

# Modelowanie ryzyka kredytowego w banku

Bartosz Jeżewski

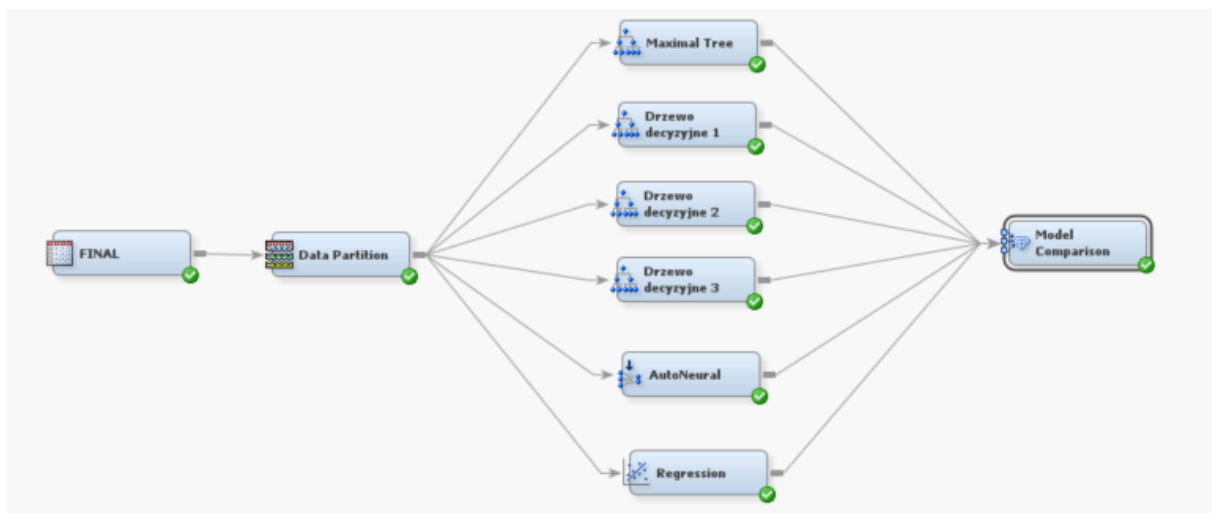
---

---

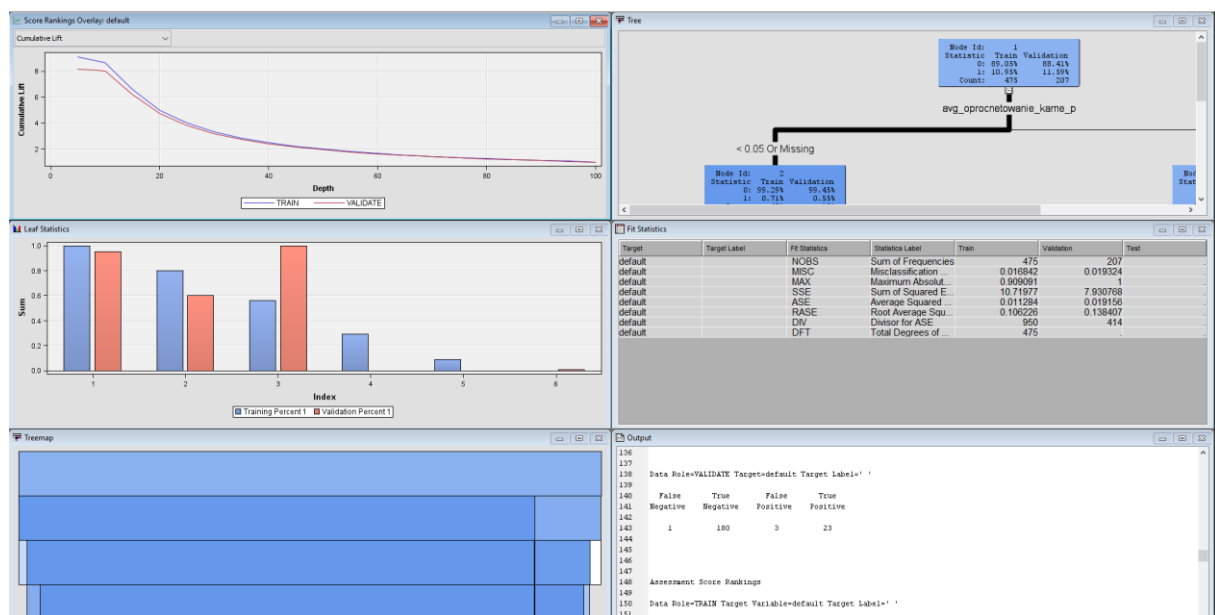
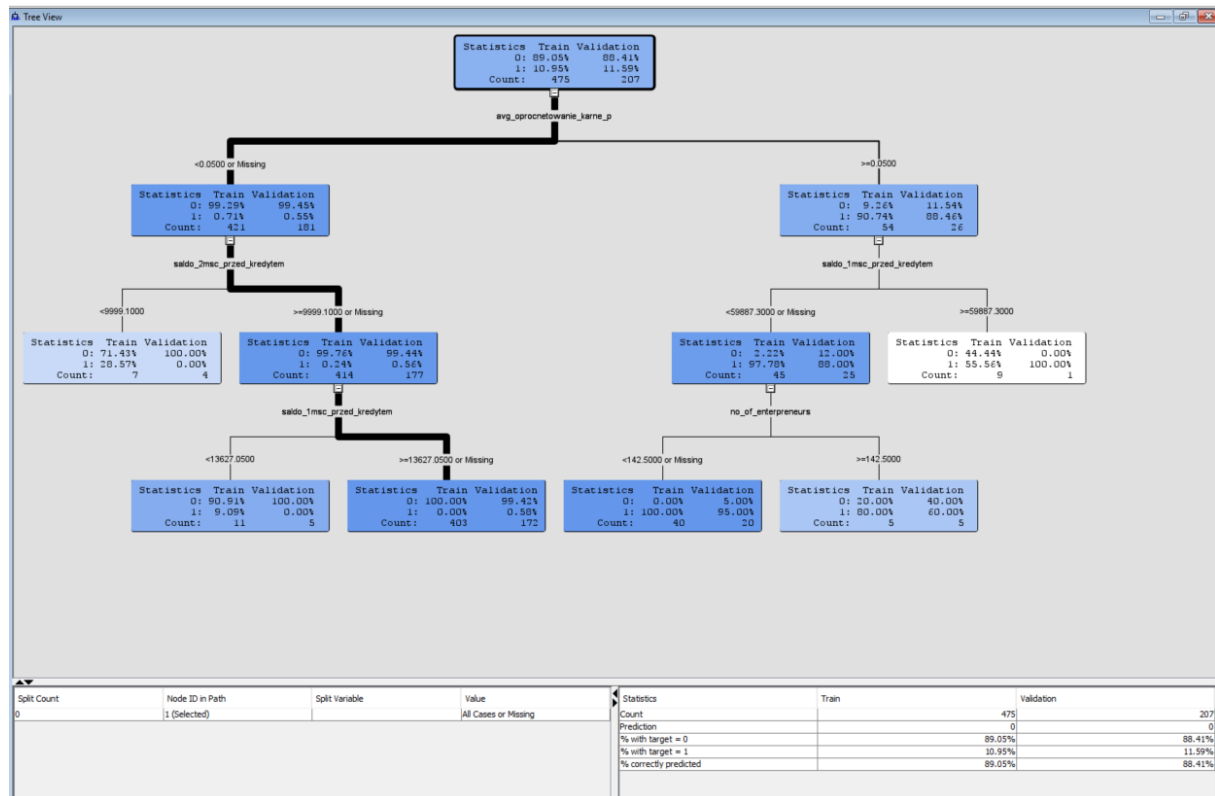
Lista zmiennych:

1. default – zmienna zależna
2. plec – płeć klienta.
3. overall\_sum\_of\_transfers – suma wszystkich płatności na koncie klienta biorącego kredyt.
4. wiek – wiek klienta w latach na dzień przyznania kredytu.
5. staz – staż (od otwarcia konta do przyznania kredytu) klienta w latach na dzień przyznania kredytu.
6. dlug\_na\_koncie – informacja czy klient posiada dług na koncie „tak” i „nie”.
7. frequency – częstość wyciągów na koncie klienta.
8. najczestsza\_kategoria – najczęstsza kategoria transakcji.
9. most\_bank\_transfers\_to – do jakiego banku najczęściej przesyłane są środki.
10. account\_peroid – czas posiadania konta.
11. card\_ownership – czy klient posiada kartę „tak” i „nie”.
12. card\_ownership\_duration – jeśli klient posiada kartę to jak długo.
13. saldo\_przy\_kredycie – saldo klienta na dzień zaciągania kredytu.
14. saldo\_1msc\_przed\_kredytem – saldo klienta na miesiąc przed kredytem.
15. saldo\_2msc\_przed\_kredytem – saldo klienta na dwa miesiące przed kredytem.
16. saldo\_3msc\_przed\_kredytem – saldo klienta na trzy miesiące przed kredytem.
17. sr\_wypłata\_region – średnia wypłata w regionie w którym mieszka klient.
18. cirmes95 – liczba popełnionych przestępstw w regionie -rok 1995.
19. cirmes96 - liczba popełnionych przestępstw w regionie -rok 1996.
20. bezrobocie95 – stopa bezrobocia regionu w roku 1995.
21. bezrobocie96 – stopa bezrobocia regionu w roku 1996.
22. mieszkancy\_regionu – ilość mieszkańców w regionie kredytobiorcy.
23. no\_of\_entrepreneurs - liczba przedsiębiorców na 1000 mieszkańców.
24. loan\_duration – czas trwania kredytu.
25. loan\_amount – kwota kredytu.
26. card\_type – typ karty ("junior", "classic", "gold").
27. avg\_transfer – ogólna średnia kwota przelewu
28. avg\_skladka\_ubezpieczenia – średnia kwota przelewów dla kategorii "składka ubezpieczenia" dla danego konta.

29. avg\_rachunki\_domowe - średnia kwota przelewów dla kategorii "rachunki domowe" dla danego konta.
30. avg\_leasing - średnia kwota przelewów dla kategorii "leasing" dla danego konta.
31. avg\_rata\_kredytu - średnia kwota przelewów dla kategorii "rata kredytu" dla danego konta.
32. avg\_splata\_kredytu\_p – średnia wartość transakcji spłaty kredytu.
33. avg\_emerytura\_p – średnia wartość transakcji emerytury.
34. avg\_wyd\_dom\_p – średnia wartość transakcji wydatków na potrzeby domowe.
35. avg\_oprocentowanie\_karne\_p – średnia wartość transakcji oprocentowania karnego kredytu przy ujemnym saldzie.
36. avg\_odsetki\_p – średnia wartość transakcji odsetek.
37. avg\_wyciag\_p – średnia wartość transakcji za wyciąg.
38. avg\_ubezpieczenie\_p – średnia wartość transakcji za ubezpieczenie.
39. liczba\_przychodow – liczba transakcji z typem przychód.
40. liczba\_wypłat – liczba transakcji z typem wypłaty.
41. wypłaty\_do\_przychodow – oznacza stosunek transakcji wypłat do transakcji przychodów.
42. payments – wielkość raty kredytu.

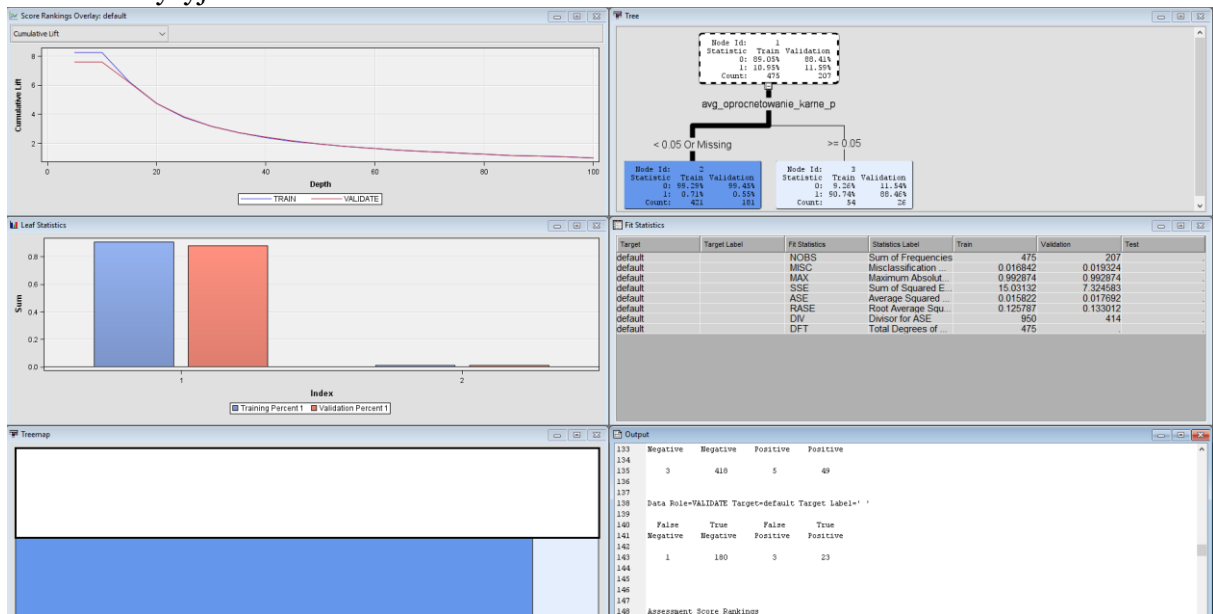


Maximal tree:



General		
Node ID	Tree	
Imported Data		...
Exported Data		...
Notes		...
Train		
Variables		...
Interactive		...
Import Tree Model	No	
Tree Model Data Set		...
Use Frozen Tree	No	
Use Multiple Targets	No	
Splitting Rule		
Interval Target Criterion	ProbF	
Nominal Target Criterion	ProbChisq	
Ordinal Target Criterion	Entropy	
Significance Level	0.2	
Missing Values	Use in search	
Use Input Once	No	
Maximum Branch	2	
Maximum Depth	6	
Minimum Categorical Size	5	
Node		
Leaf Size	5	
Number of Rules	5	
Number of Surrogate Rules	0	
Split Size		
Split Search		
Use Decisions	No	
Use Priors	No	
Exhaustive	5000	
Node Sample	20000	
Subtree		
Method	Assessment	
Number of Leaves	1	
Assessment Measure	Decision	
Assessment Fraction	0.25	
Cross Validation		
Perform Cross Validation	No	
Number of Subsets	10	
Number of Repeats	1	
Seed	12345	
Observation Based Import		
Observation Based Import	No	
Number Single Var Import	5	
P-Value Adjustment		
Bonferroni Adjustment	Yes	
Time of Bonferroni Adjustm	Before	
Inputs	No	
Number of Inputs	1	
Depth Adjustment	Yes	
Output Variables		
Leaf Variable	Yes	
Interactive Sample		
Create Sample	Default	
Sample Method	Random	
Sample Size	10000	
Sample Seed	12345	
Performance	Disk	

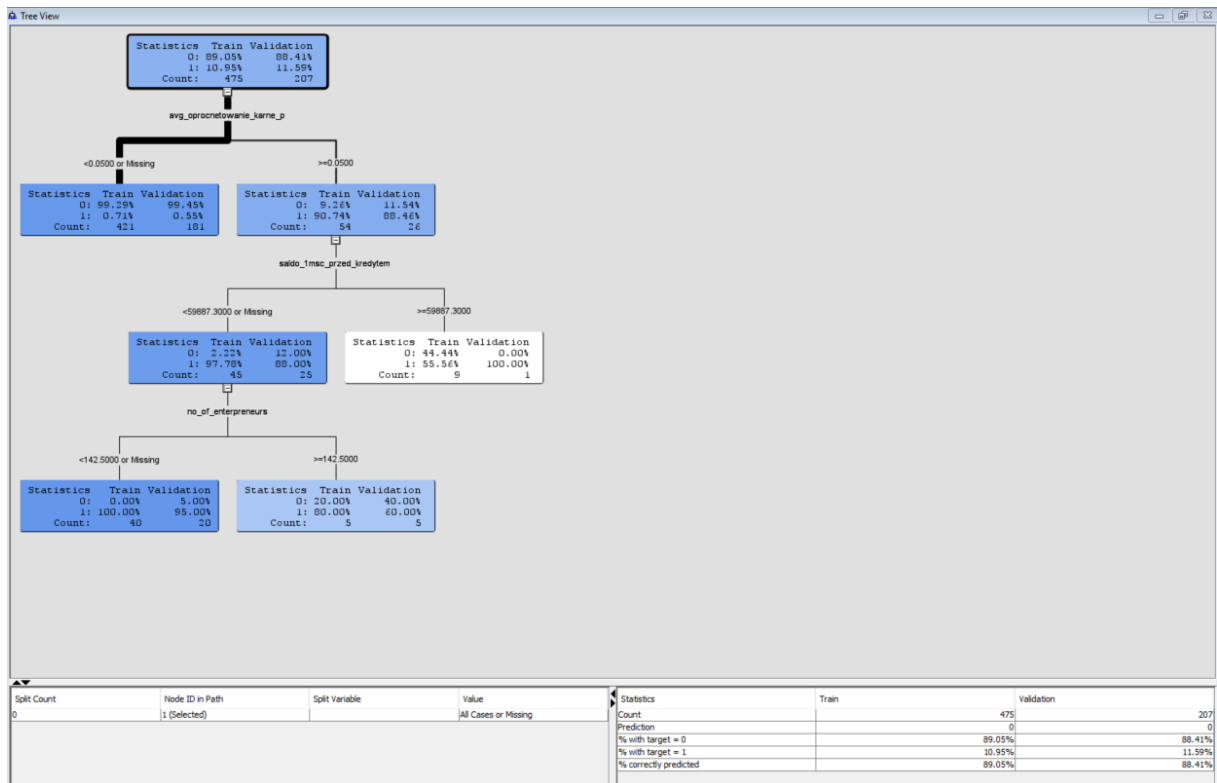
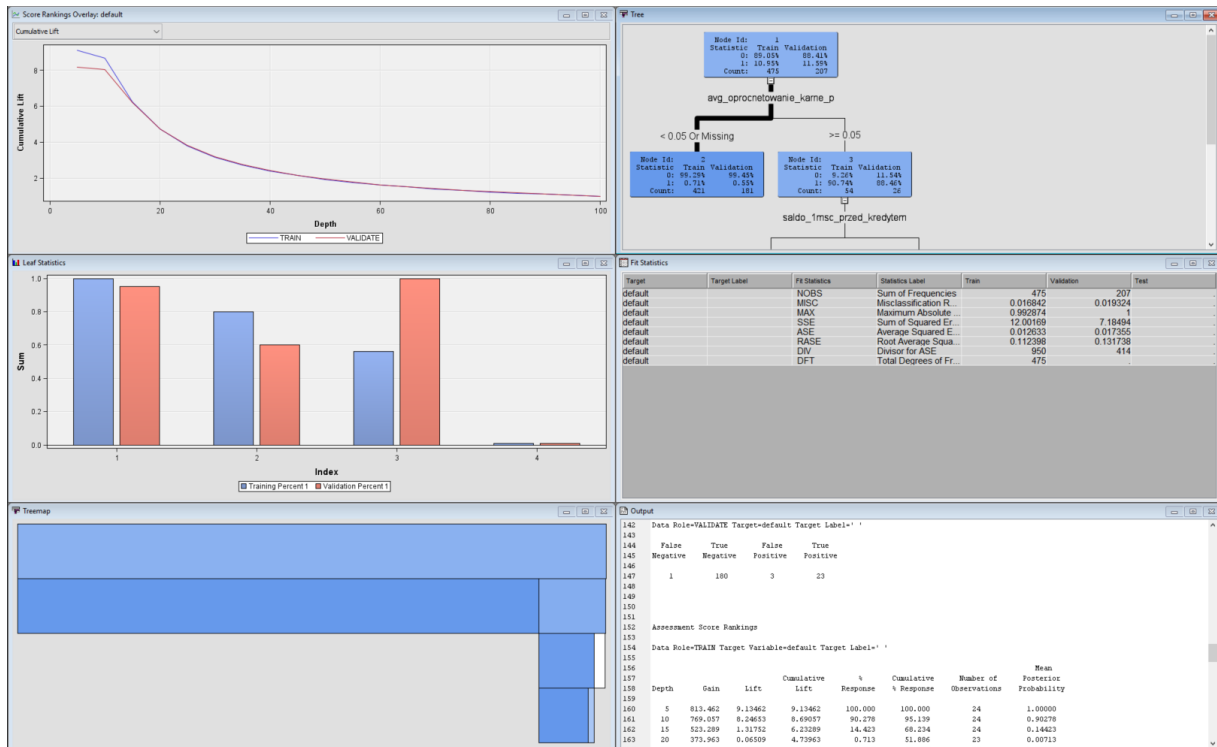
## Drzewo decyzyjne 1:

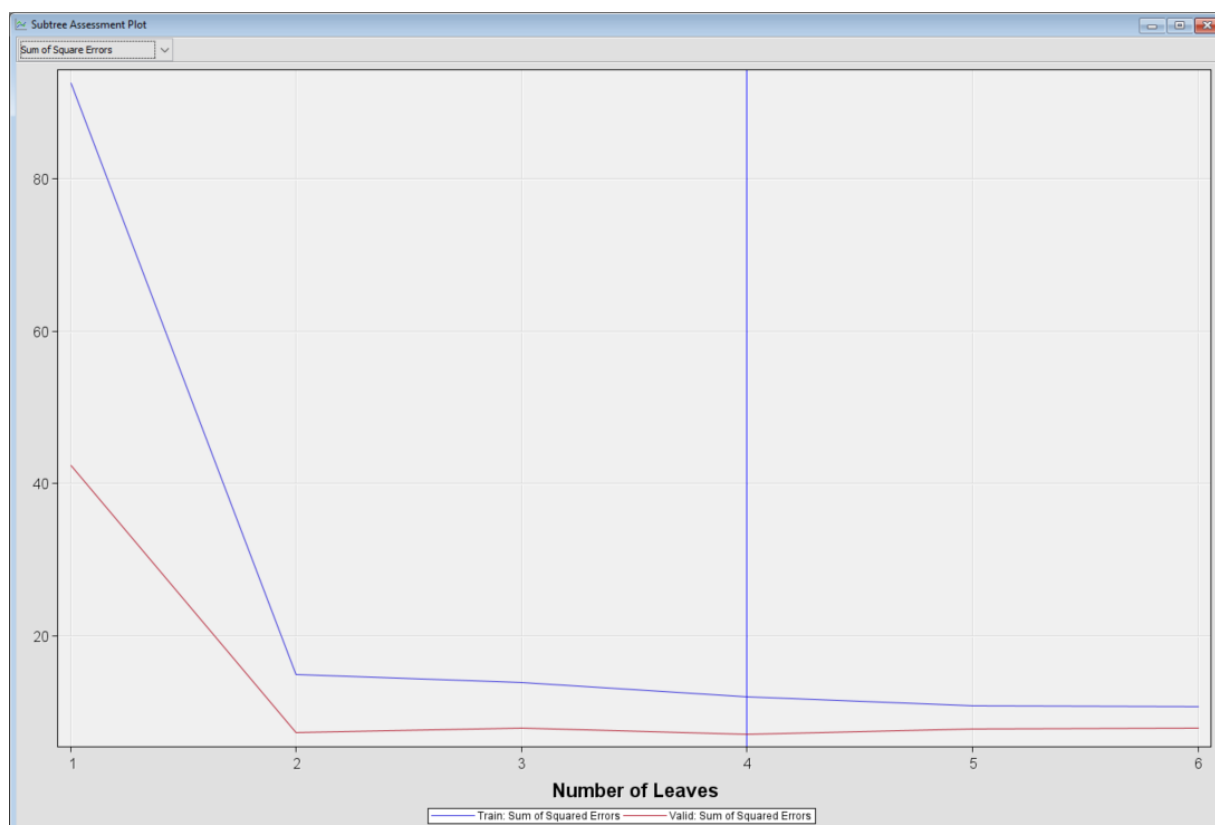
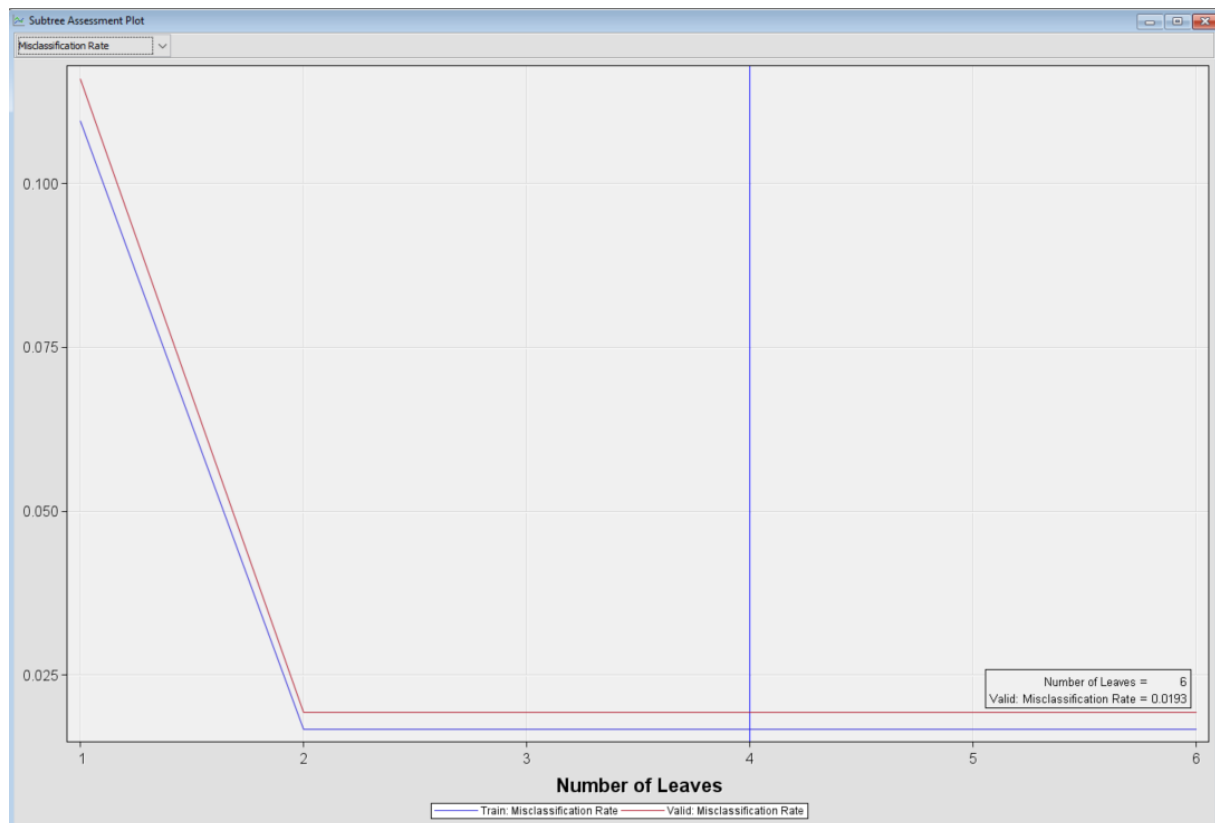


Statistics	Train	Validation
Count	475	207
Prediction	0	0
% with target = 0	89.05%	88.41%
% with target = 1	10.95%	11.59%
% correctly predicted	89.05%	88.41%

Drzewo decyzyjne 2:

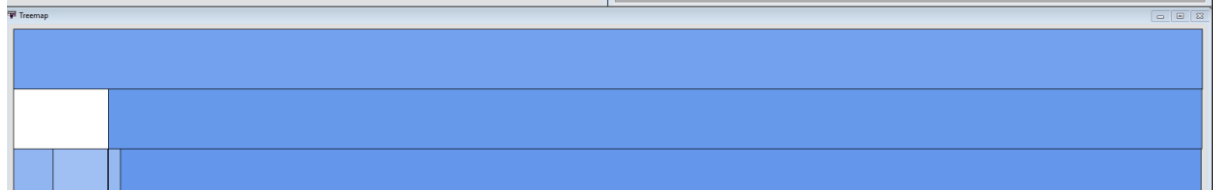
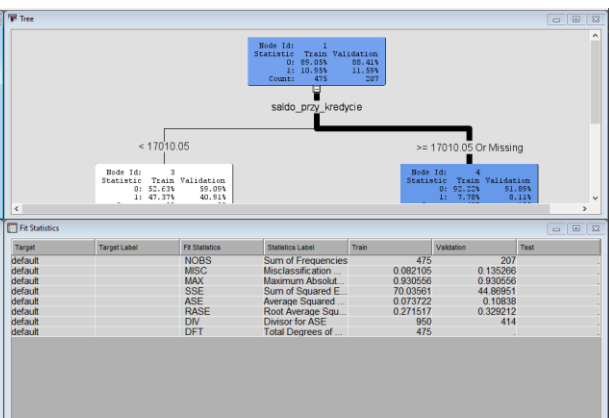
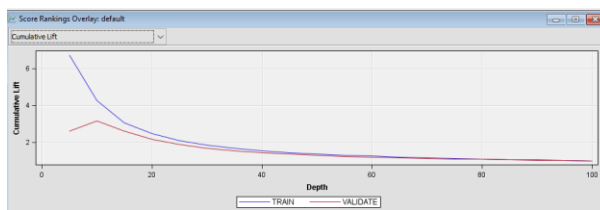
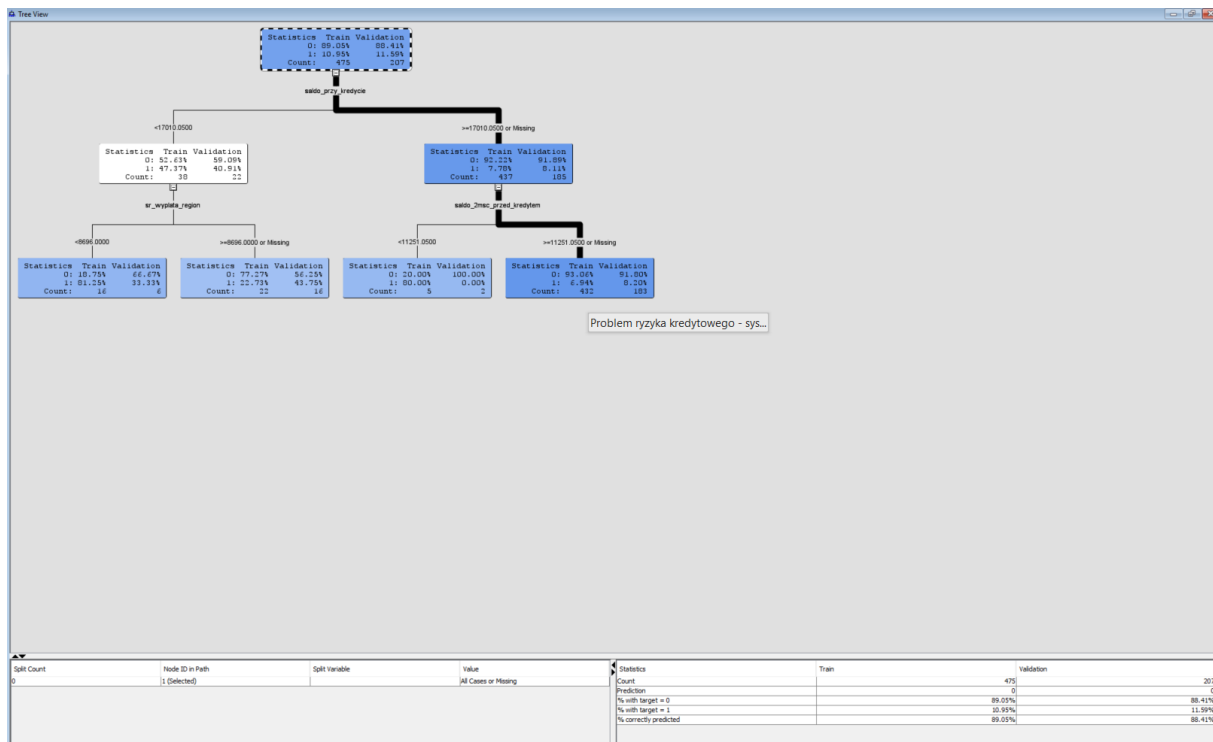
General	
Node ID	Tree3
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Interactive	...
Import Tree Model	No
Tree Model Data Set	...
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	6
Minimum Categorical Size	5
Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Average Square Error
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Bootstrap	...



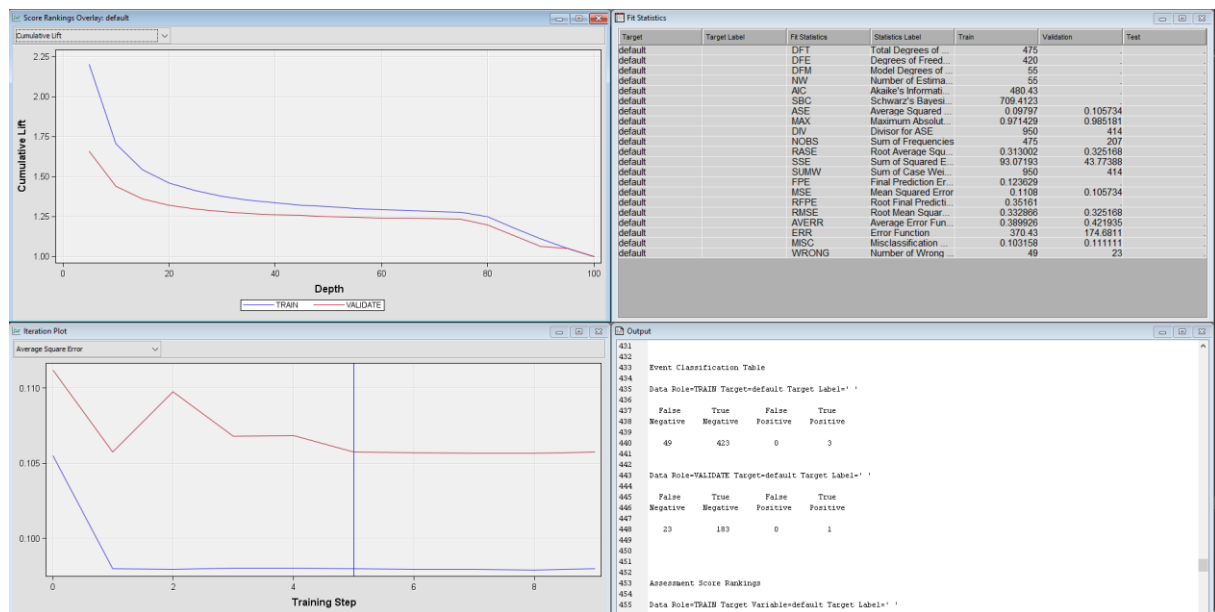


Drzewo decyzyjne 3:






## Sieć neuronowa:



<b>General</b>	
Node ID	AutoNeural
Imported Data	***
Exported Data	***
Notes	***
<b>Train</b>	
Variables	***
<b>Model Options</b>	
Architecture	Single Layer
Termination	Overfitting
Train Action	Search
Target Layer Error Function	Default
Maximum Iterations	8
Number of Hidden Units	2
Tolerance	Medium
Total Time	One Hour
<b>Increment and Search Opti</b>	
Adjust Iterations	Yes
Freeze Connections	No
Total Number of Hidden Uni	30
Final Training	Yes
Final Iterations	5
<b>Activation Functions</b>	
Direct	Yes
Exponential	No
Identity	No
Logistic	No
Normal	Yes
Reciprocal	No
Sine	Yes
Softmax	No
Square	No
Tanh	Yes
<b>Score</b>	
Hidden Units	No
Residuals	Yes
Standardization	No
<b>Status</b>	

Regresja:

 Selection Options ✕

Property	Value
Sequential Order	No
Entry Significance Level	0.1
Stay Significance Level	0.05
Start Variable Number	0
Stop Variable Number	0
Force Candidate Effects	0
Hierarchy Effects	Class
Moving Effect Rule	None
Maximum Number of Steps	0

**Entry Significance Level**

Significance level for adding variables in forward or stepwise regression.

**General**

Node ID Reg

Imported Data

Exported Data

Notes

**Train**

Variables

**Equation**

Main Effects Yes

Two-Factor Interactions No

Polynomial Terms No

Polynomial Degree 2

User Terms No

Term Editor

**Class Targets**

Regression Type Logistic Regression

Link Function Logit

**Model Options**

Suppress Intercept No

Input Coding Deviation

**Model Selection**

Selection Model Forward

Selection Criterion Default

Use Selection Defaults No

Selection Options

**Optimization Options**

Technique Default

Default Optimization Yes

Max Iterations 0

Max Function Calls 0

Maximum Time 1 Hour

**Convergence Criteria**

Uses Defaults Yes

Options

**Output Options**

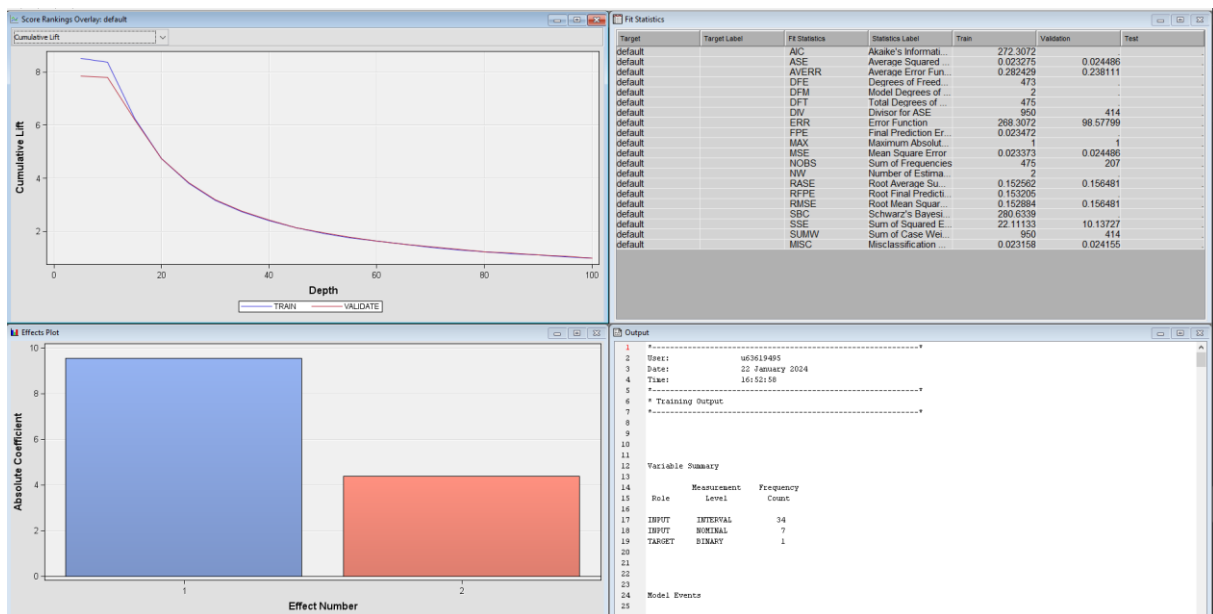
Confidence Limits No

Save Covariance No

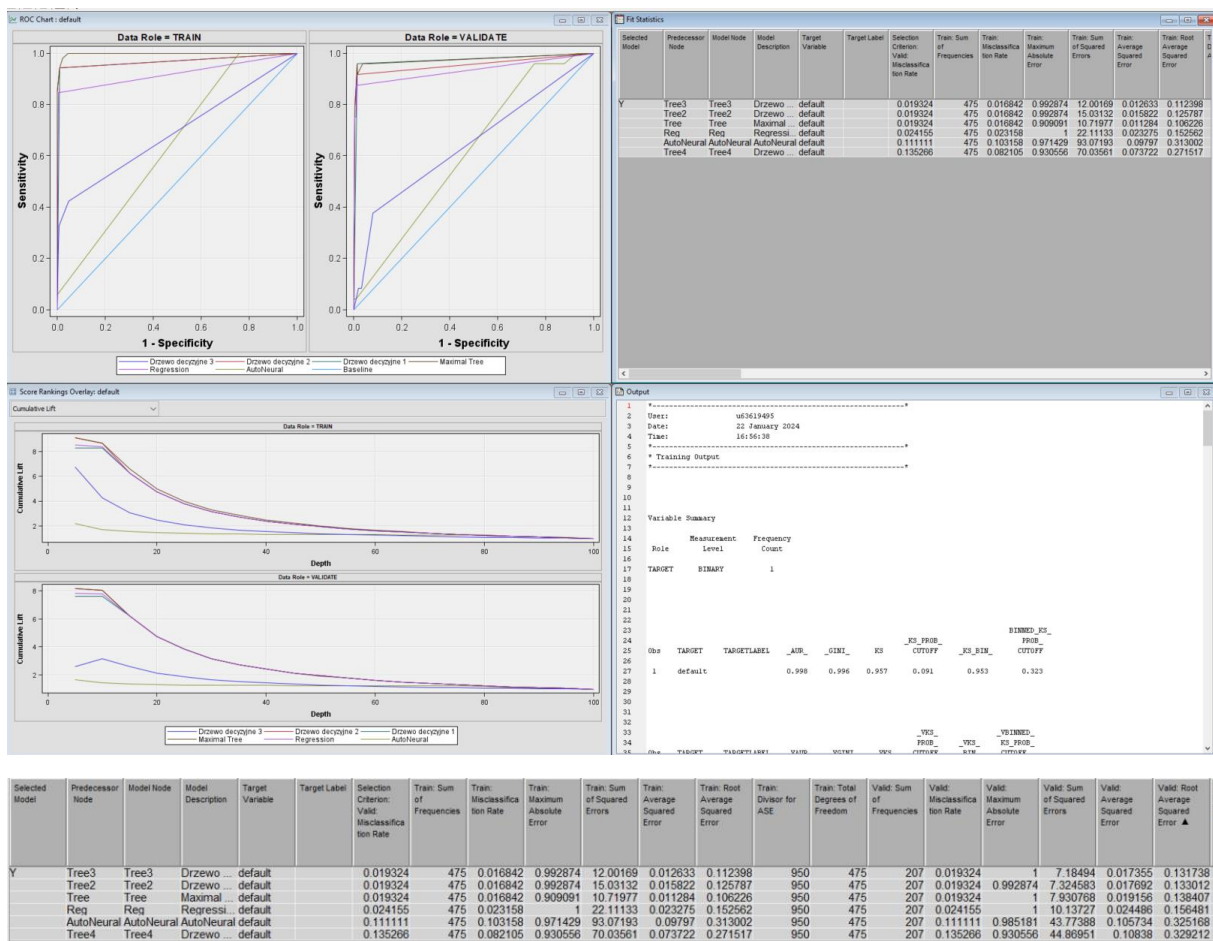
Covariance No

Correlation No

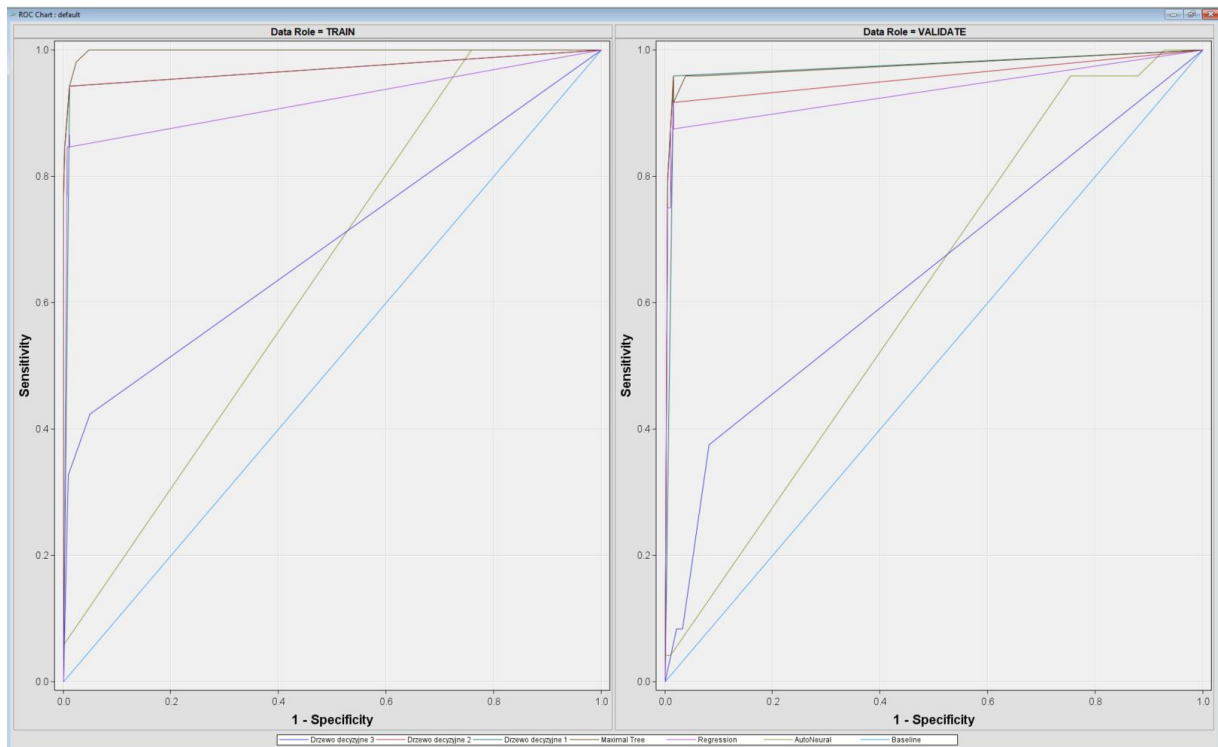
Statistics No



Podsumowanie:



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error
Y						Target Label											
	Tree3	Tree3	Drzewo ...	default		0.019324	475	0.016842	0.992874	12.00169	0.012633	0.112398	950	475	207	0.019324	
	Tree2	Tree2	Drzewo ...	default		0.019324	475	0.016842	0.992874	15.03132	0.015822	0.125787	950	475	207	0.019324	
	Tree	Tree	Maximal ...	default		0.019324	475	0.016842	0.909091	10.71977	0.011284	0.106226	950	475	207	0.019324	
	Reg	Reg	Regressi...	default		0.024155	475	0.023158	1	22.11133	0.023275	0.152562	950	475	207	0.024155	
	AutoNeural	AutoNeural	AutoNeural	default		0.111111	475	0.103158	0.971429	93.07193	0.09797	0.313002	950	475	207	0.111111	
	Tree4	Tree4	Drzewo ...	default		0.135286	475	0.082105	0.930556	70.03561	0.073722	0.271517	950	475	207	0.135286	



#### Event Classification Table

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Model Node	Model Description	Data Role	Target	Target Label	False Negative	True Negative	False Positive	True Positive
Tree	Maximal Tree	TRAIN	default		3	418	5	49
Tree	Maximal Tree	VALIDATE	default		1	180	3	23
Tree2	Drzewo decyzyjne 1	TRAIN	default		3	418	5	49
Tree2	Drzewo decyzyjne 1	VALIDATE	default		1	180	3	23
Tree3	Drzewo decyzyjne 2	TRAIN	default		3	418	5	49
Tree3	Drzewo decyzyjne 2	VALIDATE	default		1	180	3	23
Tree4	Drzewo decyzyjne 3	TRAIN	default		35	419	4	17
Tree4	Drzewo decyzyjne 3	VALIDATE	default		22	177	6	2
AutoNeural	AutoNeural	TRAIN	default		49	423	0	3
AutoNeural	AutoNeural	VALIDATE	default		23	183	0	1
Reg	Regression	TRAIN	default		8	420	3	44
Reg	Regression	VALIDATE	default		3	181	2	21