PENERAPAN SISTEM *DATA CLEANING* DALAM MASTER DATA DENGAN MENGGUNAKAN ALGORITMA *DUPLICATE COUNT STRATEGY* (STUDI KASUS: PT XYZ)

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

FILDZAH ADRA ARIFAH

NIM 1132001017

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

JAKARTA

2021

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	**: Fildzah Adra Arifah**
NIM	**: 1132001017**
Tanda Tangan	**:** 
Tanggal	**: 1 Juli 2021**

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Fildzah Adra Arifah

NIM : 1132001017

Program Studi : Informatika

Fakultas : Teknik dan Ilmu Komputer

Judul Skripsi : Penerapan Sistem *Data Cleaning* dalam Master Data dengan Menggunakan Algoritma *Duplicate Count Strategy* (Studi Kasus: PT XYZ)

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

DEWAN PENGUJI

Penguji 1 : Sondang Sibuea, S.Kom., M.Kom. (.....)

Penguji 2 : Albert A. Sembiring, S.T, M.T. (.....)

Anggota : Prof. Dr. Hoga Saragih, S.T, M.T, IPM. (.....)

Anggota : Ihsan Ibrahim, S.T., M.T. (.....)

Ditetapkan di : Jakarta

Tanggal : 25 Agustus 2021

UNGKAPAN TERIMA KASIH

Puji dan syukur atas kehadirat Allah Subhanahu wata'ala atas rahmat-Nya dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir dengan judul "**Penerapan Sistem *Data Cleaning* dalam Master Data dengan Menggunakan Algoritma *Duplicate Count Strategy*** (Studi Kasus: PT XYZ)" ini ditulis untuk memenuhi salah satu syarat dalam menyelesaikan perkuliahan pendidikan strata satu (S1) pada Program Studi Informatika, Universitas Bakrie.

Banyak pihak yang telah membantu penulis dalam penelitian dan penulisan Tugas Akhir ini, baik itu berupa bimbingan, saran, maupun dukungan secara moril dan materil. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Proposal Tugas Akhir ini, sangatlah sulit bagi peneliti untuk menyelesaikannya dan masih banyak kekurangan. Dengan segenap kerendahan hati, melalui kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Hoga Saragih, S.T., M.T., IPM., selaku Kepala Program Studi Informatika dan dosen pembimbing, yang senantiasa memberikan masukan dan motivasi kepada penulis;
2. Ihsan Ibrahim, S.T., M.T., selaku dosen pembimbing, yang telah meluangkan waktunya serta memberikan bimbingan, saran, dan perbaikan dalam menyelesaikan penelitian ini;
3. Sondang Sibuea, S.Kom., M.Kom., selaku dosen pembahas dan penguji yang memberikan saran dan perbaikan terhadap penelitian ini;
4. Albert A. Sembiring, S.T, M.T., selaku dosen pembahas dan penguji yang memberikan saran dan perbaikan terhadap penelitian ini;
5. Seluruh Bapak/Ibu dosen Program Studi Informatika Universitas Bakrie, yang telah memberikan banyak ilmu, pengetahuan, wawasan kepada penulis selama perkuliahan;
6. Keluarga tercinta, kedua Orang tua penulis (Rafles dan Ade Eriani), saudara kandung penulis (Rifdah Adra Shalihah) dan keluarga besar yang telah memberikan dukungan dan doa yang sangat berarti bagi penulis;

7. Fuaidah Hanani dan Sairam Salim, yang telah membantu dalam proses penelitian, penulisan dan memberikan dukungan juga saran;
8. Muhammad Fadhil Zulmardhiya, yang bersedia menjadi tempat curhat mengenai suka duka dalam penyusunan tugas akhir ini, juga telah memberikan doa, dukungan, motivasi, semangat dan kasih sayang hingga saat ini;
9. Seluruh teman Informatika seangkatan, senior dan junior. Terima kasih atas kebersamaan selama masa studi di kampus;
10. Seluruh pihak yang terlibat dalam penyusunan Tugas Akhir ini yang tidak dapat penulis sebutkan satu persatu;

Penulis berharap semoga Allah Subhanahu wata'ala membalas kebaikan seluruh pihak yang telah membantu penulis dalam menyelesaikan Tugas Akhir ini. Penulis berharap Tugas Akhir ini bermanfaat bagi pihak-pihak terkait ke depannya.

Jakarta, 1 Juli 2021

Fildzah Adra Arifah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama	: Fildzah Adra Arifah
NIM	: 1132001017
Program Studi	: Informatika
Fakultas	: Teknik dan Ilmu Komputer
Jenis Tugas Akhir	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

PENERAPAN SISTEM DATA CLEANING DALAM MASTER DATA DENGAN MENGGUNAKAN ALGORITMA DUPLICATE COUNT STRATEGY

(STUDI KASUS: PT XYZ)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*datebase*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 24 Agustus 2021

Yang menyatakan

(Fildzah Adra Arifah)

PENERAPAN SISTEM *DATA CLEANING* DALAM MASTER DATA DENGAN MENGGUNAKAN ALGORITMA *DUPLICATE COUNT STRATEGY* (STUDI KASUS: PT XYZ)

Fildzah Adra Arifah[1]

ABSTRAK

Adanya permasalahan berupa duplikasi data dalam sebuah master data, peneliti melakukan penerapan sistem untuk *data cleaning* yang dapat mendeteksi duplikasi data pada master data konsumen Divisi *Consumer Care* PT XYZ. Dalam penelitian ini digunakan algoritma untuk mendeteksi duplikasi data yaitu dengan menerapkan pendekatan metode *Duplicate Count Strategy* (DCS++) dan *N-Gram*. Sistem *data cleaning* diperuntukan bagi *Sales Admin* agar mempermudah dalam menemukan duplikasi data. Selain itu, sistem ini juga dibuat untuk merapikan format penulisan telepon dan fax yang ada pada master data konsumen Divisi *Consumer Care* PT XYZ. Penerapan ini dibangun dengan menggunakan bahasa pemrograman C#. Hasil dari penerapan sistem *data cleaning* yang dibangun akan dinilai seberapa efektif metode DCS++ dan N-Gram dengan menghitung nilai *recall* dan *precision* dalam mendeteksi duplikasi data.

Kata kunci: *Data cleaning*, Deteksi Duplikasi Data, *Duplicate Count Strategy*, *N-gram*

[1] Mahasiswa Program Studi Informatika, Universitas Bakrie

IMPLEMENTATION OF DATA CLEANING SYSTEM IN MASTER DATA USING DUPLICATE COUNT COUNT STRATEGY ALGORITHM (CASE STUDY: PT XYZ)

Fildzah Adra Arifah[2]

ABSTRACT

The existence of a problem in the form of duplication of data in a master data, the researchers implemented a system for data cleaning that can detect duplication of data in the consumer master data of the Consumer Care Division of PT XYZ. In this study, an algorithm is used to detect duplication of data by applying the Duplicate Count Strategy (DCS++) and N-Gram method approaches. The data cleaning system is intended for Sales Admins to make it easier to find duplicate data. In addition, this system was also created to tidy up the format of writing telephone and faxes in the consumer master data of the Consumer Care Division of PT XYZ. This application is built using the C# programming language. The results of the implementation of the data cleaning system that was built will be assessed how effective the DCS++ and N-Gram methods are by calculating the recall and precision values in detecting duplication of data.

***Keywords**: Data Cleaning, Duplicate Detection, Duplicate Count Strategy, N-Gram*

[1] *Student of Informatics Program, Bakrie University*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN