Diabetes Risk Prediction Analysis

TEAM MEMBERS

MANICHANDANA

HARSHITHA

GEETHA

PAVAN

Contents

INTRODUCTION

PROPOSED METHODOLOGY

IMPLEMENTATION

CONCLUSION

Introduction

- Diabetes is a most common disease in these world affecting major part of the population. Although it is a global concern but still most part of population are unaware to deal with it.
- The inability of the pancreas to produce the required insulin or the inability to use it properly to convert glucose into energy is the cause of Diabetes.
- Diabetes majorly of 3 types :
    1. Diabetes Mellitus
    2. Gestational Diabetes
    3. Diabetes Insipidus

Data Science

Data Science has an intersection with artificial intelligence but is not a subset of artificial intelligence. This means it is the technique that uses some techniques of artificial intelligence, machine learning, and deep learning.

By using this technique data are first visualized, then builds model, and at last do prediction.

Machine Learning

Machine Learning is a subset of Artificial Intelligence. It is a set of algorithms that train on a data set to make predictions or take actions to optimize some systems.

It is a technique in which computer program learn from experiences or past experiences. Then specifically related to that it performs the task.

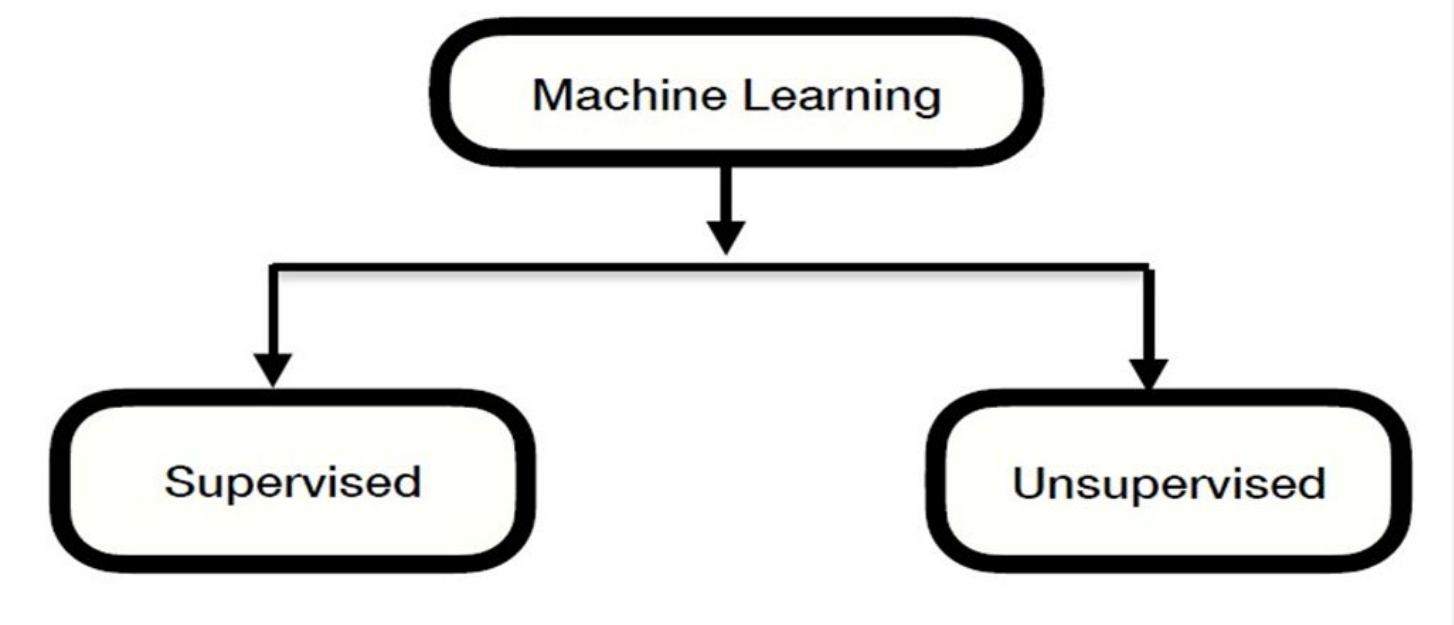
Machine Learning algorithm create model using sample data or training data.

The steps involved in machine learning are:-

- Problem
- Data Collection and validation
- Model Building
- Feedback

Machine Learning Models

- In Machine Learning, techniques, and data are more important, but identifying the type of problem (supervised or unsupervised) is equally important,



Supervised Learning

- Supervised learning is a type of machine learning where the algorithm is trained on a labeled dataset, meaning that the input data used for training is paired with corresponding output labels. The goal of supervised learning is to learn a mapping function from the input variables to the output variables, such that the algorithm can make predictions or decisions on new, unseen data.

Unsupervised Learning

- The algorithm is not provided with explicit instructions on what the correct output should be for each input during the training process. Instead, it must find patterns, relationships, or structures within the data on its own.

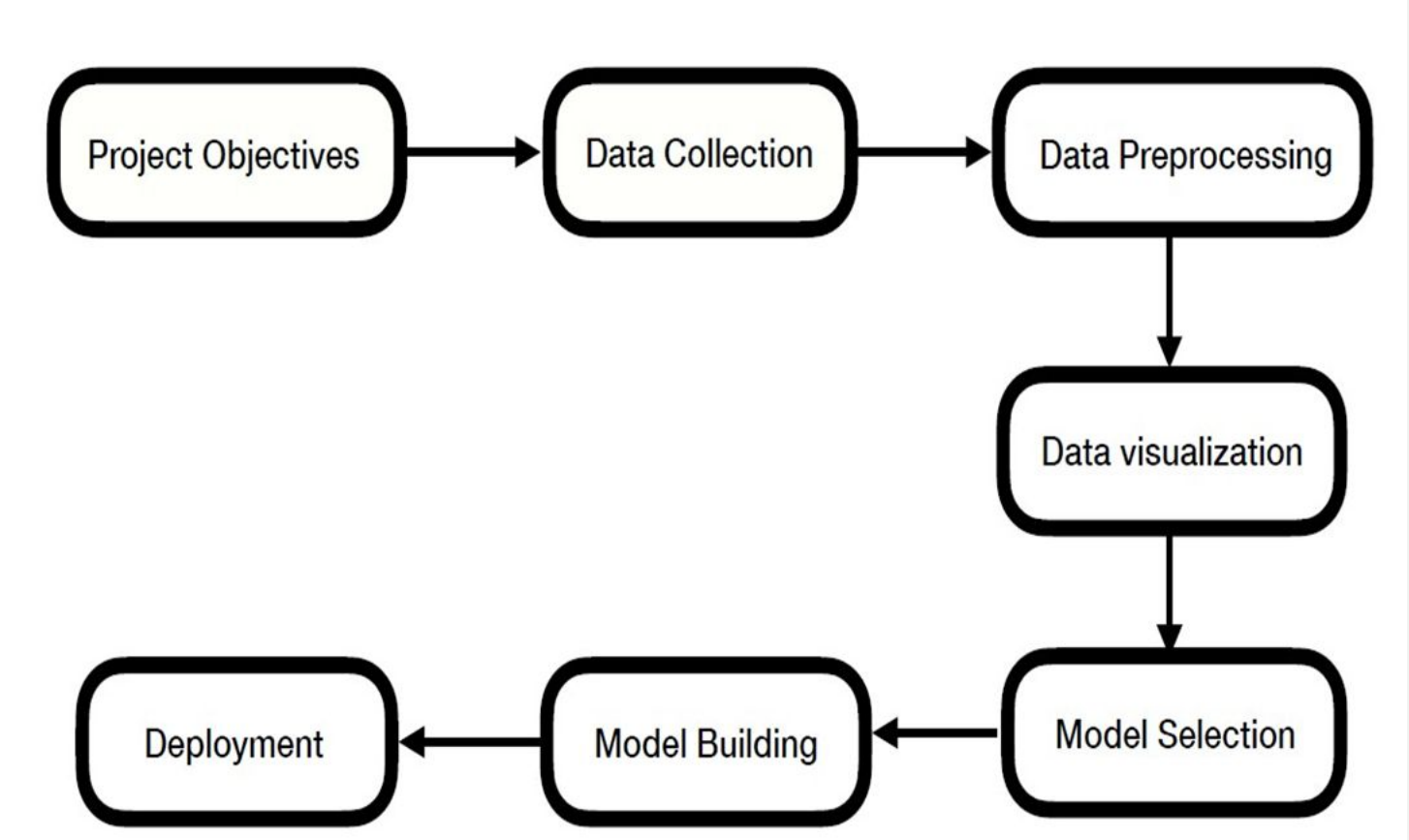
Proposed Methodology

- It is a Data science project in which we need to predict whether a given data point is diabetic or not. Since it comes under Classification technique which is a type of Supervised Machine Learning, here the model is built using different classification algorithms like K-Nearest Neighbors and Support Vector Machine.
- The model built with a Support Vector Machine gave high accuracy out of all in predicting whether the given data point is diabetic or not.

Libraries Used

- **Numpy:**
  - Here, Num means numeric and Py means python.
  - It is a scientific computing library for python.
  - It supports multi-dimensional arrays. It is used to represent large numbers of data in the form of an array
  - NumPy Library is used for numeric calculation.
  - It is used to store data as it uses less memory. It is very convenient and process fast.
- **Pandas:**
  - Pandas is a powerful python data analysis toolkit.
  - Pandas is used for data manipulation, analysis and cleaning.
- **Matplotlib:**
  - To represent data graphically to analyze the data easily.

Process Flow



Step 1: Project Objectives :

This is the first step in which business problems and requirements are specified.

Step 2: Data Collection :

In this process, the data required for the project are collected or gathered.

Step 3: Data Preprocessing :

In this process, raw data or original data are purified. For data preprocessing the python libraries like NumPy, Panda, and etc are used.

Step 4: Data Visualization :

Data visualization is a graphical representation of data. Here, data are transformed into pie charts, graphs, bar graphs, histograms, etc. For data visualization, python libraries like matplotlib, and Seaborn are used.

Step 5: Model Selection :

In this step, we have to select best machine learning model according to the data which will give us accurate results.

Step 6: Model building :

After, prediction is done from that model. If the prediction is not accurate then again model gets train and testing is done again.

Step 7: Model Deployment :

This is the last stage of the machine learning life cycle. Machine learning models are deployed into the production environment for taking decisions.

Implementation

- **Statement:**
- The goal of the project is to build a model which can give high accuracy in predicting the disease.
- **Requirements:**
- **SOFTWARE:** Jupiter notebook
- **LIBRARIES**: Pandas, Numpy, Matplotlib, Seaborn.

Implementation Steps

- Step 1:- Import all the required libraries.
- Step 2:- Read the dataset and store it into a new variable.
- Step 3:- check the shape of the dataset.
- Step 4:- check for missing values in dataset.
- Step 5:- Perform the data visualization process.
- Step 6:- Select the model(K Neighbors Classifier, Support Vector Classifier).
- Step 7:- Check for the highest accuracy among all the algorithms used.
- Step8:- After finding highest accuracy, check whether the model selected is correct or not.

Conclusion

- In conclusion, our diabetes risk prediction analysis project has provided valuable insights into the factors influencing the risk of diabetes among the study population. Through the utilization of a supervised learning approach. By using KNN and SVC algorithms. We successfully developed a predictive model capable of assessing an individual's likelihood of developing diabetes based on a set of input features.

Did you hit your project time goals?

- Meeting project time goals is crucial to the success of any team project. In our team we put a strong emphasis on project planning at the beginning, we also recognize that challenges may arise. But, regular team meetings, status updates, and a collaborative work environment help us stay informed about progress. By all these we finally made to hit our project time goals.

What has changed since the midterm?

- Since the midterm, our team has achieved significant milestones. We've successfully completed implementation and have made substantial progress towards our project goals. We revisited our initial objectives to make sure we reached our project goals. We encountered with some challenges. Finally, we have completed our project with the team collaboration.


Any Questions?

Thank You