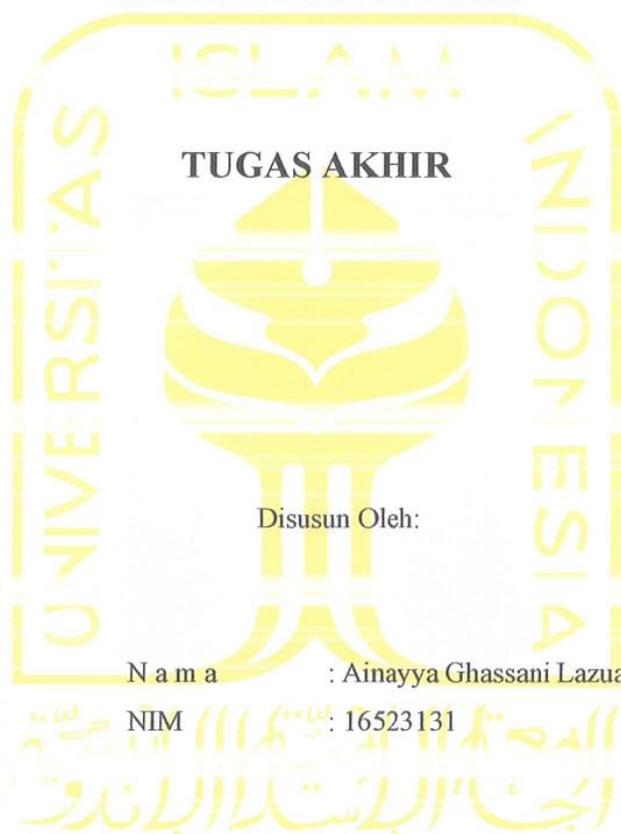


HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PENERAPAN PROSES DATA CLEANSING DENGAN  
MENGUNAKAN METODE DUPLICATE ELIMINATION,  
INCOSISTECY DETECTION, DAN HANDLING MISSING  
ENTRIES PADA SALAH SATU DATA SISTEM RUMAH  
SAKIT DI INDONESIA**



N a m a : Ainayya Ghassani Lazuardy

NIM : 16523131

Yogyakarta, 6 Januari 2020

Pembimbing,

A handwritten signature in blue ink, appearing to be 'Hari Setiaji', is written over a horizontal line.

( Hari Setiaji, S.Kom., M.Eng)

## HALAMAN PENGESAHAN DOSEN PENGUJI

**PENERAPAN PROSES DATA CLEANSING DENGAN  
MENGUNAKAN METODE DUPLICATE ELIMINATION,  
INCOSISTECY DETECTION, DAN HANDLING MISSING  
ENTRIES PADA SALAH SATU DATA SISTEM RUMAH  
SAKIT DI INDONESIA**

## TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika di Fakultas Teknologi Industri Universitas Islam Indonesia  
Yogyakarta, 6 Januari 2020

Tim Penguji

Hari Setiaji, S.Kom., M.Eng

**Anggota 1**

Fietyata Yudha, S.Kom., M.Kom.

**Anggota 2**

Andhika Giri Persada, S.Kom., M.Eng.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dr. Raden Teduh Dirgahayu, S.T., M.Sc.)

**HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan di bawah ini:

Nama : Ainayya Ghassani Lazuardy

NIM : 16523131

Tugas akhir dengan judul:

**PENERAPAN PROSES DATA CLEANSING DENGAN  
MENGUNAKAN METODE DUPLICATE ELIMINATION,  
INCOSISTECY DETECTION, DAN HANDLING MISSING  
ENTRIES PADA SALAH SATU DATA SISTEM RUMAH  
SAKIT DI INDONESIA**

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila dikemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung resiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 6 Januari 2020



(Ainayya Ghassani Lazuardy)

## **HALAMAN PERSEMBAHAN**

Saya persembahkan karya ini untuk:

Ayah dan Bunda saya atas segala kasih sayang yang sudah diberikan dan selalu bangga terhadap anaknya dalam keadaan suatu apapun. Semoga ayah dan bunda diberi rezeki yang melimpah, kesehatan dan keselamatan oleh Allah SWT.

**HALAMAN MOTO**

**“DO IT WITH YOUR PASSION OR NOT AT ALL.”**

-mediawebapps.com

## KATA PENGANTAR

Alhamdulillah atas segala nikmat Allah yang telah diberikan, saya mampu menyelesaikan penelitian ini berjudul “Data Cleansing pada data Sistem Informasi Rumah Sakit”. Dalam penyusunan tugas akhir ini tidak lepas dari banyaknya bimbingan dan bantuan dari teman – teman dan keluarga terdekat. Dengan segala hormat penelitian ini saya persembahkan kepada:

1. Allah SWT karena telah memberikan rahmatnya, ridhonya, sehingga saya dapat melangkah pada tahap ini meskipun saya masih banyak dosanya.
2. Orang tua saya, Ibu Yusnia Raya dan Bapak Iswan Gunadi karena sudah mendukung saya dalam segi emosional dan *financial*.
3. Bapak Hari Setiaji, S.Kom., M.Kom selaku pembimbing saya dalam mengerjakan tugas akhir.
4. Dosen – dosen informatika yang saya banggakan dan saya kagumi.
5. Afik, Resti, Tsaniya, Vita, Anin, Alex, Veda karena sudah mendengarkan keluh kesah saya selama kuliah dan memberikan solusi yang bagus.
6. Septi, Dhanu, Ray teman semasa SMA yang masih ada sampai sekarang selalu disisi saya
7. Gaswara, Afifah, Citra, Vani selaku teman satu bimbingan saya yang selalu saling mendukung.
8. Teman – teman Student Staff Informatika yang sudah berbagi pengalaman bekerja sama dengan saya
9. Teman – teman hexadecima yang selalu bersama – sama menghadapi perkuliahan.

Akhir kata, penulis mengucapkan Terima Kasih.

Yogyakarta, 6 Januari 2020

( Ainayya Ghassani Lazuardy )

## SARI

Data dari Sistem Informasi Rumah Sakit tidak memiliki indikasi sebagai data yang berkualitas. Hal ini disebut juga dengan data kotor. Data kotor dapat berupa duplikasi data, tidak konsistennya data, dan data kosong. Data kotor ini dikarenakan kesalahan dalam entri data, skema sistem yang tidak berkesinambungan, data yang tumpang tindih dan lain – lain. Sehingga dibutuhkan tahapan *Data Cleansing*. *Data Cleansing* digunakan untuk mengubah data kotor menjadi data yang berkualitas yang nantinya data yang berkualitas akan diterapkan dengan tahapan data mining. Untuk itu, dibutuhkan metode yang tepat untuk melakukan proses *Data Cleansing*. Ditemukan bahwa data sistem rumah sakit mengalami duplikasi data pada tabel RAWAT\_JALAN berjumlah 17 baris dan tabel PENDUDUK 6 baris, tidak konsistennya data pada tabel ASURANSI\_KEPESERTAAN\_VISIT berjumlah 55 baris, tabel KUNJUNGAN BPJS 3 baris, tabel PENDUDUK 381 baris, dan data kosong pada tabel KUNJUNGAN 137 baris dan RAWAT\_JALAN 4 baris. Maka data sistem rumah sakit ini menerapkan metode *Duplicate Elimination*, *Inconsistency Detection*, dan *Handling Missing Entries*. Ketiga metode ini dapat dijalankan secara paralel atau berurutan. Dapat dijalankan untuk metode *Duplicate Elimination* terlebih dahulu lalu *Inconsistency Detection* dan setelahnya menggunakan *Missing Entries*. Maka dari itu, dengan metode – metode tersebut data sistem rumah sakit sudah tidak mengalami duplikasi data, tidak konsistennya data dan tidak ada lagi data kosong data sistem rumah sakit.

Kata Kunci: *Data cleansing*, *Duplicate Elimination*, *Inconsistency Detection*, *Handling Missing*

## GLOSARIUM

|                   |                                                                                  |
|-------------------|----------------------------------------------------------------------------------|
| Records           | merupakan kumpulan dari elemen-elemen data yang terkait dalam sebuah basis data  |
| Import            | proses untuk memasukkan data pada sistem                                         |
| Export            | proses untuk mengeluarkan data dan mengubah data menjadi format <i>file</i> lain |
| Training data set | sekumpulan data yang dapat diolah                                                |