



# Detecting and Predicting Diabetes Using Supervised Learning: An Approach towards Better Healthcare for Women

Aakansha Rathore  
IGDTUW  
Kashmiri Gate, Delhi-India

Simran Chauhan  
IGDTUW  
Kashmiri Gate, Delhi- India

Sakshi Gujral\*  
IGDTUW  
Kashmiri Gate, Delhi- India

**Abstract:** This paper aims at Detecting Diabetes with PIMA Indian Diabetes Data-set. PIMA India is concerned with women's health. The risk of developing diabetes in Women is quite high due to various factors. Hence, the idea is to Detect and Predict this Disorder with the help of Machine Learning techniques-Support Vector Machine and Decision Trees respectively. The advantage of using these techniques is that it helps in automation of process and makes tasks like Classification, Clustering simpler. The Paper begins with the introduction and emphasize on the worst effect of the Diabetes by explaining various disorders associated with it brief Literature Survey is done to study the work done in it. Then, Section 3 describes the Proposed Approach with Pseudo Code in R Framework. The Framework is used here is R Studio for better analysis and Visualizations. Finally, Results are discussed with Conclusion and Future Scope.

**Keywords:** PIMA, R framework, SVM, Decision Tree, Diabetes.

## 1. INTRODUCTION

Present day life-style has inculcated very serious problem of Diabetes. It is the situation in which glucose is present in excess in the blood that is left unconsumed and increases insulin level in the body. This is the mother cause of many associated health issues like heart attacks, liver failure, kidney failures, nerves damages, vision loss and Polycystic ovarian Syndrome (PCOD). This PCOD is frequently occurring in women now a days because of high resistance towards insulin. As a result, even at teen age chances of being diabetic has increased a lot. Hence, diabetes detection and prediction is an important concern for better health care services especially for Women.

1.1 Type-1 Diabetes-It frequently occurs in the children where no insulin is produced in the body. Pancreatic cells are destroyed as a result no glucose is formed in the body. It is generally known as juvenile diabetes. Common symptoms are-weight loss, dehydration and damage to body parts like liver, kidney, vision loss etc.

### 1.2 Type- 2 Diabetes[1]

It generally occurs at latter stage and women are more prone to this type .This is the situation where excess amount of glucose is present in the blood that is left unconsumed. The Causes of this type of diabetes are excessive weight, no exercise, hereditary etc.

### 1.3 Disorders associated with Diabetes

#### 1.3.1 Diabetic Retinopathy-[2][3]

This is the situation where retina of the eyes is affected because of the pro-longed effect of the diabetes. At the initial stage, blur vision is noticed. Gradually, hard exudes are developed at retina that can also muscles to bleed. Hence, it will certainly lead to vision loss.

1.3.2 Diabetic Neuropathy-It deals with the problem of nerve damage due to excess sugar. Neuropathy deals with the damage of often foot -nerves. As a result, it is advised that Diabetic patients should walk with great precautions.

#### 1.3.3 Malfunctioning of Liver [4]

Liver is highly affected by high blood glucose as it is not able to synthesize bile juice properly. Hence, digestive system is affected by this. Liver shows indigestion problems as early signs first of damage but in worst case, liver paralysis occurs.

#### 1.3.4 Stomach paralysis or Gastro paresis

The worst effect of Diabetes is nerve damage of stomach. It is just like Neuropathy, damage of nerves. Due to this damage, there is feeling of nausea, bloating, indigestion even though small amount of food is eaten.

#### 1.3.5 PCOD

This is the most alarming disorder that is frequently occurring in Women. It can affect teenagers female also. It mainly occurs due to high insulin in the blood. This led to development of multiple Cysts in the ovaries. This can cause weight gain, excessive hair growth on face, acne and in worst case ovarian cancer can be developed. Hence, chances of being pregnant decrease gradually.

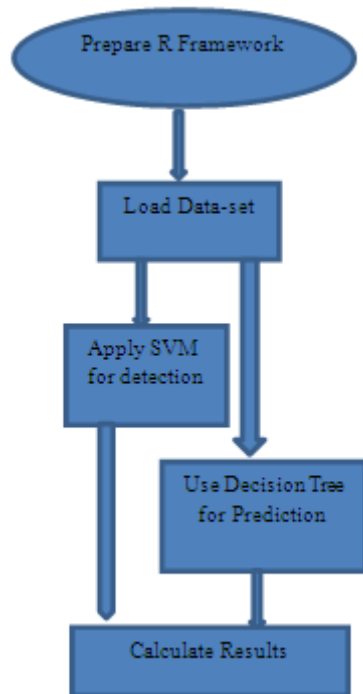
## 2. LITERATURE SURVEY

The literature surveys shows that various Data-sets are available at UCI repository and forms an important part of Diabetes detection study. Various classifiers are used to detect Diabetes. Among them some of the frequently used classifiers and clustering techniques are Random forest, K-Means, J.48 algorithm, Back-propagation, fuzzy approaches. However, there is no significant work that can predict as well as detect.[5-12]

## 3. PROPOSED WORK

This work aims at both detecting Diabetes as well as predicting the risk of diabetes in Pima Indian Diabetes data set. The framework used here is R Studio with R programming language. The selection of R framework is

done with keeping in mind that it forms an important aspect of Data Analytics and visualization studio provides a statistical tool with support of machine learning and visualization language is easy to learn, provides high code density, freeavailable, easy to install and provides sophisticated results also. It has huge web support also. One of the striking features of R Studio is that it can be integrated with Spark that is mainly required to handle big data sets.Hence, power of both big-data as well as analytics can be harnessed. The proposed work is given below-



**Fig 1-Proposed Work**

### 3.1 Data-Set Description

The proposed work makes use of Pima Indian Diabetes Data-set. This Data set is primarily concerned with the women health. Here, 768 instances of almost 21 years of age of women is collected and various parameters are defined. The 8 attributes that are defined here are-

1. Number of Times Pregnancy occurred-
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skinfold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)<sup>2</sup>)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

1. 0-It indicates False Diabetic Test
2. 1-It indicates True Diabetic Test

### 3.2 Methods Used-

1. Support Vector Machines-It is one of the supervised learning techniques that are often used for classification of the Data-sets. It aims at forming the hyper plane that is at the maximal distance from the classes during training phase. During testing phase, the new instance is calculated on the basis of the maximum distance from the hyper plane so formed.SVM classification can be extended up to multiple-class classification. In that case, formation of the Hyper-

plane is done through Kernel Formation.SVM is easy to implement and doesn't even require high processing time when dealing with the small data-set of 768 instances. Hence, it formed an important part of the Diabetes Detection. It divided the data-set into two classes that is diabetic or Non-Diabetic. The package "e1071" is used in Studio for implementation.

2. Decision Tree-It is also one of the supervised learning algorithm techniques that is frequently used for the prediction. It is capable of forming sophisticated predictive model. It forms an important decision support tool system and an integral part of operational research.Hence, selection of Decision Tree will be an appropriate choice for prediction of Diabetes symptoms among women. This is important step as early prediction will help in developing a model that is more oriented towards better healthcare for women.

1. The Decision Support System is based upon splitting the values based on attributes and conditions, so a graph like structure, tree is formed that can be traversed from root to leaf and prediction can be evaluated.
2. The packages used in Studio for implementing Decision Tree is Package "Party" has ctree function that is used to create and visualize Decision Tree.
3. The syntax used here is ctree(formula,data) where-
4. formula-It deals with predictor and variable response.
5. data-It is name of data set used.

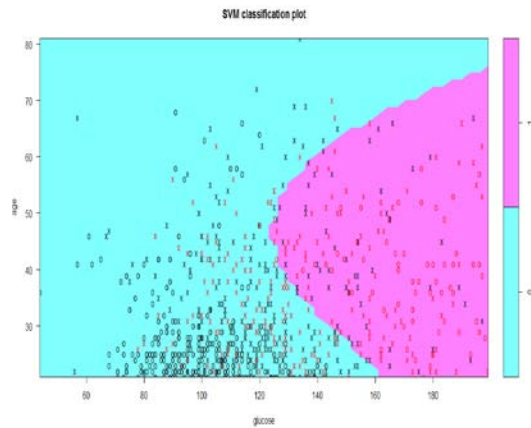
### 4.PSEUDO CODE

The R Code for detection of Diabetes is given-

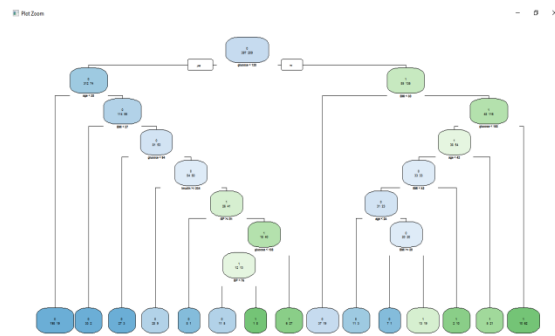
```

datadiab$factoof<- factor(datadiab$facto)
View(datadiab)
library(e1071) #library for svm
#implementing svm
svmfit1 <-
svm(factoof~age+glucose+insulin+BP+pregnant+BMI+triceps, data= datadiab)
summary(svmfit1)
plot(svmfit1, data= datadiab, age~glucose,slice =
list(pregnant = 4, glucose = 140, BP = 70, triceps = 30, BMI
= 32, insulin = 175))
RCodeFor Prediction Of Diabetes
datadiab<- fSetDiabetes
#View(datadiab)
str(datadiab)
datadiab$fac<- factor(datadiab$class)
set.seed(1234)
pd<- sample(2,nrow(datadiab),replace = TRUE,prob =
c(0.8,0.2))
train<- datadiab[pd==1,]
test<- datadiab[pd==2,]
#View(train)
#View(test)
#tree with party
library(party)
tree2 <-
ctree(fac~pregnant+glucose+BP+triceps+age+insulin+BMI
+pedigree,data = train, controls = ctree_control(mincriterion
= 0.9))
tree2
plot(tree2)
  
```

```
#predict
predict(tree2,test)
Results-
```



**Fig 2- SVM classification For Diabetic & Non-Diabetic**



**Fig 3- Decision Tree generated to Predict chances of Diabetes**

**CONCLUSION**

Diabetes Detection is done successfully with SVM Classifier with 82% accuracy. This can be further improved by using Hybrid approaches of multiple Classifiers[1][4] as well as by incorporating Fuzzy Logic. Similarly, Decision Tree is generated for prediction of the Diabetes in Women.Hence, proposed approach will yield an effective method for both prediction and detection.

**FUTURE SCOPE**

The Work can be extended for Big Data that can be analyzed with Hadoop.Hence, the project can meet the demands of future also.

**ACKNOWLEDGEMENT**

We would like to acknowledge all the authors that provide significant help in the research of Diabetes Detection.

**REFERENCES**

- [1] ChandrakarOmprakash,Dr.kumarJatinder, Saini R.,” Development of Indian Weighted Diabetic Risk Score (IWDRS) using Machine Learning Techniques for Type-2 Diabetes”,ACM COMPUTE '16, October 21-23, 2016.
- [2] Mohammed Imran, Alhanouf M. Al-Abdullatif, Bushra S. Al-Awwad, Mzoon M. Alwalmani, Sarah A. Al-Suhaibani, and Shahad A. Al-Sayah,”Towards Early Detection of Diabetic Retinopathy Using Extended Fuzzy Logic”,International Journal of Pharma Medicine and Biological Sciences Vol. 5, No. 2, April 2016.
- [3] Mohammed Imran, Alhanouf M. Al-Abdullatif, Bushra S. Al-Awwad, Mzoon M. Alwalmani, Sarah A. Al-Suhaibani, and Shahad A. Al-Sayah,”Towards Early Detection of Diabetic Retinopathy Using Extended Fuzzy Logic”,International Journal of Pharma Medicine and Biological Sciences Vol. 5, No. 2, April 2016.
- [4] R. Meza-Palacios et al., Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus, Expert Systems With Applications (2016),
- [5] Polatkemal, SalihGüne,”An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease”, Digital Signal Processing 17 (2007) 702–710.
- [6] GowdaKaregowda Ashs1 , Manjunath A.S , Jayaram M.S.,” APPLICATION OF GENETIC ALGORITHM OPTIMIZED NEURAL NETWORK CONNECTION WEIGHTS FOR MEDICAL DIAGNOSIS OF PIMA INDIANS DIABETES”, International Journal on Soft Computing ( IJSC ), Vol.2, No.2, May 2011.
- [7] ButwallMani,kumarShradha,” A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier”,International Journal of Computer Applications (0975 – 8887) Volume 120 – No.8, June 2015.
- [8] VeenaVijayan V.,C.Anjali,” Prediction and Diagnosis of Diabetes Mellitus -AMachine Learning Approach”, 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015 | Trivandrum.
- [9] E.P.Ephzibah,” COST EFFECTIVE APPROACH ON FEATURE SELECTION USING GENETIC ALGORITHMS AND FUZZY LOGIC FOR DIABETES DIAGNOSIS”, International Journal on Soft Computing ( IJSC ), Vol.2, No.1, February 2011.
- [10] G. Thippa Reddy1, a , NeeluKhare,” FFBAT- Optimized Rule Based Fuzzy Logic Classifier for Diabetes”, International Journal of Engineering Rese
- [11] VarmaKamadiV.S.R.P,RaoAllamAppab, ThummalaSitaMahalakshmia,” A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach”, Applied Soft Computing.
- [12] YoichiHayashi n, ShonosukeYukita,” Rule extraction using Recursive-Rule extraction algorithm with J48graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the Pima Indian dataset”, Informatics in Medicine Unlocked.