## Diabetics Prediction System based on Life Style

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | |  | **INTRODUCTION** | |  |  |  |
|  | |  | 1.1 Overview | |  |  |  |
|  | |  | 1.2 Purpose | |  |  |  |
| **2** | |  | **LITERATURE SURVEY** | |  |  |  |
| **3** | |  | **PROPOSED SYSTEM** | | |  |  |
| **4** | |  | **BLOCK DIAGRAM** | |  |  |  |
| **5** | |  | **FLOWCHART** | |  |  |  |
| **6** | |  | **RESULT** |  |  |  |  |
| **7** | |  | **ADVANTAGES & DISADVANTAGES** | | |  |  |
| **8** | |  | **CONCLUSION** | |  |  |  |
|  | |  |  | |  |  |  |
|  | |  |  | |  |  |  |
|  |  | |  | |  |  |  |
|  |  | |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| 1. **INTRODUCTION**   **1.1 Overview** |  |  |

Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves. About 422 million people worldwide have diabetes, the majority living in low-and middle-income countries, and 1.6 million deaths are directly attributed to diabetes each year. Both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.

Type 1 diabetes can develop at any age, but occurs most frequently in children and adolescents. When you have type 1 diabetes, your body produces very little or no insulin, which means that you need daily insulin injections to maintain blood glucose levels under control.

Type 2 diabetes is more common in adults and accounts for around 90% of all diabetes cases. When you have type 2 diabetes, your body does not make good use of the insulin that it produces. The cornerstone of type 2 diabetes treatment is healthy lifestyle, including increased physical activity and healthy diet. However, over time most people with type 2 diabetes will require oral drugs and/or insulin to keep their blood glucose levels under control.

Gestational diabetes (GDM) is a type of diabetes that consists of high blood glucose during pregnancy and is associated with complications to both mother and child. GDM usually disappears after pregnancy but women affected and their children are at increased risk of developing type 2 diabetes later in life.

* 1. **Purpose**

Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis. Machine Learning plays a significant role in healthcare industries. Healthcare industries have large volume databases. Using machine learning one can study datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In this study we have proposed a diabetes prediction model based on the lifestyle of an individual.

The aim of the study is to develop an end-to-end web application that predicts the probability of females having diabetes and to predict whether a patient has the risk of developing 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, Glucose, Blood Pressure, Skin Thickness, BMI, insulin level, age, and Diabetes Pedigree Function.

1. **LITERATU RE SURVEY**

*[Hasan, Shameem. (2018). Prediction of Diabetes Based on Artificial Intelligence Technique. 05. 11-15.]*

Hasan has done a predictive analysis of diabetes through artificial neural networks. The objective of this paper is to develop a system that can offer a precise early prediction of diabetes for a patient with the help of artificial intelligence technique. The datasets consist of several medical predictor variables and one target variable. Independent variables include the Body Mass Index (BMI), insulin level, age, number of pregnancies the patient had (for female) and some others. In this research work, many statistical approaches such as Bayesian regularization (BR), Levenberg–Marquardt algorithm (LM) and scaled conjugate gradient (SCG) have been used for the evaluation and accuracy of the performance and results. Among them the BR is considered to be the optimal one as it develops the nonlinear relationships and it has more predictive abilities. To get better and refined results, the data was tested through ten hidden layers, and on observing the results, it can be seen that the BR shows least mean square error for diabetes prediction. The nftool (neural fitting tool) of MATLAB has been used in our proposed work to determine the performance and the results.

*[A. Yahyaoui, A. Jamil, J. Rasheed and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," 2019 1st International Informatics and Software Engineering Conference (UBMYK), Ankara, Turkey, 2019, pp. 1-4, doi: 10.1109/UBMYK48245.2019.8965556.]*

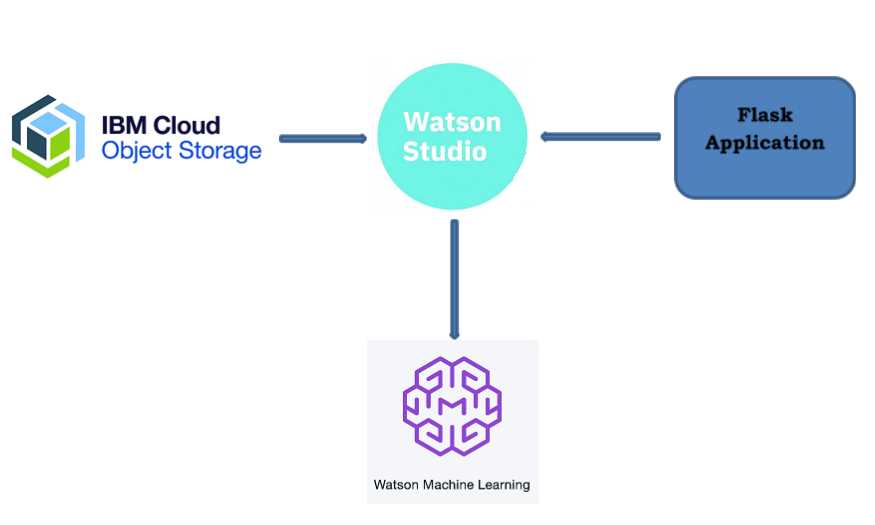
Amani and his colleagues developed a Decision Support System for Diabetes Prediction. They compared conventional machine learning with deep learning approaches. For conventional machine learning method, we considered the most commonly used classifiers: Support Vector Machine (SVM) and the Random Forest (RF). On the other hand, for Deep Learning (DL) we employed a fully Convolutional Neural Network (CNN) to predict and detect the diabetes patients. The proposed system is evaluated on publicly available Pima Indians Diabetes database which consisted of total 768 samples each with 8 features. 500 samples were labeled as non-diabetic while 268 were diabetic patients. The overall accuracy obtained using DL, SVM and RF was 76.81%, 65.38% and 83.67% respectively. The experimental results show that RF was more effective for diabetes prediction compared to deep learning and SVM methods.

1. **PROPOSED SYSTEM**

In our project we develop a Automatic AI model for the prediction of diabetes. For this we use IBM cloud platform. IBM Cloud provides a full-stack, public cloud platform with a variety of products in the catalog, including compute, storage, and networking options, end-to-end developer solutions for app development, testing and deployment, security management services, traditional and open-source databases, and cloud-native services.

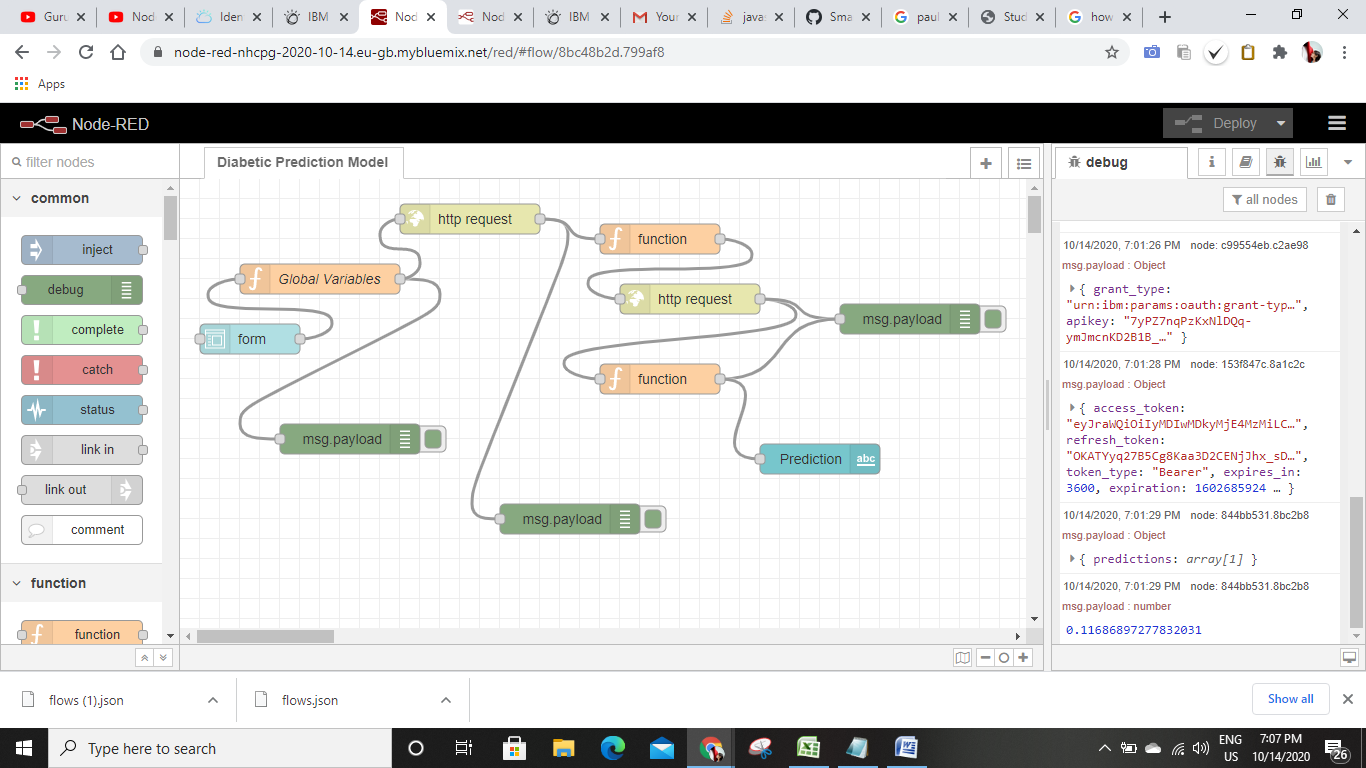
In IBM cloud we use Watson Studio, which provides a suite of tools for data scientists, application developers and subject matter experts, allowing them to collaboratively connect to data, wrangle that data and use it to build, train and deploy models at scale. We associate a Machine Learning Instance to Watson Studio. Then we upload our diabetes dataset, which consists of 768 records of women. The dataset is split up into 90/10, where 90 percent is used for training the model and the rest is used for testing. The machine learning algorithms, Decision Tree, Extra Trees classifier, Gradient Boosting, LGBM, Logistic Regression, Random Forest, XGB classifiers are applied and results are obtained. Both Cross Validation and Holdout operations are performed. After the model is developed, it is integrated to a web application. This auto AI web application is built using Node RED. Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things. The automated web application for Diabetic Prediction is built and deployed.

1. **BLOCK DIAGRAM**

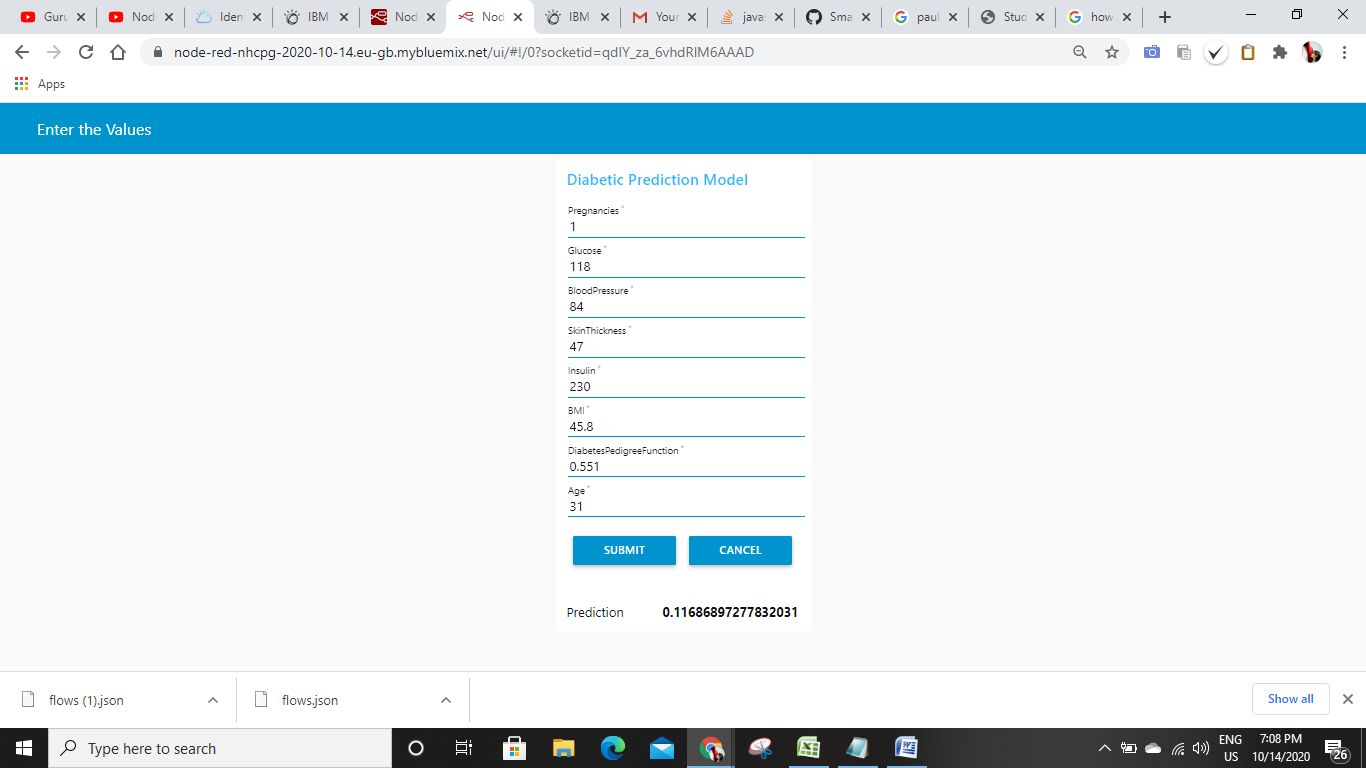


**Node- Red**

1. **FLOW CHART**



The Node Red Flowchart consists of Form node, which assigns label and variables to the attributes and grab their values. We use function node to set global variables, headers and payload for the values. The API key from the IBM cloud is given in function node, which is sent to http request node to fetch the access token. Next we add up another function node to call all the variables and send their values, payload, and headers to the scoring endpoint, which is pasted in the http request node. A text node is added at end to display the predicted value.



1. **RESULTS**

The model is trained using 90% of the dataset and tested by remaining 10%. The following machine learning algorithms are employed, Decision Tree, Extra Trees classifier, Gradient Boosting, LGBM, Logistic Regression, Random Forest and XGB classifiers. Both Holdout and Cross Validation operations are performed. In Cross Validation XGB classifier outperformed others by showing an accuracy of 77% whereas in Holdout, Gradient Boosting classifier stood first with 84%.

1. **ADVANTAGES AND DISADVANTAGES**

Machine Learning advance more information to doctors so that they can make better decisions about patient diagnoses and treatment options while understanding the possible outcomes and costs for each. The value of machine learning use cases in health insurance is its ability to process large datasets outside the scope of human capabilities, and then reliably transforms the analysis of the data into clinical insights that help doctors plan and provide care, which ultimately leads to better outcomes, costs lower than attention, and increasing patient satisfaction. By using this sophisticated type of analysis, we can provide better information to doctors at the point of patient care. Machine learning in health care helps in the customized treatments that can not only be more efficient and effective by pairing individual health with predictive analytics, but it is also ripe are for further research and better assessment of the disease.

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated. ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you. Machine Learning is autonomous but highly susceptible to errors. Suppose you train an algorithm with data sets small enough to not be inclusive. You end up with biased predictions coming from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of ML, such blunders can set off a chain of errors that can go undetected for long periods of time. And when they do get noticed, it takes quite some time to recognize the source of the issue, and even longer to correct it.

1. **CONCLUSION**

The main contribution of our study is to develop a web application for real time prediction of Diabetes, by giving the required values as input. The machine-learning techniques are employed to identify patients with high risk of developing Diabetes. These predictive models are validated using cross validation and holdout. These models can be set up in a computer program online to help physicians in assessing patient’s risk of developing Diabetes Mellitus.