

DATE	18 NOVEMBER 2022
TEAM ID	PNT2022TMID19223
PROJECT NAME	A Gesture-based Tool for Sterile Browsing of Radiology Images
MARK	2 MARKS

PROPOSED SOLUTION

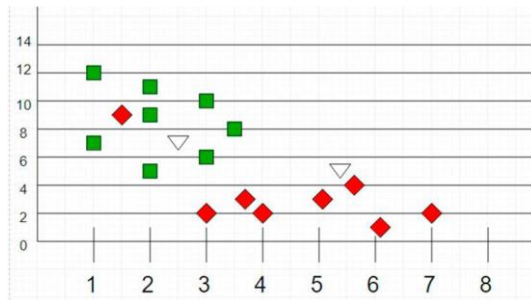
Abstract

Over 30 million people in India are suffering from diabetes and many others are under the risk. Thus, early diagnosis and treatment is required to prevent diabetes and its associated health problems. This study aims to assess the risk of diabetes among individuals based on their lifestyle and family background. The risk of diabetes was predicted using different machine learning algorithms as these algorithms are highly accurate which is very much required in the health profession. Once the model will be trained with good accuracy, then individuals can self-assess the risk of diabetes. In order to conduct the experiment, 952 instances have been collected through an online and offline questionnaire including 18 questions related to health, lifestyle and family background. The same algorithms were also applied to the Pima Indian Diabetes database. The performance of Random Forest Classifier is found to be most accurate for both datasets.

Where, y is the output which is the result of weighted sum of input variables x . If the output is more than 0.5, the output is 1 else the output is 0.

K- Nearest Neighbor Classifier

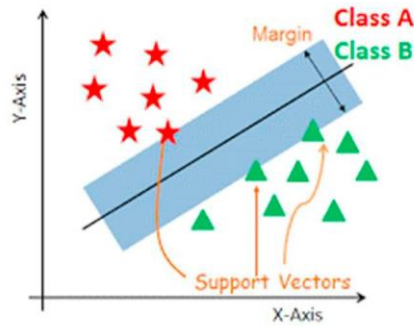
K-Nearest Neighbor (KNN) method can be used to solve problems pertaining to both regression as well as classification, though it is generally being used to solve classification problems in business. Its major advantage is simplicity of translation and low computation time. In figure 1, the points (2.5, 7) and (5.5, 4.5) will be allocated to any one of the clusters. The KNN uses Euclidean distance function to find distances between existing data points and any new data point. Thus, (2.5, 7) will belong to the green cluster, whereas, (5.5, 4.5) will belong to the red cluster .



KNN example

Support Vector Machine (SVM)

SVM is a supervised classifier in machine learning algorithms that can be used both for regression and classification. It is majorly applied in solving classification problems. The goal of SVM is to classify data points by an appropriate hyperplane in a multidimensional space. A hyperplane is decision boundary to classify data points. The hyperplane classifies the data points with maximum margin between the classes and the hyperplane. Figure 2 shows support vector machine classification .



. Support Vector Machine

Naïve Bayes Classification Method

Naïve bayes classification method is a probabilistic machine learning algorithm based on Bayes theorem described in probability. Even with its simplicity it outperforms other classifiers; hence, it is one of the best classifiers. The Bayes theorem for calculating posterior probability is given below [9]:

$$p(x|c) = \frac{p(c)p(x)}{p(x)}$$

Where,

$P(c|x)$ = Posterior Probability

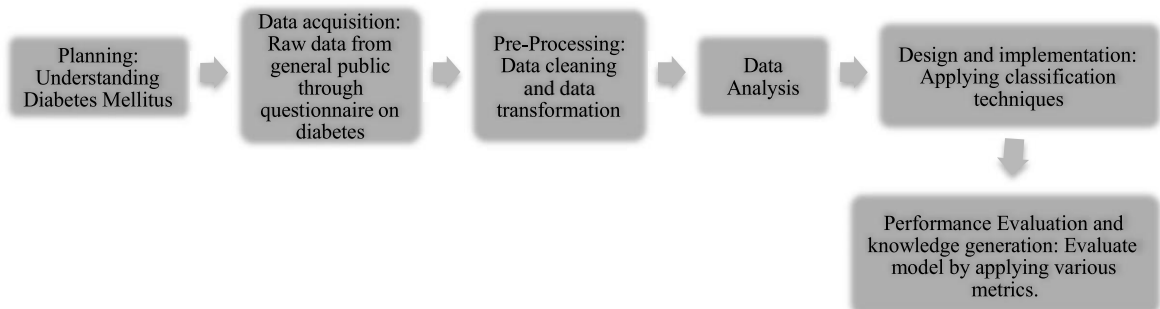
$P(x|c)$ = Likelihood

$P(c)$ = Class Prior Probability

$P(x)$ = Predictor Prior Probability

Design and Implementation

For the purpose of the study, RStudio was used for implementation and R programming language was used for coding. Machine learning algorithms like logistic regression, k-nearest neighbour, support vector machine, naive bayes classifier, decision tree and random forest classifications were implemented on the dataset collected and the Pima dataset in order to predict diabetes. All these predictions from each classifier are then compared with each other. The following are the steps to apply machine learning algorithm, shown in figure 5.



Flowchart of Research.

SUBMITTED BY

M.ABIMANYU

N.GOWTHAMAN

C.S.KRISHNAPRAKASH

B.SATHEESHKUMAR