IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-7 E-ISSN: 2347-2693

Mining Association Rule of Frequent Itemsets Measures for an Educational Environment

N. Balajiraja

Department of Information Technology,
J J College of Arts and Science (Autonomous), Pudukkottai, Tamilnadu, India

Available online at: www.ijcseonline.org

Received:02/Jun/2016 Revised: 16/Jun/2016 Accepted: 11/Jul/2016 Published: 31/Jul/2016

Abstract— This study deals with the design of an hostel inmate - informatics system, which addresses the issues to discover the fact likeness to stay. By using Data Mining (DM) techniques, the data stored in a Data Warehouse (DW) can be analyzed for the purpose of uncovering and predicting hidden patterns within the data. So far, different approaches have been proposed to accomplish the conceptual design of Data Warehouse by applying the multidimensional modeling paradigm. This paper presents a novel approach to integrating data mining model into multidimensional models in order to accomplish the conceptual design of Data Warehouse with Association Rules (AR). To this extent, the Association Rules for modeling in the conceptual level. The main advantage of our proposal is that the Association Rules rely on the goals and user requirements of the Data Warehouse, instead of the traditional method of specifying Association Rules by considering only the final database implementation structures such as tables, rows or columns. In this way to show the benefits of our approach, implementation of specified Association Rules would be created on a commercial database management server.

Keywords— SData Mining ; Data Warehousing; Multidimensional; Association rule;

I. INTRODUCTION

In the present educational environment, providing accommodation to students in the college hostel with necessary facilities, safeguarding and inducing them to concentrate on their studies hour become an arduous. Students staying in hostel differ in state, race, religion, language, food, clothe, culture and habits. For example NRI students and north Indian students especially from Manipur, Nagaland, Assam etc. Studying in tamilnadu are provided separate accommodation and facilities are given to take the food which they desire. Some college provide separate single room with attached bath room. Without minding the money, a few are ready to pay a very high amount for leading a sophisticated life.

Here carries are going to be analyzed for students' hesitation to stay in hostel in general. In a few hostels, students may be restricted to use adequate water for taking bath and washing their clothes.

In some hostels limited quantity of food may be supplied. They may not like more usage of spice, oil, butter, ghee etc for the preparation of food in the mess. Sometimes hostlers may be harshly treated by assistants or supervisors or hostel office clerks. The cleanliness maintained in the kitchen may be disliked by some.

In fact, if the Institution hostel is able to gather data coming from different students (existing databases called students databases) and to analyze them, the hostel can supply its foods the requested quality in due time and can react faster when students have problems. Then hostel could become more efforts and be able to develop more and more its quality accommodation.

Frequently, data are shared throughout different Institution hostel areas and applications and are not usually centralized in a centralized unique view. Consequently, the process of decision-making from data analysis is uneasy. Decisional information systems (DIS) have been recognized as an appropriate approach to facilitate access of data stored in different applications. The aim of DIS is to discover links between data in order to check a set of intuitive hypothesis based on experience. This way, DIS allow to understand and to evaluate the quality within environment (for example the cost-effective factor). Coupled with models, analytic tools, and user interfaces, DIS have the potential to provide information and knowledge (in offering supports to identify opportunity, to make critical decisions and to formulate strategy and to evaluate them, etc.).

Successfully supporting decision-making is based on the availability of integrated high-quality data organized and presented in a timely and easily understandable manner. Data warehouses emerged to meet this requirement. The principle of data warehouses is to gather the whole or a part of data (which are stored in a distributed, not necessary connected databases) in a centralized unique view. The main interest is to provide a global vision of the overall distributed databases in order to ensure a global consistency of DIS.

To gather all needed data into a single place in order to have a global view could be interesting. This way, Hostels can anticipate on the basis of reliable information and can propose its students a global and efficient offer. Consequently, it could be interesting:

- To analyze the students' activity and the intake of food and maintenance quality accommodation in an institution,
- To identify students' health problems concerning habitations uncertainty and the daily use water.
- To propose required information and knowledge for decision-making process about hostel management, supervisor and warden of Institutions.

In order to reach these goals, we provide tools to help the decision makers. This study takes into account the overall process of retrieval, transforming and loading data from disparate databases to the data warehouse prior to the process of Knowledge Discovery in Databases (KDD). This term (KDD) is used to describe the entire information discovery cycle. It thus includes both creation and access to major databases as well as all the procedures required in order to extract information and to be able to use it.

Our approach uses the data warehouse and KDD approaches coupled to performance evaluation by discrete event simulation. This would like to propose an operated modelling environment which kernel concerns data warehousing approach. This modelling environment should allow the hostel to optimize the better students quality and like to stay.

The DIS will provide the international company, decision aid-tools for decision makers in order to obtain information and knowledge for operational, tactical but also strategic decisions. These tools will allow the firm to: Reach the goal between the needs required for taken into facility (inmate, food, water, health care, etc.) in order to make an sufficient services to their students, Increasing the students care at right timed (the moment is identified with an analysis of evolution of performance rate), and Gain students good feedback.

This paper is organized as follows. At first, section 2 focus literature review on algorithm for mining frequent itemsets. Section 3 introduces the proposed modelling environment built with the kernel based on data warehousing. Section 4 describes the architecture of the logical model of designed data warehouse. Section 5 deals association rule mining and demonstrates a real case study concerning the relationships of entities being used in proposed methodology. At the end, the finding are concluded in research perspectives.

II. .LITERATURE REVIEW ON ALGORITHM FOR MINING FREQUENT ITEMSETS

In 2000, Attila Gyenesei put forward an important method of mining association rules for market analysis [1]. The main task of every association rule mining algorithm is to discover the sets of items that frequently appear together—the frequent itemsets. The number of database scans required for the task has been reduced from a number equal to the size of the largest itemset in Apriori [2], [3].

Agrawal explain this with good example and implementing association rules method in the customer transaction [4]. Apriori is using circulatory generation for searching frequent itemsets that produces (k+1) – itemsets from k – itemsets [5].

The Traditional association rule mining algorithms can only be used to data mining problems with categorical attribute. For a data mining problem with quantitative attribute, it is necessary to transform each quantitative attribute into discrete intervals. The first is mining frequent itemsets with Apriori, and then producing association rules according to the frequent itemsets mined [6], [7].

LI Pingxiang proposed a method explores the database to filter frequent 1-itemsets and then it obtains the candidate frequent 2-itemset, 3-itemsets up to n-itemset by evaluating their probabilities in Equation [8].

HUANG Liusheng discovering a new algorithm BitMatrix, This algorithm is compared with the previously known algorithms, the Apriori and AprioriTid algorithms [9].

By analying Carlos Ordonez several problems come up when trying to discover association rules in a high dimensional data set [10].

This paper find out and address the problem of constraint-based rule mining in dense data. For example, most association rule miners allow users to set an alternative measure such as lift (Berry and Linoff, 1997; International Business Machines, 1996) or conviction. Then, this paper pointed out additional measures for identifying interesting rules, including lift and conviction. Both lift and conviction represent the predictive advantage a rule offers over simply guessing based on the frequency of the consequent [11], [12].

Daniel kunkle focused three possible straightforward solutions. First, mine all frequent generalized itemsets, and then eliminate the non-max ones. Second, mine max frequent itemsets in the ordinary case and third choice is to dynamically browse the lattice of all generalized itemsets [13], Calvanesea et al present a methodology for data warehouse design based on a conceptual representation of the enterprise, which is exploited both in the integration phase of the data warehouse [14].No modelling environment is proposed. The main works treat only of methodological framework to optimize and to evaluate performances [15], [16].

The existing works tried to finding the strong rules but in such works didn't give the expected rules. Some of the findings made from the previous works are:

Need several iterations of data.
Rule complexity, Irrelevant rules for predictive purposes, Association rules having irrelevant or redundant groups of items and Weak implication and large number of discovered rules. Mine all frequent generalized itemsets solution is extremely inefficient and, in practical cases, infeasible. Lift and Conviction measures still fail to fully enforce Occam's razor.
As a consequence to overcome the above said drawbacks, our proposal method is based on four features namely, First data preparation and select the required data, second generate itemsets that determines the rule constraints for knowledge, third mine k-frequent itemsets using the new database and fourth generate the proposed association rule that establishes the knowledge base, it will produce strong association rules which satisfy both minimum support degree (min_sup) and the minimum confidence degree (min_conf).

After introducing the basic concepts of data warehouse and literature review of frequent itemsets, we present in the next section, the proposed modelling environment based on data warehousing approach.

III. THE MODELLING ENVIRONMENT

The general functions of hostel management are to: categorize students, health care, provide facility and allocate resources, but also to control and to manage Hostel activities.

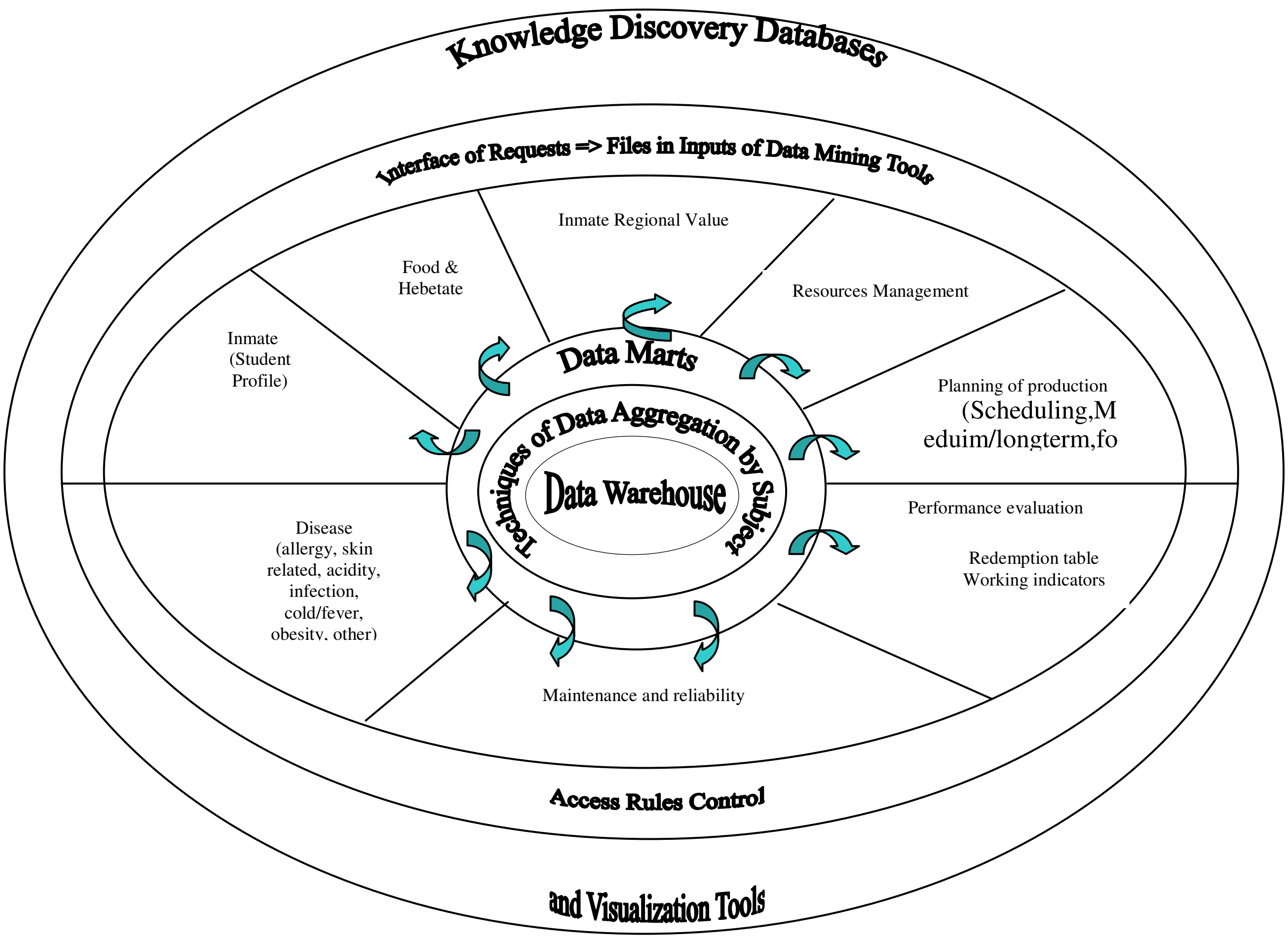

Fig. 1 The main characteristics of the proposed modelling environment

Hostels should systematically acquire the information needed to make decisions and to evaluate the effects of these decisions. To fully explore the opportunities for our approach, the propose a modelling environment based on data warehousing approach allowing to manage the resources, to elaborate planning strategy in food and Hebetate or in maintenance or in disease. So, the role of information is crucial in the methodological framework. Fig. 1 presents the main characteristics of the proposed modelling environment.

This modelling environment takes into account all the processes of extracting, transforming and loading data from disparate databases to the data warehouse until the knowledge discovery in database. The kernel of the modelling environment depends on the data warehouse architecture. From the data warehouse, we design data marts (small data warehouses) by extraction and aggregation of data according to specific subjects. Each subject is split up on modules concerning the Inmate Profile, the production planning, the Food and Hebetate, the maintenance and reliability and the performance evaluation.

Each module concerns one or more activities. All identified modules are in interaction. For example, if they interested by production, they take into account together not only the planning but also the Disease, the resource management, the maintenance, the food and Hebetate , etc. guarantee
The modelling environment will allow a step-by step process to help the hostel administration to:

- Evaluate the quality, reliability and controlling the maintenance costs of the accommodation,
- Define the inmate of students as well as their profile in order to determine inmate's regional value, Food & Hebetation, etc. and
- to plan for the medium-term and long term production,
- For the Hostel Institution, anticipate the resource needs in terms of planning, and simulate different scenarios to identify the best one.
- Study and analyze the inmates' feedback,
- Determine the impact of the inmates' behaviours on the food intake,
- etc…

The originality of our research is to treat at the same time aspects of design of data warehouse and their exploitation (using knowledge discovery in large databases) within a modelling methodological framework for optimization and performance evaluation. Very few authors look into this problem. Information sources and during the knowledge discovery activity on the information stored in the data warehouse propose a data warehouse providing the student information required to optimize the time that the Institution hostel decides to file the compound of a patent.

The kernel of our modelling environment is based on the data warehouse and the performance of the data warehouse depends on the model architecture. In the next section, we present the multidimensional models concerning our data warehouse.

IV. LOGICAL MODELS OF DATA WAREHOUSE

Gathering of data into a unique view facilitates discovery of data structures to deduce knowledge for decision making in order to:

- Identify inmate's acceptable and unacceptable hebetate to categorize the diseases,
- Regional weather value to change the likeness of staying in a chosen place,
- Measuring solid and liquid intake and analyzing the intake variations, gathering the overall infrastructure information about the considered environment (warden and deputy warden relationship, hostel police), etc.

The performance of the DIS depends on the data warehouse architecture. In the next two paragraphs, we present the architecture of the data warehouse logical models concerning our study. The proposed models concern conceptual level (conceptual representation of the corporate data). The physical models (implementation models) are not presented in this article.

Logical model of data warehouse for solid and liquid intake, habituate, identify the diseases and students like to stay hostel:

Institution administration aims to study the inmate food intake, acceptable and unacceptable habituate, and regional weather value, are the main reasons for health problem. But the hostel prepares and provides hygienic food and clean environment regularly. Many strategies followed in collecting all these inmate information, finally results to improve residence facility, inmate's likeness, etc. Fig. 2 presents the corresponding data warehouse dedicated to inmates details, habituate and diseases. Database tables are normalized in a relational model. For example, inmate master (student file), tables of student interest (habituate and intake), table of inmate's diseases, inmate's regional values.
The problem to find efficient techniques for processing complex queries has been solved.

We chose to use constellation schema [2]: Multiple fact tables share dimensional tables. We have identified four fact tables (in gray, Fig. 2):

- Inmate (this is concerning inmate details like name, address, date of birth, age, religion, city, contact number, etc.)

- Hebetate (the measurements concerning acceptable and unacceptable hebetate)
- Liquid intake(the measurement concerning per day liquid intake)
- Disease (the measurement concerning after solid and liquid intake, acceptable and unacceptable hebetate, low drinking water intake, etc.)

The other tables concern the dimensional tables (potential analysis axis).
This data warehouse also gives needed information in order to forecast production planning. In fact, with the help of inmate fact table to know the inmate personal details like height and weight calculation, age group details, then the home town weather one of the possible to anticipate to stay in outside. Hostel polices and rules are not allowing acceptable and unacceptable hebetate because its affect and encourage to do the same in other students. Inmate unacceptable hebetate, food intake, having low drinking water intake/per day, and other behaviour is most of the person leave to hostel. But it is not good for his carrier. The hebetate is main problem of inmate diseases, not the hostel environment and hostel food preparation.

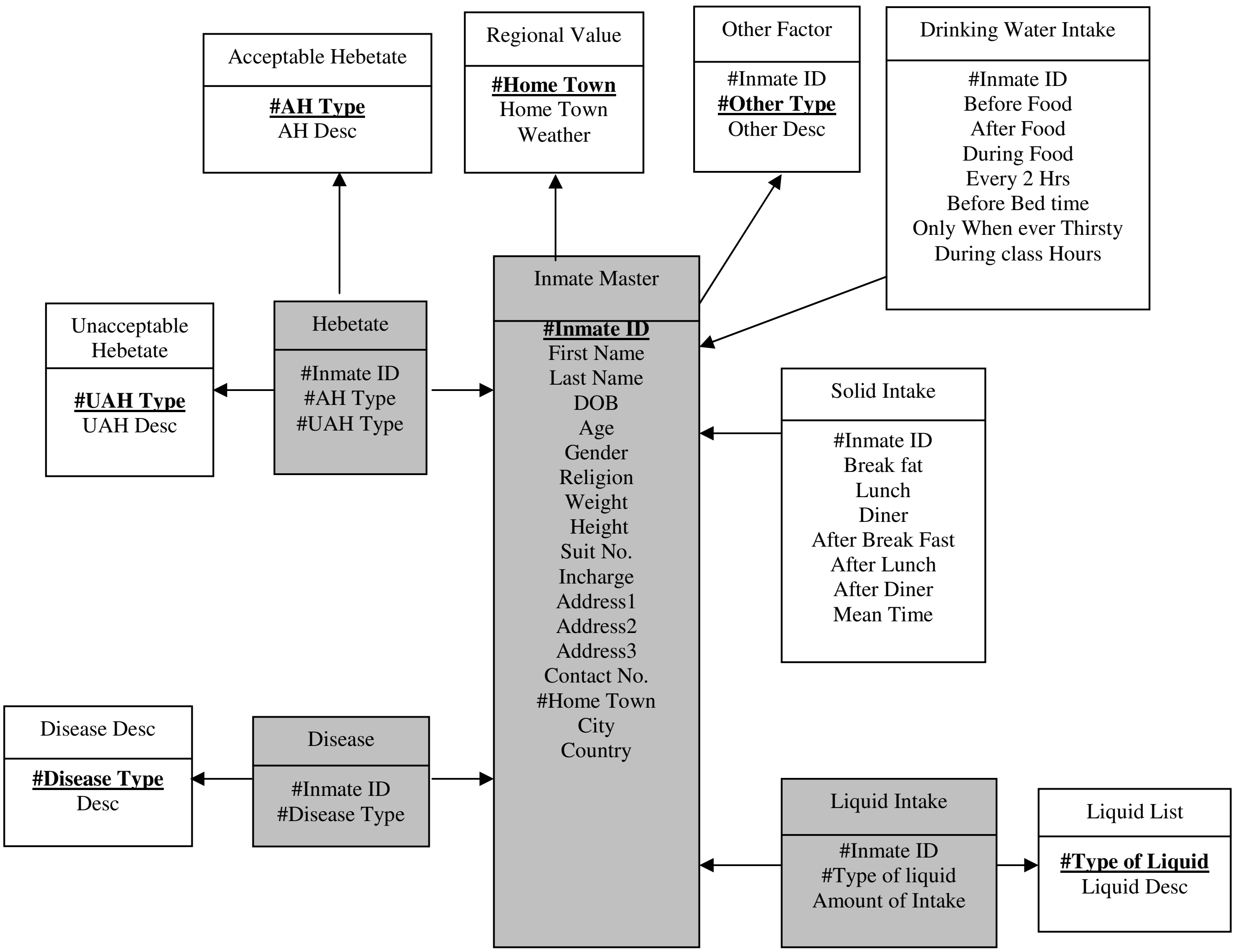

Fig. 2. A part of the logical model of data warehouse concerning

A. *Inmate* details

- Here insert the inmate personal information for further communication, if the incharge want to inform his/her parents.
- Number of time going in and out from hostel(month,week)
- Home town information to know the weather.
- Calculate body weight(compare height and wieght)

B. *Hebetate*

- This can be divided in to two table that is acceptable and unacceptable hebetate.(inmate no, AH(acceptable hebetate) type,UNH(unacceptable hebetate) type)
- here they point out daily activity (AH type,AH

  Desc)
- Here they point out unacceptable think like drugs, tobacco, etc.(UNH type,UNH desc).

C. *Liquid* Intake

- Average intake of drinking water and other liquid food per day. It is very important for human body.
- Some students very eager to take beverages liquid drinks in out side of the campus, it is not good for his/her health.
- The beverages drinks create more side effect to human body like teeth - pain, injuring, poor strength, abdomen pain, obesity problem.
- If they take more coffee and tea in his/her carrier affect gallbladder.
- The student take soups and fresh juices, it's very good for health, but most of the student not interest to take like this.

D. *Diseases*

- This section is mainly focus on students side effect.
- Students focus his/her health problem to leave hostel, but it is not true.
- Most of the health problem arrive from his/her poor hebetate only. Like some students interest chunk food, deep fry items, tawa fry items and late night diner, it is create obesity problem. After food most of them like easy desert and take heave drinking water. Sometime this type of food hebetate create heart attack disease.
- The other disease main reason is inmate daily activity and friendship, like skin disease, allergy and cold/ fever.

The queries are very simple. With this data warehouse architecture we minimize the use of join operators between tables (important interest to optimize queries). For example, the SQL query concerning ''total liquid intake per student take include liquid food and drinking water '' is the following:

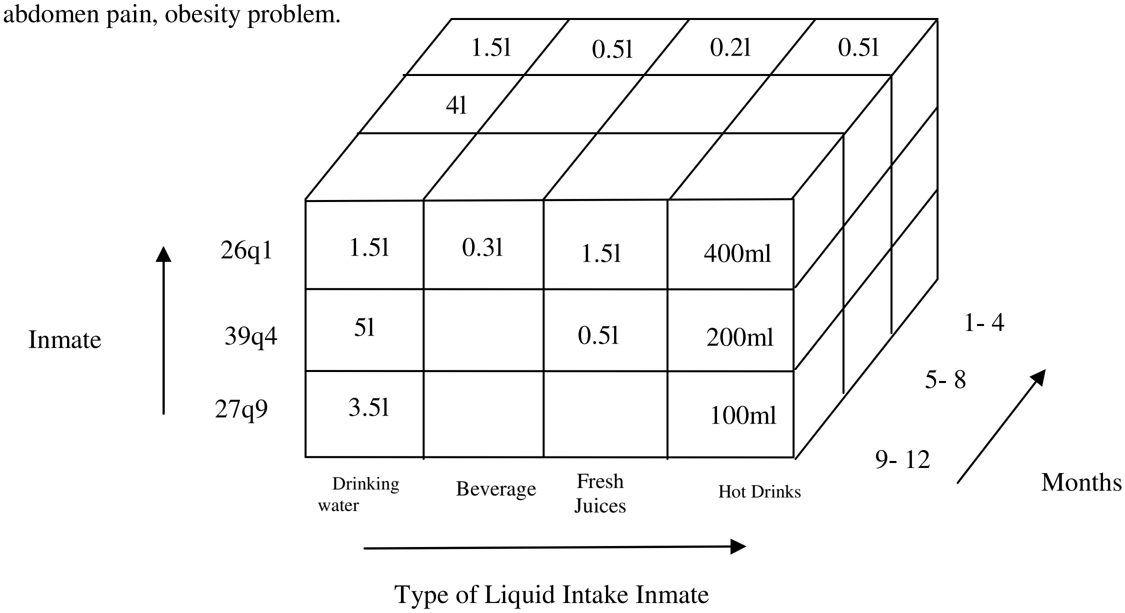

Fig. 3. Multidimensional model data cube.

V. PROPOSED METHODOLOGY

To overcome the above said drawback the proposed method incorporates the following model:

- For each frequent itemsets F, all subsets of F, i.e., F1 can be produced.
- Each non-null subset F1 of F, generates an association rule $F^1 \rightarrow (F - F^1)$, if it's

confidence degree is not less than min_conf. This is the model algorithm of association rules. Here, this method is modified in order to build up the knowledge database of expert system.

A. Establishing the knowledge database

The proposed algorithm is focused on four features, which are expressed as follows.

The task of data preparation is to transform raw transaction data according to the required structure and transfer the transformed data to the mining database. In general, the data used by rule mining are transaction data i.e., the training data set by eliminating the unwanted fields/data (like student name, reg.no, age, address.ect.) from transaction database and select the required data.
Generate itemsets that determines the rule constraints for knowledge.
Mine k-frequent itemsets using the new database.
Generate the proposed association rule that establishes the knowledge base using the proposed algorithm.
Fig. 2 shows the step by step process of the proposed model.

The proposed methodology is validated using student habitation expert system, where using the certain criteria's weather disease occurs or not is identify. First, to pre-processing the transaction data, then select the required data,

The following data is got from college student habitation

1. below 1 litre , soft drinks, tawa fry, vegetarian → Ulcer
2. below 1 litre , soft drinks, tawa fry, vegetarian → Ulcer
3. 3 to 5 litres, beverages, tawa fry, non-vegetarian → Obesity
4. 1 to 3 litres, hot drinks, tawa fry, vegetarian → Acidity
5. 1 to 3 litres , beverages, boiled, non-vegetarian → Acidity
6. 1 to 3 litres, hot drinks, fast food, vegetarian → Obesity
7. 1 to 3 litres , soft drinks, fast food, non-vegetarian → Acidity
8. 1 to 3 litres , beverages, boiled, non-vegetarian → Acidity
9. 3 to 5 litres, hot drinks, fast food, non-vegetarian → Obesity
10. below 1 litre, beverages, tawa fry, non-vegetarian → Ulcer
11. below 1 litre, hot drinks, boiled, vegetarian → Ulcer
12. below 1 litre , soft drinks, tawa fry, vegetarian → Ulcer
13. 1 to 3 litres, hot drinks, tawa fry, vegetarian → Acidity
14. 3 to 5 litres, beverages, tawa fry, non-vegetarian → Obesity
15. 1 to 3 litres, hot drinks, tawa fry, vegetarian → Acidity
16. below 1 litre, beverages, fast food, non-vegetarian → Ulcer
17. 3 to 5 litres, beverages, tawa fry, non-vegetarian → Obesity
18. 1 to 3 litres , beverages, boiled, non-vegetarian → Acidity
19. 1 to 3 litres, hot drinks, fast food, vegetarian → Obesity
20. 1 to 3 litres, hot drinks, fast food, vegetarian → Obesity

B. Item Generation

Establishing the knowledge database for student habitation expert system is based on solid and liquid food consumption habitation data. In order to use the apriori, plain water, liquid intake, solid intake, type of intake and disease name are considered as the items that cause diseases and is being named as: a, b, c, d and e respectively. Because each cause of disease has several different conditions (we call them cause values), the different conditions are also numbered. In this way, we take the form of "cause of disease name following

the cause value number" to subdivide the item. For example a1 (here "a" is cause of disease name following "1" is cause value number). The result is as follows:

a. Plain water: a1, below 1LT; a2, 1 to 3 LT; a3 3 to 5 LT;

b. Liquid Intake: b1 soft drinks; b2, Beverages; b3,hotdrinks;

c. Solid Intake: c1, tawa fry ; c2, fast food; c3, boiled;

d. Type of Intake: d1, vegetarian; d2, non-vegetarian;

e. Disease name: e1, Ulcer; e2, Acidity; e3, Obesity;

Next, we are going to take subdivided item that is "cause of disease name following cause value number" as the final item which is operated by apriori. After the preparation of these items, we begin to mine the frequent item with Apriori.

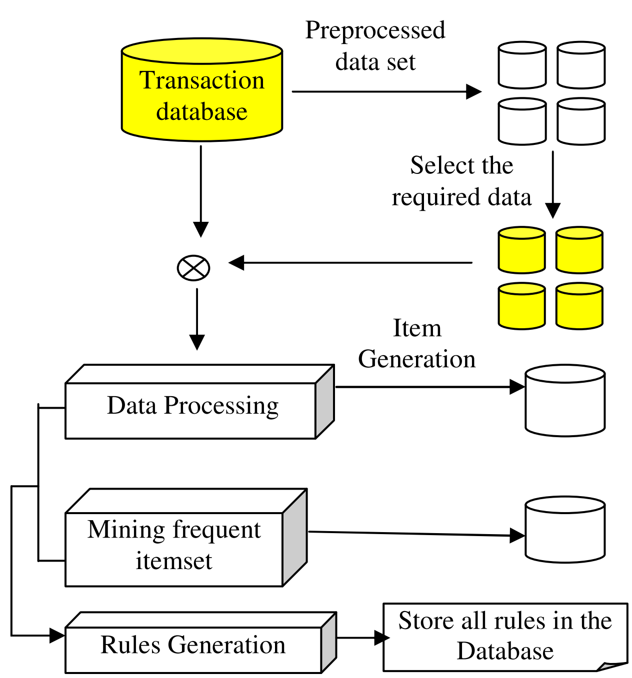

Fig 4: The proposed model

Here, we assume the min_sup and the min_conf are one and the same.

The initial database DB from the experience data is as follows (Table 1):

TABLE 1 INITIAL DATABASE DB

ID	Item
001	a1, b1, c1, d1, e1
002	a1, b1, c1, d1, e1
003	a3, b2, c1, d2, e3
004	a2, b3, c1, d1, e2
005	a2, b2, c3, d2, e2
006	a2, b3, c2, d1, e3
007	a2, b1, c2, d2, e2
008	a2, b2, c3, d2, e2
009	a3, b3, c2, d2, e3
010	a1, b2, c1, d2, e1
011	a1, b3, c3, d1, e1
012	a1, b1, c1, d1, e1
013	a2, b3, c1, d1, e2
014	a3, b2, c1, d2, e3
015	a2, b3, c1, d1, e2
016	a1, b2, c2, d2, e1
017	a3, b2, c1, d2, e3
018	a2, b2, c3, d2, e3
019	a2, b3, c2, d1, e3
020	a2, b3, c2, d1, e3

C. *Mine frequent **k**_itemsets*

Apriori algorithm is used to scan the database to obtain candidate 1-itemsets, then, select it by min-sup 1 and you can get frequent 1 – item. As follows (Table – 2):

TABLE 2 FREQUENT 1 - ITEMSETS

Item	a1	a2	a3	b1	b2	b3	c1
Support degree	6	10	4	4	8	8	10
Item	c2	c3	d1	d2	e1	e2	e3
Support degree	6	4	10	10	6	6	8

The association rule of the model Apriori algorithm links the two frequent (k – 1) – item if they have the same (k-2) items in front. For example: link "a2, b3, c1" and "a2, b3, c2", we can get "a2, b3, c1, c2". However they are not linked while we build up knowledge database. This is because"c1" and "c2" are two different values of one cause.(They both describe the solid intake of food habitation). They repel each other, so if is impossible for them to exist simultaneously. Therefore, we don't link them together. Thus, after the linking work, and through the selection of the min-sup 1, we can get the frequent k – itemsets which satisfies both min-conf and min-sup. For the above example of cause of diseases, we can get the frequent 5 – itemsets finally, using the modifying Apriori algorithm. As follows(table 3):

TABLE 3 FREQUENT 5 – ITEMSETS

Item	a1, b1, c1, d1, e1	a3, b2, c1, d2, e3
Support degree	3	3
Item	a2, b3, c1, d1, e2	a2, b2, c3, d2, e2
Support degree	3	2
Item	a2, b3, c2, d1, e3	a2, b1, c2, d2, e2
Support degree	3	1
Item	a3, b3, c2, d2, e3	a1, b2, c1, d2, e1
Support degree	1	1
Item	a1, b3, c3, d1, e1	a1, b2, c2, d2, e1
Support degree	1	1
Item	a2, b2, c3, d2, e3	
Support degree	1	

It couldn't produce new frequent itemsets any more in after mining out the frequent 5 – itemsets, so the algorithm ends. The following job is to make association rules from the frequent itemsets we have got.

D. *Model association rules Generation*

The proposed association algorithm, finds out all non_null subsets of frequent itemsets F. For each subset F1, if the confidence degree is not less than min-conf an association rule: $F^1 \rightarrow (F - F^1)$ is produced.

The process of the knowledge database establishment, decision rule is rather needed than association rule. In other words, the focal point, which pay close attention to is not the associate relationship of attributes, but the results of the combination of them. Therefore, only need to calculate the confidence degree of the subset besides the decision result's. for example, to the frequent itemsets "a2, b3, c2, d1, e3", we only need to calculate the confidence degree of "a2, b3, c2, d1", that is say f1=" a2, b3, c2 and d1"(Because e3 is the decision result item:disease name). So, when it satisfies the rule that its confidence is more than min-conf. we can produce the rule that is

$F^1 \rightarrow (F - F^1)$ i.e. "a2, b3, c2 $\rightarrow$ e3" ,. At last, compared with the primary attribute number, we can get the comprehensive rule:

1 to 3 litres $\wedge$ hot drinks $\wedge$ fast food $\wedge$ vegetarian $\rightarrow$ Obesity

. (This mean is that if the cause of disease has 1 to 3 litres and hot drinks and fast food and high cholestral vegetarian, the disease might be Obesity). The rules of the above mentioned cause of diseases example is as follows:

Rule form: (name sup as support degree and conf as confidence degree)

1. below 1 litre , soft drinks, tawa fry, vegetarian → Ulcer
sup: 0.15, conf:1
2. 3 to 5 litres, beverages, tawa fry, non-vegetarian → Obesity
sup:0.15, conf:1
3. 1 to 3 litres, hot drinks, tawa fry, vegetarian → Acidity
sup: 0.1, conf:1
4. 1 to 3 litres , beverages, boiled, non-vegetarian → Acidity
sup: 0.15, conf:0.67
5. 1 to 3 litres, hot drinks, fast food, vegetarian → Obesity
sup: 0.15, conf:1
6. 1 to 3 litres , soft drinks, fast food, non-vegetarian → Acidity
sup: 0.05, conf: 1
7. 3 to 5 litres, hot drinks, fast food, non-vegetarian → Obesity
sup: 0.05, conf: 1
8. below 1 litre, beverages, tawa fry, non-vegetarian → Ulcer
sup: 0.05, conf:1
9. below 1 letre, hot drinks, boiled, vegetarian → Ulcer
sup: 0.05, conf:1
10. below 1 litre, beverages, fast food, non-vegetarian → Ulcer
sup: 0.05, conf:1
11. 1 to 3 litres, beverages, boiled, non-vegetarian → Obesity
sup: 0.15, conf:0.33

Fig.4 shows the comparison of rule form name sup as support degree and conf as confidence degree. For example rules 6,7,8,9,10 minimum support is 0.05, rules 1,2,4,5,11 minimum support is 0.15 and rule 3 minimum support 0.1. as well as minimum confidence of all rules is 1 except 4 and 11. Fig. 5 illustrates the performance analysis for the test data set. The results show that proposed Algorithm has better performance comparing to Apriori Algorithm.

At last the knowledge database of the expert system is established and stores all the rules into the database. Thus, according to the rules table in the knowledge database, the system can output the homologous decision rules after the customer inputs some cause of disease.

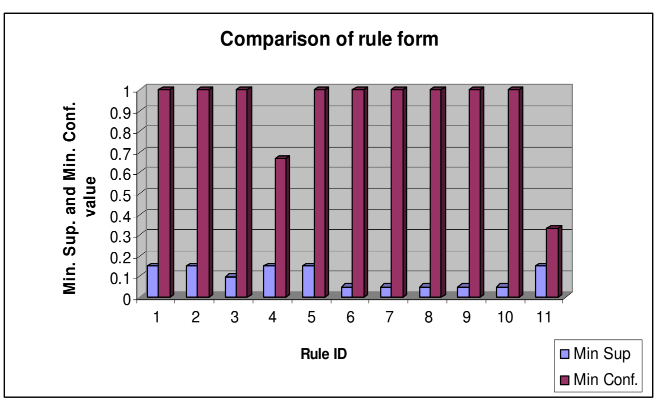

Fig. 4 Comparison of rule form

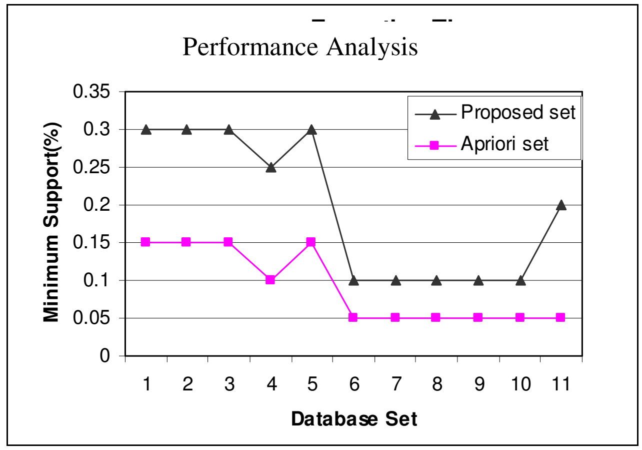

Fig. 5 Performance Analysis for Apriori and Proposed Algorithm

VI. CONCLUSION

In this paper proposed a modelling environment to support decision making process for inmates like to stay hostel and his/her problems came from himself. This environment is based on data warehouse approach and integrates eight modules: Inmate profile, Food and Hebetate, Inmate regional value, Resource management, Diseases, maintenance and reliability, planning of production, and performance evaluation. This paper is illustrated through a real case application concerning the Hostel and the optimization of international residential school and Institution. The main purpose of this paper is to show the contribution of data warehouse in order to identify the activity and the running concerning food intake hebetation and Hebetate of Inmates in an hostel. Data warehouse stores data from which we can identify necessary information and knowledge for decision-making process about management, institute and piloting of Institutions. The limits of our approach concern the disease and hebetation. Then the proposed algorithm used in mining the training data set, which can discover implicit and potential useful knowledge from large preprocessed databases. This paper provides the better knowledge base of the expert system and concludes the role of association rule in discussion. Then, this method is also used in other fields like agricultural sector, health sector and clinical research areas. So it is a better method for establishing knowledge base in expert system.

REFERENCES

[1] Hamid Mohamadlou,"A method for mining association rules in quantitative and fuzzy data" ,978-1-4244-4136-5/09, IEEE, 2009, pp.453 – 458.
[2] Agrawal, R, and Srikant, R, "Fast algorithms for mining association rules" In: Proceedings of the 20th VLDB Int'l Conf., 1994, pp 487–499.
[3] Agrawal, R, and Srikant, R, "Fast algorithms for mining association rules in large database", Technical report Fj9839, IBM Almaden Research Center, San jose, CA, jun.1994.

[4] Agrawal, R, and Imielinski, T, Swami, A, "Mining association rules between sets of items in large databases", In Proc,ACM SIGMOD, May 1993,pp.207-216.

[5] XindongWu and et al, " Top 10 algorithms in data mining", Knowl Inf Syst ,Springer-Verlag London Limited , 2008, pp 14:1–37.

[6] Attila Gyenesei, "A Fuzzy approach for mining quantitative association rules", Technical Report: TUCS-TR-336, 2000.

[7] Zhu Ming, datamining, University of Science and Technology, china Press, Hefel, 2002, pp: 115 – 126.

[8] LI Pingxiang, CHEN Jiangping and BIAN Fuling, "A Developed Algorithm of Apriori Based on Association Analysis", Geo-spatial Information Science ,Vol. 7, Issue 2, 2004, pp 108-112.

[9] HUANG Liusheng, CHEN Huaping, WANG Xun and CHEN Guoliang, "A Fast Algorithm for Mining Association Rules", J. Comput. Sci. & Technol., Vol. 15 No. 6, 2000, pp 619-624,.

[10] Carlos Ordonez, Norberto Ezquerra and Cesar, A, Santana, "Constraining and summarizing association rules in medical data", Knowl Inf Syst, 9(3), 2006, pp 259-283.

[11] Berry M.J.A and Linoff, G.S., "Data Mining Techniques for Marketing Sales and Customer Support", John Wiley & Sons, Inc., 1997.

[12] Brin, S., Motwani, R., Ullman, J., and Tsur, S., "Dynamic itemset counting and implication rules for market basket data",In Proc. of the ACM-SIGMOD Int'l Conf. on the Management of Data, 1997, PP. 255-264.

[13] Daniel Kunkle, Donghui Zhang, Gene Cooperman, "Mining Frequent Generalized Itemsets and Generalized Association Rules Without Redundancy", J. Comput. Sci. & Technol., Vol. 23(1), 2008, pp. 77-102.

[14] Calvanesea, D., Dragoneb, L., Nardib, D., Rosatib, R., and Trisolinic, S.M., Enterprise modelling and data warehousing in TELECOM ITALIA. Information Systems 3, 2006.

[15] Comelli, M., Fe´nie` s, P., Gourgand, M., and Tchernev, N., "A generic evaluation model of logistic process for cash flow and activity based costing for a company supply chain", In: International Conference on Industrial Engineering and Systems Management IESM 2005. Marrakech, Morocco, pp. 113–122, 2005.

[16] Caiyan Dai and Ling Chen, "An Algorithm for Mining Frequent Closed Itemsets with Density from Data Streams", IJCSE, pp. 40 – 48, 2016.

Authors Profile

Dr.N. BALAJIRAJA is an Assistant Professor in the Department of Information Technology at J J College of Arts and Science (Autonomous), Tamilnadu, India. He obtained his Ph.D. in computer science from Anna University, Chennai in 2015. His research interests are in the areas of DataBase and Artificial Inteligent.

