MODIFIED APRIORI FACTOR BASED FILTRATION ALGORITHM

Jyoti Sharma[1], Rajesh Kumar Chaursiya[2], Shalini[3], Dr. Swapnil Singhal[4]
[1]M.Tech Scholar, [2,3]Assistant Professor, [1,2,3]CSE dept.JIT, Jaipur, Rajasthan.

Abstract: Agriculture is the utmost important factor which forms the GDP of India. But still in an age which IT is touching the sky, the production still lies on the proper functioning of natural factors. The dissertation work, focuses on this aspect of the agriculture, as India is a vast country so applying improvement on all the factors which are responsible for the proper production as well as in all areas is practically not possible , the proposed work so to help in determining the exact factor combination which will truly increase the production in the concern area.

I. INTRODUCTION

Data mining implies gathering important information from unstructured data. So it can help accomplish particular destinations. The motivation behind a data mining exertion is regularly either to make a graphic model or a prescient model .An elucidating model presents, in succinct frame, the principle qualities of the data set. The reason for a prescient model is to enable the data digger to foresee an obscure (frequently future) estimation of a particular variable; the objective variable. The goal of prescient and illustrative model can be accomplished utilizing an assortment of data mining methods as appeared in fig 1.[1]

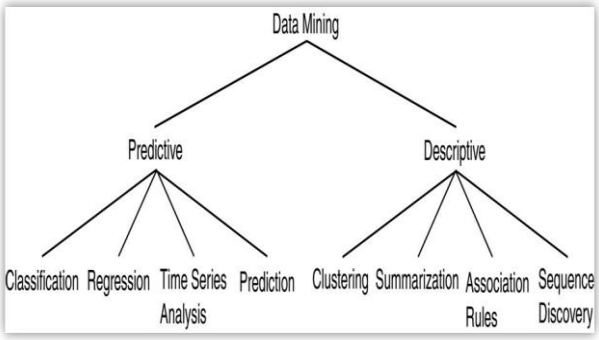

Figure 1 Data Mining Models

1. Classification: Classification in view of absolute (i.e. discrete, unordered). This procedure in light of the directed learning (i.e. wanted yield for a given info is known) .It can group the data in view of the preparation set and qualities (class mark). These goals are accomplish utilizing a choice tree, neural system and classification lead. For illustration we can apply the classification on the past record. Utilizing these methods we can without much of a stretch distinguish the execution of the understudy. [2]

2. Regression: Regression is utilized to delineate data to a genuine esteemed prediction variable. In the regression procedures target esteem are known. For instance, you can foresee the tyke conduct in light of family history.

3. Time Series Analysis: Time series analysis is the way toward utilizing measurable strategies to model and clarify a time-subordinate series of data focuses. Time series estimating is a strategy for utilizing a model to create predictions (conjectures) for future occasions in view of known past occasions. [2]

4. Prediction: It is one of a data mining systems that find the connection between free factors and the connection amongst needy and autonomous factors.

5. Clustering: Clustering is a gathering of comparative data question. Unique question is another group. It is way discovering likenesses between data as indicated by their trademark. This system in view of the unsupervised learning (i.e. wanted yield for a given information isn't known). For instance, picture handling, design acknowledgment, city arranging.

6. Summarization: Summarization is deliberation of data. It is set of pertinent errand and gives a review of data. For instance, long separation race can be compressed aggregate minutes, seconds and tallness. Affiliation Rule: Association is the most well known data mining. Affiliation endeavors to find designs in data which depend on connections between things in a similar exchange. As a result of its temperament, affiliation is sometimes alluded to as "connection procedure". This technique for data mining is used inside the market based analysis keeping in mind the end goal to distinguish a set, or sets of items that buyers regularly buy in the meantime. [2]

7. Sequence Discovery: Uncovers connections among data. It is set of protest each related with its own timeline of occasions. For instance, logical examination, catastrophic event and analysis of DNA sequence.

II. CROP PATTERN IN INDIA

The need to strengthen agricultural generation in the wake of the moderate pace of development in agricultural yield acknowledged in the current past combined with the rising interest for agricultural products, declining per capita accessibility of arable land, falling apart position of accessibility of normal assets, for example, water, and they want to keep up a high development rate of total GDP for the Indian economy has put the agricultural area back at the middle phase of India' arranging process. Without any noteworthy leap forward in agricultural generation innovation having been accomplished over the most recent , accomplishing the coveted levels of agricultural creation in the short to medium run would require attempting more purposeful endeavors towards spanning the harvest profitability holes feasible with existing innovation. They showed ability of substance manures, an essential part of the accessible agricultural generation innovation, in expanding the yield efficiency and raising the homestead productivity gives some beam of expectation. While the manure utilization, both in supreme terms and in addition on per hectare premise, has expanded complex throughout the

years, however over the couple of years the development has not been agreeable. Apart from between territorial differences in utilization of manure, there is extreme awkward nature in use of various supplements. The present manure use design in this way offers more degree for expanding the utilization of composts as well as their more proficient use and adjusted utilization of manures holds for expanded agricultural creation, efficiency; cultivate benefit and a more supportable asset base. [3]

The present examination, proposed by the Department of Fertilizers, Ministry of Chemicals and Fertilizers, Government of India has been embraced by Agro-Economic Research Centers (AERCs) situated in Delhi, Chennai, Jorhat, Ludhiana and Vishva Bharati as the example of the Directorate of Economics and Statistics, Ministry of Agriculture, Government of India. The examination has been co-ordinated by AERC, Ludhiana who has likewise given the investigation outline and the approach for the examination. The bigger examination endeavors to investigate the patterns in manure utilization and distinguish factors influencing development/stagnation in compost utilization after some time and in various locales of the nation. The examination additionally endeavors to evaluate the effect of manure use on profitability of chosen crops and the financial proficiency of compost use for vital yields in various states Based on this analysis the investigation endeavors to recommend some medicinal measures to help manure use in the nation to accomplish the objectives set for agricultural creation. The present report identifies with the territory of Haryana. The particular goals of the present investigation are:

• Analyze the patterns in compost use after some time and crosswise over various ranch estimate classifications.
• Identify the determinants of compost utilization
• Assess the effect of compost use on profitability of select harvests and furthermore the financial proficiency of manure utilize.

All inclusive 80% , for every penny of the agricultural land region is rainfed which creates 65 to 70 for each penny staple nourishments however 70 for every penny of the populace possessing in these territories are poor because of low and variable efficiency. India positions first among the rainfed agricultural nations of the world as far as both degree and estimation of create. Rainfed horticulture is polished in 66% of the aggregate trimmed region of 162 million hectares(66 per penny). Rainfed horticulture underpins 40 for each penny of the national sustenance bushel. The significance of rainfed agribusiness is clear from the way that 55 for every penny of rice, 91 for each penny coarse grains, 90 for every penny beats, 85 for every penny oilseeds and 65 for every penny cotton are developed in rainfed regions. These zones get a yearly precipitation between 400 mm to 1000 mm, which is unevenly conveyed, very questionable and flighty. In specific territories, the aggregate yearly precipitation does not surpass 500mm. Because of low and unpredictable precipitation, a noteworthy fall in sustenance creation is frequently taken note.

Inside farming, it is the rainfed horticulture that will be most affected by climate change. Temperature is an essential climate parameter that will influence efficiency of rainfed crops. The most recent three decades saw a sharp ascent taking all things together.

III. RELATED WORK

Nikita Jain, Vishal Srivastava [4] In this paper, the idea of data mining was condensed and its criticalness towards its techniques was delineated. The data mining in light of Neural Network and Genetic Algorithm is looked into in detail and the key innovation and approaches to accomplish the data mining on Neural Network and Genetic Algorithm are additionally studied. This paper additionally directs a formal audit of the zone of lead extraction from ANN and GA. [4]

Y.Elovici1, A.Kandel2, M.Last1, B.Shapira1, O. Zaafrany1 An inventive knowledge-based strategy for fear monger discovery by utilizing Web movement content as the review information is displayed. The proposed system takes in the ordinary conduct ('profile') of psychological oppressors by applying a data mining calculation to the printed substance of fear related Web locales. The subsequent profile is utilized by the framework to perform ongoing location of clients associated with being occupied with fear monger exercises. The Receiver-Operator Characteristic (ROC) analysis demonstrates that this strategy can beat a command-based interruption identification framework. [5]

Kalyani M RavalData mining is a procedure which finds helpful examples from huge measure of data The way toward removing already unknown, understandable and significant information from huge databases and utilizing it to make critical business decisions - Simoudis 1996 This data mining definition is for business enhance and for business situations. In any case, data mining is a procedure that can be connected to a data extending from climate estimating, electric load prediction, item plan, and so on. Data mining likewise can be characterized as the PC help process that burrows and breaks down tremendous sets of data and then removing the knowledge or information out of it. By its easiest definition, data mining mechanizes the discoveries of pertinent examples in database. [6].

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. [7]

IV. PROPOSED WORK

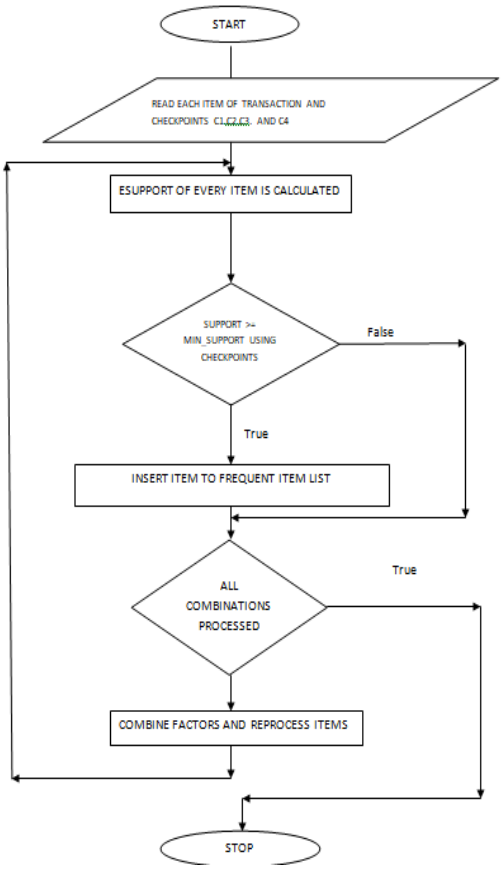

Fig 2. Proposed Algorithm

In the proposed work we have taken reviews of movies in a text file. And the text file is scanned line by line and then the whole document is processed. We have created a GUI(Graphical User Interface) in which the review file is selected and computed on the basis of selected feature from combo box.After that in output the line's score is shown after all calculation. In this process following steps are performed as:

4.1. Data Collection: -The first step is to collect movie reviews from different sites. We have collected the data from the various agriculture related websites.

4.2. Factor Fetching: All the factor data is stored in the excel file for reading all the data and filtering out only the data related to the factors.

4.3. Support Calculation: Now the occurrences of each factor are analyzed and count the number of each factor occurrence individually as well as in combination is count to calculate the support.

4.4. Minimum Transaction Support: The transactions to be evaluated is entered by the user.It is the percentage to the transactions to be taken for consideration.

Minimum Transaction Support=Percentage/100*Total Number of Transactions

4.5. CheckPoints Calculations: Checkpoints are involved for the further filtering out the data and the S represents here the minimum transaction support and the formulas for the checkpoints calculations are given bellow..

CheckPoint1 =Number of Transactions-Support Count+1
CheckPoint2 =Number of Transactions/2
CheckPoint3 =Number of Transactions/2 +1
CheckPoint4 =Support Count+1

4.6. ESupport Calculation: Now the Esupports are calculated in the two halves,

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.

4.7. Factors Priority Assignment : Factors are the prime source of deciding which combination is betters for the production enhancement.

F1 F2 F3 say suppose are the factors and P1 , P2 AND P3 are there respective priorities , to the total priority is calculated as,

F1*P1+F2*P2+F3*P3

Priority values p1,p2,p3 ranges on the categories of the factors .

4.7 Performance Evaluation Criteria

The performance evaluation of the three algorithm which are compared is based on the total number of factor combinations which they results. In this the calculation of the support and ESUPPORT involved and the Priority values in the Factor based checkpoint algorithm which is proposed in the dissertation is again checked with the threshold values to further filter out the factor combination

V. ANALYSIS OF PROPOSED WORK

The implementation proposed is simulated over the various datasets and the results obtained is shown in table 1.

	Nos of Factors Combination	Nos of Factors Combinations Filtered
Case I	103	4
Case II	270	8
Case III	480	16
Case IV	617	21
Case V	780	27
Case VI	900	34
Case VII	1123	47

Table 1 . Factor Filtration Results

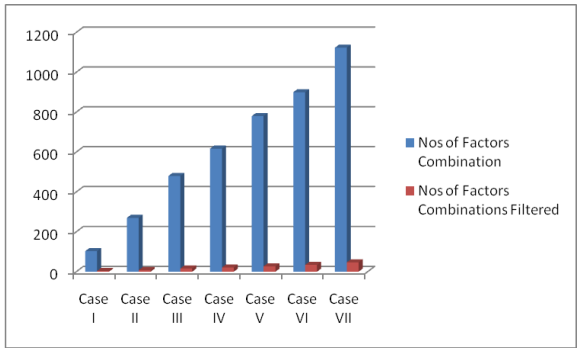

Fig 3. Factors Filteration Graph

VI. 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 further improvement in the concept is that the association concept on factor is applied , in which on the basis of the priority of the factors the value is assigned to the factors and the product of these values checked with the threshold value and the factor combination then analysed for the feasibility and is considered for further analysis.

The future degree is locating the quantity of more factors and refinement parameters with a specific end goal to improve come about.

REFERENCES

[1] Nikita Jain1, Vishal Srivastava2, "DDATA MINING TECHNIQUES: A SURVEY PAPER", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.

[2] Y.Elovici1, A.Kandel2, M.Last1, B.Shapira1, O. Zaafrany1, "Using Data Mining Techniques for Detecting Terror-Related Activities on the Web"

[3] Kalyani M Raval,"Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 10, October 2012 ISSN: 2277 128X

[4] Shen Bin, Liu Yuan, Wang Xiaoyi , "Research on Data Mining Models for the Internet of Things",Institute of Electrical and Electronics Engineers,2010

[5] 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

[6] Feng Bao, Xu He, Fengzhi Zhao, "Applying Data Mining to the Geosciences Data",International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE) ,2010

[7] 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.

