**RV College of Engineering ®, Bengaluru-59**

**(Autonomous Institution Affiliated to VTU)**

**Department of Electronics and Communication Engineering**

**16EC73P Minor Project**

**Synopsis**



**Diabetes Prediction using machine learning**

|  |  |
| --- | --- |
| **Name of the students** | **USN** |
| Manoj Naik | 1RV17EC073 |
| Mayur Raj Singh Chauhan | 1RV17EC075 |
| Paurush Gupta | 1RV17EC097 |

|  |  |
| --- | --- |
| **Guide Details** | |
| Guide Name | Dr. Rajani Katiyar |
| Guide Signature with date |  |

**INTRODUCTION:**

Diabetes is an illness that affects the ability of the body in producing the hormone insulin, which in turn makes the metabolism of carbohydrate abnormal and raises the levels of glucose in the blood. A person generally suffers from a high sugar level in blood which can have severe effects on other human organs. Insulin is an essential hormone produced by the pancreas that allows the cells to absorb glucose (blood sugar) from food supplies in order to provide them the necessary energy. Some of the symptoms of diabetes are Intensify thirst, Intensify hunger, and Frequent urination.

In medicine, doctors and current research confirm that if this disease is discovered at an early stage, the chances of recovery will be greater. But this identifying process is tedious, i.e. visiting a diagnostic center and consulting a doctor takes a lot of time and waste the budget of health care systems and people every year. But Machine learning can be used as a solution to this problem. The Machine learning algorithms use recorded datasets of former patient’s information to prepare a model and then use this model with information of an unseen patient to predict if the patient has the desired disease or not.

**LITERATURE SURVEY:**

1. 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.

* A comparative analysis of machine learning and deep learning-based algorithms for prediction of diabetes.
* It showed that RF was more effective for classification of the diabetes in all rounds of experiments which produced overall accuracy for diabetic prediction to be 83.67%. The prediction accuracy for SVM reached 65.38% while DL method produced 76.81% on dataset.

1. D. Dutta, D. Paul and P. Ghosh, "Analysing Feature Importances for Diabetes Prediction using Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018.

* It concludes that Random Forest is the most ideal algorithm for predicting Diabetes, which gives an accuracy of around 84%.
* If people want to prevent Diabetes, they should really keep their glucose level down and with increase in age they should follow a proper diet.
* Also, people born in families having a diabetic history, they should really take care of themselves.

1. P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019**.**

* SVM: Are very good when we have no idea on the data. Even with unstructured and semi structured data.
* The drawback of the SVM algorithm is that to achieve the best classification results for any given problem, several key parameters are needed to be set correctly.
* Decision tree: It is easy to understand and rule decision tree.
* Naive Bayes: It is robust, handles the missing values by ignoring probability estimation calculation.
* ANN: Gives good prediction and easy to implement. Difficult with dealing with big data with complex model. Require huge processing time.

1. R. Mirshahvalad and N. A. Zanjani, "Diabetes prediction using ensemble perceptron algorithm," 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN), Girne, 2017.

* This study proposed a learning algorithm to help the prediction of undiagnosed patients in an easy, accessible and affordable way for people around the world.
* The proposed algorithm, ensembles boosting algorithm with perceptron algorithm, which uses more than one weight vector for the test data classification.
* The proposed algorithm is validated on three different NHANES datasets confirming that AUC value improves from 0.72 to 0.75 by the proposed algorithm.
* In addition, it is shown that execution time overload of proposed algorithm is not considerably higher compared to perceptron algorithm.

1. H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani and K. Qaraqe, "Predicting Diabetes in Healthy Population through Machine Learning," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 2019.

* In this paper, they have used the linear support vector machines to construct a prediction model of future development of type-2 diabetes.
* The outcomes of the study show that high values of glucose observed at the 2 h mark during the OGTT may strongly indicate the potential risk of future development of type-2 diabetes.
* In a possible extension of this study, the prediction models may be applied on other similar datasets that include the OGTT measurements.

1. K. VijiyaKumar, B. Lavanya, I. Nirmala and S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019.

* In this study, systematic efforts area unit created in coming up with a system that finally ends up among the prediction of illness like genetic defect.
* Throughout this work Random Forest algorithms area unit studied and evaluated on varied measures.
* 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 using machine learning technique which provides advance support for predicting the accuracy rate of diabetes.

1. P. K. Saha, N. S. Patwary and I. Ahmed, "A Widespread Study of Diabetes Prediction Using Several Machine Learning Techniques," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019.

* In this study, there are several comparsion between algorithms with different preprocessing techniques and identify algorithms best performance in which preprocessing technique.
* It also shows that Neural Network was given best accuracy (80.4%) than any other methods.
* Comparing the execution time of several methods in it’s the best accuracy and found Naive Bayes was taken less execution time than any other methods.

**MOTIVATION:**

Diabetes is considered as one of the deadliest and chronic diseases and many complications may occur if it remains untreated or unidentified. In medicine, doctors and current research confirm that if this disease is discovered at an early stage, the chances of recovery will be greater. This motivated us to build a user-friendly AI model which can predict the probability for a person of having diabetes using recorded data of former patients.

**OBJECTIVES:**

* To build an efficient model that can predict the probability of having Diabetes.
* To visualize various independent variables like the number of pregnancies the patient has had, their BMI, insulin level, age, and dependent variable
* To increase the model efficiency of the model with help of data processing and hyperparameter tuning.
* To build a cloud-based user-friendly interface.

**METHODOLOGY:**

The First Step is to create a model using machine learning. This involves Importing the necessary libraries, loading the dataset, then training the model. The accuracy of the model created should be improved using hyperparameters. Then Evaluating the model. After the model is created, a web application is created using a flask and connecting it with the model. Then the project is committed to GitHub. Then the model is deployed in the cloud using Heroku.

**HARDWARE / SOFTWARE TOOLS USED FOR THE PROJECT:**

1. Python

* Python is a widely used high-level programming language for general-purpose programming.

1. Jupyter notebook

* JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.

1. Flask

* Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application.

1. GitHub

* GitHub is a web-based platform used for version control. Git simplifies the process of working with other people and makes it easy to collaborate on projects.

1. Heroku

* Heroku is a cloud Platform as a Service (PaaS). Developers use Heroku to deploy, manage, and scale modern apps.

**REFERENCES:**

|  |  |
| --- | --- |
| 1. 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. 2. D. Dutta, D. Paul and P. Ghosh, "Analysing Feature Importances for Diabetes Prediction using Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 924-928. 3. P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019. | |
|  |
|  |