

The Data Quality Model for the IoT based Real-time Water Quality Monitoring Sensors

Authors : Rabbia Idrees, Ananda Maiti, Saurabh Garg, Muhammad Bilal Amin

Abstract : IoT devices are the basic building blocks of IoT network that generate enormous volume of real-time and high-speed data to help organizations and companies to take intelligent decisions. To integrate this enormous data from multisource and transfer it to the appropriate client is the fundamental of IoT development. The handling of this huge quantity of devices along with the huge volume of data is very challenging. The IoT devices are battery-powered and resource-constrained and to provide energy efficient communication, these IoT devices go sleep or online/wakeup periodically and a-periodically depending on the traffic loads to reduce energy consumption. Sometime these devices get disconnected due to device battery depletion. If the node is not available in the network, then the IoT network provides incomplete, missing, and inaccurate data. Moreover, many IoT applications, like vehicle tracking and patient tracking require the IoT devices to be mobile. Due to this mobility, If the distance of the device from the sink node become greater than required, the connection is lost. Due to this disconnection other devices join the network for replacing the broken-down and left devices. This make IoT devices dynamic in nature which brings uncertainty and unreliability in the IoT network and hence produce bad quality of data. Due to this dynamic nature of IoT devices we do not know the actual reason of abnormal data. If data are of poor-quality decisions are likely to be unsound. It is highly important to process data and estimate data quality before bringing it to use in IoT applications. In the past many researchers tried to estimate data quality and provided several Machine Learning (ML), stochastic and statistical methods to perform analysis on stored data in the data processing layer, without focusing the challenges and issues arises from the dynamic nature of IoT devices and how it is impacting data quality. A comprehensive review on determining the impact of dynamic nature of IoT devices on data quality is done in this research and presented a data quality model that can deal with this challenge and produce good quality of data. This research presents the data quality model for the sensors monitoring water quality. DBSCAN clustering and weather sensors are used in this research to make data quality model for the sensors monitoring water quality. An extensive study has been done in this research on finding the relationship between the data of weather sensors and sensors monitoring water quality of the lakes and beaches. The detailed theoretical analysis has been presented in this research mentioning correlation between independent data streams of the two sets of sensors. With the help of the analysis and DBSCAN, a data quality model is prepared. This model encompasses five dimensions of data quality: outliers' detection and removal, completeness, patterns of missing values and checks the accuracy of the data with the help of cluster's position. At the end, the statistical analysis has been done on the clusters formed as the result of DBSCAN, and consistency is evaluated through Coefficient of Variation (CoV).

Keywords : clustering, data quality, DBSCAN, and Internet of things (IoT)

Conference Title : ICIOEA 2022 : International Conference on Internet of Everything and Applications

Conference Location : Sydney, Australia

Conference Dates : August 30-31, 2022