Data Quality Assesment

BigDataStack Software Component developed by UPRC

Description

A domain-agnostic data quality assessment and an improvement framework, that can identify valid records in a relational database and establish data veracity.

Features

The main contribution of the data quality assessment and improvement module is that can provide a domain-agnostic framework for automating the data cleaning.

Areas of Application

Data Cleaning and data analysis.

bigdatastack.eu

Data Quality Assesment

Market trends & opportunities

Data cleaning is an area that has not gone far, in contrast to other components of the data analysis pipeline, such as the modelling phase. Thus, there are a lot of opportunities to expand and create value.

Customer benefits

Data cleaning is the first step in a data analysis pipeline. It often the task engineers spend most of their time. By automating this tedious process, we can free up time to more meaningful and creative tasks.

Technological novelty

The main innovation in this component has to do with the keyword "domain-agnostic". There should be no prior knowledge about the dataset and the mechanism should not work as an expert system, but more like a probabilistic system.

TRL level: 5

Find the Open Source code here:

- https://pypi.org/project/forma/

bigdatastack.eu