Diabetic Metillus Predictions using IBM Auto AI

Team Number:- RSIP Career Basic ML 018

Name:- Preeti Meena

Learning track:- Machine Learning

College Name:- Chameli Devi Group Of Institution, Indore (M.P)

INDEX

|  |  |
| --- | --- |
| **INTRODUCTION** | |
| 1.1 Overview | |
| 1.2 Purpose | |
| **LITERATURE SURVEY** | |
| 2.1 Existing problem | | |
| 2.2 Proposed solution | | |
| **THEORITICAL ANALYSIS** | |
| 3.1 Block diagram | |
| 3.2 Hardware / Software designing | | |
| **EXPERIMENTAL INVESTIGATIONS** | | |
| **FLOWCHART** | |
| **RESULT** |  |
| **ADVANTAGES & DISADVANTAGES** | | |
| **APPLICATIONS** | |
| **CONCLUSION** | |
| **FUTURE SCOPE** | |
| **BIBILOGRAPHY** | |
| **APPENDIX-** source code | |
|  | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1. INTRODUCTION**  **1.1 Overview**  Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors as diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Data science methods have the potential to benefit other scientific fields by shedding new light on common questions. One such task is to help make predictions on medical data. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience.The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques.This project also aims to propose an effective technique for earlier detection of the diabetes disease.  **1.2 Purpose**  Classification is one of the most important decision making techniques in many real world problem. In this work, the main objective is to classify the data as diabetic or non-diabetic and improve the classification accuracy. For many classification problem, the higher number of samples chosen but it doesn’t leads to higher classification accuracy. In many cases, the performance of algorithm is high in the context of speed but the accuracy of data classification is low. The main objective of our model is to achieve high accuracy. Classification accuracy can be increase if we use much of the data set for training and few data sets for testing. This survey has analyzed various classification techniques for classification of diabetic and non-diabetic data.  **2. LITERATURE SURVEY**  **2.1 Existing problem**  Diabetes mellitus is a chronic disease characterized by hyperglycemia. It may cause many complications. According to the growing morbidity in recent years, in 2040, the world’s diabetic patients will reach 642 million, which means that one of the ten adults in the future is suffering from diabetes. There is no doubt that this alarming figure needs great attention. With the rapid development of machine learning, machine learning has been applied to many aspects of medical health for accurate predictions.   |  | | --- | | **2.2 Proposed solution** |   This project prevents the people from the avalanche by priory informing them there is a chance to the occurrence of avalanche or not. The model gets the data from the IOT based sensors. After that we want to process those data using a suitable algorithm, then our model display whether the avalanche occur or not and how strength it was. To analyse the data coming from different sensors we are applying various machine learning algorithms. If there is a chance of avalanche then the notification will be sent to people so that they can take decisions accordingly and the model is been built in Auto AI.     |  |  | | --- | --- | |  |  | | **3.THEORITICAL ANALYSIS**  **3.1 Block diagram**    **3.2 Software designing**  **IBM Cloud Computing**  Cloud computing, often referred to as simply “the cloud,” is the delivery of on-demand computing resources — everything from applications to data centers — over the internet on a pay-for-use basis. Services provided by IBM cloud computing are Iaas, Paas, Saas, serverless computing.   * Elastic resources — Scale up or down quickly and easily to meet demand * Metered service so you only pay for what you use * Self service — All the IT resources you need with self-service access     **IBM Watson Studio**  IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. With its open, flexible multicloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI, such as:  Automate AI lifecycle management with AutoAI   * Visually prepare and build models with IBM SPSS Modeler * Build models using images with IBM Watson Visual Recognition and texts with IBM Watson Natural Language Classifier * Deploy and run models through one-click integration with IBM Watson Machine Learning * Manage and monitor models through integration with IBM Watson OpenScale   **NODE RED**  Node-RED is a flow-based programming tool, originally developed by IBM’s Emerging Technology Services team and now a part of the JS Foundation.  Invented by J. Paul Morrison in the 1970s, flow-based programming is a way of describing an application’s behaviour as a network of black-boxes, or “nodes” as they are called in Node-RED. Each node has a well-defined purpose; it is given some data, it does something with that data and then it passes that data on. The network is responsible for the flow of data between the nodes.  Node-RED consists of a Node.js based runtime that you point a web browser at to access the flow editor. Within the browser you create your application by dragging nodes from your palette into a workspace and start to wire them together. With a single click, the application is deployed back to the runtime where it is run.  The palette of nodes can be easily extended by installing new nodes created by the community and the flows you create can be easily shared as JSON files. |  | | Pictorial representation:-         |  | | --- | | **4. EXPERIMENTAL INVESTIGATIONS**  The Pima Indian Diabetes Dataset consists of information on 76 patients (268 tested\_positive instances and 500 tested\_negative instances) coming from a population near Phoenix, Arizona, USA. Tested\_positive and tested\_negative indicates whether the patient is diabetic or not, respectively. Each instance is comprised of 8 attributes, which are all numeric. These data contain personal health data as well as results from medical examinations. The detailed attributes in the dataset arelisted as follows, and fig 2 shows some samples extracted from the  dataset. |  * Diastolic blood Pressure(pres) * Triceps skin fold thickness(skin) * Number of Tests(test) * Diabetes pedigree function (pedi) * Age (age) * Class variable (class) * Plasma glucose concentration at 2 hr in an oral glucose tolerance test(plas) * Number of times pregnant(preg) | | |  | | |  | |       Figure 2   |  |  |  | | --- | --- | --- | |  |  |  |   **5. FLOWCHART**    Figure.3  6. **RESULT**  Using Red-Noded UI, it was convenient for us to study the result ofthe experiment through a visualized interface. We analyzed and evaluated our model based on the following aspects. The result is shown in figure 4 and figure 5. |
| |  | | --- | | **7.Advantages & Disadvantages**  **Advantages** |  |  |  | | --- | --- | | |  | | --- | | * The advantages of this method is the whole samples in the dataset are trained and tested which can avoid the highest variance * It does not require blood sample can be tested in fasting, random, or rapid processing time, inexpensive * Easy to obtain ,no fasting is required.   **Disadvantages:-** | | |
| * Requires prompt processing thus the potential for error, point measurement can be affected by short term lifestyle changes, risks at phlebotomy. |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |  |  |
| --- | --- | --- |
|  | | |
| 8.**APPLICATION**  It is more convenient and efﬁcient for people to obtain an application about health management of DM on their mobile devices [33–37]. We are currently developing an application that will provide reasonable And rational health suggestions to the high-risk group. Diabetes patients can conveniently use this application to test their blood glucose level, blood pressure, and heart rate. Furthermore, this medical data will be saved in for  further procedures about data visualizing and mode optimization. This will not only help people understand their health conditions, but will also help them create a healthy lifestyle. | |  |
|  |  |  |
|  | | |
|  | |  |
|  | |  |
| .  **9.Conclusion**  Machine learning has the great ability to revolutionize the diabetes risk prediction with the help of advancesnced computational methods and availability of large amount of epidemiological and genetic diabetes risk dataset. Detection of diabetes in its early stages is the key for treatment. This work has described a machine learning approach to predicting diabetes levels. The technique may also help researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the disease status  **10. Future Scope**  For future work, it is necessary to bring in hospital’s real and latestpatients’ data for continuous training and optimization of our proposedmodel. The quantity of the dataset should be large enough for training  and predicting . Some advanced algorithms and models should beapplied in the study of DM. Grading forecasting standards are alsonecessary for potential diabetes patients. Developing a series of rules and  standards is a valid method to prevent people from developing DM. Basedon that, a more effective model for predicting DM and grading potentialpatients is presented. This will help to lower the growth rate of diabetesand eventually decrease the risk of developing DM.  **11. Bibliography**   * Komi, Zhai. 2017. Application of Data Mining Methods in Diabetes Prediction . * Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus, Omar Kassem Khalil Aissa Boudjella, 2016 Sixth International Conference on Developments in eSystems Engineering. * Alan Siper, Roger Farley and Craig Lombardo, “Machine Learning and Data Mining Methods in Diabetes Research”, Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 6th, 2005. * .Devi, M. Renuka, and J. Maria Shyla. "Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus." International Journal of Applied Engineering Research 11.1 (2016): 727-730. * Berry, Michael, and Gordon Linoff. Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc., 1997 * Witten, Ian H., et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016. * . Giri, Donna, et al. "Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform." Knowledge-Based Systems 37 (2013): 274-282. | |  |
| **APPENDIX**  **\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***  **A. Source code(json file)** | |  |
| [{"id":"3f7710dc.fff22","type":"tab","label":"Flow 1","disabled":false,"info":""},{"id":"54828939.24ad08","type":"function","z":"3f7710dc.fff22","name":"PreToken","func":"global.set(\"pg\",msg.payload.pg)\nglobal.set(\"pa\",msg.payload.pa)\nglobal.set(\"ps\",msg.payload.ps)\nglobal.set(\"sk\",msg.payload.sk)\nglobal.set(\"ts\",msg.payload.ts)\nglobal.set(\"ma\",msg.payload.ma)\nglobal.set(\"pd\",msg.payload.pd)\nglobal.set(\"a\",msg.payload.a)\nvar apikey=\"9xScp0T-xwdyGxRKU4GTYghvfDk0gtpJzA4qwv03SsuD\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant\_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;","outputs":1,"noerr":0,"x":255,"y":316.0000047683716,"wires":[["39212fac.36311"]]},{"id":"39212fac.36311","type":"http request","z":"3f7710dc.fff22","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":454,"y":305,"wires":[["1dd6448f.5b904b"]]},{"id":"57369875.d2d9d8","type":"inject","z":"3f7710dc.fff22","name":"","topic":"","payload":"","payloadType":"date","repeat":"","crontab":"","once":false,"onceDelay":0.1,"x":100.5,"y":230.00000286102295,"wires":[["54828939.24ad08"]]},{"id":"ddc42c8.89a7bd","type":"debug","z":"3f7710dc.fff22","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":893.0000114440918,"y":47.00000190734863,"wires":[]},{"id":"1dd6448f.5b904b","type":"function","z":"3f7710dc.fff22","name":"Pre Prediction","func":"var pg = global.get('pg')\nvar pa = global.get('pa')\nvar ps = global.get('ps')\nvar sk = global.get('sk')\nvar ts = global.get('ts')\nvar ma = global.get('ma')\nvar pd = global.get('pd')\nvar a = global.get('a')\nvar token=msg.payload.access\_token\nvar instance\_id=\"6626f614-8470-4cc6-9164-7c72f82a0f0b\"\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance\_id}\nmsg.payload={\"input\_data\": [{\"fields\": [\"preg\", \" plas\", \"pres\", \"Skin \", \"test\", \"mass\", \"pedi\", \"age\"], \"values\": [[pg,pa,ps,sk,ts,ma,pd,a]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":651.0000076293945,"y":268.00000381469727,"wires":[["a822e570.a9b388"]]},{"id":"a822e570.a9b388","type":"http request","z":"3f7710dc.fff22","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v4/deployments/9443d5ce-7bdc-431a-ae01-2847bf0e944b/predictions","tls":"","persist":false,"proxy":"","authType":"","x":744.5000076293945,"y":212.00000381469727,"wires":[["a0f2d449.c257e8"]]},{"id":"86ee3466.aeb058","type":"ui\_form","z":"3f7710dc.fff22","name":"","label":"","group":"70aa036d.e2a94c","order":0,"width":0,"height":0,"options":[{"label":"preg","value":"pg","type":"number","required":true,"rows":null},{"label":"plas","value":"pa","type":"number","required":true,"rows":null},{"label":"pres","value":"ps","type":"number","required":true,"rows":null},{"label":"Skin","value":"sk","type":"number","required":true,"rows":null},{"label":"test","value":"ts","type":"number","required":true,"rows":null},{"label":"mass","value":"ma","type":"number","required":true,"rows":null},{"label":"pedi","value":"pd","type":"number","required":true,"rows":null},{"label":"age","value":"a","type":"number","required":true,"rows":null}],"formValue":{"pg":"","pa":"","ps":"","sk":"","ts":"","ma":"","pd":"","a":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":90,"y":380,"wires":[["54828939.24ad08"]]},{"id":"4e3eb0c2.96ae7","type":"ui\_text","z":"3f7710dc.fff22","group":"70aa036d.e2a94c","order":1,"width":0,"height":0,"name":"","label":"class(1=yes,if 0=no)","format":"{{msg.payload}}","layout":"row-spread","x":920,"y":360,"wires":[]},{"id":"a0f2d449.c257e8","type":"function","z":"3f7710dc.fff22","name":"","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;","outputs":1,"noerr":0,"x":680,"y":120,"wires":[["4e3eb0c2.96ae7","ddc42c8.89a7bd"]]},{"id":"70aa036d.e2a94c","type":"ui\_group","z":"","name":"Default","tab":"b81479ef.da86d8","order":1,"disp":true,"width":"6","collapse":false},{"id":"b81479ef.da86d8","type":"ui\_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]   |  |  | | --- | --- | |  |  | |  | | |  | | | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |