Implementing Data Mining in Primary Health Center – A Review

S. Pushpalatha[1], Dr. Jagdesh Pandya[2]

[1]*(Lecturer - MCA, S.K.Patel Institute of Computer Studies,Gandhinagar – 23, Gujarat, India*
[2]*(Manager –BISAG Gandhinagar – 23, Gujarat, India*

Abstract: - *Data mining is the process of selecting, exploring and modeling a large database in order to discover model and pattern that are unknown. Enormous gathered data in Health care Information society are scattered with different archive systems which are not connected with one another. This unorganized data leads to delay in monitoring, improper planning, defocus the analysis which leads to inaccuracy in decision making. The purpose of this study is to review the relevant data mining tool and its applications in primary health center . This paper focuses on various models and techniques used in data mining for health care and its applications for better health policy-making and in decision making. It provides recommendation for future research in the application of data mining in primary health center.*

Key Words : *- Data mining tools, Health Care Information, Data Mining Health Care, Primary Health Center.*

I. INTRODUCTION

Primary Health Center (PHC) is where health related details of each individual are collected region wise. It emphasizes on health promotion, illness prevention; it includes diagnosis, treatment, and provides a link to more specialized care.

The PHC information system is a sub-set of the National Health Management System. It consists of health maps, house numbering, home-based records, Family Master, the wall chart, health facility/district referral forms, pictorial and tally sheets, Medical forms and Health Facility Registers etc. The PHC information system feeds into the health system in order to attain the desired impact. However, other parallel reporting formats such as the Disease Surveillance and monthly notification exist and continue to feed into the relevant reporting systems.

The PHC staff at the ward level will collect, analyze and use data for decision making and provide a feedback to both the communities and the next level of reporting. Data collected from communities within the ward are sent to the Ward PHC where they are gathered together with facility based data and forwarded to the Local Governance Departments of Monitoring and Evaluation for subsequent transmission to the State and zonal office. Data from the State are sent to the Ministry of Health and those from the zonal office are sent to its headquarters. In this paper the different relevant data mining tools used for PHC's are reviewed and proposes a data model for monitoring individuals information for population based health care management.

II. RELATED WORK

1. Data Mining Methods with Logistic Regression in Childhood Obesity Prediction:

Shaoyan et al. (2009) [1], applied different data mining algorithms to predict overweight and obese children in early age. They have involved genetic, sociological and even psychological factors. The highest overweight prediction rate is 55-60% in there work. They have compared different techniques to get better prediction rates, using decision trees, association rules and Bayesian networks to identify relationships among the attributes which does not give better results. SVM and Bayesian algorithms appear to be the best two algorithms for predicting overweight and obesity from their database. Generally, the results of this work have improved the prediction accuracy when compared to logistic regression and thus begun to show the value of incorporation of non-linear information in epidemiological prediction.

2. Data Mining Techniques in a Healthcare Data Warehouse:

Michael Silver et al. (2001) [2], focuses on an analysis, which was performed by a team of physicians and computer science researchers, using a commercially available on-line analytical processing (OLAP) tool in conjunction with proprietary data mining techniques. The initial objective of the analysis was to discover how to use data mining techniques to make business decisions that can influence cost, revenue, and operational efficiency while maintaining a high level of care. Another objective was to understand how to apply these techniques appropriately and to find a repeatable method for analyzing data and finding business insights. The process used to identify opportunities and effect changes is described.

3. High Performance Data Mining Using the Nearest Neighbor Join:

Christian Böhm (2000) [3], combines one point set with its k nearest neighbors in the other set. It has been shown that many standard algorithms of Knowledge Discovery in Databases (KDD) such as k-means and k-medoid clustering, nearest neighbor classification, data cleansing, post processing of sampling-based data mining etc. were implemented on top of the k-join operation to achieve performance improvements without affecting the quality of the result of these algorithms. They proposed an algorithm to compute the k-nearest neighbor join using the multipage index (MuX), a specialized index structure for the similarity join. To reduce both CPU and I/O cost, we develop optimal loading and processing strategies.

4. Describing pattern in health database:

Wilson et al (2003) [4], began to scan cases where KDD and data mining techniques were applied in health databases. They found confusion in the field regarding what constituted data mining. "Some authors refer to data mining as the process of acquiring information, whereas others refer to data mining as utilization of statistical techniques within the knowledge discovery process". Because of misconceptions still going on in the medical community about what data mining comprises, they defined data mining is the set of procedures and techniques for discovering and describing patterns and trends in data.

5. Clustering algorithm and decision support:

Some diagnostic and laboratory procedures are invasive, costly and painful to patients. An example of this is conducting a biopsy in women to detect cervical cancer. Thangavel et al (2006) [5], used the K-means clustering algorithm to analyze cervical cancer patients and found that clustering found better predictive results than existing medical opinion. They found a set of interesting attributes that could be used by doctors as additional support on whether or not to recommend a biopsy for a patient suspected of having the cervical cancer.

6. Hub and Spoke model using Artificial Intelligence:

Rajan Vohra 2009 [6] evaluated the decentralization of out patient health services at the primary health care level through the Hub and Spoke model. This model uses the dynamic load transfer and balancing and planning deployment of manpower and equipment. It uses Artificial Intelligence to convert health centers into a Hub and Spoke model. AI techniques were used to search the health network for free capacity and overloaded centers. The hub and spoke network model was used to redirect patient flow to the nearest hub. The input of hub and spoke structure is selected based on the above parameters are,

Center name, Patient load/day, No of Doctors

The hub is chosen according to highest patient load and the spoke centers are under loaded and having free capacity. The hub has to decide to service further patient requests for service. The requests can be satisfied by the hub itself, or by its spokes, if free capacity available in them. If no free capacity is available in the newly created hub and spoke network, further requests for service can be redirected to the nearest hub and spoke network. The computation of capacity and excess load is maintained be the following parameters,

Center Name, Type (Hub/spoke), free capacity available.

7. Grid Framework for Predictive purpose:

P. Vishvapathi1. et. Al [7], has shown the backbone structure for integrating different specialized hospitals with the rural Healthcare delivery system. It shows the Distributed Data Mining algorithm could be used for predictive analysis of Health data so as to give a treatment plan for most of the preventive diseases. In this they have specified grid computing to store medical data as a Grid which combines the information from a wide population to extract the knowledge that can lead to the discovery of new correlations between symptoms, diseases, genetic features or any other clinical data. Router is connected to each District Head Quartes office, from the district Hospital Server different hospitals at the District must be brought into this grid framework. Database server was placed for storing the EHR of all the patients falling under a Data Grid. These EHR data at each of the District are consolidated at the State Health Database Server. Based on these PHC Doctor has to prepare an EHR for the Patient, so the personal details from the UID Aadhaar are saved in the district database server. When a rural doctor requests a specialist advice and treatment from the different participating hospitals at the state / district in the health grid, the specialist can view the EHR record of that particular patient and suggest a treatment plan in turn providing remote specialist services to the rural poor.

III. PROBLEM STATED

National Family Health Policy identifies Primary Health Center (PHC) for delivering an effective, efficient, accessible and affordable health services to a wider population. This data may comprise thousands of records which may contain valuable patterns and dependencies hidden deep among them.

The abundant volume of the dataset and complexity of the medical domain make it very difficult for a human to analyze the data manually to extract hidden information. This may lead to an improper decision making which misinterprets the valuable suggestion provided be PHC by National health survey [8].

Based on the related work and the comparative study of various Data mining Techniques, which is helpful to solve various health related problems. This research aims in developing a data model for Public health care using Data mining techniques, where the complexity of the misinterpretation of each individual data will be minimized.

IV. PROPOSED FRAMEWORK

Population based health data management collects all individual information gathered by PHC. The collected individual's information will be recorded with unique identified number. The information about each individual must be collected and the junk information must be stored in a common data repository.

We propose to develop an appropriate data model for monitoring information of all individuals, to reciprocate with the relevant information required for population based health care management and to provide all information service to the individuals.

4.1 Define the problem:
Understanding of the application domain which is relevant to the knowledge and defining the metrics by which the model will be evaluated and defining the final objective.

4.2 Collecting Data:
Data may be scattered and stored in different formats containing inconsistencies with flawed and missing entries. Data Preparation tasks will be performed multiple times with raw data to explore the data and to refine the data. This includes table, record, attribute for cleaning of data.

4.3 Creating database for data mining:
Create database where data to be mined are stored for decision making. XML data base can be used to stored large data's. Data warehousing is an approach to store large volume of data for extracting pattern.

4.4 Creating the target data set:
Selecting the data set, examine on subset of variables and data samples of the database for determining the feasibility to solve the problem on which the discovery to be done.

4.5 Data Cleaning and Preprocessing:
Finding useful features of the data to find the desired goal. Using dimensionality reduction or transformation method to reduce the effective number of variables considered and to find the invariant representation of the data.

4.6 Selection for creating a data mining model:
Deciding whether the goal of decision support can be done based on classification, regression, clustering etc. Selecting appropriate variable to act as a predicator and new variable can be added to the existing variables to support the information.

4.7 Building a data mining model:
Creating multiple data mining models and to select a relevant model used for searching pattern from the data. Matching a particular data mining model with overall process.

4.8 Evaluating the data mining model:
To evaluate the parameter to find data accuracy in order to test the model. Errors occurred to be recorded and external validation needed to be performed to check whether selected model gives solution for health care system.

4.9 Deploy the data mining model:
Deploy the built and evaluated data mining model in external environment. A monitoring system should monitor the working of the model and generate a summarized report and analyze data in form of tables and graphs about its performance.
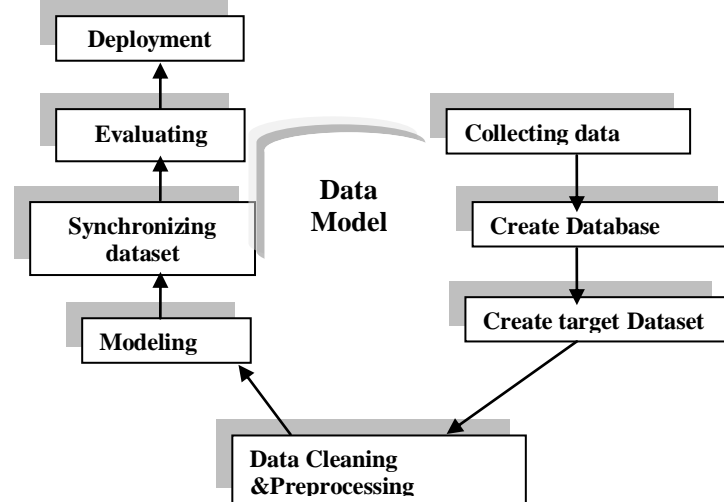
Fig1: Data model for Health care Data Management

V. CONCLUSION

This paper presents different data mining tools which are used in PHC. The proposed framework which works by defining the issue, collects data, create database and target data set, cleans, preprocesses, builds, evaluates and deploys a suitable data mining tool for the PHC. The proposed data model can be used to control the Prevalence of disease, recurrence, Predominance of major diseases in geriatrics, statistical distribution in pediatric undergoing vaccination. With the accurate data, necessary corrective decision making can be done to strengthen online disease surveillance systems for further prompt action/decision.

REFERENCES

Journal Papers:

[1]   Shaoyan Zhang, Tjortjis C., Zeng X., Qiao H., Buchan I. and Keane J., Comparing Data Mining Methods with Logistic Regression in Childhood Obesity Prediction, *Information Systems Frontiers Journal, Springer*, Vol. 11, No. 4, 2009, 449-460.

[2]   Michael Silver Taiki SakataHua.Steven B. Dolins, Michael J. O'Shea, *Journal of Healthcare Information Management,* 2001, vol. 15, no. 2.

[3]   Böhm C., Braunmüller B., Breunig M. M., Kriegel H.-P, High Performance Data Mining Using the Nearest Neighbor Join, *Int. Conf. on Information Knowledge Management (CIKM),* 2000.

[4]   Wilson A., Thabane L., Holbrook A, Application of data mining techniques in pharmacovigilance, *British Journal of Clinical Pharmacology,* (57) 2, 2003, 127-134.

[5]   Thangavel, K., Jaganathan, P.P. and Easmi, P.O, Data Mining Approach to Cervical Cancer Patients Analysis Using Clustering Technique, *Asian Journal of Information Technology (5) 4*, 413-417.

[6]   Rajan Vohra et al., Implementing a Decision Support Model in Primary Health Care*, International Journal of Information Technology and Knowledge Management*, Volume 2, No. 1, 2009, 163-166

[7]   P.Vishvapathia at. Al., A Case for using Grid Framework for Indian Rural Healthcare to Meet the Millennium Development Goals (MDGs), IJCCT, ISSN (ONLINE) : 2231–0371, ISSN (PRINT) : 0975–7449, Volume- 3, Issue-1.

[8]   http://www.who.int/healthmetrics/library/issue_1_05apr.doc