Prediction of Diabetes

**Algorithms Used:**

* Random Forest
* Decision Tree
* XgBoost Classifier
* Support Vector Machine

**Colab Notebook:** <https://colab.research.google.com/drive/1yoEyIcM8ECnnDq0GsaSsdhqGClWMSM_0?usp=sharing>

**Source:**

<https://www.analyticsvidhya.com/blog/2022/01/diabetes-prediction-using-machine-learning/>

**Dataset:**

<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

**Rows, columns:** (768, 9)

**About:**

Diabetes prediction using machine learning means using computer programs to guess if someone might get diabetes. These programs look at things like health history and lifestyle to make their guess. They learn from many examples of people with and without diabetes to make better guesses. For instance, they might look at how much sugar someone eats or if they exercise regularly. By doing this, they can give early warnings to people at risk of getting diabetes so they can take better care of themselves.

The Pima Indians Diabetes Dataset is a publicly available test dataset widely used for diabetes research and predictive modeling. It contains 768 observations of females of Pima Indian heritage aged 21 years or older. The dataset includes eight medical predictor variables and one target variable. The predictor variables are:

* Pregnancies: Number of times pregnant
* Glucose: Plasma glucose concentration over 2 hours in an oral glucose tolerance test
* Blood Pressure: Diastolic blood pressure (mm Hg)
* Skin Thickness: Triceps skinfold thickness (mm)
* Insulin: 2-hour serum insulin (mu U/ml)
* BMI: Body mass index (weight in kg/(height in m)^2)
* Diabetes Pedigree Function: A function that scores the likelihood of diabetes based on family history
* Age: Age in years
* Outcome

The target variable is ‘Outcome’’ which indicates whether the patient had diabetes (1) or not (0). This training dataset is particularly useful for testing machine learning algorithms for binary classification tasks.

**Algorithms Model and Accuracy**

|  |  |
| --- | --- |
| **Models** | **Scores** |
| Random\_forest | 0.7677 |
| Decision\_Tree | 0.7322 |
| XgBoost | 0.7401 |
| SVM | 0.7401 |

Best: **Random Forest**