International Journal of Engineering &Technology, 7 (1.7) (2018) 121-124

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Study of high yielding crops cultivation in India using data mining techniques

Dr. M J Carmel Mary Belinda [1] *, Umamaheswari. R [2], Alex David S [3]

[1] Associate Professor, Department of Computer Science and Engineering, School of Computing,Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai-62, TamilNadu, India
[2] Assistant Professor Department of Computer Science and Engineering, School of Computing,Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai-62, TamilNadu, India
**Corresponding author E-mail: carmelbelinda@gmail.com*

Abstract

Data mining in agriculture is a modern and emerging research technique. Data mining provide many techniques like k means algorithm, support vector machine, association rule mining and Bayesian belief network [1]. This technique can be used in agriculture for various purposes. This paper describes about how association rules mining and apriori algorithm can be used in agriculture field. This paper also describes about soil, its types and crops grown in each type of soil. The technique that has been used here can be a rough set study, but like this many efficient techniques can be applied to solve many problems in agriculture.

Keywords: *DataMining; Apriori; SOIL; Cultivation.*

1. Introduction

Agriculture is the backbone of a country. It is the old age practice which is followed today also. In the early stage agriculture involved only food production slowly it now changed to modern agriculture using various agriculture techniques [2]. We can say agriculture as a profession that is done with risk, as it depends on climate, geographic conditions, and economic factors. It gave strong structural and economic base to the human society for both survival and progress.

- SOIL

Soil is formed due to complex process which involves parent bedrock, climate, and slope of the land. Soil varies widely formed by many things like wind, water, temperature, and climate. Soil is the strong foundation of productive farming practices as soil provides essential nutrients that is required by the plants for their growth. Soil has three phases,

- Solid
- Liquid
- Gases

It acts as a medium for the growth of the plants. Soil component must exist in proper ratio, so that it will lead favorable condition for the growth of the plant.

- SOIL QUALITY

Soil quality is how well soli does what we want it to do. To grow crops, soil wants to hold water and nutrients which make

- Plant to take their root
- Suppress pests and weeds.

High quality soils has,

- Good soil tilth
- Good drainage system
- Sufficient depth
- Contains no chemicals
- SOIL FERTILITY

Provide nutrients for the plants for their growth

- SOIL STRUCTURE

Organization and orientation of primary and secondary practice to form a pattern is called as soil structure. It is needed and important as it influence amount and nature of sponginess to regulate water, air in the soil. It increases drainage and reduce soil erosion.

- SOIL AND ITS CLASSIFICATIONS

Soil is the most valuable resource in India as it plays a vital role in the agriculture and most of the Indian agriculture depends upon the limit and quality of the soil.

- ALLUVIAL SOIL

Alluvial soil is formed due to deposition of sediments by rivers, sea waves deposited from the rivers, winds. As the soil contains the alluvium, and the soil is named as alluvial soil. The soil is rich in humus and potassium. It is found in the Indo Ganga Brahmaputra Plains, Coastal Plains and Northern plains. Alluvial soil is the abundant harvest and high fertility that's why it is the best soil used for agriculture in India. The crops that can be cultivated in this soil are Rice, Wheat, Sugarcane, Jute oilseeds, maize's and pulses.

- RED SOILS

Red soil found in very low rainfall condition on granite and geneses rocks. The color of the soil is red due to presence of iron. It is found in Tamil Nadu, Karnataka, North and South eastern Madhya Pradesh, Jharkhand. Wheat, rice, millets, gram, pulses, oilseeds, sugarcane, tobacco and cotton are the crops that grow here.

- BLACK SOIL

Black soils are clay in nature, which becomes very sticky when wet. Extensively found upon the Lava Plateaus of Maharashtra, Gujarat, Madhya Pradesh, west Andhra Pradesh. Crops grows in Black soil are Sugarcane, wheat, tobacco, maize, ragi and groundnut.

- MOUNTAIN SOIL

Copyright © 2018 Dr. M J Carmel Mary Belinda et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mountain soil is formed because of deposition of organic matter present in the forests. This soil is rich in humus. It is found in Eastern and Western Ghats, Himalayan regions. Crops that grows in Mountain soil are wheat, tea, coffee, spices maize, barley and tropical fruits.

- LATERITE SOIL

Laterite soil is a kind of clayey rock which is formed due to high temperature and high rainfall. The soil is found in South Maharashtra, the Western Ghats in Kerala and Karnataka, parts of Assam, Tamil Nadu, and Karnataka. Crops grows on laterite soil are tea, coffee, coconut, cashew nut, rubber etc.

- DESERT SOIL

The desert soil contains 90% to 95% of sand and 5%to 10% of clay. Rajasthan, Haryana and Punjab are the areas that contain deserts oil in high amount. Desert soil yield variety of crops, like wheat, millets, barley, maize, pulses, cotton.

2. Data mining

Data mining is an emerging technique, which discover interesting and useful pattern from a large volume of database. It can't be done manually. Agriculture includes wide variety of database. It is really a challenge to obtain the desired pattern form those databases. Nowadays modern trends and technology has been used in the field of agriculture [4]. One of the efficient and effective technologies that can be applied to the agriculture field is none other than data mining, which is more effective in discovering the hidden pattern. It is useful for,

Prediction:
Reaction of attribute present in the database in the future.
Classification:
Partition (division) of data into groups.
Optimization:
Way of using the limited resources that are available.

- Sources of data

The data that is required for analysis of any data can be obtained from,

- Spatial data
- Temporal data
- Transactional data
- Data warehouse
- ASSOCIATION RULES

Association rule is used to find relationship between the two items. This relationship describes the association level, either strong link or weak link between those two related items. The associatively between the two items can be determined with the help of two variables, that is

- Support
- Confidence

Support describes about probability of two items with that of the no of items present in the list

SUPPORT=P (AUB)/N

Confidence describes about probability of two items with that of the first item that has been taken for comparison.

CONFIDENCE=P (AUB)/P (A)

Table 1: Transaction (Cultivation) of Each Crop

Soil	Rice	Wheat	Sugarcane	Jute	Pulses	Maize	Millets	Cotton	Tobacco	Tea	Coffee
Alluvial	1	1	1	1	1	1					
Red	1	1	1		1		1	1	1		
Black		1	1			1			1		
Mountain		1								1	1
Laterite											1
Desert		1				1		1			

Based on the transactions given in the table1, the support and confidence level for the crops wheat and rice has been calculated below, by considering the minimum support count value as 30%.

SUPPORT=P (AUB)/N

=P (WHEAT U RICE)/6

=2/6

=33%

CONFIDENCE=P (AUB)/P (A)

=P (WHEAT U RICE)/P (WHEAT)

=2/5

=40%

Since the calculated value of the support (33%) and confidence (40%) is greater than the given support count value (30%), it is clearly observed that the crops wheat and rice have strong relation between each other.

3. Apriori algorithm

Apriori algorithm was given by Agrawal and Srikant in the year 1994.Apriori algorithm is one of the algorithms used in the data mining [3]. This algorithm is mainly used to find the frequent items that are present in the given data base. Generally, this algorithm is used in market basket analysis, to find the set of items that are purchased together in a single transaction.

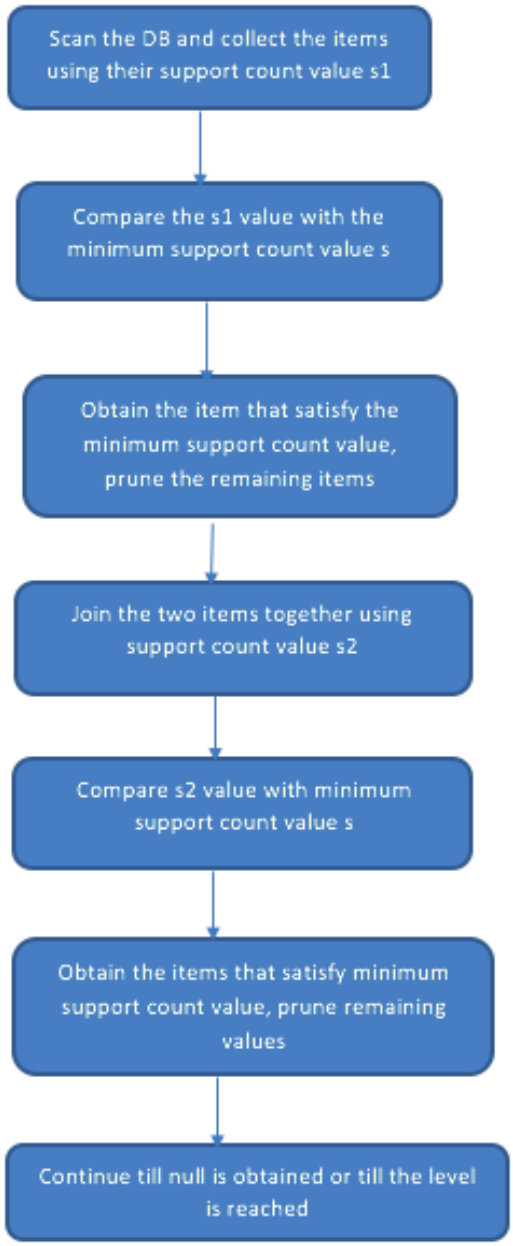


Fig. 1: Apriori Algorithm Steps.

Here the apriori algorithm is used to find the crops that are cultivated in high level in India. Therefore, this algorithm is used to determine crops that are cultivated in most of the areas in India. [5] Actually this algorithm uses the bottom up approach technique, where the frequent subsets of items are taken one item per time. This step is called as candidate key generation. Then the groups of items are tested against the data that are present in the given database. Finally, the algorithm gets ended when no successful extensions of data sets are found in the data base. The two main operation that are done here are,

- Join
- Prune

Join:

Join two items when they are purchased together. Here the crops are chosen when they are cultivated in more than one soil.

Prune:

The items will be removed from the table when their support count does not match with minimum support count. Here the minimum support count is 2, i.e., min_sup count=2.

Based on the table1, the table 2 has been derived, which include the crops that are cultivated in the different types of the soil. The table is derived based on the transactions which have the value as one.

Table 2: Types of Crops Cultivated in Various Soil

Soil	Crops	Crops	Crops	Crops	Crops	Crops	Crops
Alluvial	Rice	Wheat	Sugarcane	Jute	Pulses	Maize	
Red	Wheat	Rice	Millets	Pulses	Cotton	Sugarcane	Tobacco
Black	Sugarcane	wheat	Groundnut	Tobacco	Maize	Ragi	
Mountain	Tea	Coffee	Spices	Wheat	Maize		
Laterite	Tea	Coffee	Coconut	Cashew nut	Rubber		
Desert	Wheat	Millets	Barley	Maize	Pulses	Cotton	

From the table 2, the table 3 has been derived, which include some of the crops that are cultivated in various soils.

Table 3: Sample Crops Obtained Based on Table 2

Soil	Crops	Crops	Crops	Crops
Alluvial	Rice	Wheat	Sugarcane	Jute
Red	Wheat	Rice	Sugarcane	Pulses
Black	Sugarcane	Wheat	Maize	
Mountain	Tea	Coffee	Wheat	
Laterite	Tea	Coffee	Rubber	
Desert	Wheat	Millets		

Table 4 is obtained with the help of table3. This table includes number of times the crops has been cultivated in various soils. Example, rice cultivated in 2 different soils and wheat cultivated in 5 different soils

Table 4: Single Crops Growth in Soil

Crops	Support count value
Rice	2
Wheat	5
Sugarcane	3
Jute	1
Pulses	1
Maize	1
Tea	2
Coffee	2
Rubber	1
Millets	1

Table 5 derived from table 4, this table includes the crops which satisfy the minimum support count value. Other crops have been pruned as they not satisfy the minimum support count value.

Table 5: Crops Satisfy Minimum Support Count

Crops	Minimum support count
Rice	2
Wheat	5
Sugarcane	3
Tea	2
Coffee	2

Table 6 obtained using the table 5; this table includes the support count value of the different crops cultivated in same soil. Example rice and wheat can be cultivated in alluvial and red soil.

Table 6: Dual Crops Cultivation in Same Soil

Crops	Support count value
Rice, Wheat	2
Rice, Sugarcane	2
Rice, Tea	0
Rice, Coffee	0
Wheat, Sugarcane	2
Wheat, Tea	1
Wheat, Coffee	1
Sugarcane, Tea	0
Sugarcane, Coffee	0
Tea, Coffee	2

Table 7 derived from table 6, this table includes the crops which satisfy the minimum support count value. Other crops have been pruned as they not satisfy the minimum support count value.

Table 7: Crops Satisfy Minimum Support Count

Crops	Minimum support count
Rice, Wheat	2
Rice, Sugarcane	2
Tea, Coffee	2

Table 8 derived from table 6, in which the support count value of multiple crops has been taken.

Table 8: Multiple Crops Cultivation in Same Soil

Crops	support count
Rice, Wheat, Sugarcane	2
Rice, Wheat, Tea	0

Table 9 derived from table 8, this table includes the crops which satisfy the minimum support count value. Other crops have been pruned as they not satisfy the minimum support count value.

Table 9: Crops Satisfy Minimum Support Count

Crops	support count
Rice, Wheat, Sugarcane	2

From the table 9 it can be judged that the crops rice and wheat are cultivated in large amount in India than other crops.

4. Conclusion

With great extent and development of the data mining techniques, it can be applied to various fields. As data mining can able to act on various kinds of data and effectively it can predict the valuable pattern that is hidden in that particular data base. In this paper association rule mining and Aprior algorithm has been discussed and using the Aprior algorithm, it is possible to predict the highly cultivated crops in India. Agriculture in India faces many problems nowadays. It is difficult to solve all the problems individually. Data mining techniques can be applied to various fields of agriculture like testing soil quality, crop productivity etc so that our nature can be preserved as agriculture is the backbone of our Indian economy.

References

[1] Manisha Sahane, BalajiAgalve, RazaullahKhan, SanjaySirsat,"An Overview Of Data Mining Techniques Applied To Agriculture Soil Data", International Journal Of Agriculture Innovation and Research, volume 3,issue2,September 2014

[2] Hetal Patel,DharmendraPatel,"A Brief Survey Of Data Mining Techniques Applied To Agriculture Data" International Journal Of Computer Applications, Volume 9,june 2014

[3] P. Jaganathan, S.Vinothini,P.Bacialakshmi,"A StudyOf Data Mining Techniques to Agriculture, International Journal Of Research in Information Technology, volume 2,april 2014.

[4] UmamaheswariR,Siva Purnima S,Dr. S. SaravanaMahesanCustomerPreservence for an Organisation Using Data Mining.International Journal of Civil Engineering and Technology (IJCIET). October 2017,Volume 8, Issue 10, pp. 933–938

[5] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).

[6] Gleaso CP. Large area yield estimation/forecasting using plant process models.paper presentation at the winter meeting American society of agricultural engineers palmer house, Chicago, Illinois. 1982 Majumdar J, Ankalaki S. Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves. In: Paper presented at international conference on computational science and engineering. 2016.

[7] Jain A, Murty MN, Flynn PJ. Data clustering: a review. ACM ComputSurv. 1999; 31(3):264–323. https://doi.org/10.1145/331499.331504.

[8] Jain AK, Dubes RC. Algorithms for clustering data. New Jersey: Prentice Hall; 1988.

[9] Berkhin P. A survey of clustering data mining technique. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Berlin: Springer; 2006. p. 25–72 https://doi.org/10.1007/3-540-28349-8_2.