Prediction of Type 1 Diabetes Mellitus using Datamining Techniques

K.Saminathan

Abstract: Type1 diabetes is a sickness occurs when your immune system fighting against infection, affects and erode the insulin generating beta cells of the pancreas. In general, when the blood sugar stage increases, the pancreas makes more insulin. Insulin helps to go sugar out of the blood so it can be used for liveliness. Type 1 diabetes occurs due to the immune system which affects cells in the pancreas that make insulin. The pancreas cannot make adequate insulin, so the blood sugar level continues to increase. According to the children history of type 1 diabetes may enhance risk of their life. Type 1 diabetes cannot be cured, but it can be controlled and managed. In this study we use Naive Bayes, linear regression and k-means algorithm for data analysis and prediction. It predicts the diabetes affected children with maximum level of accuracy 96% by using of data mining algorithms.

Keywords: Type 1 Diabetes, Naive Bayes, Linear Regression, K-Means, Decision Stump.

I. INTRODUCTION

Diabetes is a chronic disease that comes when the pancreas is not able to make enough insulin, or when the body cannot fulfill good use of the insulin it produces [1]. It has three types.
- Type 1
- Type 2
- GDM

Type 1 diabetes is a disease in which the body makes little insulin or not to control blood sugar levels. Type 1 diabetes is kown as an insulin-dependent diabetes. It is otherwise named as juvenile diabetes. Digestion process converts the food which we intake is broken down into vital components. In general one of the major energy producing food material is carbohydrate. It is further refined into sugars and glucose. Human body cells obtain energy mainly from glucose. Glucose usually invades into cells to give energy to the body [2]. Insulin is a hormone made by the pancreas that acts like a key to control blood sugar. If our body doesn’t produce enough insulin to control the sugar, it may develop hyperglycemia (high blood sugar). Hyperglycemia provides long term complications. To avoid those complications, the body needs insulin. But People with type1 diabetes cannot make insulin because the pancreas are damaged and destroyed. So they need insulin injection to control blood sugar. Insulin disagreement actually had symptoms like extreme hunger, weight loss, fatigue, irritability or behavior changes [3].

Diabetes patients’ count grows day by day in all over the world. Classification techniques are preferred and extensively applied in the medical field. The collection of data are classified into various classes according to required constrains that helps to predict the disease. Researchers were conducted experiments to diagnose the diseases using different classification algorithms of machine learning approaches like J48, Support Vector Machine, Naive Bayes, Decision Tree, Decision Table. Researchers have determined that machine learning algorithms works better in diagnosing different diseases. Data Mining and Machine learning algorithms have the competence of handling abundant amount of data from several different sources and integrating the background information for study of possible cases. The proposed work focuses on children who are suffering from diabetes. In this work classification algorithms and also clustering algorithms are used to identify the prediction of diabetes from patients’ data set. Experimental performances of all the algorithms are compared on various measures and achieved good accuracy.

II. RELATED WORKS

Prediction of diabetes type-1 using classification algorithms discuss the multiple and different strategies, which are widely used in the medical field. The classification process of data into diverse classes according to derived constrains comparatively an individual classifier [4]-[7]. Diabetes prediction system has the main objective of prediction of diabetes among particular or certain age of people suffered by type–I [8]. Machine learning system is applied to design the diabetes system. The decision tree algorithm is applied. The experimental results were good enough as the designed system facilitates well in predicting the diabetes incidents at group of age, with better accuracy level using Decision tree [9]. Dataset on significant risk factors for Type 1 Diabetes are reviewed [10]. In this study explain dataset and detailed data analysis results of Type-1 Diabetes have been given. Comparison of Classifiers for the Risk of Diabetes Prediction. The diabetes mellitus prediction model based on data mining is designed to get prediction [11].
Four well known classification models that are Artificial Neural Networks, Decision Tree, Naive Bayes and Logistic Regression were tested. This model uses data mining algorithms and ensures that the dataset quality is sufficient. It gets a better accuracy level of prediction.

### III. RESEARCH TOOLS AND METHODOLOGY

![Proposed work Architecture](image)

**Fig.1. Proposed work Architecture**

The proposed work architecture is shown in figure 1. In this research many approaches such as Linear regression, Naive Bayes, Decision Stump and K-Means were experimented to find the best result.

**A. WEKA Tool**

In this work WEKA tool software is used to experiment. This tool is designed at University of Waikato in the country of New Zealand. The designed tool contains a collection of enormous machine learning techniques for regression, data classification, clustering, visualization, Association etc. The major advantage of using WEKA is that it supports customization as per the requirements of the problem. The primary goal of this study is the prediction of the patient suffered by type1 diabetes using the WEKA tool in the medical database Indian pediatrics.

<table>
<thead>
<tr>
<th>Table-I: Diabetes data set description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

In this research 102 instances with 7 attributes are experimented and verified. Four data mining algorithms were applied to verify the best which results good result.

**B. Linear Regression**

Linear regression algorithm tries to model the relationship between two variables by fitting a linear equation to observed data [11]. A linear regression line has an equation form

\[ L = m + nX \]

Where X is the explanatory variable, L is the dependent variable.

**Fig.2. Result of Linear Regression Algorithm**

**C. Bayesian Classification**

One of the widely used statistical classifiers is Bayesian classifier. The class membership probabilities are easily predicted by this classifier. The probability of a given tuple fit to a particular class. The Naive Bayes classification algorithm is a probabilistic classifier algorithm. It is based on probability models that associate strong independence assumptions.

**Fig.3. Result of Naive Bayes Algorithm**

**D. Decision Stump**

A machine learning method, which consist of a one-level decision tree is decision stump. “That is, it is a decision tree with one internal node which is root node is immediately connected to the terminal nodes such as leaf nodes.” A decision stump predictions are decided based on the value of mono input.
E. K-Means algorithm

A cluster is a collection of similar data objects. It groups the similar data for analysis [12]. K-means clustering algorithm is an unsupervised learning algorithm. The k-means clustering algorithm tries to split a given anonymous data into k clusters. Initially k chooses centroids. A centroid is a center data point (imaginary or real).

Table II: Cluster Formation and Number of Clusters

<table>
<thead>
<tr>
<th>Number</th>
<th>Label</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cluster 0</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>Cluster 1</td>
<td>64</td>
</tr>
</tbody>
</table>

IV. PERFORMANCE OF VARIOUS ALGORITHMS

Performance of four classification algorithms are depicted clearly in Figure 6. It is clear the experimental work of algorithms as Linear regression, Naive Bayes, Decision Stump and K-Means results 82, 96, 58 and 77 percentage of accuracy respectively.

V. CONCLUSION AND FUTURE WORK

One of the most widespread real-world medical hurdles is the detection and prediction of diabetes at its primary stage. In this research work, proficient efforts are used in designing a system which results in the prediction of emerging ruin out disease like diabetes. In this system design the data mining algorithms are used for diabetes prediction with maximum accuracy level of 96% is obtained. The Maximum accuracy level is obtained from Naive Bayes algorithm compared with other algorithms.

For future work, it is suggested and expected to use the hospital's real and recent patients' data for continuous training and development of our proposed model. The size of the dataset is preferred to be large volume in training phase and predicting. In future work will use Big Data Technology for analyzing large amount of data. Use effective new algorithms for accurate prediction.

REFERENCES


AUTHORS PROFILE

Dr. K. Saminathan, Assistant Professor, Department of Computer Science, A.V.V.M Sri Pushpam College, Poondi. He has thirteen years of experience and guided more than 20 projects. He has completed his Post Graduate degree in Computer Science at Periyar E.V.R College, Trichy in 2006. Master of Philosophy in Computer Science at A.V.V.M. Sri Pushpam College, Poondi in 2007. Ph.D in Computer Science at Bharathidasan University and degree awarded in the year of 2017. He published more than 10 research articles in various domains such as Image Processing, Computer Networks and Machine Learning in peer reviewed International Journals.