



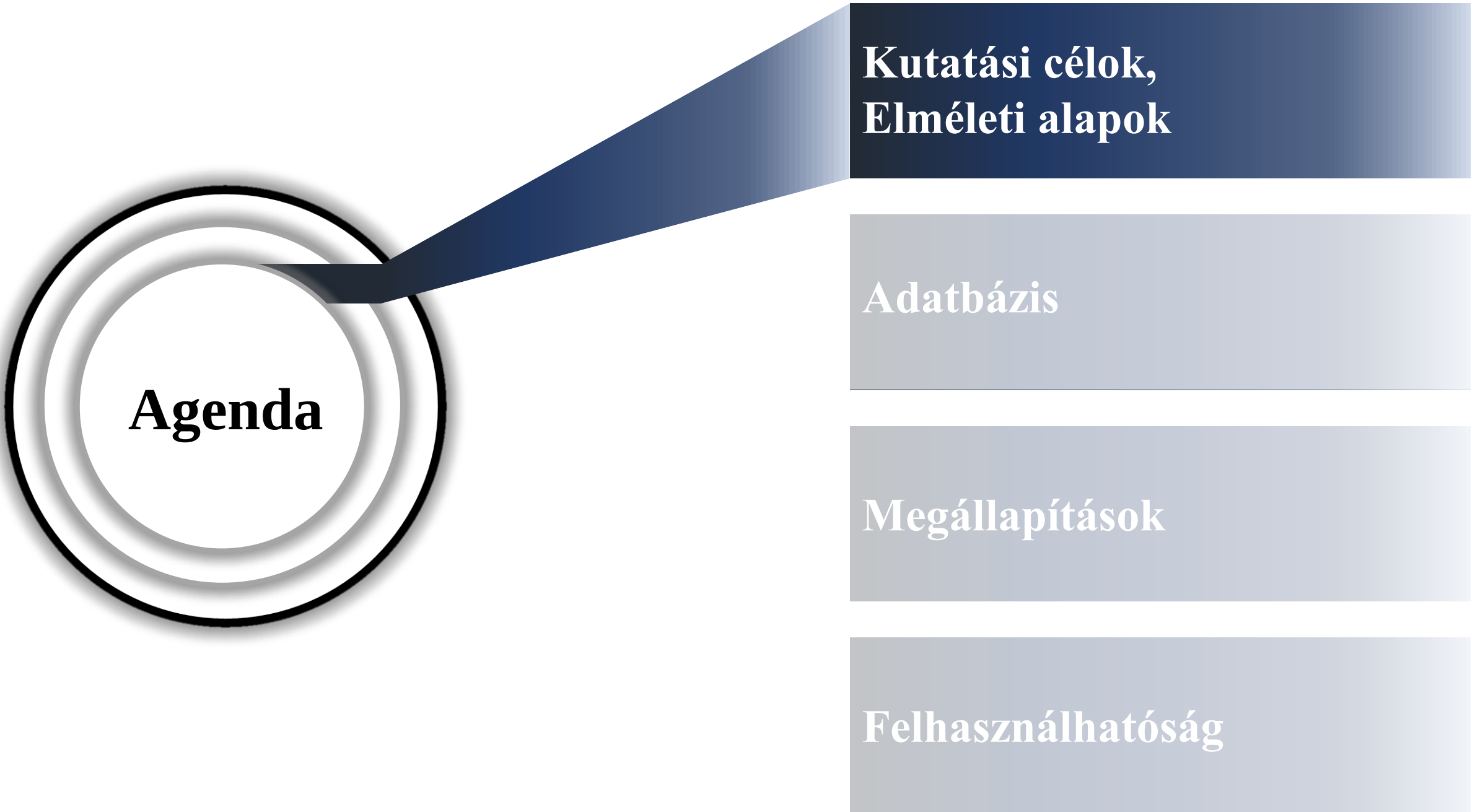
A lakossági hitelek nemfizetésének előrejelzése többváltozós statisztikai módszerekkel

Készítette:

Dr. Fodor Kitti

Budapest, 2024.02.22.

Tudományos vezető: Dr. Varga Beatrix Margit



Agenda

**Kutatási célok,
Elméleti alapok**

Adatbázis

Megállapítások

Felhasználhatóság

Témaválasztás, kutatási célok

Alacsony kockázat a nemfizetésre

Saját hitelbírálat

Nehéz beazonosítani

CÉL

Lakossági hitelek bedőlésének vizsgálata többváltozós statisztikai eljárásokon alapuló módszerek segítségével,

Annak a feltérképezése, hogy milyen tényezők járulnak hozzá ahhoz, hogy valakiből nemfizető adós váljon,

Mely többváltozós statisztikai módszer produkálja a legjobb eredményt.

A kutatás elméleti pillérei

Hitelezés története

évezredek óta az emberi kultúra része

Magyar hitelkörkép, eladósodottság

A magyar lakosság fogyasztói hiteleinek alakulása 2001-2021 (Mrd Ft)



Egyéb, hitelezéssel összefüggő, nehezen mérhető tényezők



Összességében megállapítható, hogy a különböző elemzési módszerek közül alapvetően a neurális hálók produkálták a legjobb eredményeket.

Nemzetközi és magyar csődmodell-kutatások

az első próbálkozások az 1930-as évekre

Altman, 5 mutatószám, 95%

Beaver, egyváltozós, 158 vállalat, 90%

Deakin, DA, 34 elemű mintán tesztelte, 97%

Altman (1977), 7 mutatószám, 111 elem (58 túlélő), 95%

Altman: nem reprezentatív, megközelítőleg azonos arány

Ohlson: első reprezentatív, logisztikus regresszió, 2163 vállalat, 4,85% csődös, $p=0,038$, 83%

Frydman: RPA, 80-as évek, 94%, jelentős különbség a túlélő és csődbe ment vállalatok kategorizálásában.

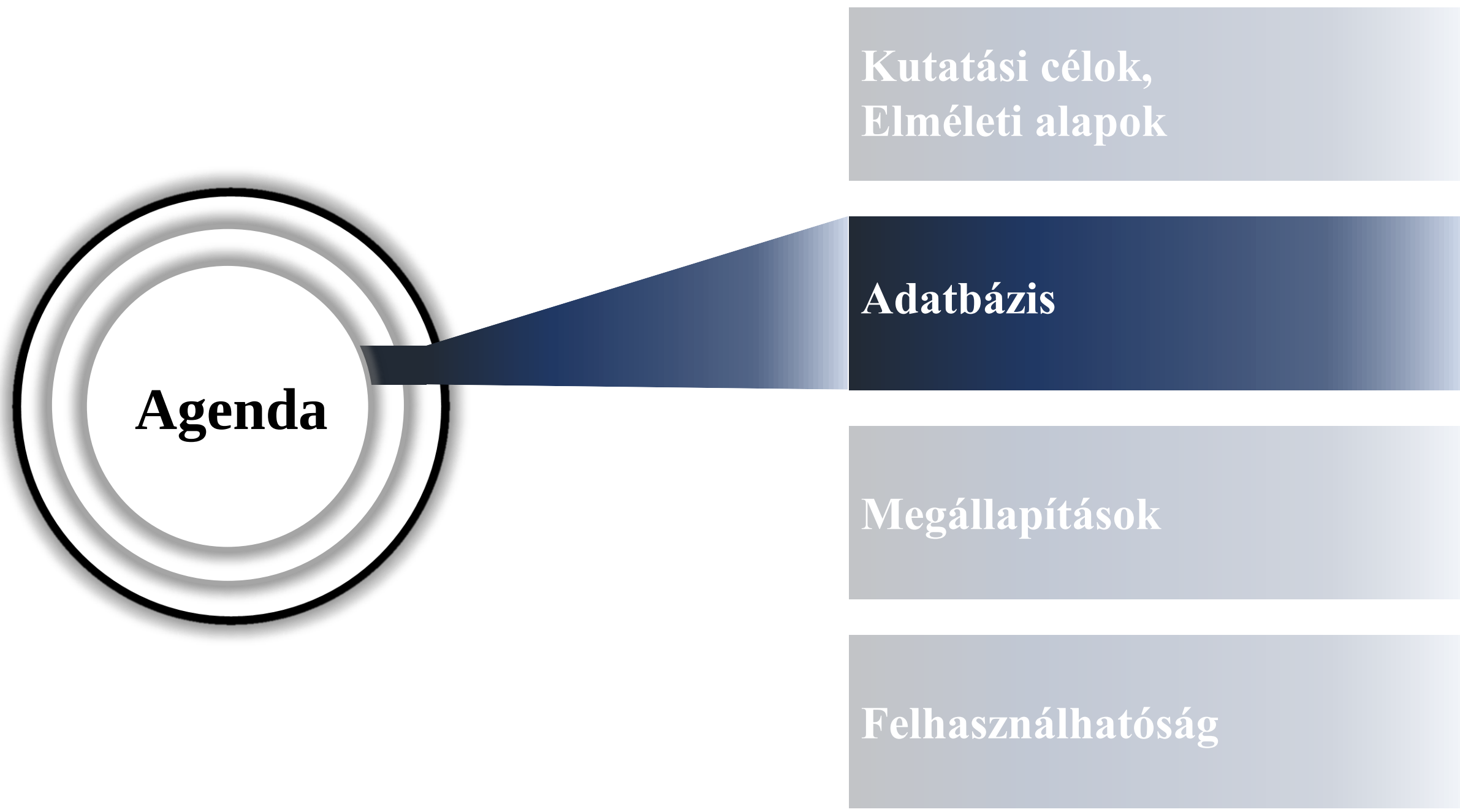
Odom & Sharda: NN, 90-es évek, 100% és 82%

Mo: 90-es években kezdődtek a kutatások

Virág Miklós és Hajdú Ottó, DA (77,92%) és log. reg. (81,8%), 154 vállalat, 50-50%,

Később NN, 156 elem, 75% tanuló, 2 rétegű háló, jobb eredmény, mint DA és log.

DA, logisztikus regresszió, probit, RPA, neurális háló



Agenda

**Kutatási célok,
Elméleti alapok**

Adatbázis

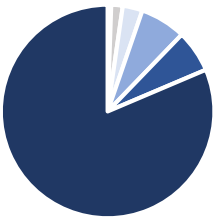
Megállapítások

Felhasználhatóság

KHR

- teljes lista
- Segíti a bankok közötti információáramlást
- Segítséget nyújt a hitelbírálat során
- Csökkenti a túlzott eladósodottság kockázatát

A fennálló mulasztások megoszlása a mulasztás időtartama szerint (%; 2021)



- 1-30 nap
- 31-90 nap
- 91-180 nap
- 181-360 nap
- 361-720 nap
- 720+ nap

2021. szeptember 30.

1 unit = 1 hitelügylet

Rendelkezésre álló változók:

- az ügyfél anonim azonosítója
- az ügyféli minőség (adós vagy adóstárs)
- az ügyfél kora: az adatbázis létrejöttkor az ügyfél kora,
- az ügyfél neve
- Szerződés torzított azonosítója:
- Szerződés típusa: hiteltípusok
- Szerződés státusza (lezárt vagy fennálló)
- A szerződés megkötésének negyedéve
- A szerződés lejáratának negyedéve
- A szerződéses összeg és devizaneme
- A tőketartozás összege és devizaneme
- A törlesztési összeg és devizaneme

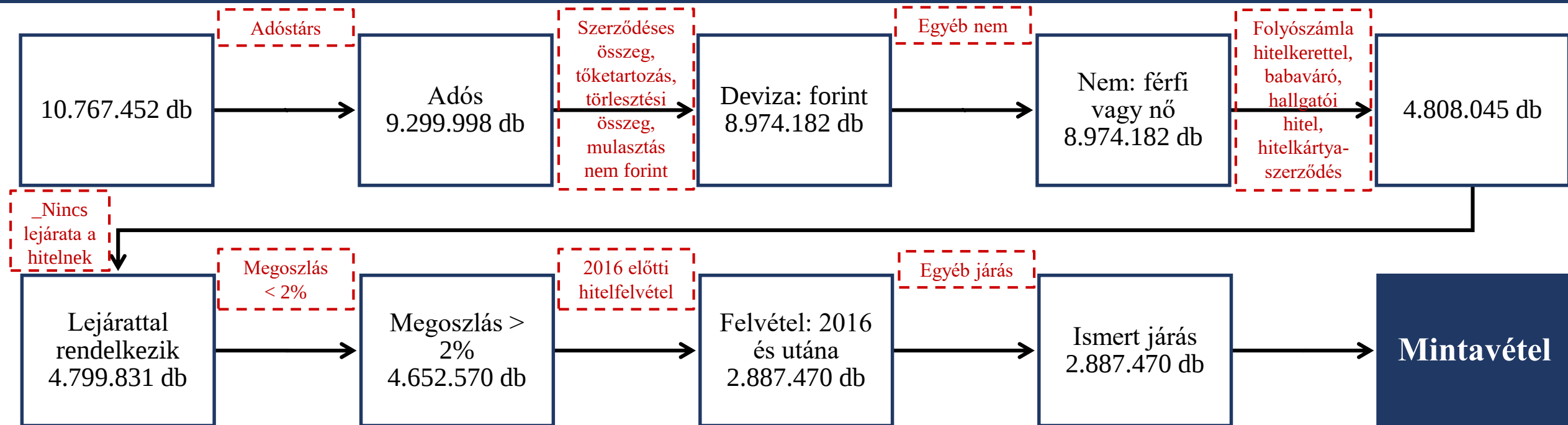
- A mulasztás összege, amennyiben van
- A mulasztás státusza (lezárt vagy fennálló)
- A mulasztás összegének devizaneme
- A mulasztás kezdetének negyedéve
- A mulasztás lezárásának negyedéve
- Járás

A fenti változókon túlmenően az alábbiakkal bővítettem az adatbázist:

- Futamidő
- Mulasztással rendelkező vagy nem rendelkező ügyletről van-e szó
- Megye és régió
- Átlagkereset (eFt):

- Munkanélküliségi ráta
- Foglalkoztatási ráta
- 1 főre jutó GDP
- Érettségizők aránya
- Az átlagkereset szerződéses összeghez viszonyított aránya
- Az átlagkereset törlesztési összeghez viszonyított aránya
- Törlesztési összeg szerződéses összeghez viszonyított aránya
- Az ügyfél hitelfelvételkori életkora

Az adatbázis szűkítése



Nemteljesítőnek azt a hitelügyletet minősítettem, mely rendelkezett mulasztási összeggel.

Minta 1.

n=500



**Tanuló és tesztelő
minta: 70-30%**

Minta 2.

n=500





Agenda

Kutatási célok,
Elméleti alapok

Adatbázis

Megállapítások

Felhasználhatóság

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Logisztikus regresszió I.

Minta 1



Feltételezések

KHR által
nyilvántartott
adatok

n=500

70%

30%

Omnibus teszt: $p < 0,001$ ✓

Hosmer and Lemeshow goodness-of-fit teszt ($p = 0,212$) ✓

Nagelkerke $R^2 = 38,8\%$

$$P_{(default)} = \frac{e^{0,032 + 1,030x_1}}{1 + e^{0,032 + 1,030x_1}}$$

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya

Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	138	4	97,2
		1	4	4	50,0
	Overall Percentage				94,7
Tanuló	Default	0	261	62	80,8
		1	4	23	85,2
	Overall Percentage				81,1

a. The cut value is ,039

Magas besorolási pontosság

Jelentős különbség a szenzitivitás és a specificitás között a teszt minta esetében



A jól teljesítő hitelek kategorizálása nagyobb arányban helyes



H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Döntési fa I.

Minta 1



Feltételezések

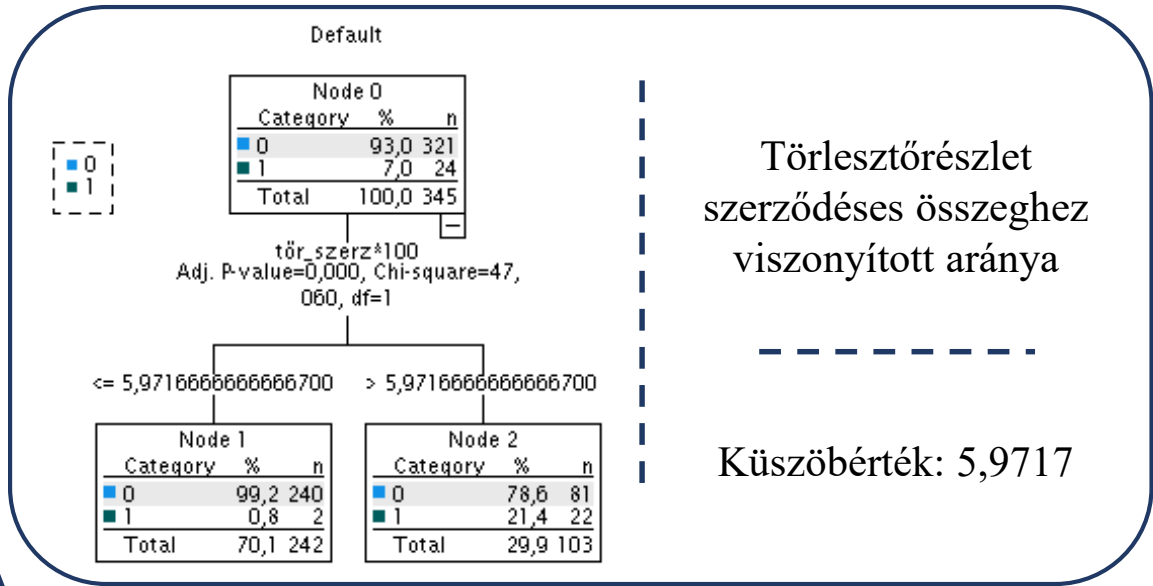
KHR által
nyilvántartott
adatok

n=500

70%

30%

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	312	0	100,0%
	1	24	0	0,0%
	Overall Percentage	100,0 %	0,0%	93,0%
Teszt	0	144	0	100,0%
	1	11	0	0,0%
	Overall Percentage	100,0 %	0,0%	92,9%



Magas besorolási pontosság

Jelentős különbség a szenzitivitás és a specificitás között



A nemteljesítő hitelek nem tudja helyesen kategorizálni.

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Neurális háló I.

Minta 1



Feltételezések

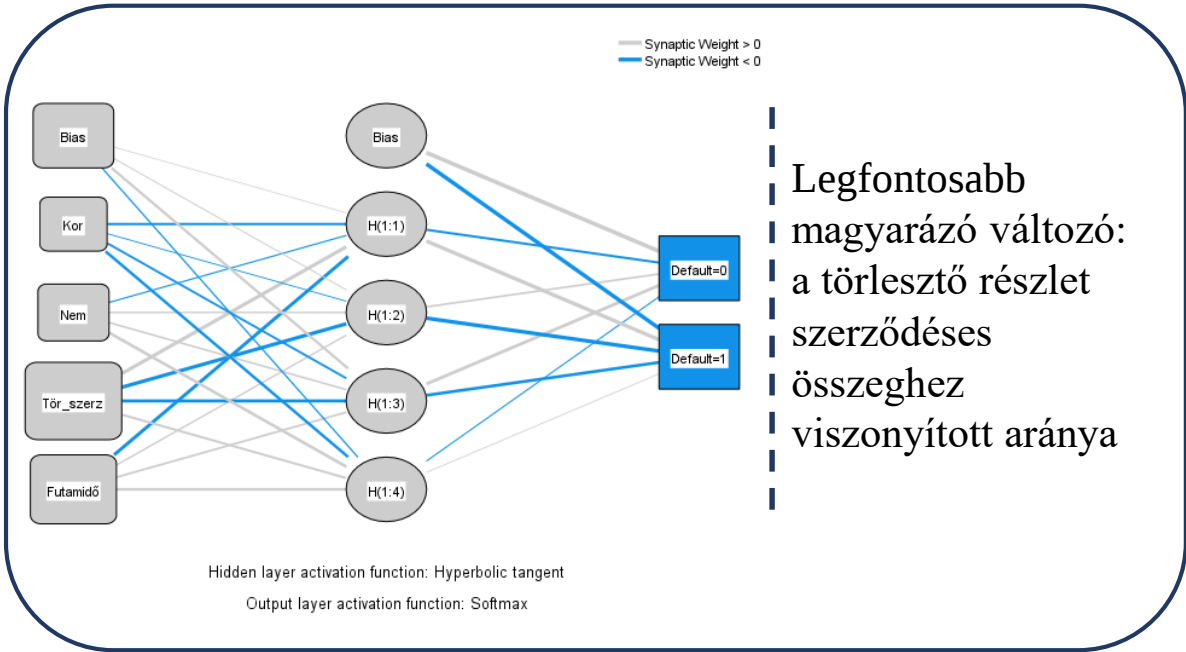
KHR által
nyilvántartott
adatok

n=500

70%

30%

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	307	9	97,2%
	1	14	12	46,2%
	Overall Percent	93,9%	6,1%	93,3%
Teszt	0	149	0	100,0%
	1	5	4	44,4%
	Overall Percent	97,5%	2,5%	96,8%



Magas besorolási pontosság

Jelentős különbség a szenzitivitás és a specificitás között



A jól teljesítő hitelek kategorizálása nagyobb arányban helyes

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Logisztikus regresszió II.

Minta 2



Feltételezések

KHR által
nyilvántartott
adatok

n=500

70% 30%

Omnibus teszt: $p < 0,001$ ✓
Hosmer and Lemeshow goodness-of-fit teszt ($p = 0,105$) ✓
Nagelkerke $R^2 = 69,3\%$

$$P_{(default)} = \frac{e^{0,516 + 1,051x_1 + 0,982x_2}}{1 + e^{0,516 + 1,051x_1 + 0,982x_2}}$$

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya
 x_2 : futamidő.

Minta típusa	Megfigyelt		Becsült			Log. reg. I.
			Default		Percentage Correct	
			0	1		
Teszt	Default	0	60	27	69,0 ↓	97,2
		1	13	63	82,9 ↑	50,0
	Overall Percentage				75,5 ↓	94,7
Tanuló	Default	0	139	24	85,3 ↑	80,8
		1	32	142	81,6 ↓	85,2
	Overall Percentage				83,4 ↑	81,1

a. The cut value is ,390



Magas besorolási pontosság

Teszt minta esetében csökkent a besorolási pontosság, míg tanuló minta esetében enyhén nőtt



Jelentős különbség a specificitás és a besorolási pontosság esetében

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Döntési fa II.

Minta 2



Feltételezések

KHR által nyilvántartott adatok

n=500

70%30%

Minta típusa	Megfigyelt	Becsült			Döntési fa I.
		0	1	Percent Correct	
Tanuló	0	132	46	74,2% 	100,0%
	1	9	165	94,8% 	0,0%
	Overall Percentage	40,1%	59,9%	84,4% 	93,0%
Teszt	0	58	14	80,6% 	100,0%
	1	6	70	92,1% 	0,0%
	Overall Percentage	43,2%	56,8%	86,5% 	92,9%

Default

Node 0

Category	%	n
0	50,6	178
1	49,4	174
Total	100,0	352

tör_szerz*100
Adj. P-value=0,000, Chi-square=204,395, df=2

<= 7,053333333333300

(7,053333333333300, 79,9592000000000000]

> 79,9592000000000000

Node 1

Category	%	n
0	93,6	132
1	6,4	9
Total	40,1	141

Node 2

Category	%	n
0	40,6	43
1	59,4	63
Total	30,1	106

Node 3

Category	%	n
0	2,9	3
1	97,1	102
Total	29,8	105

Törlesztőrészlet szerződéses összeghez viszonyított aránya

Küszöbértékek: 7,0533 és 79,9592

Csökkent a besorolási pontosság

Jelentősen nőtt a szenzitivitás értéke



Az új minta összetétele kedvezőbb

H1: A KHR által nyilvántartott információk alkalmasak arra, hogy a csődelőrejelzés során alkalmazott klasszifikációs eljárások segítségével magas megbízhatósággal előre jelezzék a nemfizetés kockázatát.

Neurális háló II.

Minta 2



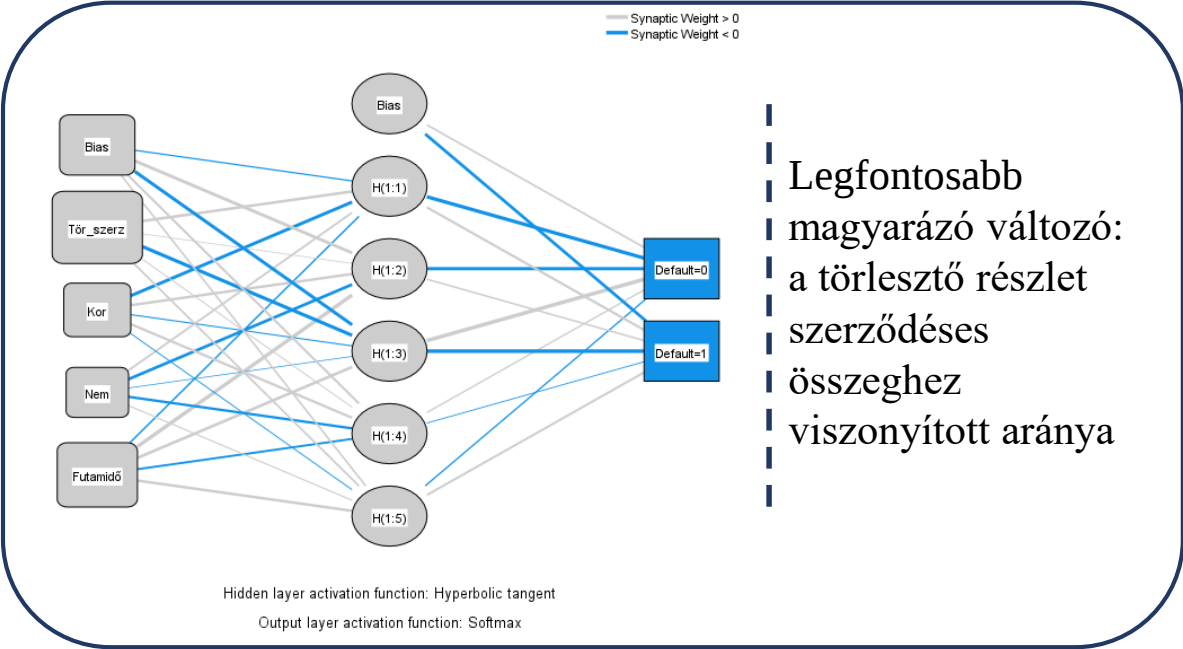
Feltételezések

KHR által nyilvántartott adatok

n=500

70%30%

Minta típusa	Megfigyelt	Becsült			Neurális háló I.
		0	1	Percent Correct	
Tanuló	0	163	14	92,1% 	97,2%
	1	53	132	71,4% 	46,2%
	Overall Percent	59,7%	40,3%	81,5% 	93,3%
Teszt	0	73	0	100,0%	100,0%
	1	19	46	70,8% 	44,4%
	Overall Percent	66,7%	33,3%	86,2% 	96,8%



Csökkent a besorolási pontosság

Nőtt a szenzitivitás értéke



Az új minta összetétele kedvezőbb

Eredmények összehasonlítása, megállapítások

	Elemzés I.			Elemzés II.		
	Log. regr. I.	Döntés i fa I.	Neuráli s háló II.	Log. regr. I.	Döntés i fa II.	Neuráli s háló II.
Tör/szerz*100	X	X	X	X	X	X
Futamidő			X	X		X
Kor			X			X
Nem			X			X

Megfigyelt	Percent correct					
	Log. regr. I.	Döntési fa I.	Neuráli s háló II.	Log. regr. I.	Döntési fa II.	Neuráli s háló II.
0	80,8%	100,0%	97,2%	85,3	74,2%	92,1%
1	85,2%	0,0%	46,2%	81,6	94,8%	71,4%
Σ	81,1%	93,0%	93,3%	83,4	84,4%	81,5%

H1

T1: A fenti elemzések alátámasztják, hogy a KHR által nyilvántartott információk alapján kialakítható egy olyan változókészlet, mely segítségével magas megbízhatósággal lehet előre jelezni a nemfizetés veszélyét.

T2: Az alkalmazott módszerek esetében beigazolódott a túltanulás jelensége. Egy olyan új minta használata, amelyikben azonos arányban szerepelnek a jól és rosszul teljesítő hitelek megoldást jelent a problémára. Megállapítható, hogy ugyan a besorolási pontosság csökken, azonban nő a szenzitivitás értéke, amely jellemző az elemzési cél szempontjából pozitív tulajdonságnak tekinthető.



- 20.1.2.8. Munkanélküliségi ráta, vármegye és régió szerint [%]
- 20.1.2.6. Foglalkoztatási ráta, vármegye és régió szerint [%]
- 20.8.2.1. A teljes munkaidőben alkalmazásban állók havi bruttó átlagkeresete a munkáltató székhelyének elhelyezkedése alapján, vármegye és régió szerint
- 21.1.2.2. Egy főre jutó bruttó hazai termék vármegye és régió szerint
- 23.1.2.4. Érettségizettek és felsőfokú oklevelet szerzettek aránya vármegye és régió szerint [%]

H2: A vállalati csődelőrejelzés során a legmagasabb besorolási pontossággal a neurális hálók segítségével lefuttatott elemzések rendelkeztek (Odom-Sharda, 1990; Olmeda-Fernandez, 1997). Feltételezésem szerint ez az állítás a lakossági hitelek esetében is igaz, tehát a lakossági hitelek kategorizálásánál szintén a neurális hálók fogják a legjobb eredményt szolgáltatni, felülmúlva minden, általam alkalmazott módszert.

H3: A KHR által nyilvántartott információk köre elmarad, a bankok által, a hiteligényléshez bekért információktól (pl. jövedelemadatok). Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Logisztikus regresszió III.

Minta 1



Feltételezések

Bővített
adatbázis

n=500

70%30%

Minta típusa	Megfigyelt		Becsült			Log. regr. I.
			Default		Percentage Correct	
			0	1		
Teszt	Default	0	128	14	90,1 	97,2
		1	1	7	87,5 	50,0
	Overall Percentage				90,0 	94,7
Tanuló	Default	0	267	56	82,7 	80,8
		1	3	24	88,9 	85,2
	Overall Percentage				83,1 	81,1

a . The cut value is ,068

Omnibus teszt: $p < 0,001$ 

Hosmer and Lemeshow goodness-of-fit teszt ($p = 0,855$) 

Nagelkerke $R^2 = 54,2\%$

$$P_{(default)} = \frac{e^{0,915 + 1,040x_1 + 0,984x_2 + 0,951x_3}}{1 + e^{0,915 + 1,040x_1 + 0,984x_2 + 0,951x_3}}$$

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya,
 x_2 : a kereset szerződéses összeghez viszonyított aránya,
 x_3 : futamidő.



Magas besorolási pontosság

Az új változók közül van, ami szignifikánsnak bizonyult.



Az új változók javítják a nemteljesítő hitelek kategorizálását.

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Döntési fa III.

Minta 1

Feltételezések

Bővített adatbázis

n=500

70%30%

Minta	Megfigyelt	Becsült			Döntési fa I.
		0	1	Percent Correct	
Tanuló	0	332	0	100,0%	100,0%
	1	25	0	0,0%	0,0%
	Overall Percentage	100,0%	0,0%	93,0%	93,0%
Teszt	0	133	0	100,0%	100,0%
	1	10	0	0,0%	0,0%
	Overall Percentage	100,0%	0,0%	93,0%	92,9%

Default

Node 0

Category	%	n
0	93,0	332
1	7,0	25
Total	100,0	357

tör. szerz*100
Adj. P-value=0,000, Chi-square=43,867, df=2

<= 2,3739285714285700

(2,3739285714285700, 6,2346774193548400]

> 6,2346774193548400

Node 1

Category	%	n
0	100,0	143
1	0,0	0
Total	40,1	143

Node 2

Category	%	n
0	97,2	104
1	2,8	3
Total	30,0	107

Node 3

Category	%	n
0	79,4	85
1	20,6	22
Total	30,0	107

Törlesztőrészlet szerződéses összeghez viszonyított aránya

Küszöbértékek: 2,3739 és 6,2347

Magas besorolási pontosság

Egy új változó sem került a modellbe

Jelentős különbség a szenzitivitás és a specificitás között

A nemteljesítő hitelek nem tudja helyesen kategorizálni.

Fodor Kitti: A lakossági hitelek nemfizetésének előrejelzése többváltozós statisztikai módszerekkel
Budapest, 2024.02.22.

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Neurális háló III.

Minta 1

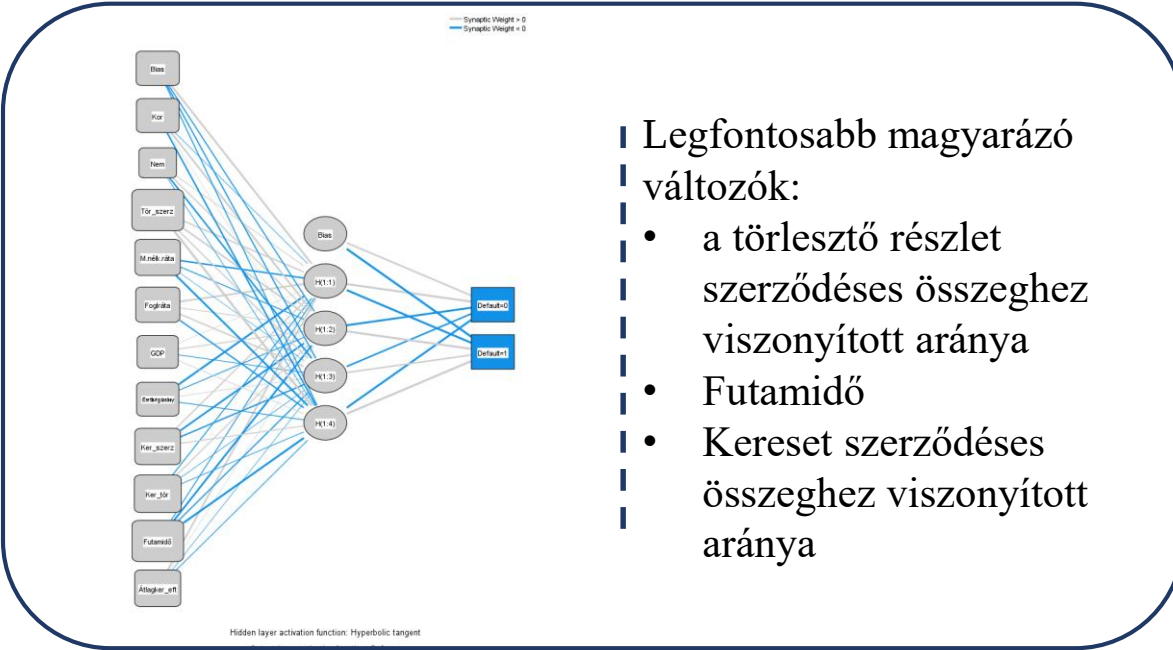
Feltételezések

Bővített adatbázis

n=500

70%30%

Minta	Megfigyelt	Becsült			Neurális háló I.
		0	1	Percent Correct	
Tanuló	0	318	1	99,7%	97,2%
	1	7	20	74,1%	46,2%
	Overall Percent	93,9%	6,1%	97,7%	93,3%
Teszt	0	145	1	99,3%	100,0%
	1	3	5	62,5%	44,4%
	Overall Percent	96,1%	3,9%	97,4%	96,8%



- Legfontosabb magyarázó változók:
- a törlesztő részlet szerződéses összeghez viszonyított aránya
 - Futamidő
 - Kereset szerződéses összeghez viszonyított aránya

Magas besorolási pontosság

Az új változók javítják a nemteljesítő hitelek kategorizálását.

A jól teljesítő hitelek kategorizálása nagyobb arányban helyes.

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Logisztikus regresszió IV.

Minta 2



Feltételezések

Bővített adatbázis

n=500

70%30%

Minta	Megfigyelt		Becsült			Log. reg. II.
			Default		Percentag e Correct	
			0	1		
Teszt	Default	0	70	17	80,5 	69,0
		1	7	69	90,8 	82,9
	Overall Percentage				85,3 	75,5
Tanuló	Default	0	131	32	80,4 	85,3
		1	11	163	93,7 	81,6
	Overall Percentage				87,2 	83,4
a. The cut value is ,338						

a. The cut value is ,338

Omnibus teszt: $p < 0,001$ 

Hosmer and Lemeshow goodness-of-fit teszt ($p = 0,218$) 

Nagelkerke $R^2 = 74,0\%$

$$P_{(default)} = \frac{e^{20,627 + 1,039x_1 + 0,999x_2 + 0,976x_3 + 0,992x_4}}{1 + e^{20,627 + 1,039x_1 + 0,999x_2 + 0,976x_3 + 0,992x_4}}$$

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya,
 x_2 : a kereset törlesztő részlethez összeghez viszonyított aránya,
 x_3 : futamidő,
 x_4 : átlagkereset (eFt).

Magas besorolási pontosság

Az új változók közül van, ami szignifikánsnak bizonyult.



Az új változók javítják a nemteljesítő hitelek kategorizálását.

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Döntési fa IV.

Minta 2



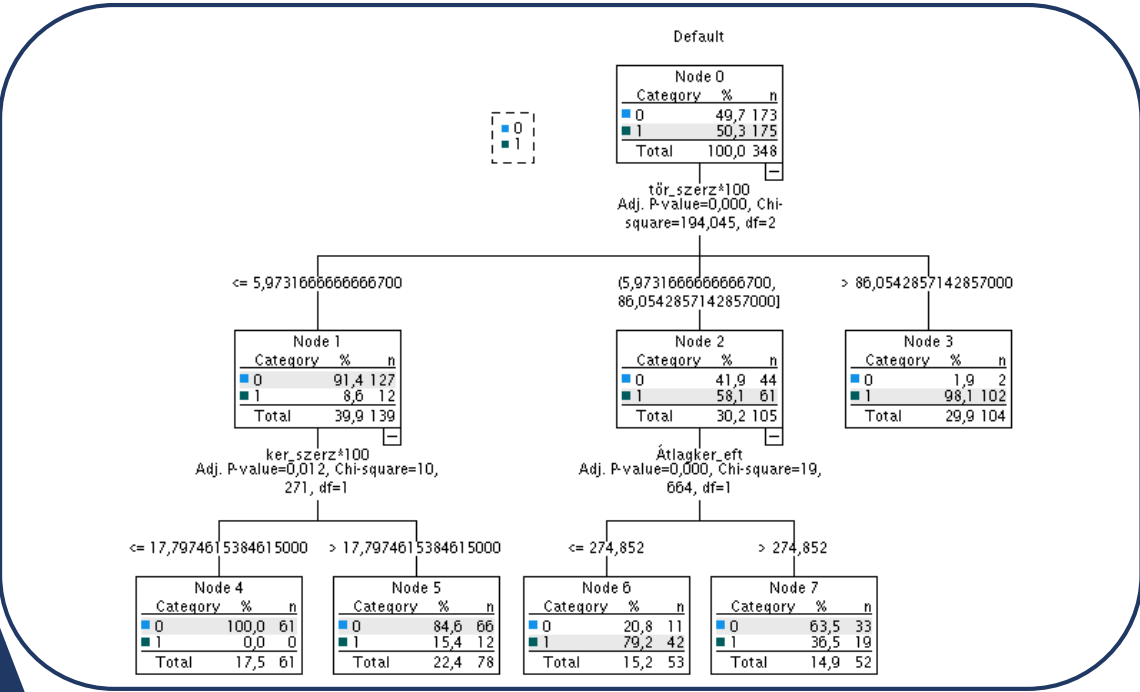
Feltételezések

Bővített adatbázis

n=500

70%30%

Minta	Megfigyelt	Becsült			Döntési fa II.
		0	1	Percent Correct	
Tanuló	0	160	13	92,5% 	74,2%
	1	31	144	82,3% 	94,8%
	Overall Percentage	54,9%	45,1%	87,4% 	84,4%
Teszt	0	66	11	85,7% 	80,6%
	1	14	61	81,3% 	92,1%
	Overall Percentage	52,6%	47,4%	83,6% 	86,5%



Az új változók szignifikánsnak bizonyultak.

Szenzitivitás értéke csökkent a Döntési fa II. modellhez képest.

Magas besorolási pontosság

H2: A neurális hálók fogják a legjobb eredményt szolgáltatni.

H3: Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Neurális háló IV.

Minta 2



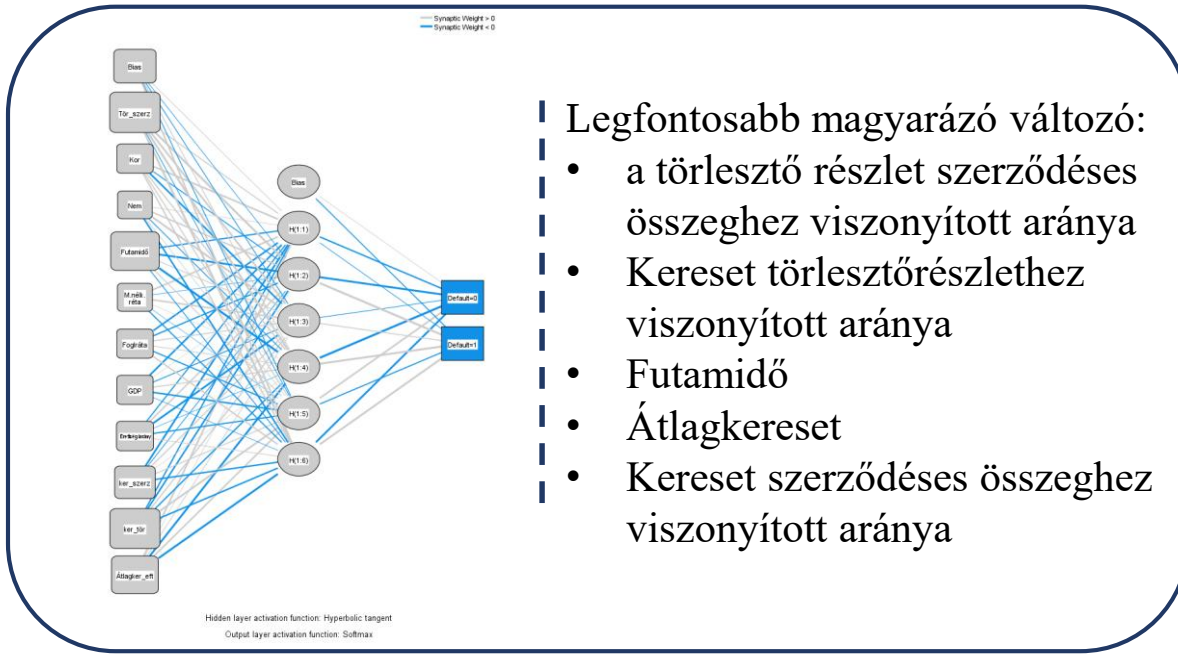
Feltételezések

Bővített
adatbázis

n=500

70% 30%

Minta	Megfigyelt	Becsült			Neurális háló II.
		0	1	Percent Correct	
Tanuló	0	161	14	92,0%	92,1%
	1	26	142	84,5%	71,4%
	Overall Percent	54,5%	45,5%	88,3%	81,5%
Teszt	0	65	10	86,7%	100,0%
	1	16	66	80,5%	70,8%
	Overall Percent	51,6%	48,4%	83,4%	86,2%



Legfontosabb magyarázó változó:

- a törlesztő részlet szerződéses összeghez viszonyított aránya
- Kereset törlesztőrészlethez viszonyított aránya
- Futamidő
- Átlagkereset
- Kereset szerződéses összeghez viszonyított aránya

Magas besorolási pontosság

Nőtt a szenzitivitás értéke

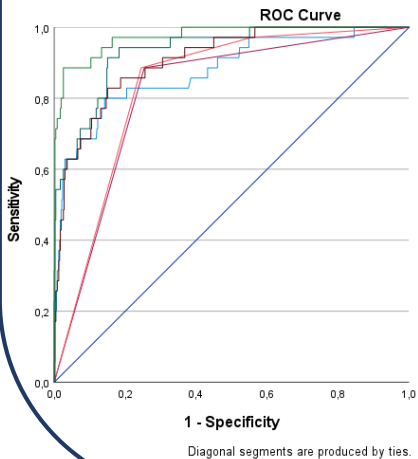


Az új változók kedvezően hatottak a modellre.



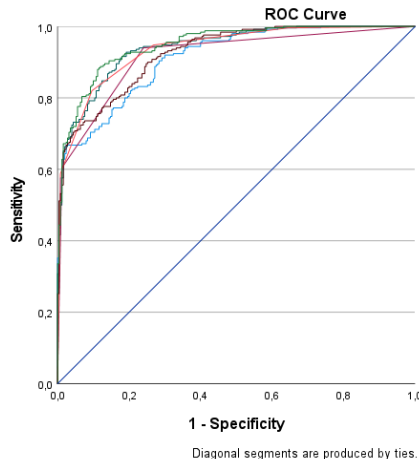
Eredmények összehasonlítása, megállapítások

		Besorolási pontosság			AUC (%)	Gini (%)
		0	1	Összesen		
Logisztikus regresszió	I.	80,8	85,2	81,1	87,7	75,4
	II.	85,3	81,6	83,4	91,2	82,4
	III.	82,7	88,9	83,1	93,4	86,8
	IV.	80,4	93,7	87,2	94,1	88,2
Döntési fa	I.	100	0	93	81,6	63,2
	II.	74,2	94,8	84,4	91,7	83,4
	III.	100	0	93	83,5	67,0
	IV.	74,2	94,8	84,4	93,7	87,4
Neurális háló	I.	97,2	46,2	93,3	90,8	81,6
	II.	92,1	71,4	81,5	92,5	85,0
	III.	97,2	46,2	93,3	97,4	94,8
	IV.	92	84,5	88,3	94,9	89,8



Source of the Curve

- Logisztikus regresszió I.
- Logisztikus regresszió II.
- Logisztikus regresszió III.
- Döntési fa I.
- Döntési fa II.
- Döntési fa III.
- Döntési fa IV.
- Neurális háló I.
- Neurális háló II.
- Neurális háló III.
- Neurális háló IV.
- Reference Line



Source of the Curve

- Logisztikus regresszió I.
- Logisztikus regresszió II.
- Logisztikus regresszió III.
- Döntési fa I.
- Döntési fa II.
- Döntési fa III.
- Döntési fa IV.
- Neurális háló I.
- Neurális háló II.
- Neurális háló III.
- Neurális háló IV.
- Reference Line

H2

H3

T3: Az új változók kedvezően hatottak minden modell esetében az elemzések teljesítményére (besorolási pontosság, AUC és Gini-koefficiens). Annak ellenére, hogy a besorolási pontosság összességében csökkent, jelentősen csökkent a specificitás és szenzitivitás közötti különbség, és ezek együttesen a modellek teljesítményének javulásához vezettek.

T4: Az alkalmazott módszerek közül a legmagasabb besorolási pontossággal a neurális hálók rendelkeznek, a legjobbnak ítélt modell pedig a Neurális háló IV., mellyel közel 90%-os besorolási pontosság érhető el.



	Logisztikus regresszió				Döntési fa				Neurális háló			
	I.	II.	III.	IV.	I.	II.	III.	IV.	I.	II.	III.	IV.
Tör/szerz	X	X	X	X	X	X	X	X	X	X	X	X
ker/szerz			X					X			X	X
ker/tör				X							X	X
Futamidő		X	X	X					X	X	X	X
Átlagker				X				X			X	X
Kor									X	X	X	X
Nem									X	X	X	X
M.nél. ráta											X	X
Fogl. ráta											X	X
GDP											X	X
Érettségizők arány											X	X

T5: A mulasztásba kerülést leginkább az alábbi változók határozták meg:

- törlesztési összeg szerződéses összeghez viszonyított aránya
- futamidő
- átlagkereset
- átlagkereset szerződéses összeghez viszonyított aránya
- átlagkereset törlesztési összeghez viszonyított aránya

H4: A Magyarország területeit jellemző eltérő társadalmi, gazdaság helyzet kapcsolatba hozható a hitelek nemfizetésével.

H5: A különböző demográfiai adatok (nem és életkor) hatással vannak a nemfizetésre.

H4: Társadalmi, gazdasági tényezők vizsgálata.

H5: Demográfiai tényezők vizsgálata.

	Logisztikus regresszió				Döntési fa				Neurális háló			
	I.	II.	III.	IV.	I.	II.	III.	IV.	I.	II.	III.	IV.
Tör/szerz	X	X	X	X	X	X	X	X	X	X	X	X
ker/szerz			X					X			X	X
ker/tör				X							X	X
Futamidő		X	X	X					X	X	X	X
Átlagker				X				X			X	X
Kor									X	X	X	X
Nem									X	X	X	X
M.nél. ráta											X	X
Fogl. ráta											X	X
GDP											X	X
Érettségizők arány											X	X

Symmetric Measures				Approximate Significance
		Value		
Nominal by Nominal	Phi	-,044		,325
	Cramer's V	,044		,325
N of Valid Cases		500		

Nem - default

Symmetric Measures				Approximate Significance
		Value		
Nominal by Nominal	Phi	,136		,231
	Cramer's V	,136		,231
N of Valid Cases		500		

Régió - default

H4

T6: Az elvégzett klasszifikációs elemzések és nemparaméteres próbák alapján azt a következtetést vontam le, hogy egyik társadalmi, gazdasági jellemző sem hozható kapcsolatba a nemteljesítés tényével területi aggregált szinten.

H5

T7: Az elvégzett klasszifikációs elemzések és nemparaméteres próbák alapján azt a következtetést vontam le, hogy a rendelkezésre álló demográfiai adatok közül egyik sem hozható kapcsolatba a nemteljesítés tényével területi aggregált szinten.

H6: A különböző hiteltípusok esetében a bankok más-más feltételt támaszthatnak az ügyféllel szemben, mely akár a nemfizetésre is hatással lehet. Feltételezésem szerint, amennyiben hiteltípusonként készítjük el az előrejelző modellt, jobb besorolási pontosság érhető el.

H6: Amennyiben hiteltípusonként készítjük el az előrejelző modellt, jobb besorolási pontosság, jobb teljesítmény érhető el.

	Egyéb hitelügylet	Fogy. és áruvásárlási hitel	Gépjármű lízing	Lakáscélú jelzálog-hitel	Szabad felhasz. jelz.-hitel	Személyi hitel	Teljes modell
0	88,60%	97,20%	90,50%	90,10%	92,10%	92,00%	92,00%
1	86,10%	95,70%	91,20%	91,20%	83,30%	86,60%	84,50%
Σ	87,30%	96,50%	90,80%	90,60%	88,00%	89,40%	88,30%
AUC	93,30%	99,80%	97,20%	93,20%	95,70%	95,40%	94,90%
Gini	86,60%	99,60%	94,40%	86,40%	91,40%	90,80%	89,80%

Egyéb hitelügylet	Fogy. és áruv. hitel	Gépjármű lízing	Lakáscélú jelz.-hitel	Szabad felh. Jelz.-hitel	Személyi hitel	Teljes modell
Tör/szerz	Tör/szerz	Átlagkereset	Tör/szerz	Ker/tör	Ker/tör	Tör/szerz
Futamidő	Ker/szerz	Tör/szerz	Ker/szerz	Tör/szerz	Tör/szerz	Ker/tör
Átlagker	GDP	Ker/tör	Ker/tör	Fogl, ráta	Átlagker	Futamidő
Kor	Ker/tör	Ker/szerz	Kor	M.nélk. ráta	Ker/szerz	Átlagker
Ker/tör	M.nélk. ráta	GDP	Futamidő	GDP	Futamidő	Ker/szerz

H6

T8: : A fogyasztási és áruvásárlási, a gépjármű lízing, a szabadfelhasználású jelzáloghitel, és a személyi hitel esetében amennyiben önállóan, csak az adott hiteltípusra végezzük az elemzéseket, jobb teljesítmény érhető el neurális háló segítségével.





Agenda

**Kutatási célok,
Elméleti alapok**

Adatbázis

Megállapítások

Felhasználhatóság

Korlátok

Szűkített adatbázis: hat leggyakoribb hiteltípus

Voltak olyan hitelek, amelyeket, azok jellemzői miatt ki kellett hagynom az elemzésekből (pl. Babaváró hitel).

Moratórium torzító hatása

Érdekes terület lehet a későbbiekben ezen hiteleknek a vizsgálata.

Az adatbázis lekérdezésekor még fennállt.
Egy későbbi kutatás során megvizsgálható, hogy más következtetések vonhatók-e le egy későbbi időpontban, amikor már nem állt fenn a moratórium, és hosszabb idő eltelt annak eltörlése óta.

Egy hiányt pótló téma, kevés kutató, kevés tudományos munka foglalkozik a lakossági hitelek nemfizetésének jellemzésével, előrejelzésével statisztikai módszerek segítségével.

Köszönöm a figyelmet!

Módszer	Előny	Hátrány
Diszkriminancia analízis	- Többváltozós modell - Egyszerű felépítés	- Független változók esetében legalább intervallumskála - Nemlineáris kapcsolat esetében nem alkalmas - Komplex feltételrendszer (egyváltozós és többváltozós normalitás, homoszkedaszticitás, multikollinearitás)
Logisztikus regresszió	- Többváltozós modell - Független változó lehet nominális skálán mért - Kevesebb feltétel, mint a DA esetében - Változók súlya könnyen értelmezhető	- Multikollinearitás - Érzékeny az outlierekre - Megfelelő mintanagyság

Módszer	Előny	Hátrány
Döntési fa	• Ötvözi az egy és többváltozós elemzéseket • A bevont változók esetében nincs megkötés • Nem feltétel a változók normális eloszlása • Bináris függő változó esetében a legegyszerűbb a használata	• Túltanulás veszélye • Előrejelzésre tesztelő és tanuló minta egyidejű alkalmazása esetében használható
Neurális háló	• Kezeli a nemlinearitást • Többváltozós modell • Nem probléma a hiányzó adat, az outlier és a multikollinearitás sem • Nagy számú változó és elemszám kezelése • Nincs szükség előzetes információra a függvényyszerű kapcsolat típusáról • Rugalmas felépítés • Nincs szükség az eloszlásról feltételezést tenni • Robosztus, nincs korlátozó előfeltétel • Nominális ismérv alkalmazható	• Nehéz értelmezni a kapcsolatok súlyait • Általában nagy mintát igényelnek • Tanításuk időigényes • Túltanulás veszélye

Többféle ajánlás az ROC görbe alapján

Youden statisztika

$szenzitivitás + specificitás - 1$

➔ MAX

Bal felső sarokhoz legközelebbi pont meghatározása

$(1 - szenzitivitás)^2 + (1 - specificitás)^2$ ➔ MIN

Alkalmazott

Youden

Bal felső sarok

Logisztikus regresszió I.

0,300

0,0389

Logisztikus regresszió II.

0,400

0,6297

0,3902

Logisztikus regresszió III.

0,060

0,0675

Alkalmazott

Classification Table^a

Minta típusa	Observed		Predicted		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	71	16	81,6
		1	9	67	88,2
	Overall Percentage				84,7
Tanuló	Default	0	139	24	85,3
		1	21	153	87,9
	Overall Percentage				86,6

a. The cut value is ,400

Youden

Bal felső sarok

Classification Table^a

Minta típusa	Observed		Predicted		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	70	17	80,5
		1	7	69	90,8
	Overall Percentage				85,3
Tanuló	Default	0	132	31	81,0
		1	11	163	93,7
	Overall Percentage				87,5

a. The cut value is ,338

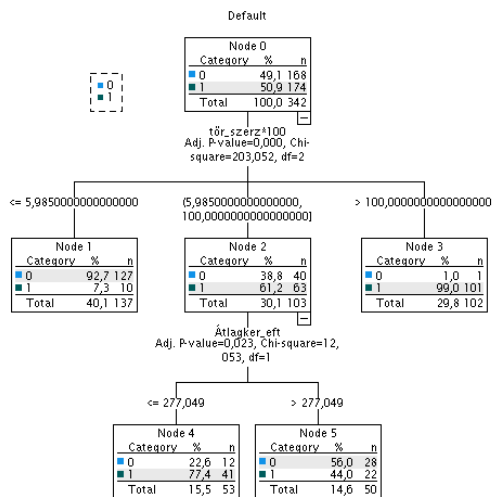
Classification Table^a

Minta típusa	Observed		Predicted		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	73	14	83,9
		1	10	66	86,8
	Overall Percentage				85,3
Tanuló	Default	0	142	21	87,1
		1	23	151	86,8
	Overall Percentage				86,9

a. The cut value is ,418

Másik döntési fa eljárás alkalmazása?

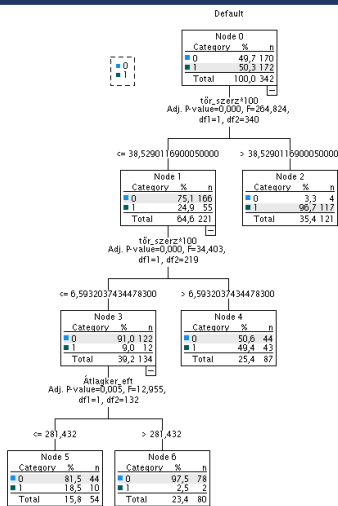
EXHAUSTIVE CHAID



Classification

Sample	Observed	Predicted		
		0	1	Percent Correct
Training	0	155	13	92,3%
	1	32	142	81,6%
	Overall Percentage	54,7%	45,3%	86,8%
Test	0	71	11	86,6%
	1	15	61	80,3%
	Overall Percentage	54,4%	45,6%	83,5%

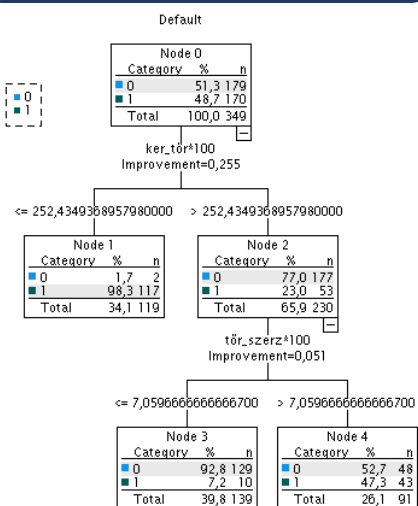
QUEST



Classification

Sample	Observed	Predicted		
		0	1	Percent Correct
Training	0	166	4	97,6%
	1	55	117	68,0%
	Overall Percentage	64,6%	35,4%	82,7%
Test	0	78	2	97,5%
	1	29	49	62,8%
	Overall Percentage	67,7%	32,3%	80,4%

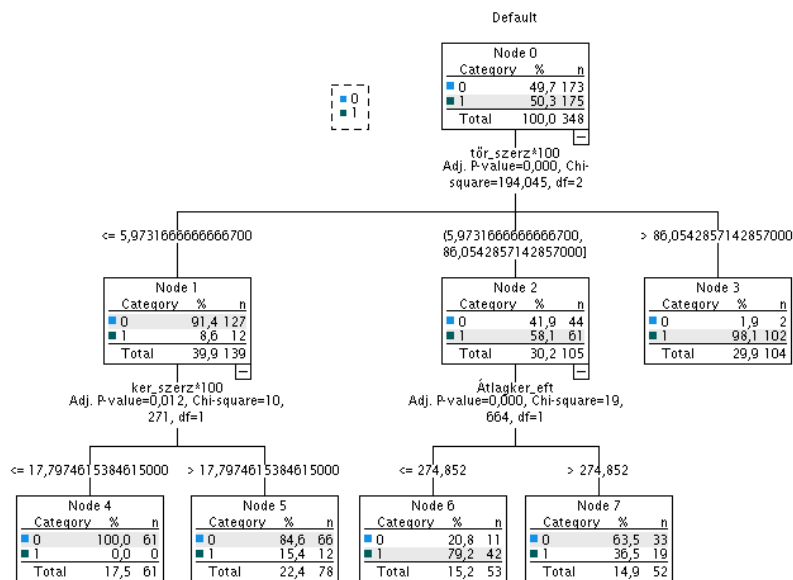
CRT



Classification

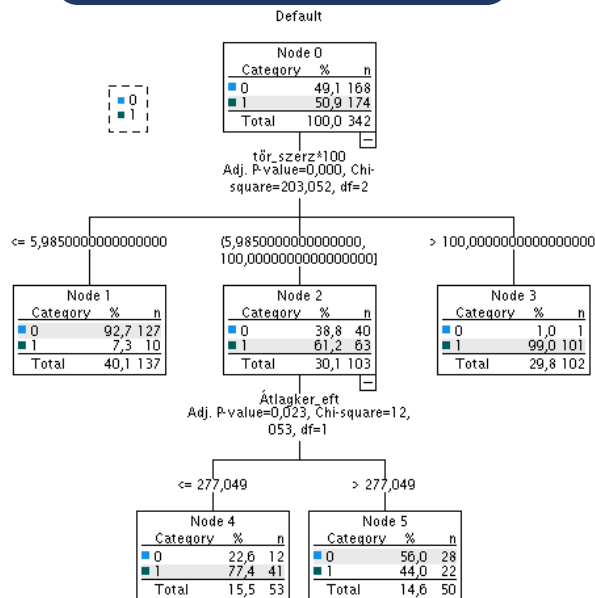
Sample	Observed	Predicted		
		0	1	Percent Correct
Training	0	177	2	98,9%
	1	53	117	68,8%
	Overall Percentage	65,9%	34,1%	84,2%
Test	0	67	4	94,4%
	1	33	47	58,8%
	Overall Percentage	66,2%	33,8%	75,5%

Döntési fa IV.
CHAID



Minta	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	160	13	92,5%
	1	31	144	82,3%
	Overall Percentage	54,9%	45,1%	87,4%
Teszt	0	66	11	85,7%
	1	14	61	81,3%
	Overall Percentage	52,6%	47,4%	83,6%

EXHAUSTIVE
CHAID



Minta	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	155	13	92,3%
	1	32	142	81,6%
	Overall Percentage	54,7%	45,3%	86,8%
Teszt	0	71	11	86,6%
	1	15	61	80,3%
	Overall Percentage	54,4%	45,6%	83,5%



Logisztikus regresszió I.

Egy szignifikáns változó



Logisztikus regresszió II.

Minta		Collinearity Statistics	
		Tolerance	VIF
Teszt	(Constant)		
	tör_szerz*100	,907	1,103
	Futamidő	,907	1,103
Tanuló	(Constant)		
	tör_szerz*100	,948	1,055
	Futamidő	,948	1,055



Logisztikus regresszió III.

Minta		Collinearity Statistics	
		Tolerance	VIF
Teszt	(Constant)		
	tör_szerz*100	,885	1,130
	ker_szerz*100	,669	1,494
	Futamidő	,739	1,353
Tanuló	(Constant)		
	tör_szerz*100	,936	1,068
	ker_szerz*100	,788	1,269
	Futamidő	,783	1,277



Logisztikus regresszió IV.

Minta		Collinearity Statistics	
		Tolerance	VIF
Teszt	(Constant)		
	tör_szerz*100	,806	1,241
	ker_tör*100	,866	1,155
	Futamidő	,844	1,185
	Átlagker_eft	,897	1,115
Tanuló	(Constant)		
	tör_szerz*100	,739	1,353
	ker_tör*100	,801	1,248
	Futamidő	,888	1,127
	Átlagker_eft	,874	1,144



	Egyéb hitelügylet	Fogy. és áruvásárlási hitel	Gépjármű lízing	Lakáscélú jelzálog- hitel	Szabad felhasz. jelzáloghitel	Személyi hitel	Teljes modell	Egyesített modell
0	88,60%	97,20%	90,50%	90,10%	92,10%	92,00%	92,00%	81,3%
1	86,10%	95,70%	91,20%	91,20%	83,30%	86,60%	84,50%	83,1%
Σ	87,30%	96,50%	90,80%	90,60%	88,00%	89,40%	88,30%	82,2%
AUC	93,30%	99,80%	97,20%	93,20%	95,70%	95,40%	94,90%	90,9%
Gini	86,60%	99,60%	94,40%	86,40%	91,40%	90,80%	89,80%	81,8%