Early Diabetes Prediction Using Machine Learning

Ambuja K [1], Anusha G S [2], Sangeetha T D [2], Sharath Kumar N [2], Sumanth K Naik [2]

[1]Asst. Prof, CSE, KS School of Engineering and Management, Bengaluru, India.
[2] Student, CSE, KS School of Engineering and Management, Bengaluru, India.
Email: ambuja@kssem.edu.in

Abstract: - Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affect other organs of human body. Diabetes can be controlled if it is predicted earlier. Diabetes is one of the fastest-growing diseases in the world and requires constant monitoring. To verify this, we are exploring different machine learning algorithms that will help with this baseline prediction. achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by constructing models from datasets collected from patients. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. We are using Random Forest (RF) for model building. The Project work gives the accurate or higher accuracy model shows that the model is capable of predicting diabetes effectively. We are expecting our result to show that Random Forest achieved higher accuracy. The accuracy of this model is above 90%.

Key Words: — *Machine learning, Random Forest, Accuracy.*

I. INTRODUCTION

Diabetes is a disease which lasts longer with the potential to cause a worldwide health care disaster. Many algorithms are used to predict diabetes, including machine learning methods like Random Forest, (KNN) K-Nearest Neighbor, Decision Tree and so on. With this machine learning techniques. Diabetes mellitus or imply diabetes is a disease caused due to the increase level of blood glucose and by the parental genetics (DNA). Traditional methods of different kinds, based on physical and chemical tests, are available for diagnosing diabetes disease. Early prediction of diabetes is bit challenging task for medical field due to complex interdependence on different parameters as diabetes affects human organs such as eye, heart, kidney, nerves and foot. Machine learning is a fast-growing scientific field in data science which imitates human intelligence and learn from past experience.

The aim of this project is to develop a model which can perform early prediction of diabetes for a patient with a greater accuracy by comparing the results of different machine learning techniques.

II. METHODOLOGY

In order to achieve our goal, our methodology comprises if few steps from which we accumulate datasets of the given attributes for the patients and we will do the pre-processing of our given attribute to apply on the given machine learning techniques find out the predictive analysis of the data.

Table.1. Dataset Description

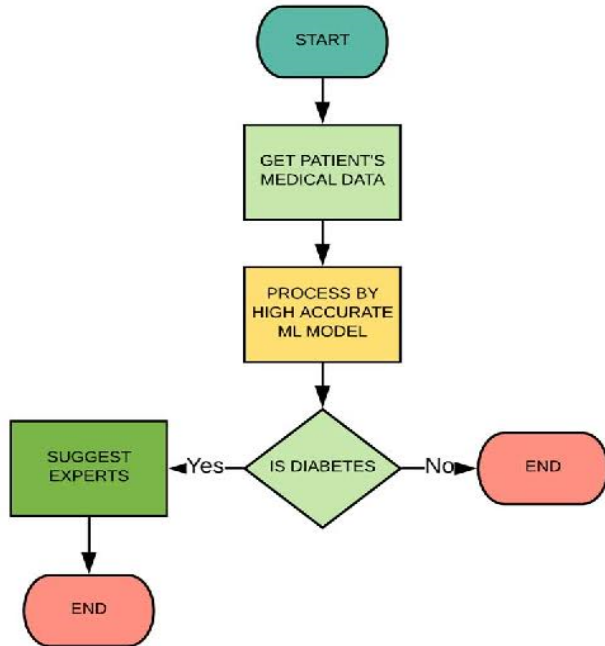| S no.	Attributes
1	Pregnancy
2	Glucose
3	Blood Pressure
4	Skin Thickness
5	Insulin
6	BMI (Body Mass Index)
7	Diabetes Pedigree Function
8	Age

Fig.1. Disease Prediction Algorithm

2.1 Dataset And Attributes

In this work, we collect diabetes data from Kaggle. The dataset consists of various attributes for diabetes mellitus for 2700 patients. the attributes are.

2.2 Applying Machine Learning Techniques

Once the data has been created for modelling, we employed Random Forest machine learning classification algorithm which we are used to implement to predict diabetes mellitus. Some overview of these techniques.

2.3 Random Forest

This an ensemble learning method for classification and regression and other task that operates by constructing a multitude of decision tree at training time and outputting the class that takes average votes of individual trees. The first algorithm for random decision forests was created by Tin Kam Ho using random subspace method. Ho established that to gain the accuracy it should over train where it can randomly restrict sensitive selected features of the given data.

Random forests consist of 4 –12 hundred decision trees, each of them built over a random extraction of the observations from the dataset and a random extraction of the features. Not every tree sees all the features or all the observations, and this guarantees that the trees are de-correlated and therefore less prone to over-fitting. Each tree is also a sequence of yes-no questions based on a single or combination of features. At each

node (this is at each question), the three divides the dataset into 2 buckets, each of them hosting observations that are more similar among themselves and different from the ones in the other bucket. Therefore, the importance of each feature is derived from how "pure" each of the buckets is.
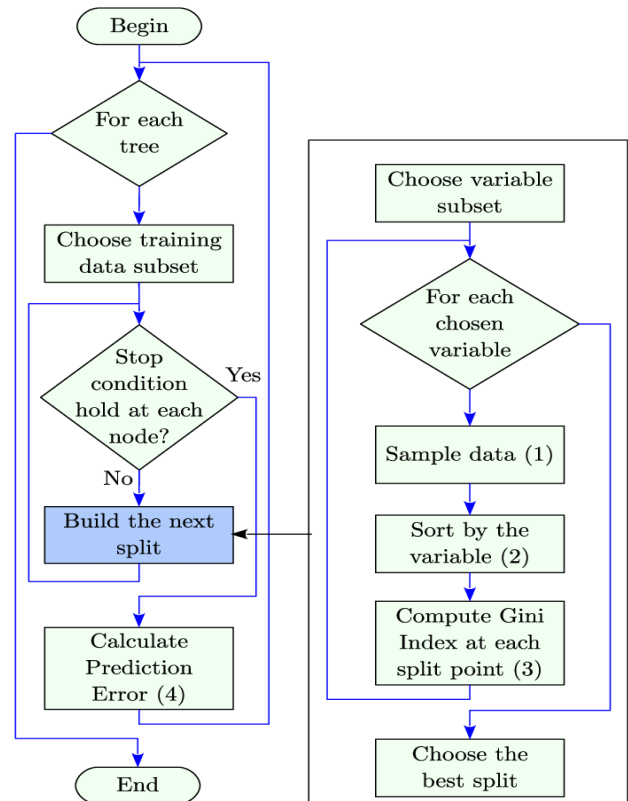

Fig.2. Working of Random Forest Algorithm

III. OTHER TECHNOLOGIES USED

3.1 DJANGO

Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation.

3.2 PYTHON

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different

programs and isn't specialized for any specific problems. The python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.
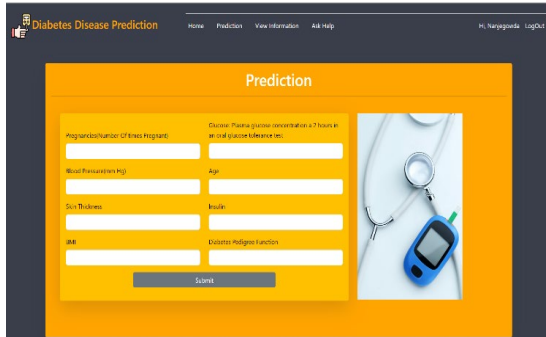
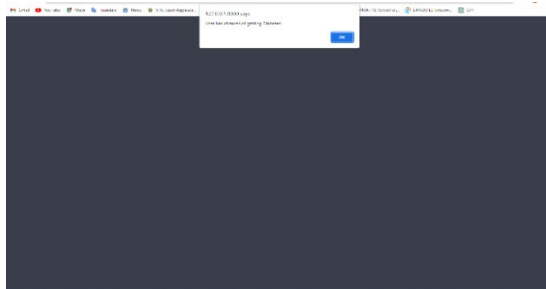## IV. SNAP SHOTS

Fig.3. Prediction page
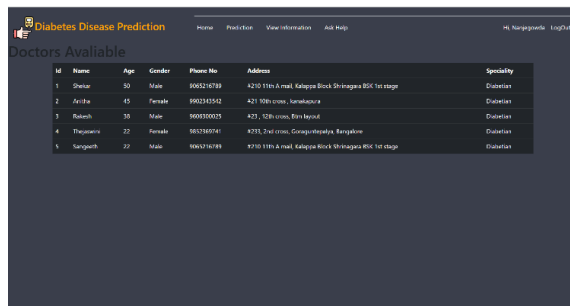

Fig.4. Prediction result page.

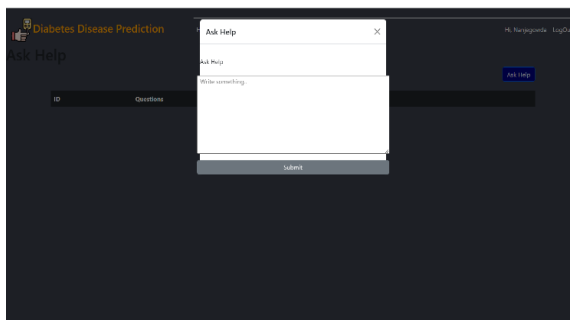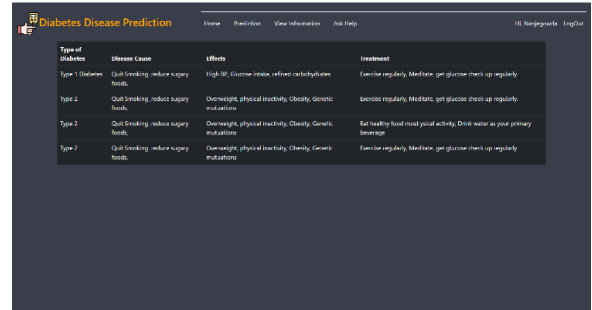Fig.5. Available Doctors information

Fig.6. Help desk.

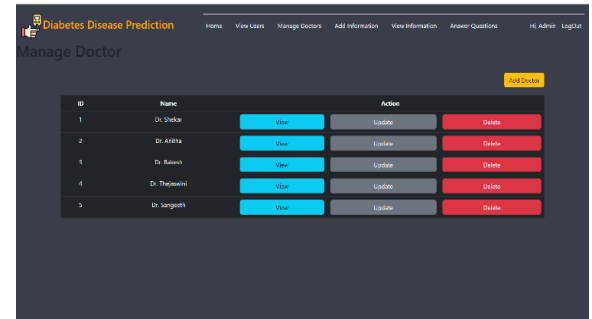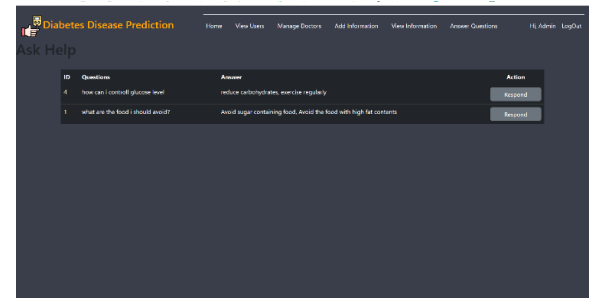Fig.7. Diabetes Types and measures.

Fig.8. Doctor information.

Fig.9. Doctor Response.

V. GOALS AND OBJECTIVES

The goal of this project is to find effective machine learning based classifier models for detecting diabetes in individuals utilizing clinical data. The results of this study suggest that an appropriate pre-processing pipeline on clinical data and applying ML based classification may predict diabetes accurately and efficiently. The derived results for predicting the diabetes will be displayed using smart web application.

[1] To collect dataset of different classes. Indian diabetes dataset named PIMA will be used for analysis for this study. It consists of seven independent attributes. [2] To improve quality and effectiveness of obtained data. In order to increase the performance of machine learning models the accuracy must be improved. [3] To predict diabetes using ML algorithms. Model building using Machine Learning algorithm Random Forest. [4] To provide information to the

patients whether they may affect diabetes or not. The technique may also help researchers to develop an accurate tool that will help them make better decision about the disease status. [5] Provide the highest accurate result for the person with controlling measures. The controlling measures like what type of treatment they should take, type of food that should have and avoid and what kind of activity they should do to control the disease. [6] Database will be maintained and Doctors information will be provided. The data of the person is maintained and the person can see the doctors available nearby and he\she can ask the question to the doctors.

VI. Contribution To Society

The model could be used in health monitoring programs, and as an automated mass population screening tool without the need for extra costs compared to traditional methods. It increases the risk of early death, and it's complications can lower quality of life. Early prediction of diabetes can help you avoid complications. Since diabetes is a non-curable disease, the present treatments and medications will only control associated symptoms. Therefore, it can significantly help diabetes research and, ultimately, improve the quality of healthcare for diabetic patients. It allows you to understand the importance of diet plan and exercise along with the medications.

APPLICATIONS:

This application provides wide range of benefits for doctors as well as patients as it provides early prediction of diabetes, it provides easy accessibility for both the doctors and patients. This system also helps patients to check their results using webpage without consulting doctors. Diabetes prediction using machine learning is to provide warnings to patients who are at the starting stage of diabetes. It helps people to take precautions when they grow older. It also helps doctors to predict diabetes via machine learning algorithms using past medical records of diabetes patients. As we are developing the application, it is accessible on user's device such as smart phones and computers. It will provide result with few known parameters instead of blood test.

VII. Conclusion

The main aim of this project is to present a clear overview of automatic Machine Learning based diabetes prediction. ML techniques has been developed for the effective and efficient prediction of diabetes. The goal of developing a diabetes prediction model is to shift from higher precision to higher reliability for real times applications. The model is trained and tested with Random Forest techniques. A model that can be used to predict diabetes in the world population is needed. Hybrid ML/combined algorithms can be introduced in the challenging complicated datasets in future research to obtain better performance. These techniques will be helping patients when performing real-world tasks in digital reality. Analysis of Diabetes disease is an attractive area for big data Analytics research for various purposes, including its significant impact on health and the demand for diabetes prediction. In the future, if more and more diabetic patients use automatic glucose sensors that continuously measure glucose levels, the amount of data related to blood sugar will greatly increase. This discussion helps to provide a clear-cut view of diabetes prediction and helps to frame better diabetes prediction techniques to overcome diabetes through timely prediction.

References

[1]. A Review of Diabetic Prediction Using Machine Learning Techniques | M. Rajeswari1, Dr. P.Prabhu 2 1(M.Phil research scholar, Department of Computer Applications, Alagappa University, 2(Assistant professor in IT (DDE), Department of Computer Applications, Alagappa University, Karaikudi.| published on 2019.

[2]. Diabetes Prediction using Machine Learning Techniques | Mitushi Soni Dept of Computer Science and Engineering Shri G.S. Institute of Technology and Science, Dr. Sunita, Shri G.S | September-2020 | Published by IJERT.

[3]. Diabetes Prediction Using Machine Learning | Kishan Patel Department of IT Engineering PHCET, Manu Nair Department of IT Engineering PHCET, Shubham Phansekar Department of IT Engineering PHCET | International Journal of Scientific & Engineering Research | Volume 12| Issue 3 | March-2021.

[4]. An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning | Umm e Laila 1, Khalid Mahboob 2, Abdul Wahid Khan 3, Faheem Khan 4, and Whangbo Taekeun 4, Department of Computer Engineering, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan;| published at sensor | Published: 13 July 2022.

[5]. Diabetes Prediction Using Machine Learning | International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653.

[6]. Dr. Vaidehi V, Aishwarya Mujumdar "Diabetes Prediction Using Machine Learning algorithms" Year: 2019 | Journal Article | Publisher: ELSEVIER

[7]. Md. Kamrul Hasan, Md. Ashraful Alma, Dola Das, Eklas Hossain, Mahmudul Hasan "Diabetes Prediction Using

Ensemble of Different Machine Learning Classifiers" IEEE Access Year: 2020 | Volume: 8, Journal Article | Publisher: IEEE.

[8]. Jitranjan Sahoo, Manoranjan Dash, Abhilash Pati "Diabetes Prediction Using Machine Learning Classification Algorithms" Year: 2020.

[9]. Leila Ismail, Huned Materwala "intelligent diabetes mellitus prediction framework using machine learning" emerald Access year:2021.

[10]. Usama Ahmed, Ghassan Issa, Muhammad adnan Khan, Shabib Aftab, Muhammad Farhan Khan, Reaed A T Said, Taher M Ghazal "Diabetes Prediction Empowered with Fused Machine Learning" | Year: 2022 | Volume :5, Journal Article | Publish: IEEE.

[11]. Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis and Konstantinos Moustakas "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction" | Year: 2021 | Volume: 8, Journal Article | Publisher: IEEE.