

## **ABSTRAK**

# **ANALISIS *SINGLE COLUMN CARDINALITIES*, *DATA PATTERN* DAN *VALUE DISTRIBUTION* UNTUK *PROFILING DATA* MENGUNAKAN *OPEN SOURCE PLATFORM***

Oleh

**SANDY RAMA AMETHYST**

**1202142181**

Pentingnya kualitas data dapat memberikan pengaruh besar terhadap proses bisnis yang ada pada suatu perusahaan. Namun masih banyak perusahaan yang belum memahami pentingnya kualitas data tersebut. Permasalahan yang sering terjadi terhadap kualitas data pada banyak perusahaan di Indonesia adalah tidak tersaringnya data yang diinputkan, sehingga terdapat permasalahan data mengenai data kosong, pola yang tidak terstandar dan penyebaran data yang tidak merata. Permasalahan ini dapat ditangani dengan beberapa proses yang bertahap, dimana salah satunya adalah data *profiling*. Data *profiling* adalah proses mengumpulkan informasi dari sebuah data sehingga dapat diukur kualitas, keakurasian dan konsistensi dari sebuah data. Untuk proses data *profiling* dapat dilakukan dengan menggunakan berbagai macam *tools* baik dari yang berbayar hingga yang berbasis *open source* dimana masing-masing memiliki keunggulan tersendiri sesuai dengan kebutuhan. Dalam penelitian ini, fokus utama pada analisis data dengan melakukan data *profiling* menggunakan metode *cardinalities*, *data pattern* dan *value distribution* menggunakan aplikasi *open source*. Hasil dari *profiling* akan diimplementasikan dalam bentuk logika pada aplikasi *open source* dan akan dikomparasi dengan aplikasi *open source* lainnya.

Kata kunci : data *profiling*, *open source*, *cardinalities*, *data pattern*, *value distribution*