A NOVEL APPROACH FOR DISEASE PREDICTION SCHEME FOR HEALTHCARE MANAGEMENT

Miss. Madhuri Raosaheb Ghatage
Department of Computer Engineering.
JSPM NTC Pune
ghatagemadhuri@gmail.com

Dr. Nihar Ranjan
Department of Computer Engineering.
JSPM NTC Pune
nihar.pune@gmail.com

***Abstract*—Health examination has performed an essential position for preserving people's fitness since it may not best assist human beings apprehend their very own health conditions simply but also avoid lacking the first-rate timing of disorder treatment. However, in current health examination systems, people get only a basic report from single health examination and no advanced health risk analysis is provided. However, because of the shortage of experienced doctors and physicians, maximum healthcare groups cannot meet the medical demand of public. Due to that public need the Medical Treatment on-line with accuracy. Now a day's, public has no time to get the health practitioner physically, and then search the web hospital near approximately the current location. With the great use of hospital information system, there is big quantity of generated information which may be used to enhance healthcare service. Therefore, extra and additional realities mining bundles are created to give individuals more prominent hand crafted medicinal services administration. The main concept is to decide medicinal ailments as in step with given side results and every day routine given by the user, at the side of the potential of offering the closest doctors facility of that gift area. The framework provides a smooth to recognize interface to examinees and specialists. Examinees can realize the diverse side effects taking place in their body at the same time as the professionals can get a number of examinees with potential hazard. In the proposed paper we utilize diverse calculation to extend the safety of sensitive information of medical hospital the board consists of professionals, patients, etc.**

***Index Terms*—Disease prediction, Safety, Data Mining, Machine Learning, Medical Treatment, Hospital Management.**

I. INTRODUCTION

A. BACKGROUND

Knowledge discovery in databases is well-defined manner consisting of several distinct steps. Data mining is the center step, which results in the discovery of hidden but useful understanding from large databases. Data mining technology presents a user-oriented approach to novel and hidden patterns within the facts. The found information can be used by the healthcare administrators to improve the great of carrier. A major task going through healthcare groups is the supply of satisfactory offerings at low priced costs. Quality carrier implies diagnosing patients efficaciously and administering remedies which are effective.

Poor clinical decisions can cause disastrous consequences which might be consequently unacceptable. Hospitals need to also decrease the fee of clinical tests. They can attain these consequences by way of employing appropriate computer-based records and/or selection help structures. Health care information is huge. It consists of affected person essential information, resource management facts and transformed data. Health care groups should have ability to research statistics. Treatment statistics of thousands and thousands of sufferers can be stored and automated and information mining techniques may help in answering numerous critical and essential questions associated with fitness care.

In this context, digital healthcare systems (EHRs) worker such policies and therefore have been classified as protection crucial systems. These structures are differentiated in one vital element to other systems, the balancing between confidentiality and availability. The anxiety between these goals is clear: while all of the patient's statistics should be available to be shared and monitored to supply expert healthcare offerings; for security reasons, part of the information can be taken into consideration private and have to now not be accessible. To this end, on this paper, we behavior an in-intensity survey study to analyze the healthcare machine's safety and privacy threats. Then, we propose a unique safety model that captures the situation of records interoperability and helps the essential security of Electronic Health Record together with the capability of providing fine-grained access manage.

Disease Predictor is a web based totally software that predicts the disorder of the consumer with appreciate to the symptoms given through the consumer. Disease Prediction system has statistics sets collected from exceptional health associated sites. With the help of Disease Predictor the person will be capable of recognize the chance of the ailment with the given symptoms

As the use of internet is growing each day, people are usually curious to recognise distinct new things. Individuals consistently attempt to allude to the web if any issue emerges. Individuals approach web than clinics and specialists.

Individuals don't have quick choice when they endure with specific sickness. Thus, this framework can be useful to the individuals as they approach web 24 hours.

B. MOTIVATION

The Motivation behind that is to deal with a huge quantity of various disease information and on that the risk prediction of disease can be examined. With the great use of hospital statistics system, there is a large amount of generated statistics which can be used to enhance fitness care services, therefore developing facts mining applications to offer human beings more custom designed health care service.

C. OBJECTIVES

1. To determine the scientific diseases according to given symptoms and daily routine of the person and when he searches the hospital, the health center that's nearest to their present day area is given as result.
2. To expect the ailment based on laboratory outcomes and diagnostic facts thinking about the phrase count.

II. REVIEW OF LITERATURE

In this paper, the author has provided an sensible and powerful heart attack prediction techniques using records mining. Firstly, it provided an green method for the extraction of extensive patterns from the heart disorder statistics warehouses for the efficient prediction of coronary heart assault Based on the calculated full-size weight age, the frequent styles having cost more than a predefined threshold were chosen for the valuable prediction of coronary heart attack. In this paper the drawbacks are for predicting coronary heart assault significantly 15 attributes are listed. Besides the 15 listed in clinical literature we also can include other facts mining strategies, e.g., Time Series, Clustering and Association Rules. [1]

In this paper, writer provided a middleware answer technique to aid facts and community protection over e-Healthcare gadget sing medical sensor networks. It has been shown that a masquerade assault may be released to the device and patients 'facts are in danger. We proposed this middleware to counter this sort of attack in which a user and all devices into the healthcare community are mutual authenticated. Finally a overall performance evaluation has been achieved with regard to masquerade attack and the end result exhibits the efficient of the proposed solution. [2]

In this paper, creator design an inference assault-resistant e-healthcare cloud system with fine-grained get admission to control. We first suggest a two-layer encryption scheme. To make certain an efficient and fine-grained get admission to control over the EHR records, we layout the first-layer encryption, in which we devise a specialized access coverage for each information attribute inside the EHR, and encrypt them personally with high efficiency. To keep the get admission to pattern of statistics attributes in the EHR, we further assemble a blind records retrieving protocol. We also exhibit that our scheme can be easily prolonged to help seek functionality. Finally, we conduct tremendous security analyses and overall performance evaluations, which verify the efficacy and performance of our schemes. [3]

In this paper the writer proposes a semantic-based totally steady discovery framework for cellular healthcare organization networks that exploits semantic metadata (profiles and policies) to permit bendy and stable carrier search/retrieval. As a key feature, this method integrates get admission to manage functionalities within the discovery framework to offer users with filtered views on available services based totally on service get right of entry to requirements and user security credentials. Identification of answers to these challenges is important if clinical choice assist is to gain its capability and improve the quality, safety and performance of healthcare. [4]

In this paper the author proposed a technique that, given a question submitted to a seek engine, shows a listing of related queries. The related queries are based in formerly issued queries, and may be issued by means of the person to the hunt engine to song or redirect the hunt procedure. The approach proposed is based on a question clustering procedure in which groups of semantically similar queries are identified. The clustering manner uses the content material of historical possibilities of person's registered inside the query log of the hunt engine. The approach not only discovers the associated queries, but also ranks them in line with a relevance criterion. Finally, we display with experiments over the question log of a search engine the effectiveness of the method. [5]

The numerous heart ailment prediction strategies are discussed and analyzed on this paper. The facts mining techniques used to predict coronary heart diseases are mentioned here. Heart ailment is a mortal disease by means of its nature. This ailment makes several problems together with heart assault and death. In the clinical domain, the importance of records mining is perceived. From the comparative look at we are able to conclude that Support Vector Machine (SVM) technique is an green approach for predicting heart disease. [6]

In this paper, the author proposed Lightweight Sharable and Traceable, a lightweight secure records sharing solution with traceability for mHealth systems. Lightweight Sharable and Traceable seamlessly integrates some of key protection functionalities, which includes fine-grained get admission to manage of encrypted statistics, keyword seek over encrypted statistics, traitor tracing, and consumer revocation into a coherent machine layout. Considering that cellular gadgets in mHealth are aid constrained, operations in data owners' and data customers' devices in Lightweight Sharable and Traceable are kept at lightweight and provide security. Further, great experiments on its performance (on both PC and cell device) confirmed that Lightweight Sharable and Traceable is very promising for sensible applications. [7]

In this paper, the writer proposes a Highly Available, Scalable and Secure distributed information storage device for excessive overall performance and stable information management. Distributed and parallel information garage or record structures along with Object-based Storage Devices and flexible key distribution schemes Data at rest (static) and in transit (dynamic) are protected with one-of-a-kind encryption strategies for privacy and integrity. Secret sharing and replication help both safety and availability. Encryption and key management aren't essential in records at rest protection. The future work includes an in depth simulation and in addition overall performance analysis. [8]

In this the writer proposed a safety scheme for customers. This scheme provides storing and sharing their intricate facts inside the Cloud environment. This scheme gives important encryption and decryption method for attaining safety on cloud application. The revocation process is an express overall performance destroyer within the get entry to control approach in cryptography. In this scheme, the unique facts is firstly separated into severa elements. Then these parts are despatched to the cloud server. Whenever a consumer revocation happens, the information owner dreams simply to retrieve one part and re-encrypt it. This scheme is based on cryptographic storage application. Furthermore techniques are implemented to enhance the security of the records. [9]

In this paper, the writer have proposed a covered multiple owner statistics sharing machine. This device is used for dynamic companies within the cloud environment. Any Cloud user can unidentified person distribute data with other users so one can enhance the signature of institution and dynamic broadcast encryption techniques. For this, the storage transparency and encryption calculation fee are self governing with admire to the quantity of customers which are revoked. Furthermore, the protection and investigation machine with actual proofs is analyzed. [10]

III. PROPOSED METHODOLOGY

The proposed gadget we build which leverages data mining methods to reveal the relationship among the regular physical examination information and the ability health hazard given by using the consumer or public. This gadget uses the Machine studying and Data Mining algorithms like Na¨ıve Bayes, Support Vector Machine are used for the sickness prediction and for the garage of the data the gadget used MYSQL database. The system offers a consumer-friendly interface for diverse customers and doctors.

In this paper, an efficient and privacy-preserving disease chance prediction scheme for e-healthcare is proposed. In the existing paper there may be a downside of safety related to patient's records. So within the proposed work we're going to use encryption technique to provide safety to the sensitive information of the patients. Compared with the prevailing work we're going to use Na¨ıve Bayes algorithm to go looking the records, and used encryption algorithm to provide protection and SVM algorithm to predict the diseases.

Advantages of Proposed System:

1. Increases human-computer interactions.
2. Recommends the hospital and doctor to patient in keeping with illnesses predicted.
3. Location of User is detected.
4. Shortest distance is provided.

A. Architecture

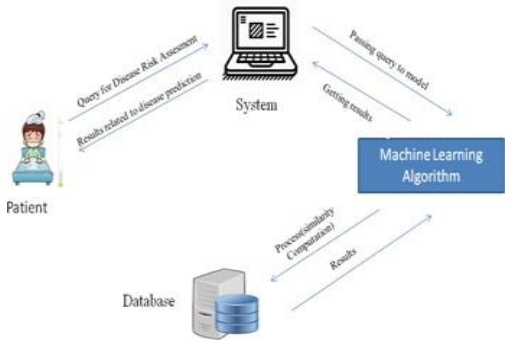

Fig. 1. Proposed System Architecture

Explanation:

In this system, first we are login to the system by using OTP. After login to the system we are going to search the hospital or doctor. Then enter the symptoms or answer the questions that is generated by the system. Then processing of the algorithms will be done and, at last, the predicted disease will be get as the final result from the system.

In this different modules are used:

Module 1:

Patient: Patient will login to the system and then enter the symptoms and get the predicted result according to its ailment.

Module 2:

Doctor: Doctor will first register and login to the system. After that doctor will view the patients symptoms and according to its ailment will give the prediction on his health that which disease the patient have.

Module 3:

Admin: Admin will manage the database of hospital as well as patients record.

B. Algorithms

1. SVM Algorithm:

Support Vector Machine (SVM) is the one of the

unsupervised learning technique.

Support Vector Machine (SVM) is used to classify the disease. SVM Support vector machines are mainly two class classifiers, linear or non-linear class boundaries.

The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.

The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class.

Steps:

Step 1: Read the test symptoms and trained symptoms.

Step 2: Check the all test symptoms of patients and also get all train symptoms.

Step 3: Consider the kernel.

Step 4: Train the SVM using both symptoms and show the output.i.e.disease.

Step 5: Classify an observation using a Trained SVM Classifier.

2. Naive Bayes Algorithm:

Naive Bayes is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms. Naive Bayes classifier is the fast, accurate and reliable algorithm. Naive Bayes classifiers have high accuracy and speed on large datasets.

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.

The Naive Bayes algorithm is called naive because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

The Naive Bayesian classifier is based on Bayes theorem with the independence guess between predictors.

A Naive Bayesian model is easy to form, with no critical iterative parameter computation which makes it particularly useful for very large data sets.

Regardless of its simplicity, the Naive Bayesian classifier often does particularly well and is widely used because it often outperforms more experienced classification methods.

Naive Bayes classifier calculates the probability of an event in the following steps:

Step 1: Calculate the prior probability for given class labels

Step 2: Find Likelihood probability with each attribute for each class

Step 3: Put these value in Bayes Formula and calculate posterior probability.

Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

C. Mathematical Model

1. Mathematical equation:

We have k sub-spaces so that there are k classification results of sub-space, called CL_SS1,CL_SS2, ..., CL_SSk. Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$\overline{CL} = \frac{1}{k}\sum_{i=1}^{k} CL\_SS_i$$

Or weighted mean value:

$$CL = \frac{1}{k}\sum_{i=1}^{k} W_i \overline{CL\_SS_i}$$

Where W_i is the weight of classification result of subspace , SS_i and satisfies:

$$\sum_{i=1}^{R} W_i = 1$$

The centroid is calculated as follows:

$$\bar{X} = \frac{\sum_{i=0} X_i}{k}, \bar{Y} = \frac{\sum_{i=0} Y_i}{k}$$

Where $(\bar{X}, \bar{Y})$ represents the centroid of the hand, X_i and Y_i are x and y coordinates of the i^{th} pixel in the hand region and k denotes the number of pixels that represent only the hand portion.

In the next step, the distance between the centroid and the finger tip was calculated. For distance, the following Euclidean distance was used:

$$\text{Distance} = \sqrt{(x2 - x1)^2 (y2 - y1)^2}$$

Where (x1, x2) and (y1, y2) represent the two co-ordinate values.

IV. SYSTEM SPECIFICATION

A. Software Requirements used

- Operating System - Windows
- Application Server - Apache Tomcat 7
- Front End - HTML, JSP, CSS
- Database - My SQL 5.0
- IDE - Eclipse Oxygen
- Coding Language - Java 1.8

B. Hardware Requirements used

- Processor - Dual core/Intel i3
- Speed - 1.8 GHz
- RAM -2 GB (Min)
- Hard Disk - 100 GB

V. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i5-6700HQ CPU @ 2.60GHz, 16GB memory, Windows 7, MySQl Server 5.1 and Jdk 1.8.

In our system, we divide the data sets into training set and testing set by the ratio 2 : 1. Then our system applies the two algorithms mentioned above in risk-prediction task of three symptoms. Few questions related to the symptoms are asked by the system and the user has to give the required

answer. In our system User search with the keyword and gets result according to the current location.

A. Results and Performance

Classification between Algorithms:

S.No	Classification Methods	No. of Samples	Predicted Samples	Rate of prediction
01	Proposed Classification	23	22	95%
02	Existing Clas-sification	25	22	92%

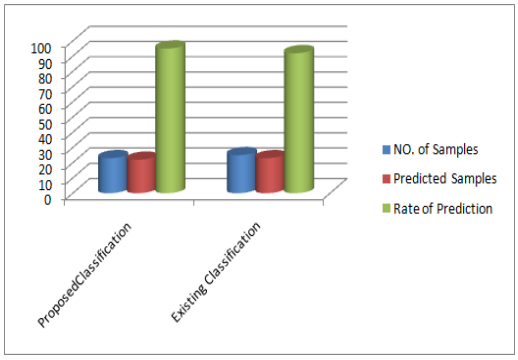

Fig. 2. Classification graph

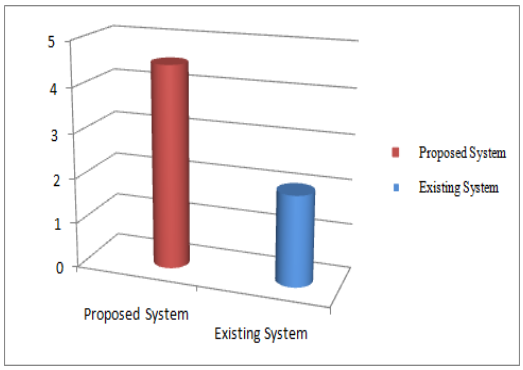

Fig. 3. Algorithms Comparison

VI. CONCLUSION

In this paper, we proposed a privacy-retaining disorder predicting system which can help physicians make a right prognosis of sickness and offer health services for sufferers anytime anywhere in a privacy-maintaining way. This task implements a sickness danger prediction scheme which leverages facts about the data mining strategies to reveal the connection between the regular physical examination statistics and the fitness danger given via the person or public. Different system getting to know algorithms are implemented to predict physical condition examine that will be in danger of physical deterioration in the subsequent year. In our technique user or patient search the health facility and the consequences given are in line with the nearest vicinity of contemporary area of consumer/patient. User / Patient give signs and symptoms and the system predicts the disease and presents the medicines. The advantages of privacy-maintaining prognosis are to maintain the balance among safety and performance which must be taken into consideration firstly. Therefore, how to optimize the model education using set for performance improvement and finding an effective manner of introducing some different advanced system gaining knowledge of strategies to build the privacy-retaining disease prediction system are worth of investigation.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and for their guidance. We are thankful to the authorities of University of Pune and concern members of CPGCON 2020 conference, organized by, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] Srinivas K, Rani B K, Govrdhan A. "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks". International Journal on Computer Science Engineering, 2010.
[2] Alessandra Toninelli, Rebecca Montanari, And Antonio Corradi "Enabling secure service discovery in mobile healthcare enterprise networks", IEEE Wireless Communications Volume: 16, Issue: 3 , June 2009.
[3] Ndibanje Bruce, Mangal Sain, Hoon Jae Lee, "A Support Middleware Solution for e-Healthcare System Security", IEEE 16th International Conference on Advanced Communication Technology.
[4] Wei Zhang, Yaping Lin, Jie Wu, Fellow and Ting Zhou "Inference Attack-Resistant E-Healthcare Cloud System with Fine-Grained Access Control", IEEE Transactions on Services Computing 2018.
[5] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query recommendation using query logs in search engines," in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588–596.
[6] Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S, "A Survey on Predicting Heart Disease using Data Mining Techniques", 2018 Conference on Emerging Devices and Smart Systems (ICEDSS).
[7] Yang Yang, Ximeng Liu, Robert H. Deng, Yingjiu Li, "Lightweight Sharable and Traceable Secure Mobile Health System", IEEE Transactions on Dependable and Secure Computing, 2017.
[8] Zhiqian Xu, Hai Jiang, "HASS: Highly Available, Scalable and Secure Distributed Data Storage Systems", 2009 International Conference on Computational Science and Engineering.
[9] Kamara, S., Lauter, K. Sion, R., Curtmola, R., Dietrich, "Cryptographic Cloud Storage", 2010 Workshops of LNCS Springer, Heidelberg, vol. 6054, pp. 136-149, 2010.
[10] Bethencourt, J., Sahai, A.,Waters, B., "Ciphertext policy attribute-based encryption", 28th IEEE Symposium on Security and Privacy, pp. 321-334, 2007.