

# Vývojová zpráva ke scoringové funkci xxx

*Jméno příjmení*

xx.xx.2009

## 1) Zdroj dat

Aplikační data žadatelů o úvěr. Zdroj: CD přiložené k publikaci Thomas, L.C., Edelman, D.B., Crook, J.N. (2002). *Credit Scoring and Its Applications*, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia.

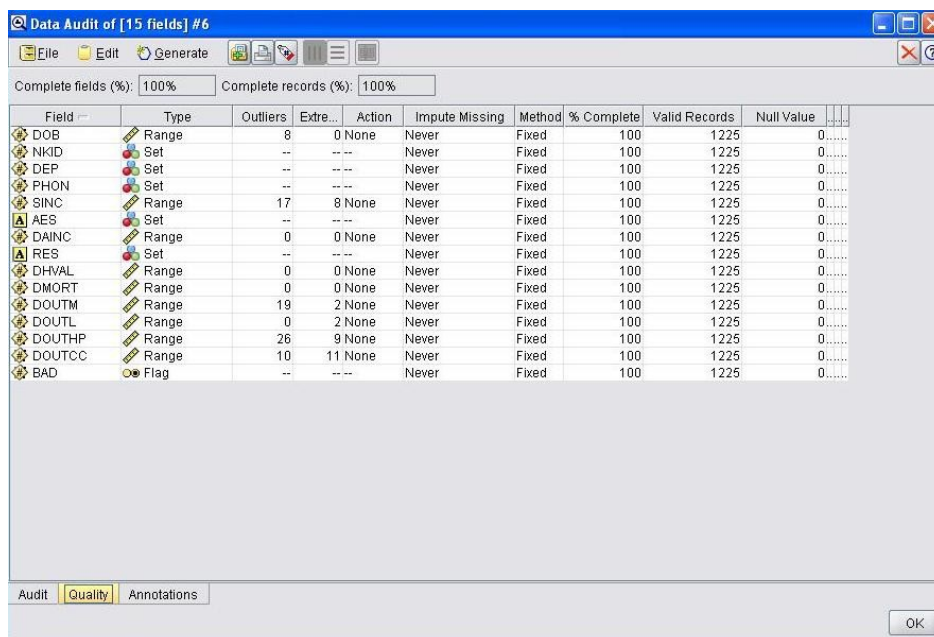
## 2) Popis a analýza dat

Variable Name	Description	Codings
dob	Year of birth	If unknown the year will be 99
nkid	Number of children	number
dep	Number of other dependents	number
phon	Is there a home phone	1=yes, 0 = no
sinc	Spouse's income	
aes	Applicant's employment status	V = Government W = housewife M = military P = private sector B = public sector R = retired E = self employed T = student U = unemployed N = others Z = no response
dainc	Applicant's income	
res	Residential status	O = Owner F = tenant furnished U = Tenant Unfurnished P = With parents N = Other Z = No response
dhval	Value of Home	0 = no response or not owner 000001 = zero value blank = no response
dmort	Mortgage balance outstanding	0 = no response or not owner 000001 = zero balance blank = no response
doutm	Outgoings on mortgage or rent	
doutl	Outgoings on Loans	
douthp	Outgoings on Hire Purchase	
doutcc	Outgoings on credit cards	
Bad	Good/bad indicator	1 = Bad 0 = Good

Datový soubor obsahoval 14 vysvětlujících proměnných a 1 cílovou proměnnou. Rozsah souboru byl 1225 řádků. Základní popis je dán následujícími tabulkami:

### Data Audit of [15 fields]

	Field	Graph	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	DOB		range	3.00 0	99.000	51.038	15.302	-0.714	--	1225
2	NKID		set	0.00 0	5.000	--	--	--	6	1225
3	DEP		set	0.00 0	2.000	--	--	--	3	1225
4	PHON		set	0.00 0	1.000	--	--	--	2	1225
5	SINC		range	0.00 0	50000.00 0	1990.085	4802.341	3.788	--	1225
6	AES		set	--	--	--	--	--	11	1225
7	DAINC		range	0.00 0	64800.00 0	21244.21 1	15896.20 8	0.502	--	1225
8	RES		set	--	--	--	--	--	5	1225
9	DHVAL		range	0.00 0	64928.00 0	15693.85 8	20736.33 2	1.000	--	1225
10	DMORT		range	0.00 0	64000.00 0	11225.69 1	18889.20 7	1.468	--	1225
11	DOUTM		range	0.00 0	3800.000	342.005	427.994	1.942	--	1225
12	DOUTL		range	0.00 0	28000.00 0	121.927	839.640	30.186	--	1225
13	DOUTH P		range	0.00 0	1600.000	28.722	119.324	6.572	--	1225
14	DOUTC C		range	0.00 0	2800.000	39.595	168.697	8.958	--	1225
15	BAD		flag	0.00 0	1.000	--	--	--	2	1225



### 3) Transformace dat

V průběhu přípravy dat byly použity tyto datové transformace:

Odstranění odlehlých hodnot / doplnění chybějících / diskretizace / agregace / kombinace proměnných / WOE / jiné...

Příklad tabulky rozdělení četností+def.rate+WOE ...tabulky rozdělení četností pro diskretní proměnné se je vhodné uvést už v bodě 2).

	Data			
RES	Počet z BAD	Počet z BAD2		
F	129	10,53%		
N	66	5,39%		
O	624	50,94%		
P	252	20,57%		
U	154	12,57%		
Celkový součet	1225	100,00%		
	Data			
RES	Počet z ID	Součet z BAD	def_rate	WOE
F	129	30	0,233	0,0725
N	66	31	0,470	-0,3933
O	624	164	0,263	0,0019
P	252	60	0,238	0,0591
U	154	38	0,247	0,0387
Celkový součet	1225	323	0,264	

## 4) Modelování

Byly testovány tyto modely....seznam všech modelů a jejich stručný popis

Např....

### Logistická regrese:

Fields

Target

BAD

Inputs

DOB

NKID

DEP

PHON

SINC

AES

DAINC

RES

DHVAL

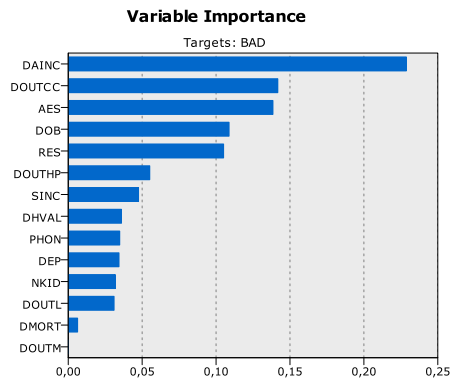
DMORT

DOUTM

DOUTL

DOUTH

DOUTCC



Build Settings

Use partitioned data: false

Calculate variable importance: false

Calculate raw propensity scores: false

Procedure: Multinomial

Base category:

Model type: Main Effects

Include constant in equation: true

Mode: Simple

Multinomial Method: Enter

Output =  $-0,01307 * DOB + -0,00004683 * SINC + -0,00002745 * DAINC +$   
 $-0,000004588 * DHVAL + 0,000002847 * DMORT + 0,0002189 * DOUTM +$   
 $0,00007665 * DOUTL + -0,001382 * DOUTH + -0,002207 * DOUTCC +$   
 $18,56 * [NKID=0] + 18,63 * [NKID=1] + 18,58 * [NKID=2] + 18,8 * [NKID=3] +$   
 $17,42 * [NKID=4] + -0,9928 * [DEP=0] + -0,9972 * [DEP=1] + 0,1379 * [PHON=0] +$   
 $0,1932 * [AES=B] + 0,007235 * [AES=E] + -0,01056 * [AES=M] +$   
 $-0,4939 * [AES=N] + -0,1896 * [AES=P] + 0,06029 * [AES=R] + -0,1853 * [AES=T] +$   
 $0,5025 * [AES=U] + -0,3596 * [AES=V] + 0,3941 * [AES=W] + 0,04138 * [RES=F] +$   
 $0,9657 * [RES=N] + 0,4718 * [RES=O] + 0,1391 * [RES=P] + -17,49$

# Strom - CART:

## Fields

Target

BAD

Inputs

DAINC

RES

DMORT

DOU TL

DOUTC C

## Build Settings

Levels below root: 5

DAINC <= 0 [ Mode: 0 ] => 0,0

DAINC > 0 and DAINC <= 34 500 [ Mode: 0 ]

RES in [ "F" "P" "U" ] [ Mode: 0 ] => 0,0

RES in [ "N" "O" ] [ Mode: 0 ]

DMORT <= 0 [ Mode: 0 ]

DAINC <= 12 000 [ Mode: 1 ] => 1,0

DAINC > 12 000 [ Mode: 0 ] => 0,0

DMORT > 0 [ Mode: 0 ]

DOUTC C <= 0 [ Mode: 0 ]

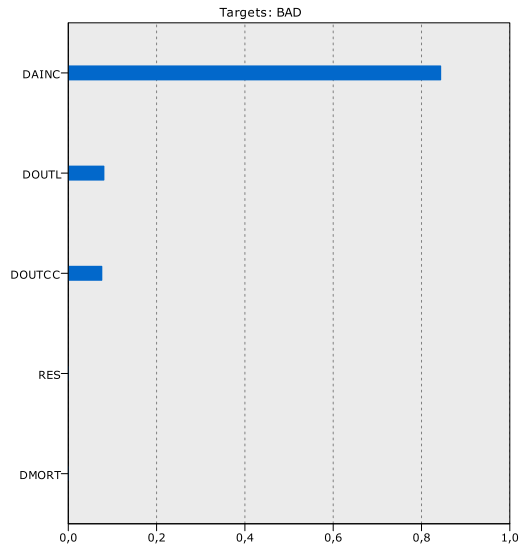
DOU TL <= 388 [ Mode: 0 ] => 0,0

DOU TL > 388 [ Mode: 1 ] => 1,0

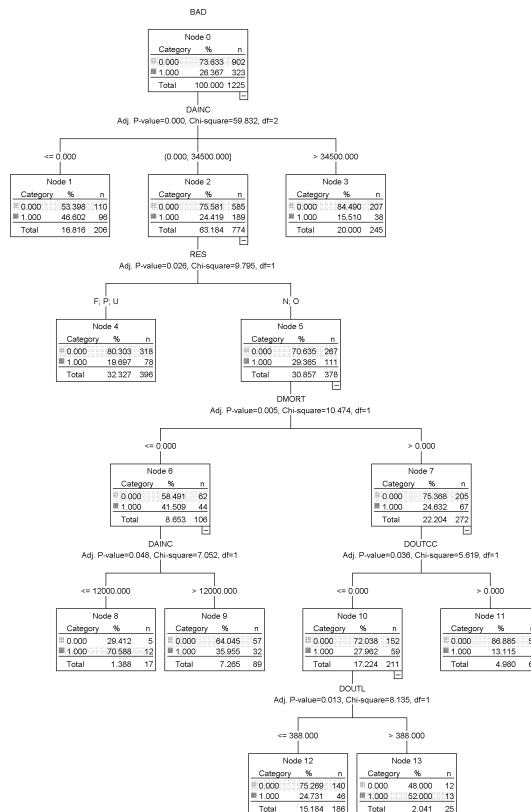
DOUTC C > 0 [ Mode: 0 ] => 0,0

DAINC > 34 500 [ Mode: 0 ] => 0,0

## Variable Importance



1:57 odp., Út. XI 17, 2009



## Neuronová síť:

### Analysis

Input Layer: 36 neurons

Hidden Layer 1: 3 neurons

Output Layer: 1 neurons

### Fields

#### Target

BAD

#### Inputs

AES

DAINC

DEP

DHVAL

DMORT

DOB

DOUTCC

DOUTHHP

DOUTL

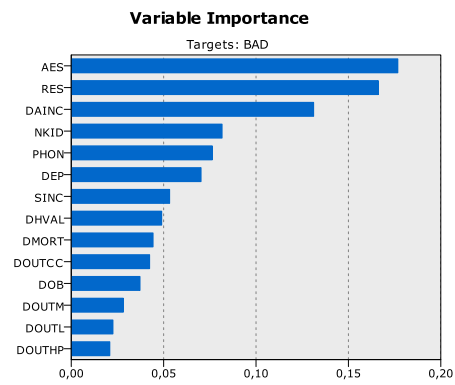
DOUTM

NKID

PHON

RES

SINC



# Analýza kvality:

Analysis

## Results for output field BAD

### Individual Models

#### Comparing \$R-BAD with BAD

Correct	910	74,29%
Wrong	315	25,71%
Total	1 225	

#### Coincidence Matrix for \$R-BAD (rows show actuals)

	0.000000	1.000000
0.000000	885	17
1.000000	298	25

#### Performance Evaluation

0.000000	0,016
1.000000	0,814

#### Confidence Values Report for \$RC-BAD

Range	0,519 - 0,857
Mean Correct	0,758
Mean Incorrect	0,689
Always Correct Above	0,857 (0% of cases)
Always Incorrect Below	0,519 (0% of cases)
90% Accuracy Above	Never reached requested level
2,0 Fold Correct Above	Never reached requested level

#### Comparing \$L-BAD with BAD

Correct	916	74,78%
Wrong	309	25,22%
Total	1 225	

#### Coincidence Matrix for \$L-BAD (rows show actuals)

	0.000000	1.000000
0.000000	868	34
1.000000	275	48

#### Performance Evaluation

0.000000	0,031
1.000000	0,798

#### Confidence Values Report for \$LP-BAD

Range	0,5 - 1,0
Mean Correct	0,764
Mean Incorrect	0,702
Always Correct Above	0,908 (5,39% of cases)
Always Incorrect Below	0,5 (0,08% of cases)
90,29% Accuracy Above	0,855
2,0 Fold Correct Above	0,876 (84,55% of cases)

#### Comparing \$N-BAD with BAD

Correct	914	74,61%
Wrong	311	25,39%
Total	1 225	

#### Coincidence Matrix for \$N-BAD (rows show actuals)

	0.000000	1.000000
0.000000	880	22
1.000000	289	34

#### Performance Evaluation

0.000000	0,022
1.000000	0,834

#### Confidence Values Report for \$NC-BAD

Range	0,017 - 0,85
Mean Correct	0,576
Mean Incorrect	0,496
Always Correct Above	0,845 (0,49% of cases)
Always Incorrect Below	0,018 (0,08% of cases)
91,49% Accuracy Above	0,822
2,0 Fold Correct Above	0,876 (79,86% of cases)

#### Agreement between \$R-BAD \$L-BAD \$N-BAD

Agree	1 102	89,96%
Disagree	123	10,04%
Total	1 225	

#### Comparing Agreement with BAD

## 5) Popis scoringové funkce xxx

Výsledná scoringová funkce je popsána násl. tabulkou (proměnné, koeficienty, signifikance)... např.:

		Parameter Estimates					95% Confidence Interval for Exp(B)		
BAD(a)		B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
		Intercept	-17,489	1,400	155,950	1	,000		
	DOB	,013	,007	3,872	1	,049	,987	,974	1,000
	SINC	,000	,000	6,627	1	,010	1,000	1,000	1,000
	DAINC	,000	,000	20,826	1	,000	1,000	1,000	1,000
	DHVAL	,000	,000	,906	1	,341	1,000	1,000	1,000
	DMORT	,000	,000	,317	1	,574	1,000	1,000	1,000
	DOUTM	,000	,000	1,122	1	,289	1,000	1,000	1,001
	DOUTL	,000	,000	,385	1	,535	1,000	1,000	1,000
	DOUTH	-,001	,001	2,511	1	,113	,999	,997	1,000
	DOUTCC	-,002	,001	5,343	1	,021	,998	,996	1,000
	[NKID=,000]	18,560	,778	568,742	1	,000	114947117,230	25006290,618	528380636,752
	[NKID=1,000]	18,630	,795	549,317	1	,000	123240531,758	25951576,580	585252638,555
	[NKID=2,000]	18,581	,789	554,119	1	,000	117385860,648	24987575,122	551451679,995
	[NKID=3,000]	18,802	,835	506,816	1	,000	146376912,596	28482598,401	752255824,393
	[NKID=4,000]	17,421	,000	.	1	.	36802358,349	36802358,349	36802358,349
	[NKID=5,000]	0(b)	.	.	0	.	.	.	.
	[DEP=,000]	-,993	,800	1,538	1	,215	,371	,077	1,779
	[DEP=1,000]	-,997	,911	1,199	1	,273	,369	,062	2,198
1.0	[DEP=2,000]	0(b)	.	.	0	.	.	.	.
	[PHON=,000]	,138	,233	,351	1	,553	1,148	,728	1,811
	[PHON=1,000]	0(b)	.	.	0	.	.	.	.
	[AES=B]	,193	,897	,046	1	,829	1,213	,209	7,037
	[AES=E]	,007	,805	,000	1	,993	1,007	,208	4,879
	[AES=M]	-,011	,919	,000	1	,991	,989	,163	5,997
	[AES=N]	-,494	1,374	,129	1	,719	,610	,041	9,012
	[AES=P]	-,190	,785	,058	1	,809	,827	,178	3,854
	[AES=R]	,060	,820	,005	1	,941	1,062	,213	5,295
	[AES=T]	-,185	,804	,053	1	,818	,831	,172	4,018
	[AES=U]	,503	1,063	,224	1	,636	1,653	,206	13,275
	[AES=V]	-,360	,797	,204	1	,652	,698	,146	3,328
	[AES=W]	,394	,869	,206	1	,650	1,483	,270	8,148
	[AES=Z]	0(b)	.	.	0	.	.	.	.
	[RES=F]	,041	,313	,018	1	,895	1,042	,565	1,924
	[RES=N]	,966	,346	7,771	1	,005	2,627	1,332	5,179
	[RES=O]	,472	,270	3,047	1	,081	1,603	,944	2,722
	[RES=P]	,139	,294	,224	1	,636	1,149	,646	2,044
	[RES=U]	0(b)	.	.	0	.	.	.	.

a. The reference category is: 0.0.

b. This parameter is set to zero because it is redundant.

## 6) Evaluace

Podrobné charakteristiky kvality... grafy/tabulky (LC, gini, ks, lift, ...)  
...na vývojovém vzorku (popř. na testovacím).