data profiling*.
2. Menggunakan modul-modul *data cleansing* yang sudah pernah dibuat.
3. Mengintegrasikan modul-modul yang sudah pernah dibuat dengan tujuan mempermudah proses pembersihan serta penggunaan yang lebih fleksibel.
4. Implementasi penerapan integrasi paket pada modul data cleansing menggunakan *open source tools* yaitu Pentaho Data Integration.
5. Sistem yang digunakan merupakan sistem eksisting, yaitu *Data Quality Management Tools* yang menggunakan *framework* pemrograman web Laravel dan RDBMS berupa Mysql.
6. Sistem dapat dijalankan pada sistem operasi Linux.

I.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain bertambahnya ilmu serta pengalaman penulis terhadap penggunaan kembali dan pengintegrasian paket pada modul *Data Cleansing*, dan dapat memberikan manfaat yang positif untuk perusahaan. Manfaat yang diharapkan untuk perusahaan antara lain meningkatnya kualitas data pada perusahaan agar dapat mempermudah bisnis dalam mengambil keputusan, serta mempermudah pengguna dalam melakukan *cleansing* data.

I.6 Sistematika Penulisan

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut:

Bab I Pendahuluan

Pada bab ini berisi mengenai penjelasan latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penelitian.

Bab II Tinjauan Pustaka

Pada bab ini menjelaskan tinjauan pustaka yang berisi teori-teori dan kajian-kajian literatur pendukung untuk riset dan beberapa riset yang pernah dilakukan oleh peneliti sebelumnya.

Bab III Metodologi Penelitian.

Pada bab ini berisi penjelasan model konseptual dan sistematika penyelesaian masalah penelitian yang akan digunakan pada riset ini.

Bab IV Analisis dan Perancangan

Pada bab ini berisi penjelasan mengenai analisa dan perancangan penelitian yang akan dilakukan.

BAB V Implementasi Dan Pengujian

Pada bab ini berisi penjelasan tentang implementasi perancangan logika yang sudah dirancang.

Bab VI Kesimpulan dan Saran

Pada bab ini menjelaskan kesimpulan dan saran yang berisikan tentang ringkasan hasil dari penelitian yang sudah dilakukan dan saran yang dapat dipertimbangkan untuk penelitian berikutnya.