MACHINE LEARNING IN MEDIACL ADVANCEMENT

Ramdhan Prajapat

Date: 29th March, 2024

1. Abstract:

Machine learning is like a super-smart computer program that learns from Medical Data. Machine learning is changing the field of Medical Sector. Machine Learning algorithms are empowering medical professionals, accelerating research and ultimately improving patient outcomes. By analyzing medical records, Machine learning helps doctors diagnose diseases earlier and predict health risks. One exiting area is diabetes prediction. Machine Learning can analyze your blood sugar, weight and family history to see how likely you are to develop diabetes. This allows for early intervention, like lifestyle changes to prevent the disease altogether.

2. Problem Statement:

Diabetes is a chronic condition affecting millions globally, often going undetected until complications arise. Early detection is crucial for effective management and preventing serious health issues. Traditional methods rely on blood tests which may not always capture the early stages of the disease. This is where machine learning comes in. We aim to develop a machine learning model that can predict the risk of developing diabetes using available patient data. This data could include factors like Blood sugar level, Body Mass Index (BMI), Age, Family history of diabetes, Blood Pressure, Cholesterol levels etc. By analyzing these factors, the model can identify individuals at high risk of developing diabetes.

3. Market / Customer/ Business Need Assessment:

Diabetes is growing health concern worldwide, leading to serious complications and increased healthcare costs. Early detection is key to managing the diseases effectively and preventing these complications.

This solution targets several key markets:

- Early detection allows them to intervene sooner, improving patient outcomes and reducing long-term healthcare costs.
- Individuals at risk of diabetes benefit from preventive measures and lifestyle changes, potentially avoiding the disease altogether.
- Early intervention can lower overall healthcare costs associated with diabetes complications.

For healthcare institutions and organizations, a machine learning-based diabetes prediction system offers several advantages. Earlier detection leads to better management and reduced complications. Early intervention can prevent expensive treatments needed for advanced stages of diabetes. Streamlining patient screening and risk assessment allows doctors to focus on high-risk individuals. Offering advanced diabetes prediction tools can attract patients and position the institution as a leader in preventative care.

Overall, I want to tell that a machine learning approach to diabetes prediction addresses a critical need in the healthcare market, benefiting patients and businesses alike.

4. Target Specifications and Characterization:

Target Specifications:

- The machine learning model should accurately predict the risk of developing diabetes.
- The system should be user-friendly and integrate seamlessly with existing healthcare workflows.
- The model's prediction should be clear and understandable to healthcare professionals.
- The system should allow for customization based on specific patient demographics and risk factors.
- The solution should meet all relevant data privacy regulations and ensure patient information is protected.

Customer Characteristics:

- Customers value tools that improve early detection and lead to better patient management.
- Customers are interested in solutions that reduce long-term healthcare costs associated with diabetes.
- Customers appreciate the objective and evidence-based approach offered by machine learning.
- These customers prioritize tools that identify individuals at risk before symptoms appear.
- These customers require solutions that meet data privacy regulations and ensure patient information is secure.

5. External Searches (Information searches):

5.1 Applications of Machine Learning in Cancer Prediction and Prognosis

Machine learning is a branch of artificial intelligence that employs a variety of statistical, probabilistic and optimization techniques that allows computers to "learn" from past examples and to detect hard-to-discern patterns from large, noisy or complex data sets. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. As a result, machine learning is frequently used in cancer diagnosis and detection. More recently machine learning has been applied to cancer prognosis and prediction. This latter approach is particularly interesting as it is part of a growing trend towards personalized, predictive medicine. In assembling this review we conducted a broad survey of the different types of machine learning methods being used, the types of data being integrated and the performance of these methods in cancer prediction and prognosis.

The conclusions were made based on the data available on the internet.

The percentages were calculated roughly in a relative manner

5.2 Machine learning-based prediction of survival prognosis

Survival prediction after first diagnosis is crucial for both disease specialists and patients or their family members. Firstly, as the survival prospects of diabetes patients significantly rely on the progression of the disease, accurately predicting the prognosis would be beneficial for estimating the severity of the condition and the timing of disease progression. On the other hand, patients and their families can establish suitable goals based on precise survival predictions. Consequently, timely preventive measures and treatments can be initiated, and the risk of inappropriate treatment decisions, such as over-treatment or delayed palliative care, can be effectively mitigated.

5.3 Diabetes dataset:

Features are computed from clinical measurements associated with diabetes diagnosis. They describe characteristics of patients' physiological parameters.

Attribute Information:

1) ID number
2) Outcome (0 = non-diabetic, 1 = diabetic)

Nine real-valued features are computed for each patient:

a) Pregnancies (number of pregnancies)
b) Glucose (plasma glucose concentration)
c) Blood Pressure (diastolic blood pressure)
d) Skin Thickness (triceps skin fold thickness)
e) Insulin (2-hour serum insulin)
f) BMI (body mass index)
g) DiabetesPedigreeFunction (diabetes pedigree function)
h) Age (age in years)

5.4 Machine learning is the future of Diabetes prediction:

Machine learning is set to change the medical industry in the coming times - it wouldn’t make sense for diabetology to not be disrupted too. Currently, ML models are still in the testing and experimentation phase for diabetes treatment. As datasets are getting larger and of higher quality, researchers are building increasingly accurate models. Machine Learning is the next step forward for us to overcome this hurdle and create a high accuracy diabetology system.

6 Benchmarking alternate products:

In a comprehensive study focused on diabetes patients, individuals who opted for alternative treatments instead of conventional medical interventions faced significantly poorer outcomes in terms of survival. Following a median duration of 5 years, those with diabetes were nearly five times more likely to experience mortality if they pursued alternative therapies as their initial approach compared to those who underwent conventional treatment. This underscores the heightened risk of mortality associated with choosing alternative medicine, underscoring the importance for patients to carefully consider their treatment decisions.

Moreover, the demographic profile of patients selecting alternative treatments tended to skew towards younger age, female gender, better overall health, and higher socioeconomic status in terms of income and education. While some of these characteristics, such as better health status, might typically improve survival prospects following a diabetes diagnosis, the choice of alternative therapies negated these potential advantages.

7 Applicable Regulations (Government and Environmental)

a. The privacy and security of protected health information by Machine Learning Model.
b. ML model need to undergo FDA (Food and Drug Administration) clearance or approval processes.
c. Creating an e-mail service to mail the report to the patient and doctor.
d. Consider to publication and reporting guidelines.
e. Consideration to Good Clinical Practice guidelines is essential.
f. Consider environmental regulations related to the use of resources, waste management, and pollution control.

8 Business Opportunity:

Diabetologists are pretty good in diagnosing diabetes while they are not so good in the prognosis of diabetes.It takes more than two weeks to identify diabetes in an individual. To overcome this hazardous circumstance, our main objective is to use Machine Learning, which not only gives faster results but also demonstrates higher accuracy in the diabetes prediction process. Developing a business model for Diabetes prediction involves several things like data acquisition, model development, deployment strategy etc.

If anyone want to check this model enter the below link:

CHECK APP

Diabetes Prediction(Above 21 Years of Age)

Name:

Plasma Glucose Concentration :

0.00 − +

Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

Diastolic blood pressure (mm Hg):

0.00 − +

BloodPressure: Diastolic blood pressure (mm Hg)

Triceps skin fold thickness (mm):

0.00 − +

SkinThickness: Triceps skin fold thickness (mm)

2-Hour serum insulin (mu U/ml):

0.00 − +

Insulin: 2-Hour serum insulin (mu U/ml)

Body mass index (weight in kg/(height in m)^2):

0.00 − +

BMI: Body mass index (weight in kg/(height in m)^2)

Diabetes Pedigree Function:

0.00 − +

DiabetesPedigreeFunction: Diabetes pedigree function

Age:

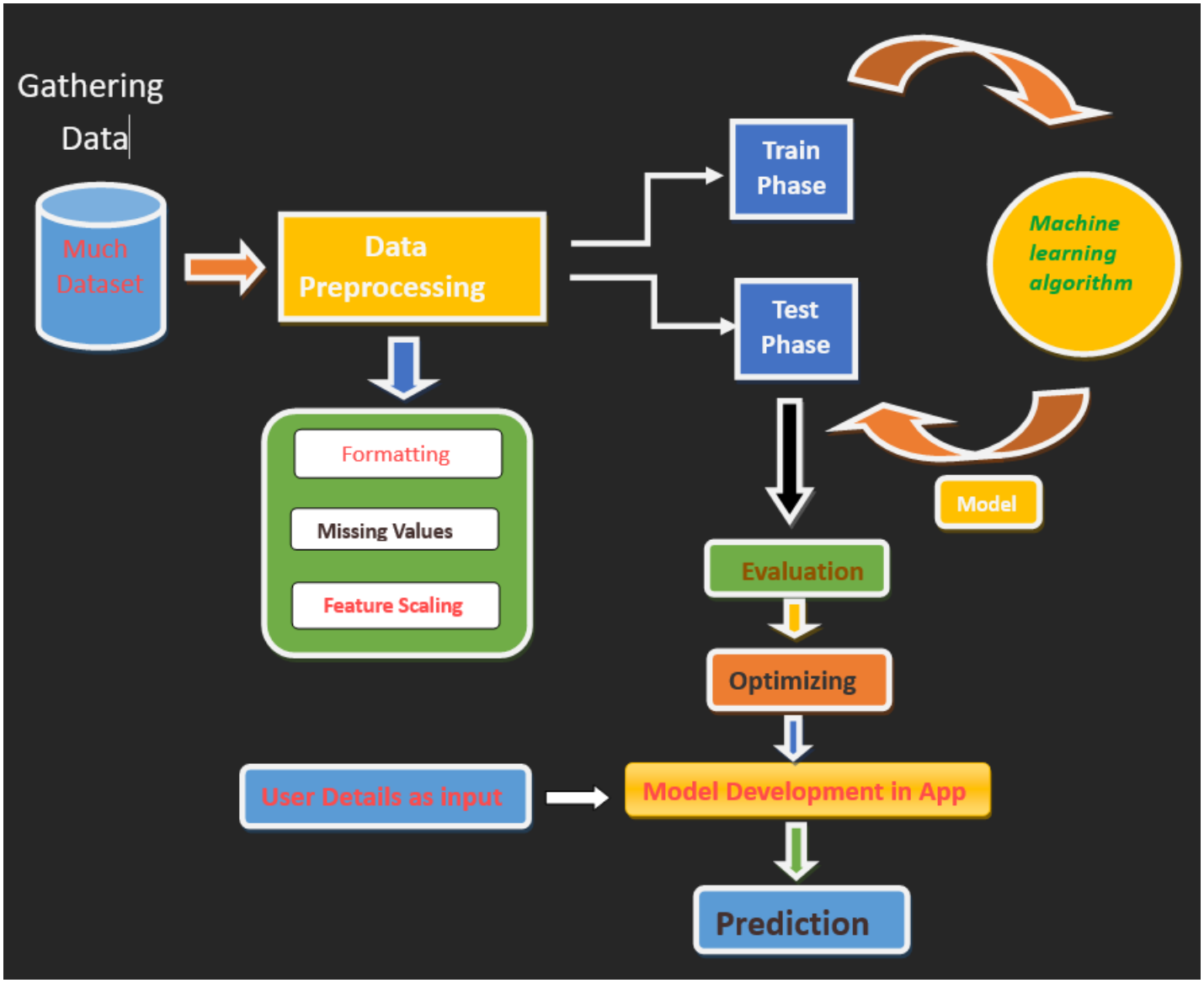
0.00 − +

Age: Age (years)

Predict

Outcome: Class variable (0 or 1)

9 Final Product prototype:



10 Product details:

Machine Learning (ML) is revolutionizing various fields by its ability to analyze the data, identify patterns, and make predictions or decisions. One significant advantage of ML is its remarkable speed compared to human capabilities. For instance, in medical diagnostics, while it may take a diabetologist days to analyze a biopsy sample, ML algorithms can process thousands of similar samples in some seconds. This rapid processing enables quicker decision-making, potentially leading to faster treatments and improved patient outcomes.

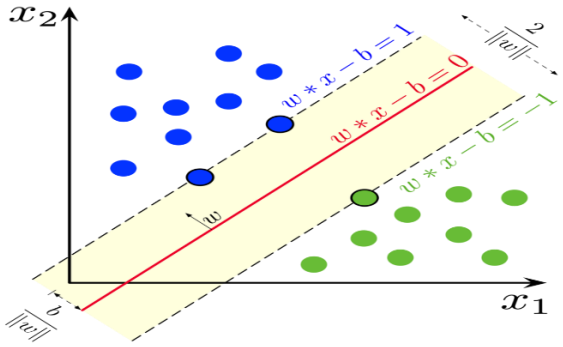
Machines have more accuracy. With the advancement of the Internet of Things technology, there is so much data out in the world that humans can't possibly go through it all. That's where machines help us to predict the diabetes prediction.

10.1 Algorithm:

Classification model contains different algorithms which can be used to classify based on the predefined diabetes dataset.

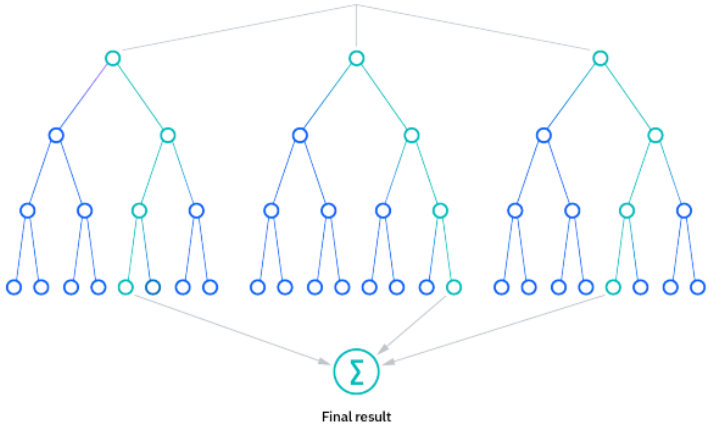
1. Logistic Regression: Logistic regression is a classification algorithm. It is used to predict a binary outcome based on a set of independent variables. A binary outcome is one where there are only two possible scenarios—either the event happens (1) or it does not happen (0). And in diabetes prediction scenario if happens 1 means that patient is suffering from diabetes and if happens 0 means that patient is not suffering from diabetes. I have gotten 75.97% accuracy and AUC is 0.79.

2. Support Vector Machine: Support vector machine is a supervised algorithm used for classification. The main objective of SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient Versatile: different Kernel functions can be specified for the decision function. I have gotten 73.37% accuracy and AUC is 0.80 .



3. RandomForest:

A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees more it will be robust. Similarly, the greater the number of trees in a Random Forest Algorithm, the higher its accuracy and problem-solving ability. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model. I have gotten 75.97% accuracy and AUC is 0.81.



10.2 Python-libraries for Cancer Prognosis:

- Pandas: Pandas is a Python library used for working with data sets. Pandas can clean noisy data sets, and make them readable and relevant. The Pandas library allows you to organize and explore data quickly and effectively. It accomplishes this by offering us Series and Data frames, which enable you to represent data and modify it in various ways effectively.

  Syntax: import pandas as pd
  df = pd.read_csv('diabetesdata.csv')

- Scikit Learn: Scikit-learn (formerly scikits. learn and also known as sklearn) is a free software machine learning library used based on python programming language in spyder IDE. It features various classification, regression and clustering algorithms including (SVM) support vector machines , random forests, gradient boosting, k-means and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

  Syntax for Logistic Regression:
  From sklearn.linear_model import LogisticRegression

  Syntax for SVM:
  From sklearn.svm import SVC

  Syntax for Random Forest:
  From sklearn.ensemble import RandomForestClassifier

- Seaborn: Seaborn is a python data visualization library based on matplotlib. It is built on top of matplotlib and closely integrated with pandas data structures. This is a visualization tool used to demonstrate the count of benign and malignant cells through the predefined dataset.

  Syntax: Import seaborn as sns

Algorithm with highest accuracy among classification algorithms is chosen as the best algorithm for Diabetes prediction.

(CODE SNIIPET)

11 Conclusion:

The utilization of artificial intelligence (AI) and machine learning (ML) in predicting diabetes presents a promising frontier in healthcare. Through the analysis of diabetes datasets and the application of machine learning algorithms, these technologies offer the potential to enhance early detection and management of diabetes, thereby improving patient outcomes and reducing healthcare costs.

Ultimately, the primary goal is to save lives and enable doctors to focus more on evaluation by using the efficiency of ML.