“Quality Data Collection Framework for Big Data Systems”

By

Ala’a Mohammad Hammad

Supervisor

Dr. Mohammad Abdallah

Abstract

The rapid increase in information technology, industries, news, educational websites, and others led to the phenomenon of big data, where big data was collected using several methods, including sensors, frameworks, systems, and web, then a set of factors and legislation appeared such as GDPR and ISO / IEC 25010 to improve the quality of data, but it has not been fully implemented in all organizations due to the diversity of the operational nature of the institutions, and it also did not cover different types of data, so a framework for collecting quality data from big systems were developed, which consists of seven main factors (data source reliability, trustworthy, data suitability, data preservation, data integrity, rapid data collection, data reusability) and twenty sub-factors. these factors are improving the efficiency of the data collection process and obtaining high-quality data, The process of collecting and analyzing high-quality data was validated using appropriate filters in the Weka tool and using Data quality dimensions (DQD) such as the calculation of missing data, but the filters did not cover two sub-factors of a reliable source and legal data so it was included in the contract between the data provider and the data receiver.

Keywords

Quality Data collection: data quality, quality data factors, big data, big data quality collection.