International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 792-794 www.ijaem.net ISSN: 2395-5252

Diabetes Prediction Using Machine Learning

Rahul Ranjan, Mr Kuldeep B.Vayadande, Rahul Kumar Sharma
Master of Computer Application, Jain(Deemed to be University), Bangalore
Assistant Professor, Jain(Deemed to be University), Bangalore
Master of Computer Application, Jain(Deemed to be University), Bangalore

Date of Submission: 15-11-2020 Date of Acceptance: 30-11-2020

ABSTRACT - Diabetes is a common disease and lots of people are suffering from this disease. There are various factors which result in the chances of having such disease like Age, obesity, lack of exercise, living style, bad diet, high blood pressure, etc. People having diabetes has various risks of diseases like eye problem, heart disease, kidney disease, stroke, etc. The big challenge for the health care industries nowadays is to give a more precise result which could easily predict whether a patient is having or diagnosed with such disease. Many AI and Machine learning models are being considered by various health care industries for better prediction. The only difference occurs in the accuracy of predicting the result. Models like K-NN, SVM, and other Classification models are mostly in use
.For this model I would be working on Jupyter notebook using a python programming language which has many inbuilt libraries like NumPy, pandas, matplotlib, etc.

KEYWORDS – Risk of Diseases, Machine learning models, Classification model, python, libraries, KNN, Feature Scaling

I. INTRODUCTION

Technology is providing us many advanced platforms such as Machine Learning, Data Mining, IoT, etc to satisfy the need of society, almost in every field of work we use this technology. It has many real-time applications such as the Internet of things, Artificial intelligence, cloud computing, etc. In Medical Science the use of such technology is increasing day by day to provide better service to the patients in a small duration of time. Considering the current scenarios, In many countries, Diabetes has become a very severe disease. According to 2017 statistics, around 425 million people suffer from this Mellitus disease.

Every year 1-2 million peoples are losing their lives due to diabetes. And by 2045 it may rise to 600 million. So the objective of this project is to build a decision support system to predict and diagnose a certain disease with an extreme amount of precision.

There is various classification model which have different accuracy of prediction. In my project, I will work on some of them and analyze which model would be better for this problem. Diabetes is one of the most widely spread diseases and every year Millions of people are getting affected. It is caused by an increase in the blood sugar level.[1] Also, it has various effects on different parts of the body which results in hypertension, Heart alignment, eye issues, and so on.[2] As per the WHO report on 14th November 2016, 422 million adults are affected by this chronic disease and around 1.6 million people have lost their lives.
Many countries and organizations are worried about this chronic disease for achieving control and preventing it from being spread.

II. LITERATURE REVIEW

In machine learning, one can select any model based on the desired problem or objective which helps in making better insights. For a given data one may come with different results as models having numerous ways of handling the problem. .
[1] In 2015 a model was proposed on classification technique to study the find the hidden patterns in the dataset. Navies Bayes and Decision Tress were used. Both the models were compared and the effectiveness of both Algorithms was shown. .
[2] Shraddha Kumar and Mani Butwall (2015) proposed a Model using a Random Forest classifier to predict diabetes behavior.
[3] K- Nearest Neighbors is an example of instance- based learning in which training data set are stored, so that classification for any new unclassified record may be easily found by comparing it to the most similar record in the training set.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 792-794 www.ijaem.net ISSN: 2395-5252

Authors	Methodologies	Results
Weifeng Xu	Adabost and RF	Better accuracy by RF and less with AdaBoost
Pradeep et	RF , KNN ,Decision Tree,SVM	Before Preprocessing, g Decision tree shows an accuracy of 79%. RF and KNN were also similar.
Wei et al	SVM	Sex and Physical activity are considerable attributes.
A. Negi, V. Jaiswal	Different global dataset	Differ in accuracy.

III. EXISTING SYSTEM

In various research articles, a lot of systems were proposed for predicting such disease. Like using data mining techniques the researcher used to collect thousands of data and using some mining algorithm they used to come up with some meaningful insights. Also using various classification regression neural networks they try to overcome the missing values and outliers which help them in increasing the Accuracy of the Model [4]. One of the models was based on soft computing which determines the risk level of having such diabetes [5]. [6] A model by Andrew Person the predictive analysis was categorized into three forms mostly the operational management, medical management, and biometric design. According to him, the predictive analysis can help in addressing the frequency of a patient being admitted for chronic diseases. [7]B. Rodriguez Sanchez in his model predicts the relation between diabetes disease and the activities through which one got affected. According to him, this disease affects the perception of people regarding the effect of their condition and mostly the people of age between 50-65 years are suffering from such disease.

IV. PROPOSED SYSTEM

In my project, I would be working on a simple Machine learning classification model. And using the model I could train my model using the data which consist of various attributes related to a diabetes patient like Age, Sex, Blood Pressure, Insulin, Skin Thickness, BMI, etc. and based on these attributes I would predict the result for a patient whether he is suffering from diabetes or not. The KNN classification model best fits this problem.

The term KNN stands for K-Nearest Neighbors and this model was developed by Thomas Cover which is mostly used for classification and regression problems. The task of this algorithm is to assign a class or group to a new data point based on its nearest neighbors which are calculated by using various distance formula techniques like (Euclidean distance, Manhattan Distance, cosine distance, etc). Here I have used a simple Euclidean distance formula which is a distance formula between two points. This algorithm is also termed a lazy learner algorithm because all the operations are done without any training operations. The idea of this algorithm is to create a set of group or clusters which are near to the original data point and accordingly assign the new data with a similar group.

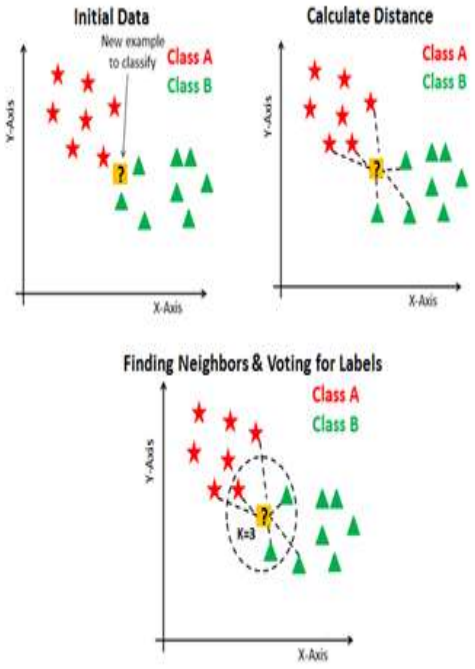

Fig 1

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 792-794 www.ijaem.net ISSN: 2395-5252

- Select the number of K of the Neighbors in the dataset.
- Calculate the distance from each neighbor using Euclidean distance.
- Take the K nearest neighbors as per the calculated Euclidean distance.
- Among this k-neighbors count the no of the data points in each category
- Assign the new data point to that category for which the number of the neighbors is maximum.

Finally, the Accuracy of my model comes close to 80 % and for any new patient it could easily predict whether the patient is having diabetes or not

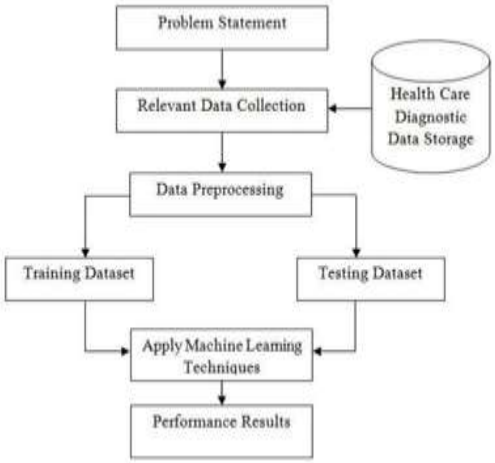

Fig 2

VI. CONCLUSION

Machine Learning plays an important role in various fields such as Healthcare, Stocks & Marketing, Banking, Weather Forecast, and so on. With the help of KNN Algorithms, it becomes easy to evaluate and fetch meaningful information from them. In KNN by using the various K- values of the K-NN classifier the accuracy of the model increases simultaneously, this study aims to accurately predict whether a given patient is suffering from diabetes or not. Finally, the Accuracy of my model comes close to 80 % and for any new patient, it could easily predict whether the patient is having diabetes or not.

REFERENCES

[1]. Utilization of data mining techniques for diagnosis of diabetes, ThirumalChandran , N.Mala Nagarajan.
[2]. Mani Butwall and Shraddha Kumar. Article:A Data Mining Approach for the Diagnosis of
[3]. Diabetes Mellitus using Random Forest Classifier. International Journal of Computer Applications120(8):36- 39, June 2015.
[4]. Karegowda , A.G Jayaram , M . A and Manjunath , A.S., 2012. Cascading k-mean clustering and K- nearest neighbors classifier for categorization of diabetic patients. Internal Journal of Engineering and Advance technology,1(3), pp.147-151
[5]. Abdullah A, Aljumah, Mohammed GutamAhmad 2012 Diabetes health care in young and old patient, journal of king Saud university, computer, and information science 25:127-136
[6]. Cichosz. 10 CGM and modulation autonomic of information Combining detection improves and prediction Enables Measurements Technol SciDiabetes J. event hypoglycemic spontaneous of .137-132
[7]. for analytics Predictive 2012 Asia Qualex, Pearson Andrew. 1 for the health industry
[8]. Rodriguez –Sanchez B, Alessie RJM, Feenstra TL ,2017 the relationship between diabetes-related compilation and productive activities among older Europeans.
[9]. Fig 1 https://medium.com/@pradeep.dhote9/k-nn-k- nearest-neighbors- 3f34c60d5f2e
[10]. Fig 2 https://www.semanticscholar.org/pa
[11]. per/Performance- Analysis-of- Machine-Learning- Techniques- Faruque- Asa duzzaman/ 0a 164c4528b9e966d 7 d76169473c7b97fd3468a6

- Euclidean distance can be calculated as:

Sqrt((x2-x1)^2+(y2-y1)^2)

V. E-R DIAGRAM

In KNN by using the various K- values of the K-NN classifier the accuracy of the model increases simultaneously, this study aims to accurately predict whether a given patient is suffering from diabetes or not.