Health Measure System for diabetic people

Khushboo Mantri
Dept of Computer and Information Science
Indiana University Purdue University, Indianapolis
kmantri@iu.edu

Samiksha Gaikwad
Dept of Computer and Information Science
Indiana University Purdue University, Indianapolis
ssgaikwa@iu.edu

***Abstract*—Nowadays, people with diabetes are one of the fastest-growing health issues all over the world. People with diabetes need to make healthy food choices, stay at, or get to a healthy weight, move more every day, and many other things. So, this project aims to develop a system to detect the health measure of people having people with diabetes considering factors like a person's diet, lifestyle, insulin, blood sugar levels, alcohol intake, etc. We try to predict the healthiness condition of the person considering all the above factors so that the person can improve his health by changing his lifestyle. Our system will keep track of patient's health in a timely manner and generate a alert when the patient's vital parameters crosses the normal value.**

I. INTRODUCTION

The project deals with collecting the nutritional information of the food that a person consumes in a day as well as the body movements done in the specific day and combining both these things to check the health conditions. Everyone has a different body type and everyone's health gets affected differently. Considering each individual's body type their health data is to be generated and the resultant output needs to be given. The statistics provided by the project will help many people suffering from diabetes or who have possibilities of facing high blood sugar level problems in the future to take the relative measures which will keep their blood sugar in a moderate range and thus can tackle the diabetes problem to a much more extent. This will also help in creating awareness among people about diabetes and inspire them to have a healthy lifestyle. No two people have the same body type, hence choosing a specific machine learning algorithm is important which could work with all the data and give comparatively better results. The body vitals contain the independent variables as they do not bias the other parameters those are the height, weight etc. Based on these parameters the actual processing has to be carried out.

At first we wanted to get to know about the various algorithms that can be used for the system so we carried out a test on sample dataset which gave us a precise algorithm to use and using that algorithm we developed a model which would predict the health of a person. For many healthcare applications, having a straight forward sensor-to-cloud architecture is not feasible, exclusively due to the certainty that most hospitals would not prefer patient data to be hoarded exterior [3].

Cloud based health monitoring system is used for a bold approach in the contemporary life. Exclusively old patients should be systematically monitored and their loved persons need to apprise about their health condition at that time. If system

disclose any abrupt shift in patient, then it signal's the user and their family member.

.

II. BACKGROUND

Nearly 26 million people in the United States currently have diabetics [1] Out of these, most of the diabetic people have people with type 2 diabetes; almost 1 million Americans have Type 1 diabetes [1] Type 1 diabetes is when your body does not make insulin, so You need to take insulin from outside sources to live. Type 2 diabetes is when our body does not use insulin well, which is produced by your body. Diabetes continues to be a growing health problem all over the world. Diabetes was the seventh leading cause of death in the U.S. in 2015[5]. Decades of intensive research have resulted in an extensively enhanced understanding of the pathophysiology and impact of diabetes, as well as a host of new and improved therapies. The lifestyle modification or treatment with metformin can delay the impact of developing diabetes by 58% [4]. Findings indicated that even modest weight loss could significantly minimize risk. [2] Also, a recent analysis of 10-year follow-up data shows that lifestyle intervention and metformin treatment were highly cost-effective treatments for patients with prediabetes.[5] Diabetes is a severe disease, but it can often be managed by diet, physical activity, and the appropriate use of insulin and other medications to control blood sugar levels. People who suffer from diabetes are at increased risk of severe health complications, including premature death, heart disease, vision loss, stroke, kidney failure, and amputation of toes, feet, or legs [2]. For many healthcare applications, having a simple sensor-to-cloud architecture is not viable, especially due to the fact that most hospitals would not prefer patient data to be stored outside [3].

III. PROPOSED SYSTEM

According to annual prevalence and incidence of diagnosed diabetes (type 1 and type 2 combined) for civilian, aged between 18 to 79 has the highest number of diabetic patients in the United States. [1] These days we have a lot of applications that track the person's diet and physical exercise routines and can show what the person is doing less or what improvements are needed in his diet and exercise. But we don't have many applications that are self-learning and can give active feedback when we enter the parameters that are needed. We decided to carry out the project in two parts:

1. Predict the algorithm which would best suit out data by comparing various methods like KNN and SVM.
2. To analyze the health of patients in a graphical manner for easy understanding and predict the health considering symptoms.

- **Part 1:**

For the purpose of looking into a best suitable algorithm we looked for KNN classifier, Naïve Bayes Classifier, Logistic Regression and SVM based on the study we carried out on these models we found out their advantages on our set of data. The comparison we carried out can be clearly shown in fig.1

A. KNN Classifier

In pattern recognition, the KNN algorithm could be a method for classifying objects supported closest training examples within the feature space. KNN could be a variety of instance-based learning or lazy learning where the function is barely approximated locally, and every one computation is deferred until classification. KNN is that the fundamental and most straightforward classification technique when there's little or no prior knowledge about the distribution of the info.[6]

B. Naïve Bayes classifier:

Naïve Bayes is super simple, we are just doing a bunch of counts. If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models like logistic regression, so you wish less training data. [2] And whether or not the NB assumption doesn't hold, an NB classifier still often does an excellent job in practice. A straight bet if want something fast and straightforward that performs pretty much. Its primary disadvantage is that it can't learn interactions between features (e.g., it can't learn that although you're keen on movies with Brad Pitt and Tom Cruise, you hate movies where they're together).[6]

C. Logistic Regression:

Logistic Regression provides lots of ways to regularize our model, and we do not must worry the maximum amount about your features being correlated as we are doing in Naive Bayes.[2] We furthermore may have a superb probabilistic interpretation, unlike decision trees or SVMs, and we'll be able to easily update your model to require in new data (using a web gradient descent method), again unlike decision trees or SVMs. Use it if you wish a probabilistic framework (e.g., to adjust classification thresholds, to mention when you're unsure, or to urge confidence intervals) or if you expect to receive more training data within the future that we want to be ready to quickly incorporate into our model.

D. Advantages of SVMs:

High accuracy, excellent theoretical guarantees regarding overfitting, and with an appropriate kernel, they'll work well whether or not you're data isn't linearly separable within the base feature space.[6] Uniquely accessible in text classification problems where very high-dimensional spaces are the norm. Memory-intensive, hard to interpret. with units.

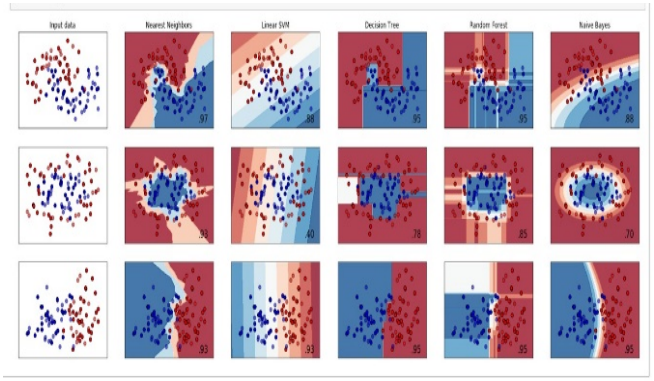

Fig 1. Results for the comparison of various algorithms on our dataset: The leftmost is the input data which is followed by the result formed various algorithms respectively(left to right input data , KNN Classifier, SVM, Decision Tree, Random Forest and Naïve Bayes)

- **Part 2:**

In the first part we saw the advantages of various classifiers which were better, we were planning to use the KNN classification or the Logistic Regression models as the target would be a binary one but after careful analysis we came across another classifier called the Gradient Boosting Classifier.

Why Gradient Boosting Classifier?

Several supervised machine learning models are founded on a single predictive model (i.e. linear regression, penalized models, naive Bayes, support vector machines). Alternatively, other approaches such as bagging and random forests are built on the idea of building an ensemble of models where each individual model predicts the outcome and then the ensemble simply averages the predicted values.[7] Boosting makes new models ensemble sequentially. It Often provides predictive accuracy that cannot be beat, has Lots of flexibility i.e. it can optimize on different loss functions and provides several hyper-parameter tuning options that make the function fit very

flexible, there's no data pre-processing required because it often works great with categorical and numerical values as is , also handles missing data whose imputation not required.[7]

Thus the steps involved in making our model work include:

1. Demonstrate how you imported, queried, and sorted your data
2. Parse: Identify any outliers, define variables
3. Mine: Perform statistical analysis, correlate data
4. Refine: Describe and plot your data
5. Model: Perform model (train subset as needed)
6. Model: Tune and evaluate the model
7. Discuss data (pipeline) processes - importing/sorting/querying/munging
8. Discuss statistical and visual analysis
9. Discuss model selection and implementation process
10. Interpret findings and relate to goals/criteria

IV. MODULES

Our system will be working on four modules which are as follows:

1. Human health data:
   The biological data of people is collected like their age, blood pressure, glucose count, etc. This will act as the dataset for training and testing the model.
2. Model implementation
   The model which we generated uses the gradient boosting algorithm
3. Cloud Storage
   Intelligent Data Storage. The user can access the cloud service to upload/download the health information via authentication. It can provide smart applications and send private health reports
4. Sending details on mobile device
   A portable computing device with a unique International Mobile Equipment Identity (IMEI) which can connect and send messages whenever the data is generated.

V. RESULTS

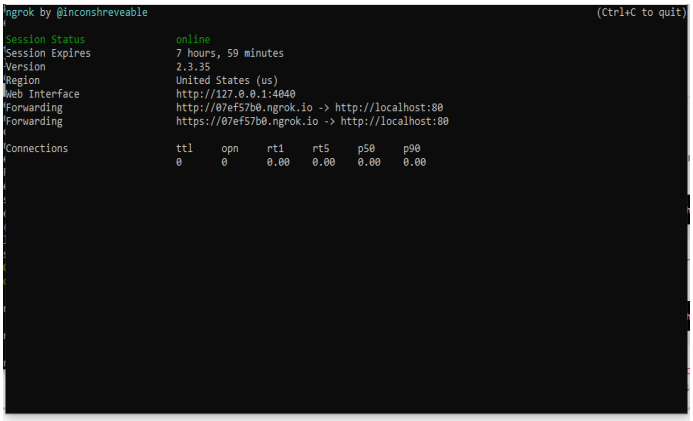

Fig 2: Starting ngrok for secure transfer of data

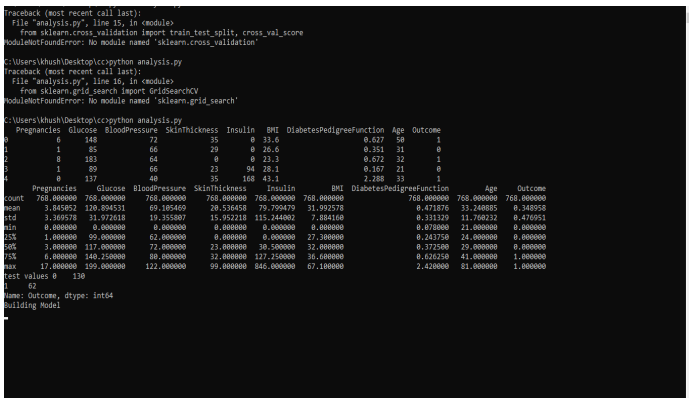

Fig 3. Model Executing

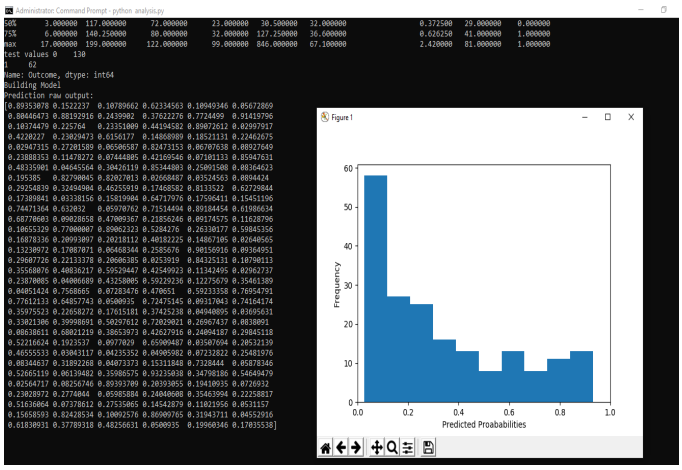

Fig 4. Result of the predictive System

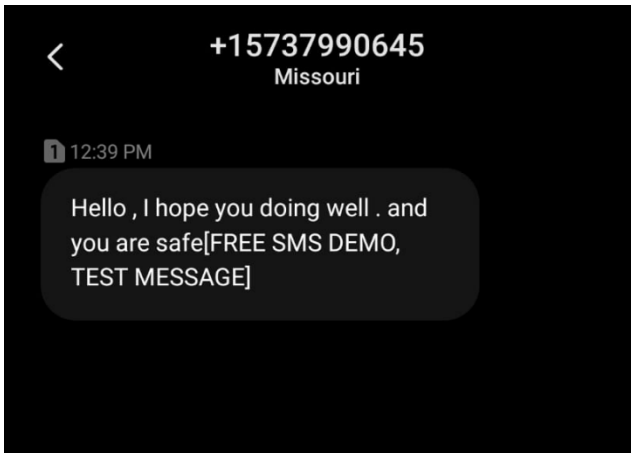

Fig 5 Message sent to the user of the system

VI. CONCLUSION

Health Monitoring is one of the most crucial things and proper care has to be taken. In this project we proposed a system based on cloud which monitors the health of people for the analysis of diabetes in them. We worked on various supervised machine learning algorithm to check which is best suitable to provide more accurate result. Once the results of prediction are calculated the results are sent via messaging to the contact provided by the user by secure transmission.

VII. REFERENCES

[1] American Diabetes Association (2015a). (5) Prevention or delay of type 2 diabetes. Diabetes Care 38 (Suppl), S31–S32.

[2] American Diabetes Association (2015b). (4) Foundations of care: education, nutrition, physical activity, smoking cessation, psychosocial care, and immunization. Diabetes Care 38 (Suppl), S20–S30.

[3] Z. Zhang and X. Hu, "ZigBee based wireless sensor networks and their use in medical and health care domain," in Proceedings of the 7th International Conference on Sensing Technology (ICST '13), pp. 756–761, Wellington, New Zealand, December 2013.

[4] Bailey, T.S., Ahmann, A., Brazg, R., Christiansen, M., Garg, S., Watkins, E., Welsh, J.B., and Lee, S.W. (2014). Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor. Diabetes Technol. Ther. 16, 277–283.

[5] Bansal, N. (2015). Prediabetes diagnosis and treatment:review. World J. Diabetes 6, 296–303.

[6] Min Chen ,Yixue Hao, Kai Hwang , Lu Wang , Lin Wang –IEEE"Disease Prediction by Machine Learning Over Big Data From Healthcare Communities"

[7] Samaher Al_Janabi , Hayder Fatlawi (2016) Integrated Prediction Model for Huge\Big Healthcare Database (5) –Research Gate