

DOI 10.14750/ME.2024.011

Miskolci Egyetem
Gazdaságtudományi Kar
Gazdaságelméleti és Módszertani Intézet

DOKTORI (Phd) ÉRTEKEZÉS
Fodor Kitti

Miskolc, 2023.

DOI 10.14750/ME.2024.011

MISKOLCI EGYETEM
GAZDASÁGTUDOMÁNYI KAR
Gazdaságelméleti és Módszertani Intézet



Fodor Kitti

**A lakossági hitelek nemfizetésének előrejelzése többváltozós
statisztikai módszerekkel**

**The modelling of household loan defaults using multivariate
statistical methods**

Témavezető: Dr. Varga Beatrix Margit

Doktori iskola vezetője: Prof. Dr. Sikos T. Tamás

Hantos Elemér Gazdálkodás – és Regionális Tudományi Doktori Iskola

Miskolc, 2023.

TUDOMÁNYOS VEZETŐ AJÁNLÁSA

Fodor Kitti Ph.D. fokozat megszerzése iránti eljárási kérelméhez

Fodor Kitti tanulmányi munkáját és tudományos tevékenységét 2012 óta kísérem figyelemmel. Alapképzési és mesterképzési tanulmányai alatt 2013. évtől az Üzleti Statisztika és Előrejelzési Intézeti Tanszék demonstrátoraként segítette a tanszék oktatási és kutatási munkáját. Tanulmányai mellett a Hantos Elemér Szakkollégium tagjaként rendszeresen vett részt különböző gazdasági, pénzügyi és számviteli esettanulmányi versenyeken, amelyeken csapatával több alkalommal is kiemelkedő helyezést ért el. Fodor Kitti számvitel mesterszakon fejezte be tanulmányait a 2017. évi tavaszi félévben, majd ezt követően felvételt nyert a Miskolci Egyetem Hantos Elemér Gazdálkodás- és Regionális Tudományi Doktori Iskolába, illetve ezzel párhuzamosan a Pallas Athéné Domus Educationis Alapítvány (PADE) ösztöndíj programjába.

Oktatási tevékenysége során Statisztika, Üzleti statisztika, Kvantitatív statisztikai módszerek tárgyakat tartott magyar és angol nyelven, amelyhez kapcsolódóan aktívan részt vett a tananyag folyamatos fejlesztésében is. A főbb tárgyai mellett két szemeszter során Mikroökonómiát is oktatott magyar nyelven.

A tématerületéhez és annak módszertanához kapcsolódóan folyamatos kutatási munkát folytat, amelynek eredményeképpen a témának megfelelően eddig 20 tudományos cikke jelent meg. Az elmúlt időszakban rendszeresen részt vett szakmai konferenciákon, ahol lehetősége nyílt tudományos kutatási eredményeinek a bemutatására.

A leírtak alapján megállapítható, hogy *Fodor Kitti* stabil, korszerű elméleti és gyakorlati alapokkal rendelkezik, tudományos munkáját az igényesség, szakmaiság, és az új ismeretek iránti elkötelezettség jellemzi. Elért eredményei, publikációi, oktatói tapasztalata alapján méltó arra, hogy a doktori cím megszerzéséhez szükséges követelmények teljesítéséhez lehetőséget kapjon.

Miskolc, 2023. június 04.

Dr. Varga Beatrix Margit Ph.D.

témavezető

Nyilatkozat

Alulírott Fodor Kitti, büntetőjogi felelősségem tudatában kijelentem, hogy „A lakossági hitelek nemfizetésének előrejelzése többváltozós statisztikai módszerekkel” c. Hantos Elemér Gazdálkodás – és Regionális Tudományi Doktori Iskolába beadott Ph.D értekezés önálló munkám eredménye, az irodalmi hivatkozások egyértelműek és teljesek.

Miskolc, 2023. május 29.

aláírás

Tartalom

1	Bevezetés	1
1.1	A kutatás aktualitása, kutatási probléma megfogalmazása	1
1.2	A disszertáció felépítése, hipotézisei	1
2	Hitelezés története, magyar hitelkörkép	4
2.1	A hitelezés története rövid története	4
2.2	Magyar hitelkörkép, eladósodottság	8
2.3	Az eladósodottság veszélyei	11
2.4	Hitelek csoportosítása lakossági hitelezés esetében	13
2.5	Hitelezéssel kapcsolatos alapfogalmak	14
2.6	Egyéb, a hitelezéssel összefüggő, nehezen mérhető tényezők	15
2.6.1	Pénzügyi inklúzió	16
2.6.2	Pénzügyi kultúra	16
2.6.3	Pénzügyi reziliencia	22
3	Csődelőrejelzés fejlődéstörténete.....	24
3.1	Nemzetközi csődmodell-kutatások	24
3.2	Magyar csődmodell-kutatások	29
4	A csődelőrejelzés módszerei	33
4.1	Diszkriminancia analízis	34
4.2	Logisztikus regresszió	37
4.3	Rekurzív particionáló algoritmus (RPA)	41
4.4	Neurális hálók	44
4.5	Módszerek alkalmazhatósága, összegzése	48
4.6	Módszerek értékelése	50
4.6.1	Klasszifikációs mátrix	50
4.6.2	ROC (Receiver Operating Characteristic) görbe	51
4.6.3	Gini-koefficiens (AR)	52
4.7	További, az elemzések során alkalmazott módszerek	54
4.7.1	Korrelációs számítás	54
4.7.2	Multikollinearitás	54
4.7.3	Keresztáblaelemzés	55
5	A kutatás adatbázisa	56
5.1	A Központi Hitelnyilvántartó Rendszer	56
5.2	Adatbázis	57

5.3	Mintavétel.....	63
6	A kutatás eredményei, hipotézisek vizsgálata	64
6.1	A kutatáshoz alkalmazott szoftver.....	64
6.2	Csőelőrejelzési módszerek használata lakossági nemfizetés előrejelzésére.....	64
6.2.1	Logisztikus regresszió I.....	64
6.2.2	Döntési fa I.....	66
6.2.3	Neurális háló I.....	68
6.2.4	Új minta kialakítása.....	69
6.2.5	Logisztikus regresszió II.....	69
6.2.6	Döntési fa II.....	72
6.2.7	Neurális háló II.....	74
6.2.8	Eredmények összehasonlítása	75
6.2.9	Az adatbázis bővítése	77
6.2.10	Logisztikus regresszió III.....	77
6.2.11	Döntési fa III.....	81
6.2.12	Neurális háló III.....	82
6.2.13	Logisztikus regresszió IV.....	83
6.2.14	Döntési fa IV.....	86
6.2.15	Neurális háló IV.....	89
6.2.16	Eredmények összehasonlítása II.....	90
6.3	Gazdasági, társadalmi, demográfiai jellemzők hatása a mulasztásra	94
6.4	Van-e különbség az egyes hiteltípusok között?.....	96
6.4.1	Egyéb hitelügylet.....	96
6.4.2	Fogyasztási, áruvásárlási hitel.....	97
6.4.3	Gépjármű lízing.....	98
6.4.4	Lakáscélú jelzáloghitel	99
6.4.5	Szabad felhasználású jelzáloghitel	100
6.4.6	Személyi hitel.....	101
6.4.7	Összegzés	102
7	A kutatás főbb megállapításai, következtetések, javaslatok.....	103
7.1	A disszertáció tézisei, új és újszerű megállapításai	103
7.2	Az eredmények felhasználhatósága, további kutatási irányok	105
8	Összefoglalás.....	106
9	Summary.....	107

10 Irodalomjegyzék.....	108
11 Mellékletek.....	118
11.1 Logisztikus regresszió I. outputjai	118
11.2 Döntési fa I. outputjai	119
11.3 Neurális háló I. outputjai	120
11.4 Logisztikus regresszió II. outputjai	122
11.5 Döntési fa II. outputjai	124
11.6 Neurális háló II. outputjai	124
11.7 Logisztikus regresszió III. outputjai	126
11.8 Döntési fa III. outputjai	127
11.9 Neurális háló III. outputjai.....	128
11.10 Logisztikus regresszió IV. outputjai	130
11.11 Döntési fa IV. outputjai.....	132
11.12 Neurális háló IV. outputjai.....	133
11.13 Egyedi modellek outputjai	135
11.13.1 Egyéb hitelügylet.....	135
11.13.2 Fogyasztási és áruvásárlási hitel.....	138
11.13.3 Gépjármű lízing	141
11.13.4 Lakáscélú jelzáloghitel	144
11.13.5 Szabad felhasználású jelzáloghitel	147
11.13.6 Személyi hitel	150

Ábrajegyzék

1. ábra: A háztartási adósságállomány alakulása 2000. 06. hó és 2004.06. hó között	5
2. ábra: A háztartások hiteleinek denominációs szerkezete	6
3. ábra: Új háztartási hitelek a teljes hitelintézeti szektorban 2008-2020 között (Mrd Ft)	7
4. ábra: A moratóriumban érintett hitelek darabszám szerinti aránya a háztartásoknál 2021 júniusában.....	8
5. ábra: A magyar háztartások adóssága a GDP arányában 2001-2021	9
6. ábra: A magyar lakosság fogyasztói hiteleinek alakulása 2001-2021 (Mrd Ft).....	9
7. ábra: A fennálló mulasztások számának változása státusz szerint 2006-2021 között.....	10
8. ábra: JTM korlátok	12
9. ábra: A felnőtt lakosságon belül a bankszámlával rendelkezők aránya 2021-ben	16
10. ábra: Helytelen válaszok aránya az Econventio tesztek esetében korcsoportonként	18
11. ábra: A pénzügyi kultúra szintjei az Európai országokban	19
12. ábra: A helyes válaszok aránya 2019-ben és 2022-ben.....	21
13. ábra: A pénzügyi tartalékok időbeli megoszlása 2020-ban.....	23
14. ábra: Döntési fa a tanulási minta alapján.....	32
15. ábra: Az empirikus vizsgálat folyamata	33
16. ábra: A diszkriminanciaelemzés menete	35
17. ábra: Logisztikus valószínűség a magyarázó változó függvényében	38
18. ábra: Tipikus döntési fa	42
19. ábra: Az egyes megoszlásokhoz tartozó entrópia értékek	43
20. ábra: Neurális hálózat egy rejtett réteggel	44
21. ábra: A tanulási és tesztelési hiba alakulása a tanítási ciklusok számának függvényében	45
22. ábra: Klasszifikációs tábla hitelnemfizetésre.....	50
23. ábra: ROC görbe.....	51
24. ábra: Néhány lehetséges ROC-görbe	52
25. ábra: Néhány lehetséges CAP görbe	53
26. ábra: Gini-koefficiens kiszámítása (AR).....	53
27. ábra: A fennálló mulasztások megoszlása a mulasztás időtartama szerint 2021-ben	57
28. ábra: Az adatbázis szűkítésének lépései	60
29. ábra: Döntési fa I. a tanuló és tesztelő minta esetében	67
30. ábra: A kialakult neurális háló a Neurális háló I. modell esetében	68
31. ábra: Döntési fa II. a tanuló és tesztelő minta esetében	73
32. ábra: A kialakult neurális háló a Neurális háló II. modell esetében	74

33. ábra: Döntési fa III. a tanuló és tesztelő minta esetében	81
34. ábra: A kialakult neurális háló a Neurális háló III. modell esetében	82
35. ábra: Döntési fa IV. a tanuló minta esetében	87
36. ábra: A kialakult neurális háló a Neurális háló IV. modell esetében	89
37. ábra: A létrehozott modellek ROC-görbéje	92

Táblázatjegyzék

1. táblázat: A kutatás hipotézisei	2
2. táblázat: A logisztikus regresszió modell változói és paramétereinek értékei	29
3. táblázat: Az egyes modellek besorolási pontossága a teszt és a tanuló minta alapján	32
4. táblázat: A struktúravizsgáló módszerek egy részének összefoglalása	34
5. táblázat: A neurális háló erősségei és gyengeségei	48
6. táblázat: Klasszifikációs eljárások előnyei és hátrányai	49
7. táblázat: Az eredeti adatbázis megoszlása hiteltípusok szerint	59
8. táblázat: A szűkebb adatbázis megoszlása hiteltípusok szerint	62
9. táblázat: A Logisztikus regresszió I. modell szignifikáns változói	65
10. táblázat: Klasszifikációs mátrix a Logisztikus regresszió I. modell esetében	65
11. táblázat: Klasszifikációs mátrix a Döntési fa I. modell esetében	67
12. táblázat: Klasszifikációs mátrix a Neurális háló I. modell esetében	69
13. táblázat: A logisztikus regresszió II. modell magyarázóereje	70
14. táblázat: A Logisztikus regresszió II. modell szignifikáns változói	70
15. táblázat: Klasszifikációs mátrix a Logisztikus regresszió II. modell esetében	71
16. táblázat: Logisztikus regresszió II. modellbe bevont magyarázó változók között fennálló kapcsolat	72
17. táblázat: A VIF mutató értékei a Logisztikus regresszió II. modell esetében	72
18. táblázat: Klasszifikációs mátrix a Döntési fa II. modell esetében	73
19. táblázat: Klasszifikációs mátrix a Neurális háló II. modell esetében	75
20. táblázat: A klasszifikációs modellek által használt változók összefoglalása	75
21. táblázat: Az alkalmazott módszerek besorolási pontossága	76
22. táblázat: A Logisztikus regresszió III. modell szignifikáns változói	78
23. táblázat: Klasszifikációs mátrix a Logisztikus regresszió III. modell esetében	79
24. táblázat: Korrelációs mátrix a multikollinearitás ellenőrzéséhez	80
25. táblázat: A VIF mutató értékei a Logisztikus regresszió III. modell esetében	80
26. táblázat: Klasszifikációs mátrix a Döntési fa III. modell esetében	81

27. táblázat: Klasszifikációs mátrix a Neurális háló III. modell esetében	83
28. táblázat: A Logisztikus regresszió IV. modell szignifikáns változói	83
29. táblázat: Klasszifikációs mátrix a Logisztikus regresszió IV. modell esetében	84
30. táblázat: Korrelációs mátrix a multikollinearitás ellenőrzéséhez	85
31. táblázat: A VIF mutató értékei a Logisztikus regresszió IV. modell esetében	86
32. táblázat: Klasszifikációs mátrix a Döntési fa IV. modell esetében	88
33. táblázat: Klasszifikációs mátrix a Neurális háló IV. modell esetében	90
34. táblázat: A létrehozott modellek magyarázó változóinak összefoglalása	91
35. táblázat: A kialakított modellek teljesítménye a különböző értékelési technikák alapján .	92
36. táblázat: A létrehozott modellek magyarázó változóinak összefoglalása, kiemelve a gazdasági, társadalmi, demográfiai változókat.....	94
37. táblázat: A nem és a nemfizetés közötti kapcsolat vizsgálata	95
38. táblázat: A nem és a nemfizetés közötti kapcsolat vizsgálata	95
39. táblázat: Az egyéb hitelügylet kategória során kialakult neurális háló értékelése	96
40. táblázat: Az egyéb hitelügylet kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása	97
41. táblázat: A fogyasztási és áruvásárlási hitel kategória során kialakult neurális háló értékelése	97
42. táblázat: A fogyasztási és áruvásárlási hitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása	98
43. táblázat: A gépjármű lízing kategória során kialakult neurális háló értékelése	98
44. táblázat: A gépjármű lízing kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása	99
45. táblázat: A lakáscélú jelzáloghitel kategória során kialakult neurális háló értékelése	99
46. táblázat: A lakáscélú jelzáloghitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása	100
47. táblázat: A szabad felhasználású jelzáloghitel kategória során kialakult neurális háló értékelése	100
48. táblázat: A szabad felhasználású jelzáloghitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása	101
49. táblázat: A személyi hitel kategória során kialakult neurális háló értékelése	101
50. táblázat: A személyi hitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása.....	102
51. táblázat: A modellek teljesítményének összehasonlítása	102

1 Bevezetés

1.1 A kutatás aktualitása, kutatási probléma megfogalmazása

Mind a vállalati, mind pedig a lakossági hitelezésnél fontos szempont a pénzügyi intézet számára, hogy olyan ügyfeleknek nyújtson hitelt, akiknél alacsony annak a kockázata, hogy az összeget nem fizetik vissza. Az intézetek ugyan rendelkeznek saját hitelbírálati folyamattal, azonban ennek ellenére előfordul, hogy nem megfelelően választják meg azokat az ügyfeleket, akik számára célszerű hitelt nyújtani.

A vállalati csődök előrejelzésének jelentős szakirodalma van, viszont a lakossági hitelígénylők besorolásával kevés irodalom foglalkozik. A vállalati szektor esetében megfigyelhető volt, hogy a módszertani lehetőségek és az informatika fejlődésével folyamatosan nőtt az alkalmazott elemzési lehetőségek köre, a kezdeti egyváltozós elemzésektől napjainkra eljutottunk a neurális háló segítségével készült modellekig.

Ez utóbbi állítás hazánkban fokozottan igaz, elenyésző tudományos munka kapcsolódik ehhez a témakörhöz. Mindez annak ellenére igaz, hogy nem is olyan régen volt a legutóbbi válság, és az utóbbi időben a hitelezés kimondottan felpörgött, úgy veszik fel sokan a hiteleket, mintha nem lenne holnap. Annak a felismerése, hogy kiből válik rossz adós, nem egyszerű. Az adatok elemzésére több módszer is kínálkozik, amik akár eltérő eredményt is mutathatnak. Az elemzés szempontjából érdekes kérdés lehet, milyen adatok kerüljenek be az adatbázisba, miket vegyünk figyelembe befolyásoló tényezőként, melyek azok a változók, amelyek a leginkább meg tudják határozni a nemfizetés valószínűségét.

A kutatásom célja a lakossági hitelek bedőlésének vizsgálata többváltozós statisztikai eljárásokon alapuló módszerek segítségével. A különböző módszerek és modellek segíthetnek annak a feltérképezésében, hogy mik járulnak hozzá ahhoz, hogy valakiből nemfizető adós váljon, továbbá választ kaphatunk arra, hogy mely többváltozós statisztikai módszer produkálja a legjobb eredményt. A módszer segítségével várhatóan létrejön egy olyan modell, mely a lehető leghatékonyabban szűri ki a rossz adósokat.

A kutatást nehezíti azonban, hogy korlátozott az elérhető adatoknak a köre, mellyel a kutatás során én is szembesültem, hosszú időre volt szükség, mire partnert találtam az adatbázis esetében.

A kutatás újszerűségét a fentebb leírtak alapján tehát az adja, hogy hazánkban ez a kutatási terület a vállalati szektorra fókuszál, az én kutatásom fókuszában pedig a lakossági szektor áll.

1.2 A disszertáció felépítése, hipotézisei

A disszertációban a bevezetés után a szakirodalmi áttekintés következik, mely során kitérek a hitelezés rövid történetére, mikorra vezethető vissza a hitelek eredete, milyen fontos események történtek az elmúlt években. Külön figyelmet fordítok arra, hogy elemezzem a magyarországi hitelezési helyzetet, illetve, hogy milyen lehetséges veszélyei lehetnek az eladósodottnak. Továbbá röviden összefoglalom milyen hiteltípusok érhetők el jelenleg, melyek azok a mindenképpen fontos alapfogalmak, amelyekkel tisztában kell lenni hitelfelvétel esetén. A

második fejezet végén térek ki arra, milyen egyéb tényezők létezhetnek, melyeknek fontos szerepe lehet a hitelezés során, kiemelve a pénzügyi inklúzió, kultúra és reziliencia jelentőségét, röviden ismertetve egy korábbi kutatásom eredményeit is.

A szakirodalmi áttekintés során vizsgálom, hogy hasonló kutatási területen milyen módszereket alkalmaztak. Ennek a megválaszolásában a csődelőrejelzés történetére támaszkodom, hiszen számos kutatás foglalkozott és foglalkozik még ma is ezzel a területtel. A teljes kép miatt megvizsgálom a nemzetközi és a hazai csődmodell-kutatásokat egyaránt. A szakirodalom ezen része segít abban, hogy kikristályosodjon, melyek azok a módszerek, amelyeket az elemzés során én is alkalmazni tudok, ezt követően pedig bemutatom az egyes módszereket.

A következő fejezet az adatbázist nyújtó BISZ Zrt. és az adatbázis bemutatásával kezdődik. Az általuk rendelkezésemre adott adatbázison végeztem el a szükséges elemzéseket. A hipotéziseim alátámasztásához több átfogó, és több egyedi modellt készítek. Ezek a modellek segítenek abban, hogy megvizsgáljam, mely változók járulnak hozzá a nemfizetés valószínűségének növekedéséhez, illetve meghatározhatóvá válik, melyik módszertan nyújtja a legpontosabb eredményt.

A kutatás során az alábbi hipotéziseket fogalmaztam meg:

1. táblázat: A kutatás hipotézisei¹

Hipotézis	Alkalmazott módszertan
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, Döntési fa, Neurális háló, ROC-görbe, Gini-koefficiens, Korreláció
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 hiteligenyléshez bekért információktól (pl. jövedelemadatok). Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.	
H4: A Magyarország területeit jellemző eltérő társadalmi, gazdaság helyzet kapcsolatba hozható a hitelek nemfizetésével.	Logisztikus regresszió, Döntési fa, Neurális háló, Keresztábra elemzés
H5: A különböző demográfiai adatok (nem és életkor) hatással vannak a nemfizetésre.	

¹ KHR: Központi Hitelnyilvántartó Rendszer

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.

Neurális háló,
ROC-görbe,
Gini-koefficiens

Forrás: Saját szerkesztés

2 Hitelezés története, magyar hitelkörkép

2.1 A hitelezés története rövid története

A hitelezés évezredek óta az emberi kultúra része. Az első szabályok Hammurápihoz köthetők, aki kőtáblákon rögzítette a kölcsönzőkre és a kölcsönvevőkre vonatkozó szabályokat. (Fekete-Tatay, 2012)

Az ókori Babilóniában, Föníciában és Görögországban a finanszírozási élet virágzott, melyhez magas kamatok társultak, és már ekkor is létezett az uzsora fogalma is. Amennyiben valaki nem fizette vissza a kölcsönt, komoly következményekkel is szembesülhetett, mint például a halál vagy a rabszolgaság.

A középkorban a hitelfelvételek fő célja a fogyasztási igények kielégítése volt. A kamathoz való hozzáállást a niceai zsinat egyértelműen rögzítette, ugyanis kimondta a kamattilalmat. A XII. században akár kitesztással is büntették a kamatszédőt. Ez az ellenállás a reformáció idején kezdett enyhülni. Hazánkban a XIV. században kezdtek ismét kölcsönügyleteket lebonyolítani, de ezt is főleg az uralkodók tették. A XV. században merült fel a gondolat, hogy a pénzt mégiscsak forgatni kell a gazdaság érdekében. Ez hazánkra is igaz volt, a Thurzó család kapcsolatban állt a Fugger és Medici családokkal, akik akkoriban ismert bankárdinasztiának számítottak. Az 1600-1800-as években a hiteleket főként gazdag birtokosok, polgárok nyújtották.

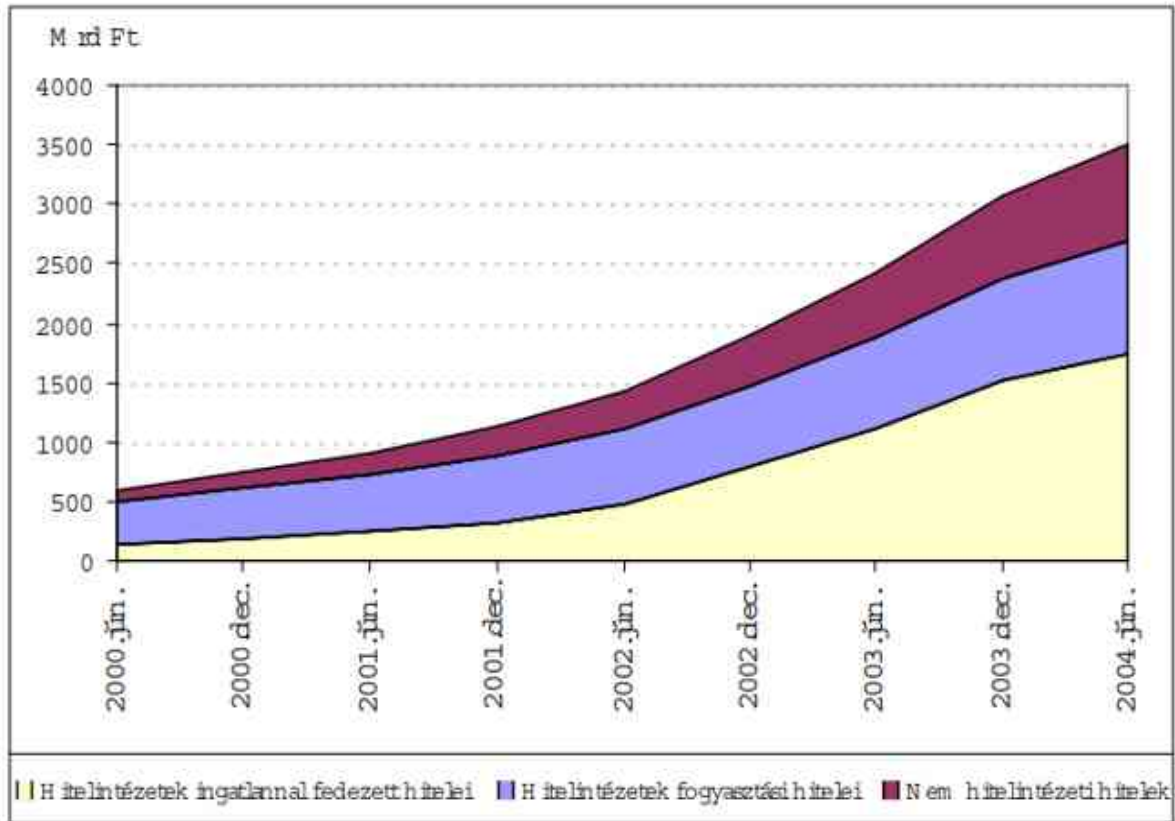
A kontinentális országokban a XIX. századtól beszélhetünk hitelezésről. Magyarországon az alapgondolat gr. Széchenyi István nevéhez köthető. Több meghatározó esemény is történt ebben az időszakban. 1830-ban megalapították az első magyar bankot. Továbbá a modern hitel fogalma is ehhez az időszakhoz, illetve az olasz polgári törvénykönyvhöz köthető. (Vértesy, 2008)

A nagy gazdasági világválság Magyarországot is erősen sújtotta. Az I. világháborút követően rengeteg földbirtokos adósodott el olyan mértékben, hogy nem tudták hiteleiket fizetni, így a jelzálogok alapját jelentő termőföldek kerültek veszélybe. A gazdák megsegítésére számtalan intézkedést vezettek be, mint például a végrehajtási moratórium vagy a védett birtokká történő nyilvánítás. Az intézkedések sikeresnek bizonyultak, a válságot követően néhány év múlva már tudták a gazdák a hiteleiket törleszteni. (Szabó, 2021)

A 40-es évekre vált érezhetővé ennek az iparágak az elszemélytelenedése, azonban egy kérdés állandó marad(t), nyújt-e a bank hitelt az igénylőnek, vagy sem. Az 1949-ben alapított Országos Takarékpénztár volt az az intézmény, amely főleg lakossági ügyfelekkel foglalkozott. A bankok korábbi monopolhelyzetét az 1987-től érvényben lévő kétszintű bankrendszer törte meg. A rendszerváltást követően voltak nehézségei a hazai bankéletnek, mint például '92-es bankcsődök. A bankcsődök háttérében az alábbi tényezők húzódtak:

- belső hitelek nyújtása
- bizonylati fegyelem hiánya
- banküzemi szabályok figyelmen kívül hatása.

A bankrendszer stabilitásának kezdete a '90-es évek közepéig azonban váratott magára. (Vértesy, 2008)

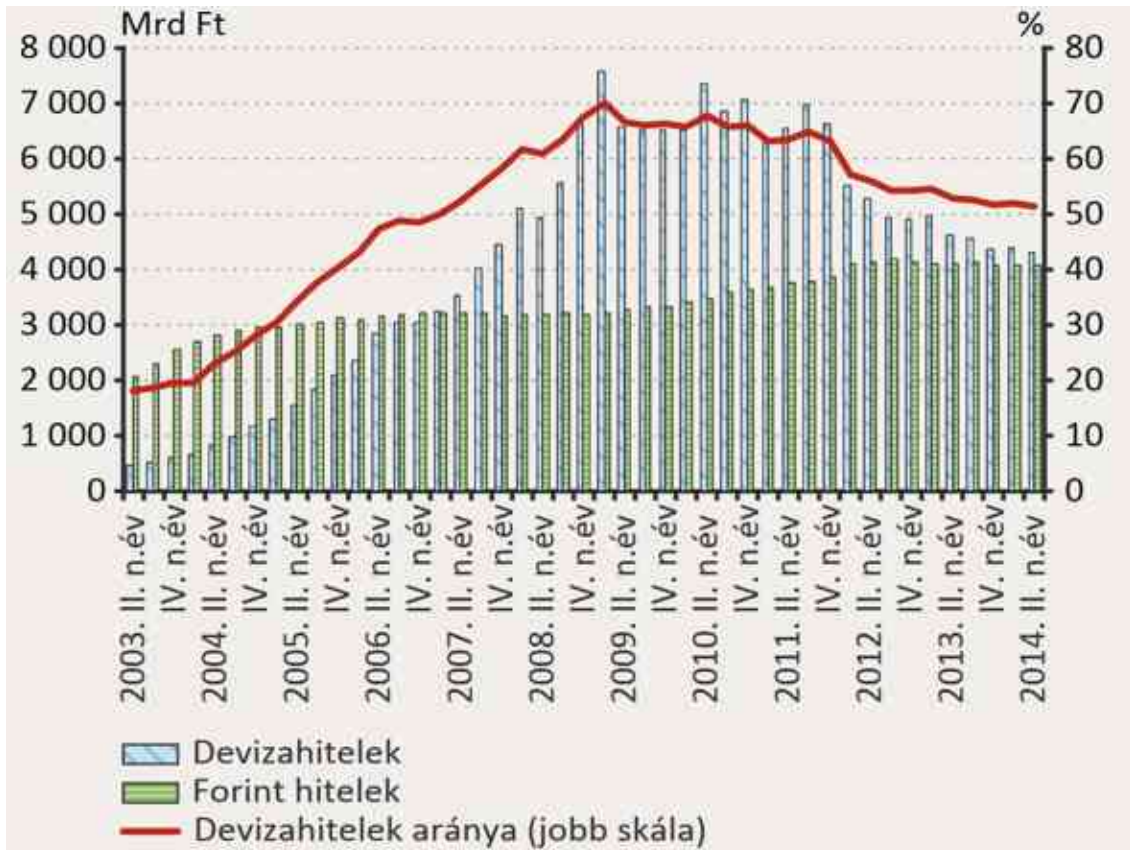


1. ábra: A háztartási adósságállomány alakulása 2000. 06. hó és 2004.06. hó között

Forrás: Dobák-Sági (2005) p3.

Hosszú ideig hitelezés esetén főként a nagyvállalati hitelezésen volt a fókusz, jelentős változás csak a 90-es évek végén történt. 2004-re a magyar lakosság hitellel rendelkező része már milliós nagyságrendet öltött, a bankok versengtek a hitelezőkért, mely a feltételek folyamatos enyhítésével járt. A gyors ütemű növekedést támasztja alá az 1. számú ábra is, ahol megfigyelhető, hogy jelentősen nőtt az adósságállomány a hitelintézetek ingatlannal fedezett, illetve fogyasztási hiteleinél, továbbá a nem hitelintézeti hiteleknél is 2000 és 2004 között. A megtakarítások ezzel ellentétesen alakultak, csökkenés volt megfigyelhető esetükben, ami akár adósságcsapdához is vezethet egyes ügyfeleknél. Már a válság kialakulása előtt voltak jelek, hogy a pénzügyi tudatosság, pénzügyi kultúra területén kedvezőtlen a helyzet a magyar lakosság esetében, amely veszélyes tényezőnek minősül a hitelezés kapcsán, és ennek a tudásnak a hiánya is vezethet adósságcsapdához.

Mindezt igazolta, hogy 2004-ben drasztikusan, 50%-kal megnőtt a nemteljesítő hiteleknek a száma. Voltak azonban olyan intézetek, amelyeknél a korábbi nemteljesítés nem volt elutasító tényező, kimondottan a BAR-listán szereplő ügyfeleket célozták meg, még nagyobb adósságba sodorva őket. (Dobák-Sági, 2005)



2. ábra: A háztartások hiteleinek denominációs szerkezete

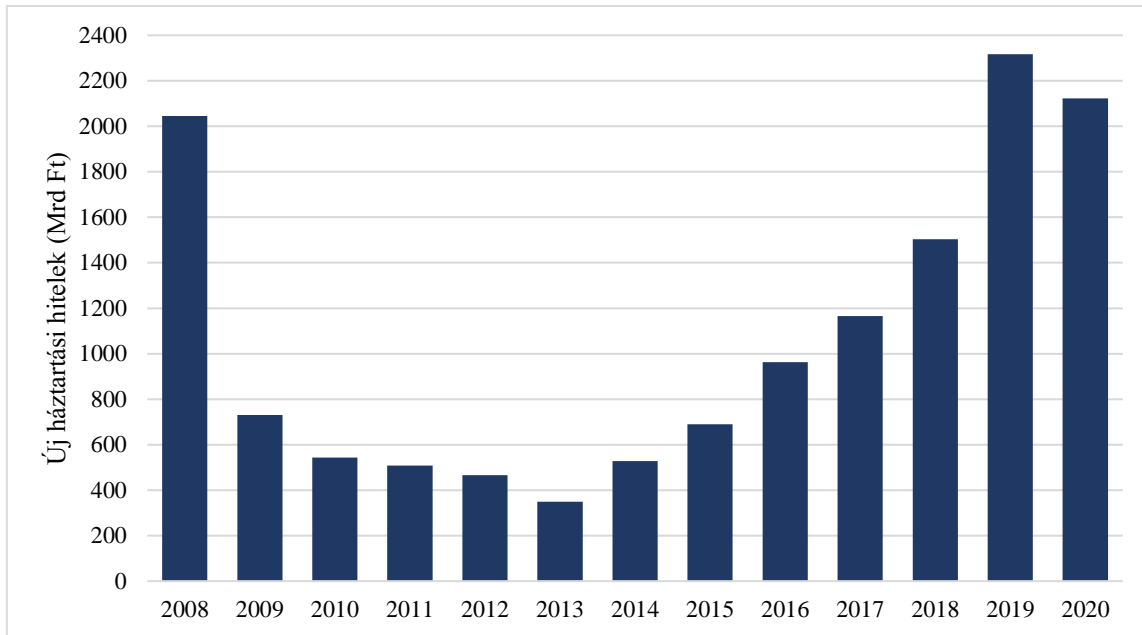
Forrás: MNB, 2014

Ahogy a 2. ábrán láthatjuk a felvett devizahitelek értéke 2003 és 2009 között gyors ütemben növekedett. A fentebb leírtak, illetve a növekvő devizahitel állomány jó táptalajt biztosítottak egy újabb „katasztrófának”. A 2008-as válság miatt a frank árfolyama jelentősen megnövekedett, ami a törlesztőrészek olyan mértékű növekedésével járt, hogy a hitelezők már nem tudtak ennek eleget tenni. A kormány igyekezett megsegíteni a bajba jutott adósokat, és az alábbi intézkedések születtek:

- Vég törlesztés adott frank vagy euró árfolyamon,
- Árfolyamgát: késelelembe jutott ügyfelek esetében 5 évig kedvezményes törlesztőrészlet fizetése,
- Gyűjtőszámla,
- NET program (Nemzeti Eszközkezelő program).

Végezetül 2014-ben kivezették a devizaalapú hiteleket.

Ahogy a 3. számú ábrán is látható 2009-et követően néhány év csökkenés figyelhető meg az új háztartási hitelek esetében, azonban 2014-től kezdődően alapvetően növekedésről beszélhetünk.

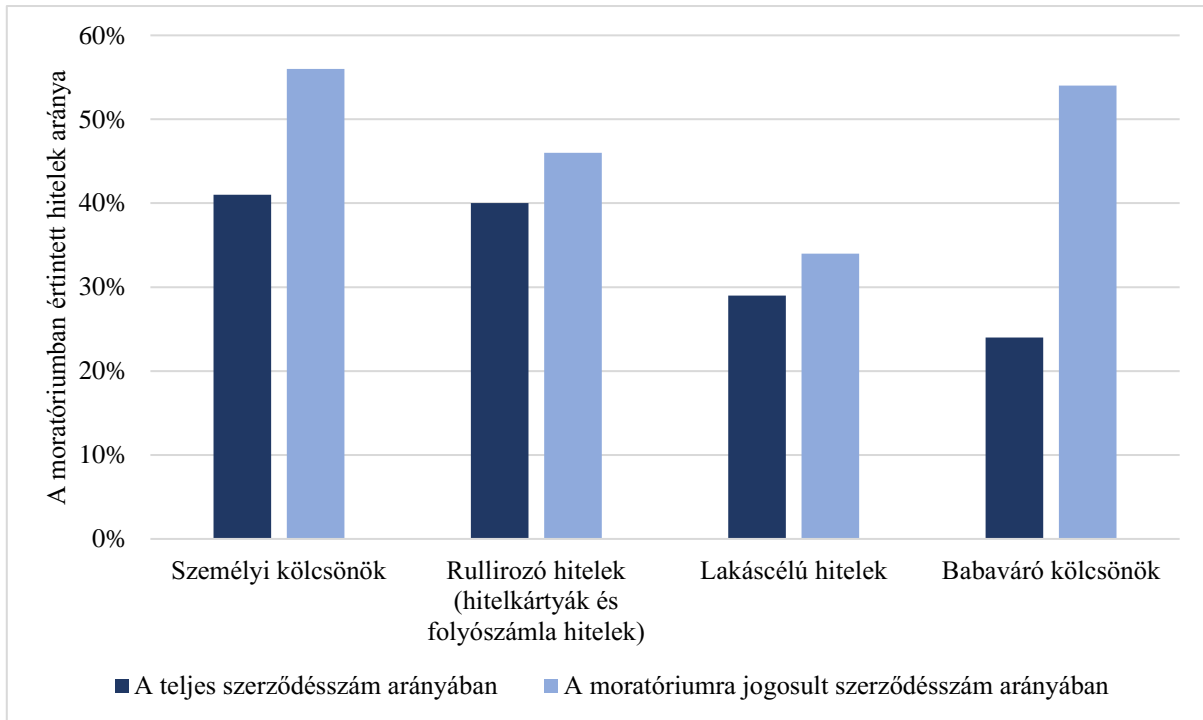


3. ábra: Új háztartási hitelek a teljes hitelintézeti szektorban 2008-2020 között (Mrd Ft)

Forrás: Saját szerkesztés MNB (2021) alapján

2015-ben egy új termékkel bővült a hitelintézetek kínálata, ami a Családi Otthonteremtési Kedvezmény (CSOK). A vissza nem térítendő támogatást és a kedvezményes hitelt gyermekkel rendelkező, vagy gyermeket vállaló ügyfelek, házaspárok igényelhetik még ma is. A kedvezményes hitelek köre 2019. év közepén a Babaváró hitellel bővült, amit pedig gyermeket vállaló házaspárok igényelhetnek. Ezen hitelek megjelenésének is hatása volt arra, hogy az utóbbi években folyamatosan nőtt a hazai hitelezés.

A 2019 év végén kiobbant koronavírus járvány azonban új fejezetet nyitott. Az intézkedések közül a törlesztési moratóriumot emelném ki, amelynek hatása lehet a nemfizetésre. A moratóriumot 2020. márciusában vezettek be, mellyel kezdetben bárki élhetett, majd az idő múlásával folyamatosan szigorodtak a feltételek. Hazánkban több, mint másfél millió ügyfél élt ezzel a lehetőséggel.



4. ábra: A moratóriumban érintett hitelek darabszám szerinti aránya a háztartásoknál 2021 június ában

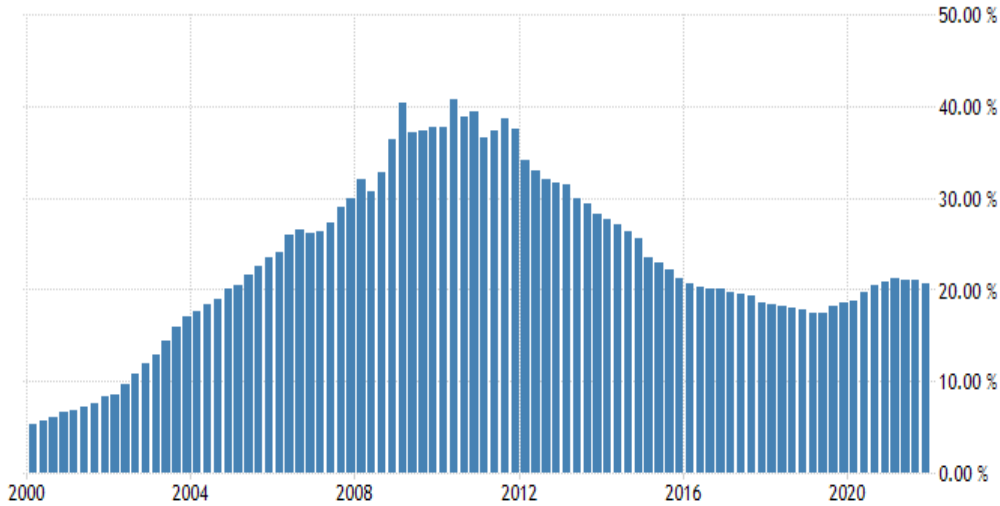
Forrás: Tasi Tibor: *Adóspárti a magyar hitelmoratórium* p30.

2022 utolsó negyedévére jelentős visszaesés volt megfigyelhető a lakossági hitelkibocsátás esetében. A személyi hitelek esetében 2022 utolsó negyedévében 18%-kal kevesebb hitelt bocsátottak ki, mint az előző év azonos időszakában, ami több, mint 20 milliárd forint értékű csökkenést jelent. A lakáscélú hitelek esetében pedig 2022 utolsó negyedévében több, mint 175 milliárd forinttal kevesebb hitelt bocsátottak ki, mint az előző év azonos időszakában, ami 54%-os csökkenésnek felel meg. (MNB, 2023a)

2.2 Magyar hitelkörkép, eladósodottság

Mivel kutatásom fő fókusza a lakossági hitelezés, illetve a lakossági hitelezés nemfizetése, így fontosnak tartottam megvizsgálni a magyar lakosság hitelezéssel kapcsolatos jellemzőit, eladósodottságát, valamint a felvett, illetve a nemteljesítő hitelek alakulását.

Az eladósodottság mérésére a háztartási hitelek GDP-hez viszonyított arányát választottam, az utóbbi 20 év alakulását figyelembe véve.



5. ábra: A magyar háztartások adóssága a GDP arányában 2001-2021

Forrás: Trading Economics

A 5. ábrán jól látható, hogy kb. a 2009-es évig egy növekvő tendencia figyelhető meg, melyet néhány év stagnálás követ. Az ezt követő években azonban a mutatószám értéke folyamatosan csökkenő tendenciát mutat, elérve az utóbbi évek mélypontját 2019-ben, ahonnan ismét enyhe növekedésnek indult.



6. ábra: A magyar lakosság fogyasztói hiteleinek alakulása 2001-2021 (Mrd Ft)

Forrás: Trading Economics

A fogyasztói hitelek alakulása ettől eltérőbb képet mutat a 6. ábra alapján. A 2009-es csúcsot követően a felvett hitelek összege folyamatosan csökkent, azonban ezt 2016 után egy gyorsütemű növekedés követte, mely feltehetően részben a 32/2014 MNB rendeletnek köszönhető. 2019-ben a felvett fogyasztói hitelek értéke meghaladta a 2009-es csúcsponti értéket, 2021-re pedig megduplázta azt.

A hitelválság során számtalan rossz emlék rögződött be az emberek gondolataiba, mely szintén a 2016-ig csökkenő hitelállomány háttérében állhat. A rossz emlékeknek köszönhetően egyre jobban nőtt a bankok, illetve a pénzügyi szektor irányába mutatkozó bizalomhiány. Lautenschläger (Lautenschläger, 2015) gondolatai szerint a bizalomhiány kialakulásának

egyik főszereplője a 2008-2009-ben bekövetkezett válság, mely nagymértékben megrendítette az emberek bizalmát, amit a mai napig sem sikerült teljes mértékben visszaépíteni. A bizalom elvesztése azonban több témát illetően is aggodalomra ad okot:

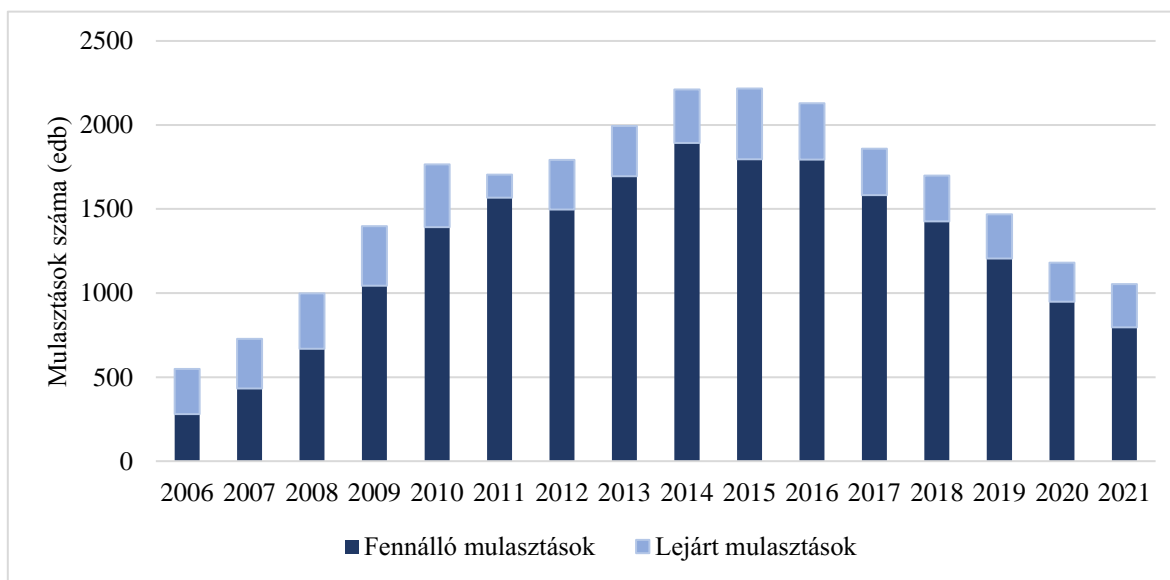
- Hátráltatja a bankok megfelelő működését és az erőforrások újraelosztását,
- Hátráltatja a növekedést,
- A stabilitás elvesztéséhez és költséges válságokhoz vezet.²

Az általános vélemény az, hogy a válság kirobbanása a visszafizethetetlen hitelekre vezethető vissza.

A hitelfelvétel alapvetően egy hosszútávú döntés, és számos kockázatot rejt, emiatt fontos a megfontolt gondolkodás, hiszen egy nem megfelelően átgondolt döntés miatt, akár az emberek megélhetése is veszélybe kerülhet. A törvényi megfogalmazás szerint mulasztásnak az alábbi minősül: „Mulasztásnak a KHR törvényben megfogalmazott, folyamatosan, 90 napnál hosszabb ideig fennálló és az aktuális legkisebb minimálbélnél nagyobb összegű törlesztést nevezük.” (KHR, 2017, p17)³

Az MNB aktuális tájékoztatása alapján: „Ha a tartozás mértéke meghaladja a késelemben esés időpontjában érvényes legkisebb összegű havi minimálbért, és ez a minimálbér összeget meghaladó késelelem folyamatosan, több mint 90 napon keresztül fennállt, a hitelezője megküldi a késelelem tényére vonatkozó információkat a Központi Hitelinformációs Rendszerbe.” (MNB, Tájékoztató a fizetési késelelemmel és a hitelszerződések felmondásával kapcsolatban.)

A fennálló és a lezárt mulasztásokat a KHR tartja nyilván, melyek az elmúlt években az alábbiak szerint alakultak.



7. ábra: A fennálló mulasztások számának változása státusz szerint 2006-2021 között

Forrás: Saját szerkesztés a KHR éves tájékoztatói alapján

² <https://www.ecb.europa.eu/press/key/date/2015/html/sp150928.en.html>

³ Központi Hitelnyilvántartó Rendszer, továbbiakban KHR

Ahogy az 7. ábrán is megfigyelhető a fennálló mulasztások száma egészen 2014-ig folyamatosan növekvő tendenciát mutat, azonban 2015. évet követően csökkenő tendencia figyelhető meg, mely a 32/2014 MNB rendeletnek köszönhető, amely meghatározta a jövedelem arányos törlesztő részlet (továbbiakban JTM) felső határát. Ezzel az addig működő hitelképességi vizsgálat mellé kötelezően beépült a JTM korlát ellenőrzés a hitelek kihelyezése során.

Késedelem szerint vizsgálva a hiteleket öt csoportot lehet elkülöníteni, melyekre vonatkozóan az MNB nyilvántartást vezet:

- késedelem nélküli hitelek
- 30 napon belüli késedelemmel rendelkező hitelek
- 31-90 nap közötti késedelemmel rendelkező hitelek
- 91-365 nap közötti késedelemmel rendelkező hitelek
- 1 éven túli késedelemmel rendelkező hitelek

Ezeket a hiteleket mind bruttó, mind pedig nettó értéken vezeti.

2.3 Az eladósodottság veszélyei

Könnyen hozhatja úgy az élet, hogy valakinek pénzügyi segítségre van szüksége. Ilyen esetekben azonban fontos, hogy megfelelően mérlegeljük, és alapos átgondolás után hozzuk meg a döntést, hiszen ezek hosszú évekre kihathatnak. Amennyiben nem a terveknek megfelelően alakulnak az események, vagy akár túlbecsültük pénzügyi teherbíró képességünket, súlyos anyagi vészhelyzetbe kerülhetünk.

A hitelfelvétel egyik legnagyobb kockázata, ha az igénylő nem tudja visszafizetni a felvett összeget. Ez több következménnyel járhat:

- késedelmi kamat felszámolása
- követeléskezelő költségeinek felszámolása
- KHR listára történő felkerülés.

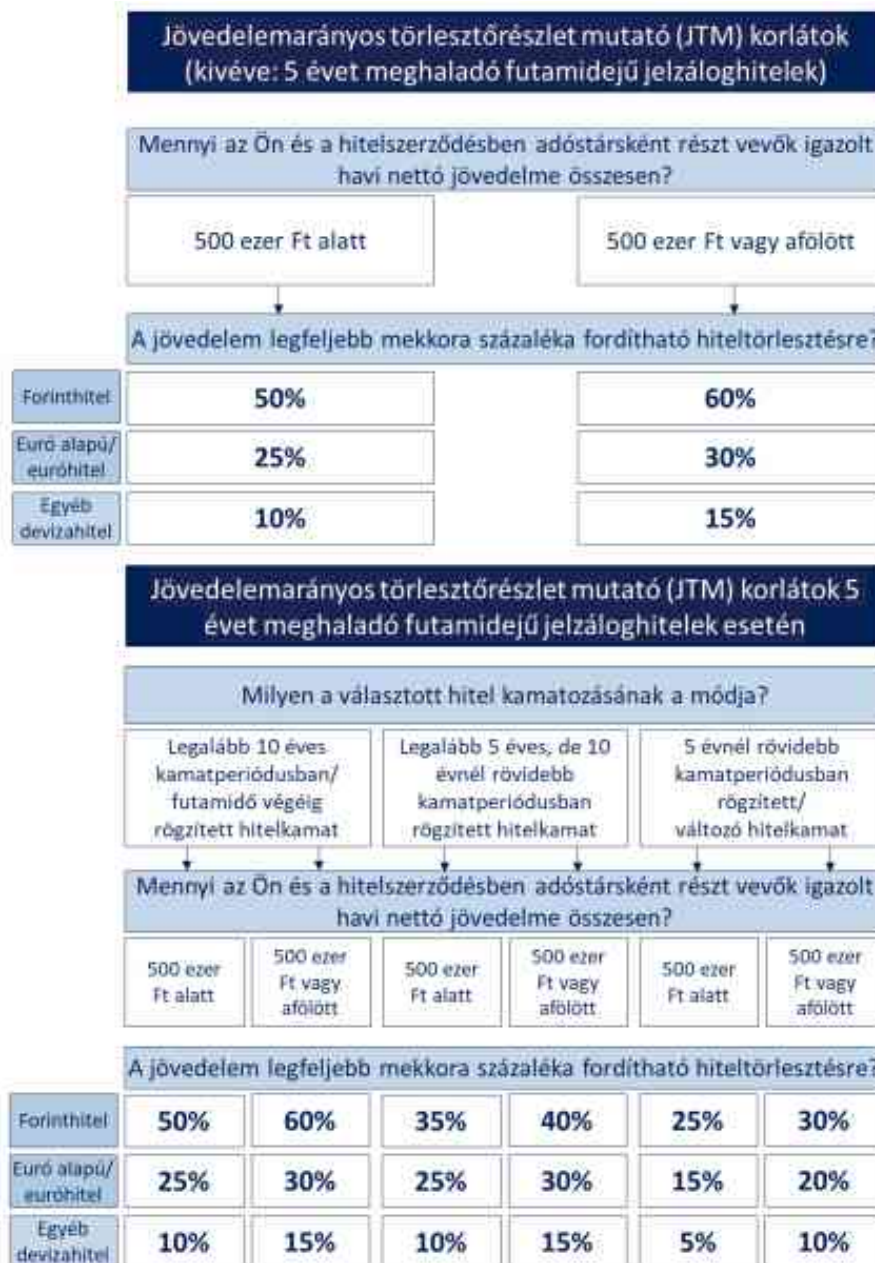
A legsúlyosabb következmény azonban a hitel fedezetül bejegyzett vagyontárgy elvesztése, amely akár lakhatási problémákat is okozhat, és amennyiben a fedezet nem elegendő a hitel teljes összegének rendezésére, szükséges lehet tovább fizetni a törlesztőrészeket.

Számos olyan intézkedést vezettek be az elmúlt években, amelyekkel a túlzott eladósodottságot igyekeztek megfékezni. 2015. január 1-jétől hatályos a JTM és HFM, melyeket a már korábban is említett 32/2014. MNB rendelet tartalmaz.

„A jövedelemarányos törlesztőrészlet-mutató (JTM) az ügyfelek rendszeres, legális jövedelmének meghatározott arányában korlátozza az új hitel felvételekor maximálisan vállalható törlesztési terheket, és ezáltal mérsékli az eladósodást.” (MNB, 2019)

„A hitelfedezeti mutató (HFM) a fedezett hiteleknél (pl. jelzáloghitelek, gépjárműhitelek) a fedezetek értékének arányában korlátozza a felvehető hitelek nagyságát.” (MNB, 2019)

Az alábbi táblázat foglalja össze a JTM-val kapcsolatos jelenleg (2023. május) érvényben lévő szabályokat.



8. ábra: JTM korlátok

Forrás: MNB tájékoztató a túlzott eladósodottság kockázatairól

Ezen korlátok meghozatala során az MNB folyamatosan figyelemmel kíséri a hitelpiac alakulását, és szükség esetén módosítja a szabályozásokat, melyre több alkalommal is volt már példa. A legújabb módosítás 2023. július 1-jén lép életbe. Az egyes változtatások háttérében az elmúlt időszak inflációs környezete és bérdinamikája áll. Jelenleg 500.000 Ft feletti nettó jövedelem esetén van lehetőség magasabb JTM limit alkalmazására, mely összeg 600.000 Ft-ra növekedik az új szabályozás hatályba lépését követően. Tehát ezzel a jövedelemmel rendelkezők jövedelmük akár 60%-át fordíthatják hiteltörlesztésre. (MNB, 2023b)

2.4 Hitelek csoportosítása lakossági hitelezés esetében

Jelenleg hazánkban számtalan fajta hitel igényelhető. A hitelek csoportosítása több szempont alapján lehetséges. A következőkben néhány lehetséges csoportosítást vizsgálunk meg.

- Időtáv szerint:
 - rövid lejáratú (1-2 év)
 - közép lejáratú (1-5 év)
 - hosszú lejáratú (5+ év)
 - lejárat nélküli hitelek.
- Folyósítás időtartama szerint:
 - határozott idejű
 - határozatlan
- Hitelfelhasználás célja szerint:
 - szabadon felhasználható
 - meghatározott célra használható
- Hitel mögötti biztosíték szempontjából⁴:
 - ingatlanfedezetes
 - nem ingatlanfedezetes
- Kamatperiódus szerint:
 - fix: a teljes futamidő alatt változatlan
 - változó: a futamidő alatt kamatperiódusonként változik
 - változtatható: kamatláb mértéke kamatperiódusonként változhat a hitel futamideje alatt.

A változó kamatozású hitelek esetében azt mindenképpen fontos kiemelni, hogy ezek nagyobb kockázatot rejtnek, hiszen amennyiben változik a hitel kamata, akkor a havi törlesztőrészlet összege is változhat, akár jelentős mértékben is. Ilyen esetekben mindig nagyon fontos a kamatkockázat mérlegelése. Változó kamatozás esetében beszélhetünk referenciakamathoz kötött, 5 évre, illetve 10 évre rögzített kamatozásról. (Bankszövetség fogalomtára)

A Magyar Nemzeti Bank rendelkezik egy hitel- és lízingtermék választó programmal, mely összefoglalja a ma Magyarországon elérhető hiteltípusokat. Alapvetően a cél a hitelek csoportosítása, azonban a későbbi elemzések szempontjából érdekes hitelek definíciója is megtalálható ebben az alfejezetben.

Lízing

A lízing különlegessége, hogy a lízing fedezetét képező eszköz a futamidő végén kerülhet a lízingbe vevő birtokába. A lízingszerződés lehet ingatlan- és gépjárműlízing.

Gépjárműlízing: „A fogyasztók részére új és használt gépjárművek megvásárlásához kínált olyan finanszírozási forma, amelynek során a lízingbeadó pénzügyi intézmény a gépjárművet szerződéssel lízingbe adja a lízingbe vevőnek, azzal, hogy a futamidő végéig, illetve az esetleges végtörlesztésig a gépjármű tulajdonjogát a pénzügyi intézmény fenntartja.” (MNB)

⁴ A hitel biztosítéka hatással van a hitel feltételeire, minél értékállóbb a biztosíték, annál kedvezőbbek a feltételek.

Hitelek

Hitelek esetében az alábbi típusok elérhetők:

- Számlahitelek
- Fogyasztási hitelek
 - Személyi kölcsön
 - Áruhitel
 - Gépjárműhitel
- Jelzáloghitelek
 - Ingatlan célú hitelek
 - Szabadfelhasználású jelzáloghitelek
- Megtakarítás fedezete melletti hitelek
- Hitelkártyák

Fogyasztási hitel: „A fogyasztási hitelek a mindennapi élet szokásos használati tárgyainak megvásárlásához, javíttatásához, illetve szolgáltatások igénybevételéhez – a természetes személy részére – nyújtott kölcsön, valamint a felhasználási célhoz nem kötött kölcsön, ha azt a természetes személy nem üzletszerű tevékenysége keretében veszi igénybe.

Jelzáloghitel: A jelzáloghitel esetén a hitelintézet a hitel fedezetéül felkínált ingatlanra jelzálogjogot jegyeztet be. Nemfizetésnél érvényesíti jelzálogjogát, azaz értékesítheti az ingatlant, a ki nem fizetett tartozást, és annak kamatait pedig az így befolyt összegből biztosítja. E fedezet mellett jellemzően nagyobb hitelösszeg vehető fel hosszú lejáratra, más hiteltermékhez képest kedvezőbb kamatozással.” (MNB)

2.5 Hitelezéssel kapcsolatos alapfogalmak

Vannak fogalmak, amelyek kiemelten fontosak a hitelezés szempontjából, így az alábbi részben ezeket gyűjtöttem össze. Ezen fogalmak ismerete a nemfizetés szempontjából is fontos lehet, és a magyar lakosság pénzügyi kultúráját érintő felmérések alapján, a lakosság egy jelentős része nincs tisztában ezen fogalmak tényleges jelentésével, jelentőségével. Az egyes fogalmak meghatározásához a Bankszövetség fogalomtárát használtam.

„Adós: Az az ügyfél (természetes személy vagy szervezet), aki/amely hitel/kölcsönszerződést köt a bankkal, vállalva a felvenni szándékozott hitel/kölcsön visszafizetését.” (Bankszövetség fogalomtára)

„Futamidő: Az az időtartam, amely alatt a kölcsönt és annak járulékait vissza kell fizetnie az adósnak a szerződésben rögzített ütemezés szerint. A futamidő elteltével megszűnik a hitel/kölcsönszerződés, akkor is, ha a visszafizetés nem történt meg teljes összegben. Az adósnak ebben az esetben is a hitel/kölcsön szerződés alapján fennálló teljes tartozását meg kell fizetnie.” (Bankszövetség fogalomtára)

„Hiteldíj: A hitelnyújtás során a bank által felszámított összeg, amelynek elemei általában a következők lehetnek, de egy hitelügylet, illetve hitelkonstrukció nem feltétlenül tartalmazza valamennyi alább felsorolt hiteldíjat:

- ügyleti kamat
- kezelési költség

- rendelkezésre tartási jutalék
- hitelbírálati díj
- folyósítási jutalék vagy szerződéskötési díj vagy hitelkeret megnyitási díj – jellemzően a hitel/kölcsön kifolyósításakor kerül felszámításra
- szerződésmódosítási díj
- bankgarancia, bankkezesesség, okmányos meghitelezés díja, (ha ilyen kibocsátása kapcsolódik a hitelszerződéshez)
- előtörlesztés, illetve végtörlesztés díja,
- egyéb, speciális hitelműveletekhez kapcsolódó díjak, jutalékok, kezelési költségek.

A hiteldíj egyes elemeinek felsorolását, mértékét, felszámításuk eseteit általában a banki kondíciós listák, hirdetések, illetve az üzletszabályzatok, általános szerződési feltételek tartalmazzák.” (Bankszövetség fogalomtára)

„Kamat: A hitel/kölcsön „ára”, amelyet az adós köteles megfizetni a hitel/kölcsön nyújtójának, az igénybe vett kölcsön használatáért. A hitel/kölcsönösszeg százalékában kerül meghatározásra, általában éves kamatláb formájában.” (Bankszövetség fogalomtára)

„Kamatperiódus: Az ügyleti kamat változásának gyakoriságát mutatja. Az az időszak, amelyen belül a felvett hitel kamatlába nem változik. A bank csak a kamatperiódus fordulónapján változtathatja meg a hitel kamatát.” (Bankszövetség fogalomtára)

„Adóstárs: Az az ügyfél (természetes személy vagy szervezet), aki/amely az adóssal együtt hitel/kölcsönszerződést köt a bankkal, kölcsön igénybevétele céljából és vállalja annak visszafizetését.” (Bankszövetség fogalomtára)

„Teljes hiteldíj mutató (THM): Teljes Hiteldíj Mutató, amelyet a fogyasztónak nyújtott hitelről szóló törvényben meghatározott valamennyi kölcsön és pénzügyi lízing esetén fel kell tüntetni. A THM egyéves időszakra vetített százalékos érték és jogszabályban meghatározott módon tartalmaz minden olyan díjat, költséget, amelyet az adósnak a tőkén felül kell megfizetnie szerződésszerű teljesítés esetén (kamat, kezelési költség, hitelbírálati díj, értékbecslés díja, hitelközvetítőnek fizetendő díj, ingatlan-nyilvántartási eljárás díja, kötelező bankszámlavezetés díja, kötelező biztosítás díja stb.) Nem tartalmazza a késedelemi kamatot, a közjegyzői eljárás díját, a futamidő hosszabbítás (prolongálás) díját, a bankszámla használatának díját stb. 2012. áprilisától a mértéke törvényben maximált. A THM szerepe az, hogy a hitelt/kölcsönt felvenni szándékozó lakossági ügyfelek számára előzetes tájékoztatást adjon, tegye lehetővé az összehasonlítást a különböző hitelezők hasonló termékei, ajánlatai között.” (Bankszövetség fogalomtára)

2.6 Egyéb, a hitelezéssel összefüggő, nehezen mérhető tényezők

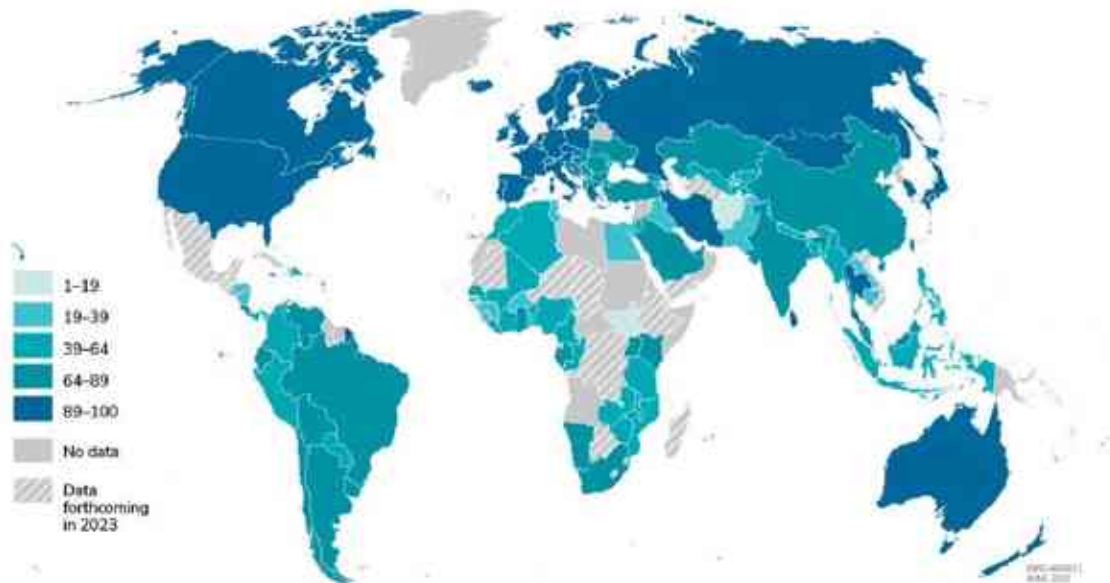
Ebben a fejezetben alapvetően a pénzügyi kultúrát vizsgálom, azonban az OECD megközelítése szerint a pénzügyi kultúra a pénzügyi inklúzió része, így elsőként ez utóbbira térek ki.

2.6.1 Pénzügyi inklúzió

A pénzügyi inklúzió egy folyamat, mely biztosítja a pénzügyi termékekhez és szolgáltatásokhoz való hozzáférést a hátrányosabb helyzetű személyek számára is megfizethető, tisztességes és átlátható módon. Egy korábbi felmérésben számos érdekes és elgondolkodtató megállapítás született 2009-ben:

- a világ felnőtt lakosságának több, mint fele nem vesz igénybe pénzügyi szolgáltatásokat sem megtakarítás sem pedig hitelezés céljából,
- ezen személyeknek közel 90%-a Afrikában, Latin-Amerikában, Ázsiában vagy a Közel-Keleten él,
- a pénzügyi szolgáltatásokat igénybe vevő felnőttek 2/3 része kevesebb, mint napi 5 dollárból él Afrikában, Ázsiában és a Közel-Keleten. (Chaia et. al., 2009)

2021-ban készült egy újabb felmérés a területtel kapcsolatosan a Világban felügyelete alatt, melyben az előzőhöz képest már egy kedvezőbb kép rajzolódott ki. A bankszámlákkal rendelkezők aránya 50%-kal nőtt, így 2021-ben a felnőtt lakosság már 76%-a rendelkezett bankszámlával, ez az arány a fejlődő országokban 71% volt. Az egyes országok esetében az arányokat a 9. ábra mutatja: (Worldbank, 2021)



9. ábra: A felnőtt lakosságon belül a bankszámlával rendelkezők aránya 2021-ben

Forrás: Worldbank: The Global Findex Database 2021

2.6.2 Pénzügyi kultúra

Béres-Huzdik úgy fogalmazott, hogy a hitelezés a gazdaság egy fontos mozgatórugója, melynek ahhoz, hogy a gazdaság megfelelő teljesítőképességgel rendelkezzen jól kell funkcionálnia. Azonban maga a hitelezés nem elegendő ehhez, fontos, hogy a hitelfelvevők tudatos pénzügyi döntéseket hozzanak. A hitelezéssel összefüggő döntéseket befolyásolhatja az ügyfél pénzügyi ismerete, illetve különböző tudatalatti mintázatok is. A pénzügyi ismeretek

vizsgálatára alkalmas eszköz lehet a pénzügyi kultúra, továbbá fontos lehet megismerni a fogyasztói szokásokat is, mert a döntéseink közel fele nem tudatos. (Béres-Huzdik, 2022)

A szerzőpáros véleményével maximálisan egyetértek, a hitelezés esetében egy fontos tényező a pénzügyi kultúra, és emiatt néhány gondolatot mindenképpen szeretnék ennek a témának szánni a disszertációmban is. Úgy gondolom, hogy jobb pénzügyi ismeretekkel jobb döntéseket tudunk hozni.

Napjainkban azonban nagyon széles és összetett a kínálat a pénzügyi piacon, és ahhoz, hogy valaki magabiztosan el tudjon igazodni az információk rengetegében, megfelelő pénzügyi ismeretek szükségesek. (Kovács et al., 2014) Elegendő egyetlen nem jól átgondolt döntés, és olyan nem várt helyzet adódhat, ami az emberek életének hosszú éveire kihat, gondoljunk csak a nem túl régen történt devizaválságra.

A pénzügyi kultúrát azonban nehéz definiálni, mert nem létezik egy általánosan elfogadott fogalom. Ez talán annak köszönhető, hogy csupán a XX. század elején kezdtek a kutatók ezzel a területtel foglalkozni, mára viszont kedvelt és fontos kutatási területté vált. Az MNB szerint a pénzügyi kultúra „a pénzügyi ismeretek és képességek olyan szintje, amelynek segítségével az egyének képesek a tudatos és körültekintő döntéseikhez szükséges alapvető pénzügyi információkat azonosítani, majd azok megszerzése után azokat értelmezni, és ez alapján döntés hozni, felmérve döntésük lehetséges jövőbeni pénzügyi, illetve egyéb következményeit.” (MNB-PSZÁF, 2008)

A következőkben néhány tanulmány fontosabb megállapításait foglaltam össze.

Remund tanulmányában a pénzügyi kultúra különböző meghatározásának elemeit 5 kategóriába sorolta:

1. Pénzügyi szolgáltatásokról szóló ismeretek: a tudás javítja az egyén pénzügyi jólétét
2. Képesség a pénzügyekről való kommunikáláshoz
3. Személyes pénzügyek kezelésére való alkalmasság
4. Képesség a megfelelő pénzügyi döntések meghozatalához
5. Magabiztosság a hatékony jövőbeni pénzügyi tervek készítéséhez. (Remund, 2010)

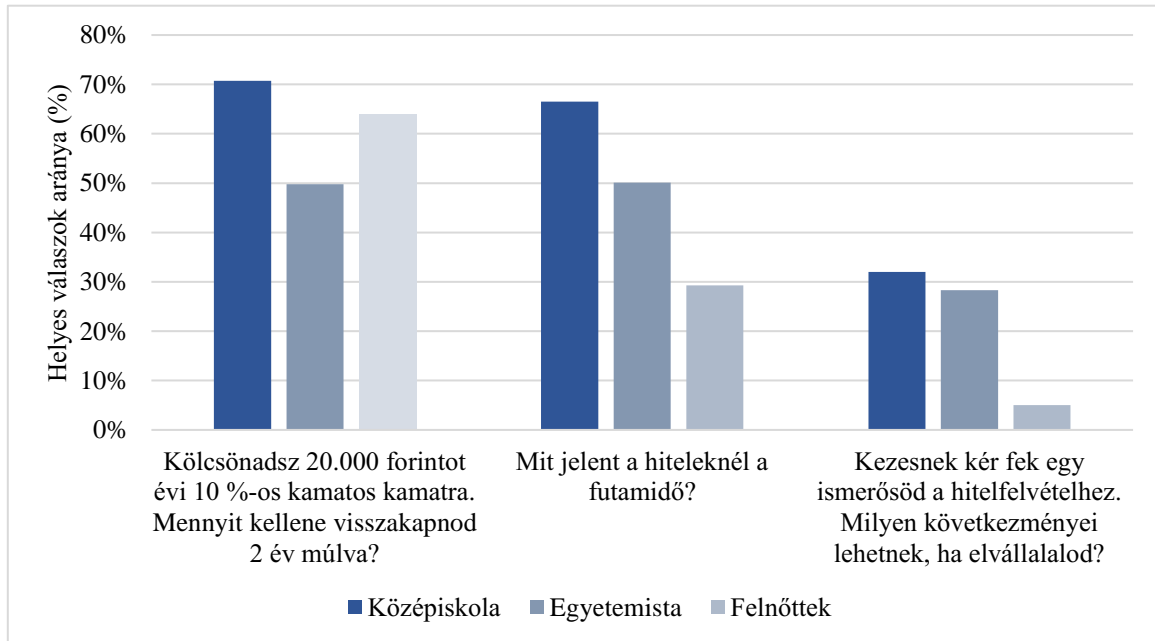
Botos et al. úgy fogalmazott, hogy hitelezés esetében a pénzügyi kultúra megnyilvánulhat abban, hogy mennyire jól mérik fel a fogyasztók a hitelezéssel kapcsolatos teherbíró képességüket, illetve a hitel kockázatait. Kutatásuk a közép-alföldi régióra fókuszált, és arra a következtetésre jutottak, hogy a fogyasztók jelentős része a devizahitel felvételekor nem volt tisztában az árfolyamban rejlő kockázatokkal. (Botos et al., 2012)

Béres és szerzőtársai úgy gondolták, hogy a pénzügyi személyiség két részből áll, és a pénzügyi személyiség segítségével „következtethetünk az adós személyiségre, viselkedésmintáira, és ami a bank szempontjából elsődleges, hogy a hitelt képes lesz-e valaki visszafizetni.” (Béres et al., 2015, p27.) A gyerekkorban megszerzett attitűdöknek fontos hatása van a felnőttkori viselkedésre a hitelfelvétel és hiteltörlesztés során is. Ezek a gyerekkorban kialakult viselkedésformák azonban csak lassan és nehezen változtathatók meg. (Nagy-Tóth, 2012)

Kovács-Révész-Ország a középiskolások pénzügyi kultúrájának mérését végezte el. Arra a következtetésre jutottak, hogy a középiskolások pénzügyi kultúrájának szintje nem tekinthető jónak, illetve tudásukat nem tudják a gyakorlatban alkalmazni. Több terület is kritikuskak bizonyult, melyek között szerepelt a hitelekkel kapcsolatos tudás is. (Kovács et al, 2014)

Az Econventio évek óta foglalkozik a pénzügyi kultúra kutatásával több korosztály esetében is. Mivel a különböző korcsoportok esetében ugyanaz a teszt került kitöltésre, így az összehasonlítás segítségével megvilágításra kerülnek az egyes csoportok esetében tapasztalható problémás területek.

Az 10. ábrán kiemeltem néhány kérdést, fogalmat, melyek a hitelezés szempontjából kiemelten fontosak.



10. ábra: Helytelen válaszok aránya az Econventio tesztek esetében korcsoportonként

Forrás: Saját szerkesztés az Econventio teszt eredményei alapján

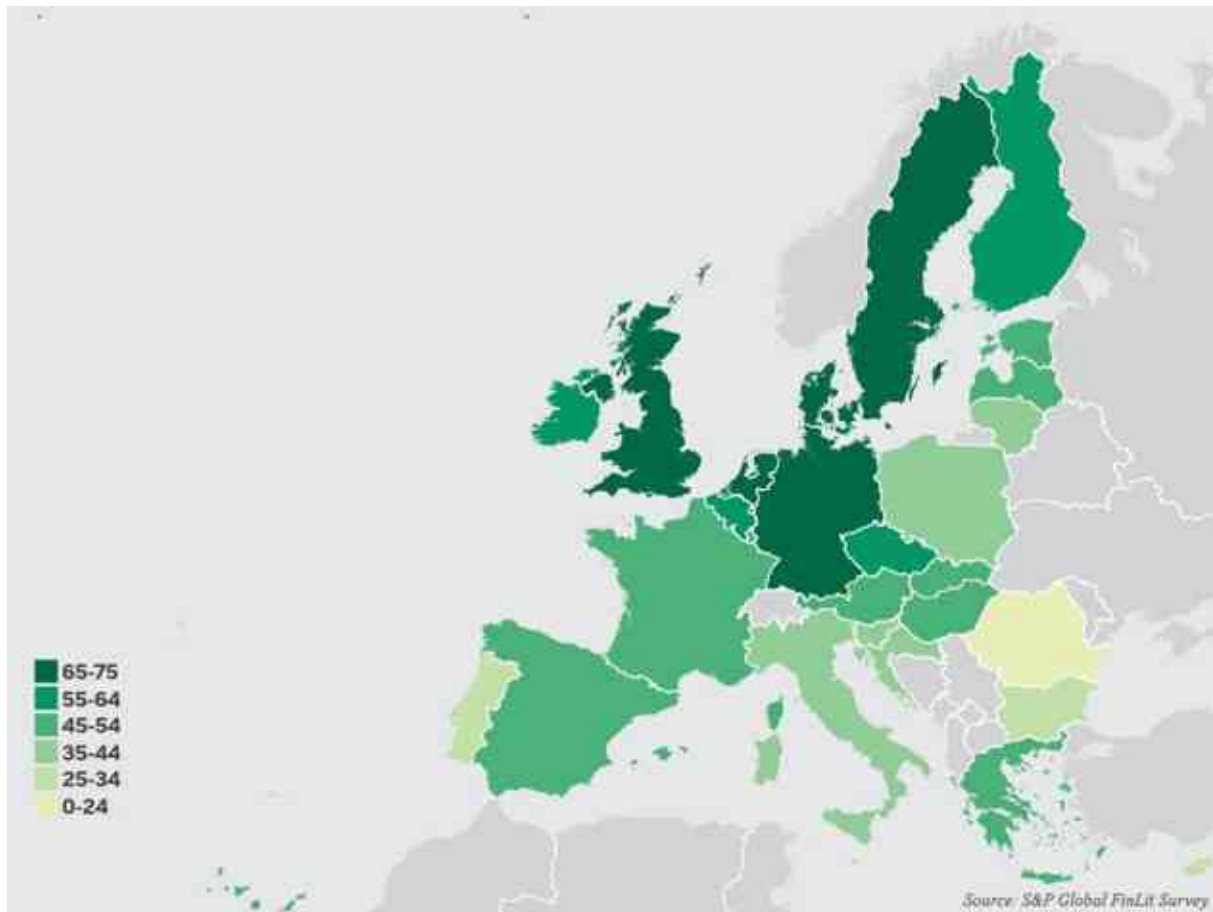
Az ábráról leolvasható, hogy a kamatos kamatszámítással minden korosztálynak problémája adódott, a legkedvezőbb arányt az egyetemisták érték el, közülük minden második fő tudta alkalmazni a kamatos kamatszámítást. Az egyetemisták a futamidő esetében is hasonló teljesítményt értek el. Meglepő azonban, hogy a felmérésben résztvevő felnőttek közel harmada nem volt tisztában a futamidő jelentésével. Felmerülhet a kérdés, ezen ismeretek nélkül lehet-e átgondolt döntést hozni. (Econventio, 2015)

Potóczki cikke alapján az mondható el, hogy a pénzügyi kultúra szintje hazánkban összességében átlag alattinak minősíthető. Ezt a következtetést az OECD által kezdeményezett nemzetközi kutatás alapján vonta le. A kutatásban részt vett 30 ország közül Magyarország a 22. helyen végzett. Az OECD felmérésében a megszerezhető 21 pontból 12,4 pontot ért el Magyarország, mellyel 0,8 ponttal elmaradt az átlagos értéktől. Az összesített eredmény három részeredményből tevődött össze. A pénzügyi hozzáállás területén átlag feletti eredmény született, a pénzügyi ismeretek átlagosnak minősültek, míg a pénzügyi magatartás átlag alattinak. (Potóczki, 2017)

A 2016-ben készült felmérést követően az OECD megvizsgálta a felnőttek pénzügyi kultúráját 2017-ben és 2020-ban is. 2020-ban Magyarország a 14. helyet érte el, tehát a középmezőnyben helyezkedett el a megszerzett 12,3 ponttal. A felmérésben résztvevők átlagos értékelése azonban jelentősen csökkent 2016-hoz képest, és Magyarország értéke már csak 0,5 ponttal

maradt el az átlagostól. Pénzügyi hozzáállás és pénzügyi ismeretek esetében átlag feletti eredmények születtek, azonban a pénzügyi magatartás területén továbbra is átlag alatti teljesítményről lehet beszélni. (OECD, 2016; OECD, 2017; OECD 2020)

Egy másik nemzetközi kutatást az S&P végzett, mely szerint a magyar lakosság 46%-a pénzügyi analfabéta. Európai szintén hazánk a középmezőnyben van. A 11. ábrán láthatjuk, hogy a különböző európai országok esetében változatosan alakul a pénzügyi kultúra szintje. A legrosszabb eredménnyel Románia rendelkezett. (Lusardi et al., 2015)



11. ábra: A pénzügyi kultúra szintjei az Európai országokban

Forrás: *Financial literacy around the World*, Klapper et al., 7., 2015

Béres és Huzdik a pénzügyi kultúra megjelenése makrogazdasági szinten című cikkében összegyűjtésre kerültek a pénzügyi kultúra demográfiai jellemzői, mely alapján magasabb pénzügyi kultúrával rendelkeznek:

- a férfiak,
- a diplomával rendelkezők,
- a magasabb jövedelemmel, illetve nagyobb megtakarítással rendelkezők,
- akiknek több munkatapasztalata van,
- a gyermektelen házaspárok. (Béres-Huzdik, 2015)

A Pénziránytű Alapítvány is végzett kutatást a területen. A felmérés során kiderült, hogy a résztvevők között a legnagyobb problémát a kamatos kamattal, illetve az egyszerű kamatszámítással kapcsolatos kérdés okozta. A felmérést 2010 és 2015-ben is elvégezték. A két időpont között történt néhány jelentős változás. „Jelentősen csökkent azok aránya, akik határozottan egyetértettek a következő állításokkal:

- Mielőtt valamit vásárolnék, alaposan megfontolom, hogy telik-e rá.
- Időben befizetem a számláimat.
- Hosszú távú célokat határozok meg és törekszem elérni őket.

Kevesebben utasították el azt az állítást, miszerint „Inkább a mának élek, holnap majd csak lesz valahogy.” Továbbá növekedett azok aránya, akik egyetértettek azzal, hogy „A pénz arra való, hogy elkölsük.” (Pénziránytű Alapítvány, 2015)

Sumit és szerzőtársai 2015-ben azt vizsgálták, hogy a pénzügyi kultúrának van-e hatása a jelzáloghitelek nemteljesítésére. A kutatásaik alapját egy amerikai, másodrendű hitelt nyújtó intézet szolgáltatta. Elemzéseik alapján azt a következtetést vonták le, hogy a jobb pénzügyi kultúra csökkenti a hitelek nemteljesítési arányát.

Összességében elmondható, hogy a pénzügyi kultúra egy nagyon fontos tényező, hiánya olyan problémákat okozhat, mint a túlzott eladósodás és a deviza problémák. A magyar lakosság pénzügyi kultúrájának szintje nem kielégítő, ami nagy kockázatot jelenthet. (Mit jelent a pénzügyi kultúra?, 2014)

Mivel ez a terület az én érdeklődésemet is felkeltette, így megvizsgáltam az általunk oktatott, gazdaságtudományi képzésen résztvevő egyetemisták pénzügyi kultúráját. A felmérés alapját az S&P kérdőíve adta. A kérdőív ezen része 5 kérdést tartalmazott, mely 4 tématerületet fedett le, ezek

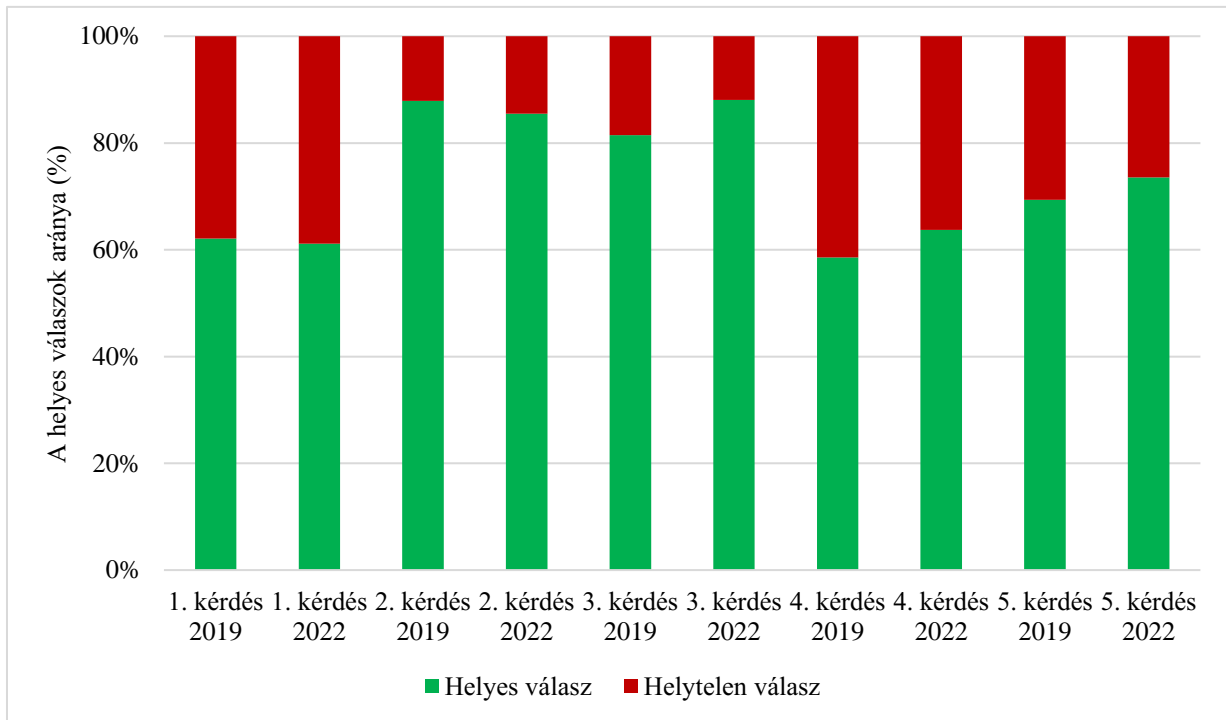
- a kamatos kamat,
- a kockázatkerülés,
- infláció,
- kamatszámítás.

A kutatásban az érdekelte ki a pénzügyileg kulturált jelzöt, aki legalább 3 tématerület esetében helyesen válaszolt. A felmérést két évben végeztem el, elsőként 2019-ben, majd pedig 2022-ben. 2019-ben 124, gazdaságtudományi képzésen résztvevő hallgató töltötte ki a felmérést, 2022-ben pedig 193. A kutatás a nappali tagozatos, alapképzéses hallgatókra fókuszált.

Elsőként a hallgatóknak arról kellett nyilatkozniuk, hogyan ítélik meg a saját pénzügyi kultúrájukat, melyet egy 1-7-ig terjedő skálán tehettek meg. A 2019-es értékelések kicsit jobban voltak, 2022-ben a hallgatók egy hetes skálán átlagosan 4,42-re értékelték a saját tudásukat.

Az első két kérdés a kamatos kamatszámítással kapcsolatos tudásra irányult. A helyes választ adók aránya szinte pontosan megegyezett a két vizsgált évben, a második kérdés esetében pedig egy apró csökkenés figyelhető meg. A harmadik kérdés a kitöltők diverzifikációját mérte fel, már 2019-ben is a kitöltők több, mint 80% kockázatkerülő volt, azonban ez az arány 2022-ban megközelítette a 90%-ot. A negyedik az inflációval kapcsolatos ismereteket vizsgálta, és megállapítható, hogy a mostani hallgatóknak nagyobb része volt tisztában az infláció fogalmával, ami véleményem szerint a jelenlegi gazdasági helyzetnek is köszönhető, azonban

ennek ellenére, erre a kérdésre érkezett a legkevesebb helyes válasz, csupán a kitöltők 64% volt tisztában a fogalommal. Az utolsó kérdés pedig az egyszerű kamatszámítás segítségével a numerikus képességeket vizsgálta, mely esetében szintén néhány százalékpontos növekedés figyelhető meg a helyes válaszok arányában. Az alábbi diagram az egyes kérdésekre adott helyes válaszok arányát foglalja össze 2019-ben és 2022-ben.⁵



12. ábra: A helyes válaszok aránya 2019-ben és 2022-ben

Forrás: Saját szerkesztés

A helyes válaszok alapján megállapítható, hogy a hallgatók 65%-a rendelkezik megfelelő pénzügyi ismeretekkel. Az S&P 2014-es felmérése alapján a magyar felnőtt lakosság 54%-a rendelkezett megfelelő szintű pénzügyi kultúrával.

Mindkét évben a cél az volt, hogy a hallgatókat homogén csoportba rendezzem, megvizsgálva, milyen az egyes csoportok pénzügyi kultúrája, és hogy milyen különbségek figyelhetők meg az egyes csoportok között. A hallgatók homogén csoportokba történő rendezését klaszterelemzéssel végeztem el. Elsőként a hierarchikus klaszterelemzésen belül a legközelebbi szomszéd módszert alkalmaztam a kiugró értékek azonosítására, majd ezt követően a Ward-módszer segítségével megállapítottam a klaszterek optimalás számát, végezetül pedig a K-means módszer segítségével létrehoztam a klasztereket.

A klaszterek képzése mindkét évben ugyanazon három változó mentén történt meg:

- Milyenek értékeli a saját pénzügyi kultúráját?
- Pénzügyek terén mennyire tartja magát kockázatvállalónak?
- Beosztja előre mire, mennyit költ (költésvetést készít)?

⁵ Az eredményeket Varga-Fodor (2022) is tartalmazza.

2019-ben négy homogén csoportot azonosítottam be, 2022-ben viszont ötöt.

Az egyes klaszterek közötti különbségeket vizsgálva hasonló eredményeket tapasztaltam én is, mint a korábbi kutatók.

A részletesebb elemzéseket, leírásokat a Fodor (2019) és Varga-Fodor (2022) tartalmazza.

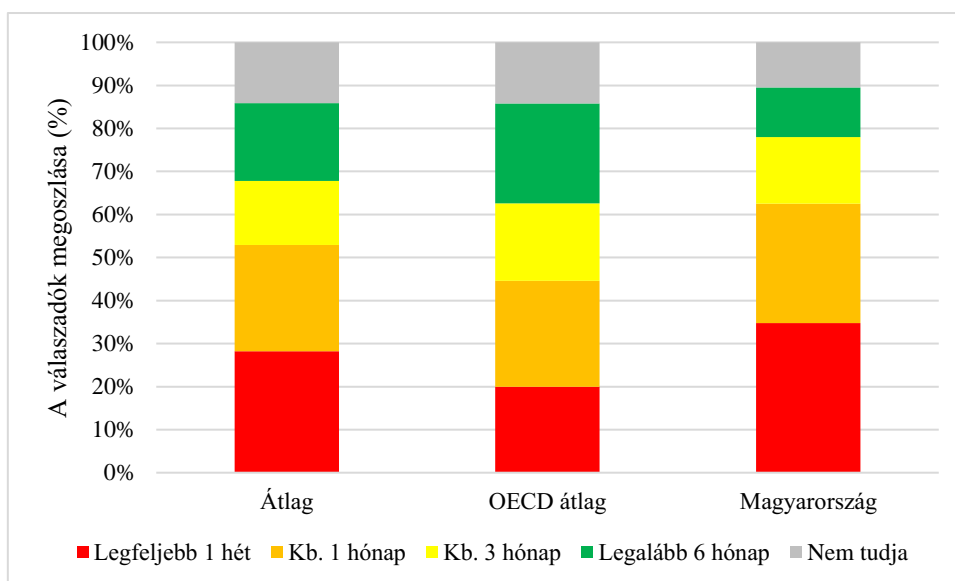
2.6.3 Pénzügyi reziliencia

A pénzügyi reziliencia azt a képességet jelenti, hogy egy egyén vagy háztartás mennyire képes ellenállni, felépülni egy negatív pénzügyi sokk esetén, ami lehet akár a munkahely elvesztése, egészségügyi probléma vagy egy nagyobb, nem várt kiadás. Úgy gondolom, ennek a tényezőnek is hatása lehet a nemfizetésre, hiszen minél felkészültebbek vagyunk, minél rugalmasabban tudunk reagálni, annál könnyebb lehet felépülni egy nem várt pénzügyi sokkból. Az OECD megközelítése szerint a pénzügyi reziliencia hat elemből áll:

A pénzmozgások nyomon követése: magában foglalja többek között a kiadások tervezését és nyilvántartását, a kiadások bevételekhez való viszonyát. Ezen a ponton sajnos Magyarország nem teljesített jól, a vizsgált országok közül az utolsó volt a rangsorban. A magyar válaszadók csupán 46,6%-a követi nyomon a pénzügyeket, ez az arány több, mint 20 százalékponttal kevesebb, mint az átlagos érték. Leginkább Thaiföldön (86,3%), Szlovéniában (84,1%) és Ausztriában (83,2%) élőkre jellemző ez a magatartás.

Átgondolt kiadások: szükségletek mérlegelése, pénzügyi kötelezettségek időben történő teljesítése tartozik ehhez az elemhez. A felmérés során megvizsgálták, hogy a válaszadók hány százaléka gondolja át alaposan a vásárlásokat. Ezen a területen Magyarország átlaghoz közeli értéket ért el, csupán kb. 5 százalékponttal maradt el tőle, ezzel a vizsgált 25 országból a 18. helyen szerepelt. Ezen a területen is Thaiföld (94,8%) eredménye volt a legkiemelkedőbb. Megvizsgálták továbbá azt, hogy a felnőttek milyen aránya fizetik ki a számlákat időben, mely esetben viszont a magyar válaszadók (81,5%) átlag feletti értéket értek el, és a 11. helyen álltak a vizsgált országok közül. A legmagasabb arányban az észtek (95,1%), míg a legalacsonyabb arányban a németek (50,4%) fizetik időben a számlákat.

Pénzügyi tartalék: azt vizsgálták, hogy a fő jövedelemforrás megszűnése esetén, mennyi időre elegendő tartalék áll rendelkezésre.



13. ábra: A pénzügyi tartalékok időbeli megoszlása 2020-ban

Forrás: Saját szerkesztés OECD (2020) alapján

Ahogy a 13. ábra is mutatja, magas azoknak az aránya, akiknek a tartalékjai maximum 1 hétre elegendők, továbbá elmarad az átlagos értékektől azoknak az aránya, akik legalább 6 hónapos tartalékkal rendelkeznek. A képzett tartalékok segíthetnek abban, hogy könnyebben át lehessen vészelní egy nem várt pénzügyi problémát, elkerülve ezzel az esetleges nemfizetővé válást.

Pénzügyi stressz: megvizsgálták, hogy a válaszadók milyen arányára volt igaz, hogy a kiadások tartósan meghaladják a bevételeket. Ez az állítás minden 5. magyar válaszadóra igaz volt, azonban ez jelentősen kedvezőbb, mint az átlagos érték. Leginkább a peruiakra igaz, hogy bevételeik nem fedezik a kiadásokat (62,4%), és legkevésbé pedig a koreaiakra igaz (11,8%).

Pénzügyi tervezés: megtakarítások képzése a nehezebb időszakokra, illetve a hosszabb távú pénzügyi célokra. Akárcsak a pénzügyi tartalékok elem, ez is fontos szerepet játszhat abban, hogy egy esetleges nem várt helyzetben elkerülhetővé váljon a nemfizető adóssá válás. A magyar válaszadók 51,3%-a rendelkezik aktív megtakarítással, azonban ez az érték jelentősen elmarad az átlagostól (70,4%). Továbbá minden harmadik magyar válaszadónak van hosszútávú pénzügyi terve, amelyre megtakarítást is képez.

Csalási tudatosság: lehetséges pénzügyi csalások ismerete, elkerülve ezzel az áldozattá válást. (OECD, 2020; OECD, 2021)

Összességében a pénzügyi reziliencia elemei között több olyan is szerepel, amelyek fontos szerepet játszhatnak a nemteljesítővé válás során, és ez a szerepük nem csak pénzügyi sokk esetén mutatkozhat meg.

3 Csődelőrejelzés fejlődéstörténete

A kutatásom fókusza ugyan a lakossági szektor, azonban lakossági területen elenyésző szakirodalom foglalkozik a nemfizetési kockázat vizsgálatával, így a módszertani alapot a vállalati szektorra készült csődmodellek jelentik. Emiatt ebben a fejezetben röviden összefoglalom a nemzetközi és hazai kutatásokat, fejlődéstörténetet.

3.1 Nemzetközi csődmodell-kutatások

A csődelőrejelzés szakirodalmát tekintve Virág Miklós neve mindenképpen említést érdemel. Szerzőtársaival együtt összefoglalta a pénzügyi elemzést és csődelőrejelzést, figyelmet fordítva utóbbi fejlődéstörténetére egyaránt. A csődelőrejelzés kutatása még nem rendelkezik 100 éves múlttal, az első próbálkozások az 1930-as évekre tehetők, azonban a mai értelemben használatos modellek csak az 1960-as években jelentek meg. A napjainkig eltelt idő alatt azonban a modellek és módszerek sokat változtak, mely az elemzési lehetőségek és az informatika fejlődésének is köszönhető.

A kezdeti „modellek” tulajdonképpen nem valódi modellek voltak még, hiszen sem a módszertani, sem pedig a technikai feltételek nem voltak adottak, így tulajdonképpen azokat a mutatószámokat keresték a kutatók, amelyek esetében a csődbe jutott és a túlélő vállalatok szignifikánsan különböznek, ezeket a mutatószámokat összehasonlították, és ez alapján próbáltak különböző összefüggéseket megállapítani. Mivel ez egy viszonylag új kutatási területnek számított, így hamar népszerűvé vált, sorra jelentek meg a témakörrel kapcsolatos tudományos munkák. Az 1930-as években ezen a területen kutatott Ramster és Foster (1931), Fitzpatrick (1931) Smith és Winakor (1935) is. Az egyes kutatók más-más mutatószám alapján készítették az elemzéseiket, azonban a legtöbbször által elfogadottnak a likviditási mutató bizonyult. A likviditási mutatón túlmenően megjelent, mint csődelőrejelzési faktor – a teljesség igénye nélkül- a jegyzett tőke/befektetett eszközök, összes adósság/jegyzet tőke is.

Az első egyváltozós elemzés Beaver nevéhez köthető. Elemzésébe 158 vállalatot vont be, és egyenlő arányban szerepeltek a csődbe ment és a túlélő vállalatok, azonban az adatok hiányossága miatt egyes esetekben ettől kevesebb elemmel végezte el elemzéseit. A legkisebb elemszám 117 volt. Az elemzések elvégzéséhez a vállalatok éves beszámolóit felhasználva 30 mutatószámot hozott létre, melyeket 6 csoportba sorolt:

- cash-flow mutatók,
- nettó jövedelem mutatók,
- az adósság összes eszközhöz viszonyított arányához kapcsolódó mutatók,
- likvid eszközök összes eszközhöz viszonyított arányához kapcsolódó mutatók,
- likvid eszközök rövid lejáratú adósságokhoz viszonyított arányához kapcsolódó mutatók,
- forgalmi mutatók.

Az elemzések célja, találni egy olyan mutatószámot, amely segítségével a lehető legpontosabban lehet kategorizálni a vállalatokat. Beaver megállapításai alapján az alábbi 3 mutató esetében figyelhetők meg a legélesebb különbségek:

- cash flow/eszközök

- cash flow/ adósság
- nettó árbevétel/adósság

A besorolást egy cut off pont segítségével végezte el. A módszerével 90%-os besorolási pontossággal tudta a vállalatokat kategorizálni. A módszer hátránya, hogy egyváltozós modellről van szó, tehát a kategorizálást egyetlen egy mutatószám alapján végzi, így, ha a különböző mutatószámok más besorolást eredményeznek, azt a módszer nem tudja kezelni. Többek között ez az oka, hogy ez a módszer nem terjedt el. (Beaver, 1966; Virág, 2004)

A módszertani fejlődéseknek köszönhetően az egyváltozós modelleket a többváltozós modellek váltották fel, melyeket a 60-as évek végén kezdtek el alkalmazni, és elsőként a diszkriminancia analízis vált népszerűvé. Az első valódi modell megalkotása Altman nevéhez fűződik, aki modelljét 5 pénzügyi mutatószámra építette fel, mely 95%-os megbízhatósággal tudta előre jelezni a fizetéseképtelenné válást.

Az Altman-féle csődmodellben az alábbi magyarázó változók szerepeltek:

- működő tőke/eszközök (X_1)
- visszatartott nyereség/eszközök (X_2),
- EBIT/eszközök (X_3),
- részvények piaci értéke/adósság (X_4),
- árbevétel/eszközök (X_5).

A mutatószámok segítségével a következő modellt alkotta meg:

$$Z = 0,012 \cdot x_1 + 0,014 \cdot x_2 + 0,033 \cdot x_3 + 0,006 \cdot x_4 + 0,099 \cdot x_5$$

A kategóriába soroláshoz a vágási pont értékét 2,675-nél húzta meg. Minél magasabb ez a Z érték, annál kisebb a valószínűsége a csődbe jutásnak. A modell alapján néhány évvel később egy bővített, hétváltozós modell jött létre. (Altman, 1968; Virág, 2004)

Altman öt és hétváltozós modelljének megjelenése között több elemző is foglalkozott a témával. Deakin szintén diszkriminancia analízist használt a csődelőrejelzéshez, és eredményeit egy 34 elemű mintán tesztelte. A modell besorolási pontossága 97% volt. (Deakin, 1972) Blum 1974-es modellje szintén 90% feletti besorolási pontossággal rendelkezett. (Blum, 1974)

Altman ötváltozós modelljének bővített változata 1977-ben jött létre, amikor Haldeman-nal és Narayanan-nal dolgozott együtt. Az új modell létrehozásának indokai között szerepelt, hogy az első modell létrehozásának körülményeihez képest megváltozott a gazdasági környezet, így indokolt volt a modell nagyvállalatokra történő kiterjesztése, illetve az iparvállalatokon túl a kereskedelmi vállalatok bevonása az elemzésbe. Mindemellett a számviteli előírásokban is változások történtek, amik szintén az új modell létrehozása mellett sorakoztattak fel érveket. Altman az első modell megalkotásához egy 66 elemű mintát használt, azonban az új modell esetében már egy ettől nagyobb, 111 elemű mintával dolgozott, melyben 58 túlélő vállalat kapott helyet, és megközelítőleg azonos arányban szerepeltek a kereskedelmi és iparvállalatok. Az elemzésbe bevont változók száma is öttel nőtt, így ezúttal 27 változó segítségével próbálta a modellt kialakítani, melyek közül végül 7 került be a végső modellbe. A ZETA-modell mutatói a következők voltak:

- eszközjövodelmezőség,
- jövodelmezőség stabilitása,

- adósságszolgálat,
- kumulatív jövedelmezőség,
- likviditás,
- kapitalizáció,
- méret.

Az egyes változókhoz tartozó súlyokat azonban napjainkig nem hozták nyilvánosságra.

Rövid távon a ZETA-modell hasonlóan teljesít, mint az 1968-as modell. Előnye a hosszútávú előrejelzési képességében rejlik. Az eredeti modell 5 évvel a bekövetkezés előtt 36%-os pontossággal tudta előre jelezni a csődöt, a ZETA modell azonban közel 77%-os megbízhatósággal tudta ezt megtenni. A ZETA-modell esetében a negatív értéket képviselő vállalatok fognak becsődölni, a pozitív értékkel rendelkezők pedig a túlélő vállalatok. Újdonság volt továbbá, hogy priori valószínűségeket és költségeket vittek be a modellbe, mellyel a kritikus értéket lehetett módosítani:

$$ZETA_C = \ln \frac{q_1 C_1}{q_2 C_2}$$

ahol a q_1 és q_2 a priori csoportba jutási valószínűség, a C_1 és C_2 pedig az első és másodfajú hibák költségei. (Virág, 2004)

Az Altman nevéhez fűződő modellek nem voltak reprezentatívak, és megközelítőleg azonos arányban szerepeltek a mintában a fizetőképes és a fizetéképtelen vállalatok. Az első reprezentatívnak tekinthető felmérés Ohlson nevéhez fűződik, ahol a fizetőképes és fizetéképtelen vállalatok aránya jól tükrözte a sokaságban előforduló arányokat. Ohlson nevéhez fűződik továbbá, hogy elsőként ő alkalmazta a csődelőrejelzési modellekben a logisztikus regressziót. A logisztikus regresszió alkalmazásának elterjedése a 80-as évekre tehető, mely kb. 10 évig a legnépszerűbb csődelőrejelzési módszernek számított. Az Ohlson által vizsgált mintába 2163 vállalat került, melynek a 4,85%-a csődbe ment, a maradék pedig túlélő volt.

Elemzéseibe az alábbi 9 mutatót vonta be:

- log (összes eszköz/GNP árszínvonal index),
- kötelezettségek/összes eszköz,
- forgótőke/összes eszköz,
- rövid lejáratú kötelezettségek/forgóeszközök,
- Dummy változó: értéke abban az esetben 1, ha a teljes tartozás értéke nagyobb, mint az összes eszköz értéke,
- nettó árbevétel/összes eszköz,
- üzleti eredmény/kötelezettségek,
- Dummy változó: értéke abban az esetben 1, ha a nettó árbevétel az elmúlt két évben negatív volt,
- nettó árbevétel változása.

Amennyiben a modell során alkotott P érték meghaladja a 0,038-as értéket, akkor a vállalat csődveszélyesnek minősül. A modell megközelítőleg 83%-os pontossággal rendelkezett. (Ohlson, 1980)

A logisztikus regresszió hosszú éveken keresztül népszerű elemzési módszer volt a csődelőrejelzés területén, a következőkben néhány nevesebb modell, illetve a fontosabb magyarázóváltozók találhatók, azonban ettől sokkal nagyobb számú kutató foglalkozott a területtel:

- Martin (1977)
 - nettó árbevétel/összes eszköz
 - bruttó leírások/nettó üzleti eredmény
 - kereskedelmi és ipari hitelek/összes hitel
 - bruttó tőke/kockázati eszközök
- Pantalone és Platt (1987)
 - nettó árbevétel/összes eszköz
 - saját tőke/összes eszköz
 - összes hitel/összes eszköz
 - kereskedelmi és ipari hitelek/összes hitel
 - lakásépítések százalékos változása
- Zavgren (1985)
 - készletek/értékesítés
 - lejárt követelések/készletek
 - (készpénz+rövid lejáratú befektetések) /összes eszköz
 - likviditási gyorsráta
 - adózás utána sajáttőke arányos nyereség
 - (kötelezettségek-passzív időbeli elhatárolások) /saját tőke
 - értékesítés/összes eszköz
- Keasey-McGuinness (1990)
 - saját tőke/kötelezettségek
 - beszerzés/szállítók
 - adózás előtti eredmény/értékesítés

A logisztikus regresszió mellett a probit analízist is igyekeztek meghonosítani a területen, azonban ez nem örvendett ekkora népszerűségnek, bár annak alkalmazása is mérföldkőnek számított módszertani szempontból. Elsőként Zmijewski alkalmazta a nyolcvanas évek közepén. Elemzéséhez egy 840 iparvállalatot tartalmazó adatbázist használt, melyben 40 csődbe ment vállalat volt. A modellbe az alábbi három változó került:

- nettó árbevétel/összes eszköz,
- összes tartozás/összes eszköz,
- rövid lejáratú kötelezettségek/forgóeszközök.

A létrehozott modellt egy, az elemzésből kihagyott részmintán tesztelte. A modell besorolási pontossága 98%. (Zmijewski, 1984)

A csődelőrejelzések esetében a következő újdonság a rekurzív particionáló algoritmusok megjelenése volt, mely a 80-as évek közepére tehető. A módszer első alkalmazói közé tartozik Altman, Frydman és Kao. Elemzéseik során több mutató mellett az alábbiakat is vizsgálták:

- nettó árbevétel/összes eszköz,
- log(forgóeszközök),
- EBIT/összes eszköz,
- log(kamatfedezet+15),
- készpénz/összes eszköz,

- likvid eszköz/összes eszköz.

A modell besorolási pontossága 94%, azonban jelentős különbség mutatkozott a helyes kategorizálást illetően a túlélő és a csődbe ment vállalatok között. A túlélő vállalatokat 99%-os pontossággal tudta a modell helyesen kategorizálni, azonban a csődbe ment vállalatok esetében ez az érték csupán 84% volt, összességében véve viszont, ez is magas aránynak tekinthető. (Frydman et al., 1985; Cséffai et al., 2019)

A 2000-es években McKee-Greenstein páros szintén ezzel a módszerrel próbált elemzéseket végezni, azonban végezetül a rekurzív particionáló algoritmusok használata nem terjedt el a szakirodalomban. (McKee-Greenstein, 2000)

A rekurzív particionáló algoritmusok után a következő nagy lépés a neurális hálók megjelenése volt, mely a 90-es évekre tehető, ezzel egyidejűleg a klaszterelemzési eljárások is egyre nagyobb szerepet kaptak. A neurális hálók első alkalmazása Odom és Sharda nevéhez köthető. Modelljüket az Altman által 1968-ban használt változókra építették. A minta 129 vállalatból állt, melyeket két csoportba sorolt. 74 vállalat alkotta a tanuló mintát, és 55 vállalat a tesztelő mintát. A tanuló minta esetében a besorolás tökéletes volt, tehát felülmúlta a diszkriminancia analízis által elért eredményeket. A teszt minta esetében a besorolás ugyan már nem volt tökéletes, de a 82%-os besorolási pontossággal jelentősen meghaladta a diszkriminancia analízis által elért besorolási pontosságot. (Odom-Sharda, 1990)

Tam-Kiang elemzéseiket bankok esetében végezték el, egy éves távlatban a neurális háló működött a legjobban, kétéves időtávban azonban a logisztikus regresszió eredményi bizonyultak a legjobbnak. (Virág-Kristóf, 2005)

A Coats-Fant páros a diszkriminancia analízis teljesítményét hasonlította a neurális háló teljesítményéhez, és az előzőkhez hasonló következtetésekre jutottak ők is.

A 90-es évek második felében Olmeda és Fernandez is foglalkozott a kutatási területtel, azonban ők spanyol bankok adatait dolgozták fel. Kutatásaikat a fentebb említett minden modell segítségével elvégezték (diszkriminancia analízis, logisztikus regresszió, döntési fa, neurális háló), melyek közül a legjobbnak a neurális háló bizonyult, 82,4%-os besorolási pontossággal. (Olmeda-Fernandez, 1997)

Zhang, Hu és Patuwo a neurális hálót a logisztikus regresszióval vetették össze. Előbbi 88,2%, míg utóbbi 78,6%-os besorolási pontosságot ért el. (Zhang et al., 1999)

A kutatási területtel, továbbá a különböző elemzési módszerek eredményeinek összehasonlításával foglalkozott a teljesség igénye nélkül:

- Fletcher-Goss (1993),
- Wilson-Sharda (1994),
- Boritz-Kennedy-Albuquerque (1995),
- Leshno-Spector (1996),
- Kiviluoto (1998),
- Laitinen-Kankaanpaa (1999),
- Tan (1999),
- Pompe-Bilderbeek (2000),
- Lin-McClean (2001),
- Charitou-Neophytou-Charalambous (2004).

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

3.2 Magyar csődmodell-kutatások

Hazánkban az 1991. évi XLIX. törvény biztosította a csőd szabályozásának jogi kereteit, ezzel lehetőséget adva a magyar kutatások megkezdéséhez. Az első publikált modell Virág Miklós és Hajdu Ottó nevéhez köthető. Elemzéseikhez diszkriminancia analízist és logisztikus regressziót alkalmaztak. Adatbázisuk 154 feldolgozóipari vállalatból állt. Fele-fele arányban szerepeltek a mintában a becsődölt és a túlélő vállalatok. Azok a vállalatok minősültek az elemzésükben csődbe ment vállalatnak, akik esetében 1992. augusztusában csődeljárás volt folyamatban. 16 mutató mentén végezték el elemzéseiket, végül az 1991. évi modell az alábbi változókat tartalmazta:

- likviditási gyorsráta (X_1),
- cash-flow/összes tartozás (X_2),
- forgóeszközök/összes eszköz (X_3),
- cash flow/összes eszköz (X_4).

A fenti változók alapján 2 modell született, melyeket egymásból kivonva, a centráló konstans elhagyva kapjuk meg az alábbi modellt:

$$z = 1,3566x_1 + 1,63397x_2 + 3,66384x_3 + 0,03366x_4$$

A 2,61612 értéknél nagyobb értéket elérő vállalatok sorolandók a túlélő vállalatok közé, és a modell 77,92%-os besorolási pontossággal kategorizált helyesen.

Logisztikus regresszió esetében a magyarázóváltozók száma bővült, a diszkriminanciaanalízishez képest, melyeket az alábbi táblázat foglal össze, a hozzájuk tartozó paraméterekkel együtt:

2. táblázat: A logisztikus regresszió modell változói és paramétereinek értékei

	Változó neve	paraméter értéke
β_0	konstans	3,432
β_1	likviditási gyorsráta (X_1),	-10,320
β_2	árbevétel arányos nyereség (X_2),	0,1439 E-01
β_3	cash flow/összes tartozás (X_3),	-4,438
β_4	forgóeszközök/összes eszköz (X_4)	-0,2292 E-01
β_5	vevők/szállítók (X_5)	8,170

Forrás: Saját szerkesztés Virág-Kristóf (2005) alapján

A modellhez tartozó vágási pont értéke 0,525, tehát ha egy vállalat esetében a csődvalószínűség nem haladja meg ezt az értéket, akkor az adott vállalatot a modell túlélő vállalatnak kategorizálja. A logisztikus regresszió teljesítménye meghaladta a diszkriminancia analízissel készült modell teljesítményét, 81,8%-ot ért el. (Virág-Kristóf, 2005)

A szerzőpáros 1996-ban bemutatott egy új csődmodell családot, mely az Early Warning System elnevezést kapta. Az elkészült modellek különlegessége, hogy az egyes nemzetgazdasági ágakra és ágazatokra specifikusan készültek el. Elemzéseiket diszkriminancia analízis segítségével végezték el, egy körülbelül tízezres nagyságrendű adatbázison. A nagy átfogó modell mellett készültek el az ágazati szintre lebontott modellek, szám szerint 30. A vizsgált ágazatok között szerepelt például:

- fafeldolgozás,
- textíliák gyártása,
- építőipar,
- ingatlanügyletek,
- közúti járműgyártás,
- víztermelés, -kezelés és elosztás,
- szárazföldi és csővezetékes szállítás.

A modell részleteit, egyenleteit a szerzők üzleti titkok miatt azonban nem publikálták, de azt tudni lehet, hogy a csődmodelles család felülmúlta a korábbi modell teljesítményét.

A későbbi években a Virág-Kristóf páros neurális hálók alkalmazásával próbálkozott, az úgynevezett „backpropagation” eljárás segítségével. Más kutatók eredményei azt mutatták, hogy a neurális háló akkor szolgáltatja a legjobb eredményt, ha a mintában (megközelítőleg) azonos arányban szerepelnek túlélő és csődbe jutott vállalatok. Elemzésüket egy 156 elemű mintán végezték el. A minta $\frac{3}{4}$ része alkotta a tanuló mintát. A véglegesnek ítélt modell két köztes réteggel rendelkezett, az első köztes réteg 6, míg a második köztes réteg 5 neuronból állt. Az elemzésük eredménye hasonló volt a nemzetközi tapasztalatokhoz, miszerint a neurális háló segítségével jobban teljesítő modell hozható létre. A neurális háló teljesítménye a diszkriminancia analízis teljesítményét 8,6 százalékponttal, a logisztikus regresszió teljesítményét pedig 4,7 százalékponttal haladta meg. (Virág-Kristóf, 2005)

Egy későbbi tanulmányban a Virág-Kristóf páros korábbi modelljüket iparági rátákkal egészítették ki. Az elemzéshez korábbi adatbázisukat használták, mely 78 túlélő és 78 csődbe ment vállalatot és 16 mutatószámot tartalmazott:

- likviditási gyorsráta,
- likviditási ráta,
- pénzeszközök aránya (%),
- CF és összes tartozás aránya,
- forgóeszközök aránya (%),
- tőkeellátottsági mutató (%),
- eszközök forgási sebessége,
- készletek forgási sebessége,
- vevők forgási sebessége,
- eladósodottság mértéke (%),
- saját vagyongaránya (%),
- bonitás,
- befektetett eszközök hosszú lejáratú hitelekkel fedezett aránya (%),

- forgóeszközök rövid lejáratú hitelekkel fedezett aránya,
- árbevétel-arányos nyereség (%),
- saját vagyonarányos nyereség (%).

Ezek a mutatók az iparági csoportnak megfelelően helyesbítettek. A minta felosztását a már korábban is alkalmazott 75-25% arányban tették meg. Az elemzéseket négy módszer segítségével hajtották végre. Diszkriminancia analízis esetében az alábbi függvényt kapták:

$$z = -14101x_1 + 33,74902x_2 - 66,36x_3 + 78,16907x_4 + 146,1178x_5 + 52,5579x_6$$

ahol,

X_1 : iparági átlaggal korrigált likviditási gyorsráta,

X_2 : iparági átlaggal korrigált forgóeszközök aránya,

X_3 : iparági átlaggal korrigált eszközök forgási sebessége,

X_4 : iparági átlaggal korrigált készletek forgási sebessége,

X_5 : iparági átlaggal korrigált saját vagyon aránya,

X_6 : iparági átlaggal korrigált saját vagyonarányos nyereség.

Amennyiben az egyenletbe helyettesítve a kapott érték kevesebb, mint 1,06252, akkor az adott vállalat túlélő vállalatnak minősül.

Logisztikus regresszió esetében az előzőtől kevesebb magyarázóváltozó bizonyult szignifikánsnak:

$$\Pr_{(fizetőképes)} = \frac{e^{0,04233+621,92432X_1-170,80129X_2-99,43514X_3-245,79408X_4}}{1 + e^{0,04233+621,92432X_1-170,80129X_2-99,43514X_3-245,79408X_4}}$$

ahol,

X_1 : iparági átlaggal korrigált likviditási gyorsráta,

X_2 : iparági átlaggal korrigált készletek forgási sebessége,

X_3 : iparági átlaggal korrigált vevők forgási sebessége,

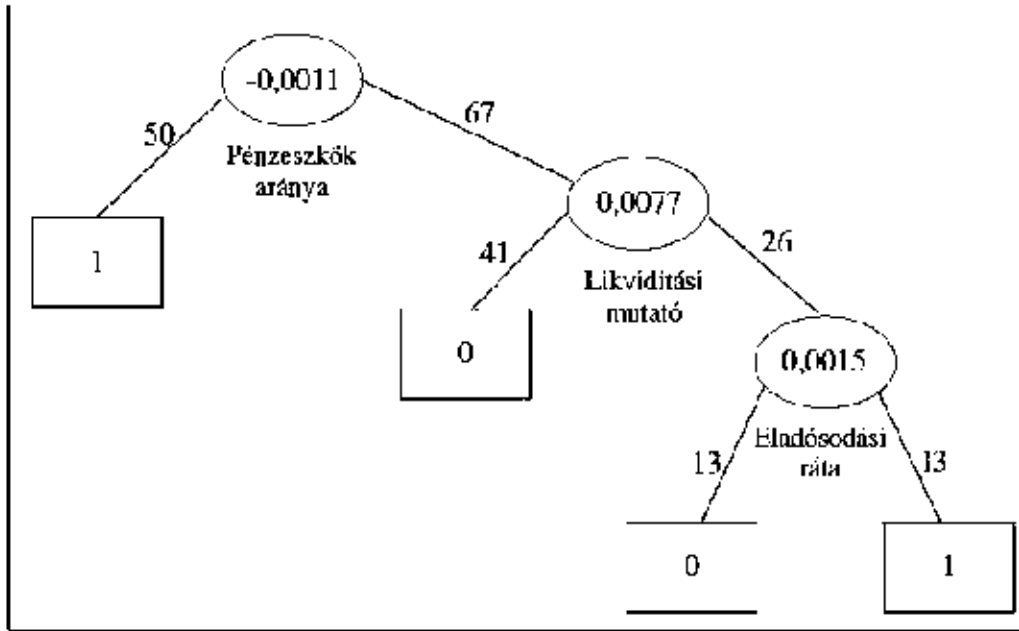
X_4 : iparági átlaggal korrigált saját vagyon aránya.

A modell vágási pontja 0,48.

A döntési fa esetében particionáló változónak az alábbiak minősültek:

- iparági átlaggal korrigált pénzeszközök aránya,
- likviditási mutató,
- eladósodási ráta.

A tanuló mintán létrejött döntési fát a 14. számú ábra tartalmazza.



14. ábra: Döntési fa a tanulási minta alapján

Forrás: Virág-Kristóf (2006) p32.

Neurális háló esetében a 2005-ben publikált eredményeket vették alapul, tehát 16 változót használtak fel, az első köztes réteg 6 neuronból állt, míg a második köztes réteg 4-ből.

A különböző modellek besorolási pontosságát a 3. számú táblázat tartalmazza. (Virág-Kristóf, 2005)

3. táblázat: Az egyes modellek besorolási pontossága a teszt és a tanuló minta alapján

	Besorolási pontosság	
	Tanuló minta	Tesztelő minta
Diszkriminancia analízis	82,91%	79,49%
Logisztikus regresszió	85,47%	74,36%
RPA	82,91%	79,49%
Neurális háló	82,91%	79,49%

Forrás: Saját szerkesztés Virág-Kristóf (2005) alapján

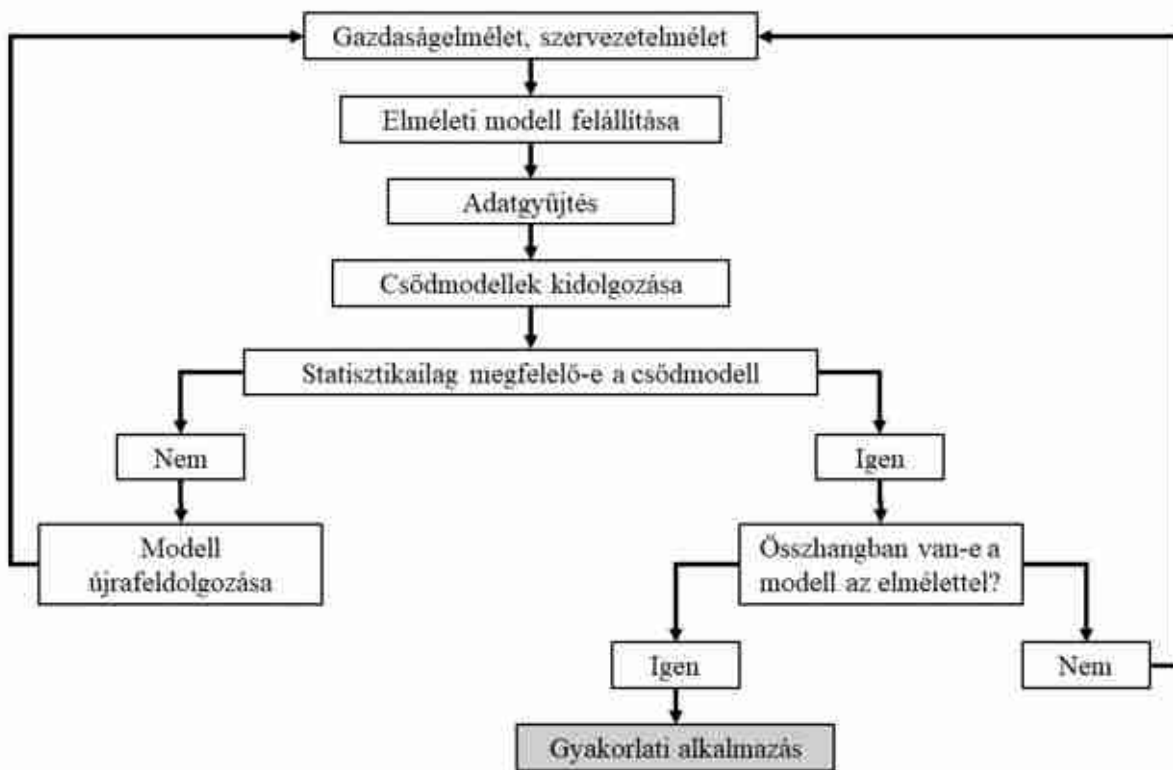
Megállapítható, hogy a tanuló minta esetében az egyes modellek közel azonos pontossággal tudták kategorizálni a vállalatokat, a tesztelő minta teljesítménye minden esetben elmaradt a tanuló minta teljesítményétől, a legnagyobb különbség a logisztikus regresszió esetében tapasztalható.

Virág és szerzőtársai úgy gondolták, további nagy hatással erre a kutatási területre Neophytou és Mar Molinero lesz, akik szerint egy paradigmaváltás fog bekövetkezni a csődjelölés tekintetében. (Virág et al., 2013)

4 A csődelőrejelzés módszerei

A lakossági hitelezés esetében a nemfizetés előrejelzésével foglalkozó szakirodalmak listája sajnos elég rövid. A kutatók már évtizedek óta foglalkoznak a vállalati csődök előrejelzésével, azonban a lakossági hiteligénylők besorolása nem kapott ugyanakkora figyelmet. Ebből kifolyólag tűztem ki a célt magam elé, hogy megpróbáljak egy olyan modellt kialakítani, amely a lakossági hiteligénylők besorolásával foglalkozik. Ehhez jó alapot nyújthat a vállalatok esetében alkalmazott módszerek köre, hiszen a vállalatok csődelőrejelzéséhez használt módszerek szakirodalma sokkal tágabb.

A csődelőrejelzés egy többlépcsős feladatnak tekinthető, melynek folyamatát az alábbi ábra szemlélteti:



15. ábra: Az empirikus vizsgálat folyamata

Forrás: Virág et al. (2013) p141.

A napjainkban rendelkezésre álló statisztikai elemzések köre nagyon tág. Annak függvényében, hogy a magyarázó és eredményváltozó milyen típusú, többféle elemzési lehetőséget tudunk megkülönböztetni. Az elemzési lehetőségek közül többet is használni fogok a kutatási hipotéziseim alátámasztásához. A függő és független változó típusa szerinti módszereket a 4. számú táblázat foglalja össze.

4. táblázat: A struktúravizsgáló módszerek egy részének összefoglalása

		Független változó (x)	
		Nem metrikus	Metrikus
Függő változó (y)	Nem metrikus	Keresztábra elemzés	Diszkriminancia elemzés, logisztikus regresszió, döntési fa, neurális háló
	Metrikus	Variancia elemzés	Korreláció, Regresszió elemzés

Forrás: Saját szerkesztés Sajtos-Mitev (2007); Hámori (2001) és Kristóf (2002) alapján

4.1 Diszkriminancia analízis

Az ismérvek közötti törvényszerűségek vizsgálatára számtalan statisztikai módszer kínálkozik. Az olyan esetekben, mely során az eredményváltozó kategória-kimenetű, klasszifikációról beszélhetünk. A klasszifikációs eljárások egy típusa a diszkriminancia analízis, ami egy olyan elemzési módszer, mely során a függő változó létező kategóriáiba kerülnek az esetek besorolásra. A függő változó általában névleges skálán mért változó, mely a nemfizetési kockázat előrejelzése során a “jó”, illetve “rossz” adós kimenettel rendelkezik, így ebben az esetben egy kétcsoportos diszkriminancia elemzésről beszélünk. A független változók esetében azonban fontos szempont, hogy ezek mérési szintje elérje legalább az intervallumskálát.

Az elemzés segítségével arra a kérdésre derülhet fény, hogy melyek azok a változók, melyek esetében a meglévő csoportok szignifikánsan eltérnek egymástól. Ennek az elemzése a kiinduláskor meglévő egyedek segítségével történik, majd az elkészült függvény segítségével megbecsülhetővé válik, hogy az újonnan megjelenő egyedek mely csoportba fognak tartozni.

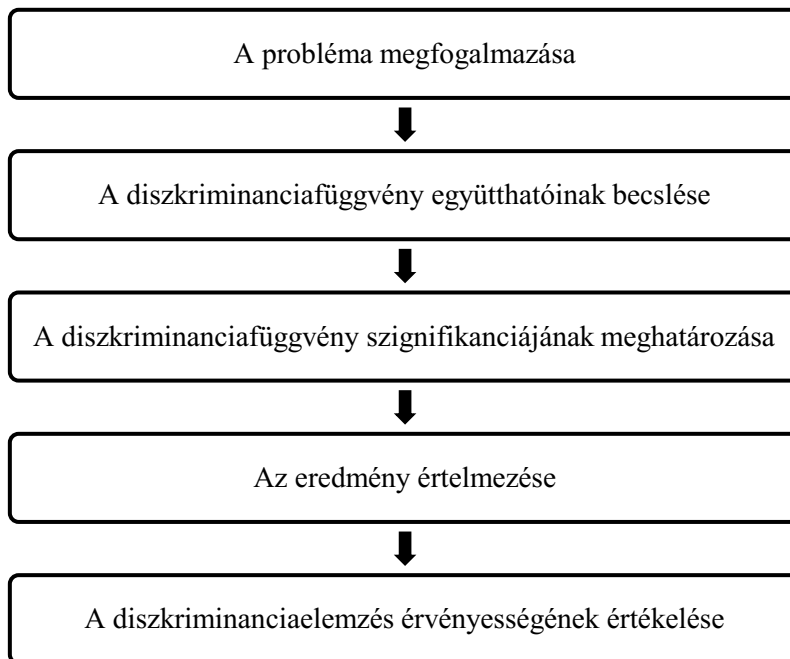
A diszkriminancia analízisnek azonban vannak hátrányai. Egyik ilyen hátránya az, hogy nem alkalmas olyan esetek elemzésére, melyek során nemlineáris kapcsolatot feltételezünk.

Ezen kívül vannak egyéb feltételek, amelyek teljesülése szükséges az elemzés elvégzéséhez. Ezeket Sajtos-Mitev (Sajtos L. – Mitev A., 2007) összegyűjtötte, mely feltételek az alábbiak:

1. A változók mérési szintje: A függő változó legyen kategorikus változó. Amennyiben nem az, akkor szükséges a függő változó kategorizálása. Az ilyen esetekben azonban mivel a kategóriák szubjektívan kerülnek kialakításra, torzításhoz vezethet. Ez az eset felmerülhet a diszkriminancia nemfizetés-előrejelzési alkalmazása során is, hiszen nincs általános definíciója annak, hogy ki nevezhető “jó”, illetve “rossz adósnak”. Az lesz-e a “rossz” adós, aki 3-6-12 hónapig nem fizet, vagy a nem fizetett összeg esetleg elér egy bizonyos összeghatárt, így ennek a pontos definiálása mindenképpen szükséges az elemzés elvégzése során. A független változóknak metrikus változóknak kell lenniük.
2. Az adatok függetlensége: A megfigyeléseknek egymástól függetlennek kell lenniük.
3. A csoportok kizárólagossága: Minden egyes megfigyelési egység kizárólag egyetlen csoportba tartozhat.
4. Csoportnagyság: Az egyes csoportoknak, azaz nemfizetés esetén a “rossz” adós és “jó” adós csoportnak megközelítőleg egyforma nagyságúnak kellene lennie.

5. Mintanagyság: Fontos a megfelelő mintaelemszám. Nagyobb minta esetén javulhat a megbízhatóság. Nincs általánosan elfogadott szabály az elemszám és a változók arányára vonatkozóan. Poulsen-French szerzőpáros szerint a független változók száma ne haladja meg az elemszám 2-vel csökkentett értékét, míg a Sajtos-Mitev páros ajánlása szerint az elemszámnak tízszer nagyobbnak kell lennie, mint az elemzésbe bevont független változók száma. (Poulsen-French, 2003; Sajtos L. – Mitev A., 2007)
6. Linearitás: az elemzés elvégzéséhez szükséges azt feltételezni, hogy a független változók között lineáris kapcsolat van.
7. Egyváltozós és többváltozós normalitás: a leggyakrabban alkalmazott eloszlás a normális, hiszen ilyen esetekben a paraméterek becslése egyszerűbb. Tesztelése történhet grafikus elemzéssel (Boxplot), illetve kvantitatív elemzések segítségével is egyváltozós normalitás esetében. Kvantitatív tesztek közül a Kolmogorov-Smirnov és a Shapiro-Wilk próba alkalmas erre. Többváltozós normalitás esetében pedig a Mahalanobis-távolságot érdemes választani. Sérülésének okozója lehet például, ha kiugró elemek szerepelnek a mintában, azonban mivel a diszkriminancia elemzés kimondottan érzékeny a kiugró elemekre, ilyen esetben célszerű lehet másfajta elemzési módszert választani.
8. Homoszkedaszticitás: A függő változó csoportjaiban a független változó varianciájának hasonlóan kell lennie. A feltétel ellenőrzésében a Box's M mutató segít. Ennek hiánya is visszavezethető a kiugró értékekre.
9. Multikollinearitás: Diszkriminanciaelemzés esetében is problémát jelenthet, ha a független változók egymással is korrelálnak, ilyenkor érdemes lehet ezeket a változókat összevonni, vagy egyeseket kivenni az elemzésből.
10. Outlierek (Sajtos-Mitev, 2007)

A diszkriminanciaelemzés menetét a 16. ábra foglalja össze.



16. ábra: A diszkriminanciaelemzés menete
Forrás: Malhotra: Marketingkutató (2008), p587

Az első lépés magában foglalja a változók meghatározását, és a validáláshoz szükséges tanuló és tesztelő minta létrehozását. A változók meghatározásánál fontos, hogy figyelembe vegyük a korábban megfogalmazott feltételeket.

A második lépés az együtthatók becslése, mely történhet közvetlen módszerrel vagy lépcsőzetes diszkriminancia elemzéssel. A közvetlen módszer “a diszkriminanciafüggvény egy megközelítése, amely úgy becsüli a diszkriminációs funkciót, hogy a független változókat egyszerre veszi figyelembe.” (Malhotra, 2008, p589.) Ebben az esetben a változók diszkrimináló képessége nem kerül figyelembevételre, mely akkor lehet célszerű például, ha a modell mögött egy már meglévő elméleti modell áll. A lépcsőzetes diszkriminanciaelemzés “olyan diszkriminanciaelemzés, ahol a független változókat egymás után veszik az elemzésbe annak alapján, hogy mennyire képesek elkülöníteni az egyes csoportokat.” (Malhotra, 2008, p589.)

A diszkriminanciaelemzés során létrejön egy diszkriminanciafüggvény, mely “a független változók olyan lineáris kombinációja, amely a legjobban szétválasztja a függő változó kategóriáit.” (Malhotra, 2008, p584.)

A diszkriminanciafüggvény a következőképpen írható fel:

$$Z = a + W_1X_1 + W_2X_2 + \dots + W_JX_J$$

ahol

a: a konstans tag

W: a diszkriminancia-együttható

X: a független változó

Z: a diszkriminanciafüggvény értéke.

A harmadik lépés a függvény szignifikanciájának meghatározása, melynek az alapja az SPSS-ben a Wilk's λ mutató, amit átranzformálunk χ^2 értékévé. Negyedik lépésként értelmezzük a kapott eredményeket, végezetül pedig megvizsgáljuk elemzésünk érvényességét, melyhez a korábban kialakított tanuló és tesztelő mintát alkalmazzuk. A besorolások pontosságáról egy klasszifikációs mátrix segítségével kaphatunk információt. (Sajtos-Mitev, 2007; Malhotra, 2008)

Annak meghatározásához, hogy az egyes esetek melyik csoportba kerülnek, szükséges egy vágási érték meghatározása, mely az alábbi formula segítségével tehető meg, abban az esetben, ha a csoportok nagysága egyenlő:

$$Z_c = \frac{Z_A + Z_B}{2}$$

ahol,

Z_c : az optimális vágási érték

Z_A : centroid az A csoport esetében

Z_B : centroid a B csoport esetében.

Nem egyenlő csoportok esetében szükséges a centroid értékek súlyozása.⁶ (Dhamnetiya et al., 2022)

Mint ahogyan a fenti felsorolás is mutatja, a diszkriminanciaelemzésnek számos fontos feltétele van, melyek együttes teljesülése nem egyszerű, azonban ennek a módszernek van egy napjainkban nagyon elterjedt alternatívája, a logisztikus regresszió, mely esetében sokkal kevesebb feltétel teljesülése szükséges, tehát egy robusztusabb próba, például nincs szükség a normál eloszlás meglétére, illetve a csoportok közötti egyenlő varianciára sem.

A rendelkezésre álló adatbázist figyelembe véve a fentebb felsorolt feltételek egyidejű teljesítése nem megoldható, így arra a döntésre jutottam, hogy a diszkriminancia analízis használatától annak komplex feltételrendszere miatt célszerű eltekinteni. Így annak bővebb szakirodalmi, elméleti áttekintése nem képezi a dolgozat részét.

4.2 Logisztikus regresszió

A cél ugyanaz, mint ami a diszkriminanciaelemzés esetében is volt, tehát a függő változó előre meghatározott csoportjaiba besorolni a megfigyelési egységeket. Logisztikus regresszió esetében megkülönböztethetünk dichotom logisztikus regressziót és politochom logisztikus regressziót. Mivel a nemfizetés esetében a függő változó továbbra is két kimenettel rendelkező binomiális változó, így a hangsúly a binomiális logisztikus regresszióra helyeződik.

A vizsgálat menete hasonló a diszkriminanciaelemzés esetében tárgyalt menethez. Elsőként fontos a probléma megfogalmazása, melyet az elemzés elvégzéséhez szükséges feltételek vizsgálata követ.

A diszkriminanciaelemzéssel ellentétben a logisztikus regresszió esetében a feltételek sora sokkal rövidebb:

1. Ami mindenképpen fontos, hogy a változók mérési szintje megfelelő legyen, mely a függő változó esetében nominális változót takar, a független változó esetében nincs megkötés, akár nominális változók is használhatók.
2. A megfigyelések függetlensége: A megfigyeléseknek egymástól függetlennek kell lenniük.
3. Megfelelő mintanagyság: legalább 60 elemű minta. A mintanagysághoz kapcsolódó feltétel továbbá, hogy minden kimenet esetében szükséges egy elégséges megfigyelési mennyiség.
4. Multikollinearitás: a magyarázó változók kizárólag az eredményváltozóval függenek össze.

Amennyiben a szükséges feltételek teljesülnek, elvégezhető az elemzés, melyet az eredmények értelmezése, majd pedig az érvényesség vizsgálata követ.

A modell alapja az "odds", melynek mértékét kalkulálja a modell. Ennek értéke tulajdonképpen a jó adós és a rossz adós szerep bekövetkezési esélyeinek hányadosát takarja. Amennyiben ezt képlettel szeretnénk kifejezni, azt a következőképpen tehetjük meg:

⁶ A centroid az eredményváltozó adott csoportjára vonatkozó diszkriminanciaértékek átlaga.

$$odds_x = \frac{P_x}{1 - P_x}, \text{ azaz } odds_x = \frac{P_{j\acute{o} \text{ ad\acute{o}s}|x}}{1 - P_{j\acute{o} \text{ ad\acute{o}s}|x}}$$

A logisztikus regresszió során azzal a feltételezéssel élünk, hogy az odds logaritmusa az meghatározható a magyarázó változók lineáris függvényeként, mely a következőképpen írható fel:

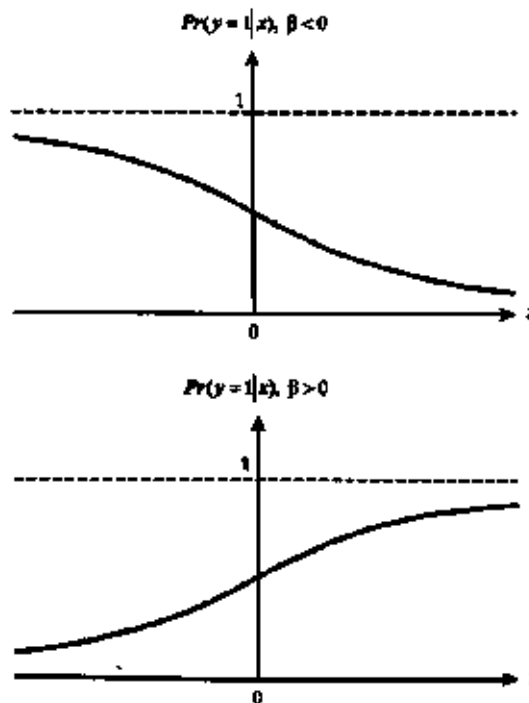
$$\ln(odds_x) = \text{logit}(P_x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

amit, ha átalakítunk, akkor a következő formát kapjuk:

$$odds_x = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p} = e^{\beta^T x}$$

További átalakítások után jutunk el oda, hogy a jó adósság valószínűsége a következő:

$$P_x = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}}$$



17. ábra: Logisztikus valószínűség a magyarázó változó függvényében

Forrás: Hajdú (2003) p292.

A klasszifikáláshoz szükség van továbbá egy úgynevezett cut point value-ra⁷, melynek szerepe, hogy amennyiben a számított P_x meghaladja ezt az értékét, akkor az adott illető rossz adóssá válik a modellben, amennyiben pedig nem, akkor jó adósként kerül besorolásra. Az érték természetesen változtatható, azonban célszerű úgy meghatározni, hogy a rossz besorolás során keletkező veszteség a lehető legkisebb legyen. Optimális értékének meghatározásában segíthet az ROC görbe, melyről a 4.6.2 fejezetben írok.

A paraméterek becslése maximum likelihood módszer szerint történik, tehát azon becsült értékekre van szükség, melyek esetében a minta likelihoodja a maximális értéket veszi fel. Ez a következőképpen néz ki: (Hajdú, 2003)

$$L = \prod_{i=1}^n \Pr(y = y_i | \mathbf{b}) = \prod_{i=1}^n P_{ib}^{y_i} (1 - P_{ib})^{1-y_i} = \prod_{i=1}^n \frac{e^{y_i b x_i}}{1 + e^{b x_i}} \rightarrow \max$$

Logisztikus regresszió esetében a β paraméter tesztelése úgynevezett Wald-statisztika segítségével történik a következőképpen:

$$Wald_i = \left(\frac{a_i}{SE_{a_i}} \right)^2$$

ahol

a_i : a logisztikus regresszió együtthatója

SE_{a_i} : pedig a logisztikus együttható standard hibája.

Az elemzés során a következő hipotézis vizsgálata történik:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Amennyiben az adott együttható szignifikánsnak tekinthető, akkor elmondható, hogy hozzájárul az elemzéshez. (Hajdú, 2003; Malhotra, 2008)

A paraméterek tesztelése mellett fontos figyelmet fordítani a modell illeszkedésére egyaránt, mely ebben az esetben a Hosmer and Lemeshow goodness-of-fit statisztika segítségével tesztelhető, mely során a nullhipotézisben azzal a feltételezéssel élünk, hogy a modell megfelelően illeszkedik. A hipotézis ellenőrzése χ^2 statisztika segítségével történik. Ehhez szükséges az egyes esetek bekövetkezési valószínűségeit növekvő sorrendre rendezni, majd csoportokra bontani, és az egyes csoportokban megvizsgálni, hogy a különböző kategóriákból hány tényleges és hány becsült eset tartozik az adott csoportba.

⁷ vágási pont

A modell illeszkedésén kívül tudjuk ellenőrizni, hogy a teljes modell szignifikáns-e, melyre az Omnibus tesztet alkalmazzuk. A teszt χ^2 próba segítségével vizsgálja, hogy van-e különbség az alap modell és az új modell log-likelihood-ja között.

Végezetül pedig az elemzés elvégzése során fontos szempont a modell magyarázóereje. Lineáris regresszió esetében a (többszörös) determinációs együttható segítségével ragadható ez meg, azonban a lineáris regresszió során a paraméterek becslése OLS segítségével történik, míg a logisztikus regresszió esetében az eredményváltozó nominális skálán mért, jelen esetben két kimenettel rendelkező változó, és a paraméterek becslésére a maximum likelihood módszert alkalmazzuk. Ennek következtében klasszikus értelemben vett R^2 értékről nem beszélhetünk, ilyen esetben ún. pszeudó R^2 -t alkalmazunk és ezek közé tartozik a McFadden R^2 , a Cox and Snell R^2 és a Nagelkerke R^2 is. Elemzéseim során ez utóbbi kettőt fogom alkalmazni. A Cox and Snell R^2 a létrehozott modell log likelihoodját az alapmodell log likelihood értékéhez viszonyítja, azonban hátránya, hogy a mutatószám értéke tökéletes modell esetén is kisebb, mint 1.

$$R_{Cox-Snell}^2 = 1 - \left(\frac{-2LL_{null}}{-2LL_{aktuális}} \right)^{\frac{2}{n}} ; R_{Nagelkerke}^2 = \frac{R_{Cox-Snell}^2}{\max R_{Cox-Snell}^2} ;$$

$$\max R_{Cox-Snell}^2 = 1 - (-2LL_{null})^{2/n}$$

Így azért, hogy a mutatószám képes legyen elérni a maximális 1 értéket, ki kell korrigálnunk, melyet az elérhető maximális értékkel tehetünk meg, így megkapva a Nagelkerke R^2 értékét. Ezek alapján a Nagelkerke R^2 használata célszerűbb. Nullához közeli érték esetében gyenge, míg egyhez közeli érték esetében erős modelltől beszélhetünk. Ezen mutatószámok értelmezése eltér a megszokott R^2 értelmezéstől, azonban ebben az esetben is a modell magyarázóereje ragadható meg vele. Egy lehetséges értelmezése, hogy az eredményváltozót hány százalékban magyarázzák a bevont magyarázó változók, míg a fennmaradó részt a modellen kívüli változók magyarázzák. (Székelyi-Barna, 2002; Malhotra, 2008; Newsom, 2008; Oravecz, 2008; Afifi et al., 2012; Hámosi, 2014; IBM)

Amennyiben a létrejött modell megfelelőnek minősül, az utolsó előtti lépés a kapott eredmények értelmezése, mely hasonlít a többváltozós regressziós modell együtthatóinak értelmezéséhez. Ezen értékeket, amennyiben az elemzés SPSS programcsomag segítségével történik, a "Variables in the Equation" nevű táblázat EXP(B) oszlopa tartalmazza, mely megmutatja, hogy az adott változó mennyivel járul hozzá a becsléshez.

Logisztikus regresszió esetében is egy klasszifikációs mátrix tartalmazza a besorolási eredményeket, pontosságot. Az utolsó lépés pedig az eredmények érvényességének vizsgálata, amely a logisztikus regresszió során is történhet tanuló és tesztelő minta alkalmazásával.

Mint minden elemzési eljárásnak, a logisztikus regresszióknak is vannak hátrányai, mely közé tartozik, hogy kimondottan érzékeny a multikollinearitásra, továbbá a kiugró értékekre, melyeket ebből kifolyólag kiemelten fontos kezelni, illetve, hogy előrejelző képessége nagy mintaelemszám esetében tekinthető kiemelkedően jónak. Előnye azonban a diszkriminanciaelemzéssel szemben, hogy kevesebb feltétel teljesülése szükséges.

4.3 Rekurzív partícionáló algoritmus (RPA)

A módszertan többféle néven is létezik, sokan egyszerűen döntési faként hivatkoznak rá. Ez az elemzési lehetőség is a klasszifikációs módszerek közé tartozik, melynek alapja egy folyamatábraszerű struktúra. (Hajdú, 2008) Kezdetben kizárólag kategorikus függő és független változó esetén alkalmazták, azonban idővel a módszert kiterjesztették metrikus változókra is. A cél a csoporton belüli variancia minimalizálása, úgy, hogy a csoportok közötti variancia a lehető legnagyobb legyen.

Csödelőrejelzésekben való alkalmazása a 80-as évekre tehető. A módszer ötvözi az egy és többváltozós elemzéseket, hiszen 1-1 felosztás egy változó szerint történik, azonban összességében több változót von be az elemzésbe. Minden egyes lépésnél igyekeznek az algoritmus csökkenti a téves besorolásokat. Az algoritmus segítségével egy iteratív folyamat megy végbe, melyet kimondottan számítógépre terveztek.

A döntési fáknak több típusa létezik, melyek közül az alábbiak rendelkeznek letisztult módszertannal:

- CHAID: Chi-squared Automatic Trees
- Exhaustive CHAID
- CRT: Classification and Regression Trees
- QUEST: Quick, Unbiased, Efficient Statistical Trees

Az elemzés nagy előnye, hogy a bevont változók esetében nincs megkötés, metrikus és nem metrikus változók is bevonhatók.

A folyamat három fő lépésből áll, mely az egyesítés, felosztás és megállás.

Egyesítés (merging)

„Minden egyes magyarázó változó esetében, a függő változóra vonatkozóan, statisztikailag független, pontosabban a statisztikailag legkevésbé összefüggő kategóriák egyesítését” jelenti. (Hámori, 2001, p704)

Adott magyarázó változók esetében minden lehetséges párosítást megvizsgál, és Pearson-féle χ^2 teszt segítségével megvizsgálja, hogy a különböző párosítások esetében milyen valószínűség mellett függetlenek az eredményváltozó kategóriái és a magyarázóváltozó kategóriapárjai, majd az algoritmus megkeresi a legmagasabb p értékkel rendelkező esetet, és az egyesítési küszöbhez hasonlítja. Mindezt addig teszi, ameddig a legmagasabb p érték nem lesz kisebb, mint az egyesítési küszöbérték, és akkor a ciklus leáll. A folyamat minden egyes magyarázó változó esetében végbemegy.

Felosztás (splitting)

„A megfigyelések, a függő változó tekintetében legkevésbé függetlennek tekinthető magyarázó változó kategóriái szerinti felosztását” jelenti. (Hámori, 2001, p704)

A magyarázó változók közül a legkisebb p értékkel rendelkező kerül kiválasztásra, mely értéket a felosztási küszöbértékhez kell viszonyítani. Amennyiben kisebb, akkor létrejön a felosztás és egészen addig ismétlődik, amíg tovább már nem felosztható az adott részadatbázis.

Megállás (stopping)

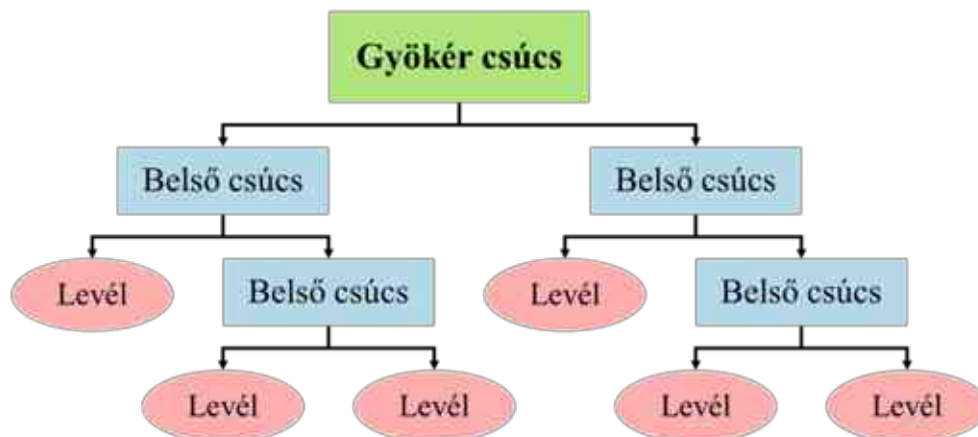
„Az algoritmus addig folytatja rekurzív módon a kategóriák egyesítését és az esetek felosztását, míg el nem ér valamely előre definiált megállítási kritériumot.” (Hámori, 2001, p704)

A felosztás-megosztás ciklus egészen addig ismétlődik, amíg valamilyen megállási kritérium be nem következik:

- p meghaladja a felosztási küszöböt,
- nincs különbség az esetek között magyarázó vagy eredményváltozó esetében,
- részadatbázis elemszáma kisebb, mint az előre megadott esetszám,
- eléri a maximális mélységét a fa.

A folyamat minden esetben egy gyökércsúcsból indul ki. A létrejött alcsoportokat csomópontnak (node) vagy belső csúcsnak nevezik és levelekben végződnek. Az előrejelzés alapját a levelek adják, melyek a fának azon része, amit nem osztunk tovább. Ezen elemek rajzolják végül ki a fát. A döntési fák a leggyakrabban fentről lefelé terjednek, azonban előfordulhat balról jobbra való irányultság is. (Hajdú, 2008)

A fentebb leírt folyamat eredményeként egy fa rajzolódik ki, melyre egy példát az alábbi ábrán láthatunk:



18. ábra: Tipikus döntési fa

Forrás: Saját szerkesztés

A döntési fákat, annak felosztását megközelíthetjük egy másik szempontból is, ami pedig az adathalmazban lévő bizonytalanság vagy rendezetlenség csökkentése. Egy adathalmaz rendezetlenségét az entrópia szó jelöli és a gyökércsomópont, valamint a belső csúcsok meghatározásában nyújt segítséget. Abban az esetben, ha a döntési fa két kimenettel rendelkezik, mint az én esetemben is, az entrópia a következőképpen határozható meg:

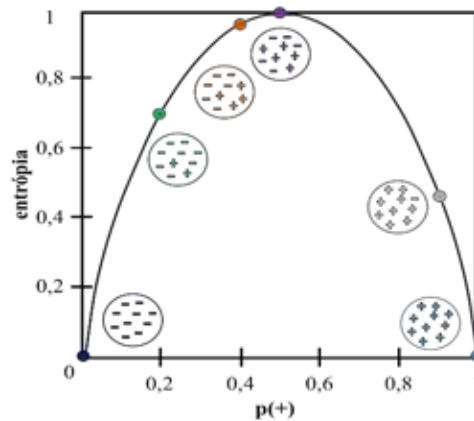
$$E(S) = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)}$$

ahol

p_+ : a pozitív csoport aránya

p_- : a negatív csoport aránya

A döntéseket könnyebb meghozni alacsonyabb entrópia esetében. Az entrópia értéke 0 és 1 között mozoghat, homogén csoport esetében veszi fel a minimális értéket és a maximális értéket pedig akkor, ha a csoportok megoszlása egyenletes.



19. ábra: Az egyes megoszlásokhoz tartozó entrópia értékek

Forrás: Saját szerkesztés Fawcett – Provost (2013) alapján

A folyamat során minden esetben egy adott csomóponthoz tartozó entrópia kerül kiszámításra, és nem az entrópia mértékében végbemenő változás. Ez utóbbira szolgál az információnyereség fogalma, mely megmutatja, hogy a kiindulási entrópia mennyivel csökkent a felosztást követően, azaz a bizonytalanság csökkenésének mértékét kalkulálja ki, és segít a megfelelő szétválasztások meghatározásában. A legnagyobb információnyereséggel rendelkező attribútum fogja a legjobb felosztást eredményezni, tehát minél nagyobb az információnyereség, annál értékesebb az adott változó a célváltozó előrejelzésében, azaz az adott változó annál közelebb helyezkedik el a gyökércsúcshoz. Ezt követően a folyamat rekurzív módon ismétlődik.

A módszer mellett szóló érv, hogy nem szerepel a feltételek között a változók normális eloszlása, kezeli a nemlinearitást, könnyen értelmezhető, nem igényel nagy számítási igényt és rugalmas. Alkalmazása akkor a legegyszerűbb, ha bináris elválasztások vannak. Más algoritmusokhoz képest kevesebb adattisztításra van szükség. A döntési fák esetében metrikus és nem metrikus változók egyaránt felhasználhatók. Végeredményként magas arányban megkapjuk a megfelelő fizetőképességi besorolást, a pontos besorolási adatok a klasszifikációs mátrixban találhatóak meg ebben az esetben is.

A módszertan hátránya, hogy sok csoport, folytonos eredményváltozó, vagy kis számú tanuló minta esetében hajlamos a hibákra. Előrejelzési célra nem alkalmazható, mivel többségében a tanuló adatbázisra specializálódik, fennállhat a tútanulás veszélye, főleg mélyebb fák esetében. Azonban a probléma orvosolható a mesterséges intelligenciamodellek tútanulás ellen kifejlesztett módszerével, azaz, ha az adatokat egy tanuló és egy tesztelő részre osztjuk és megvizsgáljuk, hogy mindkét esetben hasonló eredmények születtek-e. (Hámori, 2001; Saini, 2021; Decision Tree, 2023; IBM)

4.4 Neurális hálók

Napjaink legjobb teljesítményű számítógépének még mindig az emberi agy tekinthető. A feladatok elvégzését információ-feldolgozó egységek, neuronok segítik. „A neurális hálózatok, pontosabban a mesterséges neurális hálózatok olyan információ-feldolgozó paradigmák, amelyek ihletését az emlősök agyának igen nagy sűrűséggel összekapcsolt párhuzamos feldolgozó struktúrái és folyamatai adták. A neurális hálózatok lényegében olyan matematikai modellek, amelyek a biológiai idegrendszerek egyes információfeldolgozási elvei alapján működnek és ennek alapján adaptív tanulásra képesek.” (Ketskemény et al., 2011, p394.) A módszert számos területen alkalmazták már sikeresen, mint például:

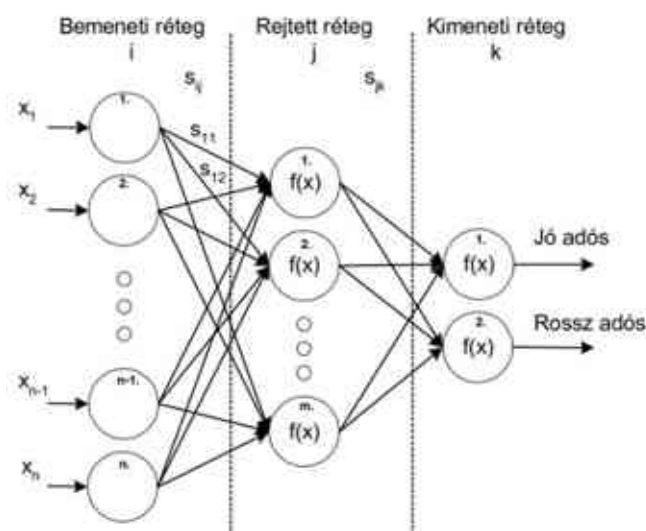
- tőzsdei indexek alakulásának előrejelzése,
- meteorológiai előrejelzések,
- robotok vezérlése,
- fogyasztói viselkedés jóslása,
- kockázatmenedzsment,
- orvosi diagnózis stb. (Ketskemény et al., 2011)

„A neurális hálók alapeleme az elemi neuron. Az elemi neuron egy több-bemenetű, egy-kimenetű eszköz, ahol a kimenet a bemenetek lineáris kombinációjaként előálló közbenső érték nemlineáris függvénye.” (Kristóf, 2002, 16. old)

A neurális hálók három fő rétegből állnak:

- bemeneti: ismert információkat, hálóba táplált változókat tartalmazó neuronok,
- köztes (rejtett): szintén neuronokból épül fel, súlyai folyamatosan változnak,
- kimeneti.: eredmény-neuronokat tartalmaz, súlyok kötik össze a rejtett réteggel. (Kristóf, 2002)

Az egy rejtett réteget tartalmazó neurális háló lehetséges felépítését szemlélteti a következő ábra.



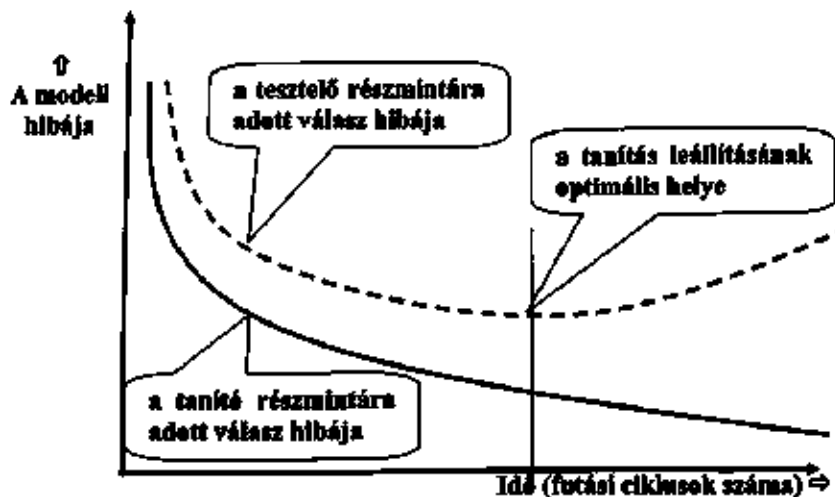
20. ábra: Neurális hálózat egy rejtett réteggel

Forrás: Kiss (2003) 62. old.

A modell általános működése könnyen érthető, amennyiben a bemenő jel meghaladja az ingerküszöböt, akkor az információ továbbításra kerül, melyek hatására a neuron folyamatosan tanul és változik. Az érkező információk a szinapszisok hatására erősödhetnek vagy gyengülhetnek, azonban ezek értékei is változnak a tanulás folyamán. (Kiss, 2003)

Az algoritmus négy elvi lépésből épül fel:

- Tanulás/gyakorlás: Ennek a fázisnak a célja, hogy a modell által meghatározott besorolás a lehető legtöbb esetben egyezzen a tényleges besorolással. A tanulás minden esetben példákon keresztül történik, és a folyamat során fontos a súlyok megfelelő meghatározása, mely egy tanuló minta alapján történik. A háttérben egy iteratív folyamat megy végbe, mely során a súlyok változhatnak.
- Tesztelés: a túltanulás elkerülése érdekében van erre szükség, melyhez egy részmintát célszerű alkalmazni. A tesztelő mintában lévő adatok függetlenek a létrejött modelltől. A tanuló és tesztelő hiba a folyamat során változik, a tanulási ciklust a tesztelő hiba minimumánál célszerű leállítani, melyet a 16. számú ábra szemléltet.



21. ábra: A tanulási és tesztelési hiba alakulása a tanítási ciklusok számának függvényében

Forrás: Ketskemény et al. (2011), p400.

- Érvényesítés/validálás: Ez történhet érvényesítő (holdout) adatok segítségével, melyek olyan adatok, amelyeket sem a tanulás, sem pedig a tesztelés során nem használtunk.
- A modell futtatása

A hálók viselkedését befolyásolják a kialakított súlyok és az input-output függvények, melyek négy típusba sorolhatók:

- lineáris,
- küszöb,
- szigmoid,
- tangens hiperbolikus.

Elemzéseim során az MLP⁸ módszert alkalmaztam, ami az egyszerű perceptront bővíti rejtett rétegekkel, melyeket a bemeneti és kimeneti réteg közé helyez, javítva a tanulási teljesítményt. A réteg között az információ áramolhat visszacsatolással és visszacsatolás nélkül. A legismertebb a back propagation hálózat, melynél a hiba visszafelé terjed, folyamatosan alakítva a súlyokat. Az optimális súly elérése az átlagos négyzetes hiba (MSE) minimalizálásával jár, azaz úgy szükséges a súlyokat meghatározni, hogy az elvárt és megfigyelt output közötti eltérés csökkenjen. Ehhez a folyamathoz szükség van a súlyok hibáinak deriváltjára (EW), mely megmutatja a hiba változásának arányát⁹ (EA), ha az egység aktivitási szintjében változás következik be. Az EA meghatározásához elsőként szükség van a rejtett réteg és a kapcsolódó output egységek közötti súlyokra, majd vesszük a súlyok és az output egységek EA értékeinek lineáris kombinációját, megismételve a lépéseket a többi réteg esetében is, visszafelé haladva. Végezetül a súlyok hibáinak deriváltját úgy kapjuk meg, hogy az EA és a bejövő kapcsolatok aktivitásának vesszük a szorzatát. A súlyok jelölése W-vel történik, értékük pozitív és negatív is lehet, pozitív érték esetén a súlynak erősítő hatása van, míg negatív érték esetében gyengítő. (Kristóf, 2002; Ketskemény et al., 2011)

Az output réteg egységének aktivitása egy kétlépcsős folyamat segítségével határozható meg, melyhez elsőként a teljes súlyozott X_j input meghatározása szükséges:

$$X_j = \sum_i y_i W_{ij}$$

ahol,

y_i : a j-edik egység aktivitási szintje az előző rétegben

W_{ij} : a kapcsolat súlya az i-edik és a j-edik egység között.

Ezt követően szükséges az y_i aktivitásának meghatározása a teljes súlyozott input függvényében:

$$y_j = \frac{1}{1 + e^{-x_j}}$$

Végezetül pedig a függvény hibájának kiszámítása következik:

$$E = \frac{1}{2} \sum_i (y_i - d_i)^2$$

ahol,

y_i : az egység aktivitási szintje az output rétegben

d_i : az i-edik egység elvárt outputja.

⁸ Multi-Layer Perceptron

⁹ az elvárt és megfigyelt outputok különbsége

A backpropagation algoritmus négy lépésből áll:

1. Meghatározza, a hiba változásának gyorsaságát a kimeneti egységben aktivitásában megkövetkezett változás hatására.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

2. A hiba változási ütemének meghatározása akkor, ha egy kimeneti egység által kapott összes bemenet változik (EI), ehhez az 1. lépés eredményét kell megszorozzuk a kimeneti egység változásának sebességével.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{\partial y_j}{\partial x_j} = EA_j y_j (1 - y_j)$$

3. A hiba változási ütemének meghatározása annak függvényében, hogy a kapcsolat súlya hogyan változik a kimenet változásával, azaz a 2. lépés eredménye megszorozva az egység aktivitási szintjével.

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = EI_j y_i$$

4. Annak meghatározása, hogy az előző rétegben az egység aktivitásában bekövetkezett változás hatására, milyen gyorsan változik a hiba. Ehhez a különálló kimeneti egységekre gyakorolt hatásokat kell összeadni, melyeket úgy kaphatunk meg, hogy a 2. lépés eredményét megszorozzuk a kimeneti egység kapcsolatainak súlyaival.

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial y_i} = \sum_j EI_j W_{ij}$$

A fenti képletek jelölései:

EA: a hiba változásának gyorsasága

E: a függvény hibája

y: egység aktivitási szintje

d: elvárt output

EI: a hiba változásának üteme, kimeneti egység által kapott össze bemenet változásának hatására

EW: súlyok hibáinak deriváltja

W: a kapcsolat súlya (Stergiou – Siganos, 1996; Kristóf, 2002)

A neurális hálók számos előnnyel rendelkeznek:

- kezeli a nemlinearitást,
- diszkrét és folytonos ismérvek is használhatók,
- nem okoz problémát a hiányzó adat
- nagy számú változó és elemszám kezelése (Kristóf, 2002)

Kristóf (2004) részletesebben is összefoglalta a neurális hálók erősségeit és gyengeségeit:

5. táblázat: A neurális hálók erősségei és gyengeségei

Erősség	Gyengeség
Képesek modellezni a nemlineáris kapcsolatokat	Nehéz értelmezni a kapcsolatok súlyait („black box” probléma)
Nem igényelnek előzetes információkat a függvényszerű kapcsolat típusáról	Nem biztos, hogy megtalálják a globális optimumot
Felépítésük nagyon rugalmas	Általában nagy mintát igényelnek
Nincs szükség az eloszlásról feltételezést tenni	Tanításuk meglehetősen időigényes
Tanulásra képes, intelligens rendszerek	Alkalmazásuk ellentmond számos közgazdasági elméletnek
	A korai leállító eljárás a kutató véletlenszerű döntését igényli

Forrás: Kristóf (2004), p28.

Ahogy a táblázatban is szerepel a kapcsolatok súlyainak értelmezése nehéz, így a kutatás során azt vizsgáltam, hogy az egyes változók mennyire minősülnek fontosnak az egyes modellekben. Ennek vizsgálatára az SPSS programcsomagban egy táblázat és egy ábra segítségével van lehetőség. A táblázat a fontosságot és a normalizált fontosságot tartalmazza, míg az ábra csak ez utóbbit. Fontos, hogy ezek a módszerek a kapcsolat irányáról nem adnak információt, tulajdonképpen egy érzékenységvizsgálat történik. (Ketskemény et al., 2011; IMB SPSS Neural Networks 26)

„Az egyes független változók normalizált fontossága (Normalized Importance) annak a mértéke, hogy a függő változó modell által számolt értékeit milyen mértékben határozzák meg az egyes független változók változásai. A normalizálás azt jelenti, hogy az egyes fontosság (Importance) értékeket az algoritmus elosztotta a legnagyobb fontossággal és a többi így a legnagyobb fontosság százalékában adja meg.” (Ketskemény et al., 2011, 416. old)

4.5 Módszerek alkalmazhatósága, összegzése

Mindegyik, fentebb bemutatott statisztikai elemzési lehetőség rendelkezik előnyökkel és hátrányokkal, melyek támaszt nyújthatnak a különböző módszerek alkalmazása során. Ezeket

az előnyöket és hátrányokat az alábbi táblázatban foglaltam össze, ezzel segítve, hogy mely módszerek alkalmazása lehetséges az elemzések során.

6. táblázat: Klasszifikációs eljárások előnyei és hátrányai

Módszer	Előny	Hátrány
Diszkriminancia analízis	<ul style="list-style-type: none"> Többváltozós modell Egyszerű felépítés 	<ul style="list-style-type: none"> Független változók esetében legalább intervallumskála Nemlineáris kapcsolat esetében nem alkalmas Komplex feltételrendszer (egyváltozós és többváltozós normalitás, homoszkedaszticitás, multikollinearitás)
Logisztikus regresszió	<ul style="list-style-type: none"> Többváltozós modell Független változó lehet nominális skálán mért Kevesebb feltétel, mint a DA esetében Változók súlya könnyen értelmezhető 	<ul style="list-style-type: none"> Multikollinearitás Érzékeny az outlierekre Megfelelő mintanagyság
Döntési fa	<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> Túltanulás veszélye Előrejelzésre tesztelő és tanuló minta egyidejű alkalmazása esetében használható
Neurális háló	<ul style="list-style-type: none"> 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 függvényeszerű kapcsolat típusáról Rugalmas felépítés Nincs szükség az eloszlásról feltételezést tenni Robosztus, nincs korlátozó előfeltétel Nominális ismérv alkalmazható 	<ul style="list-style-type: none"> Nehéz értelmezni a kapcsolatok súlyait Általában nagy mintát igényelnek Tanításuk időigényes Túltanulás veszélye

Forrás: Saját szerkesztés a 4.1-4.4 fejezet irodalmi alapján

4.6 Módszerek értékelése

A klasszifikációs modellek értékelésére több lehetőség kínálkozik:

- Klasszifikációs mátrix
- ROC-görbe
- Gini-koefficiens
- Kolmogorov-Smirnov teszt

A Kolmogorov-Smirnov teszt nem tartozik az elterjedt módszerek közé, melynek oka a módszer közelítő jellegében rejlik, így a kutatások során ezt az értékelési lehetőséget nem alkalmaztam.

4.6.1 Klasszifikációs mátrix

A felsorolt négy módszer közül a legismertebb a klasszifikációs mátrix. Ez az a módszer, mely az alapját adja a besorolási pontosságnak, mellyel az elmúlt évtizedekben készült csődmódelleket is előszeretettel jellemezték a kutatók. A klasszifikációs mátrix lényege, hogy a tényleges és becült csoporttagságokat hasonlítja össze, ezáltal meghatározva, hogy összességében az esetek milyen részét sikerül helyesen kategorizálni.

		Előrejelzett csoport		
		Jó adós	Rossz adós	
Tényleges csoport	Jó adós	VN	ÁP	specificitás
	Rossz adós	ÁN	VP	szenzitivitás
				pontosság

22. ábra: Klasszifikációs tábla hitelnemfizetésre értelmezve

Forrás: *Kvantitatív statisztikai módszerek – Logisztikus regresszió*

A mátrix alapján több, az előrejelzési képesség értelmezését segítő mutatószám kialakítható:

- Szenzitivitás: A ténylegesen rossz adósoknak hány százalékát tudja azonosítani.

$$\frac{VP}{ÁN + VP}$$

- Specificitás: A ténylegesen jó adósoknak hány százalékát tudja azonosítani.

$$\frac{VN}{VN + AP}$$

- Pozitív prediktív érték: A rossz adósok előrejelzett ügyfelek közül hány százalék volt valóban rossz adós.

$$\frac{VP}{VP + \acute{A}P}$$

- Negatív prediktív érték: A jó adósnak előrejelzett ügyfelek közül hány százalék volt valóban jó adós.

$$\frac{VN}{VN + \acute{A}N}$$

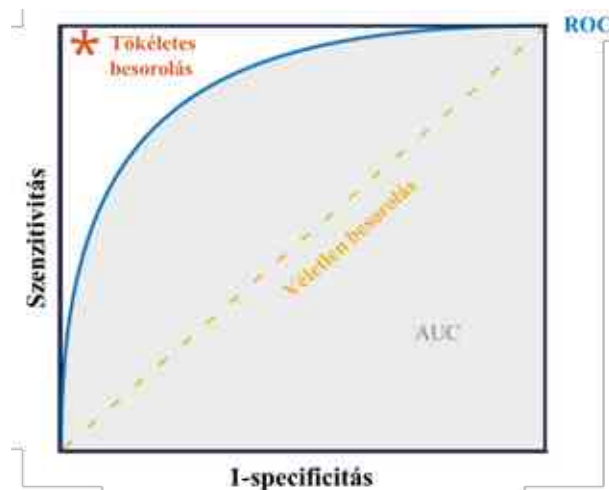
A klasszifikációs tábla az egyik legegyszerűbb módja a modellek teljesítményének mérésére.

Mindenképpen mérlegelni szükséges, hogy melyik hiba a „kedvezőbb”, az, ha egy ügyfelet nemteljesítőnek sorolnak be, és nem kap hitelt, vagy pedig, ha egy későbbi nem teljesítő ügyfelet a modell jól teljesítőnek jelez előre, és a későbbiekben mégsem tudja törleszteni a hitelt.

4.6.2 ROC (Receiver Operating Characteristic) görbe

Kezdetben, az 50-es években radarjel észleléssel kapcsolatosan alkalmazták, majd a későbbiekben kémiai, diagnosztikai elemzések hatékonyságát tesztelték a segítségével. (Fazekasné, 2002)

A görbe egy egységnyi oldalú négyzetben helyezkedik el, az egyik tengelyen a szenzitivitás a másik tengelyen pedig az 1-specificitás értékei láthatók.

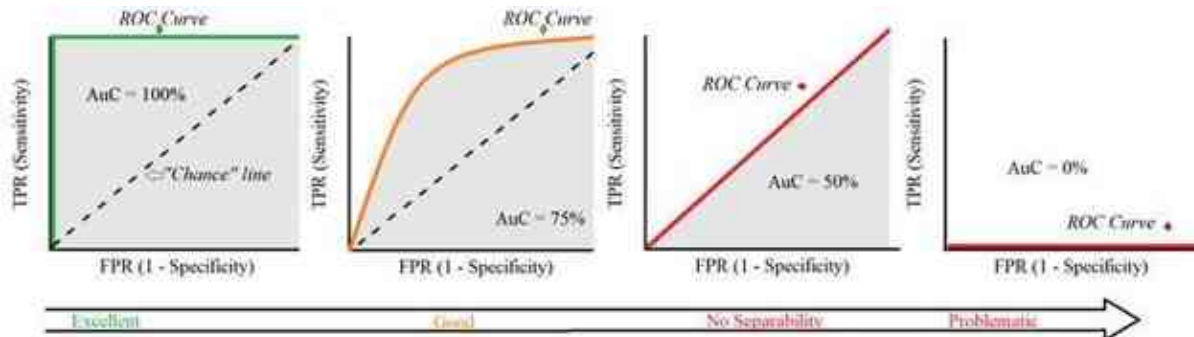


23. ábra: ROC görbe

Forrás: Saját szerkesztés <https://www.mathworks.com/help/deeplearning/ug/compare-deep-learning-models-using-ROC-curves.html> alapján

A görbe a különböző vágási pontokhoz tartozó szenzitivitás és 1-specificitás értékeit köti össze. A bal felső sarok szimbolizálja a tökéletes besorolást, tehát minél inkább közelít a görbe ehhez a ponthoz, annál jobb besorolással rendelkezik. A különböző görbék összehasonlítása az alattuk

lévő terület segítségével történhet, melyet a szakirodalom AUC néven használ (Area under the curve). Az AUC értéke 0 és 100% között mozog, ahol a 100% a tökéletes besorolást jelenti. Amennyiben az AUC értéke 80-90% körül mozog, az már kiemelkedőnek tekinthető. A következő képen erre láthatunk néhány példát.



24. ábra: Néhány lehetséges ROC-görbe

Forrás: <https://www.datasciencecentral.com/roc-curve-explained-in-one-picture/>

A klasszifikációs eljárás értékelése mellett az ROC-görbe alkalmas arra is, hogy segítségével megállapítsuk az optimális vágási értéket, melyre több ajánlás is létezik. Az elemzéseim során két ajánlást fogok figyelembe venni. Mindkét alkalmazott ajánlás a szenzitivitás és specificitás értékén alapszik.

Az első ajánlás a Youden-szabály, mely szerint a szenzitivitás+specificitás-1 értéknek a maximálisnak kell lennie. A második ajánlás pedig a bal felső ponthoz legközelebbi pont meghatározása, ami az alábbi képlet segítségével lehetséges:

$$(1 - \text{szenzitivitás})^2 + (1 - \text{specificitás})^2$$

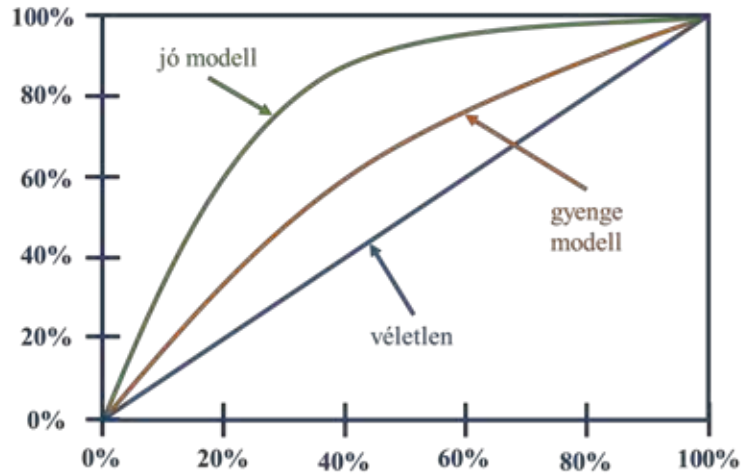
Az így kapott értéknek a minimálisnak kell lennie. (Ferenci, 2023)

Abban az esetben, ha a két ajánlás eltérő optimális vágási értéket javasol, akkor közülük a kutatási célok szempontjából kedvezőbbet fogom választani.

4.6.3 Gini-koefficiens (AR)

Mielőtt a Gini koefficiensre térnék, előtte mindenképpen fontosnak tartom, hogy ejtsek néhány gondolatot a CAP¹⁰ görbéről, melyet szintén a klasszifikációs modellek értékelésére alkalmaznak. A görbe elkészítéséhez elsőként az eseteket rangsorba szükséges rendezni a legkockázatosabbtól a legbiztonságosabbig. Amennyiben ezt a lakossági hitelekre vetítjük le, akkor az x tengelyen a hitelügyletek arányát láthatjuk, míg az y tengelyen a nemteljesítő hitelügyletek kumulált relatív gyakoriságát. Minél inkább közelít a görbe a bal felső sarokhoz, annál jobb a modell, és minél inkább közelít az átlóhoz, annál gyengébb, melyet a 25. számú ábra is szemléltet.

¹⁰ kumulált megfeleléségi profil



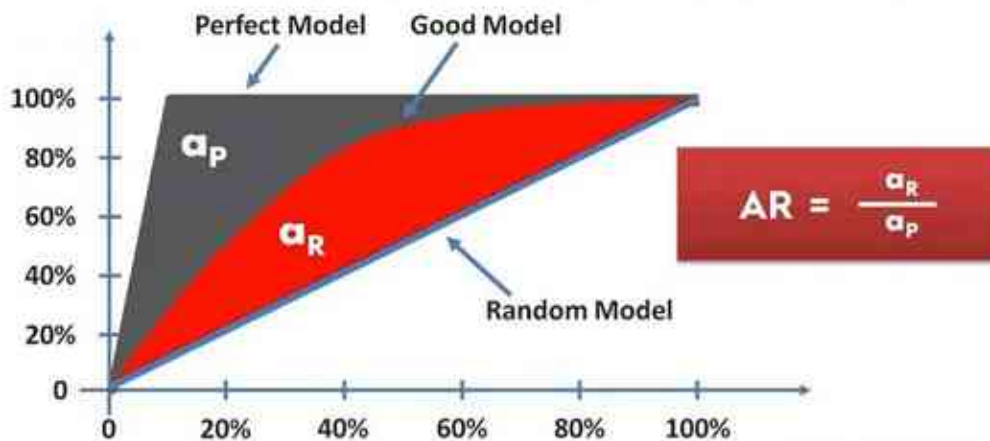
25. ábra: Néhány lehetséges CAP görbe

Forrás: Saját szerkesztés

A modell teljesítménye az AR^{11} segítségével mérhető, melyet Gini koefficiensként is használnak. A mutatószám kiszámítása a véletlen, a tökéletes és az aktuális modell segítségével történik. Az aktuális modell görbéje és a véletlen modell görbéje közötti területet viszonyítja a tökéletes modell és a véletlen modell közötti területhez. Tehát:

$$Gini = \frac{A_R}{A_P}$$

A mutatószám maximális értéke 1.



26. ábra: Gini-koefficiens kiszámítása (AR)

Forrás: Ojo Olawale: *The CAP curves* (<https://waleblaq.medium.com/the-cap-curves-the-cumulative-accuracy-profile-58a141e01fae>)

¹¹ accuracy ratio

Van más módja is a mutatószám meghatározásának, az ROC görbe esetében kiszámolt AUC értékből is meghatározható a Gini koefficiens értéke az alábbi képlet segítségével:

$$Gini = 2 (AUC - 0,5)$$

A különböző Gini-koefficiens (AR) értékek esetében az alábbi értelmezések terjedtek el:

- $AR < 60\%$: rossz modell
- $60\% < AR < 70\%$: korrekt modell
- $70\% < AR < 80\%$: jó modell
- $80\% < AR < 90\%$: legjobb előrejelző modell
- $90\% < AR < 100\%$: 97%-os értékig lehetséges, hogy van egy kiemelkedően jó modellünk, azonban érdemes ilyen esetben még egyszer felülvizsgálni az eredményeket. (Engelman et al, 2003; Olawale, 2020)

4.7 További, az elemzések során alkalmazott módszerek

4.7.1 Korrelációs számítás

A korreláció metrikus változók között fennálló sztochasztikus kapcsolatot jelent (ld. 4. számú táblázat). A lineáris korrelációs kapcsolat irányát és szorosságát a lineáris korrelációs együttható (r) segítségével mérhetjük.

A mutatószám előjele a kapcsolat irányát mutatja, értéke pedig a kapcsolat szorosságát. Minél közelebb esik az értéke a 0-hoz, annál gyengébb a kapcsolat, és amennyiben 0, akkor korrelátlanságról beszélhetünk.

4.7.2 Multikollinearitás

A multikollinearitás vizsgálata több mutatószámmal lehetséges, melyek közül a tolerancia-mutatóra és a VIF mutatóra fogok kitérni.

A variancia inflátor faktor (VIF) minden magyarázóváltozóra külön kerül meghatározásra. A mutató értékének alsó korlátja 1, ebben az esetben az adott magyarázó változó nem korrelál a többivel, felső korlátja pedig nincs. Értelmezésével kapcsolatosan általánosan elfogadott szabály nincs, azonban hüvelykujjszabály szerint amennyiben értéke 2 alatt van, akkor gyenge, 2 és 5 közötti érték esetén erős, 5 felett pedig nagyon erős, káros multikollinearitásról beszélünk. (Freund – Wilson, 1993; Hajdú, 2003; Kovács, 2008)

A tolerancia-mutató azt mutatja meg, hogy egy adott független változó varianciájából hány százalékot nem tud magyarázni a többi független változó, tulajdonképpen a VIF mutató reciprok értékét jelenti. Értéke 0 és 1 között mozog, minél közelebb van az érték egyhez, annál kevésbé fenyeget a multikollinearitás veszélye. (Székelyi-Barna, 2002)

4.7.3 Keresztáblaelemzés

A keresztáblaelemzés egyszerűségének köszönhetően egy közkedvelt, könnyen értelmezhető vizsgálat, mellyel nominális vagy ordinális változók közötti kapcsolat fennállását lehet tesztelni, erre a leggyakrabban a Pearson-féle χ^2 -t alkalmazzák. A nullhipotézisben azzal a feltételezéssel élünk, hogy nincs összefüggés a két változó között.

Az elemzés a megfigyelt gyakoriságokat a függetlenség esetére feltételezett (elvárt) gyakoriságokhoz viszonyítja, így kalkulálva a χ^2 értékét.

$$\chi^2 = \sum_{\text{minden cella}} \frac{(f_0 - f_e)^2}{f_e}$$

ahol,

f_0 : megfigyelt gyakoriság

f_e : elvárt gyakoriság,

továbbá

$$f_e = \frac{n_r n_c}{n}$$

ahol,

n_r : adott sor összege

n_c : adott oszlop összege

n : a teljes mintanagyság.

Szignifikáns kapcsolat esetében a kapcsolat erősségét több mérőszám segítségével is lehet mérni, mint például a Cramer V mutató, melynek értéke 0 és 1 között mozog, értelmezése megegyezik az előbbi fejezetben leírt lineáris korrelációs együttható abszolútértékének értelmezésével. (Sajtos-Mitev, 2007; Malhotra, 2008; Domán-Szilágyi-Varga, 2009)

5 A kutatás adatbázisa

5.1 A Központi Hitelnyilvántartó Rendszer

Magyarországon a lakossági hitelezőkkel kapcsolatos információkat a Központi Hitelnyilvántartó Rendszer, röviden KHR tartja nyilván, mely segíti a bankok közötti információáramlást a hitelezők tekintetében, segítséget nyújt a hitelbírálat során, és csökkenti a túlzott eladósodottság kockázatát. A 2011. évi CXXII. törvény szerint a KHR célja: “A Központi Hitelinformációs Rendszerben [...] nyilvántartott adatok kezelésének célja a hitelképesség megalapozottabb megítélése, valamint a felelős hitelezés feltételei teljesítésének és a hitelezési kockázat csökkentésének előmozdítása az adósok és a referenciaadat-szolgáltatók biztonságának érdekében.” (2011. évi CXXII. törvény a központi hitelnyilvántartó rendszerről) A nyilvántartó rendszer két alrendszerrel rendelkezik, az egyik a lakossági, a másik pedig a vállalkozási alrendszer. A KHR nem csak a rossz adósokról szóló információkat gyűjti, hanem úgynevezett teljes listát vezet, azaz a kötelezettségeknek időben eleget tevő ügyfelek is szerepelnek a nyilvántartási rendszerben.

Magánszemélyesek esetében a nyilvántartható adatok két csoportba sorolhatók a törvény szerint:

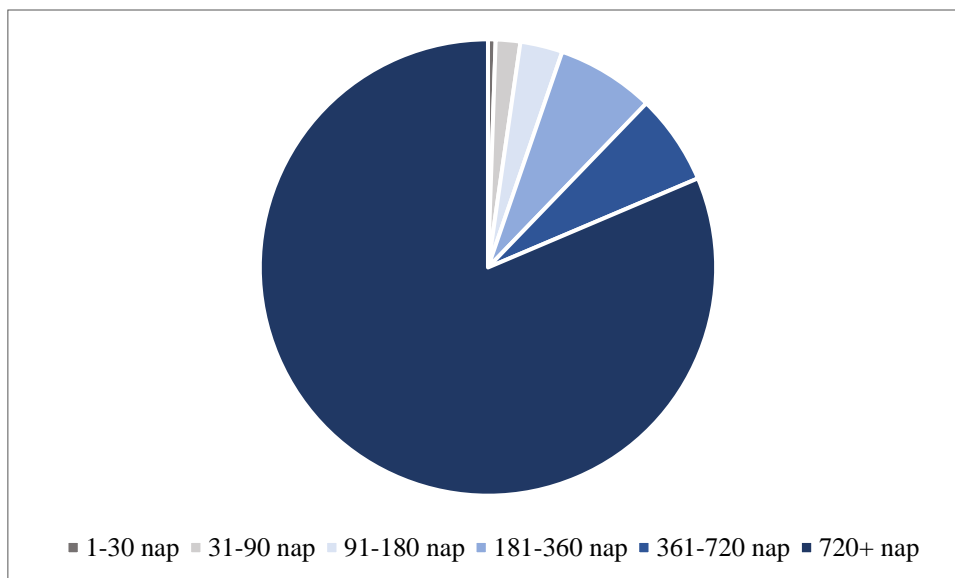
- Azonosító adatok
- Az adatszolgáltatás tárgyát képező szerződés adatai

A bankok számára a KHR-ből származó információk fontos szerepet játszanak egy hitelbírálat során, mely információkat további információkkal egészítenek ki. A KHR segítségével a bankok pontosabban fel tudják mérni ügyfeleik hitelképességét, lehetővé téve, hogy akár könnyebben jussanak hitelhez, valamint a felelős hitelezés jegyében elkerüljék a fizetésképtelen, vagy túlzottan eladósodott ügyfeleknek való hitelnyújtást.

A rendszerben az MNB aktuális tájékoztatása alapján az minősül “rossz adósnak”, aki esetében az alábbi igaz: „Ha a tartozás mértéke meghaladja a késedelembe esés időpontjában érvényes legkisebb összegű havi minimálbért, és ez, a minimálbér összeget meghaladó késedelem folyamatosan, több mint 90 napon keresztül fennállt, hitelezője megküldi a késedelem tényére vonatkozó információkat a Központi Hitelinformációs Rendszerbe.” (MNB)

A KHR minden éven elkészíti éves tájékoztatóját, melyet tanulmányozva elmondható, 2021. decemberében a rendszerben nyilvántartottak száma 5.078 millió fő volt, mely az ország felnőtt lakosságának 63,4%-át teszi ki. Ugyanezen év szeptemberében kicsit kevesebb, 5.035 millió főt tartott nyilván a rendszer. A nyilvántartott fennálló hitelszerződések számát tekintve elmondható, hogy több olyan ügyfél szerepel az rendszerben, aki egytől több hitelszerződéssel rendelkezik.

A mulasztások alakulását tekintve az állapítható meg a 2021-es évre, hogy a mulasztások száma az év során folyamatosan csökkent, januárról decemberre a fennálló tartozások száma 13,9%-kal csökkent. A fennálló tartozások alakulását törvényi módosítások is befolyásolták, mint például a törlesztési moratórium fokozatos szűkítése.



27. ábra: A fennálló mulasztások megoszlása a mulasztás időtartama szerint 2021-ben

Forrás: Saját szerkesztés a KHR adatai alapján

A mulasztások időtartamát tekintve a fennálló mulasztások 12,21%-a maximum egy éve áll fenn, 6,4% esetében a fennállás időtartama nem haladja meg a 720 napot, jelentős része, 81,39%-a, pedig meghaladja a közel 2 évet. (KHR éves tájékoztató, 2021)

5.2 Adatbázis

Az elemzések elvégzése szükséges adatbázist a BISZ Zrt. biztosította. Az adatok leválogatását 2021. szeptember 30-án végezte el a vállalat, tehát az adatbázisban az abban az időpontban nyilvántartásban szereplő személyek találhatók meg. Az adatbázis egy egysége egy hitelügyletet jelent, tehát előfordulhatnak benne olyan személyek, akik többször is szerepelnek az adatbázisban más-más hitelügylettel. Összességében ezen a napon a nyilvántartásban 10.767.452 db hitelügylet szerepelt, illetve az alábbi változók:

- az ügyfél anonim azonosítója
- az ügyféli minőség
 - adós
 - adóstárs
- az ügyfél kora: az adatbázis létrejöttkor az ügyfél kora, ez a változó az elemzés szempontjából nem releváns, helyette az ügyfél hitelfelvételkori kora került kiszámításra.
- az ügyfél neme: az adatbázis kevesebb, mint 1%-a egyéb kategóriába került. Ennek oka, hogy a változó meghatározása az MTA által megadott férfi és női keresztnévek egyezőségének vizsgálatával történik. A külföldi és adathibás nevek kerültek az „egyéb” kategóriába.
- Szerződés torzított azonosítója: KHR szerződés azonosító torzított változata
- Szerződés típusa: KHR-ben lévő szótárkódok pontos definíciói (hiteltípusok)
- Szerződés státusza

- Lezárt
- Fennálló
- A szerződés megkötésének negyedéve
- A szerződés lejáratának negyedéve
- A szerződéses összeg
- A szerződéses összeg devizaneme
- A tőketartozás összege
- A tőketartozás devizaneme
- A törlesztési összeg
- A törlesztési összeg devizaneme
- A mulasztás összege, amennyiben van
- A mulasztás státusza
 - Lezárt
 - Fennálló
- A mulasztás összegének devizaneme
- A mulasztás kezdetének negyedéve
- A mulasztás lezárásának negyedéve
- Járás: Az ügyfél lakcímének irányítószámából aggregált járási besorolás. A be nem sorolható/hibás irányítószámok az „egyéb” kategóriába kerültek.

A fenti változókon túlmenően az alábbiakkal bővítettem az adatbázist:

- Futamidő: a szerződés kezdete és vége között eltelt idő. Mivel a szerződés kezdete és vége negyedéves besorolásként ismert, így ennek a változónak a korlátja is a negyedéves pontosság. A futamidőt havi egységben mértem.
- Mulasztással rendelkező vagy nem rendelkező ügyletről van-e szó
- Megye: járás alapján a KSH információinak segítségével került meghatározásra
- Régió: a megye alapján a KSH információinak segítségével került meghatározásra
- Átlagkereset (eFt): a hitelfelvétel évének átlagkeresete az adott megyében
- Munkanélküliségi ráta: a hitelfelvétel évének munkanélküliségi rátája az adott megyében
- Foglalkoztatási ráta: a hitelfelvétel évének foglalkoztatási rátája az adott megyében
- 1 főre jutó GDP: a hitelfelvétel évének értéke az adott megyében
- Érettségizők aránya: a hitelfelvétel évének értéke az adott megyében
- Az átlagkereset szerződéses összeghez viszonyított aránya
- Az átlagkereset törlesztési összeghez viszonyított aránya
- Törlesztési összeg szerződéses összeghez viszonyított aránya
- Az ügyfél hitelfelvételkori életkora: bár az alapadatbázis is rendelkezett egy életkor változóval, ez a változó az ügyfél aktuális életkorát mutatta. Melyből visszszámítottam az ügyfél születési évét, majd ezt követően pedig a hitelfelvételkori életkorát.

A kiinduló adatbázisban az egyes hiteltípusok az 7. számú táblázatban is látható gyakorisággal fordultak elő.

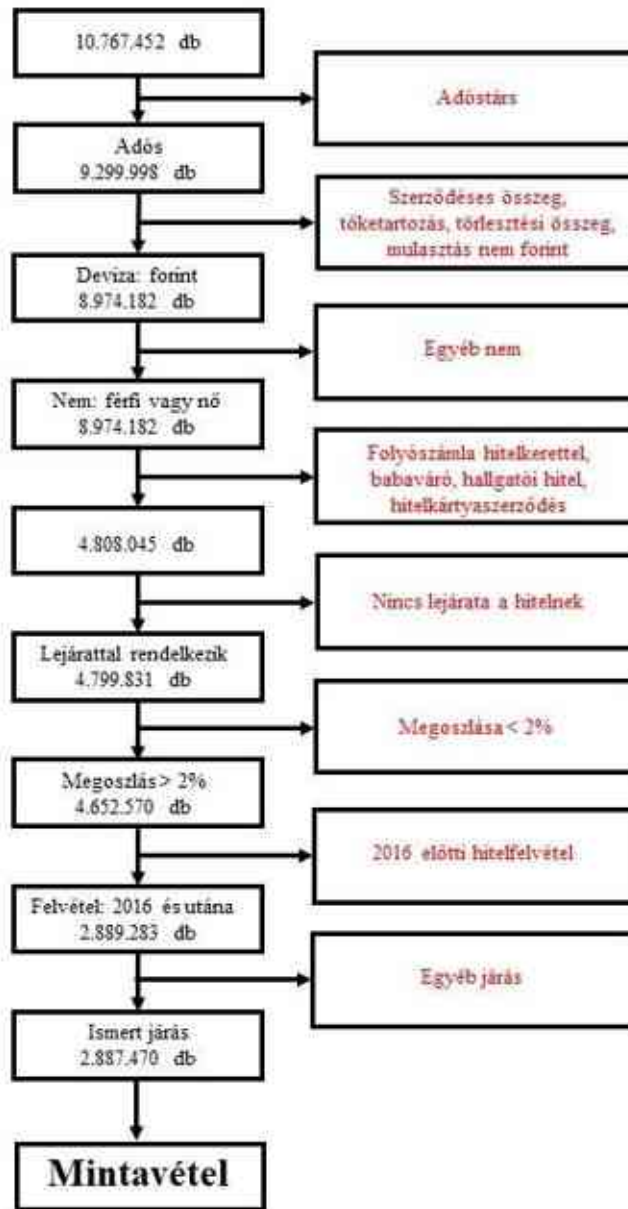
7. táblázat: Az eredeti adatbázis megoszlása hiteltípusok szerint

Szerződés típusa	Gyakoriság	Megoszlás (%)
Ismeretlen	2	,0
Babaváró hitel	310423	2,9
Bank garancia, kezesség	28	,0
Bankkártya szerződés	6	,0
Befektetési hitel	1825	,0
Egyéb hitelügylet	236176	2,2
Egyéb jelzálog fedezetű hitel	2305	,0
Egyéb készpénz hitel	121	,0
Egyéb lízing szerződés	7134	,1
Értékpapír kölcsönzés	1695	,0
Fogyasztási, áruvásárlási hitel	767012	7,1
Folyószámla hitelkerettel	2372217	22,0
Gépjármű lízing	216405	2,0
Gépjármű vásárlási hitel	117697	1,1
Gyűjtőszámla hitel	8625	,1
Hallgatói hitel	232305	2,2
Hitelkártya szerződés	1680150	15,6
Ingó jelzálog fedezetű hitel	93	,0
Kombinált egyéb hitel	4	,0
Kombinált gépjármű lízing	10742	,1
Kombinált gépjármű vásárlási hitel	14805	,1
Kombinált jelzáloghitel	168599	1,6
Kombinált lakáscélú jelzálog nélküli hitel	613	,0
Kombinált személyi hitel	58839	,5
Lakáscélú jelzálog nélküli hitel	77099	,7
Lakáscélú jelzáloghitel	1320420	12,3
Lombard hitel	3479	,0
Östermelői hitel	26546	,2
Pénzügyi faktoring ügylet	1017	,0
Szabad felhasználású jelzáloghitel	429021	4,0
Személyi hitel	2701278	25,1
Váltó leszámítolási ügylet	2	,0
Záloghitel	769	,0
Total	10767452	100,0

Forrás: Saját szerkesztés

Mivel az elemzés alapját egy ettől szűkebb bázis képezi, így a nyers adatbázis további jellemzésétől eltekintek, azonban az adatbázis szűkítése során néhány fontos jellemzőjét megismerjük.

Az elemzések megkezdése előtti első teendő az adatbázis tisztítása, a kutatási céloknak megfelelő szűkítése volt, melyet az alábbi folyamatábra foglal össze.



28. ábra: Az adatbázis szűkítésének lépései

Forrás: Saját szerkesztés

Első lépésként az ügyfélminőség szerint vizsgáltam meg az adatbázist. Ügyfélminőség szempontjából megkülönböztethetünk adós, illetve adóstárs besorolást.

A hitelügyletek 13,63%-a esetében volt az ügyfélminőség adóstárs, elemzéseim során viszont az adósokra fektetem a hangsúlyt, így első lépésként az adóstárs típusú ügyfélminőséggel rendelkező hitelügyleteket hagytam el az adatbázisból.

Második lépés a devizanemek vizsgálata, mely során kitértem a szerződéses összeg, a tőketartozás, a törlesztési összeg, valamint a mulasztási összeg devizanemére egyaránt, és az adatbázist a forint alapú hitelekre szűkítettem, azaz a kitételrel, hogy a fenti négy változónak egyidejűleg forintnak kell lennie. A hitelügyletek jelentős része megfelelt ennek a kritériumnak, 325.816 hitelügylettől váltam meg e lépés során.

A **harmadik lépés** során a nemek szerinti megoszlást vizsgáltam. Az elemzések során vizsgálni szeretném, hogy a nemnek van-e hatása a nemfizetésre, esetleg a mulasztási összeg nagyságára, emiatt az egyéb nem zavaró tényezővé válna, így ennél a lépésnél, azokat az ügyleteket, mely esetében az ügyfél neme egyéb volt, elhagytam az adatbázisból. Egyéb besorolás az ügyletek kevesebb, mint 1%-nál volt megfigyelhető.

Negyedik lépésként kivettem az alábbi hiteleket az adatbázisból:

- Hitelkártya szerződés, hallgatói hitel: mindkét esetben elmondható, hogy nincs konkrét dátum a szerződés végét illetően, tehát nem lehet meghatározni a szerződés kezdete és vége között eltelt időt, amelynek hatását szintén vizsgálni szeretném az elemzések során.
- Folyószámla hitelkerettel
- Babaváró: ez a hiteltípus 2019. július 1-től elérhető, azonban a lakosság csak egy bizonyos része számára. A hitel egy speciális konstrukcióval rendelkezik, mely szerint az első gyermek születését követően kamatmentessé válik, a felvételt követően a házaspároknak 5 év áll a rendelkezésükre ezt a feltételt teljesíteni, amennyiben ez nem sikerül, a hitel piaci kamatozásúvá válik, és a korábban igénybe vett támogatást is meg kell téríteni. Ezek alapján úgy vélem, hogy mivel ez egy annyira összetett konstrukció, mely több, nem csak pénzügyi tényezőtől függ, hogy célszerűbb az elemzések során elhagyni.

Ahogy már fentebb írtam, szeretném a futamidő hatását vizsgálni a nemfizetésre, azonban még szerepelnek olyan hitelügyletek, melyek esetében a lejárati idő nem ismert, így **ötödik lépésként** ezeket a hitelügyleteket szűrtem ki az adatbázisból.

Az ötödik lépést követően az adatbázis a kezdeti 10.767.452 db hitelügyletről 4.799.831 db hitelügyletre szűkült, mely 28 féle hiteltípusból tevődik össze. A fennmaradt hitelügyletek 93,97%-a 6 típusba sorolható, a többi hiteltípus 2%-a alatti arányban fordul elő, így az adatbázist ezen a ponton erre a 6, leggyakrabban előforduló típusra szűkítettem. Az alábbi táblázatban láthatók az egyes hiteltípusok, a zöld színnel jelöltek maradtak továbbra is az adatbázisba, az egyes típusok mellett a hozzájuk tartozó gyakoriság és arány látható.

8. táblázat: A szűkebb adatbázis megoszlása hiteltípusok szerint

Hitel típusa	Gyakoriság (db)	Megoszlás (%)
Ismeretlen	1	0,00%
Kombinált egyéb hitel	2	0,00%
Váltó leszámítolási ügylet	2	0,00%
Bankkártya szerződés	4	0,00%
Bank garancia, kezesség	19	0,00%
Ingó jelzálog fedezetű hitel	77	0,00%
Egyéb készpénz hitel	95	0,00%
Befektetési hitel	154	0,00%
Kombinált lakáscélú jelzálog nélküli hitel	301	0,01%
Záloghitel	344	0,01%
Értékpapír kölcsönzés	527	0,01%
Pénzügyi faktoring ügylet	961	0,02%
Egyéb jelzálog fedezetű hitel	1277	0,03%
Lombard hitel	3052	0,06%
Gyűjtőszámla hitel	4499	0,09%
Egyéb lízing szerződés	5235	0,11%
Kombinált gépjármű lízing	10391	0,22%
Kombinált gépjármű vásárlási hitel	14451	0,30%
Östermelői hitel	25328	0,53%
Lakáscélú jelzálog nélküli hitel	34826	0,73%
Kombinált személyi hitel	45715	0,95%
Gépjármű vásárlási hitel	52270	1,09%
Kombinált jelzáloghitel	89906	1,87%
Gépjármű lízing	200151	4,17%
Egyéb hitelügylet	211031	4,40%
Szabad felhasználású jelzáloghitel	221062	4,61%
Lakáscélú jelzáloghitel	689357	14,36%
Fogyasztási, áruvásárlási hitel	730219	15,21%
Személyi hitel	2458574	51,22%
Összesen	4799831	100,00%

Forrás: Saját szerkesztés

Végezetül pedig azokat a hiteleket választottam ki, amiket az elmúlt 5 évben vettek fel, így a végső adatbázisban a 2016-tól felvett hiteleket szerepelnek, melyeknek a száma **2.889.283 db.**

A későbbi elemzések során kiderült, hogy a járások esetében is keletkezett egy „egyéb” kategória, melyeknél így nem megadhatóak a KSH oldaláról felhasznált megyei szintű adatok, így ezt az 1.813 elemet is elhagytam az adatbázisból.

5.3 Mintavétel

Mivel a rendelkezésre álló adatbázis milliós nagyságrendű volt még a szűkítést követően is, így mindenképpen szükséges volt a mintavétel. A mintavétel során fontos szempontnak tartottam, hogy a kutatási célnak megfelelően, de minden egyes elemnek egyenlő esélye legyen a mintába kerülésre.

Az egyes mintavételi lehetőségek ismertetésébe nem megyek bele, kizárólag az általam használt technikákat emelném ki. Kutatási célom és hipotéziseim megkövetelték, hogy több mintával dolgozzak az elemzések során, így a mintavételezés kétféle technikával történt.

Az egyik technika az egyszerű véletlen mintavétel, mely során minden egyes egyednek azonos esélye van bekerülni a mintába. Ehhez az SPSS lehetőségein belül a véletlen minta opciót választottam. Az elemzések elvégzésénél fontos, hogy megfelelő méretű mintával dolgozzak, így végül egy 500 elemű mintát választottam ki elsőként.

A csődelőrejelzések során bebizonyosodott, hogy az általam is alkalmazni kívánt technikák akkor működnek kimondottan jól, ha a mintában (megközelítőleg) azonos az egyes csoportok elemszáma, így a második minta kiválasztásához az egyenletes elosztás lehetőségét választottam. A jól és a rosszul teljesítő hitelek esetén is 250-250 elemet választottam ki egyszerű véletlen mintavétellel, és így jött létre az 500 elemű minta.

A kutatás második felében hiteltípusonként is vizsgáltam a helyes kategorizálást, melyhez új mintára volt szükségem. Mivel a megelőző kutatások alátámasztották, hogy az a minta volt megfelelő, amelyben az egyes csoportok (megközelítőleg) azonos arányban szerepelnek, így ebben az esetben már csak ezzel a lehetőséggel éltem. Minden hiteltípusnál 200 elemű mintát vettem, melyből 100 hitel jól teljesítő, 100 hitel pedig nem jól teljesítő hitel volt. Az egyes csoportokon belül a mintavétel véletlenszerűen történt. (Domán-Szilágyi-Varga, 2009)

6 A kutatás eredményei, hipotézisek vizsgálata

6.1 A kutatáshoz alkalmazott szoftver

A klasszifikációs elemzések elvégzéséhez a SPSS programcsomagot alkalmaztam, mely használható a logisztikus regresszióval, a döntési fával és a neurális hálóval való elemzés elvégzéséhez is.

6.2 Csődelőrejelzési módszerek használata lakossági nemfizetés előrejelzésére

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

Ahogy az irodalmi áttekintés esetében látható volt, vállalatok esetében a legújabb módszerek közé a logisztikus regresszió, a rekurzív particionáló algoritmus és a neurális háló tartozik. Feltételezésem szerint ezek a módszerek alkalmasak arra, hogy a lakossági hitelezés esetében is előre jelezzék, mely ügyfél, hitelügylet fog nemteljesítővé válni.

Nemteljesítőnek azt a hitelügyletet minősítettem, mely rendelkezett mulasztási összeggel.

Az állítás alátámasztásához mindhárom módszerrel elkészítettem a modellt, melyhez egy 500 elemű véletlen mintát választottam első esetben. Az elemzés elvégzéséhez a KHR által átadott adatbázist használtam fel. Az elemzések során kiemelt figyelmet fordítottam a kapott eredmények validálására, melyhez a mintát egy tanuló és egy tesztelő részre osztottam. A mintának a 70%-a alkotta a tanuló adatbázist. Logisztikus regresszió esetében az általam generált particionáló változót használtam ehhez, míg a döntési fa és a neurális háló esetében az elemzésbe épített lehetőséggel éltem.

6.2.1 Logisztikus regresszió I.

Elsőként a logisztikus regressziós elemzést végeztem el. Négy magyarázóváltozó állt a rendelkezésemre. A megfelelő magyarázó változó kiválasztását enter és a backward elimináció segítségével is elvégeztem. Mindkét esetben ugyanaz az az egy magyarázóváltozó bizonyult szignifikánsnak.

Az Omnibusz teszt segítségével megállapítható, hogy a magyarázó változók szignifikáns kapcsolatban állnak-e a függő változóval. Ennek elemzéséhez χ^2 próbát alkalmazunk, és ha a szignifikancia szint kisebb, mint 5%, a modell megbízhatónak tekinthető. A létrejött modell esetében $p < 0,001$, tehát a modell megbízható, továbbá a Hosmer and Lemeshow goodness-of-fit teszt alapján pedig megállapítható, hogy a létrehozott modell megfelelően illeszkedik ($p = 0,212$). A létrehozott modell közepes magyarázó erővel rendelkezik (Nagelkerke $R^2 = 38,8\%$).

9. táblázat: A Logisztikus regresszió I. modell szignifikáns változói

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
Teszt	tör_szerz*100	,024	,007	13,249	1	<,001	1,025
	Constant	-3,594	,499	51,836	1	<,001	,027
Tanuló	tör_szerz*100	,029	,004	49,181	1	<,001	1,030
	Constant	-3,448	,308	125,008	1	<,001	,032

Forrás: Saját szerkesztés

A létrejött modell egyenlete a következő formában írható fel:

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

ahol,

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya.

Tehát amennyiben a törlesztő részlet szerződéses összeghez viszonyított arány esetében 1 százalékpontos növekedés történik, akkor átlagosan 3,0%-kal nő az esélye annak, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.

10. táblázat: Klasszifikációs mátrix a Logisztikus regresszió I. modell esetében

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

Forrás: Saját szerkesztés

A választott vágási érték eltér az alapértelmezett 0,5 értéktől (11.1 melléklet). A szakirodalmi részben tárgyalt ajánlások közül mindkét módszer ugyanazt a 0,039 értéket tartotta optimálisnak, így ezt alkalmaztam én is. Az így választott cut off érték hatására a besorolási pontosság ugyan 93,7%-ról 81,1%-ra csökkent, azonban a nem teljesítő hitelek esetében a helyes kategorizálás aránya 40,7%-ról 85,2%-ra nőtt, így ez tekinthető kedvezőbbnek.

Összességében a létrehozott modell 81,1%-os pontossággal kategorizálta helyesen a hitelügyleteket, 66 elem besorolása volt helytelen. Mivel az alkalmazott cut off érték alacsony, így a véletlen besorolás minden ügyletet nemteljesítőnek minősített, és a véletlen csoportosításhoz (7,7%) képest a 81,1%-os érték jelentős növekedésnek tekinthető.

A tanuló és tesztelő minta között jelentős különbség figyelhető meg a besorolási pontosság, a szenzitivitás és a specificitás esetében egyaránt. A tesztelő minta szenzitivitás értéke jelentősen elmarad a tanuló minta értékétől. Ennek oka lehet, hogy a mintában túlnyomórészt jól teljesítő hitelek szerepelnek, azaz a minta összetétele kedvezőtlen az elemzés számára. Ennek javítása érdekében a későbbiekben egy új mintán is el fogom végezni az elemzést, így egyelőre ezt a modellt véglegesnek tekintem.

A logisztikus regresszió, a döntési fával és a neurális hálóval ellentétben, több feltételt is támaszt az elemzés elvégzéséhez, melyekről a 4.2 fejezetben írtam korábban, így ezeket a feltételeket le kell ellenőrizni:

1. Változók mérési szintje: a függő változó kétszorosított változó, a független változók pedig bármely skálán mérhetők, így ez a feltétel teljesül.
2. Az adatok függetlensége: egy elem egy hitelügyletet jelent, mely független a többi hitelügylettől, tehát ez a feltétel is teljesül.
3. Mintanagyság: 500 elemű minta
4. Multikollinearitás: az elemzés során egy magyarázóváltozó bizonyult csak szignifikánsnak.

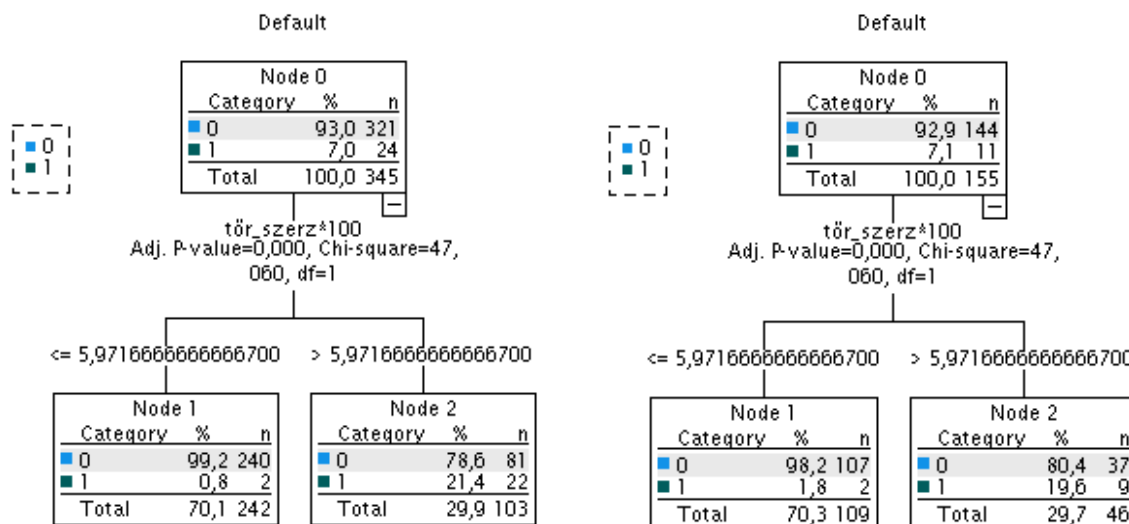
Tehát megállapítható, hogy a létrehozott modell minden feltételnek megfelel, valamint a validálás is sikeres volt.

Mivel az első három feltétel teljesülésére kiemelt figyelmet fordítottam, így a későbbiekben részletesen kizárólag a multikollinearitás tesztelésére térek ki.

6.2.2 Döntési fa I.

Az elemzés megkezdése előtt mindenképpen fontos megemlíteni, hogy a döntési fa hátrányai közé tartozik, hogy hajlamos a túltanulásra, melynek veszélye jelen esetben is fennáll, hiszen a mintában túlnyomó részt jól teljesítő hitelek (93%) szerepelnek.

A döntési fa esetében az algoritmusnak ugyanazon négy magyarázó változó állt a rendelkezésére, mely közül a törlesztő részlet szerződéses összeghez viszonyított aránya bizonyult jól szétválasztó változónak az algoritmus alapján. A tanuló és tesztelő mintán lefuttatott döntési fát a 29. ábra tartalmazza.



29. ábra: Döntési fa I. a tanuló és tesztelő minta esetében

Forrás: Spss output, saját szerkesztés

Ez a döntési fa egy 0. szintből és egy 1. szintből állt. A 0. szinten a teljes adatbázis egyben látható, továbbá látható a függő változó egyes kategóriáinak megoszlása, elemszáma. Ezt követi egy iteratív folyamat, az algoritmus minden magyarázó változó esetében elvégzi az elemzést, majd kiválasztja azt, amelyik a legnagyobb befolyással rendelkezik. Ez a változó jelen esetben a törlesztő részlet szerződéses összeghez viszonyított aránya. Amennyiben ezt követően az algoritmus talál még szignifikáns változókat, akkor a fa további szintekkel bővül, amennyiben nem, akkor pedig az adott szinten véget ér a fa.

Látható, hogy abban az esetben, ha az adott változó értéke kevesebb, mint 5,9717, akkor elenyésző a nemteljesítő hitelek száma.

A besorolások pontosságáról a klasszifikációs mátrix ad információt.

11. táblázat: Klasszifikációs mátrix a Döntési fa I. modell esetében

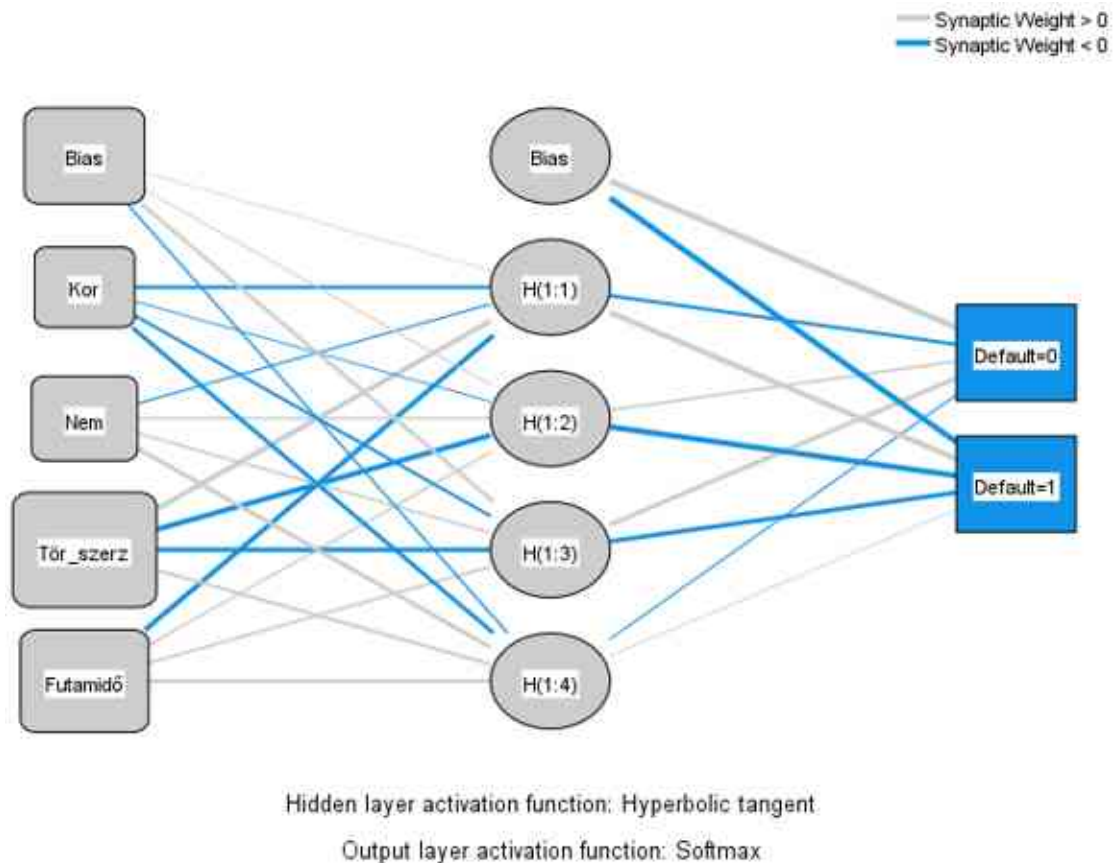
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	312	0	100,0%
	1	24	0	0,0%
	Overall Percentage	100,0%	0,0%	93,0%
Teszt	0	144	0	100,0%
	1	11	0	0,0%
	Overall Percentage	100,0%	0,0%	92,9%

Forrás: Saját szerkesztés

A tanuló adatbázis esetében 93,0%-os besorolási pontosságot ért el a modell, azaz a hitelügyletek 7%-át nem tudta helyesen kategorizálni, azonban a nemteljesítő hitelek közül egyet sem kategorizált be helyesen. Ennek oka, hogy a nemteljesítő hitelek túl alacsony számban fordultak elő a mintában, így az algoritmus a jól teljesítő hitelek besorolást túltanulta. A problémára megoldást jelenthet egy olyan minta kialakítása, melynek a jól és rosszul teljesítő hitelek (megközelítőleg) azonos arányban fordulnak elő.

6.2.3 Neurális háló I.

Neurális hálók esetében a csődelőrejelzés során elterjedt „Multi-Layer Perceptron” lehetőséget választottam. Ahogy a módszertani részben olvashattuk a neurális háló semmilyen előfeltétellel nem él, egy robosztus elemzés. Azonban a döntési fához hasonlóan szintén fennáll a túltanulás veszélye. A rendelkezésre álló változók közül minden változó bevonható az elemzésbe, így az alábbi háló jött létre:



30. ábra: A kialakult neurális háló a Neurális háló I. modell esetében

Forrás: Spss output, saját szerkesztés

A normalizált fontosság értékek alapján (11.3 számú melléklet) az algoritmus legfontosabb magyarázó változónak a törlesztő részlet szerződéses összeghez viszonyított arányát választotta. A létrejött modell magas besorolási pontossággal rendelkezik, a tanuló adatbázis

esetén 6,7%-ban, a tesztelő adatbázis esetében pedig 3,2%-ban csoportosította rosszul a hitelügyleteket, mely részletes értékei az alábbi klasszifikációs táblában láthatóak:

12. táblázat: Klasszifikációs mátrix a Neurális háló I. modell esetében

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%

Forrás: Saját szerkesztés

A klasszifikációs mátrixban láthatjuk, hogy a szenzitivitás értéke elmarad a specificitás értékétől, tehát ez az elemzési módszer is, akárcsak a döntési fa, a jól teljesítő hitelek besorolásánál ér el nagyobb pontosságot. Ebben az esetben is megoldást jelenthet egy olyan minta alkalmazása, melyben a jól és nem jól teljesítő hitelek (megközelítőleg) azonos arányban fordulnak elő.

6.2.4 Új minta kialakítása

Az első három modell eredményeit összevetve megállapítható, hogy a logisztikus regresszióval és a neurális hálóval megalkotott modellek kedvezőbbek, azonban még tovább javíthatók ezek a modellek is.

Ez alapján elmondható, a modellek számára nem kedvező, ha valamelyik csoport jelentősen nagyobb arányban fordul elő a mintában. Ezzel a problémával a csődelőrejelzés kutatói is szembesültek, így elemzéseiket olyan mintákon végezték el, mely esetében az egyes csoportok aránya megegyezett, például Beaver (1966), Altman (1968), Deakin (1972), Odom és Sharda (1990). Ennek megfelelően kialakítottam egy új mintát, amiben szintén 500 elem volt, és ebben a mintában 50%-50% arányban fordultak elő a jól és nem jól teljesítő hitelek.

6.2.5 Logisztikus regresszió II.

Újból lefuttattam az elemzést, ezúttal az újonnan létrehozott mintán. Az Omnibusz teszt alapján egy megbízható modellt kaptam ebben az esetben is ($p < 0,001$), továbbá a Hosmer and Lemeshow goodness-of-fit teszt alapján pedig megállapítható, hogy a létrehozott modell megfelelően illeszkedik ($p = 0,105$). A modell magyarázó erejét tekintve jelentős különbség figyelhető meg a tanuló és a teszt minta esetében.

13. táblázat: A logisztikus regresszió II. modell magyarázóereje

Minta típusa	-2 Log likelihood	Cox & Snell R ²	Nagelkerke R ²
Teszt	137,594 ^a	,416	,555
Tanuló	219,627 ^b	,520	,693

Forrás: Saját szerkesztés

Ahogy a 13. táblázatban látható, a tanuló mintán létrehozott modell magyarázóereje jelentősen meghaladja a teszt minta magyarázó erejét. Az előző esetben szignifikánsnak bizonyult változó, ebben az esetben is annak bizonyult ($p < 0,001$), továbbá bővült a modell a futamidő változóval.

14. táblázat: A Logisztikus regresszió II. modell szignifikáns változói

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
Teszt	tör szerz*100	,031	,007	17,529	1	<,001	1,031
	Futamidő	-,023	,008	8,190	1	,004	,977
	Constant	-,066	,380	,031	1	,861	,936
Tanuló	tör szerz*100	,050	,009	29,528	1	<,001	1,051
	Futamidő	-,019	,006	9,418	1	,002	,982
	Constant	-,662	,287	5,309	1	,021	,516

Forrás: Saját szerkesztés

A létrejött modell egyenlete a következő formában írható fel:

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

ahol,

x₁: a törlesztő részlet szerződéses összeghez viszonyított aránya.

x₂: futamidő.

Tehát ez alapján az állapítható meg, hogy ha törlesztő részlet szerződéses összeghez viszonyított arány esetében 1 százalékpontos növekedés történik, akkor átlagosan 5,1%-kal nő az esélye annak, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett. A futamidő esetében pedig elmondható, hogy ha a futamidő 1 hónappal nő, akkor átlagosan 1,8%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.

Végezetül pedig megvizsgáltam a besorolási pontosságot, melynek részletes értékeit az alábbi klasszifikációs mátrix tartalmazza.

15. táblázat: Klasszifikációs mátrix a Logisztikus regresszió II. modell esetében

Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	60	27	69,0
		1	13	63	82,9
	Overall Percentage				75,5
Tanuló	Default	0	139	24	85,3
		1	32	142	81,6
	Overall Percentage				83,4
a. The cut value is ,390					

Forrás: Saját szerkesztés

Az optimális vágási érték esetében jelentős különbség mutatkozott a két ajánlás között, így megvizsgáltam, hogy az alapértelmezett 0,5 érték, vagy pedig a két ajánlás kedvezőbb-e a kutatási cél szempontjából (11.4 melléklet). A felsoroltak közül a bal felső sarokhoz tartozó optimum bizonyult a legjobbnak, hiszen a besorolási pontosságok között jelentős különbség ugyan nem volt, azonban a szenzitivitás értéke jelentősen magasabb volt ebben az esetben.

Összességében a létrehozott modell 83,4%-os pontossággal kategorizálta helyesen a hitelügyleteket, 56 elem besorolása volt helytelen. A modell segítségével elért pontosság jelentősen meghaladja a véletlen kategorizálással elért eredményt (51,6%).

A klasszifikációs mátrix alapján megállapítható, hogy a tanuló minta esetében magasabb besorolási pontosságot ért el a modell, ami annak köszönhető, hogy a jól teljesítő hiteleket a tanuló minta esetében nagyobb pontossággal azonosította a modell, azonban mivel ez nem a végleges modell, így ebben az esetben a validálást elfogadottnak tekintem.

A logisztikus regresszió esetében egy fontos feltétel, hogy a magyarázó változók függetlenek legyenek, ne álljon fenn multikollinearitás, melyet két módon vizsgáltam. Egyrészt megvizsgáltam a változópárok közötti kapcsolatot lineáris korrelációs együttható segítségével, másrészt pedig a VIF mutatót alkalmaztam.

16. táblázat: Logisztikus regresszió II. modellbe bevont magyarázó változók között fennálló kapcsolat

		tör szerz*100	Futamidő
tör_szerz*100	Pearson Correlation	1	-,248**
	Sig. (2-tailed)		<,001
	N	500	500
Futamidő	Pearson Correlation	-,248**	1
	Sig. (2-tailed)	<,001	
	N	500	500
**. Correlation is significant at the 0.01 level (2-tailed).			

Forrás: Saját szerkesztés

A korrelációs mátrix alapján elmondható, hogy ugyan a két változó között szignifikáns kapcsolat van ($p < 0,001$), azonban ez a kapcsolat egy gyenge, negatív irányú kapcsolat.

17. táblázat: A VIF mutató értékei a Logisztikus regresszió II. modell esetében

Minta típusa		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

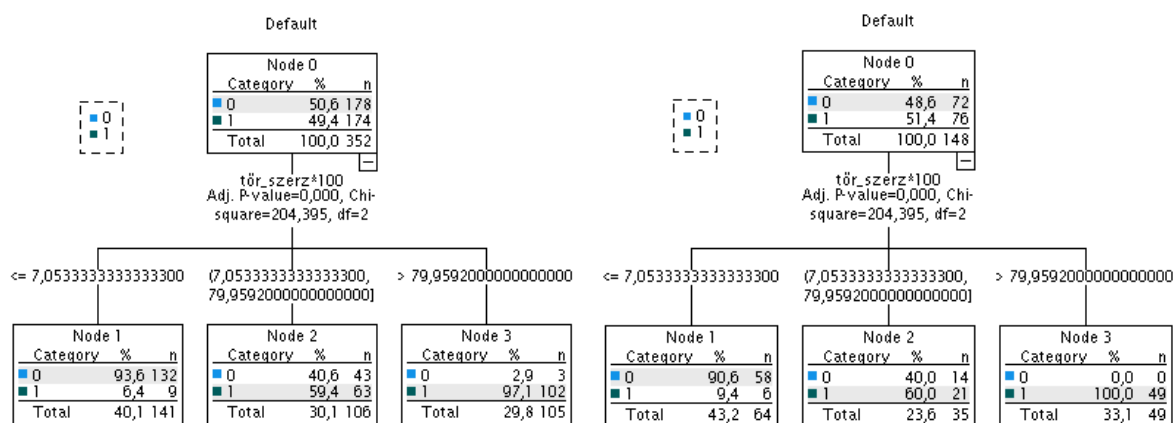
Forrás: Saját szerkesztés

A VIF mutató értéke minden egyes vizsgált esetben 1-hez közeli, továbbá a toleranciamutató értéke is szintén egyhez közeli. A kapott értékek gyenge multikollinearitásra vagy annak hiányára utalnak.

A számított mutatószámok alapján tehát megállapítható, hogy a logisztikus regresszió ezen feltétele teljesül. Az elemzés többi feltételét a 6.2.1 pont alapján teljesítettnek tekintem.

6.2.6 Döntési fa II.

Ismételten ugyanabból a négy magyarázóváltozóból indult ki az algoritmus, és ebben az esetben is ugyanaz a változó bizonyult szignifikánsnak, mint az első esetben, mely a 31. számú ábrán látható. Ez a változó a törlesztő részlet szerződéses összeghez viszonyított aránya



31. ábra: Döntési fa II. a tanuló és tesztelő minta esetében

Forrás: Spss output, saját szerkesztés

Ez a döntési fa is egy szinttel rendelkezik, azonban ebben az esetben a korábbi 2 levél helyett, már 3 alakult ki.

A létrejött fa alapján megállapítható, hogy ha a változó értéke nem haladja meg a 7,053 értéket, akkor alacsony a nemteljesítő hitelek aránya, abban az esetben viszont, ha meghaladja a 79,959 értéket szinte biztos, hogy az adott hitel nemteljesítő hitel.

A besorolások pontosságáról a klasszifikációs mátrix ad információt.

18. táblázat: Klasszifikációs mátrix a Döntési fa II. modell esetében

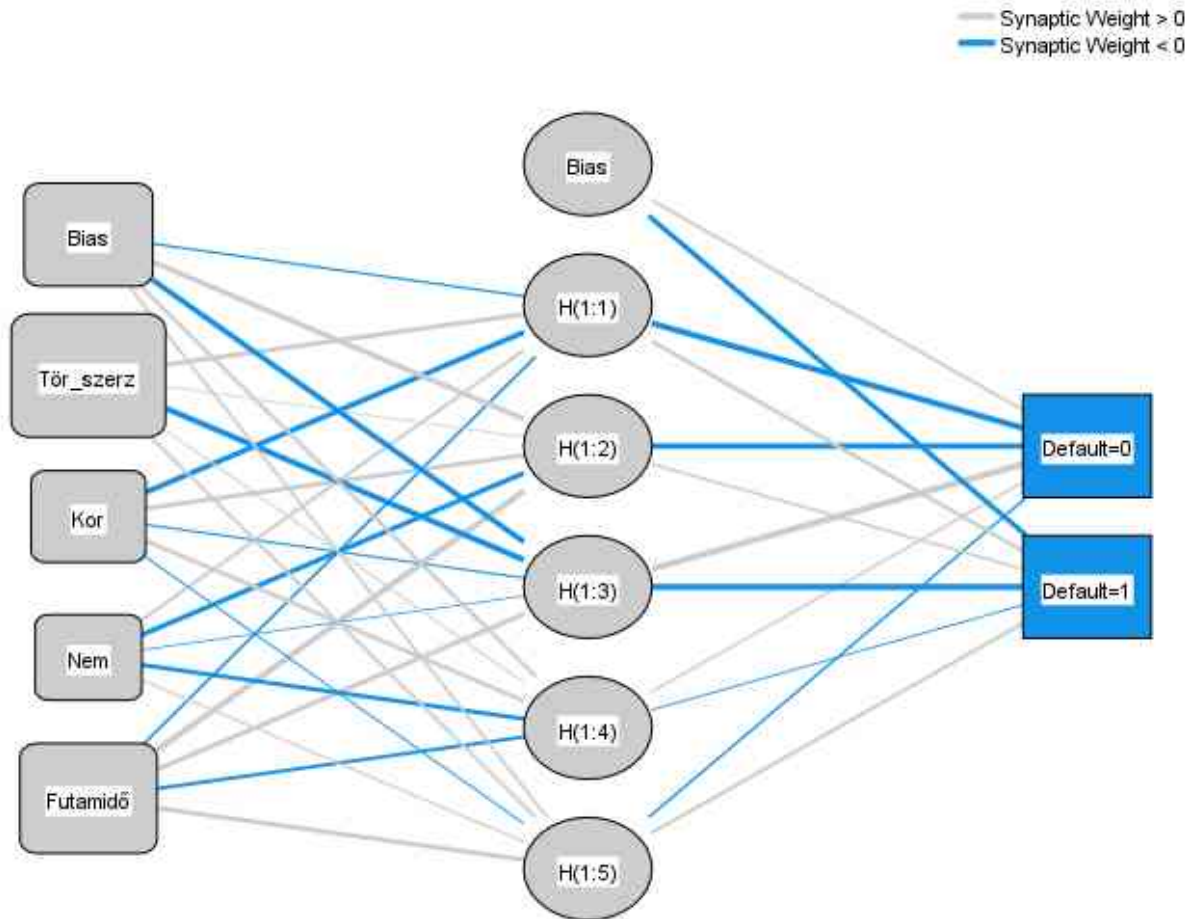
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	132	46	74,2%
	1	9	165	94,8%
	Overall Percentage	40,1%	59,9%	84,4%
Teszt	0	58	14	80,6%
	1	6	70	92,1%
	Overall Percentage	43,2%	56,8%	86,5%

Forrás: Saját szerkesztés

A létrejött modell 86,5%-os pontossággal kategorizálta helyesen a hitelügyleteket, 20 ügylet esetében tévedett csupán. Az első modell besorolási pontosságához képest (93,0%) ugyan romlott ebben az esetben az érték, azonban az első modell esetében az elemzés céljának szempontjából kedvezőtlen volt, hogy egy nem jól teljesítő hitelügyletet sem tudott bekegerezálni. Ebben az esetben viszont a nemteljesítő hitelek jelentős részét a megfelelő csoportba sorolta, továbbá a tanuló és a tesztelő mintán elért eredmény nem különbözik jelentősen egymástól, így a kapott eredményeket érvényesnek tekintem.

6.2.7 Neurális háló II

A kiválasztott független változók köre nem változott az előző modellhez képest, azonban a létrejött neurális háló képe jelentősen eltér, az új háló több neuronnal rendelkezik a rejtett rétegben.



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

32. ábra: A kialakult neurális háló a Neurális háló II. modell esetében

Forrás: Spss output, saját szerkesztés

Amennyiben a különböző független változók jelentőségét vizsgáljuk, továbbra is az állapítható meg, hogy a módszer a legjelentősebb változónak a törlesztőrészlet szerződéses összeghez viszonyított arányát ítéli (11.6 számú melléklet), azonban a további három változó esetében a sorrendet illetően változás történt. Az első modell esetében a legkevésbé jelentős változónak az életkor bizonyult, ebben az új modellben már a nem volt a legkevésbé jelentős magyarázó változó.

A besorolási pontosság ugyan összességében csökkent, azonban a specificitás és szenzitivitás értéke közelít egymáshoz, valamint a nemteljesítő hitelek besorolása jelentősen javult. Összességében az ügyletek 86,2%-ának kategorizálása volt helyes, és 19 ügylet esetében tévedett a modell.

19. táblázat: Klasszifikációs mátrix a Neurális háló II. modell esetében

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

Forrás: Saját szerkesztés

A modell előrejelző képessége hasonló a teszt és a tanuló minta esetében, tehát a modell véglegesíthető.

6.2.8 Eredmények összehasonlítása

A kezdetleges modelleket négy magyarázó változóra építtem, és az alábbi táblázat összefoglalja, hogy a különböző elemzési módszerek mely magyarázó változókat találták szignifikánsnak.

20. táblázat: A klasszifikációs modellek által használt változók összefoglalása

	Elemzés I.			Elemzés II.		
	Log. regr. I.	Döntési fa I.	Neurális háló II.	Log. regr. I.	Döntési fa II.	Neurális háló II.
Tör/szerz*100	X	X	X	X	X	X
Futamidő			X	X		X
Kor			X			X
Nem			X			X

Forrás: Saját szerkesztés

A fentiek alapján megállapítható, hogy a KHR által nyilvántartott adatok közül a hitelek nemfizetését tekintve legjelentősebbnek a törlesztő részlet szerződéses összeghez viszonyított aránya tekinthető.

A különböző minták alapján alkalmazott módszerek besorolási pontosságát tekintve is jelentős különbség figyelhető meg, melyet az alábbi táblázat foglal össze.

21. táblázat: Az alkalmazott módszerek besorolási pontossága

Megfigyelt	Percent correct					
	Logisztikus regresszió I.	Döntési fa I.	Neurális háló I.	Logisztikus regresszió II.	Döntési fa II.	Neurális 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%
Összesen	81,1%	93,0%	93,3%	83,4	84,4%	81,5%

Forrás: Saját szerkesztés

A fentiek alapján megállapítható, hogy minden választott módszer sikeresen, magas pontossággal alkalmazható a nemfizetések előrejelzésére. A kezdeti modellek besorolási pontossága ugyan magasabb volt majdnem minden esetben, azonban jelentős különbség figyelhető meg az egyes csoportok kategorizálásánál. Az újonnan kialakított minta esetében ugyan az előrejelzési képesség csökkent, azonban jelentősen javult a nemteljesítő hitelek helyes kategorizálása, illetve csökkent a szenzitivitás és specificitás közötti különbség, így ezen modellek jobbnak tekinthetők.

A 35. táblázatban szereplő AUC és Gini-koefficiensek alapján minden modell kiemelkedőnek értékelhető, erős modell.

Az elemzések alapján a **H1 hipotézist elfogadom.**

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 beigazolódtott a tútanulás 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.

6.2.9 Az adatbázis bővítése

Hiteligénylés esetén a hitelt nyújtó intézet számos információt kér a potenciális ügyfélről, mely alapján el tudja dönteni, nyújt-e az igénylőnek hitelt vagy sem, ilyen adat például az igénylő igazolt jövedelme. A KHR nyilvántartásában ezek az információk nem szerepelnek, tehát az általam felhasznált információk köre szűkebb, mint amilyen információk alapján a bankoknak meg kell hozniuk a döntést. Ebből kifolyólag döntöttem úgy, hogy szükséges az adatbázis új információkkal való bővítése. Mivel ezek szenzitív adatok, így a tényleges értékek nem voltak elérhetőek számomra, emiatt a KSH adatait használtam fel. A hiteligénylő lakóhelyének járása alapján meghatároztam melyik megyében, illetve régióban él. Az adatbázis bővítéséhez az alábbi táblázatokat használtam fel:

- 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 [%]

Továbbá ezen új adatok segítségével arányszámokat képeztem. Az adatbázis részletes leírása a 5.2 pontban olvasható.

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.

A bevont új változók esetében is mindkét mintán elvégeztem minden egyes elemzést.

6.2.10 Logisztikus regresszió III.

Ismét elvégeztem az elemzést, ezúttal a már bővített adatbázison. Az újonnan létrehozott modell egy megbízható ($p < 0,001$), jól illeszkedő ($p = 0,855$) modell, mely 54,2%-os magyarázóerővel rendelkezik. A tanuló és tesztelő minta esetében jelentős különbség nem figyelhető meg.

A szignifikánsnak ítélt változók köre azonban bővült az eredeti modellhez képest. Az alábbi táblázatban lévő magyarázó változók kerültek be a végleges modellbe:

22. táblázat: A Logisztikus regresszió III. modell szignifikáns változói

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
Teszt	tör szerz*100	,058	,018	10,915	1	<,001	1,060
	ker szerz*100	-,034	,013	6,242	1	,012	,967
	Futamidő	-,067	,030	5,109	1	,024	,935
	Constant	1,158	1,593	,529	1	,467	3,183
Tanuló	tör szerz*100	,039	,007	32,258	1	<,001	1,040
	ker szerz*100	-,016	,004	14,017	1	<,001	,984
	Futamidő	-,050	,016	10,065	1	,002	,951
	Constant	-,089	,758	,014	1	,907	,915

Forrás: Saját szerkesztés

A létrejött modell egyenlete a következő formában írható fel:

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

ahol,

x₁: a törlesztő részlet szerződéses összeghez viszonyított aránya,

x₂: a kereset szerződéses összeghez viszonyított aránya,

x₃: futamidő.

A fentiek alapján megállapítható, hogy

- ha törlesztő részlet szerződéses összeghez viszonyított arány esetében 1 százalékpontos növekedés történik, akkor átlagosan 4,0%-kal nő az esélye annak, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.
- ha az átlagkereset szerződéses összeghez viszonyított arány esetében 1 százalékpontos növekedés történik, akkor 1,6%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.
- A futamidő esetében pedig elmondható, hogy ha a futamidő 1 hónappal nő, akkor átlagosan 8,5%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.

Az új változók bevonása kedvezően hatott a nemteljesítő hitelek besorolására

23. táblázat: Klasszifikációs mátrix a Logisztikus regresszió III. modell esetében

Minta típusa	Megfigyelt		Becsült			Log. regr. I. értékei
			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 ¹²						

Forrás: Saját szerkesztés

Akárcsak a korábbiakban, ebben az esetben is megvizsgáltam mindkét korábban tárgyalt ajánlást az optimális vágási érték meghatározásához. Amennyiben az alapértelmezett 0,5 értéket 0,068-ra csökkentem, kedvezően hat a nemteljesítő hitelek helyes kategorizálására mind a tanuló, mind pedig a tesztelő minta esetében (11.7 melléklet). A 23. számú táblázatban látható, hogy az új változók bevonásának hatására megnőtt a nemteljesítő hitelek esetében a helyes kategorizálás aránya.

Összességében a létrehozott modell 83,1%-os pontossággal kategorizálta helyesen a hitelügyleteket, 58 elem besorolása volt helytelen, ami 8 ügylettel kevesebb, mint ami a Logisztikus regresszió I. esetében volt tapasztalható. Mivel az alkalmazott cut off érték alacsony, így a véletlen besorolás minden ügyletet nemteljesítőnek minősített, és a véletlen csoportosításhoz (7,7%) képest az elért besorolási pontosság jelentős növekedésnek tekinthető.

A multikollinearitás esetében megállapítható, hogy ugyan a futamidő és a kereset szerződéses összeghez viszonyított aránya között közepesen szoros, negatív irányú és szignifikáns kapcsolat van (24. táblázat), azonban bármely mutató elhagyásának hatására jelentősen csökkent a modell magyarázó ereje.

¹² Alacsony vágási pont, azonban a csődelőrejelzések során is volt erre példa, Ohlson (1989) esetében $p=0,038$, az általa használt mintában a csődbe jutott vállalatok aránya 4,85% volt.

24. táblázat: Korrelációs mátrix a multikollinearitás ellenőrzéséhez

		tör_szerz*100	ker_szerz*100	Futamidő
tör_szerz*100	Pearson Correlation	1	,239**	-,201**
	Sig. (2-tailed)		<,001	<,001
	N	500	500	500
ker_szerz*100	Pearson Correlation	,239**	1	-,460**
	Sig. (2-tailed)	<,001		<,001
	N	500	500	500
Futamidő	Pearson Correlation	-,201**	-,460**	1
	Sig. (2-tailed)	<,001	<,001	
	N	500	500	500

** . Correlation is significant at the 0.01 level (2-tailed).

Forrás: Saját szerkesztés

A multikollinearitást VIF-mutató segítségével vizsgálva elmondható, hogy kizárólag gyenge multikollinearitás figyelhető meg, hiszen a mutatók értéke minden esetben a 2 alatti tartományban maradt, így végezetül minden változót benne hagytam a végleges modellben.

25. táblázat: A VIF mutató értékei a Logisztikus regresszió III. modell esetében

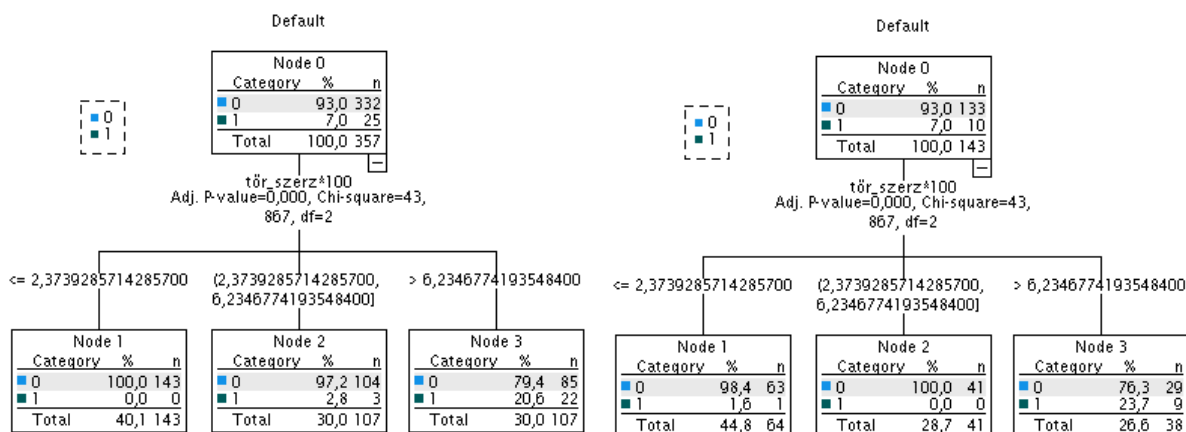
Minta típusa		Collinearity Statistics	
		Tolerance	VIF
Teszt	(Constant)		
	tör_szerz*100	,885	1,130
	ker_szerz*100	,669	1,494
	Futamidő	,739	1,353
Tanuló	(Constant)		
	tör_szerz*100	,936	1,068
	ker_szerz*100	,788	1,269
	Futamidő	,783	1,277

Forrás: Saját szerkesztés

Az elemzés többi feltételét a 6.2.1 pont alapján teljesítettnek tekintem.

6.2.11 Döntési fa III.

A döntési fa esetében az algoritmus nem épített be új változót az eredeti mellé, így szinte teljesen ugyanaz a modell (Döntési fa I.) jött létre ebben az esetben is. A különbség egyetlen oka, hogy az elemzés során a beépített validálás lehetőséget választottam, mely esetben mindig újra és újra alkotja a szoftver a tanuló és tesztelő mintát. A bevont magyarázó változó ugyanaz, mint az első esetben volt, azonban ebben a mintafelosztásban három levéllel rendelkezik a fa.



33. ábra: Döntési fa III. a tanuló és tesztelő minta esetében

Forrás: Spss output, saját szerkesztés

A döntési szabályok alapján az mondható el, hogy ha a változó értéke meghaladja a 6,235-ös értéket, akkor a legnagyobb az esélye a nemteljesítő hitellé válásnak.

A három levél létrejötte nem befolyásolja a modell besorolási pontosságát. Összességében a Döntési fa III. modell a hitelügyletek 93,0%-t kategorizálta helyesen, azonban a rosszul teljesítő hitelek közül egyet sem tudott az algoritmus helyesen kategorizálni.

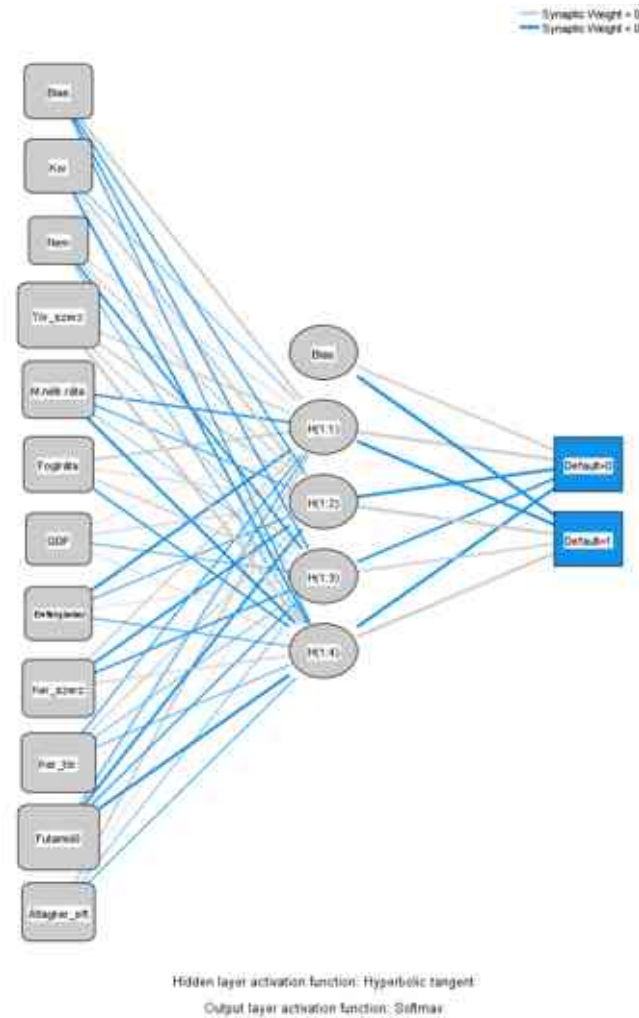
26. táblázat: Klasszifikációs mátrix a Döntési fa III. modell esetében

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	332	0	100,0%
	1	25	0	0,0%
	Overall Percentage	100,0%	0,0%	93,0%
Teszt	0	133	0	100,0%
	1	10	0	0,0%
	Overall Percentage	100,0%	0,0%	93,0%

Forrás: Saját szerkesztés

6.2.12 Neurális háló III.

A neurális háló esetében nincs változószelekció, így tulajdonképpen minden új változó beépült a modellbe, ezzel egy sokkal nagyobb neurális hálót létrehozva, mely 4 neuronnal rendelkezik a rejtett rétegben.



34. ábra: A kialakult neurális háló a Neurális háló III. modell esetében

Forrás: Spss output, saját szerkesztés

A független változók jelentőségét vizsgálva megállapítható, hogy az algoritmus a legjelentősebb 3 változónak a törlesztőrészlet szerződéses összeghez viszonyított arányát, a futamidőt és a kereset szerződéses összeghez viszonyított arányát választotta, melyek a logisztikus regresszió során is szignifikáns változónak bizonyultak. A független változók fontosságát a normalizált fontosság alapján határoztam meg ebben az esetben is, az ehhez tartozó output a 11.9 mellékletben található. A legkevésbé fontos magyarázóváltozónak a nem bizonyult.

27. táblázat: Klasszifikációs mátrix a Neurális háló III. modell esetében

Minta típusa	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%

Forrás: Saját szerkesztés

Az új modell besorolási pontossága növekedett a Neurális háló I.-hez képest az új változók hatására, így már a hitelügyletek 97,7%-át kategorizálta helyesen, és csupán 8 esetben tévedett, azonban a specificitás és a szenzitivitás között még mindig jelentős különbség figyelhető meg.

6.2.13 Logisztikus regresszió IV.

A legutolsó logisztikus regresszió modell új változókkal bővült az előzőkhez képest. A modell megbízható (Omnibus teszt $p < 0,001$), valamint egy jól illeszkedő (Hosmer and Lemeshow goodness-of-fit teszt $p = 0,218$). Mindemellett megállapítható, hogy magas magyarázó erővel rendelkezik (Nagelkerke $R^2 = 74,0\%$). A Logisztikus regresszió II. modellben a törlesztőrészlet szerződéses összeghez viszonyított aránya, valamint a futamidő szerepelt, mint szignifikáns magyarázó változó. Ezek a változók ebben a modellben is helyet kaptak, azonban soruk további kettővel bővült, mely az átlagkereset, illetve a kereset törlesztési összeghez viszonyított aránya.

28. táblázat: A Logisztikus regresszió IV. modell szignifikáns változói

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
Teszt	tör_szerz*100	,022	,007	8,414	1	,004	1,022
	ker_tör*100	-,001	,000	7,493	1	,006	,999
	Futamidő	-,030	,009	10,530	1	,001	,970
	Átlagker (eft)	-,008	,004	5,427	1	,020	,992
	Constant	3,781	1,172	10,405	1	,001	43,868
Tanuló	tör_szerz*100	,038	,008	21,074	1	<,001	1,039
	ker_tör*100	-,001	,000	7,369	1	,007	,999
	Futamidő	-,024	,007	12,388	1	<,001	,976
	Átlagker (eft)	-,008	,003	7,041	1	,008	,992
	Constant	3,027	1,015	8,885	1	,003	20,627

Forrás: Saját szerkesztés

A létrejött modell egyenlete a következő formában írható fel:

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

ahol,

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).

A logisztikus regressziós modellek közül ez az utolsó, továbbá ez rendelkezik a legtöbb független változóval, melyek alapján az alábbi következtetéseket lehet levonni:

- ha törlesztő részlet szerződéses összeghez viszonyított arány esetében 1 százalékpontos növekedés történik, akkor átlagosan 3,9%-kal nő az esélye annak, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.
- ha a kereset törlesztő összeghez viszonyított aránya 1 százalékponttal nő, akkor 0,1%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.
- ha a futamidő 1 hónappal nő, akkor átlagosan 2,4%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.
- ha az átlagkereset 1000 forinttal nő, akkor átlagos 0,8%-kal csökken annak az esélye, hogy az adott ügylet nemteljesítő ügylet lesz, minden más változatlansága mellett.

Az új változók a besorolási pontosságra is kedvezően hatottak.

29. táblázat: Klasszifikációs mátrix a Logisztikus regresszió IV. modell esetében

Minta típusa	Megfigyelt		Becsült			Logisztikus regresszió II.
			Default		Percentage 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						

Forrás: Saját szerkesztés

A választott cut off érték esetében kedvezőbben alakult a nemteljesítő hitelek kategorizálása, mint az alapértelmezett 0,5 érték esetében (11.10 melléklet). A teljes modell esetében javult a besorolási pontosság. Mindemellett a nemteljesítő hitelek esetében is magasabb arányban kategorizálta be a modell az ügyleteket helyesen, tehát az új változók figyelembevétele pozitív hatással volt a modellre.

Összességében a létrehozott modell 87,2%-os pontossággal kategorizálta helyesen a hitelügyleteket, 43 elem besorolása volt helytelen, ami a logisztikus regressziós modellek közül a legkedvezőbb. A modell segítségével elért pontosság jelentősen meghaladja a véletlen kategorizálással elért eredményt (51,6%).

A változók közötti multikollinearitás esetében megállapítható, hogy ugyan szignifikáns kapcsolat figyelhető meg, azonban ez a kapcsolat a legtöbb esetben gyenge.

30. táblázat: Korrelációs mátrix a multikollinearitás ellenőrzéséhez

		tör_szerz*100	ker_tör*100	Futamidő	Átlagker (eft)
tör_szerz*100	Pearson Correlation	1	-,351**	-,248**	-,282**
	Sig. (2-tailed)		<,001	<,001	<,001
	N	500	500	500	500
ker_tör*100	Pearson Correlation	-,351**	1	-,073	,187**
	Sig. (2-tailed)	<,001		,102	<,001
	N	500	500	500	500
Futamidő	Pearson Correlation	-,248**	-,073	1	,217**
	Sig. (2-tailed)	<,001	,102		<,001
	N	500	500	500	500
Átlagker (eft)	Pearson Correlation	-,282**	,187**	,217**	1
	Sig. (2-tailed)	<,001	<,001	<,001	
	N	500	500	500	500
**. Correlation is significant at the 0.01 level (2-tailed).					

Forrás: Saját szerkesztés

31. táblázat: A VIF mutató értékei a Logisztikus regresszió IV. modell esetében

Minta típusa		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

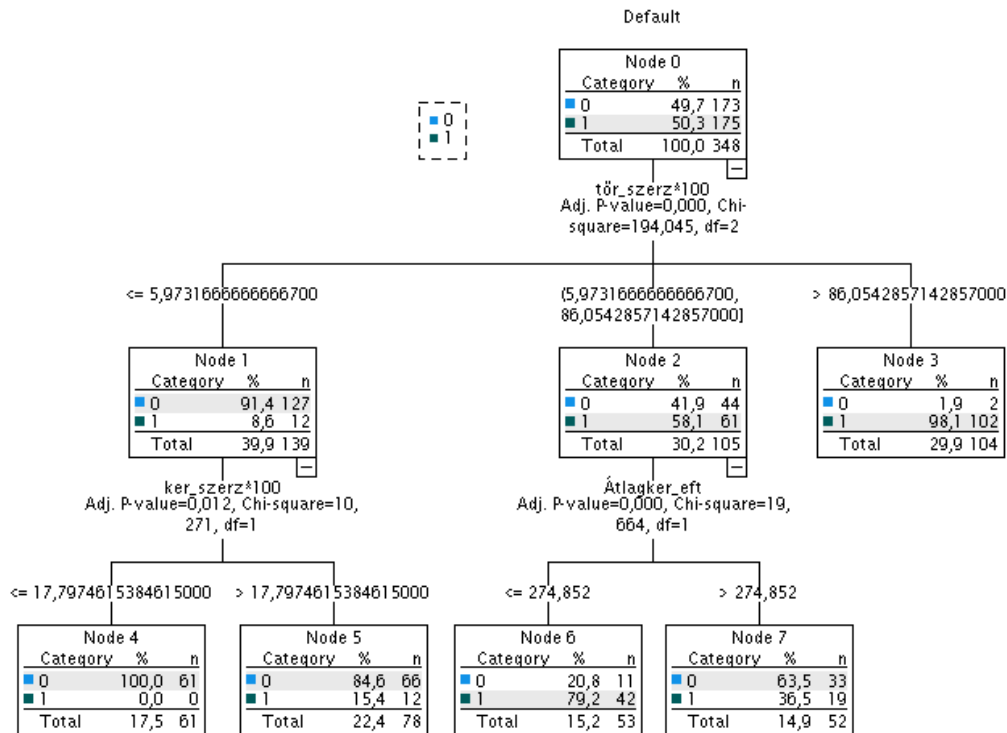
Forrás: Saját szerkesztés

Ez utóbbi esetben is a VIF mutató értékei a 2 érték alatti tartományban maradtak, továbbá a toleranciamutató értékei is magasak, tehát nem sérül a logisztikus regressziószámítás ezen feltétele.

Az elemzés többi feltételét a 6.2.1 pont alapján teljesítettnek tekintem.

6.2.14 Döntési fa IV.

Az utolsó döntési fa esetében az algoritmus legelső szétválasztó változónak az eddig is használt, törlesztőrészlet szerződéses összeghez viszonyított arányát választotta, azaz a fa első szintjén e szerint három csoportba sorolta a hitelügyleteket.



35. ábra: Döntési fa IV. a tanuló minta esetében

Forrás: Spss output, saját szerkesztés

A bal oldali belső csúcs esetében (Node 1) a következő szétválasztó változó a kereset szerződéses összeghez viszonyított aránya, míg a középső belső csúcs (Node 2) esetében az átlagkeresetben különböztek leginkább a két csoport értékei.

A döntési fa alapján megállapítható, hogy ha

- a törlesztő részlet szerződéses összeghez viszonyított aránya kisebb, mint 5,973 és a kereset szerződéses összeghez viszonyított aránya kisebb, mint 17,798, akkor a hitel jól teljesítő hitel. Amennyiben a kereset szerződéses összeghez viszonyított aránya meghaladja a fenti értéket, akkor van esély arra, hogy a hitel nemteljesítővé válik;
- a törlesztő részlet szerződéses összeghez viszonyított aránya nagyobb, mint 86,054, akkor szinte biztos, hogy az adott hitel nemteljesítő hitel;
- a törlesztő részlet szerződéses összeghez viszonyított aránya a fenti két érték közé esik és az ügyfél átlagkeresete kevesebb, mint 274.852 Ft, akkor a hitel nagyobb valószínűséggel nemteljesítő hitel.

32. táblázat: Klasszifikációs mátrix a Döntési fa IV. modell esetében

Minta típusa	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%

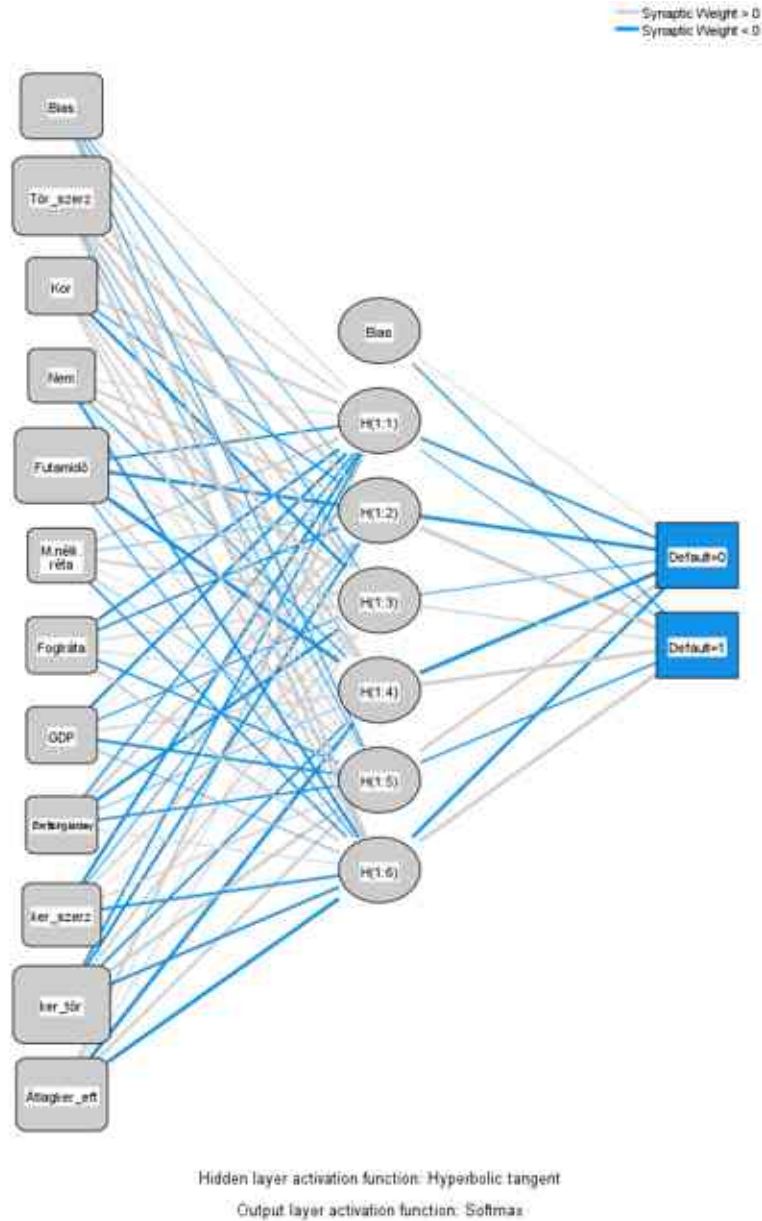
Forrás: Saját szerkesztés

A döntési fa II. modellhez képest javult a besorolási pontosság (32. táblázat), ez a modell az ügyletek 87,4%-át kategorizálta helyesen. A specificitás és szenzitivitás értékei közelebb kerültek egymáshoz, ez alapján ez utóbbi modellt kedvezőbbnek ítélem. A tanuló és a teszt mintán elért eredmény nem különbözik jelentősen egymástól, így a kapott eredményeket érvényesnek tekintem.

A Döntési fa IV. modell esetében megvizsgáltam, hogy más döntési fa eljárás alkalmazása más, esetleg kedvezőbb eredményre vezetne-e, és azt állapítottam meg, hogy az SPSS által támogatott eljárások közül (CHAID, Exhaustive CHAID, QUEST, CRT) a legmagasabb besorolási pontossággal a CHAID alapú eljárás rendelkezett. Az Exhaustive Chaid és a QUEST esetében ugyanúgy a törlesztőrészlet szerződéses összeghez viszonyított aránya bizonyult a legelső szétválasztó változónak, míg a CRT esetében megelőzte ezt a változót a kereset törlesztőrészlethez viszonyított aránya.

6.2.15 Neurális háló IV.

A legösszetettebb neurális háló modellnek a legutolsó bizonyult, 6 neuronnal a rejtett rétegben.



36. ábra: A kialakult neurális háló a Neurális háló IV. modell esetében

Forrás: Spss output, saját szerkesztés

A magyarázó változók jelentőségét vizsgálva az alábbi 5 magyarázó változó bizonyult a legfontosabbnak:

- törlesztőösszeg szerződéses összeghez viszonyított aránya
- kereset törlesztőösszeghez viszonyított aránya
- futamidő

- átlagkereset
- kereset szerződéses összeghez viszonyított aránya.

Ezek mellett azonban a többi változó is fontos az algoritmus számára, hiszen amennyiben elhagyjuk őket, csökken a modell besorolási pontossága. Legkevésbé fontos változónak a nemet ítélte az algoritmus. (11.12 számú melléklet)

A végleges modell magas (88,3%) besorolási pontossággal rendelkezik, azaz a hitelügyletek 88,3%-át kategorizálta a modell helyesen, és 40 hitelügylet esetében tévedett.

33. táblázat: Klasszifikációs mátrix a Neurális háló IV. modell esetében

Minta típusa	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%

Forrás: Saját szerkesztés

Az új változók kedvezően hatottak a nemteljesítő hitelek kategorizálásra, azaz a szenzitivitás értéke növekedett. A besorolási pontosság közel azonos a tanuló és a tesztelő minta esetében, így a kapott modell véglegesíthető.

6.2.16 Eredmények összehasonlítása II.

Összességében tehát 12 darab modell készült, melyből az első hat (Logisztikus regresszió I-II.; Döntési fa I-II.; Neurális háló I-II.) kizárólag a KHR által nyilvántartott adatokat tartalmazta, majd ezt követően bővítettem a modelleket a KSH által nyilvántartott, megyei szintű adatokkal, illetve a belőlük számított arányszámokkal.

Az alábbi táblázatban összefoglaltam, hogy a létrehozott modellek mely magyarázó változókat találták szignifikánsnak.

34. táblázat: A létrehozott modellek magyarázó változóinak összefoglalása

	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

Forrás: Saját szerkesztés

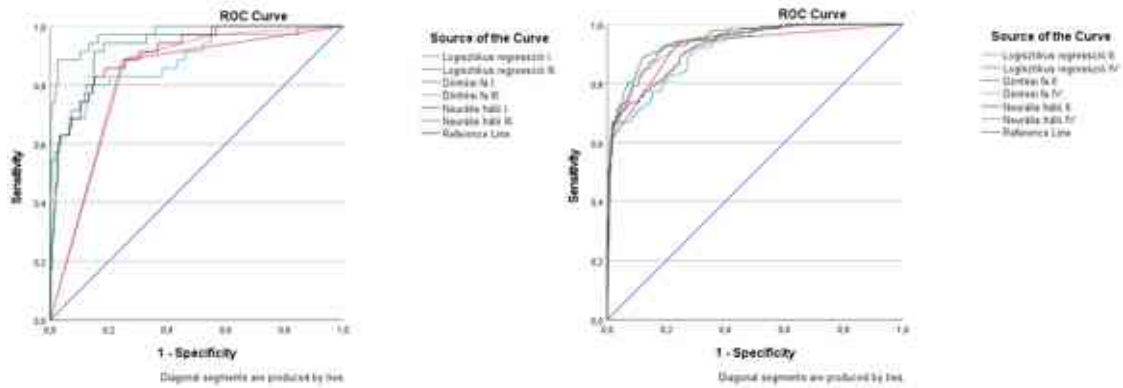
Azon elemzések esetében, melyeknél már a rendelkezésre álltak az új változók, 6 modellből 5-ben szignifikánsnak is bizonyultak, kizárólag a Döntési fa III. modell nem talált közöttük szignifikáns változót, melynek oka feltehetően a minta kedvezőtlen összetétele.

A 34. számú táblázatban megfigyelhető, hogy a törlesztő részlet szerződéses összeghez viszonyított aránya változó minden egyes modellben szerepel, mint magyarázó változó. Ezen kívül több alkalommal bizonyult szignifikánsnak a futamidő, az átlagkereset, a kereset törlesztőrészlethez, illetve szerződéses összeghez viszonyított aránya.

Azonban az új változók beépítése még nem feltétlenül garancia arra, hogy ténylegesen javítja is a modellt. Így mindenképpen fontosnak tartottam, hogy a klasszifikációs módszerek értékeléséhez használt mutatószámokat összehasonlítsam.

Elsőként az ROC görbéket vizsgáltam meg, melyek itt két csoportra osztva láthatók.¹³

¹³ Az egyes elemzések egyedi ROC-görbéje a mellékletben megtalálható.



37. ábra: A létrehozott modellek ROC-görbéje

Forrás: Spss output, saját szerkesztés

Az I. és III. számú modellek esetében jelentősebb különbség figyelhető meg. Az ábra alapján leggyengébb modellnek a Döntési fa I. és Döntési fa III. bizonyul. A jobb oldali ábrán ennyire erős különbségek már nem tapasztalhatók.

Ahhoz, hogy egyszerűbb és egyértelműbb legyen a különböző modellek rangsorolásra, egy táblázatban gyűjtöttem össze a különböző értékelési technikák során kapott eredményeket.

35. táblázat: A kialakított modellek teljesítménye a különböző értékelési technikák alapján

		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

Forrás: Saját szerkesztés

Az AUC érték¹⁴ esetében, a 80-90% körüli érték már kiemelkedőnek tekinthető. Minden létrehozott modell AUC értéke jelentősen meghaladja a 80%-ot, sőt a legtöbb esetben a 90%-ot is.

Hasonló konklúzió vonható le a Gini-kefficiens esetében is, ennél a mutatónál hüvelykujjszabály szerint a 70% feletti érték utal kifejezetten erős modellre. A Döntési fa I. és Döntési fa III. modellt leszámítva, minden esetben ettől magasabb érték tapasztalható.

Mindhárom vizsgált szempont szerint a legjobb modellnek a Neurális háló III. számítana, azonban jelentős különbség figyelhető meg a jól és rosszul teljesítő hitelek kategorizálása során, így ez a modell nem a legideálisabb. A fenti értékek alapján a neurális háló IV. modell besorolási pontossága 88,3%, AUC értéke 94,9%, Gini-koefficiens értéke pedig 89,8%. Ezek alapján ez a modell a legideálisabb a lakossági hitelek nemfizetésének előrejelzéséhez. Második legjobbnak pedig a Logisztikus regresszió IV. bizonyult, melynek értékei alig maradnak el a neurális háló értékeitől. Mindkét modell a bővített adatbázison érte el ezt a teljesítményt, tehát megállapítható, pozitív hatása volt az új változók bevonásának.

Az elemzések alapján tehát a **H2 és H3 hipotézist is elfogadom**, továbbá megállapításaimat az alábbi tézisek foglalják össze:

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 specifitás 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.

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

A csődelőjelzés során alkalmazott módszerekkel hasonló besorolási pontosság érhető el a lakossági hitelek esetében is.

¹⁴ ROC görbe alatti terület nagysága

6.3 Gazdasági, társadalmi, demográfiai jellemzők hatása a mulasztásra

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.

A feltételezések ellenőrzéséhez elsőként az eddig elvégzett klasszifikációs modelleket veszem alapul. Azokat vizsgáltam meg, melyek a teljes állományra vonatkoztak.

36. táblázat: A létrehozott modellek magyarázó változóinak összefoglalása, kiemelve a gazdasági, társadalmi, demográfiai változókat

	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

Forrás: Saját szerkesztés

A logisztikus regresszió és döntési fa modelljei közül mindkét esetben a legutolsó volt az, amelyiknél az átlagkereset szignifikáns magyarázóváltozónak bizonyult. Ezen kívül ezekben a modellekben semmilyen, a KSH oldaláról gyűjtött adat nem szerepelt szignifikáns változóként. A neurális hálók esetében a nehézséget az jelenti, hogy az algoritmus feltétel nélkül minden változót bevon az elemzésbe, tehát a különböző gazdasági, társadalmi, demográfiai változókat egyaránt.

Azt követően a feltételezés megerősítéséhez vagy cáfolásához nemparaméteres próbákat alkalmaztam. A kérdés olyan szempontból érdekes, hogy milyen formában szeretném a nemteljesítést bevonni az elemzésbe. Ugyanolyan formában, mint eddig is használtam (0= jól teljesítő, 1= mulasztásba került), vagy pedig a mulasztás összegét szeretném bevonni. Mivel a

kutatás során a fókusz csak a nemteljesítésen volt, és nem pedig annak nagyságán, így előbbi t választottam, és keresztábra elemzéssel vizsgáltam meg, hogy megfigyelhető-e kapcsolat a régió, illetve a nem esetében.

Elsőként a nem és a nemfizetés kapcsolatát vizsgáltam meg.

37. táblázat: A nem és a nemfizetés közötti kapcsolat vizsgálata

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

Forrás: Spss output, saját szerkesztés

A 37. számú táblázat alapján megállapítható, hogy a nem és a nemfizetés között nincs szignifikáns kapcsolat ($p=0,325$).

A régió és a nemfizetés esetében az elemzés eredményét a 38. táblázat tartalmazza.

38. táblázat: A nem és a nemfizetés közötti kapcsolat vizsgálata

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

Forrás: Spss output, saját szerkesztés

Megállapítható, hogy a vizsgált változók függetlenek egymástól ($p=0,231$).

Tehát összességében mind a nem, mind pedig a régió esetében arra a következtetésre jutottam, hogy nincs szignifikáns kapcsolat ($p>0,05$).

A fentiek alapján a **H4** és **H5** hipotézist elutasítom.

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.

T7: Az elvégzett klasszikációs 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.

6.4 Van-e különbség az egyes hiteltípusok között?

A különböző hiteltípusok esetében a bankok más-más feltételt támaszthatnak az ügyféllel szemben, mely akár a nemfizetésre is hatással lehet. Így szeretném megvizsgálni, hogy jobb modellt kapunk-e abban az esetben, ha hiteltípusonként készítünk elemzést, és milyen különbségek figyelhetők meg a bevont magyarázó változók fontossága között.

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.

A korábban elvégzett elemzések alapján arra a döntésre jutottam, hogy a kérdés megválaszolására azt a mintát használom, melyben a jól és rosszul teljesítő hitelek azonos arányban fordulnak elő, és felhasználom a KSH oldaláról származó magyarázóváltozókat is. A módszerek közül pedig a neurális hálót választottam, hiszen ez produkálta a legjobb eredményt minden tekintetben. Azaz az új modelleket a Neurális háló IV. modellhez fogom hasonlítani.

6.4.1 Egyéb hitelügylet

Akárcsak a teljes modell esetében, itt is a minta 30%-át kiválasztottam validálás céljából. A létrehozott neurális háló 6 neuronnal rendelkezik. A tanuló és tesztelő minta esetében hasonló besorolási pontosságot kaptam. Összességében az egyedi modell az ügyletek 87,3%-t kategorizálta helyesen.

39. táblázat: Az egyéb hitelügylet kategória során kialakult neurális háló értékelése

	Egyéb hitelügylet	Teljes modell
0	88,6%	92,0%
1	86,1%	84,5%
Σ	87,3%	88,3%
AUC	93,3%	94,9%
Gini	86,6%	89,8%

Forrás: Saját szerkesztés

Ahogy a 39. számú táblázatban látható, ha önállóan vizsgáljuk az egyéb hitelügyleteket, nem érhető el jobb eredmény sem a besorolási pontosság, sem az AUC érték és a Gini-kefficiens esetében sem.

Mivel a neurális hálók esetében minden változó a modell részét képezi, így az egyes változók fontosságát vizsgáltam meg. A fontosnak ítélt magyarázó változók között az öt legfontosabb változót emeltem ki. Látható, hogy az egyéb hitelügyletek esetében a kor is megjelent, mint lényeges magyarázó változó, ami a teljes modellben nem szerepelt az öt leglényegesebb változó között.

40. táblázat: Az egyéb hitelügylet kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Egyéb hitelügylet	Teljes modell
Tör/szerz	Tör/szerz
Futamidő	Ker/tör
Átlagker	Futamidő
Kor	Átlagker
Ker/tör	Ker/szerz

Forrás: Saját szerkesztés

6.4.2 Fogyasztási, áruvásárlási hitel

A fogyasztási és áruvásárlási hitel esetében jött létre az eddigi legnagyobb neurális háló, 9 neuronnal. A tanuló és tesztelő minta besorolási pontossága teljesen megegyezik, és az eddigi legmagasabb besorolási pontosságot érte el a modell, az ügyletek 96,5%-a került a helyes csoportba, tehát ebben az esetben javasolt lehet egy önálló modell alkalmazása.

41. táblázat: A fogyasztási és áruvásárlási hitel kategória során kialakult neurális háló értékelése

	Fogyasztási és áruvásárlási hitel	Teljes modell
0	97,2%	92,0%
1	95,7%	84,5%
Σ	96,5%	88,3%
AUC	99,8%	94,9%
Gini	99,6%	89,8%

Forrás: Saját szerkesztés

Az algoritmus olyan változókat ítélt meg fontosnak, melyet egy korábbi modell sem, ez a GDP, illetve a munkanélküliségi ráta. Ennek háttérében feltehetően az állhat, hogy ezt a hiteltípust

könnyen lehet igényelni, és kevesebb feltétel teljesülési szükséges a folyósításhoz. Alapvetően a társadalomnak egy bizonyos rétegére jellemző ezen hitelek felvétele.

42. táblázat: A fogyasztási és áruvásárlási hitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Fogyasztási és áruvásárlási hitel	Teljes modell
Tör/szerz	Tör/szerz
Ker/szerz	Ker/tör
GDP	Futamidő
Ker/tör	Átlagker
Munkanélküliségi ráta	Ker/szerz

Forrás: Saját szerkesztés

6.4.3 Gépjármű lízing

Gépjármű lízing esetében az figyelhető meg, hogy a rosszul teljesítő hitelek besorolása jobb, mint az alap modell esetében, és a modell teljesítménye is jobb, tehát ennél a hitelnél is célszerű lehet a különálló modell alkalmazása. Az önálló modell 90,8%-os besorolási pontossággal rendelkezik, azaz az ügyletek kevesebb, mint 10%-ban volt a kategorizálás téves.

43. táblázat: A gépjármű lízing kategória során kialakult neurális háló értékelése

	Gépjármű lízing	Teljes modell
0	90,5%	92,0%
1	91,2%	84,5%
Σ	90,8%	88,3%
AUC	97,2%	94,9%
Gini	94,4%	89,8%

Forrás: Saját szerkesztés

A változók esetében a futamidőt kiszorította az adott megye 1 főre jutó GDP-jének az értéke, amely a terület gazdasági helyzetéről nyújthat információt.

44. táblázat: A gépjármű lízing kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Gépjármű lízing	Teljes modell
Átlagkereset	Tör/szerz
Tör/szerz	Ker/tör
Ker/tör	Futamidő
Ker/szerz	Átlagkereset
GDP	Ker/szerz

Forrás: Saját szerkesztés

6.4.4 Lakáscélú jelzáloghitel

Lakáscélú hitelek esetében egy kisebb neurális háló jött létre, mely két neuronnal rendelkezik a rejtett rétegben. Ennek a modellnek a teljesítménye hasonló az alap modell teljesítményéhez, elenyésző különbség figyelhető meg csupán. A szenzitivitás értéke viszont jelentősen javult, ami kedvező. Összességében az ügyletek 90,6%-t kategorizálta rendesen a modell.

45. táblázat: A lakáscélú jelzáloghitel kategória során kialakult neurális háló értékelése

	Lakáscélú jelzáloghitel	Teljes modell
0	90,1%	92,0%
1	91,2%	84,5%
Σ	90,6%	88,3%
AUC	93,2%	94,9%
Gini	86,4%	89,8%

Forrás: Saját szerkesztés

A modell alapvetően ugyanazokat a változókat választotta egy kivételével, mely az átlagkereset, helyette az életkor szerepel az öt legfontosabb változó között.

46. táblázat: A lakáscélú jelzáloghitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Lakáscélú jelzáloghitel	Teljes modell
Tör/szerz	Tör/szerz
Ker/szerz	Ker/tör
Ker/tör	Futamidő
Kor	Átlagker
Futamidő	Ker/szerz

Forrás: Saját szerkesztés

6.4.5 Szabad felhasználású jelzáloghitel

Szabad felhasználású hitel esetében az önálló modell teljesítménye szinte teljesen megegyezik az alap modell teljesítményével, így sem előnnyel, sem hátránnyal nem jár az önálló modell használata, az ügyletek 88,0%-t kategorizálta helyesen.

47. táblázat: A szabad felhasználású jelzáloghitel kategória során kialakult neurális háló értékelése

	Szabad felhasználású jelzáloghitel	Teljes modell
0	92,1%	92,0%
1	83,3%	84,5%
Σ	88,0%	88,3%
AUC	95,7%	94,9%
Gini	91,4%	89,8%

Forrás: Saját szerkesztés

Azonban az önálló modell esetében más változók bizonyultak fontosnak, melyek főként az adott megye gazdasági helyzetét jellemzik.

48. táblázat: A szabad felhasználású jelzáloghitelek kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Szabad felhasználású jelzáloghitelek	Teljes modell
Ker/tör	Tör/szerz
Tör/szerz	Ker/tör
Foglalkoztatási ráta	Futamidő
Munkanélküliségi ráta	Átlagker
GDP	Ker/szerz

Forrás: Saját szerkesztés

6.4.6 Személyi hitel

Az utolsó csoport pedig a személyi hitelek. A személyi hitelek esetében az állapítható meg, mint a szabad felhasználású hitelek során. Közel azonos besorolási pontosság jött létre, a rosszul teljesítő hitelek besorolása kis mértékben haladta meg az alap modell kategorizálását. Az új modell az ügyletek 89,4%-ot sorolta be helyesen.

49. táblázat: A személyi hitel kategória során kialakult neurális háló értékelése

	Személyi hitel	Teljes modell
0	92,0%	92,0%
1	86,6%	84,5%
Σ	89,4%	88,3%
AUC	95,4%	94,9%
Gini	90,8%	89,8%

Forrás: Saját szerkesztés

A magyarázó változók köre teljes mértékben megegyezik az alap modell magyarázó változóival.

50. táblázat: A személyi hitel kategória során alkalmazott 5 legfontosabb magyarázó változó és a teljes modell 5 legfontosabb magyarázó változójának összehasonlítása

Személyi hitel	Teljes modell
Ker/tör	Tör/szerz
Tör/szerz	Ker/tör
Átlagker	Futamidő
Ker/szerz	Átlagker
Futamidő	Ker/szerz

Forrás: Saját szerkesztés

6.4.7 Összegzés

Az alábbi táblázat foglalja és hasonlítja össze az egyes modellek és a teljes modell által elért eredményeket.

51. táblázat: A modellek teljesítményének összehasonlítása

	Egyéb hitelügylet	Fogy. és áruvásárlási hitel	Gépjármű lízing	Lakáscélú jelzáloghitel	Szabad felhasz. jelzáloghitel	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%

Forrás: Saját szerkesztés

Megállapítható, hogy az egyéni modellek közül egyik sem marad el jelentősen a teljes modell teljesítményétől, közel azonos, vagy jobb teljesítmények születtek. A fogyasztási és áruvásárlási, a gépjármű lízing és a lakáscélú jelzáloghitel esetében a szenzitivitás értéke számottevően növekedett, tehát a létrejött egyéni modellek nagyobb pontossággal tudták beazonosítani a nemteljesítő hiteleket ebben az esetben.

A legtöbb hiteltípus esetében a modellek mérésére alkalmazott mutatószámok közül legalább 2 meghaladja az alapmodell értékét, tehát azok jobbnak tekinthetők.

A fenti elemzések alapján a **H6 hipotézist elfogadom**.

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.

7 A kutatás főbb megállapításai, következtetések, javaslatok

7.1 A disszertáció tézisei, új és újszerű megállapításai

A kutatás alapjául szolgáló adatbázist a BISZ Zrt. biztosította, így röviden bemutattam a KHR-t, annak tevékenységét. A kezdeti, a BISZ Zrt. által átadott adatbázis több, mint 10.000.000 hitelügyletet tartalmazott, így mindenképpen szükség volt az adatbázis szűkítésére, melyet a kutatási céloknak megfelelően meg is tettem. A mintavétel alapjául végül egy 2.887.470 hitelügyletet tartalmazó adatbázis szolgált, melyből a kutatási céloknak megfelelően kétféle technikával vettem mintát.

A kutatás főbb eredményeit a megfogalmazott hipotézisek mentén foglalom össze:

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

Nemteljesítőnek azt a hitelügyletet minősítettem, mely rendelkezett mulasztási összeggel.

Az állításom alátámasztásához elsőként egy 500 elemű véletlen mintát használtam, melyen végrehajtottam a klasszifikációs eljárást mindhárom elemzési eszköz segítségével. Ugyan a modellek magas besorolási pontossággal rendelkeznek, azonban nagy különbség alakult ki a specificitás és a szenzitivitás között, így a modellek nem voltak optimálisnak tekinthetők, mely a minta tulajdonságaira vezethető vissza. Ezért egy új mintát alakítottam ki, melyben azonos arányban szerepeltek a jól teljesítő és a nemteljesítő hitelek. Az új minta kedvezően hatott a nemteljesítő hitelek besorolására. Így az elemzések alapján a H1 hipotézist elfogadtam, és az alábbi megállapításokat fogalmaztam meg:

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.

Mivel az adatbázis nem tartalmazott minden olyan információt, amit a hitelt nyújtó intézetek kérhetnek, így szükségesnek ítéltam az adatbázis bővítését, melyhez a KSH oldalán elérhető adatokat használtam.

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 hitelgényléshez bekért információktól (pl. jövedelemadatok). Új, a KHR által nem nyilvántartott változók bevonásával javítható a besorolási pontosság.

Majd ezt követően mindkét minta esetében ismételt elvégeztem a klasszifikációs elemzéseket, és látható volt, hogy kedvezően hatott az elemzésekre az új változók bevonása. Legjobb modellnek végül a Neurális háló IV. modellt ítélt meg, mely 88,3%-os besorolási pontosságot ért el. Tehát a **H2 és H3 hipotézist is elfogadtam.**

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 háló rendelkezik, a legjobbnak ítélt modell pedig a Neurális háló IV., mellyel közel 90%-os besorolási pontosság érhető el.

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

A csődelőrejelzés során alkalmazott módszerekkel hasonló besorolási pontosság érhető el a lakossági hitelek esetében is.

A kutatás következő állomásaként kitértem arra is, hogy a különböző gazdasági, társadalmi, demográfiai mutatóknak van-e jelentősége a nemfizetés során.

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.

A hipotézisek alátámasztásához a korábban végrehajtott elemzéseket vettem alapul, valamint keresztábrák elemzést végeztem. Mindkét elemzési lehetőség alapján a **H4 és H5 hipotézist elutasítottam.**

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.

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.

Mivel a különböző típusú hitelek felvétele eltérő feltétellel járhat, így utolsó lépésként megvizsgáltam, megfigyelhető-e különbség abban az esetben, ha egy-egy mintát kizárólag egy hiteltípus alkot.

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.

Az elemzést 6, előre kiválasztott hiteltípus esetében végeztem el, és egyes típusoknál megfigyelhető volt, hogy jobb eredményt produkáltak ezek az új, egyéni modellek. Tehát a **H6 hipotézist elfogadtam.**

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.

7.2 Az eredmények felhasználhatósága, további kutatási irányok

A megállapításaimat a 2016-2021 évek adatai alapján vontam le.

A kutatás során korlátot jelent, hogy az elemzéseket, a szűkített adatbázis hat leggyakoribb hiteltípusa esetében végeztem el, továbbá voltak olyan hitelek, amelyeket, azok jellemzői miatt ki kellett hagynom az elemzésekből (pl. Babaváró hitel). Érdekes terület lehet a későbbiekben ezen hiteleknek a vizsgálata.

A kapott eredményeket torzíthatja, hogy a COVID-19 vírus miatt 2020. március 19-én az állam bevezette a törlesztési moratóriumot, mely az adatbázis lekérdezésekor még fennállt. Emiatt 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. Az OECD 2021. évben készült felmérésében is említette, hogy valószínűsíthető a nemteljesítő hitelek számának növekedése a moratórium eltörlését követően. (OECD, 2021b)

A kutatás és az eredmények hasznosíthatóságát tekintve úgy gondolom, hogy ez egy hiányt pótló téma, kevés kutató, kevés tudományos munka foglalkozik a lakossági hitelek nemfizetésének jellemzésével, előrejelzésével statisztikai módszerek segítségével.

8 Összefoglalás

Doktori kutatásom során a lakossági hitelek nemfizetését, illetve a nemfizetés előrejelzését vizsgáltam. Célom egy olyan modell kialakítása volt, melynek a segítségével magas pontossággal lehet kategorizálni a nemfizető hiteleket.

A kutatás első felében a témához kapcsolódó szakirodalmat tanulmányoztam, mely során kitértem a hitelezés történetére egy rövid áttekintés segítségével, megvizsgáltam a jelenlegi hitelezési helyzetet Magyarországon, illetve azt, hogy jelenleg milyen az eladósodottság. Ezt követően kitértem arra, milyen szempont szerint lehet a lakosság részére elérhető hiteleket csoportosítani, valamint melyek azok az alapfogalmak, amelyek mindenképpen nagy jelentőséggel bírnak a hitelezés során. Továbbá röviden megvizsgáltam, melyek azok a tényezők, amelyeknek fontos szerepe lehet a hitelezési magatartás esetében, annak ellenére, hogy ezek olyan tényezők, amelyek mérése nehéz.

Mivel a lakossági hitelek besorolásával elenyésző mennyiségű szakirodalom foglalkozik, így a kutatás alapjául a csődelőrejelzési modellek szolgáltak. Ezen tudományág megismeréséhez egy időbeli és módszertani áttekintést végeztem. Kiindultam a legelső, még egyváltozós csődelőrejelzési elemzésekből, megvizsgáltam Altman híres modelljét, azt, hogy milyen fejlődésen ment keresztül ez a tudományág, egészen a napjainkban használatos neurális hálóig.

Miután feltérképeztem a leggyakrabban használt módszertanok körét, egy rövid módszertani áttekintés következett, amelyben kitértem az egyes elemzési lehetőségek alkalmazásának korlátjaira, előnyeire, hátrányaira, majd végül arra a következtetésre jutottam, hogy az elemzési lehetőségek közül a logisztikus regressziót, a döntési fát és a neurális hálót szeretném a kutatásaim során alkalmazni. Szintén ebben a fejezetben foglaltam össze, hogy melyek azok a technikák, amelyek segítségével a klasszifikációs eljárások értékelhetők, összehasonlíthatók.

A kutatás alapjául szolgáló adatbázist a BISZ Zrt. biztosította, így röviden bemutattam a KHR-t, annak tevékenységét. A kezdeti, a BISZ Zrt. által átadott adatbázis több, mint 10.000.000 hitelügyletet tartalmazott, így mindenképpen szükség volt az adatbázis szűkítésére, melyet a kutatási céloknak megfelelően meg is tettem. A mintavétel alapjául végezetül egy 2.887.470 hitelügyletet tartalmazó adatbázis szolgált, melyből a kutatási céloknak megfelelően kétféle technikával vettem mintát.

Az elemzések alapján arra a következtetésre jutottam, hogy az alkalmazott módszerek abban az esetben működnek optimálisan, ha a minta összetétel esetében (megközelítőleg) azonos arányban figyelhetők meg a jól teljesítő és a nemteljesítő hitelek. A legpontosabb kategorizálást a neurális háló módszer segítségével sikerült elérni, mely nem csak a KHR által nyilvántartott adatokból indult ki, hanem felhasználta a KSH oldaláról összegyűjtött információkat is.

Megállapítottam továbbá, hogy a különböző gazdasági, társadalmi, demográfiai mutatók nem állnak kapcsolatban a nemfizetéssel.

Végezetül pedig megállapítottam, hogy vannak olyan hiteltípusok, amelyek esetében javasolt lehet az elemzéseket önállóan elvégezni. Ennek háttérében az egyes típusok folyósításának követelményei állhatnak esetlegesen, hiszen vannak olyan hitelek, amelyeknek felvételéhez kisebb korlátokat támasztanak a szolgáltatók.

Úgy gondolom, hogy a kutatás elején kitűzött célokat teljesíteni tudtam, és bízom benne, hogy a kapott eredmények mások számára is hasznosnak bizonyulnak.

9 Summary

My PhD research investigated the default and prediction of default on household loans. My aim was to develop a model that can be used to categorise defaulted loans with high accuracy.

In the first half of the research, I reviewed the literature on the topic. I gave a brief overview of the history of lending, the current lending situation in Hungary and the current level of indebtedness. I have explained how the loans available to the public can be categorised, and what are the basic concepts that are of great importance when it comes to lending. I have also briefly examined the factors that may play an important role in lending behaviour, despite the fact that these are factors that are difficult to measure.

Since there is a scarcity of literature on the classification of retail loans, the research was based on bankruptcy prediction models. To get to know this discipline, I conducted a chronological and methodological review. Starting from the first univariate bankruptcy prediction analyses, I examined Altman's famous model and the evolution of this discipline up to the neural network used today.

After having identified the most commonly used methods, a short methodological overview followed, in which I discussed the limitations, advantages and disadvantages of each of the analysis options, and finally I came to the conclusion that I would like to use logistic regression, decision tree and neural network in my research. Also in this chapter I summarised the techniques that can be used to evaluate and compare the classification procedures.

The database on which the research was based was provided by BISZ Ltd., so I briefly introduced the KHR and its activities. The initial database provided by BISZ Ltd. contained more than 10.000.000 loan transactions, so it was necessary to narrow down the database, which I did according to the research objectives. Finally, the sampling was based on a database of 2,887,470 loan transactions, from which I took a sample using two different techniques.

Based on the analyses, I conclude that the methods used work optimally when the sample composition has (approximately) the same proportion of performing and non-performing loans. The most accurate categorisation was achieved by using the neural network method, which not only used the data recorded by the KHR, but also used information collected from the KSH.

I also found that various economic, social and demographic indicators are not related to non-payment.

And finally, I have found that there are some types of loans for which it may be advisable to carry out the analyses independently. This may be due to the disbursement requirements of each type, as there are some loans for which providers have lower restrictions.

Úgy gondolom, hogy a kutatás elején kitűzött célokat teljesíteni tudtam, és bízom benne, hogy a kapott eredmények mások számára is hasznosnak bizonyulnak.

I believe that I was able to meet the objectives I set at the beginning of the research and I am confident that the results will be useful to others.

10 Irodalomjegyzék

1. Afifi, A., May, S., Clark, V.A. (2012): Practical Multivariate Analysis 5th edition, CRC Press, New York
2. Ajitesh, K. (2023): Logit vs Probit Models: Differences, Exmples <https://vitalflux.com/logit-vs-probit-models-differences-examples/>
3. Altman, E. I. (1968): Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, The Journal of Finance, Vol. 23. No. 4. 589-609. old.
4. Altman, E. I., Haldeman, R., Narayanan, P. (1977): Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations. Journal of Banking & Finance, 1, 1977.
5. Andersen, H. (2008): Failure prediction of Norwegian banks: A logit approach. Norges Bank Working Paper, No. 2., 2008.
6. ASIC (2011): Financial literacy and behavioural change. Australian Securities and Investments Commission March 2011 <http://download.asic.gov.au/media/1343582/rep230-financial-literacy-and-behavioural-change.pdf>
7. Atkinson, A., Messy, F. (2012), “Measuring Financial Literacy: Results of the OECD / International Network on Financial Education (INFE) Pilot Study”, OECD Working Papers on Finance, Insurance and Private Pensions, No. 15, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5k9csfs90fr4-en>
8. Bánfi, Z. (2013): Lehet-e túlzott mértékű a lakossági eladósodottság? Pénzügyi Szemle Online, 2013. április 24. <https://www.penzugyiszemle.hu/vitaforum/lehet-e-tulzott-merteku-a-lakossagi-eladosodottsag>
9. Bankmonitor.hu. (2014). Eladósodott magyarok? Hol állunk Európában? Bankmonitor: <https://bankmonitor.hu/cikk/eladosodott-magyarok-hol-allunk-europaban/>
10. Bankmonitor.hu. (2015). Soha nem gondolná: a magyarok a legkevésbé eladósodottak Közép-Európában. Bankmonitor.hu: <https://bankmonitor.hu/cikk/soha-nem-gondolna-a-magyarok-a-legkevesbe-eladosodottak-kozep-europaban/>
11. Bankszövetség fogalomtár: <https://www.bankszovetseg.hu/fogalomtar/h.cshtml>
12. Bankweb: Hogyan kellene döntenünk? <https://www.bankweb.hu/hogyan-kellene-dontenunk/>
13. Bareith, T., Koroseczné Pavlin, R., Kövér, Gy. (2014): Felszámolások vizsgálata a Nyugat-dunántúli régióban E-conom 2014 III. /2. p102-124
14. Brassai, S. T. (2019): Neurális hálózatok és a fuzzy logika, Scientia Kiadó, 2019, Kolozsvár
15. Beaver, W. H. (1966): Financial ratios as predictors of failure, Empirical Research in Accounting: Selected Studies, Journal of Accounting Research, Supplement to Vol. 5. 71-111. old.
16. Benedek, G. (2000–2001): Evolúciós alkalmazások előrejelzési modellekben, I–II. Közgazdasági Szemle, 12. sz. 988–1007. o. és 1. sz. 18–30. o.
17. Benedek, G. (2003): Evolúciós gazdaságok szimulációja. PhD-értekezés, BKÁE matematikai közgazdaságtan és ökonometria tanszék, Budapest.

18. Béres, D. (2013). Pénzügyi kultúra - mi is ez valójában? Forrás: Pénzügyi Szemle online: <https://www.penzugyiszemle.hu/vitaforum/a-penzugyi-kultura-mi-is-ez-valojaban>
19. Béres, D., Huzdik, K. (2015). A pénzügyi kultúra megjelenése makrogazdasági szinten. Pénzügyi Szemle. Forrás: https://asz.hu/storage/files/files/penzugyiszemle/2012/322_336_beres_huzdik.pdf
20. Béres, D., Huzdik, K., Németh, E., Zsótér, B. (2015). Pénzügyi személyiség - A magyar lakosság pénzügyi magatartása, szokásai és attitűdjei. Forrás: Pénziránytű Alapítvány: https://www.penziranytu.hu/sites/default/files/csatolmany/penzugyi_szemelyiseg_kutatasi_jelentes_2015_11_02.pdf
21. Béres, D., Huzdik K. (2022): A fiatal felnőttek és a hitelek (kockázatkerülési csapda) Pénzügyi szemle, 2022/1 33-55
22. Engelmann, B., Hayden, E., Tasche, D. (2003): Testing rating accuracy, 2003 január, Credit Risk, www.risk.net.
23. Besenyey, L., Gidai, E., Nováky, E. (1982): Előrejelzés, megbízhatóság, valóság Közgazdasági és Jogi Kiadó, Budapest
24. Blum, M. (1974): Failing company discriminant analysis. Journal of Accounting Research, Vol. Spring pp.1-21.
25. Boritz, J.E., Kennedy, D.B., Albuquerque, A.M. (1995): Predicting Corporate Failure Using a Neural Network Approach. Intelligent Systems in Accounting, Finance and Management, vol. 4, no. 2, pp. 95-111.
26. Botos, K., Boros, J., Béres, D., Csernák, J., Németh, E. (2012/3): Pénzügyi kultúra és kockázatvállalás a közép-alföldi háztartásokban Pénzügyi Szemle 2012/3 (p291-309.) https://asz.hu/storage/files/files/penzugyiszemle/2012/291_309_botos_es_tsai36.pdf
27. Brown, M., Henchoz, C., Spycher, M. (2017): Culture And Financial Literacy. Working Papers on Finance No. 2017/03. Swiss Institute of Banking And Finance (S/Bf – Hsg)
28. Charitou, A., Neophytou, E., Charalambous, C. (2004): Predicting Corporate Failure: Empirical Evidence for the UK. European Accounting Review, vol. 13, no. 3, pp. 465-497.
29. Collins, R., Green, R. (1982): Statistical methods for bankruptcy prediction. Journal of Economics and Business, 34(4), pp. 349–354.
30. Curott, N. A. (2016): Studies in Applied Economics, Adam Smith’s Theory of money and banking (2016 febr.) http://sites.krieger.jhu.edu/iae/files/2017/04/Curott_Smiths_Theory.pdf
31. Chaia, A, Dalal, A., Goland, T., Gonzalez, M. (2009). Half the World is Unbanked. https://www.findevgateway.org/sites/default/files/publications/files/mfg-en-paper-half-the-world-is-unbanked-oct-2009_0.pdf
32. Cséffai, P., Almási, B., Dr. Bárczi, J. (2019) Közvetlen versenytási összehasonlítás a csődelőrejelzésben, JURA, 2019/1., 219-233.o. <https://szakcikkadatbazis.hu/doc/7903086>
33. Csikós-Nagy, B (2002): Közgazdaságtan a globalizáció világában I. MTA Társadalomkutató Központ, Budapest
34. Dancsik, B., Fábrián, G., Fellner, Z., Horváth, G., Lang, P., Nagy, G., Winkler, S. (2015). A nemteljesítő lakossági jelzáloghitel-portfólió átfogó elemzése mikroszintű adatok segítségével. MNB.

35. Deakin, E.B. (1972): A Discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 10-1 (Spring 1972) pp. 167-179
36. Decision Tree (2023) <https://www.geeksforgeeks.org/decision-tree/>
37. Dhamnetiya, D., Goel, M. K., Jha, R. P., Shalini, S., Bhattacharyya, K. (2022). How to Perform Discriminant Analysis in Medical Research? Explained with Illustrations. In *Journal of Laboratory Physicians* (Vol. 14, Issue 04, pp. 511–520). Georg Thieme Verlag KG. <https://doi.org/10.1055/s-0042-1747675>
38. Dobák, P., Sági, J. (2005). Fogyasztási hitelek: növekvő eladósodottság? Forrás: Hitelintézeti Szemle: http://www.bankszovetseg.hu/Content/Hitelintezeti/51Dobak_Sagi.pdf
39. Domán, Cs., Szilágyi, R., Varga B. (2009): Statisztikai elemzések alapjai II., Közgazdasági-módszertani képzés fejlesztéséért Alapítvány, Miskolc
40. Econventio Teszt. (2015). Forrás: <http://www.econventio.hu/#letolthetodokumentumok/econventio-teszt-2015-tanulmanyok>
41. Farkas, B. (2016): A várakozások szerepe a közgazdasági gondolkodásban Közgazdasági szemle, LXIII. évf, 2016. november p.1177-1191. http://real.mtak.hu/42261/1/02_Farkas_BeataA_u.pdf
42. Fawcett, T. (2004): ROC Graphs: Notes and Practical Considerations for Researchers. HP Laboratories, Technical Report, Palo Alto, CA USA. March 2004. http://home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf
43. Fawcett, T., Provost, F. (2013): Data Science for Business, O'Reilly Media, Inc., Sebastopol, Kalifornia
44. Fazekasné Kis, M. (2002): ROC analízis alkalmazása, *Acta Agraria Debreceniensis*, 2002.március 27.
45. Fekete, O., Tatay, T. (2012): Hitelezők és adósok kapcsolatának szabályozási kérdései https://kgk.sze.hu/images/dokumentumok/kautzkiadvany2012/penzugy/fekete_tatay.pdf
46. Felföldi-Szűcs, N.Á. (2011): Hitelezés vevői nemfizetés mellett (Ph.D. értekezés, BCE, 2011)
47. Ferenci, T. (2023): Kategoriális eredményváltozó modellezése: a logisztikus regresszió és változatai <https://www.medstat.hu/oktatas/RegressziosModellezes/FerenciTamas-RegressziosModellezes-KategorialisEredmenyValtozoALogisztikusRegresszio-slides.pdf>
48. Fish, K., Blodgett, J. (2003): A visual method for determining variable importance in an artificial neural network model: An empirical benchmark study. *J Target Meas Anal Mark* **11**, 244–254. <https://doi.org/10.1057/palgrave.jt.5740081>
49. Fitzpatrick, P. (1932): A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies. The Accountants Publishing Company, 1932.
50. Freund, R. J., Wilson, W. J. (1993): *Statistical methods*, Academic Press, Inc.; San Diego
51. Frydman, H., Altman, E. I., Kao, D. L. (1985): Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress, *The Journal of Finance*, Vol. 40. No. 1. 303-320. old.
52. GFLEC. (n.d.) S&P Global Finlit Survey. Global Financial Literacy Excellence Center, The George Washington University. <https://gflec.org///initiatives/sp-global-finlit-survey/>
53. Hajdu, O. (2003): Többváltozós statisztikai számítások, Központi Statisztikai Hivatal, Budapest

54. Hajdu, O. (2004): A csődesemény logit-regressziójának kismintás problémái, Statisztikai Szemle, 82. évf. 4. sz. 392-422. old.
55. Hajdú, O. (2018): Többváltozós statisztikai R Open alkalmazások (2018. szept 28.) Statisztikai szemle
56. Halas, G. (2004): Hogyan mérjük nemfizetési valószínűséget? Módszertani összehasonlítás. Mérés Határ Műhely, 2004. szeptember.
57. Hámori, G. (2001): A CHAID alapú döntési fák jellemzői, Statisztikai Szemle, 79. évf. 8. sz. 703-710. old.
http://www.ksh.hu/statszemle_archive/2001/2001_08/2001_08_703.pdf
58. Hámori, G. (2001): A fizetésképtelenség előrejelzése logit-moddal. Bankszemle, 45. évf, 1-2 szám, 65-87. old.
59. Hámori, G. (2014): Predikciós célú klasszifikáló statisztikai modellek gyakorlati kérdései, Ph.D értekezés
60. Hazel, K., Onaga, E. (2003): Experimental social innovation and dissemination: the promise and its delivery, American Journal of Community Psychology 32(3-4), pp. 285-294. <http://dx.doi.org/10.1023/b:ajcp.00000004748.50885.2e>.
61. Hornyák, A. (2015): Attitűdök és kompetenciák a középiskolás diákok, mint potenciális Banki ügyfelek körében, PhD értekezés, Nyugat-magyarországi Egyetem, Sopron http://doktori.nyme.hu/493/1/Hornyak_Andrea_Disszertacio.pdf
62. Horrigan, J. L. (1966): The determination of long term credit standing with financial ratios. Journal of Accounting Research, Vol.4. 1966.
63. Horváth, G. (2006): Neurális hálózatok, Panem Könyvkiadó, Budapest
64. How behavioral economics explains 6 common money mistakes <https://www.creditcards.com/credit-card-news/behavioral-economics-6-money-mistakes-1267.php>
65. Hosszú, Z. (2011). A lakosság fogyasztási viselkedése és annak jövedelem szerinti heterogenitása a válság előtt mikrostatistikák alapján. Forrás: MNB-Szemle: <https://www.mnb.hu/letoltes/hosszu.pdf>
66. IBM: Decision Trees, <https://www.ibm.com/topics/decision-trees>
67. IBM SPSS Neural Networks 26: https://www.ibm.com/docs/en/SSLVMB_26.0.0/pdf/en/IBM_SPSS_Neural_Network.pdf
68. IBM Pseudo R Square: <https://www.ibm.com/docs/en/spss-statistics/saas?topic=model-pseudo-r-square>
69. Jia, C., Jiajun, J., Yu-jane, L. (2018): Financial literacy and gender difference in loan performance, Journal of Empirical Finance, vol 48., 2018. szept., p307-320. <https://www.sciencedirect.com/science/article/abs/pii/S0927539818300422>
70. Weeks, J. (1998): A neoklasszikus közgazdaságtan kritikájára, Aula Kiadó Kft
71. Kahneman, D. (2013): Gyors és lassú gondolkodás, HVG Kiadó, Budapest
72. Karajz, S.: Közgazdasági elméletek előadás anyaga
73. Keasey, K., McGuinness, P. (1990): The failure of UK industrial firms for the period 1976-1984: Logistic analysis and entropy measures. Journal of Business Finance and Accounting, vol. 17. no. 1. pp. 119-135.
74. Ketskemény, L., Izsó, L., Könyves Tóth, E. (2011): Bevezetés az IBM SPSS Statistics programrendszerbe, Artéria Stúdió Kft, Budapest

75. KHR éves tájékoztatók. (2011-2022). Forrás: <https://www.bisz.hu/dokumentumtar>
76. Kiss, F. (2003): A credit scoring fejlődése és alkalmazása. PhD. doktori értekezés, Budapesti Műszaki Egyetem, 2003.
77. Kiviluoto, K. (1998): Predicting Bankruptcies with the selforganizing map. *Neurocomputing* 21 (1998) pp. 191-201.
78. Kosztópulosz, A.: A pénzügyi válság és következményei: monetáris politikai és szabályozási kihívások
79. Kovács, E. (2009): Pénzügyi adatok statisztikai elemzése, Tanszék Kft., Budapest
80. Kovács, L. (2015). A pénzügyi kultúra kutatása és aktuális feladataink. Forrás: <http://bankszovetseg.hu/Public/gep/2015/Kovacs-Levente-79-88.pdf>
81. Kovács, L. (2017). A pénzügyi kultúra fejlesztése, mint önérdék. In: Péter, Zs. (szerk.) 2017: A pénzügyi kultúra aktuális kérdései, különös tekintettel a banki szolgáltatásokra. Miskolc: A közgazdaságtani-módszertani képzés fejlesztéséért Alapítvány.
82. Kovács, L., Pál, Zs. (2017): A pénzügyi kultúra vizsgálatának sarokkövei, In: Veresné, Somosi Mariann; Lipták, Katalin (szerk.) „Mérleg és Kihívások” X. Nemzetközi Tudományos Konferencia = „Balance and Challenges” X. International Scientific Conference : Konferenciakiadvány: A közgazdászok képzés elindításának 30. évfordulója alkalmából, p. 344-353.
83. Kovács, L., Terták, E. (2019): *Financial literacy - Theory and Evidence*, Verlag Dashöfer, Bratislava
84. Kovács, P. (2008): A multikollinearitás vizsgálata lineáris regressziós modellekben, *Statisztikai Szemle*, 86 évf. 1. szám, 38-67pp
85. Kovács, P., Révész, B., Ország, G. (2014). A pénzügyi kultúra és attitűd mérése. Forrás: www.eco.u-szeged.hu/download.php?docID=40014
86. Kovács, P. (2017): Fiatalok pénzügyi kultúra mérésének fejlesztésének lehetőségei (Miskolc, 2017, A pénzügyi kultúra aktuális kérdései, különös tekintettel a banki szolgáltatásokra) 31-37
87. Kovács, P., Kuruczleki, É. (2017). A magyar lakosság pénzügyi kultúrája (Financial literacy of hungarian population). In V. D. Mariann (Ed.), "Mérleg és Kihívások" X. Nemzetközi Tudományos Konferencia Miskolc-Lillafüred.. pp. 336-343.. <http://gtk.uni-miskolc.hu/files/12432/Konferencia%20k%C3%B6tet%202017%20b.pdf>
88. Koltay G., Vincze J.: (2009): Fogyasztói döntések a viselkedési közgazdaságtan szemszögéből (*Közgazdasági Szemle*, LVI. évf., 2009. június (495-525. o.))
89. Klapper, L., Lusardi A., Oudheusden P. (2015) *Financial Literacy Around the World: Insights from the Standard & Poor's Rating Services*, S&P
90. Kristóf, T. (2002): A mesterséges neurális hálók a jövőkutatás szolgálatában. *Jövöelméletek*, 9. BKÁE Jövőkutatási Kutatóközpont, Budapest.
91. Kristóf, T. (2004): Mesterséges intelligencia a csődelőrejelzésben. *Jövőtanulmányok*, 21. BKÁE Jövőkutatási Kutatóközpont, Budapest.
92. Kristóf, T. (2008): A csődelőrejelzés és a nem fizetési valószínűség számításának módszertani kérdéseiről, *Közgazdasági Szemle*, LV. évf. 2008. május., p441-461.)
93. KSH táblázatok:
 - 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 [%]
94. Laitinen, T., Kankaanpää, M. (1999): Comparative analysis of failure prediction methods: the Finnish case. *The European Accounting Review* 8:1 (1999) pp. 67-92.
 95. Lautenschläger, S. (2015). Reintegrating the banking sector into society: earning and reestablishing trust. Forrás: ECB: <https://www.ecb.europa.eu/press/key/date/2015/html/sp150928.en.html>
 96. Leshno, M., Spector, Y. (1996): Neural network prediction analysis: The bankruptcy case. *Neurocomputing*, Vol. 10. March 1996, pp. 125-147.
 97. Libby, R. (1975): Accounting ratios and the prediction of failure: Some behavioral evidence. *Journal of Accounting Research*, Spring 1975, pp. 150-161.
 98. Lin, F.Y., McClean, S. (2001): A Data Mining Approach to the Prediction of Corporate Failure. *Knowledge-Based Systems*, vol. 14, no. 3, pp. 189-195.
 99. Lusardi, A., Klapper, L., Oudheusden, P. (2015). Financial literacy around the World (S&P). Forrás: http://gflec.org/wp-content/uploads/2015/11/Finlit_paper_16_F2_singles.pdf
 100. Malhotra, N. K. (2008): *Marketingkutató* (Akadémiai Kiadó, Budapest)
 101. Martin, D. (1977): Early-warning of Bank Failure: A Logit Regression Approach. *Journal of Banking and Finance*, 1977, no.1, pp. 249-276.
 102. McKee, T.E., Greenstein M. (2000): Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 2000, no.19. pp. 219-230.
 103. Mérő, K. (2003). A gazdasági növekedés és a pénzügyi közvetítés mélysége. *Közgazdasági Szemle*, 50.(július-augusztus), 590–607. oldal
 104. Mészáros, A. (2017): pénzügyi attitűdök és pénzügyi viselkedés. In: Péter, Zs. (szerk.) 2017: A pénzügyi kultúra aktuális kérdései, különös tekintettel a banki szolgáltatásokra. Miskolc: A közgazdaságtani-módszertani képzés fejlesztéséért Alapítvány.
 105. Meyer, S (2010): Insights from Behavioral Economics for Personal Finance https://www.newyorkfed.org/medialibrary/media/banking/MeierSlides_NYU.pdf
 106. Mill, J. S. (1875): *A nemzetgazdaságtan alapelvei*, Budapest
 107. Mindy, F. J. (2002): *On Becoming a Consumer: Development of Consumer Behavior Patterns in Childhood*. Elsevier: New York
 108. Mit jelent a pénzügyi kultúra? Miért van rá szükséged? (2014). Forrás: <http://www.penzvilaga.hu/mit-jelent-a-penzugyi-kultura/>
 109. MNB. (dátum nélk.). Tájékoztató a fizetési késedelemmel és a hitelszerződések felmondásával kapcsolatban. Forrás: MNB: https://www.mnb.hu/archivum/Felugyelet/root/fooldal/fogyasztoknak/fizetesi_nehezsegek/ha_mar_baj_van/fizetesi_kesedelem_120312
 110. MNB: Hitel- és lízingtermék választó https://alk.mnb.hu/fogyasztoknak/bal_menu/ptilekerdezo/12/info.html
 111. MNB. (2006). A lakosság pénzügyi kultúrájának felmérése. Forrás: MNB: <https://www.mnb.hu/letoltes/mnb-penzugyi-kultura-kvalitativ-1.pdf>
 112. MNB (2014): Pénzügyi stabilitási jelentés

113. MNB. (2015) A nemteljesítő lakossági jelzálog-portfólió átfogó elemzése mikroszintű adatok segítségével
114. MNB. (2018): Tájékoztató a túlzott eladósodottság kockázatairól
115. MNB (2021): Hitelezési folyamatok 2021. március
116. MNB (2023a): Hitelezési folyamatok 2023. március
117. MNB (2023b): Tovább erősíti a hitelpiac egészséges hosszú távú működését az MNB
<https://www.mnb.hu/sajtoszoba/sajtokozlemenyek/2023-evi-sajtokozlemenyek/tovabb-erositi-a-hitelpiac-egeszseges-hosszu-tavu-mukodeset-az-mnb>
118. MNB-PSZÁF. (2008). Együttműködési megállapodás a pénzügyi kultúra fejlesztése területén. Forrás: MNB: <https://www.mnb.hu/letoltes/0415mnbpszafmegallpodas-penzugyi-kultura-fejleszte.pdf>
119. Nagy P., Tóth Zs. (2012): Értelmelem és érzelem. A lakossági ügyfelek gazdasági magatartása és a bankokkal szembeni attitűdjei. Hitelintézeti Szemle, 2012. évi 11. Különszám 13-24. p.
<http://unipub.lib.uni-corvinus.hu/1039/1/13-24-ig-nagy-toth.pdf>
120. Neter, J. (1966): Discussion of financial ratios as predictors of failure. *Journal of Accounting Research*, Vol.4. 1966, pp. 112-118.
121. Neulinger, Á. (2015). Megtakarítás és hitelfizetés a hazai családi életciklusokban. Forrás: <https://emok.hu>
122. Newsom, J. T. (2008): *Categorical Data Analysis*
<https://web.pdx.edu/~newsomj/cdaclass/>
123. Odom, M.D., Sharda, R. (1990): A Neural Network Model for Bankruptcy Prediction. *IJCNN International Joint Conference on Neural Networks*, Vol. 2, San Diego, CA, 1990. pp. 163-1
124. OECD (2016), “OECD/INFE International Survey of Adult Financial Literacy Competencies”, OECD, Paris, www.oecd.org/finance/OECD-INFE-International-Survey-of-Adult-Financial-Literacy-Competencies.pdf
125. OECD (2017), G20/OECD INFE report on adult financial literacy in G20 countries, <https://www.oecd.org/daf/fin/financial-education/G20-OECD-INFE-report-adult-financial-literacy-in-G20-countries.pdf>
126. OECD (2020): OECD/INFE 2020 International Survey of Adult Financial Literacy, www.oecd.org/finance/education/launchoftheoecdinfe-global-financial-literacy-survey-report.htm
127. OECD (2021): G20/OECD-INFE Report on supporting financial resilience and transformation through digital financial literacy, www.oecd.org/finance/supporting-financial-resilience-and-transformation-through-digital-financial-literacy.htm
128. OECD (2021b): OECD Economic Surveys, Hungary
<https://www.oecd.org/economy/surveys/Hungary-2021-OECD-economic-survey-overview.pdf>
129. Ohlson, J. (1980): Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, Vol. 18. No. 1. 109-131. old.
130. Ojo Olawale: The CAP curves (<https://waleblaq.medium.com/the-cap-curves-the-cumulative-accuracy-profile-58a141e01fae>) (2020.07.06.)

131. Olmeda, I., Fernandez, E. (1997): Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction. *Computational Economics*, Vol. 10. Issue 4. pp. 317–352.
132. Oravecz, B. (2007): Credit scoring modellek és teljesítményük értékelése, *Hitelintézeti szemle*, 6. évf. 6. szám, 607-627. old
133. Oravecz, B. (2008) Szelekciós torzítás és csökkentése az adóminősítési modelleknél, Ph.D. értekezés
134. Pál, Zs. (2017) A pénzügyi kultúra aktuális kérdései, különös tekintettel a banki szolgáltatásokra (Tanulmánykötet) <http://www.penz7.hu/uploads/images/ckfinder/files/tanulmanykotetfinal.pdf>
135. Pantalone, C. C., Platt, M.B. (1987): Predicting Commercial Bank Failure Since Deregulation. *New England Economic Review*, 1987, Jul-Aug, pp. 37-47.
136. Pénziránytű Alapítvány. (2015). A Pénziránytű Alapítvány által lebonyolított pénzügyi kultúra kutatás fő eredményeinek bemutatása. Forrás: MNB: <https://www.mnb.hu/letoltes/a-penziranytu-alapitvany-altal-lebonyolított-penzugyi-kultura-kutatas-fo-eredmenyeinek-bemutatasa.pdf>
137. Pizarroso, J., Portela, J., & Muñoz, A. (2022). NeuralSens: Sensitivity Analysis of Neural Networks. *Journal of Statistical Software*, 102(7), 1–36. <https://doi.org/10.18637/jss.v102.i07>
138. Pompe, P.P.M., Bilderbeek, J. (2000): Prestaties van modellen en ratio's bij het voorspellen van faillissementen. *Maandblad voor Accountancy en Bedrijfseconomie*, Oktober 2000. pp. 465-472.
139. Potóczki, J. (2017). A magyar lakosság pénzügyi kultúrájának szintje az öngondoskodás tükrében - nemzetközi és hazai kutatási eredmények. Forrás: JATEPress, Szeged: http://acta.bibl.u-szeged.hu/49707/1/gtk_2017_157-170.pdf
140. Rabin M. (2008): *Pszichológia és közgazdaságtan*, Alinea Kiadó, Budapest
141. Raiffeisen Bank (2017): *Hitelkiskosok*
142. Raiffeisen Bank: *Banki alapismeretek*
143. Ramster, J., Foster, L. (1931): *A Demonstration of Ratio Analysis*
144. Remund, D. L. (2010). *Financial Literacy Explicated: The Case for a Clearer Definition in an Increasingly Complex Economy*. Forrás: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1745-6606.2010.01169.x>
145. Roedder John, R. D. (1999): Consumer Socialization of Children: A Retrospective Look at Twenty-Five Years of Research. *Journal of Consumer Research*, vol. 26, 183–213.
146. Saini, A. (2021): *Decision Tree Algorithm – A Complete Guide*, <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>
147. Sáfrány, A. (2012): *A lakossági energiahatékonyság növelésének lehetőségei a viselkedési közgazdaságtan eredményeinek tükrében (szakdolgozat, BCE)*
148. Sajtos L., Mitev A. (2007): *SPSS kutatási és adatelemzési kézikönyv* (Alinea Kiadó, Budapest)
149. Stergiou, C., Siganos, D. (1996): *Neural Networks* <https://srii.sou.edu.ge/neural-networks.pdf>
150. Sima J. Gandhi (2017): *Understanding students from a behavioral economics perspective: how accelerating student loan subsidies generates more bang for the buck* https://law.ku.edu/sites/law.drupal.ku.edu/files/docs/law_journal/v17/gandhi.pdf

151. Sumit, A., Souphala, C., Yunqi, Z. (2015): How does financial literacy affect mortgage default?
152. Süveges, G. (2018): Fókuszban a Pénzügyi Kultúra Miskolcon, Számvitel, Adó, Könyvvizsgálat: Szakma 60., 45-47
153. Szabó, D. M. (2021): A lakossági hitelezés fejlődés és a pénzügyi fogyasztó védelem
154. Székelyi, M., Barna, I. (2002): Túlélőkészlet az SPSS-hez, Typotex Kiadó
155. Szemán, J., Süveges, G. (2017): A pénzügyi kultúra jellemzői a gazdasági felsőoktatásban tanuló fiatalok esetén, X. Számvitel Tudomány-Szakma-Oktatás Nemzetközi Konferencia (Papp László Számviteli Konferencia) „A Számviteli Rendszerben És Környezetében Bekövetkezett Változások – Múlt, Jelen, Jövő”, Pécs
156. Szepesi, Gy. (2013): Géniuszok párharca (Közgazdasági szemle, LX. évf, 2013. június (633-649. o)
http://epa.oszk.hu/00000/00017/00204/pdf/EPA00017_kozgazdasagi_szemle_2013_06_633-649.pdf
157. Tam, K.Y., Kiang, M.Y. (1992): Managerial Applications of Neural Networks - The Case of Bank Failure Predictions. Management Science, Vol. 38, No. 7. pp. 926-947.
158. Tan, C.N.W. (1999): An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System. Bond University, Australia, 1999.
159. Tasi, T. (2021): Adóspárti a magyar hitelmoratórium Bonum publicum, p30-31, Planet Budapest
160. Thaler, R. H. (2009): Nudge- Improving Decisions about health, wealth, and happiness
161. Thaler, R. H. (2016): Rendbontók – A viselkedési közgazdaságtan térnyerése, HVG Kiadó, Budapest
162. The World Bank <https://www.worldbank.org/>
163. Trading Economics. (dátum nélkül.). Forrás: <https://tradingeconomics.com/>
164. Tooth R. (2015): Behavioural economics and the regulation of consumer credit
<http://www.srgexpert.com/wp-content/uploads/2015/08/Consumer-Credit-Behavioural-Economics-Case-Study-2012-Final.pdf>
165. Tóth Sz. (2015): Bolti akciók és ezek eladásra gyakorolt hatásai a viselkedési közgazdaságtan szemszögéből (szakdolgozat, BCE)
166. Varga B., Szilágyi R. (2011): Kvantitatív információképzési technikák
https://www.tankonyvtar.hu/hu/tartalom/tamop425/0049_08_kvantitativ_informaciokepzesi_technikak/3268/index.html
167. Vértesy, L. (2008): A pénzügyi intézmények finanszírozási tevékenységének jogi szabályozása Magyarországon Ph. D értekezés
168. Virág, M., Kristóf, T., Fiáth, A., Varsányi, J. (2013): Pénzügyi elemzés, csődelőrejelzés, válságkezelés, Kossuth Kiadó, Budapest
169. Virág, M., Hajdu, O. (1996): Pénzügyi mutatószámokon alapuló csődmodell-számítások, Bankszemle, 15. évf. 5. sz. 42-53. old.
170. Virág, M. (2004): A csődmodellek jellegzetességei és története, Vezetéstudomány, 35. évf. 10. sz. 24-32. old.
171. Virág, M., Kristóf, T. (2005): Az első hazai csődmodell újraszámítása neurális hálók segítségével. Közgazdasági Szemle, 52. évf. 2. sz. 144-162. old.

172. Virág, M., Kristóf, T. (2006): Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel, *Vezetéstudomány*, 37. évf. 1. sz. 25-35. old.
173. Washington Post <https://www.washingtonpost.com>
174. What is Classical Theory of Interest? <http://www.preservearticles.com/2012020222452/what-is-classical-theory-of-interest.html>
175. Wilcox, J. W. (1971). A simple theory of financial ratios as predictors of failure. *Journal of Accounting Research*, Autumn 1971, pp. 389-395.
176. Wilcox, J. W. (1973): A prediction of business failure using accounting data. *Journal of Accounting Research*, Spring 1973, pp. 163-179.
177. Wilson, R. L., Sharda, R. (1994): Bankruptcy Prediction Using Neural Networks. *Decision Support Systems*, vol. 11, no. 5, pp. 545-557.
178. Winakor, A., Smith, R. (1935): Changes in the Financial Structure of Unsuccessful Industrial Corporations. Bureau of Business Research, University of Illinois, Urbana, IL, Bulletin 51.
179. Worldbank (2021): The Global Findex Database
180. World Health Organization <https://www.who.int/>
181. Zavgren, C. (1985): Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Business Finance and Accounting*, vol.12, no. 1, pp. 19-45.
182. Zhang, G., Hu, M., Patuwo, B. (1999): Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, Vol. 116. pp. 16–32.
183. Zmijewski, M.E. (1984): Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 1984, no. 22, pp. 59-82.
184. Zsótér et al (2017): The Impact of Changes in the Socio-Economic Environment on Financial Literacy. Forrás: Pénziránytű Alapítvány https://www.penziranytu.hu/sites/default/files/csatoImany/zsoter_2017_2_a.pdf
185. 2011. évi CXXII. törvény
186. 32/2014 MNB rendelet

11 Mellékletek

11.1 Logisztikus regresszió I. outputjai

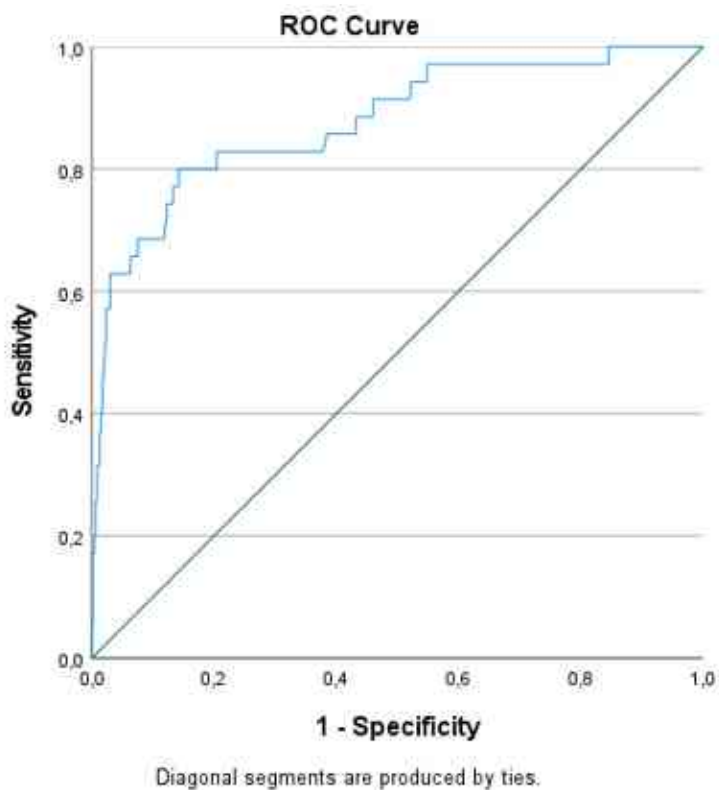
Omnibus Tests of Model Coefficients					
Minta típusa			Chi-square	df	Sig.
teszt	Step 1	Step	16,744	1	<,001
		Block	16,744	1	<,001
		Model	16,744	1	<,001
tanuló	Step 1	Step	62,222	1	<,001
		Block	62,222	1	<,001
		Model	62,222	1	<,001

Model Summary				
Minta típusa	Step	-2 Log likelihood	Cox & Snell R ²	Nagelkerke R ²
teszt	1	45,721 ^a	,106	,310
tanuló	1	127,992 ^b	,163	,388

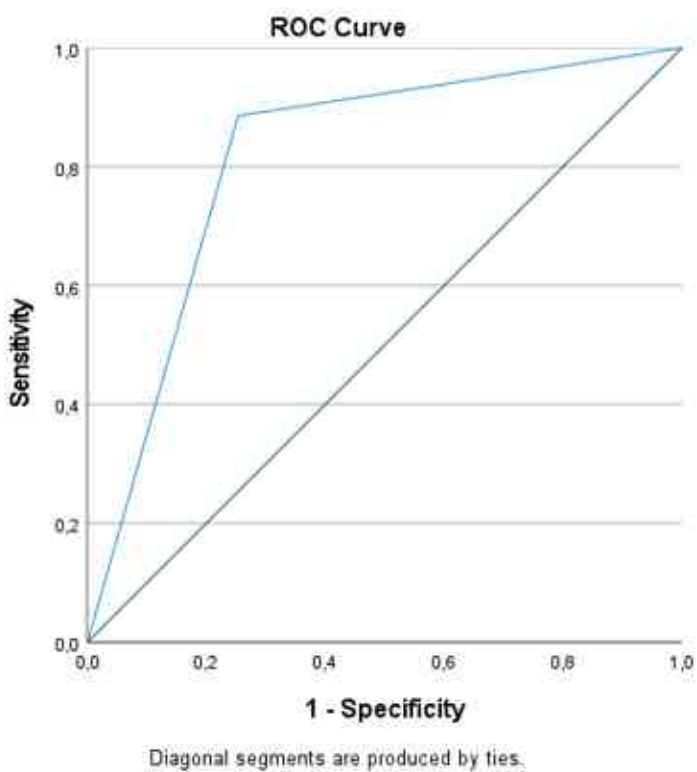
Hosmer and Lemeshow Test				
Minta típusa	Step	Chi-square	df	Sig.
teszt	1	17,357	8	,027
tanuló	1	10,815	8	,212

Classification Table (alapértelmezett)					
Minta típusa	Megfigyelt	Becsült			
		Default		Percentage Correct	
		0	1		
Teszt	Default	0	141	1	99,3
		1	5	3	37,5
	Overall Percentage				
Tanuló	Default	0	317	6	98,1
		1	16	11	40,7
	Overall Percentage				

a. The cut value is ,500



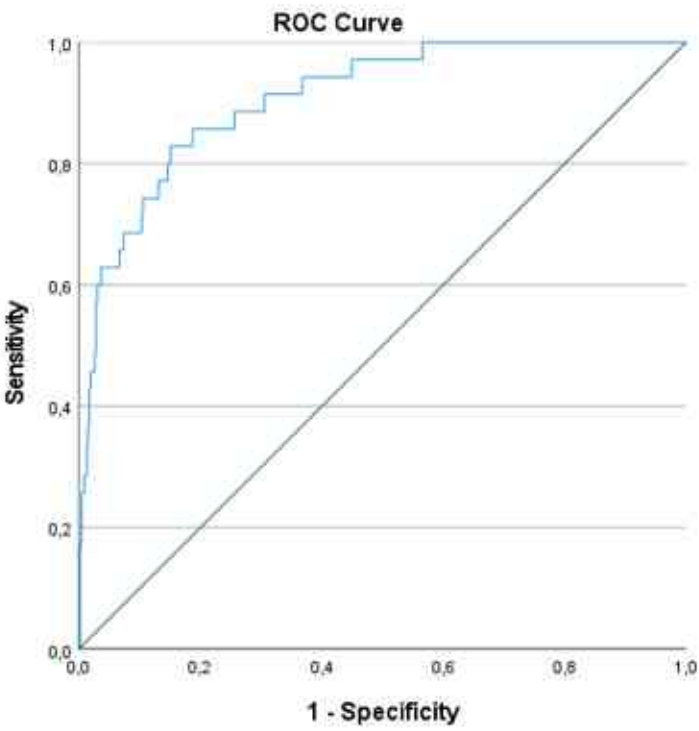
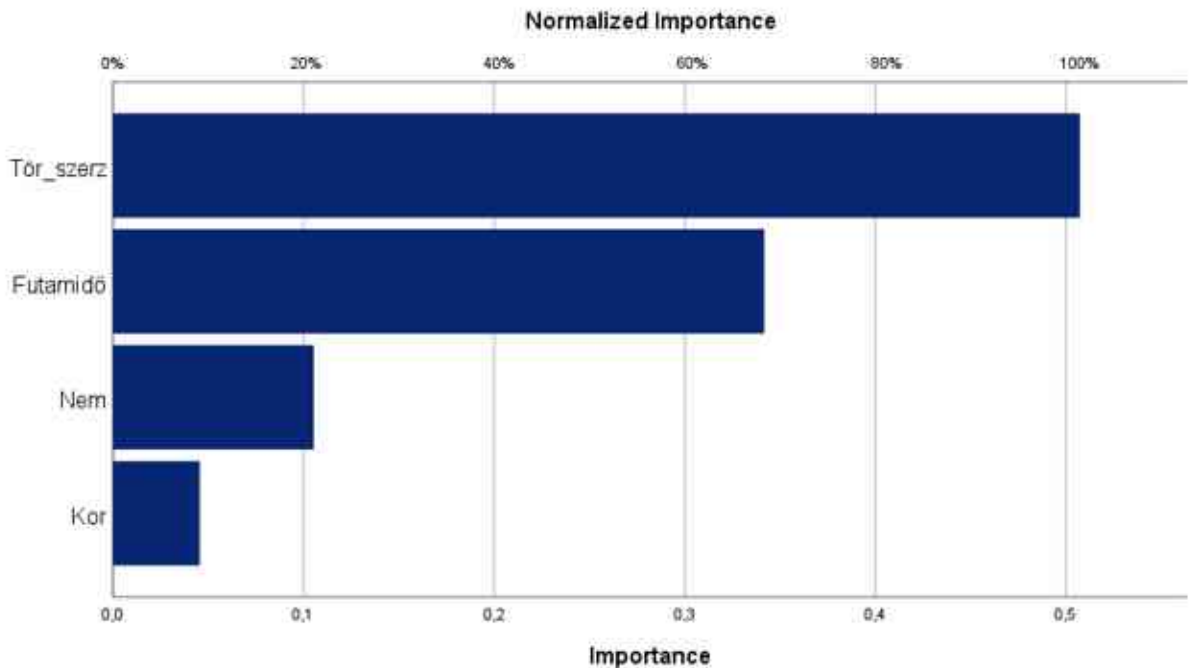
11.2 Döntési fa I. outputjai



11.3 Neurális háló I. outputjai

Network Information			
Input Layer	Covariates	1	Kor
		2	Nem
		3	tör_szerz*100
		4	Futamidő
	Number of Units ^a		4
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

Model Summary		
Tanuló	Cross Entropy Error	59,461
	Percent Incorrect Predictions	6,7%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,07
Teszt	Cross Entropy Error	19,756
	Percent Incorrect Predictions	3,2%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



11.4 Logisztikus regresszió II. outputjai

Omnibus Tests of Model Coefficients					
Minta típusa			Chi-square	df	Sig.
Teszt	Step 1	Step	87,629	2	<,001
		Block	87,629	2	<,001
		Model	87,629	2	<,001
Tanuló	Step 1	Step	247,196	2	<,001
		Block	247,196	2	<,001
		Model	247,196	2	<,001

Model Summary				
Minta típusa	Step	-2 Log likelihood	Cox & Snell R ²	Nagelkerke R ²
Teszt	1	137,594 ^a	,416	,555
Tanuló	1	219,627 ^b	,520	,693

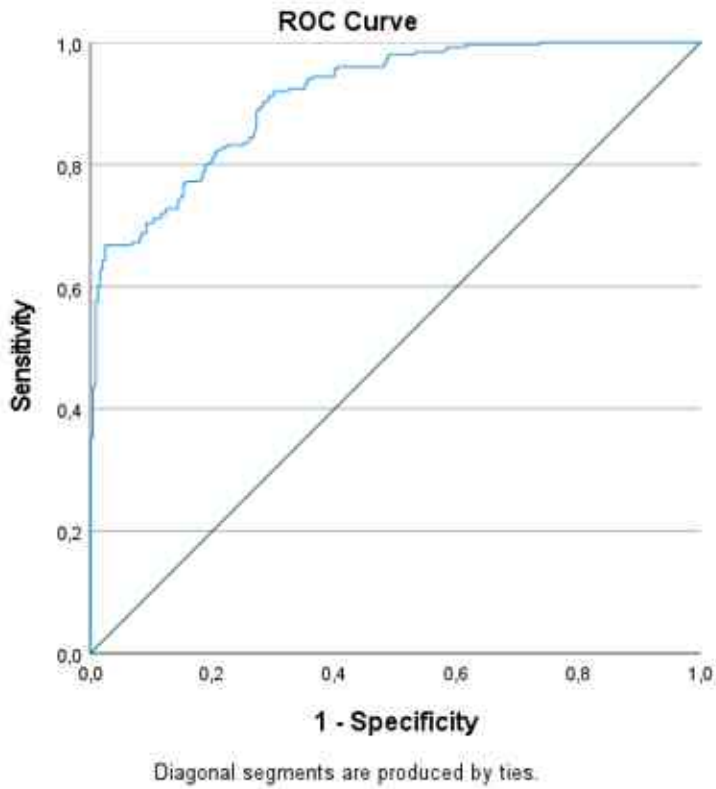
Hosmer and Lemeshow Test				
Minta típusa	Step	Chi-square	df	Sig.
Teszt	1	12,608	8	,126
Tanuló	1	13,214	8	,105

Classification Table (alapértelmezett)					
Minta típusa	Megfigyelt	Becsült			
		Default		Percentage Correct	
		0	1		
Teszt	Default	0	79	8	90,8
		1	33	43	56,6
	Overall Percentage				
Tanuló	Default	0	155	8	95,1
		1	50	124	71,3
	Overall Percentage				

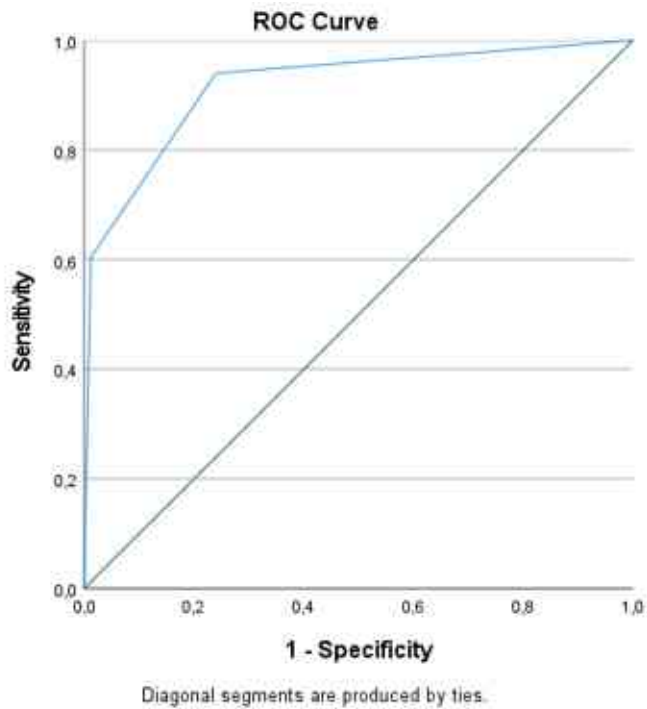
a. The cut value is ,500

Classification Table (Youden szabály szerinti optimális)					
Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	85	2	97,7
		1	33	43	56,6
	Overall Percentage				78,5
Tanuló	Default	0	159	4	97,5
		1	50	124	71,3
	Overall Percentage				84,0

a. The cut value is ,630



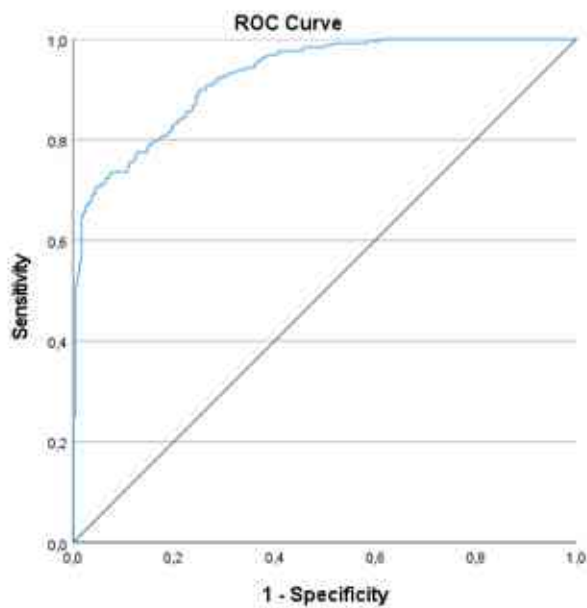
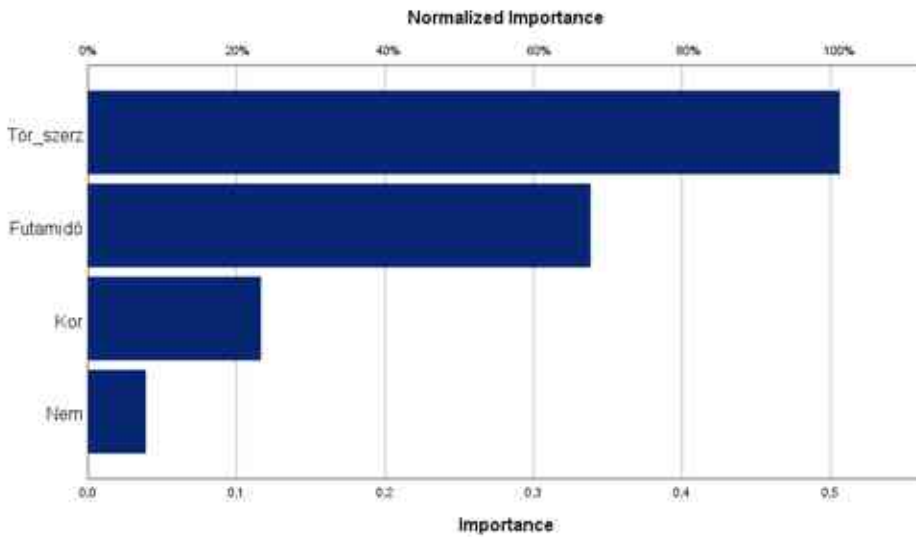
11.5 Döntési fa II. outputjai



11.6 Neurális háló II. outputjai

Network Information				
Input Layer	Covariates	1	tör_szerz*100	
		2	Felv kor	
		3	Nem	
		4	Futamidő	
	Number of Units ^a		4	
	Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1	
	Number of Units in Hidden Layer 1 ^a		5	
	Activation Function		Hyperbolic tangent	
Output Layer	Dependent Variables	1	Default	
	Number of Units		2	
	Activation Function		Softmax	
	Error Function		Cross-entropy	
a. Excluding the bias unit				

Model Summary		
Tanuló	Cross Entropy Error	129,453
	Percent Incorrect Predictions	18,5%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,07
Teszt	Cross Entropy Error	41,612
	Percent Incorrect Predictions	13,8%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



11.7 Logisztikus regresszió III. outputjai

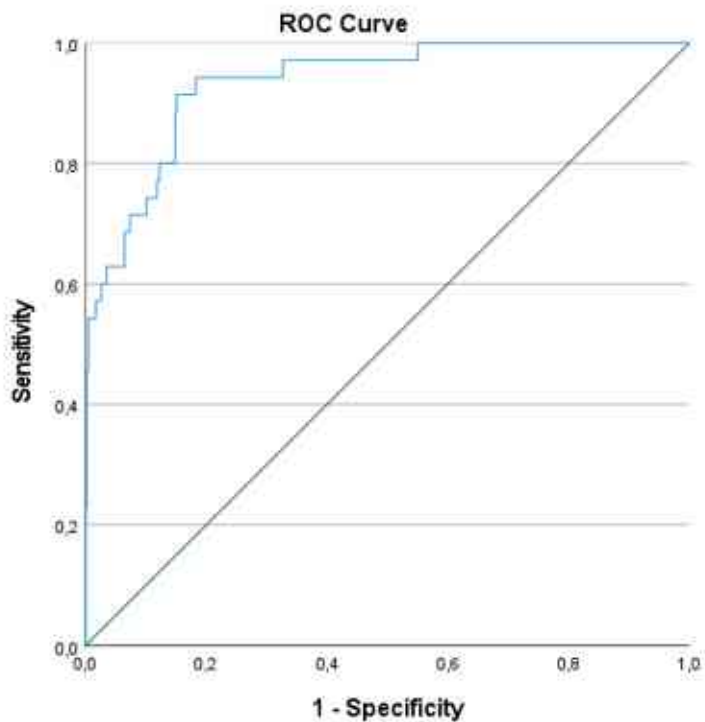
Omnibus Tests of Model Coefficients					
Minta típusa			Chi-square	df	Sig.
Teszt	Step 1	Step	29,235	3	<,001
		Block	29,235	3	<,001
		Model	29,235	3	<,001
Tanuló	Step 1	Step	90,165	3	<,001
		Block	90,165	3	<,001
		Model	90,165	3	<,001

Model Summary				
Minta típusa	Step	-2 Log likelihood	Cox & Snell R ²	Nagelkerke R ²
Teszt	1	33,230 ^a	,177	,520
Tanuló	1	100,049 ^b	,227	,542

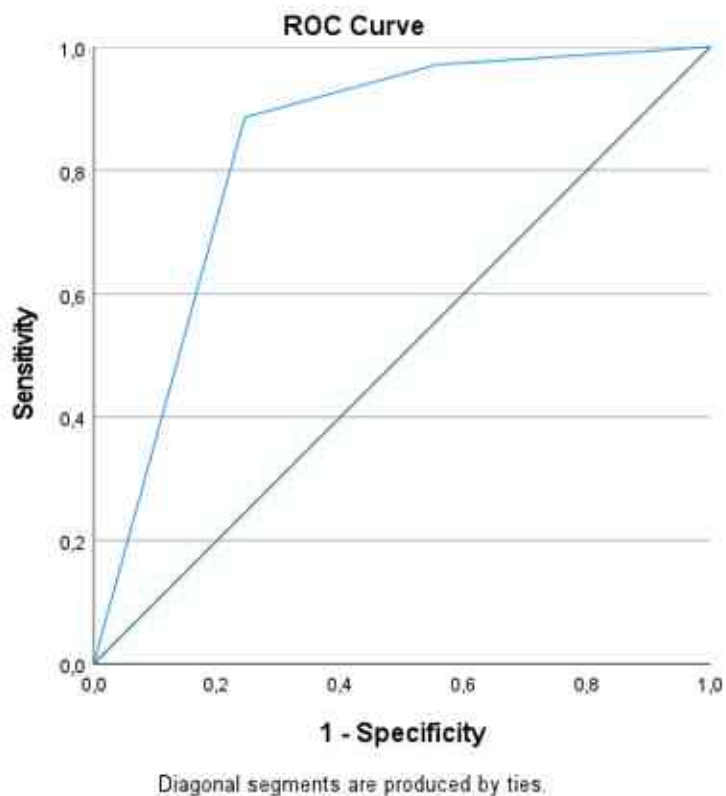
Hosmer and Lemeshow Test				
Minta típusa	Step	Chi-square	df	Sig.
Teszt	1	12,617	8	,126
Tanuló	1	4,019	8	,855

Classification Table (alapértelmezett)					
Minta típusa	Megfigyelt	Becsült			
		Default		Percentage Correct	
		0	1		
Teszt	Default	0	140	2	98,6
		1	4	4	50,0
	Overall Percentage				96,0
Tanuló	Default	0	319	4	98,8
		1	12	15	55,6
	Overall Percentage				95,4

a. The cut value is ,500



11.8 Döntési fa III. outputjai

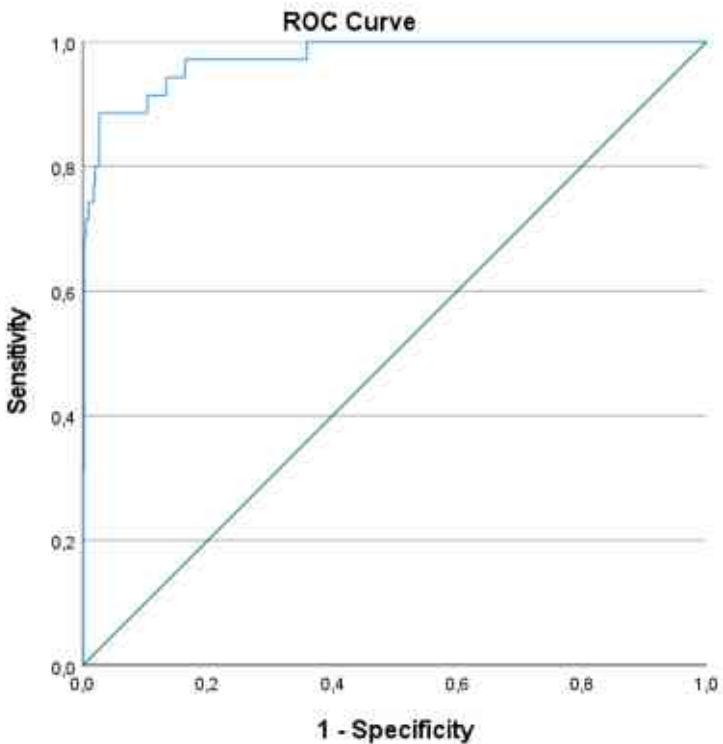
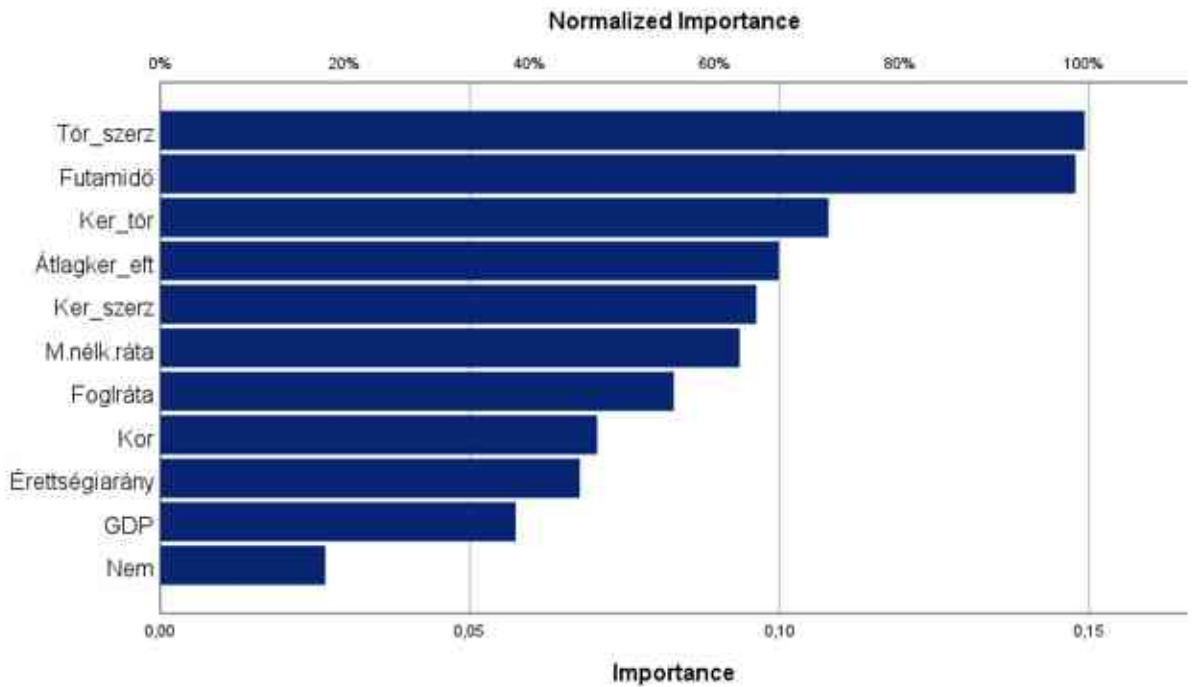


11.9 Neurális háló III. outputjai

Network Information			
Input Layer	Covariates	1	Kor
		2	Nem
		3	tör_szerz*100
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1 főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker_tör*100
		10	Futamidő
		11	Átlagker eft
	Number of Units ^a		11
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

Model Summary		
Tanuló	Cross Entropy Error	24,781
	Percent Incorrect Predictions	2,3%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,13
Teszt	Cross Entropy Error	16,764
	Percent Incorrect Predictions	2,6%

Dependent Variable: Default
 a. Error computations are based on the testing sample.



11.10 Logisztikus regresszió IV. outputjai

Omnibus Tests of Model Coefficients					
Minta típusa			Chi-square	df	Sig.
Teszt	Step 1	Step	108,466	4	<,001
		Block	108,466	4	<,001
		Model	108,466	4	<,001
Tanuló	Step 1	Step	272,543	4	<,001
		Block	272,543	4	<,001
		Model	272,543	4	<,001

Model Summary				
Minta típusa	Step	-2 Log likelihood	Cox & Snell R ²	Nagelkerke R ²
Teszt	1	116,757 ^a	,486	,649
Tanuló	1	194,279 ^b	,555	,740

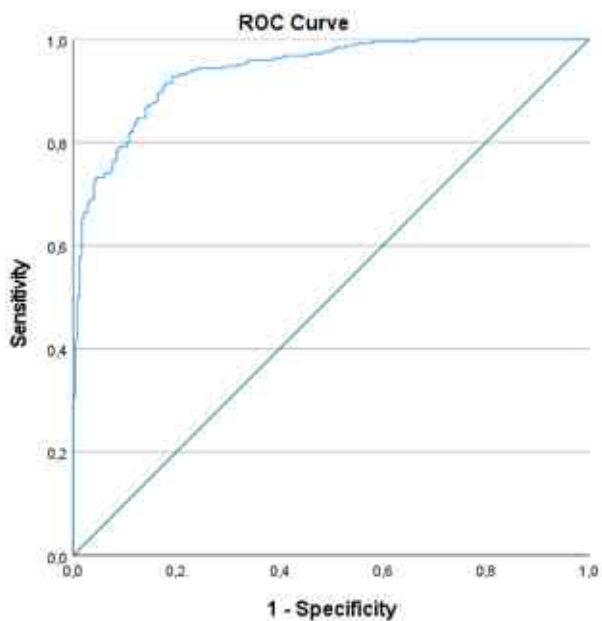
Hosmer and Lemeshow Test				
Minta típusa	Step	Chi-square	df	Sig.
Teszt	1	8,231	8	,411
Tanuló	1	10,725	8	,218

Classification Table (alapértelmezett)					
Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Teszt	Default	0	76	11	87,4
		1	14	62	81,6
	Overall Percentage				84,7
Tanuló	Default	0	147	16	90,2
		1	34	140	80,5
	Overall Percentage				85,2

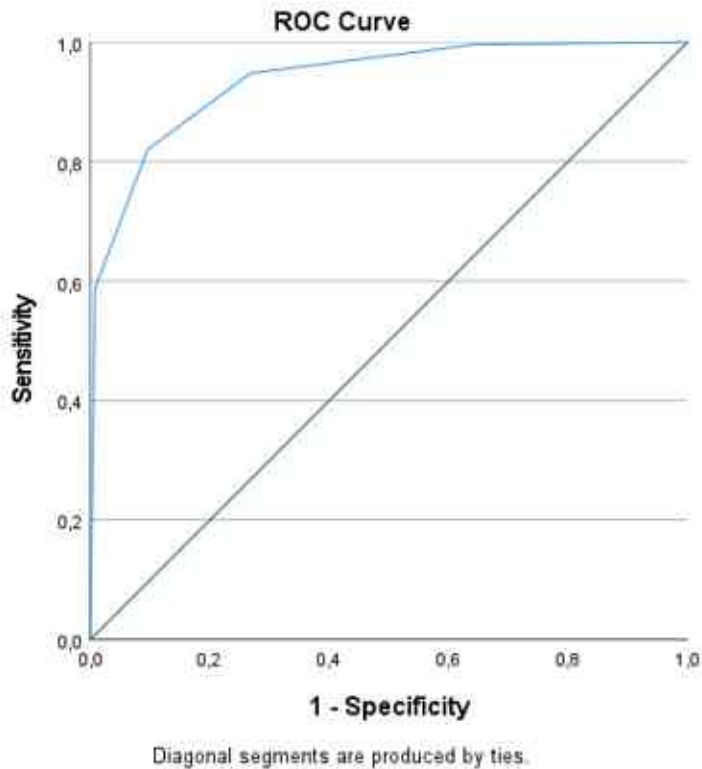
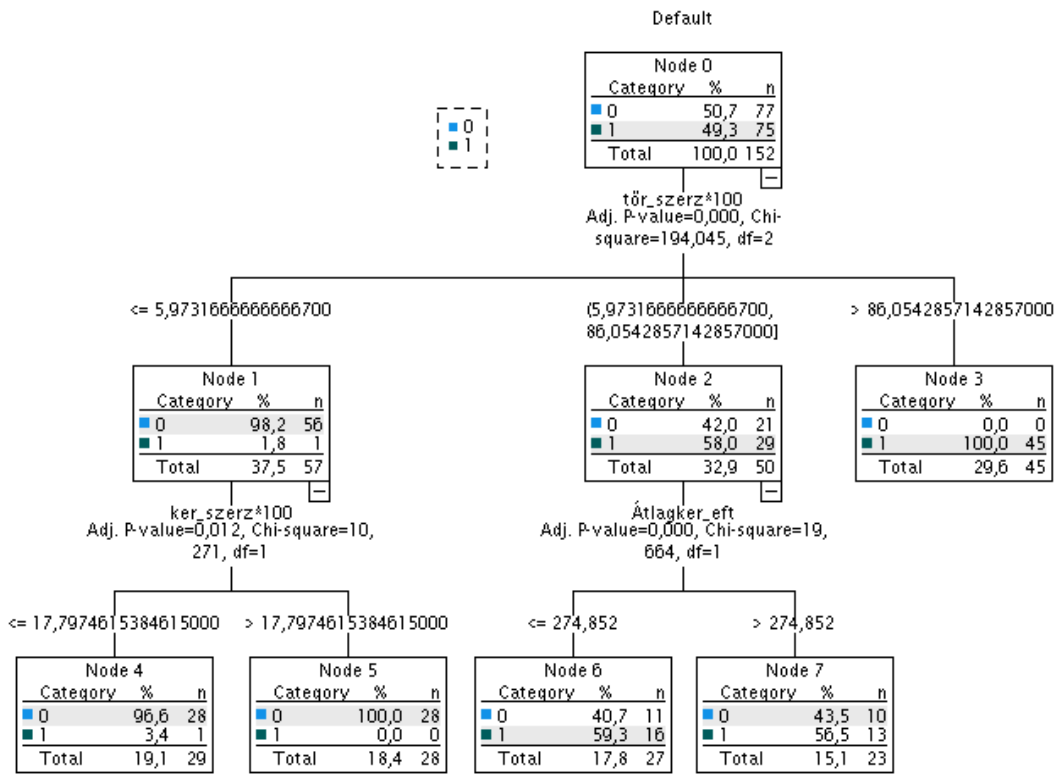
a. The cut value is ,500

Classification Table (Bal felső sarok szerinti optimális)					
Minta típusa	Megfigyelt		Becsült		
			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



11.11 Döntési fa IV. outputjai



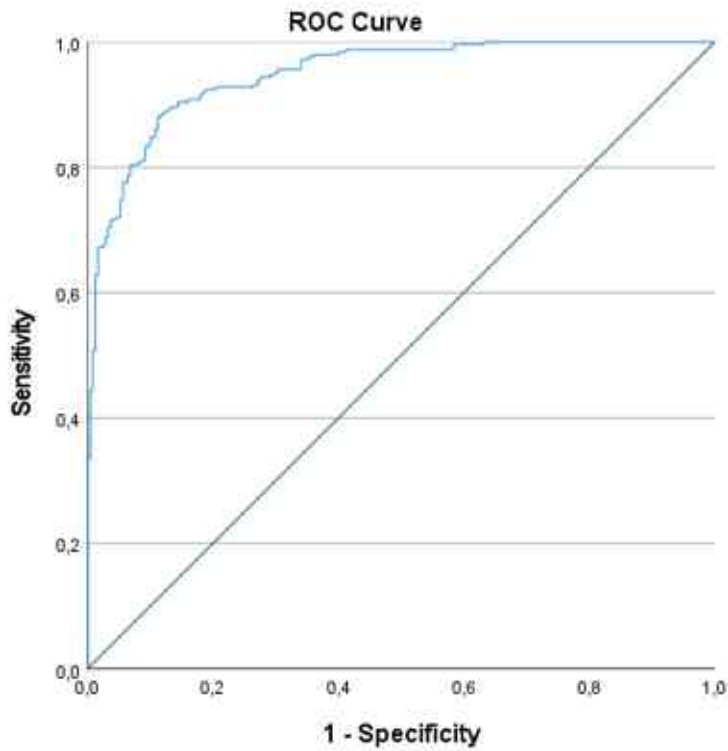
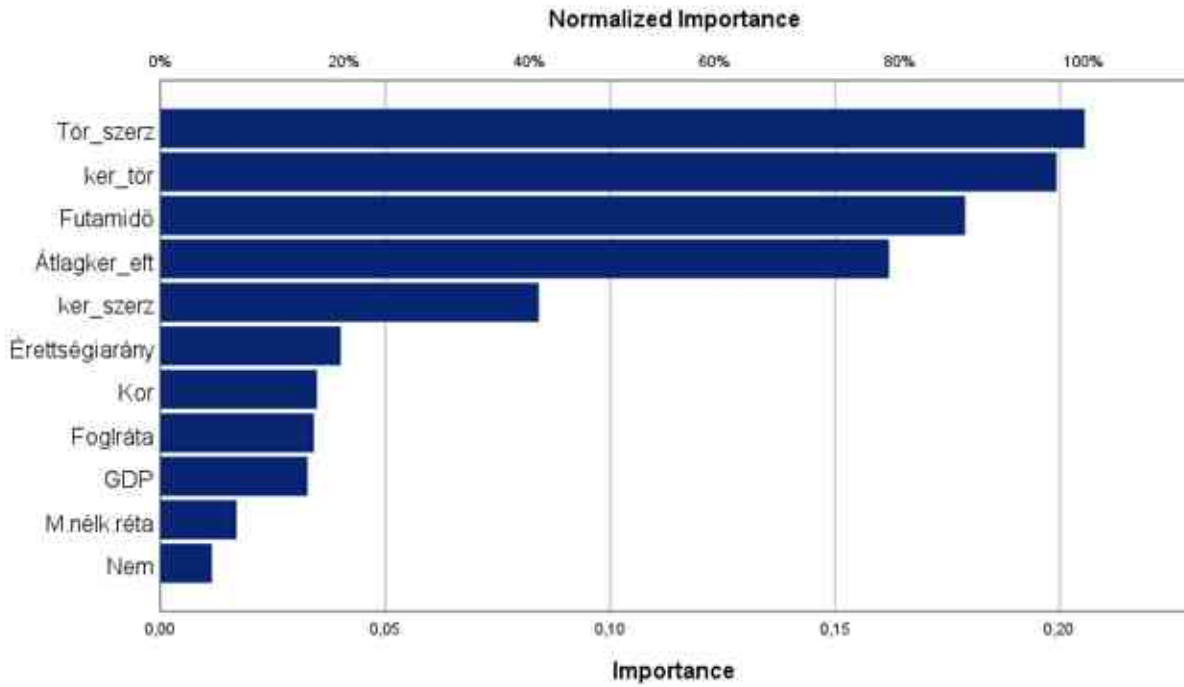
11.12 Neurális háló IV. outputjai

Network Information			
Input Layer	Covariates	1	tör_szerz*100
		2	Felv kor
		3	Nem
		4	Futamidő
		5	Munkanélküliségi ráta
		6	Foglalkoztatási ráta
		7	1főre jutó GDP
		8	Érettségi arány
		9	ker_szerz*100
		10	ker tör*100
		11	Átlagker eft
	Number of Units ^a		11
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		6
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

Model Summary		
Tanuló	Cross Entropy Error	91,863
	Percent Incorrect Predictions	11,7%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,11
Teszt	Cross Entropy Error	53,432
	Percent Incorrect Predictions	16,6%

Dependent Variable: Default

a. Error computations are based on the testing sample.



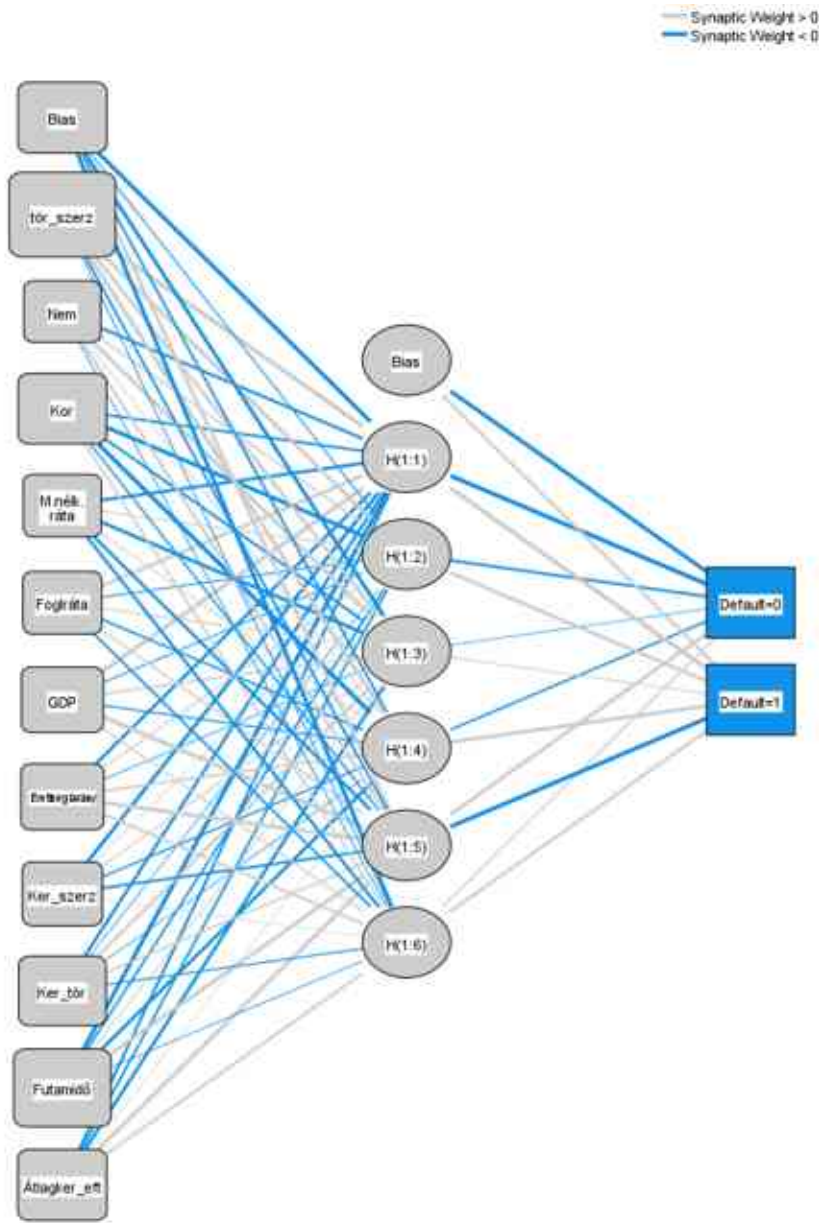
11.13 Egyedi modellek outputjai

11.13.1Egyéb hitelügylet

Network Information			
Input Layer	Covariates	1	tör_szerz*100
		2	Nem
		3	Kor
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker_tör*100
		10	Futamidő
		11	Átlagker_eft
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		6
	Activation Function		Hyperbolic tangens
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

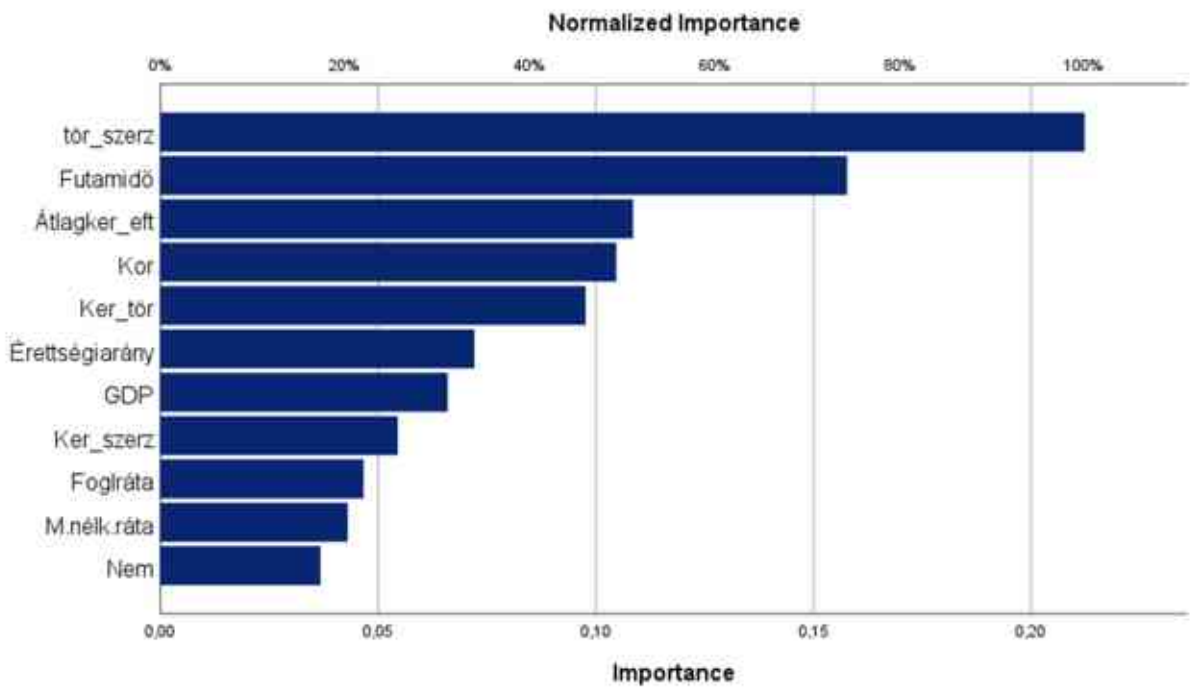
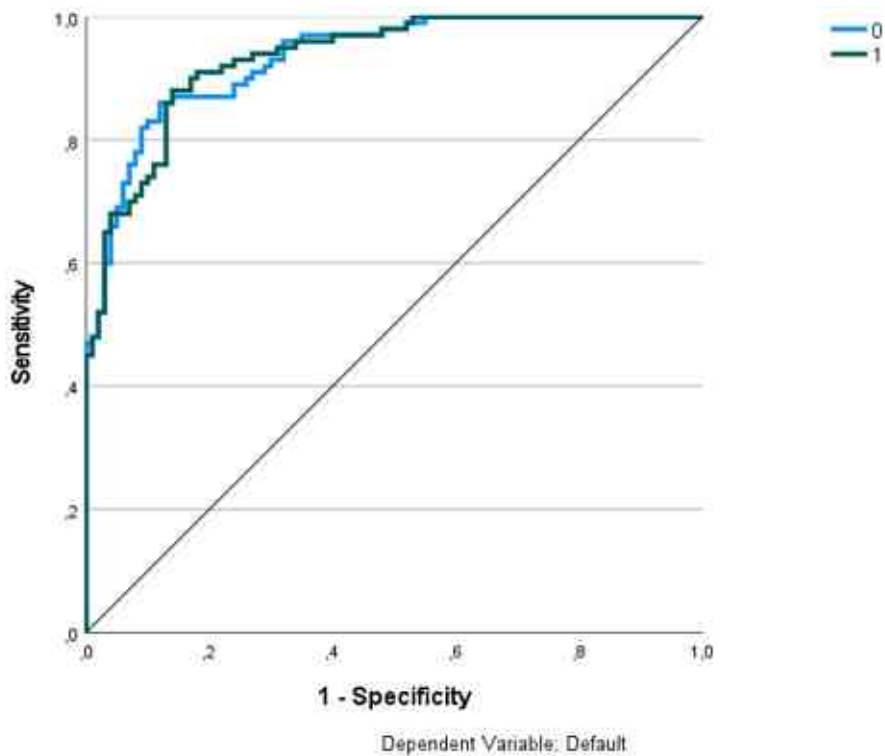
Model Summary		
Tanuló	Cross Entropy Error	44,982
	Percent Incorrect Predictions	12,7%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,06
Teszt	Cross Entropy Error	20,149
	Percent Incorrect Predictions	13,8%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

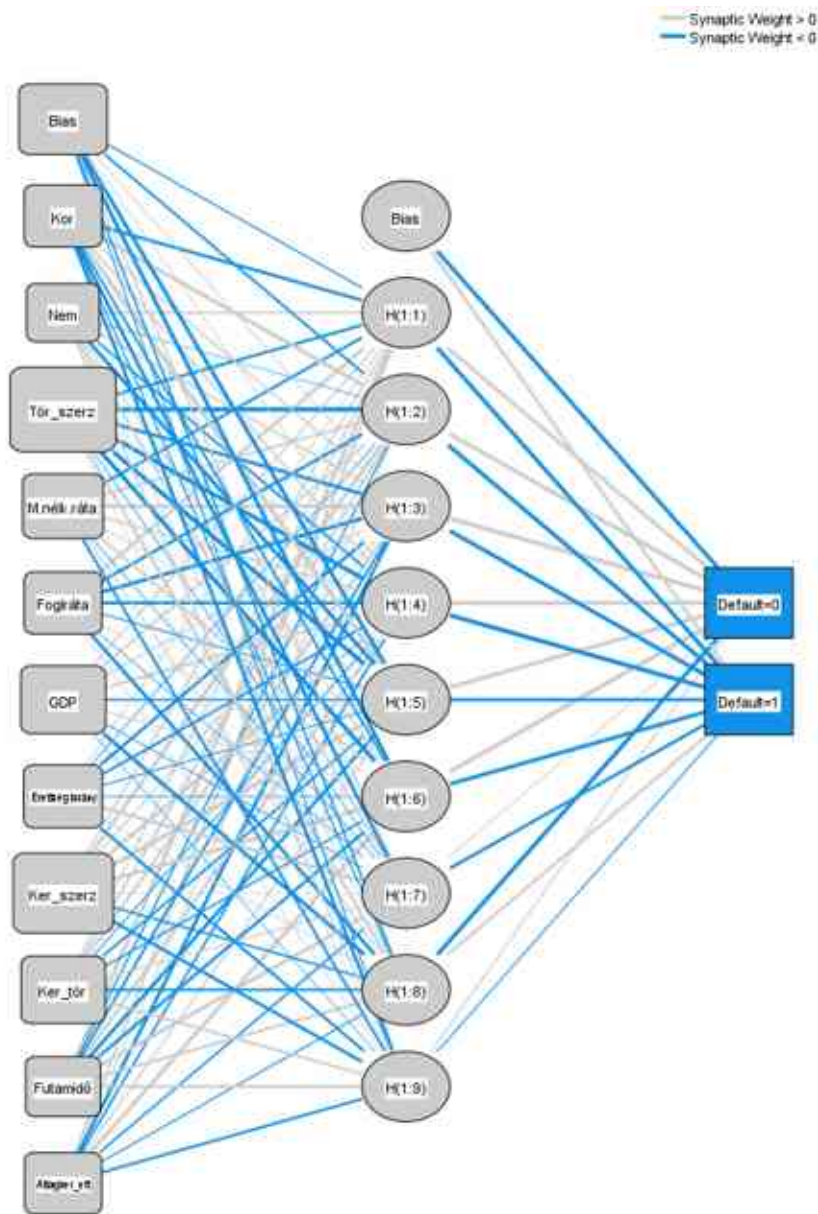
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	62	8	88,6%
	1	10	62	86,1%
	Overall Percent	50,7%	49,3%	87,3%
Teszt	0	24	6	80,0%
	1	2	26	92,9%
	Overall Percent	44,8%	55,2%	86,2%
Dependent Variable: Default				



11.13.2 Fogyasztási és áruvásárlási hitel

Network Information			
Input Layer	Covariates	1	Kor
		2	Nem
		3	tör_szerz*100
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1 főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker_tör*100
		10	Futamidő
		11	Átlagker eft
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		9
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

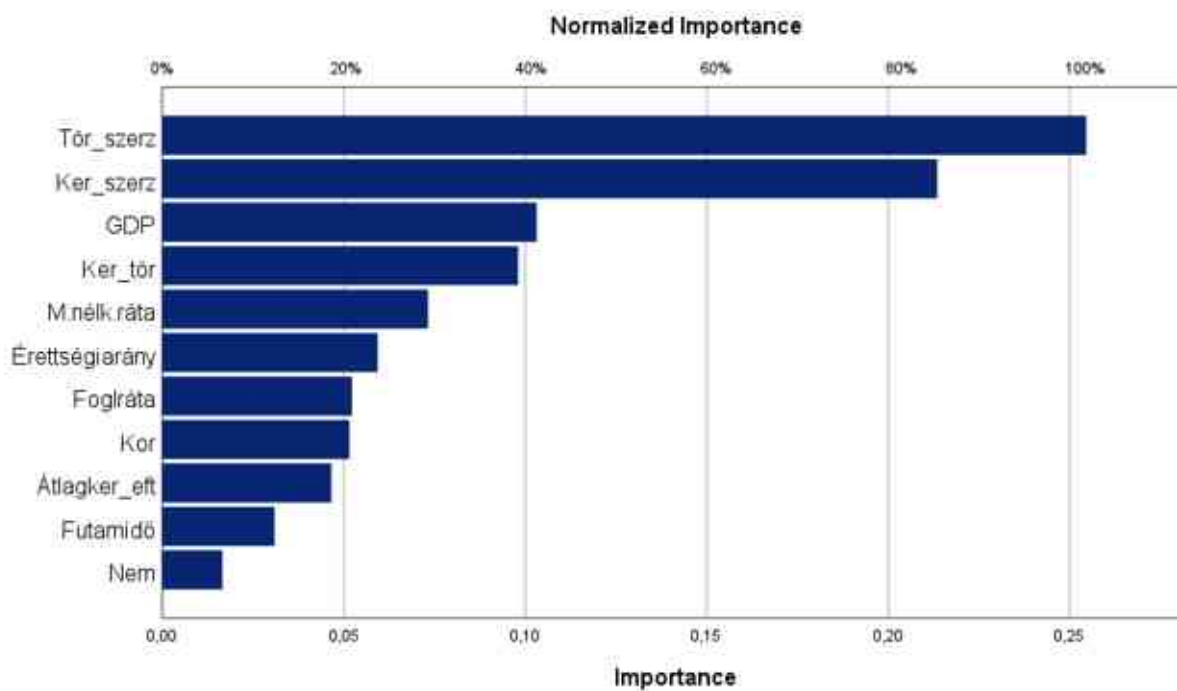
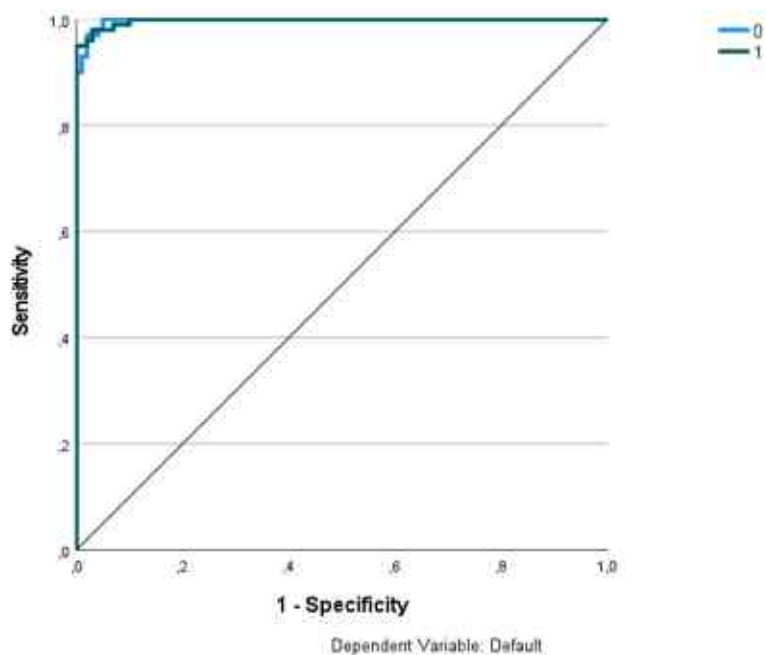
Model Summary		
Tanuló	Cross Entropy Error	10,604
	Percent Incorrect Predictions	3,5%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,06
Teszt	Cross Entropy Error	3,329
	Percent Incorrect Predictions	3,4%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

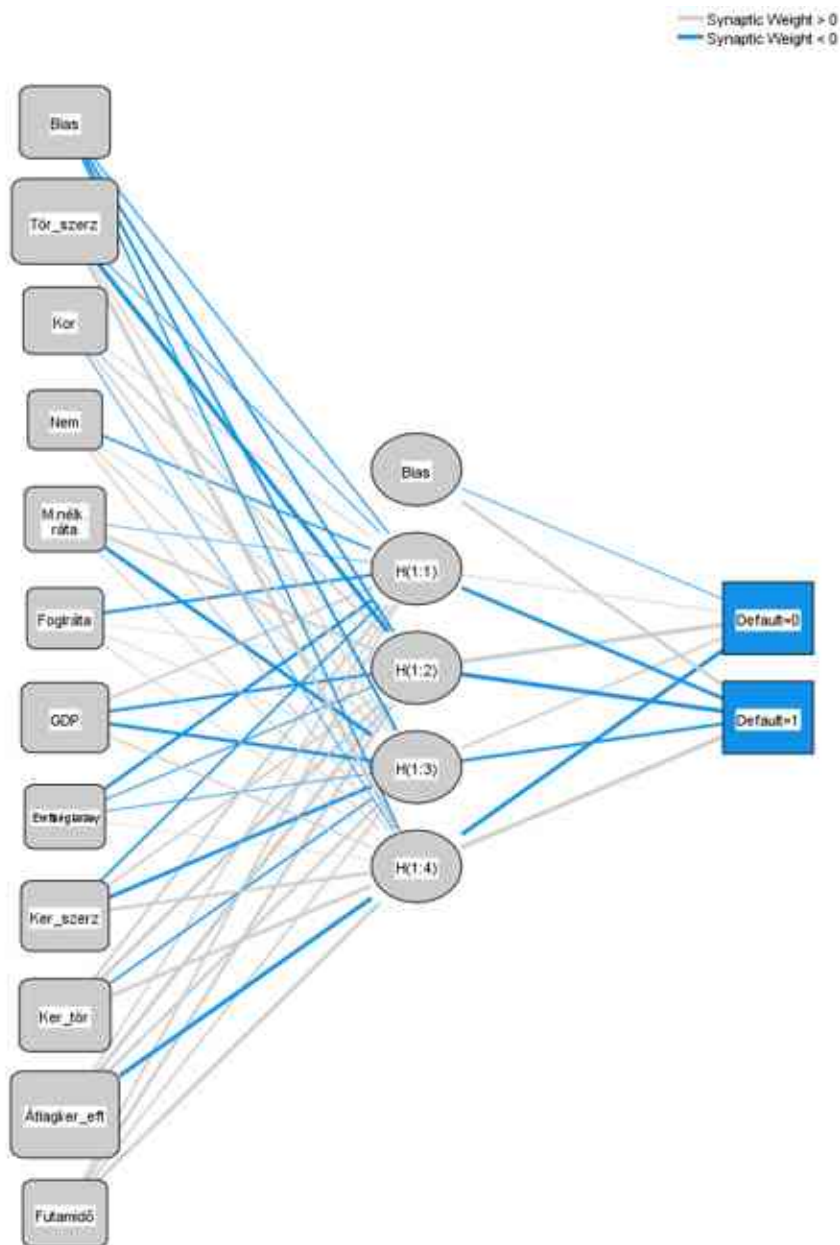
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	70	2	97,2%
	1	3	67	95,7%
	Overall Percent	51,4%	48,6%	96,5%
Teszt	0	27	1	96,4%
	1	1	29	96,7%
	Overall Percent	48,3%	51,7%	96,6%
Dependent Variable: Default				



11.13.3Gépjármű lízing

Network Information			
Input Layer	Covariates	1	tör szerz*100
		2	Kor
		3	Nem
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker tör*100
		10	Átlagker_eft
		11	Futamidó
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

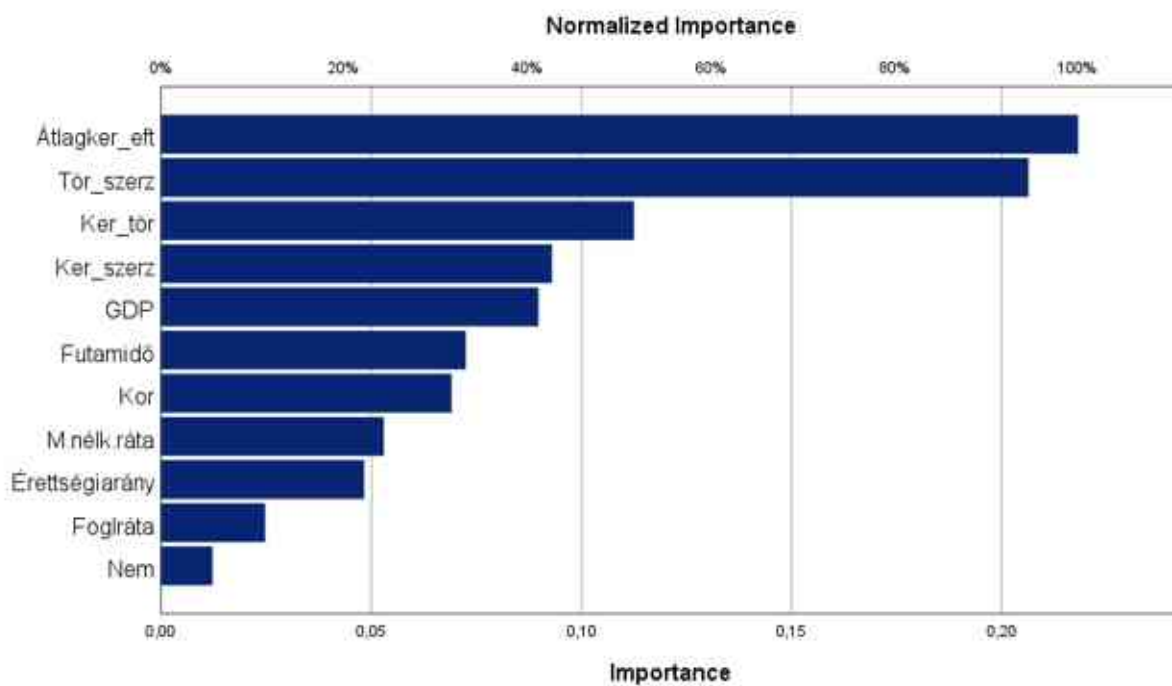
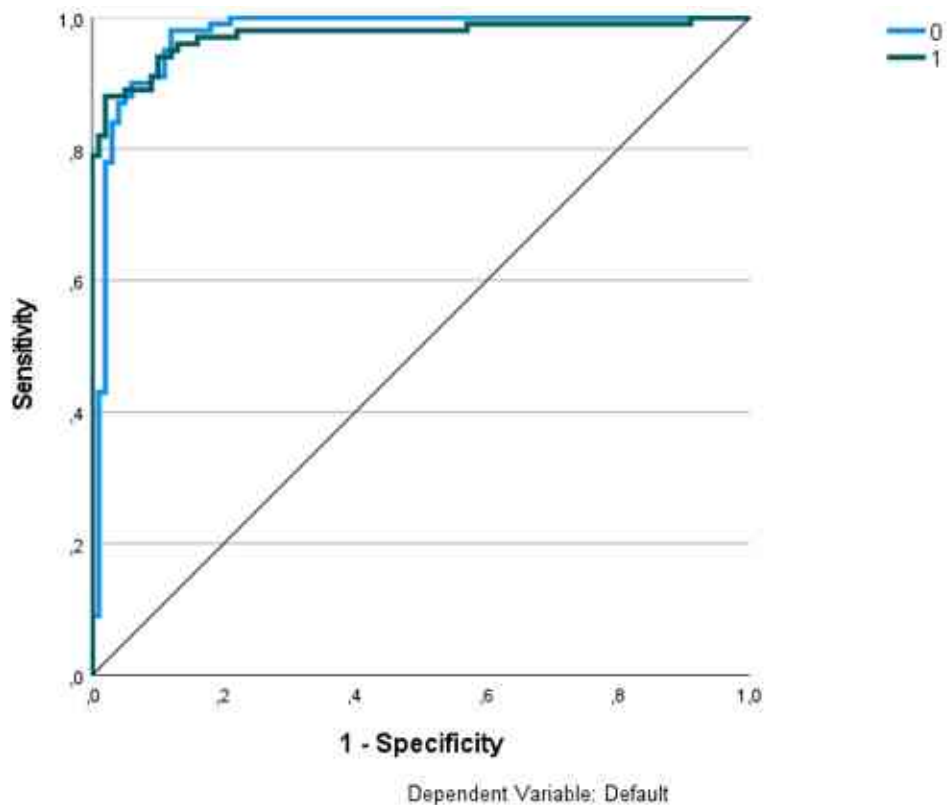
Model Summary		
Tanuló	Cross Entropy Error	24,873
	Percent Incorrect Predictions	9,2%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,07
Teszt	Cross Entropy Error	15,606
	Percent Incorrect Predictions	8,6%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent
 Output layer activation function: Softmax

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	67	7	90,5%
	1	6	62	91,2%
	Overall Percent	51,4%	48,6%	90,8%
Teszt	0	26	0	100,0%
	1	5	27	84,4%
	Overall Percent	53,4%	46,6%	91,4%

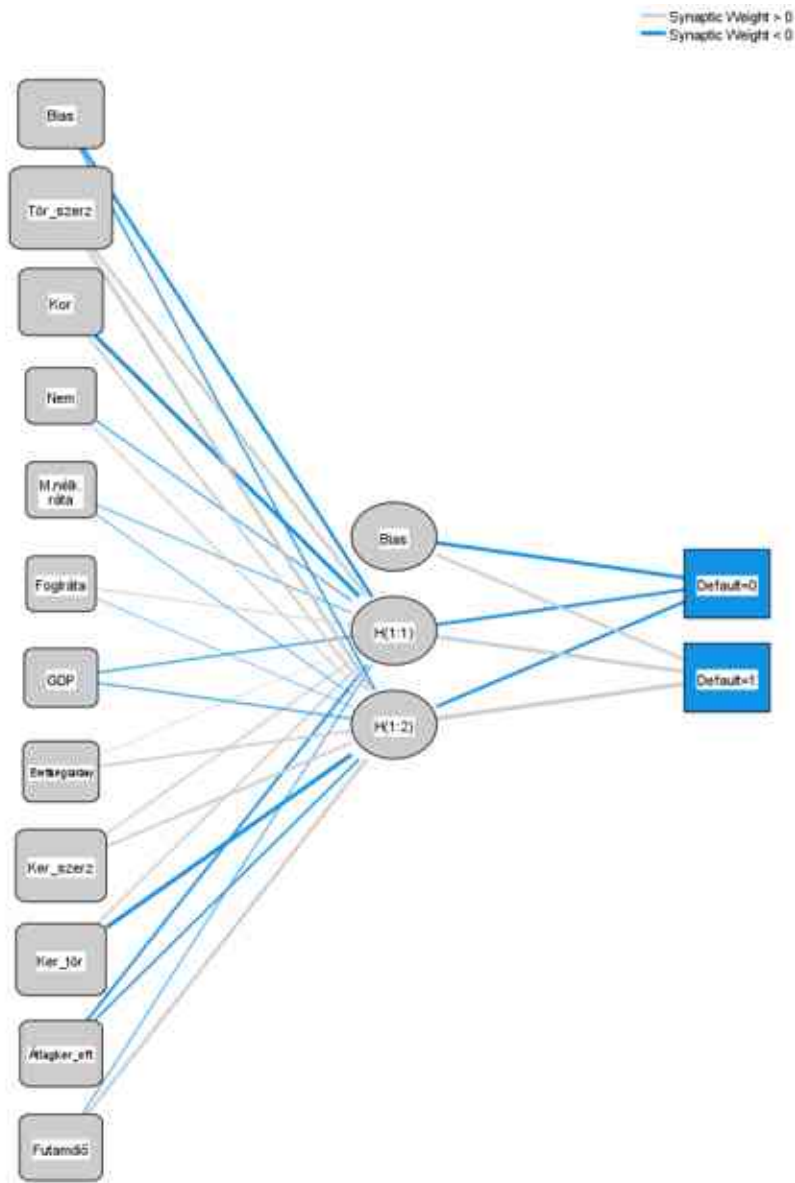
Dependent Variable: Default



11.13.4Lakáscélú jelzáloghitel

Network Information			
Input Layer	Covariates	1	tör szerz*100
		2	Kor
		3	Nem
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker tör*100
		10	Átlagker_eft
		11	Futamdió
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		2
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

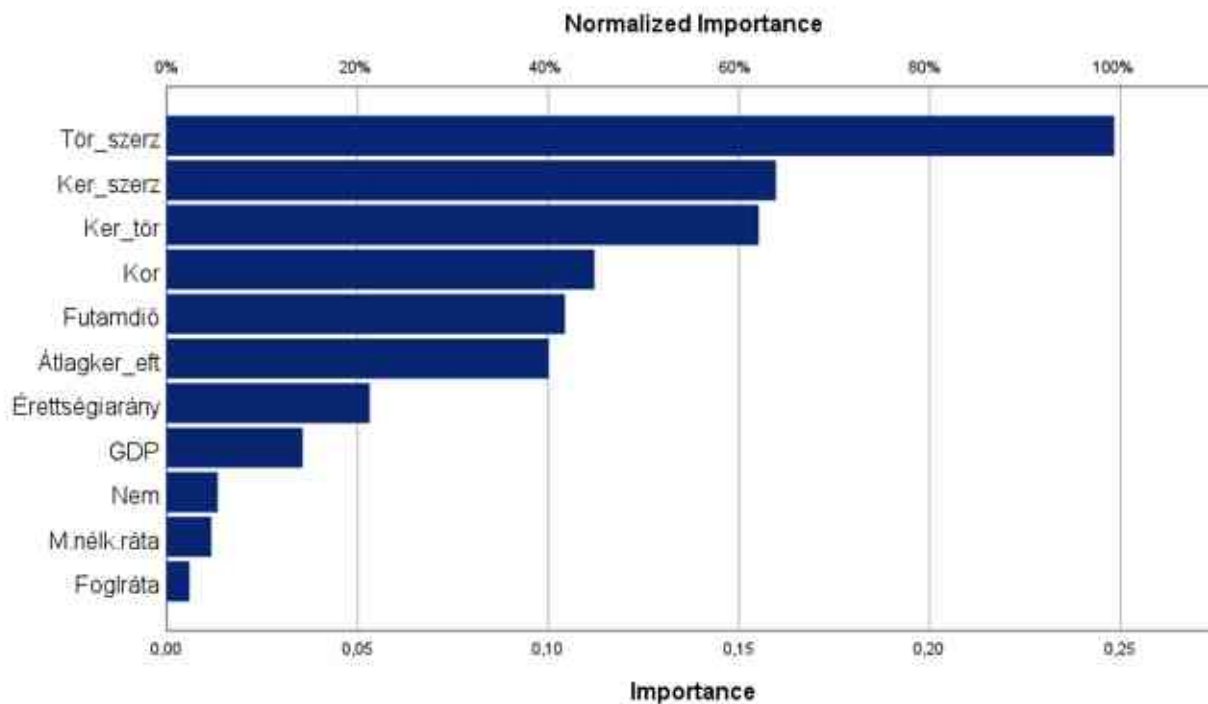
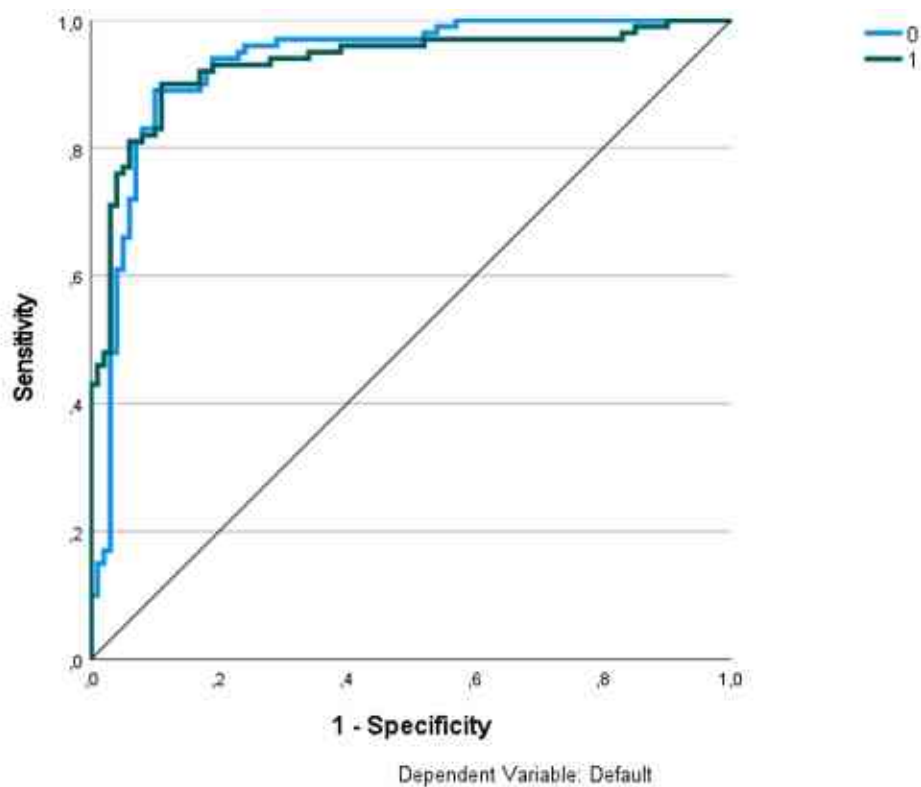
Model Summary		
Tanuló	Cross Entropy Error	42,100
	Percent Incorrect Predictions	9,4%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,12
Teszt	Cross Entropy Error	22,780
	Percent Incorrect Predictions	13,1%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent
 Output layer activation function: Softmax

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	64	7	90,1%
	1	6	62	91,2%
	Overall Percent	50,4%	49,6%	90,6%
Teszt	0	25	4	86,2%
	1	4	28	87,5%
	Overall Percent	47,5%	52,5%	86,9%

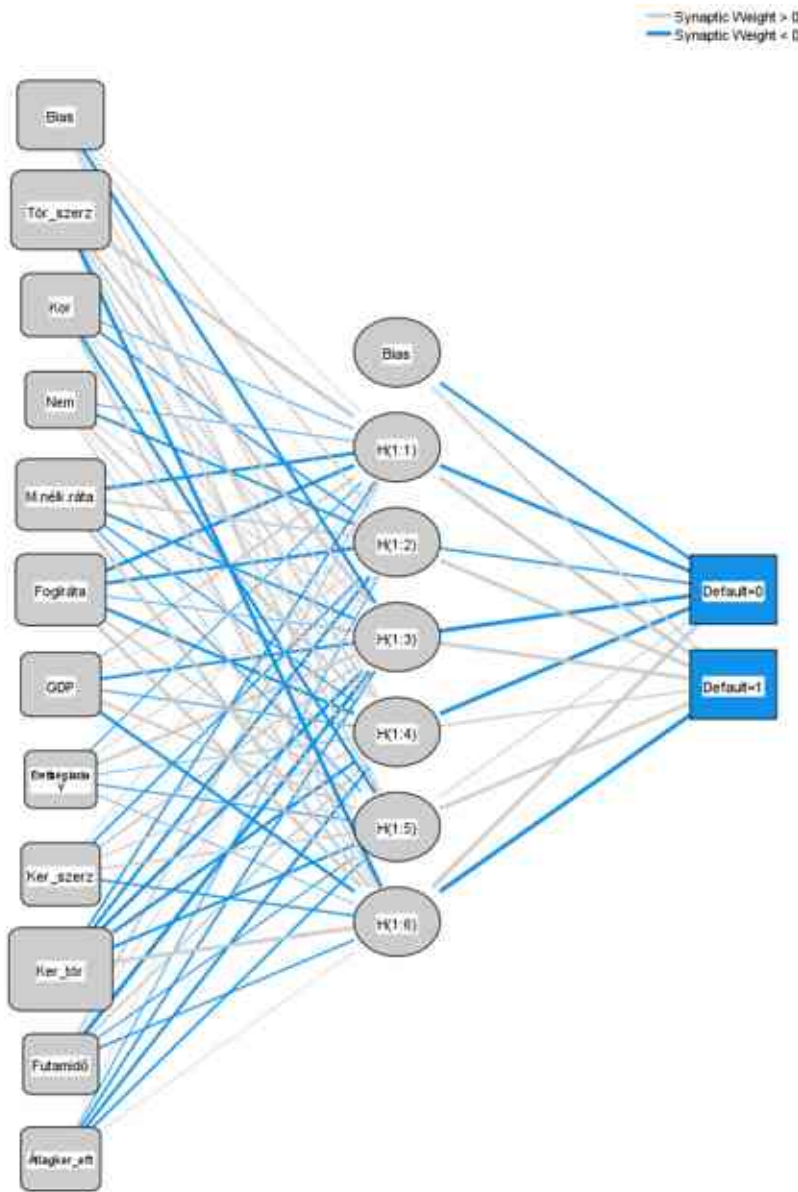
Dependent Variable: Default



11.13.5Szabad felhasználású jelzaloghitel

Network Information			
Input Layer	Covariates	1	tör_szerz*100
		2	Kor
		3	Nem
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1 főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker tör*100
		10	Futamidő
		11	Átlagker eft
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		6
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

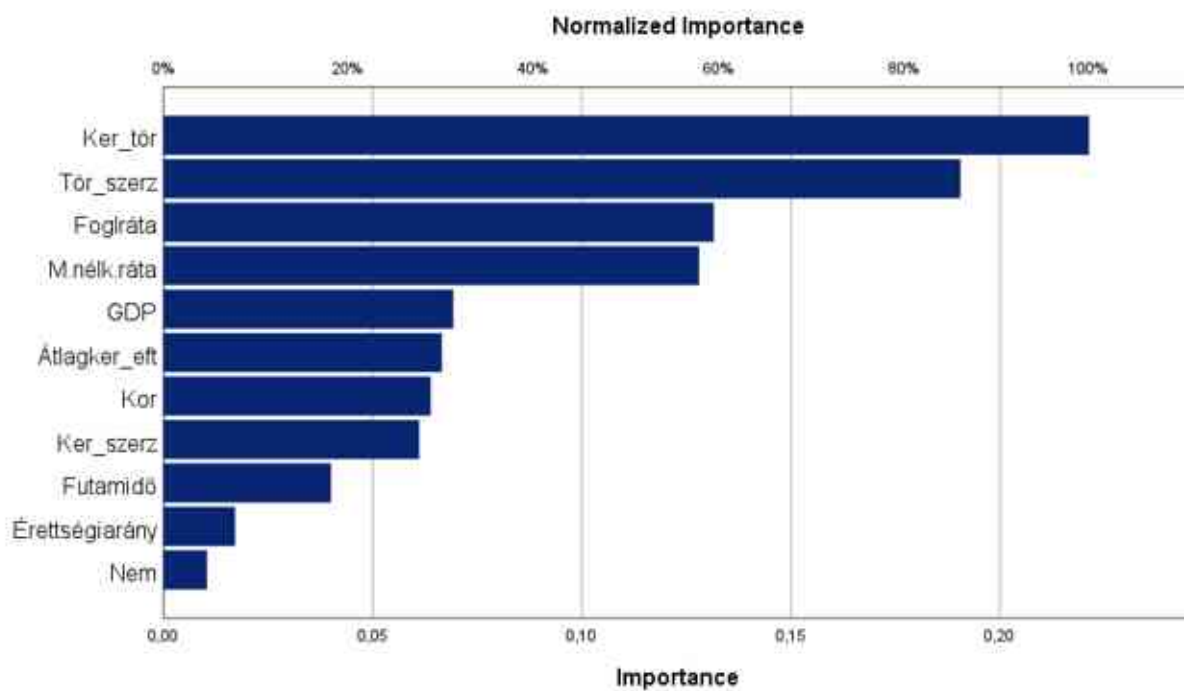
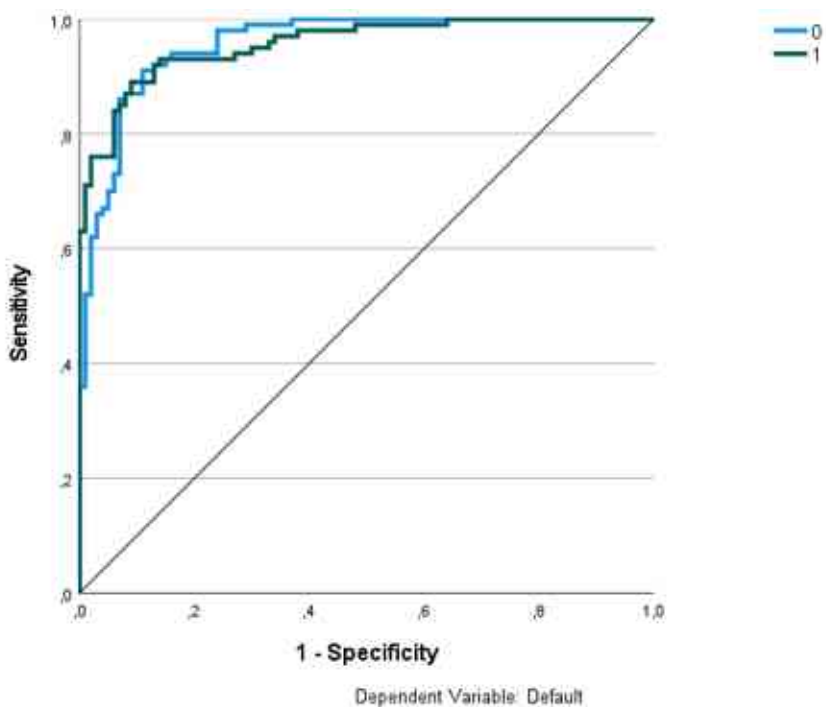
Model Summary		
Tanuló	Cross Entropy Error	38,387
	Percent Incorrect Predictions	12,0%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,07
Teszt	Cross Entropy Error	13,611
	Percent Incorrect Predictions	10,3%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent
 Output layer activation function: Softmax

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	70	6	92,1%
	1	11	55	83,3%
	Overall Percent	57,0%	43,0%	88,0%
Teszt	0	23	1	95,8%
	1	5	29	85,3%
	Overall Percent	48,3%	51,7%	89,7%

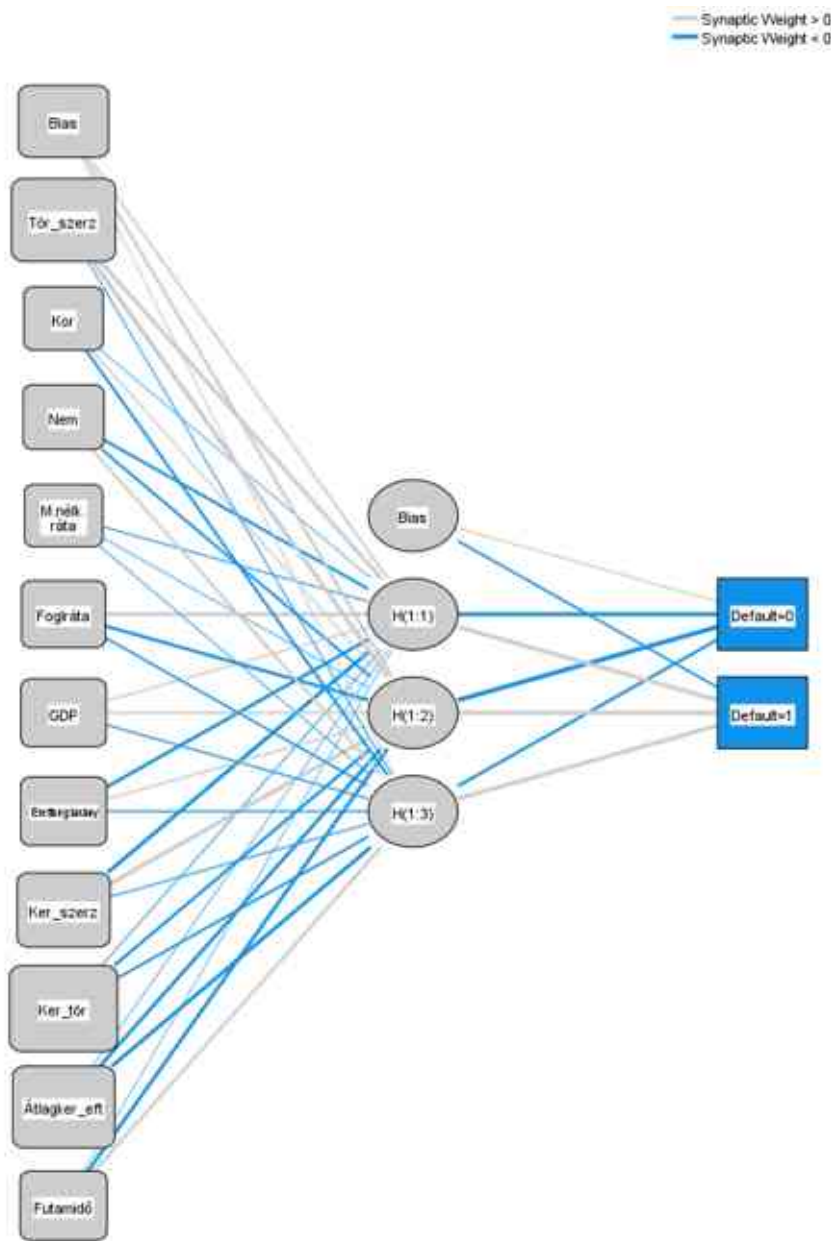
Dependent Variable: Default



11.13.6Személyi hitel

Network Information			
Input Layer	Covariates	1	tör_szerz*100
		2	Kor
		3	Nem
		4	Munkanélküliségi ráta
		5	Foglalkoztatási ráta
		6	1 főre jutó GDP
		7	Érettségi arány
		8	ker_szerz*100
		9	ker_tör*100
		10	Átlagker_eft
		11	Futamidő
Number of Units ^a		11	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		3
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Default
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

Model Summary		
Tanuló	Cross Entropy Error	37,757
	Percent Incorrect Predictions	10,6%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,08
Teszt	Cross Entropy Error	20,052
	Percent Incorrect Predictions	19,0%
Dependent Variable: Default		
a. Error computations are based on the testing sample.		



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	69	6	92,0%
	1	9	58	86,6%
	Overall Percent	54,9%	45,1%	89,4%
Teszt	0	19	6	76,0%
	1	5	28	84,8%
	Overall Percent	41,4%	58,6%	81,0%
Dependent Variable: Default				

