International Journal of Engineering and Techniques - Volume 7 Issue 3, May 2021

RESEARCH ARTICLE OPEN ACCESS

Diabetes Prediction Using Data Mining

Prof.Prashant Ahire[1],Mr.Suraj Kshirsagar[2],Mr.Saurabh Chauhan[3],Mr.Sandesh Kolhe[4],Mr.Vaibhav Chavan[5]

1(Dept of Computer Engineering, Dr.D.Y.Patil Institute of Technology, Pimpri ,Pune

Email: ksprashantahire@gmail.com)

2,3,4,5 (Student at Dr.D.Y.Patil Institute of Technology, Pimpri ,Pune

Email: surajkshirsagar485@gmail.com)

Abstract:

This research paper and project was conducted on the design and implementation of a diabetes prediction system, a case study of Fudawa Health Centre. This research will help in prediction of diabetes even before we specifically check for diabetes clinicals. The current protransparent. The cess of carrying this activity is manually which tends not to analyze data flexible for the doctors, and transmission of information is in prediction system was designed using Programming Language the existing system was taking in order to meets the demands of this system Python and MySQL (Microsoft Structured Query at Language) as the back end and a strategic approach to analyze and solve the problems of the existing system by implementing the naïve bayes classifier. The implementation of this new system will help to reduce the stressful process, doctors' face during prediction of diabetes, the result of the experiment shows that the propose better prediction in terms of accuracy.

Keywords *— **Diabetes , Machine Learning , Predict Dataset.***

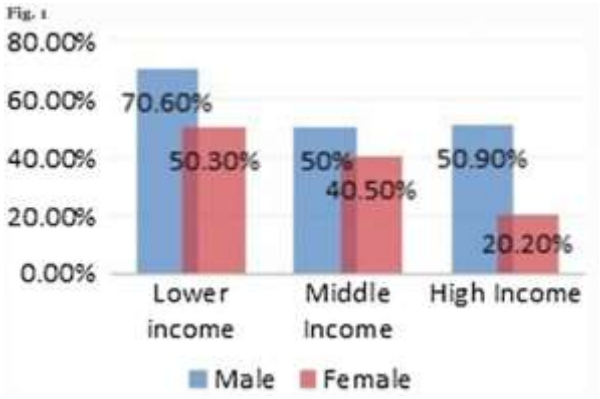
I. INTRODUCTION

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" Although data mining has been around for more than two decades, its potential is only being realized now. Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases. The two most common modelling objectives are classification and prediction. Classification models predict categorical labels (discrete, unordered) while prediction models predict continuous-valued functions. Decision Trees and Neural Networks use classification algorithms while Regression, Association Rules and Clustering use prediction algorithms is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website.

II. LITERATURE REVIEW

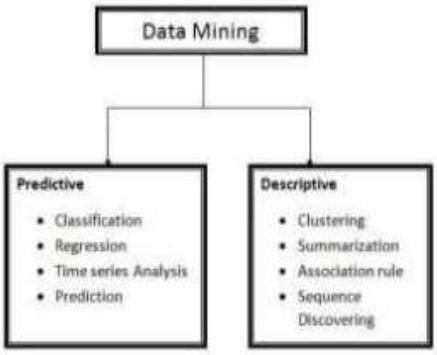
Data Mining is the process of extraction of hidden patterns from previously unknown and imaginably useful information from huge amounts of data. The diagnosis of Disease Is one of the major applications where data mining tools are showing

successful results. Lifestyle Diseases linked with the way people live their life. Numerous works have been done related to lifestyle disease diagnosis using different data mining techniques. The dataset, algorithms, methods used by the authors and the observed results along with the future work is carried out in finding out efficient methods of medical diagnosis for various lifestyle diseases. The annual report of World Health Association, add up to the number of individuals experiencing diabetes is 422 million the year. Consistently, there is a significant increment in the number individuals experiencing diabetes in different healing centers .The world health organization(WHO) reports on "Diabetes Care 2021" by American Diabetes Association and Standards for Medical care in Diabetes, a study for correlation between diverse races and their pay. Figure.1 demonstrates the diverse individuals (gender and wage) matured between 29 and 70 years, level of passing because of hypertension.



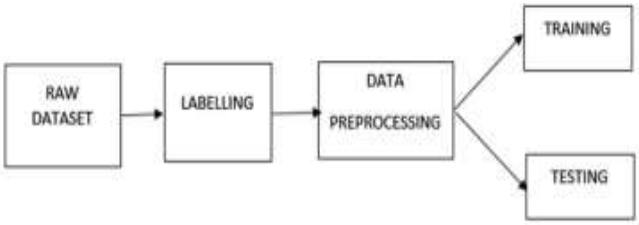
III. SYSTEM ARCHITECTURE

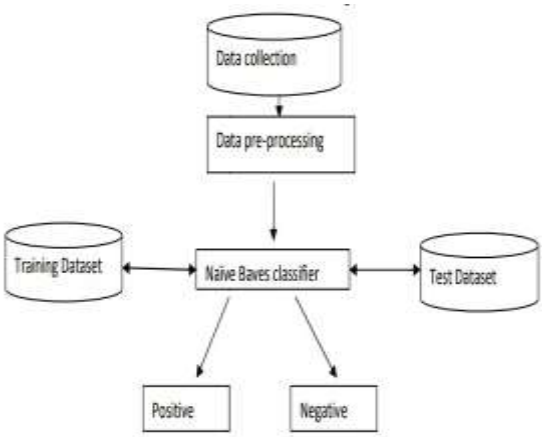
System architecture is a conceptual model that defines the structure and behavior of the system. It comprises the system components and the relationship describing how they work together to implement the overall system. System design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements .Systems design could be seen as the application of systems theory to product development. Object- oriented analysis and methods are becoming the most widely used methods for computer systems design. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user. The UML has become the standard language in object oriented analysis and design.



IV. DATA FLOW DAIGRAM

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing. A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored.





A. Other Specifications

This application would be a tremendous asset for doctors who can have structured specific and invaluable information about their patients / others so that they can ensure that their diagnosis or inferences are correct and professional. Finally, the huge appreciations received from the doctors on having such software prove that in a place like, where diseases are on the rise, such applications should be developed to cover the entire state. The common person stands to benefit from doctors having such a tool so that he/she can be better knowledgeable as far as personal health and wellbeing is concerned.

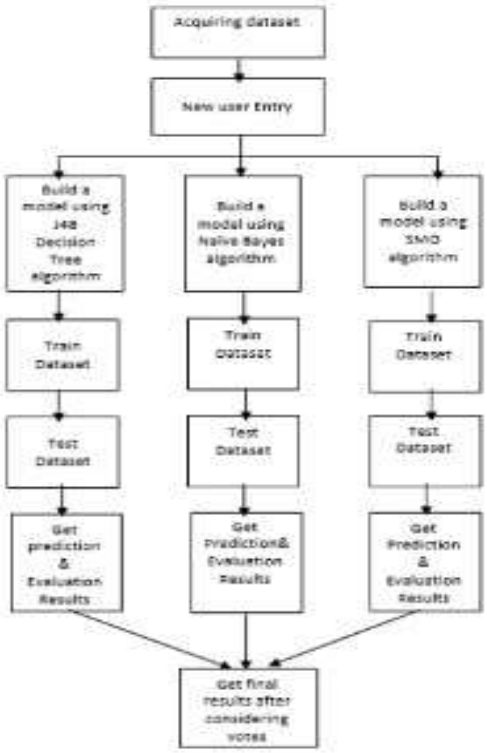
1) ***Advantages :***

- *Fast prediction of diabetes over a conventional method.*
- *User friendly design and keeps data secured.*
- *Checks diabetes by comparing different data like Blood Pressure , Skin thickness, Insulin, Glucose, Pregnancies etc. for accurate results.*

2) ***Limitations :***

- *As this is a Medicle Project, so sometimes it may not give accurate results.*

B. Entity Relationship Diagram



C. System Implementation Plan

1. Login *:*

a) Add/View Training data
b) Add/View Doctor data
c) View user details
d) View feedback

2. User :

a) Register with basic details like Name, Mobile number, Age etc.
b) Login
c) Check diabetes by comparing details with collected data.

V. CONCLUSIONS

An Application using a data mining algorithm of classes' comparison has been developed to predict the occurrence of or recurrence of diabetes risks. In addition, the result of the application shows that the predictions system is capable of predicting diabetes effectively, efficiently and most importantly, timely. That means the application is capable of helping a physician in making decisions towards patient health risks. It generates results that make it closer to real

life situations. That makes data mining more helpful in the health sector, which means that it is necessary for knowledge discovery in the healthcare's sector. Much more than huge savings in costs in terms of medical expenses, loss of duty time and usage of critical medical facilities, The naïve bayes classifier based system is very useful for diagnosis of diabetes. The system can perform good predictions with less error and this technique could be an important tool for supplementing the medical doctors in performing expert diagnosis. In this method the efficiency of forecasting was found to be around 95%. This application would be a tremendous asset for doctors who can have structured specific and invaluable information about their patients / others so that they can ensure that their diagnosis or inferences are correct and professional. Finally, the huge appreciations received from the doctors on having such software prove that in a place like, where diseases are on the rise, such applications should be developed to cover the entire state. The common person stands to benefit from doctors having such a tool so that he/she can be better knowledgeable as far as personal health and wellbeing is concerned.

ACKNOWLEDGMENT

This has been the light of the day due to invaluable contribution of certain individuals whose constant guidance, support and encouragement resulted in the realization of our project.

We are grateful to our Guide **Prof.Prashant Ahire** and HOD of computer department **Dr. S. V. Chobe** for providing us the necessary help and encouragement whenever we needed, which has resulted in the success of our project.

We would also like to thank all the staff members of our department, without whose constructive suggestions and valuable advice, the simple idea, which had borne by us, would not have been able to blossom forth to give such a beautiful bloom.

Last but not the least; we are grateful to all our friends and our parents for their direct or indirect constant moral support throughout the course of this project.

REFERENCES

[1]. Acharya, Rajendra, U and Yu, Wenwei, (2010).Data Mining Techniques in Medical Informatics. The Open Medical Informatics Journal, PMCID: PMC2916206.

[2]. Aflori C., and Craus, M., (May 2007) Grid Implementation of the Apriori algorithm Advances in Engineering Software, 38(5), pp. 295-300. A. J.T. Lee, Y.H. Liu, H.Mu Tsai, H. [3]. Anbarasi M., (2010). 'Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm,' International Journal of Engineering Science and Technology, 2(10), 5370- 5376.

[4]. Bronzino, D. Joseph, Medical Devices and Systems, 2006

[5]. Chauraisa V., and Pal, S.,(2013). 'Data Mining Approach to Detect Heart Diseases', International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2, (4), pp 56-66.

[6]. Clifton, Christopher (2010), Encyclopedia Britannica: Definition of Data Mining Retrieved 2016.

[7]. Data Mining Curriculum, ACM SIGKDD, 2006-04-30, retrieved 2016

[8]. Fayyed, Usama., (15 June1999), First Editorial By Editor-InChief, SIGKDD Explorations 1:1, doi: 10.1145/2207243.2207269

[9]. Fayyed, Usama., Piatetsky- Shapiro, Gregory., Smyth, Padhraic.,(1996) From Data Mining to Knowledge Discovery in Databases.

[10]. Han J.,and Kamber, M. (2010). Data Mining: Concepts and Techniques, 2nd ed., the Morgan Kaufmann Series.

[11] Y Cai, D Ji, Dong-feng Cai, "A KNN Research Paper Classification Method Based on Shared Nearest Neighbor", at Shenyang Institute of Aeronautical Engineering.

[12] K Saxena1, Dr. Z Khan2, S Singh3," Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm", at Invertis University