

Application of Market–Basket Analysis on Healthcare of Children in Schools of Rural South Kashmir

M. Maqbool Pala¹  and Ajaz Ahmad Lone^{*2} 

¹Department of Computer Applications, Government Degree College (GDC), Bhaderwah, Jammu and Kashmir, India

²Department of Education, Government Degree College (GDC), Chenani, Jammu and Kashmir, India

Citation: M. Maqbool Pala and Ajaz Ahmad Lone (2026). Application of Market–Basket Analysis on Healthcare of Children in Schools of Rural South Kashmir. *Journal of Business, IT, and Social Science*. DOI: <https://doi.org/10.51470/BITS.2026.05.02.19>

Corresponding Author: **Ajaz Ahmad Lone** | E-Mail: 1986lone@gmail.com

07 May 2026: Received | 09 June 2026: Revised | 13 July 2026: Accepted | 06 August 2026: Available Online

Copyright: This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Data analysis plays an important role in the present time to understand the patterns by way of exploring it in meaningful ways. Market–basket is one of these methods thickly used to find frequently occurring items in a transactional database and many researchers use the Apriori algorithm for this purpose. This paper presents the application of Market Basket Analysis to the healthcare of learners of schools of rural South Kashmir. This work tries to find frequent infections that occur together in such schools by using the Apriori algorithm. This could help the stakeholders of such institutions to be more cautious about the frequently occurring diseases and take all possible precautionary measures to safeguard the health of its learners so that regular presence of these students is ensured. In addition, it could also help the doctors reach to the accurate diagnosis which would help in speedy recovery of ailing student/s.

Keywords: Frequent disease Frequent itemset Healthcare Market-basket analysis Common symptom.

1. Introduction

In general, hundreds of transactions are recorded in our day-to-day life, which includes, fields such as marketing, finance, health care, etc. In modern times, the usage of manual records for storing any sort of information has diminished to a larger extent due to advancements in technology. Every Organization, Educational Institution, Company are maintaining their digital database to store their records. Every database related to that domain has an ample amount of data related to it, for example, educational institution database has the information related to students of their institution, which helps to maintain the records of all students' activities, their academic progress, sports activities as well as their health conditions. Teachers monitor health condition of students by observing certain visual changes or by monitoring attendance records. As these databases have thousands of records, it is impossible to read the database manually and retrieve the required information out of it. So, various data mining techniques (association, clustering, classification, and prediction) are used to extract hidden and useful information from these databases. Market Basket Analysis is one of such techniques of data mining that focuses on "particular pattern" of itemset. The aim of this analysis is to identify the interesting patterns by considering the things happening together. To identify the patterns, analysis of the itemset in the database is performed. This requires details on the recorded transactions, which includes the information such as, the coexisting or interrelated transactions in a database. Data mining has immense potential for identifying the meaningful and hidden patterns in the medical database also. In medical cases most of the pathogenic infections lead to secondary infection of some opportunist organisms.

In other words, most of the times more than one pathogen attack a body at a time. In the present work, an online repository of healthcare maintained by PHCs of Rural South Kashmir followed by occasional health camps conducted by them, dataset which includes Electronic Health Records (EHR) of several student patients, the information about their details from different schools, diseases a student has suffered/are suffering from, year in which got affected by the diseases etc. is used. The current work identifies all the frequent diseases that are occurring in such areas by analyzing this medical data set.

2. Literature review

Market Basket Analysis – a technique of data mining, generally used to identify meaningful associations in a customer purchased data. Market Basket Analysis is done using different data mining algorithms. Market Basket Analysis can be implemented by using Apriori algorithm [1-3]. This algorithm determines the frequency of data or itemset from the transaction database using candidate itemset generation. Market Basket Analysis can also be implemented using the FP-tree algorithm [4]. Data here is structured in the form of a tree. Algorithm creates an FP tree using two iterations over data. This algorithm generates frequent itemset without the generation of candidate itemset. Market Basket Analysis using the FP-growth algorithm [5-6], uses a technique in which it finds long frequent patterns with several short recursive modes; it connects the least frequent itemset to supply good selectivity. It reduces the search overhead. Market Basket Analysis is implemented using the FP Bonsai algorithm [7]. It is implemented to mine the frequent patterns that result in efficient frequent itemsets which highlights similar constraints.

The usage of the FP bonsai algorithm has proven to be having greater efficiency and better results in lesser execution time. The Apriori and FP growth are combined to simplify mining of frequent pattern of itemsets (Singh et al. 2014). It focusses on the web patterns of server log files. This combination of FP Growth and Apriori Algorithm can be applied to application of Market Basket Analysis in any social sector [8]. Many authors are having different opinion on using different algorithms or at times a combination of one or more algorithms for analyzing databases of various sectors [9] uses multiple mining techniques to implement Market Basket Analysis. [10] uses an approach of single-layer feed-forward partially connected neural network. It takes less time as repeated scanning of the database is avoided and helps in improving the efficiency. Apriori is also used in medical research with the aim of reducing the manual way of analyzing the data in medical domain. For example, to detect the Hepatitis A [11] and to find frequent diseases Hepatitis and Gastritis [12]. An efficient technique for Apriori algorithm in medical data mining is proposed by [9]. Another novel approach proposed for disease prediction using weighted Association rule and by [13].

Table A, summarizes various approaches proposed by various authors for such research. This table clearly shows various applications based on the market-basket analysis. Analysis of medical data, however, becomes all the more important as is believed to ease the work of doctors. The existing work focuses on finding frequent diseases, medication and to predict the disease based on the symptoms etc. Focus is also on finding the diseases that could occur together. Hence, the present work considers the application of market-basket analysis to health care, in particular to find frequently occurring diseases. Accordingly, a disease dataset with 12 several types of diseases is considered. As the number of diseases is 12, the traditional Apriori algorithm is used.

Table 1: Literature review summary

Algorithm	Purpose	Remarks
Apriori [3-4]	Market-basket analysis	Finds frequent item sets. Requires a greater number of database scans
FP-tree [5-6]	Market-basket analysis	Finds frequent item sets. Visualizes data as a tree. Requires a smaller number of database scans
The Apriori and FP growth are combined [7]	Application of Market Basket Analysis	Finds the frequent web patterns
Neural network based [1;8]	Market-basket analysis	Avoids repeated database scans

3. Methodology

Multiple approaches were used for this study. Approach proposed in [8] also proved helpful. The philosophies, approaches, strategies, choices, timeline, techniques and procedures which are useful for the present work are highlighted in Fig 1. The method used in the paper is the Apriori. Apriori is applied to the dataset, and reduction is applied to mine frequent patterns among the diseases that are occurring in a particular year and in a particular region.

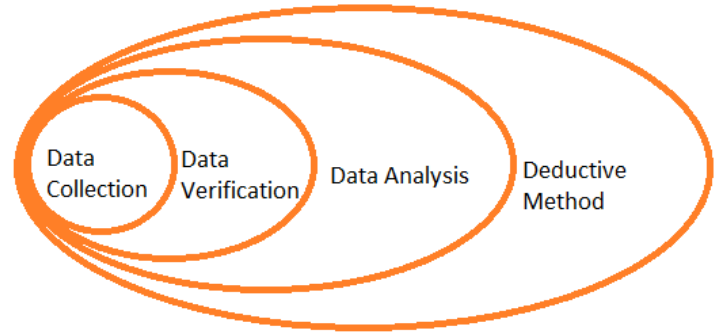


Fig 1: Data Collection, verification, analysis and Methods

The experiment consists of two phases. They are Candidate Itemset generation and Pruning. The Candidate Itemset generation, generates all candidate Itemsets from the given set of items, and Pruning is carried out based on the user defined minimum support. Pruning method follows a property in which an itemset is frequent if and only if all its subsets are frequent. The dataset, which is going to be analyzed, is found to fit into the existing algorithm in using inductive logic. The results obtained in the experiment gives frequent diseases occurring in a particular year or in a particular region. Data collection is done using Electronic Medical Record of various Primary Health Centres located at various regions of rural South Kashmir, through interaction with certain doctors having specialization in community medicine and from school administration. The focus was on the selective diseases (12) that have been found prevalent and recurring in those areas. Candidate itemset was decided and generated for analysis. Frequency of diseases and pattern of simultaneous infections was identified using Apriori. The Pruning step removes all infrequent items based on the subsets of an itemset. This process is repeated till entire database is scanned for such combinations/ patterns. Finally, frequent itemsets of all the passes of the algorithm are combined to give global frequent itemset reflecting frequent diseases occurring in a particular year/ region.

4 Result and observations

After applying the Apriori algorithm to the dataset taken, frequent disease itemsets are found. All frequent itemsets: 1-item, 2-item, 3-item, etc. are generated from the candidate itemsets. These frequent disease itemsets help in taking preventive measures to overcome diseases as are presented with common symptoms.

Table 2 shows the sample input data set and Table 3 shows the sample output obtained. The minimum support count considered for this experiment is 92%.

Table 2: shows sample input dataset

ID	Disease Name
1	Measles
2	Malaria
3	Hepatitis C
4	Chicken Pox
5	Gastritis
6	Mumps
7	Cholera

Consider the 1-item frequent set from the result, which refers to the diseases, which are occurring often (Row 1 of Table 2). Consider the 2-item frequent set {1,4} from the result indicate diseases at respective positions in Table 2, these diseases have common symptoms like abdominal pain, rashes, nausea, and fever, diarrhoea etc. Consider the 3-item frequent set {3,5,7} from the result, where 3 refers to Hepatitis C, 5 to Gastritis, 7 to Cholera etc.

Table 3: Frequent itemsets

S No	Frequency of itemset size	Frequent sets of diseases
1	1	{1}, {2}, {3}, {4}, {5}, {6}, {7}, {8}, {9}, {10}, {11}, {12}
2	2	{1,4}, {2,8}, {3,5}, {9,12}, {10,11}
3	3	{1,4,8}, {3,5,7}, {5,9,12}, {8,10,11}
4	4	{1,4,6,8}, {3,5,7,9}, {8,10,11,12}

Graphical Representation of Frequent Disease Itemsets

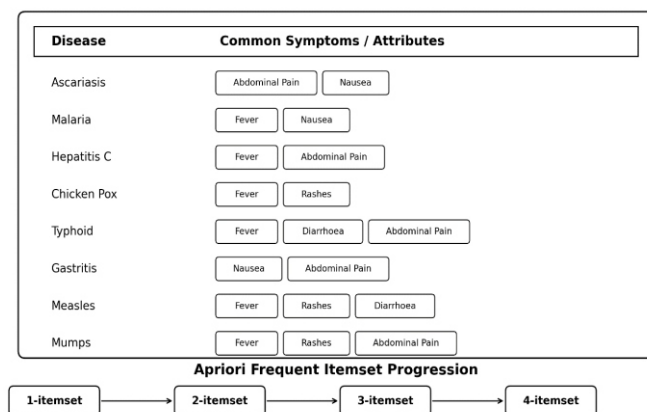


Fig. 2. Graphical Representation of Frequent Disease Itemsets

Fig. 2 A graphical representation of frequent diseases with the associated symptoms. Ascariasis is always presented with abdominal pain and loss of appetite etc., while as Malaria, Typhoid share common symptoms like Abdominal pain, Nausea, fever etc. Similarly, diseases like Hepatitis C, Gastritis and Typhoid share symptoms like Abdominal pain, Nausea and Diarrhoea. These common symptoms also help to identify the diseases that occur or may occur simultaneously. Most of the student patients in rural south Kashmir schools were found to suffer from such patterns of diseases. Responsible factors include contaminated water, open defecation, and tight spaces to share. Poor hygiene and improper vaccination and poor nutrition add to the misery. For the diseases in the less frequent itemsets like Ascariasis, the common symptoms of the diseases will be more specific than the higher frequent itemsets. Similarly, all other diseases that occur frequently together are found by the current work as shown in Table 3. By looking into this information, doctors can plan the required necessary medications well in advance to ensure the safety of these students.

The study can also be used to predict and prepare management for other people living in those areas as are sharing same conditions. Therefore, Health workers can create awareness among common masses regarding the same so that they can protect themselves from being affected with possible diseases. Thus, the Table 3 gives information regarding the frequently occurring diseases as well as diseases that occur together frequently.

References

1. Ali Y, Farooq A, Alam TM, Farooq MS, Awan MJ, Baig TI (2019) Detection of schistosomiasis factors using association rule mining. IEEE Access 7:186108–186114
2. Bhargav A, Mathur RP, Bhargav M (2014) "Market basket analysis using artificial neural network," International Conference for Convergence for Technology, pp. 1–6
3. Borgelt C, Kruse R (2002) Induction of association rules: apriori implementation. Compstat 1:395–400 Chavan G, Gaikwad N, Samal T, Sonule A, Palivela H, and Patil AC (2014) "Various mining techniques defined for mining product valuation instances in market basket data," International Conference on Green Computing Communication and Electrical Engineering (ICGCCCE), pp. 1–6
4. Gayathri B (2017) "Efficient market basket analysis based on FP Bonsai," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 788–792
5. Harahap M, Husein AM, Aisyah S, Lubis FR, Wijaya BA (2018) Mining association rule based on the diseases population for recommendation of medical need. J of Phys Conf Series 1007:12–17
6. Ilayaraja M, and Meyyappan T (2013) "Mining medical data to identify frequent diseases using Apriori algorithm," 2013 International conference on Pattern Recognition, Informatics, and Mobile Engineering, pp.194–199
7. Khan MA, Pradhan SK, Fatima H (2019) An efficient technique for Apriori algorithm in medical data mining. Lecture Notes Netw. Syst. Springer Nature 32:187–195
8. L. Min, W. Chunyan, and Y. Yuguang, "The Research of FP-Growth Method Based on Apriori Algorithm in MDSS," 2010 International Conference on Digital Manufacturing & Automation, Changsha, pp. 770–773, 2010.
9. Lakshmi KS, Vadivu G (2019) A novel approach for disease comorbidity prediction using weighted association rule mining. J Ambient Intell Humanized Comput. <https://doi.org/10.1007/s12652-019-01217-1>
10. Nengsih W (2015) "A comparative study on market basket analysis and apriori association technique," 3rd International Conference on Information and Communication Technology (ICoICT), 461–464
11. Saunders M, Lewis P, and Thornhill A (2009) "Research methods for business students", Fifth Edition, Pearson Education Limited, England
12. Singh AK, Kumar A, and Maurya A K (2014) "An empirical analysis and comparison of apriori and FP- growth algorithm for frequent pattern mining," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, pp. 1599–1602, 2014.
13. Yongmei L and Yong G (2009) "Application in market basket research based on FP-growth algorithm," 2009 WRI World Congress on Computer Science and Information Engineering, pp. 112–115