

Diabetes Prediction System Using Gaussian Algorithm

Sarvesh Wadekar¹, Nihar Yelve², Kuldip Yadav³, Prof. Urjashree Patil⁴

UG Students, Department of Computer Engineering^{1,2,3}

Professor, Department of Computer Engineering⁴

Shivajirao S. Jondhale College of Engineering, Dombivli, Thane, Maharashtra, India

Abstract: *Diabetes is among critical diseases and lots of people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes. To deal with this problem we have created a system which can predict Diabetes at early stages. We have made it in the form of a mobile app so everyone can use anywhere with just a touch on their mobile. The system is capable of predicting Diabetes by using users data such as age, BMI, insulin, blood pressure, glucose, pregnancy and with the help of our trained model which is using Gaussian algorithm we are predicting if the user has Diabetes or not with high accuracy. This system helps people to save their money and time from the tests that they have to undergo.*

Keywords: Gaussian NB, Diabetes Prediction, App, Machine Learning

I. INTRODUCTION

Healthcare sectors have large volume databases. Such databases may contain structured, semi-structured or unstructured data. Big data analytics is the process which analyses huge data sets and reveals hidden information, hidden patterns to discover knowledge from the given data. Considering the current scenario, in developing countries like India, Diabetic Mellitus (DM) has become a very severe disease. Diabetic Mellitus (DM) is also known as Non-Communicable Disease (NCB).

A technique called, Predictive Analysis, incorporates a variety of machine learning algorithms, data mining techniques and statistical methods that uses current and past data to find knowledge and predict future events. By applying predictive analysis on healthcare data, significant decisions can be taken and predictions can be made.

Predictive analytics can be done using machine learning and regression technique. Predictive analytics aims at diagnosing the disease with best possible accuracy, enhancing patient care, optimizing resources along with improving clinical outcomes. Machine learning is considered to be one of the most important artificial intelligence features supports development of computer systems having the ability to acquire knowledge from past experiences with no need of programming for every case.

Machine learning is considered to be a dire need of today's situation in order to eliminate human efforts by supporting automation with minimum flaws. Existing method for diabetes detection uses lab tests such as fasting blood glucose and oral glucose tolerance. However, this method is time consuming.

There has been drastic increase in rate of people suffering from diabetes since a decade. Current human lifestyle is the main reason behind growth in diabetes. In order to avoid or reduce severity of such impact, there is a need to create a system using machine learning algorithm and data mining techniques which will provide accurate results and reduce human efforts. Our project will be basically focusing on diabetes prediction using Gaussian algorithm.

II. LITERATURE SURVEY

Diabetes prediction using Machine learning was discussed by Jingyu Xue, FanchuMin[1] as they used few machine learning algorithms to predict diabetes at younger stages. Here they used SVM and Light GBM and they found that the SVM algorithm gives the highest accuracy and hence they proved that machine learning can also be an integral part of medical science.

Mitushi Soni and Dr. Sunita Varma [2] proposed KNN, Random Forest and SVM algorithms and showed results using popular algorithms here the performance of these algorithms was shown. In this she compared the algorithms and showed their accuracy. S.R. Priyanka Shetty and Sujata Joshi [3] used some data mining techniques and showed that whether the user is diabetic or no diabetic the also added few features to show the blood sugar levels and spread some awareness.

Huaping Zhou, Raushan Myrzashova and Rui Zheng [4] used some enhanced deep neural networks to predict diabetes. They used DNN model as their prediction system they used different techniques to increase the accuracy of model like Hyper parameter Tuning

N.Sneha and Tarun Gangil [5]also used SVM, Random Forest, KNN, Naive Bayes and showed performance and compared the accuracy.

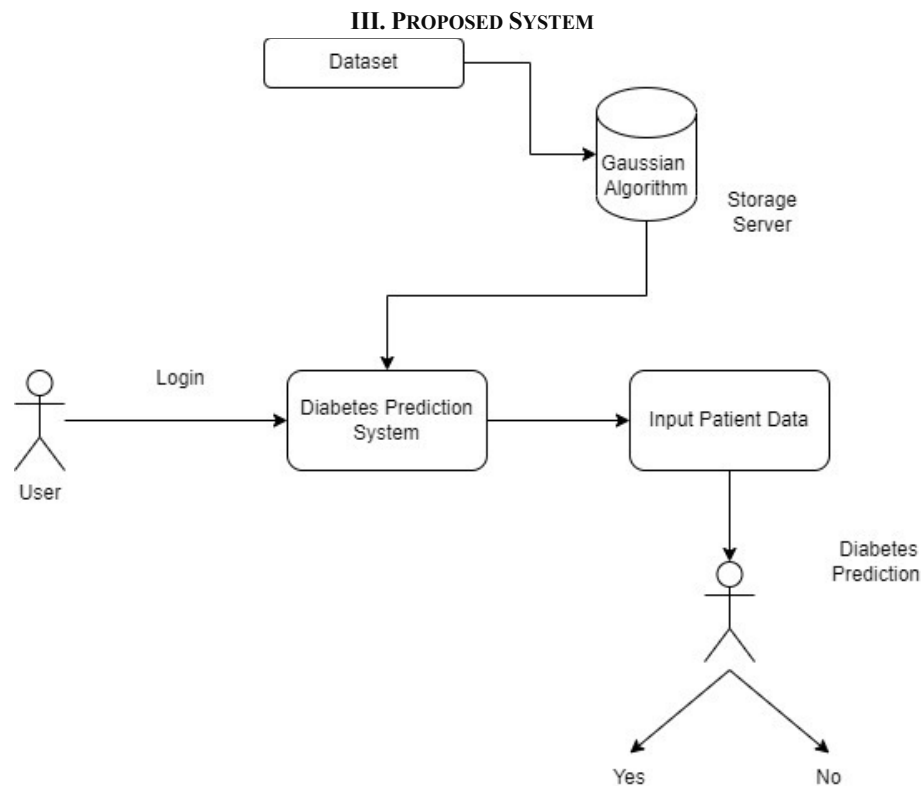


Figure 1: Proposed System

The data set in this project comes from the open source standard test data set website Kaggle. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. The dataset is trained and tested using gaussian algorithm and is stored in a server. When a user enters the Diabetes Prediction System after login id it cannot change it's username and password. After login the user can predict diabetes by entering the appropriate values. In the end the user will get to know if it has diabetes or not.

IV. METHODOLOGY

Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian distribution. This fast and flexible model gives highly reliable results. It works well with large data and it will be used for classifying large surveys in future. There is no need to spend much time for training as it is very easy to use. It provides better grading performance by eliminating insignificant specifications.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x)$ = probability density function

σ = standard deviation

μ = mean

4.1 Gaussian Naive Bayes

When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed to be-

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Sometimes assume variance

- is independent of Y (i.e., σ_i),
- or independent of X_i (i.e., σ_k)
- or both (i.e., σ)

Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution.

An approach to create a simple model is to assume that the data is described by a Gaussian distribution with no co-variance (independent dimensions) between dimensions. This model can be fit by simply finding the mean and standard deviation of the points within each label, which is all what is needed to define such a distribution.

4.2 Flowchart for Machine Learning Algorithm

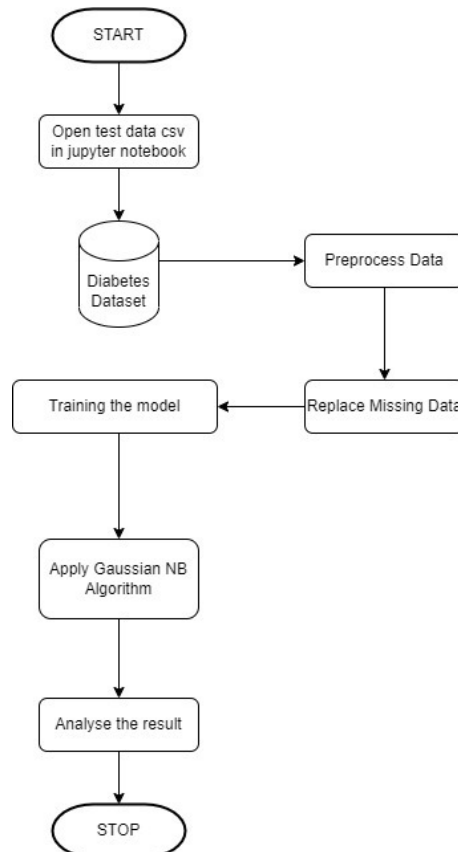


Figure 2: Flowchart for ML algorithm

Above figure shows working of our machine learning algorithm. The process starts by loading the dataset into our machine learning model, it uses this dataset to train the model after that it applies Gaussian NB algorithm on data and result analysis takes place.

4.3 Flowchart for App

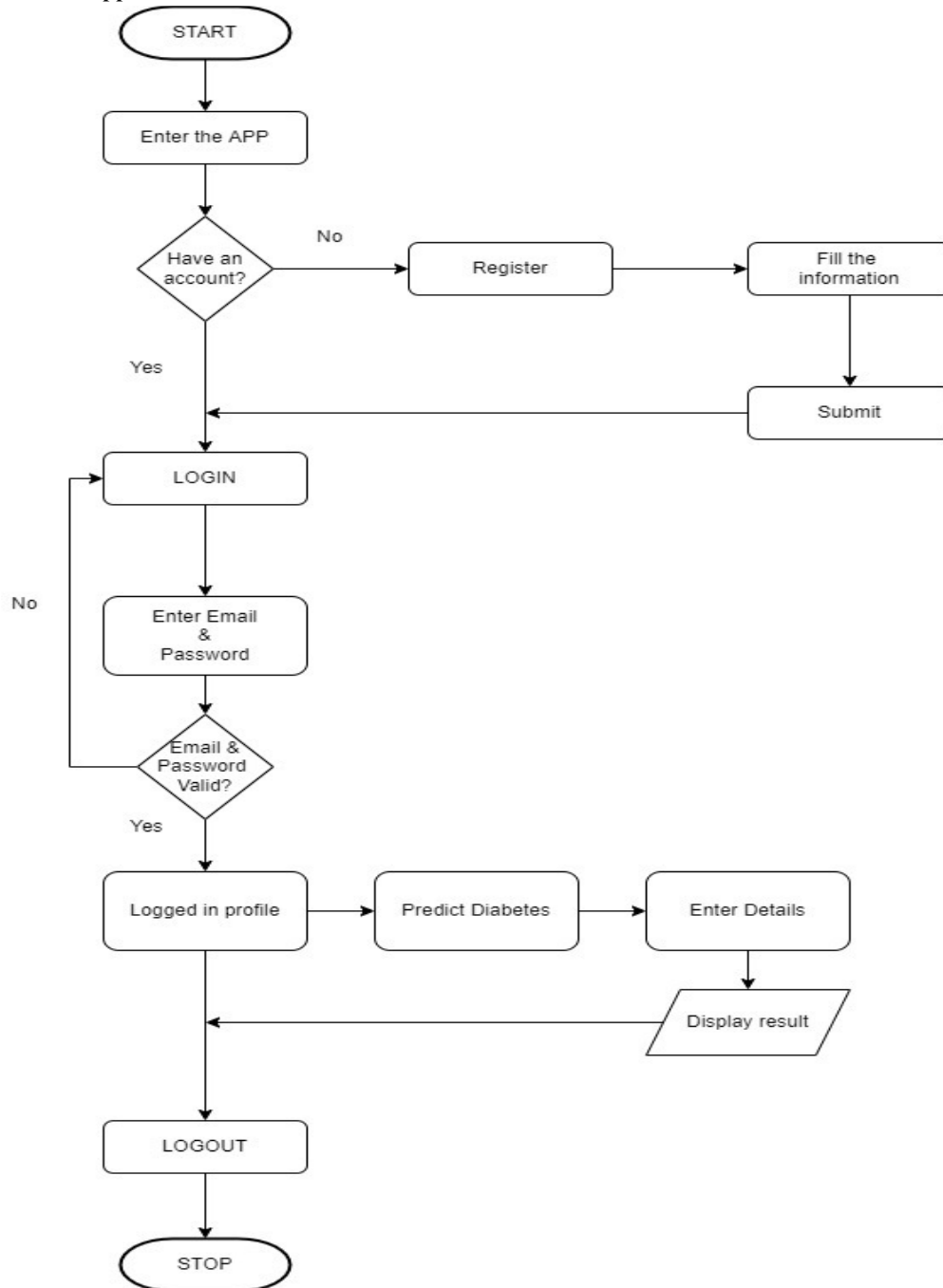


Figure 3: Flowchart for App

Above figure shows working of the app. After entering the app, user will be greeted with login screen, if it is an existing user, user can login by entering their credentials. If it is a new user, then user can register by entering their personal details, and then login to the app. After login, they could predict diabetes by tapping the predict button and entering the desired information. Then the result is displayed and then they can logout from the app.

V. RESULT ANALYSIS

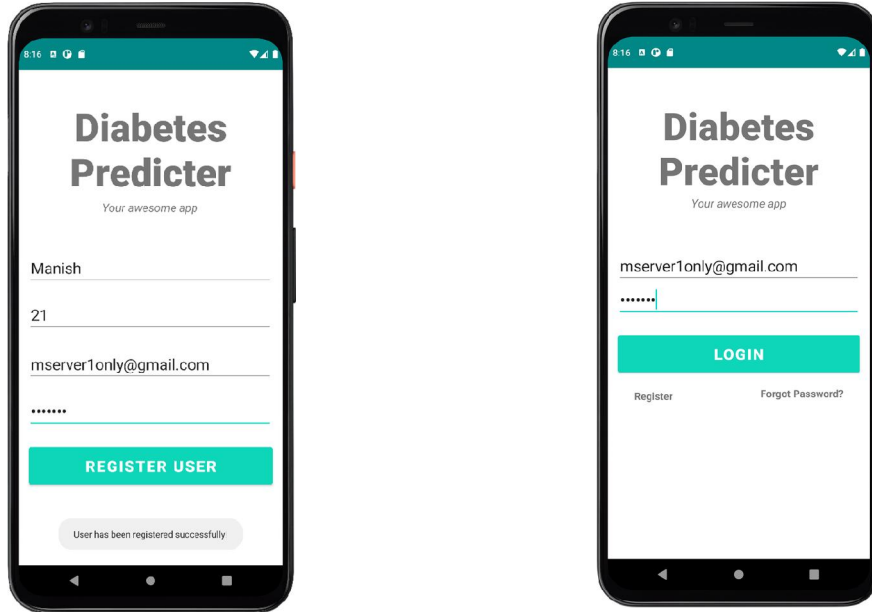


Figure 4: Login Screen

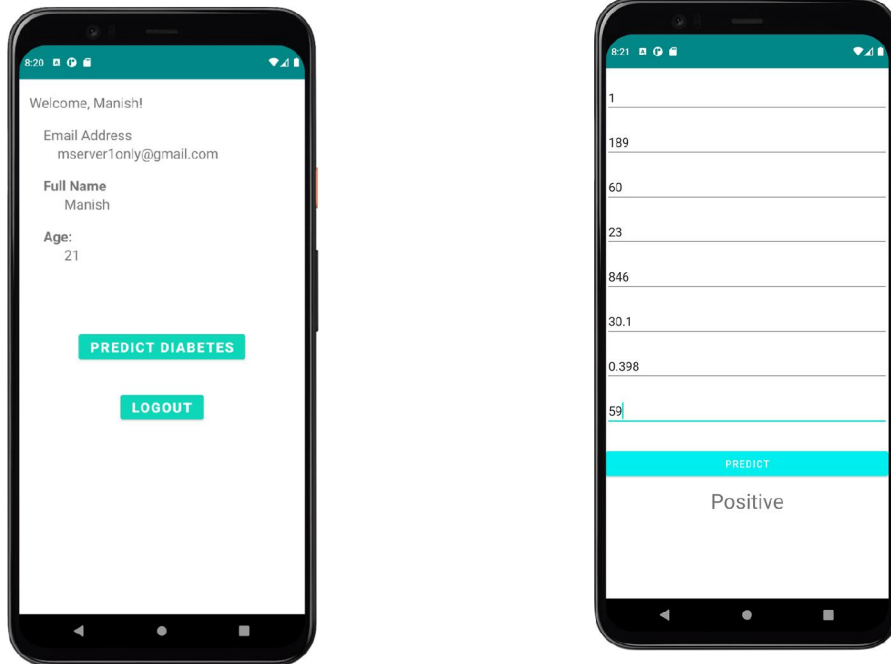


Figure 6: Prediction Screen

VI. APPLICATIONS

1. We can use this system at home if you have to check it without wasting time or if you don't want to go to hospital for checkup.
2. We can use this in hospitals too where there is lack of facility for checking diabetes.
3. Medical scientists can also use this algorithm and can do research on diabetes.
4. Doctors can use this system when they don't have kit to check diabetes by taking only user's data.
5. Patient's data are easy to send to doctor, so patients and doctor can work together to adjust treatment and better manage patient's diabetes based on real, accurate information.

VII. CONCLUSION AND FUTURE WORK

We have successfully developed the desired system to predict diabetes. The main aim of the system was to detect diabetes at early stage. Our system helps the user to predict if they have diabetes or not by entering data asked by the system. We also observed that if glucose and blood pressure is high at that time the possibility of diabetes is high. Skin thickness is also considerably decisive in diabetes prediction. Machine learning is valuable in disease diagnosis. The capability to predict diabetes early assumes a vital role for the patients appropriate treatment procedure. In future, new parameters could be added to get more accurate and precise results.

ACKNOWLEDGMENT

It has been a great learning and experience for us throughout the whole time so far specially the main project. We weren't sure if we could pull this off as it was more challenging than we were prepared for. However, with the support and guidance of **Prof. Urjashree Patil**.

We were able to complete the project on time. She is really patience and helpful person we wouldn't be able to finish my project without her guidance and support we feel it wouldn't have been possible. As the project begun, we were in touch with guide. Sometime often as we were facing difficulties at some stages, we have also consulted **Prof. Reena Deshmukh**. We also express our deepest thanks to our H.O.D. **Dr. Uttara Gogate** who is benevolent helps us making available the computer facilities to us for our project in our laboratory and making it true success.

Lastly, we would like to thank our College Principal **Dr. P. R. Rodge** for encouraging us and giving moral support. We would also like to thank our colleagues who helped us directly or indirectly during our project.

REFERENCES

- [1]. Jingyu Xue , Research on Diabetes Prediction Method Based on Machine Learning. To cite this article: 2020 *J. Phys.: Conf. Ser.* **1684** 012062
- [2]. Mitushi Soni , Dr. Sunita Varma ,Diabetes Prediction using Machine Learning Techniques Dept of Computer Science and Engineering , International Journal of Engineering Research & Technology (IJERT) ,Vol. 9 Issue 09, September-2020 ,ISSN: 2278-0181 ,IJERTV9IS090496
- [3]. S. R. Priyanka Shetty , Sujata Joshi , A Tool for Diabetes Prediction and Monitoring Using Data Mining Technique ,I.J. Information Technology and Computer Science, 2016, 11, 26-32
- [4]. Huaping Zhou, Raushan Myrzashova* and Rui Zheng Zhou Diabetes prediction model based on an enhanced deep neural network, EURASIP Journal on Wireless Communications and Networking, (2020) 2020:148
- [5]. N. Sneha and Tarun Gangil ,Analysis of diabetes mellitus for early prediction using optimal features selection, <https://doi.org/10.1186/s40537-019-0175-6>, Sneha and Gangil *J Big Data (2019) 6:13*
- [6]. Dost Muhammad Khan1, Nawaz Mohamudally2, "An Integration of K-means and Decision Tree (ID3) towards a more Efficient Data Mining Algorithm ", Journal Of Computing, Volume 3, Issue 12, December 2011.
- [7]. K. Rajesh and V. Sangeetha, "Application of Data Mining Methods and Techniques for Diabetes Diagnosis", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.
- [8]. Humar Kahramanli and Novruz Allahverdi,"Design of a Hybrid System for the Diabetes and Heart Disease", Expert Systems with Applications: An International Journal, Volume 35 Issue 1-2, July, 2000.