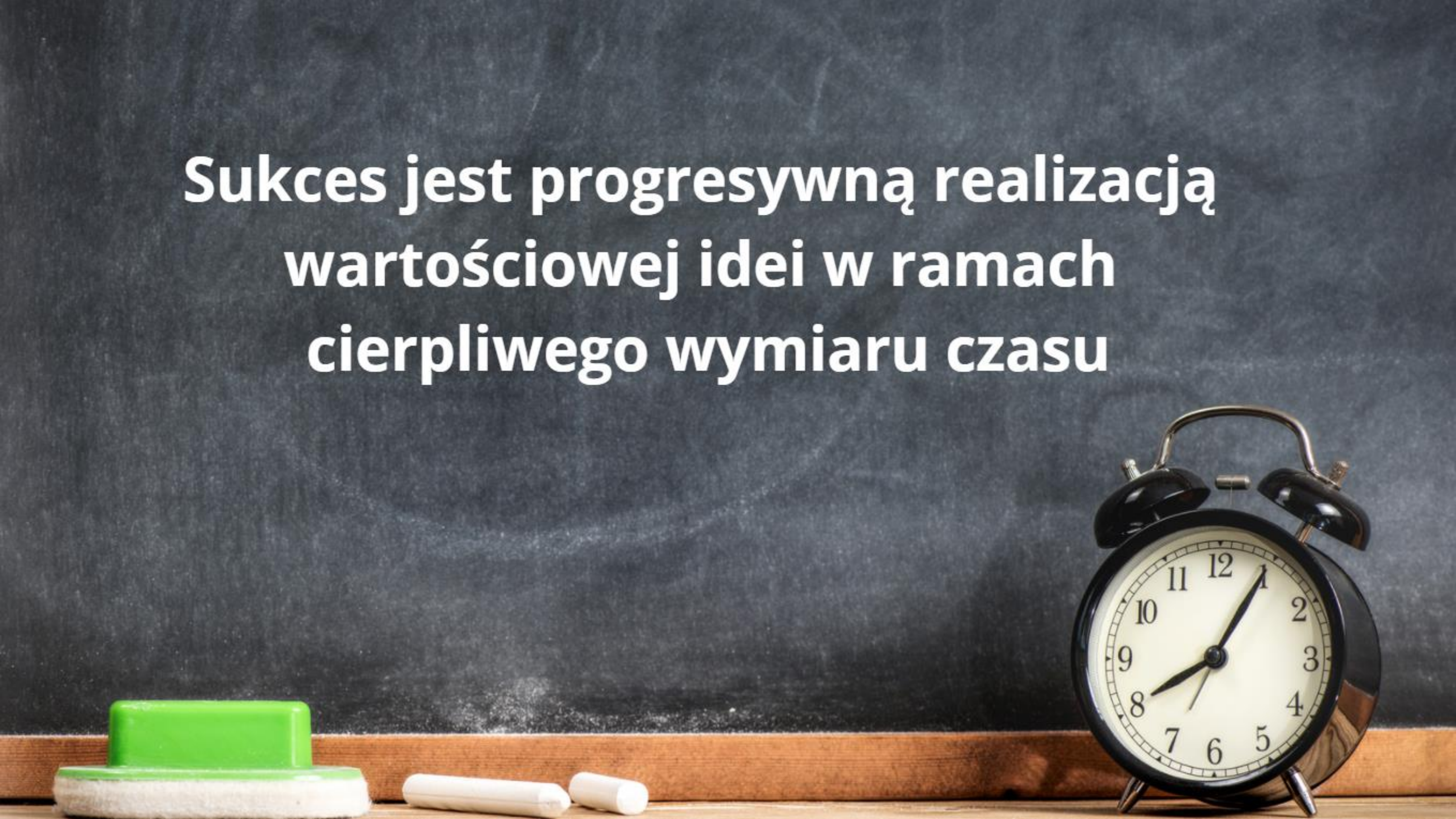
GRADIENT

KOŁO NAUKOWE

Warsztaty Sztucznej Inteligencji

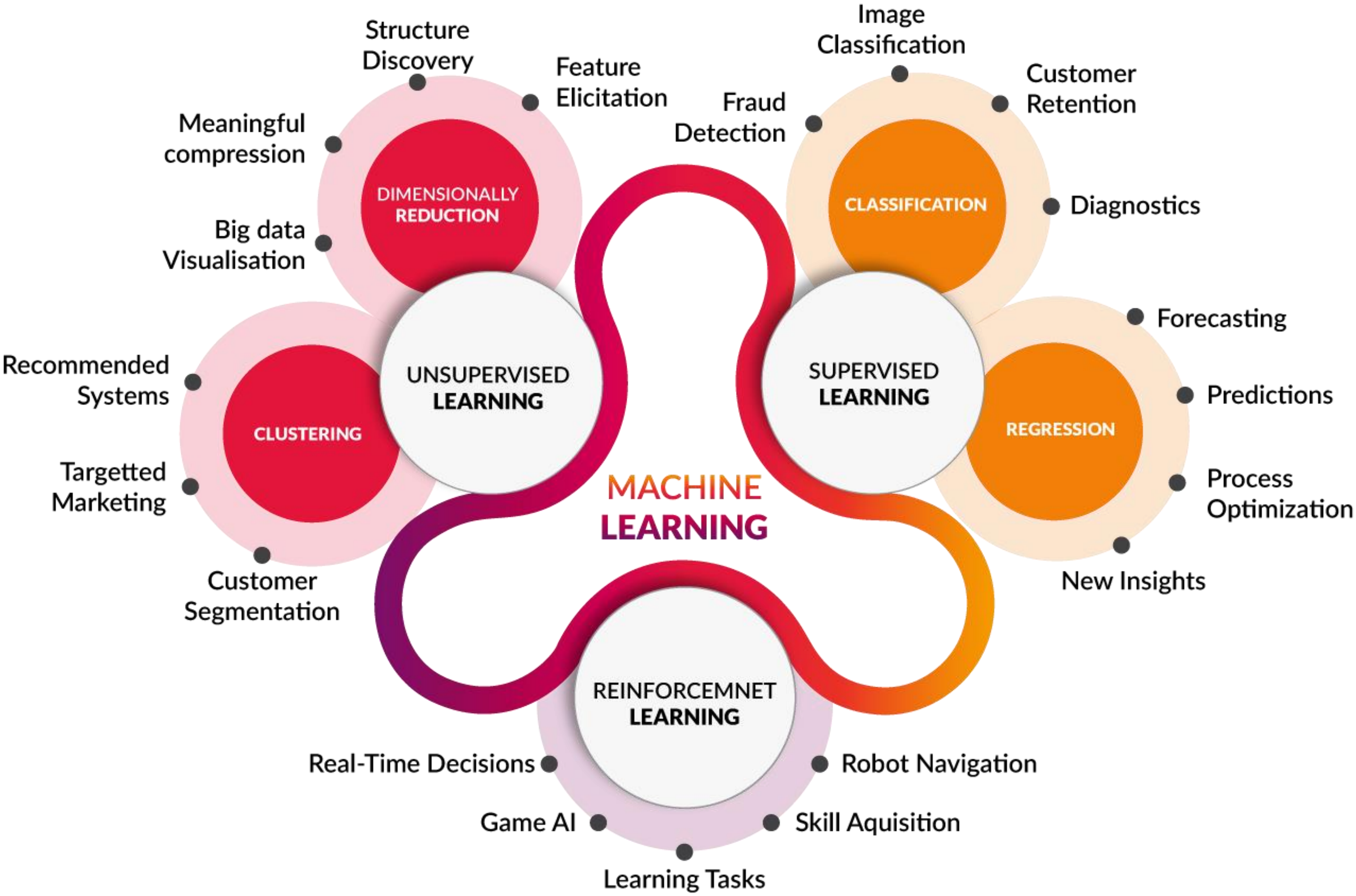
Środa 18:00 EA AUD 2

10.11.2021 - Wprowadzenie do uczenia maszynowego 1
24.11.2021 - Wprowadzenie do uczenia maszynowego 2
01.12.2021 - Wprowadzenie do uczenia głębokiego
08.12.2021 - Wizja komputerowa
15.12.2021 - Sieci rekurencyjne i ich zastosowania
05.01.2021 - Uczenie ze wzmocnieniem
12.01.2021 - Implementacja uczenia maszynowego
19.01.2021 - Modele ML w środowisku produkcyjnym

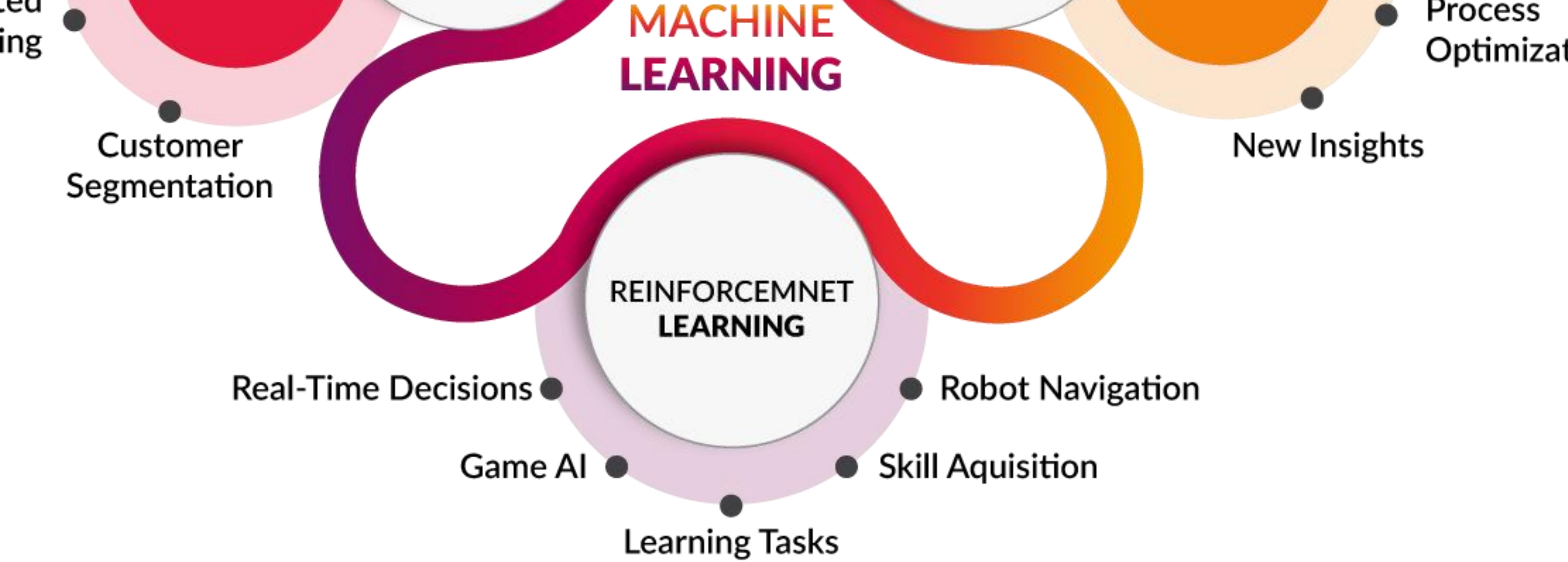

Sukces jest progresywną realizacją
wartościowej idei w ramach
cierpliwego wymiaru czasu

Agenda

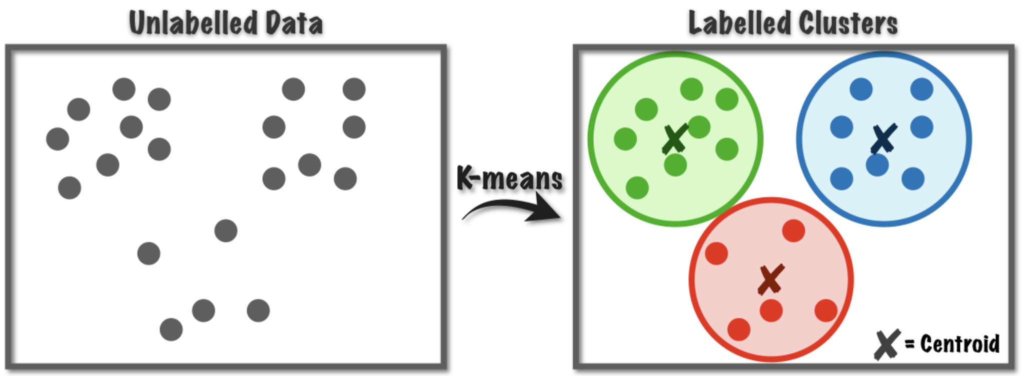
1. Rodzaje uczenia maszynowego (5min)
2. Zastosowanie regresji wielorakiej (10min)
3. Budowa drzewa klasyfikującego (20 min)
4. Pytania rekrutacyjne + QUIZ (15min)

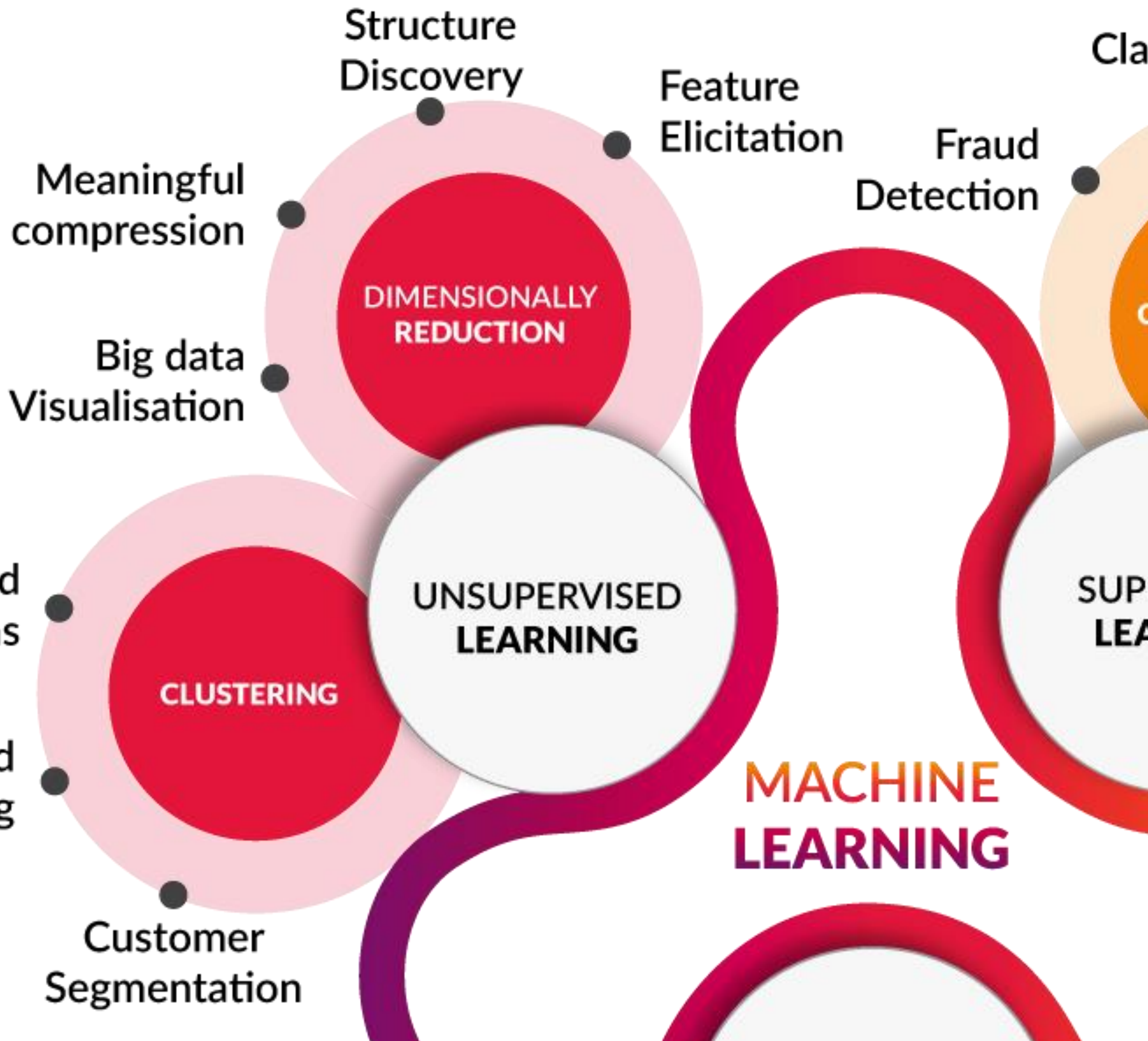


MACHINE LEARNING
UNSUPERVISED LEARNING
DIMENSIONALLY REDUCTION
Structure Discovery
Feature Elicitation
Meaningful compression
Big data Visualisation
CLUSTERING
Recommended Systems
Targetted Marketing
Customer Segmentation
SUPERVISED LEARNING
CLASSIFICATION
Image Classification
Customer Retention
Fraud Detection
Diagnostics
REGRESSION
Forecasting
Predictions
Process Optimization
New Insights
REINFORCEMNET LEARNING
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks


MACHINE
LEARNING
Customer
Segmentation
Process
New Insights
REINFORCEMNET
LEARNING
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks


FlappyBird


Unlabelled Data
K-means
Labelled Clusters
= Centroid


Structure Discovery
Feature Elicitation
Fraud Detection
Meaningful compression
DIMENSIONALLY REDUCTION
Big data Visualisation
UNSUPERVISED LEARNING
Recommended Systems
CLUSTERING
Targetted Marketing
Customer Segmentation
MACHINE LEARNING

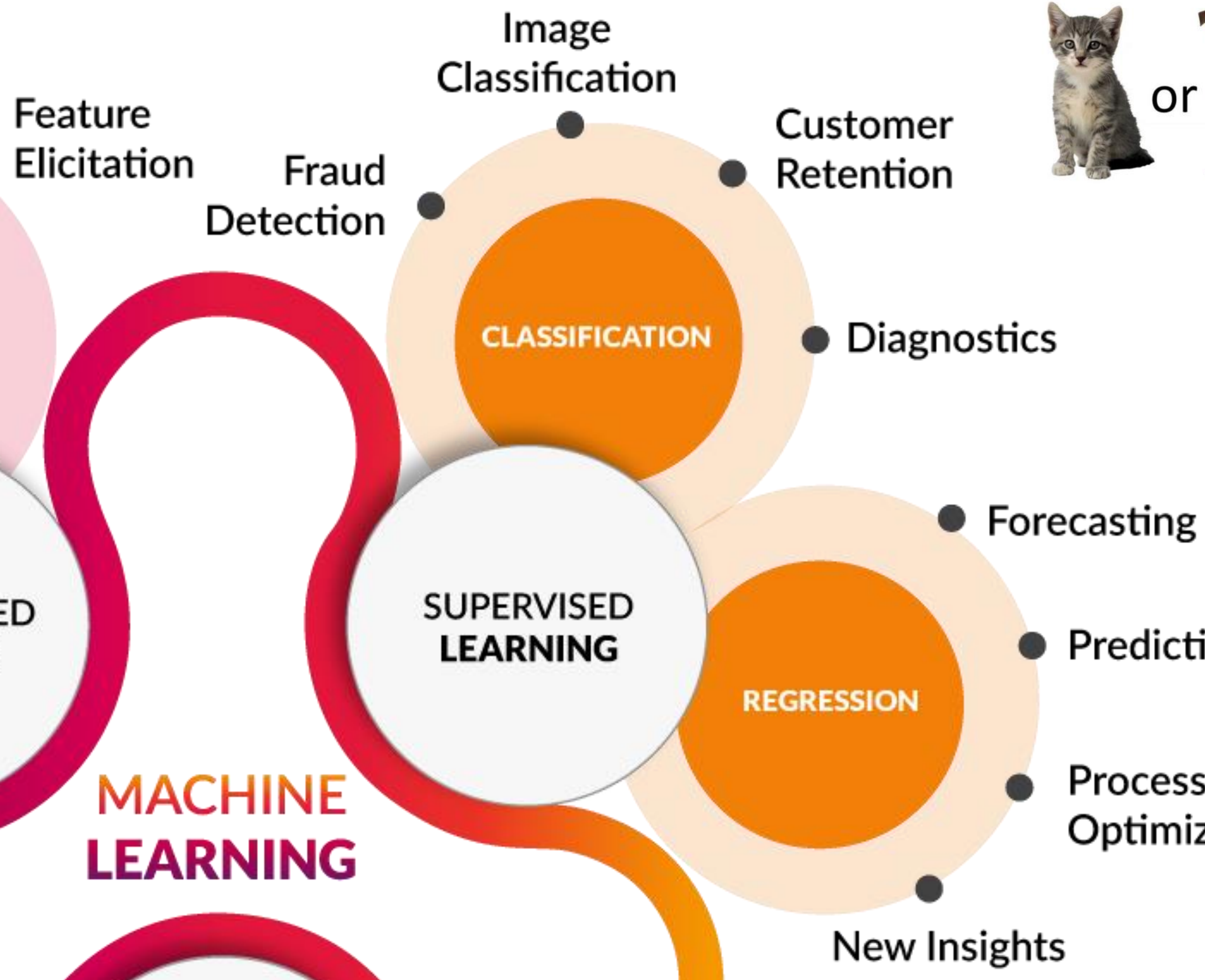
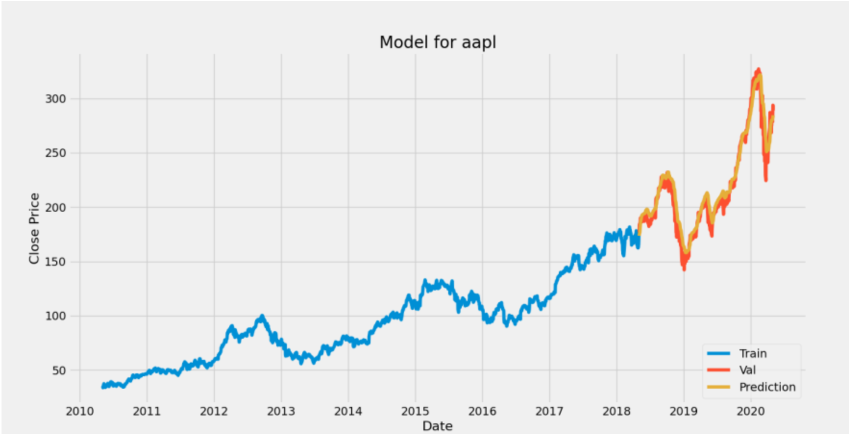
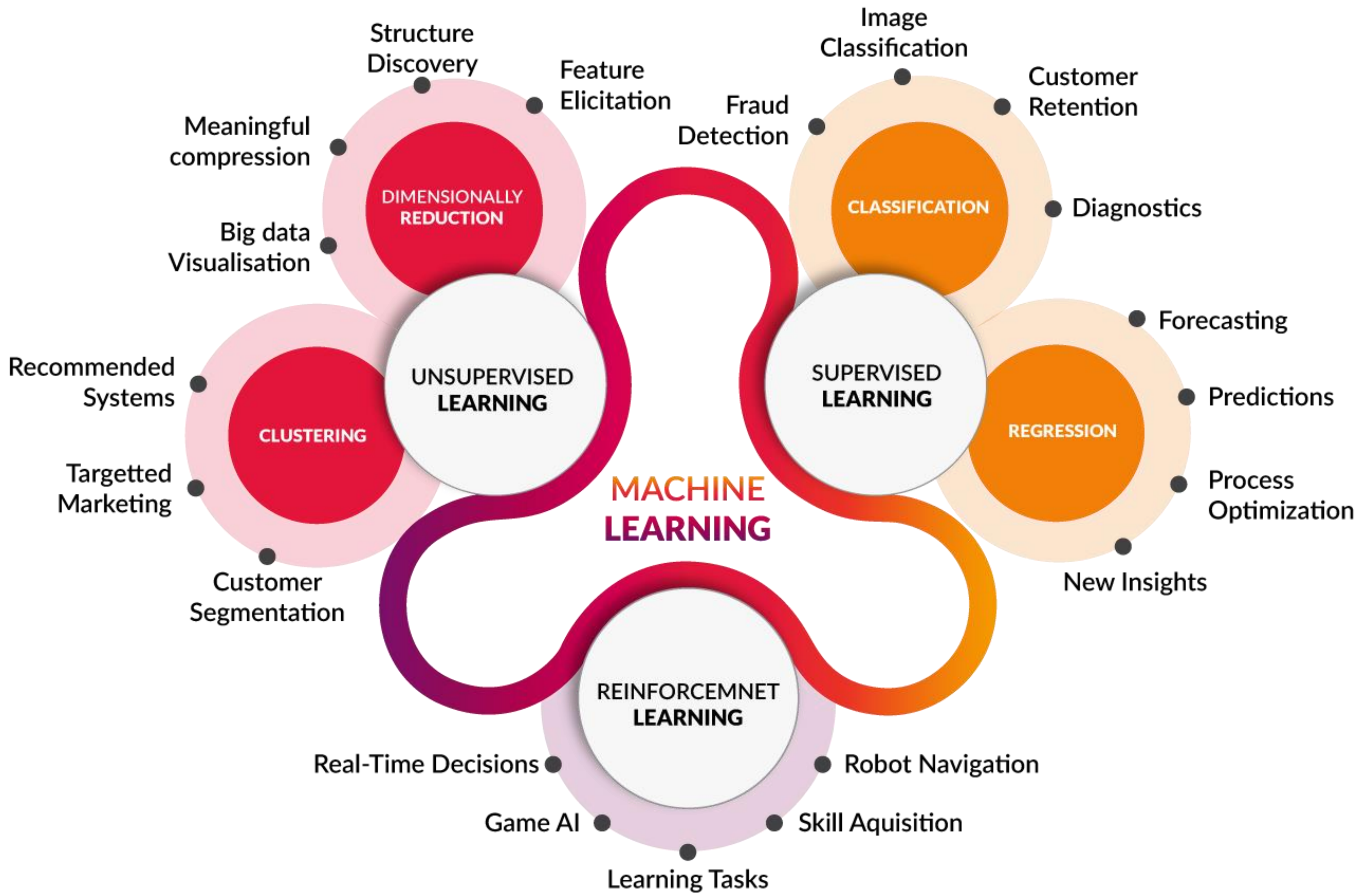


Image
Classification
Feature
Elicitation
Fraud
Detection
Customer
Retention
CLASSIFICATION
Diagnostics
SUPERVISED
LEARNING
Forecasting
REGRESSION
Predictions
Process
Optimization
New Insights
ED
MACHINE
LEARNING


or


Model for aapl
Close Price
300
250
200
150
100
50
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date
Train
Val
Prediction

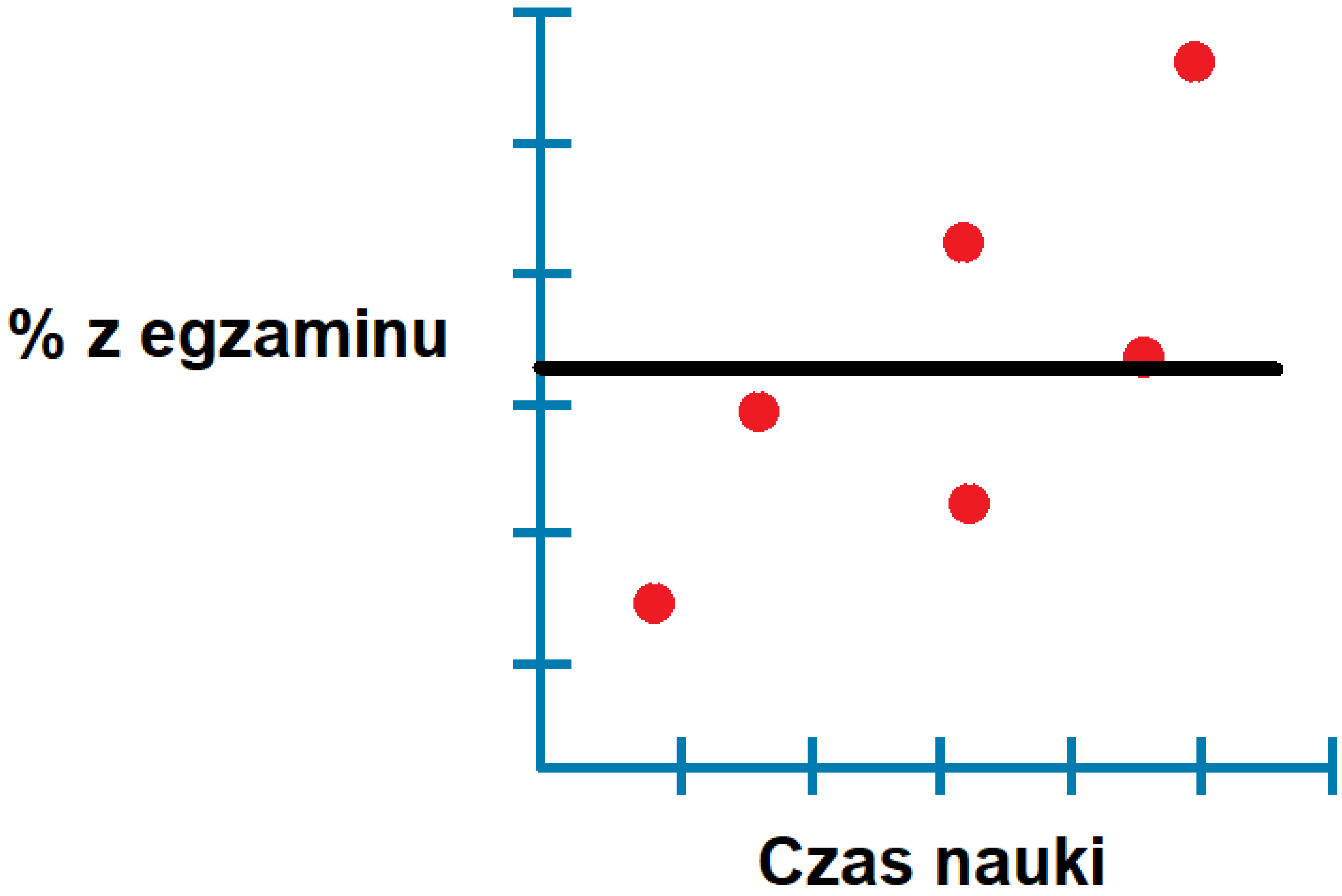
kNN


MACHINE LEARNING
UNSUPERVISED LEARNING
DIMENSIONALLY REDUCTION
Structure Discovery
Feature Elicitation
Meaningful compression
Big data Visualisation
CLUSTERING
Recommended Systems
Targetted Marketing
Customer Segmentation
SUPERVISED LEARNING
CLASSIFICATION
Image Classification
Customer Retention
Fraud Detection
Diagnostics
REGRESSION
Forecasting
Predictions
Process Optimization
New Insights
REINFORCEMNET LEARNING
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

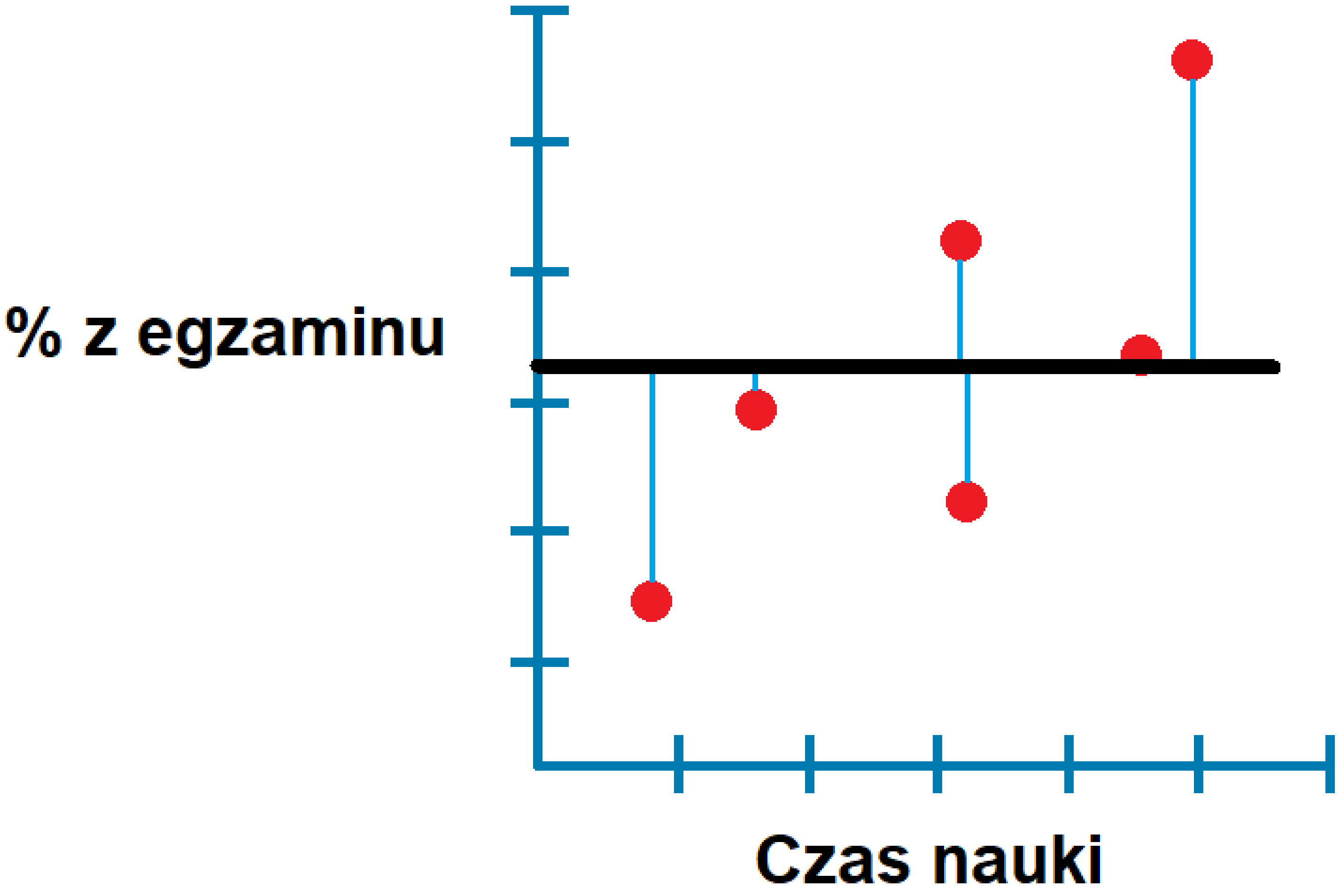
Z poprzedniego wykładu

- Hipoteza
- Funkcja kosztu
- Gradient

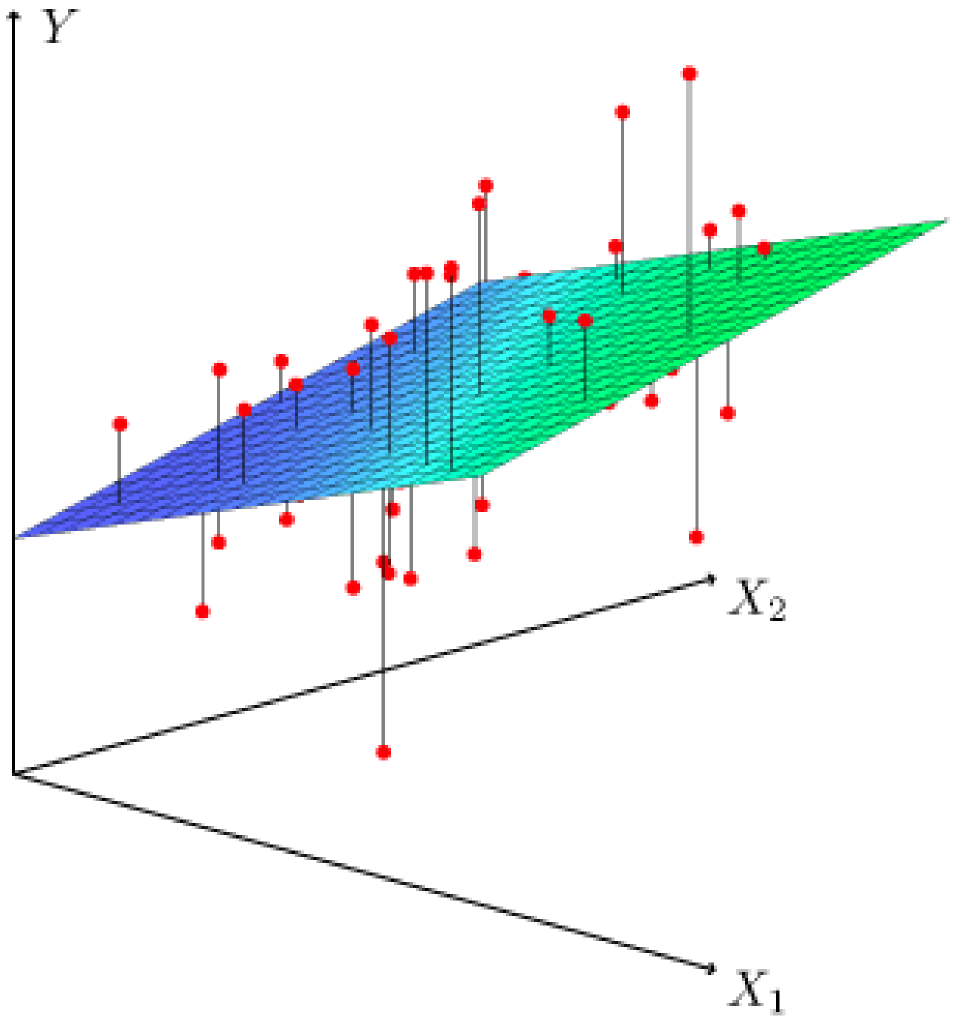
Regresja liniowa



Regresja liniowa



Regresja wieloraka

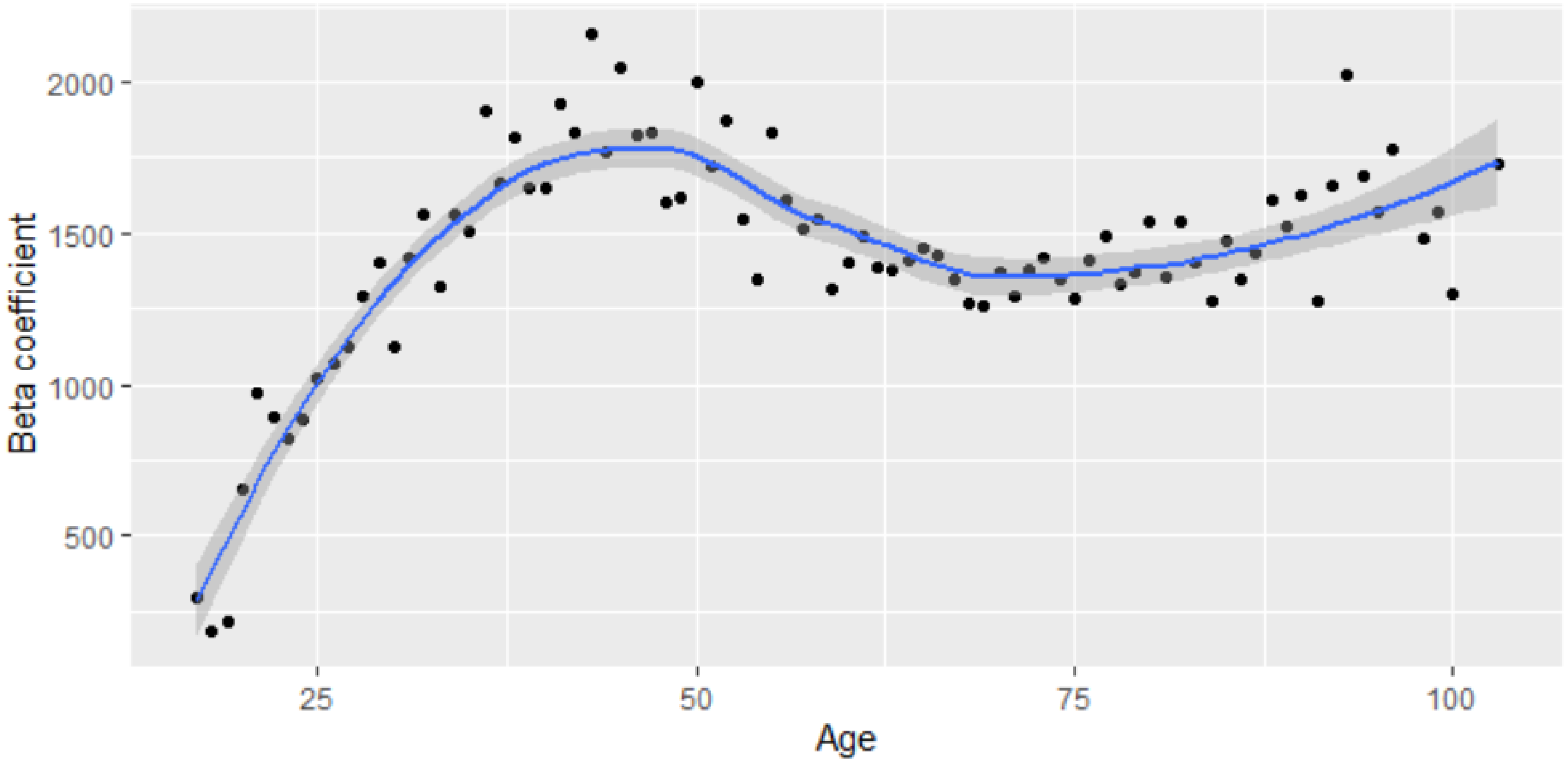


$$Y = aX_1 + bX_2 + c$$

```r
lm(income~sex, data=extract, weights=waga)
```

Coefficients:	ΔR^2	Estimate	Std. Error	t value	P-Value
Step 1	0.04				<0.001
(Intercept)		1749.59	16.15	108.32	<0.001
Gender		580.51	23.26	24.96	<0.001

Wiek jako czynnik



```r
lm(income~sex+ageabove16upto40+age40flag+ageafter40+age60plusflag+
  +height_ms+height_fs, data=extract, weights=waga)
```

Coefficients:	ΔR^2	Estimate	Std. Error	t value	P-Value
Step 2 with age and height	0.09				<0.001
(Intercept)		758.10	55.88	13.57	<0.001
Gender		533.56	22.76	23.44	<0.001
Height (men)		28.28	2.39	11.81	<0.001
Height (women)		12.73	2.62	4.87	<0.001
Age between 16 and 40		70.74	3.35	21.14	<0.001
Age 40-60 (Flag)		1546.49	67.24	23.00	<0.001
Age between 40 and 60		-36.06	3.24	-11.81	<0.001
Age after 60 (Flag)		755.10	59.03	12.79	<0.001

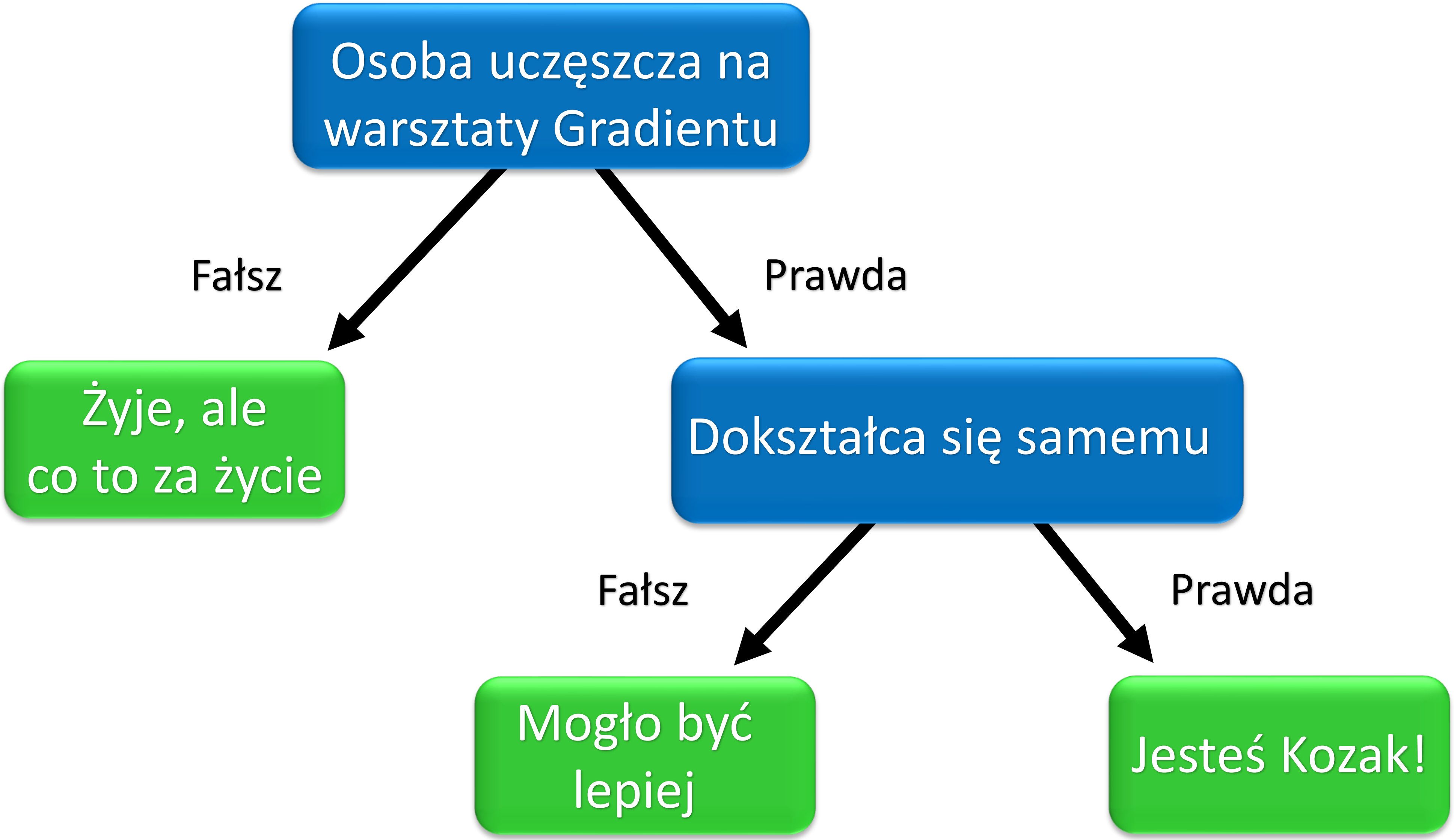
```
lm(income~sex+ageabove16upto40+age40flag+ageafter40+age60plusflag
   + height_ms+ height_fs+relevel(education, ref = " without")
   + relevel(class, ref = " village"), data=extract, weights=waga)
```

Step 3 with education and class	0.25				<0.001
(Intercept)		65.27	118.95	0.55	.584
Gender		671.34	21.23	31.62	<0.001
Age between 16 and 40		52.70	3.28	16.07	<0.001
Age 41-60 (Flag)		1447.85	65.53	22.10	<0.001
Age between 41 and 60		-28.91	65.53	-9.76	<0.001
Age after 60 (Flag)		825.23	59.25	13.93	<0.001
Height (men)		17.33	2.19	7.90	<0.001
Height (woman)		0.05	2.40	0.019	0.985
PhD		2739.23	157.07	17.44	<0.001
Master		1430.77	109.60	13.05	<0.001
Bachelor		917.71	115.82	7.92	<0.001
Postsecondary		594.42	120.06	4.95	<0.001
Vocational		504.80	108.11	4.67	<0.001
Secondary		498.22	111.79	4.46	<0.001
Basic vocational		128.28	107.45	1.19	.233
Middle school		311.28	144.44	2.16	.031
Primary		-28.59	108.25	-0.26	.792
City bigger than 500 thou.		671.10	36.27	18.50	<0.001
City between 200 and 500 thou.		316.06	37.80	8.36	<0.001
City between 100 and 200 thou.		257.96	41.44	6.23	<0.001
City between 20 and 100 thou.		149.93	28.94	5.18	<0.001
City less than 20 thou.		206.68	34.37	0.01	<0.001

Age between 16 and 40		52.70	3.28	16.07	<0.001
Age 41-60 (Flag)		1447.85	65.53	22.10	<0.001
Age between 41 and 60		-28.91	65.53	-9.76	<0.001
Age after 60 (Flag)		825.23	59.25	13.93	<0.001
Height (men)		17.33	2.19	7.90	<0.001
Height (woman)		0.05	2.40	0.019	0.985
PhD		2739.23	157.07	17.44	<0.001
Master		1430.77	109.60	13.05	<0.001
Bachelor		917.71	115.82	7.92	<0.001
Postsecondary		594.42	120.06	4.95	<0.001
Vocational		504.80	108.11	4.67	<0.001
Secondary		498.22	111.79	4.46	<0.001
Basic vocational		128.28	107.45	1.19	.233
Middle school		311.28	144.44	2.16	.031
Primary		-28.59	108.25	-0.26	.792
City bigger than 500 thou.		671.10	36.27	18.50	<0.001
City between 200 and 500 thou.		316.06	37.80	8.36	<0.001
City between 100 and 200 thou.		257.96	41.44	6.23	<0.001
City between 20 and 100 thou.		149.93	28.94	5.18	<0.001
City less than 20 thou.		206.68	34.37	6.01	<0.001

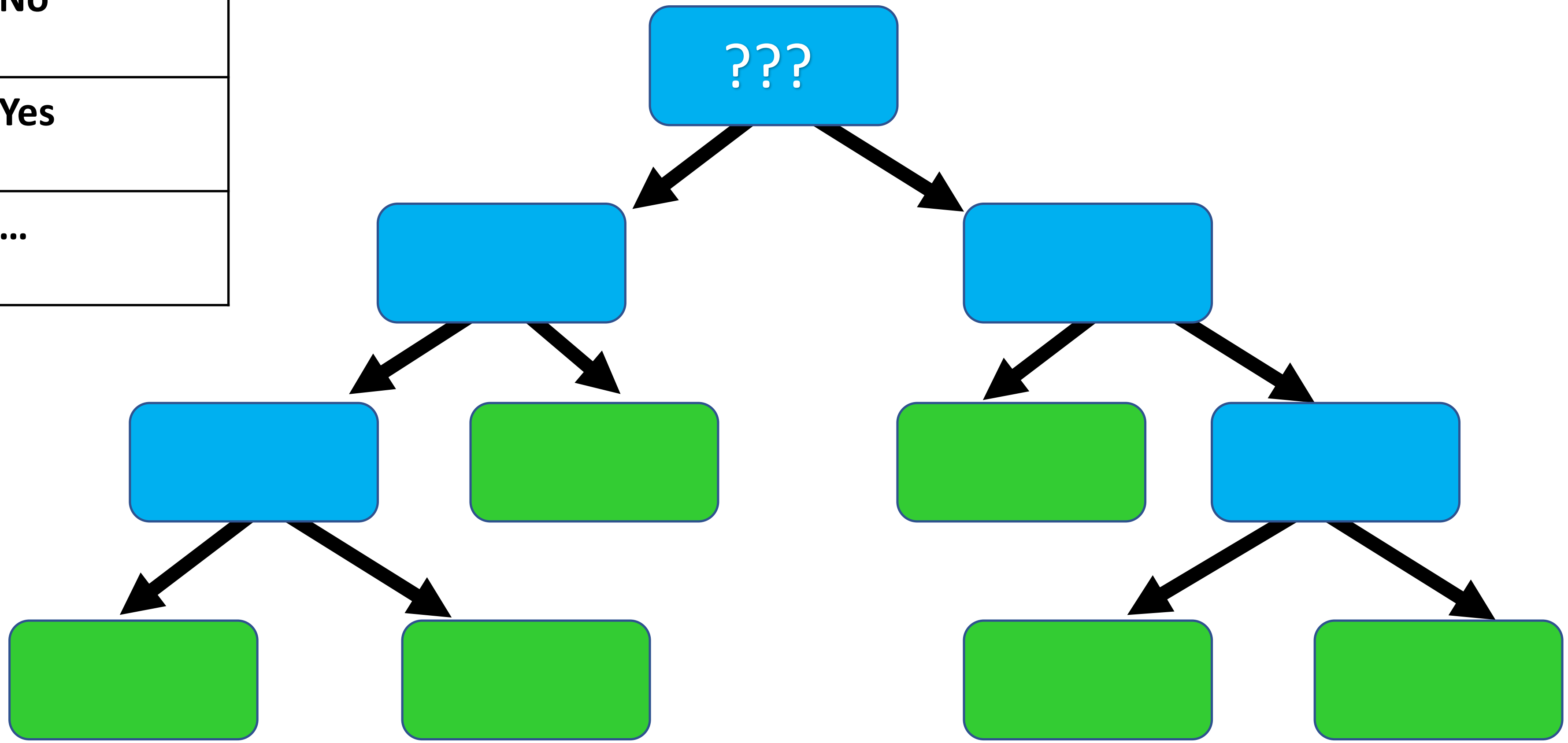
$$Income = 65.27 + 671.34(Gender) + 52.70(Age\ 16\text{-}40) + 1447.85(Age\ 41\text{-}60\ Flag) - 28.91(Age\ 40\text{-}60) + 825.23(Age\ after\ 60\ Flag) + 17.33(Height\ men) + 0.05(Height\ woman) + 2739.23(PhD) + 1430.77(Master) + 917.71(Bachelor) + 594.42(Postsecondary) + 504.80(Vocational) + 498.22(Secondary) + 128.28(Basic\ vocational) + 311.28(Middle\ school) - 28.59(Primary) + 671.10(City\ more\ than\ 500\ thou.) + 316.06(City\ 200 - 500\ thou.) + 257.96(City\ 100\text{-}200\ thou.) + 149.93(City\ 20\text{-}100\ thou.) + 206.68(City\ less\ than\ 20\ thou.),$$

Drzewa decyzyjne

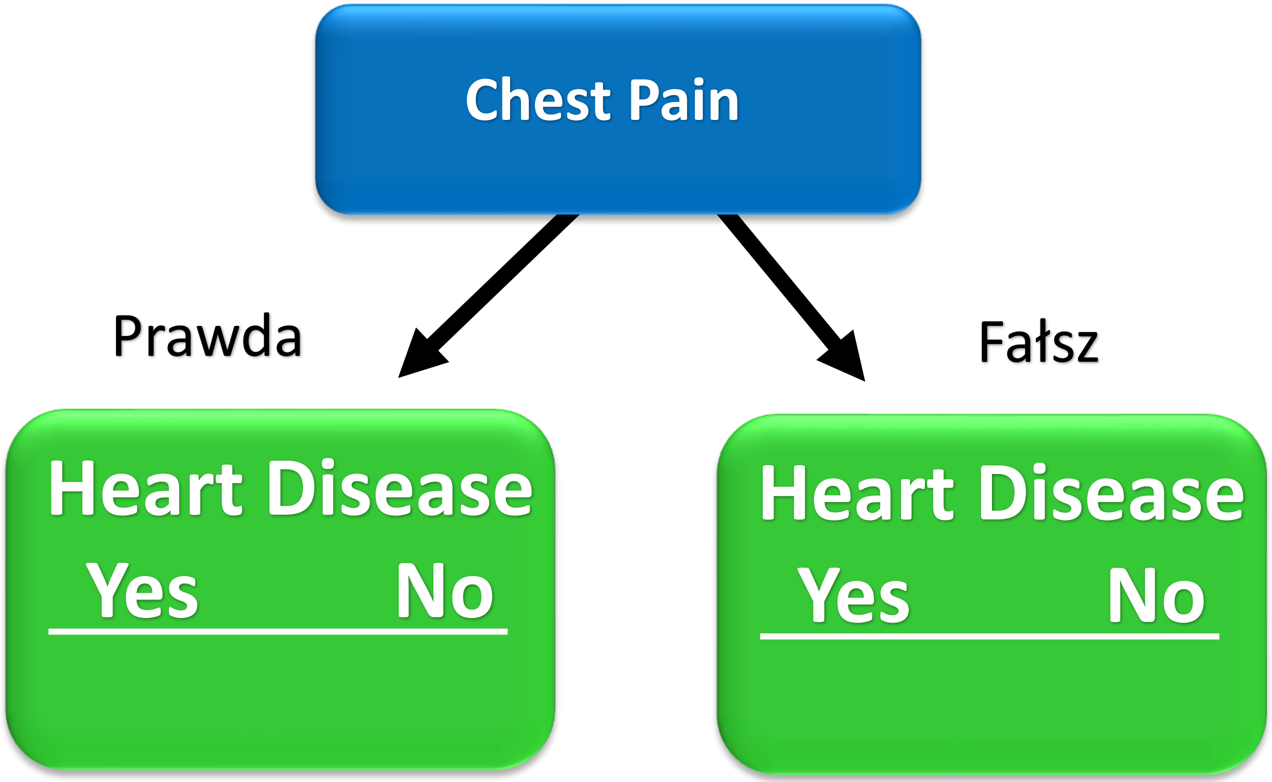

Osoba uczęszcza na warsztaty Gradientu
Fałsz
Prawda
Żyje, ale co to za życie
Dokształca się samemu
Fałsz
Prawda
Mogło być lepiej
Jesteś Kozak!

Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…

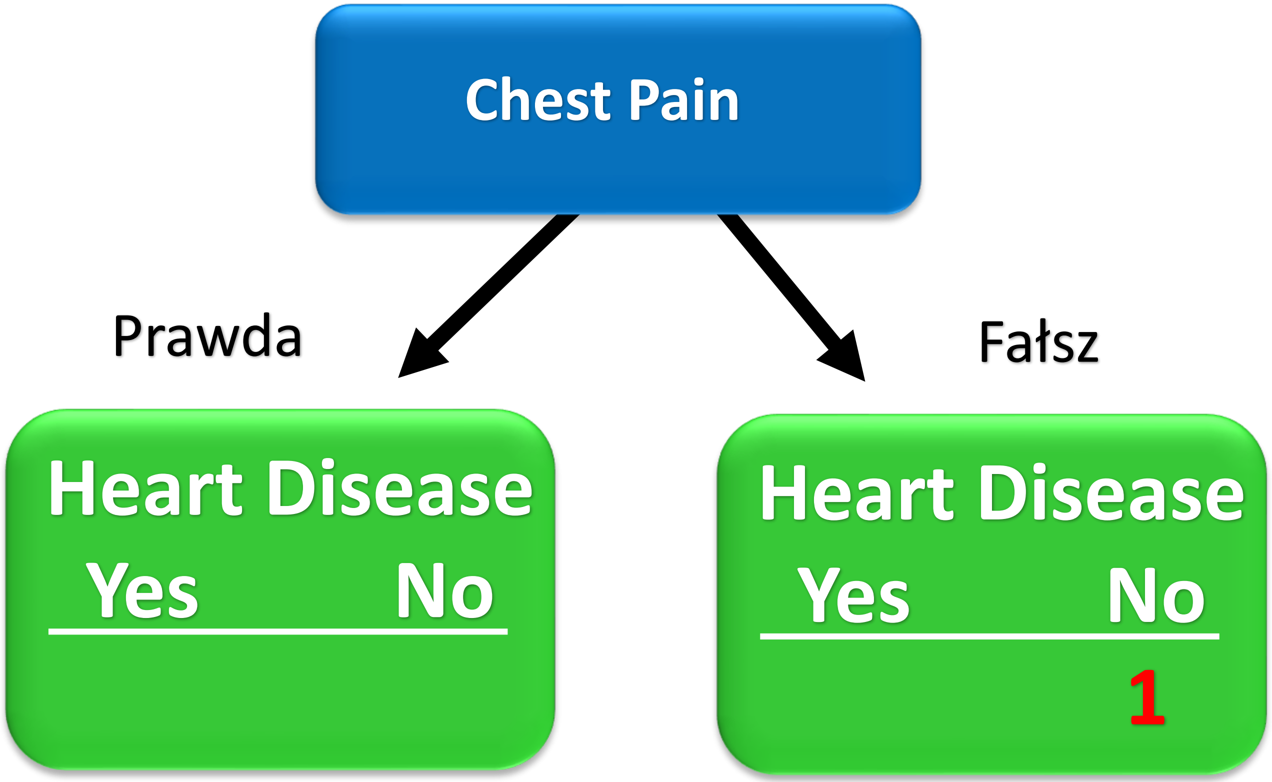
Co powinno być „root node”?



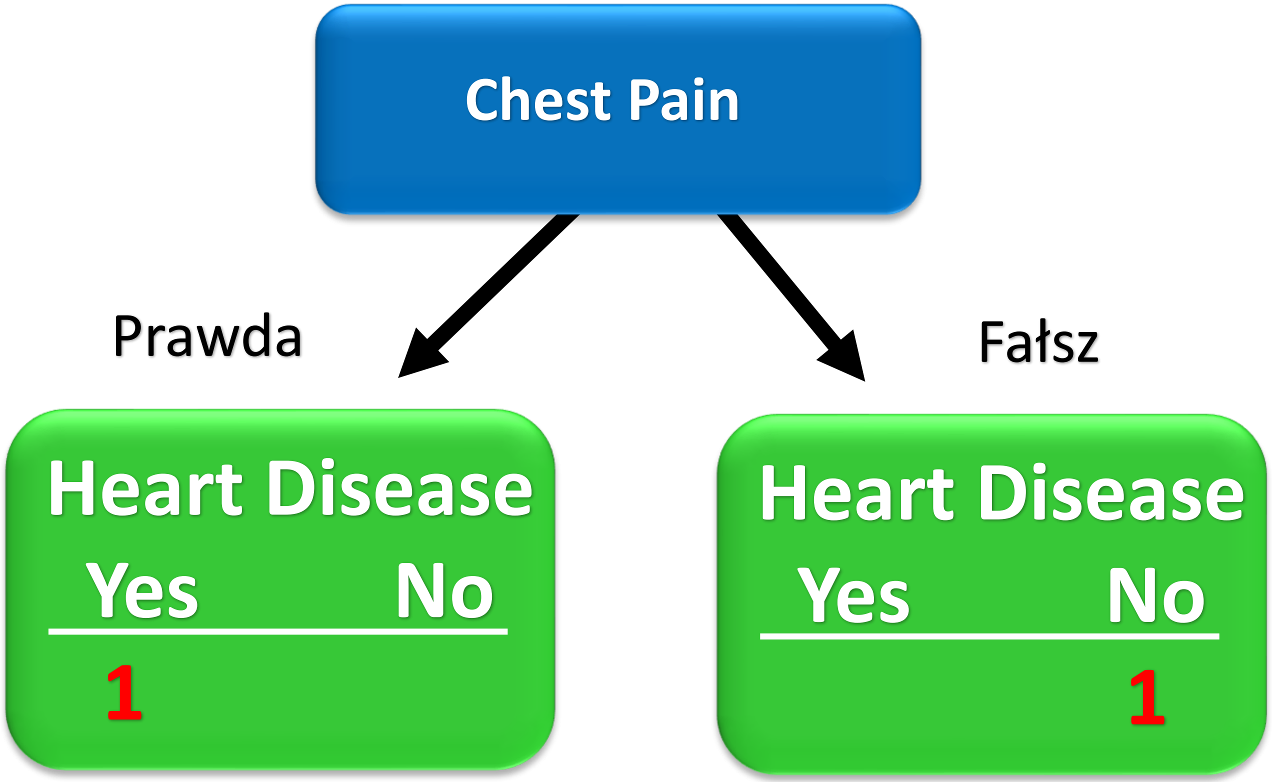
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



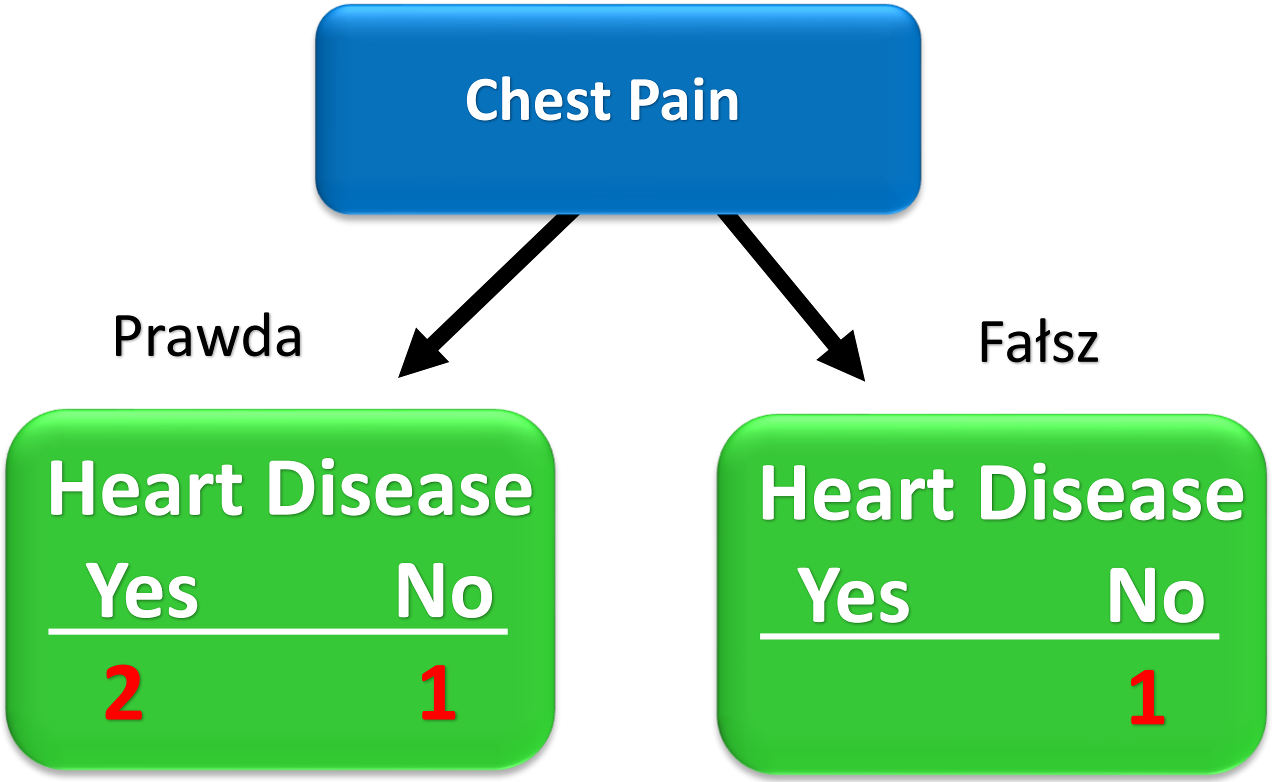
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



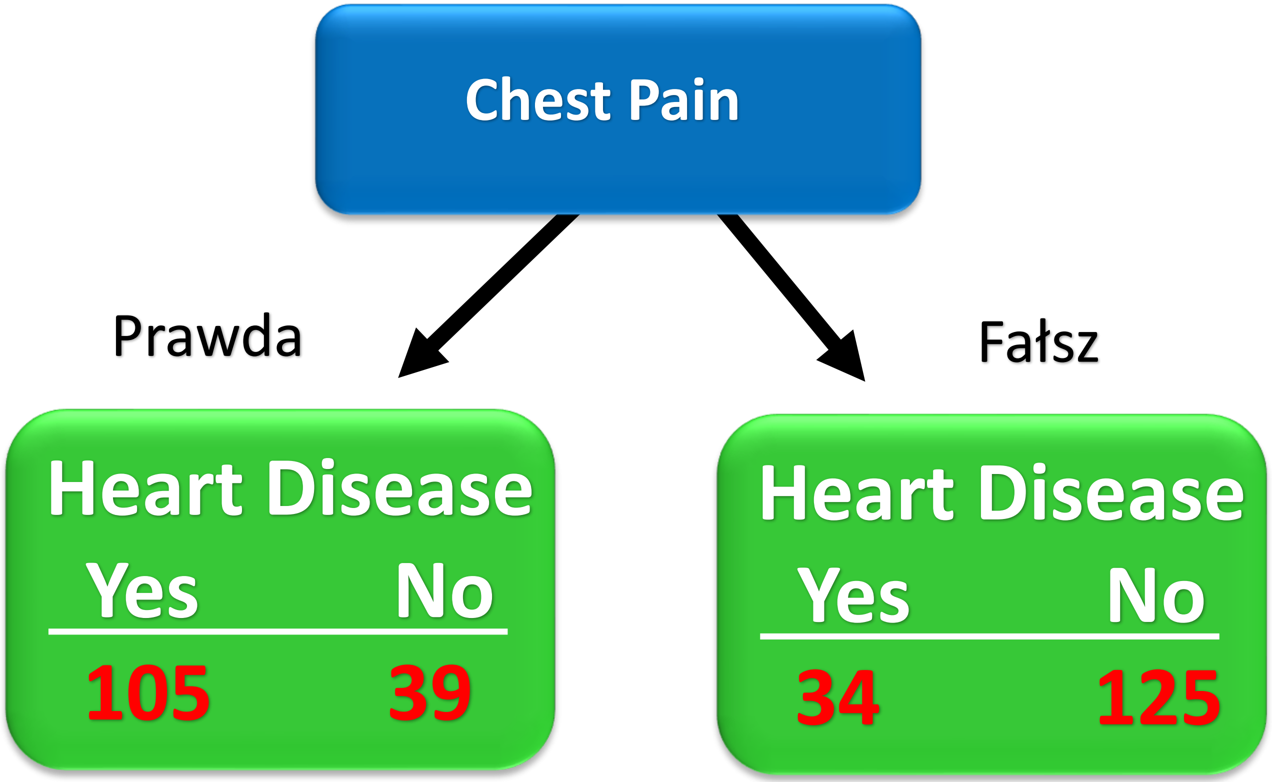
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…

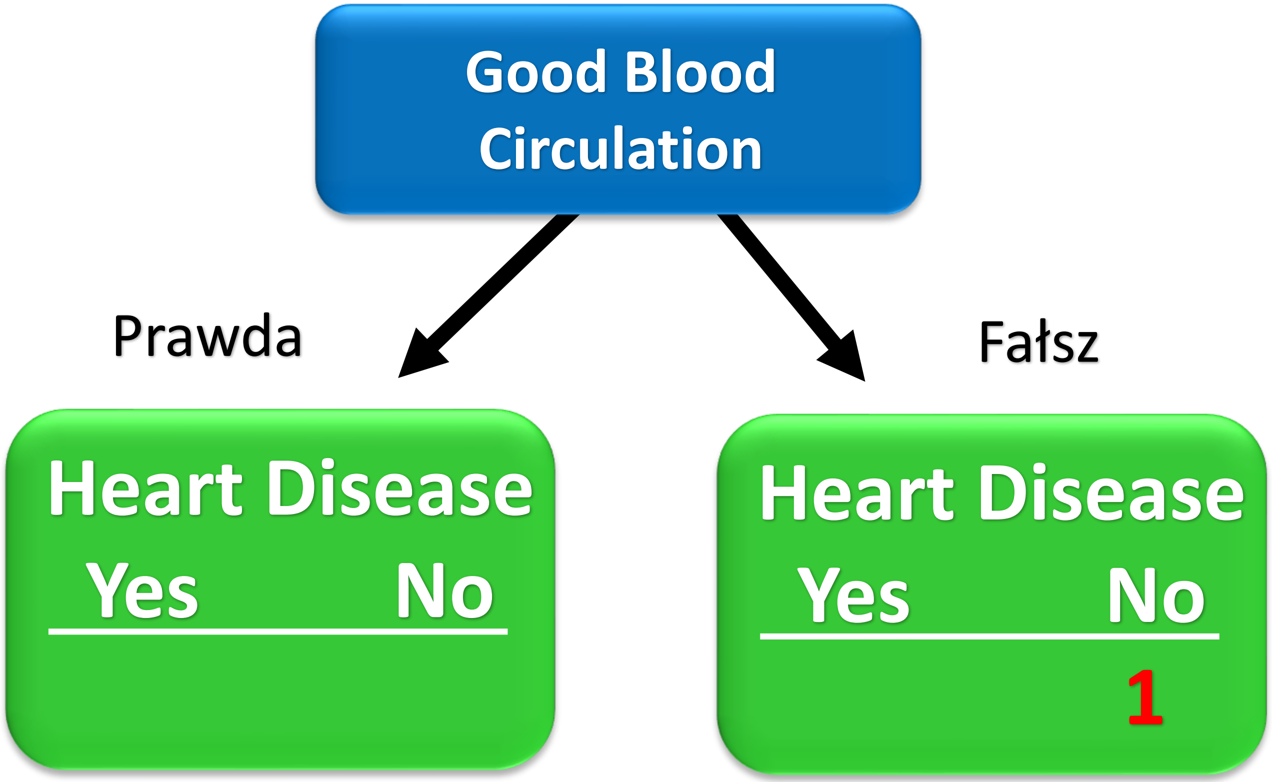


Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…

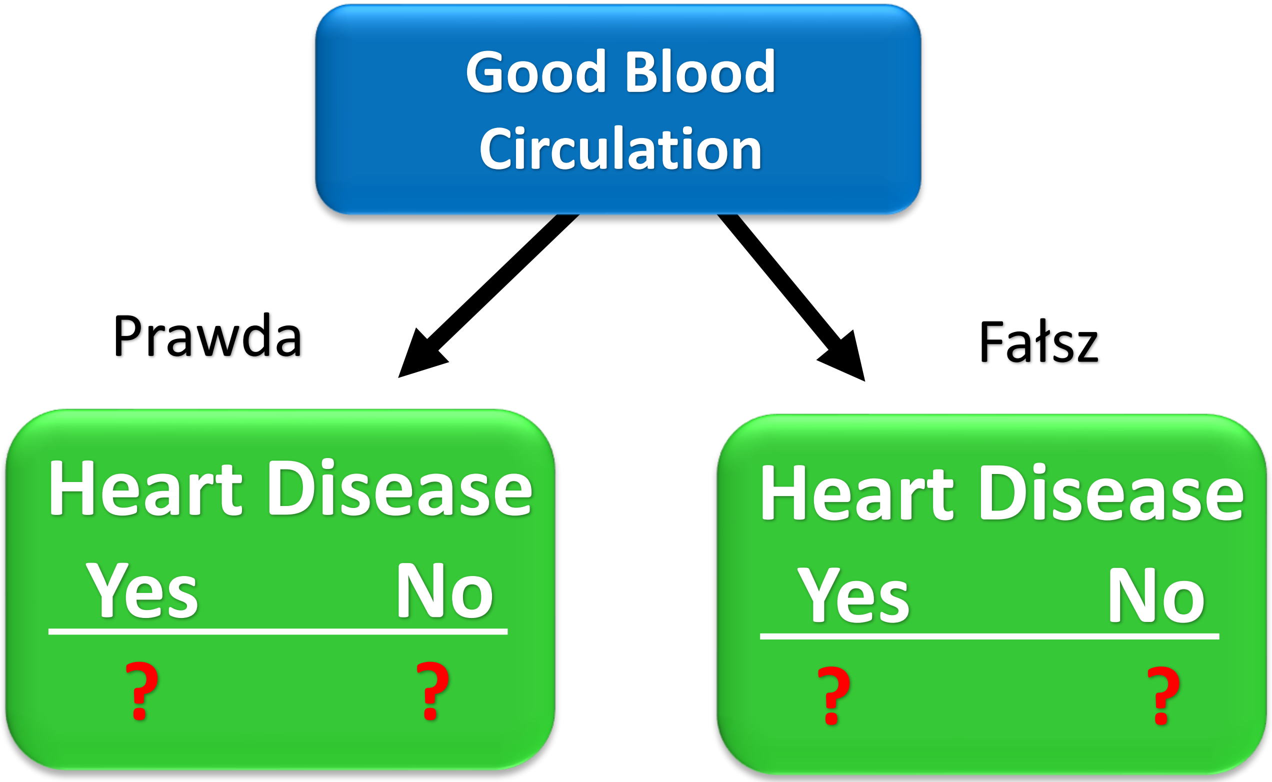


Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…

Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
...	...	...	...

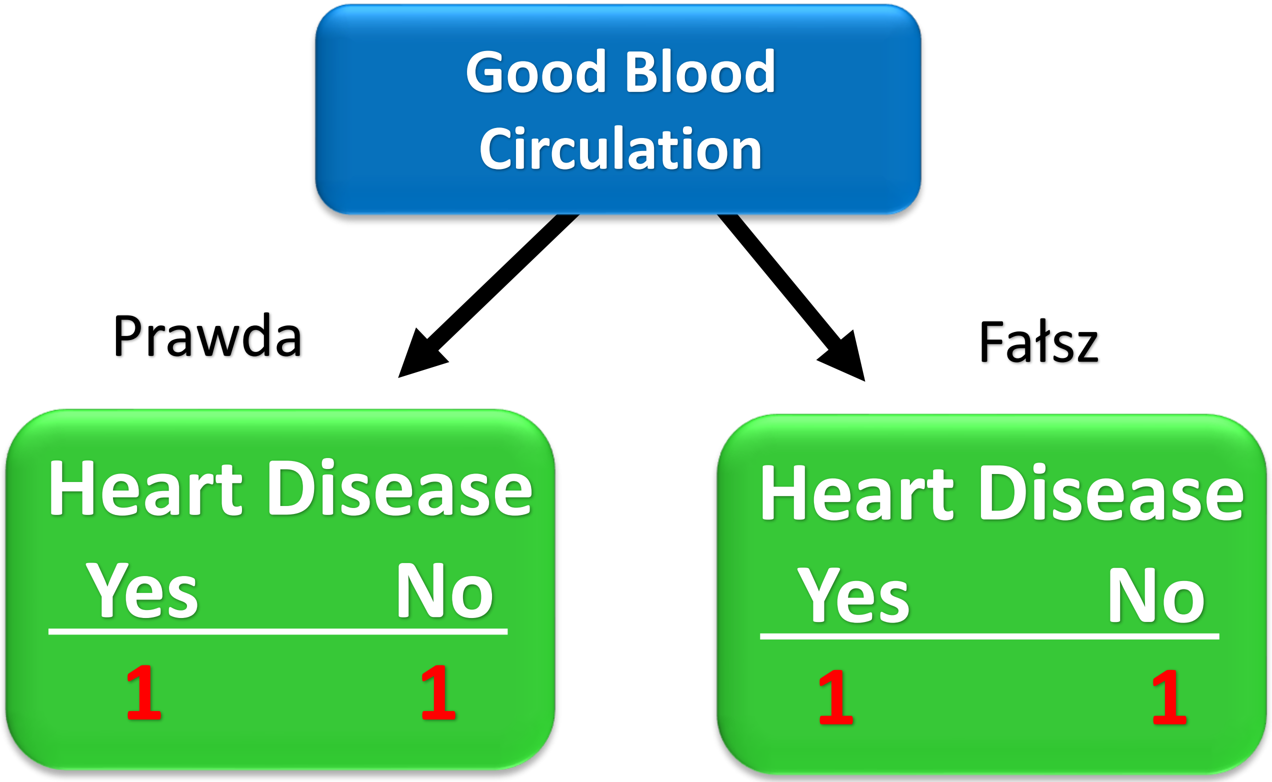


Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…

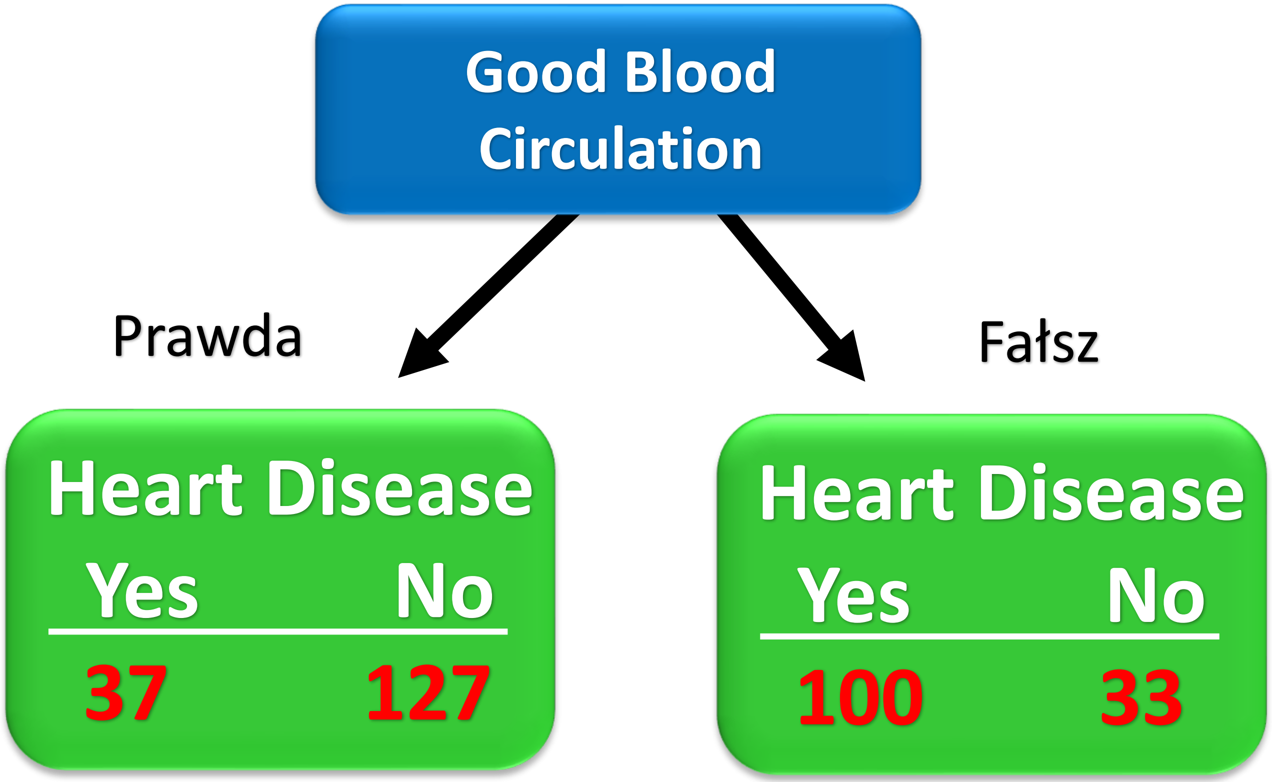


Pytanie kontrolne!

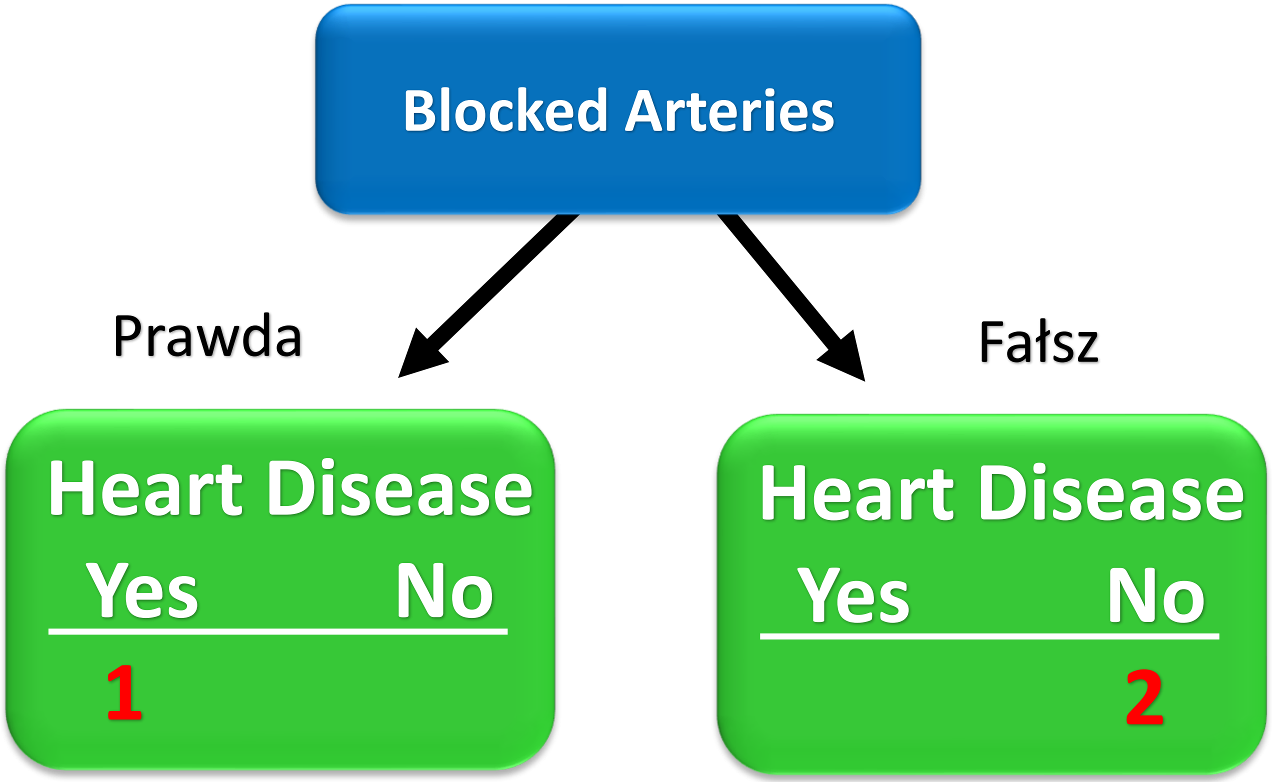
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



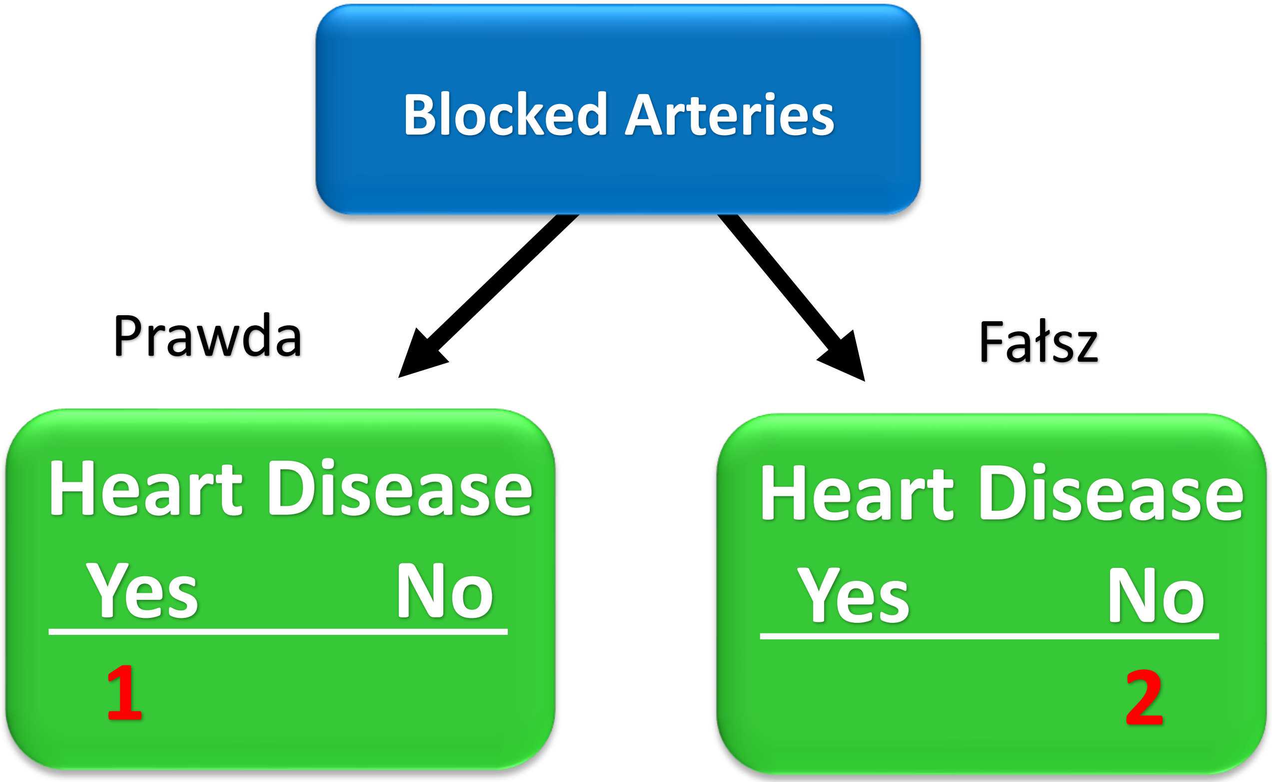
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Nan	Yes
…	…	…	…



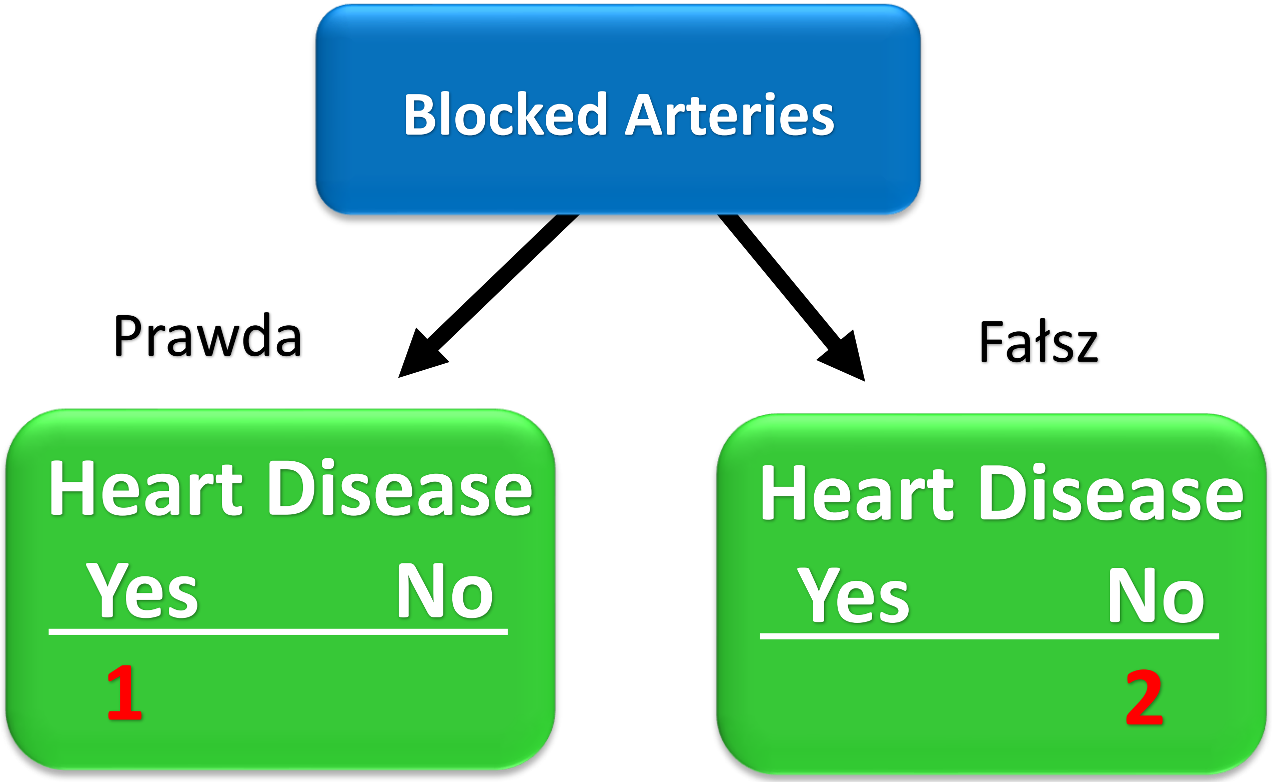
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	NA	Yes
…	…	…	…



NA vs NaN

Pytanie kontrolne!

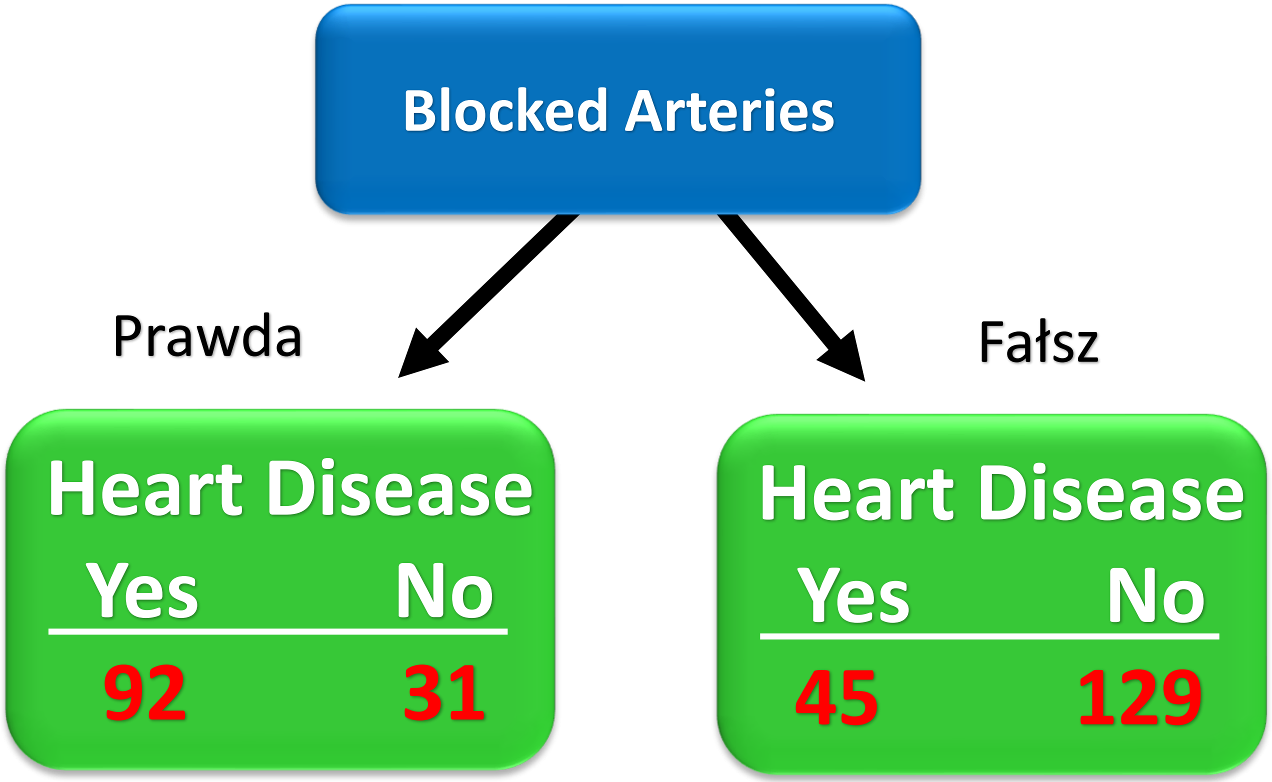
Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	NA	Yes
...	...	...	...



Dealing with missing data

Skip

Chest Pain	Good Blood Ciculation	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	NA	Yes
…	…	…	…



Idealnie by było…

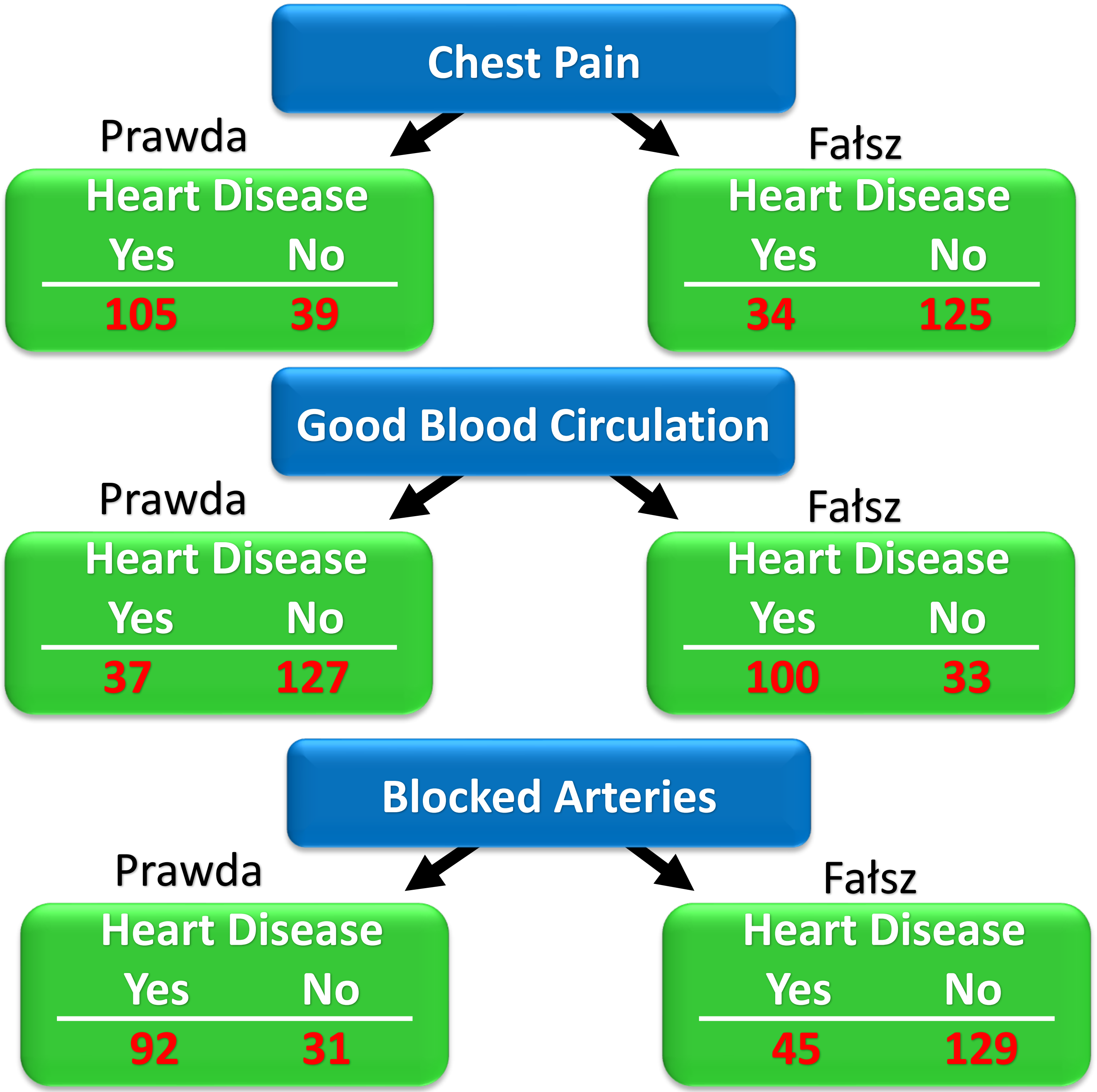
100% vs 0%

Jest nieczysto

„impure”

Jak zmierzyć

Impurity?

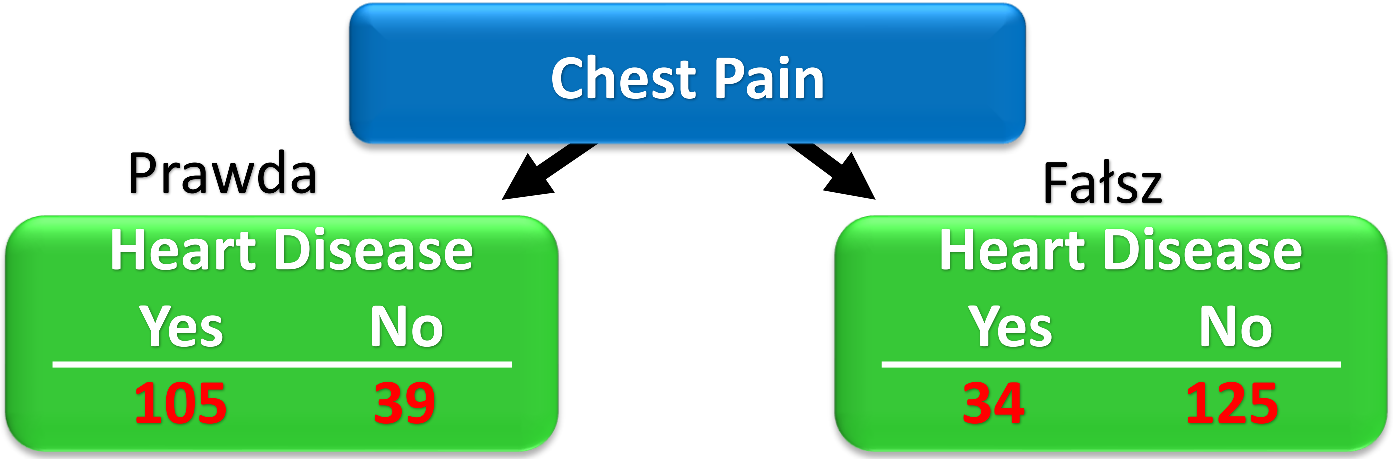


$$\mathrm{gini}(S) = 1 - \sum \mathrm{p_j}^2$$

S – *zbiór przykładów należących do n k*las

$\mathrm{p_j}$ – względna częstość wystepowania klasy j w S

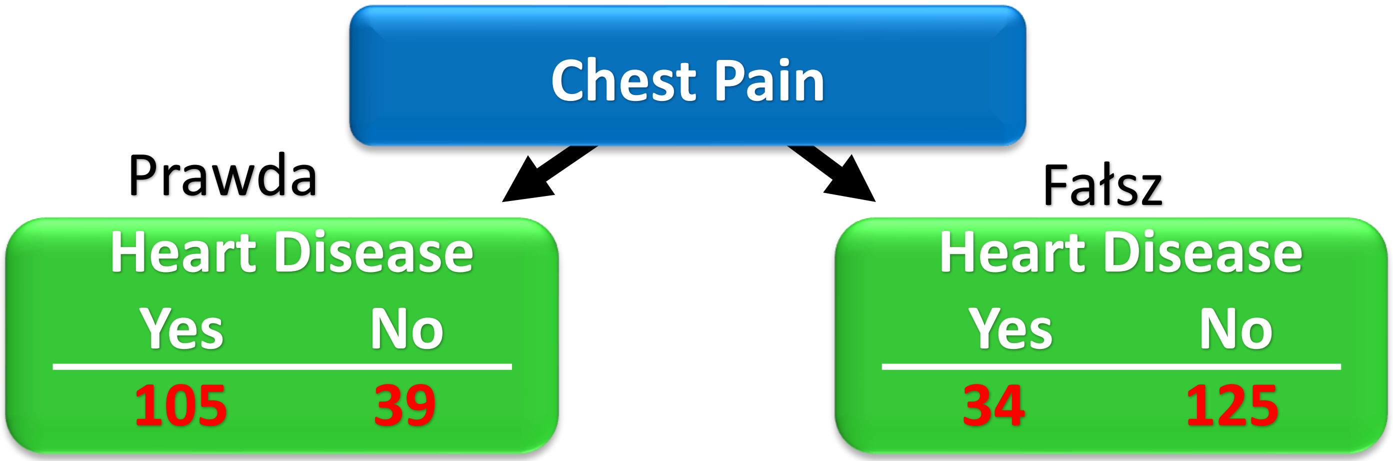
$$\text{Gini} = 1 - (p_{yes})^2 - (p_{no})^2$$



$$\text{Gini}_1 = 1 - \left(\frac{105}{105 + 39}\right)^2 - \left(\frac{39}{105 + 39}\right)^2 = 0.395$$

$$\text{Gini}_2 = 1 - \left(\frac{34}{34 + 125}\right)^2 - \left(\frac{125}{34 + 125}\right)^2 = 0.336$$

$$\text{Gini} = 1 - (p_{yes})^2 - (p_{no})^2$$

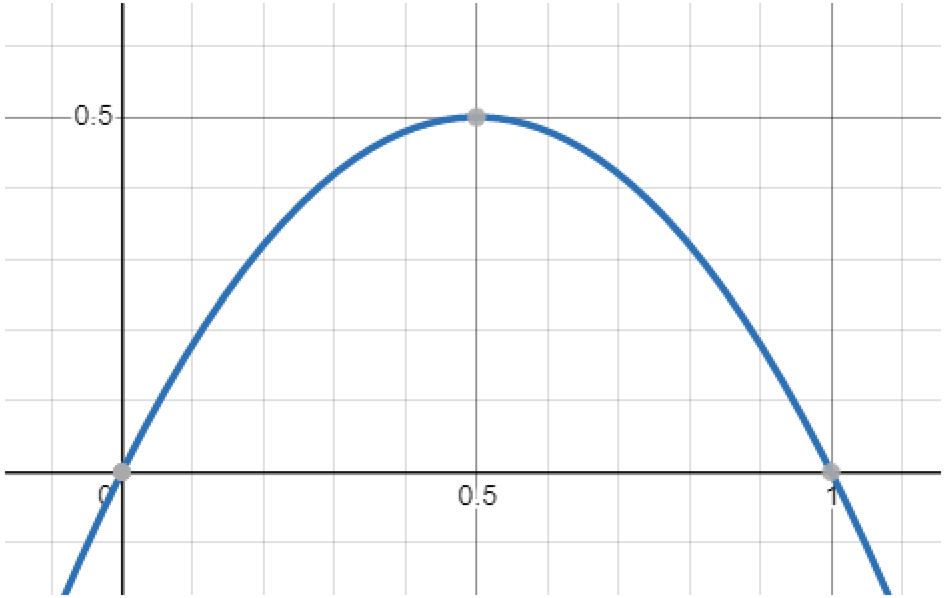


0.395 0.336

Gini całkowite = średnia ważona

$$\text{Gini}_C = \frac{144*0.395 + 159*0.336}{144 + 159} = 0.364$$

$$Gini = 1 - (p_{yes})^2 - (p_{no})^2$$



$$\mathrm{Gini}_{\mathrm{CP}} = 0.364$$

Chest Pain

Prawda: Heart Disease Yes	Prawda: Heart Disease No	Fałsz: Heart Disease Yes	Fałsz: Heart Disease No
105	39	34	125

$$\boxed{\mathrm{Gini}_{\mathrm{GBC}} = 0.360}$$

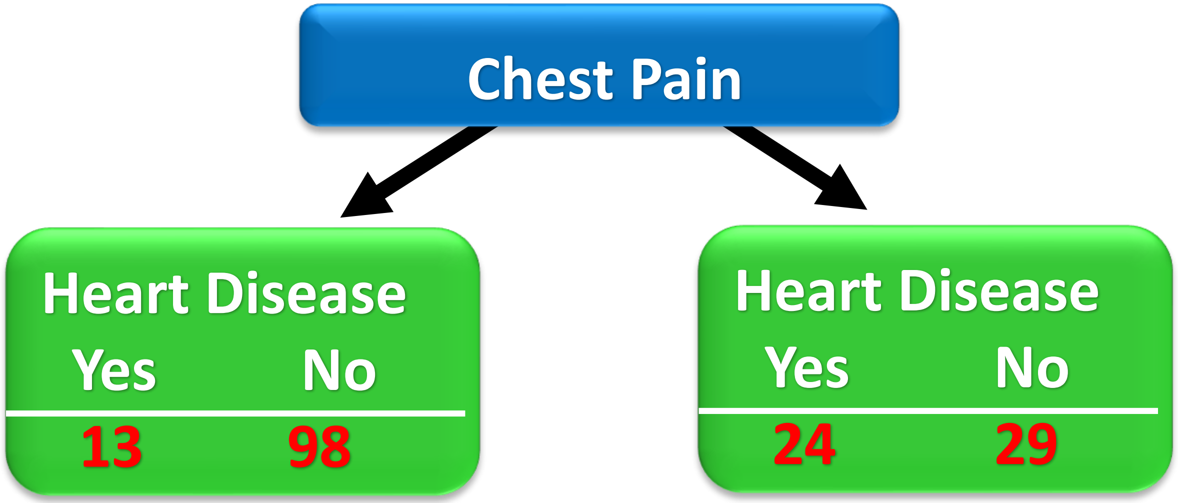
Good Blood Circulation

Prawda: Heart Disease Yes	Prawda: Heart Disease No	Fałsz: Heart Disease Yes	Fałsz: Heart Disease No
37	127	100	33

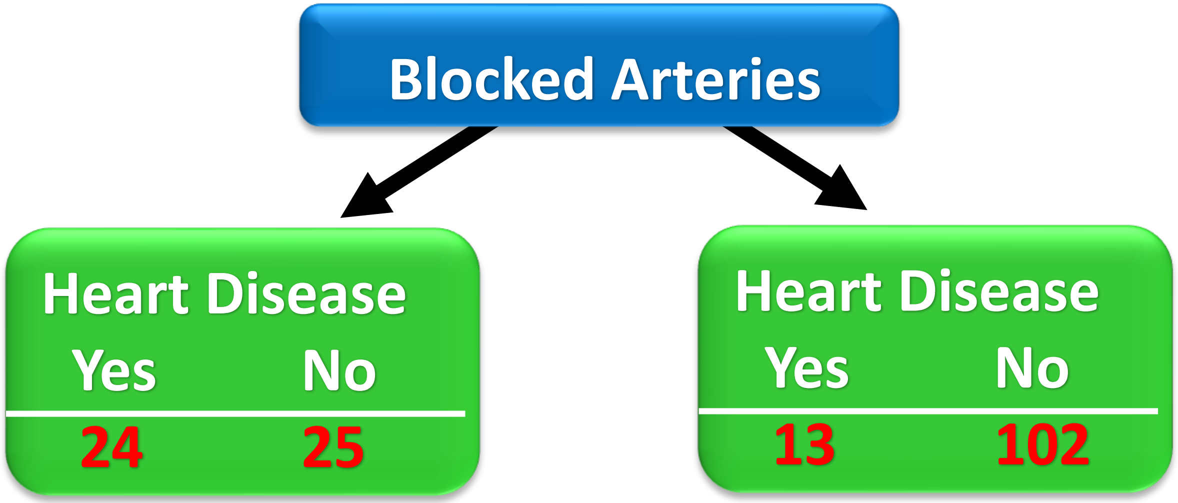
$$\mathrm{Gini}_{\mathrm{BA}} = 0.381$$

Blocked Arteries

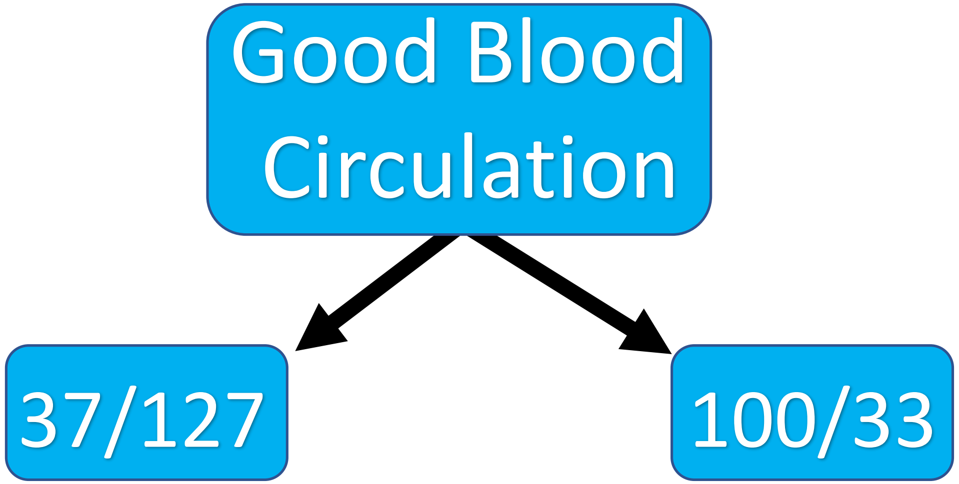
Prawda: Heart Disease Yes	Prawda: Heart Disease No	Fałsz: Heart Disease Yes	Fałsz: Heart Disease No
92	31	45	129

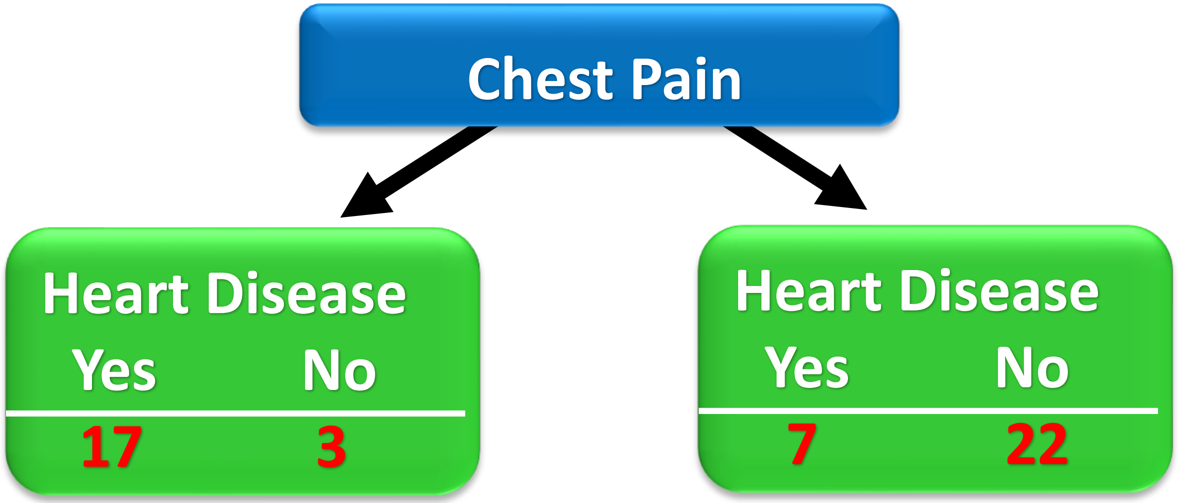


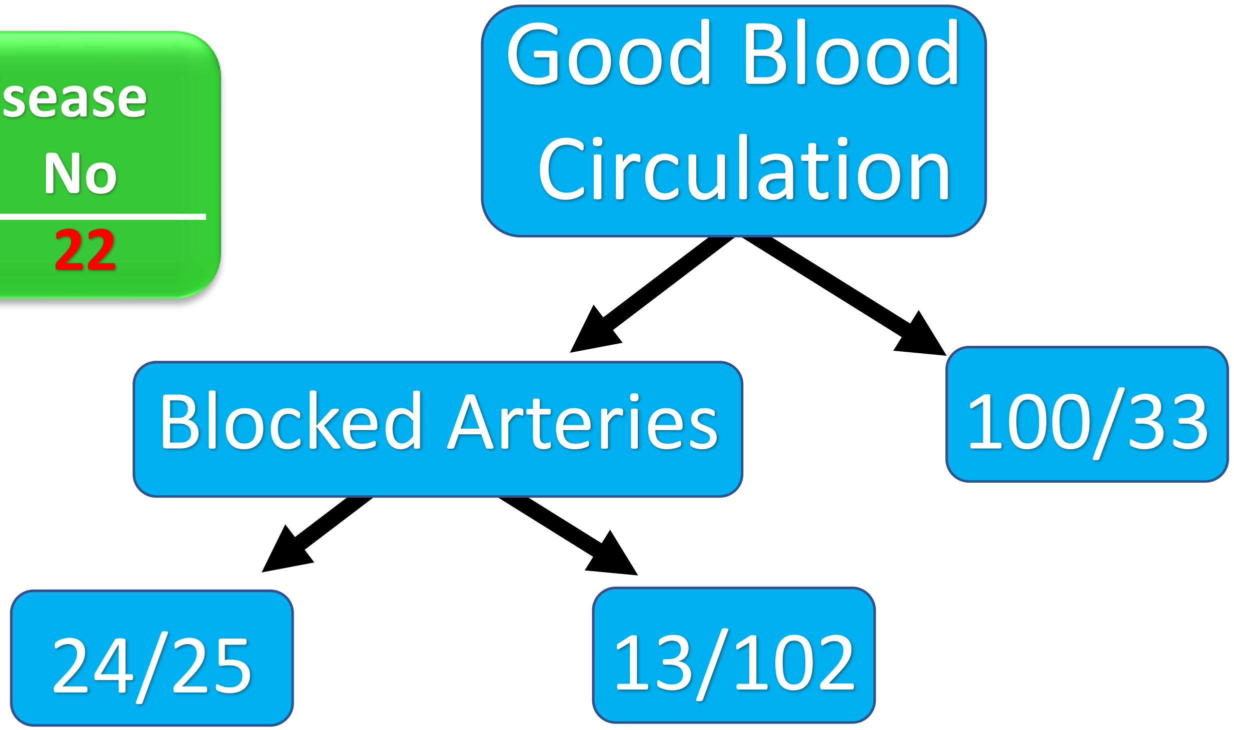
$$\text{Gini}_{\text{CP}} = 0.3$$

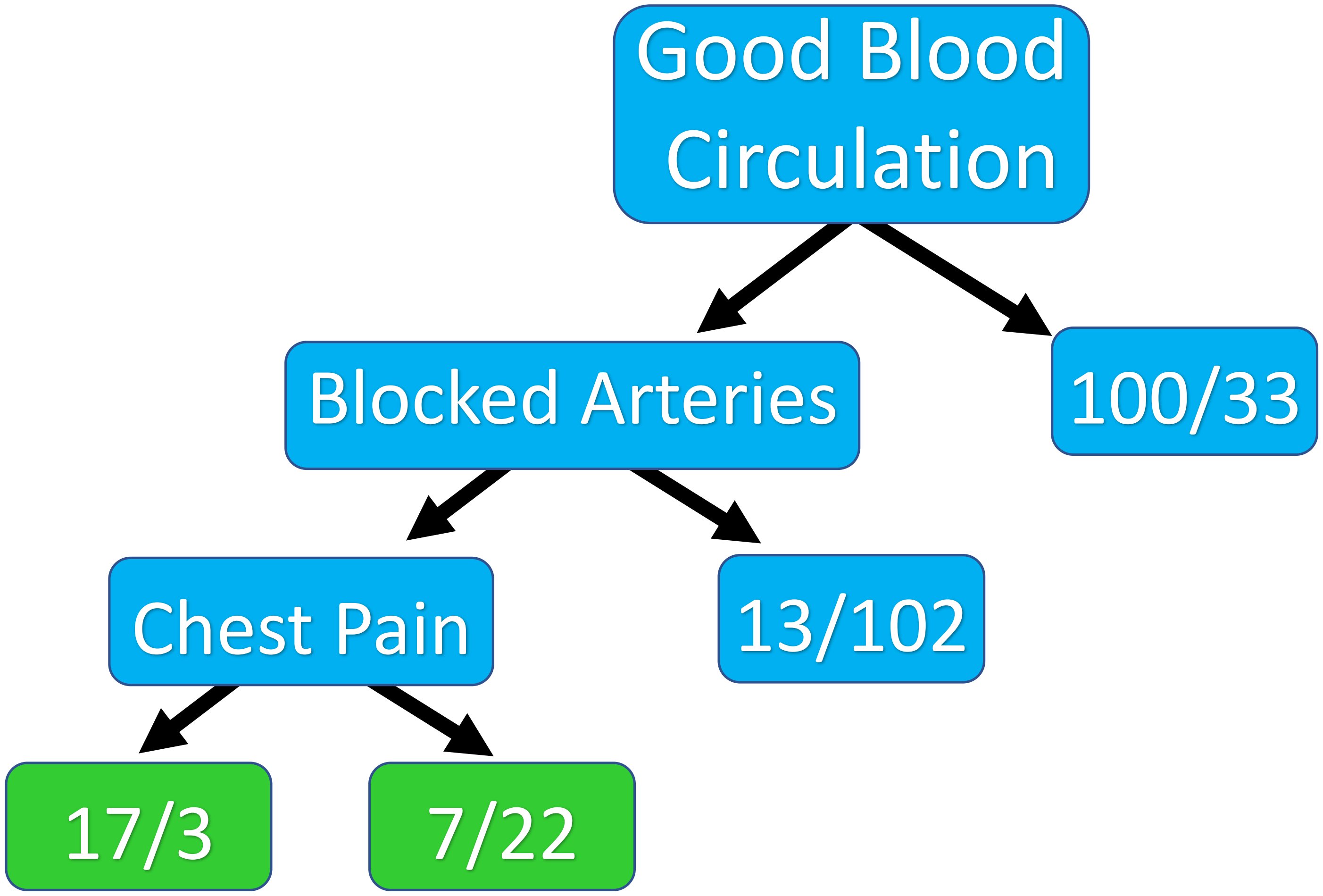


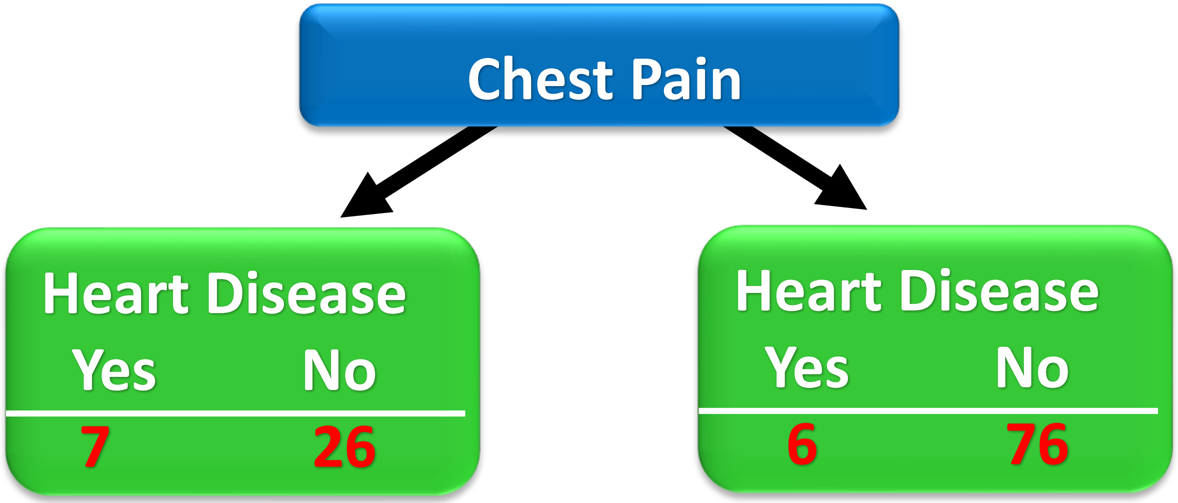
$$\boxed{\text{Gini}_{\text{BA}} = 0.290}$$



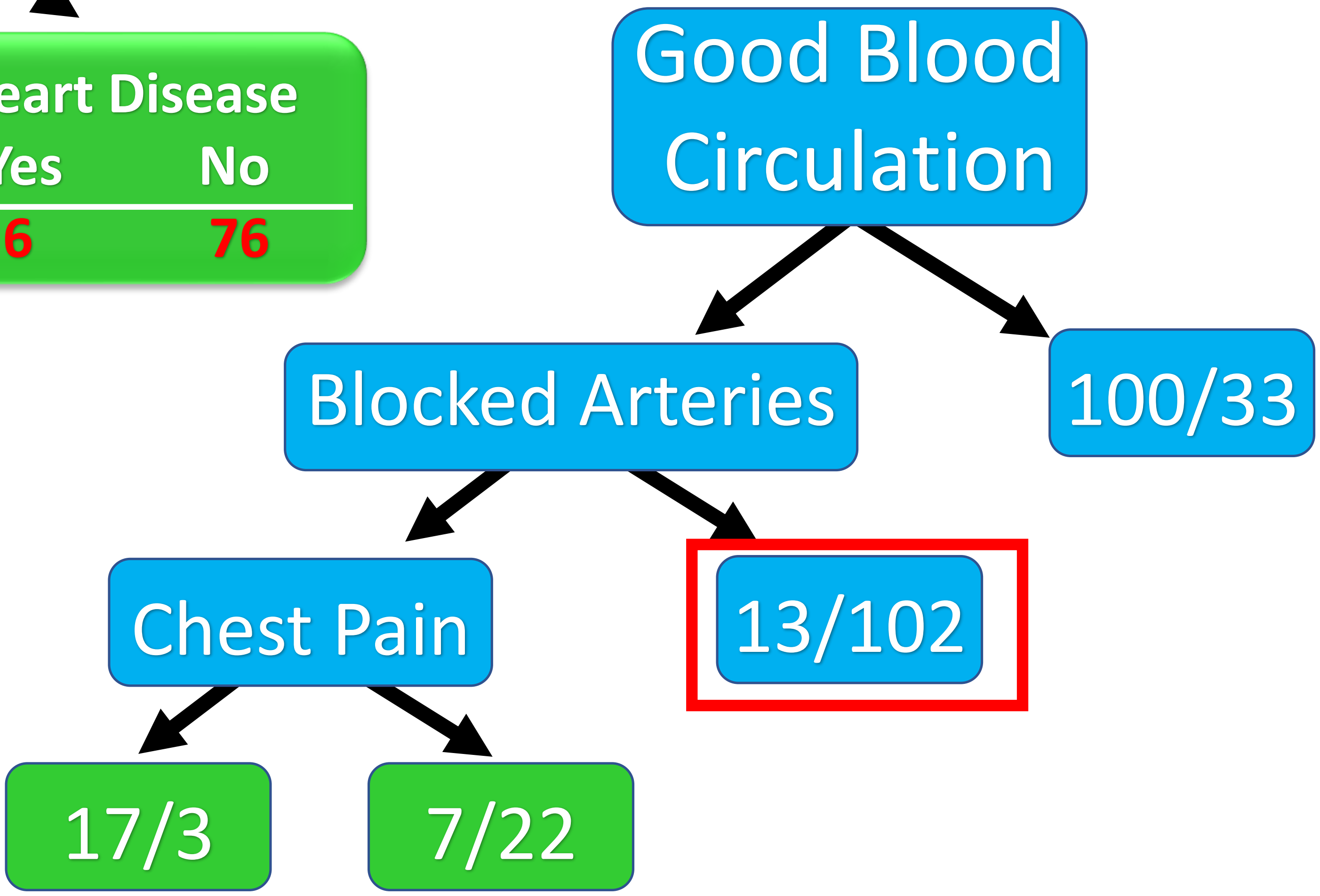

Chest Pain
Heart Disease
Yes No
17 3
Heart Disease
Yes No
7 22


Good Blood Circulation
Blocked Arteries
100/33
24/25
13/102

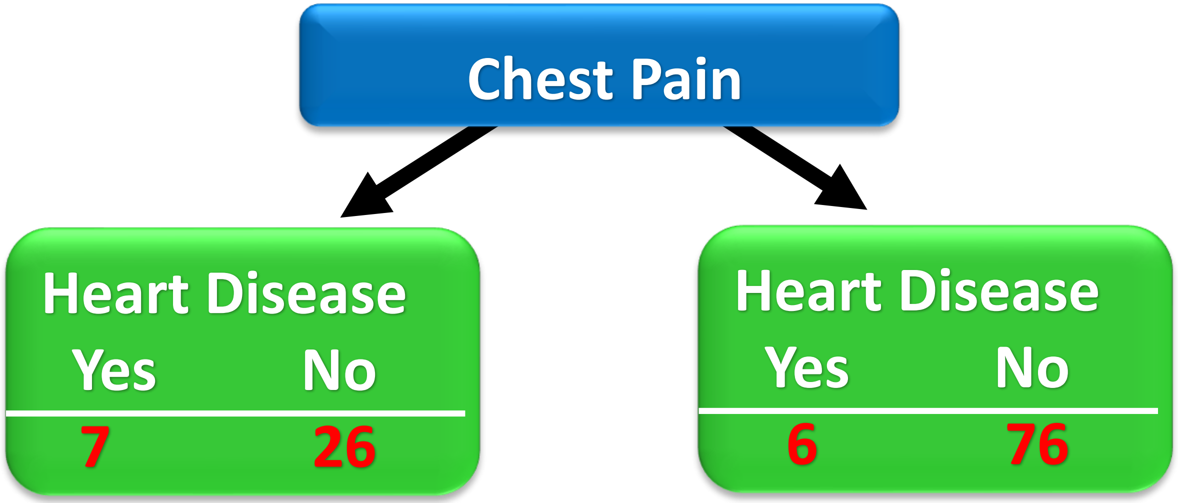

Good Blood Circulation
Blocked Arteries
100/33
Chest Pain
13/102
17/3
7/22



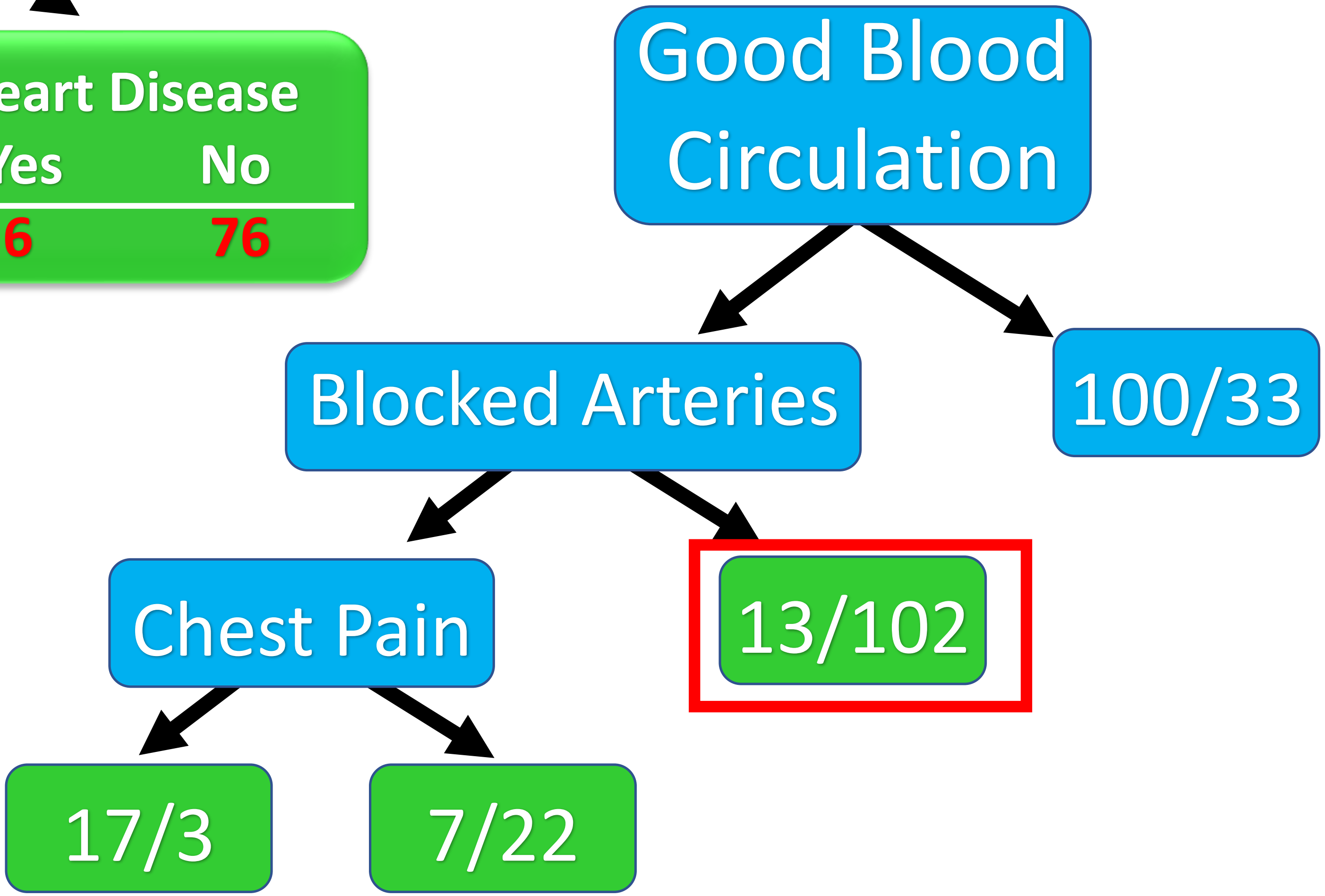
Gini = 0.29



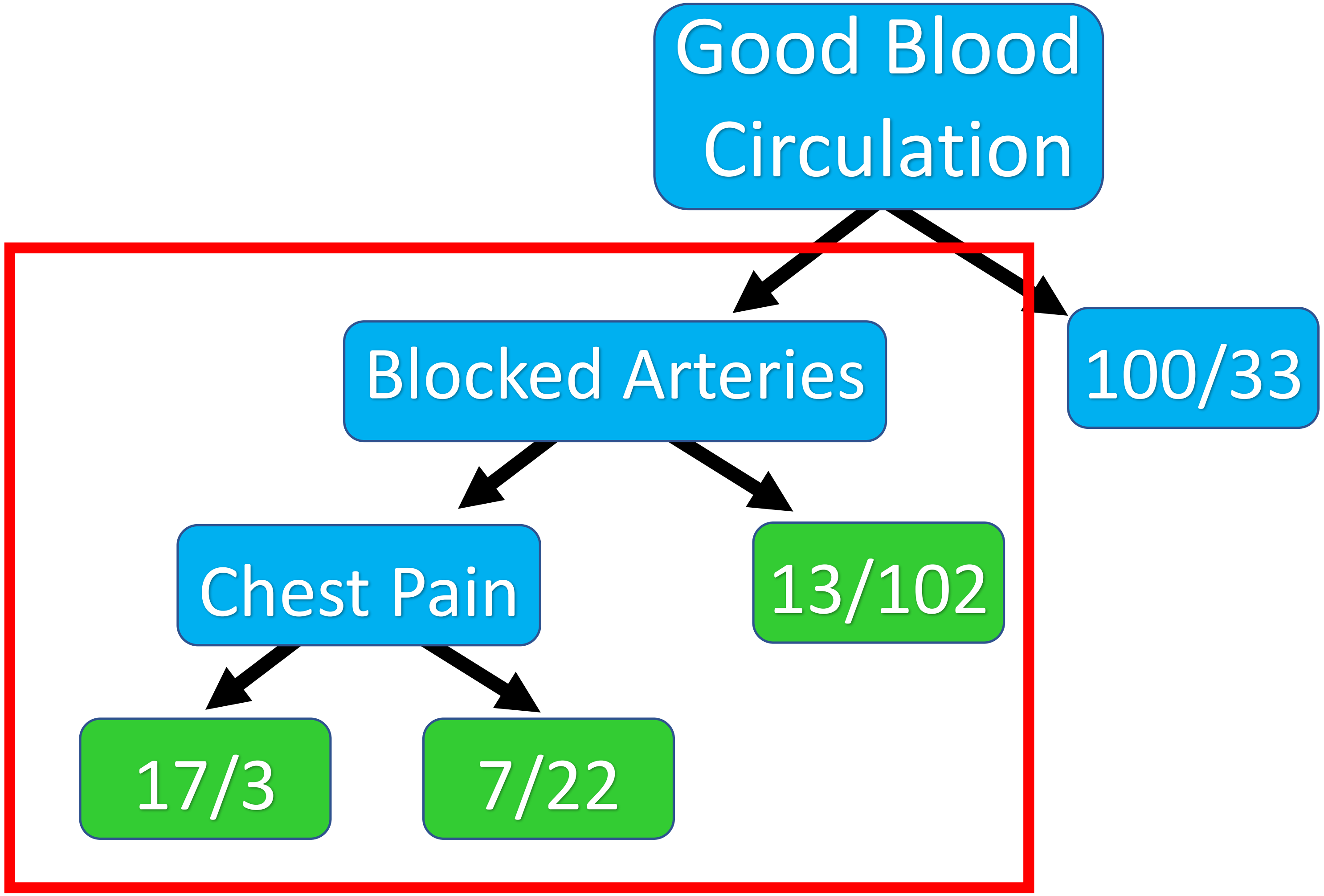
$$\text{Gini}_0 = 1 - \left(\frac{13}{13+102}\right)^2 - \left(\frac{102}{13+102}\right)^2 = 0.2$$

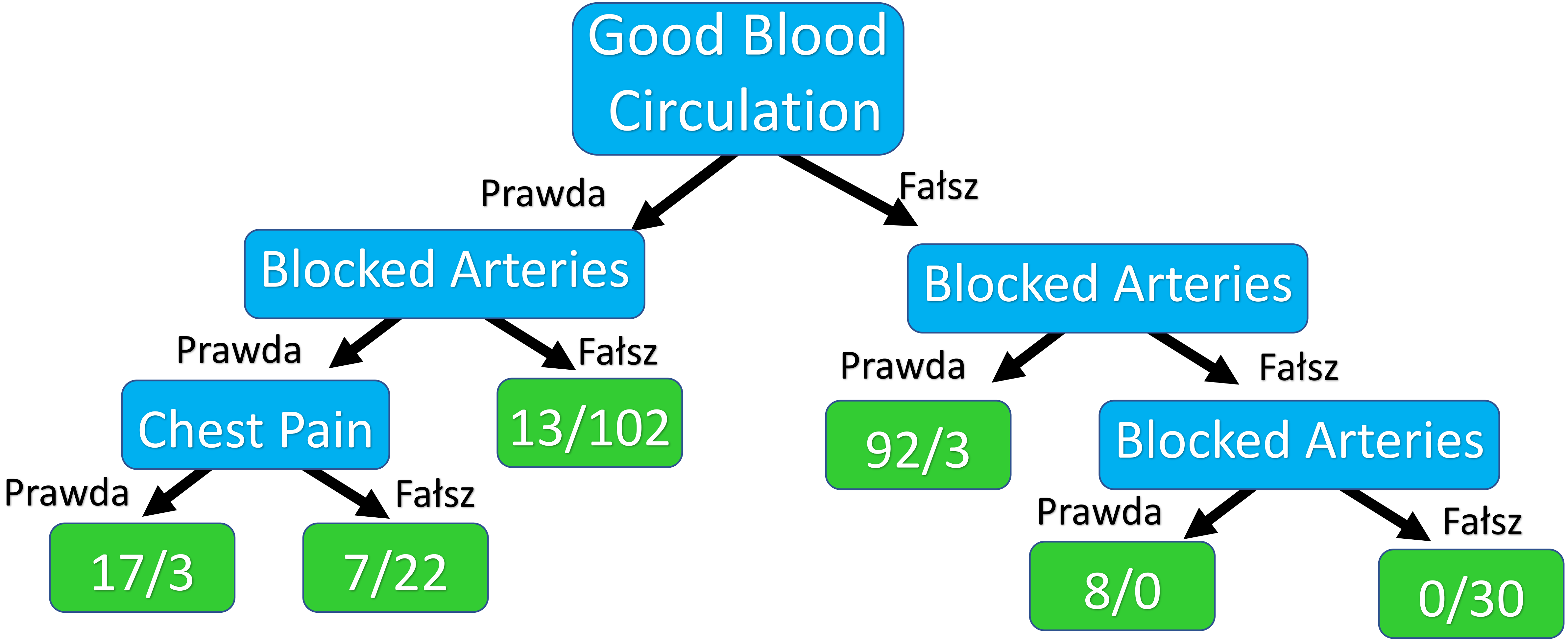


Gini = 0.29



$$\text{Gini}_0 = 1 - \left(\frac{13}{13 + 102}\right)^2 - \left(\frac{102}{13 + 102}\right)^2 = 0.2$$

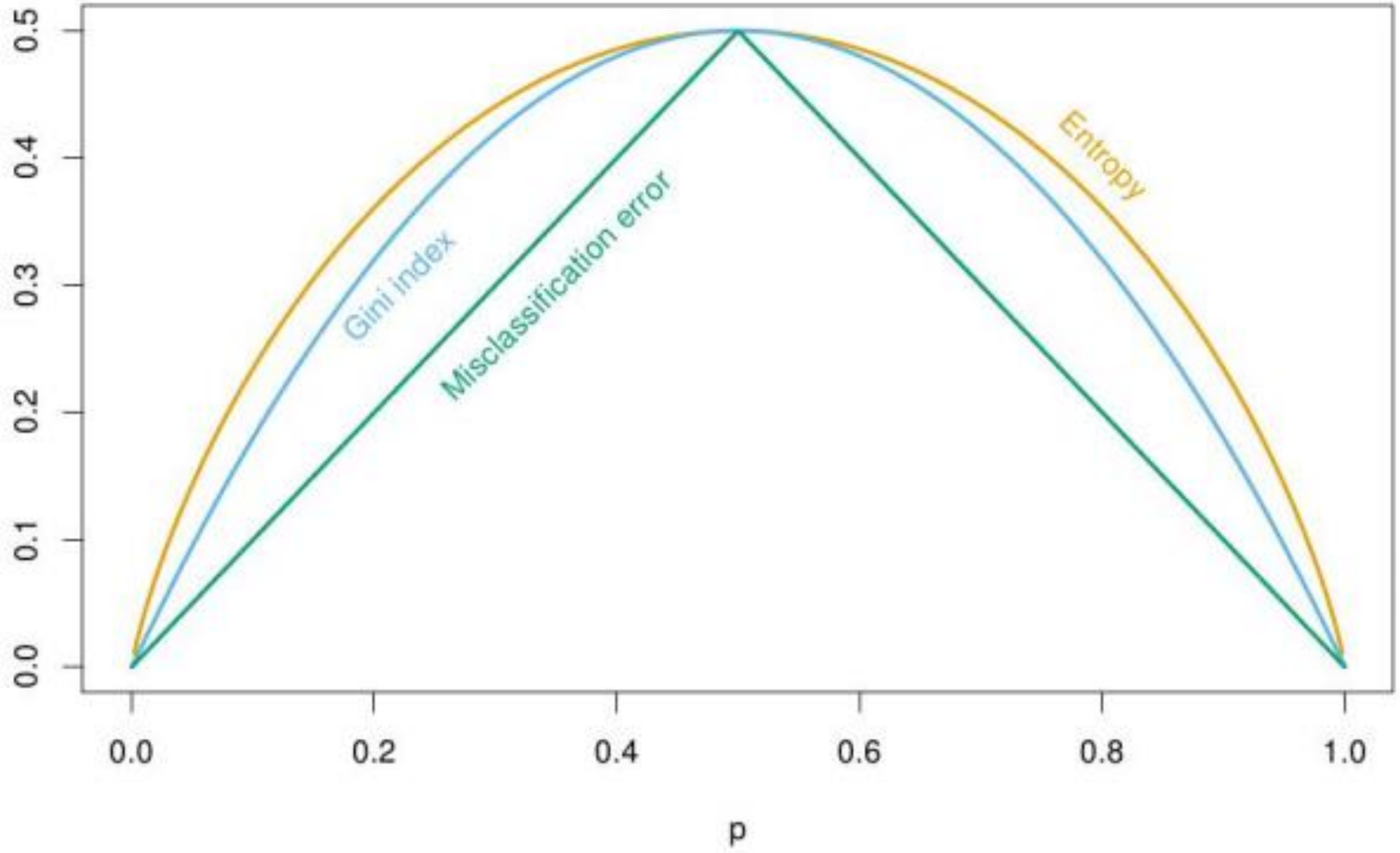

Good Blood Circulation
Blocked Arteries
100/33
Chest Pain
13/102
17/3
7/22


Good Blood Circulation
Prawda
Fałsz
Blocked Arteries
Blocked Arteries
Prawda
Fałsz
Prawda
Fałsz
Chest Pain
13/102
92/3
Blocked Arteries
Prawda
Fałsz
Prawda
Fałsz
17/3
7/22
8/0
0/30

1, Classification error: $1 - \max_k p_k$

2, Gini Index: $1 - \sum_k p_k^2$

3, Entropy: $-\sum_k p_k \ln p_k$



Pytania rekrutacyjne

Algorytmy i struktury danych

Uczenie nadzorowane vs nienadzorowane

Zadania z prawdopodobieństwa

Metody numeryczne

Task 1

For a Poisson random variables X and Z, the probability mass function was given by:

$$P(X = x_i; \lambda_i, z_i) = \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} = \frac{(\theta_0 + \theta_1 z_i)^{x_i} e^{-(\theta_0+\theta_1 z_i)}}{x_i!}, x_i \in \{0, 1, ..., \infty\}, 0 < \lambda_i = (\theta_0 + \theta_1 z_i), z_i \in \{0, 1\}$$

As for the likelihood function, considering an independent sample $x_1, x_2, \ldots, x_n$, from a Poisson variables

$$L = \prod_{i=1}^{N} \frac{(\theta_0 + \theta_1 z_i)^{x_i} e^{-(\theta_0+\theta_1 z_i)}}{x_i!} = \frac{\prod_{i=1}^{N} (\theta_0 + \theta_1 z_i)^{x_i}}{\prod_{i=1}^{N} x_i!} e^{-\sum_{i=1}^{N} (\theta_0+\theta_1 z_i)} \tag{1}$$

Applying natural logarithm to likelihood function

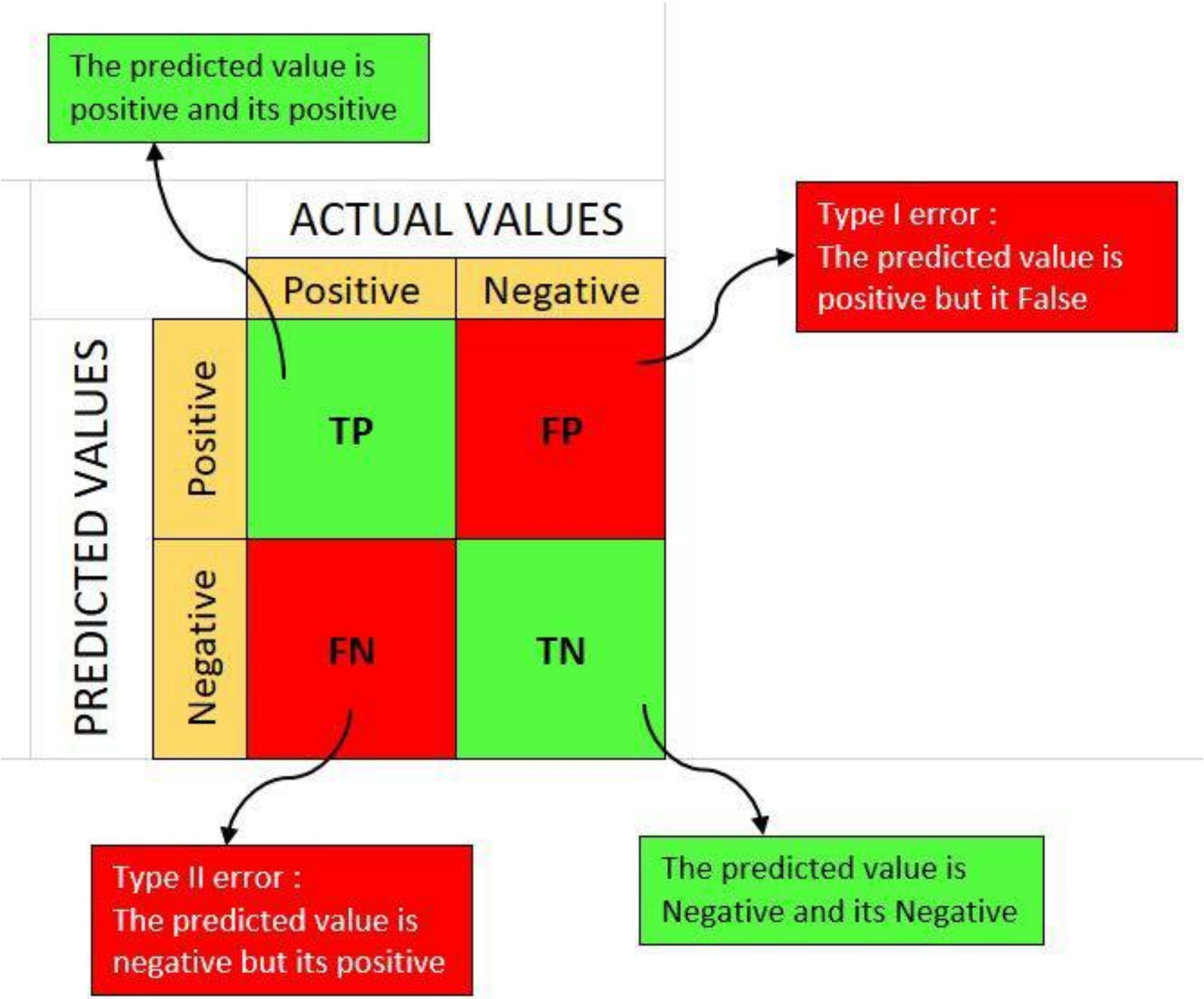
$$\begin{aligned} lnL = ln(\prod_{i=1}^{N} (\theta_0 + \theta_1 z_i)^{x_i}) + ln(e^{-\sum_{i=1}^{N} (\theta_0+\theta_1 z_i)}) - ln(\prod_{i=1}^{N} x_i!) = \\ = \sum_{i=1}^{N} x_i ln(\theta_0 + \theta_1 z_i) - \sum_{i=1}^{N} (\theta_0 + \theta_1 z_i) - \sum_{i=1}^{N} ln(x_i!) \end{aligned} \tag{2}$$

Calculating gradient

$$\frac{\partial lnL}{\partial \theta_0} = \sum_{i=1}^{N} \frac{x_i}{\theta_0 + \theta_1 z_i} - n \quad \frac{\partial lnL}{\partial \theta_1} = \sum_{i=1}^{N} \frac{x_i z_i}{\theta_0 + \theta_1 z_i} - \sum_{i=1}^{N} z_i \tag{3}$$

Calculating Hessian matrix

$$\begin{bmatrix} \frac{\partial^2 lnL}{\partial \theta_0^2} = \sum_{i=1}^{N} \frac{-x_i}{(\theta_0+\theta_1 z_i)^2} & \frac{\partial^2 lnL}{\partial \theta_0 \partial \theta_1} = \sum_{i=1}^{N} \frac{-x_i z_i}{(\theta_0+\theta_1 z_i)^2} \\ \frac{\partial^2 lnL}{\partial \theta_1 \partial \theta_0} = \sum_{i=1}^{N} \frac{-x_i z_i}{(\theta_0+\theta_1 z_i)^2} & \frac{\partial^2 lnL}{\partial \theta_1^2} = \sum_{i=1}^{N} \frac{-x_i z_i^2}{(\theta_0+\theta_1 z_i)^2} \end{bmatrix}$$

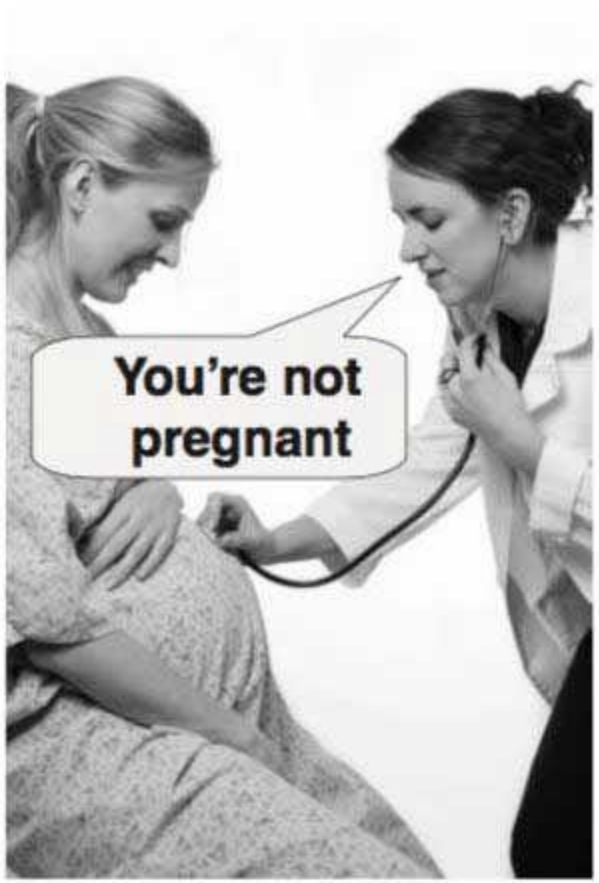


$$P = \frac{T_p}{T_p + F_p}$$

$$R = \frac{T_p}{T_p + F_n}$$

Type I error
(false positive)

Type II error
(false negative)




THANK YOU