

DIABETES PREDICTION USING MACHINE LEARNING ALGORITHM

Ms. Archana Ugale. Ms. Sakshi Wagh. Ms. Pooja Sonawane. Ms. Swati Rane. Ms. Snehal
Pardeshi.

Department Of IT Engineering
SVIT Nashik, Maharashtra, India

Abstract:- *the purpose of the research is to make use of different machine learning algorithm to predict the diabetes. the PIMA Indian diabetes dataset (PIDD) is been used to predict diabetes disease. this paper discussed approaches for diabetes prediction and by using different approaches like KNN, RF, SVM and comparing all three results. the main purpose behind is result accuracy or correctness. As diabetes is a common disease affecting individuals globally.it is not a short-term disease so it can increase complications for long time in future. an comprised our model with previous study it shows proposed model can provide better accuracy depend on dataset used and ML algorithm which we used for predicting the project work can gives the accurate model shows that the model is proficient of predicting diabetes effectively. According to our result or study of previous research our deduction shows that random forest achieve higher accuracy compare to other machine learning technique and attributes used for this graphical user interface based diabetes prediction application are glucose, blood pressure, skin thickness, insulin, hemoglobin etc.*

Keywords :- Random Forest Method(RF), k-nearest neighbor Method(KNN), Support Vector
Machine(SVM), Machine Learning, Prediction.

Introduction :- Diabetes is a pernicious disease in the word. Diabetes can caused of high glucose level or stoutness and so forth. It affect the insulin hormone level in patient body which results in untypical metabolism of crabs and increases the level of sugar in the blood. It occurred when body is not to support for make enough insulin. According WHO approximately 422 millions people worldwide has suffer from diabetes and 1.5 millions death per year. According to GDP (Global Diabetes Prevalence) in 2019 the percentage is 9.3 means approx.463M and about to conclusion that it would rather increases to 10.2% (578 million) up to 2030 However the pervasiveness had been found among various countries like Canada, China, and India etc. early prediction of disease like diabetes can controlled and save the human life to execute this work the analyzed the prediction of diabetes by just taking some parameter like age, pregnancy etc. the IDF said that about 52% of Indians are not aware that they suffered by High blood sugar .the PIMA dataset is used for analysis and for training the ML model over that data .the dataset originated form National institute of diabetes and kidney Disease.

Objective :-

The preponderant purpose of this project is to predictive model to predict diabetes at early stage using the clinical and the e-diabetic data. The objective of purposed are framed below.

- To predict diabetes by general patterns.
- To create graphical user interface based on e-diabetic portal.
- To apply ML (machine learning) algorithm to derive patterns
- Beforehand divination(prediction) of diabetes for treatment.
- By apply at a time three ML algorithm to get the most accurate result by comparing three results.

Problem Statement :- to apply systematic techniques to reduce the hospital readmission of diabetic patients classification of data done by SVM, KNN, and RF as the diabetes has approx. 93% affected people worldwide and it has long time impact disease .so the early prediction is main objective.

Literature review :-

- 1) Mitushi soni and Dr, sumita varma proposed svm, knn, and logistic regression for prediction of diabetes develop a system which can perform early prediction. In that work they conclude that random forest classifier achieved better results compare other. The paper “Diabetes prediction using machine learning techniques.” ISSN: 2278-0181.
- 2) The anatomization to related work gives results on various healthcare datasets, where analysis and prediction were carried out using various methods and techniques various prediction model had been developed and implemented by various researches using variants of data mining.
- 3) Darcy a. Davis proposed individual risk prediction based on medical history. their paper also predict each patients greatest disease risk based on their own medical history data. dataset are used for medical coding and collaborative assessment and recommendation engine (CARE) information technique from this it is observed that the ML algorithms place a significant role in knowledge. Discovery from the databases especially in medical diagnosis with the medical data.
- 4) V. Ranjani and M. Duralraj discussed the prospective use of classification- based data mining techniques such as rule-based methods, decision tree algorithm, naïve bayes and artificial neural networks to the massive volume of healthcare data in this research medical problems have been analyzed and evaluated such as heart disease and blood pressure.

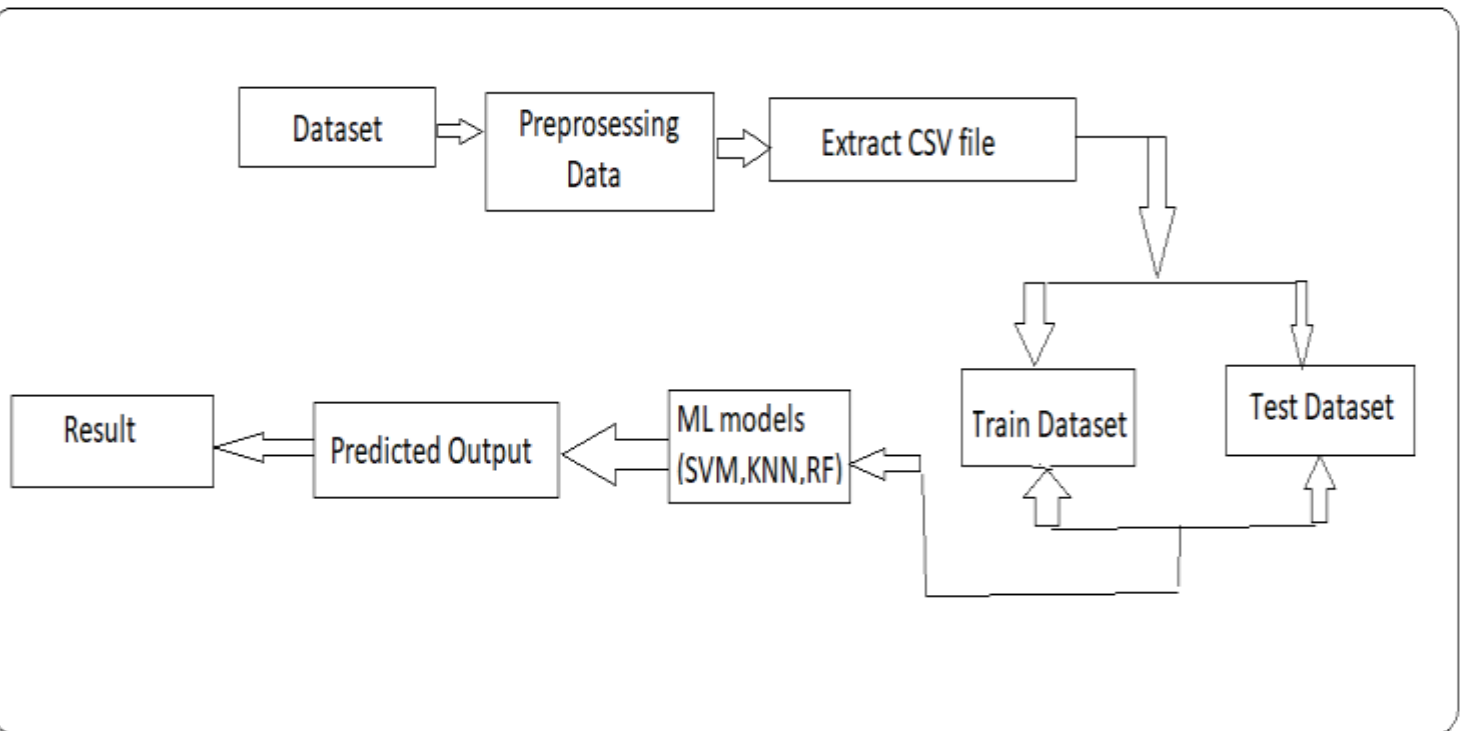
Implementation :-

Methodology :-

Shows architecture diagram for prediction of diabetes model.

This model has five modules

1. Collection of Data.



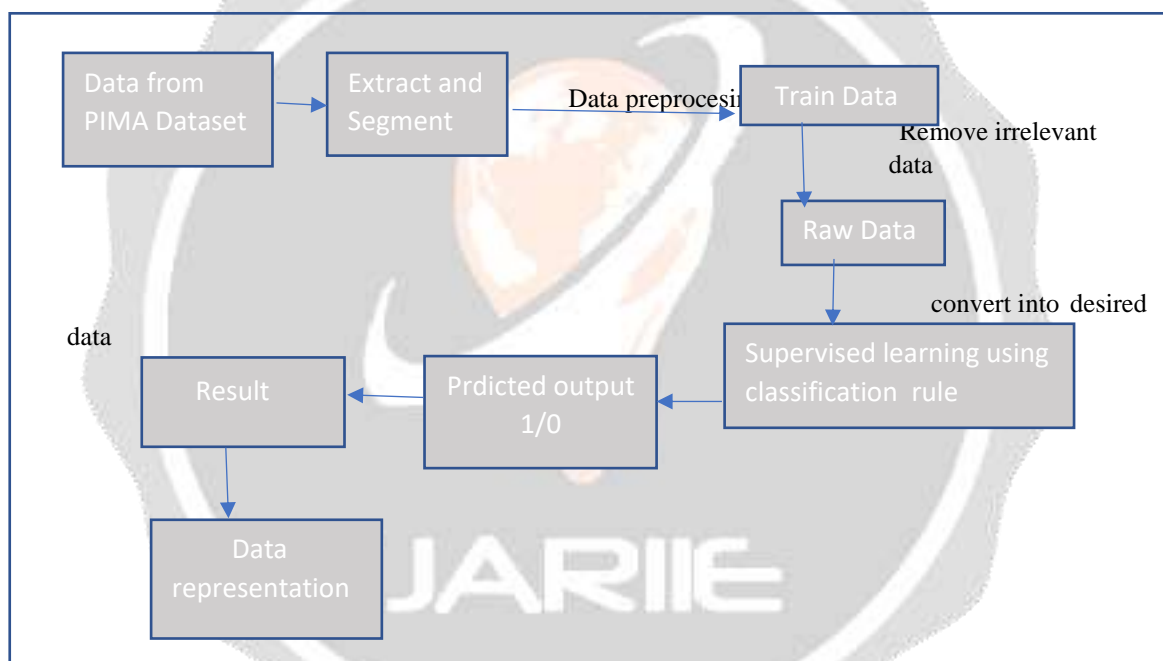
2. Preprocessing over Data.
3. Clustering.
4. Model Building.
5. Evaluation.

Dataset Collection

This module includes collection of data and understanding the data to study trends which helps in prediction and access the results, Dataset description is given below.

<u>Parameter (Attributes)</u>	<u>Type</u>
Number of Pregnancies	N
Glucose Level	N
Blood Pressure	N
Insulin	N
BMI (body mass index)	N
Skin Thickness	N
Blood Pressure	N
Diabetes Pedigree	N
Outcomes	N

- 1) **Data pre-processing**
 This module consists of process of preparing the raw data and making it suitable for a machine learning model. (i.e. cleaning and organizing) to make it suitable for a building and training ML model.
- 2) **Clustering**
 In this phase, we have a machine learning technique which groups the unlabeled dataset. It is a way of grouping the data points into different clusters.
- 3) **Model Building**
 This phase includes building model for prediction of diabetes we have implementing various machine learning algorithms. This algorithm includes support vector classifier, Random Forest, K-nearest neighbor.
- 4) **Evaluation: -**
 This final step of prediction model here we evaluate the prediction results using various evaluations metrics like classification, accuracy, confusion matrix.

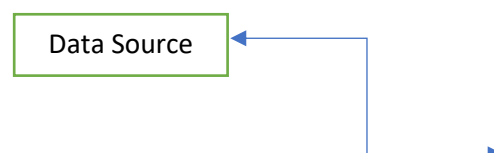


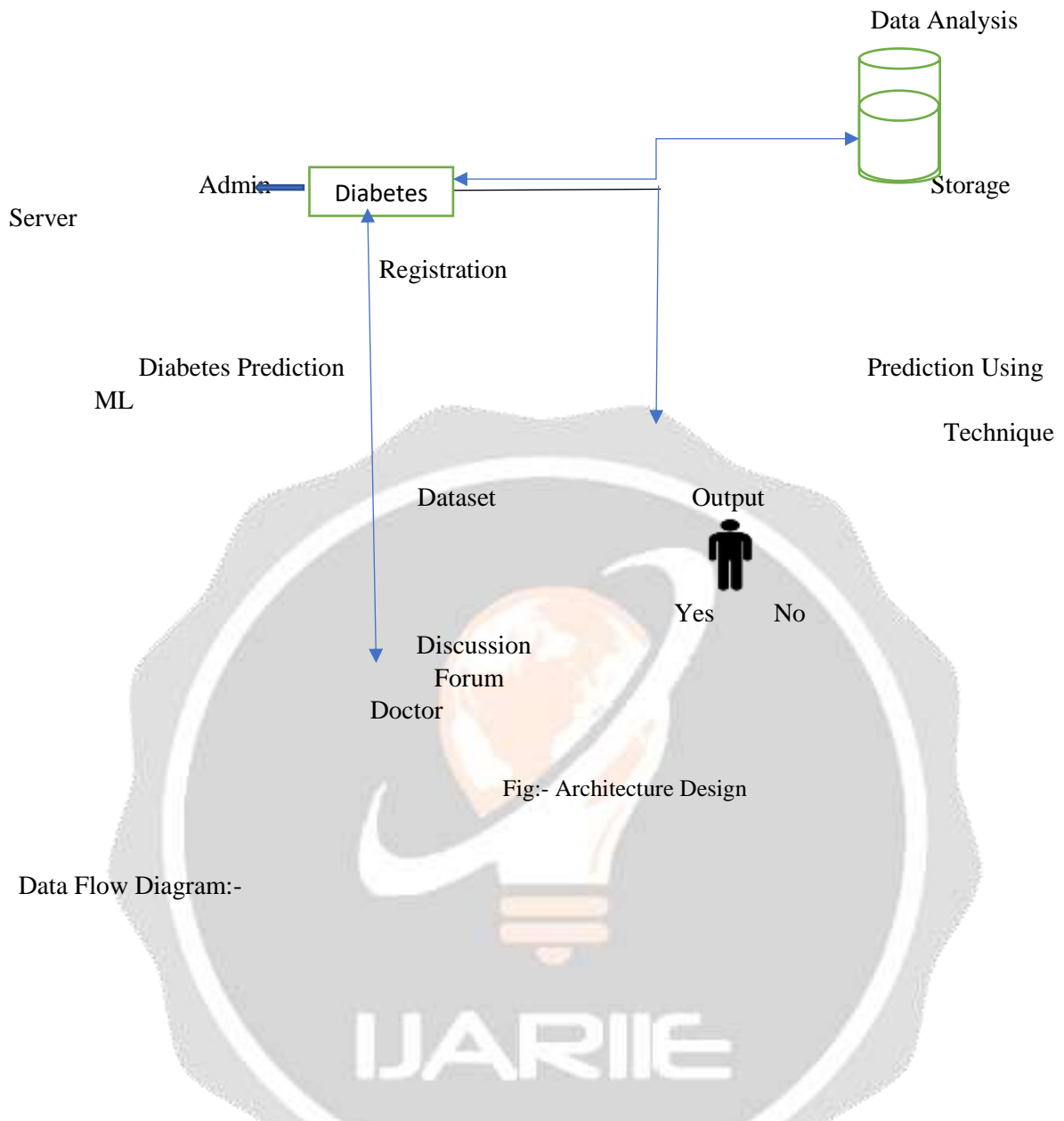
Proposed Work:-

The forecast or prediction of diabetes is the major and concerning. It is severe the complications the diabetes is plays very important role for the human life because it leads to death. This system module is offered for initial prediction of diabetes and time prediction where time prediction means when the Diabetes patient will help to improve the habit of patient.

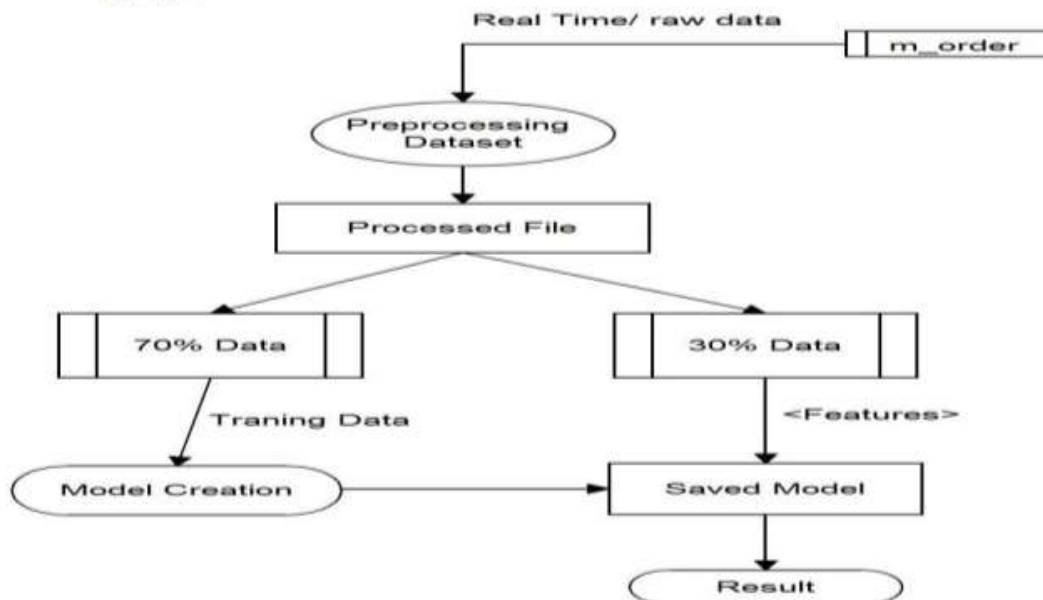
System Design: -

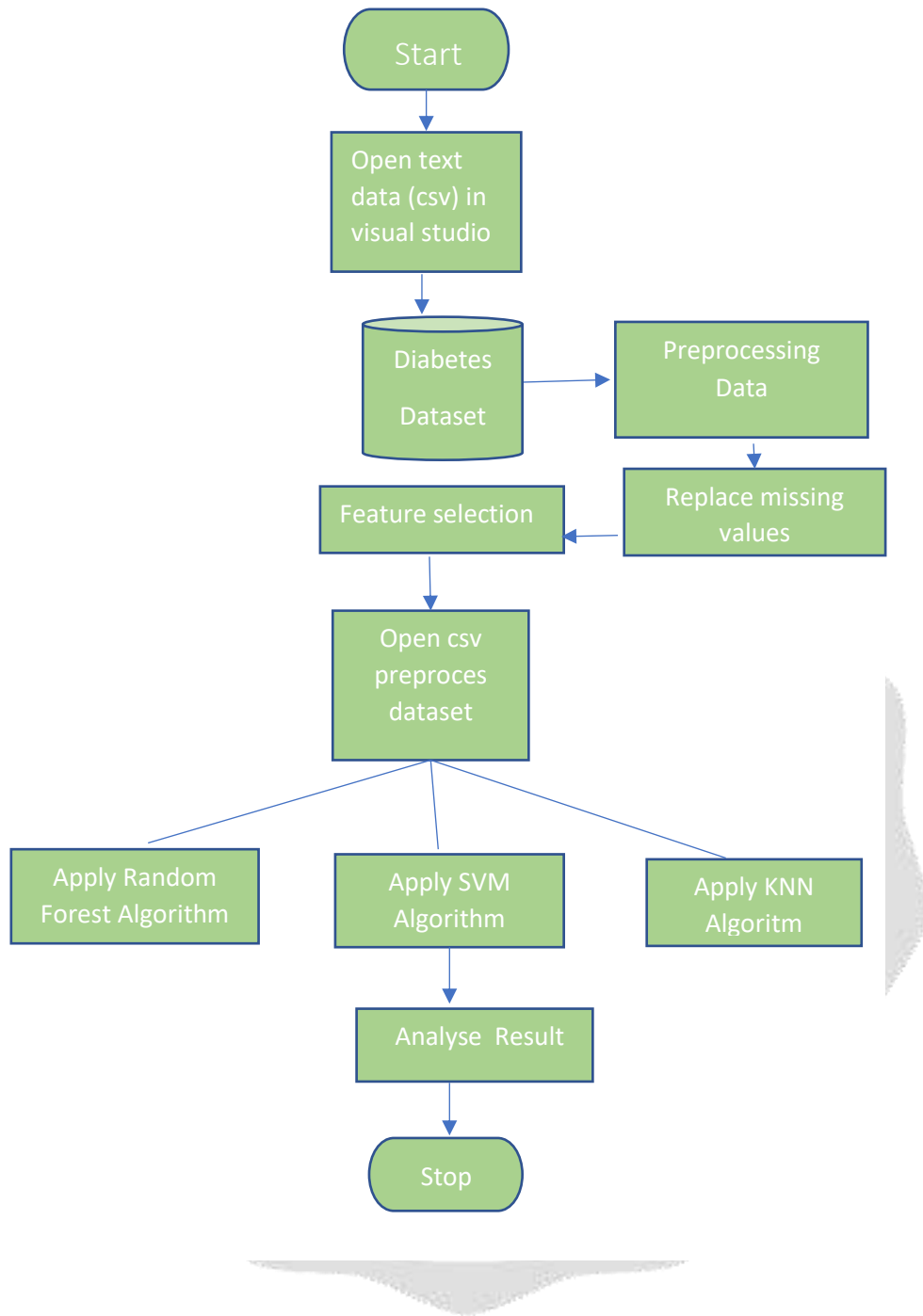
System design is the process which is used for defining the interface, modules and data for a system to specified the demand to satisfy system design is a seen as the application of the system theory



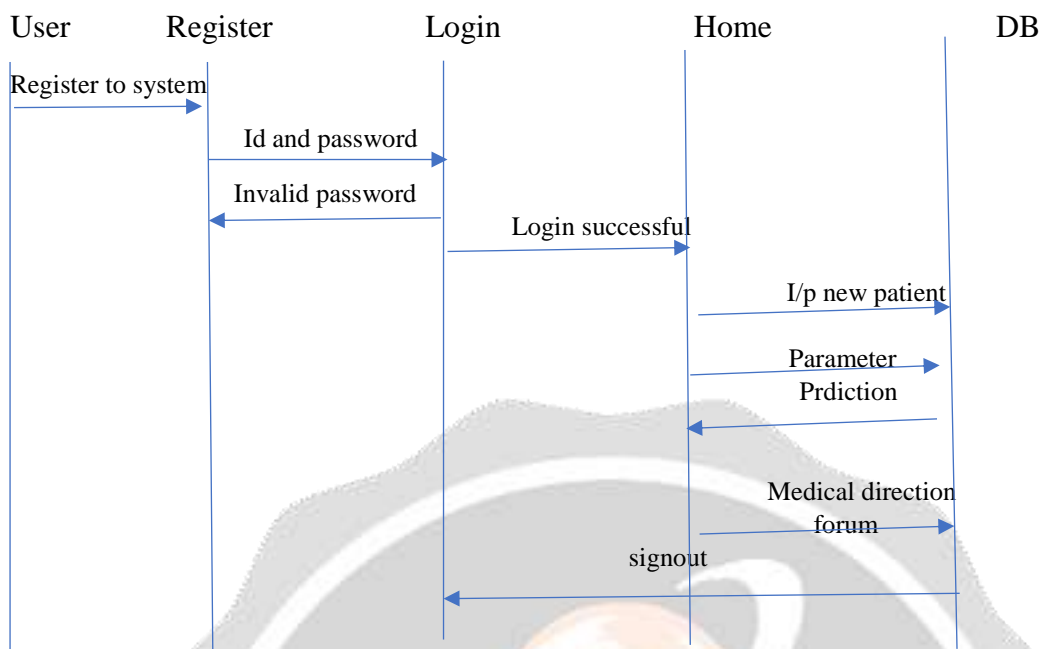


DFD 1





Sequence Diagram:-



Model Used:-

Spiral model-

Logarithmic spiral model was developed by D.A.Cole and P.V.lade. in 1984. For the purpose of prediction. It is the system development lifecycle (SDLC) method used for risk management that combines iterative development process model. The four main points in the spiral model are –

1. Identify objectives.
2. Perform risk analysis.
3. Develop and test.
4. Revive and evaluate.

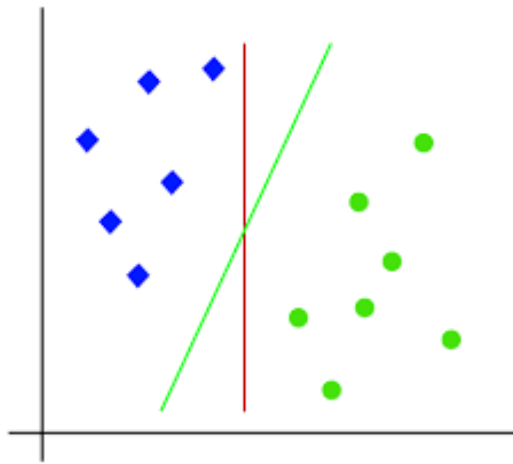
Algorithms used for building this module:-

1) Random forest algorithm:-

Random forest is a supervised machine learning algorithm that is used widely in classification and regression. Probably it is the one of the best algorithm for classification. It is able to classifying large data with accuracy.

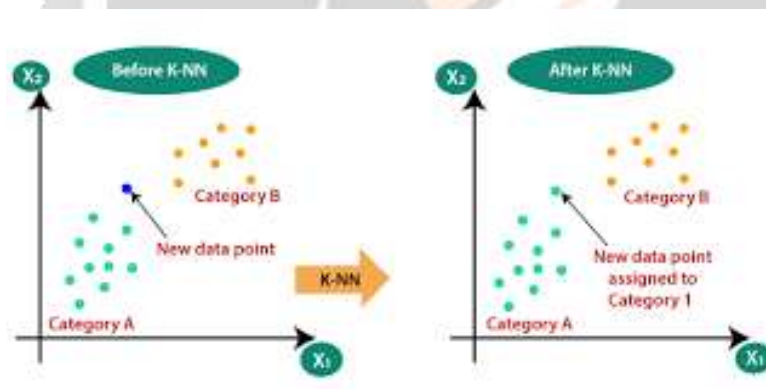
2) Support vector machine :-

We have used svm for identifying for classification of genes, patients on the basis of genes it can be used for classification and regression purpose. If a no of features are high compared to a number of data points in the dataset.



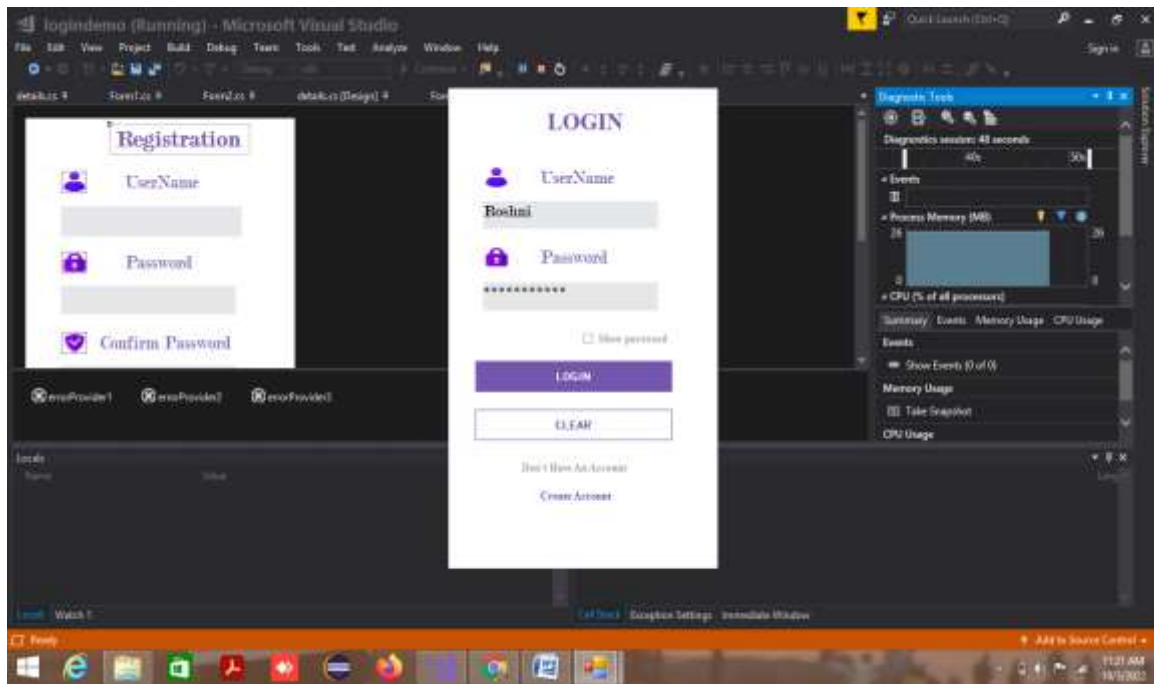
3) K-nearest neighbor method/ algorithm (KNN):-

The abbreviation KNN stands for k-nearest neighbor .it is supervised machine learning algorithm also used for classification and regression. this algorithm at the training phase only save the datasets and after getting new data then it classifies that data into category.

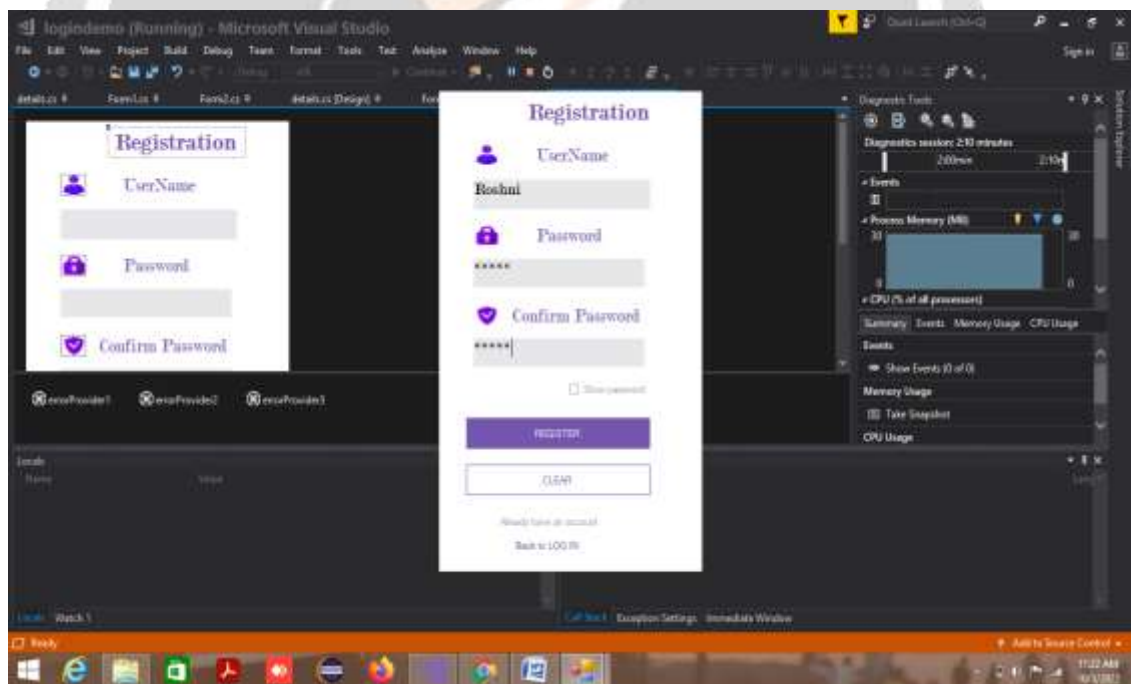


Model Appearance

Login Page:-



Registration Page:-



Form page

Conclusion:-

The prediction of diabetes is vital in today's condition and this is related to its cause and its threatening complications that's why the earlier prediction of diabetes is most important. The system module uses some of the machine learning algorithms for prediction so as obtain more accurate results according to our conclusion of analysis we see that from the three algorithm used in this module (i.e. SVM, KNN, random forest) out of that the random forest method gives the more accurate results. The accuracy is upto 99 % also the KNN can gives accuracy 98% and this model helps for earlier prediction when the illness is identified at early stage proper treatment becomes available machine learning algorithms will improve prediction techniques.

References:-

1. Diabetes care decision support system”2nd International conference on industrial and information system IEEE-2018.
2. J. Maria shyala and Dr. M. Renuka Devi “ Analysis of various data mining techniques to predict diabetes mellitus”. IJ.ISSN 0973-4562 volume 11.
3. Avantika Nahar ¹, Dr. Ajay bala², and Saurabh Sharma ³ “ diabetes prediction using machine learning “,IJRDET ISSN .2347-64352,volume 9.
4. “Performance analysis of machine learning techniques to predict Diabetes mellitus” Md Faisal Faruque; Asaduzzamn Iqbal N.Sarker, IEEE 2019.
5. Aishwarya Mujumdar^a and Dr. Vaidehi v^b “Diabetes prediction using machine learning algorithms. ICRIAC 2019.
6. Mani Butwall and Shraddha Kumar. “A Data mining approach for the Diagnosis of Diabetes mellitus”. International Journal of computer applications, volume-120-november 8,2015.
7. V. Sangeetha and K Rajesh “Application of data mining Methods and Techniques for diabetes Diagnosis , International Journal of engineering and innovative technology (IJETT) volume 2,ISSUE 3,September 2021.
8. Anjali cand Veena v, prediction of Diagnosis of Diabetes Mellitus computational system (RAICS) [Trivendrum.