Metabolic Syndrome Prediction

Asma Ben Hmidene

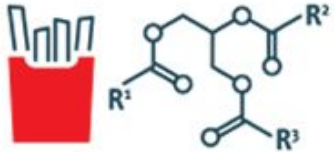
METABOLIC SYNDROME

Cluster of health conditions

VISCERAL OBESITY

HYPERTENSION

INSULIN RESISTANCE



HIGH TRIGLYCERIDES

LOW HDL-CHOLESTEROL

Source: https://www.starhealth.in/blog/metabolic-syndrome

Addressed problem and Objectives:

Metabolic syndrome (MetSyn) is a cluster of risk factors that raise medical costs.

Developing a prediction model that can quickly identify persons at high risk of MetSyn and offer them a treatment plan is crucial.

Early prediction of metabolic syndrome will highly impact the quality of life of patients as it gives them a chance for making a change to the bad habit and preventing a serious illness in the future.

Targeted stakeholders:

Doctors, Hospitals, Pharmaceutical companies, Research Institutes and Ministry of Health (social aid, health assurance, national projects,...)

How to diagnose Metabolic Syndrome?

The following parameters are done to check if a person has metabolic syndrome.

- Checking Blood pressure
- Measuring waist circumference
- Fasting blood glucose level (can also include HbA1C level)
- Measuring the level of Triglycerides and High-density lipids
- Liver function test

If three or more factors are outside the normal range, you have metabolic syndrome.

Data:

Categorical problem with 14 variable features, 1 target and 2401 entries.

Personal data

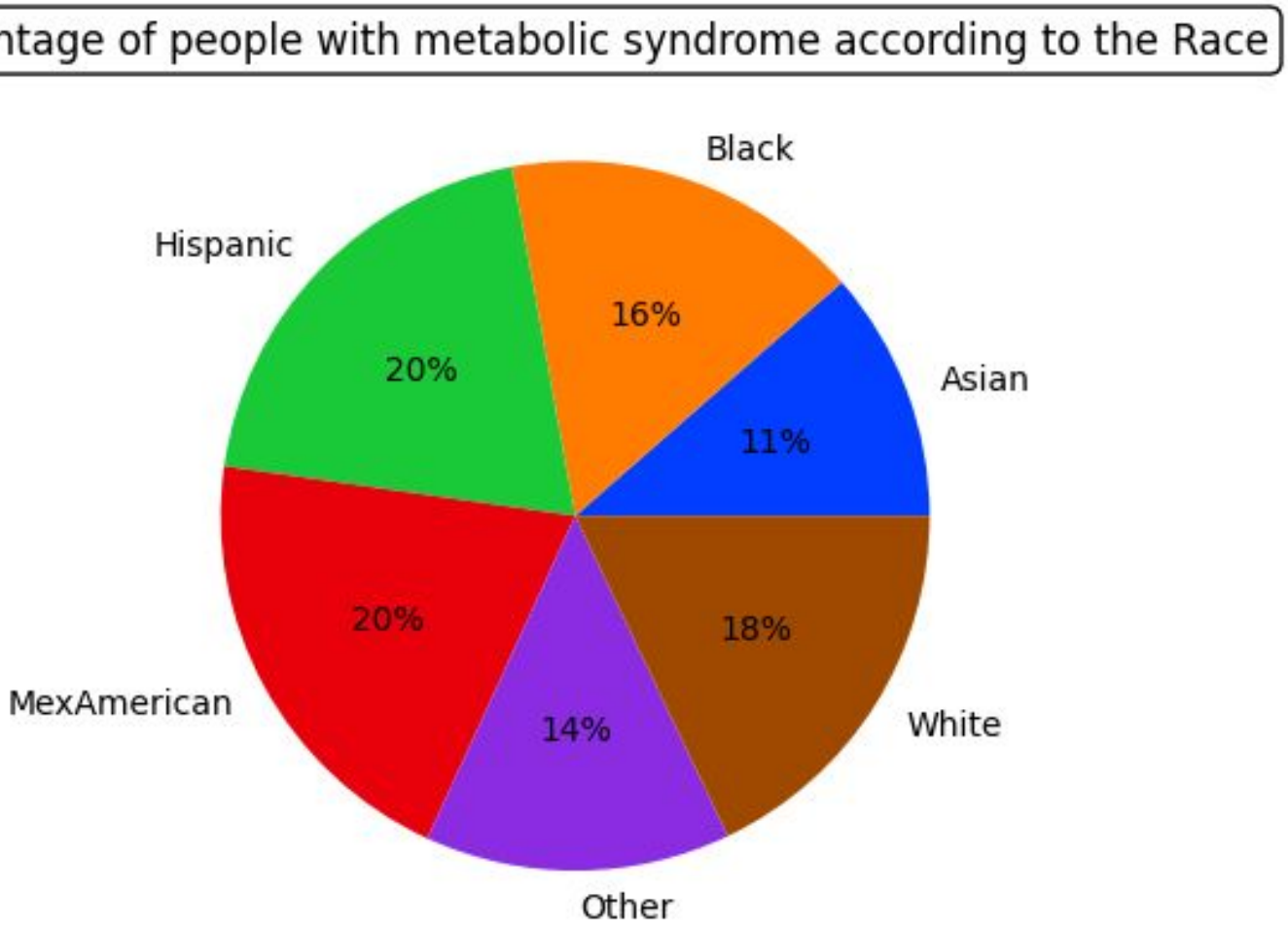
- Age
- Sex
- Marital status
- Race
- Income

Medical records

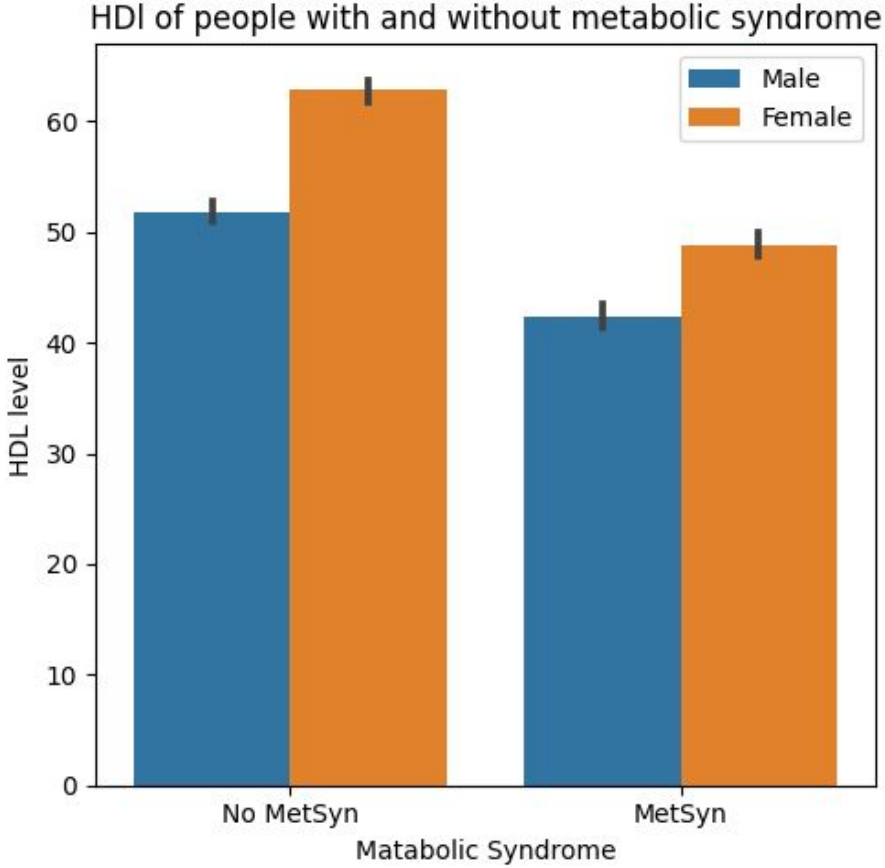
- Waist circumference
- Blood Glucose
- High-density lipids (HDL)
- Triglycerides

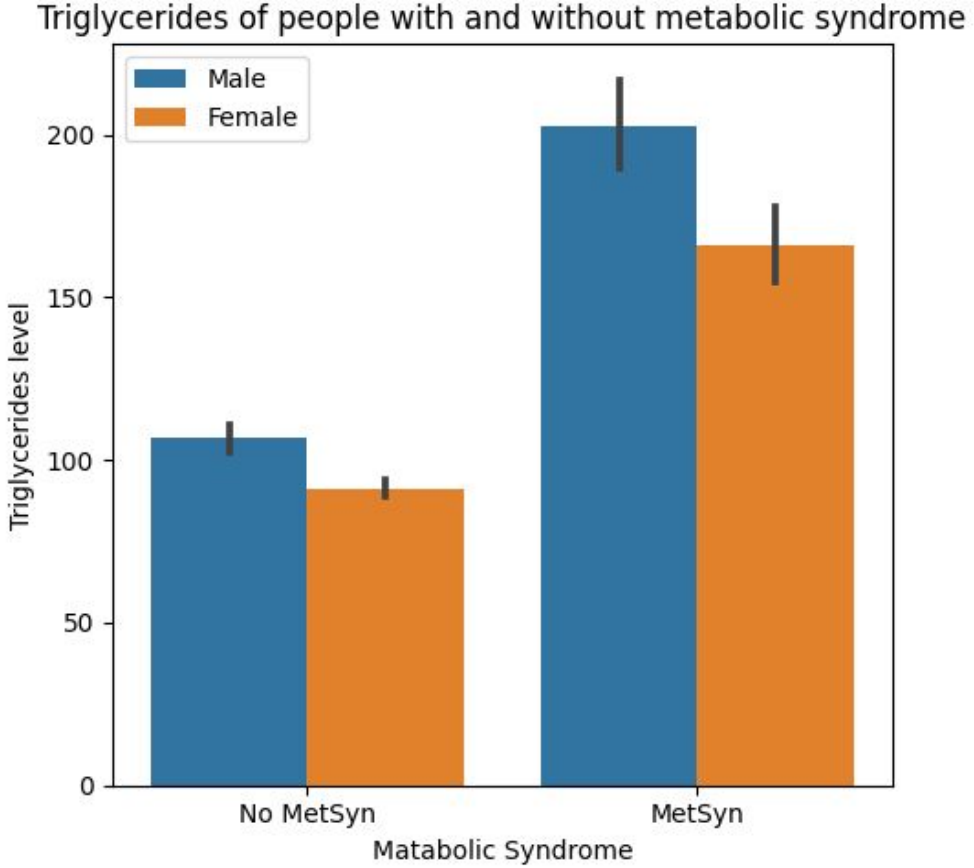
- BMI (Body Mass Index)
- Albuminuria (kidney disease)
- UrAlbCr (Urine albumin-creatinine, kidney disease)
- Uric Acid

Data:


Percentage of people with metabolic syndrome according to the Race
Black
16%
Asian
11%
White
18%
Other
14%
MexAmerican
20%
Hispanic
20%

Data:


HDl of people with and without metabolic syndrome
Male
Female
HDL level
0
10
20
30
40
50
60
No MetSyn
MetSyn
Matabolic Syndrome


Triglycerides of people with and without metabolic syndrome
Male
Female
Triglycerides level
0
50
100
150
200
No MetSyn
MetSyn
Matabolic Syndrome

Prediction Model:

Three models were tested:

- LogisticRegression
- k-nearest neighbors
- Random Forest

All the 3 models were tuned to get a better metrics results.

→ Tuned Random Forest model gave the best results with:

- an accuracy of 88%.
- Number of patients that are healthy but predicted to have MetSyn (false positive): 25/601
- Number of patients that are not healthy but predicted to have no MetSyn (false negative): 50/601

Final recommendations:

- Data is unbalanced:1579 person without MetSyn Vs 822 person with MetSyn.

- It is better to include more data entries.

- Limitation of using PCA: pca in this problem did not ameliorate the accuracy of the selected model. However, we got a slightly more balanced model (less overfitting problem).

Final recommendations:

- Choice of variable features: It is better to include parameters that are more related to the diagnosis of the Metabolic Syndrome. In such case, it will be possible to do, for example, feature engineering :

siMS score = 2*Waist/Height + Gly/5.6 + Tg/1.7 + TAsystolic/130—HDL/1.02 or 1.28 (for male or female subjects, respectively).

siMS risk score = siMS score * age/45 or 50 (for male or female subjects, respectively) * family history of cardio/cerebro-vascular events (event = 1.2, no event = 1).

Thank you for your attention