

# AI-Based Diabetes Prediction System

Presented by

A.Afroze  
V.Kalpana  
S.Karkuzhali  
S.keerthana

# ABSTRACT

- *Diabetes is a very common disease affecting individuals worldwide. Diabetes increases the risk of long-term complications including heart disease, and kidney failure among others. People might live longer and lead healthier lives if this disease is detected early.*
- *Different supervised machine learning models trained with appropriate datasets can aid in diagnosing the diabetes at the primary stage. The goal of this work is to find effective machine-learning-based classifier models for detecting diabetes in individuals utilizing clinical data. The machine learning algorithms to be trained with several datasets in this article include Decision tree (DT), Naive Bayes (NB), k-nearest neighbor (KNN), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR) and Support Vector Machine (SVM).*
- *Further, using various feature selection approaches, we have identified and prioritized a number of risk factors. Extensive experiments have been conducted to analyze the performance of the model using two different datasets.*

# CONTENT

- *The disease “Diabetes Mellitus” is one of the most common critical diseases in the world. According to the World Health Organization (WHO), diabetes affects 8.5% of adults over the age of 18 and is responsible for 1.6 million deaths worldwide.*
- *Although the rate of diabetes-related premature death in many developing countries fell from 2000 to 2010, the statistics again increased between 2010 and 2016. The four primary diseases, namely cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes, kill over 18% of people worldwide and have become a serious public health concern. For example, in 2000, deaths from diabetes climbed by 70%, and in 2020, mortality among males are expected to grow to 80%. Diabetes mellitus can result from obesity, age, lack of exercise, lifestyle, hereditary diabetes, high blood pressure, poor diet, etc. Over time, people with diabetes have a high risk of diseases such as heart disease, stroke, kidney failure, nerve damage, eye issues, etc*

# LITRATURE SURVEY

- *In this section, some closely related works are discussed briefly. In most of the reseach works, Pima Indians Diabetes Dataset (PIDDD) have been used by many researchers for diabetes prediction. Various supervised machine Learning algorithms were used to predict diabetes .*
- *Radial basis function (RBF) kernel SVM artificial newral network (ANN), multifactor dimensionality reduction (MDR), linear SVM and k-NN are some of them to mention. Based on p value and odds ratio (OR), Logistic Regression(LR) has been used to recognize the risk factors for diabetes .*
- *Four classifiers have been adopted to predict diabetic patients, such as*
  - *NB, DT, Adaboost, and RF. Partition protocols like-K2, K5, and K10 were also adopted, repeating these protocols into 20 trails. For the performance measurement of the accuracy (ACC) and area under the curve (AUC) were analyzed.*

# METHODOLOGY

- Firstly, we pre-process two separate datasets. In the pre-processing stage, correlation between attributes of the datasets is analyzed for finding useful features in detecting diabetes. After that, the data is divided into two sets: training and testing. The training set is utilized to develop predictive ML models using a variety of machine learning algorithms. Next, we assess the proposal's performance with respect to different metrics. Finally, the best ML model is deployed in a web application using flask. Following this, we describe the workflow of each part briefly:
- *Outliers Removal - Attributes' values that are beyond acceptable boundaries and have high variation from the rest of the respective attribute's value might be present in the dataset. Such attributes' value might degrade the machine learning algorithm's performance. To eliminate such outliers, we applied the IQR (Inter-quartile Range) approach.*

# Adopted machine learning algorithms

- *In this section, we will describe various machine learning algorithms that are used in the predictive model.*
  - *Naive Bayes - The Naive Bayes classified model is designed based on Bayes' Theorem. The assumption of independence among predictors is considered in this model. Naive Bayes can classify a given problem instance using a conditional probability model. The probability of an entity having  $x=(x_1, x_2, x_3, \dots x_n)$ ,  $n$  number of features (independent variables) is calculated as  $P(C_k|x_1, x_2, \dots x_n)$  for each of  $K$  possible outcomes or classes  $C_k$ .*
- *K-Nearest Neighbor- K-NN is one of the mostly used basic Machine Learning algorithms based on the Supervised Learning methodology. K-NN is used for both Regression and mostly classification . The K-NN algorithm considers the similarity between the new case/data and the existing case/data. The new case is then assigned to the category that is the most similar among the available options.*



# Experimental results analysis

- *Our proposed model is tested and evaluated in this section using a variety of machine learning algorithms, including NB, DT, RF, SVM, LR, GB, and K-NN. To find the effectiveness we have used 4 different datasets and each of them contains different types and number of attributes.*
- *The proposed model is built in Python and executes on a computer having an Intel Core i7 processor with a 4 GB graphics card, 16GB RAM and a 64-bit Windows operating system running at 1.80 GHz. To test the efficiency of our model, we have used a 10-fold cross validation process. The dataset is shuffled and divided into 10 segments at random, with one segment serving as the test set and the others serving as the training set in turn. The average of the results from multiple experiments is considered as the final output of the experiment.*

# CONCLUSION

- *In our research, firstly, we have adopted several machine learning algorithms and evaluated their performances to predict the diabetes of individuals. Secondly, we have conducted several experiments and evaluated the performances of the proposal. We found that SVM outperforms the other algorithms. Finally, based on our observed results, a smart web application is developed for predicting the diabetes accordingly. Any individual can submit clinical data to this web application, which can then forecast the existence or absence of diabetes. Individuals who are unsure or simply want a routine checkup may consider this application. Our model is compared with two recent studies, and the findings reveal that, depending on the dataset and the ML method used, the suggested model can offer greater accuracy ranging from 2.71% to 13.13%.*





THANK YOU ....