e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 5, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.54

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

Diabetes Prediction Using Machine Learning

S. Narendra, N. Durga Deepika, S. Pavithra, P. Susmitha, P. Chetankumar

Assistant Professor, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Andhrapradesh, India

Student, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Andhrapradesh, India

Student, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Andhrapradesh, India

Student, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Andhrapradesh, India

Student, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Andhrapradesh, India

ABSTRACT*:* The diabetes is one of terminal diseases in the world. It is additional an inventor of various varieties of disorders for example: coronary failure, blindness, urinary organ diseases etc. In such conditions the patient is required to visit a hospital or diagnostic centre, for check-up to get reports. They need to invest time and money every time because of this traditional method. But the growth of Machine Learning algorithm finds a solution for this conventional issue, we have advanced system mistreatment information processing that could forecast whether the patient has polygenic illness or not. Furthermore, forecasting the sickness initially ends up in providing the patients before it begins vital. The aim of this analysis is to develop a model which might predict the diabetic risk level of a patient with a better accuracy. The model can be developed by using machine learning algorithms like Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, XG Boost. Particularly we are focusing more on KNN to develop our model.

KEYWORDS: Diabetes, Decision trees, Support vector machines, Diseases, Machine learning algorithms, Classification algorithms, Training

I. INTRODUCTION

Diabetes is the chronic disease will be long lasting till the end of life. Diabetes occurs when glucose levels in blood are too high and it is also called Blood sugar. It effects the mechanism of body to turn food into energy. The main cause of diabetes is overweight, obesity and inactive lifestyle. There are three types of diabetes like Type-1 diabetes, Type-2 diabetes, and Gestational diabetes. There is also special type in diabetes called Prediabetes occurs when glucose level in blood is higher than normal but less than the level for type-2 diabetes. There are various symptoms used to detect diabetes disease like weight loss, blurry vision, sores, frequent urination, extreme fatigue, increase in thirst and hunger.

II. RELATED WORK

K. Vijayakumar et al. developed a system for diabetes Prediction using random Forest algorithm which might perform early prediction of diabetes for a patient with a higher accuracy by using Random Forest algorithm in machine learning technique. The proposed model gives the best results for diabetic prediction and the result showed that the diabetes prediction system can predict the diabetes disease effectively and efficiently. NonsoNnamoko et al. presented predicting diabetes onset: they used an ensemble supervised learning approach which has five widely used classifiers are used for the ensembles and a meta-classifier is used to combine their outputs. The results are compared with similar studies which used the same dataset within the literature. It resulted as by using the proposed method, diabetes onset prediction can be done with higher accuracy. Tejas N. Joshi proposed Diabetes Prediction Using Machine Learning Techniques for diabetes prediction using three different machine learning algorithms like: SVM, Logistic regression, and ANN. This system proposed an effective technique for earlier detection of the diabetes. Deeraj Shetty proposed diabetes disease prediction using data mining that gives analysis of diabetes

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

malady utilizing diabetes patient's database. In this system, they the used machine learning algorithms like Bayesian and KNN to apply on diabetes patient's database and analyse them by taking various attributes of diabetes for prediction of diabetes disease. Muhammad AzeemSarwar proposed study on prediction of diabetes using machine learning algorithms in healthcare they applied six different machine learning algorithms performance and accuracy of the applied algorithms is discussed and compared. In this study different machine learning algorithms are compared and found which algorithm is best suited for prediction of diabetes. Based on previous research work, it has been observed that the classification process is not much improved. Hence a system is required for Diabetes Prediction to handle the issues identified in the previous research work.

III. PROPOSED METHODOLOGY

A. Dataset Description:

The dataset used to develop our machine learning model is Pima Indian Diabetes Dataset gathered from UCI repository. The dataset consisting of information about 768 patients.

S.NO	ATTRIBUTES
1.	Pregnancies
2.	Glucose
3.	Blood Pressure
4.	Insulin
5.	Body Mass Index
6.	Diabetes pedigree function
7.	Age
8.	outcome

Table 1: Dataset Description

The value of outcome is either 0 or 1 indicates non-Diabetic person and Diabetic person, respectively.

Distribution of Diabetic and Non-Diabetic Patients:

We are developing a model based on the dataset collected from UCI repository named Pima Indian Diabetes. The dataset consists of information about 500 non diabetic patients and 268 diabetic patients representing outcome as 0 and 1, respectively. The model generates accurate results only when the dataset is balanced. Here in our case the dataset is slightly unbalanced so that it would not affect our results.

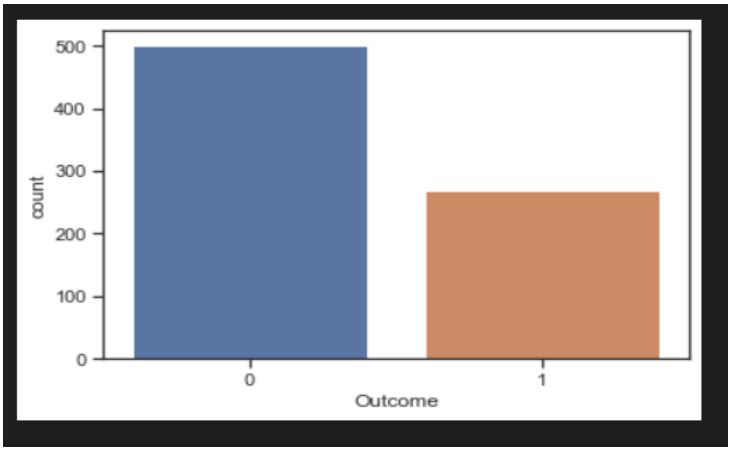

Figure 1: Diabetic and Non-Diabetic Patient Ratio

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

B. Data Pre-processing:

Data pre-processing plays a significant role when it comes to handling enormous data. Data related to medical field consists of missing values and some other inconsistent data which will affect the performance of the model.
So, after we are done with data mining, we will perform data pre-processing. To make sure that the machine learning models will give accurate results with more concise predictions, applying this process is mandatory.
we are required to perform data pre-processing in 2 steps for Pima Indian diabetes dataset.

1). **Removal of missing values**- In this step, we will try to remove all those instances which have zero as value. Instances with zero as value is not possible. So those kinds of instances are eliminated. Also, we make feature subset by removing the features or instances that are not relevant. This procedure is termed as features subset selection, which works faster by reducing dimensionality of the data.

2). **Data splitting**- When we are done with cleaning the data, then the data will undergo normalization for training and testing the model. The entire data is divided into 2 parts: One for training the model and the other for testing the model. We will start with training data by keeping testing data aside. The main aim of normalization is to bring all the attributes to same level or same scale. A training model will be produced under this training process based on some algorithms, logic and by considering the values of the features in the training dataset.

C. Applying Machine Learning:

We apply machine learning techniques when the data is ready for data pre-processing. We can use many different classification and ensemble techniques to predict diabetes disease. To develop our model, we used K- Nearest neighbour Machine Learning algorithm. It is one of the simplest supervised machine learning algorithms. KNN algorithm draws the similarity between existing data and new data and put the new data into one of existing categories that most suits it. It is non parametric and called lazy learner algorithm because it does not learn immediately from training dataset. It performs action on the dataset at the time of classification. The main objective of applying machine learning algorithm is to improve analysing and performance of model for accurate results in prediction of disease.

K-Nearest Neighbour Algorithm:

```
m = [KNN ()]
model = m;
model. fit ();
model.predict( );
```

Step 1: Consider sample dataset of rows and columns as Pima Indians dataset.
Step 2: Consider test dataset consisting of attributes and rows.
Step 3: Calculate Euclidean distance with the help of formula
Step 4: Find the K value where is the number of nearest neighbours. We used Hyperparameter optimization technique to find out accurate value of K.

```
error= []
fromsklearn.neighbors import KNeighborsClassifier
#Calculating error for K values between 1 and 25
fori in range (1, 25):
   knn=KNeighborsClassifier(n_neighbors=i)
   knn.fit(x_train,y_train)
   pred_i=knn.predict(x_test)
   error.append(np.mean(pred_i != y_test))
```

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

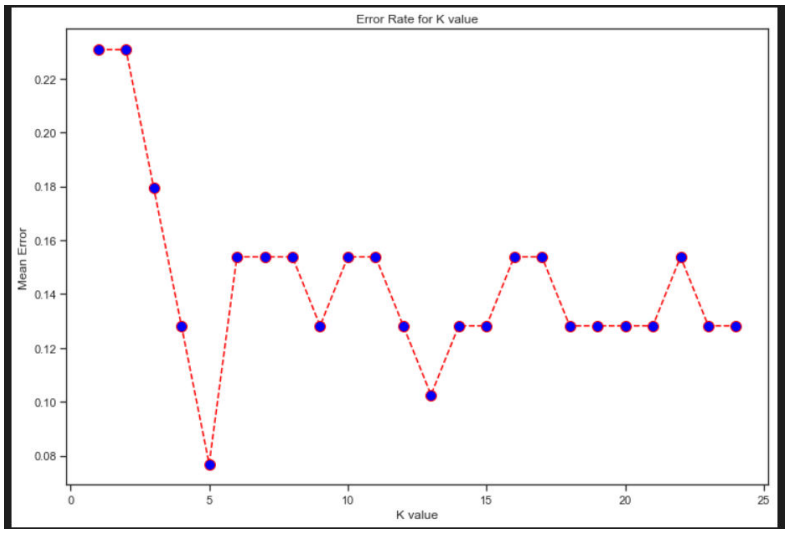

Figure 2: Error rate for K value

Step-5: Among these k neighbours, count the number of the data points in each category.
Step-6: Assign the new data points to that category for which the number of the neighbour is maximum.
Step-7: Our model is ready.

```
#Model Building using KNN
fromsklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=5, metric='minkowski')
knn.fit(x_train, y_train)
```

IV. MODEL BUILDING

Model Building is the most important phase in the prediction of diabetes using Machine learning. Here we implemented KNN algorithm discussed above for prediction of diabetes.

Procedure of Proposed Methodology:
Step 1: Import Libraries and dataset.
Step 2: Remove missing values from data pre-processing
Step 3: Split 95% of data for training and 5% for testing for better accuracy of the model.
Step 4: Select KNN Machine learning algorithm.
Step 5: For the mentioned machine learning algorithm build the classifier model based on the training dataset.
Step 6: For the mentioned machine learning algorithm test the classifier model based on the test set.
Step 7: Perform comparison between experimental performance results obtained for each classifier.
Step 8: Analyse the performance of the model based on various measures.

V. RESULTS AND DISCUSSION

By applying K- Nearest Neighbour Algorithm our model secured an accuracy of 92.3% in the prediction of diabetes disease. The accuracy of the model depends upon K value used while training the model using KNN algorithm and on the percentage of data used for training and testing. The data in the dataset must be balanced to generate accurate results. The following are the cases observed while developing the model for diabetes prediction using machine learning.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

K-Value	Training Data	Testing Data	Accuracy
11	65%	35%	76.2%
12	65%	35%	78%
13	75%	25%	79.6%
20	90%	10%	80.5%
12	70%	30%	80%
15	80%	20%	83.1%
5	95%	5%	92.3%

Table 2: Observations

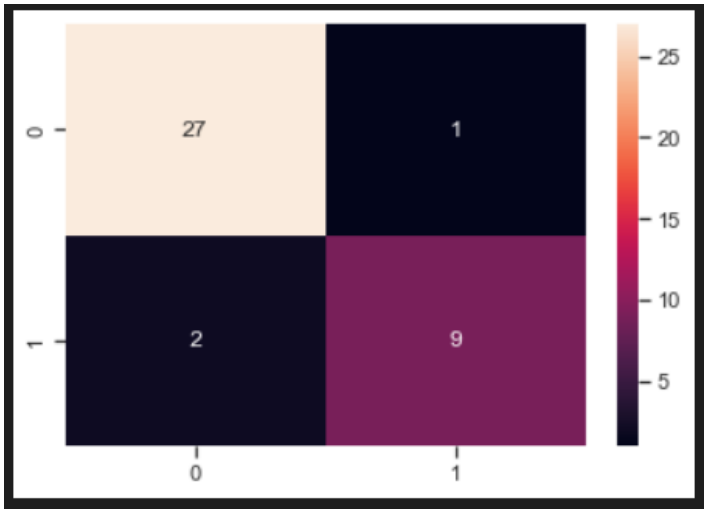

Figure 3: Confusion Matrix

VI. CONCLUSION

The main idea of the project was to design and execute Diabetics Prediction using K- Nearest Neighbour machine learning Algorithm and to check the performance Analysis of it and has been achieved successfully. The proposed approach uses the best supervised machine learning algorithm which is K- Nearest Neighbouring algorithm and we acquired accuracy of 92.3%. The experimental results can be used in health care to prognosticate early and make opinions to detect diabetes in the early stage. Diabetes diseases when aggravated become beyond control. Diabetes diseases are complicated and take away lots of lives every year. When the early symptoms of diabetes diseases are ignored, the patient might end up with drastic consequences. The Diabetes disease can be kept under control if it is detected early.

REFERENCES

1. Kumari, Sonu, and Archana Singh .2013. A data mining approach for the diagnosis of diabetes mellitus. Intelligent Systems and Control (ISCO), 7th International Conference on. IEEE.
2. T. Jayalakshmi and Dr. A. Santhakumaran, 2010. A Novel Approach for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks," International Conference on Data Storage and Data Engineering,159-163.
3.] Perveen, S., Shahbaz, M., Guergachi, A., &Keshavjee, K. 2016. Performance analysis of data mining classification techniques to predict diabetes. Procedia Computer Science, 82, 115-121.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506020 |

4. S. sa'di, A. Maleki, R. Hashemi, Z. Panbechi, and K. Chalabi.2015. Comparison of Datamining Algorithms in the Diagnosis of Type II Diabetes, IJCSA, vol. 5, no. 5.
5. Prema N S, Varshith V, Yogeswar J.2019. Prediction of Diabetes using Ensemble Techniques from International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878,
6. J. Pradeep Kandhasamy, S. Balamurali .2015. Performance Analysis of Classifier Models to Predict Diabetes Mellitus, Procedia Computer Science
7. Sonali Vyas, Rajeev Ranjan, Navdeep Singh, Arohan Mathur.2019. Review of Predictive Analysis Techniques for Analysis Diabetes Risk.
8. Hang Lai1, Huaxiong Huang, Karim Keshavjee, Aziz Guergachi1 and Xin Gao.2019. Predictive models for diabetes mellitus using machine learning techniques, BMC Endocrine Disorders Article.
9. Md. Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus, IEEE
10. Prema N S, Varshith V, Yogeswar J.2019. Prediction of Diabetes using Ensemble Techniques from International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878

Impact Factor
7.54

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com