Diabetes Miletus prediction

By:
Abhishek Patnaik

INDEX:

1. ABSTRACT:

In today's world diabetes is the major health challenges in India. It is a group of a syndrome that results in too much sugar in the blood. It is a protracted condition that affects the way the body mechanizes the blood sugar. Prevention and prediction of diabetes mellitus is increasingly gaining interest in medical sciences. The aim is how to predict at an early stage of diabetes using different machine learning techniques. we use well-known classification that are Decision tree, K-Nearest Neighbors and Random forest. These classification techniques used with Pima Indians diabetes dataset. Therefore, we predict diabetes at different stage and analyze the performance of different classification techniques. We also proposed a conceptual model for the prediction of diabetes mellitus using different machine learning techniques. In this paper we also compare the accuracy of the different machine learning techniques to finding the diabetes mellitus at early stage.

1. INTRODUCTION

2.1 INTRODUCTION FOR PYTHON:

Python is developed by **Guido van Rossum**. Guido van Rossum started implementing Python in 1989. Python is a very simple programming language so even if you are new to programming, you can learn python without facing any issues.

Interesting fact: Python is named after the comedy television show Monty Python's Flying Circus. It is not named after the Python snake.

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-oriented way or a functional way.

2.2 INTRODUCTION FOR MACHINE LEARNING:

Machine Learning (ML) is a subset of AI . Machine Learning is a latest buzzword floating around. It deserves to, as it is one of the most interesting subfield of Computer Science

Machine Learning is the most in-demand technology in today's market. Its applications range from self-driving cars to predicting deadly diseases such as ALS. The high demand for Machine Learning skills is the motivation behind this blog.

Here's a list of reasons why Machine Learning is so important:

- **Increase in Data Generation:** Due to excessive production of data, we need a method that can be used to structure, analyze and draw useful insights from data. This is where Machine Learning comes in. It uses data to solve problems and find solutions to the most complex tasks faced by organizations.
- **Improve Decision Making:** By making use of various algorithms, Machine Learning can be used to make better business decisions. For example, Machine Learning is used to forecast sales, predict downfalls in the stock market, identify risks and anomalies, etc.

- **Uncover patterns & trends in data**: Finding hidden patterns and extracting key insights from data is the most essential part of Machine Learning. By building predictive models and using statistical techniques, Machine Learning allows you to dig beneath the surface and explore the data at a minute scale. Understanding data and extracting patterns manually will take days, whereas Machine Learning algorithms can perform such computations in less than a second.
- **Solve complex problems:** From detecting the genes linked to the deadly ALS disease to building self-driving cars, Machine Learning can be used to solve the most complex problems.

2. LITERATURE SURVEY

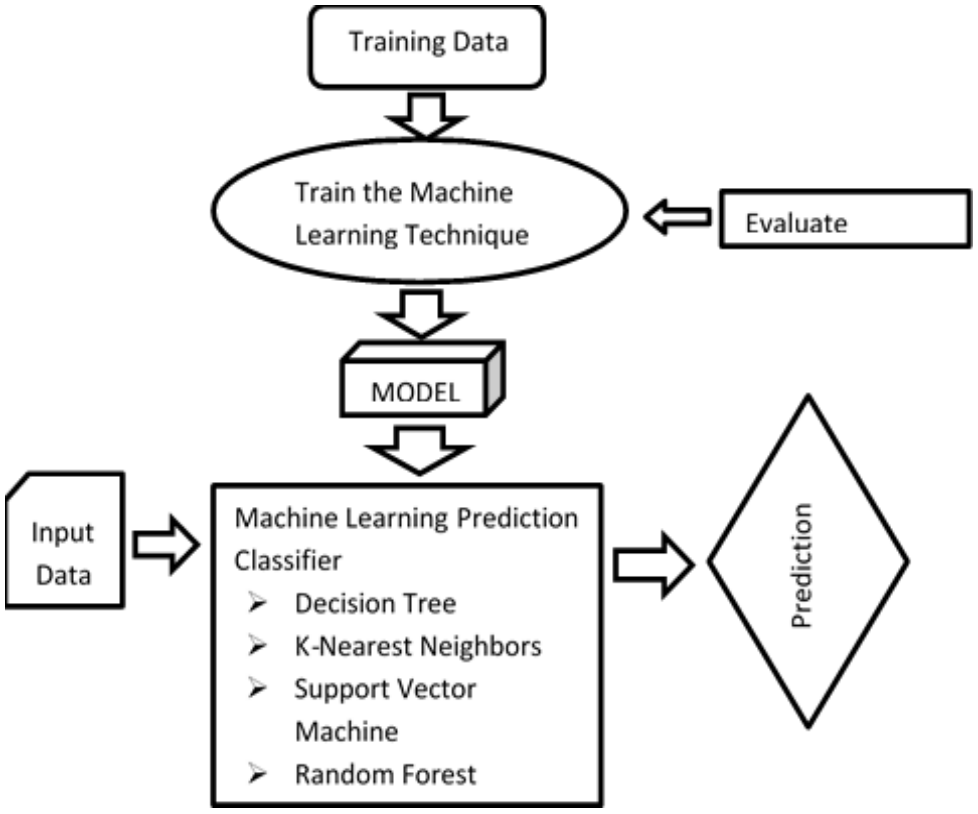
- In 2017, National Diabetes Statistic Report for Center Disease Control and Prevention (CDC), gives the facts give an account of the United States that 30.3 million individuals have diabetes, among that 23.1 are analyzed and 7.2 million are undiscovered individuals.

- In 2018, the American Diabetes Association models of therapeutic care in diabetes discharges a report about “Order and finding of diabetes” which incorporates the arrangement of diabetes, diabetes

care, treatment objectives, criteria for conclusion test ranges and dangers esteems, chance engaged with diabetes.

- In 2017, Global provides details regarding Diabetes by world wellbeing association , it expresses the weight of diabetes, hazard components and inconveniences of diabetes. Likewise, gives the data about counteracting diabetes in individuals with high hazard and overseeing diabetes at beginning times with fundamental solutions to be taken.

3. METHODOLOGY



4.1 DATA SET: The main Objective of using this dataset was to predict through diagnosis whether a patient has diabetes, based on certain diagnostic measurements included in the dataset. The Pima Indian Diabetes (PID) dataset having: 769 records describing female patients (of which there were 500 negative instances (65.1%) and 268 positive instances (34.9%)).

4.2. Data preprocessing: In real-world data there can be missing values and/or noisy and inconsistent data. If data quality is low then no quality results may be found. It is necessary to preprocess the data to achieve quality results. Cleaning, integration, transformation, reduction, and discretization of data are applied to preprocess the data.

4.3 Decision tree: Decision Tree is a classifier using the classification regression trees (CART) algorithm that is capable of handling both classification and regression unlike simple decision tree algorithm. It does not have a computational set of rules.

Input: a. Set of input data are training samples.

b. Set of attributes from input samples.

c. Splitting the attributes by best partitioning criteria.

Output: A decision tree

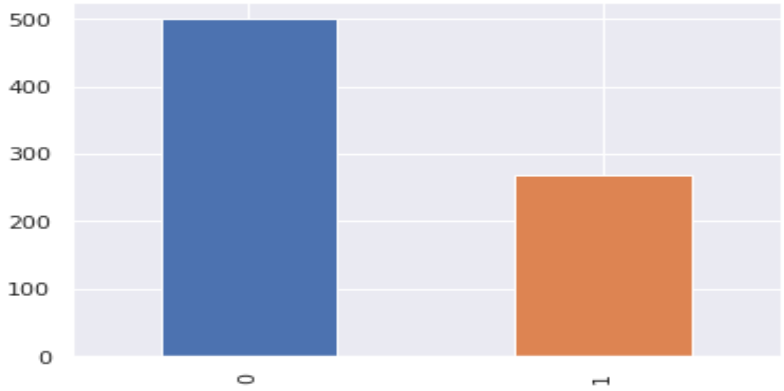
4.4 RANDOM FOREST: Random Forest is the one of the Classifier which is used for Classifications problems. Random Forest is ensemble classifier made using many decision trees where ensemble means that uses multiple machine learning algorithm to obtain the predictive performance It is better than other for the prediction of diabetes mellitus**.**

4.5 FUTURE SCOPE: This work will be considered as basement for the health care system for Diabetes patients using machine learning Algorithms.

4. DATA ANAYLSIS:

	A	B	C	D	E	F	G	H	I	J	K
1	preg	plas	pres	Skin	test	mass	pedi	age	class		
2	6	148	72	35	0	33.6	0.627	50	1		
3	1	85	66	29	0	26.6	0.351	31	0		
4	8	183	64	0	0	23.3	0.672	32	1		
5	1	89	66	23	94	28.1	0.167	21	0		
6	0	137	40	35	168	43.1	2.288	33	1		
7	5	116	74	0	0	25.6	0.201	30	0		
8	3	78	50	32	88	31	0.248	26	1		
9	10	115	0	0	0	35.3	0.134	29	0		

- ✓ This initial data is collected from Pima Indians diabetes dataset. It will be used to comparative analysis of different machine learning techniques



- ✓ The above graph shows that the data is biased towards datapoints having outcome value as 0 where it means that diabetes was not present actually. The number of non-diabetics is almost twice the number of diabetic patients

5. CONCLUSION

In this we have studied different ML algorithms. The main motto is to prevent and cure diabetes and to improve the lives of all people affected by diabetes. The advantage of this system is that, the prediction process is less time consuming. It will help the doctors to start the treatments early for the Diabetes patients.

6. Reference Link

Dataset downloaded from:

https://www.kaggle.com/akhilalexander/diabeticprediction