IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105200 |

Disease Prediction Exercise and Medicine Recommendation Using Machine Learning

Pradip Ugale[1], Pratik Ingale[2], Ganesh Vairagar[3], Sunny Shirsath[4]

UG Student, Dept. of I.T., PREC, SPPU University, Maharashtra, India

Assistant Professor, Dept. of I.T., PREC, SPPU University, Maharashtra, India

ABSTRACT: Health and drug is gaining a lot of significance in moment's advancing world, where evolving technology is being used to combat nearly all the known conditions. These days application of Big Data is expanding in biomedical and mortal services groups, exact disquisition of medicinal information benefits beforehand malady discovery, quiet care and group administrations. fractured remedial information lessens examination perfection. The machine learning computations are proposed for successful anticipation of ceaseless infection. To beat the trouble of deficient information. The dataset comprise of structured data and unshaped data. To prize features from unshaped data algorithm will be employed. Framework proposes SVM computation and Naive Bayesian computation for sickness anticipation exercising unshaped and structured information collectively from sanitarium information. On that we're working on tradition grounded on attained result. system is also proposed which will prevision the inquiry and answers and will give proper responses to the guests. For that, two computations are proposed KNN and SVM. KNN algorithm will perform bracket on answers and SVM computation will perform bracket on answers. It'll help customer to discover stylish inquiries and answers linked with infections. Technologies similar as data mining and recommender technologies give possibilities to explore implicit knowledge from opinion history records and help croakers to diagnose the medical complaint and define drugs rightly to drop drug error effectively. This design proposes a system which takes as input the symptoms of the case to prognosticate the complaint, which is followed by recommending the applicable drug. This system consists of a database system module, data medication module, complaint vaticination module, drug recommendation module, model evaluation and data visualization module..

KEYWORDS: disease prediction, data analytics; Machine Learning; Healthcare, medicine recommendation

I. INTRODUCTION

With the enhancement of living norms, the prevalence of habitual complaint is adding . It's essential to perform threat assessments for habitual conditions. With the growth in medical data, collecting electronic health records(EHR) is decreasingly accessible. Proposed a healthcare system using smart apparel for sustainable health monitoring had completely studied the miscellaneous systems and achieved the stylish results for cost minimization on tree and simple path cases for miscellaneous systems. Cases ' statistical information, test results and complaint history are recorded in the EHR, enabling us to identify implicit data- centric results to reduce the costs of medical case studies. Proposed an effective inflow estimating algorithm for the health pall system and designed a data consonance protocol for the PHR(Personal Health Record)- grounded distributed system. pall system and designed a data consonance protocol for the PHR(Personal Health Record) grounded distributed system. Proposed six operations of big data in the field of healthcare but these schemes have characteristics and blights also.

The data set is generally small, for cases and conditions with specific conditions, the characteristics are named through experience. still, thesepre-selected characteristics perhaps not satisfy the changes in the complaint and its impacting factors. With the development of big data analytics technology, further attention has been paid to complaint vaticination from the perspective of big data analysis, colorful inquiries have been conducted by opting the characteristics automatically from a large number of data to ameliorate the delicacy of threat bracket, rather than the preliminarily named characteristics. still, those being work substantially considered structured data. For unshaped data, for illustration, using Convolutional neural network to prize textbook characteristics automatically has formerly attracted wide attention and also achieved veritably good results. likewise, there's a large difference between conditions in different regions, primarily because of the different climate and living habits in the region. therefore, threat bracket grounded on big data analysis, the following challenges remain How should the missing data should be addressed? How should the main habitual conditions in a certain region and the main characteristics of the complaint in the region be determined? How can big data analysis technology be used to dissect the complaint and produce a better model? To break these problems, proposed System combines the structured and unshaped data in healthcare field to assess the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105200 |

threat of complaint. First, Weak machine learning tool will 0% Plagiarized 100% Unique Characters:3100 Words:445 Sentences:21 Speak Time: 4 Min Page 1 of 2 be to reconstruct the missing data from the medical records collected from a sanitarium. To handle structured data, I consult with sanitarium experts to prize useful features. For unshaped textbook data, System selects the features automatically using textbook point birth system, that folds together Unicode conversion, forced lowercasing, word boundary discovery, and string hash calculation.

II. METHODOLOGY

The thing of this paper is to prognosticate whether a case is amongst the cerebral infarction high- threat population according to their medical history. further formally, we regard the threat vaticination model for cerebral infarction as the supervised literacy styles of machine literacy, i.e., the input value is the trait value of the case, which includes the case's particular information similar as age, gender, the frequence of symptoms, and living habits(smoking or not) and other structured data and unshaped data. The affair value indicates whether the case is amongst the cerebral infarction high- threat population. According to the different characteristics of the case and the discussion with croakers , we will concentrate on the following three datasets to reach a conclusion. Structured data(S- data) use the case's structured data to prognosticate whether the case is at high- threat of cerebral infarction. Text data(T- data) use the case's unshaped textbook data to prognosticate whether the case is at high- threat of cerebral infarction. Structured and textbook data(S&T- data) use the S- data and T- data above tomulti- dimensionally fuse the structured data and unshaped textbook data to prognosticate whether the case is at high- threat of cerebral infarction. In this paper, for S- data, according to the discussion with croakers and Pearson's correlation analysis, we prize the case's demographics characteristics and some of the characteristics associated with cerebral infarction and living habits(similar as smoking). We'll introduce machine literacy and deep literacy algorithms used in this work. For S- data, we use three conventional machine learning algorithms, This is because these three machine literacy styles are extensively used. For T- data, we propose CNN- grounded unmoral complaint threat vaticination(CNN- UDRP) algorithm to prognosticate the threat of cerebral infarction complaint. For complaint threat modeling, the delicacy of threat vaticination depends on the diversity point of the sanitarium data, i.e., the better is the point description of the complaint, the advanced the delicacy will be. For some simple complaint,e.g., hyperactive- lipidemia, only a many features of structured data can get a good description of the complaint, performing in fairly good effect of complaint threat vaticination. But for a complex complaint, similar as cerebral infarction mentioned in the paper, only using features of structured data isn't a good way to describe the complaint. thus, in this paper, we work not only the structured data but also the textbook data of cases grounded on the proposed CNN- MDPR algorithm. We find that by combining these two data.

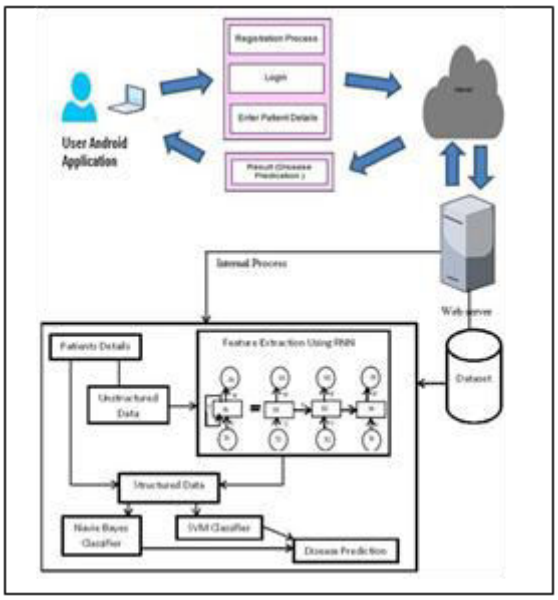

Fig: Architecture Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105200 |

III. PROPOSED ALGORITHM

1 Machine Learning :

Machine Learning is such a _eld which gives an ability to learn without being explicitly programmed. They mainly focus on the prediction. Statistics and Machine Learning are closely related fields. They can be divided into three categories: 1) Supervised Learning 2) Unsupervised Learning 3) Reinforcement Learning

2 Healthcare:

Healthcare is preserving or improving the health through prevention, diagnosis and treatment of that particular disease. Healthcare contributes beyond the delivery of services to the patients. It contributes to the part of country economy. It is mainly regarded as to determine in promoting the physical and mental health around the world.

3 Naive Bayesian :

The Naive Bayes is a classification technique based on the Bayes Theorem. They are easy to build and useful for large data set. It is even a highly sophisticated classification method. It is used to predict the multi-class prediction. It performs well for categorical input variables. Naive Bayes is a simple technique for constructing classifier: models that assign class label to problem instances, represented as vector of feature values, where the class labels are drawn from some finite set. The application is the real time prediction, text classification, multiclass prediction.

4 K Nearest Neighbor :

The k-Nearest Neighbor stores all the cases and classifies new class based on the similarity measures. The output of k-NN is the class membership. The object is been classified on the basis of majority votes of its neighbors. The values of output are by averaging the values of its k nearest neighbor. It is a special case of variable bandwidth. They are used for both classification and regression. The neighbors are taken from the set of objects for which the class or the object property value is known. It is sensitive to the local structure of the data.

IV. SIMULATION RESULTS

Software testing is really an arrangement of different errands whose main role is to completely practice the computer based system. Although each test has a different reason, all work to confirm that the framework components have been appropriately incorporated and perform assigned errands. Strategy A test case has parts that depict an input, action or event and an expected response, to decide whether an element of an application is working accurately. There are levels in which each experiment will fall so as to stay away from duplication effort.

Data mining supports many different techniques for knowledge discovery and prediction such as classification, clustering, sequential pattern mining, association rule mining and analysis. Data mining is extensively used in business analysis, strategic decision making, financial forecasting, future sales prediction etc. machine learning algorithms are proposed for effective prediction of chronic disease.

To extract feature from unstructured data RNN algorithm will be used. Here , user will upload the test file i.e previous health record . RNN algorithm extract the features from that file .and pass that features to the Naïve SVM algorithm for disease predication .System proposes Naive Bayesian algorithm to predict the disease using structured data. System allows user to select the symptoms . System passes that symptoms to the Navie Bayes algorithm to perform disease prediction.

Community question answering system (CQA) is also proposed in this paper. it predicts the question and answers and provides appropriate answers to the users. For that two algorithms are proposed KNN and SVM . KNN algorithm performs feature extraction and classification on questions and SVM algorithm performs classification on answers. It will help user to find best questions and answers related to the chronic diseases.

1. Dataset

Dataset is collected from the Hospital . Dataset contains 4 classes(Asthma, Hypertension, Cerebral Infarction, Diabetes) and 28 attributes (Shortness of breath, Coughing, Wheezing, Chest tightness, Stress, Chest pain, Blood in the urine, Irregular heartbeat, limb numbness, dizziness, aphasia and clumsiness, headache, increased thirst, hypertension, hunger,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105200 |

frequent urination , Fever, Very rapid breathing, sighing, Pale, sweaty face, difficulty concentrating, tiredness, runny nose, Trouble sleeping, nasal congestion, Chest pain and pressure, aphasia and clumsiness)

2. Methodology of Evaluation

The designed model is trained on the training data and evaluated on the test set. The performance of the trained model is evaluated as the Mean Absolute Error (MAE) between the predicted AES and the expert's manual echo scores (MES).Proposed System uses Naïve Bayes and SVM classifier to predict the disease. The number of instances correctly predicted as required following table shows the Evaluation of the system.

Correctly Classified Instances 1026 97 %

Incorrectly Classified Instances 0 3 %

Mean absolute error 0.25

Root mean squared error 0.3118

Relative absolute error 66.6668 %

Root relative squared error 72.0083 %

Total Number of Instances 1026

KNN vs. SVM:

While searching question with KNN and SVM.

Sr.no	KNN	SVM
1	100 Question	21 Question
2	362 Question	36 Question

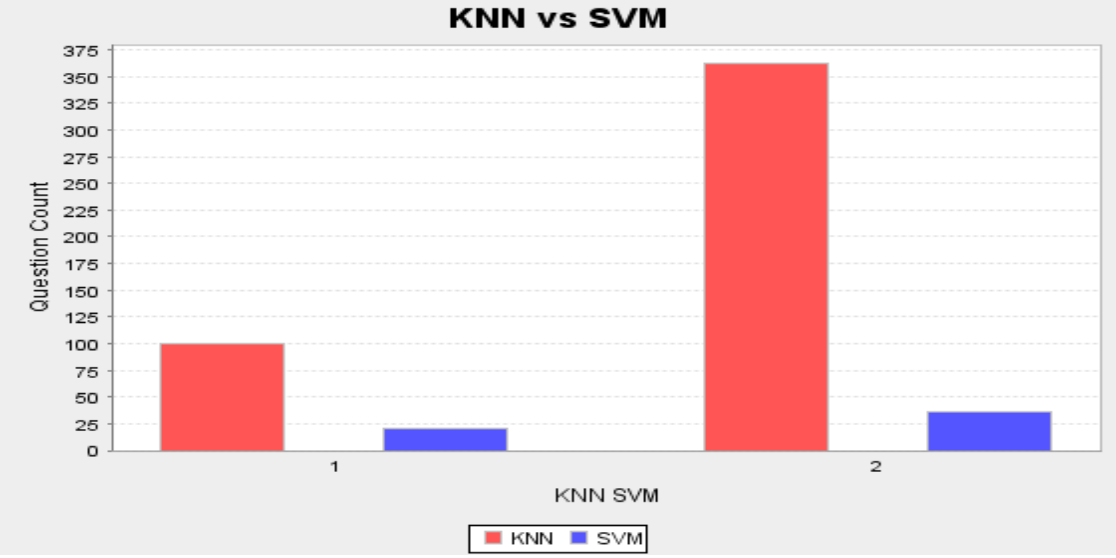

Fig. knn vs svm

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105200 |

V. CONCLUSION AND FUTURE WORK

As chronic disease has increased, a new conventional neural network based multimodal disease risk prediction (CNNMDRP) algorithm in which structured and unstructured data from hospital is being used. In this structured and unstructured data, the personal information and detail history of the patient is being stored. In this CNN-MDRP both data are being used for predicting the chronic disease in that particular patient. In unstructured data patients may have missing data. So, the missing data of that particular patient can also retrieve through the genetic algorithm. The featured from unstructured dataare been extracted correctly. Then the extracted features are structured data. Both Structured data and extracted structured data are used for predicting the exact chronic disease with Naive Bayes classifier and the SVM classifier. Community question answering system (CQA) is also proposed to help user to post the questions and answers related to the disease. To propose CQA system KNN and SVM algorithms are used.

Making correction in big data of hospital in chronic disease.Limitation of observation and contents to the public regarding chronic disease from government department.Regarding, solution to the chronic disease is to be accordingly to the state or region or group of people with common symptoms. The database is growing tremendously via use of the Internet. Thus, our future work would be to extend the k-anonymity and the l-diversity model using the BFO algorithm to the dynamic and distributed database.

REFERENCES

[1].Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K.Mago,2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map",IEEE International Conference on Fuzzy Systems (FUZZ),pp. 1377-1382.

[2].Ashish Chhabbi, LakhanAhuja, SahilAhir, and Y. K. Sharma,19 March 2016,"Heart Disease Using Data Mining Techniques", International Journal of Research in Advent Technology,E-ISSN:2321-9637,Special Issue National Conference "NCPC-2016", pp. 104-106.

[3].Liu Y, Teverovskiy L, Lopez O, Aizenstein H, Meltzer C,Becker J(2007) Discovery of biomarkers for alzheimer's disease prediction from structural mr images. In: 2007 IEEE international symposium on biomedical imaging, April.

[4] Sathyabama Balasubramanian, Balaji Subramani. SYMPTOM'S BASED DISEASES PREDICTION IN MEDICAL SYSTEM BY USING K-MEANS ALGORITHM, International Journal of Advances in Computer Science and Technology, Vol.3, No.2, February 2014.

[5]. Rahul Isola, Rebeck Carvalho and Amiya Kumar Tripathy.Knowledge discovery in Medical system by using Differential Diagnosis, AMSTAR and K-NN, IEEE Transaction on Information Technology in Biomedicine, Vol.16, No.6, November 2012.

[6]. S.Fox and M. Duggan. Health online 2013.Pew Internet and American Life Project. http://pewinternet.org/Reports/2013/Health- online.aspx,2013.

[7].X. Guo, J. Lu, Intelligent e-government services with personalized recommendation techniques, International Journal of Intelligence Systems, 2007,401–417

[8].T. Lee, J. Chun, J. Shim, S.-G. Lee, An ontology-based product recommender system for B2B marketplaces, International Journal of Electronic Commerce, 2006,125–155.

[9].J.B. Schafer, J. Konstan, J. Riedl, E-commerce recommendation applications, Applications of Data Mining to Electronic Commerce ,Springer, US 2001, 115–153.

[10].T.Hung-Wen,S.VonWun,A personalized restaurant recommender agent for mobile eservice, 2004 IEEE International Conference on e- Technology, e-Commerce and e-Service. EEE, 2004, 259–262.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details