Validity as a Measure of Data Quality in Internet of Things Systems

Rishabh Deo Pandey1 · Itu Snigdh1

Accepted: 7 May 2022 / Published online: 8 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Data quality became significant with the emergence of data warehouse systems. While accuracy is intrinsic data quality, validity of data presents a wider perspective, which is more representational and contextual in nature. Through our article we present a different perspective in data collection and collation. We focus on faults experienced in data sets and present validity as a function of allied parameters such as completeness, usability, availability and timeliness for determining the data quality. We also analyze the applicability of these metrics and apply modifications to make it conform to IoT applications. Another major focus of this article is to verify these metrics on aggregated data set instead of separate data values. This work focuses on using the different validation parameters for determining the quality of data generated in a pervasive environment. Analysis approach presented is simple and can be employed to test the validity of collected data, isolate faults in the data set and also measure the suitability of data before applying algorithms for analysis. On analyzing the data quality of the two data sets on the basis of above-mentioned parameters. We show that validity for data set 1 was found to be 75% while it was found to be 67% only for data set 2. Availability and data freshness metrics performance were analyzed graphically. It was found that for data set 1, data freshness was better while availability metric was found better for data set 2. Usability obtained for data set 2 was 86% which was higher as compared to data set 1 whose usability metric was 69%. Thus, this work presents methods that can be leveraged for estimating data quality that can be beneficial in various IoT based industries which are essentially data centric and the decisions made by them depends upon the validity of data.

Keywords Internet of things (IoT) · Data quality · Metric · Validity · Data freshness · Completeness · Availability · Usability

Abbreviations
IoT Internet of Things
\(t_i\) Time at which update occurred

* Rishabh Deo Pandey
phdcs10056.19@bitmesra.ac.in

Itu Snigdh
itusnigdh@bitmesra.ac.in

1 Department of CSE, BIT Mesra, Mesra, India
1 Introduction

Internet of Things (IoT) is a network of physical devices or things embedded with electronic devices such as sensors, actuators, software and an interconnecting network which enables it to collect and manage data without involving human interaction [1]. Common to all the mentioned applications, the IoT application comprises of four major segments [2]. The sensor and actuators with the communication capabilities are the integral part and mostly hardware oriented. The software segment consists of the user interface modules, connectivity and adaptability of the apps across different machine capabilities and interfaces [3]. IoT combines various technologies into a single intelligent unit. This intelligence in IoT networks is attributed to the amount and quality of data. All IoT devices generate varied amounts of data at varied levels of scalability and time dependence like a point of sales terminal, distributed sensors or industrial machinery [4]. These data range from being mission critical to custom business logic. Also, with the proliferation of organizations in to the IoT domains and its enablement through hardware, software services and connectivity, the complexity of the applications have increased. In addition to handling increased software complexity, IoT performance also depends on its responsiveness and real time digital services. Since an important part of an IoT application is the software in addition to its hardware, evaluation of the performance of an IoT application becomes necessary [5]. Software metrics therefore need to be developed for IoT applications, in order to monitor IoT performance for fulfilling the following requirements:

- Able to monitor devices that run on different processor architecture.
- Able to monitor IoT application written in different programming languages.
- Overhead in monitoring should be minimal.
- Able to receive data generated.

It is recognized that Software metrics is a measure of characteristics of the software that are quantifiable or countable. Also, the main objective of any software metric is to analyze the product or process; to determine its quality and suggest improvements as well as predict when the software development process is over [6–8].

In this paper, we try to determine data quality in an IoT application on the basis of different metrics namely, validity, data freshness, completeness, usability and availability. These metrics have been studied using two separate data sets. The major focus of this paper is to verify these metrics on aggregated data set instead of separate data values. In order to determine performance of a software model, we have predefined ISO-9126 Software Quality Characteristics which is used for evaluating the software model [9].
1.1 Role of Data and its Quality in IoT Applications

Analysis of IoT data can be helpful in many ways such as optimizing operations, efficient use of energy and resources. Combining IoT with data analytics becomes effective in real-time applications like healthcare, telematics, smart cities and the like [10]. We have a large number of real-time applications available in the market which are fully dependent on the data for its functioning [11, 12]. Therefore, if the data quality degrades it will have a negative impact on the end user relationship. To implement any IoT based application, a large number of sensors are deployed which further generate huge amounts of data. This data generally falls under the category of big data since it justifies the properties of big data namely volume, velocity and variety. Data obtained from the sensors needs to be validated before using it for any decision-making process [13]. Also, IoT application domains like utilities, mobile and enterprises require continuous monitoring of the data as any type of fault can lead to wrong analysis, thereby leading to a decrease in trust by the end user. Apart from affecting the reasoning capability, it also leads to wastage of energy since in case of wireless sensor nodes, every bit of transmission consumes energy [14]. In addition, IoT-based monitoring applications require alerting capabilities that are sophisticated and customizable considering the most decentralized network of devices or the inherent excessive dynamics of the network and devices. The volume of telemetry data that is collected from IoT devices is impacted by the intermittent connectivity issues [15]. Therefore, the traditional alerting tools face challenges which can be overcome to an extent by measuring the quality of the data collected. We are also aware that every IoT application is unique with a said functionality and rarely uses off-the-shelf products [16]. Therefore, a generalized solution “one-size-fits-all” is not applicable.

2 Methodology for Analysis of Metrics

We consider “Temperature” as a common attribute that forms a part of most applications varying from personalized healthcare, to environment monitoring. The data quality dimensions related to parameter “temperature” can be illustrated as Figure 1.

Whenever we apply the concept of IoT for studying environmental behavior in any geographical location, large number of sensors are deployed [17–19]. The data obtained from these sensor nodes go through various issues such as hardware & software problem, poor connectivity, environmental effects [20]. Thus, the data samples obtained needs to be...
validated. Several types of faults occur which needs to be sorted for data validation. In this paper, we focus on two classes of faults; out-of-range faults and struck-at faults.

1. Out-of–range faults occur when the data obtained from the sensors does not lie within an expected range of values. Generally, this type of faults are determined by comparing data using a threshold limit which is usually based on domain knowledge.

```plaintext
Input: Actual and Obtained maximum temperature for different months in a year.
Output: Number of invalid entries that suffer out of range faults.

\[ \text{count}_{\text{invalid}} = 0; \]
\[ n = \text{number of months}; \]
\[ \text{for } i \leftarrow 0 \text{ to } n \text{ do } i \leftarrow i+1 \]
\[ \text{difference}[i] = | \text{actual}[i] - \text{obtained}[i] | \]
\[ \text{end for} \]
\[ // \text{Average and SD is obtained for difference data set and limits are defined.} \]
\[ \delta_{\text{lower}} = \text{Average} - \text{SD}; // \text{lower threshold value} \]
\[ \delta_{\text{upper}} = \text{Average} + \text{SD}; // \text{upper threshold value} \]
\[ \text{for } i \leftarrow 0 \text{ to } n \text{ do } i \leftarrow i+1 \]
\[ \text{if } (\text{difference}[i] < \delta_{\text{lower}} \text{ or difference}[i] > \delta_{\text{upper}}) \]
\[ \text{then } \text{count}_{\text{invalid}} \leftarrow \text{count}_{\text{invalid}} + 1 \]
\[ \text{end if} \]
\[ \text{end for} \]
\[ \text{end} \]
```

2. Struck-at faults or constant faults occur when the data obtained from the sensor nodes show no change or very little change for a significant amount of time. It behaves as if data has reached a frozen state and shows no variation.
Validity as a Measure of Data Quality in Internet of Things Systems

3 Metrics used

The methodology and quality metrics considered in this research are depicted through Figure 2. Our analysis starts with number and maximum value of updates on a time-series framework. The metrics used on the data is dependent on the frequency of updates received are data freshness, completeness and availability as shown in Figure 2. On the contrary, validity is determined on the maximum of the update value of the parameter considered.

3.1 Validity

When we deal with data related to pervasive environment validity is defined on the basis of some practical considerations [21]. To determine validity of any measure, various techniques are used. For example, a database application comprising of student details, can be validated by cross-checking from various other available data sources [22]. However, when we consider the case of IoT applications, it is not easy to obtain correctness of data since these data are generated for a short span of time. Therefore, for defining validity, we first need to define validity rules. Validity rule (V) can be expressed as a Boolean function. If validity rule V is satisfied by the data object, then its value is 1 else 0. These rules are domain specific [23–28]. The two most common validity rules are:

```
Input: Maximum temperature readings of a month.
Output: Number of updates which acts as a parameter to determine struck-at faults or constant faults.

\[ \text{count}_{\text{update}} = 0; \]
\[ n = \text{number of days in a month}; \]
\[ // \text{Average and SD is obtained for a month data set.} \]
\[ \delta_{\text{lower}} = \text{Average} - \text{SD}; // \text{lower threshold value} \]
\[ \delta_{\text{upper}} = \text{Average} + \text{SD}; // \text{upper threshold value} \]
for i ← 0 to n do i ← i+1
    if (temperature[i] < \delta_{\text{lower}} \text{ or } \text{temperature[i]} > \delta_{\text{upper}})
        then \text{count}_{\text{update}} ← \text{count}_{\text{update}} + 1 \quad // \text{more the value of count}_{\text{update}} less will be the chances of constant fault. 
    end if
end for
```


a. Static rules that are satisfied by verifying the data set. For example temperature in Polar region will be negative or around zero degree Celsius.

b. Dynamic rules are used to verify the changes in the data i.e. drop of 10 degrees is not possible in half an hour.

Validity can also be expressed as a stochastic measure when it is evaluated for a particular observation period. Let $n$ be the number of updates that have occurred and $n_{\text{valid}}$ be the total valid instances; probability of validity can be expressed as:

$$\text{Prob. Validity} = \frac{n_{\text{valid}}}{n}$$

In research application areas, validity refers to how accurately a particular method measures any quantity. So, when we are dealing with data related to pervasive environment, validity is defined on the basis of some practical considerations that we try to explore in our work. The requirements of IoT applications are varied and the data quality dimensions align accordingly. For example, a timely data may not be accurate while an accurate data may not be complete or in time. These tradeoffs force us to outline a metric system that could map single attribute “validity” in terms of its timeliness, completeness and its usability [29].

We assume IoT monitoring applications which employ several sensors that collect temperature data for storage and analysis. For the ease of elaboration, our focus of study considers ‘Maximum’ temperature. Since we consider data for an entire year, we also focus on evaluating metrics for aggregated data in our study. We usually obtain large set
of temperature data because it keeps on fluctuating. Therefore, large volume of data will be obtained for a year. The data sets [30, 31] which has been used for evaluating validity is obtained from the meteorological lab, in the district of Ranchi, India [23.3441° N, 85.3096° E]. It consists of maximum and minimum temperature of all months for the year 2012. We have considered maximum temperature as a measure for determining its validity. For deciding the validity of a dataset, first of all we need to calculate the update_value which serves as an important entity for validation of any data. To obtain update_value, we followed following steps:

- **Step 1:** Obtain mean of maximum temperature for each month and calculate Standard Deviation (SD) for the dataset.
- **Step 2:** Set the limits with the help of obtained mean and SD values.
- **Step 3:** To determine updates in a month, check those temperature values which doesn’t lie in the limit.
- **Step 4:** Calculate the different update_value for each month which is the difference between the consecutive updates in terms of the duration.

On the basis of these values, max update_value for each month was determined which is depicted in Fig. 3. From the figure we see that there are two discrepancies in the graph, in the month of February and July which should ideally be aligned as they belong to the same period and same geographical location. Spikes obtained in the graph for these months are due to the significant difference between the maximum update_value of the two data sets which creates confusion regarding the correctness of data set. Thus, any irrelevance is an anomaly.

To determine validity, we used weather data [32] and our algorithm compared the max temperature value. Then, we isolated the particular day in a month where max-update was found and compared max temperature for that particular day with reading of the reference website. When there was a considerable difference, we calculated average and standard deviation for those differences. As, we have earlier stated that to determine validity, firstly we need to find those validity rules that are true (1) or false (0). They are represented as bool values.

![Maximum update_value obtained](image)

*Fig. 3* shows maximum update value determined for each month
4 Result 1

To determine probability validity, we used formula mentioned in (1). Formula was applied on the two datasets as shown in Tables 1 and 2. Prob. Validity obtained for the two data sets was 75% and 67% as mentioned in Table 3, which shows that data set 1 has higher chances of validity than data set 2.

| UpDATES      | Actual | Obtained | Difference | Bool |
|--------------|--------|----------|------------|------|
| 07-01-2012   | 18     | 18       | 0          | 0    |
| 09-02-2012   | 23     | 23.6     | 0.6        | 1    |
| 10-03-2012   | 28     | 29.6     | 1.6        | 0    |
| 15-04-2012   | 34     | 34.8     | 0.8        | 1    |
| 23-05-2012   | 41     | 41.7     | 0.7        | 1    |
| 19-06-2012   | 31     | 31.8     | 0.8        | 1    |
| 31-07-2012   | 27     | 27.2     | 0.2        | 1    |
| 14-08-2012   | 25     | 25.4     | 0.4        | 1    |
| 30-09-2012   | 32     | 32.1     | 0.1        | 1    |
| 27-10-2012   | 28     | 28       | 0          | 0    |
| 18-11-2012   | 24     | 23.9     | 0.1        | 1    |
| 22-12-2012   | 22     | 21.8     | 0.2        | 1    |

| UpDATES      | Actual | Obtained | Difference | Bool |
|--------------|--------|----------|------------|------|
| 18-01-2012   | 24     | 26.8     | 2.8        | 1    |
| 22-02-2012   | 31     | 35.2     | 4.2        | 1    |
| 12-03-2012   | 31     | 32.78    | 1.78       | 1    |
| 06-04-2012   | 33     | 40.14    | 7.14       | 0    |
| 31-05-2012   | 39     | 47.76    | 8.76       | 0    |
| 23-06-2012   | 25     | 27.65    | 2.65       | 1    |
| 25-07-2012   | 29     | 28.23    | 0.77       | 0    |
| 22-08-2012   | 25     | 30.72    | 5.72       | 1    |
| 15-09-2012   | 28     | 26.72    | 1.28       | 0    |
| 23-10-2012   | 28     | 32.72    | 4.72       | 1    |
| 12-11-2012   | 25     | 30.78    | 5.78       | 1    |
| 10-12-2012   | 27     | 30.40    | 3.40       | 1    |

| Measure              | Data set 1 | Data set 2 |
|----------------------|------------|------------|
| Average              | 0.46       | 4.08       |
| Standard deviation   | 0.47       | 2.44       |
| Upper limit          | 0.93       | 6.52       |
| Lower limit          | 0.01       | 1.64       |
| Probability validity | 75%        | 67%        |
4.1 Data Freshness

Timeliness of data processing has also been referred to as “perishable insights”. If data is not collected and analyzed within the real time constraints, it no longer remains valid and is unusable. So, data freshness becomes one of the most important attributes for measuring the quality of data. It’s a highly beneficial measure especially in case of an IoT application due to the heterogeneity of data sources, high interoperability and rapid generation of data from the sensors. Playing a major role for the analysis of data, it mainly refers to how often a new data is obtained.

As we have considered two separate data sets, so we obtained number of updates occurring in each month for both data sets shown in Fig. 3. Through this figure, we can clearly observe the number of updates occurring in each month for the two data sets. It was observed that in the months of March, August and October, difference in the frequency of updates in the data sets were quite significant. Since validity of data also depends upon the frequency of updates occurring in the data set i.e. more the number of updates more is the validity of data, the guarantee for data freshness is established.

5 Result 2

From the Fig. 4 we see that for the majority of months, number of updates is more for data set 1 as compared to data set 2. Thus, we found data freshness more in data set 1 than data set 2.

5.1 Availability

Data availability refers to the process of conforming that data is available to the user when it is required [33, 34]. As we consider pervasive environment, we define availability with respect to an observation period and interval between updates.

\[
\text{Availability} = 1 - \frac{\max(0, \sum_{i=0}^{n} (t_{i+1} - t_{i}) - T^{exp})}{OP}
\]  

(2)

![Graph showing number of updates per month for both data sets](image)

**Fig. 4** Represents the number of updates occurring in each month for both data sets
Where $n$ is the number of updates occurring in a month, $t_i$ is the time at which update occurred, $t_{i+1}$ is the time at which next update occurred. $OP$ refers to the observation period (30 in our case), while $T^{exp}$ is the expiration time for evaluating availability of the data and is user dependent. As we can clearly observe from Eq. (2), that availability is directly proportional to the number of updates. More the value of $n$, more will be the value of $\sum_{i=0}^{n} (t_{i+1} - t_i)$. Also, as it covers all the updates that are occurring for a given $T^{exp}$, to evaluate availability for a given expiry time, it should accommodate all the updates and they should be additive in nature. The reason for implementing $\max(0, \sum_{i=0}^{n} (t_{i+1} - t_i) - T^{exp})$ is because if in any case 0 is returned then it shows that very less amount of update in the data set has taken place which is not good for any data set.

We can view the performance of availability metric of two data sets for two different expiration times through Fig. 5a, b. Availability metric obtained using (2) for the two data sets were plotted for two different expiration times. The two curves in both the plots demonstrates the calculated value of availability for an entire year for the two data sets. Apart from the month of November, and May, the two data sets behaved similarly. The reason for such deviation is due to the difference in the update intervals for the two data set for the month of November and May. It shows that similar updates was not registered under the same set of conditions and hence can be treated as anomalous if accuracy is a determining criteria.

![Availability measure for $T_{exp}=10$](image1)

![Availability measure for $T_{exp}=15$](image2)

Fig. 5 Availability metric obtained a expiration time:10 days b expiration time:15 days
The results of computing availability, through Eq. (2) for each month for two different expiry times for both data sets are shown in Tables 4 and 5. Thus, from observing the availability metric obtained from Fig. 5a, b and Tables 4 and 5, it was concluded that data set 2 was better in terms of availability as compared to data set 1.

### 6 Result 3

In case of contextual and representational quality of data, concise representation, completeness, value added, and relevancy are most important. Completeness generally refers to the property of any system that doesn’t suffer from data loss [35]. It defines that whether the data source is capable of providing all the information which it mentioned and is required by the application [36]. Completeness can be evaluated at different levels such as:

| Month | $T^{exp} = 10$ days | $T^{exp} = 15$ days |
|-------|---------------------|---------------------|
| Jan | 0.50 | 0.67 |
| Feb | 0.40 | 0.57 |
| March | 0.37 | 0.53 |
| April | 0.47 | 0.63 |
| May | 0.47 | 0.63 |
| June | 0.53 | 0.70 |
| July | 0.30 | 0.47 |
| Aug | 0.33 | 0.50 |
| Sep | 0.33 | 0.50 |
| Oct | 0.30 | 0.47 |
| Nov | 0.63 | 0.80 |
| Dec | 0.40 | 0.57 |

### 6.1 Completeness

| Month | $T^{exp} = 10$ | $T^{exp} = 15$ |
|-------|----------------|----------------|
| Jan | 0.53 | 0.70 |
| Feb | 0.40 | 0.57 |
| March | 0.40 | 0.57 |
| April | 0.53 | 0.70 |
| May | 0.30 | 0.47 |
| June | 0.50 | 0.67 |
| July | 0.33 | 0.50 |
| Aug | 0.30 | 0.47 |
| Sep | 0.33 | 0.50 |
| Oct | 0.30 | 0.47 |
| Nov | 0.93 | 1.00 |
| Dec | 0.43 | 0.60 |
- Scheme Completeness (entities and attributes are missing)
- Column Completeness (column values are missing)
- Population Completeness (column does not contain all the samples of population)

Mathematically, it can be defined as:

\[
\text{Completeness} = 1 - \frac{\text{no. of incomplete items}}{\text{total no. of items}}
\]  \hspace{1cm} (3)

In our study, completeness is directly proportional to the number of updates. With a greater number of updates, the chances of missing the updated data will be less. As we are computing completeness annually, it will depend on the update values of individual month(s) and will be calculated as a relative term with respect to the maximum number of updates found in a yearly dataset i.e.

\[
\text{Completeness} = \frac{\sum u_j}{\sum u_{\text{max}}}
\]  \hspace{1cm} (4)

Where: \( u_j \) = no. of updates occurring in a month

\( u_{\text{max}} \) = Maximum no. of updates occurring in a month.

7 Result 4

For evaluating completeness for the data sets, formula mentioned in Eq. (4) was used. Number of updates obtained for each month can be observed in Table 6. Using these updates values, for evaluating completeness, it was found that data set 1 had relative completeness measure of 75% while data set 2 had 69%. Thus, it was concluded that data set 1 is more complete than data set 2.

| Month | Updates in Dataset 1 | Updates in Dataset 2 |
|-------|----------------------|----------------------|
| Jan   | 11                   | 9                    |
| Feb   | 11                   | 11                   |
| March | 13                   | 8                    |
| April | 10                   | 9                    |
| May   | 8                    | 6                    |
| June  | 7                    | 6                    |
| July  | 10                   | 6                    |
| Aug   | 9                    | 12                   |
| Sep   | 7                    | 6                    |
| Oct   | 13                   | 10                   |
| Nov   | 8                    | 6                    |
| Dec   | 10                   | 10                   |
7.1 Usability

Data may be incomplete by design or due to operational requirements. Moreover, when data quality is measured on the basis of context, the same data is needed for multiple tasks but different required characteristics. As per application requirements, the same data representation may render useless if it requires aggregates or fields of data that do not exist. This causes poor relevancy to the application and is considered incomplete for analysis, thereby reducing its usability. Usability has also been correlated with the interpretability and ease of understanding of a data that may require expert opinion due to the enormous volume or simply the lack of common interpretability. We measure the usability of data by measuring the deviation from their aggregates to the individual values. The degree of aggregation and variation in their representations have been computed and as their maximum tolerable anomaly which is represented in Table 7.

8 Result

To decide usability, a month was selected randomly (February in our case). All the entries of the month were compared with the entries in Ref. [21] for the same month. Set of valid entries were determined by the interval limits mentioned in Table 7. A usability rate of 69% and 86% was found for the data sets respectively, which shows data set 2 had better usability than data set 1.

9 Conclusion

Our research work presents an approach to study data quality characteristics specifically to IoT applications. We present a method to measure sensor data quality which can in turn be correlated to estimating accuracy. We also show that datasets could be compared through metrics and dependent on the aIoT application requirements, end users could choose an appropriate dataset for further analysis. In this work we define metrics for measuring the data quality for an IoT application that uses weather data and display significance of the analysis. Existing IoT applications are fully dependent on the accuracy of data for the purpose of correct decision making, which is being generated by the sensors. As we know that such type of data falls under the category of Big data, therefore validating such type of data is not an easy task. Also, there is not a unique solution to validate the data because these data vary application wise. Therefore, we try to present methods to determine data validity in a uniform method. In addition, contrary to earlier research works, we propose metrics for data quality for aggregated data. We used five metrics namely validity, data freshness, availability, completeness and usability for observing the data quality on a collective basis. The result shows that validity, freshness and availability metrics for data set 1 is more than data set 2. We also depicted that

| Table 7 | Represents usability metrics obtained for both data sets for February |
|---------|---------------------------------------------------------------|
| Criteria | Data set 1                   | Data set 2                   |
| upper_L= | 0.892681             | 5.982644             |
| lower_L= | 0.169388             | 1.573908             |
| Usability= | 69%                  | 86%                  |
completeness and usability of data collected may vary but is always aligned with the validity of data. A more valid data will be more complete, more usable, more available and up to date. Through this work, data accuracy for IoT applications can be estimated using previously defined metrics. Since IoT being a major part of Industry 4.0, validity of data is essential for the effective establishment of any IoT based business model. Therefore, this approach can be used to verify the data sets so that decisions made upon by analyzing the data would be correct. In future we aim to implement other context-based metrics on data obtained from IoT based applications.

References

1. Chen, H., Jia, X., & Li, H. (2011). A brief introduction to IoT gateway. In: IET international conference on technology and application (ICCTA 2011). IET.
2. Alam, M., Nielsen, R. H., & Prasad, N. R. (2013) The evolution of M2M into IoT. In: 2013 First international Black Sea conference on communications and networking (BlackSeaCom). IEEE.
3. Udoh, I. S., & Kotonya, G. (2018). Developing IoT applications: Challenges and frameworks. IET Cyber-Physical Systems: Theory and Applications, 3(2), 65–72.
4. Babovic, Z. B., Protic, J., & Milutinovic, V. (2016). Web performance evaluation for internet of things applications. IEEE Access 4, 6974–6992
5. Fenton, N. E., & Neil, M. (2000). Software metrics: roadmap. In: Proceedings of the conference on software engineering.
6. Miguel, J. P., Mauricio, D., & Rodríguez, G. (2014). A review of software quality models for the evaluation of software products. arXiv preprint: arXiv:1412.2977.
7. Babbar, H., Parthiban, S., Radhakrishnan, G. & Rani, S. (2022). A genetic loading based algorithm to improve the QoS metrics for software defined networking for multimedia applications. Multimedia Tools and Applications, pp.1–19.
8. Dalla Palma, S., Di Nucci, D., Palomba, F., & Tamburri, D. A. (2020). Toward a catalog of software quality metrics for infrastructure code. Journal of Systems and Software, 170, 110726.
9. Bhatti, S. N. (2005). Why quality? iso 9126 software quality metrics (functionality) support by uml suite. ACM SIGSOFT Software Engineering Notes, 30(2), 1–5.
10. Power, A., & Kotonya, G. (2019). Providing fault tolerance via complex event processing and machine learning for iot systems. In: Proceedings of the 9th International Conference on the Internet of Things.
11. Elijah, O., Rahman, T. A., Orikumhi, I., Leow, C. Y., & Hindia, M. N. (2018). An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges. IEEE Internet of Things Journal, 5(5), 3758–3773.
12. Tiwari, R., Sharma, N., Kaushik, I., Tiwari, A. & Bhushan, B., (2019). Evolution of IoT & data analytics using deep learning. In:2019 international conference on computing, communication, and intelligent systems (ICCCIS) (pp. 418–423). IEEE.
13. Nieh, J., & Lam, M. S. (2003). A SMART scheduler for multimedia applications. ACM Transactions on Computer Systems (TOCS), 21(2), 117–163.
14. Kuemper, D. et al. (2018). Valid. IoT: a framework for sensor data quality analysis and interpolation. In: Proceedings of the 9th ACM Multimedia Systems Conference.
15. Jayashree, L. S., Arumugam, S., & Meenakshi, A. R. (2008). A communication-efficient framework for outlier-free data reporting in data-gathering sensor networks. International Journal of Network Management, 18(5), 437–445.
16. Sampathkumar, A., Murugan, S., Elningar, A.A., Garg, L., Kanmani, R. & Malar, A., (2020). A novel scheme for an IoT-based weather monitoring system using a wireless sensor network. In Integration of WSN and IoT for smart cities (pp. 181–191). Springer, Cham.
17. Kishorebabu, V., & Sravanthi, R. (2020). Real time monitoring of environmental parameters using IoT. Wireless Personal Communications, 112(2), 785–808.
18. Tiwari, M. M., Narang, D., Goel, P., Gadhwal, A., Gupta, A., & Chawla, A. (2020). Weather monitoring system using IoT and cloud computing. Weather, 29(12s), 2473–2479.
19. Nisa, I. U., & Ahsan, S. N. (2015). Fault prediction model for software using soft computing techniques. In: 2015 International Conference on Open Source Systems and Technologies (ICOSST). IEEE.
Validity as a Measure of Data Quality in Internet of Things Systems

20. Ravichandran, J., & Arulappan, A. I. (2013). Data validation algorithm for wireless sensor networks. In: International Journal of Distributed Sensor Networks 9(12), 634278.
21. Luo, T., Huang, J., Kanhere, S. S., Zhang, J., & Das, S. K. (2019). Improving IoT data quality in mobile crowd sensing: A cross validation approach. IEEE Internet of Things Journal, 6(3), 5651–5664.
22. Kumari, A., Tanwar, S., Tyagi, S., & Kumar, N. (2019). Verification and validation techniques for streaming big data analytics in internet of things environment. IET Networks, 8(3), 155–163.
23. Klima, M. et al. (2020). Quality and reliability metrics for IoT systems: a consolidated view. International Summit Smart City 360°. Springer, Cham.
24. Karkouch, A. et al. (2016). Data quality in internet of things: A state-of-the-art survey. Journal of Network and Computer Applications 73(2016), 57–81.
25. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211–218.
26. Byabazaire, J., O’Hare, G., & Delaney, D. (2020). Data quality and trust: Review of challenges and opportunities for data sharing in IoT. Electronics, 9(12), 2083.
27. Kaiser, M., Klier, M., & Heinrich, B. (2007). How to measure data quality? A metric-based approach. ICIS 2007 Proceedings, pp. 108.
28. Buchholz, T., Küpper, A., & Schifflers, M. (2003). Quality of context: What it is and why we need it. Workshop of the HP OpenView University Association.
29. Thomas, M. O., & Rad, B. B. (2017). Reliability evaluation metrics for internet of things, car tracking system: a review. Int. J. Inf. Technol. Comput. Sci.(IJITCS) 9(2), 1–10.
30. https://mausam.imd.gov.in/
31. http://amsskolkata.gov.in/mc/ranchi
32. https://www.timeanddate.com, last accessed: July 2020.
33. Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. PLoS ONE, 13(5), e0194768.
34. Larsen, M. A. D., Petrovic, S., Engström, R. E., Drews, M., Liersch, S., Karlsson, K. B., & Howells, M. (2019). Challenges of data availability: Analysing the water-energy nexus in electricity generation. Energy Strategy Reviews, 26, 100426.
35. Xiong, J., Ren, J., Chen, L., Yao, Z., Lin, M., Wu, D., & Niu, B. (2018). Enhancing privacy and availability for data clustering in intelligent electrical service of IoT. IEEE Internet of Things Journal, 6(2), 1530–1540.
36. Akhrif, O., Benfaress, C., Jai, M.E., El Idrissi, Y.E.B. & Hmina, N. (2021). Completeness based classification algorithm: a novel approach for educational semantic data completeness assessment. Interactive Technology and Smart Education.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mr. Rishiabh Deo Pandey is currently pursuing his Ph.D. under the guidance of Dr. Itu Snigdha from Dept. of Computer Science & Engineering Birla Institute of Technology, Mesra, Ranchi in the field of IoT. He received his M.Tech degree in (of Computer Science & Engineering) from B.I.T Mesra in 2019 and B.Tech degree in Computer Science & Engineering from Shaheed Bhagat Singh State technical Campus, Ferozepur, Punjab in 2016. His area of interests includes IoT and Software Metrics.
Dr. Itu Snigdh received her Ph. D from B.I.T Mesra in the area of Wireless sensor networks, Department of Electronics and Communication Engineering. She received her master’s degree in (Software Engineering) from B.I.T Mesra (Ranchi), India in 2002 and her bachelor’s degree in Electrical Engineering from B.I.T Sindri, India in 1996. She joined the Department of Computer Science and Engineering at BIT Mesra in 2003 and is currently working as an Assistant Professor. She is currently a member of the Board of Governors of BIT Mesra. She has authored and coauthored several journal articles, book chapters and conference papers. Her areas of interest include Cyber physical systems, IoT, Software Engineering, Database Management Systems and Wireless Sensor Networks.