Title : **Diabetics Prediction System based on Life Style**
Name of the Faculty : Ms sheetal Kusal
Designation : Research Scholar
College Name: Symbiosis International University

1 INTRODUCTION

1.1 Overview

Diabetes is a common, chronic disease. Prediction of diabetes at an early stage can lead to improved treatment. Data mining techniques are widely used for prediction of disease at an early stage. In this project, diabetes is predicted using significant attributes, and the relationship of the differing attributes is also characterized. Various tools are used to determine significant attribute selection, and for clustering, prediction, and association rule mining for diabetes.

1.2 Purpose

Diabetes prediction system is very useful system in the healthcare field. A system for diabetes prediction is proposed. The proposed system used classification algorithm for eliminating the undesired data, thus reducing the processing time.

2 LITERATURE SURVEY

2.1 Existing problem

Several researcher have develop and design a systems for diabetes prediction based on various algorithms and methods. V. Ahuja et al. [1] proposed a system for diabetes disease classification using Support Vector Machine (SVM). The authors used Pima Indian diabetes dataset for evaluation. The obtained accuracy was 78% base on using the Radial Basis Function (RBF) kernel of SVM as the classifier. Aiswarya et al. [2] used J48 Decision Tree and Naïve Bayes as a classifiers for classify the diagnosis of diabetes. Pima Indian diabetes dataset is used in the proposed system and the classification results was 74.8%, 79.5% for J48 Decision Tree and Naïve Bayes respectively. Rajesh et al. [3] proposed a system for diabetes classification based on using C4.5 algorithm for classification. The authors achieved classification rate of 91% by evaluating the training data through data feature relevance analysis. Harlem et al. [4] Proposed a system based on a technique in data mining for diabetes disease prediction. The proposed system has three main steps which are: preprocessing, feature extraction and parameter evaluation. In preprocessing step, the empty and anomalies sets are removed from the used dataset. Besides that, the helpful hidden patterns and relationships of the dataset are explored in the feature extraction step in order to improve the decision making result. Furthermore, the proposed system evaluated based on using J48, Naive Bayes and the achieved rates are 73.8%, 76.3% respectively. In addition, Ravi et al. [5] Fuzzy c means clustering and support vector machine for developing diabetes mellitus prediction. The authors used a dataset that consists of 768 cases and the obtained result was 59.5%. In [6] Krishnaveni et al. proposed a six various techniques to predict diabetic disease. The used techniques are Discriminant analysis, KNN Algorithm, Naïve Bayes, SVM with Linear Kernel function, and SVM with RBF Kernel function. The obtained results of the proposed system for the used techniques are 76.3% using discriminant analysis, 71.1% using KNN Algorithm, 76.1% using Naïve Bayes, 74.1% using SVM with Linear Kernel function, 74.1% using SVM with RBF Kernel function. However, several authors are used various methods in order to get the

best prediction rate.

2.2 Proposed solution

The proposed diabetes prediction system has two main stages that work together to achieve the desired results. The first stage of the proposed system is the data preparation, and the second one is the classification. However, the input into the system is the PID dataset and the output will be one class which represents the diabetic or the non-diabetic. In this, we need to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The datasets consist of several medical predictor variables and one target variable, Diabetes. Predictor variables include the number of pregnancies the patient has had, Plasma glucose concentration, Diastolic blood pressure, Triceps skin fold thickness, their BMI, Serum insulin level, Diabetes pedigree.

3 THEORITICAL ANALYSES

3.1 Block Diagram
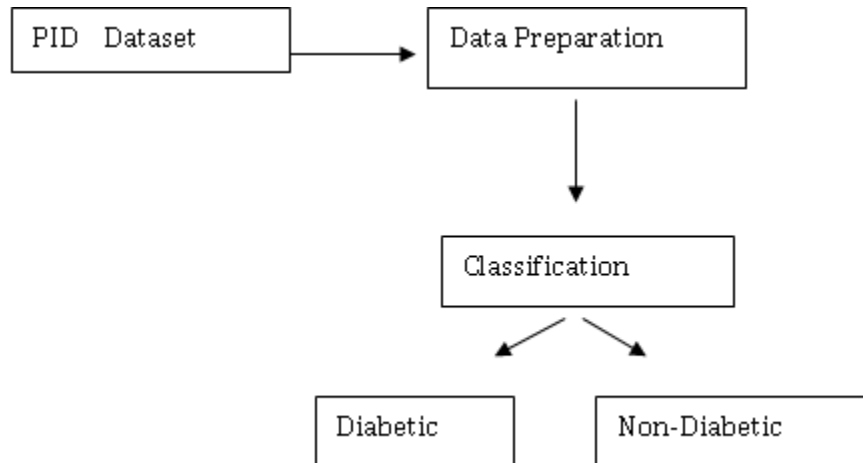
3.2 Hardware / Software designing.
Develop an end-to-end web application that predicts the probability of females having diabetes. The application must be built with Node Red for building UI with the machine learning model trained & deployed on IBM Watson Studio. IBM Node red, IBM Watson Studio, IBM Machine Learning, IBM Cloud Object Storage.
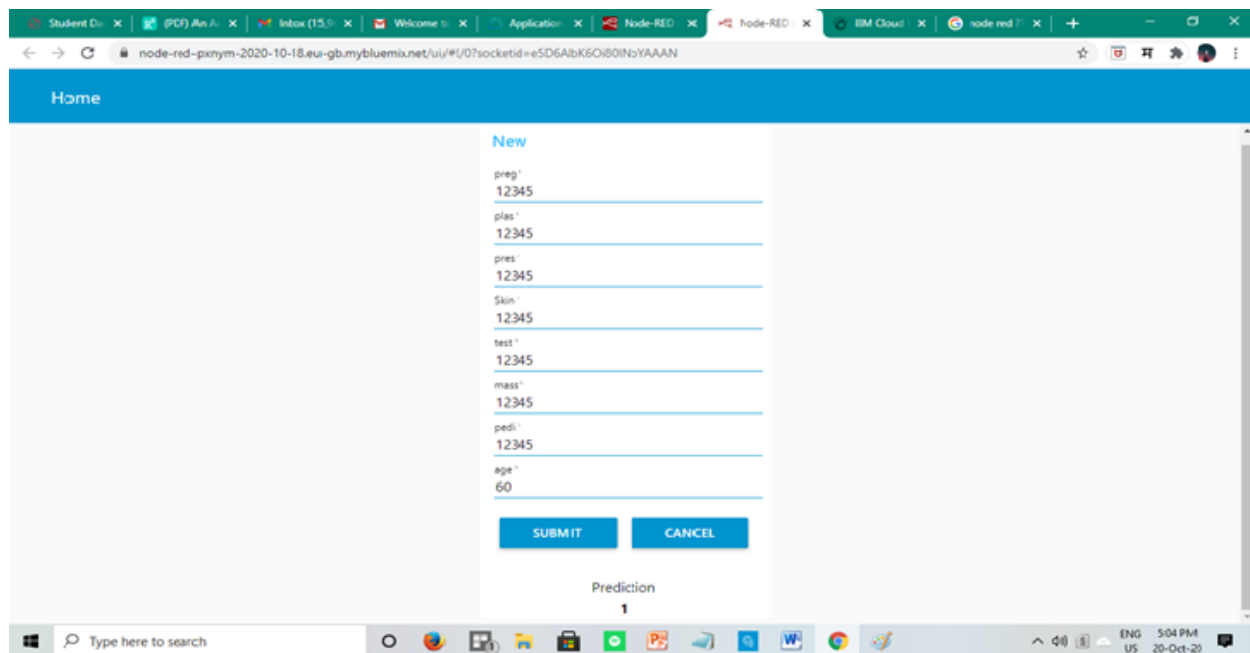
4 EXPERIMENTAL INVESTIGATIONS

5 FLOWCHARTS

IBM Cloud Object Storage → Watson Studio → User Interface Using IBM Node Red
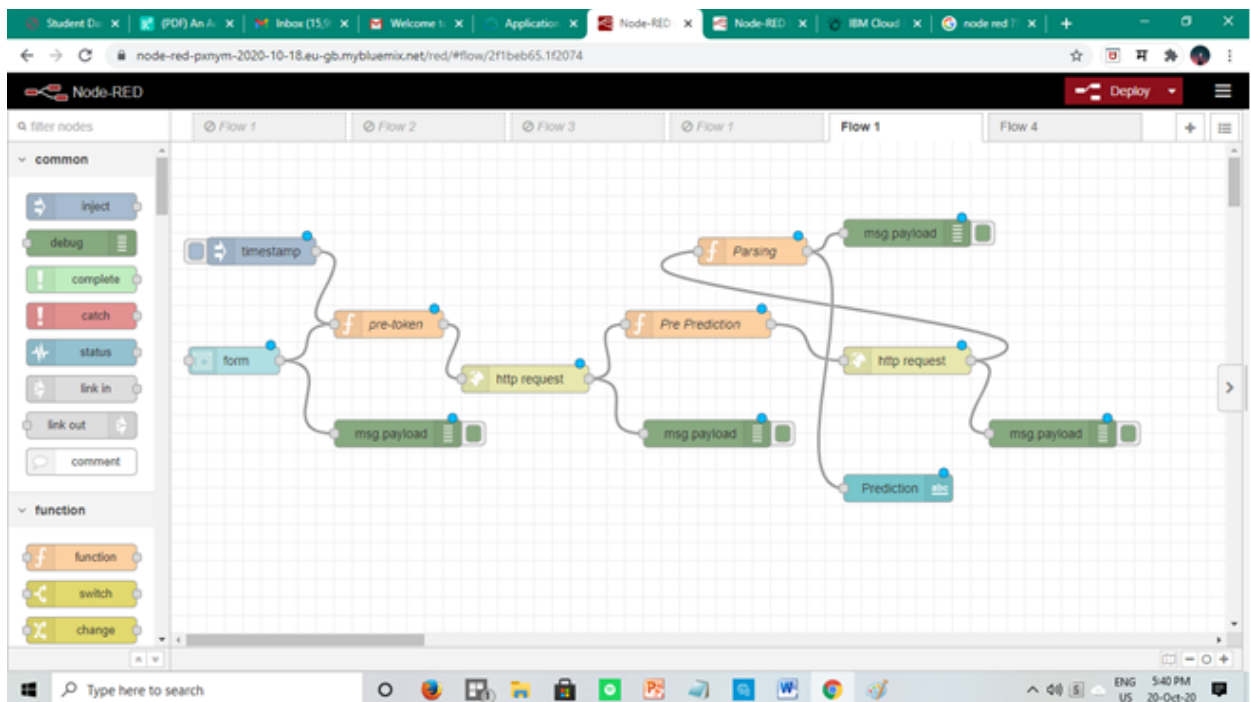
Watson Machine Learning
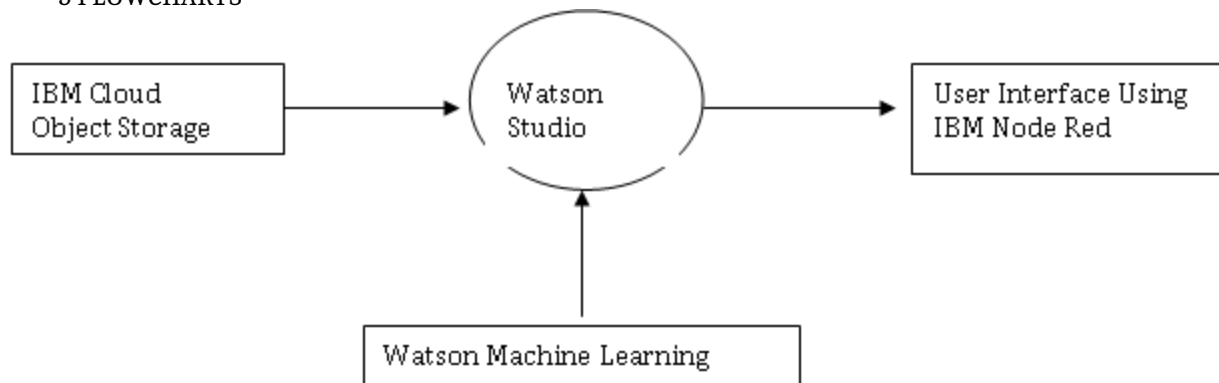
6 RESULTS

ADVANTAGES & DISADVANTAGES

Machine learning algorithms are used to classify and diagnosis the diseases, in order to eliminate the problem and reduce the required cost. Besides that, using the machine learning algorithm lead to meaningful and accurate decisions. Medical data sets that contain irrelevant data (noise) are used to train and evaluate the machine learning algorithms. The noise are affects the decision results of the used algorithm.

APPLICATIONS

Diabetes prediction system is very useful system in the healthcare field.

CONCLUSION

A diabetes prediction system is proposed. The proposed system used 768 instances within 8 attributes for each one of PID dataset. The used data is preprocessed in order to remove the unwanted data, and lead to faster processing time. Moreover, the dividing technique of the dataset into subset, made an optimal classification result. The proposed system focused on the features analysis and classification parts. The experiment illustrate the effects of using the algorithms of the proposed system through achieving a higher classification rate that the other systems.

FUTURE SCOPE

Different machine learning algorithms can used to increase the prediction rate.

BIBILOGRAPHY

1. V. Anuja and R.Chitra., "Classification Of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications (IJERA), vol.3,Issue 2, pp. 1797-1801, 2013.
2. Aiswarya I., S. Jeyalatha and Ronak S., "Diagnosis Of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol.5,

,No. 1, pp. 1- 1-14, 2015.

3. K.Rajesh and V.Sangeetha,"Application of Data Mining Methods and Techniques for Diabetes Diagnosis," in proceedings of International journal of Engineering and Innovative Technology, vol.2, Issue 3, pp. 43-46, 2012.

4. Harleen and Dr. Pankaj B.,"A Prediction Technique in Data Mining for Diabetes Mellitus," Journal of Management Sciences and Technology, vol. 4, Issue 1, pp. 1-12, 2016.

5. Ravi S. and Smt T., "Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine," International Journal of Computer Trends and Technology (IJCTT), vol. 11, No. 2, pp. 94-98, 2014.

6. G. Krishnaveni*, T. Sudha," A Novel Technique To Predict Diabetic Disease Using Data Mining Classification Techniques" in International Conference on Innovative Applications in Engineering and Information Technology (ICIAEIT-2017), vol. 3, Issue 1, pp. 5-11, 2017

APPENDIX