iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5273

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Review of Data Analysis Algorithm and its Applications

Kavyashree C[1], Kavya Shree S[2], Likhitha Singh M[3], Vaishnavi J[4], A RafegaBeham[5]
[1, 2, 3, 4]*Student, Dept. of ISE, New Horizon College of Engineering, Bangalore, India*
[5]*Senior Assistant Professor, Dept. of ISE, New Horizon College of Engineering, Bangalore, India*

Abstract: ***As of late, the web application and correspondence have seen a considerable measure of advancement and notoriety in the field of Information Technology. These web applications and correspondence are consistently producing the vast size, distinctive assortment and with some real troublesome multifaceted structure information called huge information. Thus, we are currently in the time of huge programmed information gathering,methodically getting numerous estimations, not knowing which one will be applicable to the marvel of intrigue. This paper advances the 2 most utilized data mining algorithms utilized as a part of the examination field which are: SVM and Apriori. With every calculation, an essential clarification is given with a continuous case, and every calculation advantages and disadvantages are weighed exclusively. Numerous analysts are doing their exploration in dimensionality lessening of the enormous information for powerful and better investigation report and information representation.***
Keywords: Data, Analysis, algorithm, application, intelligent system.

I. INTRODUCTION

Information is delivered in such plentiful sums that today the need to break down and comprehend this information is of the substance. The gathering of information is accomplished by bunching calculations and would then be able to additionally be broke down by mathematicians and by enormous information investigation strategies. This bunching of information has seen a wide scale use in interpersonal organization imaging investigation, statistical surveying, restorative and so on. Today, framework and individuals utilize the web with an exponential age of huge size of information. The extent of information on the web is estimated in Exabyte (EB) and Petabytes (PB). By 2025, the expectation is that the Internet will outperform the cerebrum size of everybody living in the entire world. This firm development of information is a result of advances in computerized sensors, calculations, correspondences, and capacity that have made extensive social affairs of information. [8] The name BigData had been concocted, by Roger Magoulas a scientist, to portray this peculiarity. Enormous information, by definition, is a term used to depict an assortment of information - organized, semi-organized and unstructured, which makes it an unpredictable information foundation. This paper plans to examine a portion of the distinctive investigation strategies and devices which can be connected to huge information, and additionally the open doors gave using huge information examination in different choice areas.

II. ALGORITHMS FOR DATA ANALYSIS

An algorithm in information mining (or machine learning) is an arrangement of heuristics and estimations that makes a model from information. To make a model, the calculation initially breaks down the information you give, searching for particular sorts of examples or patterns. The calculation utilizes the consequences of this investigation over numerous emphases to locate the ideal parameters for making the mining model. These parameters are then connected over the whole informational collection to remove noteworthy examples and itemized measurements.

The mining model that a calculation makes from your information can take different structures, including:
An arrangement of groups that depict how the cases in a dataset are connected. A choice tree that predicts a result, and portrays how unique criteria influence that result. A numerical model that figures deals. An arrangement of tenets that depict how items are assembled together in an exchange, and the probabilities that items are obtained together.

A. Bolster vector machine (SVM)

It characterizes the information into two classes from the hyperplane. SVM doesn't utilize choice tree. A hyperplane is a capacity like the condition of a line, y=mx+b. It's a straightforward order of an undertaking with only two highlights. SVM makes sense of the perfect hyperplane which isolates the information into two classes. Case for SVM calculation: There are a bundle of red and blue balls on the table. The balls aren't excessively combined and you could take a stick and separate the balls without moving the stick. So by along these lines when another ball is added to the table, by knowing which side of the stick the ball is on, the shade of the new ball can be anticipated. Also, the balls speak to the information focuses and the red and blue balls speak to the two classes. The hyperplane is the stick. SVM makes sense of the capacity of the hyperplane independent from anyone else. The issue is the point at which the balls are blended and a straight stick won't work. Here is the arrangement. Toss the balls noticeable all around and utilize a paper to partition the balls noticeable all around. Lifting up the table is proportional to mapping your information in higher measurements. In such cases, we go from two measurements to three measurements. The second measurement is the table surface and the third measurement is the balls noticeable all around. SVM does this by utilizing part which works in higher measurement. The vast sheet of paper is a capacity for a plane as opposed to a line. In this manner, SVM maps the things into higher measurements and finds a hyperplane to isolate the classes. Edges are for the most part connected with SVM. It is the separation between the hyperplane and the two nearest information focuses from the individual class. Preferred standpoint: Produce extremely exact classifiers, Less finished fitting and powerful to clamor. The restrictions are: SVM is a double classifier. To complete a multi-class grouping, pairwise orders can be utilized (one class against all others, for all classes). Computationally costly, in this way runs slow.[13]

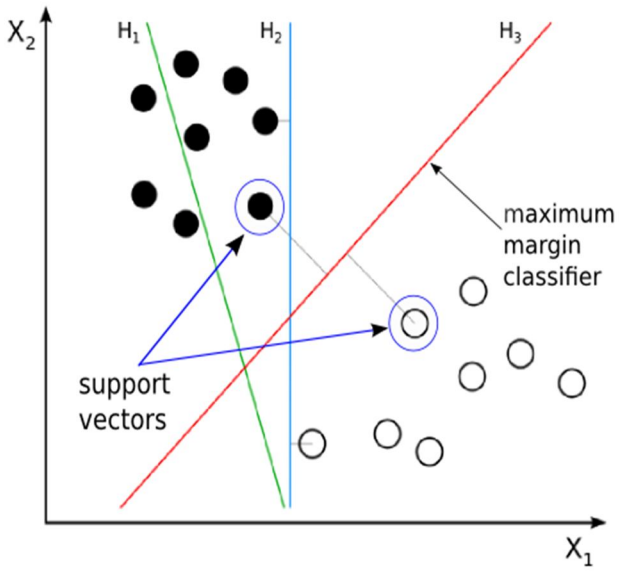


Figure1 It is the straightforward model for speaking to help vector machine procedure. The model comprises of two distinct examples and the objective of SVM is to isolate these two examples.

The help vector machine more often than not manages design grouping that implies this calculation is utilized for the most part to classify the diverse kinds of examples. Presently, there is distinctive sort of examples i.e. Direct and non-straight. Straight examples are designs that are effectively recognizable or can be effortlessly isolated in low measurement though non-direct examples are designs that are not effectively discernable or can't be effectively isolated and henceforth these sort of examples should be additionally controlled with the goal that they can be effortlessly isolated. [14]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

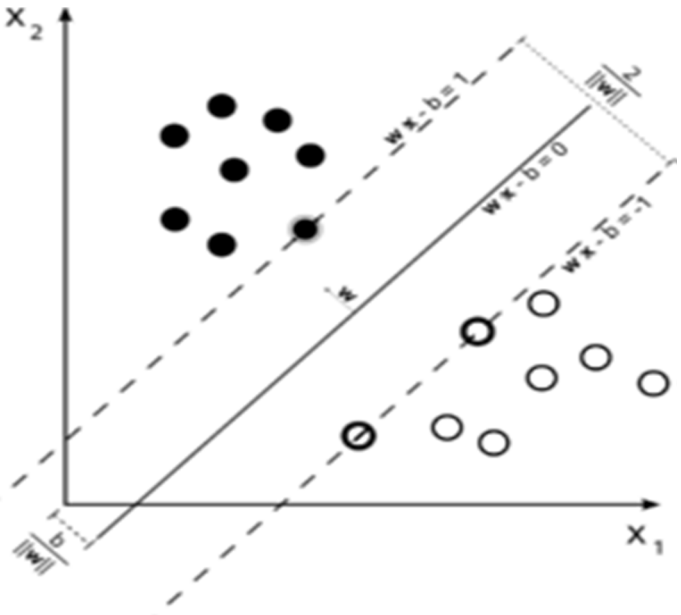


Figure 2 The model consists of three different lines. The line w.x-b=0 is known as margin of separation or marginal line.

B. The Apriori algorithm

A standout amongst the most well known data mining approaches is to discover frequent item sets from an exchange informational collection and infer association rules. Finding frequent item sets (item sets with recurrence bigger than or equivalent to a client determined minimum support) isn't paltry in view of its combinatorial explosion. Once frequent item sets are acquired, it is direct to produce association rules with certainty bigger than or equivalent to a client indicated least confidence.Apriori is a fundamental calculation for finding frequent item sets utilizing candidategeneration[1]. It is portrayed as a level-wise finish seek calculation utilizing hostile to monotonicity of item sets, "if an item set isn't frequent, any of its super set is never frequent". By convention,Apriori accept that items inside an exchange or item set are arranged in lexicographic request. Let the set of frequent itemsets of size *k* be *Fk*and their candidates be *Ck* .Apriori first sweeps he database and searches for frequent itemsets of size 1 by accumulating the count for eachitem and collecting those itemsthat satisfy the minimum support requirement. It then iterates onthe following three steps and extracts all the frequent item sets.

1) Generate *Ck*+1, candidates of frequent itemsets of size *k* +1, from the frequent item setsof size *k*

2) Sweep the database and calculate the support of each candidate of frequent itemsets.

3) Add those itemsets that satisfies the minimum support requirement to *Fk*+1. The Apriori algorithm is shown in Figure. 3. Function Apriori-gen in line 3 generates *Ck*+1from *Fk*in the following two step process:

4) Join step: Generate RK+1, the initial candidates of frequent itemsets of size k + 1 by taking the union of the two frequent itemsets of size k, Pk and Qkthat have the first k−1 elements in common.

RK+1 = Pk ∪ Qk= {iteml, . . . ,itemk−1, itemk, itemk }

Pk = {iteml,i tem2, . . . , itemk−1, itemk}

Qk= {iteml,i tem2, . . . , itemk−1, itemk }

where, iteml<i tem2 <· · · <*itemk*<*itemk*

Prune step: Check if all the itemsets of size *k* in *Rk*+1 are frequent and generate *Ck*+1 byremoving those that do not pass thisrequirement from *Rk*+1. This is because any subsetof size *k* of *Ck*+1 that is not frequent cannot be a subset of a frequent itemset of size*k* + 1. Function subset in line 5 finds all the candidates of the frequent itemsets included in transaction*t*. Apriori, then, calculates frequency only forthe candidates generated this way bysweeping the database.It is evident that Apriori sweeps the database at most *k*max+1 times when the maximum sizeof frequent itemsets is set at *k*max.TheApriori achieves good performance by reducing the size of candidate sets (Figure. 3).However, in situations with very many frequent itemsets, large itemsets, or verylow minimumsupport, it still suffers from the cost of generating a huge number of candidate setsand scanning the database repeatedly to check a large set of candidate itemsets. In fact, it isnecessary to generate 2100 candidate itemsets to obtain frequent itemsets of size 100.

C. Algorithm Apriori

Algorithm 1 *Apriori*

F_l=(Frequent itemsets of cardinality 1);
for($k = 1; F_k \neq \phi; k++$) do begin
  C_{k+1} = apriori-gen(F_k); //New candidates
  for all transactions $t \in$ Database do begin
    C'_t = subset(C_{k+1}, t); //Candidates contained in t
    for all candidate $c \in C'_t$ do
      $c.count++$;
    end
    $F_{k+1} = \{C \in C_{k+1} \quad | c.count \geq$ minimum support $\}$
  end
end
Answer $\cup_k F_k$;

Figure. 3

A considerable lot of the example discovering calculations, for example, decision tree, classification rules and clustering techniques that are much of the time utilized as part of data mining have been created in machine learning research.

III. APPLICATIONS

A. Urban Intelligent Transportation System

At present, there are different levels of traffic congestion in major cities. The existence of such problems has an adverse effect on peoples travel experience while increasing the traffic risk. We have mainly analysed the value and characteristic of big data technology and analysed the urban intelligent transportation system from GPS technology, GIS technology and structure. The technology and structure of urban intelligent transport system are as follows:

B. The key Technologies in Urban in Telligent Transportation System

1) *GPS technology:* In the city smart correspondence framework, the capacity of GPS innovation is essentially to furnish clients with exact situating capacity to meet the constant route need of client
2) GIS technology: In real utilize, GIS innovation can be founded on the real need of clients, with the assistance of GPS innovation to change over the significant geographic data into instinctive frame and sustain it back to clients, so clients can be clients can settle on fitting choices
3) *Communication technology:* The way toward utilizing the city clever transportation framework can be viewed as the procedure of information transmission amongst framework and client through framework channel..[7]

C. The analysis of Intelligent Transportation System Structure

The feasible urban intelligent transportation system structural elements design patterns are as follows:

1) The central system elements- Its functions are charging management, traffic management etc
2) The remote system elements- Its function is to provide real data support for users in remote areas.
3) The elements of the road system- It is mainly responsible for inspecting various commercial vehicles, assessing the road conditions and carrying out parking management
4) The vehicle system elements- Its functions are data acquisition and information management of different vehicles such as buses. [7]

D. The Application of big Data Technology in Urban Intelligent Transportation System

1) The Practical Application Demand Of Urban Intelligent Transportation System Based On Big Data

a) Inclusive needs and meet traffic request demand : It is essentially the same as the fundamental information handling undertakings of big data technology.

b) Modular needs: The structure of astute transportation framework ought to have great particular qualities with a specific end goal to utilize diverse modules to play distinctive capacities.

2) Big data Technology Application Function

a) Massive data acquisition function: To mitigate activity weight and enhance the nature of urban movement administration, the quantity of data gathering gadgets, for example, video checking in urban rush hour gridlock organize has expanded all together

b) Mass data computing capabilities: The use of huge information innovation can make utilization of distributed computing bunch, through the appropriated approach to finish the gigantic information rapid figuring.

c) Massive data retrieval function: It alludes to the attributes of the business information question and the genuine movement information use prerequisite of the clients and tweaking the web search tool of the smart transportation framework and utilizing the huge information innovation to upgrade the inquiry speed of the framework. [7]

E. Robots and Data

Nowadays we are getting closer to what Sci-fi industry taught us with respect to robots, their structure and applications. Regardless of in front of timetable days of mechanical innovation, they are not simply proposed to perform repetitive and non-shrewd errands. By virtue of various circumstances of enhancements in hardware, kinematics, control, etcetera and furthermore movements in Al and related fields, we are seeing awesome progress in the zone. In such way, the present apply independence outfits individuals with workplaces and help with a broad extent of utilization domains from surgery and therapeutic exercises to examination of room and interior sea [1]. Overall, robots may be thought about from a couple of perspectives. each robot administrator for playing out its doled-out errands and exercises ought to oversee data in some casing. Specifically, and as demonstrated by their described purposes, robots:

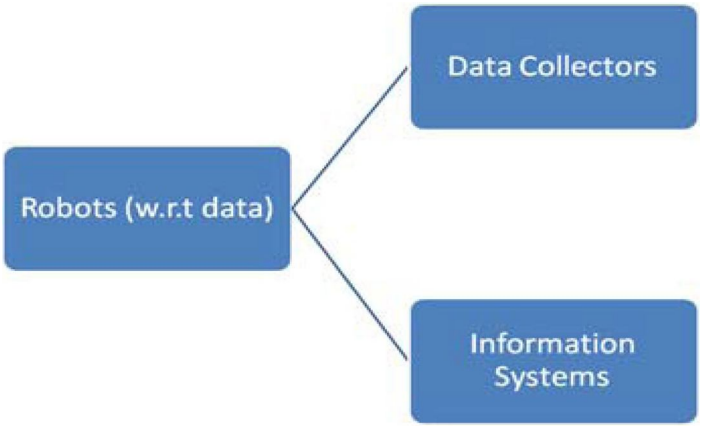


Figure 4. General classification of robots respecting their relationship with data

Robots that have a place with the highest point of the line are a subset of second ones; however as a result of their specific applications it is more quick witted to depicted in an alternate social affair. As a couple of instances of this kind of robots, submerged robots [2], Unmanned Aerial Vehicles (UAVs) [3] and fiasco information gatherer robots [4] could be determined. Robots of second kind may be considered as a kind of complex information system that game plans with different sorts of data and takes preferences of taking care of them to improve its execution in various points. This social event contains a far reaching extent of mechanical masters including self-overseeing and shrewd ones and oversee broad measure of data from various sources. Considering this thought and concerning their relationship with data, robots could be assembled in two significant classes (Figure.4):

1) Robots as data (information) experts (gatherers)

2) Robots as information systems

3) Data mining for robot soccer:

Since the start of mechanical innovation, one of its most captivating and surely understood spaces is robot soccer. The RoboCup 2050 vision - to beat champ of the most recent FIFA World Cup by a gathering of totally self-administering humanoid robot authorities [46]-is considered as a strong motivation for researchers in the field. From a data driven viewpoint, there are two important classes of data in a given Robot Soccer organize: 1) data related to the gathering and 2) data related to the adversary (gathering). Such data consolidate records about limit, position and execution of each pro particularly and gatherings (as Multi administrator systems) when all is said in done, as log archives. (Figure.5) To be more specific, unpretentious components of uses of data mining process in a given robot soccer circumstance may be cleared up as takes after:

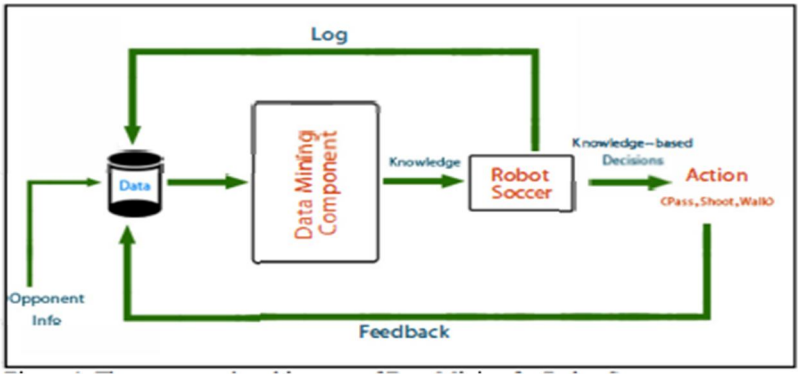


Figure5. The conceptual architecture of Data Mining for Robot Soccer

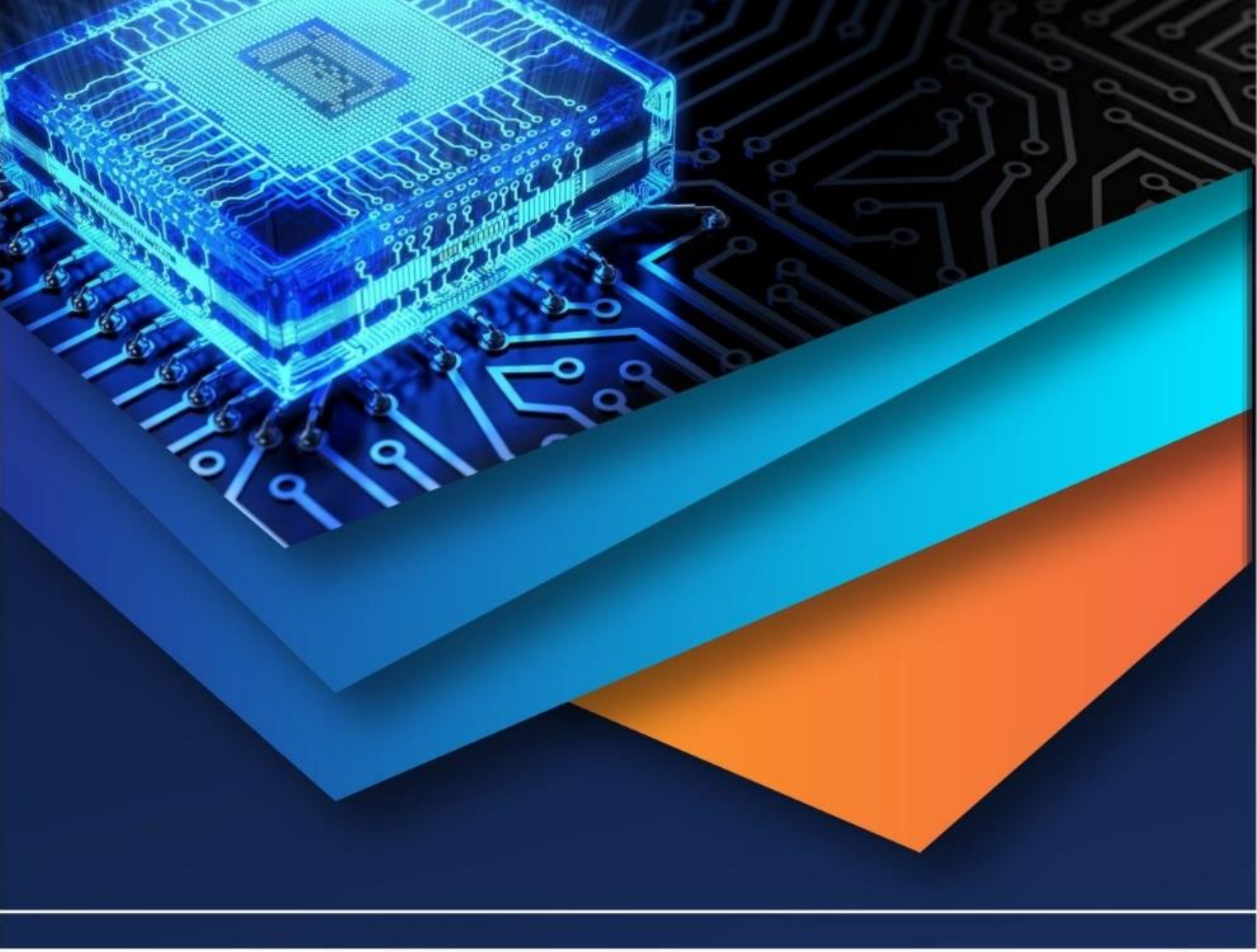
Inside the Game: 1) web mining of the enemy's cases to achieve adaption in method and systems. 2) burrowing the gathering's status for organizing with predefined plans. After the Game: 1) separating gathering's execution all things considered what's more, masters particularly to evaluate the accomplishment rate. 2) using information (events, correspondences, etcetera.) recorded in the sort of log by process mining to think about the structure viability, accuse acknowledgment and conclusion etcetera. The results gotten in the stage will be used to energize the principle arrange (before the entertainment) analytic process.

IV. CONCLUSION

Data mining is a broad area that integrates techniques from several fields including machine learning, statistics, pattern recognition, artificial intelligence, and database systems, for the analysis of large volumes of data. There have been many data mining algorithms rooted in these fields to perform different data analysis tasks. The above analysis shows that the application of big data technology has significantly enriched the practical value of urban intelligent transport system. With the help of cloud computing and clustering mechanism, big data can generate a good amount of data acquisition, mass data retrieval and other functions.

REFERENCES

[1] Giulianotti, et aI., "Robotics in general surgery: personal experience in a large community hospital," Archives of surgery, vol. 138, no. 7, pp. 777-784, July 2003.
[2] S. Zhao, and 1. Yuh, "Experimental study on advanced underwater robot control," Robotics, IEEE Trans Robot, vol. 21, issue. 4, pp. 695-703, August 2005.
[3] P. Corke, et aI., "Autonomous deployment and repair of a sensor network using an unmanned aerial vehicle," In Proceedings of IEEE International Conference on Robotics and Automation, ICRA'04, 2004, Vol. 4, pp. 3602-3608.
[4] F. Matsuno, and S. Tadokoro, Rescue robots and systems in Japan, In Proceedings of IEEE International Conference on Robotics and Biomimetics, ROBIO, 2004, pp. 12-20.
[5] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB conference, pp 487–499
[6] Bayardo Jr, Roberto J. "Efficiently mining long patterns from databases"
[7] Bigdata technology and its analysis of applications in urban intelligent transport system, Liu Yang.
[8] A survey paper on big data analytics, M. D. AntoPraveena; B. Bharathi, 2017 International Conference on InformationCommunication and Embedded Systems (ICICES)
[9] Algorithms in data mining", Springer-Verlag London limited, 2007.
[10] http://rayli.net/blog/data/top-10-data-mining-algorithms-inplain-english/
[11] http://www.kdnuggets.com/2015/05/t[1] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, QiangYang,HiroshiMotoda, "Top 10 op-10-data-miningalgorithms-explained.html
[12] http://www.slideshare.net/Tommy96/top-10-algorithms-in-datamining
[13] http://ijcsit.com/docs/Volume%207/vol7issue1/ijcsit2016070166
[14] http://ijarcsse.com/Before_August_2017/docs/papers/Volume_4/12_December2014/V4I12-0492



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY