

Data Quality Assessment

BigDataStack Software Component developed by UPRC

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

Initial TRL

Basic Technology Research

Final TRL

Technology Demonstration

End Users

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

Key Features and Benefits

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

Data Quality Assessment

Essential Information for Users

The data assessment and improvement framework consume a respectable amount of data in order to train its model and discover intricate relationships between the values in a data set. The use of GPU can also facilitate this process by allowing the algorithm to leverage the multicore architecture of such hardware.

Standards involved in the development of the component

Not applicable

Implementation in BigDataStack Use Cases

Data Cleaning component is targeting all BigDataStack use cases: connected consumer, real-time shipping and smart insurance.

How can the BigDataStack component contribute to Standardization foundations or initiatives?

Not applicable

Differentiators from competitors in the market

Domain agnostic