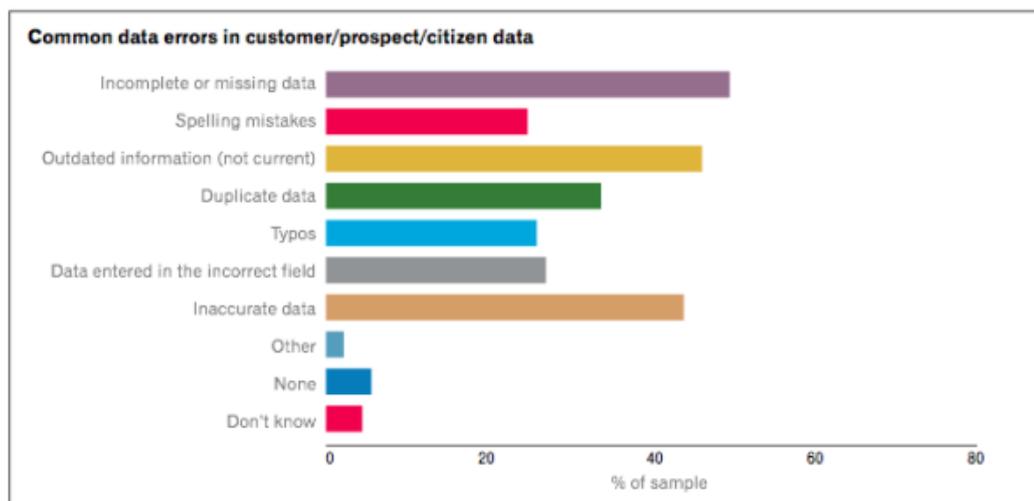


Bab I Pendahuluan

I.1 Latar Belakang

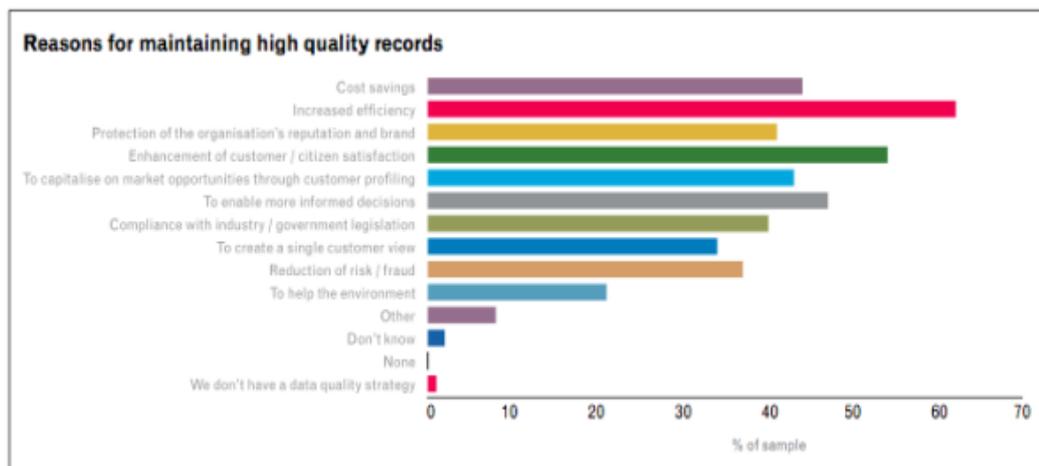
Kemajuan teknologi dan informasi adalah sesuatu yang tidak bisa kita hindari lagi dalam kehidupan ini, kemajuan teknologi dan informasi akan berjalan beriringan dengan ilmu pengetahuan. Setiap inovasi dan pengembangan terus dilakukan untuk memberikan manfaat serta kemudahan bagi perusahaan. Salah satu manfaat dan kemudahan terbesar yang diperoleh perusahaan adalah berkembangnya sebuah data, data merupakan fakta mentah yang belum diolah yang sering kali tidak langsung dapat dimengerti oleh penerima data tersebut, maka dari itu data tersebut harus diolah terlebih dahulu sebelum akhirnya disajikan sebagai informasi untuk dapat diterima dan dimengerti dan dimanfaatkan oleh perusahaan untuk mendukung proses bisnisnya. Dengan demikian, data harus memiliki kualitas yang baik, sehingga dapat memberikan nilai manfaat terhadap perusahaan (Amethyst, 2018).



Gambar I-1 Kontributor kesalahan pada data (Kramer, 2015)

Pada gambar 1-1. hasil penelitian menyoroiti tentang kesalahan pada data yang umum terjadi di perusahaan yang ada di UK, Eropa, dan USA, dari gambar tersebut menunjukkan bahwa banyak aspek yang mempengaruhi kualitas data. Persentase terbesar penyumbang kesalahan data adalah *incomplete or missing data*, data yang belum lengkap akan membuat perusahaan kehilangan pendapatannya karena kualitas data tidak lengkap akan mempengaruhi produktivitas dan pengambilan keputusan. Pada masa kini, data banyak digunakan dalam suatu perusahaan sebagai acuan dalam

menjalankan roda perusahaan, tidak hanya untuk dioperasikan tetapi juga digunakan untuk level yang strategis. Kualitas data yang buruk akan berdampak negatif bagi organisasi seperti keputusan yang buruk saat membuat dan merencanakan strategi. Oleh karena itu, manajemen kualitas data atau DQM menjadi masalah yang berkembang saat ini, kualitas data yang buruk berarti informasi yang tersedia tidak akurat dan dapat membahayakan organisasi serta bisa mengakibatkan sumber daya menjadi terbuang. Selain itu juga dapat mempengaruhi hubungan organisasi dengan pelanggannya. Kualitas data identik dengan kualitas informasi karena kualitas data yang buruk menghasilkan informasi yang tidak valid dan bisnis yang buruk kinerjanya. Dua masalah paling umum yang disebabkan oleh kualitas data yang buruk pertama adalah diperlukan waktu tambahan untuk merekonsiliasi data, kedua adalah kehilangan keandalan dalam sistem atau aplikasi karena kesalahan yang ada (Sabtania dkk, 2017).



Gambar I-2 Penelitian tentang alasan menjaga kualitas data (Kramer, 2015)

Berdasarkan gambar 1-2. memperlihatkan bahwa data dianggap sebagai aspek berharga oleh perusahaan. Bisnis yang tidak mempertimbangkan untuk menggunakan data dalam pemasaran mereka sekarang telah menyadari pentingnya data dalam meningkatkan efektivitas pemasaran mereka dan dalam membuat keputusan bisnis yang lebih baik. Namun, data yang buruk akan merugikan sebuah perusahaan. Setiap tahun *Experian Data Quality* melakukan penelitian tentang kualitas data di perusahaan-perusahaan di Inggris, Eropa dan Amerika Serikat (Kramer, 2015).

Informasi yang berkualitas dihasilkan dari sebuah proses yang membutuhkan kemampuan untuk menganalisis dan mengelola sebuah data yang nantinya bisa

digunakan untuk menunjang kebutuhan dan meningkatkan *value* perusahaan. Untuk mendapatkan kualitas data yang sesuai kebutuhan perusahaan, data perlu dilakukan pengoreksian, atau dihapus data-data yang salah, tidak lengkap, tidak akurat, atau memiliki format yang salah, maka dari itu dibutuhkan sebuah metode yang dinamakan *data cleansing*. Sebagian besar *tools* yang bisa digunakan untuk melakukan penelitian ini berbayar, dengan anggaran yang relatif tinggi, maka dari itu sulit diimplementasikan dalam skala kecil hingga perusahaan menengah di negara berkembang (Kusumasari & Fitria, 2016). Sebaliknya, kami menemukan *open source tools* yang membutuhkan anggaran jauh lebih kecil. Fokus dari penelitian ini adalah untuk menyediakan proses *data cleansing* menggunakan *open source tools*, yang merupakan *tools* yang tidak berbayar. Salah satu *tools* yang digunakan adalah *Pentaho Data Integration* (PDI), sebuah aplikasi dengan kemampuan mengolah data dalam skala besar, menyediakan proses ETL, dan juga mengubah data dari satu format ke format lain.

Pada penelitian sebelumnya yang dilakukan oleh Khoirunisa mengatakan bahwa pendekatan *Data Quality Management* dilakukan dengan menggunakan pendekatan *data cleansing (DC)*. Pengolahan *Data Quality Management* dengan mengimplementasikan algoritma menggunakan salah satu *tools open source* yaitu *Pentaho Data Integration*. *Pentaho Data Integration* dapat membantu menampung serta memberikan fungsi untuk melakukan perancangan arsitektur aplikasi sebagai wadah untuk menjalankan algoritma agar dapat diintegrasikan menjadi satu fitur *data quality* (Khoirunisa & Kusumasari, 2017).

Pada penelitian ini, badan pemerintah yang bertanggung jawab dalam pengecekan keaslian sebuah makanan, minuman, obat-obatan, dan kosmetik di Indonesia memiliki masalah yaitu banyaknya kesalahan input pada data yang menyebabkan data yang tersimpan didalam database memiliki format data yang berbeda-beda dan tidak terisi atau masih kosong, kemudian untuk mendapatkan hasil *cleansing* yang diinginkan perusahaan, belum ada aturan khusus untuk memfilter data yang akan dilakukan proses *cleansing*, maka dari itu dibuatlah modul *cleansing* secara *generic*, modul ini akan membantu perusahaan untuk menyeragamkan *pattern* NIE (Nomor Izin Edar) yang menjadi patokan bahwa produk sudah terdaftar di BPOM, dengan cara melakukan sebuah proses *cleansing* secara fleksibel dimana user bisa memilih modul-modul yang

ada pada proses *data cleansing* untuk memproses NIE. Perusahaan bisa lebih spesifik untuk mendapatkan data hasil pembersihan dan diharapkan dapat meningkatkan pengambilan keputusan selanjutnya. Selain itu perubahan aturan-aturan bisnis yang sering terjadi mempengaruhi format penamaan produk yang terdaftar. Akibatnya data menjadi tidak akurat dan kualitas kurang baik. Kesalahan input yang terjadi juga akan berpengaruh pada penamaan pada produk dan akan menimbulkan persepsi berbeda untuk suatu produk. Masalah-masalah ini akan menimbulkan masalah berkelanjutan karena data yang masuk akan meningkat di masa depan. Maka dari itu diperlukan pengembangan *data quality management* untuk menjaga dan meningkatkan kualitas data agar lebih baik (Dwiandriani, 2017).

I.2 Rumusan Masalah

Berdasarkan uraian masalah yang telah dijelaskan pada latar belakang, maka dapat diperoleh permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana proses perancangan dan desain komponen modul pemrosesan *data cleansing* secara *generic* pada *data quality management* untuk menunjang perbaikan kualitas data?
2. Bagaimana mengembangkan dan menyesuaikan penelitian terkait untuk mengimpleptasikan modul *data cleansing* secara *generic* pada *data quality management dashboard*?

I.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Melakukan proses perancangan dan pengembangan modul *data cleansing* secara *generic* pada aplikasi *data quality management* untuk mendukung proses perbaikan kualitas data.
2. Mengimplementasikan dan melakukan penyesuaian dengan algoritma penelitian sebelumnya terkait *data cleansing* untuk mengembangkan *data quality management dashboard*.

I.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi manfaat secara keilmuan dan manfaat secara teknikal. Manfaat keilmuan yang diharapkan adalah dapat menerapkan modul secara *generic* dari arsitektur *data cleansing* yang di integrasikan dengan

konsep *data quality management* berbasis *open source*, yang mana pada penelitian tersebut dapat terus dimanfaatkan sebagai solusi dalam menangani permasalahan data yang ada.

Kemudian manfaat secara teknis diharapkan dengan adanya tugas akhir ini dapat berkontribusi untuk membantu menyelesaikan permasalahan yang dihadapi oleh perusahaan saat ini, dengan semakin banyaknya perusahaan mulai melakukan tata kelola data, diharapkan perusahaan dapat memanfaatkan pengelolaan *data quality management* dengan teknologi *open source* ini untuk menekan biaya dan memanfaatkan secara maksimal *output* yang diperoleh dari hasil penelitian ini.

I.5 Batasan Masalah

Batasan dalam melakukan penelitian ini adalah sebagai berikut:

1. Studi kasus yang digunakan untuk penelitian dilakukan pada BPOM.
2. Penelitian ini hanya dilakukan pada bidang *data cleansing*.
3. Membuat sebuah modul *data cleansing* dengan tujuan mempermudah proses pembersihan agar pemrosesan lebih fleksibel.
4. Implementasi penerapan *data cleansing* menggunakan *open source tool*.
5. Algoritma pada penelitian ini berfokus pada *data cleansing pattern* dan *data null*.
6. Tabel yang digunakan pada penelitian ini adalah *tb_ereg_merk* berisi nomor izin edar dari merk dan *tb_ereg_pattern* berisi nomor izin edar (NIE) obat dan makanan.
7. *Cleansing data pattern* mengacu pada metode *pattern frequency analysis* yang dilakukan untuk menyeragamkan pola data Nomor Izin Edar. *Cleansing data null* berfokus untuk mengisi kolom yang masih kosong dan mengisi dengan kolom referensi yang ada.

I.6 Sistematika Penulisan

Sistematika penulisan ini terbagi menjadi beberapa bab dari pokok pembahasan, secara umum dapat dijabarkan sebagai berikut:

Bab I Pendahuluan

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

Bab II Tinjauan Pustaka

Pada bab ini berisi penjelasan kajian – kajian literatur pendukung untuk riset dan beberapa *related work* yang pernah dilakukan oleh peneliti sebelumnya.

Bab III Metodologi Penelitian

Pada bab ini berisi penjelasan mengenai metodologi penelitian. Metodologi penelitian membahas mengenai konseptual dan sistematika dari penelitian yang dilakukan. Pemilihan dan penjelasan dari metode yang digunakan akan dijabarkan pada bab ini

Bab IV Analisis Sistem

Pada bab ini berisi tentang model dari sistem atau penelitian yang akan dilakukan.

Bab V Implementasi Dan Pengujian

Pada bab ini berisi tentang implementasi pembuatan logika, pengujian, menganalisa dari hasil analisis dan evaluasi.

Bab VI Kesimpulan Dan Saran

Pada bab ini dijelaskan kesimpulan hasil dari penelitian yang dilakukan dan saran yang dapat dipertimbangkan untuk penelitian berikutnya.