Urszula Szczęsna
Gaspar Sekula
Mateusz Deptuch

# Obesity risk prediction

# Cel projektu

Celem projektu jest przewidywanie ryzyka otyłości u pacjentów na podstawie ich ogólnego stanu zdrowia i stylu życia.

# Zastosowanie biznesowe

Nasz model może być używany przez lekarzy w szpitalach i klinikach, a także przez zwykłych ludzi dbających o swoje zdrowie, chcących zapobiegać nadwadze I otyłości.

# Dane

20758

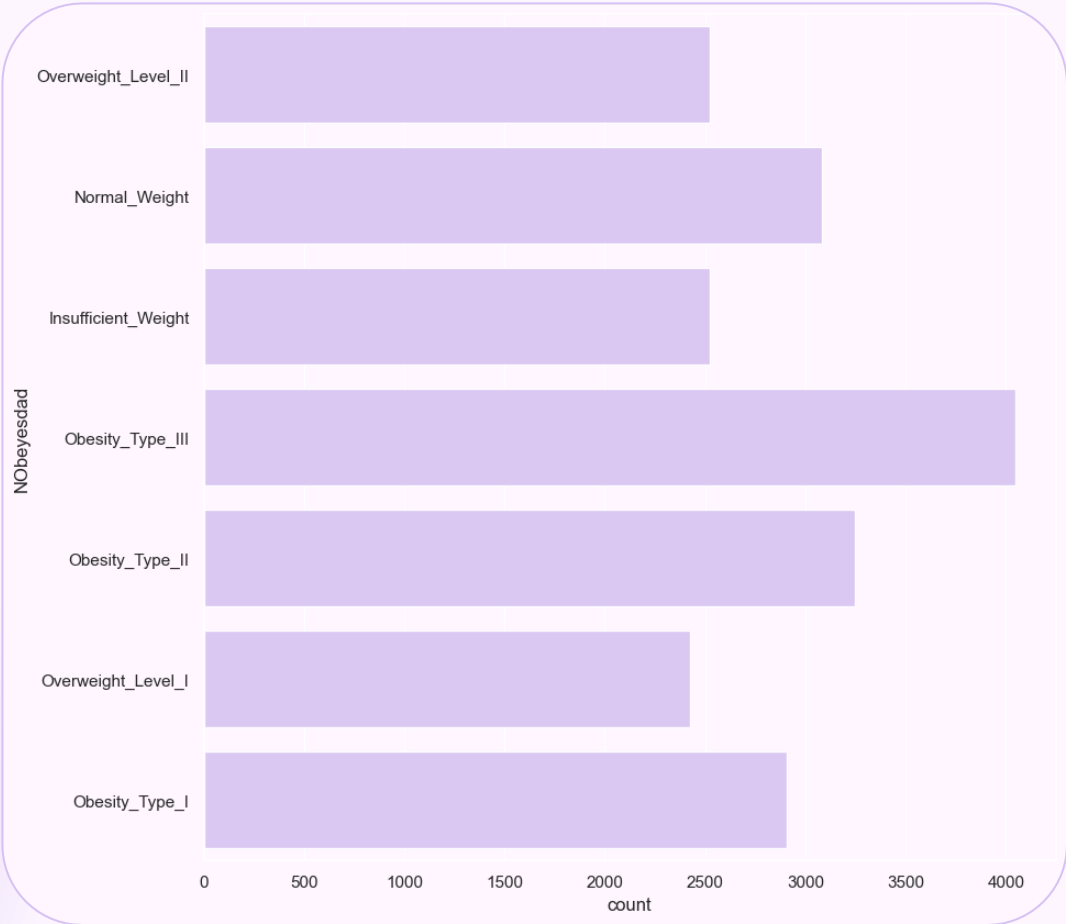
wierszy

18

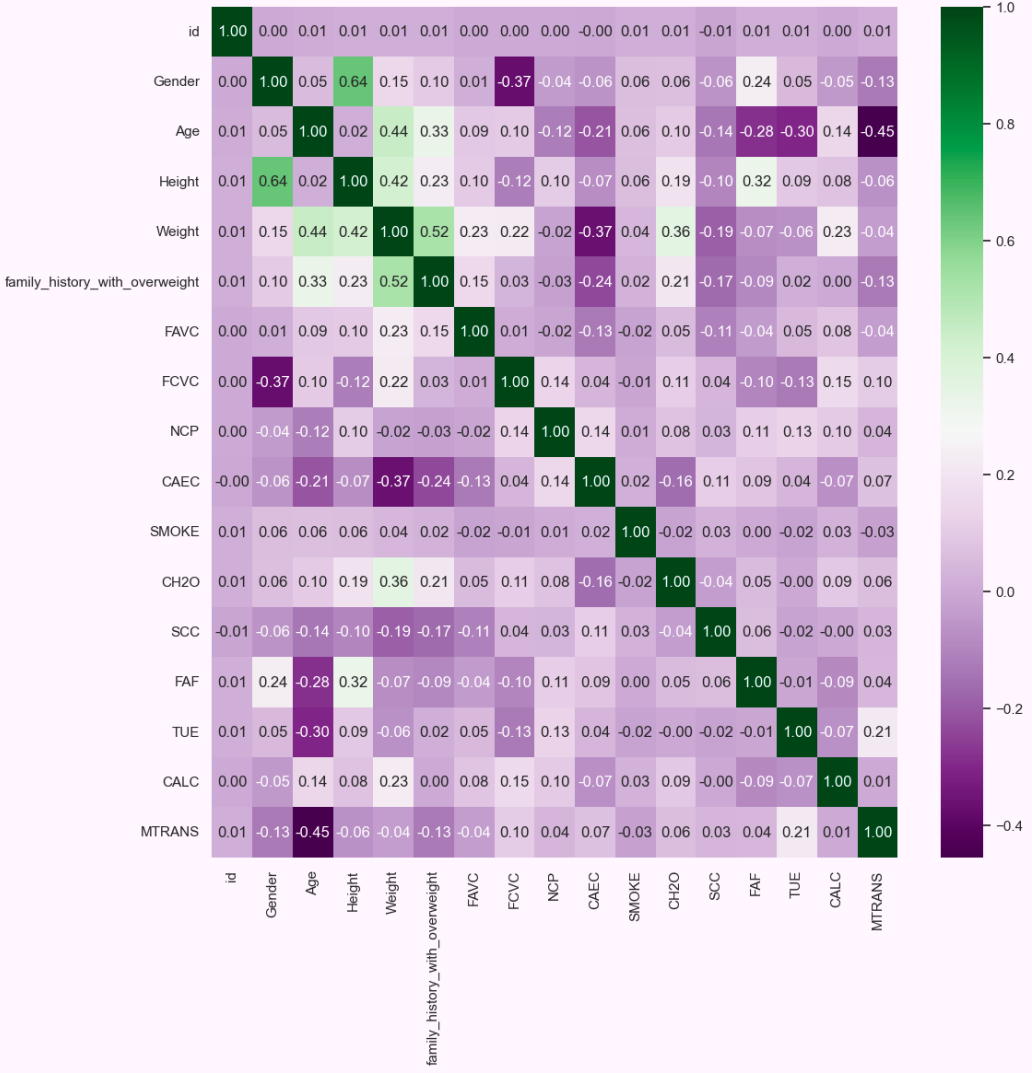
kolumn

KLASA TARGET:

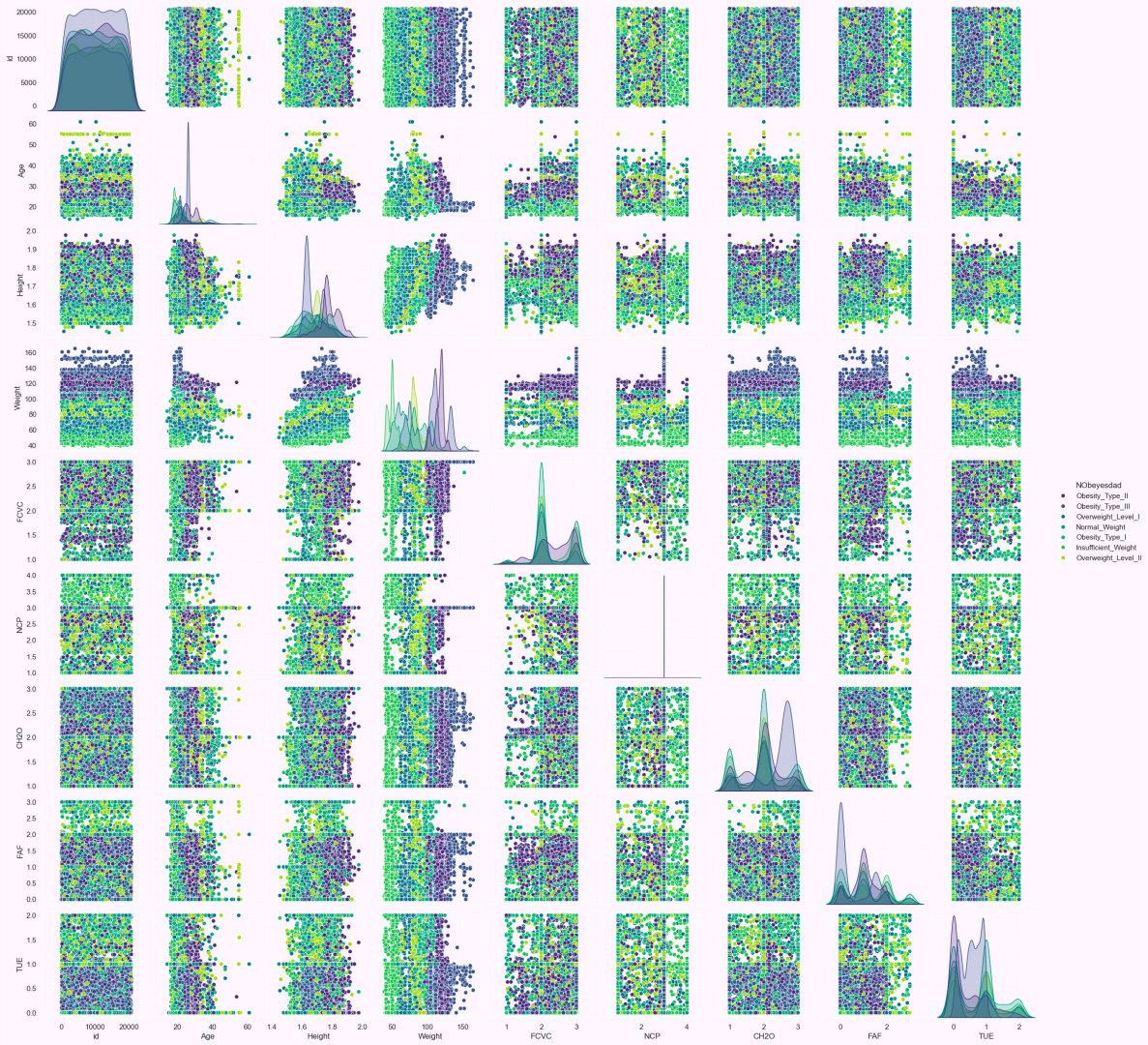
- Obesity_Type_III
- Obesity_Type_II
- Obesity_Type_I
- Overweight_Level_II
- Overweight_Level_I
- Normal_Weight
- Insufficient_Weight



# EDA


id 1.00 0.00 0.01 0.01 0.01 0.01 0.00 0.00 0.00 -0.00 0.01 0.01 -0.01 0.01 0.01 0.00 0.01
Gender 0.00 1.00 0.05 0.64 0.15 0.10 0.01 -0.37 -0.04 -0.06 0.06 0.06 -0.06 0.24 0.05 -0.05 -0.13
Age 0.01 0.05 1.00 0.02 0.44 0.33 0.09 0.10 -0.12 -0.21 0.06 0.10 -0.14 -0.28 -0.30 0.14 -0.45
Height 0.01 0.64 0.02 1.00 0.42 0.23 0.10 -0.12 0.10 -0.07 0.06 0.19 -0.10 0.32 0.09 0.08 -0.06
Weight 0.01 0.15 0.44 0.42 1.00 0.52 0.23 0.22 -0.02 -0.37 0.04 0.36 -0.19 -0.07 -0.06 0.23 -0.04
family_history_with_overweight 0.01 0.10 0.33 0.23 0.52 1.00 0.15 0.03 -0.03 -0.24 0.02 0.21 -0.17 -0.09 0.02 0.00 -0.13
FAVC 0.00 0.01 0.09 0.10 0.23 0.15 1.00 0.01 -0.02 -0.13 -0.02 0.05 -0.11 -0.04 0.05 0.08 -0.04
FCVC 0.00 -0.37 0.10 -0.12 0.22 0.03 0.01 1.00 0.14 0.04 -0.01 0.11 0.04 -0.10 -0.13 0.15 0.10
NCP 0.00 -0.04 -0.12 0.10 -0.02 -0.03 -0.02 0.14 1.00 0.14 0.01 0.08 0.03 0.11 0.13 0.10 0.04
CAEC -0.00 -0.06 -0.21 -0.07 -0.37 -0.24 -0.13 0.04 0.14 1.00 0.02 -0.16 0.11 0.09 0.04 -0.07 0.07
SMOKE 0.01 0.06 0.06 0.06 0.04 0.02 -0.02 -0.01 0.01 0.02 1.00 -0.02 0.03 0.00 -0.02 0.03 -0.03
CH2O 0.01 0.06 0.10 0.19 0.36 0.21 0.05 0.11 0.08 -0.16 -0.02 1.00 -0.04 0.05 -0.00 0.09 0.06
SCC -0.01 -0.06 -0.14 -0.10 -0.19 -0.17 -0.11 0.04 0.03 0.11 0.03 -0.04 1.00 0.06 -0.02 -0.00 0.03
FAF 0.01 0.24 -0.28 0.32 -0.07 -0.09 -0.04 -0.10 0.11 0.09 0.00 0.05 0.06 1.00 -0.01 -0.09 0.04
TUE 0.01 0.05 -0.30 0.09 -0.06 0.02 0.05 -0.13 0.13 0.04 -0.02 -0.00 -0.02 -0.01 1.00 -0.07 0.21
CALC 0.00 -0.05 0.14 0.08 0.23 0.00 0.08 0.15 0.10 -0.07 0.03 0.09 -0.00 -0.09 -0.07 1.00 0.01
MTRANS 0.01 -0.13 -0.45 -0.06 -0.04 -0.13 -0.04 0.10 0.04 0.07 -0.03 0.06 0.03 0.04 0.21 0.01 1.00
id Gender Age Height Weight family_history_with_overweight FAVC FCVC NCP CAEC SMOKE CH2O SCC FAF TUE CALC MTRANS
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

# EDA


id
Age
Height
Weight
FCVC
NCP
CH2O
FAF
TUE
NObeyesdad
Obesity_Type_II
Obesity_Type_III
Overweight_Level_I
Normal_Weight
Obesity_Type_I
Insufficient_Weight
Overweight_Level_II

usunięcie ”ID”

dodanie kolumny “BMI”

encoding i standaryzacja

# Feature Engineering

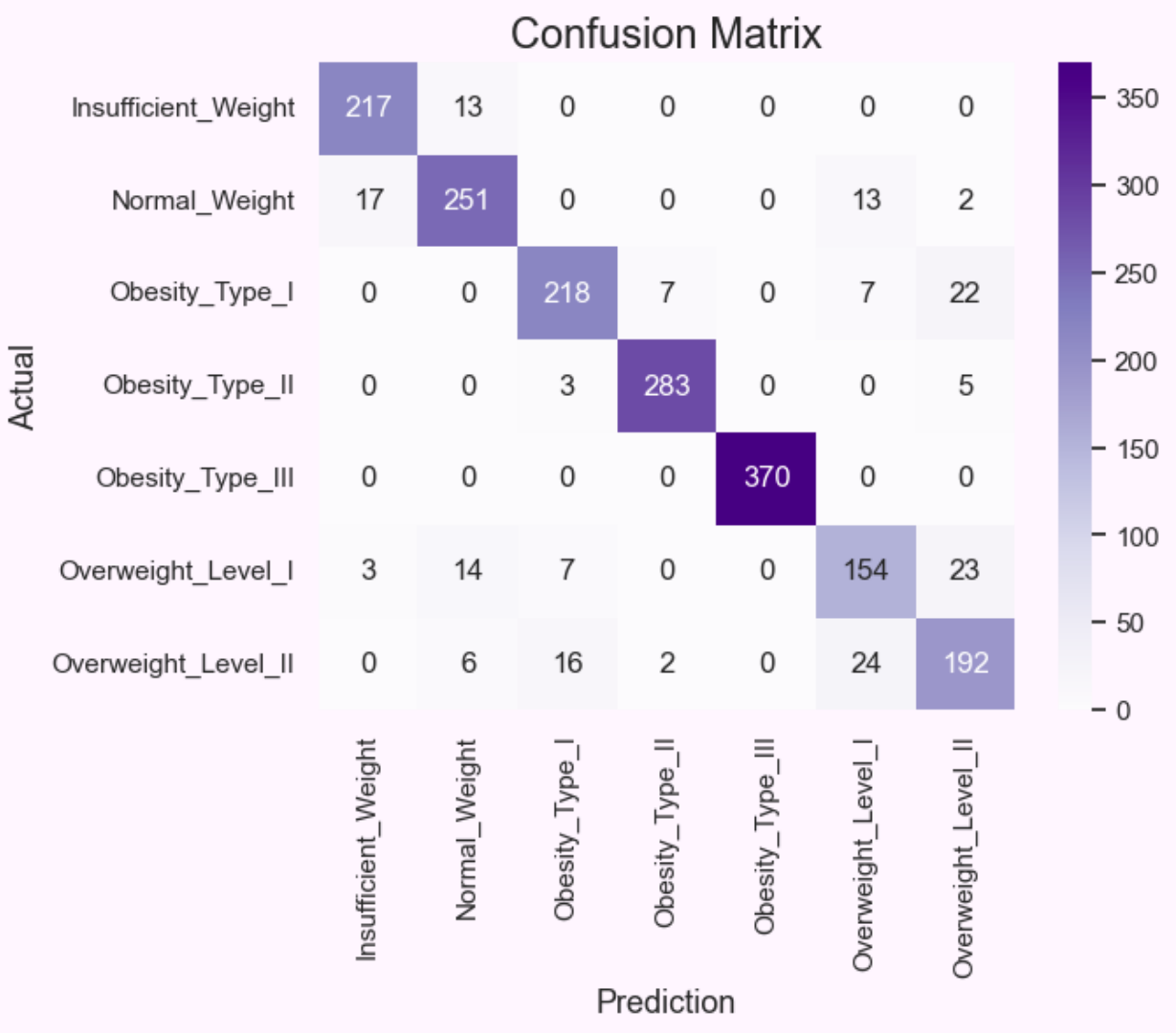
# → Modelowanie

| Model | Train Accuracy | Test Accuracy |
|---|---|---|
| Random Forest | 0.89 | 0.88 |
| Decision Tree | 0.86 | 0.86 |
| SVC | 0.86 | 0.86 |
| Logistic Regression | 0.85 | 0.85 |
| Naive Bayes | 0.74 | 0.75 |
| XGBoost | 0.89 | 0.89 |
| SGD | 0.73 | 0.75 |
| K-Neighbors | 0.73 | 0.74 |
| AdaBoost | 0.68 | 0.56 |
| Stacking | 0.90 | 0.88 |
| TPOT | 0.89 | 0.88 |

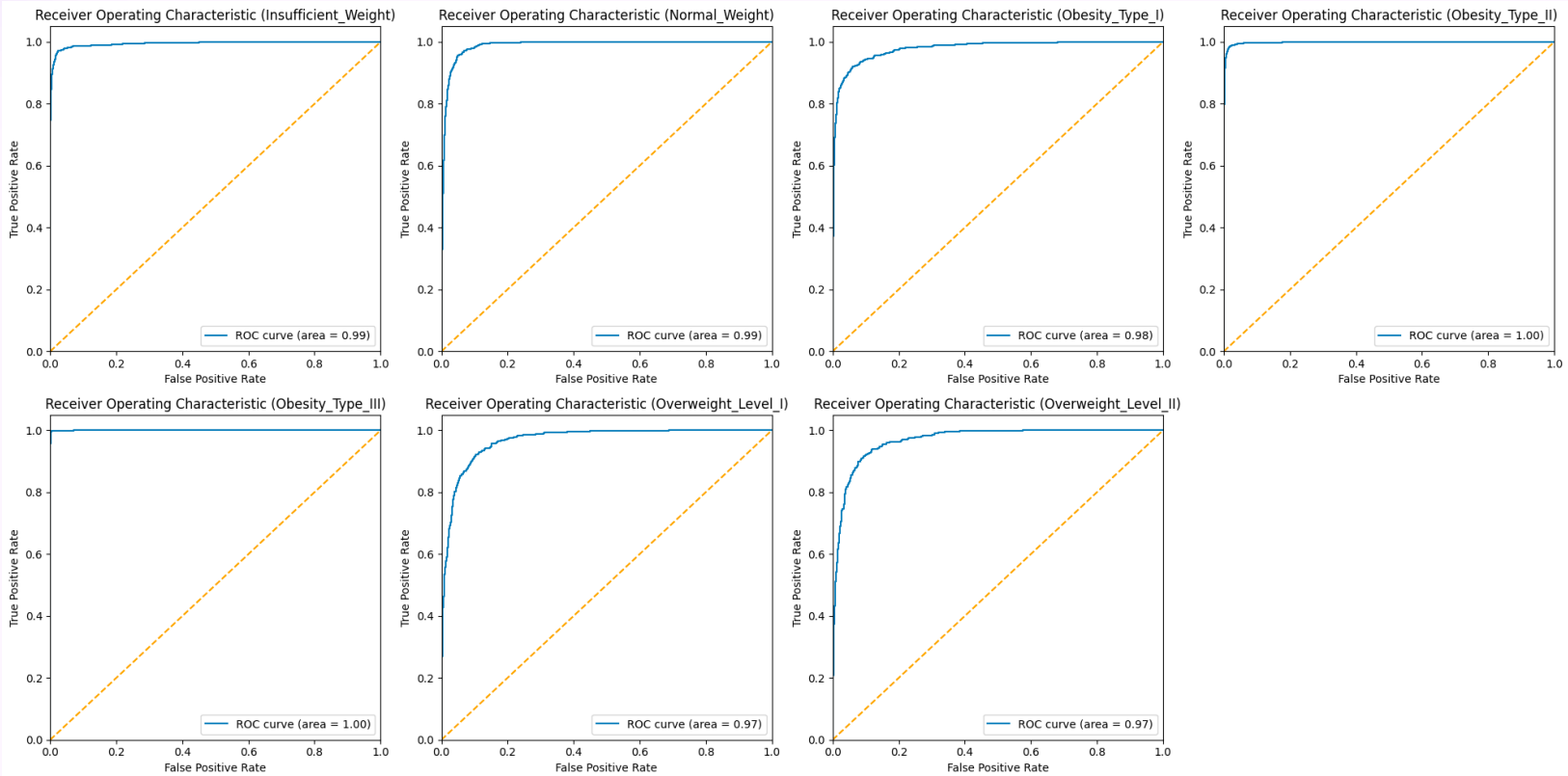
# → Finalny model

| Class | Precission | Recall | F1-Score |
|---|---|---|---|
| Insufficient Weight | 0.94 | 0.92 | 0.93 |
| Normal Weight | 0.85 | 0.88 | 0.87 |
| Obesity Type I | 0.88 | 0.84 | 0.86 |
| Obesity Type II | 0.98 | 0.98 | 0.98 |
| Obesity Type III | 0.99 | 1.00 | 0.99 |
| Overweight Level I | 0.77 | 0.75 | 0.76 |
| Overweight Level II | 0.76 | 0.77 | 0.76 |
| **Accuracy** | | 0.89 | |
| **Macro avg** | 0.88 | 0.88 | 0.88 |
| **Weightrd avg** | 0.89 | 0.89 | 0.89 |

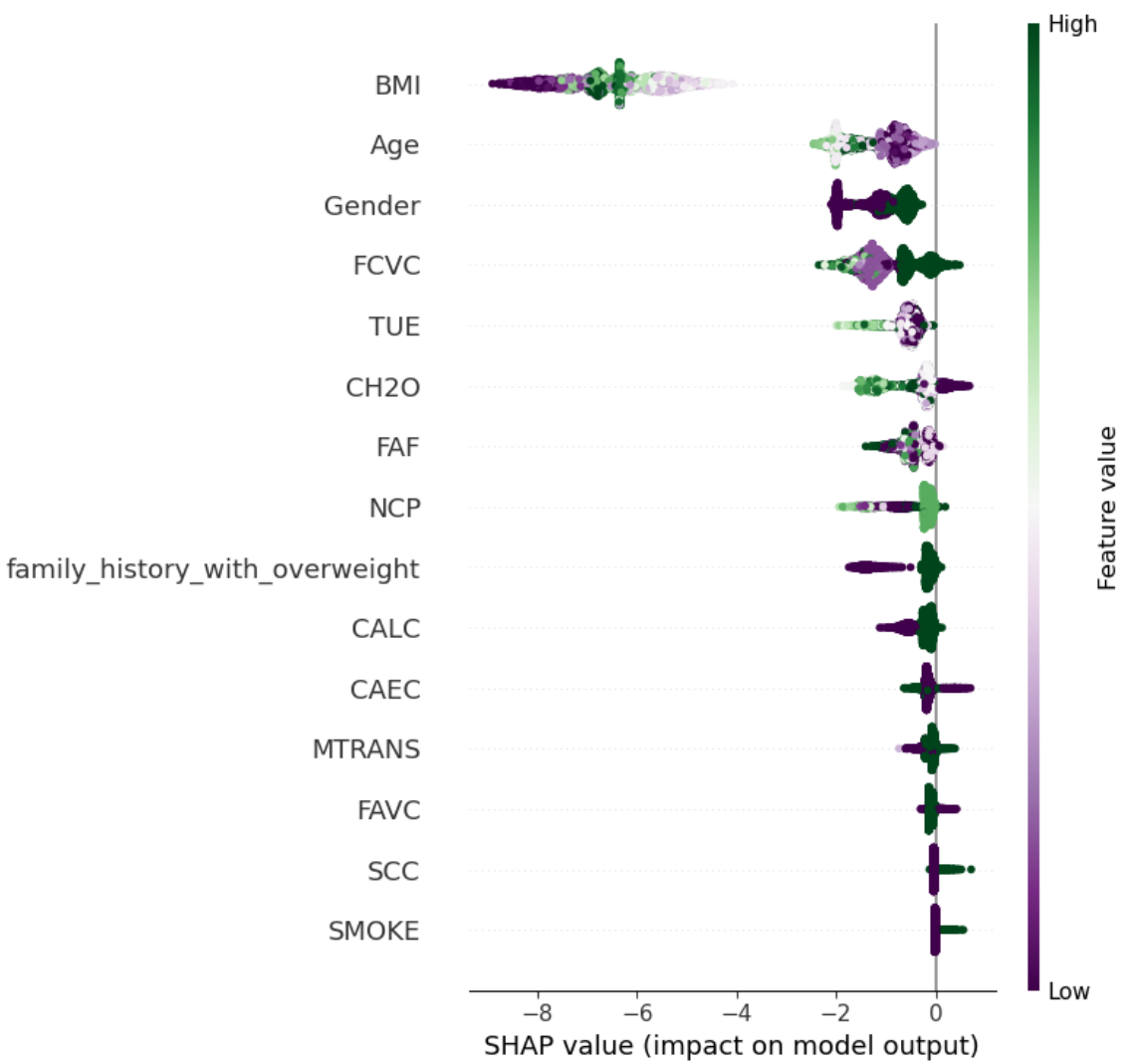
# Finalny model



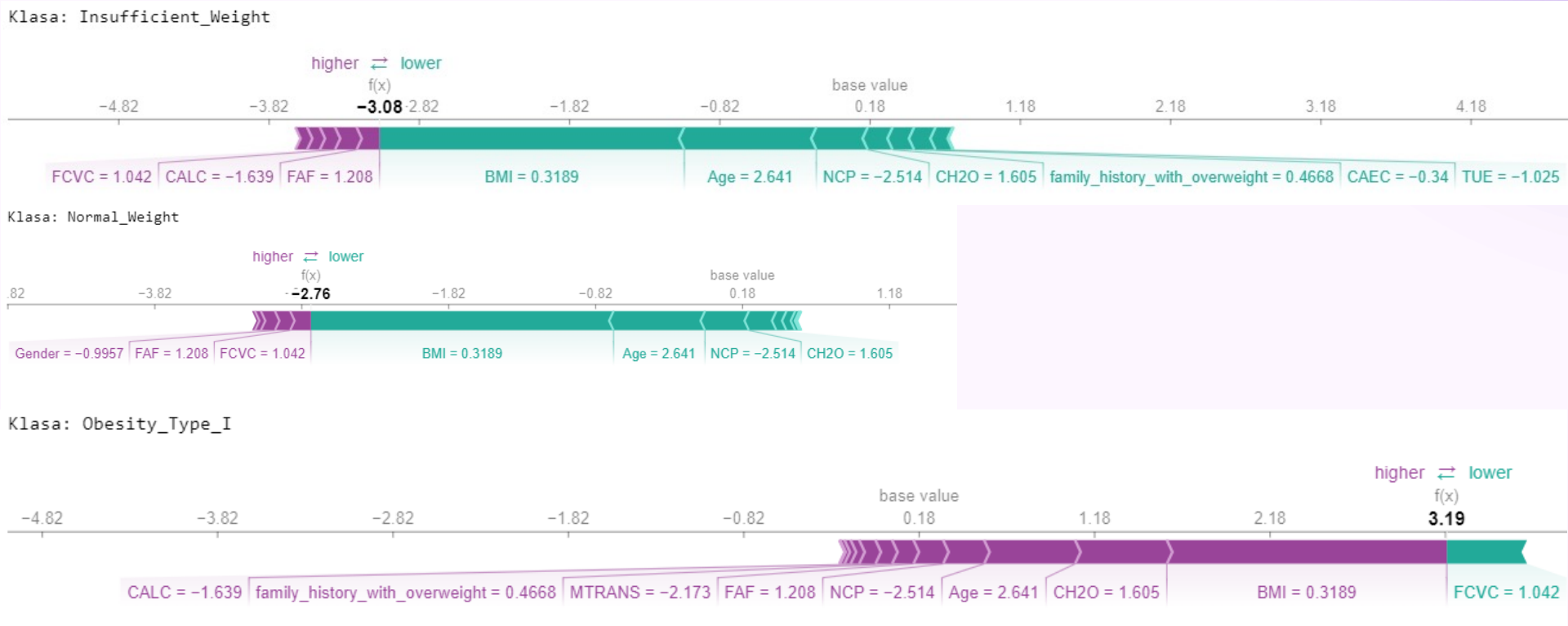
# ROC Curves


Receiver Operating Characteristic (Insufficient_Weight)
ROC curve (area = 0.99)
Receiver Operating Characteristic (Normal_Weight)
ROC curve (area = 0.99)
Receiver Operating Characteristic (Obesity_Type_I)
ROC curve (area = 0.98)
Receiver Operating Characteristic (Obesity_Type_II)
ROC curve (area = 1.00)
Receiver Operating Characteristic (Obesity_Type_III)
ROC curve (area = 1.00)
Receiver Operating Characteristic (Overweight_Level_I)
ROC curve (area = 0.97)
Receiver Operating Characteristic (Overweight_Level_II)
ROC curve (area = 0.97)
True Positive Rate
False Positive Rate

# SHAP

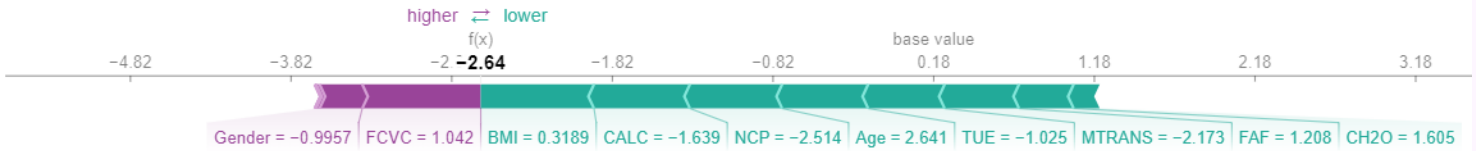

BMI
Age
Gender
FCVC
TUE
CH2O
FAF
NCP
family_history_with_overweight
CALC
CAEC
MTRANS
FAVC
SCC
SMOKE
−8
−6
−4
−2
0
SHAP value (impact on model output)
High
Feature value
Low

# Przykład: Pacjent nr 1 ma Obesity Type I

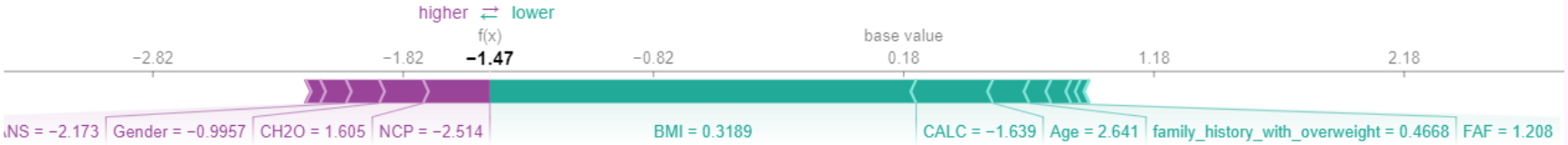


Klasa: Obesity_Type_II

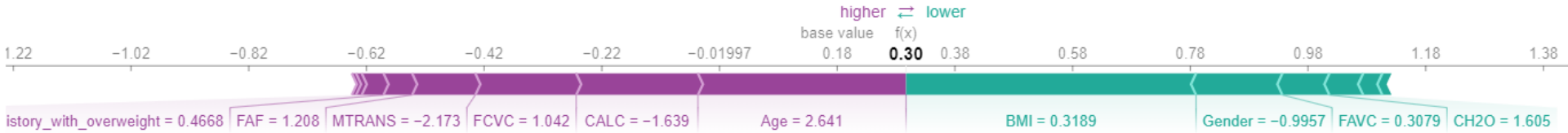

Klasa: Obesity_Type_III




Klasa: Overweight_Level_I




Klasa: Overweight_Level_II



Normal weight

Normal weight

Normal weight

# Testy na ludziach

Urszula Szczęsna
Gaspar Sekula
Mateusz Deptuch

# Obesity risk prediction