Journal of Engineering Sciences

Vol 13, Issue 05, MAY/2022
ISSN NO:0377-9254

PREDICTION OF MULTIPLE DISEASES USING MACHINE LEARNING & API

K.Venkateswara Rao[1],V.Naga Sriram[2], SK.Shazreen[3],V.Chaitanya[4],V.Mahesh[5]
[1] Professor, Department of Computer Science and Engineering
[2,3,4,5]Student,Department of Computer Science and Engineering
[1,2,3,4,5]Vignan's Lara Institute of Technology & Science, Andhra Pradesh, 522213, India
*nagasriram04@gmail.com

Abstract:

Many contemporary machine learning models for health care data analysis are centered on a specific disease.For example,one analysis could be for diabetes, another for Kidney,and an other for Heart problems. There is no universal approach that can forecast multiple diseases using a single analysis. This article proposes a system for predicting a wide range of diseases using the Flask API. This article investigated diabetes, diabetic retinopathy, heartdisease, and kidney disease. Other disorders, such as skin conditions, fever analysis, and avariety of others,can be added later.Multiple sickness analysis was implemented using machine learning techniques, tensorflow, and the Flask API. Python pickling and unpickling are used to save and restore the model's behaviour. The behaviour of the final model will be saved in a python pickle file. The Flask API has been created. The user must send the disease's parameters as well as the disease's name when using this API. The Flask API will call the corresponding model and return the status of the patient.There levance of this research is to analyze the most common diseases so that the patient's status can be monitored and patients can be warned a head of time , lowering the deathrate.

Keywords:

Xgboost , Random Forest , Diabetes , Mortality , Flask API

Introdution:

During many analyses of existing health-care systems, just one disease was evaluated at a time. For instance, articles are used to examine diabetes,diabetes retinopathy,and heart disease prediction.
The majority of articles concentrate on a single ailment. When a corporation wants to analyze the health records of its customers, it must utilize a range of models. The existing system's technique is useful for analyzing only a single ailment.

Today's mortality rate has risen as a result of a lack of precise disease identification. Even if a patient has been healed of one ailment, they may still be suffering from another. In real life, I was in a similar position. After the accident, my father was able to recuperate. My father was discharged from the hospital, but he passed away a few days later.Internally, I'm having heart problems that I'm not awareof. Many such incidents have been reported in many people's life stories. When analyzing the disease, several current systems used only a few factors. As a result, identifying the diseases that will be caused as a result of the disease's influence may be impossible.

Diabetes can cause heartdisease, neuropathy, retinopathy, hearingloss, and dementia, for example. Data sets for diabetes analysis, diabetic retinopathy, heart disease, and kidney detection were examined in this work.Many other disorders, such as skin diseases, fever-related diseases, and others, may be included in the future. This study is adaptable. It has since been expanded to cover a wide range of illnesses. The developer must also include the model file for the new disease's analysis when adding a new illness analysis to this API.

In order to save model behaviour, the developer must plan for python choosing while establishing a new disease. The developer can load a pickled file to obtain model behaviour while using the Flask API. When a user wants to look into a patient's health, they may either predict a specific disease or see if the report has any parameters that might be utilised to forecast other diseases.

This analysis will yield the most relevant diseases. The goal of this essay is to prevent the death rate from rising day by day by notifying patients ahead of time about their medical conditions. The cost of patient analysis can be decreased because various illnesses models and forecasts are done in onespot.

Literature Review:

"Prediction of Cardiovascular Disease Using Machine Learning Algorithms" (2018). This paper contributes the correlative application and analysis of distinct machine learning algorithms in R software, which provides a quickway for users to employ machine learning algorithms in R software for forecasting the cardiovascular disease [1]. "A Proposed Model for Lifestyle Disease Predict using Support Vector Machine" (2018). The goal of this research is to better understand support vector machines and how they might be used to forecast lifestyle diseases in people [2]."Multi Disease Prediction Using Data Mining Techniques" (2017Two distinct data mining n order to determine the best classifier, their performance was compared. In data mining and machinelearning, creating precise and computationally efficient classifiers for medical applications is a key challenge [3]. "Machine Learning Algorithms for Heart Disease Prediction" (2018).classification algorithms were employed to forecast various diseases in this study, Two supervised data mining techniques, Nave Baye's Classifier and Decision tree classification, were used to predict the chance of a patient developing heart disease in this study. The Nave Baye's classifier successfully predicted heart disease patients 87% of the time, while the decision tree model properly predicted heartdisease patients 91% of the time [4]."Analysis of Heart Disease Prediction Using Data MiningTechniques"(2017) Heart disease is one of the main causes of death around the world, and early detection of heart disease is critical. The proposed new algorithm outperforms another algorithm interms of accuracy, according to this study [5]. "Medical Disease Symptoms Prediction Review "Using Data Mining Technique"(2017) The performance of medical disease prediction using datamining techniques is evaluated in this research. The medical diagnostic of disease data, such as cancer, liverdisease, and heartdisease,was classified by the classifier.In comparison to the standard cluster ensemble technique,the SVM algorithm identified data better[6].

Methodology:

Dataset Preparation:

The Pima Indian Diabetes Dataset, collected from a Frankfurt, Germany hospital, was initially used for diabetes research. Over 150 GB of picture data from the UCI machine learning collection is being used to research diabetic retinopathy. The study included data from people with cardiac disease in Cleveland, Hungary, and Switzerland. In addition to those supplied by visiting affiliated hospitals, other live data sets were employed in the current investigation.

This analysis is significantbecause itgathered the necessary variables thatwill cause the diseaseand, as a result of that disease, any other conditions that may arise.After performing this study,there is a chance of lowering the death rate since itwill bepossible to forecast the maximumnumber of

diseasesthat will occur, enable patients to be informed about their treatment options inadvance Industry standards are followed when designing train sets and tests. Using the Scikit-learntrain test split method, the data is divided into 70 percent for training and 30 percent for testing.
Diabetesfeaturetrain,diabetesfeature_test,diabeteslabeltrain,diabeteslabel_test=traintestsplit;
diabeteslabeltraindiabeteslabel_test=traintestsplit;diabeteslabeltrain,
diabeteslabel_test=traintests(diabetesfeatures,diabeteslabel,testsize=0.3,randomstate=0)
(diabetesfeatures,diabeteslabel,testsize=0.3,randomstate=0)
(diabetesfeatures,diabeteslabel,testsize=0.3,randomstate=0)

MachinelearningandDeeplearning:

The main purpose of this paper is to develop a multi-disease prediction model,therefore the machine learning and deep learning techniques use dare briefly explained.Different machine learning and deep learning approaches are used to analyze diabetes,heartdisease prediction, and kidney detection. Navie Bayes categorization is similar to logistic regression algorithm, Methods such as SVM, Decision Tree Algorithm, Random Forest Algorithm, and others are used to identify the patient's status.The accuracy of logistic regression for diabetes analysis is 92 percent,Random forest for heartdisease classification is 95 percent, The diabetes retinopathy analysis includes retinal pictures.To analyze the photos, I utilized the Python Tensorflow package. Tensorflow convolution neural networks are used to build the model,which is then tested using the test set.The developed model was 89 percent accurate.

SAVINGMODELBEHAVIORWITHPYTHONPICKLING

Pythonpicklingforheartdiseasepredictiondataset:

After the data set has been processed using the training and testsets, the optimal algorithm with the highest accuracy is chosen. Python pickling may be used to store model behaviour. To serialise and de-serialize the Python object structure, the pickle model is employed. An object can be pickled and saved to disc in Python. A pickle file is a character stream in Python that provides all of the information needed to recreate an object in an other script.

MODELFORPREDICTIONOFMULTIPLEDISEASES

Loadingpicklefiletopredictthedisease:

The model behaviour is saved as a pickle file once the model is finished. Four disorders were considered in this study, resulting in four pickle files. Load all picklefiles into the python script that will perform the multi-disease analysis before starting to analyse the disease.

Dataflow:

The data flow in analysis is depicted in Figure 1. The initial stage in the analysis is to pre-process the data. Because the blood pressure of a living human being cannot be zero, pre-processing is required. Such records should be pre-processed. After pre-processing, data set preparation, and model building for various diseases, with available data sets.Pickle files include all of the model's behaviour. FlaskAPI was created.

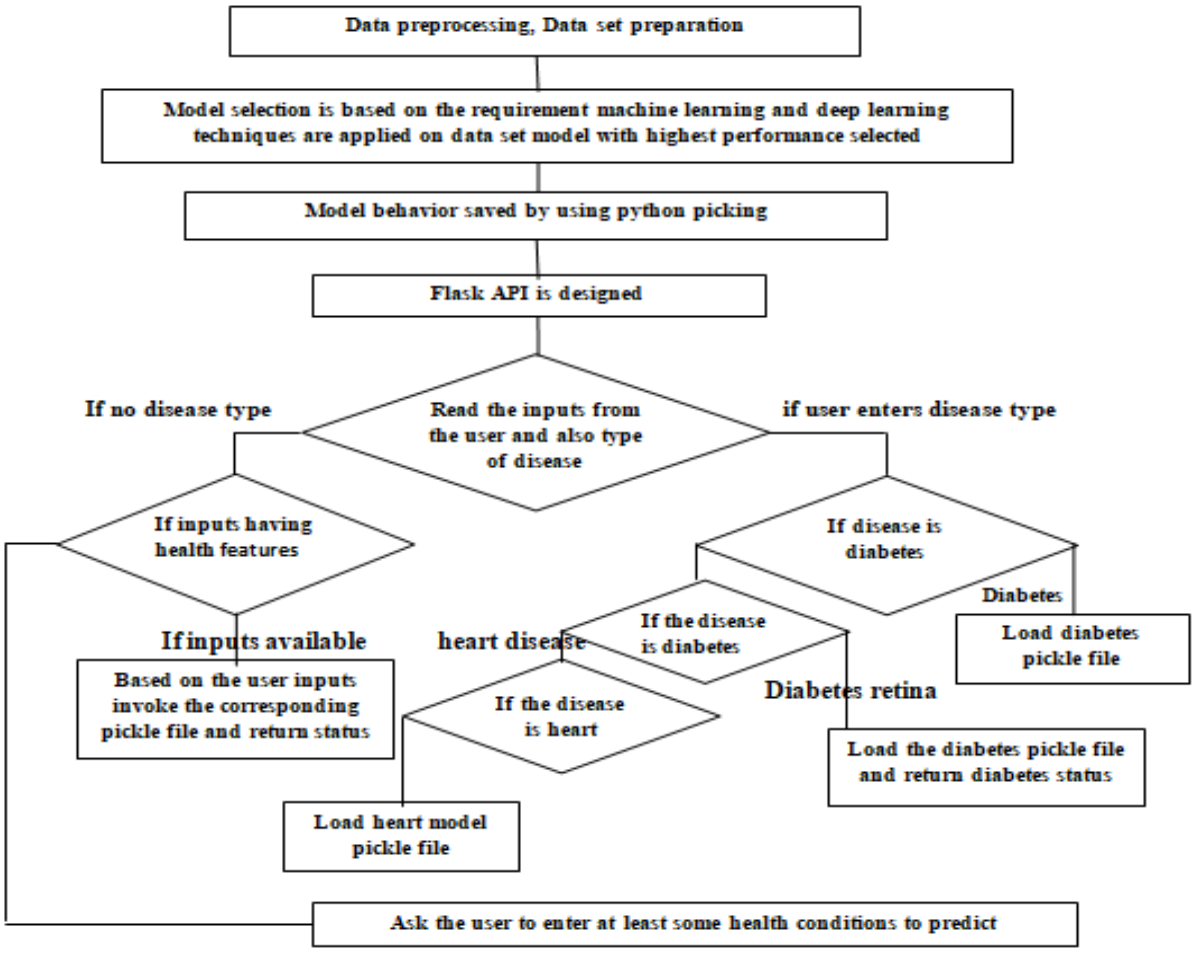
The training of models is an important phase in Machine Learning projects. The two basictechniques

to machine learning are supervised and unsupervised learning. Initially, ourmodellargelyuses thefirststrategy.

In Supervised Learning, such as the Training Set, are used to train the system,and then the model is requested to do a task.Using the test set, forecast new values.The splitting of a dataset is critical for achieving high model accuracy. The most commonly used percentage. In our system, we seek to apply many algorithms to the training dataset first, then select the optimal model algorithm based on the model's Confidence and testing dataset accuracy, Then, to get correct findings, apply it to the testing dataset.

The data flow is as following steps:

1. Datapreprocessing , DataPreparation
2. ModelSelection
3. Modelbehaviour Saved by using python pickling
4. Flask Api is designed
5. Read input from user & also type of disease
6. Appropriate Model is loaded
7. Ask the user to enter at least some health conditions to predict



XGBOOSTAlgorithm :

XGBoost is a gradient boosting-based decision-tree-based ensemble Machine Learning approach. In

unstructured data prediction, artificial neural networks outperform all other algorithms or frameworks (pictures,text,etc.). Decision tree-based algorithms, on the other hand, are now considered best-in-class for small to medium structured tabular data. The graph below depicts how tree-based algorithms have evolved over time.

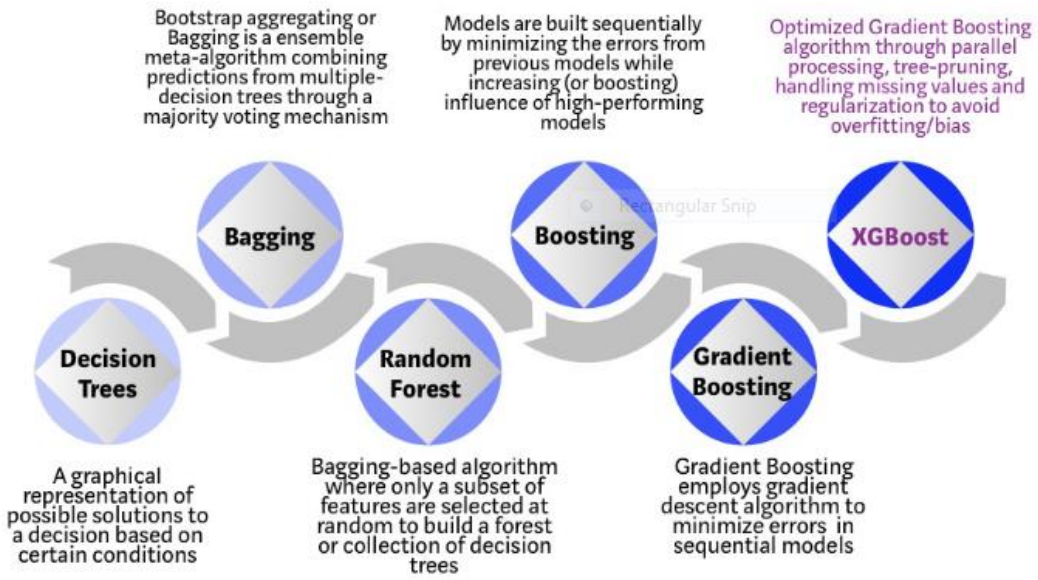


Fig. The XGBoost algorithm evolved from Decision Trees.

Decision Tree:
Every recruiting manager has a set of standards, such as educational attainment, years of experience, and interview performance. A hiring manager conducting interviews with candidates basedon his or her own set of criteria is analogous to a decision tree.
Bagging:
Imagine that instead of a single interviewer, there is a panel of interviewers, each of whom gets a vote. Bagging, also known as bootstrap aggregation is the act of merging the input so fall interviewees in order to reach a final judgment through a democratic voting procedure.
Random Forest:

It's a bagging-based strategy, but only a subset of features is randomly selected. To put it another way, each interviewer will only assess the applicanton a set of qualifications chosen at random(e.g.a technical interview for testing programming skills and a behavioral interview for evaluating non-technicalskills.
Boosting:

This is a different strategy in which each interviewer changes the evaluation criteria based on the preceding interviewer's response.By adopting a more dynamic evaluation procedure, this 'boosts' the efficiency of the interview process.
Gradient Boosting:
A type of boosting that uses a gradient descent technique to eliminate errors; for example,strategy consulting firms use case interviews to select out less qualified candidates.

XGBoost:

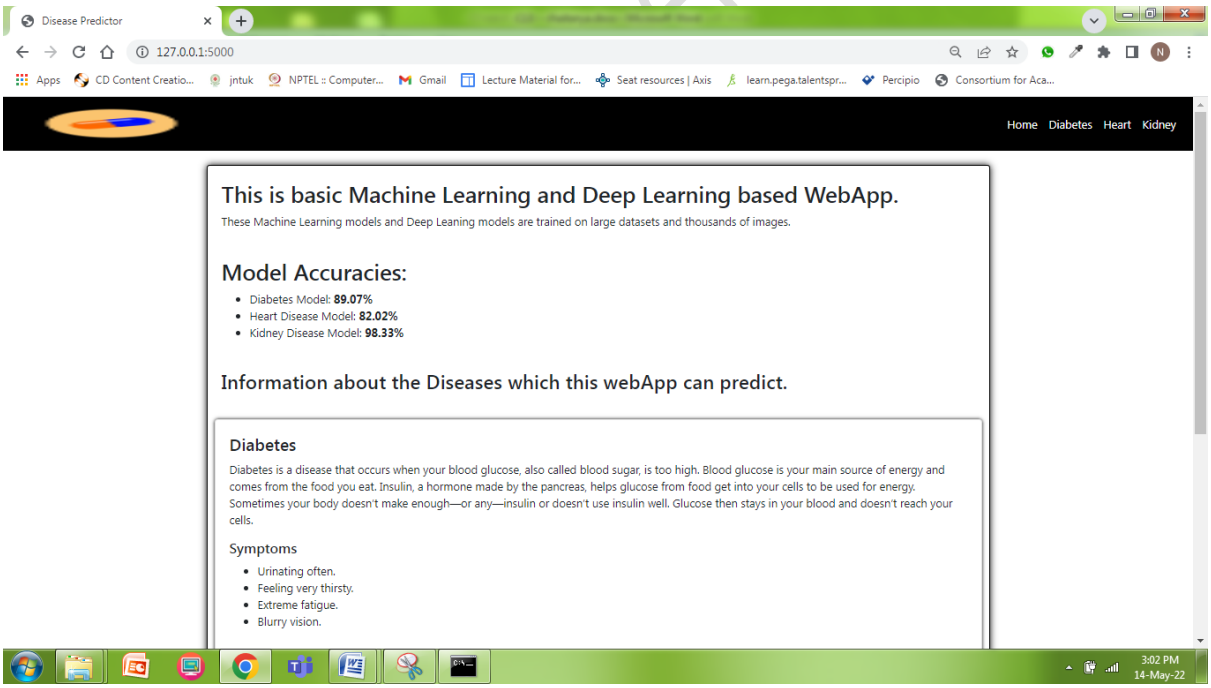
Consider XGBoost to be gradient boosting on steroids (after all, it's called "Extreme Gradient Boosting" for a reason!). It is a perfect combination of software and hardware optimization techniques that produce superior results in the shortest amount of time with the least amount of computing resources.

RandomForestAlgorithm:

Randomforest is a learning algorithm that is supervised.It creates a"forest"out of an ensemble of decision trees,which are commonly trained using the "bagging" method.The bagging method's basic premise is that combining several learning models improves the overall output.

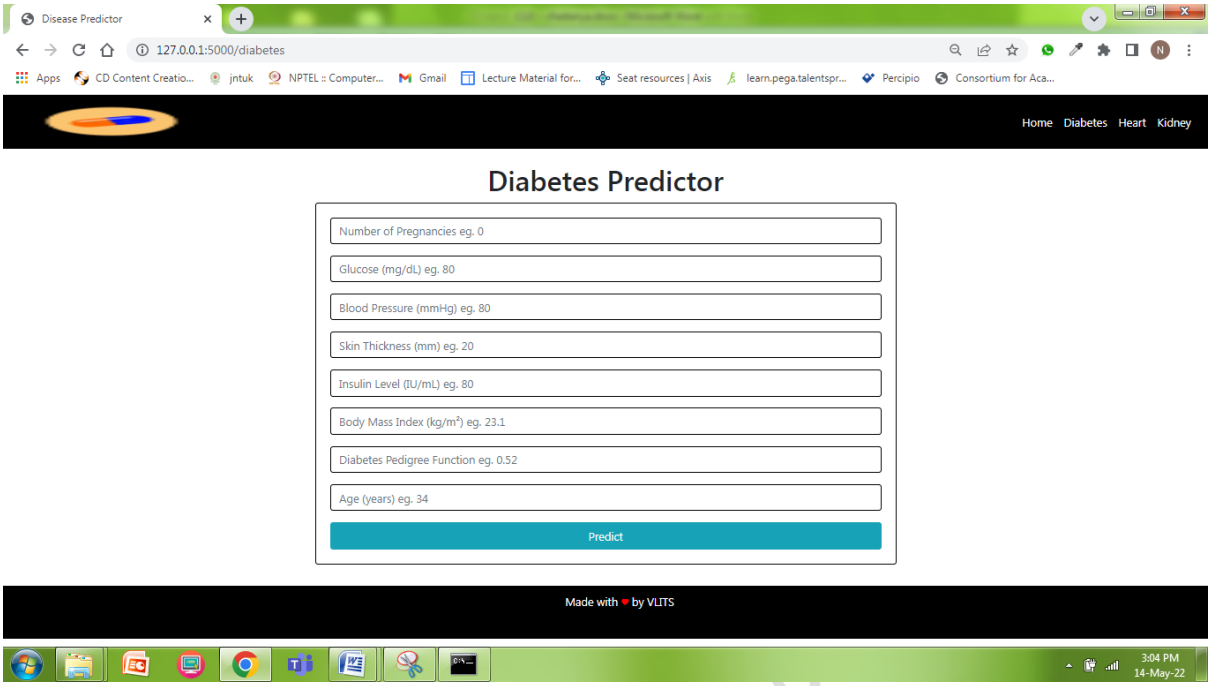
We've also created a web application using Flask to predict term deposit subscriptions, with the following web page.
The Home page of website is as follows:

HOME PAGE:



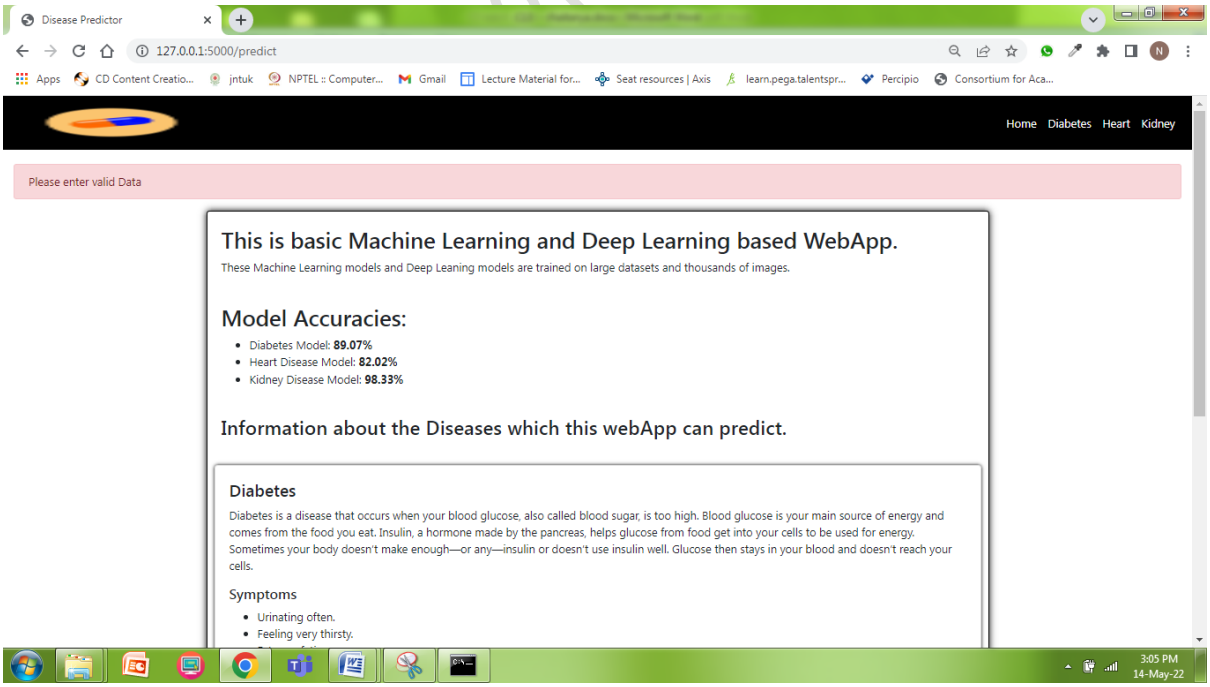
By selecting an appropriate disease the appropriate parameters will be displayed accordingly.
Out of three diseases namely diabetes , Kidney , Heart .The Diabetes Parameters are shown below
That heart & kidney also have respective parameters to take the parameters as a input.
The follwing shows input Parameters of Diabetes Disease.

DIABETES PARAMETERS



After giving values to the above form and then by clicking the predict button we get the outcome as follows,if we enter invalid input to the parameters:

RESULT:



Conclusion:

A multi-disease prediction model predicts multiple diseases at the same time. The disease will be

predicted based on the user's input..The user will be offered the option.The corresponding disease model will be triggered and predicted based on the user's inputs if the user wishes to predict a specific disease or if the user doesn't pick a disease category.The advantage of a multi-disease prediction model is that it can anticipate the likelihood of multiple diseases occurring in the future,as well as minimize the mortality ratio.

References:

[1]. Jaymin Patel, Prof.Tejal Upadhyay, Dr. Samir Patel "Heart disease prediction usingMachine learning and Data Mining Technique" Volume 7.Number1 Sept 2015March2016.
[2]. G. Parthiban, S.K.Srivasta "Applying Machine learning methods in Diagnosing HeartdiseaseforDiabeticPatients"International Journal of AppliedInformation Systems(IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USAVolume3–No.7,August2012.
[3].Thenmozhi.K and Deepika.P,HeartDisease Prediction using classification withdifferentdecisiontreetechniques.InternationalJournalofEngineeringResearch&GeneralScience,Vol2(6),pp6-11,Oct2014.
[4]."DiseasePredictionUsingMachineLearningOverBigData"VinithaS,SweetlinS, VinushaHandSajiniS(2018).
[5]."HeartDiseasePredictionusingLogisticRegressionAlgorithmusingMachineLearning"ReddyPrasad,Pidaparthi Anjali,S.Adil,N.Deepa(2019)
[6]."HeartDiseasePredictionUsingDataMiningTechniques"H.BenjaminFredrickDavidandS.AntonyBelcy(2018).
[7]."MultiDiseasePredictionUsingDataMiningTechniques"K.Gomathi,Dr.D.ShanmugaPriyaa(2017).
[8] " Rotating Solar Trees ", Lecture Notes in Electrical Engineering 601, Springer Nature Singapore Pte Ltd. 2020, Page No: 482-487.
[9] "Prediction of Dengue Disease Cases by ML Techniques", International Journal of Data Science and Machine Learning (IJDSML), ISSN : 2692-5141, Vol-1 Issue-1, Sep 2020, Page No: 1-6.
[10] "An Approach For Detecting Phishing Attacks Using Machine Learning Techniques", Journal of Critical Reviews (JCR), ISSN : 2394-5125, Vol-7 Issue-18, Jun 2020, Page No: 321-324.
[11] "Disease Prediction and Diagnosis Implementing Fuzzy Neural Classifier based on IoT and Cloud", International Journal of Advanced Science and Technology (IJAST), ISSN : 2005-4238, Vol-29 Issue-5, May 2020, Page No: 737-745.