Predictive Model for Diabetes using Machine Learning Algorithms

Ananya V[1], Sushmitha N[2] and Byra Reddy[3]

[1,2] Department of ISE, RVCE
Bangalore, Karnataka, India

[3] Business Solution, Aarika Infosystems Pvtl Ltd
Bangalore, Karnataka, India

Abstract

In current trend of automation, Machine Learning is widely used for automation and predictive model designs, we tried to predict the patient is diabetic or not by considering various parameters and used various models to check the accuracy of the prediction. This paper presents various Machine Learning techniques to predict diabetes by reading the patient's data. Various algorithms have been employed to regulate proper detection. Algorithms like KNN, SVC, logistic regression, naïve Bayes and decision tree are being used to read and detect the patient's data.

Keywords: Machine Learning, Predictive Model, KNN, SVC.

1. Introduction

Diabetes is an ailment whereby (glucose) isn't processed in the body. This builds the glucose in the blood to alarmingly abnormal states. This is known by the name hyperglycemia. In this condition, body is unfit to create adequate insulin. The other plausibility is that body can't react to the delivered insulin.

Diabetes is hopeless; it must be controlled. A diabetic individual can create serious intricacies like nerve harm, heart assault, kidney disappointment and stroke. As indicated by insights in 2017, an expected 8.8% of worldwide populace has diabetes. This is probably going to increment to 9.9% constantly (Navoneel Chakrabarty et al., 2018). Hyperglycemia brought about by diabetes, make irregularities in the cardiovascular framework free of the conceivable nearness of dyslipidemia, blood vessel hypertension and so on. Diabetic Mellitus is a lot of related ailments wherein the human body is unfit to control the amount of sugar in the blood [2]. There it is expected to examinations the effectively accessible gigantic diabetic informational indexes to find some amazing certainties which may help in creating some expectation model. The center is to build up the forecast models by utilizing certain AI calculations [5]. AI accentuations on the improvement of PC programs that can instruct themselves to change and develop when uncovered to new or inconspicuous information[1,4].

2. Literature Review

Numerous organizations and research bunches are attempting to treat diabetes, however avoiding the illness will greatly affect wellbeing in danger gatherings[3]. Diabetes expectation makes an accuracy prescription way to deal with aversion of diabetes that steers endeavors towards the individuals who are at most astounding danger of building up the illness and who might profit most from medication treatment or preventive way of life methodologies[7]. There are plans to make this expectation framework, to order and anticipate kind of diabetes the patient might experience the ill effects of, i.e if an individual is experiencing type 1 diabetes or type 2 diabetes, etc[8,9]. The venture is very important in the present situation since it causes individuals to realize what is their odds of having diabetes and in like manner they can take estimates adjusting their way of life with the goal that they can lead solid life[10].

a. K- Nearest Neighbors Algorithm:

The k-Nearest neighbors (KNN) calculation is a basic, simple to-actualize directed AI calculation

that can be utilized to take care of both order and relapse issues. A directed AI calculation (rather than an unsupervised AI calculation) is one that depends on named input information to get familiar with a capacity that creates a fitting yield when given new unlabeled information[11].

b. SVC (Support Vector Classifier):

A Support Vector Machine (SVM) is a discriminative classifier formally described by a secluding hyperplane. Accordingly, given stamped getting ready data (oversaw learning), the figuring yields a perfect hyperplane which arranges new models. In two dimensional spaces this hyperplane is a line isolating a plane in two segments where in each class lay in either side. The most pertinent AI calculation for our concern is Linear SVC.

c. Logistic Regression:

Determined backslide is another technique gotten by AI from the field of experiences. Determined Function: Logistic backslide is named for the limit used at the focal point of the strategy, the key limit. The determined limit, in like manner called the sigmoid limit was made by experts to portray properties of masses advancement in science, rising quickly and amplifying at the passing on utmost of the earth.

3. Predictive Model Design

We emphasis on detailed description of individual module for implementation of Diabetes Prediction using Machine Learning.
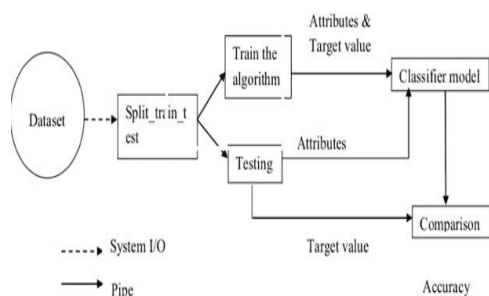
a. System Architecture

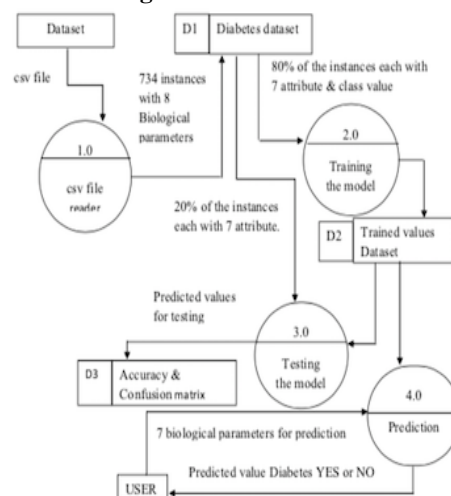Fig-1: System Architecture

b. Detailed Design

Fig-2: Detailed Design

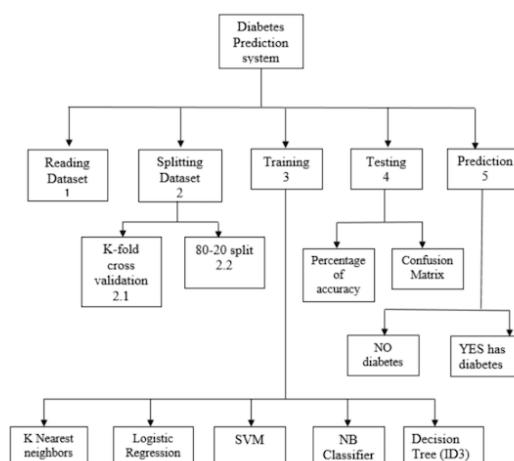### c. Structure chart of Diabetes prediction system

Fig-3: Predictive Model

Organized stream diagram gives in general methodology for organizing the program. It gives insights regarding every module advanced amid detail plan and aides in coding. The modules and their plan for the venture is appeared in the underneath figure - 3Data investigation is a strategy of inspecting, filtering, changing, and showing data with the goal of discovering profitable information, teaching closures, and supporting essential administration. Data examination has different perspectives and approachs, including grouped techniques under a collection of names, and is used in different business, science, and human science territories. In the present business

world, data examination accept an occupation in settling on decisions dynamically intelligent and helping associations work even more effectively.

4.Results and Discussion

The user information is accepted for evaluation.These include the parameters such as the name of the patient,their age,gender,glucose levels,blood pressure levels,insulin levels,BMI value and the diabetes pedigree function value.If the gender of the patient is given as female,the number of pregnancies undergone are taken into account as well.We take the information from the user and pass it to the function as a parameter and provide it to the trained model for classification.

a. Male patient with optimum levels of all parameters considered

```
Prediction Using User Data
name = input("Name Of The Patient : ").capitalize()
a = int(input("Age : " ))
gender = input("Gender : ").lower()
p = 0
if(a < 12):
    print(name ,"Your Not Eligible For This Test")
else:
    if(gender == "male"):
        p = 0
    else:
        p = float(input("number of pregnancies undergone :"))

    g = float(input("Glucose Level in mg/dL : "))
    b = float(input("Blood Pressure Level : "))
    i = float(input("Insulin Level : "))
    bm = float(input("BMI :"))
    d = float(input("Diabetes Pedigree Function : "))


Name Of The Patient : ANMOL
Age : 18
Gender : MALE
Glucose Level in mg/dL : 88
Blood Pressure Level : 44
Insulin Level : 66
BMI :44
Diabetes Pedigree Function : 66
```

Fig-4: Male Patient

b. Different Parameters

```
GLUCOSE LEVEL :
        less than Optimal level
        optimal glucose level :   90.0-125.0
        your glucose level :      88.0


BLOOD PRESSURE :
 Blood Pressure level is less than Optimal level
        optimal blood pressure level :  80.0-120.0
        your blood pressure level :       44.0


INSULIN :
        optimal insulin level :  50.0-276.0
        your insulin level :     66.0


BMI :
        more than Optimal level
        optimal BMI level :      18.0-276.0
        your BMI level :         44.0
```
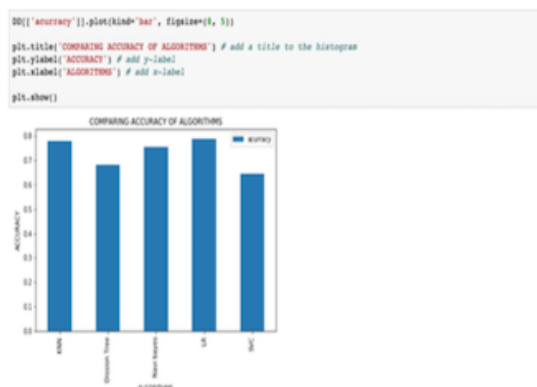
Fig-2: Different Parameters

c. Comparative Analysis with Different Algorithms

	Cross validation	R^2 value	acurracy
KNN	0.767003	0.701823	0.780669
Dission Tree	0.703110	0.776042	0.680297
Navi bayes	0.757809	0.751302	0.754647
LR	0.770882	0.776042	0.788104
SVC	0.651059	0.751302	0.644981

d. Graphical Representation

```
DD[['acurracy']].plot(kind='bar', figsize=(8, 5))

plt.title('COMPARING ACCURACY OF ALGORITHMS') # add a title to the histogram
plt.ylabel('ACCURACY') # add y-label
plt.xlabel('ALGORITHMS') # add x-label

plt.show()
```

6. Conclusions

The diabetes prediction system is developed using five machine supervised/classification modelling techniques. These models are trained and validated against a test dataset w.r.t accuracy score, cross-validation and r-square value. All five models are able to extract patterns in response to the predictable states. Although not the most effective model, the Logistic regression result is easier to read and interpret, what is more, the training over Logistic regression is very efficient. It is certain that machine learning is being integrated into the medical profession in more ways than we can think. From accurate diagnosis to finding better treatments and suggesting cost effective ways to cure the illness, the emerging technology have become a go-to solution for medical needs across the globe.

Now, as we speak of machine learning and artificial intelligence being introduced in medical practices, the objective of this project is not to replace the medical professionals but enhance the doctor's medical expertise and scale it to unprecedented levels. As the medical field is getting flooded with massive amounts of data from hospital records, patient history and others, the relevance of machine learning is only obvious. As these technologies take a leap from assisting doctors in mundane tasks to suggesting solutions and remedies, Machine learning and AI are definitely painting a larger picture when it comes to medical science.

Acknowledgments

References

[1]. Berina Alić ; Lejla Gurbeta ; Almir Badnjević, Machine learning techniques for classification of diabetes and cardiovascular diseases, 2017 6th Mediterranean Conference on Embedded Computing (MECO), Year: 2017,Pages: 1 - 4

[2]. Navoneel Chakrabarty, A Deep Learning Method for the detection of Diabetic Retinopathy, 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2-4 Nov. 2018, **DOI:** 10.1109/UPCON.2018.8596839

[3]. A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management Kevin Plis, Razvan Bunescu, Cindy Marling, Jay Shubrook, Frank Schwartz.

[2]. Improved J48 Classification Algorithm for the Prediction of Diabetes Gaganjot Kaur, Amit Chabra,Department of Computer Science and Engineering GNDU, Amritsar (Pb.), India.

[3]. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes Wei YuEmail author, Tiebin Liu, Rodolfo Valdez, Marta Gwinn and Muin J Khoury,BMC Medical Informatics and Decision Making (2010).

[4]. Hypoglycemia Prediction Using Machine Learning Models for Patients With Type 2 Diabetes Bharath Sudharsan, MS, Malinda Peeples, RN, MS, CDE, Mansur Shomali, MD, CM.

[5]. Type 2 Diabetes Risk Forecasting from EMR Data using Machine Learning -Subramani Mani, MBBS PhD, Yukun Chen, MS, Tom Elasy, MD MPH, Warren Clayton, MD, and Joshua Denny, MD MS.

[6]. Hybrid prediction model for Type-2 diabetic patients B.M.PatilR.C.Joshi Durga Toshniwal.

[7]. The use of the area under the ROC curve in the evaluation of machine learning algorithms,Andrew P.Bradley.

[8]. World Health Organization, "Report of a study group: Diabetes Mellitus", World Health Organization Technical Report Series, pp. 727, 1985.

[9]. Kemal Polat, Salih Gunes, Ahmet Arslan, "A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine", Expert Systems with Applications, vol. 34, no. 1, pp. 482-487, January. 2008.

[10]. B. J. Lee, B. Ku, J. Nam, D. D. Pham, J. Y. Kim, "Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes", IEEE J. Biomed. Health Informat., vol. 18, no. 2, pp. 555-561, Mar. 2014.

[11]. Manaswini Pradhan, Dr. Ranjit Kumar Sahu, " Predict the onset of diabetes disease using Artificial Neural Network (ANN) " International Journal of Computer Science &Emerging Technologies (E-ISSN: 2044-6004) 303 Volume 2, Issue 2, April 2011.