

## 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\_KEPESEERTAAN\_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*