IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Symptom based Disease Prediction, Drug recommendation and Drug Recommendation

SK. Heena[1], V. Surya Kiran[2], S. Susheel Kumar[3], Y. Rahul[4].

[1] *Assistant Professor, Department of Computer Science and Engineering, Qis college of Engineering and Technology,Ongole, Andhra Pradesh, India*
[2,3,4,5]*Student, Department of Computer Science and Engineering, Qis college of Engineering and Technology,Ongole, Andhra Pradesh, India*

Abstract— Data analysis and exploration are necessary for machine learning in order to identify significant trends and patterns. Everyone on Earth relies on allopathic therapies and medications in the modern era. Medical datasets can benefit from machine learning techniques since they offer a wide range of opportunities for both textual and visual data. Data mining is the key to extracting insight from these voluminous, enigmatic data that are present in medical services. This paper presents an API for prescribing medications to users with a certain ailment, which would also be diagnosed by the framework through the analysis of the user's symptoms using machine learning techniques. Here, we make use of some insightful data pertaining to the mining process to identify the most precise illness that may be associated with symptoms. The condition is easily identifiable by the sufferer. Patients may clearly identify their disease by merely attributing their symptoms, and the program interface indicates what disease the user may be infected with. The framework will show complacency in urgent instances where the patient cannot go to a doctor's office or in circumstances where professionals are available in the area. By considering numerous features in the database, predictive analysis on the illness would be carried out, leading to the recommendation of drugs to the user. The outcomes of the experiment can also be applied to healthcare tools and future research.

***Keywords*:- Machine learning, Drug recommendation, knowledge base, Logistic Regressions, symptoms, Support Vector Machine, Disease Predictions.**

I. INTRODUCTION

Although surgeons still need technology in a variety of ways, such as surgical representation and x-ray photography, it has perceptually lagged behind since the advent of powerful computing. Due to additional considerations including weather, atmosphere, blood pressure, and a wide range of other variables, the procedure still requires the doctor's knowledge and expertise. No model has been able to successfully examine the vast quantities of variables that are considered to constitute the entire variables needed to understand the complete functioning process itself. Medical decision support systems must be employed to address this problem. The doctors might get help from this system in making the right choice. The term "medical decision support system" refers to both the method used to try to identify potential diseases or disorders and the conclusion drawn from this method. The diagnostic opinion in the sense that it denotes a type of abnormality within a categorization or a degree of abnormality on a continuum. It is affected by non-medical elements including patient or doctor financial motivations and power ethics. It could be a succinct summary or a lengthy formulation, or it might even take the form of a tale or a metaphor. It could be a channel of communication, like computer code, that launches a payment, prescription, notification, informational request, or piece of advice. Knowing what is typical and assessing the patient's current condition are indications of medical diagnostics. Rule-based systems that automatically offer answers to routine management issues are known as automated decision support systems.

Due to various considerations or in the event of uncommon disorders, making medical decisions can be a very skilled and challenging task. Alternative variables include stress and tired misdiagnosis, which can result from medical professionals' ignorance and insufficient data. The full set of variables, including the current history of medical records, the history of family records, various aspects relating to patient records, and the vast quantity of accessible hidden factors, may be examined using a standard algorithm. When several alternatives are feasible, the presence of an entity can be detected using differential diagnosis techniques, which also includes the candidate alternatives. This strategy requires an information gathering or process of elimination that reduces the likelihood of candidate conditions to insignificant levels. There are four steps in it: 1) The doctor compiles all pertinent patient data and compiles a list of symptoms. 2) All potential reasons for the symptoms should be listed by the doctor. 3) The doctor should order the list such that the most serious potential causes of the symptoms are listed first. 4) The medical professional should rule out or treat all potential causes,

starting with the most serious emergencies. Use the test method or another scientific method to "rule out" a possibility. Using tests that should yield separate results and deleting the diagnosis from the list if there won't be any such diagnoses will rely on which diagnosis is correct. Based on the doctor's training and expertise, this is possible. This approach is incredibly simple to use.

to use the K-Means technique to decrease the enormous number of variables and identify the diseases that are most likely to occur. This method is better suited to grouping many diseases together. The clustering problem is solved using K-Mean, one of the unsupervised learning techniques. Finding the k centroids, one for each cluster, is the major goal. The many tests that were administered to the patients will be used as clustering attributes. Using this approach results in fewer iterations, clearly defined cluster boundaries without overlap, and accurate results for each and every diagnosis. Anyone with internet access can utilize this system's Service Oriented Architecture (SOA), which employs LAMSTAR Network to compute weight, improve algorithm correctness, assess overall performance,and deliver better results..

I.1 Problem statement

Although surgeons still need technology in a variety of ways, such as surgical representation and x-ray photography, it has perceptually lagged behind since the advent of powerful computing. Due to additional considerations including weather, atmosphere, blood pressure, and a wide range of other variables, the procedure still requires the doctor's knowledge and expertise. No model has been able to successfully examine the vast quantities of variables that are considered to constitute the entire variables needed to understand the complete functioning process itself. Medical decision support systems must be employed to address this problem. The doctors might get help from this system in making the right choice. The term "medical decision support system" refers to both the method used to try to identify potential diseases or disorders and the conclusion drawn from this method.

I.2 Motivation

Due to many considerations or in the event of uncommon disorders, making medical decisions can be an exceedingly specialized and challenging task. The other variables include stress; fatigued misdiagnosis may result from doctors' ignorance and poor information. The entire set of variables, including the current circumstances, the history of the patient's medical records, the history of the patient's family records, and numerous factors pertaining to the patient records, as well as the sheer quantity of attainable hidden factors, may be examined by a standard algorithm. When several alternatives are feasible, the presence of an entity can be detected using differential diagnosis techniques, which also includes the candidate alternatives. This strategy requires an information gathering or process of elimination that reduces the likelihood of candidate conditions to insignificant levels.

I.3. LITERATURE SURVEY

Using association rule mining and classification to combine to create a prediction model that achieves maximum accuracy, associative classification is a very new and lucrative technology. When maximal accuracy in a prediction model is required, associative classifiers are particularly well suited for use. Maximum model accuracy is needed in many fields, including the medical field. Heart disease is the leading cause of death in affluent nations and a major cause of disease burden in emerging nations. According to mortality data from the registrar general of India, heart disease is a leading cause of death in that country. In the state of Andhra Pradesh, coronary heart disease accounts for over 30% of rural fatalities. Therefore, a decision support system for diagnosing a patient's cardiac ailment is required. In this study, we suggest an effective genetic approach-based associative classification method for the prediction of heart disease. The primary drivers for utilising a genetic algorithm to find high level prediction rules are the rules' high levels of understandability, predictive accuracy, and interestingness. According to experimental findings, the majority of classifier rules aid in the most accurate diagnosis of heart disease and even assist physicians in their decision-making.

I.3.1 Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques

Authors: V.Krishnaiah , Dr.G.Narsimha, Dr. N. Subhash Chandra

Abstract: Cancer is the top cause of death in both men and women. Early cancer identification can help the disease be completely eradicated. Therefore, techniques to detect cancer nodules at an early stage are becoming more and more necessary. The disease of lung cancer is frequently misdiagnosed. If lung cancer is not found in its early stages, it might cause additional major problems that hasten mortality. Early detection and diagnosis of the illness play a significant role in the possibility of a cure and the prognosis. One of the most common types of medical malpractice committed worldwide is a diagnostic error. In the business and scientific realms, data mining and knowledge discovery have various applications. The use of data mining techniques in the healthcare system can produce insightful information. The possible application of rule-based, decision tree, naive Bayes, and artificial neural network classification-based data mining methodologies to a sizable amount of healthcare data is briefly evaluated in this research. The industry collects enormous amounts of healthcare data, but they are regretfully not "mined" to uncover hidden information. For data preprocessing and effective decision-making, One Dependency Augmented Naive Bayes Classifier (ODANB) and Naive Creedal Classifier 2 (NCC2) are utilized. With the aim of creating accurate classifications even when working with small or insufficient data sets, this is a naive Bayes extension to imprecise probabilities. Underutilization of underlying links and patterns is common. Complex "what if" questions cannot be answered by conventional decision support systems, but lung cancer disease detection can. Age, sex, wheezing, shortness of breath, and discomfort in the shoulder, chest, or arm are some common lung cancer symptoms that can be used to predict a person's probability of having the disease. The study's goal is to offer a framework for early disease identification and precise explains. Virtual bots and other models with CUI diagnosis, which will help the doctor save the patient'slife.

I.3.2 Data Mining Techniques to Predict Chronic Kidney Disease

Authors: Swaroopa Shastri, Surekha,

Abstract: Chronic renal disease includes the condition in which the kidneys stop working and lessen their ability to keep a person with the disease healthy. Wastes in the blood start to build in high levels as the kidneys' health deteriorates. The current trend for achieving analytic results is data mining. The human services sector collects enormous amounts of untapped data in order to uncover hidden information for insightful analysis and fundamental leadership. Data mining is the method for removing hidden information from huge databases. Our paper's objective is to predict CKD using the classification method Nave Bayes. In light of the Glomerular Filtration Rate, the stages of CKD are anticipated (GFR). One of the most common diseases in the US is chronic kidney disease (CKD). Twenty-six million persons in the United States have CKD, and another million are at elevated risk, according to recent figures. A kidney tissue sample is taken and tested together with blood and urine tests to determine the clinical diagnosis of CKD. The prevention of renal failure by early diagnosis and detection of kidney illness is crucial. By analyzing previous patient data and diagnosis records, data mining and analytics approaches can be used to predict CKD. In this study, CKD is predicted using predictive analytics methods like Decision Trees, Logistic Regression, Naive Bayes, and Artificial Neural Networks. Data pre-processing is done to find the factors that should be taken into account in the prediction models and to impute any missing data. Based on prediction accuracy, various predictive analytics methods are evaluated and contrasted.

II. PROPOSED SYSTEM

to use the K-Means technique to decrease the enormous number of variables and identify the diseases that are most likely to occur. This technique is more suited for clustering more diseases. One of the unsupervised learning algorithms used to address the clustering issue is K-Mean. Finding the k centroids, one for each cluster, is the major goal. The many tests that were administered to the patients will be used as clustering attributes.

Work that is paperless benefits the environment. • To improve the patients' efficiency and accuracy in order to aid them in the future. • Managing disease-related information.

III. DESIGN

System Architecture

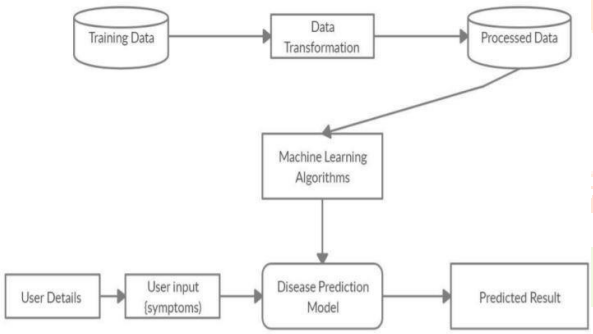


Fig 1: System design

flowchart

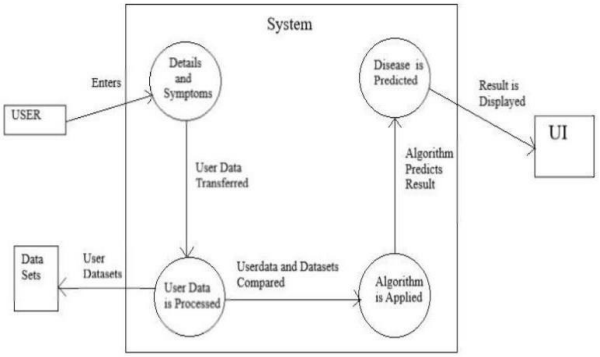


Fig 2: Class diagram

Sequence diagram

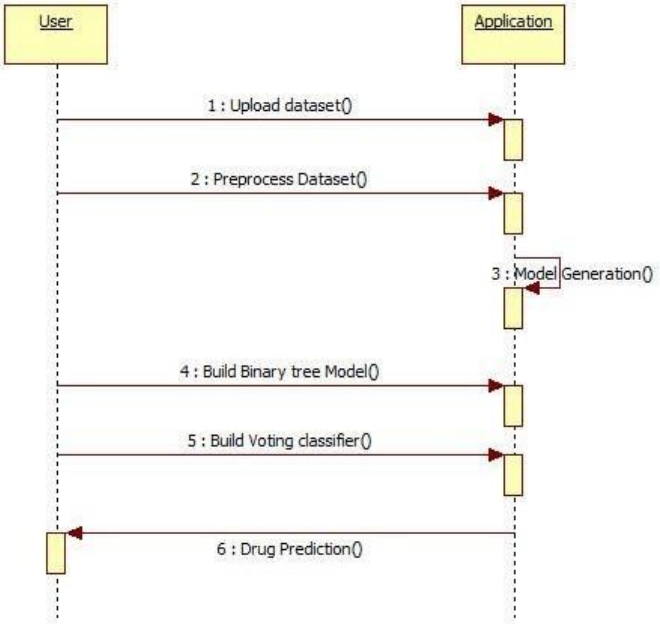
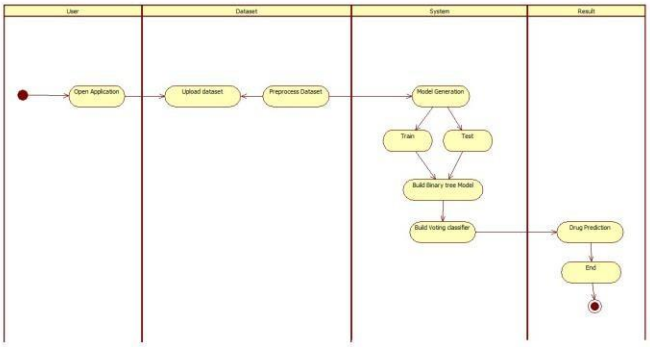
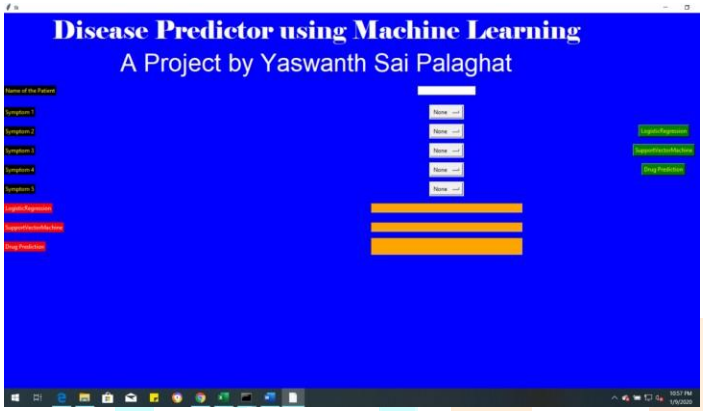


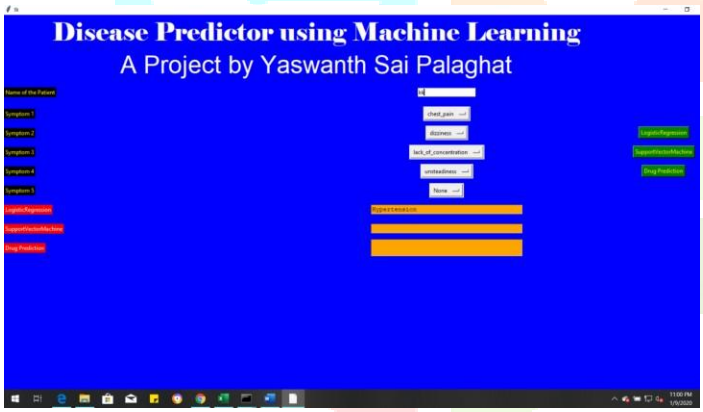
Fig 3: Sequence diagram

Activity Diagram:



A. Output screenshots:





B. CONCLUSIONS

This paper provided a flowchart of how information machine learning techniques are used in the regulatory, clinical, investigative, and educational aspects of clinical predictions. While there is currently limited practical use of information machine learning in health-related concerns, this paper established that information mining systems have a great potential to improve several aspects of clinical predictions. Additionally, the inevitable growth of clinical data will increase the possibility for data mining systems that improve social insurance quality and lower costs.

C. FUTURE ENHANCEMENTS

This paper provided a flowchart of how information machine learning techniques are used in the regulatory, clinical, investigative, and educational aspects of clinical predictions. While there is currently limited practical use of information machine learning in health-related concerns, this paper established that information mining systems have a great potential to improve several aspects of clinical predictions. Additionally, the inevitable growth of clinical data will increase the possibility for data mining systems that improve social insurance quality and lower costs.

REFERENCES

1. Mário W. L. Moreira, Joel J. P. C. Rodrigues, Antonio M. B. Oliveira, Ronaldo F. Ramos, Kashif Saleem, "A preeclampsia diagnosis approach using Bayesian networks", Communications (ICC) 2016 IEEE International Conference on, pp. 1-5, 2016, ISSN 1938- 1883.

2. Ameera M. Almasoud, Hend S. Al-Khalifa, Abdulmalik Al-Salman, "Recent developments in data mining applications and techniques", Digital Information Management (ICDIM) 2015 Tenth International Conference on, pp. 36-42, 2015

3. K Sowjanya, Ayush Singhal, Chaitali Choudhary, "MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices", Advance Computing Conference (IACC) 2015 IEEE International, pp. 397-402, 2015.

4. Bharathan Venkatesh, Danasingh Asir Antony Gnana Singh, Epiphany Jebamalar Leavline, Advance in Intelligent Systems and Computing, vol. 517, pp. 633, 2017, ISSN 2194-5357, ISBN 978- 981-10-3173-1.

5. Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In Proceedings of the 2018 International Conference on Digital Health (DH '18). ACM, New York, NY, USA, 121-125.

6. Effieient Algorithms to find Frequent Itemset Using Data MiningSagar Bhise1, Prof. Sweta Kale International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056.

7. Text Data Mining of Care Life Log by the Level of Care Required Using KeyGraph Muneo Kushima, Kenji Araki, Tomoyoshi Yamazaki, Sanae Araki, Taisuke Ogawa, Noboru Sonehara Proceedings of the International MultiConference of Engineers and Computer Scientists 2017 Vol I, IMECS 2017, March 15
- 17, 2017, Hong Kong.

8. Chen, W., Guo, H., Zhang, F., Pu, X., and Liu, X. (2012). Mining Schema Matching Between Heterogeneous Databases. In Consumer Electronics,

Communications and Networks (CECNet), 2012 2ndInternational Conference on, pages 1128–1131. IEEE.

9. Han, J. and Gao, J. (2009). Research Challenges for Data Mining in Science and Engineering. Next Generation of Data Mining, pages 1– 18.

10. H. Abe, S. Hirano and S. Tsumoto, "Evaluation Temporal Models Based on Association Mining From Medical Documents," Japan Association for Medical Informatics, vol. 27, no. 1, pp. 33-38, 2007.

11. Sun, Y., Han, J., Yan, X., and Yu, P. S. (2012). Mining Knowledge from Interconnected Data: A Heterogeneous Information Network Analysis Approach. Proceedings of the VLDB Endowment.

12. M. Akhil jabbar & Dr. Priti Chandrab "Heart Disease Prediction System using Associative Classification and Genetic Algorithm" International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT, 2012.

13. Nikhil N. Salvithal" Appraisal Management System using Data mining "International Journal of Computer Applications (0975 – 8887) Volume 135 – No.12,February 2016.

14. Tanvi Sharma, Anand Sharma & Vibhakar Mansotra "Performance Analysis of Data Mining Classification Techniques on Public Health Care Data" International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 6, June 2016.

15. V.Krishnaiah , Dr.G.Narsimha, Dr.N.Subhash Chandra "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques"(IJCSIT)International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 - 45.

16. Jaimini Majali, Rishikesh & Niranjan, Vinamra Phatak "Data Mining Techniques For Diagnosis And Prognosis Of Cancer" International Journal of Advanced Research in Computer and CommunicationEngineering Vol. 4, Issue 3, March 2015.

17. Data Mining Techniques to Predict Diabetes Influenced Kidney Disease Swaroopa Shastri, Surekha, SaritaInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307.

18. DeFronzo Ralph. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. Diabetes. 2009;58:773-95.