Data Mining Applications for Healthcare Sector Using a Weighted Support Vector Regression (WSVR) Techniques

R.Kavitha[#1] and S.BharathBabu[*2]

[#] *M.Phil Research Scholar, K.M.G College of Arts & Science, Gudiyatham.*

[*] *Asst.Prof., Department Of MCA, Sri Balaji Chockalingam Engineering College, Arni.*

Abstract— **Weighted Support Vector Regression (WSVR) which used weight factor on the basis of sensor reading providing continuous monitoring the patients to provide better health care services. In this method it as two regression types: Logistic Regression it is used to estimate of relative risk for various medical conditions. This is used to predict the risk factors such as Diabetes, Angina, Stroke etc. Here present a regression Decision tree algorithm is regression used in order to predict the number of number of hospitalization days in a population. Proposed algorithm used to claim the health insurance of patients. This algorithm improves the accuracy level of mining and avoid the noisy, irrelevant and massive too. The accuracy of data mining increased with the technique of WSVR and Decision tree algorithm. In technique to look at the changed information mining application in the human services segment for separating helpful data. The expectation of infections utilizing Data Mining applications is a testing errand yet it radically decreases the human exertion and expands the indicative exactness. The expectation of infections utilizing Data Mining applications is a testing under taking yet it radically lessens the human exertion and builds the symptomatic exactness.**

Index Terms— **WSVR, Logistic Regression, Decision tree algorithm.**

I. Introduction

Data mining applications in the scientific side. Scientific data mining distinguishes itself in the sense that the nature of the datasets is often very different from traditional market driven data mining applications. In this work, a detailed survey is carried out on data mining applications in the healthcare sector, types of data used and details of the information extracted. Data mining algorithms applied in healthcare industry play a significant role in prediction and diagnosis of the diseases. There are a large number of data mining applications are found in the medical related areas such as Medical device industry, Pharmaceutical Industry and Hospital Management. To find the useful and hidden knowledge from the database is the purpose behind the application of data mining. Popularly data mining called knowledge discovery from the data. This paper mainly compares the data mining tools deals with the health care problems. The comparative study compares the accuracy level predicted by data mining applications in healthcare.

Infertility is on the rise across the globe and it needs the sophisticated techniques and methodologies to predict the end results of infertility treatments particulars This comparative study could be useful for aspiring researchers in the field of data mining by knowing which data mining tool gives an accuracy level in extracting information from healthcare data.

II. THE PROCESS OF DATA MINING

The application of data mining tools in spreadsheet of the program that analyzes data to identify patterns and relations, user profiling and development of business strategies can be started . Most data mining software include online analytical processing, traditional statistical methods, such as cluster analysis, discriminate analysis and regression analysis, and non-traditional statistical analysis such as neural networks, decision trees, link analysis and association analysis. data mining process is inextricably linked to computers. With the help of special software, a big computer systems analyze data from different angles, find a hypothesis, experiment with them and learn from previous experiences. Unlike humans, which would let the obvious connection between the two data missed because it is beyond their expectations, such an error cannot happen to a computer. Also a human can be a victim of the conditionality with previous experience, which can be both positive and negative, but in any case impossible to avoid.

III. VARIOUS METHODS IN DATA MINING

Data mining consists of various methods. Different methods serve different purposes, each method has its advantages and disadvantages. Data mining tasks can be divided into descriptive and predictive. While descriptive tasks have a goal on finding a human interpreted forms and associations, after reviewing the data and the whole construction of the model, prediction tasks tend to predict an outcome of interest. Although the goals of description and prediction tasks may overlap, the main difference is that the predictive ones require that data include a special variable of response. The response can be categorical or numeric, further classifying data mining as classification and regression.

IV. PREDICTIVE AND DESCRIPTIVE DATA MINING TASKS

A. Classification and Regression

Identification of new templates with predefined objectives; these tasks are predictive and they include the creation of models to predict target, or dependent variable from the set of explained or independent variables. Classification is the process of finding a function that allows the classification of data in one of several classes. For classification tasks, the target variable usually has a small number of discrete values, while with the regression tasks, target variable is continuous.

B. Association rule

Association rule analysis is descriptive data mining task which includes determining patterns, or associations, between elements in data sets. Associations are represented in the form of rules, or implications.

C. Cluster analysis

Descriptive data mining task with the goal to group similar objects in the same cluster and different ones in the different clusters. Process of grouping determines groups of data that are similar, but different than other data. In this process variables are identified by which the best grouping is being realized.

D. Text mining

Most of the available data is in the form of unstructured or partially structured text, and it is different from conventional data that are completely structured. Text is unstructured if there is no previously determined format, or structure in data. Text is partially structured if there is a structure linked with the parts of data. While text mining tasks usually fall under classification, clustering and association rule data mining categories, it is the best to observe them separately, because unstructured text demands a specific consideration. In particular, method for representation of textual data is critical.

V. APPLICATION OF DATA MINING IN HEALTHCARE

There is a great potential for data mining application in healthcare. Healthcare institutions are very oriented on use of patient's information. Ability to use a data in databases in order to extract useful information for quality health care is a key of success of healthcare institutions. Healthcare information systems contain large volumes of information that include information on patients, data from laboratories that are constantly growing. With the use of data mining methods, useful patterns of information can be found in this data, that will later be used for further research and report evaluation. A very important issue is how to classify large volumes of data. Automatic classification is done based on the similarities that are present in data. This type of classification is useful only if the conclusion, that is drawn, is acceptable for the doctor or the end user. Data mining provides support for identification of reliable relations between treatment and outcome.

VI. DECISION TREE ALGORITHM

Is a graphical representation of the relations that exist between the data in the database. It is used for data classification. The result is displayed as a tree, hence the name of this technique. Decision trees are mainly used in the classification and prediction. It is a simple and a powerful way of representing knowledge. The models obtained from the decision tree are represented as a tree structure. The instances are classified by sorting them down the tree from the root node to some leaf node.

The nodes are branching based on if-then condition. Tree view is a clear and easy to understand, decision tree algorithms are significantly faster than neural networks and their learning is of shorter duration. Decision tree is a tree where each (non-terminal) node represents a test or decision on the item of information that is listed for consideration.

The choice of a particular industry depends on the outcome of the test. In order to classify the data, process is starting from the root node and following the argument down until it reaches the final node, at which time a decision is made. Decision tree can also be interpreted as a special form of a rule set, which is characterized by its hierarchical organization of rules.

VII. EXISTING SYSTEM

A. Application of rough set theory for medical informatics data analysis

The research work aims to analyze the medical data by applying Rough Set Theory of data mining approach. The data reduction process has been done using rough set theory reduction algorithm. Rough set is mainly used to reduce the attributes without compromising its knowledge of the original. To analyze the fertilization data, ROSETTA tool kit reduction algorithm is used in this work to produce the optimal reduct set without affecting the original knowledge

B. Artificial neural network in classification and prediction

The artificial neural network is constructed with multi-layer perception and back-propagation training algorithm, and constructed network is trained, tested and validated using patients 'This work finally sample IVF data compares the success rate between desired output which is field recorded data and actual output which is predicted output of neural network. In the comparison between desired and actual output of the neural network is illustrated.

VIII.   PROPOSED SYSTEM

A.   *Regression in Health Care*

The proposed Weighted Support Vector Regression (WSVR) which is used to weight factor on the basis of sensor reading for providing continuous monitoring to patients in order to provide them better healthcare services. In this research, on the basis of experimental results, it has been found the proposed approach had better accuracy than simple vector regression**.**

In this proposed work to look at the changed information mining application in the human services segment for separating helpful data. The expectation of infections utilizing Data Mining applications is a testing errand yet it radically decreases the human exertion and expands the indicative exactness. The expectation of infections utilizing Data Mining applications is a testing under taking yet it radically lessens the human exertion and builds the symptomatic exactness. In existing work data found are noisy, irrelevant and massive too. The accuracy level of data mining will be increased with this technique Weighted Support Vector Regression (WSVR) while comparing with existing work.
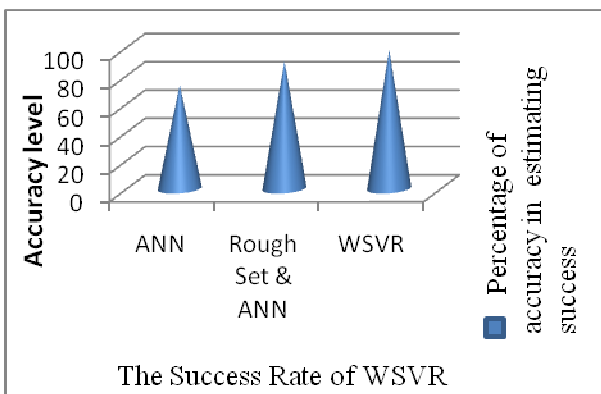

Fig 1. Chart for Comparison of ANN, Rough set and Weighted Support Vector Regression for the Accuracy level of data mining.

IX.   DATA MINING APPLICATIONS IN HEALTHCARE SECTOR

Healthcare industry today generates large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records, medical devices etc. Larger amounts of data are a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining applications in healthcare can be grouped as the evaluation into broad categories.

A.   *Treatment effectiveness*

Data mining applications can develop to evaluate the effectiveness of medical treatments. Data mining can deliver an analysis of which course of action proves effective by comparing and contrasting causes, symptoms, and courses of treatments.

B.   *Customer relationship management*

Customer relationship management is a core approach to managing interactions between commercial organizations-typically banks and retailers-and their customers, it is no less important in a healthcare context. Customer interactions may occur through offices, billing departments, inpatient settings, and ambulatory care settings.

C.   *Healthcare Management*

Data mining applications can be developed to better identity and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims to aid healthcare management. Data mining used to analyze massive volumes of data and statistics to search for patterns that might indicate an attack by bio-terrorists.

D.   *Hospital Management*

Organizations including modern hospitals are capable of generating and collecting a huge amount of data. Application of data mining to data stored in a hospital information system in which temporal behaviour of global hospital activities is visualized. Three layers of hospital management:
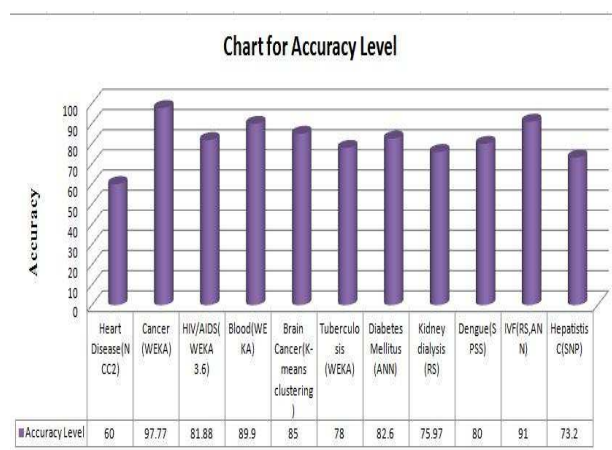
Services for hospital management
Services for medical staff
Services for patients

S.No	Type of disease	Data mining tool	Technique	Algorithm	Traditional Method	Accuracy level(%) from DM application
1	Heart Disease	ODND, NCC2	Classification	Naïve	Probability	60
2	Cancer	WEKA	Classification	Rules. Decision Table		97.77
3	HIV/AIDS	WEKA 3.6	Classification, Association Rule Mining	J48	Statistics	81.8
4	Blood Bank Sector	WEKA	Classification	J48		89.9
5	Brain Cancer	K-means Clustering	Clustering	MAFIA		85
6	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	Probability, Statistics	78
7	Diabetes Mellitus	ANN	Classification	C4.5 algorithm	Neural Network	82.6
8	Kidney dialysis	RST	Classification	Decision Making	Statistics	75.97
9	Dengue	SPSS Modeler		C5.0	Statistics	80
10	IVF	ANN, RST	Classification			91
11	Hepatitis C	SNP	Information Gain	Decision rule		73.20

TABLE 1.DATA MINING APPLICATIONS IN HEALTHCARE

	Rough Set	ANN	Rough Set & ANN (Hybrid)
Percentage of Accuracy in Estimating Success	47	73	90

TABLE 2. COMPARISON BETWEEN THREE DIFFERENT DATA MINING APPLICATIONS

Fig 2. Chart for Accuracy Level of using Data mining tools for diagnosis

X. RESULTS OF COMPARATIVE STUDY

A comparative study of data mining applications in healthcare sector by different researchers given in detail. Mainly data mining tools are used to predict the successful results from the data recorded on healthcare problems. Different data mining tools are used to predict the accuracy level in different healthcare problems. In this study, the following list of medical problems has been analyzed and evaluated.

Heart Disease
Cancer
HIV/AIDS
Blood
Brain Cancer
Tuberculosis
Diabetes Mellitus
Kidney dialysis
Dengue

XI. CONCLUSION

In this paper to look at the changed information mining application in the human services segment for separating helpful data. The expectation of infections utilizing Data Mining applications is a testing errand yet it radically decreases the human exertion and expands the indicative exactness. The expectation of infections utilizing Data Mining applications is a testing under taking yet it radically lessens the human exertion and builds the symptomatic exactness. In existing work data found are noisy, irrelevant and massive too. The accuracy level of data mining will be increased with this technique "Weighted Support Vector Regression" while comparing with existing work. By using the sensor readings we give the better treatment for patients in their healthcare center.

REFERENCES

[1] HianChyeKoh and Gerald Tan, Applications in Healthcare Information Management –Vol No 2.
[2] JayanthiRanjan, —Applicati techniques in pharmaceuticJournalof Theoretical and Applied Technology, (2007).
[3] RubanD.Canlas Jr., MSIT., Healthcare: Current application.
[4] EliasLemuye,tatusPredictive—Hiv ModelingS Using Data Mining Technology.
[5] Arvind Sharma and P.C. Gu Number of Blood Donors through their Age and Blood Group by using DaInternational Journal of Communication and Computer Technologies Volume 01 –No.6, Issue: 02 September 2012.
[6] Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence , 37-54.
[7] Kusiak, A., Kernstine, K., Kern, J., McLaughlin, K., & Tseng, T. (2000). Data Mining: Medical and Engineering Case Studies. Industrial Engineering Research 2000 Conference, (pp. 1-7). Cleveland, Ohio.
[8] David Page and Mark Craven,Biological — Applications of MultiRelationalData Mining
[9] D.Hand, H.Mannila and P.Smyth, "Principles of data mining", MIT,(2001).
[10] H.C. Koh and G Tan "Data mining Applications in Healthcare" Journal of Healthcare Information Management\, vol 19, no.2,(2005)