ABSTRACT

AI-Assisted Data Quality and Anomaly Detection on Hadoop-Based Healthcare Operational Data

In most healthcare organizations, there are copious amounts of data generated during the operation, and the organization utilizes different types of systems in the process. Data quality plays a major role in the analysis of data and also in the decision-making process, but the conventional approach to validating the data using several rules becomes complicated when dealing with increased volumes of data or complex data patterns. Here, the aim is to implement the AI-based approach for assessing the data quality using the Hadoop framework for big data.

The proposed system has been devised by taking advantage of the scalability provided by Hadoop, at the same time, machine learning techniques have also been employed, aiming at identifying data consistency, missing information, and data anomalies. Moreover, this system incorporates data processing, conducted in a distributed manner, in order to efficiently handle high-velocity as well as high-volume healthcare data. By incorporating AI technology regarding data anomalies, the system ensures efficient data processing, leading to a high degree of data trustworthiness, thereby enabling more effective decisions in the healthcare domain.

Team Members:

Shreya Phadke (2320030194)

Mrinalini Banala (2320030309)

Rushank Chaudhari (2320030105)