

Diabetes prediction system via ML algorithms

Qavamodin Soleimani

Supervisor: Dr. Fereshteh Dehghani

January 6, 2023



Data mining levels

- 1 Introduction
- 2 Data collection
- 3 Data cleaning
- 4 Visualization
- 5 Data modeling
- 6 Results

Introduction

What is diabetes?

Global statistics

About 422 million people worldwide have diabetes and 1.5 million deaths are directly attributed to diabetes each year.

What to do?

Methods of preventing diabetes.

Predicting diabetes

They found that BMI, waist-hip ratio (WHR), age, systolic and diastolic blood pressure, and a family history of diabetes were the most significant predictive features for T2D and prediabetes (Lama et al., 2021)

Data collection

Supervised vs. unsupervised learning: Which is best for you?

- Evaluate your input data: Is it labeled or unlabeled data? Do you have experts that can support additional labeling?
- Define your goals: Do you have a recurring, well-defined problem to solve? Or will the algorithm need to predict new problems?
- Review your options for algorithms: Are there algorithms with the same dimensionality you need (number of features, attributes or characteristics)? Can they support your data volume and structure?

Data cleaning

Missing data

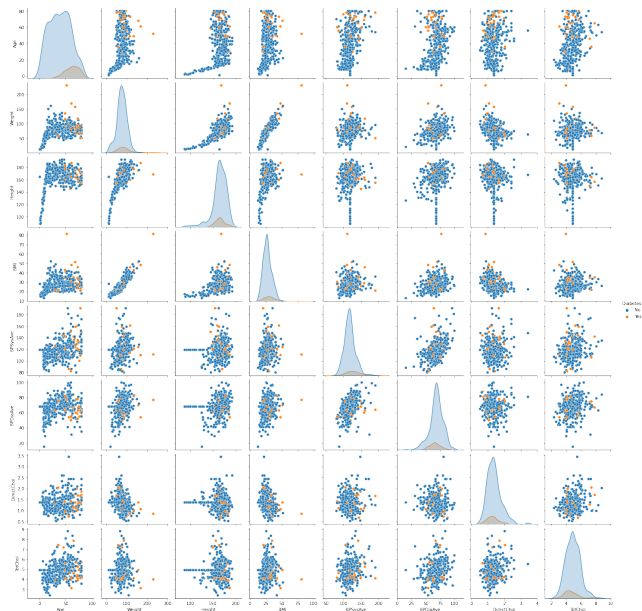
0 Values

Remove rows or replacing

NaN Values

Remove rows or replacing

Visualization



Data modeling

1 LR

2 RF

Results

H
eat map