RANCANG BANGUN SISTEM *DATA CLEANING* UNTUK MASTER DATA KONSUMEN DI PT XYZ DENGAN MENERAPKAN METODE *SORTED NEIGHBOURHOOD* DAN METODE *N-GRAM*

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

RAHMA MUALIFA
1112001011

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2016

RANCANG BANGUN SISTEM *DATA CLEANING* UNTUK MASTER DATA KONSUMEN DI PT XYZ DENGAN MENERAPKAN METODE *SORTED NEIGHBOURHOOD* DAN METODE *N-GRAM*

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

RAHMA MUALIFA
1112001011

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2016

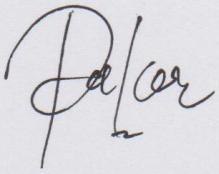
HALAMAN PERNYATAAN ORISINALITAS

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

Nama : **Rahma Mualifa**

NIM : **1112001011**

Tanda Tangan :



Tanggal 10 Agustus 2016

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama	:	Rahma Mualifa
NIM	:	1112001011
Program Studi	:	Informatika
Fakultas	:	Teknik dan Ilmu Komputer
Judul Skripsi	:	Rancang Bangun Sistem *Data Cleaning* Untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode *Sorted Neighbourhood* dan *N-Gram*

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

DEWAN PENGUJI

Pembimbing : Yusuf Lestanto, S.T., M.Sc. (.....)

Penguji 1 : Prof. Dr. Hoga Saragih, S.T., M.T. (.....)

Penguji 2 : Gun Gun Gumilar, S.Kom., MMSI (.....)

Ditetapkan di : Jakarta

Tanggal : 10 Agustus 2016

UNGKAPAN TERIMA KASIH

Puji dan syukur kehadirat Allah SWT karena atas rahmat-Nya dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir dengan judul "Rancang Bangun Sistem *Data Cleaning* Untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode *Sorted Neighbourhood* dan Metode *N-Gram*" 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. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Hoga Saragih, S.T., M.T., selaku Kepala Program Studi Informatika, yang senantiasa memberikan masukan dan motivasi kepada penulis;
2. Yusuf Lestanto, S.T., M,Sc., selaku dosen pembimbing, yang telah meluangkan waktunya serta memberikan bimbingan, saran, dan perbaikan dalam menyelesaikan penelitian ini;
3. Prof. Dr. Hoga Saragih, S.T., M.T. dan Gun Gun Gumilar, S.Kom., MMSI selaku dosen pembahas dan penguji yang memberikan saran dan perbaikan terhadap penelitian ini;
4. Seluruh Bapak/Ibu dosen Program Studi Informatika UB, yang telah memberikan banyak ilmu, pengetahuan, wawasan kepada penulis selama perkuliahan;
5. Keluarga tercinta, kedua Orang tua penulis (Ariyanto dan Siti Rohmatun) dan saudara kandung penulis (Rahma Maulida dan Rahman Abid) yang telah memberikan dukungan dan doa yang sangat berarti bagi penulis.;

6. Tim Magang (Mario Joel, Aulia Syarifuddin, Mega Silviana, Celina Maya Cantika, M. Fadil). Terima kasih telah memberikan semangat, motivasi, dukungan, suka cita dan kebersamaan selama ini;
7. Teman VMG (Destalia Dianing Putri, Septy Dwi Aryani, Eka Juniar, Andining Tyas, Diti Puspa Permata). Terima kasih telah memberikan semangat, motivasi, dukungan, suka cita dan kebersamaan selama ini;
8. Tim Seperjuangan (Sawitri Sadanti, Stefanny Uliarta, Sarah Putri, Rien Pratama, Rahmad Pratama Dita, Fazz Faidurrahman, Evi Margaretha, Rizky Akbarie, Chandra Setiawan). Terima kasih telah menjadi teman yang selalu memberikan semangat, motivasi, dukungan dan suka cita selama lebih dari 4 tahun.
9. Teman-teman TIF 2011 senasib seperjuangan. Terima kasih sudah menemani dan bekerja sama selama lebih dari 4 tahun masa studi di UB;
10. Seluruh pihak yang terlibat dalam penyusunan Tugas Akhir ini yang tidak dapat penulis sebutkan satu persatu;

Dengan segala keterbatasan yang ada, penulis menyadari bahwa penyusunan tugas akhir ini masih jauh dari kesempurnaan. Untuk itu, saran dan kritik akan selalu diterima agar penulis dapat memperbaiki setiap kekurangan untuk kesempurnaan dimasa mendatang.

Akhirnya, penulis menyampaikan ucapan terima kasih dan semoga Allah SWT membalas segala kebaikan serta melimpahkan berkat dan rahmat-Nya kepada semua pihak yang telah membantu selama ini. Penulis berharap semoga Tugas Akhir ini berguna dan bermanfaat bagi kita semua.

Jakarta, 10 Agustus 2016

Rahma Mualifa

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Rahma Mualifa

NIM : 1112001011

Program Studi : Informatika

Fakultas : Teknik dan Ilmu Komputer

Jenis Tugas Akhir : Rancang Bangun

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:

Rancang Bangun Sistem *Data Cleaning* Untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode *Sorted Neighbourhood* dan *N-Gram*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), 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 : 10 Agustus 2016

Yang menyatakan

Rahma Mualifa

Rancang Bangun Sistem *Data Cleaning* Untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode *Sorted Neighbourhood* dan *N-Gram*

Rahma Mualifa

ABSTRAK

Penelitian ini membahas tentang rancang bangun sistem *data cleaning* untuk dapat mendeteksi duplikasi data yang ada pada master data konsumen Divisi *Consumer Care* PT XYZ. Metode yang digunakan dalam penelitian ini untuk mendeteksi duplikasi data adalah dengan menerapkan pendekatan metode *Sorted Neighbourhood* (SNM) dan *N-Gram*. Sistem *data cleaning* ini bertujuan membantu *user* untuk dapat mempermudah menemukan duplikasi data. Selain itu, sistem ini juga dapat membantu *user* untuk dapat merapikan format penulisan telepon dan fax yang ada pada master data konsumen Divisi *Consumer Care* PT XYZ. Sistem yang akan dibangun adalah sistem *web based* dengan menggunakan bahasa pemrograman C#. Hasil dari sistem *data cleaning* yang dibangun kemudian akan diuji coba kepada *user* dan dinilai seberapa efektif metode SNM dan N-Gram dalam mendeteksi duplikasi data dengan menghitung nilai *recall* dan *precision* terhadap hasil proses deteksi duplikasi data.

Kata kunci: *Data cleaning*, Deteksi Duplikasi Data, *Sorted Neighbourhood*, *N-gram*

Rancang Bangun Sistem *Data Cleaning* Untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode *Sorted Neighbourhood* dan *N-Gram*

Rahma Mualifa

ABSTRACT

This research discusses data cleaning system to detect data duplication are exists in customer master data at Consumer Care Division PT XYZ. Method will be used in this research to detect data duplication is by implementing Sorted Neighbourhood Method (SNM) and N-Gram. This system aims to assist user to find duplicate data easier. Moreover, it also can assist user to fix the format number both phone and fax number within customer master data at Consumer Care Division PT XYZ. System will be developed as web-based system by using C# programming language. The result of this system development will be tested by user and rated its effectiveness through implementation of SNM and N-Gram. To compute effectiveness will be obtained recall and precision value to determine how effective this system in detecting duplication data.

Kata kunci: *Data cleaning, Duplicate Detection, Sorted Neighbourhood, N-gram*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN

DAFTAR SINGKATAN

IDE	Integrated Development Environment
IGASIS	Intra-Governmental Access To Shared Information System
KDD	Knowledge Discovery in Databases
OOP	Object Oriented Programming
SNM	Sorted Neighbourhood Method
UML	Unified Modelling Language