Diabetes Risk Prediction using Machine Learning

Francisco Montero

Overview of Diabetes

- A chronic disease that occurs when the pancreas does not produce enough insulin which leads to high glucose levels.

- Three types of diabetes are Type I-II and Gestational.

- Common symptoms of diabetes include extreme fatigue, tingling/numbness in the hands/feet, and unexpected weight loss.

Problem Area

- Reducing the risk of diabetes by identifying factors that lead to the disease and use that information formulate prevention methods.

- Project seeks to address the relationship between a patient's biomarkers and lifestyle habits that are considered unhealthy.

Proposed Data Science Solution

- Develop models to predict the risk of diabetes using:
  - Logistic Regression
  - Random Forest
  - Neural Network
- These models will:
  - Provide insight of what is influencing diabetes.
  - Identify strategies to manage the risk of the disease.

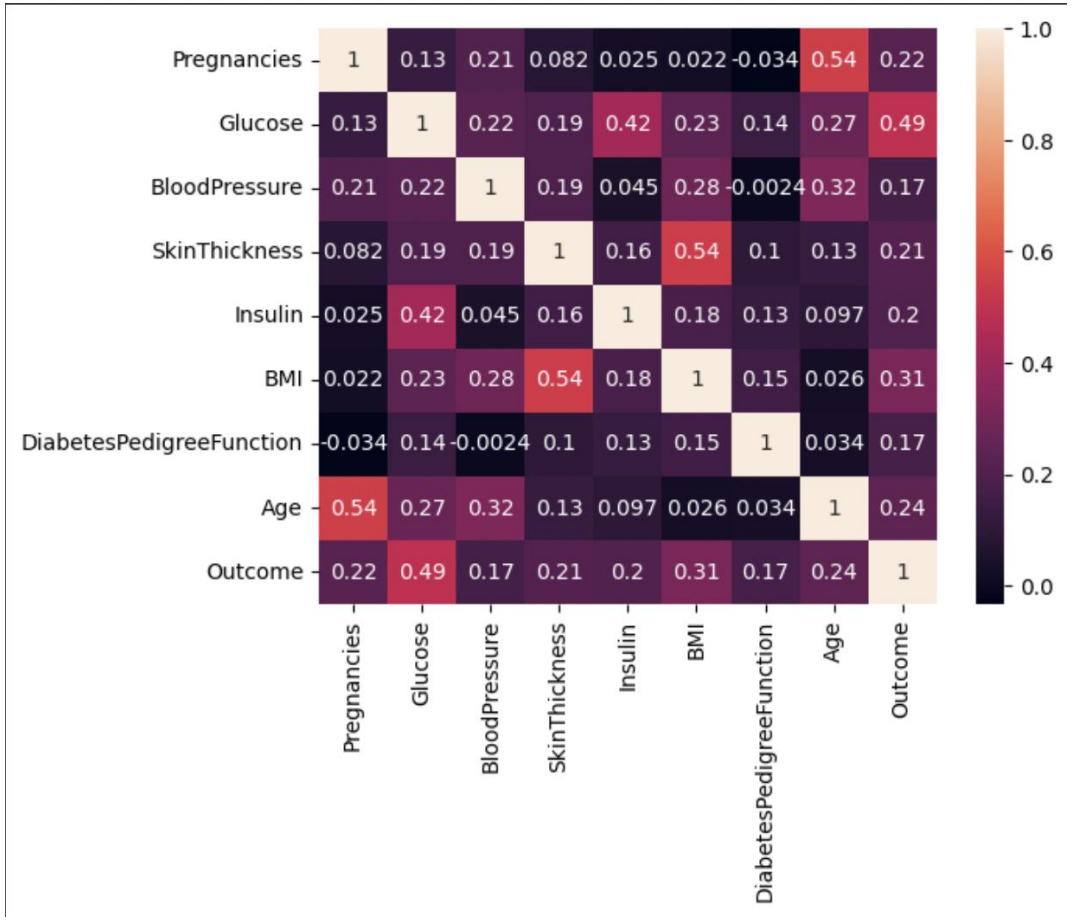
The Data

- Data Overview: Females from the Gila River Indian Community in Arizona.

- Key Variables: Age, Glucose (mg/dl), Blood Pressure (mmHg), Skin Thickness (mm), Insulin, Body Mass Index (kg/m^2), Diabetes Pedigree Function.

- Data Quality Concerns: Missing values and outliers on some of the independent variables.

The Data


Pregnancies 1 0.13 0.21 0.082 0.025 0.022 -0.034 0.54 0.22
Glucose 0.13 1 0.22 0.19 0.42 0.23 0.14 0.27 0.49
BloodPressure 0.21 0.22 1 0.19 0.045 0.28 -0.0024 0.32 0.17
SkinThickness 0.082 0.19 0.19 1 0.16 0.54 0.1 0.13 0.21
Insulin 0.025 0.42 0.045 0.16 1 0.18 0.13 0.097 0.2
BMI 0.022 0.23 0.28 0.54 0.18 1 0.15 0.026 0.31
DiabetesPedigreeFunction -0.034 0.14 -0.0024 0.1 0.13 0.15 1 0.034 0.17
Age 0.54 0.27 0.32 0.13 0.097 0.026 0.034 1 0.24
Outcome 0.22 0.49 0.17 0.21 0.2 0.31 0.17 0.24 1
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
1.0
0.8
0.6
0.4
0.2
0.0

Impact

- Supporting and growing a healthier community.
- Better health education to any affected communities.
- Healthcare workers can focus on developing resources for those who are at risk.

Next Steps

- Handling outliers.
- Train a chosen model.
- Expand modeling techniques.

Thank you!