Diabetes Prediction using Fuzzy Logic

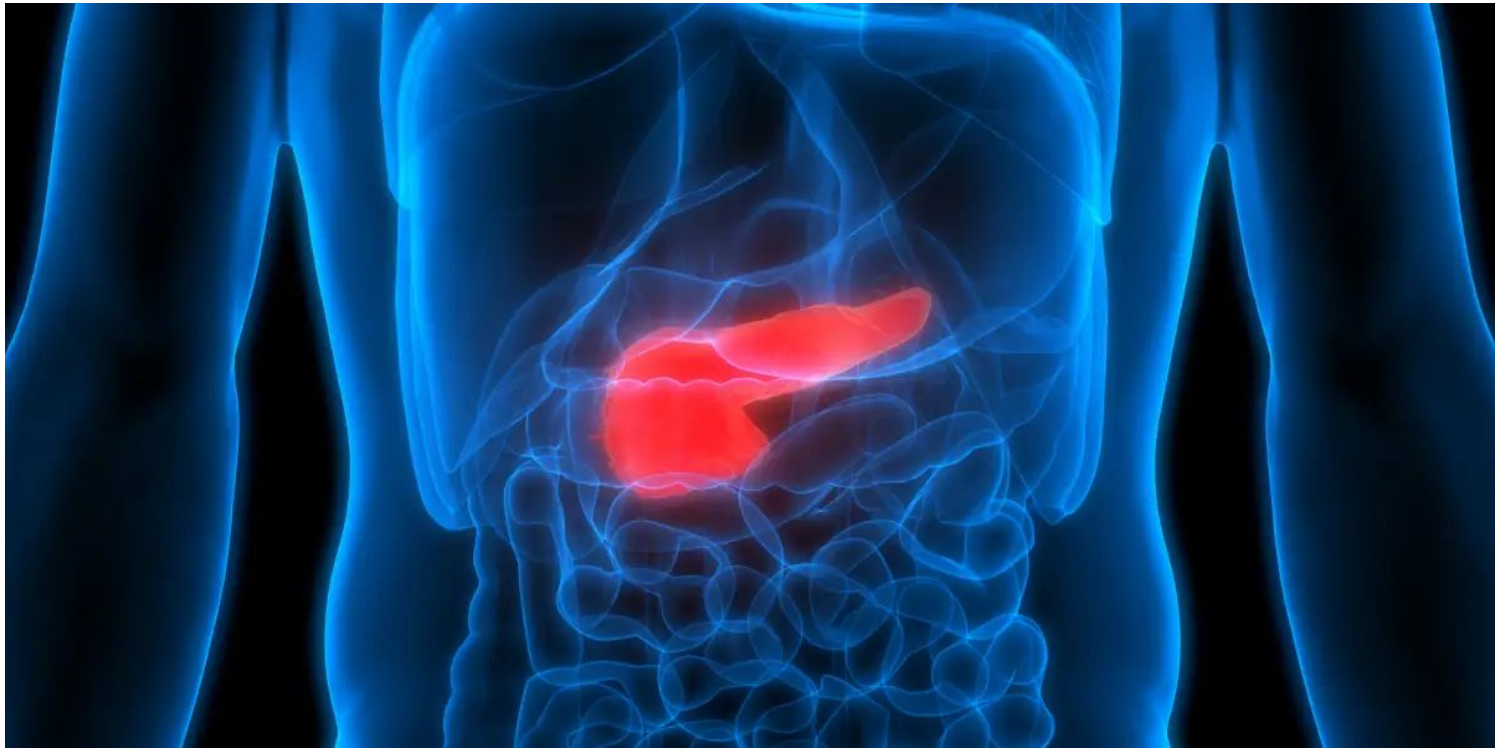
Group Fuzzing

Presenters:
Ting Wei Jing 17205876
Lee Cheng Hui 17204205
Stanley Quek 17204457
Soh Zhi Chen 17204929

01

Introduction

Problem Statement



Source: https://www.endocrineweb.com/endocrinology/overview-pancreas

Problem Statement

Source: https://www.healthline.com/health/diabetes/types-of-diabetes

Source: https://hopkinsdiabetesinfo.org/insulin-treatment-in-type-2-diabetes/

Problem Statement

Introduction

The National Diabetes Statistics Report provides up-to-date information on the prevalence and incidence of diabetes and prediabetes, risk factors for complications, acute and long-term complications, deaths, and costs.

These data can help focus efforts to prevent and control diabetes across the United States. This report is continually updated as data become available and is intended for a scientific audience.

Information about the methods is available here.

Fast Facts on Diabetes

Diabetes

- **Total:** 37.3 million people have diabetes (11.3% of the US population)
- **Diagnosed:** 28.7 million people, including 28.5 million adults
- **Undiagnosed:** 8.5 million people (23.0% of adults are undiagnosed)

Prediabetes

- **Total:** 96 million people aged 18 years or older have prediabetes (38.0% of the adult US population)
- **65 years or older:** 26.4 million people aged 65 years or older (48.8%) have prediabetes

(CDC, 2022)

Introduction

Diabetes mellitus is one of the most serious worldwide public health issues, posing a significant global burden on both public health and socioeconomic development. Although the incidence of diabetes has begun to decline in some nations, diabetes prevalence has climbed in most other developing and developed countries in recent decades [1–6]. According to the International Diabetes Federation (IDF), 9.3 percent (463 million) of adults worldwide have diabetes in 2019. The number is expected to rise to 10.2% (578 million) by 2030 and 10.9% (700 million) by 2045 if effective prevention methods are not implemented [7,8]. Furthermore, in 2017, nearly half of all people with diabetes (50.1%) were undiagnosed, approximately 374 million individuals (18–99 years) [8]. Similarly, prediabetes is estimated to affect 374 million (7.5%) of the global population in 2019 and is expected to increase to 8.0 percent (454 million) by 2030 and 548 million (8.6%) by 2045, with 48.1% of individual with prediabetes are under the age of 50 [8]. Type-2 diabetes reduces the average lifespan by around ten years [9].

Malaysia has the highest rate of diabetes in Western Pacific region and one of the highest in the world and costing around 600 million US dollars per year [10,11]. The prevalence of diabetes raised from 11.2% in 2011 to 18.3% in 2019, with a 68.3% increase [12]. According to a national survey report, in Malaysia in 2019, 3.6 million adults (18 and above years) had diabetes, 49% (3.7 million) cases were undiagnosed [13]. Diabetes is expected to affect 7 million Malaysian adults aged 18 and older by 2025, posing a major public health risk with a diabetes prevalence of 31.3% [12]. The prevalence of diabetes in Malaysia, based on published articles, ranges from 7.3% to 23.8% [14,15]. The increasing trend is a result of a variety of causes, including population expansion, population ageing, urbanization, and rising rates of obesity and physical inactivity [16]. The alarming prevalence of diabetes and its complications in Malaysia prompted this study to systematically identify, summarize available evidence on the prevalence of diabetes and prediabetes, and to estimate the pooled prevalence of diabetes and prediabetes in Malaysia. To our knowledge, no prior effort has been made to combine existing data on the prevalence of diabetes and prediabetes in Malaysian populations.

(Akhtar et al. , 2022)

Problem Statement

Diabetes often goes undetected and untreated during its early stages, which can cause serious harm to the body and makes it difficult to treat down the line.

Objectives

1. To identify individuals at high risk for developing diabetes, so that they can be targeted for preventive measures such as lifestyle interventions.

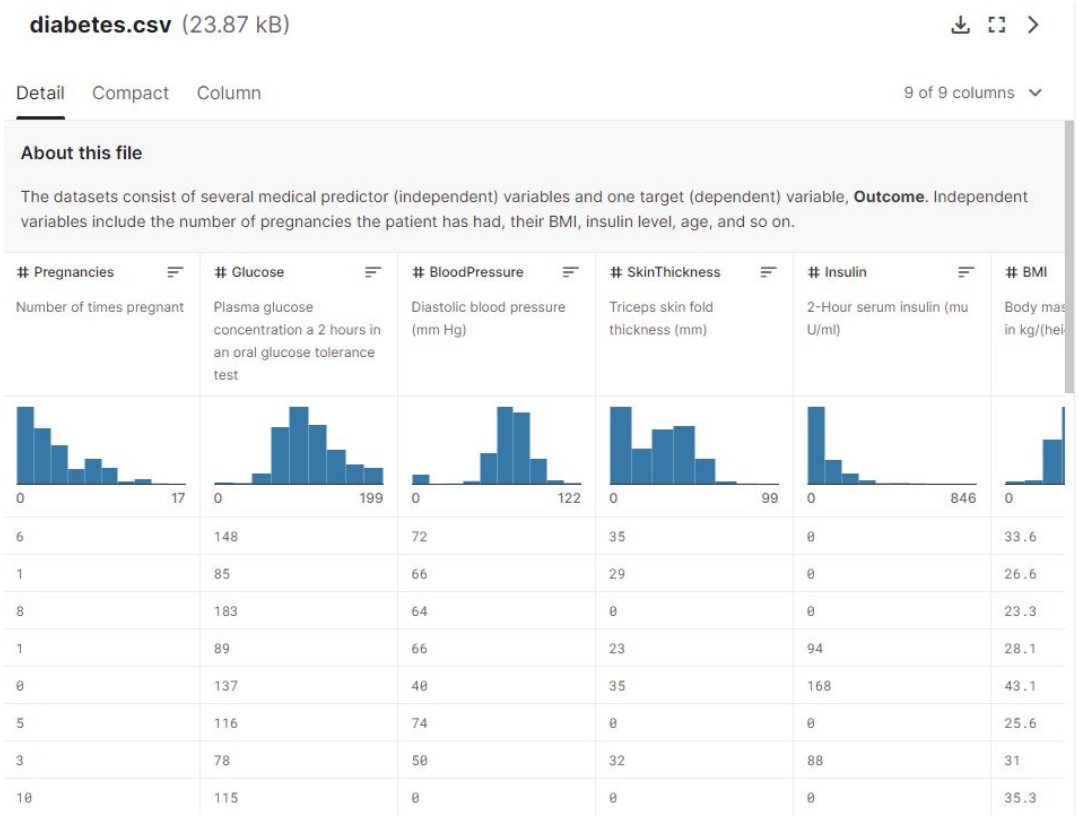
2. To diagnose diabetes at an early stage, so that people with the condition can receive treatment as soon as possible and minimise the risk of complications.

3. To assist medical professionals in the diagnosis of diabetes patients and to reduce the number of false negatives and false positive cases.

02

Describing our data

Dataset



diabetes.csv (23.87 kB)

Detail Compact Column 9 of 9 columns

About this file

The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, **Outcome**. Independent variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

# Pregnancies	# Glucose	# BloodPressure	# SkinThickness	# Insulin	# BMI
Number of times pregnant	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2-Hour serum insulin (mu U/ml)	Body mas in kg/(hei
0 17	0 199	0 122	0 99	0 846	0
6	148	72	35	0	33.6
1	85	66	29	0	26.6
8	183	64	0	0	23.3
1	89	66	23	94	28.1
0	137	40	35	168	43.1
5	116	74	0	0	25.6
3	78	50	32	88	31
10	115	0	0	0	35.3

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

Dataset Features

Column	Description
Number of pregnancies	Number of times the person gets pregnant
Plasma glucose concentration (Glucose)	Plasma glucose concentration in a person's body
Diastolic blood pressure	Diastolic blood pressure in (mm Hg)
Triceps skinfold thickness	Triceps skinfold thickness in (mm)
Serum insulin	2-Hour serum insulin (μU/ml)
Body mass index (BMI)	Body mass index (weight in kg/(height in $m)^2$)
Diabetes pedigree	History of diabetes associated with a particular person
Age	Age of a person
Class variable	Yes (diabetic) and No (non-diabetic)

Missing Features

- Glucose
- Blood Pressure
- Skin Thickness
- Insulin
- BMI

Sample size before/after removal: **768** ⟶ **392**

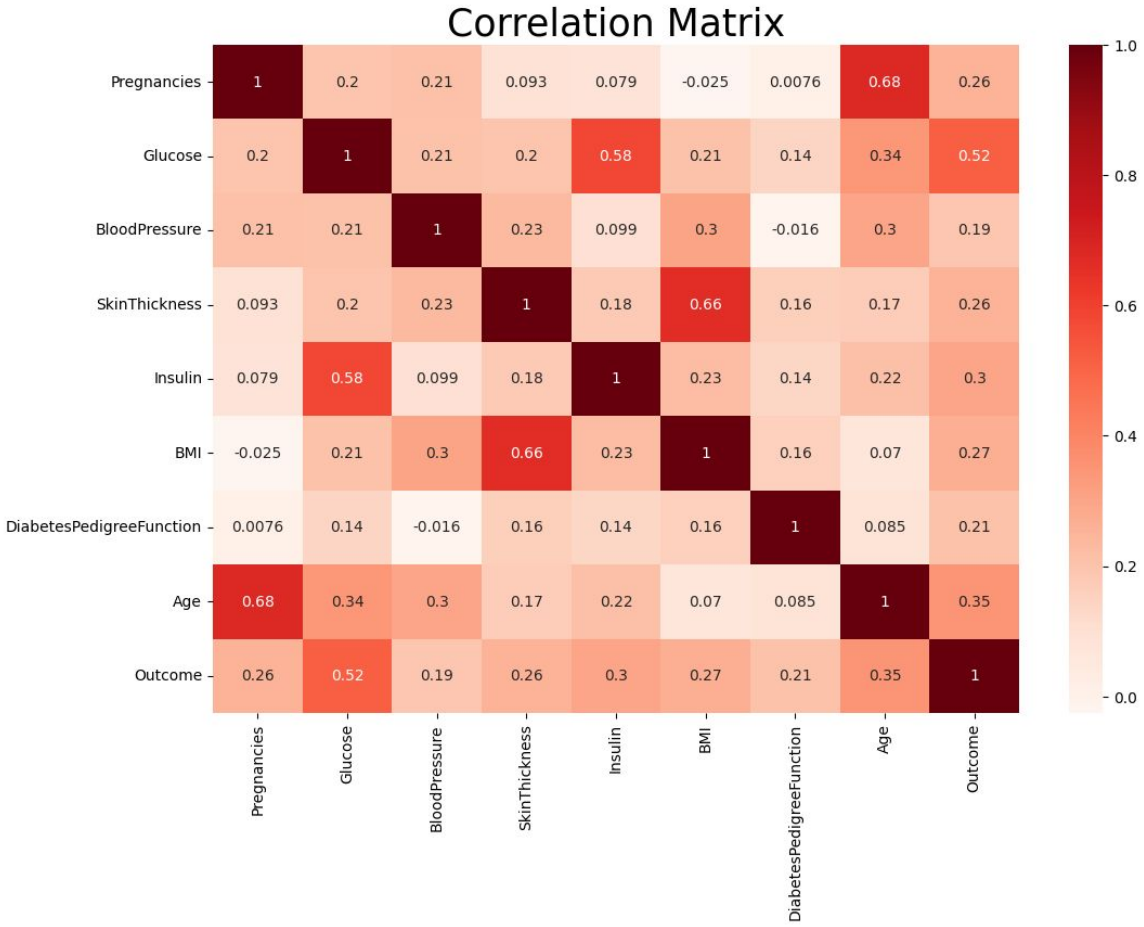
Descriptive Statistics

	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Pedigree Function	Age	Outcome
count	392	392	392	392	392	392	392	392	392
mean	3.30102	122.628	70.6633	29.1454	156.056	33.0862	0.52305	30.8648	0.33163
std	3.21142	30.8608	12.4961	10.5164	118.842	7.02766	0.34549	10.2008	0.4714
min	0	56	24	7	14	18.2	0.085	21	0
25%	1	99	62	21	76.75	28.4	0.26975	23	0
50%	2	119	70	29	125.5	33.2	0.4495	27	0
75%	5	143	78	37	190	37.1	0.687	36	1
max	17	198	110	63	846	67.1	2.42	81	1

03

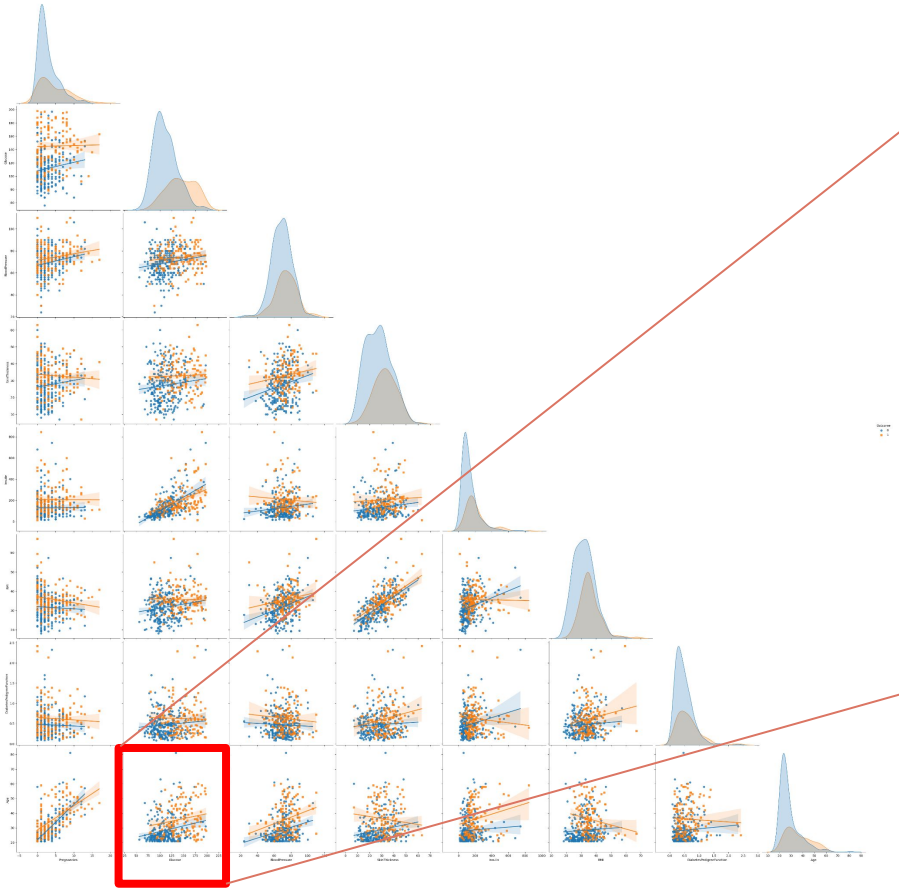
Feature Selection

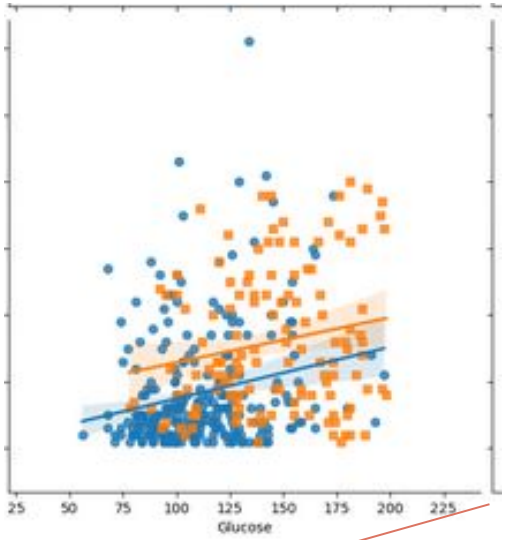
Correlation Matrix



Darker regions = Highly Correlated

Overall Pairplot

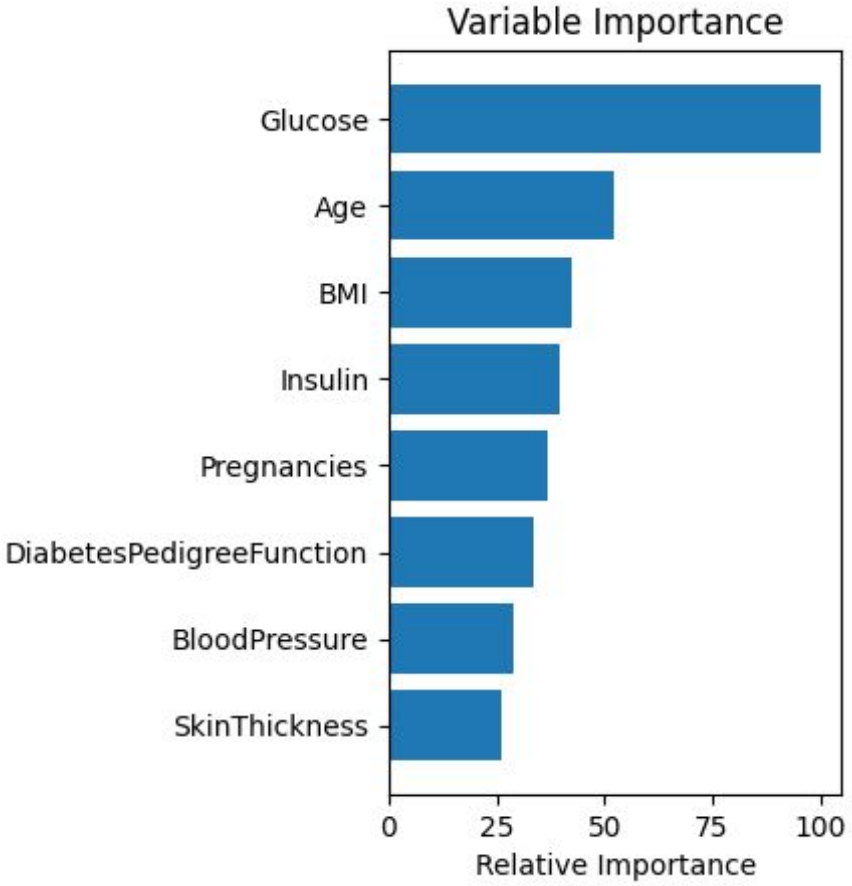




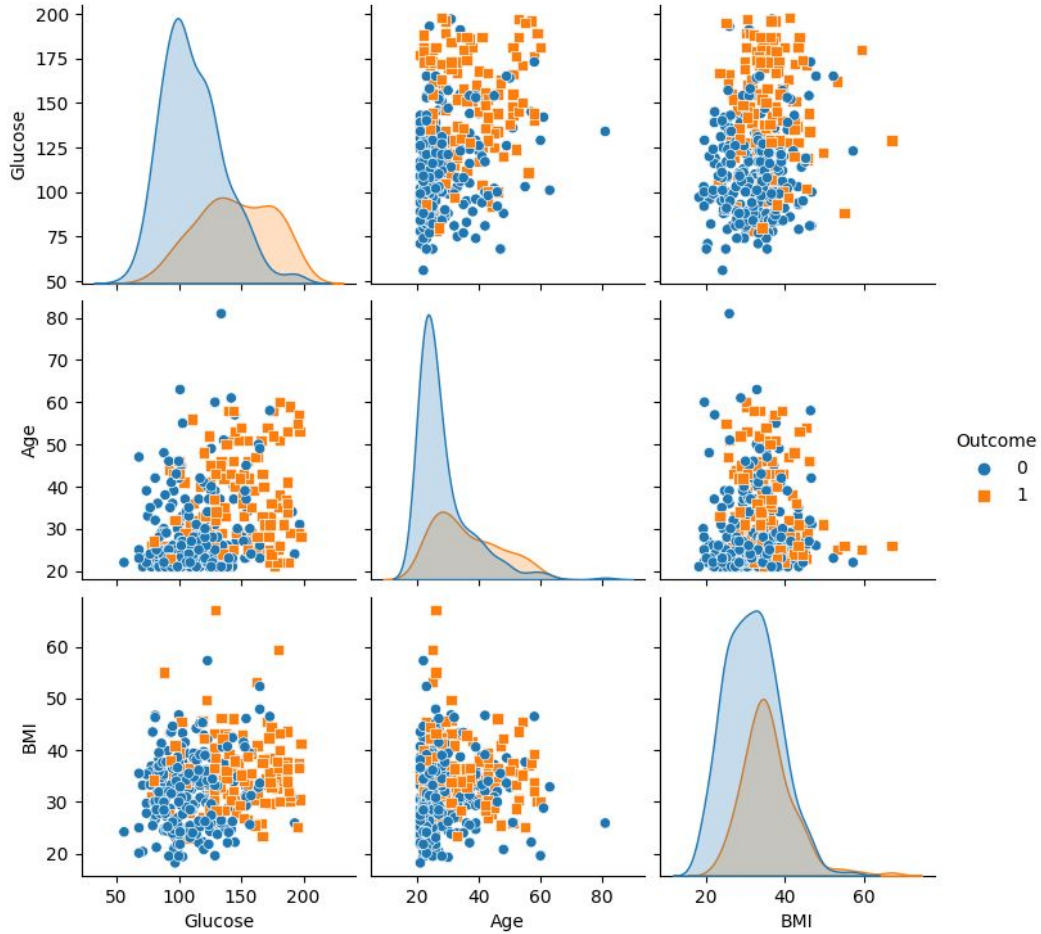
- To view how data is distributed by only considering 2 features

Feature Selection

- To determine which feature(s) contribute the most to the output.
- Calculate Feature Importance
- Glucose, Age, BMI are selected

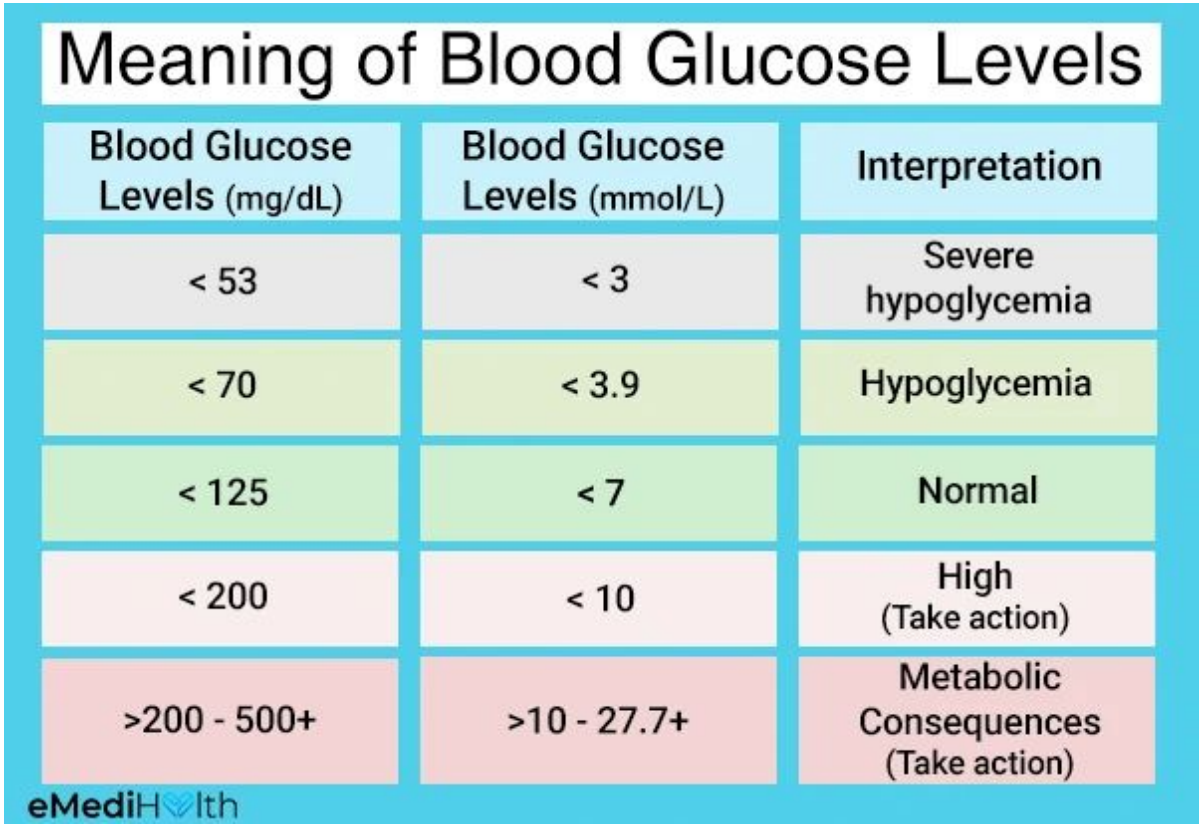


Selected Feature Pairplot



- Glucose
- Age
- BMI

Glucose Linguistic Terms



Source: https://www.emedihealth.com/glands-hormones/diabetes/manage-diabetes

Linguistic Terms	Value Range
Very Low	53 - 70
Low	70 - 97.5
Medium	97.5 - 125
High	125 - 162.5
Very High	162.5 - 200

BMI Linguistic Terms

BMI	Weight status
Below 18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Overweight
30.0-34.9	Obesity class I
35.0-39.9	Obesity class II
Above 40	Obesity class III

Linguistic Terms	Value Range
Underweight	0 - 18.5
Normal weight	18.5 - 25
Overweight	25 - 30
Obesity 1	30 - 35
Obesity 2	35 - 40
Obesity 3	40 - 68

Source:
https://theconversation.com/body-mass-index-may-not-be-the-best-indicator-of-our-health-how-can-we-improve-it-143155

Age Linguistic Terms

Age Group	N	Mean Age +/- SD	Age Range
Children	27	11.04 +/- 1.72	8.0 to 13.99
Adolescents	32	16.38 +/- .78	14.0 to 17.99
Young Adults	18	20.25 +/- 1.15	18.0 to 21.99
Middle Age Adults	25	28.49 +/- 5.40	22.0 to 42.99
Older Adults	43	63.67 +/- 5.46	51.0 to 79.99

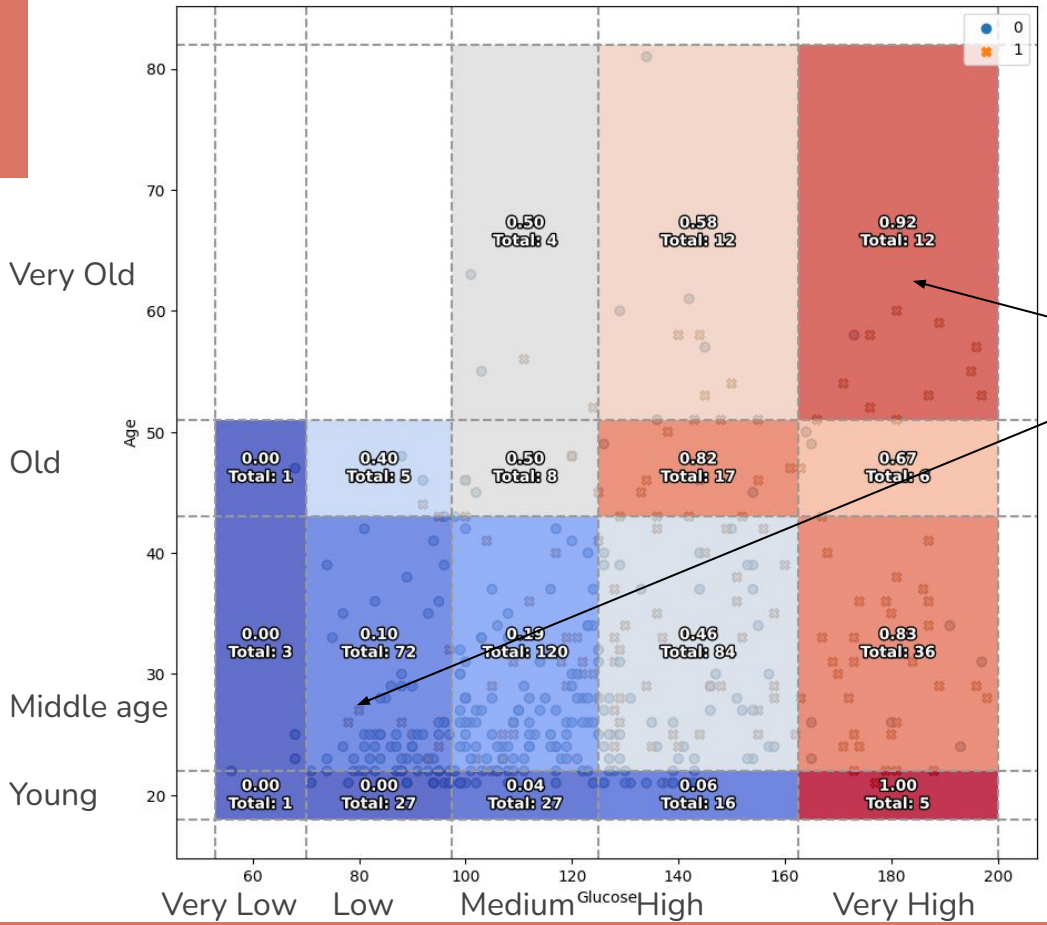
Linguistic Terms	Value Range
Young	18 - 22
Middle Age	22 - 43
Old	43 - 51
Very Old	51 - 82

Source: https://www.researchgate.net/figure/Subject-Age-Group-Information_fig4_279304799

04

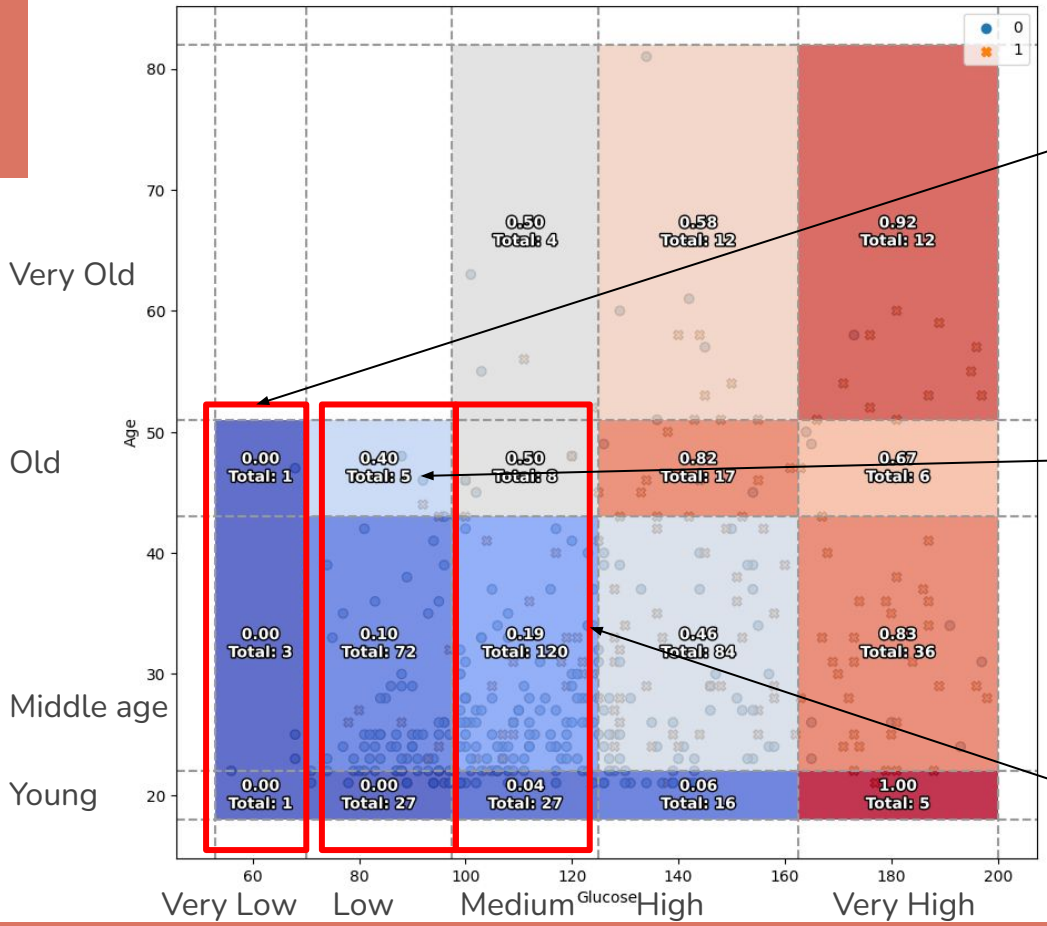
Rules Construction

Glucose vs Age rules construction



1. All the points are plotted
2. Drew boundary lines
3. Calculate ratio
4. Select dark regions (<= 0.2 or >= 0.8)

Glucose vs Age rules construction

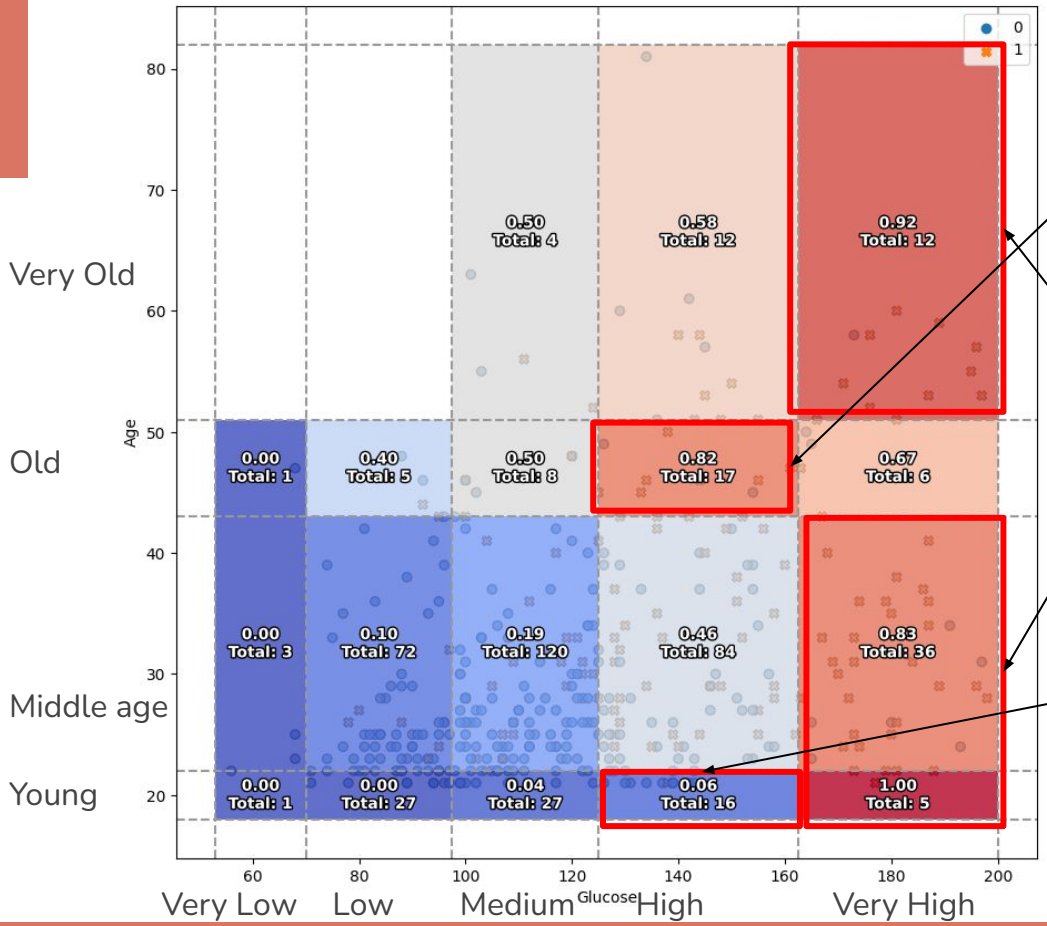


*1. If the glucose level is **very low** AND the age is **young** OR **middle age** OR **old**, then the prediction for diabetes is **no**.*

*2. If the glucose level is **low** AND the age is **young** OR **middle age**, then the prediction for diabetes is **no**.*

*3. If the glucose level is **medium** AND the age is **young** OR **middle age**, then the prediction for diabetes is **no**.*

Glucose vs Age rules construction

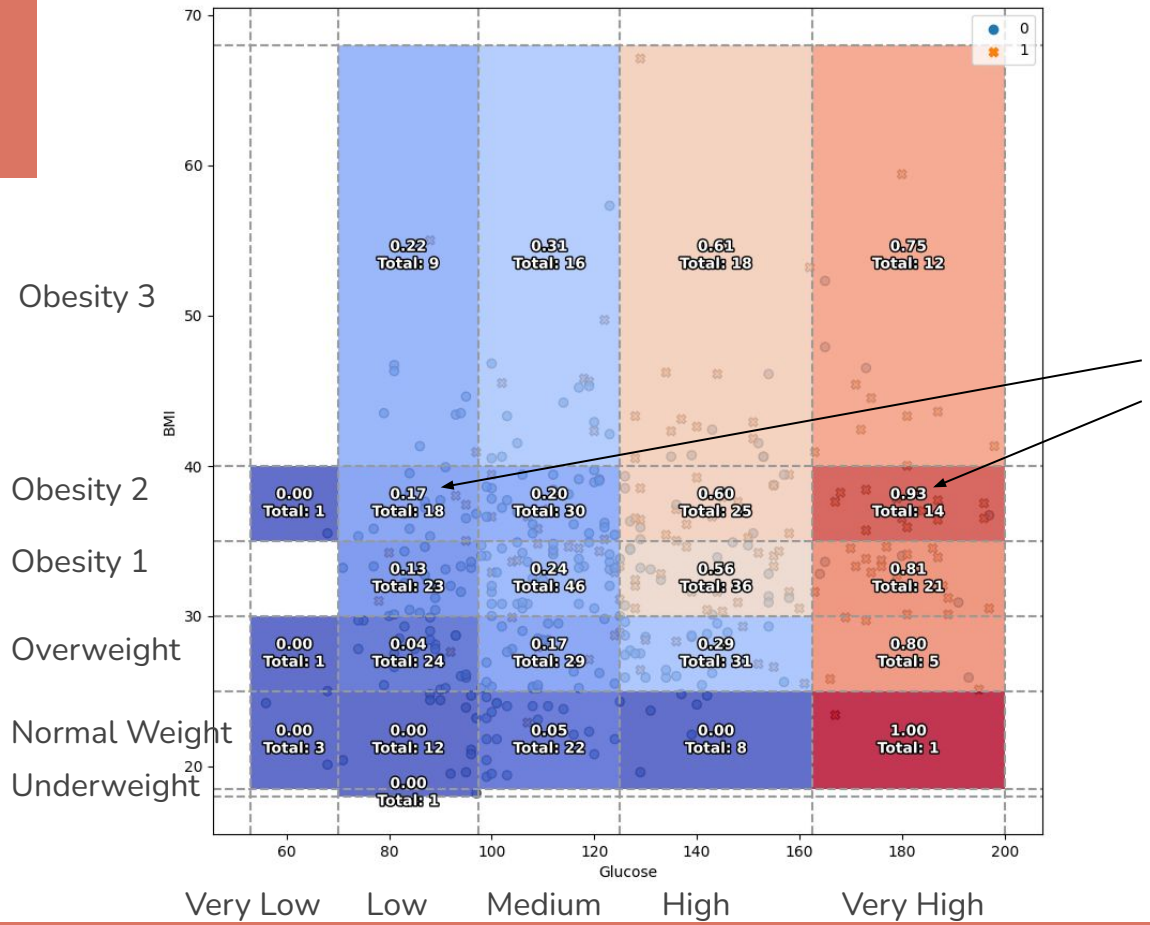


*4. If the glucose level is **high** AND the age is **old**, then the prediction for diabetes is **yes**.*

*5. If the glucose level is **very high** AND the age is **young** OR **middle age** OR **very old**, then the prediction for diabetes is **yes**.*

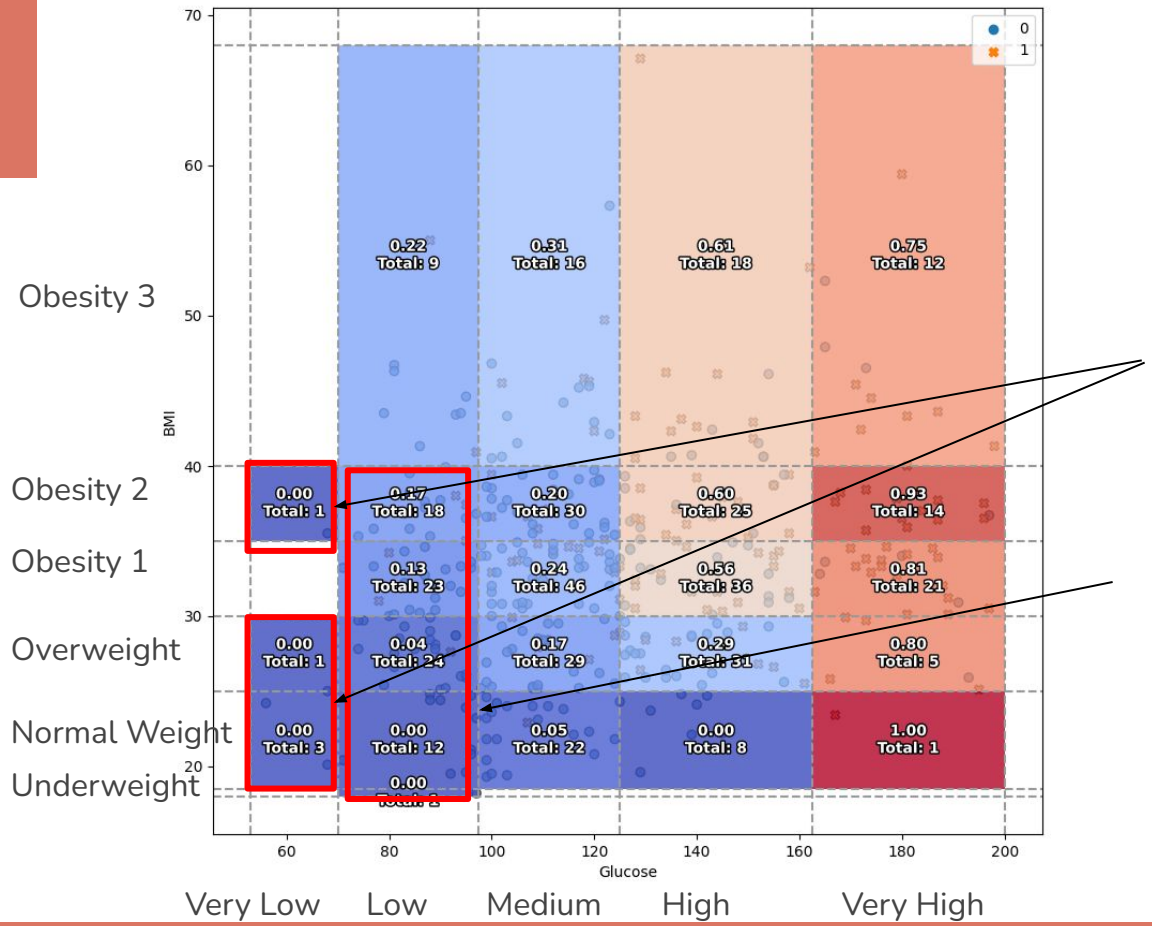
*6. If the glucose level is **high** AND the age is **young**, then the prediction for diabetes is **no**.*

Glucose vs BMI rules construction



1. All the points are plotted
2. Drew boundary lines
3. Calculate ratio
4. Select dark regions (<= 0.2 or >= 0.8)

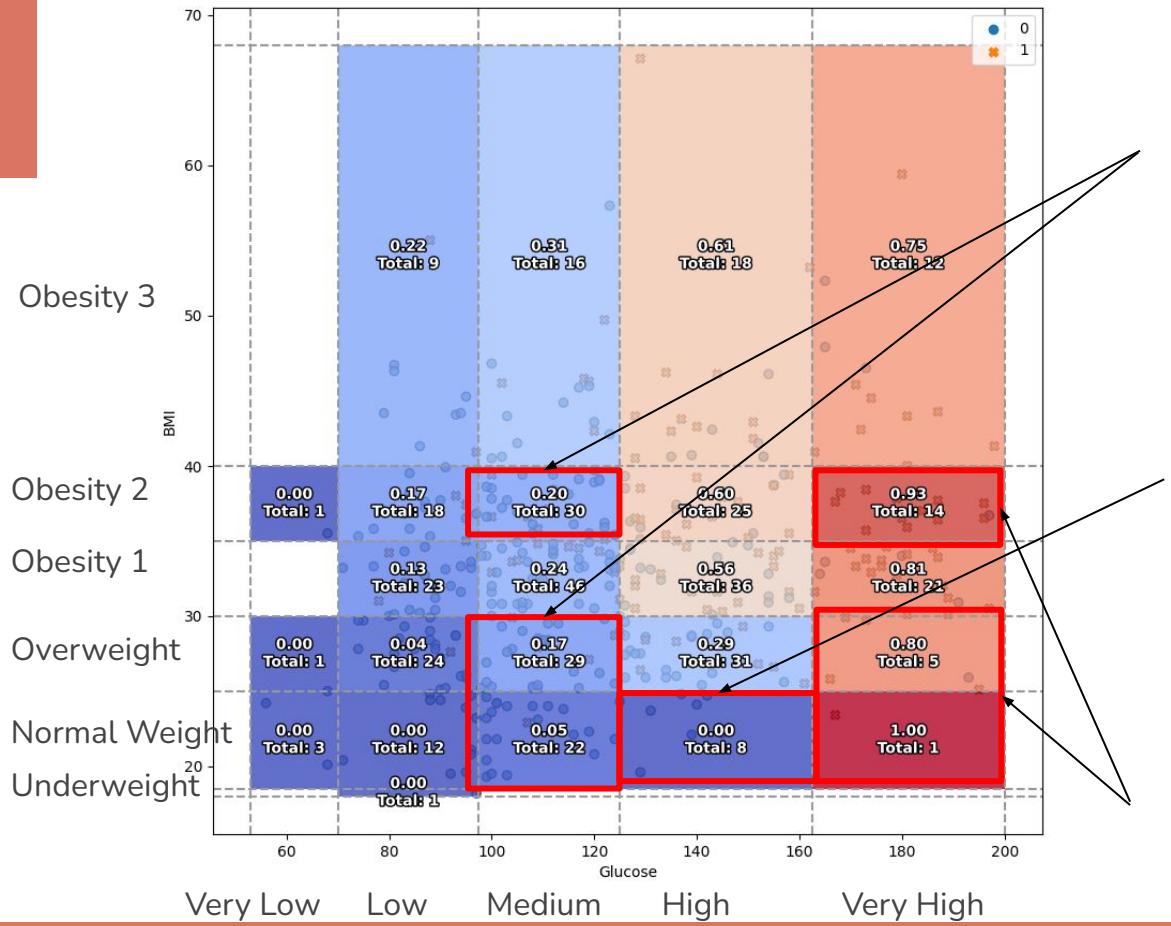
Glucose vs BMI rules construction



7. If the glucose level is ***very low*** *AND the BMI is* ***normal weight*** *OR* ***overweight*** *OR* ***obesity 2****, then the prediction for diabetes is* ***no****.*

8. If the glucose level is ***low*** *AND the BMI is* ***underweight*** *OR* ***normal weight*** *OR* ***overweight*** *OR* ***obesity 1*** *OR* ***obesity 2****, then the prediction for diabetes is* ***no****.*

Glucose vs BMI rules construction



*9. If the glucose level is **medium** AND the BMI is **normal weight** OR **overweight** OR **obesity 2**, then the prediction for diabetes is **no**.*

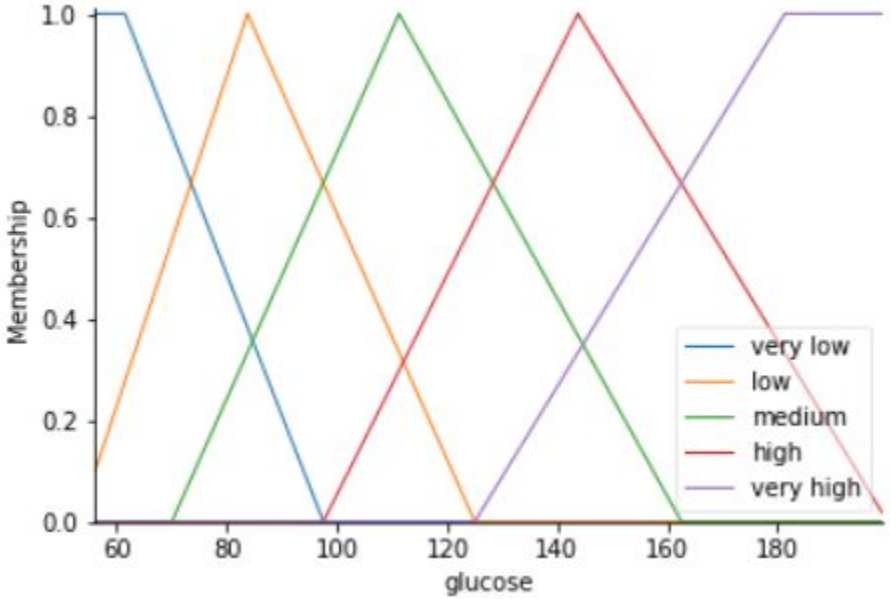
*10. If the glucose level is **high** AND the BMI is **normal weight**, then the prediction for diabetes is **no**.*

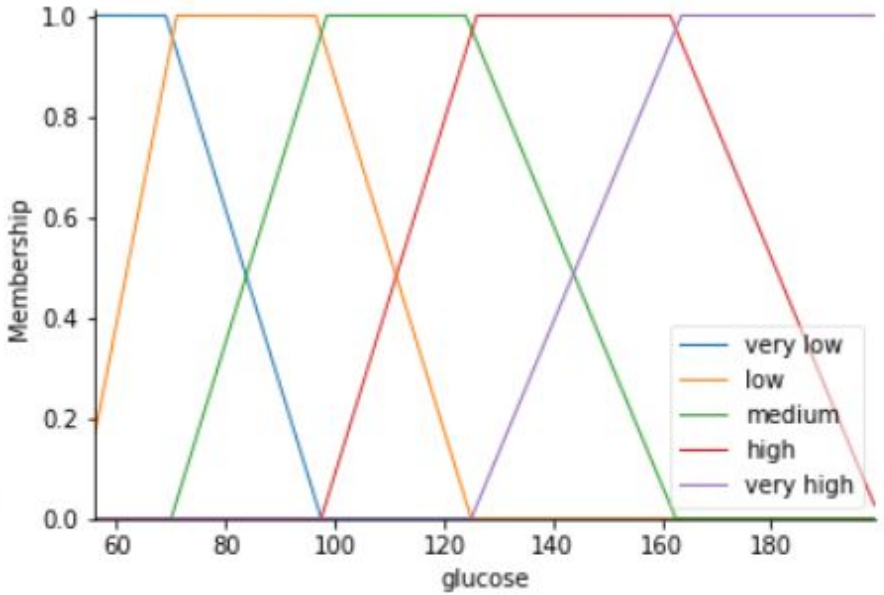
*11. If the glucose level is **very high** AND the BMI is **normal weight** OR **overweight** OR **obesity 1** OR **obesity 2**, the prediction for diabetes is **yes**.*

Why Mamdani over Sugeno?

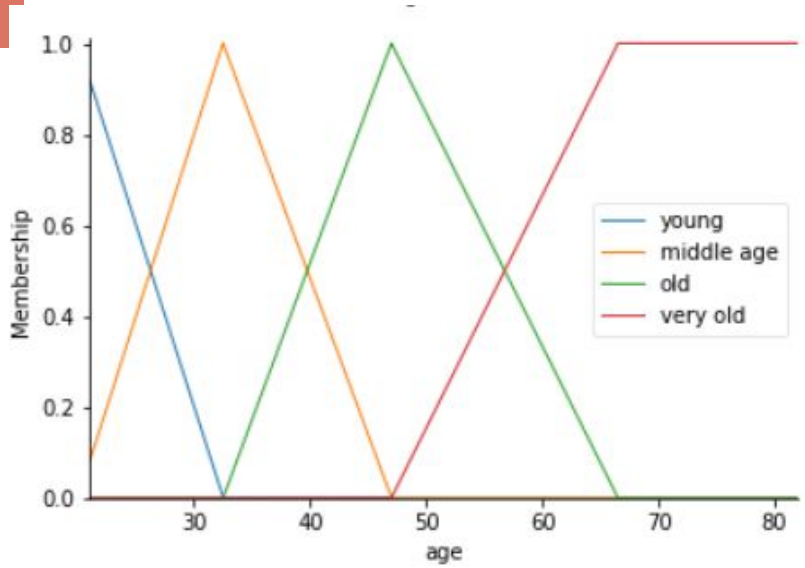
- Easier rules construction due to available researches and knowledge base
- Well interpretable
- No need to design extra crisp functions for consequent

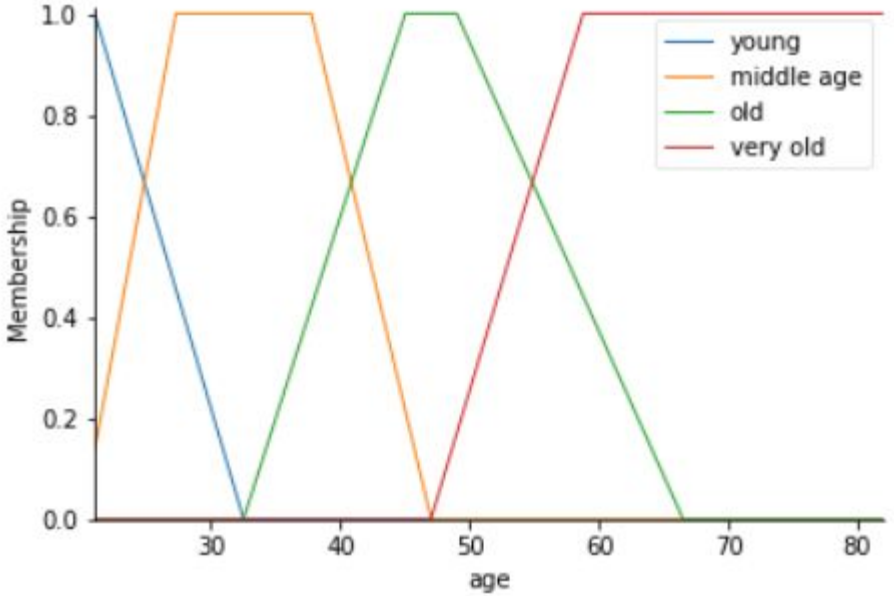
Glucose Membership Function



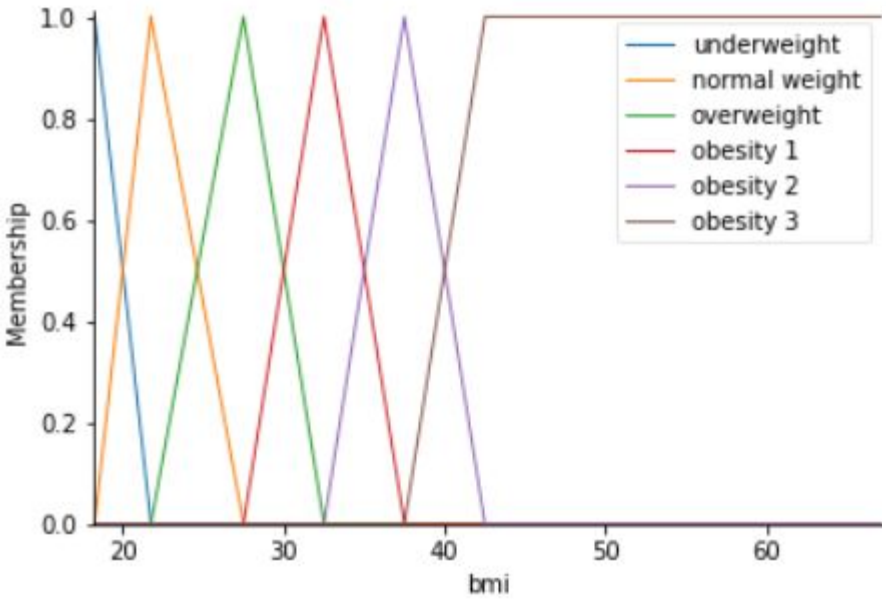


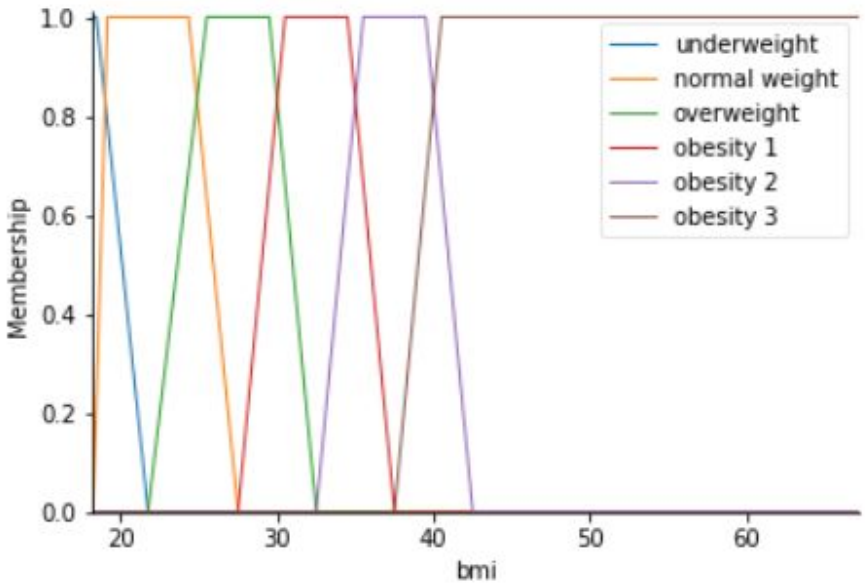
Age Membership Function



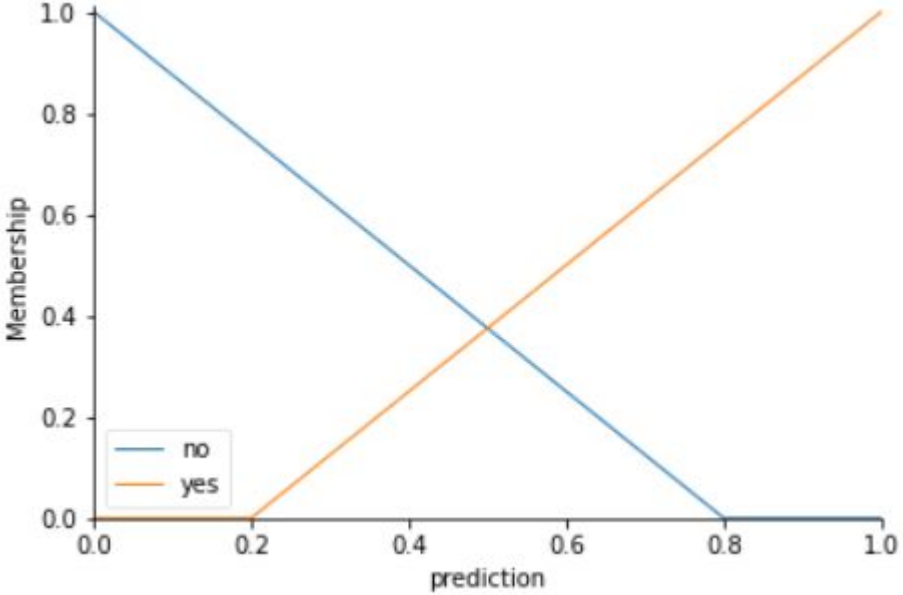


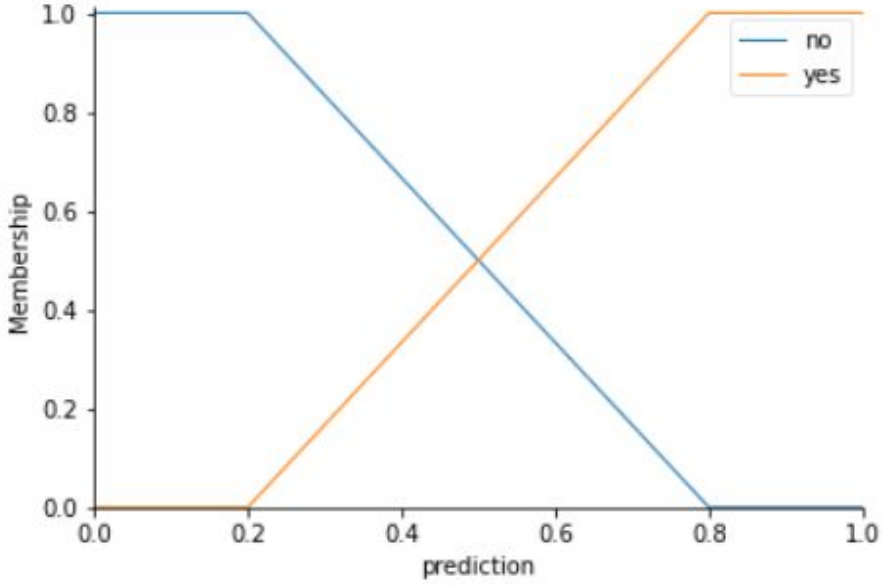
BMI Membership Function





Prediction Membership Function





05

Experiment & Fine Tuning

Combination of Membership Function Shape

Glucose	BMI	Age	Prediction	Accuracy (%)
Triangle	Triangle	Triangle	Triangle	79.85
Trapezium	Triangle	Triangle	Triangle	78.83
Triangle	Trapezium	Triangle	Triangle	80.36
Trapezium	Trapezium	Triangle	Triangle	79.85
Triangle	Triangle	Trapezium	Triangle	79.85
Trapezium	Triangle	Trapezium	Triangle	78.57
Triangle	Trapezium	Trapezium	Triangle	80.36
Trapezium	Trapezium	Trapezium	Triangle	79.34
Triangle	Triangle	Triangle	Trapezium	79.85
Trapezium	Triangle	Triangle	Trapezium	78.83
Triangle	Trapezium	Triangle	Trapezium	80.36
Trapezium	Trapezium	Triangle	Trapezium	79.85
Triangle	Triangle	Trapezium	Trapezium	79.85
Trapezium	Triangle	Trapezium	Trapezium	78.57
Triangle	Trapezium	Trapezium	Trapezium	80.36
Trapezium	Trapezium	Trapezium	Trapezium	79.34

Defuzzification Method

Method	Accuracy (%)
Centroid	80.36
Bisector of Area	80.36
Mean of maximum	80.36
Min of maximum	80.36
Max of maximum	80.10

06

Demonstration