

NOVEL 4 CHECKPOINT ANALYSIS ALGORITHM FOR ANALYSIS FACTORS AFFECTING AGRICULTURE PRODUCTION

Abdul Javed¹, Manoj Singh²

¹M.Tech Scholar, ²Head of Department, Department of Computer Science,
Gurukul Institute of Engineering and Technology, Kota, Rajasthan, India.

Abstract: *Agriculture is the foundation of the Indian country. Despite the way that huge regions in India have been brought under water system, just a single third of the edited part is flooded. The efficiency of agriculture is low. So as the request of sustenance is expanding, the specialists, ranchers, agricultural researchers and government are endeavoring to put additional exertion and systems for more generation. Furthermore, subsequently, the agricultural data builds step by step. As the volume of data builds, it requires automatic route for these data to be extricated when required. Still today, a not very many agriculturists are really utilizing the new strategies, instruments and system of cultivating for better creation. Data mining can be utilized for anticipating the future patterns of agricultural procedures. In our proposed work, we have devised an algorithm using the checkpoints to find out the combinations which are hampering the production in this way rather than working on the all possible factors, working on the specific factors will help us to maintain the stabilize and better agricultural production.*

Keywords: *Data Mining, Agricultural production, Apriori Algorithm.*

I. INTRODUCTION

Data mining is the procedure that outcomes in the discovery of new examples in expansive data sets. The goal of the data mining process is to extricate learning from a current data set and change it into a human justifiable development for propel utilize. It is the way toward breaking down data from alternate points of view and compressing it into valuable information. There is no limitation to the kind of data that can be examined by data mining.

The data can be examined in a social database, a data distribution center, a web server log or a basic content document. Analysis of data in successful way requires comprehension of fitting strategies of data mining. The aim of this paper is to give insights about various data mining strategies in context of agriculture space so analysts can get insights about proper data mining methods in setting to their work region.

Data mining in agriculture is an extremely late research point. It comprises in the use of data mining methods to agriculture. This data mining procedures utilized as a part of agriculture for prediction of issue, infection recognition, streamlining the pesticide et cetera. Recent advancements are these days ready to give a ton of information on agricultural-related exercises, which would then be able to be examined keeping in mind the end goal to discover critical information and to gather pertinent information. This data mining

procedures are utilized for malady location, design acknowledgment by utilizing different application. Data mining is going to recognize the likenesses between looking through the profitable business information from the huge database frameworks, for example, finding connected items in gigabytes of store scanner data or the mining a mountain for a vein of important dataset. Both sort of procedures required either moving through a tremendous measure of material, or to play out the inquiry wisely with the goal that precisely match will be performed. Data mining should be possible on a database whose size and quality are adequate. The innovation of data mining can create new business openings by giving these capacities:

- Automated prediction and analysis of different patterns and practices - Data mining itself computerize the procedure by getting the prescient information from expansive databases. It first setup the inquiries and after that gives the relative arrangements. An ordinary case of such a prescient framework is in the showcasing field. Data mining utilizes data on the chronicled special mailings to catch the objectives adequately with the goal that the most extreme come back from market will be accomplished. Other prescient issues incorporate the recognition of liquidation or potentially the cheats.

- Another utilization of data mining is robotized discovery of recorded examples progressively. The exhibited Data mining framework can clear finished the databases to recognize the concealed examples. One of such case of example discovery is the analysis of retail deals data to recognize the apparently random items with the goal that the powerful buy should be possible. Other example discovery analysis incorporates the recognition of false charge card and in addition the exchanges to distinguish the abnormal data.

Data mining methods can get the advantages of computerization on existing programming and equipment stages that can be actualized on new frameworks which can be overhauled and new items can be produced. At the point when data mining apparatuses are characterized on superior parallel frameworks, they can be examined with enormous databases in minutes. Speedier handling is required in such framework to get the powerful outcomes from complex frameworks. Rapid preparing and precise result from the framework makes it workable for clients to examine extensive arrangement of information. Bigger databases, thusly, gives more enhanced predictions.

Agricultural associations store tremendous measures of data as harvest databases. Patterns in these databases can be distinguished utilizing data mining rehearses, which sort and model the data with a specific end goal to land at a

conclusion. The data mining applications exhibit the data as data bazaars.

In the agricultural business, in any case, the absence of standard vocabulary has thwarted the procedure of data mining to a specific degree. This could prompt pointless issues, amid the procedure of data mining. The expansion in the utilization of institutionalized terms will decrease the level of blunders in the data mining process.

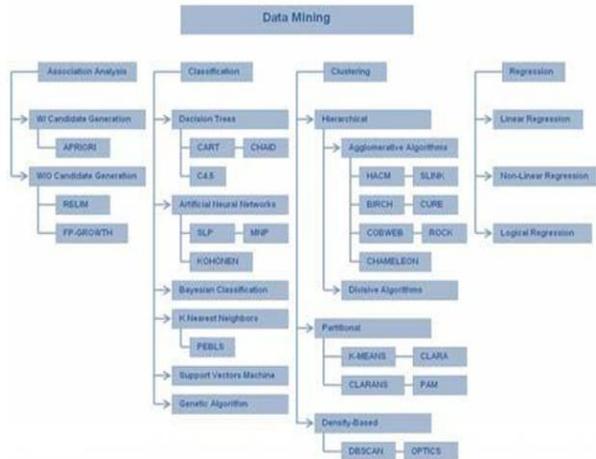


Figure 1. Data Mining Techniques

Data mining errands can be grouped into two classes: Descriptive data mining and Predictive data mining. Illustrative data mining assignments portray the general properties of the data in the database while prescient data mining is utilized to foresee express esteems in view of examples decided from known outcomes. Prediction includes utilizing a few factors or fields in the database to foresee obscure or future estimations of different factors of intrigue. To the extent data mining system is worry; in the a large portion of cases prescient data mining approach is utilized. Prescient data mining system is utilized to foresee future harvest, climate estimating, pesticides and composts to be utilized, income to be produced et cetera.

The primary methods for data mining incorporate Association rules, Classification, Clustering and Regression. The distinctive data mining strategies utilized for taking care of various agricultural issue has been talked about [3]. The graphical portrayal of various data mining strategies is appeared in figure 1.

II. RELATED STUDY

Shen Bin, Liu Yuan, Wang Xiaoyi propose four data tunneling models for the Internet of Things, which are multi-layer data mining model, circumnavigated data mining model, Grid based data mining model and data mining model from multi-improvement joining point of view. Among them, multi-layer model unites four layers: 1) data gathering layer, 2) data association layer, 3) occasion prepare layer, and 4) data mining association layer. Passed on data mining model can manage issues from securing data at various goals. Cross area based data mining model licenses Grid structure to understand the segments of data mining. Data mining model

from multi-advancement trade off viewpoint portrays the relating structure for the future Internet. [1]

Xindong Wu 2, Gong-Qing Wu 2, and Wei Ding 2 shows a HACE theory that depicts the parts of the Big Data change, and proposes a Big Data handling model, from the information mining perspective. This information driven model incorporates ask for driven aggregate of data sources, mining and examination, customer energy illustrating, and security and assurance considerations. They examine the testing issues in the information driven model moreover in the Big Data change. [2]

Feng Bao 3, Xu He 3, Fengzhi Zhao 3, addresses the segments of the petro physical information, logging information, seismic information and geographical information in light of the thoughts of the information mining. The mining musings as for the petro physical and logging information, seismic information and land information are made in perspective of their components. They uses particular mining ways to deal with handle the contrasting information, and delineates the result from the perspective of the components of information mining.

By data mining frameworks, the petro physical data are related with discover the relations and figure vault the logging data will be utilized to assess the delicate stores and see the fit supplies in dumbfounded geographical conditions; the space mining aftereffect of the 3D seismic data; the outlines and substance mining deferred consequences of the land data. [3]

Ms Shweta 4, Dr.KanwalGarg 4 thinks about data (bank data) and tries to secure the outcome utilizing Weka a data mining instrument. Collusion rule estimations are utilized to locate the best mix of various qualities in any data. In this paper maker utilizes Apriori to discover association run the show. Here maker consider three association standard figurings: Apriori Association Rule, Predictive Apriori Association Rule and Tertius Association Rule. Ms Shweta, Dr.KanwalGarg breaks down the postponed consequence of these three estimations and presents the outcome. By result got utilizing data mining instrument maker find that Apriori Association calculation performs superior to the Predictive Apriori Association Rule and Tertius Association Rule figurings. [4]

"Examination of Data Mining Tools in Knowledge Discovery Process" By Y. Ramamohan, K. Vasantharao, C. KalyanaChakravarti, A.S.K.Ratnam

Information mining, the extraction of covered prognostic data from huge databases, may well be a strong new development with charming potential to empower associations focus on most key data in their data dissemination focuses. It uses machine learning, associated science and picture techniques to disclosure and gift data in a structure that essential cognizant to individuals. varied all around delighted in information mining instruments are get-

fit these days. information mining mechanical assemblies expect future examples and works on, allowing associations to make proactive, learning driven decisions. information mining instruments can answer business addresses that for the most part were too much time overwhelming, making it difficult to decide.

In its slightest complex kind, information mining robotizes the discovery of noteworthy cases in a greatly data, utilize plot systems and computations to examine show and chronicled information which would then be able to be analyzed to envision future examples. as a consequences of information perusing in order to mine gadgets suspect future examples and practices through databases for covered illustrations, they allow relationship to make proactive, learning driven picks and answer addresses that were forerunner too long to resolve.[5]

Krutika. K .Jain and Anjali . B. Raut Propose Because of the fast development in overall information, productivity of affiliation rules mining (ARM) has been concerned for quite a while. Affiliation run mining has crucial influence in knowledge mining. The troublesome task is finding knowledge or valuable guidelines from the substantial number of tenets produced for diminished support In this paper, in view of the Apriori calculation affiliation rules depends on intriguing quality measures, for example, support, certainty and so on. Certainty esteem is a measure of lead's quality, while support esteem relates to factual centrality. Customary affiliation govern mining strategies utilize predefined support and certainty esteems. In any case, determining least support estimation of the mined standards ahead of time frequently prompts either an excessive number of or excessively few guidelines, which adversely impacts the execution of the general System. In this calculation, we will make affiliation rules relying on the dataset accessible in the database. The calculation significantly works on finding the insignificant certainty and so affiliation rules which as often as possible utilized and take after the base certainty. So the examination part of this paper is this by changing the estimation of least certainty, gives distinctive affiliation rules. The estimation of least certainty is high at that point rules separated more accurately.[6]

Neelamadhab Padhy1, Dr.Pragnyaban Mishra 2, and Rasmita Panigrahi3 In this paper we have centered an assortment of strategies, approaches and distinctive zones of the examination which are useful and marked as the essential field of data mining Technologies. As we know that many MNC's and vast associations are worked in better places of the diverse nations. Each place of operation may create extensive volumes of data. Corporate decision makers require access from every such source and take vital decisions .The data stockroom is utilized as a part of the critical business esteem by enhancing the adequacy of administrative decision-making. In an indeterminate and exceedingly focused business condition, the estimation of key information frameworks, for example, these are effortlessly perceived however in the present business condition, productivity or

speed isn't the main key for intensity. This sort of immense measure of data's are accessible as tera-to peta-bytes which has definitely changed in the territories of science and designing. To dissect, oversee and make a decision of such sort of immense measure of data we require strategies called the data mining which will changing in many fields. This paper confers more number of utilizations of the data mining and likewise o centers extent of the data mining which will accommodating in the further research.[7]

Akshita Bhandari¹, Ashutosh Gupta², Debasis Das³ There are a few mining calculations which have been created throughout the years. Apriori Algorithm is a standout amongst the most vital calculation which is utilized to separate successive itemsets from expansive database and get the affiliation manage for finding the knowledge. It essentially requires two critical things: least support and least certainty. Initially, we check whether the things are more noteworthy than or equivalent to the base support and we locate the incessant itemsets individually. Furthermore, the base certainty requirement is utilized to frame affiliation rules. In light of this calculation, this paper shows the constraint of the first Apriori calculation of sitting around idly and space for filtering the entire database looking on the incessant itemsets, and presents a change on Apriori by diminishing that dawdled relying upon checking just a few exchanges by executing a numerical recipe which at first segments the set of exchanges into bunches and select one specific group out of this. Our Algorithm can be utilized as a part of the library for finding the book that is most much of the time read and it can likewise be utilized as a part of the basic supply shop database by the shopkeeper for finding the itemsets which are every now and again sold as this takes lesser time and it's anything but difficult to discover the things so shopkeeper can make benefit by getting the information of those things which are regularly sold. It gives this outcome just by utilizing parallel calculation. The code is executed in java and the stage utilized is obscure. This current calculation's outcome is created on Mac utilizing parallel calculation else it is like the outcomes produced so far by numerous others. That is the way the outcomes are appeared and the data structure utilized as a part of this approach is the incessant example tree which can likewise be utilized to create restrictive examples and reasonable trees can be drawn for all the items.[8]

III. PROPOSED CONCEPT

The goal of the proposed strategy is to lessen CPU time which is spared by decreasing candidate set size. On the off chance that candidate set size is not as much as time required to compute the support of every candidate is less. We have proposed Method that lessens the quantity of candidate produced and time required to compute the support of every candidate so as to diminish the CPU Time we will present the checkpoints for the support esteems and likewise we additionally alter the support count plot so as to additionally channel the mixes or the components influencing the yield.

Here we will likewise endeavour to expand the quantity of

check points with a specific end goal to get the earlier outcomes. Here the fundamental motivation behind utilizing the Modified Apriori Algorithm is to discover the components blend in which we will get the best yield and the variables which will influence the yield or agriculture generation most

Base Paper Alogirithm (FApriori):-

The goal of the proposed strategy is to decrease CPU time which is spared by diminishing candidate set size. On the off chance that candidate set size is not as much as time required to ascertain the support of every candidate is less. We have proposed Method that decreases the quantity of candidate produced and time required to compute the support of every candidate. With a specific end goal to lessen CPU times, we have characterized two sort of checkpoint in dataset in light of support esteem:

checkpoint1=Totaltransaction-support count+1; $\forall \text{min_sup}$;
 checkpoint2=support count+1; if $\text{min_sup} > 50$.

Support property: If support check is n than any item to be visit it must show up in at any rate n exchanges in the dataset.

Every single new candidate after checkpoint1 can't be visit in light of support property. At checkpoint1 if $\text{min_sup} \leq 50$ or checkpoint2 if $\text{min_sup} > 50$, filter the candidate set once and check the support estimation of all candidate. The evaluated support esteem is utilized to evacuate rare itemsets at checkpoint.

Estimated Support value(Esupport)= support of candidate + Totaltransaction - checkpoint1; if $\text{min_sup} \leq 50$.

Estimated Support value(Esupport)= support of candidate + Totaltransaction - checkpoint2; if $\text{min_sup} > 50$.

If Estimated Support value < support count then that candidate will be removed from the candidate set.

Consider a dataset as show in TABLE 3.1. There are 10

transactions, namely $|D|=10$.

Modified Apriori Algorithm (Proposed Concept)

In this we have recommended the new adjusted calculation for data mining which depends on th FApriori calculation keeping in mind the end goal to refine the outcome we have contrived the calculation which works in the accompanying way.

Stage 1 : First the Minimum Support count is taken.

Stage 2: Then we will CheckPoint1 , CheckPoint2 , CheckPoint3,CheckPoint4 utilizing the accompanying recipes,

CheckPoint1 =Number of Transactions-Support Count+1

CheckPoint2 =Number of Transactions/2

CheckPoint3 =Number of Transactions/2 +1

CheckPoint4 =Support Count+1

Stage 3: Calculate the Support of DataSet utilizing the Apriori calculation and then the Esupport in first half is ascertained based on Checkpoint1 and checkpoint2 and Esupport on the second half is computed based on Checkpoint3 and Checkpoint4.

Step4: Find the Combinations based on the Esupport and frame the outcome.

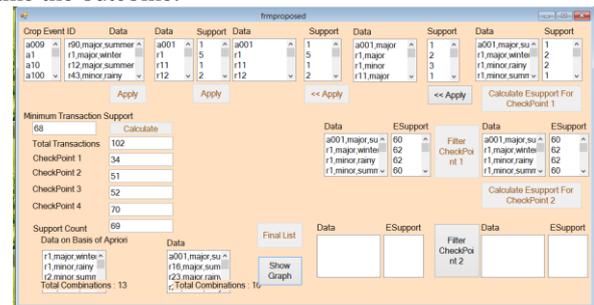


Fig 2. Proposed Implementation in VS 2010

IV. ANALYSIS OF PROPOSED WORK

Sample Data:-

We have taken a sample data for the analysis purpose in the excel sheet. And the excel sheet information is shown below,

TABLE 1 SAMPLE DATA FOR ANALYSIS

cropeventid	areaid	rainintensity	date_estimation	time_estimation	weather
a1	r1	major	22\1\1990	2:00am	winter
a2	r2	minor	24\3\1989	3:00pm	summer
a3	r1	major	1\2\2001	4:00pm	winter
a4	r6	major	24\2\2004	8:00am	winter
a5	r3	major	11\2\1994	6:00pm	summer
a6	r2	minor	23\9\1998	7:00pm	summer
a7	r35	minor	28\5\1996	8:00pm	rainy
a8	r47	minor	10\08\1997	5:00pm	winter
a9	r24	minor	25\08\1990	11:00am	summer
a10	r12	major	19\06\1991	12:00pm	summer
a11	r2	major	11\5\1989	7:00pm	rainy

a12	r18	major	14\12\2005	8:00am	winter
a13	r19	minor	21\09\1999	9:00am	summer
a14	r1	minor	24\05\1998	10:00am	rainy
a15	r46	minor	23\09\2001	7:00pm	rainy
a16	r78	major	12\01\1991	7:00am	summer
a17	r35	minor	24\09\1998	2:00am	winter
a18	r5	major	26\03\2001	3:00pm	summer
a19	r28	minor	30\01\1992	4:00pm	winter
a20	r29	minor	20\07\1997	8:00am	rainy
a21	r1	minor	29\09\1998	6:00pm	summer
a22	r20	minor	01\02\2004	7:00pm	winter
a23	r22	major	05\06\1994	8:00pm	rainy
a24	r23	major	19\06\2008	5:00pm	winter
a25	r24	major	25\08\1989	11:00am	winter
a26	r25	minor	16\10\2007	12:00pm	summer
a27	r5	minor	30\10\2003	7:00pm	summer
a28	r36	minor	17\08\1998	8:00am	rainy
a29	r38	major	18\10\2007	9:00am	winter
a30	r2	major	24\04\2004	10:00am	summer
a31	r48	major	29\08\1992	2:00am	rainy
a32	r4	minor	11\03\2004	3:00pm	winter
a33	r50	minor	20\08\2004	4:00pm	rainy
a34	r52	minor	10\08\1997	8:00am	summer
a35	r45	minor	05\06\1994	6:00pm	summer
a36	r35	major	18\10\2007	7:00pm	winter
a37	r23	major	20\08\2004	8:00pm	rainy
a38	r24	major	19\06\2008	5:00pm	summer

Data Analysis using the Apriori Algorithm

The sample data used for the analysis is analyzed using the Apriori algorithm, and the result obtained after the analysis is shown in the table 5.1 and the graph showing the comparison of the total combinations under the analysis and the total factor recommended for the betterment of the production.

TABLE 2 APRIORI ALGORITHM DATA ANALYSIS

	Total Factors	Factor Recommended for Betterment of Production
Apriori Algorithm	103	25

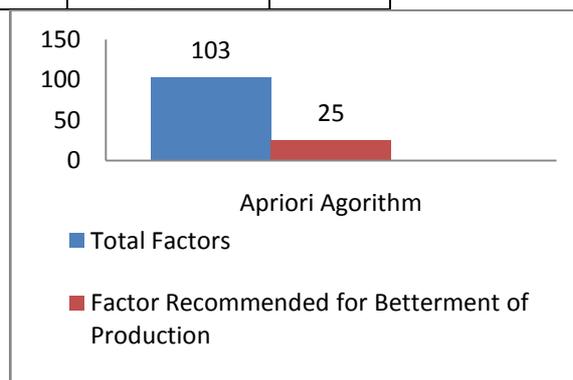


Fig 3 Comparison chart using Apriori Algorithm

Data Analysis using the FApriori Algorithm

The sample data used for the analysis is analyzed using the FApriori algorithm, and the result obtained after the analysis is shown in the table 5.2 and the graph showing the comparison of the total combinations under the analysis and the total factor recommended for the betterment of the production.

TABLE 3 FAPRIORI ALGORITHM DATA ANALYSIS

	Total Factors	Factor Recommended for Betterment of Production
FApriori Algorithm	103	19

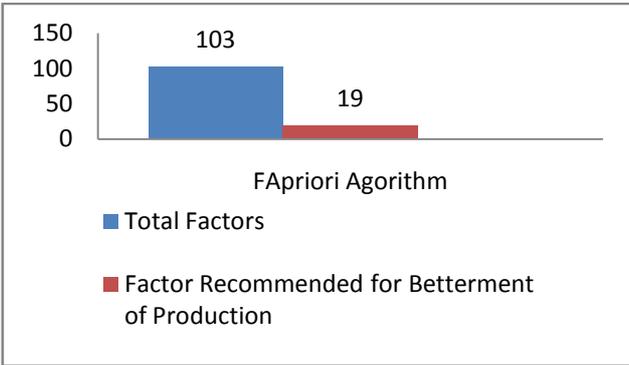


Fig 4 Comparison chart using FApriori Algorithm

5.4 Data Analysis using the Proposed Algorithm

The sample data used for the analysis is analyzed using the Proposed algorithm, and the result obtained after the analysis is shown in the table 5.4 and the graph showing the comparison of the total combinations under the analysis and the total factor recommended for the betterment of the production.

TABLE 5.4 PPROPOSED ALGORITHM DATA ANALYSIS

	Total Factors	Factor Recommended for Betterment of Production
Proposed Algorithm	103	10

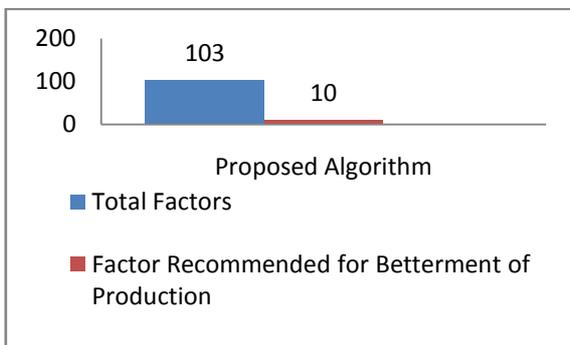


Fig 4. Comparison chart using Proposed Algorithm

Data Analysis using the Proposed Algorithm

The sample data used for the analysis is analyzed using the Apriori ,FApriori and Proposed algorithm , and the result obtained after the analysis is shown in the table 5.5 and the graph showing the comparison of the total combinations under the analysis and the total factor recommended for the betterment of the production.

TABLE 4 ALL ALGORITHM DATA ANALYSIS

	Total Factors	Factor Recommended for Betterment of Production
Apriori Algorithm	103	25
FApriori Algorithm	103	19
Proposed Algorithm	103	10

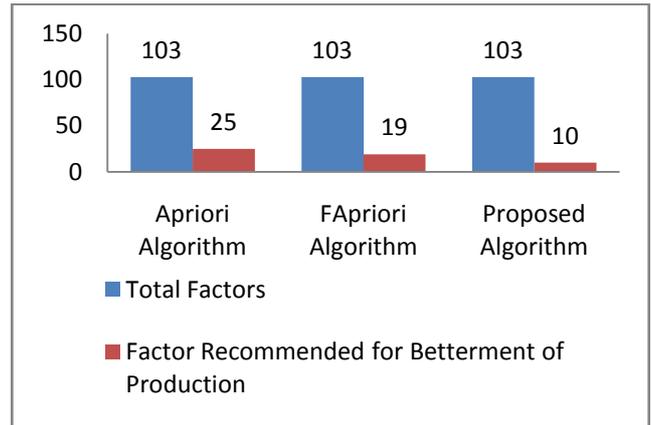


Fig 5 Comparison chart using Proposed Algorithm.

V. CONCLUSION

India is an agriculture based nation and there are number of factors which impacts the production and if those factors are distinguished properly we can enhance the yield and production level. The proposed calculation will enable us to refine the factors and discover the correct combination of the factors in which we to need to work out altogether enhance the level of crop production. The future degree is locating the quantity of more factors and refinement parameters with a specific end goal to improve comes about.

REFERENCES

- [1] Shen Bin, Liu Yuan, Wang Xiaoyi , "Research on Data Mining Models for the Internet of Things",Institute of Electrical and Electronics Engineers,2010
- [2] XindongWu,Gong-Qing Wu, and Wei Ding, , "Data Mining with Big Data",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014
- [3] Feng Bao, Xu He, Fengzhi Zhao, "Applying Data Mining to the Geosciences Data",International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE) ,2010
- [4] Ms Shweta, Dr. Kanwal Garg "Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013.

- [5] Krutika. K .Jain and Anjali . B. Raut, "Review paper on finding Association rule using Apriori Algorithm in Data mining for finding frequent pattern", *International Journal of Engineering Research and General Science* Volume 3, Issue 1, January-February, 2015 ISSN 2091-2730.
- [6] J. Han and M. Kamber, "Data Mining Concepts and Techniques, "second Edition, Morgan Kaufmann Publishers, 2006.
- [7] R. Agrawal and R. Srikant, "Fast algorithms for mining associationrules," In Proc. 20th Int. Conf. Very Large Data Bases, pp. 487-499,1994.
- [8] J.S. Park, M.S. Chen and P.S. Yu, "An Effective hash-based algorithmfor mining association rules," in Proceedings of ACM SIGMOD, pp.175–186, 1995.
- [9] S.A. Ozel, and H.A. Guvenir, "An algorithm for mining associationrules using perfect hashing and database pruning," in: Springer 10thTurkish Symposium on Artificial Intelligence and Neural Networks,Gazimagusa, pp. 257-264, 2001.
- [10] M. J. Zaki, "Scalable algorithms for association mining," *IEEE Trans.Knowledge and Data Engineering*, pp.372–390, 2000.
- [11] J. Han, J. Pei and Y. Yin, "Mining frequent patterns without candidategeneration", Proceedings of SIGMOD, 2000.
- [12] F. Wang and Y. Li, "An Improved Apriori Algorithm Based on theMatrix," 2008 International Seminar on Future BioMedical InformationEngineering, pp. 152–155, Dec. 2008.
- [13] R. Chang and Z. Liu, "An improved apriori algorithm," Proceedings of2011 International Conference on Electronics and Optoelectronics, Jul.2011.
- [14] Smita1, Priti Sharma2, "Use of Data Mining in Various Field: A Survey Paper", *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. V (May-Jun. 2014), PP 18-21.
- [15] Richard A. Huebner and Norwich University, "A survey of educational data-mining research", *Research in Higher E ducation Journal*.