# DIABETES MELLITUS DISEASE PREDICTION USING MACHINE LEARNING.

[1]Harshitha G P, [2]Prasanna P

[1]PG Scholar, [2]Associate Professor
[1]Department Of Computer Science and Engineering,
[1]PESCE, Mandya, India

***Abstract :*** Diabetes mellitus, commonly make reference to diabetes, is a kind of metabolic issue which has influenced a huge number of individuals, in this disorder the person contain high glucose level over long period of time. It is one of the most common health conditions experienced around the globe. The number of population of diabetic patient has been increasing from 108 million in 1980 to 422 million in 2014. Diabetes may leads to kidney failure, blindness, heart attack, lower limb amputation. Therefore the diagnosis of diabetes in early stages plays a vital role. The main purpose of this project is to predict whether the patient has diabetes, if not it also predict the risk of after how many years the patient might affected by diabetes. In this project, we are diagnosing diabetes using Naive Bayes and C4.5 algorithm which is among one of the major important technique of machine learning. The data set is taken from UCI machine learning repository.

***IndexTerms* - Diabetes Mellitus, Dataset, Machine Learning, Naive Bayes algorithm, C4.5 algorithm**

## I. INTRODUCTION

Diabetic is a severe condition which has impact on entire body. Insulin is a hormone in a human body which is secreted by pancreas. Insulin can generate energy by converting sugar into simple molecules, that molecules are used by body cells to induce energy. This metamorphosis has an impact due to the lack of insulin leads to the sugar gathering in the flow of blood stream. This results in the incrimination of the glucose level in the blood and the person originates Diabetes mellitus. This disease demands personal care and if problem emerge, diabetes can cause affect on status of life and reduces life span of the person. At present there is no cure for diabetes mellitus, the person may live delightful life by knowing the circumstances of the disease and maintain the circumstances effectively.

The two main types of diabetes mellitus are Type1, Type2. In Type1 diabetes mellitus condition the pancreas has little or no insulin. In Type2 diabetes mellitus conditions cells cannot use blood glucose effectively for the production of energy. This occurs when the cells become unresponsive to insulin and then the blood glucose level gradually rises high. Gestational diabetes mellitus is identified by carbohydrate intolerance of varied extremity with beginning or first identification during pregnancy. Carrying ladies with a background marked by GDM are expanded danger of future diabetes so are the children, predominantly type2 diabetes mellitus. The level of risk depends on diagnostic criteria used to recognize GDM.

All Diabetes types have something in common. Commonly, food consumed will be separated into starches and sugar, which transforms into explicit sugar called as glucose. The cells in the body will be stimulated by this glucose, in any case, these cells need insulin for suck in glucose and to discharge energy. Each type of diabetes mellitus can be treated as insulin is obtainable since 1921. Type1 and Type2 diabetes are immedicable situation and are not able to cure. In the treatment of Type1 Diabetes mellitus the transplantation of pancreas were tried but the success rate is disappointing and the second treatment was gastric bypass which is gainful in different ways. Diabetes mellitus results in a direct indication of huge glucose level in blood, alongside certain side effects including expanded starving, weight loss, frequent urination during night, feeling dizzy and increased thirst. The diabetic affected patient need persistent nursing, otherwise, it will feasibly leads to numerous risky perilous intricacies. The diabetes mellitus is identified by the two hour post-load plasma glucose level at least 200mg/dl, and the essential requirement of recognizing diabetes timely calls in numerous works about the identification of diabetes. The main cause for diabetes is represented in Fig. 1.

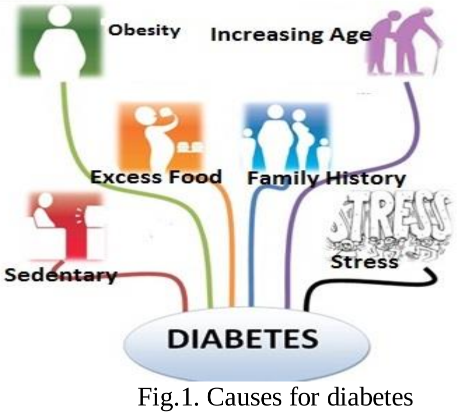


Fig.1. Causes for diabetes

Machine learning belongs to the scientific field in which machines grasp from experiences. Many scientists say that the word “Machine Learning” is similar to the word “Artificial Intelligence”, because probability of studying is the major aspect of an entity called intelligent in the broadest sense of the word. The intention of machine learning is to construct computer systems that can adapt and grasp from the data available. In recent years there is quickened advancement in the Machine learning field, many algorithms and techniques are present in Machine Learning which can be efficiently used for identification and prediction of different types of disorders. In this project, we are diagnosing diabetes using Naive Bayes and C4.5 algorithm which is one of the most important technique in Machine Learning.

Through this experiment, we predict whether the patient has diabetes, if not it also predict the risk of after how many months or year the patient might affected by diabetes. Here, we are diagnosing diabetes using Naive Bayes and C4.5 algorithms. These algorithms in Machine Learning are beneficial and convenient for doctors and diabetic mellitus patients. This system predicts the diabetes based on the attributes such as Age, Gender, Relation, Plasma Glucose Test Value, Symptoms, Family history of diabetes, Smoking, Alcohol consumption, Height, Weight and Blood pressure etc. This project uses previous diabetic patient data for the prediction of diabetes for new patient. Here we get the data set from the UCI machine learning repository.

## II. LITERATURE SURVEY

A research paper given by Berina Alic, Lejla Gurbetal, Almir Badnjevihas presents the paper about Machine learning techniques in classification of diabetes and cardiovascular diseases by Bayesian network and Artificial Neural Network[4]. Classification of diabetes and cardiovascular diseases has been explained in this paper and then it also compare results of applications of two machine learning techniques, Bayesian Network and Artificial Neural Network. In this paper for classification of diabetes disease they obtained accuracy between 72.2% and 99% by Artificial Neural Network and then they obtained accuracy between 71% and 99.51% by Bayesian Network. So, according to the results obtained it could be said that the greater possibility to get high accuracy in classification of diabetes is when Bayesian Network is used. Then again, the most commonly utilized sort of Bayesian Network is Naive Bayes sysrem which appeared most astounding precision esteems for classification of diabetes and CVD, 99.51% and 97.92% respectively, which shows that higher probability to acquire progressively precise outcomes in diabetes or potentially CVD classification is when it is applied.

A research paper given by Muhammad Shahbaz, ShoaibFaruq, Muhammad Shaheen, Syed Ather Masood distinguishes various sorts of cancer on the genes dataset by utilizing data mining classification tools to form a decision support system [5]. Cancer is a set of sickness where a few cells of the body develop strangely. These cells at that point demolish other encompassing cells and their usual task. Cancer may spread all through the human body. Since it is an exceptionally deceptive infection its determination is significant. So the determination of cancer in a early stages plays an vital role. The test is to initially analyze the fundamental sort and afterward its subtypes. Data mining technique basically support to breaking down the information, which helps in distinguishing cancer patients and this method recognises cancer disease patients. They have implemented tests utilizing three algorithms Naive Bayes, K Nearest Neighbours and SVM. As per results above Naive Bayes classification has the most precise expectation for leukaemia dataset tests. Naive Bayes has obtained 95% accuracy. It has just blunder rate of 5%. Naive Bayes is the best strategy for classifying microarray genes expression data. The exactness estimation of KNN technique is 90.72%. For SVM classification technique the exactness is 90.27%.

A research paper given by Krati Saxena, Dr. Zubair Khan, Shefali Singh utilizes the most significant methods of Artificial Intelligence for the diagnosis of Diabetes Mellitus. The Artificial Intelligence field has a fast progression [6]. A.I. has a large number of strategies and algorithms which can be adequately utilized for the prediction and determination of different diseases. These algorithms are man-made reasoning demonstrate to be practical and efficient for diabetic patients and specialists. In this paper, they are diagnosing Diabetes mellitus utilizing K-Nearest Neighbour algorithm which is standout amongst the most significant procedures. In this they consider one example preparing dataset containing 100 lines and 11 segments of the attributes. Here they consider two set data sample each consisting of 50 rows. The K-Nearest Neighbour algorithm is applied on the sample test dataset and acquired outcomes for various estimations of K which is number of closest neighbours. They have determined the exactness and blunder rates of K which is number of closest neighbours. They have determined the exactness and blunder rates for K=3 and K=5. The outcomes have been assessed utilizing MATLAB. The outcome demonstrates that as the estimation of K expands, exactness rate and blunder rate will likewise increment. KNN is a standout amongst the best Artificial Intelligence algorithms that is generally utilized for diagnosing purposes. Progressively exact and productive outcomes can be gotten through KNN.

A research paper given by Mukeshkumari, Dr.RajanVohra, Anshularora helps in anticipating diabetes by applying data mining strategy. The disclosure of knowledge from medical datasets is significant so as to make viable therapeutic determination [7]. The point of data mining is to obtain knowledge from dataset and to get recognizable descriptions of patterns. Utilizing data mining strategies to help individuals to anticipate diabetes has increased significant fame. In this paper, they proposed Bayesian Network Classifier to identify whether the people has diabetic or not. The dataset utilized is gathered from an healthcare, which gathers the data of people with and without diabetes. For the test and examination they have utilized the Weka tool. Bayesian network demonstrates the best exactness, 99.51 percent and blunder rate and blunder rate in the classification is 0.48% when the outcomes were contrasted with clinical determination.

## III. SOFTWARE REQUIREMENTS AND SPECIFICATIONS

This section of Software Requirement Specification describes all general factors of the product and its requirements.

### A. User Requirements:

- **Admin:** Administrator is a one who maintains the entire application. Administrator is an owner of the application.
- **Physician/Doctor (Diabetes Disease Specialist):** Doctor is a one who specifies the necessary inputs for disease prediction. Doctor is a service receiver. The key service given by the system is "*Diabetes Disease Prediction*" based on the medical data.
- **Receptionist:** Receptionist is one who maintains the patient registration, billing and treatment details.
- **Patient:** Patient is a one who receives the services from the application. Patients can access to treatment details.

### B. Hardware Requirements:

- Minimum of 2GB RAM
- Pentium IV or higher
- Minimum of 40GB HARD DISK
- Standard PC configuration to carryout challenging computing.

### C. Software Requirements:

PLATFORM**:** DOT NET -- VISUAL STUDIO 2008/2010

**Table 1:** Software Requirements

| Layer | Technology Used | Technology | Vendor |
|---|---|---|---|
| *Presentation Layer* | *ASP.NET 4.0 (Visual Studio 2010)* | *Microsoft. NET* | *Microsoft* |
| *Business Layer (language used)* | *C#.NET* | *Microsoft. NET* | *Microsoft* |
| *Database* | *MS Access/SQL Server* | | |
| *Data Access* | *ADO.NET* | *Microsoft. NET* | *Microsoft* |
| *Other Technologies* | *HTML,CSS,JAVA SCRIPT,AJAX* | | |

### D. Functional requirements:

- **Staff Creation Module (Admin):** Administrator of the system creates the staffs (specialist, receptionist) and manages the staffs and sets the unique id and password for each staff.
- **Patient Registration Module (Receptionist):** Receptionist of the hospital registers the patients by collecting the patient details such as name, address, contact no, email id etc… receptionist sets the patient Id and password for each patient for future use.
- **Parameters Module (Receptionist):** Receptionist of the hospital manages the different constraints required for the prediction of diabetes disease. Basically there are n number of constraints related to diabetes disease prediction.
- **Dataset Module (Receptionist):** Receptionist manages the dataset required for the diabetes disease prediction. Here receptionist uploads the old data into server which includes diabetes disease patients data with related constraints/parameters and results.
- **Input Module - New Patient (Physician):** Diabetes Disease Specialist uploads the new patient constraints, based on these constraints system will predict the output.
- **Prediction Module (Physician):** This is the core module of the project where system accepts the input given by the disease specialist. This module predicts the final output weather patient is classified to "Yes" or "No" and time prediction. We make use of classification rules technique - "*Naive Bayes Algorithm*" and *"C4.5 Algorithm"* for the output prediction which is one the efficient algorithm which works fine for small dataset as well as huge dataset.
- **Treatment Module (Physician, Patients):** This module maintained by the Specialist where specialist uploads the treatment details for the patients and patients can view the treatment details.
- **Account Module (Admin, Receptionist, Physician, Patient):** This is a common module of all the actors where they can manage their profile by updating and changing passwords.

### E. Communication Interface Requirements:

- Hyper Text Transfer Protocol (HTTP) is used to transmit documents around network.
- The HTTP protocol will be used to facilitate communications between the client and server.
- Mainly the website is implemented using SQL server 2005 as back end and ASP.NET as front end.
- The website is based on three tier architecture with data server, application server and a client.

## IV. SYSTEM ARCHITECTURE

This reference system architecture explains the process of converting input data into desired output. We collect large amount of data set from UCI machine learning repository then pre-processing is applied on dataset which removes irrelevant data and extract relevant data from dataset. These relevant data are considered training data. The diabetes disease prediction model is trained based on the training dataset to get the desired output. The result which is gained through prediction model is represented in a user friendly manner.

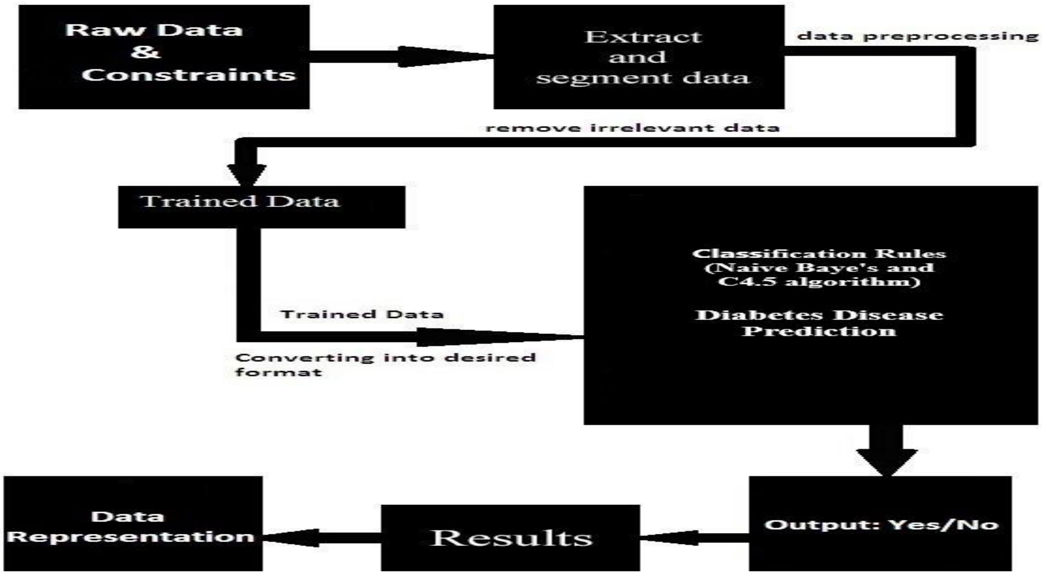


**Fig 2: System architecture.**

## V. METHODOLOGY

This web application is implemented using object oriented programming language. It is an approach which presents a method of standard programs. These standard programs generate subdivided memory area for functions and data which can be utilized as templates for generating copies of the modules on commanded.

Features of Object Oriented paradigm:

- Data has more prominence relatively than procedure.
- Programs are subdivided as objects.
- Data structures are constructed in such a way that they distinguish the objects.
- Objects can convey with one another by making use of methods.
- Whenever required newly discovered data and methods can be added effortlessly.
- While designing a program it follows bottom-up approach.

The proposed system is implemented using three tier architecture. ASP.NET is used in the presentation layer, C# classes are used in the Business logic, Table adapter is used in the data tier. Table Adapter elects a stand of database connection and data commands which are utilized to complete the dataset and modernize a SQL server database. MS SQL server (database) 2005 is used as the backend.

**ASP.NET:** ASP.NET is a platform that combines web development applications, which contributes necessary services to construct web applications for enterprise class. It is mainly syntax adaptable along with Active Service Pages. ASP.NET impact a latest infrastructure and programming model which permits you to generate a powerful latest class of applications. ASP.NET is a division of .NET Framework which permits you to take complete advantage of the common language runtime characteristics.

ASP.NET is supported for both client and server applications on Windows 2000, Windows XP Professional and Windows Server 2003. The following software is also needed to implement ASP.NET server applications.

- Windows 2000 Server or Advanced Server with Service Pack 2, Windows XP Professional or 64-Bit Edition, or one of the Windows Server 2003 family products.
- MDAC 2.7 for Data.
- Internet Information Services.

**ADO.NET- Database Connectivity:** Many of the applications require data access at some point of time generating an essential part while applications are active. The application communes with a database by making use of data access, where data is saved. Data access needs various requirements for different applications. ASP.NET utilizes ADO.NET as data access and handles protocol which authorizes us to work with data on the web.

The platform between the front end controls and backend data base is implemented by ADO.NET. All data access services and controls communes with the objects to present data which are encapsulated by ADO.NET, consequently the movement of data details are masked. Data access in ADO.NET depends on Data set and Data provider components.

*Data set:* Dataset is disorganised, in memory presentation of data. It tends to be considered as a neighbourhood duplicate of the significant segments of the database. The Dataset is preserved in memory and data in this memory may be operate and modernized. At the point when the utilization of these dataset is done, modification can be assemble back to the main database for modernisation. Data contains in the dataset may be stacked from any substantial data source like Microsoft SQL server database, Microsoft Access database or from an oracle database.

*Data Provider*: The Data Provider is in charge of contributing and preserving is associated with the database. It contains a lot of relevant components that cooperate to give data proficient and execution driven way. The .NET Framework presently occur with two data providers, the SQL data provider which is developed only to operate with Microsoft SQL Server 7.0 or beyond and the OLE DB Data provider which gives permission to commune with variant types of databases such as Access and Oracle. All Data provider comprises of the accompanying four core objects:

- Connection object: It allocates connection to the database.
- Command object: For execution of commands it makes use of command object
- Data Reader object: It provides a read-only, forward-only stream of data which is stored in data source.
- Data Adapter object: It populates a disorganised dataset and achieve upgrade for data.

**SQL Server:** Microsoft developed a Microsoft SQL Server which is a relational database management system. It is a software product which contains essential functions of ordering and recovering data as demanded by other software applications, which can execute either on same PC or on different PC over a network.

**Pseudo code**:

- *Pseudo code for Connection String:*

```
<connectionStrings>
<add name="DemandConnectionString" connectionString="Data Source=HOME-1765628F9E\SQLEXPRESS;
Initial Catalog=forecastingDB;Integrated Security=True"
 providerName="System.Data.SqlClient" />
 </connectionStrings>
```

This code is written in the web.config file.

- *Pseudo code for SQL Connection:*

```
Using (SqlConnection connection = new SqlConnection(connectionString))
   {
       connection.Open();
       // Do work here;
      connection.Close();
   }
```

It will execute the method and returns the result to the business logic. Business Logic method will return the result to the Presentation Layer.

## VI. ALGORITHM

### A. Naive Bayes Algorithm

Naive Bayes is a data mining classification technique and it is used as a classifier. This classifier is used for probability prediction if a sample belongs to particular class. The quality of Naive Bayes is high accuracy and fastest to train data. It is usually used on very large datasets. The Naive Bayes Algorithm is a probabilistic algorithm that is sequential, following steps of execution, classification, estimation and prediction. There are various data mining existing solution for finding relations between the diseases, symptoms and medications, but these algorithms have their own limitations; numerous iterations, high computational time and binning of the continuous arguments etc. Naive Bayes overcomes various limitations and can be applied on a large dataset in real time. Naïve Bayes algorithm is used to predict whether the patient has diabetes or not. The proposed system can predict for individual patient and also it can predict for group of patients. For group of patients we consider the testing dataset this data set can be added by the doctor. For finding the accuracy of algorithm we consider 3 kinds of dataset such as training data set, actual data set and testing dataset. The training dataset is used to train the algorithm which contains results of the patients. The testing data set does not contains result of patients, then we find the result for the testing data set using this algorithm and we compare these results with the actual data set, which have the correct result of the patients. By knowing that how many patients are correctly classified, we get to know accuracy of the algorithm.

**Algorithm:**

**Step 1:** Scan the dataset (storage servers)

Retrieval of required data for mining from the servers such as database, cloud, excel sheet etc.

**Step 2:** Calculate the probability of each attribute value. [n, n_c, m, p] Here for each attribute we calculate the probability of occurrence using the following formula.

**Step 3:** Apply the formulae

P( attributevalue (ai) / subjectvalue (vj) )=(n_c + mp)/(n + m)

*Where:*

- n = the number of training examples for which v = vj
- n_c = number of examples for which v = vj and a = ai
- p = a priori estimate for P(aij, vj)
- m = sample size

**Step 4:** Multiply the probabilities by p
For each class, here we multiple the results of each attribute with p and final results are used for classification.
**Step 5:** Compare the values and classify the attribute values to one of the predefined set of class.

### C4.5 Algorithm

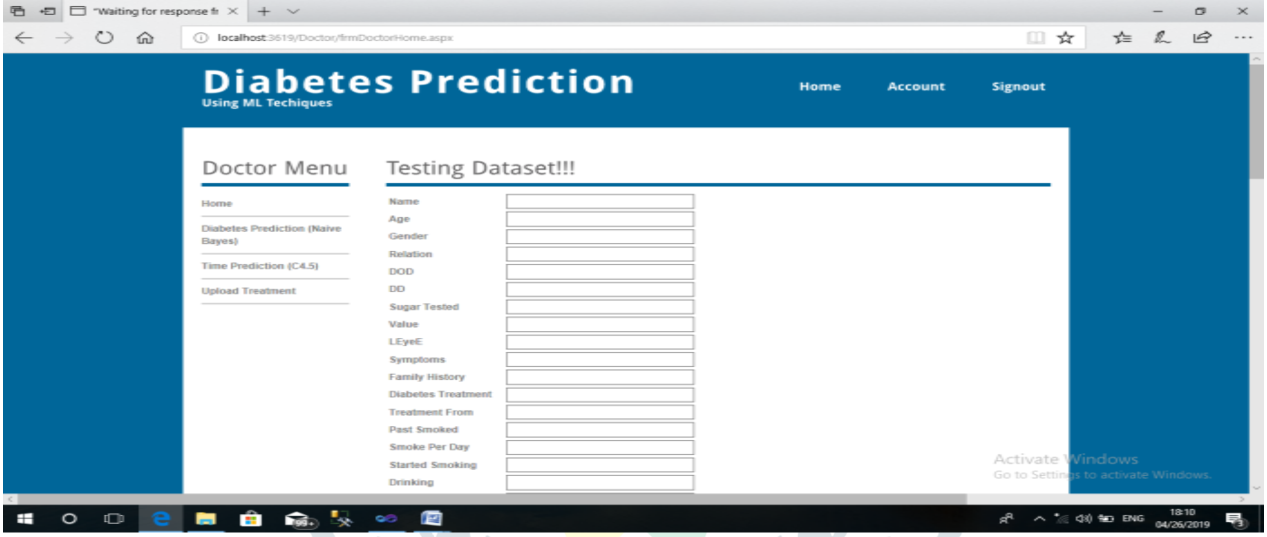
C4.5 is also referred as statistical classifier. For classification and prediction decision trees are incredible and famous tools. Decision trees generate rules, which can be induced by people and utilized in information framework, for example database. C4.5 is an algorithm which structures decision trees. It is an augmentation of ID3 algorithm and it was planned by Quinlan. C4.5 structures decision tree using a training data. Discrete and continuous attributes are handled by C4.5 algorithm. C4.5 is one of generally utilized learning algorithms. Using the concept of information entropy the decision trees are constructed from a set of training data.C4.5 algorithm is used for the time prediction.

### Algorithm:

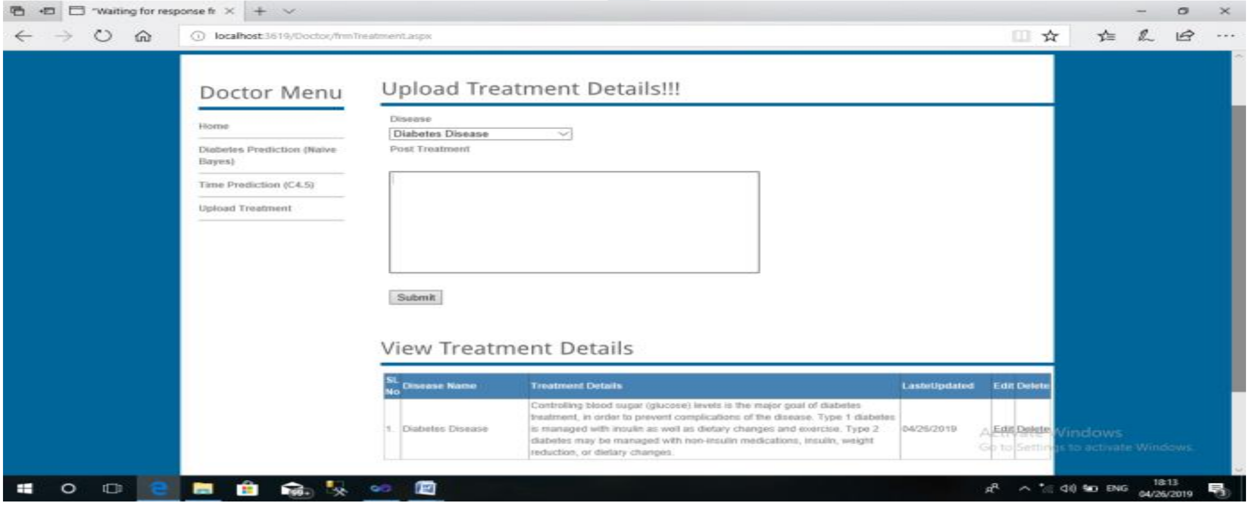
```
Step 1: Scan the dataset (storage servers)
Step 2: for each attribute a, calculate the gain [number of occurrences]
Step 3: Let a_best be the attribute of highest gain [highest count]
Step 4: Create a decision node based on a_best
Step 5: Recur on the sub-lists obtained by splitting on a_best, and add those nodes as children of node.
```

## VII. SNAPSHOTS

### A. Home page of doctor, where he can add testing dataset.

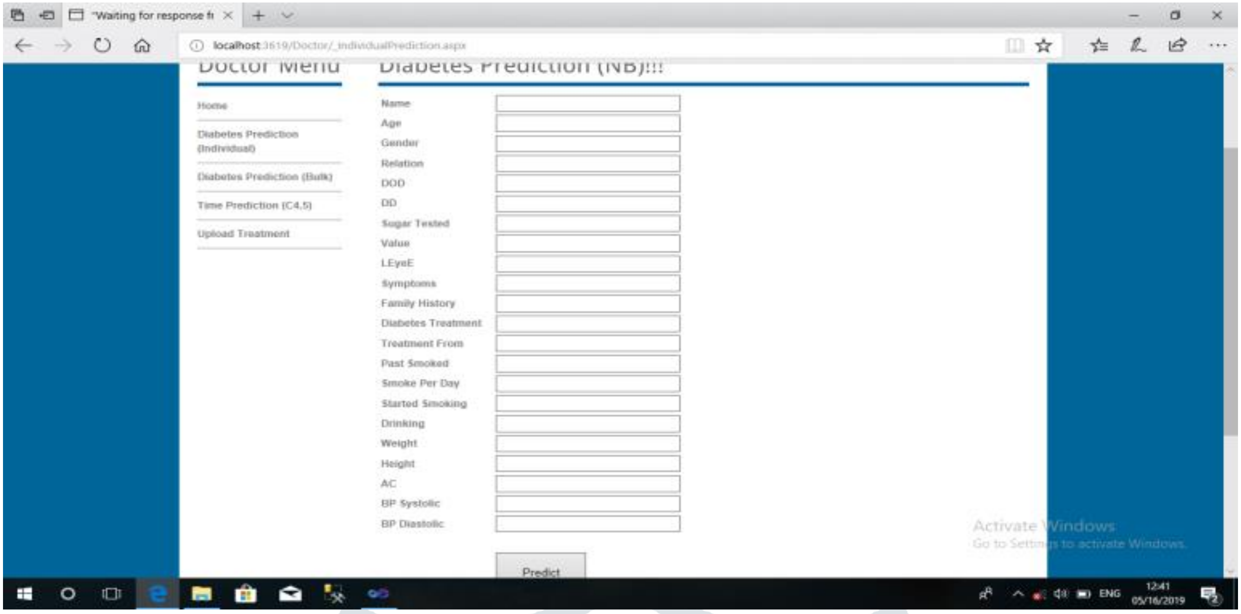


### B. Doctor can upload treatment details in this page.



| SL No | Disease Name | Treatment Details | LasteUpdated | Edit | Delete |
|---|---|---|---|---|---|
| 1. | Diabetes Disease | Controlling blood sugar (glucose) levels is the major goal of diabetes treatment, in order to prevent complications of the disease. Type 1 diabetes is managed with insulin as well as dietary changes and exercise. Type 2 diabetes may be managed with non-insulin medications, insulin, weight reduction, or dietary changes. | 04/26/2019 | Edit | Delete |

### C. Diabetes disease prediction for individual patient.

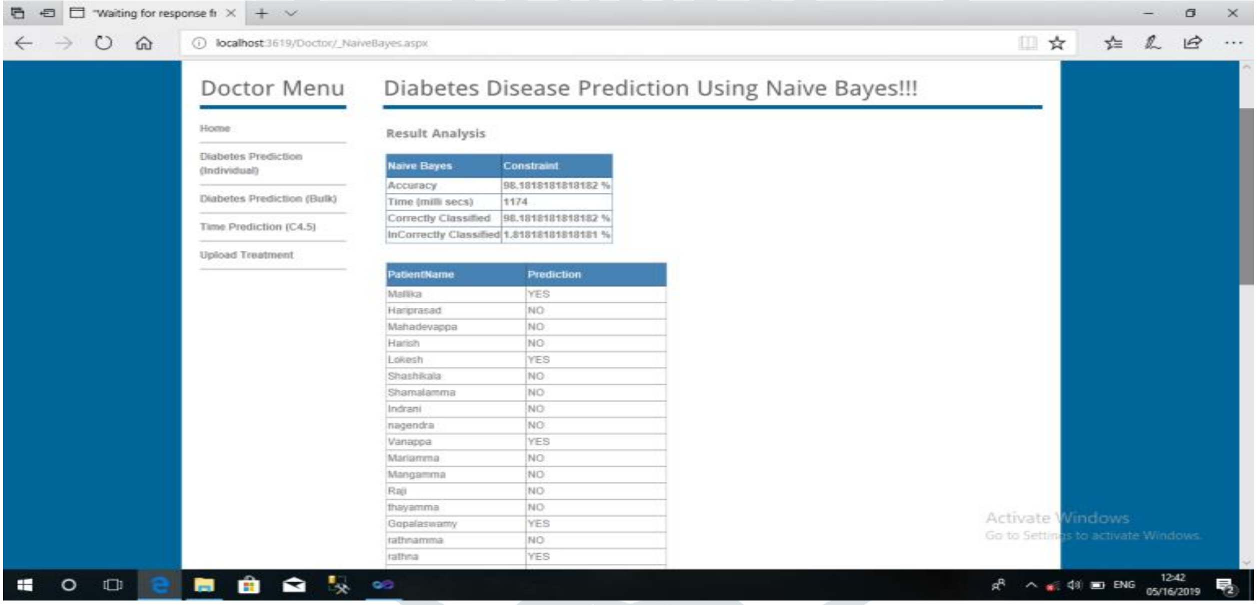

"Waiting for response fr
localhost:3619/Doctor/_individualPrediction.aspx
Doctor Menu
Home
Diabetes Prediction (Individual)
Diabetes Prediction (Bulk)
Time Prediction (C4.5)
Upload Treatment
Activate Windows
Go to Settings to activate Windows.
12:41
05/16/2019

Diabetes Prediction (NB)!!!

Name
Age
Gender
Relation
DOD
DD
Sugar Tested
Value
LEyeE
Symptoms
Family History
Diabetes Treatment
Treatment From
Past Smoked
Smoke Per Day
Started Smoking
Drinking
Weight
Height
AC
BP Systolic
BP Diastolic

Predict

### D. Diabetes disease prediction for testing dataset and then it compares results with the actual dataset.


"Waiting for response fr
localhost:3619/Doctor/_NaiveBayes.aspx
Doctor Menu
Home
Diabetes Prediction (Individual)
Diabetes Prediction (Bulk)
Time Prediction (C4.5)
Upload Treatment
Activate Windows
Go to Settings to activate Windows.
12:42
05/16/2019

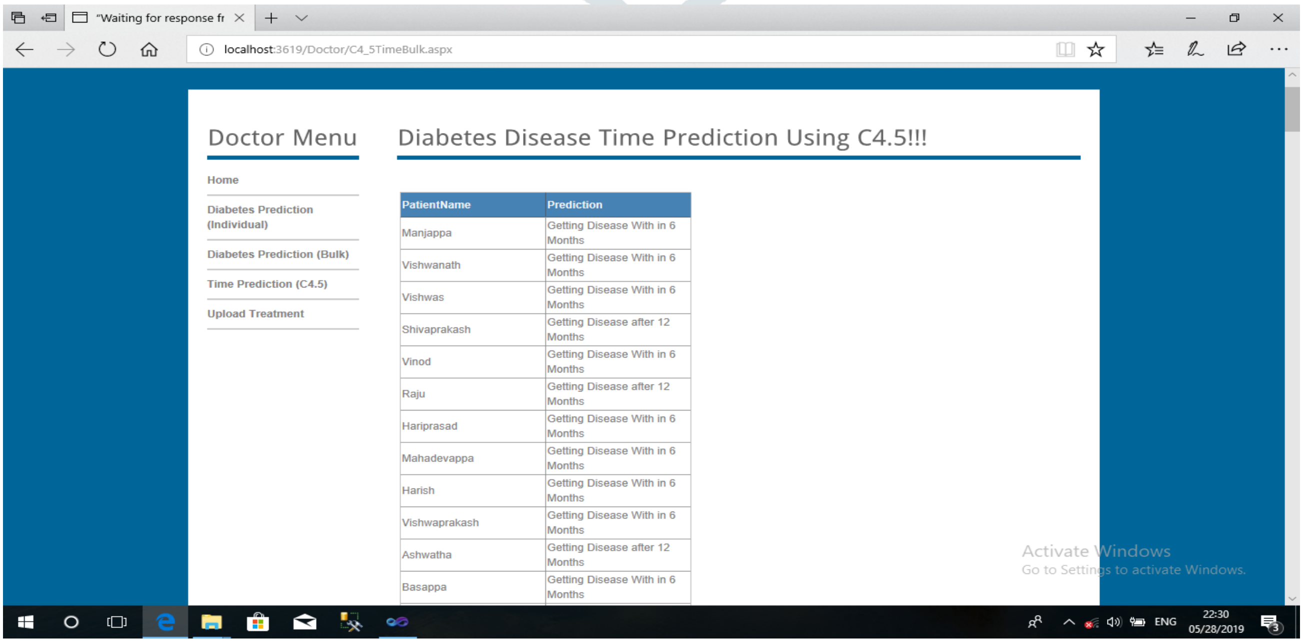
Diabetes Disease Prediction Using Naive Bayes!!!

Result Analysis

| Naive Bayes | Constraint |
|---|---|
| Accuracy | 98.1818181818182 % |
| Time (milli secs) | 1174 |
| Correctly Classified | 98.1818181818182 % |
| InCorrectly Classified | 1.81818181818181 % |

| PatientName | Prediction |
|---|---|
| Mallika | YES |
| Hariprasad | NO |
| Mahadevappa | NO |
| Harish | NO |
| Lokesh | YES |
| Shashikala | NO |
| Shamalamma | NO |
| Indrani | NO |
| nagendra | NO |
| Vanappa | YES |
| Mariamma | NO |
| Mangamma | NO |
| Raji | NO |
| thayamma | NO |
| Gopalaswamy | YES |
| rathnamma | NO |
| rathna | YES |

### E. Diabetes disease time prediction.


"Waiting for response fr
localhost:3619/Doctor/C4_5TimeBulk.aspx
Doctor Menu
Home
Diabetes Prediction (Individual)
Diabetes Prediction (Bulk)
Time Prediction (C4.5)
Upload Treatment
Activate Windows
Go to Settings to activate Windows.
22:30
05/28/2019

Diabetes Disease Time Prediction Using C4.5!!!

| PatientName | Prediction |
|---|---|
| Manjappa | Getting Disease With in 6 Months |
| Vishwanath | Getting Disease With in 6 Months |
| Vishwas | Getting Disease With in 6 Months |
| Shivaprakash | Getting Disease after 12 Months |
| Vinod | Getting Disease With in 6 Months |
| Raju | Getting Disease after 12 Months |
| Hariprasad | Getting Disease With in 6 Months |
| Mahadevappa | Getting Disease With in 6 Months |
| Harish | Getting Disease With in 6 Months |
| Vishwaprakash | Getting Disease With in 6 Months |
| Ashwatha | Getting Disease after 12 Months |
| Basappa | Getting Disease With in 6 Months |

## CONCLUSION

Diabetic mellitus is a disorder, which can induce numerous problems. The prediction and diagnosis diabetes mellitus disease using machine learning technique is advantageous. Proposed system is a medical sector application which is useful to physicians (diabetic doctors) in identifying the disease. The main purpose of this project is to predict whether the patient has diabetes, if not it also predict the risk of after how many years the patient might affected by diabetes. In this project, we are diagnosing diabetes using Naive Bayes and C4.5 algorithm which is one of the most important technique in Machine Learning. The accuracy of a proposed algorithm reaches 98.18%.

The dataset considered in this system does not consider certain parameters which relate to type of diabetes the person is affected, so the system will predict whether the patient has type1, type2 or gestational diabetes. In future, this system can be used in hospitals as a complete Medicaid diagnosis system.

## REFERENCES

**[1]** World Health Organization, “Report of a study group: Diabetes Mellitus,” World Health Organization Technical Report Series, Geneva, 727, 1985.

**[2]** Kemal Polat, Salih Gunes, and Ahmet Arslan, “A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine,” Expert Systems with Applications, vol. 34. 1, January. 2008, pp. 482-487.

**[3]** Kayaer K and Yildirim T, “Medical diagnosis on Pima Indian diabetes using general regression neural networks,” Proceedings of the international conference on artificial neural networks and neural information processing, 2003, pp. 181-184.

**[4]** Berina Alic, Lejla Gurbetal and Almir Badnjevi, “Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases”, 6th Mediterranean Conference On Embedded Computing (Meco), 11-15 JUNE 2017.

**[5]** Muhammad Shahbaz, Shoaib Faruq, Muhammad Shaheen and Syed Ather Masood, “Cancer Diagnosis Using Data Mining Technology”, Life Science Journal, 2012.

**[6]** Krati Saxena, Dr. Zubair Khan and Shefali Singh, “Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm”, International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 4, July-Aug 2014.

**[7]** Mukesh kumari, Dr. Rajan Vohra and Anshul arora, ” Prediction of Diabetes Using Bayesian Network”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014.