

Perancangan Kualitas Data di Bank XYZ dengan Metode Data Quality Management Steps

Designing Data Quality In XYZ Bank Using Data Quality Management Steps Method

San Karya¹

¹Information System, School of Information System, Bina Nusantara University

Jl. KH. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta Barat, Jakarta

Telp : (021) 534 5830

E-mail : san.karya@binus.ac.id¹

Abstract

As technology advances and time progresses, Bank XYZ faces an increasing need for high-quality data to support business processes and manage large volumes of data. This study presents a data quality design aimed at providing a solution for Bank XYZ by analyzing the cleanliness of individual customer data using the data quality management steps method and Talend data quality tools for data profiling. The design employs a data quality matrix for data measurement, which is presented through a visual dashboard as the final output. The data quality dashboard is intended to provide clear insights into the data quality levels at Bank XYZ, facilitating understanding and supporting decision-making processes.

Keywords: Data Quality, Data Quality Management Steps, Data Quality Matrix, Talend Data Quality, Dashboard

Abstrak

Seiring dengan perkembangan teknologi dan berjalannya waktu terjadi peningkatan akan kebutuhan di bank XYZ, terutama kebutuhan akan kualitas data yang baik untuk menunjang proses bisnis dan pengelolaan volume data yang besar. Perancangan kualitas data disini hadir sebagai solusi untuk bank XYZ dengan tujuan untuk membantu melakukan analisa terhadap tingkat kebersihan data nasabah individu menggunakan metode data quality management steps, serta menggunakan tools Talend Data Quality dalam melakukan profiling data. Perancangan ini menggunakan matriks untuk pengukuran data yang disajikan dalam bentuk visualisasi dashboard sebagai hasil akhir dari perancangan ini. Dengan dashboard kualitas data diharapkan mampu memberikan informasi mengenai tingkat kualitas data di bank XYZ agar mudah dipahami dan mendukung dalam pengambilan keputusan.

Kata kunci: Kualitas Data, Tahapan Manajemen Kualitas Data, Matrik Kualitas Data, Talend Kualitas Data, Dashboard

1. Pendahuluan

Pada era digital ini, teknologi berkembang pesat dan telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari [1]. Perkembangan ini membawa peluang dan tantangan signifikan dalam dunia ekonomi dan bisnis, baik secara lokal maupun internasional [2]. Teknologi kini digunakan untuk mengembangkan dan meningkatkan proses bisnis serta mendukung perekonomian [3]. Salah satu aspek penting dalam perkembangan teknologi adalah data, yang kini dianggap sebagai 'emas baru' karena mengandung informasi berharga [4].

Data adalah fakta dan angka yang dapat digunakan untuk menyusun informasi, sedangkan informasi adalah hasil pengolahan data untuk keperluan tertentu. Banyak

perusahaan berlomba-lomba mengumpulkan dan mengolah data untuk meningkatkan kualitas bisnis dan menciptakan persaingan ekonomi [5]. Oleh karena itu, data menjadi sangat berharga, dengan ratusan hingga jutaan data yang dianalisis untuk meningkatkan keputusan manajemen perusahaan.

Salah satu elemen penting dalam tata kelola data adalah kualitas data [6]. Manajemen kualitas data mencakup integritas, konsistensi, dan tingkat kepercayaan data. Data berkualitas tinggi mendukung pengambilan keputusan, mengurangi anomali, dan meningkatkan efisiensi organisasi [7]. Kualitas data mencakup akurasi, kelengkapan, ketepatan waktu, dan konsistensi yang relevan dengan kebutuhan bisnis.

Dalam dunia perbankan, seperti di bank XYZ, manajemen kualitas data sangat penting. Bank memiliki data nasabah, data transaksi, dan data produk yang harus dikelola dengan baik. Data ini juga harus dilaporkan secara rutin kepada pihak berwenang seperti OJK dan BI [8]. Kualitas data yang buruk dapat merugikan bank, baik dari segi kepatuhan hukum maupun reputasi. Oleh karena itu, diperlukan perancangan kualitas data yang baik untuk memastikan data yang dilaporkan akurat dan sesuai dengan peraturan.

Proses perancangan kualitas data melibatkan beberapa tahapan, mulai dari menentukan definisi dan tujuan hingga pengembangan data quality dan output laporan kualitas data [9]. Hasil akhir dari perancangan ini adalah laporan kualitas data yang salah satunya berupa visualisasi dashboard. Dashboard ini merepresentasikan kualitas data dari berbagai aspek, seperti konsistensi dan keakuratan data, yang memudahkan evaluasi dan pelaporan [10]. Dengan perancangan kualitas data yang baik, diharapkan bank XYZ dapat mengimplementasikan kualitas data yang membantu proses bisnis dan mempersiapkan data untuk analisis di masa depan. Implementasi ini juga dapat menjadi contoh bagi bank lain dalam meningkatkan kualitas data untuk mendukung bisnis yang berjalan.

2. Tinjauan Pustaka

Perancangan kualitas data menjadi krusial dalam pengelolaan data pada institusi keuangan seperti bank XYZ. Data yang diolah dalam perancangan ini berasal dari data warehouse bank, yang bersumber dari tiga core banking bank XYZ. Data ini mencakup data nasabah (master customer) sebanyak 1000 row, yang mewakili sekitar 5% dari total data master nasabah di bank XYZ. Data ini telah diacak (scrambled) untuk mematuhi regulasi ketat yang diterapkan oleh pihak bank.

Kualitas data adalah ukuran kondisi data yang didasarkan pada faktor-faktor seperti akurasi, kelengkapan, konsistensi, reliabilitas, dan kebaruan [11]. Mengukur tingkat kualitas data membantu organisasi mengidentifikasi kesalahan data yang perlu diselesaikan dan menilai apakah data dalam sistem IT mereka sesuai untuk memenuhi tujuan yang dimaksudkan. Menurut DAMA DMBOK, istilah kualitas data mengacu pada karakteristik yang terkait dengan data berkualitas tinggi dan proses yang digunakan untuk mengukur atau meningkatkan kualitas data. Data berkualitas tinggi harus memenuhi ekspektasi dan kebutuhan konsumen data.

Penekanan pada kualitas data dalam sistem perusahaan telah meningkat seiring dengan semakin eratnya hubungan antara pemrosesan data dan operasi bisnis. Organisasi semakin banyak menggunakan analisis data untuk mendorong keputusan bisnis. Kualitas data yang buruk akan mengakibatkan kualitas informasi yang

buruk pula, yang pada gilirannya menyebabkan kurangnya pengetahuan yang dapat ditindaklanjuti dalam operasi bisnis. Kesalahan dalam pengetahuan ini dapat berdampak signifikan pada bisnis yang berkaitan.

Data yang berkualitas buruk dapat mengakibatkan konsekuensi bisnis yang serius. Kualitas data yang rendah sering dianggap sebagai sumber gangguan operasional, analisis yang tidak akurat, dan kesalahpahaman dalam penyusunan strategi bisnis. Terdapat dua definisi utama mengenai kualitas data: pertama, data dinyatakan berkualitas tinggi jika sesuai dengan tujuan penggunaan yang dimaksudkan; kedua, data dinyatakan berkualitas tinggi jika secara tepat dan akurat merepresentasikan realitas yang dijelaskan oleh data tersebut.

Talend Open Studio for Data Quality adalah perangkat lunak yang berguna untuk analisis data dan profiling data. Profiling data, menurut Wikipedia, adalah proses memeriksa data yang tersedia dari sumber informasi yang ada (misalnya, database atau file) dan mengumpulkan statistik atau ringkasan informatif tentang data tersebut. Profiling data digunakan untuk menilai apakah data yang ada dapat dengan mudah digunakan untuk tujuan tertentu, menilai kualitas data, serta memastikan apakah data tersebut sesuai dengan standar atau pola tertentu. Talend Data Quality menyediakan berbagai fitur untuk validasi data dengan merancang aturan berbasis sintaks SQL maupun lainnya [12].

Hasil penelitian sebelumnya yang berjudul “Tinjauan umum kualitas data: Meneliti Dimensi, Anteseden, dan Dampak Kualitas Data” yang disusun oleh Peigong Li, Zhenxing Lin, Jingran Wang, Yi Liu dan Stavros Sindakis ditahun 2024. Penelitian ini mendapatkan hasil bahwa kualitas data memiliki lima dimensi penting, yaitu akurasi, kelengkapan, konsistensi, ketepatan waktu dan relevansi. Penelitian ini juga menghasilkan dua faktor penting yang mempengaruhi kualitas data, yaitu tekanan waktu dan pengalaman pengguna data [13].

Dengan memahami dan menerapkan konsep-konsep kualitas data, organisasi dapat memastikan data yang mereka gunakan akurat, lengkap, dan konsisten, sehingga dapat mendukung pengambilan keputusan yang lebih baik dan meningkatkan efisiensi operasional.

3. Metodologi Penelitian

Metode perancangan kualitas data yang diterapkan dalam penelitian ini mengacu pada buku DAMA DMBOK 2nd Edition (2017) Data Management Body of Knowledge. Prosesnya menggunakan langkah-langkah manajemen kualitas data sebagai berikut:

3.1 Define High Quality Data

Tahap ini bertujuan untuk memahami kebutuhan dan permasalahan yang dihadapi perusahaan terkait kualitas data. Definisi kualitas data, standar, serta toleransi terhadap kualitas data yang buruk harus ditetapkan sesuai dengan kebijakan perusahaan. Beberapa poin penting dalam tahap ini meliputi:

- a. Dampak kualitas data yang rendah terhadap operasi dan strategi bisnis.
- b. Definisi kualitas data yang tinggi dari perspektif perusahaan.
- c. Toleransi perusahaan terhadap kualitas data yang rendah.

3.2 Define Data Quality Strategy

Kualitas data harus sejalan dengan strategi bisnis perusahaan. Merancang framework kualitas data membantu dalam pencapaian strategi dengan mengukur progres dan dampak yang dihasilkan. Tahap ini mencakup:

- a. Memahami dan memprioritaskan kebutuhan bisnis.
- b. Menentukan standar organisasi terkait kualitas data.
- c. Mengidentifikasi dan memprioritaskan isu-isu yang pernah terjadi.

3.3 Identify Critical Data and Business Rules

Data diprioritaskan berdasarkan faktor seperti persyaratan peraturan, nilai finansial, dan dampak terhadap pelanggan. Dalam konteks perbankan, data pelanggan merupakan data kritis yang harus sesuai dengan aturan yang ada, seperti nama pelanggan yang tidak boleh salah dan tanggal lahir yang tidak boleh kosong. Tahap ini meliputi:

- a. Mendefinisikan aturan bisnis dalam perusahaan.
- b. Menentukan atribut yang menjadi elemen data kritis.

3.4 Perform Initial Data Quality Assessment

Tahap ini melibatkan analisis struktur data yang akan diolah, termasuk hubungan data dan ketergantungan data. Identifikasi masalah-masalah utama dilakukan pada tahap ini. Beberapa poin penting adalah:

- a. Menentukan tujuan dari assessment.
- b. Mengidentifikasi risiko dan dampak bisnis dari data tersebut.
- c. Mendeskripsikan struktur dan hubungan data.

3.5 Identify and Prioritize Improvement

Pada tahap ini, identifikasi dan prioritas peningkatan dilakukan dengan profiling untuk memahami permasalahan. Profiling data membantu mengidentifikasi masalah, namun tidak dapat mengidentifikasi akar penyebab dan dampaknya terhadap proses bisnis. Masukan dari stakeholder diperlukan untuk menentukan dampak yang terjadi.

3.6 Define Goals for Data Quality Improvement

Setelah analisis profil data, tujuan dan rencana peningkatan kualitas data ditetapkan. Remediasi dan rencana peningkatan harus memperhitungkan hasil cepat dengan biaya rendah serta perubahan strategis jangka panjang untuk mengatasi akar penyebab masalah dan mencegah masalah serupa di masa depan.

3.7 Develop and Deploy Data Quality Operations

Untuk mempertahankan kualitas data, rencana yang memungkinkan tim mengelola aturan dan standar kualitas data harus diterapkan. Tahap ini mencakup beberapa sub-tahapan:

- a. Manage Data Quality Rules: Mengelola aturan dan ketentuan mengenai kualitas data.
- b. Measure and Monitor Data Quality: Mengukur hasil profiling berdasarkan aturan bisnis yang ditetapkan.
- c. Develop Operational Procedure for Managing Data Issues: Mendiagnosa masalah kualitas data, identifikasi penyebab, dan evaluasi dampaknya.
- d. Establish Data Quality Service Level Agreements: Menetapkan ekspektasi organisasi untuk respons dan perbaikan masalah kualitas data.
- e. Develop Data Quality Reporting: Merancang laporan akhir yang mencakup data quality dashboard dan kesimpulan dari hasil perancangan.

Dalam penelitian ini, tools Talend Data Quality digunakan untuk proses profiling data dan analisa business rules. Data nasabah bank XYZ dianalisis menggunakan fitur seperti column profiling dan business rules analysis. Hasil profiling data disajikan dalam bentuk dashboard yang dirancang menggunakan Tableau, dengan sumber data dari file excel yang memuat hasil profiling.

4. Hasil dan Pembahasan

Hasil penerapan dari langkah-langkah manajemen kualitas data berdasarkan DAMA DMBOK 2nd Edition (2017) sebagai berikut:

4.1 Define High Quality Data

PT Bank XYZ adalah perusahaan perbankan yang telah lama beroperasi di Indonesia, mengelola data sensitif dengan risiko tinggi yang diawasi ketat oleh otoritas seperti Otoritas Jasa Keuangan (OJK). Data berkualitas tinggi di Bank XYZ didefinisikan sebagai data yang akurat, bersih, dan memenuhi kriteria regulator, sehingga dapat dipercaya untuk pelaporan, modeling, dan analisis. Kualitas data yang buruk dapat menimbulkan risiko besar, seperti kesalahan pelaporan yang mengakibatkan penalti dan merusak reputasi bank.

Dalam pengolahan data nasabah, kesalahan sering terjadi, baik dari sistem maupun human error. Bank XYZ tidak mentoleransi data yang kotor dalam pelaporan kepada regulator. Namun, untuk kebutuhan internal seperti pemasaran, toleransi terhadap kesalahan data dapat berbeda-beda. Sebagai contoh, kesalahan pada data alamat atau nomor telepon dapat menghambat kegiatan pemasaran produk bank.

4.2 Define Data Quality Strategy

Strategi kualitas data Bank XYZ tidak hanya berfokus pada pelaporan ke regulator tetapi juga memenuhi kebutuhan bisnis internal. Data nasabah digunakan dalam berbagai proses bisnis, dan kualitas data yang buruk dapat menghambat operasional dan pengambilan keputusan yang tepat. Strategi yang dirancang meliputi:

- a. Kebutuhan Pelaporan Regulator: Data dilaporkan kepada OJK dan Bank Indonesia. Data diolah dari tiga sumber core banking dan dimasukkan ke dalam data warehouse sebelum dilaporkan melalui Sistem Layanan Informasi Keuangan (SLIK).
- b. Kebutuhan Internal: Data digunakan untuk analisis internal dan pengambilan keputusan di berbagai departemen, termasuk pemasaran dan layanan pelanggan.

4.3 Identify Critical Data and Business Rules

Tahap ini menentukan data kritis dan aturan bisnis dalam pelaporan di Bank XYZ. Critical data dan business rules ditetapkan oleh OJK dan berlaku untuk semua bank di Indonesia. Data kritis tidak semua atribut ditampilkan karena privasi dan ketentuan OJK. Berikut atribut critical data dan rules di Bank XYZ:

1. Nomor CIF:
 - a. Harus unik dan tidak boleh kosong
 - b. Berisi 8 digit angka
2. Kode Tipe Customer dan Deskripsi:
 - a. Mandatory, kode dan deskripsi harus sesuai (contoh: "I" untuk Individual)
3. Jenis Kelamin:

- a. Mandatory, hanya berisi kode "F" atau "M"
4. Nama Nasabah:
 - a. Mandatory dan unik
5. Tempat dan Tanggal Lahir:
 - a. Mandatory, tidak boleh kosong
6. Alamat dan Kode Pos:
 - a. Mandatory, informasi lengkap termasuk kelurahan dan provinsi
7. NPWP:
 - a. Berisi 15 digit, unik, dan tidak mengandung karakter selain angka dan alfabet
8. Pendidikan dan Status Pekerjaan:
 - a. Mandatory, berisi kode dan deskripsi yang sesuai
9. Status Perkawinan:
 - a. Opsional, berisi salah satu dari lima kategori
10. Legal ID No dan Tipe Dokumen:
 - a. Mandatory, berisi 16 digit untuk KTP
11. Kode dan Nama Negara:
 - a. Mandatory, berisi kode dua digit dan deskripsi

4.4 Perform Initial Data Quality Assessment

Tahap ini meliputi penentuan tujuan assessment, deskripsi data, dan identifikasi risiko serta dampaknya terhadap bisnis Bank XYZ. Tujuannya adalah memastikan perancangan sesuai kebutuhan bank, serta menilai kebutuhan tambahan untuk keberhasilan implementasi. Data nasabah bersumber dari tiga core banking system dan diakses menggunakan Oracle SQL Developer. Atribut data meliputi:

1. Nomor CIF:
 - a. Primary key, harus unik
2. Tipe Customer dan Deskripsi:
 - a. Kode dan deskripsi jenis nasabah
3. Jenis Kelamin:
 - a. Jenis kelamin nasabah
4. Nama Customer, Tgl Lahir, Tempat Lahir:
 - a. Informasi identitas nasabah, mandatory
5. Alamat, Kode Pos:
 - a. Informasi alamat lengkap, mandatory
6. NPWP, Status Pekerjaan, Pendidikan:
 - a. Informasi pajak, status pekerjaan, dan pendidikan nasabah

Risiko kualitas data yang buruk meliputi duplikasi nomor CIF, kesalahan nama nasabah, dan informasi alamat yang tidak lengkap. Data yang tidak akurat dapat menghambat validasi, pelaporan, dan berbagai operasi bank, serta menimbulkan risiko kepatuhan terhadap regulasi.

Profiling data dilakukan dengan tools Talend Data Quality, menganalisis kolom seperti null count, unique count, dan pattern analysis. Hasil profiling disajikan dalam bentuk dashboard dengan visualisasi diagram batang menggunakan tools Tableau. Hal ini membantu

memonitor dan meningkatkan kualitas data secara efektif.

4.5 Identify and Prioritize Improvement

Analisis telah dilakukan terhadap data yang telah dijelaskan sebelumnya menggunakan metode profiling data. Profiling data adalah proses ekstraksi informasi tertentu dari data sesuai kebutuhan. Dalam penelitian ini, menggunakan alat Talend Data Quality dengan fitur analisis profil kolom dan analisis aturan bisnis.

1. Nomor CIF

Berikut ini merupakan hasil profiling dari setiap rulesnya:

- CIF Number harus bersifat unik, artinya dalam satu tabel tidak ada yang sama atau duplikat. Atribut CIF Number terdapat 1000 row yang unik artinya tidak ada yang duplikat dalam kolom ini.
- CIF Number harus diisi, artinya kolom ini tidak boleh ada data yang null. Hasilnya menggambarkan bahwa tidak ada baris yang null, artinya tidak ada row di atribut CIF Number yang kosong.
- CIF Number harus berisi 8 digit value. Hasil menunjukkan bahwa seluruh row CIF Number sudah sesuai dengan rules yaitu berisi 8 digit.

Rekomendasi: Sistem harus terintegrasi antara core banking untuk mencegah duplikasi CIF.

2. Kode Tipe Customer

Berikut ini merupakan hasil berdasarkan rules dari atribut Customer Type Code:

- Kolom ini tidak boleh kosong. Berdasarkan hasil profiling dinyatakan atribut ini tidak ada yang kosong dengan null count sebanyak nol.
- Kolom ini hanya berisi kode sebanyak satu sampai dua digit saja. Hasil menunjukkan bahwa semua row sudah sesuai dengan ketentuan yang berlaku.
- Kolom ini hanya berisi nilai 'I' dimana menandakan hanya nasabah individual saja. Hasil menunjukkan bahwa ada 3 row yang tidak sesuai sedangkan sisanya 997 row sesuai.

Rekomendasi: Pembagian kategori antara nasabah individu dan non-individu harus dilakukan pada sistem.

3. Jenis Kelamin

Berikut ini merupakan hasil berdasarkan rules dari atribut Gender:

- Kolom ini tidak boleh kosong. Hasilnya menunjukkan bahwa hanya 992 row yang sesuai dengan ketentuan yaitu tidak boleh null
- Kolom ini hanya boleh berisi satu digit kode. Hasilnya menunjukkan bahwa hanya sebanyak 992 row yang sesuai dengan ketentuan.
- Kolom ini hanya berisi dua opsi yaitu F (Female) dan M (Male). Berdasarkan hasil profiling menunjukkan bahwa terdapat 992 row sudah sesuai.

Rekomendasi: Data gender harus diverifikasi dengan KTP nasabah atau langsung terhubung dengan situs dukcapil untuk verifikasi.

4. Nama Nasabah

Berikut ini merupakan hasil berdasarkan rules dari atribut Nama Nasabah:

- Kolom ini tidak boleh kosong. Berdasarkan hasil analisa profiling tidak ada row yang null untuk kolom nama nasabah.
- Nama nasabah harus unik tidak boleh sama. Berdasarkan hasil analisa profiling dinyatakan bahwa hanya 512 row yang memenuhi syarat.
- Nama nasabah harus terdiri dari alphabet tidak boleh ada angka. Berdasarkan hasil analisa profiling dikatakan bahwa semua row nama nasabah sudah sesuai dengan ketentuan.

Rekomendasi: Verifikasi data dengan data dukcapil sebelum input.

5. Tempat Lahir

Berikut ini merupakan hasil berdasarkan rules dari atribut Tempat Lahir:

- Kolom ini tidak boleh kosong. Berdasarkan hasil profiling pada gambar 11 terdapat 3 row yang null sehingga sebanyak 997 (99.70%) sisa row dinyatakan sesuai dengan rules.
- Kolom ini hanya berisikan alphabet saja. Berdasarkan gambar 12 dinyatakan bahwa hanya 992 row yang sesuai.

Rekomendasi: Verifikasi data dengan data dukcapil sebelum input.

6. Nomor Identitas Resmi

Berikut ini merupakan hasil berdasarkan rules dari atribut Legal ID:

- Berdasarkan hasil profiling dapat dilihat bahwa tidak terdapat row yang null dalam kolom ini.
- Berdasarkan hasil profiling hanya 278 row baris yang unik.
- Berdasarkan hasil profiling dapat dilihat bahwa terdapat 973 row yang sesuai yaitu harus mengandung 16 digit angka.

- Berdasarkan hasil profiling dinyatakan bahwa sebanyak 851 row yang null dalam kolom ini.
- Kolom ini harus unik karena setiap NPWP nasabah pasti berbeda. Berdasarkan hasil profiling hanya 103 row yang bersifat unik.
- Kolom ini harus diisi dengan 15 (lima belas) digit NPWP. Berdasarkan hasil profiling bahwa sebanyak 137 row sesuai dengan ketentuan.

Rekomendasi: Inputan data NPWP harus divalidasi di sistem terkait agar NPWP harus diisi dan isinya harus unik.

Rekomendasi: Verifikasi data dengan data dukcapil sebelum input.

7. Tipe Dokumen

Berikut ini merupakan hasil berdasarkan rules dari atribut Doc Type:

- Kolom ini mandatory artinya tidak boleh kosong. Berdasarkan hasil profiling dapat dilihat bahwa terdapat 59 row yang null, dan 941 row yang tidak null.
- Kolom ini hanya boleh berisikan kode dengan nilai '1' saja dimana menandakan KTP. Berdasarkan hasil profiling dikatakan bahwa hanya sebanyak 168 row yang sesuai.

Rekomendasi: Hanya beri dua pilihan jenis dokumen saat input.

Melalui hasil profiling, dapat disimpulkan bahwa beberapa atribut memerlukan perhatian lebih dalam validasi data untuk meningkatkan kualitas data nasabah. Langkah-langkah perbaikan dan verifikasi yang diusulkan diharapkan dapat meningkatkan integritas dan keakuratan data nasabah di Bank XYZ.

4.6 Define Goals for Data Quality Improvement

Dalam tahap ini, tujuan dari perancangan kualitas data ditentukan untuk meningkatkan kualitas data di Bank XYZ. Berdasarkan hasil data profiling sebelumnya, ditemukan bahwa beberapa atribut kritis masih tidak sesuai sehingga menyebabkan kualitas data yang buruk pada tabel data nasabah. Berikut adalah tujuan peningkatan kualitas data berdasarkan hasil dari profiling data:

- Tujuan dari perancangan ini salah satunya adalah agar tabel CIF nasabah ini memiliki tingkat kualitas data yang bersih pada saat pelaporan ke pihak OJK melalui SLIK. Data nasabah terutama atribut CIF Number menjadi prioritas utama sehingga atribut tersebut harus bersifat unik dan tidak terjadi duplikat, apabila terjadi duplikat maka akan membuat data menjadi kurang valid dalam pelaporan ke SLIK.
- Perancangan kualitas data ditujukan tidak hanya untuk meningkatkan kualitas data saja namun juga dari sisi data governance dimana kualitas data ini merupakan salah satu bagian dari data governance itu sendiri, sehingga jika dilihat dari segi data governance kualitas data akan dapat memberikan analisa atau masukan terhadap peraturan, rules, dan ketentuan lain dalam bagian tata kelola data sehingga dapat menjadi saran dan pedoman bagi divisi data governance di bank XYZ ini.
- Perancangan ini diharapkan dapat meningkatkan efektivitas dan efisiensi proses bisnis di perusahaan bank XYZ serta dapat membantu dalam pengambilan keputusan dari pihak manajerial bank XYZ.

8. Alamat Dati 2

Berikut ini merupakan hasil berdasarkan rules dari atribut Doc Type:

- Kolom ini tidak boleh kosong. Berdasarkan hasil profiling dinyatakan bahwa terdapat 49 row yang null dalam kolom dati 2 address.
- Kolom ini hanya berisi 4 digit angka. Berdasarkan hasil profiling dinyatakan bahwa hanya sebesar 951 row yang sesuai.
- Kolom ini berisi empat digit angka dimana angka keempat angka tersebut adalah angka satu sampai sembilan. Berdasarkan hasil profiling dinyatakan bahwa tidak ada row yang sesuai.

Rekomendasi: Validasi alamat berdasarkan alamat yang terdaftar di KTP.

9. NPWP

Berikut ini merupakan hasil berdasarkan rules dari atribut NPWP:

4.7 Develop and Deploy Data Quality Operations

4.7.1 Manage Data Quality Rules

Dalam tahap ini, penulis menentukan matriks yang digunakan dalam perancangan kualitas data dan melakukan analisa untuk beberapa rules di Bank XYZ mengenai kualitas data. Berikut adalah beberapa analisa terhadap rules yang ada:

- a. Untuk ketentuan dan rules atribut CIF beberapa sudah tepat dengan hasil kualitas data yang baik, mungkin hanya ditambahkan rules seperti atribut CIF harus berisi delapan digit agar panjang data lebih tepat lagi, dan tentunya mencegah atribut digit yang kurang.
- b. Untuk atribut customer type code masih ada nilai yang tidak sesuai seperti masih ada non individual customer, diharapkan untuk kolom ini ditambahkan rules berupa hanya boleh melakukan input dengan jenis customer individual, begitupun dengan sebaliknya agar lebih mudah untuk dilakukan analisa untuk perkategori nasabah.
- c. Untuk atribut data 2 address masih banyak data yang tidak akurat seperti berisikan empat digit angka '0000' atau '1111', diharapkan kolom ini ditambahkan rules untuk memvalidasi data input agar sesuai dengan data mutlak setiap daerah di Indonesia.

Setiap hasil dari masing-masing atribut akan dikategorikan ke dalam matriks kualitas data. Matriks ini selain menjadi tolak ukur dalam melihat seberapa tinggi kualitas data, juga menjadi elemen atau dimensi dalam kualitas data. Matriks yang digunakan dalam tahap ini adalah:

1. Completeness: Mengukur persentase data dari kolom yang sudah diisi atau tidak kosong.
2. Accuracy: Mengukur seberapa tepat atau akurat nilai yang ada dalam sebuah atribut sesuai dengan ketentuan.
3. Uniqueness: Mengukur seberapa unik nilai suatu data.
4. Conformity: Mengukur apakah panjang value sudah sesuai dengan ketentuan.
5. Data Quality Score: Merupakan gabungan hasil rata-rata dari matriks completeness, accuracy, conformity, dan uniqueness.

4.7.2 Manage Data Quality Rules

Dalam sub-tahap ini, penulis melakukan pengukuran terhadap kualitas data berdasarkan hasil profiling dengan matriks yang telah dibuat. Tujuannya adalah mendeskripsikan secara jelas elemen atau matriks dalam kualitas data. Contoh pengukuran dari beberapa

atribut menunjukkan bahwa beberapa hasil masih di bawah 100%, yang artinya perlu perbaikan.

4.7.3 Develop Operational Procedures for Managing Data Issues

Pada tahap ini, penulis memberikan saran dan masukan untuk mengurangi atau mencegah masalah berdasarkan hasil profiling data sebelumnya. Hal yang perlu diperhatikan adalah prosedur untuk memperbaiki data yang kotor. Setelah mendapatkan informasi dari hasil profiling bahwa terdapat 30% data yang kotor, informasi ini akan disampaikan kepada data owner atau pihak berwenang untuk melakukan perubahan atau update data. Dengan update data yang telah diperbaiki, kualitas data akan semakin meningkat dari waktu ke waktu, terutama dilihat dari data quality score.

4.7.4 Establish Data Quality Service Level Agreements

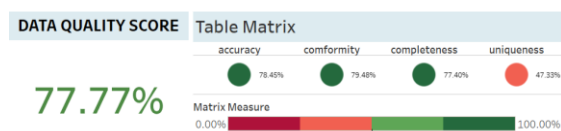
Tahap ini menentukan ketentuan mengenai tingkat kualitas data berdasarkan hasil profiling. Dari hasil matriks pada setiap kolom di tabel akan di rata-rata berdasarkan tiap matriks sehingga akan terlihat hasil total matriks dari seluruh tabel, hasil yang didapat antara lain accuracy 78.45%, conformity 79.48%, completeness 77.40%, dan uniqueness 47.33%. Dapat disimpulkan bahwa matriks uniqueness menjadi yang paling terkecil, artinya tingkat keunikan dari beberapa atribut yang harus unik dalam tabel masih sangat rendah, dan memang data tersebut secara hasil persentase belum dapat diterima. Hal yang sama juga dapat dilihat dari hasil matriks yang lain bahwa persentase yang di dapat masih dibawah 100%. Data yang masih kotor akan dikembalikan ke data owner atau pihak operasional untuk diperbaiki kembali, sehingga persentase data quality score dan matriks lainnya akan semakin membaik di masa mendatang.

4.7.5 Develop Data Quality Report

Tahap ini merancang data quality report dalam bentuk dashboard yang menyajikan informasi hasil profiling dari semua atribut atau kolom dalam tabel nasabah. Hasil profiling dimasukkan ke dalam Excel dan kemudian divisualisasikan menggunakan tools Tableau.

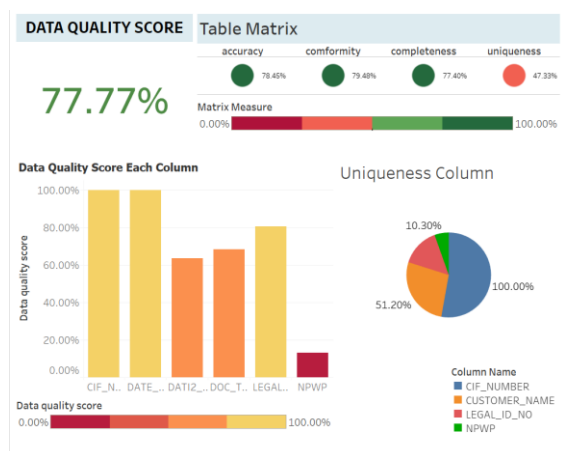
Gambar 1 menunjukkan bagian dari keseluruhan dashboard kualitas data. Pada bagian sebelah kiri, angka 77.77% menggambarkan hasil data quality score dari keseluruhan tabel nasabah. Di bagian kanan, terdapat tabel matriks yang memuat empat matriks utama beserta masing-masing nilainya, yang diwakili dengan ukuran warna. Matriks-matriks tersebut meliputi: accuracy dengan nilai 78.45%, conformity

dengan nilai 79.48%, completeness dengan nilai 77.40%, dan uniqueness dengan nilai 47.33%.



Gambar 1. Dashboard Table Matrix

Gambar 2 menyajikan keseluruhan tampilan dashboard yang terdiri dari empat grafik. Melalui visualisasi dashboard ini, pengguna dapat melihat data quality score secara keseluruhan serta memperoleh informasi yang lebih rinci mengenai data quality score untuk setiap atribut.



Gambar 2. Dashboard kualitas data

5. Kesimpulan

5.1 Simpulan

Berdasarkan hasil perancangan kualitas data dari tahap awal hingga tahap penyusunan data quality report, beberapa kesimpulan dapat ditarik dari keseluruhan perancangan ini:

1. Perancangan kualitas data di Bank XYZ dengan menggunakan metode data quality management step telah berhasil memberikan gambaran yang komprehensif mengenai tingkat kualitas data di Bank XYZ. Setiap langkah dalam metode ini memberikan analisis mendalam terhadap atribut per kolom dari data, mengidentifikasi banyaknya data yang masih kosong, kurang akurat, dan tidak unik. Hasil ini tidak hanya mencerminkan angka-angka tingkat kebersihan data atau atribut tertentu, tetapi juga memberikan beberapa masukan dan saran untuk perbaikan di masa mendatang,

khususnya yang berkaitan dengan tata kelola data dari segi kualitas data.

2. Hasil output dari perancangan ini berupa data quality report yang disajikan dalam bentuk visualisasi dashboard. Dashboard ini mempermudah dalam memantau tingkat kualitas data di Bank XYZ, khususnya untuk kategori nasabah individual. Beberapa metrik yang digunakan dalam dashboard ini mencakup completeness, accuracy, conformity, uniqueness, dan data quality score. Dengan demikian, dashboard ini diharapkan dapat mempermudah pihak yang membutuhkan informasi untuk tata kelola data di Bank XYZ serta mendukung proses pengambilan keputusan.

5.2 Saran

Berikut ini beberapa saran yang ingin disampaikan oleh penulis mengenai perancangan kualitas data ini:

1. Untuk kedepannya diharapkan dapat lebih memperluas ruang lingkup untuk perancangan kualitas data ini dengan menggunakan data-data lain di bank XYZ agar perancangan ini dapat mencakup seluruh proses bisnis di bank XYZ, dan juga lebih memenuhi seluruh kebutuhan dalam tata kelola data di bank XYZ.
2. Diharapkan agar tools Talend Data Quality yang penulis gunakan dapat melakukan generate hasil profiling ke dalam bentuk excel secara otomatis, jadi tidak secara manual lagi memasukan angka ke dalam excel, ataupun dapat menggunakan tools lain berbasis data profiling atau data quality yang dapat mengotomisasikan hasil sehingga akan mempercepat proses.

Daftar Rujukan

- [1] Qureshi, Z., 2022. How digital transformation is driving economic change. Brookings Institution.
- [2] Calvino and C. Criscuolo, F., 2019. Overcoming regulatory barriers: Challenges and opportunities for entrepreneurship in the digital age. Journal of Innovation and Entrepreneurship.
- [3] Brynjolfsson and A. McAfee, E., 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.
- [4] S. Kambatla, G. Kollias, V. Kumar, and A. Grama., 2014. Trends in big data analytics. Journal of Parallel and Distributed Computing, vol. 74 (7), pp. 2561-2573

- [5] T. McGuire, J. Manyika, and M. Chui, 2019. Why Big Data is the new competitive advantage. Ivey Business Journal.
- [6] M. Ovaska, 2021. Improving Data Quality and Governance. *Journal of Data and Information Quality*, 12 (4), pp. 23-36.
- [7] Medeiros, D.D., 2019. Big data and decision-making in healthcare: Challenges and opportunities. *Journal of Ambient Intelligence and Humanized Computing*, 10 (5).
- [8] Otoritas Jasa Keuangan, 2017. Peraturan Otoritas Jasa Keuangan Republik Indonesia Nomor 18/POJK.03/2017 Tentang Pelaporan dan Permintaan Informasi Debitur Melalui Sistem Layanan Informasi Keuangan. Available at: <https://ojk.go.id/id/kanal/perbankan/regulasi/peraturan-ojk/Pages/POJK-tentang-Pelaporan-dan-Permintaan-Informasi-Debitur-melalui-Sistem-Layanan-Informasi-Kuangan.aspx>
- [9] DAMA International., 2017. DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. Technics Publications.
- [10] S. Few., 2013. Information Dashboard Design: The Effective Visual Communication of Data. 2nd ed. O'Reilly Media.
- [11] Cai, Li, and Yangyong Zhu., 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14, pp. 1–10.
- [12] Talend., 2020. Talend Data Quality: User Guide. Talend Inc.
- [13] Peigong Li, Zhenxing Lin, Jingran Wang, Yi Liu dan Stavros Sindakis. 2024, Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *Journal of the Knowledge Economy*, 15, pp. 1159–1178.