

ABSTRACT

Profiling has an important role in the process of data so that the data used is data of good quality. Many found in company problems, that the data to be processed is data that is still dirty. This makes the data not processed or analyzed optimally. If the existing data is dirty, then the decision to be taken by the company is not necessarily correct.

In various sources, there are profiling classifications with different names but have the same purpose. Profiling is divided into 3 major sections, namely Single Column, Multi-Column, and Dependency. Single column is a data profiling analysis of one column only. Multi column using profiling analysis on two or more different columns. Meanwhile, dependency using profiling analysis on a table that is interdependent.

In this research, a new package is created in the form of Cardinalities, domain analysis, and data types which are included in a single column. This study uses a database and parameters that can be adapted to the needs of the company. The results of this study can be used in various analysis needs of the company.

Keywords— profiling, cardinalities, domain analysis, data type