



DIABETES PREDICTION

Name: Sanket Tejrao Deshmukh

Phone: 9130350331

Email: sanket7deshmukh2501@gmail.com



➤ **Domain :** Data Science.

➤ **Project Discription:**

Diabetes is a type of chronic disease which is more common among the of all age groups. Predicting this disease at an early stage can help necessary precautions and change his/her prevent the occurrence of this disease or control the disease(For people who already have the disease).

➤ **Table of Contents:**

1)	Necessary Libraries.
2)	Data Description.
3)	Data Exploration & Prepration
4)	Exploratory Data Analysis (EDA)
5)	Data Preprocessing.
6)	Data Modeling.
7)	Data Evalution.

➤ **Introduction:**

Diabetes Mellitus (DM), commonly known as diabetes, is a group of metabolic disorders characterized by high blood sugar levels over a prolonged period. Diabetes is due to either the pancreas not producing enough insulin, or the cells of the body not responding properly to the insulin produced. There are three main types of diabetes mellitus:

Type 1 diabetes results from the pancreas's failure to produce enough insulin due to loss of beta cells. This form was previously referred to as "insulin-dependent diabetes mellitus" (IDDM) or "juvenile diabetes". The cause is unknown.

Type 2 diabetes begins with insulin resistance, a condition in which cells fail to respond to insulin properly. As the disease progresses, a lack of insulin may also develop. This form was previously referred to as "non insulin-dependent diabetes mellitus" (NIDDM) or "adult-onset diabetes". The most common cause is a combination of excessive body weight and insufficient exercise.

Gestational diabetes is the third main form, and occurs when pregnant women without a previous history of diabetes develop high blood sugar levels.

Prevention and treatment involve maintaining a healthy diet, regular physical exercise, a normal body weight, and avoiding use of tobacco. Control of blood pressure and maintaining proper foot care are important for people with the disease. Type 1 diabetes must be managed with insulin injections. Type 2 diabetes may be treated with medications with or without insulin. Insulin and some oral medications can cause low blood sugar. Weight loss surgery in those with obesity is sometimes an effective measure in those with type 2 diabetes. Gestational diabetes usually resolves after the birth of the baby.

➤ **Existing Method:**

We'll be using Machine Learning to predict whether a person has diabetes or not, based on information about the patient such as blood pressure, body mass index (BMI), age, etc. The tutorial walks through the various stages of the data science workflow. In particular, the tutorial has the following sections.

- ✧ Data Description.
- ✧ Data Exploration.
- ✧ Data Preparation.
- ✧ Training and Evaluating the Machine Learning Model.
- ✧ Interpreting the ML Model.
- ✧ Saving the Model.
- ✧ Making Predictions with the Model.

➤ **Proposed method with Architecture:**

The data was collected and made available by “National Institute of Diabetes and Digestive and Kidney Diseases” as part of the Pima Indians Diabetes Database. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here belong to the Pima Indian heritage (subgroup of Native Americans), and are females of ages 21 and above.

We'll be using Python and some of its popular data science related packages. First of all, we will import pandas to read our data from a CSV file and manipulate it for further use. We will also use numpy to convert out data into a format suitable to feed our classification model. We'll use seaborn and matplotlib for visualizations. We will then import K-nearest neighbors (KNN algorithm from Scikit-Learn . This algorithm will help us build our classification model. Lastly, we will use joblib available in Scikit-Learn to save our model for future use.

➤ **Methodology:**

We'll be using Python and some of its popular data science related packages. First of all, we will import `pandas` to read our data from a CSV file and manipulate it for further use. We will also use `numpy` to convert out data into a format suitable to feed our classification model. We'll use `seaborn` and `matplotlib` for visualizations. We will then import K-nearest neighbors (KNN algorithm from `Scikit-Learn` . This algorithm will help us build our classification model. Lastly, we will use `jobli` available in `Scikit-Learn` to save our model for future use.

➤ **Implementation:**

Data Description

We have our data saved in a CSV file called `diabetes.csv`. We first read our dataset into a `pandas` dataframe called `data`, and then use the `head()` function to show the first five records from our dataset.

Data Exploration

Let us now explore our data set to get a feel of what it looks like and get some insights about it.

Let's start by finding correlation of every pair of features (and the outcome variable), and visualize the correlations using a heatmap.

Dataset Preparation (splitting and normalization)

When using machine learning algorithms we should always split our data into a training set and test set. (If the number of experiments we are running is large, then we can should be dividing our data into 3 parts, namely — training set, development set and test set). In our case, we will also separate out some data for manual cross checking.

The data set consists of record of 767 patients in total. To train our model we will be using 650 records. We will be using 100 records for testing, and the last 17 records to cross check our model.

Training and Evaluating Machine Learning Model

We can now train our classification model. We'll be using a machine simple learning model called *logistic regression*. Since the model is readily available in sklearn, the training process is quite easy and we can do it in few lines of code. First, we create an instance called diabetesCheck and then use the fit function to train the model.

➤ **Conclusion**

The results can be improved by applying the feature scaling and data cleaning. From this project we predict the type 2 diabetes, commonly called as diabetes mellitus. As a result it can help to improve their health conditions.

...