iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Health Prediction System by using Data Mining

Puja Sarage[1], Trupti Sudrik[2], Kalyani Zodage[3]
[1, 2, 3] *Computer Engineering Student, Imperial College of Engineering and Research, Pune, India.*

Abstract: This system supports an end user and online consultation. Here we propose a framework that enables clients to get moment direction on their medical problems through an astute social intelligent health care system online. The framework is bolstered with different symptoms and the disease or illness associated with those systems. Also the system allows user to share their symptoms and issues. Data Mining as a field of research has already well proven capabilities of identifying hidden patterns, analysis and knowledge applied on different research domains, now gaining popularity day by day among researchers and scientist towards generating novel and deep insights of these large biomedical datasets also. Uncovering new biomedical and healthcare related knowledge to support clinical decision making, is another dimension of data mining. Through massive literature survey, it is found that early disease prediction is the most demanded area of research in health care sector. As health care domain is bit wider domain and having different disease characteristics, different techniques have their own prediction efficiencies, which can be enhanced and changed in order to get into most optimize way.
Keywords: Data mining, Embedded system, Hill climbing, SVM, JAVA, MySQL, Eclipse.

I. INTRODUCTION

Information Mining is a non-unimportant extraction of verifiable, already obscure and potential valuable data about information. To put it plainly, it is a procedure of examining information from alternate point of view and assembling the learning from it. The found information can be utilized for various applications for instance social insurance industry. These days social insurance industry creates substantial measure of information about patients, sickness finding and so on. Information mining gives a set of strategies to find concealed examples from information. A significant test confronting Healthcare industry is nature of benefit. Nature of administration infers diagnosing sickness accurately and gives viable medications to patients. Poor determination can prompt unfortunate outcomes which are unsatisfactory. The analysis of infections is a key and perplexing employment in prescription. The acknowledgment of coronary illness from various highlights or signs is a multi-layered issue that isn't free from false suspicions and is every now and again joined by imprudent impacts. The medicinal services industry gathers immense measure of social insurance information which shockingly are "not mined" to find concealed data for successful basic leadership. Number test ought to be performed to determination the coronary illness and it frames different examinations with the information accessible and it is changed from patient to persistent for medicinal services associations, information mining methods are connected to foresee the Heart Ailment and helpful in planning learning base for encourage treatment. It was discovered that more than one out of three grown-ups are discovered ailing due to heart issues according to the reports of the World Health Organization (W.H.O.).

II. GOALS AND OBJECTIVES

A. Goals

To predict various diseases using Data mining.

B. Objectives

To predict diseases using Data mining applications is challenging task but it drastically reduces the human efforts and increases the diagnostic accuracy.
Go Green and reduce paper work.

III. MOTIVATION

Different data mining tool are used to predict the accuracy level in different health care problem, so we decided to implement system which will helpful to identify and predict the multiple diseases.

IV. PROBLEM STATEMENT

To design a Health Prediction System for medical data classification and early disease prediction by using SVM and Hill-climbing algorithm. It might have happened so many times that you or someone need doctor’s help immediately, but they are not available

©IJRASET (UGC Approved Journal): All Rights are Reserved

due to some reason. People cannot identify his symptoms and take medicines without consulting doctors. Some medicines are very much harmful to health. So user needs online consultation.

V. LITERATURE SURVEY

[1] The prediction of survival of Coronary heart disease (CHD) has been a challenging research problem for medical society. The goal of this paper is to develop data mining algorithms for predicting survival of CHD patients based on 1000 cases .We carry out a clinical observation and a 6-month follow up to include 1000 CHD cases. The survival information of each case is obtained via follow up. Based on the data, we employed three popular data mining algorithms to develop the prediction models using the 502 cases. We also used 10-fold cross-validation methods to measure the unbiased estimate of the three prediction models for performance comparison purposes. The results indicated that the SVM is the best predictor with 92.1 % accuracy on the holdout sample artificial neural networks came out to be the second with91.0% accuracy and the decision trees models came out to be the worst of the three with 89.6% accuracy. The comparative study of multiple prediction models for survival of CHD patients along with a 10-fold cross-validation provided us with an insight into the relative prediction ability of different data.

[2] Decision tree is one kind of inductive learning algorithms that offers an efficient and practical method for generalizing classification rules from previous concrete cases that already solved by domain experts. It is considered attractive for many real-life applications, mostly due to its interpretability. Recently, many researchers have been reported to endow decision trees with incremental learning ability, which is able to address the learning task with a stream of training instances. However, there are few literatures discussing the algorithms with incremental learning ability regarding the new attributes. In this paper, i^+Learning (Intelligent, Incremental and Interactive Learning) theory is proposed to complement the traditional incremental decision tree learning algorithms by concerning new available attributes in addition to the new incoming instances. The experimental results reveal that i^+Learning method offers the promise of making decision trees a more powerful, flexible, accurate and valuable paradigm, especially in medical data mining community.

[3] The data mining comprises of analysis of large data from various perspectives and obtaining summary of useful information. The information can be transferred into knowledge regarding future trends and history. Data mining has a very important role in the information technology domain. Huge amounts of complex data is generated by health care sector today. These data includes details about diseases, patients, diagnosis methods, electronic patients details hospitals resources etc,. The data mining methods are very helpful in making medicinal decisions in disease curing. The vast data collected by healthcare industry are not mined and hence information is hidden. And as a result the decision making is not effective. The knowledge discovered can be used by the healthcare administrators for enhancing the service quality. In this paper, a method for identifying frequency of diseases in particular geographical location for a given period of time using Apriori data mining technique based on association rules is proposed.[4] Medical data mining has been a popular data mining topic of late. Especially, diagnosing of the heart disease is one of the important issue and many researchers investigated to develop intelligent medical decision support systems to help the physicians. In this paper, we propose the use of decision tree C4.5 algorithm, bagging with decision tree C4.5 algorithm and bagging with Naïve Bayes algorithm to identify the heart disease of a patient and compare the effectiveness, correction rate among them. The data we study is collected from patients with coronary artery disease.

[5] Naive Bayes classifier has gained wide popularity as a probability-based classification method despite its assumption that attributes are conditionally mutually independent given the class label. This paper makes a study into discretization techniques to improve the classification accuracy of Naive Bayes with respect to medical datasets. Our experimental results suggest that on an average, with minimum description length (MDL) discretization the Naive Bayes classifier seems to be the best performer compared to popular variants of Naive Bayes as well as some popular non-Naive Bayes statistical classifiers.

[6] Clinical databases store large amounts of information about patients and their medical conditions. Data mining techniques can extract relationships and patterns holding in this wealth of data, and thus be helpful in understanding the progression of diseases and the efficacy of the associated therapies. A typical structure of medical data is a sequence of observations of clinical parameters taken at different time moments. In this kind of contexts, the temporal dimension of data is a fundamental variable that should be taken in account in the mining process and returned as part of the extracted knowledge. Therefore, the classical and well established framework of sequential pattern mining is not enough, because it only focuses on the sequentiality of events, without extracting the typical time elapsing between two particular events. Time-annotated sequences (IAS), is a novel mining paradigm that solves this problem. Recently defined in our laboratory together with an efficient algorithm for extracting them, IAS are sequential patterns where each transition between two events is annotated with a typical transition time that is found frequent in the data. In this paper we report a real-world medical case study, in which the IAS mining paradigm is applied to clinical data regarding a set of patients in

©IJRASET (UGC Approved Journal): All Rights are Reserved

the follow-up of a liver transplantation. The aim of the data analysis is that of assessing the effectiveness of the extracorporeal photo pheresis (ECP) as a therapy to prevent rejection in solid organ transplantation. For each patient, a set of biochemical variables is recorded at different time moments after the transplantation. The IAS patterns extracted show the values of interleukins and other clinical parameters at specific dates, from which it is possible for the physician to assess the effectiveness of the ECP therapy. We believe that this case study does not only show the interestingness of extracting IAS patterns in this particular context but, more ambitiously, it suggests a general methodology for clinical data mining, whenever the time dimension is an important variable of the problem in analysis.

VI. SYSTEM ARCHITECTURE

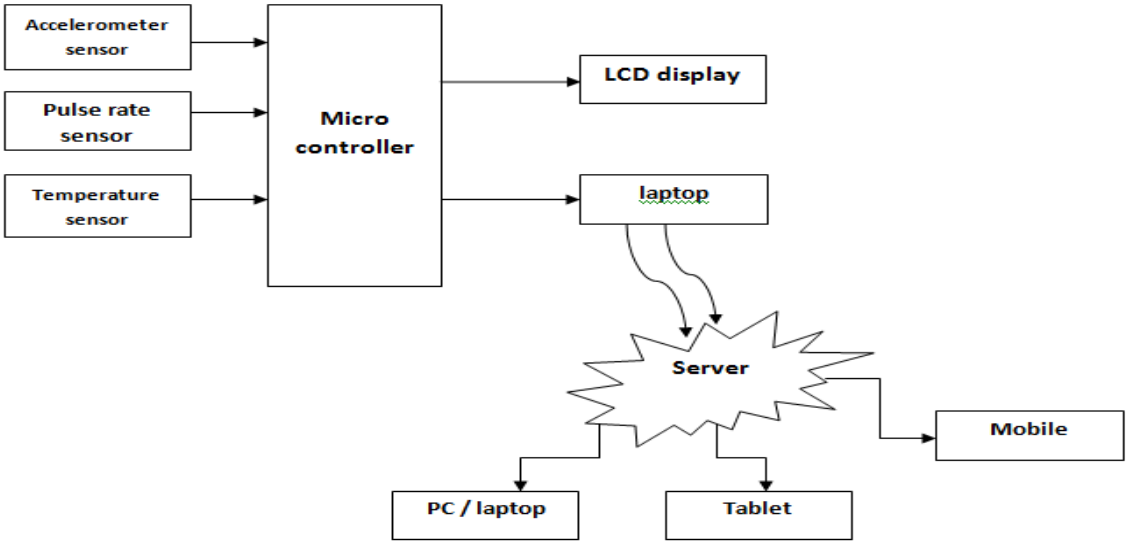


Fig.1. System Architecture

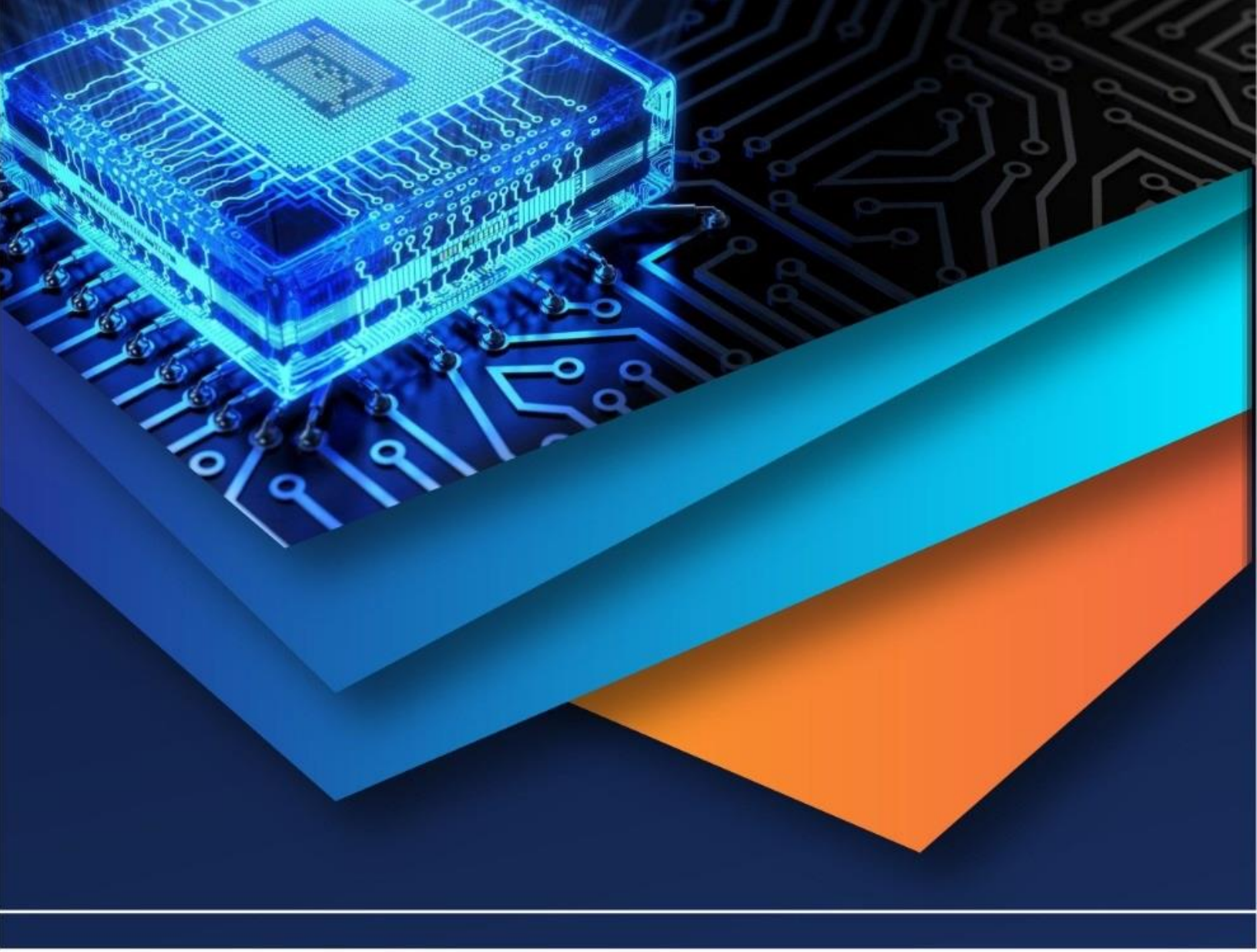
VII. CONCLUSION

We are going to design biomedical sensor based health prediction system which will help to older people, children, etc. Biomedical sensors will sense body parameters real time and server system will give predictions and recommendations to tackle with the conditions.

REFERENCES

[1] Yanwei Xing, Jie Wang and Zhihong Zhao, “Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease” In IEEE proc of International Conference on Convergence Information Technology (ICCIT 2007), pp. 868 – 872, ISBN: 0-7695-3038-9, 2007
[2] SAM CHAO, FAI WONG, “An Incremental Decision Tree Learninng Methodology Regarding Attributes in Medical Data Mining”, In IEEE proceedings of the Eighth International Conference on Machine Learning and Cyb ernetics, Baoding, 12-15 July 2009, pp. 12-15, vol. 3, ISBN: 978-1-4244-3702-3
[3] Gitanjali J. et al., “APRIORI algorithm based medical data mining for frequent disease identification”, IIJIT, Vol. 2, Issue 4 pp. 1-5, 2014, ISSN 2321-5976 2010.
[4] My Chau Tu AND Dongil Shin, “A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms” In IEEE proc of Eighth International Conference on Dependable, Autonomic and Secure Computing, pp. 183-187, ISBN: 978-1-4244-5421-1, 2009
[5] Ranjit Abraham, Jay B.Simha, Iyengar, “A comparative analysis of discretization methods for Medical Datamining with Naïve Bayesian classifier”, In proc of IEEE international conference on information Technology, pp. 235 - 236, 2006, ISBN: 0-7695-2635-7
[6] Michele Berlingeri, “Mining Clinical Data with a Temporal Dimension: a Case Study”, In proc. of IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), pp. 429 – 436, ISBN: 0- 7695-3031-1, 2007

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)