

MISKOLCI EGYETEM
GAZDASÁGTUDOMÁNYI KAR



FODOR KITTI

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

Doktori (Ph.D) értekezés tézisei

Miskolc, 2023.

MISKOLCI EGYETEM
GAZDASÁGTUDOMÁNYI KAR
HANTOS ELEMÉR GAZDÁLKODÁS- ÉS REGIONÁLIS TUDOMÁNYI DOKTORI
ISKOLA

FODOR KITTI

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

Doktori (Ph.D) értekezés tézisei

A DOKTORI ISKOLA VEZETŐJE:

Prof. Dr. Sikos T. Tamás

TUDOMÁNYOS VEZETŐ:

Dr. Varga Beatrix Margit

Miskolc, 2023.

Tartalom

1	Témaválasztás indoklása	1
1.1.	A kutatás aktualitása, kutatási probléma megfogalmazása	1
1.1	A disszertáció felépítése, hipotézisei	1
2	A kutatás elméleti alapjai	3
2.1	Magyar hitelkörkép, eladósodottság.....	4
2.2	Egyéb, a hitelezéssel összefüggő, nehezen mérhető tényezők	5
2.3	Nemzetközi csődmodell-kutatások.....	6
2.4	Magyar csődmodell-kutatások.....	7
2.5	A csődelőrejelzés módszerei	7
3	A kutatás adatbázisa	9
3.1	A Központi Hitelnyilvántartó Rendszer	9
3.2	Adatbázis	9
4	A kutatás eredményei	12
4.1	Az eredmények felhasználhatósága, további kutatási irányok	33
5	Összefoglalás.....	34
6	Irodalomjegyzék.....	35
7	A szerző publikációi	37

Ábrajegyzék

1. ábra: A magyar lakosság fogyasztói hiteleinek alakulása 2001-2021 (Mrd Ft).....	4
2. ábra: A pénzügyi kultúra szintjei az Európai országokban	5
3. ábra: A fennálló mulasztások megoszlása a mulasztás időtartama szerint 2021-ben	9
4. ábra: Az adatbázis szűkítésének lépései	11
5. ábra: Döntési fa I. a tanuló minta esetében	13
6. ábra: A kialakult neurális háló a Neurális háló I. modell esetében	14
7. ábra: Döntési fa II. a tanuló minta esetében	16
8. ábra: A kialakult neurális háló a Neurális háló II. modell esetében.....	17
9. ábra: Döntési fa III. a tanuló minta esetében.....	21
10. ábra: A kialakult neurális háló a Neurális háló III. modell esetében	22
11. ábra: Döntési fa IV. a tanuló minta esetében	25
12. ábra: A kialakult neurális háló a Neurális háló IV. modell esetében	26
13. ábra: A létrehozott modellek ROC-görbéje	28

Táblázatjegyzék

1. táblázat: A Logisztikus regresszió I. modell szignifikáns változói	12
2. táblázat: Klasszifikációs mátrix a Logisztikus regresszió I. modell esetében.....	13
3. táblázat: Klasszifikációs mátrix a Döntési fa I. modell esetében	14
4. táblázat: Klasszifikációs mátrix a Neurális háló I. modell esetében	15
5. táblázat: A Logisztikus regresszió II. modell szignifikáns változói.....	15
6. táblázat: Klasszifikációs mátrix a Logisztikus regresszió II. modell esetében	16
7. táblázat: Klasszifikációs mátrix a Döntési fa II. modell esetében.....	17
8. táblázat: Klasszifikációs mátrix a Neurális háló II. modell esetében.....	18
9. táblázat: A klasszifikációs modellek által használt változók összefoglalása	18
10. táblázat: Az alkalmazott módszerek besorolási pontossága.....	19
11. táblázat: A Logisztikus regresszió III. modell szignifikáns változói	20
12. táblázat: Klasszifikációs mátrix a Logisztikus regresszió III. modell esetében.....	21
13. táblázat: Klasszifikációs mátrix a Döntési fa III. modell esetében	22
14. táblázat: Klasszifikációs mátrix a Neurális háló III. modell esetében	23
15. táblázat: A Logisztikus regresszió IV. modell szignifikáns változói	23
16. táblázat: Klasszifikációs mátrix a Logisztikus regresszió IV. modell esetében.....	24
17. táblázat: Klasszifikációs mátrix a Döntési fa IV. modell esetében	25
18. táblázat: Klasszifikációs mátrix a Neurális háló IV. modell esetében	27
19. táblázat: A létrehozott modellek magyarázó változóinak összefoglalása	27
20. táblázat: A kialakított modellek teljesítménye a különböző értékelési technikák alapján .	28
21. táblázat: A modellek teljesítményének összehasonlítása	32

1 Témaválasztás indoklása

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 mely tényezők 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.

1.1 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 csődmódel-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 készítetek. 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:

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

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

A H1, H2, H3 hipotézis tesztelésére az alábbi eszközöket alkalmaztam:

- Logisztikus regresszió,
- Döntési fa,
- Neurális háló,
- ROC-görbe,
- Gini-koefficiens,
- Korreláció

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 H4 és H5 hipotézis tesztelésére az alábbi eszközöket alkalmaztam:

- Logisztikus regresszió,
- Döntési fa,
- Neurális háló,
- Keresztábra elemzés

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 H6 hipotézis tesztelésére az alábbi eszközöket alkalmaztam:

- Neurális háló,
- ROC-görbe,
- Gini-koefficiens

2 A kutatás elméleti alapjai

Kutatásom első lépése a magyar és nemzetközi szakirodalom tanulmányozása volt. Melyet röviden ebben a fejezetben foglalok össze. Úgy gondolom, ez hozzájárulhat a kutatásom, és annak eredményeinek a megértéséhez.

A hitelezés évezredek óta az emberi kultúra része. Az első szabályok Hammurápihoz köthetőek, aki kőtáblákon rögzítette a kölcsönzőkre és a kölcsönvevőkre vonatkozó szabályokat. (Fekete-Tatay, 2012) Az ókorban a hitelek nemfizetése komoly következményekkel járt, akár halállal is büntették. A középkorban a niceai zsinat kamattilalmat írt elő, és ez az ellenállás egészen a reformációig fennállt. A XV. században merült fel a gondolat, hogy a pénzt mégiscsak forgatni kell a gazdaság érdekében. Az 1600-1800-as években a hiteleket főként gazdag birtokosok, polgárok nyújtották. A kontinentális országokban a XIX. századtól beszélhetünk hitelezésről. (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. (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. A 90-es évek elején több bankot is érintett csőd, a bankrendszer stabilitásának kezdete a '90-es évek közepéig azonban váratott magára. (Vértesy, 2008)

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 megtakarítások ezzel ellentétesen alakultak. Már a válság kialakulása előtt voltak jelek, hogy a pénzügyi tudatosság, pénzügyi kultúra esetén kedvezőtlen a helyzet a magyar lakosság esetében, amely veszélyes tényezőnek minősülhet a hitelezés kapcsán, és ennek a tudásnak a hiánya akár adósságcspadához is vezethet.

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) 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 törlesztőrészeket olyan mértékű növekedésével járt, hogy a hitelezők már nem tudták ennek eleget tenni. Végezetül 2014-ben kivezették a devizaalapú hiteleket.

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

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)

2.1 Magyar hitelkörkép, eladósodottság

A fogyasztói hitelek esetében a 2009-es csúcst követően a felvett hitelek összege folyamatosan csökkent, azonban ezt 2016 után egy gyorsütemű növekedés követte, 2019-ben meghaladva a 2009-es csúcsponti értéket, 2021-re pedig megduplázva azt. Ennek alakulása részben a törvényi módosításnak is köszönhető.



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

Forrás: Trading Economics

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, 2015)

A hitelfelvétel egyik legnagyobb kockázata, ha az igénylő nem tudja visszafizetni a felvett összeget. Ennek következménye lehet a késedelmi kamat felszámolása, követeléskezelő költségeinek felszámolása, vagy a KHR listára történő felkerülés. A legsúlyosabb következmény azonban a hitel fedezetéül 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 32/2014. MNB rendelet tartalmaz. 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, 2023)

2.2 Egyéb, a hitelezéssel összefüggő, nehezen mérhető tényezők

A disszertációban a területtel kapcsolatos kutatásaim kiterjedtek a pénzügyi inklúzió, kultúra és reziliencia területére, azonban itt röviden kizárólag a pénzügyi kultúrára térek ki.

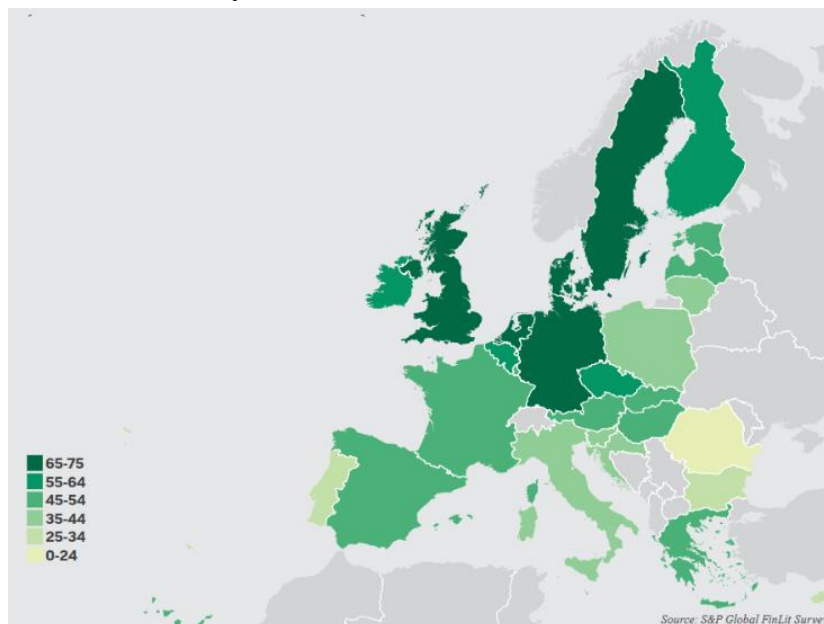
Béres-Huzdik úgy fogalmazott, hogy a hitelezés a gazdaság egy fontos mozgatórugója, melynek ahhoz, hogy a gazdaság megfelelő teljesítőképességgel rendelkezzen jól kell funkcionálnia. Azonban maga a hitelezés nem elegendő ehhez, fontos, hogy a hitelfelvevők tudatos pénzügyi döntéseket hozzanak. A hitelezéssel 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. 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)

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.)

Potóczki cikke alapján az mondható el, hogy a pénzügyi kultúra szintje hazánkban átlag alattinak minősíthető. Ez a következtetés az OECD által kezdeményezett nemzetközi kutatás alapján vonta le. A kutatásban részt vett 30 ország közül Magyarország a 22. helyen végzett. (Potóczki, 2017)

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



2. ábra: A pénzügyi kultúra szintjei az Európai országokban
Forrás: *Financial literacy around the World*, Klapper et al., 7., 2015

2.3 Nemzetközi csődmodell-kutatások

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, azokat a mutatószámokat keresték a kutatók, amelyek esetében a csődbe jutott és a túlélő vállalatok szignifikánsan különböznek, ezeket a mutatószámokat összehasonlították, és ez alapján próbálták különböző összefüggéseket megállapítani.

Az első egyváltozós elemzés Beaver nevéhez köthető. Elemzésébe 158 vállalatot vont be, egyenlő arányban szerepeltek a csődbe ment és a túlélő vállalatok. A módszerével 90%-os besorolási pontossággal tudta a vállalatokat kategorizálni. A módszer hátránya, hogy egyváltozós modelltől 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)

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. 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, az új modell esetében már egy nagyobb, 111 elemű mintával dolgozott, melyben 58 túlélő vállalat kapott helyet.

Az Altman nevéhez fűződő modellek nem voltak reprezentatívak, és megközelítőleg azonos arányban szerepeltek a mintában a fizetőképes és a fizetéseképtelen vállalatok. Az első reprezentatívnak tekinthető felmérés Ohlson nevéhez fűződik. Ohlson nevéhez fűződik továbbá, hogy elsőként ő alkalmazta a csődelőrejelzési modellekben a logisztikus regressziót. Az általa vizsgált mintába 2163 vállalat került, melynek a 4,85%-a volt csődbe ment. Amennyiben a modell során számított P érték meghaladja a 0,038-as értéket, akkor a vállalat csődveszélyesnek minősül. A modell megközelítőleg 83%-os pontossággal rendelkezik. (Ohlson, 1980)

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. A modell besorolási pontossága 94%, azonban jelentős különbség mutatkozott a helyes kategorizálást illetően a túlélő és a csődbe ment vállalatok között. (Frydman et al., 1985)

A következő nagy lépést a neurális hálók megjelenése jelentette, mely a 90-es évekre tehető. A neurális hálók első alkalmazása Odom és Sharda nevéhez köthető. Modelljüket az Altman által 1968-ban használt változókra építették. A minta 129 vállalatból állt. A tanuló minta esetében a besorolás tökéletes volt, tehát felülmúlta a diszkriminancia analízis által elért eredményeket. A teszt minta esetében a 82%-os besorolási pontossággal jelentősen meghaladta a diszkriminancia által elért eredményt. (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 spanyol bankok adatait dolgozták fel. Kutatásaikat a fentebb említett minden modell segítségével elvégezték, melyek közül a legjobbnak a neurális háló bizonyult, 82,4%-os besorolási pontossággal. (Olmeda-Fernandez, 1997) Zhang-Hu-Patuwo (1999) a neurális hálót a logisztikus regresszióval vetette össze. Előbbi 88,2%, míg utóbbi 78,6%-os besorolási pontosságot ért el. (Zhang et al., 1999)

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

2.4 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. A diszkriminancia analízis 77,92%-os besorolási pontossággal kategorizált helyesen. A logisztikus regresszió teljesítménye meghaladja a diszkriminancia analízissel készült modell teljesítményét, 81,8%-ot ért el. (Virág-Kristóf, 2005)

A szerzőpáros 1996-ben 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 modell részleteit, egyenleteit a szerzők üzleti titkok miatt azonban nem publikálták.

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)

2.5 A csődelőrejelzés módszerei

A csődelőrejelzés módszerei közül négy elemzési lehetőséget vizsgáltam meg részletesebben. A klasszifikációs eljárások egy típusa a **diszkriminancia analízis**, ami egy olyan elemzési módszer, mely során a függő változó létező kategóriáiba kerülnek az esetek besorolásra. Az elemzés segítségével arra a kérdésre derülhet fény, hogy melyek azok a változók, melyek esetében a meglévő csoportok szignifikánsan eltérnek egymástól. A módszer egyik hátránya, hogy nem alkalmas olyan esetek elemzésére, melyek során nemlineáris kapcsolatot feltételezünk, továbbá 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. (Malhotra, 2008; Sajtos-Mitev, 2007)

A **logisztikus regresszió** esetében a cél ugyanaz, mint ami a diszkriminanciaelemzés esetében, azonban a diszkriminanciaelemzéssel ellentétben a logisztikus regresszió esetében a feltételek sora sokkal rövidebb. (Hajdú, 2003; Malhotra, 2008)

A **rekurzív particionáló algoritmus** többféle néven is ismert, sokan egyszerűen döntési fáként hivatkoznak rá. A létrejött alcsoportokat csomópontnak (node) nevezik. Az előrejelzés alapját a levelek adják, melyek a fának azon része, amit nem osztunk tovább. (Hajdú, 2008) A cél a csoporton belüli variancia minimalizálása, úgy, hogy a csoportok közötti variancia a lehető legnagyobb legyen. Az algoritmus segítségével egy iteratív folyamat megy végbe, melyet kimondottan számítógépre terveztek.

A döntési fának több típusa létezik, melyek közül a csodelőrejelzés során a CHAID¹ terjedt el. Az elemzés nagy előnye, hogy a bevont változók esetében nincs megkötés, metrikus és nem metrikus változók is bevonhatók. A módszer mellett szóló érv, hogy nem szerepel a feltételek között a változók normális eloszlása. 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 előrejelzési célra nem alkalmazható, mivel többségében a tanuló adatbázisra specializálódik. 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)

Az utolsó módszer pedig a **neurális háló**. 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.” (Ketskeméty et al., 2011, p394.)

Elemzéseim során az MLP² 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. (Ketskeméty et al., 2011)

A neurális hálók számos előnnyel rendelkeznek:

- kezeli a nemlinearitást,
- nem okoz problémát a hiányzó adat
- nagy számú változó és elemszám kezelésére alkalmas. (Kristóf, 2002)

A modellek értékelésére és összehasonlítására a klasszifikációs mátrixot, ROC-görbét és Gini-koefficiens alkalmaztam. Az ROC-görbe alapján kiszámított AUC érték esetében, amennyiben értéke 80-90% körül mozog, az már kiemelkedőnek tekinthető. A Gini-koefficiens esetében szintén a 80-90% közötti érték már nagyon jó előre jelző modellre utal.

¹ Chi-squared Automatic Trees

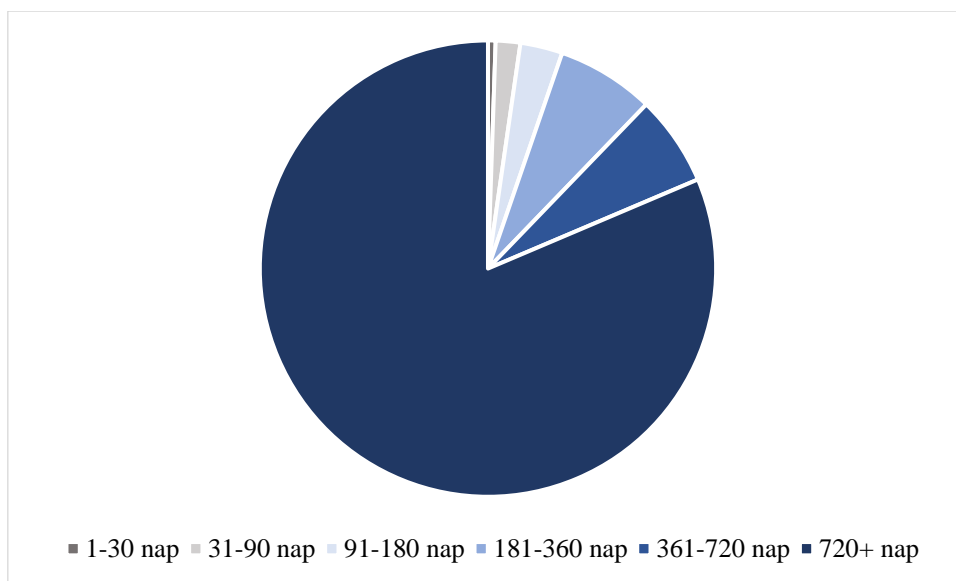
² Multi-Layer Perceptron

3 A kutatás adatbázisa

3.1 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 KHR úgynevezett teljes listát vezet, azaz a kötelezettségeknek időben eleget tevő ügyfelek is szerepelnek a nyilvántartási rendszerben.

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.



3. 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)

3.2 Adatbázis

Az elemzések elvégzéséhez 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 találhatóak 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 keresztnemek 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.

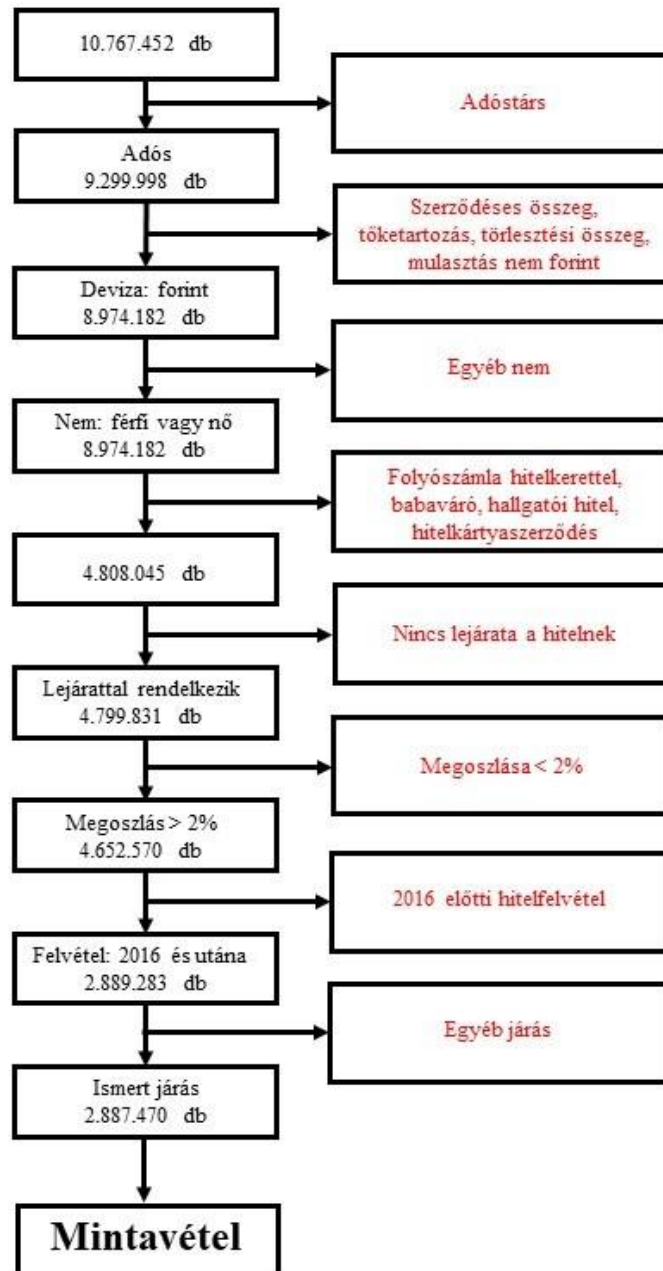
A fenti változókon túlmenően az alábbiakkal bővíttem az adatbázist:

- 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 járásban
- Munkanélküliségi ráta: a hitelfelvétel évének munkanélküliségi rátája az adott járásban
- Foglalkoztatási ráta: a hitelfelvétel évének foglalkoztatási rátája az adott járásban
- 1 főre jutó GDP: a hitelfelvétel évének értéke az adott járásban
- Érettségizők aránya: a hitelfelvétel évének értéke az adott járásban
- 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 visszaszámítottam az ügyfél születési évét, majd ezt követően pedig a hitelfelvételkori életkorát.

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:



4. ábra: Az adatbázis szűkítésének lépései

Forrás: Saját szerkesztés

4 A kutatás eredményei

A disszertációmban hat hipotézist fogalmaztam meg. Kutatásom új és újszerű megállapításait az alábbi fejezetben mutatom be.

A kutatási céloknak, kérdéseknek megfelelően kétféle mintával dolgoztam a teljes modellek létrehozása során, és ezeket bővítettem a későbbiekben. Mindkét esetben 500 elemű volt a minta. Az első mintát egyszerű véletlen mintavétellel készítettem el.

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

A legújabb, csődelőrejelzéshez használt 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 magas pontossággal előre jelezzék, mely ügyfél, hitelügylet fog nemteljesítővé válni.

Nemteljesítőnek azt a hitelügyletet minősítettem, mely rendelkezett mulasztási összeggel.

Az állítás alátámasztásához mindhárom módszerrel elkészítettem a klasszifikációs modellt. Az elemzés elvégzéséhez a KHR által átadott adatbázist használtam fel, és az eredmények validálása miatt a mintát egy tanuló és egy tesztelő részre osztottam. A tanuló mintába az elemek 70%-a került.

Logisztikus regresszió I.

Elsőként a logisztikus regressziós elemzést végeztem el. A rendelkezésre álló magyarázóváltozók közül kizárólag a törlesztőrészlet szerződéses összeghez viszonyított aránya bizonyult szignifikánsnak. Az Omnibusz teszt ($p < 0,001$) és a Hosmer and Lemeshow goodness-of-fit teszt ($p = 0,212$) alapján egy megbízható, megfelelően illeszkedő modell jött létre. A létrehozott modell közepes magyarázó erővel rendelkezik (Nagelkerke $R^2 = 38,8\%$).

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

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
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.

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

Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Tanuló	Default	0	261	62	80,8
		1	4	23	85,2
	Overall Percentage				81,1

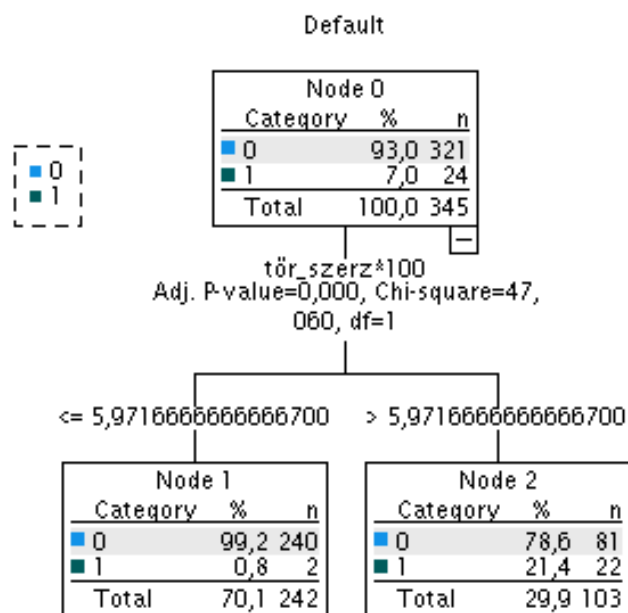
a. The cut value is ,039

Forrás: Saját szerkesztés

Összességében a létrehozott modell 81,1%-os pontossággal kategorizálta helyesen a hitelügyleteket, 66 elem besorolása volt helytelen. A tanuló és tesztelő minta esetében hasonló besorolási pontosság érhető el, azonban a jól teljesítő hitelek kategorizálása sokkal nagyobb arányban helyes. 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 másik mintán is elvégeztem az elemzést. A tanuló és tesztelő mintán létrejött modellek hasonló besorolási pontossággal rendelkeznek, így a létrejött modellt elfogadtam.

Döntési fa I.

Előjáróban is fontosnak tartom megemlíteni, hogy a döntése fa hátrányai közé tartozik, hogy hajlamos a túltanulásra, melynek veszélye jelen esetben is fennáll, hiszen a mintában túlnyomórészt jól teljesítő hitelek (93%) szerepelnek. A döntési fa megalkotásához ugyanazt a négy magyarázóváltozót alkalmaztam, mint a Logisztikus regresszió I. modell esetében, és ebben az esetben is a törlesztő részlet szerződéses összeghez viszonyított aránya bizonyult jól szétválasztó változónak az algoritmus alapján.



5. ábra: Döntési fa I. a tanuló minta esetében

Forrás: Spss output, saját szerkesztés

A létrejött döntési fa egy szintből áll, amely alapján elmondható, 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.

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

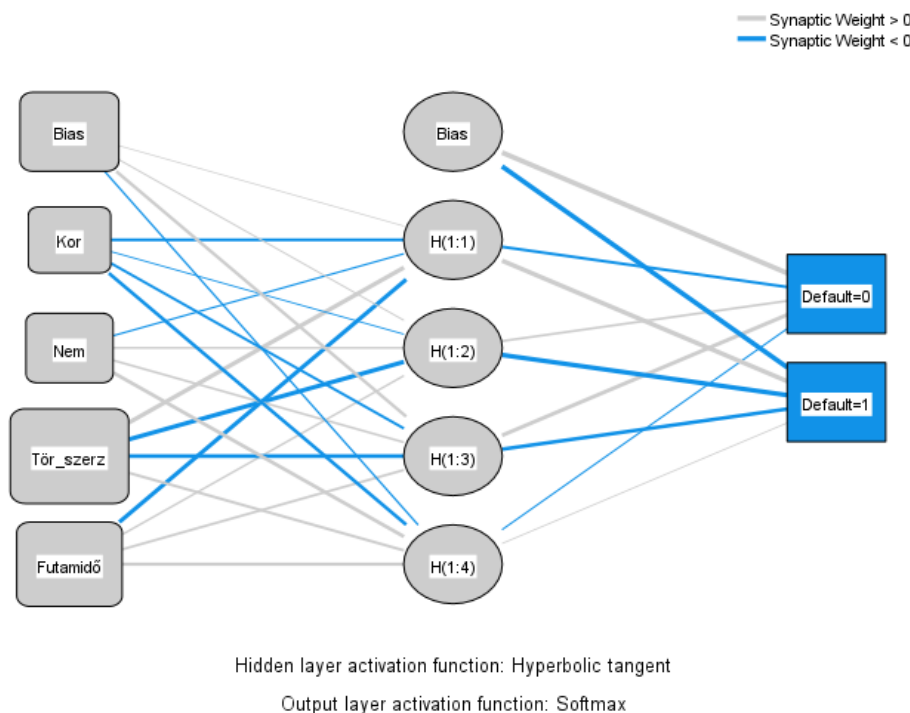
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	312	0	100,0%
	1	24	0	0,0%
	Overall Percentage	100,0%	0,0%	93,0%

Forrás: Saját szerkesztés

A tanuló adatbázis esetében 93,0%-os besorolási pontosságot ért el a modell, azaz az ügyletek 93,0%-t helyesen kategorizálta, 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, melyben a jól és rosszul teljesítő hitelek (megközelítőleg) azonos arányban fordulnak elő.

Neurális háló I.

Neurális háló esetében a csődelőrejelzés során elterjedt „Multi-Layer Perceptron” lehetőséget választottam. A neurális háló semmilyen előfeltétellel nem él, 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:



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

Forrás: Spss output, saját szerkesztés

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.

4. 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%

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.

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. 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.

Logisztikus regresszió II.

Az új mintán létrehozott modelltől elmondható az Omnibus teszt alapján, hogy egy megbízható modell jött létre ($p < 0,001$), 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 (Nagelkerke $R^2 = 69,3\%$) és a teszt (Nagelkerke $R^2 = 55,5\%$) minta esetében. Az előző esetben szignifikánsnak bizonyult változó, ebben az esetben is annak bizonyult ($p < 0,001$), továbbá bővült a modell a futamidő változóval.

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

Minta típusa	B	S.E.	Wald	df	Sig.	Exp(B)	
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,

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya.

x_2 : futamidő.

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

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

Minta típusa	Megfigyelt		Becsült		
			Default		Percentage Correct
			0	1	
Tanuló	Default	0	139	24	85,3
		1	32	142	81,6
	Overall Percentage				83,4

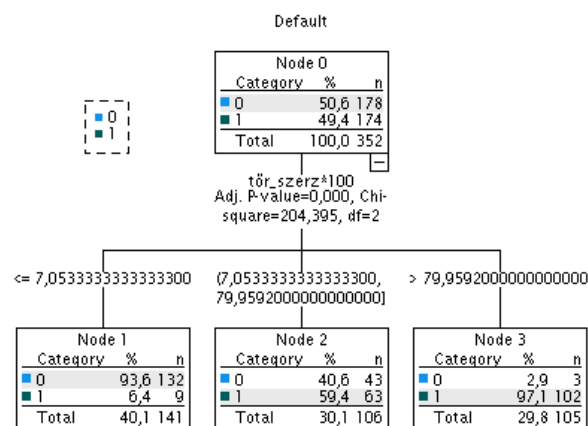
a. The cut value is ,390

Forrás: Saját szerkesztés

Összességében a létrehozott modell 83,4%-os pontossággal kategorizálta helyesen a hitelügyleteket, 56 elem besorolása volt helytelen. A klasszifikációs mátrix alapján megállapítható, hogy a tanuló minta esetében magasabb besorolási pontosságot ért el a modell, ami annak köszönhető, hogy a jól teljesítő hitelek a tanuló minta esetében nagyobb pontossággal azonosította a modell, azonban mivel ez nem a végleges modell, így ebben az esetben a validálást elfogadottnak tekintem.

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, ez a változó a törlesztő részlet szerződéses összeghez viszonyított aránya



7. ábra: Döntési fa II. a tanuló 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.

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

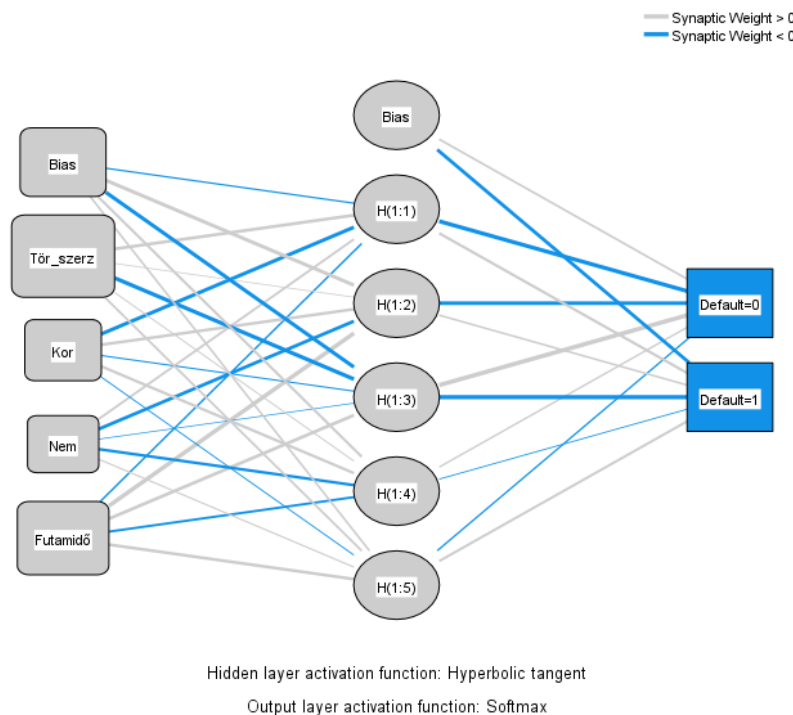
Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	132	46	74,2%
	1	9	165	94,8%
	Overall Percentage	40,1%	59,9%	84,4%

Forrás: Saját szerkesztés

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 bekategorizálni. Ebben az esetben viszont a nemteljesítő hitelek jelentős részét a megfelelő csoportba sorolta, továbbá a tanuló és a teszt (86,5%) mintán elért eredmény nem különbözik jelentősen egymástól, így a kapott eredményeket érvényesnek tekintem.

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.



8. 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 legjelentősebb változó a törlesztőrészlet szerződéses összeghez viszonyított aránya.

A besorolási pontosság ugyan összességében csökkent, azonban a specificitás és szenzitivitás értéke közelít egymáshoz, valamint a nemteljesítő hitelek besorolása jelentősen javult.

8. 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%

Forrás: Saját szerkesztés

A modell előrejelző képessége hasonló a teszt (86,2%) és a tanuló minta esetében, tehát a modell véglegesíthető.

Eredmények összehasonlítása

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

9. 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.

10. 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 20. táblázatban szereplő AUC és Gini-koefficiensek alapján minden modell kiemelkedőnek értékelhető, erős modell.

Az elemzések alapján a **H1 hipotézist elfogadom**, és az alábbi téziseket fogalmazom 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 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.

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

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

Logisztikus regresszió III.

Ezúttal a bővített adatbázison és az első mintán végeztem el az elemzést. 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:

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

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
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,

x_1 : a törlesztő részlet szerződéses összeghez viszonyított aránya,

x_2 : a kereset szerződéses összeghez viszonyított aránya,

x_3 : futamidő.

Az új változók bevonása kedvezően hatott a nemteljesítő hitelek besorolására, és a létrejött modell az ügyletek 83,1%-át kategorizálta helyesen.

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

Minta típusa	Megfigyelt		Becsült			Log. regr. I. értékei
			Default		Percentage Correct	
			0	1		
Tanuló	Default	0	267	56	82,7	80,8
		1	3	24	88,9	85,2
	Overall Percentage				83,1	81,1

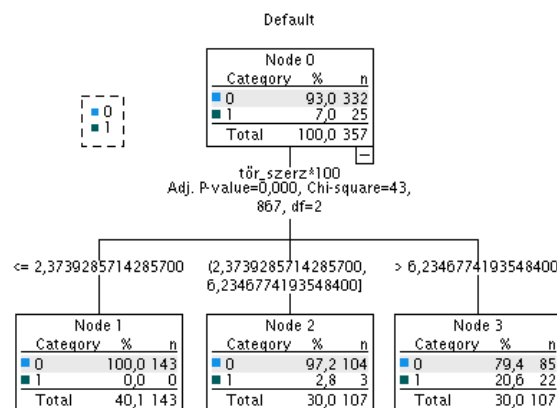
a. The cut value is ,068³

Forrás: Saját szerkesztés

A 12. számú táblázatban látható, hogy az új változók bevonásának hatására jelentősen megnőtt a nemteljesítő hitelek esetében a helyes kategorizálás aránya.

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.



9. ábra: Döntési fa III. a tanuló minta esetében

Forrás: Spss output, saját szerkesztés

³ Alacsony vágási pont, azonban a csödelőrejelzések során is volt erre példa, Ohlson (1989) esetében $p=0,038$, az általa használt mintában a csödbe jutott vállalatok aránya 4,85% volt.

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 létrejötté nem befolyásolja a modell besorolási pontosságát. A rosszul teljesítő hitelek közül egyet sem tudott az algoritmus helyesen kategorizálni.

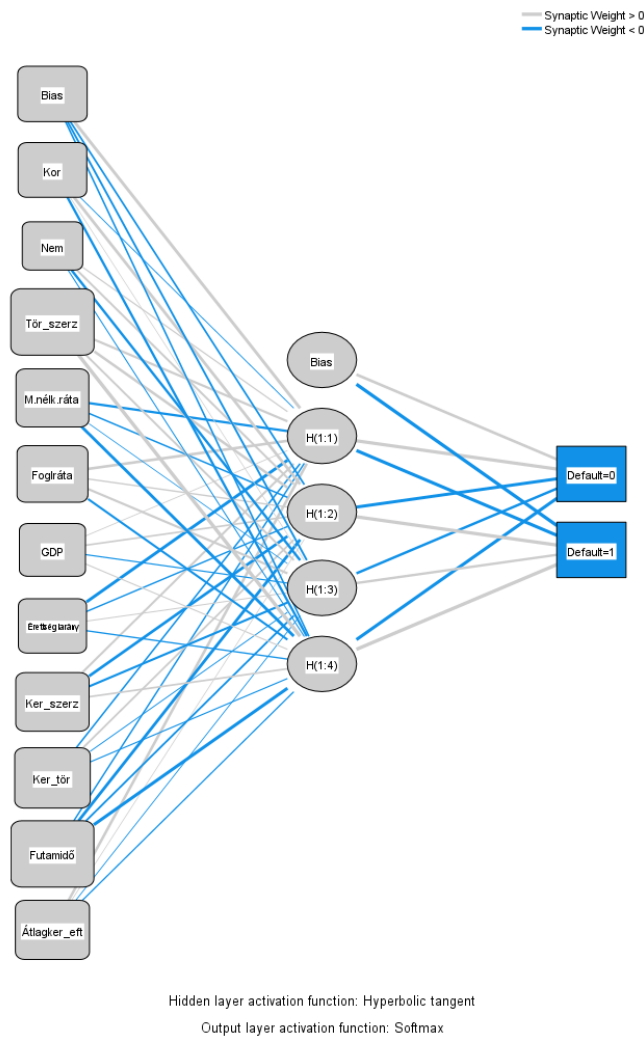
13. táblázat: Klasszifikációs mátrix a Döntési fa III. modell esetében

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%

Forrás: Saját szerkesztés

Neurális háló III.

A neurális háló esetében minden új változó beépült a modellbe, ezzel egy sokkal nagyobb neurális hálót létrehozva, mely 4 neuronnal rendelkezik.



10. 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.

14. 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%

Forrás: Saját szerkesztés

Az új modell besorolási pontossága növekedett a Neurális háló I.-hez képest az új változók hatására, az új modell az ügyletek 97,7%-t kategorizálta helyesen, azonban a specificitás és a szenzitivitás között még mindig jelentős különbség figyelhető meg.

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.

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

Minta típusa		B	S.E.	Wald	df	Sig.	Exp(B)
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,

x₄: 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 forintra nő, akkor átlagosan 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.

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

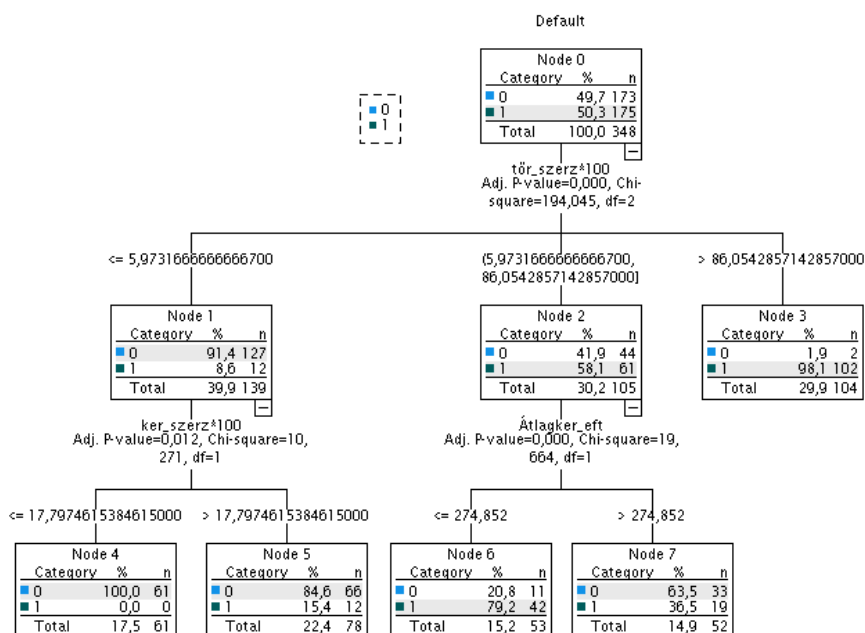
Minta típusa	Megfigyelt		Becsült			Logisztikus regresszió II.
			Default		Percentage Correct	
			0	1		
Tanuló	Default	0	131	32	80,4	85,3
		1	11	163	93,7	81,6
	Overall Percentage				87,2	83,4
a. The cut value is ,338						

Forrás: Saját szerkesztés

A teljes modell esetében javult a besorolási pontosság, és már az ügyletek 87,2%-a került helyes kategóriába. Mindemellett a nemteljesítő hitelek esetében is magasabb arány kategorizálta be a modell az ügyleteket helyesen, tehát az új változók figyelembevétele pozitív hatással volt a modellre.

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.



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

Forrás: Spss output, saját szerkesztés

A bal oldali belső csúcs esetében (Node 1) 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.

17. 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%

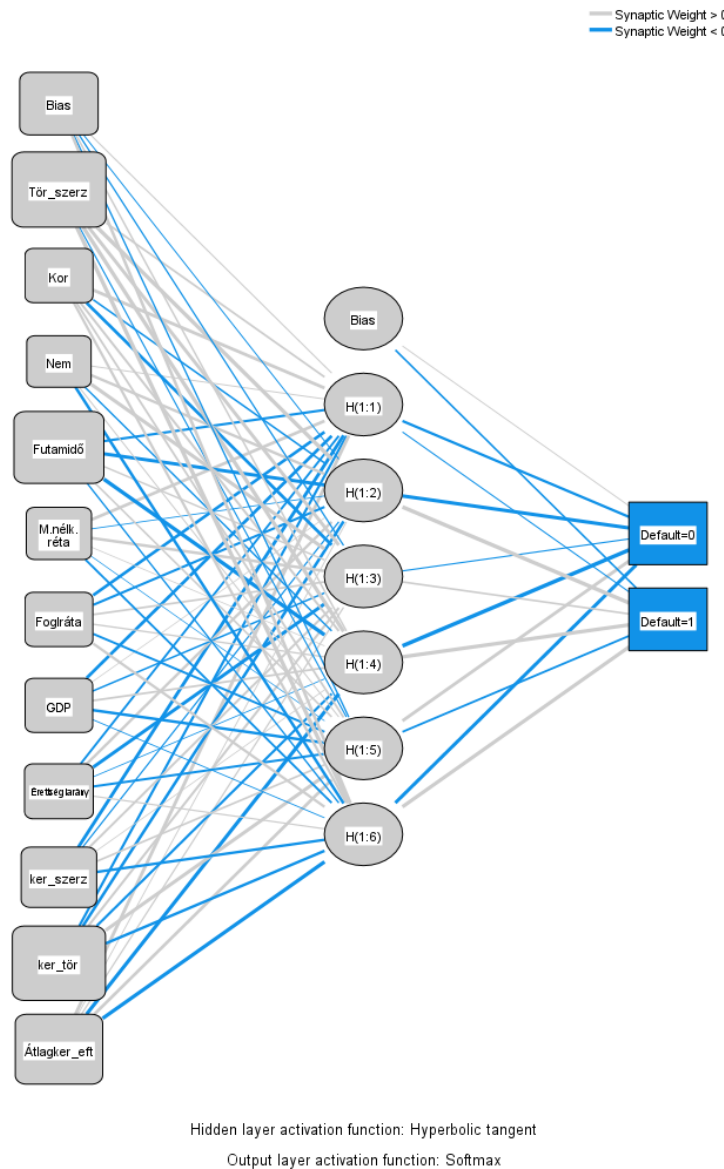
Forrás: Saját szerkesztés

A döntési fa II. modellhez képest javult a besorolási pontosság. A specificitás és szenzitivitás értékei közelebb kerültek egymáshoz, ez alapján ez utóbbi modellt kedvezőbbnek ítélem. A

tanuló és a teszt (83,6%) mintán elért eredmény nem különbözik jelentősen egymástól, így a kapott eredményeket érvényesnek tekintem.

Neurális háló IV.

A legösszettebb neurális háló modellnek a legutolsó bizonyult, 6 neuronnal.



12. 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.

A végleges modell magas (88,3%) besorolási pontossággal rendelkezik.

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

Minta típusa	Megfigyelt	Becsült		
		0	1	Percent Correct
Tanuló	0	161	14	92,0%
	1	26	142	84,5%
	Overall Percent	54,5%	45,5%	88,3%

Forrás: Saját szerkesztés

A besorolási pontosság közel azonos a tanuló és a tesztelő (83,4%) minta esetében, így a kapott modell véglegesíthető, és az ügyletek 88,3%-t kategorizálta helyesen a modell.

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.

19. 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

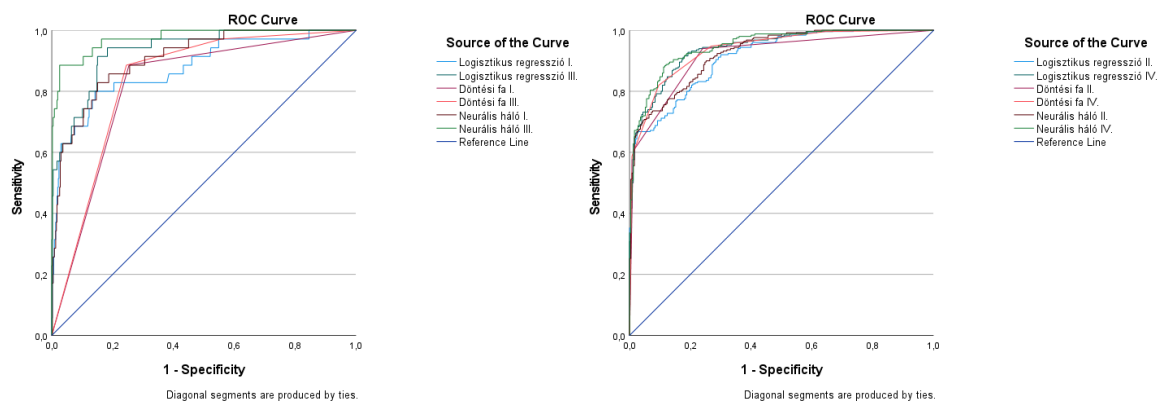
Forrás: Saját szerkesztés

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

A 19. 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 javítják 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. 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.

20. 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 érték⁴ 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 specificitás és szenzitivitás közötti különbség, és ezek együttesen a modellek teljesítményének javulásához vezettek.

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

⁴ ROC görbe alatti terület nagysága

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.

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. 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 vizsgálni. Mivel a kutatás során a fókusz csak a nemteljesítésen volt, és nem pedig annak nagyságán, így előbbi választottam, és keresztábra elemzéssel vizsgáltam meg, hogy megfigyelhető-e kapcsolat a régió, illetve a nem esetében. Mind a nem, mind pedig a régió esetében arra a következtetésre jutottam, hogy nincs szignifikáns kapcsolat ($p > 0,05$).

Ez 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 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.

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.

Egyéb hitelügylet

Akárcsak a teljes modell esetében, itt is a minta 30%-át alkotta a tesztelő mintát. A létrehozott neurális háló 6 neuronnal rendelkezik. A tanuló és tesztelő minta esetében hasonló besorolási pontosságot kaptam. 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. Az egyéb hitelügyletek esetében a kor is megjelent, mint lényeges magyarázó változó, ami a teljes modellben nem szerepelt az öt leglényegesebb változó között.

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. Tehát ebben az esetben javasolt lehet egy önálló modell alkalmazása.

Az algoritmus olyan változókat ítélt meg fontosnak, melyet egy korábbi modell sem, ez a GDP, illetve a munkanélküliségi ráta. Ennek hátterében 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.

Gépjármű lízing

Gépjármű lízing esetében azt tapasztaltam, hogy a rosszul teljesítő hitelek besorolása jobb, mint az alap modellnél, és a modell teljesítménye is jobb, tehát ennél a hitelnél is célszerű lehet a különálló modell alkalmazása.

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

Lakáscélú jelzáloghitel

Lakáscélú hitelek esetében egy kisebb neurális háló jött létre, mely két neuronnal rendelkezik. Ennek a modellnek a teljesítménye hasonló az alap modell teljesítményéhez, elenyésző különbség figyelhető meg csupán. A szenzitivitás értéke viszont jelentősen javult, ami kedvező.

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

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. Azonban az önálló modell esetében más változók bizonyultak fontosnak, melyek főként az adott megye gazdasági helyzetét jellemzik, ezek a foglalkoztatási ráta, a munkanélküliségi ráta és az 1 főre jutó GDP.

Személyi hitel

Az utolsó csoport pedig a személyi hitelek. A személyi hitelekkel kapcsolatosan 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 besorolási értékét. A magyarázó változók köre teljes mértékben megegyezik az alap modell magyarázó változóival.

Az alábbi táblázat foglalja és hasonlítja össze az egyes modellek és a teljes modell teljesítményét.

21. 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 érteke 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.

4.1 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.

5 Összefoglalás

Doktori kutatásom során a lakossági hitelek nemfizetését, illetve a nemfizetés előrejelzését vizsgáltam. Céloom 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 csödelőrejelzési modellek szolgáltak. Ezen tudományág megismeréséhez egy időbeli és módszertani áttekintést végeztem. Kiindultam a legelső, még egyváltozós csödelőrejelzési 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 háttérben 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.

6 Irodalomjegyzék

1. 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.
2. 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.
3. 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.
4. 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
5. Béres, D., Huzdik K. (2022): A fiatal felnőttek és a hitelek (kockázatkerülési csapda) Pénzügyi szemle, 2022/1 33-55
6. Blum, M. (1974): Failing company discriminant analysis. *Journal of Accounting Research*, Vol. Spring pp.1-21.
7. Deakin, E.B. (1972): A Discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 10-1 (Spring 1972) pp. 167-179
8. Dobák, P., Sági, J. (2005). Fogyasztási hitelek: növekvő eladósodottság? Forrás: Hitelintézeti Szemle: http://www.bankszovetseg.hu/Content/Hitelintezeti/51Dobak_Sagi.pdf
9. Fekete, O., Tatay, T. (2012): Hitelezők és adósok kapcsolatának szabályozási kérdései https://kgk.sze.hu/images/dokumentumok/kautzkiadvany2012/penzugy/fekete_tatay.pdf
10. 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.
11. 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/>
12. Hajdu, O. (2003): Többváltozós statisztikai számítások, Központi Statisztikai Hivatal, Budapest
13. 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
14. KHR éves tájékoztatók. (2011-2022). Forrás: <https://www.bisz.hu/dokumentumtar>
15. Ketskemény, L., Izsó, L., Könyves Tóth, E. (2011): Bevezetés az IBM SPSS Statistics programrendszerbe, Artéria Stúdió Kft, Budapest
16. Kovács, P., Révész, B., Ország, G. (2014). A pénzügyi kultúra és attitűd mérése. Forrás: www.eco.u-szeged.hu/download.php?docID=40014
17. Kristóf, T. (2002): A mesterséges neurális hálók a jövőkutatás szolgálatában. *Jövőelméletek*, 9. BKÁE Jövőkutatási Kutatóközpont, Budapest.
18. 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 [%]
19. 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>
 20. 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
 21. Malhotra, N. K. (2008): Marketingkutatás (Akadémiai Kiadó, Budapest)
 22. MNB (2023): Hitelezési folyamatok 2023. március
 23. MNB (2023): 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>
 24. 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
 25. Ohlson, J. (1980): Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research, Vol. 18. No. 1. 109-131. old.
 26. 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.
 27. Potóczyki, 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
 28. Sajtos L., Mitev A. (2007): SPSS kutatási és adatelemzési kézikönyv (Alinea Kiadó, Budapest)
 29. Szabó, D. M. (2021): A lakossági hitelezés fejlődés és a pénzügyi fogyasztóvédelem
 30. Vértesy, L. (2008): A pénzügyi intézmények finanszírozási tevékenységének jogi szabályozása Magyarországon Ph. D értekezés
 31. Virág, M. (2004): A csődmodellek jellegzetességei és története, Vezetéstudomány, 35. évf. 10. sz. 24-32. old.
 32. Virág, M., Kristóf, T. (2005): Az első hazai csődmodell újraszámítása neurális hálók segítségével. Közgazdasági Szemle, 52. évf. 2. sz. 144-162. old.
 33. 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.

7 A szerző publikációi

1. Fodor, K., & Varga, B. (2023). A pénzügyi kultúra és egyes társadalmi tényezők kapcsolata. In *What will our Future be Like? 2 essays in German, 7 in English, 30 in Hungarian language* (Német, angol és magyar nyelvű esszék) (pp. 191–199).
2. Kitti, F., & Beatrix, V. (2023). CAN HOMOGENEOUS GROUPS OF STUDENTS BE FORMED IN THE CASE OF FINANCIAL ATTITUDES? In *Building Up Financial Literacy* (pp. 35–48).
3. Fodor, K. (2023). Logistic Regression or Neural Network? Which Provides Better Results for Retail Loans? *THEORY METHODOLOGY PRACTICE: CLUB OF ECONOMICS IN MISKOLC*, 19(1), 53–62. <http://doi.org/10.18096/TMP.2023.01.05>
4. Beatrix, V., & Kitti, F. (2022). HAS THE FINANCIAL LITERACY OF STUDENTS CHANGED? In *Emerging Markets Economics and Business. Contributions of Young Researchers*. (pp. 105–108).
5. Varga, B., Fodor, K., Szilágyi, R., & S., G. A. (2022). A hallgatói lemorzsolódás jelensége és okai, tapasztalatok a Miskolci Egyetemen (The Phenomenon and Causes of Student Dropout, Experiences at the University of Miskolc). In “*Mérleg és Kihívások - Fenntarthatóság*” XII. Nemzetközi Tudományos Konferencia (pp. 604–613).
6. Varga, B., & Fodor, K. (2022). Változott-e az egyetemisták pénzügyi attitűdje? (Has the financial attitude of students changed?). In “*Mérleg és Kihívások - Fenntarthatóság*” XII. Nemzetközi Tudományos Konferencia (pp. 594–603).
7. Fodor, K., & Varga, B. (2022). Kritikus fontosságú nyersanyagok a történelemben és napjainkban. In *Doktoranduszok Fóruma, Miskolc, 2020. november 19.* (pp. 73–79).
8. Varga, B., & Fodor, K. (2021). Technological Innovation and the Critical Raw Material Stock. In *Data Analysis and Rationality in a Complex World* (pp. 321–329). http://doi.org/10.1007/978-3-030-60104-1_35
9. Beatrix, V., & Kitti, F. (2020). ECONOMIC STRENGTH – CRITICAL RAW MATERIAL STOCK. In *Proceedings of the Miskolc IPW- IV. Sustainable raw materials international project week* (pp. 268–276).
10. Szilágyi, R., Molnár, L., Lengyel, L., Fodor, K., & Tóthné, K. A. (2020). Munkaerő-piaci kompetencia-igény prognózis. In *Társadalmi és gazdasági folyamatok elemzésének kérdései a XXI. században* (pp. 61–79). <http://doi.org/10.14232/tgfe21sz.5>
11. Varga, B., & Fodor, K. (2020b). ANALYSIS OF CRITICAL RAW MATERIALS THROUGH THE EYES OF ECONOMISTS. *GEOSCIENCES AND ENGINEERING: A PUBLICATION OF THE UNIVERSITY OF MISKOLC*, 8(12), 94–104.
12. Varga, B., & Fodor, K. (2020a). A TECHNOLÓGIAI INNOVÁCIÓ ÉS A KRITIKUS NYERSANYAGOK. In *TÁRSADALMI, TECHNOLÓGIAI, INNOVÁCIÓS HÁLÓZATOK ASPEKTUSAI* (pp. 146–162).
13. Kitti, F. (2020). Using Multivariate Statistical Methods for Analysing Financial Literacy, as a Possible Appearance of Social Innovation. *THEORY METHODOLOGY PRACTICE: CLUB OF ECONOMICS IN MISKOLC*, 16(1), 11–18. <http://doi.org/10.18096/TMP.2020.01.02>
14. Fodor, K., & Dr. Varga, B. (2019b). The technology innovation and the critical raw material stock. In *16th Conference of the International Federation of Classification Societies, IFCS2019* (pp. 142–142).
15. Fodor, K., & Dr. Varga, B. (2019a). Kann die logistische Regression zur Klassifizierung kritischer Rohstoffe verwendeter werden? In *Society, Economics, Law Conference, Program, Abstracts* (pp. 23–23).

16. Kitti, F. (2019). MONEY OR LIFE? – THE FINANCIAL LITERACY OF STUDENTS. In „MÉRLEG ÉS KIHÍVÁSOK” XI. NEMZETKÖZI TUDOMÁNYOS KONFERENCIA *Ph.D. szekciók tanulmánykötete* (pp. 65–74).
17. Beatrix, V., & Kitti, F. (2019a). Development of Transport and Critical Raw Materials. In *Emerging Markets Economics and Business. Contributions of Young Researchers. Proceedings of the 10th International Conference of Doctoral Students and Young Researchers* (pp. 135–139).
18. Fodor, K., & S., G. A. (2019). Eladósodottság, mint a pénzügyi kultúra állapotának mutatója. In *Doktoranduszok Fóruma 2018: A Gazdaságtudományi Kar szekciókiadványa* (pp. 29–36).
19. Varga, B., & Fodor, K. (2019a). A FENNTARTHATÓ FEJLŐDÉS ÉS A KRITIKUS ELEMÉK. In “Mérleg és Kihívások” XI. Nemzetközi Tudományos Konferencia (pp. 393–403).
20. Lengyel, L., & Kitti, F. (2019). ANALYSIS OF EFFECTS ON LIFE EXPECTANCY AT BIRTH. In “Mérleg és Kihívások” XI. Nemzetközi Tudományos Konferencia (pp. 225–233).
21. Varga, B., & Fodor, K. (2019b). Electric Vehicles and Critical Raw Materials. *ANNALS OF THE UNIVERSITY OF ORADEA ECONOMIC SCIENCE*, 28(2), 444–455.
22. Beatrix, V., & Kitti, F. (2019b). Kann die logistische Regression zur Klassifizierung kritischer Rohstoffe verwendet werden? In *People and their values in the society* (pp. 15–25).
23. Kitti, F., Andrea, S. G., & Andrea, K. K. (2019). Financial Literacy and Indebtedness. In *MultiScience - XXXIII. microCAD International Multidisciplinary Scientific Conference*. <http://doi.org/10.26649/musci.2019.104>
24. Fodor, K., Lengyel, L., Molnár, L., Szilágyi, R., & Tóthné, K. A. (2018). Duális kompetencia igényfelmérés és prognózis.
25. Lengyel, L., & Fodor, K. (2018). Analysis of effects on life expectancy at birth. In *Multiscience XXXII. MicroCAD International Multidisciplinary Scientific Conference*.
26. Fodor, K. (2017). STATISTICAL TESTS OF MARKET EFFICIENCY. In „Mérleg és Kihívások” X. Nemzetközi Tudományos Konferencia = „Balance and Challenges” X. *International Scientific Conference* (pp. 778–787).