

# Advanced data analysis and data mining model for Internet of Things in smart city: A Survey

Dr. Seema.S

Prof., Dept. of Computer Science and Engineering  
M.S.Ramaiah Institute of Technology  
Bengaluru, India  
[seemas@msrit.edu](mailto:seemas@msrit.edu)

Sushma.B

Dept. of Computer Science and Engineering  
M.S.Ramaiah Institute of Technology  
Bengaluru, India  
[sushma.b3@gmail.com](mailto:sushma.b3@gmail.com)

**Abstract**— as information is a vital asset, it is important to improve information quality and increment the adequacy of the information. Presently a day's exercises and basic leadership in an association depends on information and data acquired from information examination, which gives different administrations to developing solid and exact process. In any case, recognizing different parts of information quality from definition, measurements, types, methodologies, strategies are basic to prepare techniques and procedures for enhancing information. The genuine reason for poor information quality can be ascribed to an absence of supporting business forms and inadequate investigation procedures. Superb information can expand open door for accomplishing top administrations in an association. In the meantime information quality appraisal is a precondition for advising the clients about the conceivable employments of the information. This paper audits the work done on information quality evaluation to enhance the nature of the information and considered multi layer information digging model for Internet of Things to address the issues in information accumulation. Give answer for the open difficulties in information gathering layer by embracing proposed information quality evaluation process with input system for information quality that has great expansibility and versatility and can address the issues of IoT information quality appraisal.

**Keywords:** Information; Information quality; quality assessment; quality dimensions

## I. INTRODUCTION

Numerous critical innovative changes have happened in the data innovation industry since the start of 21st century, for example, distributed computing, the web of things, and informal communication. The advancement of these advances has made the measure of information increment ceaselessly and aggregate at exceptional speed. Presently the measure of worldwide information is developing exponentially. The information is never again the GB and TB, yet the PB, EB and ZB [1]. Moreover the improvement and use of IoT information have been spread broadly in medicinal services, ventures, shrewd city, agribusiness, poultry and cultivating and so on. IoT is required to empower clients on mission basic administrations like vitality, portability, legitimate and law based soundness. Brilliant items are not just drivers for change as far as substance, and applications. Given their capacity to possibly change in capacity since they can be carefully improved and redesigned they may get troublesome conceivably that could prompt genuine repercussions. IoT can be relied upon to contain tremendous quantities of sensors gathering and passing on information about ecological conditions, physiological

estimations, and machine operational information. Notwithstanding the figuring gadgets that clients utilize today, for example, workstations, recreations consoles and advanced mobile phones there will be numerous gadgets and machines with installed processors running applications (keen things) that individuals make utilization of. Many shrewd things will likewise be fit for activation to accept physical activities because of utilization control. Brilliant things are imagined to give social insurance, local capacities. Savvy city is an incredible utilization of IoT [3]. Savvy reconnaissance, computerized transportation, more astute vitality the board frameworks, water dissemination, urban security and natural checking all are instances of IoT applications for shrewd city. IoT will take care of real issues looked by the general population living in urban areas like contamination, traffic clog and lack of vitality supplies and so forth items like cell correspondence empowered savvy paunch waste will send cautions to metropolitan administrations when a canister should be purged. By introducing sensors and utilizing web applications subjects can discover free accessible stopping openings over the city. Likewise the sensors can distinguish meter altering issues, general breakdowns and any establishment issues in the electric framework [5]. By quickly procuring and investigating IoT information from different sources [3] and with different utilizations analysts have understood that this monstrous measure of data has benefits for understanding client needs, enhancing administration quality, and foreseeing and avoiding dangers. Anyway the utilization and investigation of IoT information must be founded on precise and fantastic information. In this way we broke down the difficulties looked by IoT information quality and proposed a quality evaluation system for it [1].

## II. BACKGROUND

### A. History of information quality

In 1950's scientists started to think about quality issues, particularly for the quality items and a progression of definitions, for instance, quality is "how much a lot of intrinsic attributes satisfy the necessities" and "readiness for use" were distributed. Later with the quick improvement of data innovation, look into swung to the investigation of information quality. Research on information quality began abroad during the 1990s, and numerous researchers proposed diverse meanings of information quality and quality measurements. The Total Data Quality Management Group of MIT University driven by Professor Stuart Madnick has done top to bottom research in the information quality zone. They characterized information quality as readiness for use

and proposed the information quality judgment relies upon information customers. In the meantime, they characterize information quality measurement as a lot of information quality characteristics that speak to a solitary angle or build of information quality [1]. They utilized a two phase review to recognize four classifications (Intrinsic DQ, Accessibility DQ, Conceptual DQ and Representational DQ) [5] containing sixteen information quality measurements (Accuracy, Objectivity, Believability, Reputation, Access, Security, Relevancy, Value included, Timeliness, Completeness, Amount of information, Ease of control, Interpretability, Ease of understanding, Concise portrayal, Consistent portrayal) [2].

#### *B. Papers on information quality*

In 1999, Alexander and Tate have portrayed six assessment criteria (Authority, Accuracy, Objectivity, Currency, Coverage gathering of people, and communication/exchange highlights for web information. For data recovery Gauch in 2000 proposed six quality measurements including Currency, accessibility, Information to communication proportion, Authority, Popularity, and Cohesiveness [2].

Shank and Corbitt in 1999 considered information quality and set up an emotive based edge work for information quality with four dimensions and aggregate of eleven measurements and the recurrence with which they are incorporated into the distinctive information quality/data quality edge works. At that point they displayed IQIP (Identify, Quantify, Implement and Impeccable) as a way to deal with dealing with the decision and execute of value related calculations of a web slithering internet searcher [3].

As indicated by the U.S. National Institute of Statistical Sciences (NISS) in 2001, the standards of information quality are: 1. Information are an item, with clients, to whom they have both expense and esteem; 2. As an item information have quality, coming about because of the procedure by which information are produced; 3. Data quality relies upon numerous components including the reason for which information are utilized, the client, the time and so on [1].

Research in China on information quality started later than research abroad. The 63rd Research Institute of the PLA General Staff Headquarters made information quality research bunch in 2008. They talked about essential issues with information quality, for example, Definition, Error sources, enhancing approaches [1], and so on.

The PC Network Information Center of the Chinese Academy of Sciences proposed information quality appraisal techniques and file framework (Data Application Environment Construction and Service of the Chinese Academy of Sciences, 2009) [8] in which information quality is isolated into three classes including outer shape quality, content quality, and the utility of value. Every class is subdivided into quality attributes and an assessment record [1].

Xi'an Jiaotong University (2011) set up an exploration gathering of data quality that examined the difficulties and significance of guaranteeing the nature of enormous

information and reaction measures in the parts of process, innovation and the executives (Zong and Wu 2013) [1].

Educator Batini Carlo, and group from Italy (2011), proposed a Heterogeneous Data Quality Methodology (HDQM) for information quality appraisal and enhancement that considers a wide range of information oversight in an association, specifically organized information spoke to in databases, semi organized information typically spoke to in XML, and unstructured information spoke to in archives [12].

Pascal Hitzler (2012), USA, present the consequences of an efficient survey of methodologies for evaluating the nature of Linked Data. The improvement and institutionalization of semantic web advances has brought about a remarkable volume of information distributed on the web as Linked Data. Characterized 18 quality measurements and 69 measurements for quality appraisal of connected information [14].

Fatimah Sidi and group, University Putra, Malaysia (2013), proposed a structure to build information quality measurements connections and can be utilized to assess information quality in a data framework to enhance the included procedure [15]. The examination centers around four basic quality measurements; exactness, fulfillment, consistency and convenience [15].

Lulea University, Sweden (2016), directed research on information quality appraisal, as per them DQ appraisal is the logical and factual assessment of information to decide whether they meet the arranging goals of a venture and are of right sort, quality and amount to help their proposed use. DQ evaluation is process is partitioned into emotional and target appraisals. Abstract appraisal depends on client assessment and reviews while target evaluation depends on measurements that can be determined by estimating DQ traits [18].

Vivek Teegalapally and group from India (2016), considers information quality appraisal by and by can address the quality issues, for example, conflicting information and missing information [13]. Yet, to investigate significantly more quality issues they thought about after quality measurements; solid and precise, usefulness, openness, methodological soundness and trustworthiness [13]. To utilize abstract and also target quality evaluation they proposed the accompanying advances; 1. Emotional and target tests on the quality evaluations of information and functionality, 2. Acquiring the disparities and deciding their causes, 3. Finding a way to enhance the information and enhance it. The utilization and investigation of information must be founded on exact and fantastic information, which is a vital condition for creating an incentive from the information [13].

In outline the current examinations center around arrangement of investigations of web information quality and concentrates in explicit regions, for example, science, prescription, geophysics, media communications, logical information and so forth. Huge information as a developing innovation so as IoT information procures increasingly more consideration yet additionally needs inquire about outcomes in setting up IoT information quality and evaluation

strategies under multi-source and multi-demonstrate conditions [1].

### III. MOTIVATION

Low information quality is unsafe in light of the fact that it can prompt wrong or missing choices, techniques and tasks. It can log jam advancement procedures, and misfortunes for associations caused by low information quality are assessed to lie more than billions of dollars for every year. Awful information is an enormous issue 60% of undertakings experiences the ill effects of information quality issues [13], 10 – 30% of information in authoritative databases is off base and individual reports of inadequate, mistaken and uncertain hierarchical information are various. In the time of IoT the measure of information is expanding hugely and separating pertinent and top notch information gets increasingly more essential [3]. While the requirement for information quality appraisal and enhancement methodologies for IoT information was perceived no solid way to deal with evaluating the nature of IoT information was recommended yet [13]. We fill this hole and give information quality measurements and markers for information.

### IV. INTERNET OF THINGS DATA MINING MODEL

As indicated by the design of IoT and information mining outline work of RFID [3], there exists a following multi layer information digging model for IoT as appeared in Fig 1, which is isolated into four layers: information gathering layer, information the executives layer, occasion preparing layer and information mining administration layer [2]. Among them information accumulation layer embraces gadgets, for example, RFID Reader and sinks and so on. To gather different article's information, which are RFID stream information, GPS information, satellite information, sensor information and so forth [2]. Diverse sort of information requires distinctive information gathering technique. During the time spent information accumulation there are a few difficulties that should be tended to.

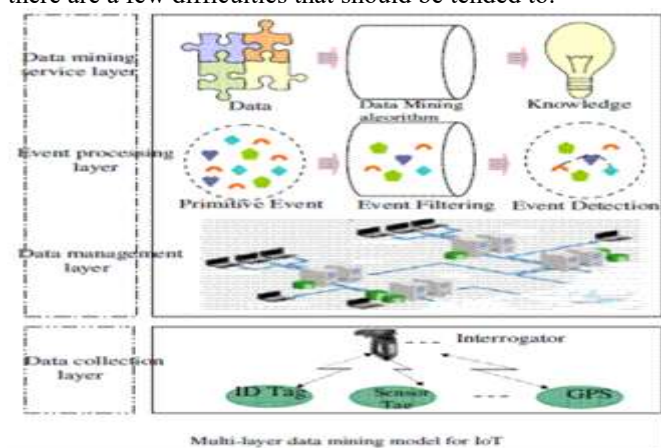


Fig1. IoT Multi-layer data mining model

### V. OPEN ISSUES IN DATA/INFORMATION COLLECTION

To get to, removing expansive scale information from various information stockpiling areas need to manage the assortment, heterogeneity, and clamor of the information, and it is a major test to discover the blame and

significantly harder to address the information [14]. When we lead information gathering from keen objects of IoT, the exceptional needs of savvy articles ought to be considered. For instance, on the off chance that we need to gather information from dispersed sensor organize, vitality productivity, versatility and adaptation to internal failure ought to be considered [2].

- The information got may have quality issues, for example, information blunders, missing data, irregularities, clamor and so forth [1]. The IoT will deliver gigantic information of brilliant items. Hence, it is important to think about how to oversee information of IoT viably and how to actualize preparing advantageously [2]. Information of brilliant items has its own attributes: 1. In nature of IoT, gadgets, for example, RFID and sensors will deliver gigantic information streams, 2. Information of savvy objects is probably going to incorrect, and normally is time-related or area related, 3. Information of shrewd articles will in general have its very own understood semantics. So it is important to perceive certain semantics of information [9].

- Poor information quality will prompt low information use productivity and notwithstanding bring genuine basic leadership botches [1]. Low quality information result in client disappointment, lost income and greater expenses related with the extra time required to accommodate information. This can prompt a decrease in the framework's believability and increment the danger of resistance with controls. It additionally builds buyer costs, increments charges, decline investor esteem and can cause mission disappointment because of basic leadership botches [9].

- The decent variety of information sources brings plenteous information types and complex information structures and builds the trouble of information coordination [1]. Previously, endeavors just utilized the information produced from their very own business frameworks, for example, deals and stock information. However, presently, information gathered and investigated by endeavors has outperformed this extension. Huge information sources are wide, including: 1. Informational collections from the web and versatile system; 2. Information from the Internet of things; 3. Information gathered by different businesses; 4. Logical test and observational information, for example, high-vitality material science test information, organic information, and space perception information. These sources create rich information types. One information type is unstructured information, for instance reports, video, sound and so forth [1]. The second sort is semi organized information, including: programming bundles/modules, spread sheets and monetary reports. The third is organized information. The amount of unstructured information involves over 80% of the aggregate sum of information in presence. There are clashes and conflicting marvels among information from various sources. On account of little information volume, the information can be checked by a manual pursuit or programming, even by ETL (Extract, Transform, and Load) or ELT (Extract, Load, and

Transform). Anyway these strategies are pointless when handling PB level even EB level information volume [1].

- Data volume is gigantic and it is exceptionally hard to pass judgment on information quality inside a sensible measure of time [1]. After the mechanical upset, the measure of data ruled by characters multiplied like clockwork. After 1970, the measure of data multiplied at regular intervals. Today, the worldwide measure of data can be multiplied at regular intervals. In 2011, the measure of worldwide information made and duplicated came to 1.8ZB. it is hard to gather, clean, incorporate, lastly get the fundamental fantastic information inside a sensible time period [1]. Since the extent of unstructured information in IoT information is high. It will require a great deal of investment to change unstructured sorts into organized kinds and further process the information. This is incredible test to the current methods of information handling quality [1].

- Data change quickly and the opportuneness of information is short, which requires higher prerequisites for handling innovation. Due the quick changes in IoT information the 'convenience' of a few information is short [1]. On the off chance that organizations can't gather the required information continuously or manage the information needs over quite a while, at that point they may acquire obsolete and invalid data. Handling and examination dependent on this information will create futile or deceiving ends, in the end prompting basic leadership botches. At present, constant handling and investigation programming for IoT is still in enhancement stages; extremely powerful business items are few [1].

## VI. INFORMATION QUALITY

So as to survey information quality, above all else an unmistakable picture of information quality is required [16]. Quality methods the totality of the attributes of a substance that bear on its capacity to fulfill expressed and inferred needs [8]. Numerous analysts saw that the advantages of information depend intensely on culmination, accuracy, consistency and opportuneness. These properties are known as information quality measurements [6]. The real test of information quality research is to characterize information quality from the buyer's perspectives as far as qualification for use and to recognize measurements of information quality as indicated by that definition. In excess of a hundred of information quality measurements were revealed in early research works, however a significant number of these measurements are much valuable in numerous application territories. A portion of the numerous critical proportions of information quality are pertinence, exactness, opportuneness, availability, utility, interoperability, fulfillment, intelligence and similarity. Then again, we require a technique to evaluate information quality and after that convey the outcomes to the client as report. Quality revealing is tied in with giving the reports that contain data about the nature of information and gives data on the lion's share quality properties of a data item so the client would have the capacity to assess item quality. In the ideal case

quality reports depend on quality markers [19]. An information quality appraisal format is normally required that determines the data every commentator ought to gather for the objective information framework [7]. There are numerous quality appraisal techniques, cross space investigation can be connected to information mix situations with many source frameworks. It empowers the recognizable proof of repetitive information over the tables from various, and now and again even the equivalent, sources [7]. Information approval calculations check if esteem or a lot of qualities is found in a reference informational index. Space examination can be connected to check if explicit information esteem is inside a specific area of qualities. Coordinating calculations are utilized to recognize copies. What's more, numerous different techniques that are as of now being used [7].

There are a few elements which make information quality issues. Some of them are database combination, information cleansing and constant interfacing [3]. In this way it is basic to keep up information quality. Following are the four different ways utilized in associations to enhance information quality 1. Information profiling 2. Information standardization 3. Semantic metadata the executives 4. Information quality firewall. [13]. Information quality devices are additionally used to address information quality issues. Coming up next are the at present utilized information quality devices in the business Parsing and standardization

1. Generalized cleansing
2. Matching
3. Profiling
4. Monitoring
5. Enrichment

## VII. INFORMATION QUALITY ASSESSMENT

An information quality measurement is a trademark or part of data for characterizing data and information prerequisites. It offers a path for estimating and overseeing information quality and in addition data [8]. From the exploration point of view, there are different quantities of measurements for information quality, we consider the accompanying information measurements best to enhance information quality [8]. In the meantime, each measurement was separated into numerous common components related with it, and every component has its very own comparing quality markers. Along these lines various leveled quality principles were utilized for assessment [1].

TABLE I INFORMATION QUALITY DIMENSIONS

Dimension	Definition
Availability	Degree of accommodation for clients to get information and related data [11].
Usability	Whether the information are helpful and address clients issues.
Reliability	Whether we can confide in the information [7].

Relevance	Describes the level of relationship between's information substance and clients desires or requests.
Presentation Quality	Valid portrayal technique for the information, which enables clients to completely comprehend the information [1].

TABLE II INFORMATION QUALITY ELEMENTS

Elements	Definition
Accessibility	It alludes to the trouble level for clients to get the information [1].
Timeliness	It's a deferral from information age and procurement to use. Information ought to be accessible inside this deferral to take into consideration significant investigation [1].
Authorization	It alludes to whether an individual or association has the privilege to utilize information [1].
Credibility	It is utilized to assess non numerical information. It alludes to the target and abstract parts of the acceptability of source or message [1].
Definition/Documentation	Consists of information particular, which incorporates information name, definition, scopes of substantial qualities, standard configurations, business rules and so forth [1].
Metadata	With the expansion of information sources and information types, since information buyers twist the importance of normal wording and ideas of information, utilizing information may bring dangers. In this manner information makers need to give metadata portraying distinctive parts of the datasets to lessen the issues caused by misconception or irregularities [1].
Accuracy	It is contrasted with realized reference esteem. In a few circumstances, exactness can be effectively estimated, for example, sex, which has just two clear qualities. In any case, in different cases, there is no realized reference esteem, making it hard to quantify precision [1].
Consistency	Whether the sensible connection between associated information is right and finishes [9].
Integrity	It has generally extraordinary implications relying upon the explicit setting. In a database, information with uprightness are said to have a total structure. In data security information respectability implies keeping up and guaranteeing the exactness and consistency of information over its whole life cycle. Means information can't be changed in an unapproved or undetected way [1].

Completeness	It implies that the estimations of all parts of a solitary datum are legitimate [1].
Auditability	From the point of view of review application, the information life cycle incorporates three stages; information age, information gathering and information use. However, here auditability implies that reviewers can decently assess information precision and trustworthiness inside normal time and labor limits amid the information use stage [1].
Fitness	This has two dimension necessities; 1. The measure of got to information utilized by clients and 2. How much the information created matches clients needs in the part of pointer definition, components, arrangement, and so on [1].
Readability	Defined as the capacity of substance to be effectively disclosed by known or all around characterized terms, properties, units, codes, truncations, or other data [1].
Structure	More than 80% of all information is unstructured, in this way organized alludes to the dimension of trouble in changing semi organized or unstructured information to organized information through innovation [1].

TABLE III INFORMATION QUALITY INDICATORS

Elements	Indicators
Accessibility	Whether an information get to interface is given [4]
	Information can be effectively made open or simple to buy [4]
Timeliness	Within a given time whether the information touch base on time [4]
	Regardless of whether information are routinely refreshed [4]
Credibility	Experts or masters routinely review and check the accuracy of the information content [1]
	Information exists in the scope of known or worthy qualities [1]
Accuracy	Data gave are exact [1]
	Data or information portrayal won't cause vagueness [1]
Consistency	After information have been handled, their ideas, esteem spaces, designs still match as before preparing [4]
	Amid certain time, information stay steady and unquestionable [4]

Integrity	Data design is clear and meets the criteria [4]
	Information are predictable with auxiliary trustworthiness [1]
Completeness	Whether the inadequacy of a segment will affect utilization of the information for information with multi segments [1]
	Regardless of whether the lack of a segment will affect information precision and honesty [1]
Fitness	Most datasets recovered are inside the recovery topic clients require [4]
	Data subject furnishes matches with clients recovery topic [4]
Readability	Data are clear and justifiable [4]
	It is anything but difficult to pass judgment on that the information gave address issues [4]

At last a suitable quality evaluation strategy for IoT information is important to make substantial determinations [1]. There exists a proposition for compelling information quality appraisal process with a dynamic input instrument dependent on enormous information's own attributes as appeared in figure [2].

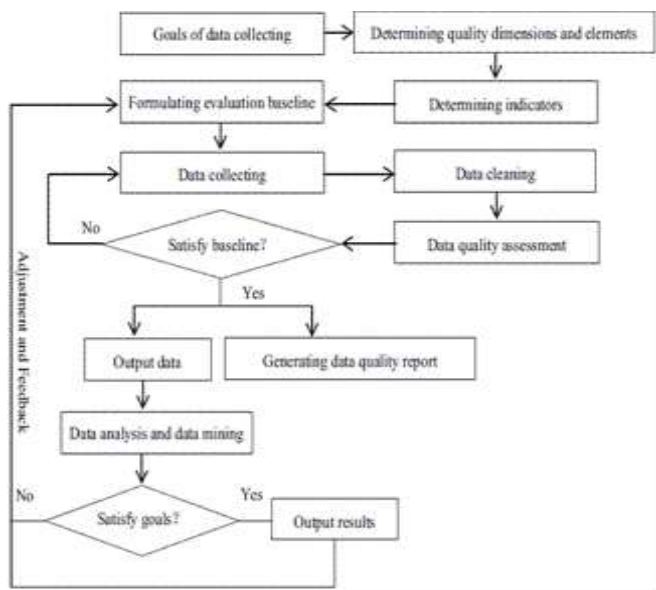


Fig2. Information Quality Assessment Process

Process begins by deciding the objectives of information gathering. In various business situations, the determination of information quality components will vary. For instance, for internet based life information auspiciousness and exactness are two imperative quality highlights. Be that as it may, on the grounds that it is hard to straightforwardly pass judgment on precision, some extra data is expected to pass judgment on the crude information, and other information sources fill in as enhancements or proof. Along these lines, validity has turned into an imperative quality measurement [1]. So as to promote quality evaluation, we have to pick quality markers for each measurement. These require the

information to conform to explicit conditions or highlights. The plan of appraisal pointers additionally relies upon the genuine business condition [1].

Every quality measurement needs extraordinary estimation instruments, procedures, and procedures, which prompt contrasts in appraisal time, expenses, and HR. Picking measurements, address the issues that can well characterize task's degree. The fundamental evaluation aftereffects of information quality measurements decide the gauge while the rest of the appraisal as a major aspect of the business procedure is utilized for nonstop location and data enhancement [1].

After the quality appraisal planning is finished, the procedure enters the information obtaining stage. There are numerous approaches to gather information, including information incorporation, web crawlers, operator strategies and so on a significant part of the information gathered isn't in every case great so we have to enhance information quality beyond what many would consider possible under these conditions without huge increment in securing cost. IoT information sources are wide and information structures are mind boggling [1]. The information got may have quality issues, for example, information mistakes, missing data, irregularities, and clamor and so on the reason for information cleaning is to recognize and expel blunders and irregularities from information so as to enhance their quality [1].

At that point procedure enters information quality evaluation and observing stages. The center of appraisal is the manner by which to assess each measurement. The present strategy has two classes, subjective and quantitative strategies [1]. After evaluation the information can be contrasted and the gauge for the quality appraisal built up above. In the event that the quality accords with the gauge standard, a subsequent information examination stage can be entered, and a quality report will be created [1]. Generally if the quality neglects to fulfill the benchmark standard, it is important to gain new information. On the off chance that the investigation results meet the objective, the outcomes are yielded and sustained back to the quality appraisal framework in order to give better help to the following round of evaluation [1]. On the off chance that outcomes don't achieve the objective, the information quality appraisal pattern may not be sensible, and we have to modify it in a convenient manner so as to acquire results in accordance with our objectives [1].

Our commitment to the current proposition will be to develop an information quality evaluation demonstrate for IoT information and arrangement of weight coefficient for every appraisal marker. Also, to build up a calculation used to make a functional evaluation of IoT information quality that can be connected to any smart city application [11].

#### CONCLUSION

The present Internet of Things stages and shrewd city applications regularly center on giving network and correspondence and furthermore bolster information gathering and utilize information investigation. In this paper we checked on the work done on information quality

appraisal to enhance the nature of the information. We considered multi layer information digging model [2] for Internet of Things to address the issues in information accumulation and gives answer for the open difficulties in information gathering layer by receiving proposed information quality appraisal process with criticism component and further we stretch out research to build an information quality evaluation display for IoT information and arrangement of weight coefficient for every evaluation pointer [11]. What's more, to build up a calculation used to make a down to earth evaluation of IoT information quality [11] that can be connected to any smart city application.

#### REFERENCES

- [1] Cai, L and Zhu, Y 2015 The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14: 2, pp. 1-10, DOI: <http://dx.doi.org/10.5334/dsj-2015-002>.
- [2] Alexander, J. E., & Tate, M. A. *Web wisdom: How to evaluate and create information on the web*, Mahwah, NJ: Erlbaum.
- [3] Shanks, G., & Corbitt, B. (1999) Understanding data quality: Social and cultural aspects. *Proceedures of the 10th Australasian Conference on Information Systems*, Wellington: MCB University Press Ltd., pp 785–797.
- [4] Cooper J, James “A. Challenges for Database Management in the Internet of Things,” *IETE Tech Rev.* 2009. 26:320-9.
- [5] P. Borek, A.Woodall. A classification of data quality assessment methods. *international conference on information quality*, 2011.
- [6] Hector Gonzalez, Jiawei Han, Xiaolei Li. “FlowCube Constructuing RFID FlowCubes for Multi-Dimensional Analysis of Commodity Flows,” *VLDB 2006*: 834-845.
- [7] Shahram Dustdar, Reinhard Pichler, Vadim Savenkov, and Hong-Linh Truong. Quality-aware service-oriented data integration: requirements, state of the art and open challenges. *ACM SIG-MOD Record*, 41(1):{11,19}, 2012.
- [8] M. Heravizadeh, J. Mendling, M. Rosemann, "Dimensions of business processes quality (QoBP)," 2009, pp. 80-91.
- [9] D. McGilvray, *Executing data quality projects: Ten steps to quality data and trusted information*: Morgan Kaufmann, 2008.
- [10] Joshua Huang. (2009) “RFID Data Mining: Opportunities and Challenges,” [Online]. Available: <http://homepage.fudan.edu.cn/~wdzhao/rfid.html>.
- [11] Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009). “Methodologies for data quality assessment and improvement”. *ACM Computing Survey*, Vol. 41, No. 3, pp1-52.
- [12] Batini Carlo<sup>1</sup>, Barone Daniele<sup>1</sup>, Cabitza Federico<sup>1</sup> and Grega Simone<sup>2</sup> : A DATA QUALITY METHODOLOGY FOR HETEROGENEOUS DATA, *International Journal of Database Management Systems ( IJDMs )*, Vol.3, No.1, February 2011 DOI: 10.5121/ijdms.2011.3105 60
- [13] Vivek Teegalapally<sup>1</sup>, Kiran Dhote<sup>2</sup>, Vamsi S. Krishna<sup>3</sup>, Shubham Rao<sup>4</sup>: Survey on Data Profiling and Data Quality Assessment for Business Intelligence . *International Research Journal of Engineering and Technology (IRJET)* e-ISSN: 2395 -0056 Volume: 03 Issue: 11 | Nov -2016 [www.irjet.net](http://www.irjet.net) p-ISSN: 2395-0072 , 2016, IRJET | Impact Factor value: 4.45 | ISO 9001:2008 Certified Journal | Page 1246.
- [14] Pascal Hitzler, Quality Assessment for Linked Data: A Survey, A Systematic Literature Review and Conceptual Framework. *Semantic Web 1* (2012) 1–5 1 IOS Press.
- [15] Payam Hassany Shariat Panahy, Fatimah Sidi, Lilly Suriani Affendey, Marzanah A. Jabar, Hamidah Ibrahim and Aida Mustapha: A Framework to Construct Data Quality Dimensions Relationships. *Indian Journal of Science and Technology* | Print ISSN: 0974-6846 | Online ISSN: 0974-5645, [www.indjst.org](http://www.indjst.org) | Vol 6 (5) | May 2013
- [16] Shen Bin, Liu Yuan, Wang Xiaoyi : Research on Data Mining Models for the Internet of Things, 978-1-4244-5555-3/10/\$26.00, 2010 IEEE.
- [17] Panahy, Payam Hassany Shariat, Fatimah Sidi, Lilly Suriani Affendey, and Marzanah A. Jabar. "The effect of information quality measurements on business process enhancement", 2014 fourth World Congress on Information and Communication Technologies (WICT 2014), 2014
- [18] Amrapali Zaveri, Andrea Maurino, Laure-Berti Equille. "Web Data Quality", *International Journal on Semantic Web and Information Systems*, 2014.
- [19] Carlo Batini. "Techniques for information quality evaluation and enhancement", *ACM Computing Surveys*, 07/01/2009
- [20] Baba Piprani. "A Model for Data Quality Assessment", *Lecture Notes in Computer Science*, 2008
- [21] Anisa Rula."DC Proposal: Towards Linked Data Assessment and Linking Temporal Facts", *Lecture Notes in Computer Science*, 2011