# Diabetes Prediction Using Machine Learning Algorithms

Alişan Köroğlu
2023900252

# **Agenda**

- What is Diabetes?

- Type 2 Diabetes

- Diabetes Dataset

- Feature Selection

- Performance of Models

# What is Diabetes?

- A condition when body cannot properly regulate blood sugar(glucose) levels.

- This happens because the body either doesn't produce enough insulin or cannot effectively use the insulin it produces.

Types of Diabetes:

- Type 1 Diabetes: The body's immune systemattacks and destroy insülin cells in the pancreeas. Patients need lifelong insulin therapy.

- Type 2 Diabetes: The body becomes resistant to insülin or doesn't produce enough. Often managed through diet, exercise, medication, and sometimes insulin.

- Gestationan Diabetes: Occurs during pregnancy and usualy resolves after childbirth. Increases the risk of developing Type 2 diabetes later in life.

- Prediabetes: Blood sugar levels are higher than normal but not high enough to be diagnosed as diabetes. A warning sign that Type 2 diabetes may develop without lifestyle changes.
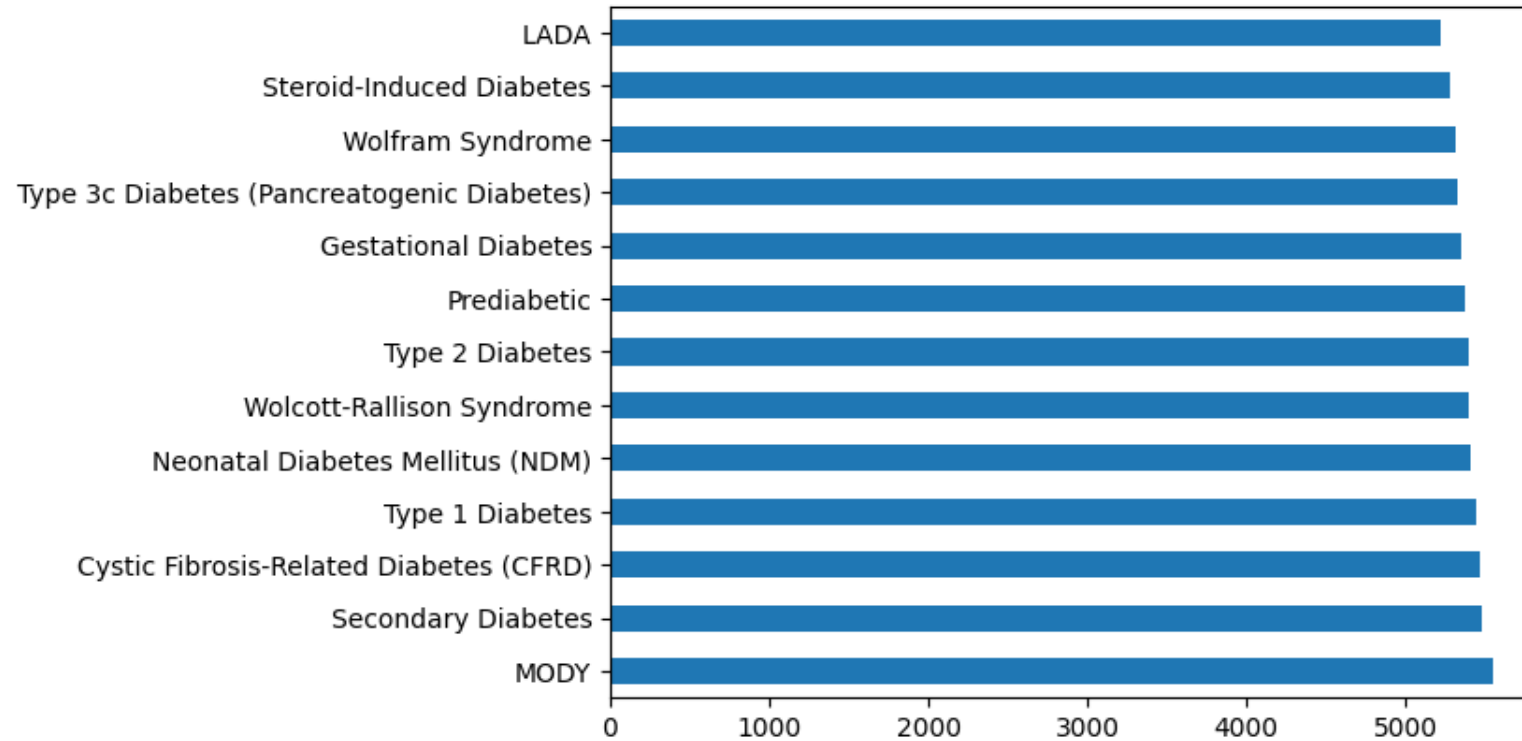
# Type 2 Diabetes

- Type 2 diabetes is a significant global health concern, with its prevalence increasing rapidly over the past few decades.

- **Prevalence**: As of 2024, approximately 537 million adults (20-79 years) worldwide are living with diabetes, with over 90% of these cases being type 2 diabetes. (IDF Diabetes Atlas)

- **Projected Increase**: The number of adults with diabetes is expected to rise to 783 million by 2045, indicating a significant global health challenge. (IDF Diabetes Atlas)

- **Mortality**: Diabetes was responsible for 6.7 million deaths in 2024, equating to 1 death every 5 seconds. (IDF Diabetes Atlas)
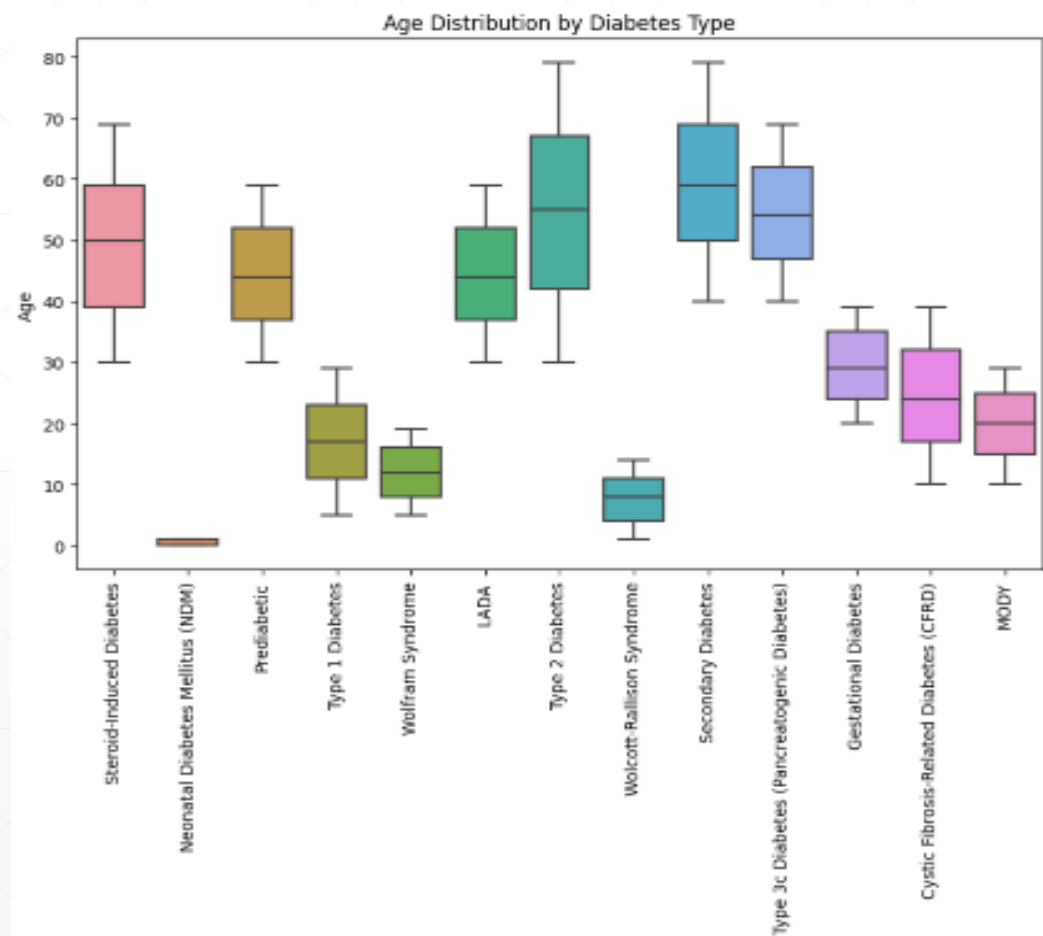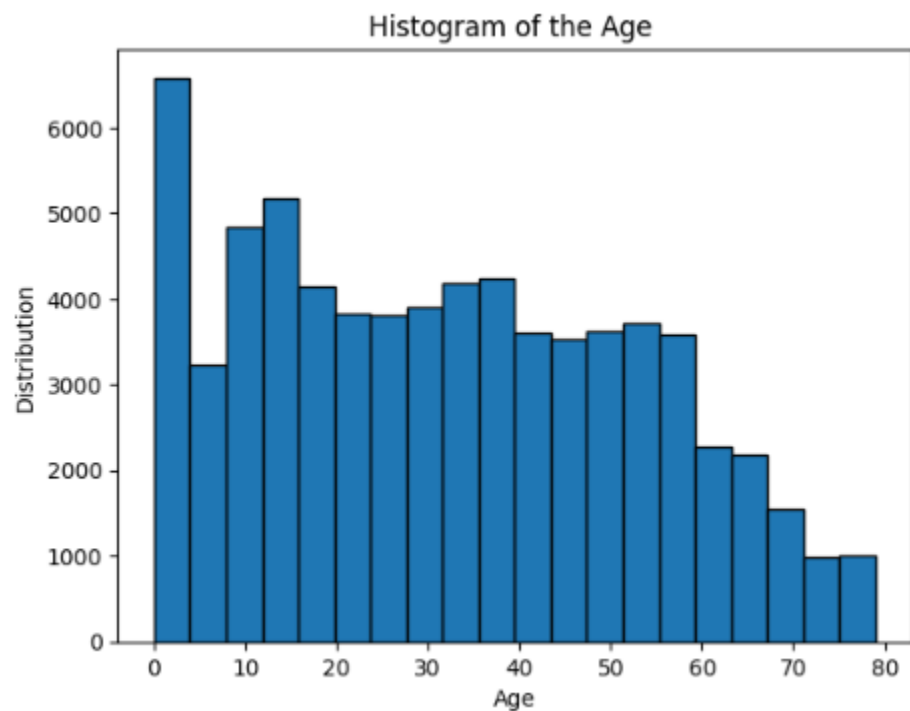
# Diabetes dataset

- The **Diabetes Dataset** available on Kaggle, contributed by Ankit Batra, is a comprehensive collection designed to facilitate the development of predictive models for diabetes diagnosis. The dataset is publicly accessible on Kaggle and can be downloaded for analysis. (https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset)

- **Dataset Characteristics:**

  - **Number of Instances (Samples):** 70k

  - **Number of Features (Columns):** 33 (plus 1 target variable)

    - 20 Categorical

    - 13 numerical

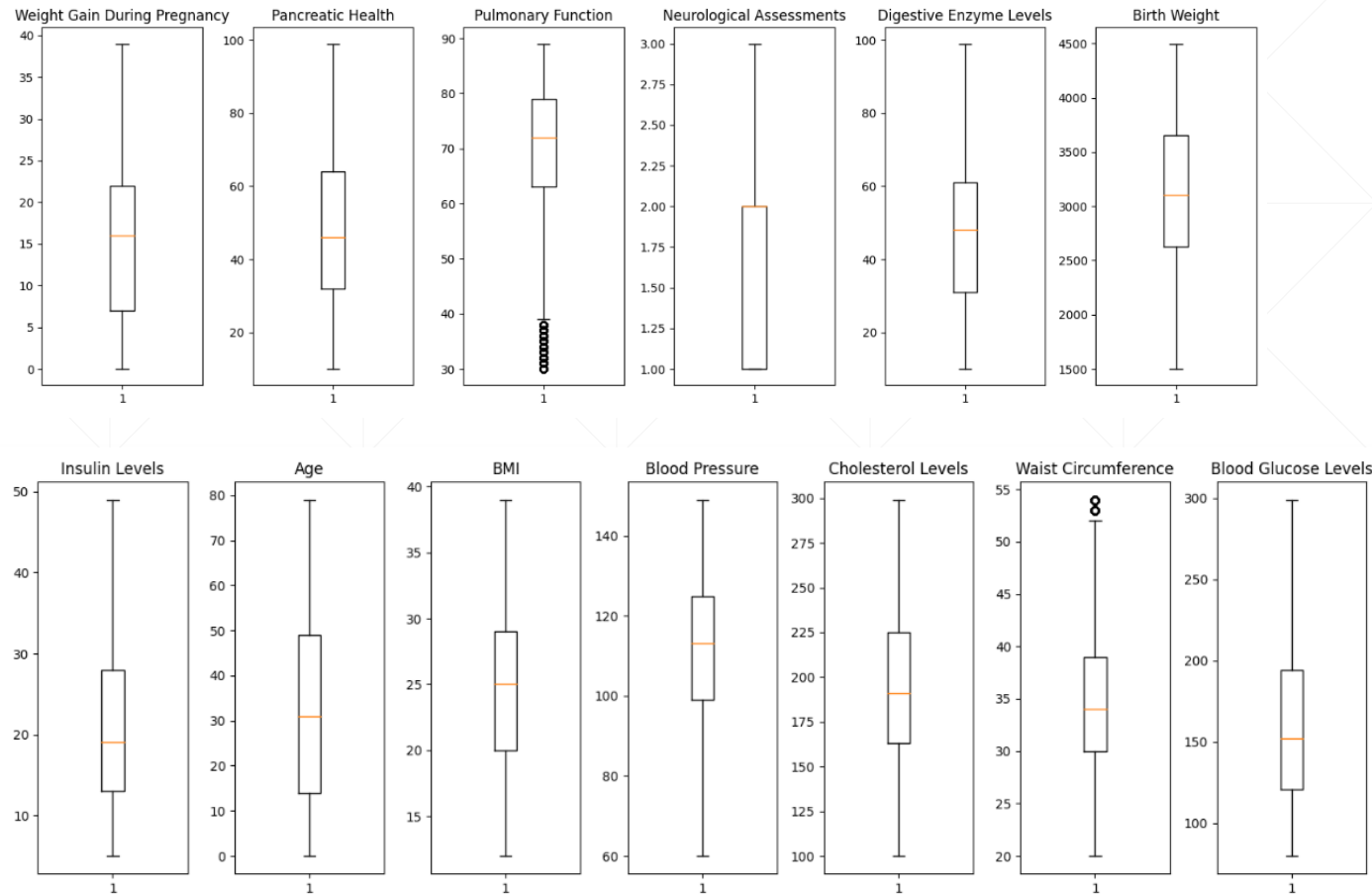  - **Target Variable (Outcome):** 8 different class of diabetes
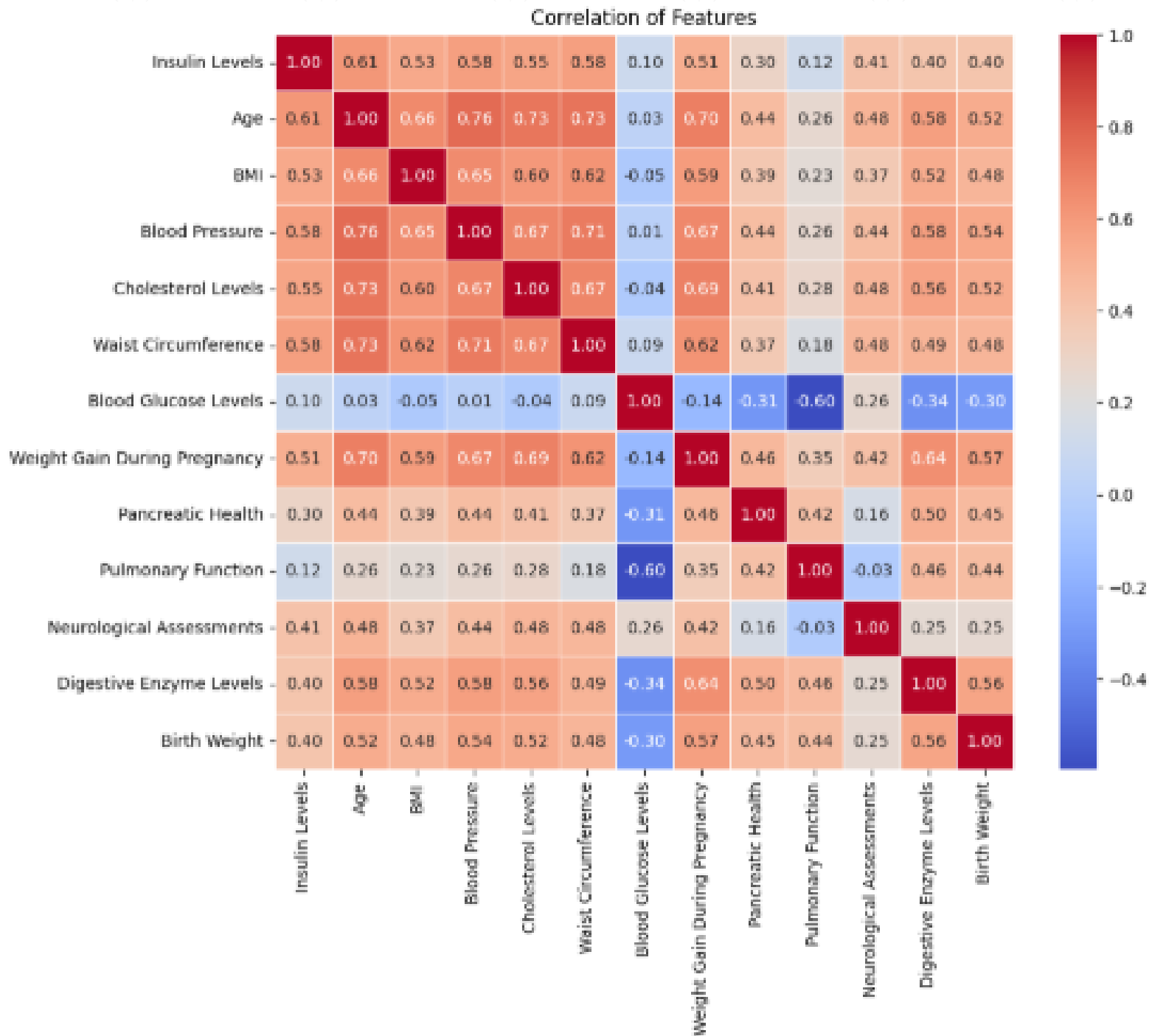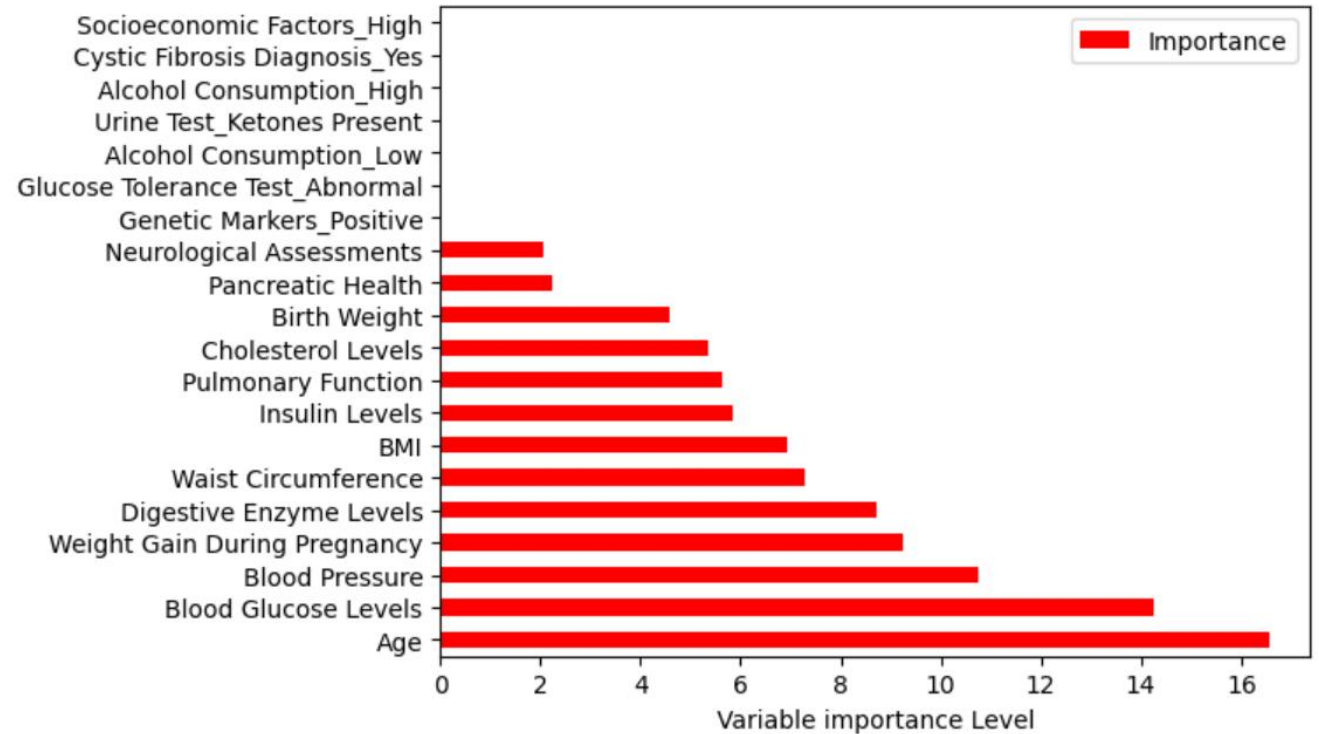
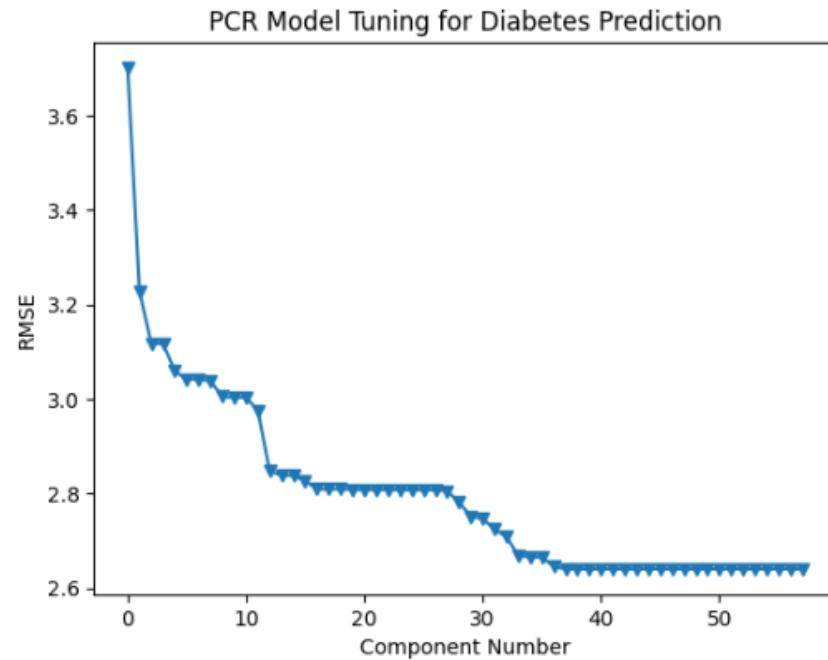# Diabetes dataset

# Diabetes dataset

# Diabetes dataset

Correlation of Features

# Feature Selection

# Performance of The Models