Scientific Journal Impact Factor (SJIF) : 3.518
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

Association Rules Based Analysis for Arthritis Patients' Data

Deepti Deshpande[1] and Dr.B.S.Anami[2]
[1,2]*Department of CSE, KLEIT,HUBBALLI,*

Abstract—Data mining in health care is also known as predictive analysis and has been an area for research.Association rules is one of the means in data mining for mining frequent patterns from transactional data set to discover interesting association and correlations with data.association rules are one of the means to uncover interesting relationships among data that are stored in data repository.This paper involves implementation of association rules using Apriori Algorithm for depicting the co-occurances between particular Arthritis and their factors.This paper targets to contribute absolute information about Arthritis and its factors.
Keywords—Data mining,Association rules,Apriori Algorithm,Support,Confidence.

I. INTRODUCTION

In a given data set mining frequent patterns searches recurring relationships.In large transactional data sets mining frequent itemset leads to discovery of associations and correlations amount item sets.These co-relations finds application in the area of medical science,customer shopping, market analysis.Data mining applications also include Business intelligence which is the technology for providing predictive views of business operations,for acquiring a better understanding of the commericial context of organization.Web search engine is another important application of data mining that searches for information on the web [1].

Association rules are representation of patterns that occur frequently in a dataset.Support-confidence framework is being employed by association rules data mining algorithms.Frequent itemsets are the outcomes of Association rule mining.Suppose for example,we have two data items M and N which are satisfying mininmum support threshold, a strong association rules in the form of M N are generated.Further,statistical co-relations between M and N are conveyed to futher analyze correlation rules[1].In a data repository consisiting of set of data items association rules is a vital approach of data mining to find out frequent patterns.Association rules in health care may be used significantly for identifying relationships among diseases.The two key terms used in identifying correlation among data items are support and confidence.Association rules are of of the means to identifying fraud in health care insurance,helps to predict co-occurance of diseases.The process of generating association rules is divided into two steps:Firstly To find all the frequent itemset in database that has minimum support.Secondly,Constrains to form association rules such as frequent itemsets and mininmum confidence constraint should be statisfied[4]. Association rules associates some importance in discovering relationships between different data items in large data repository,in data classification,for discovering important and useful information for improving quality of business or medical field,to handle high dimensionality of data,helps in detection of fraud in case of insurance[6].Health care is one of the important field where data mining is applied.Data mining helps to analyze patients records present in health care repository and identify patterns of their disease using various data mining techniques such as association rules etc,.and take measures to prevent such diseases.Patterns of data which change as data change is the main concern.Apriori Algorithm is the tool which is used to overcome these chanlleges.The two key terms in Apriori Algorithm are support and confidence which are widely used for discovering interesting relationship when mining health care data in health care repository[7].Further,this paper discuss about the literature survey carried out,proposed methodology ,results and discussions.

II. RELATED WORK

In order to know the state-of-art,tools and techniques used for analysis of data,a literature survey is as follows:

R.Karthiyayini et al.,[2015],focuses on implementation of Apriori Algorithm to discover interesting patterns and association rules in chronic diseases.For each symptoms of chronic diseases percentage of probability is calculated for effective decision making.

Gitangali J et al.,[2014], proposes Aprior Algorithm based data mining technique for generating frequency of diseases that the patients are affected in various geographical region and at various time periods.The analyses concludes with the result that patients are affected with 4 different diseases in a particular geographical area during a particular year.

Shelly Ahuja et al.,[2014],presents efficient mining based algorithm for rule generation amd uses Apriori Algorithm to improve precision and recall and F-measure value.Association rules are optimised using Ant colony Optimization.

M Durairaj et al.,[2013],focused on comparing the data mining techniques for predictive diseases in healthcare sector.The study shows that 97.7% accuracy for cancer prediction.

Sunita Soni et al.,[2010], discusses about advanced associative classifiers to improve the prediction for the patients of different age group.

From,the above literature survery, we conclude that Apriori algorithm is an efficient algorithm for predicting occurance of frequent patterns in healthcare and also used in measuring interestingness of frequent patterns by means of support and confidence.

III. PROPOSED METHODOLOGY

Apriori Algorithm was discovered by R.Agarwal et al., in 1994. Discovering interesting patterns among given data set is the main objective of Apriori Algorithm.It uses an Iterative approach.Two important measures that are taken into consideration in Apriori Algorithm are support and confidence.To find frequent itemset and association rules from frequent itemset generally two steps are important and it is also known as two step process.An Association rule is said to be Boolean,if the rule is concerned about the association between the presence or absence of items[8].

Figure 1: Block diagram

The methodology is being divided into three steps namely**:**Data Acquisition,Applying Apriori Algorithm,Association rules as shown in Figure 1 :

- **Data acquisition**:Dataset related to Arthritis patients related to their risk factors were being collected from a local clininc.Study was carried out on risk factors that were the main cause of occurance od Arthritis.Risk factors such genetics is the may factor for Rheuamtoid Arthritis.Their food habits,the environment in which the patients work etc were considered.
- **Applying Apriori Algorithm:**Apriori algorithm is the basic algorithm used for improving efficiency of generating frequent itemset.The two main steps used in Apriori Algorithm are join step and prune step[1].In this paper we use Apriori algorithm to mine the frequent patterns of particular arthritis and its factors.

Two measures of interesting are support and confidence and these are defined as

Let us consider a transaction T which consists of itemset T={X,Y}.

- **Support:** ⟶ Rule X Y holds in transaction set T with support s where s is the percentage of transaction in T that contains X U Y.
- **Confidence:** ⟶ Rule X Y has confidence c in the transaction set T, where c is the percentage of transaction in T containing X that also contain B.

In our work we may consider X as Arthristis which may Rheumatoid,Osteroarthritis etc.,and Y as factors such as food,gender etc.In such case association rules are Rheumatoid→Female.

- **Apriori Algorithm** .Finding frequent itemsets

Input: A transactional database T
Setting a minimum support count threshold.
Output:F,frequent itemsets
begin
Steps 1.Scan all the transactions to count the number of occurances of each particular type of Arthritis.
Step 2. Determining L1-set of frequent 1st itemset.
Step3.Determining C2-2 itemset candidates and scan D for support count.
Step4. Determining L2-set of frequent 2nd itemset.
Step5. Determining C3-3 itemset candidates and scan D for support count.
end
The Apriori algorithm stops its iteration when minimum threshold count is satisfied

- **Association Rules**

The association rules are as follows as shown in below table:

Table 1 Transactional Dataset

Arthritis & Affected Parts	Occurrence/Frequency
Osteroarthritis--Knee	5
Osteroarthritis--Back & knee	1
Osteroarthritis--Joints	1
Rheumatoid--Knee joints	2
Rheumatoid--Knee	10
Rheumatoid--Joints	7
Rheumatoid--Wrist & heel	1
Rheumatoid--Shoulder	1
Rheumatoid--Joins	1
Rheumatoid--Hands Joints	1
Rheumatoid--Back	1
Rheumatoid--Shoulder & wrist joints	1
Rheumatoid--Knee & shoulder	2
Rheumatoid--Fingers	1
Rheumatoid--knee & fingers	1
Rheumatoid--elbow & knee	1
SPA--Knee	2
SPA--Back and Leg joints	1
SPA--Wrist & heel	1
SPA--Knee & heel	1
SPA--Joints	1
SPA--Knee & Shoulder	1

Table 2. Itemsets

Arthritis & Affected Parts	Occurrence/Frequency
Osteroarthritis--Knee	5
Rheumatoid--Knee	10
Rheumatoid--Joints	7

IV. RESULTS AND DISCUSSION

The results of the anaylzes which is carried on Arthritis patient and its factors are as shown below:

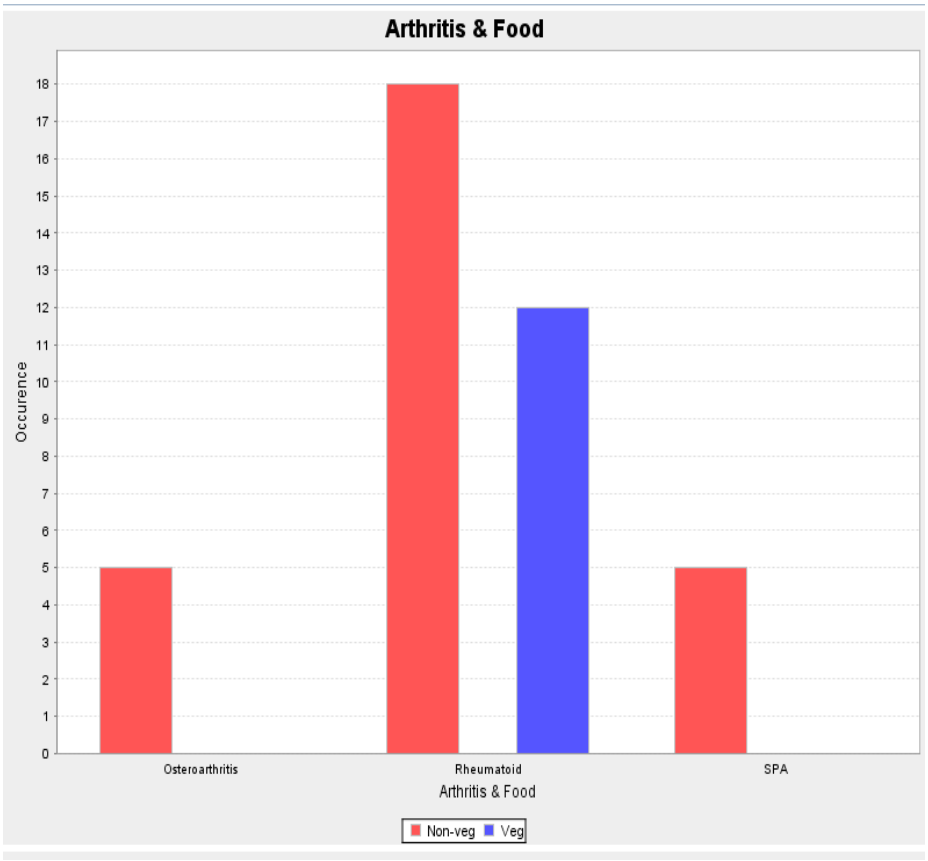

Figure 2. Occurances of arthritis and its Food habits in detail.

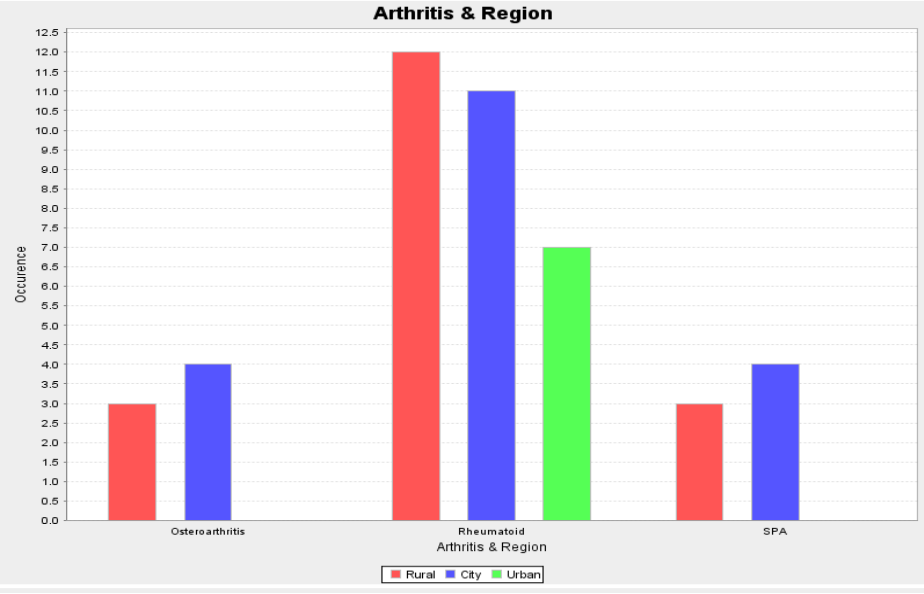

Figure 3 Occurances of arthritis and its Region in detail

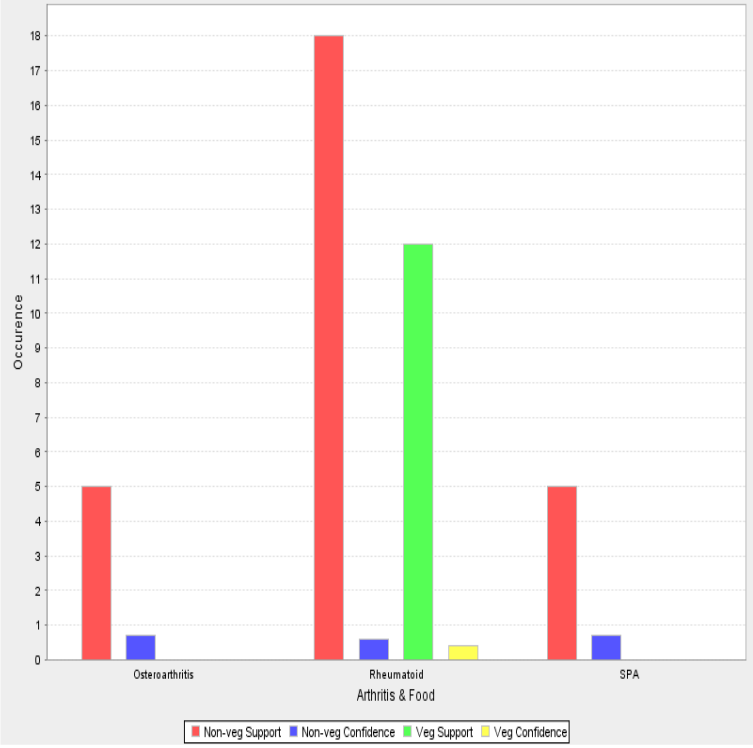

Figure 4 Results of support and confidence

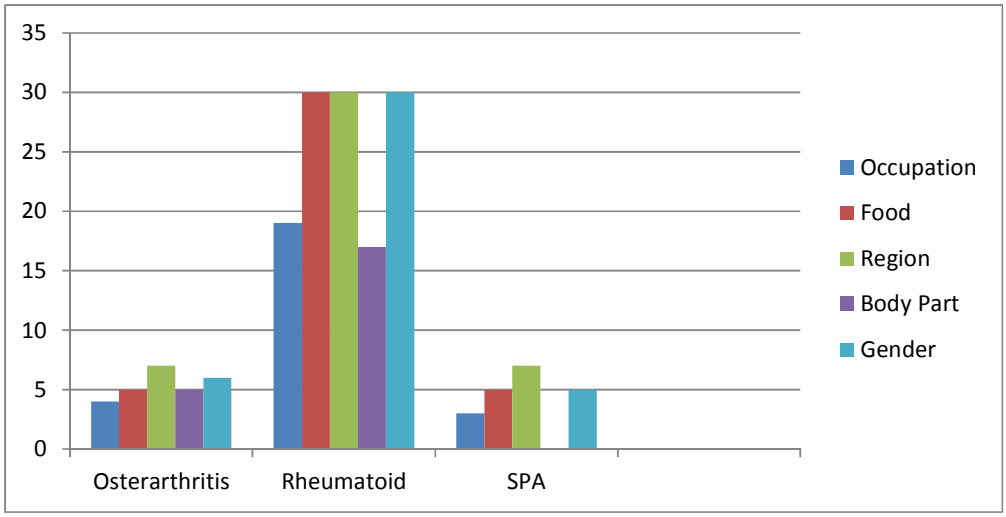

Figure 5.Occurance of of arthritis and its factors.

From Figure 2 we are able to depict more clarity in this research that the type of food which is the causing factor. We can also analysis that region the people in rural area also a factor in Rheumatoid Arthritis.From Figure 3 detail study of specific region that must be considered is depicted.In similar way other factors may be analyzed .

From the above Figure 2 we observer that in a total count of food habit(30) Rheumatoid suffering patients, 18 people consume Non-veg and 12 people consume vegetarian food.Similarly form Figure 5. we observe that patients suffering from Rheumatoid Arthritis Region factor(30) is also the same as food,and from Figure 3 we observe that patients suffering from Rheumatoid Arthritis,patients belonging to rural area are 12 in number and people belonging to city are 11 in number and urban 7 in number.Figure 4 is the study of support and confidence.

V. CONCLUSION

In this paper,we have implemented Apriori Algorithm.To depict association rules by applying Apriori algorithm on data set of Arthritis patients'.Further,support and confidence of dataset is also shown in pictorial representation.The implementation succeeds,in finding correlation between particular arthritis and its factors such as food,region.

However,in future one may opt for implementation of Aprior algorithm on Arthritis patients and their co-related diseases.

REFERENCES

[1] Jiawei Han,Micheline Kamber,Jian Pei,"Data mining Concepts and Techniques",3rd edition,USA:Morgan Kaufmann Publishers,2012,pp 243-250.

[2] R.Karthiyayini,J.Jayaprakash,"Associative Technique on Prediction of Chronic Diseases Using Apriori Algorithm",International Journal of Innovative Research in Science,Engineering and Technology,vol.4(**2015**).

[3] Gitanjali J, C.Ranichandra,M.Pounambal,"APRIORI algorithm based medical data mining for frequent disease identification",International Journal of Information Technology,vol.2(**2014).**

[4] Shelly Ahuja,Gurpreet Kaur,"Review of Association Rule Mining Using Apriori Algorithm",International Journal of Advanced Research in Computer Science and Software Engineering,vol.4(**2014**).

[5] M.Durairaj, V.Ranjani,"Data mining Applications In Healthcare Sector:A study",International Journal of Scientific and Technology Research,vol.2(**2013**)

[6] Rajdeep Kaur Aulakh,"Association Rules Mining Using Effective Algorithm:A Review",International Journal of Advanced Research in Computer Science and Software Engineering,vol.5,(**2015**).

[7] Divya Jain,Sumanlata Gautam,"Implementation of Apriori Algorithm in Health care Sector: A Survey",International Journal of Computer Science and Communication Engineering vol.2,(**2013**).

[8] Smitha T.,V.Sundaram,"Association Models for Prediction with Apriori Concept",International Journal of Advances in Engineering and Technology,vol.5(**2012**).

[9] Sunita Soni,O.P.Vyas,"Using Associative Classifiers for Predictive Analysis in Health Care Data Mining",International Journal of Computer Applications,vol. 4(**2010**).

