

BUDAPESTI CORVINUS EGYETEM

**DINAMIKUS PÉNZÜGYI MUTATÓSZÁMOK ALKALMAZÁSA A
CSŐDELŐREJELZÉSBEN**

DOKTORI ÉRTEKEZÉS

Témavezető: Prof. Dr. Virág Miklós, CSc

Nyitrai Tamás

Budapest, 2015

Nyitrai Tamás

Dinamikus pénzügyi mutatószámok alkalmazása a csődelőrejelzésben

Vállalkozások Pénzügyei Tanszék

Témavezető: Prof. Dr. Virág Miklós, CSc

Budapesti Corvinus Egyetem
Gazdálkodástani Doktori Iskola

**Dinamikus pénzügyi mutatószámok alkalmazása a
csődelőrejelzésben**

Doktori értekezés

Nyitrai Tamás

Budapest, 2015

TARTALOMJEGYZÉK

Tartalomjegyzék.....	5
Ábrák jegyzéke.....	8
Táblázatok jegyzéke.....	9
1. Bevezetés.....	10
1.1. A témaválasztás indokolása	10
1.2. A csődelőrejelzés fogalmi keretei	13
1.3. Az értekezés tartalma	18
2. Szakirodalmi áttekintés	22
2.1. A szakirodalmi áttekintés lehetséges módjai	22
2.2. A csődelőrejelzés szakirodalmának kronologikus áttekintése	23
2.3. A csődelőrejelzés szakirodalmának „keresztmetszeti” bemutatása	27
2.3.1. Az adatgyűjtés.....	29
2.3.1.1. A vizsgált vállalkozások tevékenységi köre	29
2.3.1.2. A mintavétel által lefedett időszak.....	33
2.3.1.3. A mintavétel földrajzi vonatkozásai.....	35
2.3.1.4. A dichotóm klasszifikáció problémái.....	35
2.3.1.5. A minta felosztása a két csoport között.....	38
2.3.2. Az adatelőkészítés	40
2.3.2.1. Számítási problémák	41

2.3.2.2. Az iparági sajátosságok figyelembevétele	41
2.3.2.3. Az adatok modellezésre történő előkészítése.....	44
2.3.2.4. A statikus csődmodellek kérdése	48
2.3.3. Változószelekció	50
2.3.3.1. A változószelekció nehézségei.....	50
2.3.3.2. A csődelőrejelzés további információforrásai.....	51
2.3.3.3. Változószelekciós módszerek	54
2.3.4. A csődmodell felállítása.....	58
2.3.4.1. Többváltozós statisztikai módszerek.....	60
2.3.4.2. Nemparaméteres, gépi tanulásra épülő eljárások	62
2.3.4.3. Módszerkombinációk és meta módszerek	73
2.3.4.4. Módszertani összehasonlító elemzések eredményei	77
2.3.5. A modell teljesítményének értékelése.....	80
2.3.5.1. A klasszifikációs teljesítmény mutatói.....	80
2.3.5.2. Validációs eljárások	84
2.4. A csődelőrejelzés hazai szakirodalma.....	89
2.4.1. A csődelőrejelzés kezdetei Magyarországon	90
2.4.2. Neurális hálók a hazai csődelőrejelzésben	90
2.4.3. A legkorszerűbb módszerek megjelenése a hazai csődelőrejelzésben.....	91
3. Kutatási hipotézisek	94
3.1. Nyitott kutatási kérdések a szakirodalomban.....	94
3.1.1. Dinamikus pénzügyi mutatók a csődelőrejelzésben	95

3.1.2. A kiugró értékek kezelése	98
3.2. Az értekezésben vizsgált kutatási hipotézisek	100
4. Az empirikus vizsgálathoz felhasznált adatok	102
4.1. Általános mintavételi kérdések	102
4.2. Az adatgyűjtés folyamata.....	104
5. Empirikus vizsgálat.....	111
5.1. Az első kutatási hipotézis vizsgálata.....	111
5.2. A második kutatási hipotézis vizsgálata	116
5.3. A harmadik kutatási hipotézis vizsgálata.....	119
5.4. Az empirikus vizsgálatok eredményeinek összegzése.....	126
6. Összegzés	128
6.1. Az értekezés hozzájárulása a tudományterület fejlődéséhez	129
6.2. Az értekezésben bemutatott elemzés korlátai	132
6.3. Lehetséges jövőbeli kutatási irányok	134
Függelék.....	137
Hivatkozások.....	139
A témában megjelent saját publikációk	158

ÁBRÁK JEGYZÉKE

1. ábra: A csődmmodell-építés folyamata	28
2. ábra: A mutatószámok eloszlásában bekövetkező változások hatása	42
3. ábra: A csődelőrejelzés klasszifikációs módszereinek csoportosítása	59
4. ábra: A döntési fák általános alakja	63
5. ábra: A neurális hálók általános felépítése	65
6. ábra: Az SVM módszer alapgondolata	68
7. ábra: Az SVM eljárás működése kétdimenziós esetben	69
8. ábra: Klasszifikációs mátrix	80
9. ábra: ROC görbe	82
10. ábra: A Cash flow/Adósság mutató időbeli alakulása.....	97
11. ábra: Egy kiugró értéket tartalmazó mutatószám-idősor	98

TÁBLÁZATOK JEGYZÉKE

1. táblázat: Módszertani összehasonlító elemzések eredményei a nemzetközi szakirodalomban	78
2. Táblázat: Az empirikus vizsgálatban felhasznált mutatószámok neve és számításmódja.....	107
3. Táblázat: A CHAID módszerrel felállított döntési fák találati aránya tízszeres keresztvalidáció alkalmazásával	118
4. Táblázat: Egy kiugró értéket tartalmazó mutatószám idősor	120
5. Táblázat: A kiugró értéket tartalmazó idősor értékei a helyettesítést követően...	123
6. Táblázat: Az empirikus vizsgálat keretében felállított modellek találati aránya .	125
F.1. Táblázat: Az empirikus vizsgálatban használt pénzügyi mutatók alapstatisztikai mutatói	137

1. BEVEZETÉS

Munkámban a PhD tanulmányaim során folytatott kutatásaim eredményeit foglalom össze. A doktori képzést a Budapesti Corvinus Egyetem Gazdálkodástani Doktori Iskolájának Vállalati pénzügyek specializációján folytattam Dr. Virág Miklós témavezetése mellett. Kutatási témám a csődelőrejelzés volt, mely napjainkban az egyik legfontosabb üzleti klasszifikációs probléma (Hu-Tseng [2007]). Dolgozatom bevezető fejezetében az alábbiakról olvashat:

- a témaválasztás indokolása;
- a kutatási terület fogalmi kereteinek ismertetése;
- az értekezés tartalmának bemutatása.

1.1. A témaválasztás indokolása

A vállalkozások létrejötte és megszűnése a gazdaság működésének természetes velejárója. A vállalkozások megszűnésére sor kerülhet önkéntesen és kényszerűen. Az első esetben a gazdálkodó egység tulajdonosai önként döntenek a tevékenység megszüntetéséről, rendezik a vállalkozás adósságait és felosztják az azt követően megmaradt vagyont. Erre a hazai jogrendben a végelszámolási eljárás intézménye biztosít lehetőséget.

A vállalati tevékenység megszűnésének azonban vannak kényszerű formái is. Ezekre akkor kerül sor, amikor az érintett gazdálkodó egység nem képes fizetési kötelezettségeinek rendezésére, ami miatt hitelezői felszámolási eljárást kezdeményezhetnek az adóssal szemben. Ebben az esetben bírósági felügyelet

mellett kerül sor az érintett vállalat kintlévőségeinek rendezésére. Amennyiben az adós vagyona erre nem nyújt elegendő fedezetet, a hitelezők elvesztik követeléseiket, vagy azok egy részét. Felszámolás esetén a tulajdonosokat biztosan kár éri, mivel ők a vállalatba fektetett tőke összegét, vagy annak egy részét veszítik el.

Fontosnak tartom azonban felhívni a figyelmet arra, hogy a vállalkozások fizetéseképtelensége és megszűnése nem csak a hitelezők és a tulajdonosok számára jelenthet költséget. Ilyen esetben kár érheti az érintett társaság vezetőségét és munkavállalóit, akik munkahelyüket veszíthetik el; illetve a fizetéseképtelenné vált vállalkozás vevőit és szállítóit is, akik üzleti partnereiktől kénytelenek megválni.

A gazdálkodó egység fizetéseképtelenségéből adódó kényszerű megszűnése a tevékenység valamennyi érintettje számára költséggel járhat, melynek összege nemzetgazdasági szinten is jelentős lehet. Emiatt a csődelőrejelzés az egyik legfontosabb terület a pénzügyi és számviteli szakirodalomban a Kim-Kang [2012] szerzőpáros véleménye szerint. Hasonlóképp vélekedik erről Shetty et al. [2012] is, akik szerint a nemzetgazdaságok számára az egyik legnagyobb fenyegetést az azokban működő vállalatok csődje jelenti, mivel a csődesemény nagyszámú előfordulása gazdasági és szociális problémákhoz vezethet.

Különböző mértékben ugyan, de szinte valamennyi nemzetgazdaság kénytelen volt szembesülni az előbb említett következményekkel a 2008-ban kezdődött világgazdasági visszaesés során. Egyes kutatók szerint e recesszió alapja az volt, hogy a hitelezők nem mérték fel kellő pontossággal ügyleteik kockázatát (Riberio et al. [2012]). Emiatt a pénzügyi problémák és a csőd előrejelzésének fontosságát a válság korábban soha nem látott magasságba emelte (Cao [2012a]).

Hasonlóképp vélekedik a csődelőrejelzés fontosságáról Chen et al. [2013] is, akik szerint a tudományterület ugyan már régóta kutatott, a válság azonban újra felhívta a tudományos érdeklődés figyelmét a még meg nem válaszolt kutatási kérdésekre. A recessziós gazdasági környezetben fontossá vált a potenciális csődvesztélyben lévő vállalatok minél hatékonyabb felismerése az érintett gazdálkodó egység csődjéből eredő költségek minimalizálása szempontjából.

A válság mellett a Bázel 2-es tőkeegyezmény bevezetése is új lendületet adott a csődelőrejelzés fejlődésének, ugyanis az egyezmény előírásai miatt a pénzügyi intézmények érdekelték lettek abban, hogy a lehető legpontosabb hitelscoreing technikát alkalmazzák belső minősítési rendszerükben (Brown-Mues [2012]), melyben a csődelőrejelző modellek kulcsfontosságú szerepet töltenek be. Ennek oka, hogy egy jó modell lehetővé teszi a hitelezők számára, hogy minimalizálják a szavatoló tőkében tartandó összeg nagyságát és növeljék a források kihelyezéséből származó profitjukat (Sánchez-Lasheras et al. [2012]).

A fentiek alapján megalapozottnak tűnik Wang-Ma [2012] azon megállapítása, amely szerint a csődelőrejelzés egy „forró téma” az üzleti tudományok területén, amely iránt az elmúlt 50 évben egyre intenzívebb az érdeklődés a szakirodalomban. Ennek háttérében a tudományterületet támogató informatika fejlődése, az adatokhoz történő egyre könnyebb hozzáférés és a még nyitott kutatási kérdések relatíve nagy számossága áll. Ezeket alapul véve esett az én választásom is erre a témakörre. A fenti változások ugyanis hazánkban is megfigyelhetők, az abból származó kutatási lehetőségek azonban máig kiaknázatlanok ezen a területen. Ezt a hiányt igyekeztem pótolni a csődelőrejelzés területén folytatott kutatásaim során.

1.2. A csődelőrejelzés fogalmi keretei

A tudományterület elnevezése nyelvtani értelemben egy összetett szó, amelynek mindkét tagját pontosan definiálni szükséges a dolgozat vizsgálati körének behatárolása érdekében.

Szószerinti értelmezésben a szakterület a gazdálkodó egységek csődjével foglalkozik. Ez a megállapítás azonban nem teljesen helytálló. Az angol nyelvű szakirodalomban éppúgy elterjedt a „pénzügyi nehézségek előrejelzése” (financial distress prediction), illetve az „üzleti kudarc előrejelzése” (business failure prediction) elnevezések használata is.

A szakterület hagyományos elnevezésének (csődelőrejelzés) problémája, hogy túlságosan leszűkíti a vizsgálat tárgyát a gazdálkodó egységek (tipikusan társas vállalkozások) csődjére. Az előbb bemutatott másik két elnevezés tágabban értelmezi az előrejelzés tárgyát. Ezek hátránya azonban a szubjektivitás, ugyanis az „üzleti kudarc”, illetve a „pénzügyi nehézségek” nem egzakt fogalmak, emiatt azok külön definiálásra szorulnak az egyértelműség érdekében.

A csődelőrejelzés szakirodalma nem tekinthető egységesnek az előrejelzés tárgya tekintetében. Mindössze abban mutatkozik konszenzus, hogy a szakterületen megjelenő publikációk célja olyan események bekövetkezésének előrejelzése, amelyek a vállalati tevékenység szempontjából releváns érdekcsoportok (jellemzően a hitelezők és a tulajdonosok) számára veszteséget okozhatnak.

Egy vállalkozás létesítésekor a tulajdonosok magánvagyonukat fektetik be és kockáztatják annak reményében, hogy befektetett tőkéjük megtérülése meghaladja a máshol elérhető hozamot. A hitelezők az általuk nyújtott kölcsön összegét veszíthetik el, ha az adós vállalkozás fizetéképtelenné válik. Ebben az esetben a tulajdonosokat

is kár éri, mivel a vállalkozás felszámolása esetén a hitelezők követeléseai megelőzik a tulajdonosok vagyonának felosztását a kielégítési rangsorban.

Abban az esetben, ha a vállalat tevékenységét önként szünteti meg, akkor végelszámolási eljárás keretében rendezik a társaság adósságait, majd ezt követően felosztják a megmaradt vagyont a tulajdonosok között. Ebben az esetben legfeljebb a vállalat tulajdonosait érheti veszteség, ha a megszűnéskor a vállalatba fektetett vagyon piaci értéke elmarad attól az összegtől, amelyet a tulajdonosok befektettek; ez a kár azonban „tervezhető”, ugyanis a tulajdonosok a vállalatba fektetés időpontjában vállalták magánvagyonuk esetleges elvesztésének kockázatát a máshol elérhető hozamhoz képest magasabb nyereség érdekében. Emiatt a vállalatok önkéntes megszűnésének előrejelzése nem képezi a csődelőrejelzés és így dolgozatom tárgyát sem.

A vállalatokkal szembeni csőd, illetve a felszámolási eljárás megindítása pénzügyi vonatkozású esemény, ugyanis mindkettő a vállalat fizetőképességéhez kötődik. Azok a vállalkozások, amelyek hosszabb ideig nem képesek fizetési kötelezettségeik teljesítésére, a csődeljárásról és a felszámolási eljárásról szóló 1991. évi XLIX. törvény 1. §-ának (1) bekezdése értelmében csődeljárás formájában fizetési haladékokat kérhetnek csődegyezség megkötése érdekében. A csődeljárás célja tehát, hogy a vállalat időt nyerjen fizetési kötelezettségeinek rendezésére, majd annak megtörténtét követően folytassa tevékenységét. A felszámolási eljárás célja, hogy a fizetéseképtelen adós jogutód nélküli megszüntetése során a hitelezők a törvényben meghatározott módon kielégítést nyerjenek.¹

¹ Forrás: A csődeljárásról és felszámolási eljárásról szóló 1991. évi XLIX. törvény 1. §-ának (3) bekezdése

Tekintettel arra, hogy mindkét előbb bemutatott eljárás megindításának alapja, hogy az adós nem teljesíti fizetési kötelezettségeit; a csődeljárás, illetve a felszámolási eljárás is jelentős kockázatot jelent a hitelezőknek követeléseik érvényesíthetősége; a tulajdonosoknak pedig tőkéjük megtérülése, illetve elvesztése szempontjából. Szélsőséges esetben az is bekövetkezhet, hogy a hitelezők és a tulajdonosok (egy része) kintlévőségének, illetve befektetett tőkéjének akár teljes összegét is elveszti az érintett vállalat fizetéseképtelensége miatt.

Értekezésemben az előrejelzés tárgyát a hazánkban bejegyzett társas vállalkozások fizetéseképtelenségének azon esetei képezik, amelyek miatt csőd, vagy felszámolási eljárás megindítására került sor. Meg kell jegyezni, hogy ezek kimenetele nem egyértelmű: a csődeljárás kezdeményezése nem szükségképpen jelenti az adós „túlélését”; továbbá a felszámolási eljárás végződhet a fizetéseképtelen vállalat fennmaradásával és a tartozások rendezésével is. Ebből adódóan a dolgozatban csődelőrejelzés alatt bizonyos értelemben fizetéseképtelenség-előrejelzést értek a továbbiakban.

Ez a meghatározás a „csődelőrejelzés” kifejezés szó szerinti értelmezéséhez képest egy tágabb értelmezés, egyúttal azonban korlátozást is jelent, mivel a csőd, illetve felszámolási eljárások megindítására jellemzően olyan vállalkozásokkal szemben kerül sor, amelyek kötelezettségeiknek már hosszabb ideje nem tesznek eleget, azaz már egy ideje fizetéseképtelenek. E szituáció végső következménye valamelyik eljárás megindítása. Ebből adódóan kimaradnak a vizsgálatból azok a vállalkozások, amelyek már egy ideje fizetéseképtelenek, de velük szemben még nem került sor valamely eljárás megindítására. Az ilyen vállalkozásokat tekinti a csődelőrejelzés szakirodalma pénzügyi problémákkal küzdő cégeknek (Gilbert et al. [1990]). Az e körbe tartozó vállalkozások jövőbeli fizetőképességének előrejelzését az nehezíti,

hogy helyzetükre vonatkozóan nem áll rendelkezésre objektív információ. A csőd, illetve felszámolási eljárás ténye a cégjegyzék adataiból hozzáférhető, azonban a „pénzügyi problémákra” vonatkozóan nincs egzakt definíció a szakirodalomban, ezáltal pedig e problémák fennállása sem azonosítható objektíven.

A csődelőrejelzés fogalmi kereteinek bemutatásánál szót kell még ejteni az „előrejelzés” szó kapcsán is. Módszertani szempontból a tudományterület egy dichotóm klasszifikációs probléma, melynek lényege, hogy minél pontosabban különbséget tudjunk tenni a vállalatok két csoportja: a fizetőképés és a fizetéseképtelenek között. A két csoportot az értekezésben működő, illetve csődös vállalatoknak is nevezem a továbbiakban.

Korábban utaltam rá, hogy a fizetéseképtelenség ténye objektíven csak a cégjegyzék adatai alapján állapítható meg, így a rövid ideje fizetéseképtelen vállalatok a működő vállalatok csoportjába soroltatnak. Az ilyen megfigyelések esetén ugyan már fennáll a fizetéseképtelenség ténye, de az még kérdéses, hogy végül megkezdődik-e valamely eljárás velük szemben. Ez attól függ, hogy az érintett vállalkozás képes-e megoldani rövid időn belül pénzügyi problémáit; illetve hitelezői mennyi ideig tudnak/hajlandók várni kintlévőségeik beérkezéséig. Az ilyen „átmeneti állapotban” lévő vállalkozások helyes besorolása különösen nehéz, ugyanis a pénzügyi problémák jelei már mutatkoznak pénzügyi adataikon, azonban még működő vállalatok és reális lehet az esélye annak, hogy továbbra is azok maradnak.

A fenti problémából kiindulva a szakirodalomban számos szerző a csődelőrejelző modelleket korai figyelmeztető rendszereknek (EWS: Early Warning System) nevezi. A hagyományos értelemben vett csődelőrejelző modellek a megfigyelt vállalatokat a csődös, vagy a működő vállalatok csoportjába sorolják, ami az előbb

említett előrejelzési nehézség miatt problematikus lehet. Az EWS modellek ugyanezt a klasszifikációt valósítják meg, a különbség csak annyi, hogy abban az esetben, ha egy adott vállalkozás a modell által például a csődös megfigyelésekhez soroltatott, akkor abból nem azt a következtetést vonják le, hogy a vizsgált vállalkozás biztosan fizetéseképtelenné fog válni, hanem azt, hogy a figyelembe vett adatai alapján jobban hasonlít a csődbe ment vállalatokhoz, mint a működőkhöz (Virág-Hajdu [1996]). Ez azt is jelenti, hogy ilyen esetben az érintett vállalkozás vonatkozásában jelentős a jövőbeli fizetéseképtelenség kockázata, amelyet a döntéshozóknak érdemes figyelembe venni. Ebből kiindulva Agarwal-Taffler [2007] a csődmodelleket a lázmérőhöz hasonló eszközöknek tekinti, amelyek nem mondják meg, hogy milyen betegségben szenved a beteg, azonban jelzést adnak a probléma fennállásáról és mértékéről, valamint felhívják a figyelmet a beavatkozás fontosságára; de nem jelzik előre egyértelműen a gyógyulást.

Ez a megközelítés tehát elhatárolódik a fizetéseképtelenség előrejelzésétől, mint kutatási céltól; azonban hátránya, hogy alkalmazásával nem lehet objektíven mérni a modellek teljesítményét, ami elfogadhatatlan a tudományos gyakorlatban. Emiatt értekezésemben magam is a hazai vállalatok jövőbeli fizetéseképtelenségének előrejelzését jelölöm meg kutatási célként – egyetértve egyúttal azzal a megközelítéssel, amely a modellek eredményeit előrejelzés helyett csak korai figyelmeztető jelzésnek tekinti a fizetéseképtelenség kockázatának vonatkozásában.

A csődelőrejelzés előbbiekben definiált célját csődmodellek felállításával érhetjük el, amelyek statisztikai összefüggéseket képeznek le a modellezéshez használt független változók és a fizetéseképtelenség ténye, mint függő változó között. Előbbiek körében hagyományosan a vállalkozások számviteli információs rendszeréből kinyerhető hányados típusú pénzügyi mutatókat alkalmazzák a kutatók és a gyakorlati

szakemberek egyaránt. Ebből adódóan magam is ezeket rátákat használom az értekezésben.

A csődelőrejelzésben vizsgálni kívánt kapcsolatrendszer korábban a matematikai-statisztika klasszifikációs eszközeivel, napjainkban pedig egyre inkább a mesterséges intelligenciára épülő adatbányászati módszerekkel próbálják feltárni a kutatók. E módszerek elméleti hátterét és működését az értekezés 2.3.4. alfejezetében mutatom be részletesebben.

Mindezek alapján a csődelőrejelzés a vállalati pénzügyek, illetve a statisztika (adatbányászat) határtudományának tekinthető, amely a pénzügyi mutatószámokat magyarázó változóként felhasználva tesz kísérletet a vállalatok jövőbeli fizetéseképtelenségének előrejelzésére valamely arra alkalmas többváltozós módszer alkalmazásával.

1.3. Az értekezés tartalma

A csődelőrejelzés tudományos vizsgálatának kezdetei az 1960-as évek második felére nyúlnak vissza; amikor amerikai tőzsdei vállalatok pénzügyi adatait kezdték vizsgálni abból a szempontból, hogy azok alapján van-e lehetőség a vizsgált társaságok csődjének előrejelzésére. A szakirodalomban még napjainkban is elterjedt gyakorlat a nyilvánosan működő vállalatok csődjének előrejelzése. Ennek oka, hogy a tőzsdei jelenlétből adódóan a kutatók számára az adatok könnyen és szabadon hozzáférhetőek; továbbá fontos szempont az ebbe a körbe tartozó vállalatok adatainak megbízhatósága is, ugyanis a nyilvános társaságok éves beszámolóit könyvvizsgálatnak kötelesek alávetni.

Újabb terület a csődelőrejelzésben a mikro, kis és közepes vállalkozások (KKV) csődkockázatának becslése, ahol külön kihívást jelent az adatok hozzáférhetősége a zártkörű működésből adódóan. Továbbá szót kell ejteni az adatok megbízhatóságának problémájáról is, ugyanis az e körbe tartozó társaságok általában nem kötelesek az éves beszámolóikban szereplő adatokat könyvvizsgálatnak alávetni. Ennek ellenére a hitelezők részéről éppúgy felmerül az igény a kisebb vállalkozások csődkockázatának értékelése szempontjából is (Riberio et al. [2012]).

A KKV-k forráshoz jutása makrogazdasági szempontból is kiemelt jelentőségű, ugyanis e vállalati kör a legtöbb országban – hazánkban pedig kiemelten – fontos szerepet játszik a foglalkoztatás, a jövedelemtermelés, az innováció és a növekedés szempontjából (Koyuncugil-Ozgulbas [2012]). E vállalkozások jövőbeli fizetőképességének előrejelzésére pedig a nyilvános társaságok minősítésénél használt modellek nem feltétlenül alkalmazhatóak.

Tekintettel arra, hogy Magyarországon a tőzsdei cégek aránya a működő vállalatok teljes számához viszonyítva meglehetősen alacsony, a zártkörűen működő társaságok jövőbeli fizetőképességének előrejelzése a tudományterület hazai tudományos kutatása és gyakorlati alkalmazása szempontjából kiemelt jelentőséggel bír. A magyar tőzsde méretéből adódóan nincs lehetőség a nyilvánosan működő vállalatok adatai alapján csődmodellek felállítására és használatára; a zárt társasági formában tevékenykedő társaságok esetén pedig 2009-ig az adatok hozzáférhetősége képezett jelentős akadályt a magyar vállalkozások csődjének tudományos vizsgálata előtt.

2009-től kezdődően a Magyarországon bejegyzett társaságok kötelesek éves beszámolóikat elektronikus formában visszamenőlegesen is közzétenni a Közigazgatási és Igazságügyi Minisztérium Céginformációs és az Elektronikus

Cégeljárásban Közreműködő Szolgálatánál, amely azokat honlapján² bárki számára hozzáférhetővé és kereshetővé teszi a 2000. üzleti évig visszamenőleg.

PhD tanulmányaim során folytatott kutatásaim alapját az adatokhoz történő szabad hozzáférés képezi, amely lehetőséget nyújt a hazai vállalkozások fizetéseképtelenségének – a korábbi magyar kutatásokhoz képest szélesebb körű adatbázison alapuló – tudományos vizsgálatára. Az elmúlt években folytatott kutatómunka eredményeit foglalom össze a következő fejezetekben.

Doktori tanulmányaim során kiemelt figyelmet szenteltem a csődelőrejelzés témakörében megjelent hazai és nemzetközi publikációk áttekintésének, ugyanis az eddig elért tudományos eredmények ismeretében van csak lehetőség releváns kutatási kérdések megfogalmazására. A szakirodalom bemutatása a csődelőrejelzésben jellemzően kronologikus rendező elv szerint történik. Értekezésemben azonban kísérletet teszek a szakterületen eddig elért tudományos eredmények újszerű bemutatására, melynek lényege, hogy a témakörben megjelent publikációkat a csődmodell-építés egyes fázisaihoz rendelem. Ennek alapját az adja, hogy mára a témakör szakirodalma oly mértékben kiterjedt, hogy a modellépítés egyes fázisai önálló kutatási területnek is tekinthetők. Ez az újszerű megközelítés véleményem szerint egyszerre ad átfogó képet a csődelőrejelzés egyes részterületeinek aktuális kutatási kérdéseiről, valamint azok fejlődéstörténetéről egyaránt.

A feldolgozott szakirodalom alapján megfogalmazom a vizsgálni kívánt kutatási kérdéseket. Az azokra adott válaszok reményeim szerint túlmutatnak majd az elmúlt

² <http://e-beszamolo.kim.gov.hu/kereses-Default.aspx>

közel 50 évben a szakterületen eddig elért eredményeken és újabb jövőbeli kutatási kérdéseket vetnek fel.

A következő fejezetben bemutatom a csődelőrejelzés szakirodalmának legfontosabb hazai és nemzetközi publikációit és azok kutatási eredményeit. Az ott tárgyaltakra alapozva a harmadik fejezetben megfogalmazom a vizsgálni kívánt kutatási hipotéziseket, majd a negyedik fejezetben bemutatom az ezek vizsgálata céljából összeállított adatbázist, részletesen ismertetve az adatgyűjtés módját is. Az ötödik szakaszban az empirikus vizsgálatokat, valamint azok eredményeit fejtem ki. A hatodik, záró fejezet összegzi az értekezésben bemutatott elemzések eredményeit, az azokból levonható következtetéseket, továbbá a lehetséges jövőbeli kutatási irányokat.

2. SZAKIRODALMI ÁTTEKINTÉS

A fejezet a csődelőrejelzés szakirodalmát kívánja röviden bemutatni. Bár terjedelmét tekintve ez a dolgozat legnagyobb része, ennek ellenére fel kell hívni a figyelmet arra, hogy a szakirodalom bemutatása nem tekinthető teljesnek. A témában megjelenő publikációk száma az utóbbi években rendkívüli ütemben növekszik, így terjedelmi korlátok miatt nincs lehetőség az egyes vizsgálati szempontok valamennyi kutatási kérdése tekintetében az összes empirikus vizsgálat és azok eredményeinek részletes bemutatására. Emiatt a dolgozatban felvetett kutatási kérdések mindegyike esetén csak néhány „reprezentatívnak” mondható kutatás eredményére térek ki tételesen, míg a hasonló célzatú vizsgálatokra vonatkozóan csak hivatkozás szintjén utalok – ezzel ajánlva az Olvasó figyelmébe azokat a műveket, amelyekben hasonló kutatási kérdést elemeztek, illetve ahol további, a dolgozatban nem részletezett empirikus vizsgálatokra történő hivatkozások is megtalálhatók.

2.1. A szakirodalomi áttekintés lehetséges módjai

A csődelőrejelzés fontosságát a bevezetésben hivatkozott szerzők idézetei mellett az is igazolja, hogy mára a szakterületen megjelenő publikációk számossága olyan nagy, hogy külön tanulmányok foglalkoznak a már megjelent folyóiratcikkek legfontosabb gondolatainak összefoglalásával. Abdou-Pointon [2011] a hitelscoreing modellek szakirodalmának aktuális helyzetét mutatják be, melyben felhívják a figyelmet a kutatási terület kihívásaira és javaslatot tesznek a lehetséges jövőbeli kutatási irányokra is. Hasonlóképp külön tanulmányokat szentelnek a csődelőrejelzésben alkalmazható módszerek bemutatására (lásd például Jones [1987]

munkáját); illetve azok alkalmazásának lehetséges problémáira (Balcean-Ooghe [2006]).

A csődelőrejelzés szakirodalmának bemutatását a kutatók többsége jellemzően történeti kontextusba helyezi. A témakörben mérföldkönek számító írások kronologikus bemutatása adekvát ugyan, de a szakterületen megjelenő publikációk számának dinamikus növekedése miatt ez a megközelítés véleményem szerint egyre kevésbé alkalmas arra, hogy az Olvasóban a kronologikus szakirodalmi áttekintés alapján teljes kép alakuljon ki a csődelőrejelzés tudományterületének jelenlegi állásáról és annak legfőbb kutatási kérdéseiről.

Ezt a célt sokkal inkább szolgálhatja az a megközelítés, amelyik a szakirodalmi eredményeket a csődmodellek építésének egyes fázisai köré csoportosítja, ugyanis napjainkban a publikációk jellemzően csak egy-egy részterületéhez kötődően fogalmaznak meg kutatási kérdéseket – szemben a csődelőrejelzés úttörő munkáival, amelyek a modellépítés folyamatát még „globálisan” szemlélték.

A fenti okok miatt munkámban a feldolgozott szakirodalmat a csődmodellek építésnek egyes szakaszaihoz rendelve mutatom be. Tekintettel azonban arra, hogy a publikációs gyakorlatban még napjainkban is a történeti áttekintés domináns, röviden időrendben is bemutatom a csődelőrejelzés fejlődéstörténetének legfontosabb mérföldköveit, valamint a tudományterület fejlődésének főbb motivációit.

2.2. A csődelőrejelzés szakirodalmának kronologikus áttekintése

A vállalatok jövőbeli fizetőképességével kapcsolatos bizonytalanság már a 20. század első felében is fontos szerepet játszott az üzleti élet szereplőinek döntéseiben.

Az első tudományos módszertannal végzett elemzést az 1960-as évek közepén Beaver [1966] publikálta. Munkájában a leggyakrabban használt pénzügyi-számviteli mutatószámokat hasonlította össze az amerikai tőzsdén jegyzett iparvállalatok adatai alapján. Harminc pénzügyi mutatószám előrejelző képességét vetette össze a csődöt megelőző öt éves időszak vonatkozásában. Az egyes évekre külön elvégzett vizsgálat eredményei alapján az idézett szerző számos olyan következtetést vont le, melyek máig meghatározó szerepet játszanak a csődelőrejelzés tudományos kutatásában. Ebből adódóan szinte alig találni olyan publikációt, amely ne említene meg ezen úttörő munkát a szakirodalmi áttekintésben.

Beaver munkássága nem csak amiatt figyelemre méltó, mert az elsők között vizsgálva a vállalatok jövőbeli fizetőképességének előrejelezhetőségét lerakta a tudományterület tudományos vizsgálatának alapjait, hanem azért is, mert megállapításainak többsége napjainkig megállja a helyét, továbbá az idézett mű számos olyan kutatási irányt vet fel, amelyek vizsgálata tovább növelheti a csődelőrejelző modellek pontosságát. Erre szolgál példaként saját kutatásom is, melynek legfontosabb hipotézisét szintén az imént idézett szerző kutatási eredményeire alapozva fogalmaztam meg.

A Beaver [1966] munkájában alkalmazott módszert napjainkban egyváltozós diszkriminancia-analízis néven ismeri a szakirodalom. A módszer lényege, hogy egyidejűleg csak egyetlen pénzügyi mutató értékét veszi alapul annak eldöntésére, hogy egy kérdéses vállalkozás a jövőben fizetőképessé lesz-e. E módszert a későbbiekben azonban élesen bírálták annak tudománytalansága miatt, továbbá problémát jelentett, hogy az a vállalat, amelyik egy adott mutató tekintetében csődösnek minősül, egy másik vonatkozásában nem feltétlenül; emiatt a módszer

alkalmazásakor könnyen adódhatnak ellentmondások a különböző mutatók között, amire maga a szerző is felhívta a figyelmet munkájában.

Az imént említett problémák kiküszöbölését teszi lehetővé a többváltozós matematikai-statisztikai módszerek alkalmazása, melyek közül először a többváltozós diszkriminancia-analízis jelent meg a csődelőrejelzés területén Altman [1968] világhírűvé vált munkájában, mely napjainkban már az egyetemi vállalati pénzügyek kurzusok tankönyveiben is megtalálható (Brealey-Myers [2005]). A módszer elméleti hátterét az értekezés 2.3.4. alfejezete ismerteti részletesebben.

Az alkalmazott módszer annak köszönhetné népszerűségét, hogy választ tudott adni a csődelőrejelzés három alapvető kérdésére (Virág [2004]):

- mely mutatószámok bírnak statisztikailag szignifikáns diszkrimináló erővel a csődös és a működő vállalatok csoportjai között;
- milyen súllyal célszerű e változókat figyelembe venni a modellekben;
- hogyan lehet ezeket a súlyokat objektív módon meghatározni.

Az Altman [1968] által felállított modell 95 %-os pontossággal volt képes a 33 működő és 33 csődös amerikai iparvállalatot tartalmazó mintában a megfigyelések helyes besorolására, melynek köszönhetően az 1980-as évekig a diszkriminanciaanalízis lett a domináns módszertani megoldás a csődelőrejelzésben (lásd például Deakin [1972] és Altman et al. [1977] munkáit).

Altman [1968] modelljének népszerűségét az is jelzi, hogy még napjainkban is összehasonlító elemzések tárgyát képezi (lásd például Altman [2002], Philosphov-Philosphov [2002], Hillegeist et al. [2004], Neves-Vieira [2006], Parnes [2007], Agarwal-Taffler [2008], Sueyoshi-Goto [2009] munkáit), melyek több esetben

meglepően jó előrejelző képességről számolnak be a közel 50 évvel ezelőtt publikált modell kapcsán.

A 70-es évek végétől kezdődően napjainkig a tudományterület fejlődését a klasszifikációs feladatok megoldására alkalmas matematikai-statisztikai eljárások, valamint az azt támogató informatikai háttér korszerűsödése dominálja. A logisztikus regressziót a csődelőrejelzésben először Martin [1977] alkalmazta. Munkáját követően ez a módszer vált általánossá annak köszönhetően, hogy alkalmazása nem követeli meg a diszkriminancia-analízis azon szigorú statisztikai feltevéseinek teljesülését, mint a változók normális eloszlása, vagy a vizsgált csoportok variancia-kovariancia mátrixainak egyezősége (Ohlson [1980], Zavgren [1985]).

Hasonló okok miatt kezdték alkalmazni a logisztikus regresszió mellett a probit regressziót (Zmijewski [1984], Sjøvoll [1999], Bongini et al. [2000], Bernhardsen [2001], Grunert et al. [2005]), valamint a döntési fákat (Frydman et al. [1985]). Utóbbiak előnye, hogy semmilyen statisztikai előfeltevés teljesülését nem igénylik a módszer alkalmazásához, így népszerűségük napjainkig töretlen.

Az idegrendszer működési elvét szimuláló neurális hálókat Odom-Sharda [1990] alkalmazta először a csődelőrejelzésben. Munkájuk azonban nem csak ebből a szempontból tekinthető mérföldkőnek a csődelőrejelzés fejlődéstörténetében. A neurális hálók megjelenésével ugyanis a hagyományos matematikai-statisztikai módszerek (diszkriminancia-analízis, logisztikus regresszió), valamint a döntési fák mellett lehetséges alternatívaként megjelent a mesterséges intelligencia alkalmazásának lehetősége is, mely a kutatókban felvetette a kérdést: melyik módszer tekinthető a leghatékonyabbnak a csődelőrejelzésben?

A kérdésre még napjainkban sincs egyértelmű válasz a szakirodalomban. A kutatási eredmények ugyanis azt mutatják, hogy egyes esetekben az egyik, más esetekben pedig egy másik módszer hoz jobb eredményt (Oreski et al. [2012]). Marqués et al. [2012a] is arra a következtetésre jutott, hogy nagy valószínűséggel nem létezik egyetlen „legjobb” módszer. Ennek ellenére a 90-es évek kezdetétől a módszertani összehasonlító tanulmányok dominálják a csődelőrejelzés szakirodalmát a „legjobb” módszer után kutatva. Tekintettel az e témakörben megjelenő publikációk jelentős számára (Sánchez-Lasheras et al. [2012]) ez a kutatási terület tekinthető a tudományterület vizsgálatában a főáramnak.

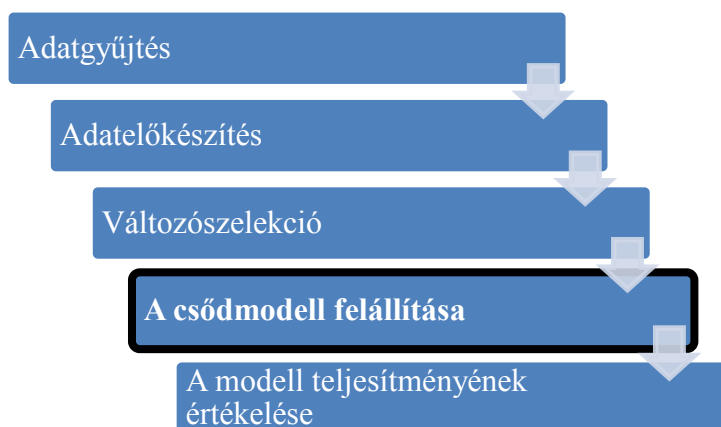
Fontos azonban hangsúlyozni, hogy a főáram mellett léteznek alternatív kutatási irányok is a csődelőrejelzésben, amelyek a modellek megbízhatóságát nem újabb klasszifikációs módszerek alkalmazásával kívánják javítani. Az e körbe sorolható publikációk száma azonban jóval kisebb, továbbá ezek jellemzően „szétszórtan” jelentek meg a csődelőrejelzés fejlődéstörténetének eddigi 50 évében. Ebből adódóan kevésbé illeszthetők be abba a folyamatos trendbe, amely a klasszifikációs módszerek korszerűsödése által dominált főáram fejlődését jellemzi. Emiatt célszerűnek tűnik a tudományterület fejlődését egy másik, „keresztmetszeti” aspektusból szemlélni, amelyben a csődelőrejelzés főárama mellett megjelennek a hasonlóan fontos szerepet betöltő egyéb kutatási irányok is.

2.3. A csődelőrejelzés szakirodalmának „keresztmetszeti” bemutatása

Munkámban ezt a megközelítést tartom célszerűbbnek a kronologikus megoldással szemben, ugyanis a „keresztmetszeti” aspektusnak köszönhetően véleményem szerint az Olvasó átfogóbb képet kaphat a csődmodell-építés folyamatáról, annak

szakaszairól, az azokhoz kötődő legfontosabb kutatási kérdésekről, illetve az ott elért eredményekről. Ezt a folyamatot szemlélteti az 1. ábra.

1. ábra. A csődmodell-építés folyamata



Az ábra a modellépítés legfontosabb fázisait szemlélteti a gyakorlati megvalósítás időbeli sorrendjében. A csődelőrejelzés fejlődéstörténetében született valamennyi publikáció hozzárendelhető az 1. ábrán szereplő fázisok valamelyikéhez³ annak függvényében, hogy a szerzők miben határozták meg tanulmányuk célját. Az ábrán a szakterület főáramának tekinthető módszertani összehasonlító kutatások a fekete körvonallal kiemelt „A csődmodell felállítása” névvel illetett kategóriába sorolhatók. Szintén e körbe tartozik a különböző adatbányászati módszerek paramétereinek optimalizációja is.

Fontos megjegyezni, hogy a csődelőrejelzés szakirodalmi áttekintésének a modellépítés egyes fázisai szerint történő csoportosítása nem igényli a kronologikus sorrendről történő lemondást. Az egyes fázisokhoz kötődő publikációkat is célszerű

³ A kiterjedtebb elemzések természetesen egyszerre több fázishoz is hozzárendelhetők.

lehet időrendi sorrendben tárgyalni annak érdekében, hogy az egyes részterületek szakirodalmának fejlődéstörténete is kirajzolódhasson a főáramba tartozó kutatási eredmények ismertetése mellett.

2.3.1. Az adatgyűjtés

Az 1. ábrán a csődmódel-építés első fázisaként az adatgyűjtés feladata szerepel. E körben két fő aspektust kell figyelembe venni: a megfigyelések számát és az azok jellemzésére használt mutatók körét. Alapvető követelmény, hogy elegendő megfigyelés álljon rendelkezésre a csődös és működő vállalatok vonatkozásában egyaránt ahhoz, hogy statisztikai modellek felállítására legyen lehetőség. E nyilvánvaló követelmények teljesülése mellett azonban a szakirodalomban számos olyan egyéb kutatási irány is körvonalazható, amely az adatgyűjtés lépéséhez kapcsolható. Ezek közül a legfontosabbakat ismertetem röviden a következő alfejezetekben.

2.3.1.1. A vizsgált vállalkozások tevékenységi köre

Lényeges szempont lehet a mintába került vállalkozások tevékenységi köre. A pénzügyi szektorban működő vállalkozásokat jellemzően kihagyják a csődmódellek adatbázisából azok eltérő sajátosságaira hivatkozva. Ezekre vonatkozóan specializált csődmódellek szükségesek, amelyek elkülönült szakirodalommal rendelkeznek (lásd például korábban Pantalone-Platt [1987a], Tam-Kiang [1992], Sarkar-Sriram [2001], valamint Kolari et al. [2002] munkáit; újabban pedig Ruzgar et al. [2008], Boyacioglu et al. [2009], illetve Quek et al. [2009] tanulmányait). Kivételként említhető Jones-Hensher [2004] és Duan et al. [2012] munkája, akik a pénzügyi

ágazatban működő vállalatokat sem hagyták ki a csődmodellek felállítására szolgáló adatbázisból.

A 2008-ban kezdődött gazdasági válság negatív hatásainak számos pénzügyi intézmény áldozatul esett világszerte. Ebből adódóan a szabályozó hatóságok részéről fokozódott az igény a pénzügyi intézmények csődjét előrejelző modellek aktualizálása iránt. A válság idején fizetéseképtelenné vált bankok adatain készített csődmodellt Ioannidis et al. [2011], valamint Serrano-Cinca – Guittérez-Nieto [2013]. Japánban Harada et al. [2013] tett kísérletet a pénzügyi intézmények csődjének előrejelzésére. A biztosítók csődjét pedig Segovia-Vargas et al. [2003] és Salcedo-Sanz et al. [2004] próbálta modellezni.

A tudományterületen elterjedt klasszifikációs módszerek azonban nem csak a fizetéseképtelenség előrejelzésére használhatók. Jagtiani et al. [2003] például a bankok pénzügyi mutatóit alapul véve annak előrejelzésére tett kísérletet, hogy a vizsgált hitelintézetek tőke-megfelelőségi mutatójának értéke eléri-e a következő évben a szabályozó hatóság által meghatározott mértéket. Bellotti et al. [2011] a Fitch bankminősítéseinek előrejelezhetőségét vizsgálta. Feki et al. [2012] pedig tunéziai bankok nem teljesítő követeléseinek arányát próbálta becsülni a csődelőrejelzésben elterjedt módszerek alkalmazásával.

A szakirodalomban konszenzus mutatkozik abból a szempontból, hogy a pénzügyi szektorban tevékenykedő vállalatok csődvészélyét iparág specifikus modellek segítségével szükséges előrejelezni. Más termelő és szolgáltató ágazatok tekintetében azonban már nem tapasztalhatunk ilyen egységes álláspontot. Egyes szerzők a mintavétel során csak egyetlen, vagy egymáshoz hasonló jellegű tevékenységi körbe tartozó vállalkozások sokaságát veszik alapul, míg mások kevésbé vannak tekintettel

az iparági hovatartozásra. Ezt a kutatási kérdést vizsgálta Gaganis et al. [2007]. Munkájukban arra a kérdésre keresték a választ, hogy szükséges-e minden iparágra külön csődmodellt felállítani, vagy elegendő egy általános modell az iparági hovatartozás figyelembevétele nélkül. Empirikus eredményeik ellentmondásosak voltak, így nem tudtak egyértelmű következtetést levonni a feltett kutatási kérdés vonatkozásában. Chava-Jarrow [2004] szintén ezzel a kérdéssel foglalkozott. Az általuk vizsgált vállalatok vonatkozásában csak kis mértékben javult a modellek előrejelző képessége abban az esetben, amikor a mintában szereplő vállalatok tevékenységi körét dummy változók formájában beépítették a csődmodellekbe. Neophytou-Mar Molinero [2005] az eddigiekkel szemben szignifikáns javulást tapasztalt az iparági hovatartozás figyelembevétele esetén, amelyből azt a következtetést vonták le, hogy különböző iparágak tekintetében különböző csődmodellekre van szükség. Ezt az eredményt erősíti Glennon-Nigro [2005] kutatása is, melyben azt tapasztalták, hogy a kereskedelemmel foglalkozó vállalkozások jellemzően nagyobb csődkockázatot mutatnak, mint a szolgáltatói szektorban tevékenykedők.

Rösch [2003] Németországban nemzetgazdasági szinten vizsgálta az egyes ágazatokban a csődbe ment vállalatok arányát, amelyek közt szignifikáns különbséget tapasztalt. A szerző szerint ennek hátterében az egyes tevékenységi körök eltérő kockázati szintje is állhat. Harada-Kageyama [2013] az egyes ágazatok csődarányait modellezte makrogazdasági változók függvényében. A modellek közt tapasztalt kisebb eltérések szintén az iparági szempontok fontosságára utalnak a csődkockázat tekintetében

Más kutatók (mint például Manjón-Antolin – Arauzo-Carod [2008], illetve Chen [2012]) egyértelműen arra az álláspontra helyezkednek, hogy a csődelőrejelzés során

a modelleket csak egyetlen iparág megfigyeléseire kell szűkíteni. Ennek szükségességét Huynh et al. [2010] például azzal indokolja, hogy ha különböző iparágakból származó vállalkozások szerepelnek a csődmodellekben, akkor könnyen előfordulhat, hogy két vállalkozás közül az egyik működő, míg a másik fizetéseképtelen annak ellenére, hogy mutatószámaik azonos értéket vesznek fel. Ebben az esetben kizárólag az érintett vállalatok tevékenységi köre bírhat diszkrimináló erővel a két, egyéb tekintetben teljesen azonos megfigyelés között. Beaver et al. [2005] szerint a közszolgáltatók és a pénzügyi szektorban tevékenykedő vállalatokat célszerű kihagyni a csődmodellek mintájából, s azokra külön modellt készíteni. Demers-Joos [2007] a hagyományos és a high-tech iparágak megkülönböztetését tartja lényegesnek a csődelőrejelző modellek szempontjából. Bose [2006] például kizárólag a dotcom cégek vizsgálatára helyezte a hangsúlyt.

Annak ellenére, hogy a szakirodalomban nem mutatkozik a konszenzus a modellezés egyetlen iparágra történő szűkítésének hatékonysága szempontjából, gyakori eset a nemzetközi szakirodalomban, hogy a szerzők egy-egy konkrét iparág vonatkozásában állítanak fel csődmodelleket: Foreman [2003] például az amerikai telekommunikációs cégekre szűkítette le vizsgálati körét, Pindado-Rodrigues [2004] cipőgyártással foglalkozó portugál cégek adatait vette alapul, Lin et al. [2009] és Yeh et al. [2010] a tajvani tőzsdén jegyzett elektronikai cégeket, Chen [2012] építőipari vállalkozásokat, Shetty et al. [2012] pedig kizárólag indiai IT vállalatokat vizsgált.

Az adatgyűjtés tekintetében érdekes aspektusra hívta fel a figyelmet Deawelheyns-Van Hulle [2006]. Az idézett szerzők kutatási eredményei arra utalnak, hogy jelentős különbség van a leányvállalatok és a teljesen önálló társaságok csődjének előrejelezhetősége között, így a gyakorlati modellezés során érdemes lehet figyelmet fordítani erre a szempontra is.

2.3.1.2. A mintavétel által lefedett időszak

Az adatgyűjtéshez kapcsolódó kutatási terület annak vizsgálata is, hogy a gazdaság általános állapota milyen hatást gyakorol a csődmodellek pontosságára. A csődelőrejelzés tudományának alapköveit a 60-as-70-es években angolszász területen fektették le, s ezt követően még évtizedekig az USA, illetve Nyugat-Európa tőzsdei adatait használták fel a témakör kutatói. Mivel ebben az időszakban jelentősebb gazdasági visszaesés nem sújtotta ezeket a régiókat, így az ott dolgozó kutatóknak nem is volt lehetőségük a recesszió hatásainak elemzésére. A kutatási irány a távol-keleten merült fel először. Sung et al. [1999] a 90-es évek végén Kelet-Ázsiában kialakult recesszió során azt tapasztalta, hogy az azt megelőző konjunktúra idején felállított modellek teljesítménye jelentősen romlott a válság időszakában. Kutatási eredményeik arra engedtek következtetni, hogy konjunktúra és visszaesés idején külön modellekre van szükség a vállalatok jövőbeli fizetőképességének minél pontosabb előrejelzése érdekében.

Más aspektusból elemezte a gazdasági recesszió hatását Bongini et al. [2000]. Kutatásaik során azt vizsgálták, hogy a válságnak milyen típusú vállalatok esnek nagyobb eséllyel áldozatul. A gazdasági visszaesések pozitív hatásai közt tartják számon azok piactisztító hatását, melynek következtében a kevésbé hatékony vállalatok elhagyják a piacot. Alacsony tőkeerejük miatt a kisebb vállalkozások is nagyobb eséllyel mennek csődbe nehéz gazdasági körülmények között. Ennek ellenére az idézett szerzők arra jutottak, hogy a kelet-ázsiai válságnak több relatíve jó pénzügyi helyzetben lévő vállalat is áldozatul esett. Az idézett szerzők elemzése az mutatták, hogy a 90-es évek végén tapasztalt kelet-ázsiai válságot a legnagyobb eséllyel a konglomerátumban működő vállalatok élték túl.

Szintén a kelet-ázsiai válság nyújtott alapot Nam-Jinn [2000], valamint Nam et al. [2008] vizsgálatainak is. Utóbbiak a recesszió hatására megváltozott makrogazdasági körülmények hatásait építették be csődelőrejelző modelljükbe, melynek köszönhetően szignifikánsan javult annak előrejelző képessége.

A napjainkban tapasztalt világgazdasági recesszió az amerikai és a nyugat-európai tőzsdéken is éreztette hatását. Du Jardin-Séverin [2012] francia vállalatok adatain hasonló következtetésre jutott, mint a kelet-ázsiai válság hatását kutató szerzők, ugyanis ebben az esetben is azt találták, hogy a konjunktúra idejéből származó adatok alapján felállított modellek a válság idején jóval gyengébb teljesítményt mutattak. Az idézett szerzők azt is vizsgálták, hogy ez a kapcsolat fordítva is érvényes-e: az eredmények arra utaltak, hogy a válság adataira épülő modellek előrejelző képessége is romlik, ha konjunkturális gazdasági környezetben kerülnek alkalmazásra. Ezek az eredmények pedig újabb empirikus bizonyítékkal szolgálnak arra, hogy gazdaság általános állapotának függvényében más csődmodellekre van szükség.

Az előzővel némileg ellentmondásos Pompe-Bilderbeek [2005] kutatási eredménye, amely szerint a konjunktúra idején készített modellek gyengébb teljesítménye recesszió idején a csőd előrejelezhetőségének csökkenésével magyarázható és nem a modellek illeszkedésének problémáival.

A makrogazdasági helyzet csődmodellekre gyakorolt hatása tekintetében az előzőekhez képest új aspektusra hívta fel a figyelmet Gersbach-Lipponer [2003], akik szimulációs vizsgálatot folytattak annak elemzésére, hogyan befolyásolja a banki tevékenység kockázatát a különböző vállalatok csődeseményei közti korreláció. A szerzők álláspontja szerint ugyanis egy negatív makrogazdasági sokk nem csak az

egyedi vállalatok szintjén mért csődkockázatot növelheti. Tekintettel arra, hogy a vállalatok nem tekinthetők teljesen függetlennek egymástól, fontos szempont lehet a vállalatok csődjeinek korrelációja, ami egy esetleges recesszió következtében szintén növekedhet, fokozva ezzel a hitelintézetek kockázatát is.

2.3.1.3. A mintavétel földrajzi vonatkozásai

A szakirodalom kronologikus bemutatása során említettem, hogy a kutatók még napjainkban is referenciaként használják Altman [1968] modelljét. Az ott hivatkozott szerzők az esetek egy részében meglehetősen jó előrejelző képességet tapasztalnak, azonban konszenzus látszik kialakulni abban kérdésben, hogy egy adott ország vállalati adathalmazán becsült modell más nemzetgazdaság cégei esetén kevésbé használható. Erre vonatkozóan Ooghe et al. [1999] végzett összehasonlító elemzéseket, melyek megerősítették az előző hipotézist. Moro et al. [2011] 12 ország vállalati adatbázisát elemezve arra a következtetésre jutott, hogy a csődesemény bekövetkezése és a pénzügyi mutatók közti kapcsolatrendszerben országonként is tapasztalhatók eltérések. Bankcsődök modellezése során hasonló eredményre jutott Ioannidis et al. [2011] is. Sueyoshi-Goto [2009] pedig japán vállalatok kapcsán hívta fel a figyelmet a vállalati gazdálkodás ország specifikus sajátosságaira, amelyek nem hagyhatók figyelmen kívül a csődelőrejelzésben sem.

2.3.1.4. A dichotóm klasszifikáció problémái

A bevezető fejezetben szó esett a csődelőrejelzés tudományának egyik legkritikusabb kérdéséről, az előrejelzésről. Munkámban magam is a vállalatok jövőbeli fizetéseképtelenségének előrejelzésére teszek kísérletek a rendelkezésre álló mintában

szereplő fizetőképes és fizetéseképtelen vállalatok pénzügyi adatainak felhasználásával. Szakkifejezéssel élve: a csődelőrejelzés egy dichotóm klasszifikációs feladat, melynek egyik legnagyobb nehézsége, hogy a vállalatokat két egymást kizáró csoportba (csődös, működő) sorolja.

E megközelítés problémája, hogy nem veszi figyelembe, hogy a két osztály nem feltétlenül zárja ki egymást az azonosításuk céljából felhasznált mutatók tekintetében. Ennek oka, hogy a korábban fizetőképes vállalatok válhatnak a későbbiekben fizetéseképtelenné. Általános tapasztalat, hogy a vállalatok fizetéseképtelensége nem egy hirtelen bekövetkező esemény, annak a legtöbb esetben vannak előjelei (Lin et al. [2011]). Ennek hátterében vélhetően az áll, hogy a fizetéseképtelenség bekövetkezése egy hosszabb-rövidebb pénzügyi nehézséggel jellemzett időszak végső stádiuma. Ebben a periódusban pedig reális feltételezés, hogy a még működő, de már pénzügyi gondokkal küzdő vállalkozások számviteli mutatói hasonló, vagy akár rosszabb értéket vesznek fel, mint a mintában szereplő fizetéseképtelen vállalatok rátái. Ez a csoport tehát átmenetet képez a csődös és a működő vállalatok között.

Hasonló „köztes állapot” jellemzi azokat a gazdálkodó egységeket, amelyek valamilyen előre nem látható, hirtelen esemény (például természeti katasztrófa) miatt válnak fizetéseképtelenné. Ebben az esetben egy egyébként stabil pénzügyi helyzettel jellemezhető vállalat is fizetéseképtelenné válhat annak ellenére, hogy azt pénzügyi mutatói nem tükrözik abból adódóan, hogy azok csak több mint fél éves késleltetéssel válnak nyilvánossá. Az ilyen esetek azonban jóval ritkábban fordulnak elő, mint az előző szakaszban bemutatott példa.

A két átmenetet képező csoportba tartozó vállalatokat nevezte Altman [1968] szürke zónának. Az ide sorolható vállalkozásokat ugyanis nem lehet egyértelműen besorolni egyik, vagy másik csoportba. Ebből a problémából indult ki Gilbert et al. [1990], akik a dichotóm klasszifikáció helyett háromcsoportos megközelítés alkalmazását javasolták a csődelőrejelzésben. Ennek lényege, hogy a csődös és a működő vállalatok mellett célszerű lehet megkülönböztetni a még működő, de már pénzügyi problémákkal küzdő vállalatok csoportját. Az idézett szerzők kutatási eredményei azt mutatták, hogy e köztes csoport előrejelzése jóval nehezebb, mint a kiemelkedően jó pénzügyi helyzetben lévő működő, illetve a kirívóan gyenge teljesítményt nyújtó csődös vállalatok megkülönböztetése. Ebből indult ki Anandarajan et al. [2001] is, amikor a működő cégek csoportjánál csak olyan megfigyeléseket vettek figyelembe, amelyek olyan események bekövetkezése esetén is tovább tudtak működni, mint a banki kölcsönnel kapcsolatos fizetési késedelem, vagy negatív cash flow-val zárt üzleti év.

A többcsoportos megközelítés létjogosultságához nem férhet kétség, gyakorlati megvalósíthatósága azonban korlátozott. Ennek oka, hogy nem áll rendelkezésre objektív definíció a pénzügyi nehézségek vonatkozásában, a szubjektív definíciók pedig nem állják ki a tudományosság próbáját. Ennek ellenére a későbbiekben más szerzők is kísérletet tettek a kérdéskör vizsgálatára: lásd például Altman et al. [1994], Slowinski-Zouponidis [1995], Bioch-Popova [2001] illetve Jones-Hensher [2007] munkáját. Utóbbiak négycsoportos megközelítés alkalmazását tartották célszerűnek.

Az objektív definíció hiányából eredő problémára különböző megoldási lehetőségeket vetettek fel a kutatók. Yang [2001] és Du Jardin [2010] megközelítésének lényege, hogy a csődös és működő vállalatok körén belül vizsgáljuk meg, hogy találhatók-e olyan csoportok, amelyek szignifikánsan

különböznek. Ez a megoldás lehetőséget nyújthat a csődös és a működő vállalatok közti átmenetet képező megfigyelések azonosítására objektív definíció hiányában. Sueyoshi-Goto [2009] azt a megoldást választotta, hogy először azonosította a dichotóm klasszifikációs modellek szürke zónáját, majd külön modellt állított fel az oda sorolt vállalatok minél pontosabb minősítése érdekében. Alfaro et al. [2008] pedig kihagyta a működő vállalatok csoportjából azokat a megfigyeléseket, amelyek három egymást követő évben veszteségesek voltak, mivel ezek vélhetően a pénzügyi problémákkal küzdő cégek körébe tartoznak, melyek inkább hasonlítanak csődös vállalatokra, mint a működőkre, így mintában történő szerepeltetésüknek torzító hatása lenne a modellekre.

Ebben a tekintetben is található ellentmondásos eredmény a szakirodalomban: Gruszczynski [2004] jobb találati arányt tapasztalt a dichotóm modellek alkalmazása esetén, mint abban az esetben, amikor a pénzügyi problémákkal küzdő cégek külön csoportot képeztek az adathalmazban.

2.3.1.5. A minta felosztása a két csoport között

Az adatgyűjtés témakörébe sorolható kérdés, hogy a modellezéshez használt minta milyen arányban tartalmazzon csődös, illetve működő vállalatokat. Statisztikai szempontból praktikus választásnak tűnhet a reprezentatív minta alkalmazása, azonban annak számos hátrányos aspektusa van a csődelőrejelzés területén:

- a reprezentativitás nem csak a két vizsgált csoport (csődös, működő) aránya tekintetében lehet érdekes; hasonlóan fontos szempont lehet a vizsgált vállalatok kora, mérete és tevékenységi köre, amelyek jelentős hatást gyakorolhatnak a vállalatok csődcockázatára. E szempontok szerint is

célszerű lenne a reprezentatív minta alkalmazása, a gyakorlatban azonban rendkívül nehéz egyidejűleg biztosítani, hogy a minta mind a négy szempont szerint megfeleljen a sokasági arányoknak, ugyanakkor kellően nagy is legyen ahhoz, hogy abból statisztikai következtetéseket lehessen levonni;

- a fizetésektelen vállalatok aránya jellemzően jóval alacsonyabb a fizetőképesekhez viszonyítva. Ez azért jelenthet problémát, mert még egy relatíve nagy elemszámú minta esetén is alacsony lehet a csődös vállalatok száma. Ebben az esetben pedig nem biztosítható, hogy a minta elegendő információt tartalmaz a fizetésektelen vállalatok lehető legpontosabb azonosításához;
- a korszerű adatbányászati módszerek hajlamosak a többségi csoport sajátosságaira specializálódni (Chen et al. [2013]), emiatt az ilyen eljárások alkalmazása esetén a szakirodalom az olyan minták használatát javasolja, amelyek fele-fele arányban tartalmaznak megfigyeléseket a két csoportból;
- általános sajátosság, hogy a csődös vállalatok pénzügyi mutatói jellemzően nagyobb szóródást mutatnak, mint a működőké, ami szintén az arányosan rétegzett minta alkalmazásával áll szemben a mintavételi hiba negatív hatásainak mérséklése érdekében.

A fenti érvek miatt gyakori a szakirodalomban a csődös vállalatok felülreprezentálása. Ezt egyrészt úgy valószínűsítjük meg, hogy a minta összeállítása során több fizetésektelen megfigyelést vesznek figyelembe azok sokasági arányához képest. A másik megközelítés ragaszkodik a reprezentatív minta alkalmazásához, és a csődös vállalatok adatainak nagyobb súlyt adva oldják meg azt a problémát, hogy az adatbányászati módszerek hajlamosak a többségi csoport sajátosságaira specializálódni. Ezen a területen Kennedy et al. [2013] folytatott kutatásokat.

Eredményeik azt mutatták, hogy a hagyományos statisztikai modellek körébe tartozó logisztikus regresszió találati arányát ugyan nem, de az adatbányászati módszerek eredményét érdemben javítja a fizetésektelen megfigyelések felülreprezentálása. Hasonló következtetésre jutott portugál építőipari vállalatok adatain Horta-Camanho [2013] is. Sánchez-Lasheras et al. [2012] a probléma megoldásának egy másik módjára hívják fel a figyelmet: a két csoport közti kiegyensúlyozott arány a működő cégek alulreprezentálásával is megoldható, amit az idézett szerzők úgy valósítottak meg, hogy azokat a működő vállalkozásokat, amelyek hasonló pénzügyi mutatókkal rendelkeztek, egyetlen reprezentatív megfigyeléssel helyettesítették.

2.3.2. Az adatelőkészítés

A csődmodellek független változóit már az 1960-as évektől kezdődően a megfigyelt vállalatok számviteli információs rendszeréből kinyerhető hányados típusú pénzügyi mutatószámok képezik (Chen [2012]). Ugyan e változók sok esetben rendkívül hatékonyan képesek a fizetőképes és a fizetésektelen vállalkozások megkülönböztetésére, szót kell ejteni azok korlátairól is, melyek közül az egyik legfontosabb az adatok múltorientációja. A számviteli információk egy lezárt üzleti év végén aktuális állapotot tükröznek, így az azokra épülő modellek előrejelző képessége korlátozott (Lin et al. [2014]).

A pénzügyi mutatószámokat évtizedek óta „nyers” formában használják a modellezés során. Ez azt jelenti, hogy azokat abban a formában alkalmazzák, ahogy a mérleg, illetve eredménykimutatás adatai alapján meghatározásra kerültek. Ez a megközelítés pedig tovább korlátozza a pénzügyi mutatók előrejelző képességét. Napjainkra a szakirodalomban konszenzus alakult ki az adatok modellezésre történő

előkészítésének fontossága kapcsán, melynek köszönhetően ez a kérdés önálló vizsgálati terület a csődelőrejelzés szakirodalmában.

2.3.2.1. Számítási problémák

Fontossága ellenére a csődelőrejelzés szakirodalmában szinte alig találni olyan publikációt, amely foglalkozna azokkal a Kristóf [2008] által felvetett számítási problémákkal, amelyek a hányados típusú pénzügyi mutatók esetében gyakran felmerülnek. Az idézett szerző az alábbi két problémára hívja fel a figyelmet:

- a nevező nulla értéke;
- az az eset, amikor a számláló és a nevező is negatív értéket vesz fel.

A pénzügyi ráták meghatározását ugyan csak az első teszi lehetetlenné; a második figyelmen kívül hagyása hasonlóan súlyos következményekkel járhat. Tekintsük példának a saját tőke arányos nyereség (ROE) mutatóját egy olyan vállalkozás esetén, melynek a vizsgált évben az adózás utáni nyeresége és a saját tőkéje is negatív volt. Ebben az esetben a pénzügyi mutató értéke teljesen félrevezető képet mutat a vállalat jövedelmezőségére vonatkozóan. Ha a problémát valamilyen módon nem korrigálják, az e mutatót is tartalmazó modellek megbízhatósága erőteljesen kétségessé válik.

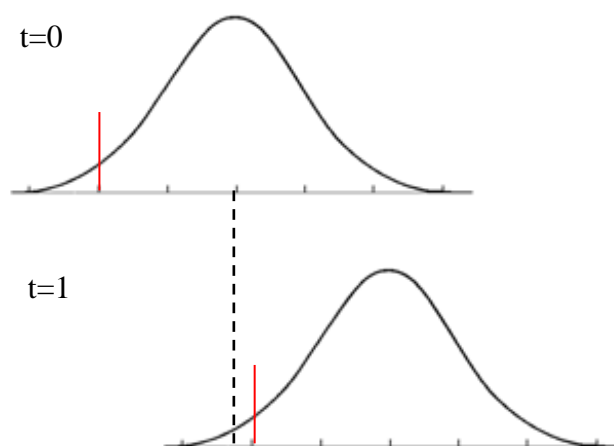
2.3.2.2. Az iparági sajátosságok figyelembevétele

A csődelőrejelzés fejlődéstörténetében évtizedekig komoly problémát jelentett a modellek időbeli instabilitásából eredő probléma. A kutatók gyakran szembesültek azzal, hogy egy adott időpont, illetve időszak adatain felállított modellek az azok

felállításához használt adatok időintervallumában megbízható eredményt mutattak; az idő múlásával azonban a modellek jelentősen veszítettek előrejelző képességükből.

A probléma megoldására Platt-Platt [1990] javasolt hatékony megoldást. Véleményük szerint a modellek időbeli instabilitásának az képezi az alapját, hogy idővel megváltozik az egyes mutatószámok középértéke, melynek következtében a modellépítés időpontjában egy átlaghoz közeli érték egy későbbi időszakban már szélsőségesen alacsony, illetve magas érték lehet az adott iparágban bekövetkező változások hatására. Normális eloszlás esetén ez az alábbi ábrán szemléltethető.

2. ábra. A mutatószámok eloszlásában bekövetkező változások hatása



Forrás: Platt-Platt [1990] alapján saját szerkesztés

A 2. ábra azt mutatja be, hogy ha egy normális eloszlású mutató középértéke ceteris paribus megnövekszik, akkor ennek hatására a $t=0$ időpont átlagos értéke a $t=1$ időpontban már az eloszlás szélén helyezkedik el, azaz akkor meglehetősen alacsonynak tekinthető az iparági átlaghoz képest. Az ábrán az eloszlásokon

feltüntetett folytonos függőleges vonal azt az értékét jelzi, amelytől alacsonyabb értékek a csődös vállalatokat jellemzik.

Egy ilyen változás jelentősen csökkentheti a $t=0$ időpont adatain felállított modell előrejelző képességét. Ha feltételezzük, hogy a vizsgált mutató tekintetében az átlagos érték a működő vállalkozások sajátosságának tekinthető, akkor a modell a $t=1$ időpontban is működőnek fogja minősíteni az adott megfigyelést annak ellenére, hogy akkor már a változó adott értéke a fizetéseképtelen vállalatokra jellemző. Ezt a problémát Platt-Platt [1990] az iparági relatív mutatók alkalmazásával oldotta meg, amelyet az alábbi képlet alapján számszerűsíthetünk:

$$\frac{\text{vállalati mutató értéke}}{\text{iparági átlag értéke}} \times 100$$

A fenti formula lényege, hogy az egyes vállalatok mutatóit azon iparág átlagához viszonyítja, amelyben az érintett vállalkozás is tevékenykedik. A megközelítés előnye, hogy alkalmazása esetén nem okoz problémát a változók középértékében idővel bekövetkező változás, ugyanis az iparági relatív mutatók a vállalati értékeknek az iparági átlagától való eltérését számszerűsítik, ebből adódóan invariánsak a középérték nagyságában bekövetkező változásokra (Platt-Platt [1990]).

Amellett, hogy az iparági relatív mutatók megoldást kínálnak az adatok időbeli instabilitásából eredő problémákra, említést kell tenni egy másik fontos sajátosságukról is. A csődelőrejelzés szakirodalmában még napjainkban is gyakori, hogy több különböző iparág megfigyeléseit szerepeltetik egyidejűleg a modellépítéshez használt mintában. Ilyen esetekben azonban kérdéses lehet, hogy két különböző iparágban tevékenykedő vállalkozás egy konkrét mutatója milyen mértékben vethető össze egymással; ugyanis a pénzügyi elemzéssel foglalkozó

tankönyvek alapelveként tárgyalják, hogy a pénzügyi mutatószámok értéke nem jelent abszolút kritériumot (Virág et al. [2013]), azokat csak valamilyen viszonyítása alapvonatkozásában lehet objektíven megítélni. Az iparági átlag erre a célra kiválóan alkalmas lehet, melyek alkalmazása lehetővé teheti a különböző iparágban tevékenykedő vállalkozások megfelelő pénzügyi mutatóinak összehasonlítását.

Az iparági relatív mutatók alkalmazása a bemutatott kedvező tulajdonságok ellenére is csak meglehetősen ritkán fordul elő a szakirodalomban. A kutatómunka során feldolgozott több száz folyóiratcikk közül csak három esetben (Dewaelheyns-Van Hulle [2004], Hillegeist et al. [2004], Berg [2007]) találkoztam a pénzügyi mutatók iparági átlaghoz történő viszonyításával. A szakirodalombeli alulreprezentáltság lehetséges okai:

- a modellezők nem szükségképpen ismerik az egyes iparágakban tevékenykedő vállalatok pénzügyi mutatóinak átlagos nagyságát;
- számos pénzügyi mutató (tipikusan a jövedelmezőségi mutatók) csak intervallum skálán mérhetők, ahol problémát okozhat az iparági relatív mutatók hányados alakban számítható formulája;
- kérdéses lehet, hogy a különböző iparágakban az egyes mutatók iparági középértéktől való eltérései mennyire összehasonlíthatóak.

2.3.2.3. Az adatok modellezésre történő előkészítése

A pénzügyi mutatók statisztikai sajátosságait már Beaver [1966] is elemezte. Az egyes pénzügyi mutatók eloszlását hisztogrammon ábrázolva arra a megállapításra jutott, hogy a legtöbb változó eloszlása aszimmetrikus. A normális eloszlástól való eltérés pedig súlyosan érinti a lineáris statisztikai modellek teljesítményét, ami a

csődelőrejelzés területén azért kiemelt jelentőségű, mert McLeay-Omar [2000] szerint a pénzügyi mutatószámok esetén a kiugró (outlier) értékek jelenléte, valamint a normális eloszlástól való eltérés inkább tekinthető alapvető sajátosságnak, mint ritka kivételnek. Emiatt a problémák kezelése kulcsfontosságú kérdés a modellek teljesítményének maximalizálása szempontjából.

A statisztikai modellek alkalmazásánál sokszor feltételezés, hogy a magyarázó változók matematikailag függetlenek egymástól. Tekintettel azonban arra, hogy e mutatókat egyazon mérleg, illetve eredménykimutatás megfelelő értékeinek felhasználásával számítják, irreális feltételezés, hogy azok függetlenek egymástól. E problémára a leggyakrabban alkalmazott megoldás a főkomponens-elemzés, amely matematikailag független mesterséges változókba tömöríti a vizsgálni kívánt változók információtartalmát. Ezt a megoldást alkalmazta Li-Sun [2011a], valamint Xiaosi et al. [2011] is.

A multikollinearitás kérdésének főkomponensekkel történő megoldása tekintetében sem mutatkozik konszenzus a szakirodalomban. E módszer kapcsán súlyos hiányosságának tekinti Wang [2004] és Huang et al. [2012], hogy nem tesz különbséget a két vizsgált csoport (csődös, működő) között. Az idézett szerzők feltevése szerint abban az esetben, ha a vizsgált sokaságon belül több csoport is megkülönböztethető, akkor a főkomponens-elemzés kevésbé megbízható eszköz. E hiányosság orvoslására kerestek módszertani megoldást, melynek alkalmazásával a csődmodellek teljesítménye mindkét szerző esetén szignifikánsan javult a hagyományos főkomponensekkel elérhető eredményhez képest. A főkomponens-elemzés információsűrítő eszközének hatékonyságát kérdőjelezi meg a csődelőrejelzésben Erdal-Ekinci [2012] is, akik arra jutottak, hogy jobb csődmodell

állítható fel az eredeti változók alkalmazásával, mint az azokból képzett 1-nél nagyobb sajátértékű⁴ főkomponensek használata esetén.

Azokat az értékeket, amelyek jelentős mértékben eltérnek a többi megfigyelés megfelelő értékeitől, kiugró értéknek, idegen szóval outlier-nek nevezi a statisztikai szakirodalom. Az ilyen megfigyelések jelenléte is jelentősen ronthatja a statisztikai modellek teljesítményét. A probléma kezelésére több lehetséges megoldás is rendelkezésre áll.

- a legegyszerűbb megközelítés a kiugró értékkel bíró megfigyelések kihagyása a modellépítéshez felhasznált mintából, amelyre még napjainkban is találunk példát a szakirodalomban (Min et al. [2011], Cao [2012b], Sánchez-Lasheras et al. [2012], Fedorova et al. [2013]);
- az outlier értékek elhagyásával azonban McLeay-Omar [2000], valamint Balcean-Ooghe [2006] szerint⁵ a modellezés szempontjából fontos információtartalmat veszünk el, így ehelyett a változók transzformációja (logaritmizálás, gyöktranszformáció) a preferált megoldás;
- megoldás lehet még az adatok kategorizálása (Sun-Shenoy [2007]), melynek lényege, hogy az egyes változók értékkészletén belül valamilyen szempont szerint osztályközöket képeznek, s a modellezés során csak ezen osztályközök sorszámain – mint ordinális skálán mért változókat – alkalmazzák. A megközelítés előnye, hogy alkalmazásával nem jelentenek problémát a kiugró értékek, hiszen azok valamely szélső osztályközben jelennek meg. Ezzel egyidejűleg azonban le kell mondanunk az egyedi adatok

⁴ A hüvelykujjszabály használata általános a matematikai-statisztikában. Lásd például Chen [2011] munkáját is.

⁵ A szerzők hipotézisét a későbbiekben Yu et al. [2014] eredményei támasztották alá, akik azt találták, hogy pusztán a Saját tőke/Eszközök mutató alapján nagyon magas találati aránnyal minősíthető az általuk vizsgált tanuló minta abban az esetben, ha a klasszifikációt aszerint végezzük el, hogy az adott változó tekintetében outlier értékkel bír-e egy adott megfigyelés.

közt tapasztalt különbözőségekből származó információról, amelynek jelentős szerepe lehet a modell teljesítménye szempontjából;

- gyakran alkalmazott megközelítés, hogy az outlier értékeket a megfigyelések eloszlásának valamely szélső percentilisével helyettesítik. Duan et al. [2012] például az eloszlás 0,5, illetve 99,5 %-os percentiliséhez igazította az e terjedelmen kívül található értékeket. Cubiles-De-La-Vega et al. [2013] ugyanerre a célra az 1, illetve 99 %-os, Pindado-Rodrigues [2004] pedig az 5, illetve 95 %-os értéket választotta.
- a kiugró értékek jelenlétéből eredő problémákra az is megoldást jelenthet, ha a megfigyelt változók értékkészletét a $[0;1]$ intervallumba⁶ transzformáljuk. Ezt a megközelítést alkalmazta például Hu [2009], Li-Sun [2011a], Swiderski et al. [2012], Riberio et al. [2012], Kennedy et al. [2013], Saberi et al. [2013] is.

Kérdéses azonban, hogy mikor tekinthető egy adott érték outliernek. Erre vonatkozóan nincs egyértelmű definíció a szakirodalomban. Objektív meghatározás hiányában gyakori a statisztikai hüvelykujjszabályok használata. Az adatelőkészítő feladatoknak sok esetben részét képezi az adatok standardizálása a megfigyelt vállalatok mutatóinak átlagával és szórásával (lásd például Hájek [2012] munkáját). A gyakorlatban azon megfigyeléseket tekintik outlier-nek, amelyek standardizált értéke az 5, illetve 3 szórás terjedelmen kívül találhatók. Utóbbira példaként Van Gestel et al. [2006] munkája említhető.

⁶ A szakirodalomban találkozhatunk még a $[-1; 1]$ intervallumba történő transzformációval is például Zhong et al. [2014] munkájában.

2.3.2.4. A statikus csődmodellek kérdése

A modellek input változóit a csődelőrejelzés megjelenésétől kezdve napjainkig leggyakrabban a számviteli információs rendszerből származó hányados típusú pénzügyi mutatószámok közül válogatják. Az ezekre épített modellek azonban időben nem mutattak stabil teljesítményt.⁷ Ez a probléma hívta fel először a figyelmet a pénzügyi mutatók nyers értékeire épített csődmodellek statikus jellegéből adódó korlátokra, melyek megoldása érdekében kezdetben a kutatók a modellek rendszeres újrabecslését javasolták (Pantalone-Platt [1987b]). Ez azonban nem jelentett érdemi megoldást a problémára.

A kérdés fontosságát és aktualitását jelzi, hogy szakirodalmi összefoglalásában Abdou-Pointon [2011] fontos jövőbeli kutatási területként nevezi meg a modellek dinamizálását. A kutatók egy része módszertani megoldást keresett e hiány pótlására. Először a túlélési modellek alkalmazása jelent meg, mely feloldja a hagyományos statisztikai modellek statikus jellegéből eredő korlátokat (Hillegeist et al. [2004]). Korszerűbb módszereket alkalmazott Du Jardin-Séverin [2012], akik abból a feltételezésből indultak ki, hogy a pénzügyi mutatók értékeinek alakulása a makrogazdasági változások hatására időbeli instabilitást mutat, viszont a változások iránya és mértéke kevésbé. Ebből kiindulva az idézett szerzők nem a változók statikus értékei alapján jelezték előre a vállalkozások jövőbeli fizetőképességét, hanem annak figyelembevételével, hogy a megfigyelések pénzügyi mutatói hogyan változtak a tárgyévet megelőző hat évben. Az idézett szerzők kutatásuk során több ezer francia vállalkozás adatait használták fel, melyek tekintetében elemezték az egyes pénzügyi mutatók időbeli alakulását a vizsgált hat éves időintervallumban. Du Jardin-Séverin [2012] e trendek közt nagyfokú hasonlóságot tapasztalt, így azokat

⁷ A problémát részletesebben a 2.3.2.2. szakasz tárgyalta.

klaszterezték annak érdekében, hogy feltárják, hogyan alakulnak az időben azon vállalatok pénzügyi mutatói, amelyek csődbe mentek, illetve azoké, amelyek tovább működtek. Tekintettel arra, hogy a csődmodellek egyidejűleg több mutatót is alapul vesznek, a többdimenziós problémát az idézett szerzők önszerveződő térképek segítségével képezték le két dimenzióba, ahol vizuálisan megjeleníthetők a jellegzetes „csőd utak”, amelyek a szemléltetés mellett előrejelzési céllal is felhasználhatók. Eredményeik szerint a dinamikus megközelítés előrejelző képessége felülmúlta a statikus modellek teljesítményét a legtöbb elterjedt klasszifikációs módszerrel szemben. A későbbiekben hasonló vizsgálatot végzett vizualizációs céllal Chen et al. [2013] is.

A kutatók másik csoportja a pénzügyi mutatók „dinamizálásával” tett kísérletet a modellek statikus jellegének feloldására: Berg [2007] az egyes pénzügyi mutatók nyers értékei mellett azt is figyelembe vette modelljében, hogy az érintett mutatók milyen mértékben változtak egyik évről a másikra. Az idézett szerző vizsgálta a „több időszakos” csődmodellek alkalmazási lehetőségeit is. Ennek lényege, hogy a vállalatok t -edik időpontbeli fizetőképességét előrejelző modellek input változói között nem csak a $t-1$ -edik időszak adatait veszi alapul, hanem azok mellett szerepelteti a $t-2$ -edik és a $t-3$ -adik időszak megfelelő mutatóit is. Az utóbbi megközelítéssel felállított modell előrejelző képessége szignifikánsan meghaladta a statikus modellekkel elérhető eredményt. Duan et al. [2012] is végzett vizsgálatokat ezen a területen: tőzsdei cégek esetén az egyes mutatók legutoljára megfigyelt értékének az azt megelőző 12 hónap átlagától vett eltérése kapcsán talált szignifikáns különbséget a csődös és a működő vállalatok között.

2.3.3. Változószelekció

A csődelőrejelzésben általános gyakorlat a hányados típusú pénzügyi-számviteli mutatószámok⁸ alkalmazása a modellek független változói között. Ezek forrása a vállalkozások által kötelezően elkészítendő számviteli dokumentumok (mérleg, eredménykimutatás) adatai. Du Jardin [2010] szerint azonban csaknem határtalan a lehetséges mutatószámok köre. Nem közömbös azonban a modellek teljesítménye szempontjából, hogy ezek közül melyeket alkalmazzuk, ugyanis már a tudományterület atyjának tekintett Beaver [1966] is megállapította, hogy a mutatók közt jelentős eltérés mutatkozik előrejelző képességük szempontjából. Nikolic et al. [2013] szerint a változószelekció célja a lehetséges magyarázó változók közül a redundáns, illetve a klasszifikációs feladat szempontjából irreleváns változók kiszűrése a modell pontosságának növelése, szerkezetének egyszerűsítése, valamint a felállításához szükséges számításigény mérséklése érdekében. Fontosságából adódóan a témakör kutatása önálló területté vált a csődelőrejelzés szakirodalmán belül.

2.3.3.1. A változószelekció nehézségei

A csődelőrejelzés súlyos hiányossága, hogy máig nem áll rendelkezésre olyan elméleti háttér, amely meghatározná, hogy mely mutatókat kell figyelembe venni a vállalatok jövőbeli fizetőképességének előrejelzése céljából (Nikolic et al. [2013]). A teoretikus keret hiányát már Beaver [1966] is felismerte, melyet egy egyszerű modellel kívánt pótolni. Ennek lényege, hogy a vállalkozást likvid eszközök

⁸ Idővel a szakirodalom felismerte e mutatók hiányosságait, melyből adódóan a későbbiekben alternatívaként jelentek meg az olyan értékalapú mutatók, mint például a gazdasági hozzáadott érték (részletesebben lásd Virág et al. [2013] munkáját). Ezek csődmodellben történő alkalmazására tett kísérletet Pasaribu [2008], azonban kutatási eredményei szerint e mutatók nem bizonyultak hatékonyak a vállalatok jövőbeli fizetőképességének előrejelzése szempontjából.

víz tározójaként értelmezi, amelyben a tározó kiszáradását tekinti fizetési képtelenségnek. A tározó vízellátását a vállalat bevételei biztosítják, a kiadásokat pedig a tározóból „kifolyó” víz szimbolizálja. Bár mai szemmel a modell meglehetősen egyszerű, továbbá ez sem ad objektív iránymutatást a modellekben szerepeltetendő változók körére, mindezek ellenére napjainkig gyakran hivatkozott: Laitinen-Laitinen [2000] is ezen elméletet vette alapul a modelljük input változóinak szelekciója során.

Elméleti modell hiányában gyakran alkalmazott megközelítés azon pénzügyi mutatók használata, amelyek más tanulmányokban eredményesnek bizonyultak (Du Jardin [2010]). Ennek hátránya, hogy különböző szerzők különböző változókat találtak szignifikánsnak a csődös és a működő vállalatok megkülönböztetése szempontjából (Lin et al. [2011]) a csődelőrejelzés tudományos szakirodalmának eddig eltelt közel 50 évében.

2.3.3.2. A csődelőrejelzés további információforrásai

Nyilvánosan működő társaságok esetén a tőkepiaci adatok széles köre is rendelkezésre áll a csődmodellek lehetséges input változóiként. A szakirodalomban máig nincs konszenzus arra vonatkozóan, hogy a számviteli vagy a tőkepiaci adatok tekinthetők-e jobbnak a jövőbeli fizetési képtelenség előrejelzése szempontjából. Mossman et al. [1998] a számviteli mutatókat, a cash flow változókat⁹, valamint a tőkepiaci hozamokat, illetve azok szórását hasonlították össze a vállalatok jövőbeli fizetőképességének előrejelzése szempontjából. Eredményeik szerint a csőd előtti évben a számviteli mutatók, 2-3 évvel a csőd előtt pedig a cash flow változók a

⁹ Az idézett szerzők számviteli mutatók alatt a mérlegből, illetve az eredménykimutatásból számítható hányados típusú pénzügyi mutatókat; míg cash flow változók alatt olyan változókat értenek, mint például a cash flow kimutatásban szereplő működési, vagy adófizetési cash flow.

leghatékonyabbak. Meglepő eredményük, hogy a tőkepiaci hozamok és azok szórása nem volt szignifikáns változó egyik időtávon sem. Agarwal-Taffler [2008] Altman [1968] modelljét állította szembe a piaci modellekkel. Eredményük szerint a számviteli adatokra épülő modell jobb eredményt mutatott.

Az előzővel szemben Atiya [2001] azt találta, hogy a számviteli mutatók mellett a tőkepiaci adatok szerepeltetése javítja a modellek találati arányát. Hasonlóképp kiegészítő jellegű kapcsolatot talált a két változócsoporthoz Das et al. [2009] is, akik a CDS (credit default swap) felárakat modellezték.

Beaver et al. [2005] azt vizsgálta, hogy világhírű cikkének megjelenése óta eltelt közel 40 évben veszítettek-e a számviteli mutatók előrejelző képességükből. A szerzők végkövetkeztetése az, hogy a tőkepiaci adatokhoz képest a számviteli mutatók nem bírnak többletinformációval. Hillegeist et al. [2004] is arra a következtetésre jutott, hogy a tőkepiaci adatokra épített modell jobb előrejelző képességgel bír, mint a kizárólag számviteli mutatókon alapuló. Hasonló eredmény adódott Chawa-Jarrow [2004] kutatásából is.

A számviteli adatokból kalkulálható pénzügyi mutatók mellett a szakirodalomban számos szerző hívja fel a figyelmet a nem pénzügyi változók fontosságára, melyek jelentősen növelhetik a modellek előrejelző képességét (Barniv et al. [2002], Balcean-Ooghe [2006], Lensberg et al. [2006], Sohn-Kim [2007], Manjón-Antolin – Arauzo-Carod [2008]). Dambolena-Khuory [1980] például arra mutatott rá, hogy bizonyos mutatók időbeli szóródásának is jelentős diszkrimináló ereje van a csődös és a működő vállalatok között.

Bankok csődjét vizsgálva Pantalone-Platt [1987b], vállalkozások vonatkozásában pedig Duffie et al. [2007] arra a következtetésre jutott, hogy a jövőbeli

fizetőképesség előrejelzése szempontjából a makrogazdasági változók nem tekinthetők hatékonyak. Ezzel szemben egy svéd bank hitelportfólióját vizsgálva Carling et al. [2007] szignifikáns magyarázó változókat talált a makrogazdasági változók között. A vállalatok korának figyelembevételét Sjøvoll [1999] javasolta. Bonfim [2009] makrogazdasági tényezőket és a vállalat korát is szignifikáns változónak találta modelljében. Utóbbiak fontosságát azzal indokolta, hogy a fiatalabb társaságok esetén más kockázati tényezőket kell figyelembe venni, mint az idősebb cégek vonatkozásában.

A makrogazdasági tényezők figyelembevételére elsősorban a panel ökonometriai modellek közé tartozó túlélési modellek keretében van lehetőség. E módszertani megoldást alkalmazva Nam et al. [2008] számos makrogazdasági változót talált szignifikánsnak a kelet-ázsiai válság időszakából vett minta adatai alapján. A makrogazdasági változók és a csődkockázat közti kapcsolat létezését nemzetgazdasági szintű adatok alapján végzett kutatások is megerősítik (Rösch [2003], Koopman-Lucas [2005], Harada-Kageyama [2011]), így szükségességük nagy valószínűséggel megalapozott a túlélési modellek alkalmazása esetén.

A nem pénzügyi változók alkalmazása a csődelőrejelzésen kívüli, de ahhoz nagyon hasonló jellegű kutatási területeken is hasznosnak bizonyult. Gaganis [2009] a számviteli beszámolókkal kapcsolatos csalásokat próbálta modellezni pénzügyi mutatókkal. A modellek teljesítményénél az idézett szerző jelentős javulást tapasztalt abban az esetben, amikor nem pénzügyi mutatókat is beemelt az input változók körébe. Sinha-Zhao [2008] autóhitelek bedőlését vizsgálta. Modelljükben a hagyományos mutatók mellett szerepeltették egy pénzügyi szakértő szubjektív megítélését is, amely a vizsgált esetek többségében javította az objektív változókra épített modell előrejelző képességét.

Balcean-Ooghe [2006] szerint a pénzügyi mutatókban nem tükröződik a vállalatok jövőbeli fizetőképességére vonatkozóan az összes lényeges információ, ezért fontosnak tartották a kvalitatív magyarázó változók alkalmazását a csődmodellekben, amire többek közt Slowinski-Zouponidis [1995], Anandarajan et al. [2001], Becchetti-Sierra [2003], Grunert et al. [2005], Alfaro et al. [2008], Huynh et al. [2010], Blanco et al. [2013], Horta-Camanho [2013] tett kísérletet. Az idézett szerzők a minőségi változók közül a vállalati vezetés hatékonyságát mérő mutatókat, a vállalat piaci helyzetét, társasági formáját, tevékenységének földrajzi helyét, valamint a társaságot hitelező bankot, mint független változókat alkalmazták sikerrel. E körbe sorolható Beynon-Peel [2001], valamint Demers-Joos [2007] tanulmánya, akik az egyes cégeket auditáló vállalkozásokat, mint kvalitatív magyarázó tényezőket is felhasználták modelljükben.

2.3.3.3. Változószelekciós módszerek

Du Jardin [2010] szerint a csődelőrejelzésben gyakori eset, hogy a lehetséges magyarázó változók körét azokra szűkítik, amelyek a korábbi kutatások során eredményesnek bizonyultak. Ez a megközelítés lényegében a régi modellek egyfajta újrabecslésének tekinthető későbbi adatokon. Bár az eredmények sokszor meglepően jók, számos kutató, köztük Lin [2009] arra hívja fel a figyelmet, hogy célszerűbb a rendelkezésre álló adatok függvényében új modellt építeni, melyekben a korábbi modellek szignifikáns változói csak lehetséges alternatívaként jelennek meg.

Feki et al. [2012] a változószelekcióra alkalmas módszereket két csoportba sorolják. Az egyik a szűrő (filter) módszer, amelybe az önálló változószelekciós eljárások tartoznak. Ezek klasszifikációs módszertől függetlenül vizsgálják, hogy mely

változók tekintetében különböznek leginkább a megfigyelt vállalatok egyes csoportjai. A módszer előnye az egyszerűség és az alacsony számításigény, viszont hátránya, hogy alkalmazásával nem garantált, hogy a választott klasszifikációs módszer diszkrimináló ereje maximális lesz. A másik megközelítést lefedő (wrapper) módszernek nevezik. Ez a választott klasszifikációs módszer keretein belül keresi a független változók azon kombinációját, amely maximalizálja annak diszkrimináló képességét. E megközelítés számításigénye jóval nagyobb, viszont lehetővé teszi, hogy a választott klasszifikációs módszer teljesítménye a lehető legnagyobb legyen. A lefedő megközelítést alkalmazó eljárások vizsgálata még rendkívül új terület a csödelőrejelzésben, ahol az elmúlt években Li-Sun [2011b], Oreski et al. [2012], valamint Lin et al. [2014] végzett összehasonlító elemzéseket. A jelentős számításigényre való tekintettel napjainkban még a szűrő módszerek alkalmazása a legelterjedtebb.

A hagyományos statisztikai módszerek¹⁰ (diszkriminancia-analízis, logisztikus regresszió) alkalmazása esetén jellemzően a stepwise (forward, backward) eljárásokat használják a szignifikáns változók azonosítása céljából. E technikák napjainkban már beépített funkcióként érhetők el az elterjedt statisztikai elemző szoftverekben, ahol a felhasználónak lehetősége van meghatározni, hogy mi az az elvárt szignifikancia-szint, amelynek alkalmazásával a változók beléphetnek, illetve kikerülnek a modellből. Hasonló okok miatt népszerűek a döntési fák is, amelyek a fa ágaztatását mindig a legnagyobb diszkrimináló erővel bíró változó értékei alapján végzik, így a változószelekció ebben az esetben is automatikus folyamat.

¹⁰ A csödelőrejelzésben elterjedt klasszifikációs módszerek elméleti háttéréről a 2.3.4. fejezet ad áttekintést.

A hagyományos klasszifikációs technikák esetén azért van lehetőség „beépített” változószelekciós eljárások alkalmazására, mert azoknak nincsenek olyan paraméterei, amelyet a felhasználónak kellene beállítania. A gépi tanulásra épülő adatbányászati módszerek többségénél azonban erre is szükség van, ami azért lényeges kérdés, mert e paraméterek optimális értékére nincs objektív iránymutatás a szakirodalomban, így azokat minden esetben optimalizálni kell a vizsgált problémának, illetve a rendelkezésre álló adatoknak megfelelően. Ez az aspektus pedig rendkívül megnehezíti a statisztikai módszereknél általánosan elterjedt forward, illetve backward eljárások alkalmazását, ugyanis a magyarázó változók körének módosulásával változik a paraméterek optimális értéke is (Ping-Yongheng [2011]).

A problémát kezdetben a szűrő módszerek alkalmazásával oldották meg. Ennek lényege, hogy valamely változószelekcióra képes eljárás szignifikáns változóit használták fel az adatbányászati modellek input változóinak körében. A legegyszerűbb megközelítést alkalmazta Li-Sun [2011a], akik azon változókat használták fel modelljükben, amelynek átlaga szignifikáns különbséget mutatott a csődös és a működő cégek csoportjai között a két független mintás t-próba eredménye alapján. E statisztikai teszt nemparaméteres megfelelőjét alkalmazta változószelekciós céllal Pindado-Rodrigues [2004]. A logisztikus regresszió szignifikáns változóit alkalmazta az SVM módszerhez Kim-Sohn [2010], ugyanezen technika esetén az RST (rough sets theory) módszer alkalmazásával szelektált változókat használta Tilmoudi et al. [2011]; Yazici [2011] a diszkriminancia-analízis szignifikáns változóit használta fel a neurális hálók input változóinak szelekciója céljából, Min et al. [2011] az általános additív modell (generalized additive model) segítségével kereste a szignifikáns változókat az SVM módszer alkalmazásával

felállított csődmmodelljéhez, a klasszifikációs és regressziós fák módszerét pedig Brezigar-Masten – Masten [2012] alkalmazta változószelekció céljából.

Bár a megközelítés alkalmazása máig gyakori, a későbbiekben a kutatói érdeklődés a genetikus eljárások felé fordult. Ezek az eljárások a természetes szelekció elvét szimulálva törekednek valamely optimalizációs feladat globális optimumának meghatározására. A csődelőrejelzés kontextusában a feladat a modellek előrejelző képességének maximalizálása, amelynek megoldására a rendelkezésre álló változók különböző kombinációi hivatottak. Ezek a változókombinációk alkotják a gének sokaságát, melynek tagjai versengenek a túlélésért. Ennek feltétele a modellező által meghatározott feladat minél hatékonyabb teljesítése. A természeti evolúcióhoz hasonlóan azok a gének, amelyek nem képesek a feladat elfogadható szintű megvalósítására, elpusztulnak, s átadják helyüket a feladatot abszolváló gének utódainak, amelyek a szülő gének kombinációjával, illetve mutációjával jönnek létre. E szimulációs megközelítés több esetben nagyon hatékonyan volt képes azonosítani az adatbányászati módszerek alkalmazása során felhasználandó független változók körét (Back et al. [1996], Lensberg et al. [2006], Hájek [2011], Oreski-Oreski [2014]). Meg kell még említeni Chi-Hsu [2012] munkáját, akik a statisztikai modellek körébe sorolható logisztikus regresszió magyarázó változóit válogatták genetikus algoritmus segítségével.

Du Jardin [2010] arra vonatkozóan végzett összehasonlító elemzéseket, hogy a csődelőrejelzésben leggyakrabban alkalmazott módszerek (neurális hálók, diszkriminancia-analízis, logisztikus regresszió) vonatkozásában melyik változószelekciós eljárás biztosítja a legjobb előrejelző teljesítményt. Eredménye arra engedett következtetni, hogy a stepwise módszerek a paraméteres eljárások esetén a

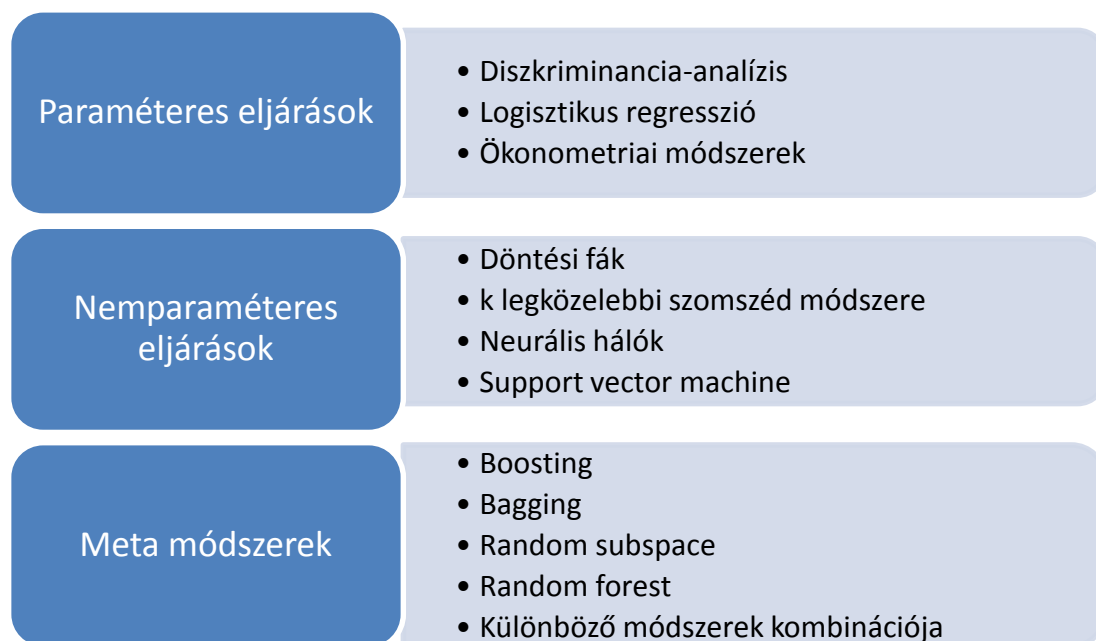
leghatékonyabbak; a neurális hálók pedig a genetikus algoritmussal szelektált változók felhasználásával mutatják a legjobb teljesítményt.

A későbbiekben történtek kísérletek az egyes adatbányászati módszerekre specializált változószelekciós eljárások kidolgozására is, azonban ezek alkalmazása a későbbiekben nem vált általánossá a szakirodalomban. Hardle et al. [2009] például az SVM eljárás vonatkozásában ismertetett egy lehetséges változószelekciós eljárást; Wang et al. [2014] pedig az információs hasznon alapuló változószelekciót próbálta integrálni a meta módszerekbe, amelyekről a következő alfejezet ad áttekintést.

2.3.4. A csődmódel felállítása

A minta összegyűjtését, a modellépítés során alkalmazni kívánt változók körének meghatározását, valamint az adatok modellezésre történő előkészítését követően a csődmódel-építés következő lépése az alkalmazni kívánt klasszifikációs módszer kiválasztása. A választás nem könnyű, ugyanis Du Jardin [2010] kutatása szerint az elmúlt 50 év szakirodalmában a kutatók több mint 50-féle módszert alkalmaztak a vállalatok jövőbeli fizetéseképtelenségének előrejelzése céljából. Ezeket az eljárásokat a szakirodalom három nagy csoportba sorolja (Kim-Kang [2012]).

3. ábra: A csődelőrejelzés klasszifikációs módszereinek csoportosítása



A rendelkezésre álló módszerek számossága joggal veti fel azt a kutatási kérdést, hogy vajon melyiket célszerű alkalmazni, ha célunk a modellek előrejelző képességének maximalizálása. A kérdés fontosságát jelzi, hogy napjainkban a témakörben megjelenő legtöbb tanulmány e módszerek összehasonlításával foglalkozik (Sánchez-Lasheras et al. [2012]). Ennek ellenére máig nincs konszenzus a szakirodalomban arra vonatkozóan, hogy melyik módszer tekinthető legjobbnak, ugyanis gyakori eset, hogy különböző kutatók különböző adathalmazokat vizsgálva más és más eljárások esetén tapasztalják a legjobb klasszifikációs teljesítményt.

Terjedelmi okok miatt a fejezetben csak a csődelőrejelzésben leggyakrabban alkalmazott klasszifikációs módszerek elméleti hátterét ismertetem részletesebben, kiemelt hangsúlyt helyezve a hazai szakirodalomban még kevésbé elterjedt legkorszerűbb adatbányászati eljárások bemutatására. Ezek mellett röviden említést

teszek még olyan egyéb klasszifikációs technikákról is, amelyek szakirodalmi súlya jóval kisebb. Ezek részleteit a hivatkozott publikációkban ismerheti meg az Olvasó.

2.3.4.1. Többváltozós statisztikai módszerek

A csődelőrejelzésben a többváltozós módszerek közül a lineáris diszkriminancia-analízis jelent meg először. Az eljárás eredményeképp kapott modell általános alakja:

$$Z = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

ahol az x_i az egy magyarázó változókat, a β_i pedig az egyes magyarázó változók becsült együtthatóit jelöli¹¹. A modell outputját (Z) diszkriminancia-értéknek nevezzük. Ez a mutató sűríti a független változók hatását egyetlen értékbe, amely alapján a klasszifikáció elvégezhető. Ehhez szükség van egy úgynevezett vágópont (cut value) alkalmazására, amely alapján elválasztjuk egymástól a két csoport megfigyeléseit. Ennek nagyságát a felhasználó szabályozhatja. A modell független változóit azok a mutatók adják, amelyek tekintetében a leginkább különböznek a vizsgált csoportok.

A módszer előnyei:

- könnyen hozzáférhető a statisztikai szoftverekben;
- alkalmazása gyors és egyszerű;
- lehetőséget nyújt a magyarázó változók és a csődesemény közti ok-okozati összefüggések elemzésére;
- a releváns változók automatikusan szelektálhatók a stepwise módszerekkel.

¹¹ részletesen lásd Virág et al. [2013]

A módszer hátrányai:

- alkalmazási feltételei (többdimenziós normalitás, variancia-kovariancia mátrixok azonossága, a magyarázó változók függetlensége) jellemzően nem teljesülnek (Du Jardin [2010]);
- az előző okok miatt klasszifikációs teljesítménye gyakran elmarad más módszerekhez képest (Wang-Ma [2011]);
- közvetlenül¹² nem képes a csődkockázat kardinális¹³ mérésére.

A fenti hátrányokból adódóan fordult a kutatók figyelmé a logisztikus regresszió felé, melyet a csődelőrejelzésben először Martin [1977] alkalmazott. A modell általános alakja az alábbi:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

ahol az x_i az egyes magyarázó változókat, a β_i pedig azok becsült együtthatóit jelöli.

A modell függő változóját gyakran nevezik logit értéknek. Az abban szereplő p változó a csőd bekövetkezésének valószínűségét szimbolizálja.

Az eljárás előnye, hogy nem támaszt olyan szigorú feltevéseket a vizsgált adathalmazzal szemben, mint a diszkriminancia-analízis; ebből adódóan klasszifikációs teljesítménye általában magasabb. A módszer előnyei megegyeznek az előző eljárás kedvező sajátosságaival, továbbá a logisztikus regresszió közvetlenül képes a csőd valószínűségét becsülni (Li-Sun [2011a]). Ezeknek, illetve relatíve

¹² A diszkriminancia-analízis outputja matematikailag csődvalószínűséggé konvertálható például a Deakin [1972] által bemutatott módon.

¹³ A csődkockázat kardinális mérése azt jelenti, hogy a csődmodell outputja arra a kérdésre is választ ad, hogy mennyivel kockázatosabb egyik megfigyelés, mint egy másik. Ezt a funkciót a diszkriminancia-analízis közvetlenül nem tölti be, ugyanis annak eredményváltozója bármilyen értéket felvehet a valós számok halmazán, így e módszer eredményváltozója csak arra alkalmas, hogy annak alapján a megfigyeléseket sorba rendezzük kockázatuk alapján (ordinális mérőszám).

magas klasszifikációs képességének köszönhetően napjainkig népszerű eljárás a csődelőrejelzésben (Nikolic et al. [2013]).

Mindkét statisztikai modell hátránya, hogy lineáris kapcsolatot feltételez a függő és a független változók között, ami a gyakorlatban jellemzően nem érvényes. A kapcsolat pontos típusa ismeretlen, csak annyi bizonyos, hogy jellemzően nemlineáris összefüggés tapasztalható a csőd bekövetkezése és a pénzügyi mutatók értékei között (Du Jardin [2010]). Ezek ugyan leképezhetők a változók interakcióinak, illetve különböző transzformációinak modellbe építésével, ezek tekintetében azonban a lehetőségek száma végtelen, így gyakorlati megvalósíthatósága korlátozott (Blanco et al. [2013]).

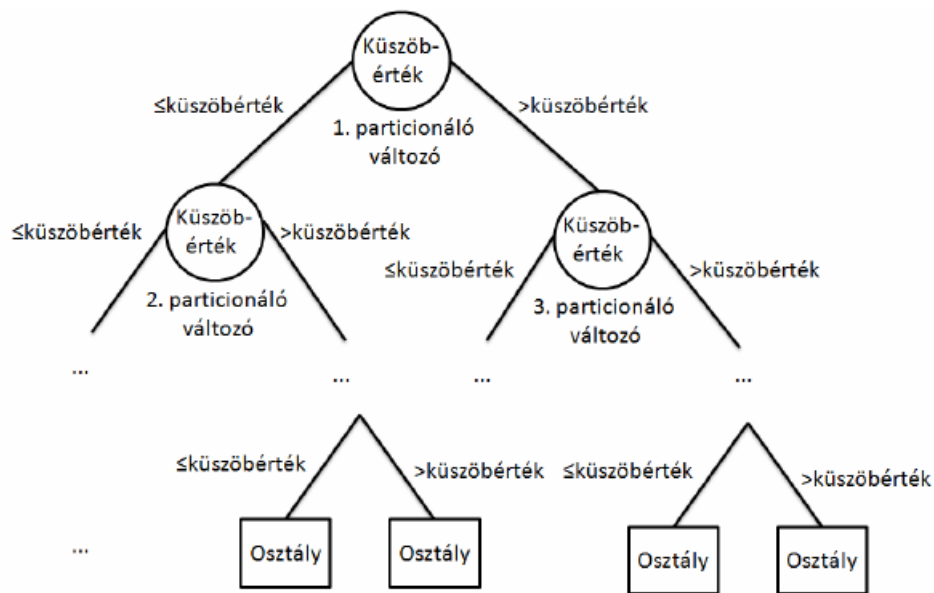
Végezetül említést kell tenni az ökonometriai módszerek alkalmazásáról is a csődelőrejelzésben. A panel ökonometriai eljárások közé tartozó túlélési modelleket (survival models) még napjainkban is gyakran használják. Lásd például Orbe et al. [2001], Cantner et al. [2006], Bharath-Shumway [2008], Löffler-Maurer [2011], Lyandres-Zhdanov [2013] munkáit. A logisztikus regresszió továbbfejlesztett változatát alkalmazta Heiss-Köke [2004], a Markov láncok csődelőrejelzésben történő adaptációjára pedig Elliott et al. [2014] tett kísérletet.

2.3.4.2. Nemparaméteres, gépi tanulásra épülő eljárások

A paraméteres módszerek előző szakaszban bemutatott hátrányaiból kiindulva a csődelőrejelzés szakirodalma az 1980-as évek közepétől a nemparaméteres eljárások felé fordult, melyek közül először a döntési fák jelentek meg. Napjainkra e módszernek számos változata jött létre és terjedt el a gyakorlatban. Ezek között

jellemzően az a különbség, hogy milyen feltétel szerint történik a fa ágaztatása. A döntési fák általános alakját az alábbi ábra szemlélteti.

4. ábra: A döntési fák általános alakja



Forrás: Kristóf [2008]

Az eljárás lényegében ötvözi az egy-, és többváltozós módszerek működési elvét. A fa első elágaztatása azon változó azon értéke szerint történik, amely szerint csoportosítva a megfigyeléseket a lehető leghomogénebb csoportok jönnek létre. Ezt követően az eljárás a már partícionált adathalmazban keresi, hogy van-e olyan változó, illetve azoknak van-e olyan értéke, amely szerint tovább finomítva a csoportokat, azok még homogénebbek lesznek a függő változó (csődös/működő) szempontjából.

A döntési fák módszerei közül gyakran alkalmazzák a klasszifikációs és regressziós fákat (CART), valamint a CHAID módszert (Koyuncugil-Ozgulbas [2012]); utóbbi a

particionálást a függő és a független változók szerint képzett csoportok közt vizsgált χ^2 -alapú függetlenségvizsgálat eredményei alapján végzi. Népszerű alternatíva ebben a módszercsoportban a C4.5 eljárás alkalmazása, amely a fa ágaztatását információelméleti alapon hajtja végre (Olmeda-Fernandez [1997], Baesens et al. [2003], Kiang [2003], Brown-Mues [2012], Abellán-Mantas [2014]). Ezek mellett napjainkban gyakran használt technika a Rough Set Theory (RST), amely egy halmazelméleten alapuló megközelítés. A módszer matematikai hátterét részletesen tárgyalja Ahn et al. [2000], McKee [2003], és McKee-Lensberg [2002].

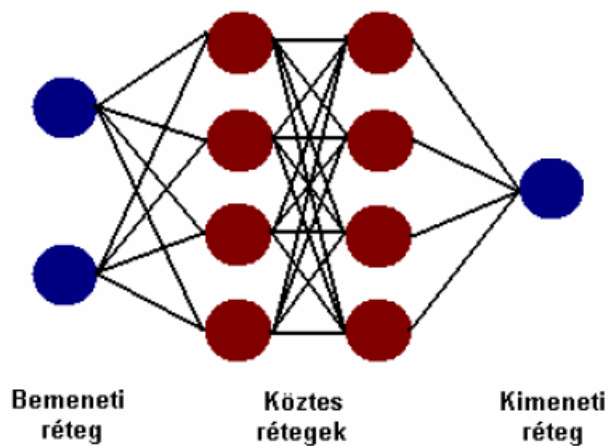
A statisztikai alapon nyugvó megközelítések problémája, hogy előre definiált kapcsolatrendszerrel feltételeznek a csődkockázat és a magyarázó változók között. A reláció formája azonban nem definiálható előzetesen egzakt matematikai formula segítségével. A probléma megoldása érdekében a csődelőrejelzés olyan módszerek felé fordult, amelyek a kapcsolatrendszerrel képesek a rendelkezésre álló adatok alapján feltárni.

Erre a célra alkalmazhatók olyan egyszerűbb eljárások is, mint a k legközelebbi szomszéd módszere, melynek lényege, hogy az egyes megfigyelések abba a csoportba kerülnek besorolásra, ahová a k darab legközelebbi szomszédjuk többsége tartozik. A megfigyelések távolságát leggyakrabban a köztük mért euklideszi távolság segítségével határozzák meg. A módszer egyszerűsége ellenére rendkívül jó klasszifikációs teljesítményre képes. Részletesebben lásd például Paleologo et al. [2010] és Gárcia et al. [2012] munkáit. Hasonló elvet követ a szó szerinti fordításban „eset alapú érvelés” (case based reasoning), melynek lényege, hogy az újabb megfigyelési egységeket abba az osztályba soroljuk be, amelyikbe a korábban megfigyelt elemek közül a vizsgált vállalathoz leginkább hasonló cég is tartozik

(Wang-Ma [2011]). Az eljárást számos szerző (például Vukovic et al. [2012], Cao [2012a]) alkalmazta sikerrel a nemzetközi szakirodalomban.

Az eddig bemutatott eljárások azonban csak másodlagos szerepet töltenek be a neurális hálók és a support vector machine mellett. Utóbbira nincs pontos magyar megfelelő, így a továbbiakban az SVM rövidítéssel utalok az eljárásra. A neurális hálók az agy tanulási módszerét szimulálják. Felépítésüket az alábbi ábra szemlélteti.

5. ábra: A neurális hálók általános felépítése



Forrás: Kristóf [2008]

A neurális hálók három fő részből állnak:

- a bemeneti réteg a magyarázó változókat tartalmazza;
- a köztes rétegben lévő elemek (neuronok) végzik az információfeldolgozást, ami az input változók értékeinek súlyozott összegén végrehajtott matematikai (általában logisztikus) transzformáció;
- a kimeneti rétegben lévő neuronok adják meg a neurális hálók outputját.

Az 5. ábrán az egyes rétegek közti vonalak a súlyokat szimbolizálják. A neurális hálók tanítása során e súlyok folyamatosan módosulnak oly módon, hogy a neurális háló outputja minél jobban megközelítse a célfüggvény értékét. A háló tanulási folyamatának befejezését az befolyásolja, hogy a felhasználó milyen leállási kritériumokat határoz meg. A tanulási folyamat véget érhet, ha

- a tanulási ciklusok száma eléri a modellező által meghatározott felső határt;
- a súlyok módosulásának mértéke kellően alacsony nem lesz;
- a tesztelő mintán elért hiba növekedni kezd.

A neurális hálók alkalmazásának legnagyobb előnye, hogy elméletileg bizonyított, hogy bármely folytonos függvény tetszőleges pontossággal közelíthető egy olyan egyetlen köztes réteget tartalmazó neurális hálóval, amelyben a neuronok száma kellően nagy (Cao [2012a]). Abban az esetben, ha a modellezőnek sikerül a neurális háló felépítését (topológiáját) optimálisan meghatározni, a rendszer képes a tanulási mintát hibátlanul minősíteni. Ez a tulajdonság azonban egyszerre előnye és hátránya is a technikának, mivel a tanuló minta hibátlan minősítése a túltanulás jele, ami azt jelenti, hogy a modell a minta sajátosságait „tanulta meg”, s emiatt mintán kívüli megfigyelések tekintetében gyenge teljesítményt mutat. A túltanulást a gyakorlatban úgy kerülük el, hogy a háló tanulási folyamata során nyomon követik a rendszer teljesítményét egy, a tanulási mintától független tesztelő minta adatain. Ennek alapján a tanítási folyamatot akkor állítják le, amikor a tesztelő mintán elért eredmény romlani kezd (Blanco et al. [2013]). A megközelítés számos előnye mellett szót kell ejteni McKee-Greenstein [2000] kritikájáról, melynek lényege, hogy a neurális hálók tanításának előbb bemutatott elve a modellező által választott tesztelő mintára optimalizálhatja a hálót, ami kérdésessé teheti, hogy az elemzéshez használt tanuló és tesztelő mintákon kívül is elfogadható teljesítményt nyújt-e a modell.

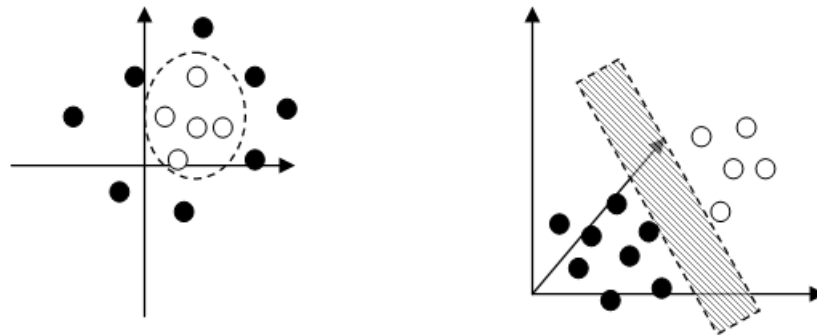
Egy másik gyakori probléma, hogy a neurális hálók tanulási folyamata hajlamos leállni a modell hibafüggvényének lokális minimumánál, melynek elkerülésére jelenleg nem áll rendelkezésre egzakt módszer (Blanco et al. [2013]).

A bemutatott hátrányok ellenére a módszer rendkívül népszerű a szakirodalomban még napjainkban is. Virágkorát az 1990-es években, illetve a 21. század első évtizedében élte. Az utóbbi 5-10 évben azonban presztízse csökkenni kezdett az előbbieken ismertetett problémákból, valamint az azokat kiküszöbölni igyekvő új módszerek megjelenéséből adódóan (Jeong et al. [2012]). Az imént idézett szerzők szerint a neurális háló továbbra is a legjobb előrejelző képességgel jellemezhető módszer, melynek teljesítményét napjaink szakirodalma alulbecsüli, mivel a felhasználók gyakran nem fordítanak elég gondot a paraméterek optimalizálására.

A módszer hátrányainak kiküszöbölése, illetve teljesítményének növelése érdekében a mesterséges intelligencia kutatói kísérletet tettek a módszer továbbfejlesztésére, melyek eseti jelleggel ugyan megjelentek (Neves-Vieira [2006], Hájek [2011], Pendharkar [2011]), általánosan azonban nem terjedtek el a szakirodalomban.

A neurális hálók vezető pozícióját az SVM módszer vette át az elmúlt egy évtizedben a csődelőrejelzés területén. A módszer lényegét a következő ábra szemlélteti.

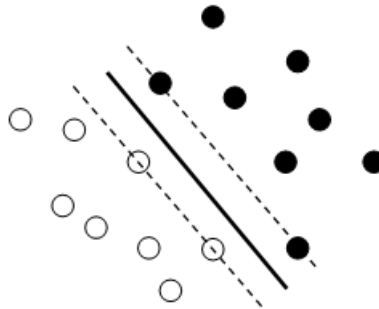
6. ábra: Az SVM módszer alapgondolata



A 6. ábra bal oldalán egy olyan eset látható, amely a csődelőrejelzésben gyakori: egy olyan nemlineáris összefüggés feltárása a feladat, amely képes az ábrán kétféle körrel jelölt megfigyelések szétválasztására. Az SVM módszer lényege, hogy az eredeti dimenziószámában megfigyelt vállalkozásokat egy kernel függvény segítségével kivetíti egy magasabb dimenziószámú térbe, ahol a két csoport lineárisan szeparálható. Az ott felállított hipersíknak az eredeti dimenziószámú térben megfelel egy nemlineáris összefüggés, amely képes a megfigyelések hiba nélküli klasszifikációjára.

Alapvető különbség a neurális hálók és az SVM módszer között, hogy előbbi az empirikus, utóbbi pedig a strukturális kockázat minimalizálásának (SRM) elvére épül (Lin et al. [2011]). Az empirikus kockázat minimalizálása azt jelenti, hogy a modell a tanulási minta alapján mért hibát igyekszik a lehető legalacsonyabb szinten tartani. Ezt a hiányosságot próbálja kiküszöbölni az a gyakorlati megoldás, hogy a tanulási folyamatot a tesztelő mintán elért eredmény romlásakor leállítják. Ezzel szemben az SRM elv a tanuló minta adatai alapján próbálja a teljes hibát minimalizálni. Kétdimenziós esetben az SVM módszer működését az alábbi ábra szemlélteti.

7. ábra: Az SVM eljárás működése kétdimenziós esetben



Az ábra szaggatott vonalai azt a területet határolják a térben, ahol egyik csoportból sem találunk megfigyelést. E két vonal közt elméletileg végtelen sok olyan egyenes létezik, amely hiba nélkül választja el a két csoportot. Ezek közül az SVM azt tekinti optimálisnak, amely a lehető legtávolabb van mindkét csoporttól, ugyanis vélhetően ez sorolja be helyesen a tanuló mintán kívüli megfigyeléseket is a legnagyobb valószínűséggel. Ez a strukturális kockázat minimalizálását célzó elv lényege. E megközelítésnek köszönhetően az SVM módszer olyan kiemelkedő általánosítási képességgel bír (Lin et al. [2011]), mely sok esetben még a neurális hálók teljesítményét is meghaladja. Különösen fontos, hogy az SRM elvből adódóan az SVM technika hatékonyabban kerüli el a tútanulást, s így kis elemszámú minta esetén is jó előrejelző képességgel jellemezhető modellek felállítását teszi lehetővé (Cao [2012b]); továbbá fontos sajátossága, hogy robusztus az outlier-ekkel szemben (Lin et al. [2011]). E tekintetben Shin et al. [2005], valamint Doumpos et al. [2005] végzett kutatásokat, melyek eredményei azt mutatták, hogy a tanuló minta méretének csökkenésekor az SVM előrejelző képessége relatíve stabil marad, míg a neurális hálóké szignifikánsan gyengül.

A 7. ábrán látható esetben a két csoport hiba nélkül szétválasztható egy lineáris függvény segítségével. A gyakorlatban azonban számos esetben előfordul, hogy ez

még a magasabb dimenziószámú térben sem valósítható meg hiba nélkül. A módszer kiterjeszthető az ilyen esetekre is oly módon, hogy az eljárás – a dolgozatban részletesen nem ismertetett – matematikai háttérénel bevezetésre kerül egy C paraméter, amely a téves besorolást bünteti. A paraméter értékére vonatkozóan elméleti kikötés nincs, annak nagyságát a felhasználónak kell szabályoznia annak megfelelően, hogy mennyire tolerálja a tanuló mintában elkövetett hibát. Minél nagyobb a C értéke, annál inkább törekszik a módszer a tanuló minta elemei esetén elkövetett hibák számának minimalizálására. Ennek azonban az az „ára”, hogy itt is felmerülhet a túltanulás problémája, ami a tesztelő mintán elért klasszifikációs teljesítményt gyengítheti.

Mint minden matematikai modellnek, így az SVM eljárásnak is megvannak a maga korlátai. Korábban említettem, hogy az optimális szeparáló hipersík meghatározása érdekében a módszer a megfigyeléseket kivetíti egy magasabb dimenziószámú térbe. E kivetítés a kernel függvények segítségével történik, amelyek közül az alábbiakat alkalmazzák a leggyakrabban.

- polinomiális:

$$k(x_i, x_j) = (x_i \cdot x_j)^d$$

- radial basis (Gauss) függvény:

$$k(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$$

- hiperbolikus tangens:

$$k(x_i, x_j) = \tanh(\kappa x_i \cdot x_j + c)$$

- ANOVA:

$$k(x_i, x_j) = \sum_{k=1}^n e^{[-\sigma(x_i^k - x_j^k)^2]^d}$$

A függvényekben az x_i és az x_j az egyes megfigyeléseket jelölik, a többi paraméter értékét – melyek a kivetítés dimenziószámát, illetve annak komplexitását szabályozzák – azonban a felhasználónak kell meghatároznia az általa vizsgált klasszifikációs probléma sajátosságainak megfelelően. Ezek értékeire vonatkozóan elméleti iránymutatás nem áll rendelkezésre. Ezt az empirikus feladatot jellemzően a táblázatos keresésnek nevezett módszerrel valósítják meg, melynek lényege, hogy kijelölnek n db lehetséges értéket a C paraméter nagyságára, valamint m db lehetséges értéket a kernel függvény beállításaira vonatkozóan és az azokból képezhető $(n \times m)$ -es mátrix valamennyi kombinációjának megfelelően felállítanak egy-egy modellt, melyek közül a legjobb teljesítményt nyújtó paraméterkombinációt választják ki a végső modell felállításához (lásd például Yeh et al. [2010] munkáját). Egy másik lehetséges megoldás a genetikusan algoritmus alkalmazása az optimális paraméterek meghatározására (Hsieh et al. [2012]).

További problémát jelent, hogy nagyméretű adathalmaz esetén rendkívül megnő a módszer futási ideje, emiatt az elmúlt években a kutatók kísérletek tettek az eljárás matematikai háttérének egyszerűsítésére, illetve klasszifikációs képességének pontosítására (Yang [2007], Peng et al. [2008], Yu et al. [2011], Kim et al. [2012]).

A neurális hálókat és az SVM módszert a szakirodalomban a mesterséges intelligenciára épülő eljárások közé sorolják. Ezek előnye, hogy anélkül képesek a függő és a független változók közti komplex nemlineáris kapcsolatrendszer feltárására, hogy annak formájára vonatkozóan bármilyen előzetes feltevessel éljenek. Más szóval ezek a technikák a kapcsolatrendszert az adatok alapján próbálják feltárni (Marqués et al. [2012b]). Ennek köszönhetően ez a két technika a hagyományos statisztikai modellek klasszifikációs teljesítményét jelentősen meghaladja, azonban közös sajátosságuk a fekete doboz jelleg, amely arra utal, hogy

a modellező csak a felhasznált adatokat és az azok eredményeképp kialakuló előrejelzést ismeri, arra a kérdésre nem kap választ, hogy mely mutatók milyen súllyal játszanak szerepet az előrejelzésben. A probléma Martens et al. [2010] szerint olyan súlyos, hogy emiatt a magasabb klasszifikációs teljesítmény ellenére sem várható, hogy ezek a módszerek általánosan elterjednének a hitelezési döntések meghozatalánál. Hasonló véleményt fogalmazott meg Lee-Choi [2013] is, akik szerint az értelmezhetőség hiánya jelentősen behatárolja e módszerek gyakorlati alkalmazhatóságát a vezetői döntések meghozatalában. Vélhetően e negatívumnak tudható be, hogy a bankok még napjainkban is az egyszerűbb hagyományos statisztikai módszereket alkalmazzák belső minősítési rendszereikben (Riberio et al. [2012]).

A modellek interpretálhatóságából eredő probléma az SVM eljárásnál megoldatlan, a neurális hálónál azonban történtek kísérletek a hiányosság áthidalására. Akkoc [2012] például a fuzzy megközelítéssel kombinálta a módszert; a már elkészült neurális hálók súlyaiból kalkulálható mutatók alapján az egyrétegű hálók értelmezésére pedig Trinkle-Baldwin [2007] tett kísérletet. A próbálkozások ugyan ígéretes eredményeket hoztak, ez a kutatási irány azonban a csődelőrejelzés területén nem képvisel domináns súlyt, s így a neurális hálók továbbra is a fekete dobozként jellemezhető eljárások közé soroltatnak.

Az egyedi klasszifikációs módszerek ismertetésénél végezetül meg kell még említeni Serrano-Cinca – Gutiérrez-Nieto [2013] munkáját, akik a különböző klasszifikációs módszerek segítségével felállított modellek előrejelzései alapján magukat a módszereket klaszterezték azzal a céllal, hogy megvizsgálják, mutatnak-e hasonlóságot előrejelzéseiket illetően. Eredményeik arra engednek következtetni,

hogy a hasonló elven működő módszerek hasonló vállalatokat sorolnak az egyes csoportokba. Az idézett szerzők által azonosított klaszterek:

- döntési fát felállító módszerek;
- lineáris klasszifikációra épülő eljárások: diszkriminancia-analízis, SVM;
- logisztikus regresszió, neurális hálók.

2.3.4.3. Módszerkombinációk és meta módszerek

Az előző szakaszokban bemutattam a csődelőrejelzésben leggyakrabban alkalmazott klasszifikációs módszerek elméleti alapjait, valamint azok előnyeit és hátrányait. Matematikai modellekről lévén szó, mindegyik eljárás szenved valamilyen hiányosságtól, és mindegyik bír előnyös sajátossággal is. Ez a tény pedig felveti a módszerek kombinációjában rejlő lehetőségeket, amelyek kutatása napjainkban egy intenzíven fejlődő trend a csődelőrejelzésben (Cao [2012a]). Az e körbe tartozó kutatások alapvetően két csoportba sorolhatók:

- meta módszerek: *egyetlen* kiválasztott klasszifikációs eljárást alkalmaznak többször, majd azok eredményeit kombinálva próbálják javítani a választott eljárás önálló használatával kapott eredményt;
- módszerkombinációk: *több különböző* klasszifikációs módszer eredményeit integrálják a még pontosabb előrejelzés érdekében.

Kim-Kang [2012] megfogalmazását alapul véve: a meta módszerek olyan gépi tanulási technikák, amelyek az önállóan gyengébb teljesítményt mutató klasszifikációs módszerek hatékonyságát fokozzák azok többszöri alkalmazásával. A két leggyakrabban alkalmazott meta módszer: a Boosting és a Bagging. Pontos

magyar megfelelő hiányában az angol elnevezéseikkel utalok ezekre az eljárásokra a továbbiakban.

A Bagging eljárás lényege, hogy a rendelkezésre álló tanuló adathalmazból véletlenszerűen kiválasztásra kerül (visszatevéssel) n darab részminta, amelyeken az eljárás lefuttatja a felhasználó által alkalmazni kívánt klasszifikációs módszert. Az n darab részmintán felállított modell besorolásainak egyszerű számtani átlagából adódik a végső modell előrejelzése az egyes megfigyelések vonatkozásában.

A Boosting eljárás az előzőtől csak abban különbözik, hogy a részminta-vétel szekvenciálisan történik oly módon, hogy az első részmintán futtatott modell által rosszul besorolt megfigyelések a következő részminta-vételi körben a helyesen besorolt megfigyelésekhez képest nagyobb esélyt kapnak arra, hogy újra bekerüljenek a mintába. A részminták számát a felhasználó szabályozhatja. Az egyes megfigyelések mintába kerülési esélye ennél a módszernél folyamatosan ingadozik annak megfelelően, hogy a legutóbb felállított modell az adott megfigyelést helyesen sorolta-e be. A Boosting eljárás módszere tehát arra törekszik, hogy a későbbi futtatások során „kijavítsa” az egyes lépések helytelen besorolásait. Ez önmagában előnyös tulajdonság, azonban a gyakorlati alkalmazás során az eljárás teljesítménye gyakran elmarad a Bagging-hez képest. Ennek oka, hogy előbbi hajlamos túlságosan nagy súlyt adni az előző lépésekben helytelenül minősített – gyakran outlier – megfigyeléseknek, amelyek általában nehezen klasszifikálhatók; ez pedig az egész modell „torzulásához” vezet, gyengítve annak diszkrimináló erejét (Kim-Kang [2012]).

A meta módszerek hatékonyságának forrása a részmintákon felállított modellek diverzitása. Az előbbieken bemutatott két megközelítés ezt a változékonyságot a

megfigyelési egységek köréből vett véletlen mintákból nyeri. Erre azonban akkor is lehetőség van, ha nem a megfigyelési egységekből, hanem a lehetséges magyarázó változók köréből veszünk részmintákat és azok alkalmazásával állítunk fel modelleket, melyek eredményeinek egyszerű számtani átlagaként ezúttal is meghatározható az egyes megfigyelések végső minősítése. E csoportban is két módszer alkalmazása terjedt el a gyakorlatban: a véletlen alterek (random subspace) és a random forests¹⁴ eljárás. Ezek módszertani hátterét terjedelmi okok miatt nem fejtem ki részletesebben, viszont az alfejezetben hivatkozott tanulmányok részletes áttekintést adnak ezekről a módszerekről is.

Napjaink aktuális kutatási iránya a különböző meta módszereket kombinálja a még jobb teljesítmény elérése érdekében. Marqués et al. [2012a] több lehetséges kombináció közül a Bagging és a véletlen alterek együttes alkalmazását találta a legjobbnak. Ugyanezt a párosítást alkalmazta Wang-Ma [2012] is az SVM módszer teljesítményének további fokozása érdekében. Ugyanez a szerzőpáros korábban folytatott kísérleti kutatásokat a Boosting és a véletlen alterek kombinációjából eredő lehetőségek vonatkozásában is (Wang-Ma [2011]).

Kérdéses lehet azonban, hogy melyik klasszifikációs módszert érdemes alkalmazni a meta módszereken belül. A választ Marqués et al. [2012b] próbálta megadni. Kutatásuk során számos klasszifikációs algoritmust kipróbáltak több különböző adathalmazon. Számításaik végül arra az eredményre vezettek, hogy a legjobb teljesítményt a C4.5 eljárás mutatja, így indokolt lehet alkalmazásuk a meta módszerek esetén. Az eredmény összhangban van Kim-Kang [2012] azon megállapításával, mely szerint a meta módszereket az instabilabb klasszifikációs teljesítményt mutató módszereknél lehet hatékonyan alkalmazni. Az egyes

¹⁴ Az eljárás nevére tudomásom szerint nincs pontos magyar megfelelő.

részmintákon adódott besorolások korrelációjának elemzése alapján az idézett szerzők arra jutottak, hogy a neurális hálók és az SVM jellemzően stabilabb besorolási teljesítményt mutat, a döntési fákra épülő eljárások esetén azonban nagyobb mértékű szóródás adódott. Feltételezésüket, miszerint a döntési fák esetében a meta módszerek hatékonyabbak, empirikus kutatási eredményeik is alátámasztották.

A másik megközelítés (a módszerkombinációk) a különböző klasszifikációs eljárások eredményeit kombinálja az egyedi módszerekkel elérhető pontosság növelése érdekében. Az ötletet már Olmeda-Fernandez [1997] korábban felvetette, intenzívebben azonban csak napjainkban foglalkozik a kérdéssel a szakirodalom. A későbbiekben Kiang [2003] is javasolta, hogy abban az esetben, ha az egyes módszerek különböző megfigyeléseket tévesztenek, lehetőség nyílik azok kombinációjára. Yim-Mitchell [2005] eredményei alátámasztották az előtte idézett két szerző feltevését. A diszkriminancia-analízis, a logisztikus regresszió és a probit regresszió előrejelzéseit használta fel a neurális hálók input változóiként, s az így kapott modell teljesítménye meghaladta az egyedi módszerekkel elért eredményeket. Napjainkban Cao [2012b] végzett kutatásokat e kérdés vonatkozásában. Az idézett szerző számos klasszifikációs módszer outputját kombinálta össze egyetlen eredménnyé, amely szignifikánsan magasabb találati arányt eredményezett, mint bármelyik eljárás önálló alkalmazása.

2.3.4.4. Módszertani összehasonlító elemzések eredményei

A statisztikai módszerek helyét az előző szakaszban bemutatott nemparaméteres technikák a 90-es évektől kezdték átvenni (Feki et al. [2012]). Körülbelül erre az időszakra tehető a csődelőrejelzésben a kutatási főáram kialakulása, amely azzal a kérdéssel foglalkozik, hogy a számos klasszifikációs alternatíva közül melyik nyújtja a legjobb teljesítményt. Annak ellenére, hogy egyre gyakrabban olvasható az a szakirodalomban, hogy vélhetően nem létezik „legjobb” megközelítés (Marqués et al. [2012a]), a legtöbb tanulmány még napjainkban is a különböző technikák összehasonlító elemzésével foglalkozik (Sánchez-Lasheras et al. [2012]). Ezek közül az alábbi táblázatban foglalom össze néhányat a teljesség igénye nélkül.

1. táblázat: Módszertani összehasonlító elemzések eredményei a nemzetközi szakirodalomban

Szerző	Alkalmazott módszerek ¹⁵	Legjobb teljesítményt nyújtó eljárás ¹⁶
Salchenberger et al. [1992]	NN, LA	NN
Tam-Kiang [1992]	NN, DA, LA, KNN	NN
Coats-Fant [1993]	NN, DA	NN
Altman et al. [1994]	NN, DA	NN
Wilson-Sharda [1994]	NN, DA	NN
Alici [1995]	NN, DA, LA	NN
Back [1996]	NN, DA, LA	NN
Leshno-Spector [1996]	NN, DA	NN
Tan [1999]	NN, PA	NN
Zhang et al. [1999]	NN, LA	NN
Fan-Palaniswami [2000]	NN, DA, SVM, LVQ	SVM
McKee-Greenstein [2000]	NN, LA, ID3	ID3
Neophytou et al. [2000]	NN, DA, LA	NN
Huang et al. [2004]	NN, LA, SVM	SVM
Min-Lee [2005]	NN, DA, LA, SVM	NN, SVM
Ding et al. [2008]	NN, DA, LA, SVM	SVM
Yoon et al. [2008]	NN, DA, LA, CART, C5	NN, SVM
Kim-Sohn [2010]	NN, LA, SVM	SVM
Lee-To [2010]	NN, SVM	SVM
Moro et al. [2011]	LA, SVM	SVM
Xiaosi et al. [2011]	NN, LA, SVM	SVM
Bae [2012]	NN, DA, LA, SVM, C5, BC	SVM
Erdal-Ekinci [2012]	NN, SVM	SVM
Lee-Choi [2013]	DA, NN	NN

A táblázatban közölt kutatási eredmények felsorolása közel sem tekinthető teljesnek, ugyanis a főáramba sorolható kutatási témakörben megjelenő publikációk száma hatalmas. Ez a rövid időrendi válogatás azonban hűen tükrözi a szakirodalomban tapasztalt legfontosabb tendenciákat. A 2000-es évek elejéig a neurális hálók teljesítményét állították szembe a hagyományos statisztikai modellek

¹⁵ Az egyes módszerek rövidítéseinek jelentése:

DA: diszkriminancia-analízis

NN: neurális hálók

LA: logisztikus regresszió

C5: a C4.5 módszer továbbfejlesztett változata

BC: Bayes klasszifikáció

CART: klasszifikációs és regressziós fák

ID3: a C4.5 eljárás kezdeti változata

LVQ: tanuló vektor kvantálás (learning vector quantization)

KNN: k legközelebbi szomszéd

¹⁶ Ahol két eljárás is szerepel, ott nem mutatkozott szignifikáns eltérés a módszerek teljesítménye között.

teljesítményével. A kutatási eredmények túlnyomó többségében a neurális hálók nyújtottak jobb teljesítményt. Az SVM módszer 2000-es évek elején kezdett megjelenni a csődelőrejelzésben. Ettől kezdődően a fő kutatási irányt e két mesterséges intelligenciára épülő eljárás összehasonlítása képezte, melynek eredményei döntő többségben az SVM módszert mutatta preferáltabb választásnak. A mesterséges intelligencia eszközeit alkalmazó tanulmányokat tekinti Yu et al. [2014] a csődelőrejelzés főáramának. Fontos azonban itt szót ejteni Serrano-Cinca – Guttiérrez-Nieto [2013] észrevételéről, melyben arra hívják fel a figyelmet, hogy a szakirodalom felülreprezentálja azokat a kutatási eredményeket, amelyek arról számolnak be, hogy a korszerűbb módszerek teljesítménye meghaladja a régebbi eljárásokkal elérhető eredményeket. Továbbá figyelembe kell venni a mesterséges intelligenciára épülő módszerek összehasonlító elemzéseinél azt is, hogy azok teljesítménye nagymértékben függ a felhasználó által beállítandó paraméterek értékétől. Ebben az esetben viszont kérdéses lehet, hogy mennyire fektetett hangsúlyt a kutató az egyes eljárások esetén arra, hogy meghatározza azokat a paramétereket, amelyek maximalizálják a felállított modellek találati arányát.

Az 1. táblázat az elmúlt 3-4 év fejleményeit nem tartalmazza. Az utóbbi időben a módszerkombinációk és a meta módszerek kutatása vált dominánssá. Ezen összetett technikák pontossága jellemzően meghaladja az egyedi eljárásokkal elérhető találati arányt. Így összességében napjainkban ezek tekinthetők a csődelőrejelzés legkorszerűbb módszertani megközelítéseinek.

2.3.5. A modell teljesítményének értékelése

A csődelőrejelző modellek teljesítményének mérése alapvető követelmény a tudományban és a gyakorlati alkalmazás szempontjából egyaránt annak érdekében, hogy megítélhessük, hogy a felállított modellek teljesítménye meghaladja-e a véletlen találgatással elérhető eredményeket, illetve hogy képet kaphassunk arról, hogyan viszonyul egy újabb modell teljesítménye a már meglévő modellekéhez képest. A szakirodalomban a modellek teljesítménymutatóinak széles köre áll rendelkezésre. Terjedelmi korlátok miatt a dolgozatban csak a legfontosabb, illetve leggyakrabban alkalmazott mutatószámokat ismertetem részletesebben.

2.3.5.1. A klasszifikációs teljesítmény mutatói

A csődmodellek jellemzően előrejelzési céllal készülnek, így teljesítményük legfontosabb mutatója a találati arány, amelyet a helyesen besorolt megfigyelések összes megfigyeléshez viszonyított arányaként határozhatunk meg. E mutatót jellemzően a klasszifikációs mátrix értékei alapján kalkulálják, melynek általános alakja az alábbi.

8. ábra: Klasszifikációs mátrix

		Előrejelzett csoport	
		Csődös	Működő
Megfigyelt csoport	Csődös	a	b
	Működő	c	d

A modell találati arányát az alábbi összefüggés adja meg:

$$\frac{a + d}{a + b + c + d}$$

A mátrix értékei alapján további modellteljesítmény-mutatók is kalkulálhatók. A szakirodalom kiemelt figyelmet szentel az elsőfajú hiba nagyságának, amelyet az alábbi aránnyal adhatunk meg:

$$\frac{b}{a + b}$$

Ez a hányados azt számszerűsíti, hogy a csődös megfigyeléseknek hány százalékát sorolja be a modell helytelenül. Du Jardin [2010] szerint a csődelőrejelzésnek elsősorban e hiba minimalizálására kell törekednie, mivel a csődös cégek működő csoportba történő besorolása a hitelezők szempontjából rendkívül költséges döntések meghozatalát eredményezheti abban az esetben, ha az érintett vállalkozás a későbbiekben valóban csődbe megy. Sokkal alacsonyabb költség társul a másodfajú hibához, melynek nagyságát az alábbi mutató fejezi ki:

$$\frac{c}{c + d}$$

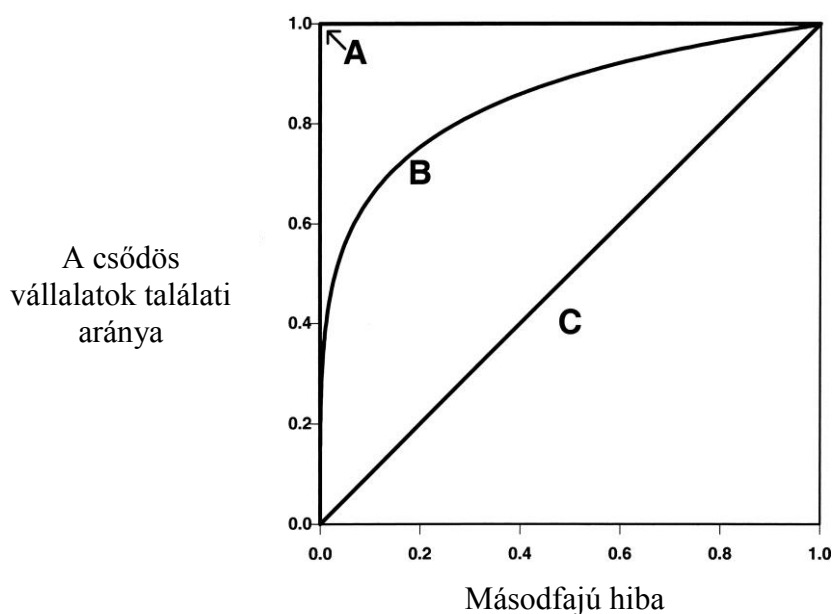
Ez a működő vállalatokon belül azon megfigyelések aránya, amelyet a modell csődösnek minősített. Hitelezői szempontból ez a hiba kevésbé költséges, mivel hitel kihelyezési döntések vonatkozásában ez azzal a következménnyel járhat, hogy az érintett vállalkozásnak nem nyújtanak hitelt, így másodfajú hiba esetén a hitelező az ügylet lehetőségköltségét veszíti el; szemben az elsőfajú hibával, ahol a hitel bedőlése esetén a hitelező a kihelyezett összeg egy részét, de akár egészét is elvesztheti. A két hiba költségeinek arányára vonatkozóan csak becslések vannak a nemzetközi szakirodalomban, ahol a modellezők többsége – pontos adatok

hiányában – a modellépítés során nem veszi figyelembe a kétféle hiba eltérő költségét, viszont tekintettel arra a tényre, hogy a hibákhoz kapcsolódó költségek aszimmetrikusak, a kutatók törekszenek olyan modellek felállítására, amelyek a lehető legkisebb mértékben követik el a költségesebb elsőfajú hibát.

Meg kell említeni azonban néhány kutatót, akik a csődelőrejelzésben elkövethető hibák aszimmetrikus költségének fontosságára hívják fel a figyelmet oly módon, hogy modelljük céljaként nem a találati arány maximalizálását, hanem a téves besorolás (várható) összköltségének minimalizálását tűzik célul. Ebből a szempontból többek között Nanda-Pendharkar [2001] és Chen et al. [2011] munkái említhetők.

A csődmodellek klasszifikációs teljesítményét gyakran szemléltetik és mérik a ROC¹⁷ görbével melynek általános alakját az alábbi ábra mutatja.

9. ábra: ROC görbe



¹⁷ Receiver Operating Characteristic

A B-vel jelölt ROC görbe a modell által elkövetett másodfajú hiba és a csődös vállalatok találati aránya közti kapcsolatot számszerűsíti. Két modell közül azt tekinthetjük jobbnak, amelyik a másodfajú hiba egy adott szintje mellett a lehető legnagyobb találati arányt képes elérni a csődös vállalatok körében. Ilyen esetben az ábrán a B-vel jelölt görbe eltávolodik a C-vel jelölt átlótól, amely a véletlen találgatásnak felel meg. Minél nagyobb az eltávolodás mértéke, annál jobb klasszifikációs teljesítmény jellemzi a modellt. Ennek számszerűsítésére gyakran használt mutató a ROC görbe alatti terület (lásd például Blanco et al. [2013] munkáját), amely a ROC görbe és a C-vel jelölt 45 fokos egyenes közti terület nagysága. A ROC görbe alatti terület az A-val jelölt görbe esetén maximális, ekkor tökéletes modelltől beszélhetünk.

Ezek mellett gyakori még a Brier pontérték használata is, melyre akkor van lehetőség, amikor a klasszifikációs eljárás segítségével a csőd valószínűsége is számszerűsíthető. A Brier pontérték az alábbi formula szerint kalkulálható (Lin et al. [2011]):

$$BS = \frac{\sum_{i=1}^n (p_i - o_i)^2}{n}$$

ahol a p_i a modell által becsült csődesélyt, az o_i a megfigyelés eredeti csoporttagságát (1 csődös, 0 működő vállalat esetén), az n pedig a megfigyelések számát jelöli.

A bemutatott teljesítménymutatók a leggyakrabban alkalmazott eszközök a modellek klasszifikációs képességének megítélésében. Ezek mellett azonban számos egyéb mutató is rendelkezésre áll, illetve elterjedt a szakirodalomban, azonban ezekre terjedelmi okok miatt nincs lehetőségem részletesen kitérni. Említés szintjén azonban

szót érdemel a Gini mutató és a CAP¹⁸ görbe is, ugyanis ezek a gyakorlati modellezésben, illetve a tudományos szakirodalomban is gyakran használt mérőszámok.

2.3.5.2. Validációs eljárások

Az előbb bemutatott teljesítménymutatókat alkalmazhatjuk a modellépítéshez felhasznált tanulási mintán, valamint egy olyan mintán, melynek elemei nem szerepeltek a tanuló mintában. Utóbbit tesztelő mintának nevezik a szakirodalomban. Előrejelző modellekről lévén szó, a modellek predikciós erejét a tesztelő mintán elért eredmény alapján becsülhetjük. Fontos azonban hangsúlyozni, hogy ez csak becslés, mivel az ott kapott eredmény annak függvénye, hogy milyen megfigyelések szerepelnek a tesztelő mintában. A modellek előrejelző képességének értékelését validáció néven ismeri a szakirodalom, így az erre a célra használt mintát szokás validációs mintának is nevezni.

Abban az esetben, ha modellek teljesítménymutatóit a tanuló mintára vonatkozóan számszerűsítjük, a modell ex post validációjáról, a tesztelő a minta esetén a modell ex ante validációjáról beszélünk. Utóbbi csoportban az tekinthető a modell előrejelző képességére vonatkozó legszigorúbb próbának, amely a modellépítéshez felhasznált adatoknál későbbi megfigyelésekre épül. Ez azt jelenti, hogy ha a modellépítéshez például 2010-2012 közti adatokat használunk, akkor a modell teljesítményét célszerű 2013-as adatokon ellenőrizni. A csődelőrejelzés első évtizedeiben a kutatók gyakran azzal szembesültek, hogy a modellek időben későbbi adatok esetén jóval gyengébb eredményt mutatnak (Platt-Platt [1990]), melyre magyarázatként a pénzügyi mutatók

¹⁸ Cumulative Accuracy Profiles

időbeli instabilitását¹⁹ jelölték meg; a kérdés még napjainkban is aktuális, melyre Berg [2007] hívta fel a figyelmet, azonban a csődelőrejelzés főárama gyakran nem fordít gondot erre a problémára.

Fontos felhívni a figyelmet arra, hogy a csődelőrejelzésben kétféle modellezési megközelítés van jelen. Az egyik a gyakorlati adósmínősítési folyamathoz áll közelebb. Ebben a megközelítésben a csődelőrejelzés egy adatbányászati (adatillesztési) feladatként értelmezhető (Darayseh et al. [2003]), melynek célja a lehető legmagasabb előrejelző képességű modell felállítása. A másik megközelítés inkább az akadémiai világban elterjedt, ugyanis a tudományos vizsgálatok célja nem feltétlenül egy konkrét előrejelző modell felállítása, hanem gyakran valamely koncepció hatékonyságának demonstrációja a csődelőrejelzésben. Ilyenkor a kutatók azt kívánják igazolni, hogy az általuk javasolt koncepcióval felállított modell teljesítménye nem csak egy egyedi tanuló és tesztelő minta esetén jobb, mint az alternatív modelleké, hanem nagy valószínűséggel ez a tendencia érvényes akkor is, ha más tanuló és tesztelő mintákra alkalmazzák a javasolt új módszert vagy megközelítést. Az ilyen jellegű kutatások esetén a mintát többször is felosztják véletlenszerűen. A szakirodalomban ez a többszörös validáció az alábbi háromféle formában jelenik meg.

1. Egyet kihagyó módszer (leave one out, Jackknife módszer)

Nevéből adódóan az eljárás az n elemű tanuló minta adatai közül $n-1$ darabot használ fel tanuló mintaként és a kihagyott egy darabot tesztelő elemként. Az eljárást n -szer ismételjük annak érdekében, hogy minden megfigyelés szerepeljen egyszer a tesztelő

¹⁹ Lásd részletesebben a 2.3.2.1. fejezetet.

elem szerepében. A modell klasszifikációs teljesítményét pedig a kihagyott elemekre adott előrejelzések alapján számított besorolási pontosság alapján ítéldhetjük meg.

Az eljárás leginkább akkor előnyös, ha a megfigyelések száma alacsony, ami miatt nincs lehetőség külön tesztelő minta alkalmazására. Wang [2004] a módszer másik előnyös sajátosságának tekinti, hogy alig csökkenti a tanuló minta méretét, ami azért hasznos, mert több megfigyelés alapján megbízhatóbb modellek felállítására van lehetőség. Ez inkább a neurális hálók esetén fontos, melynek teljesítménye érzékenyebb a megfigyelések számára (Leshno-Spector [1996]). A validációs technika hátránya, hogy nagyméretű adathalmaz, illetve bonyolult klasszifikációs módszer alkalmazása esetén rendkívül számításigényes. Emiatt a gyakorlatban ritkán alkalmazzák.

2. Keresztvalidációs eljárás

A legelterjedtebb megközelítés a modellek teljesítményének becslésére a keresztvalidációs eljárás. Ennek lényege, hogy a rendelkezésre álló mintát n egyenlő részre osztják fel, melyek közül $n-1$ részminta szolgál a modell felállítására, míg a kihagyott részlet a modell teljesítményének értékelésére. Az eljárást n -szer alkalmazzák annak érdekében, hogy minden egyes részminta szerepeljen egyszer a tesztelő minta szerepében. Ezt követően a modell teljesítményét az n darab tesztelő mintán elért találati arány átlagával számszerűsíthetjük. A megközelítés alkalmazására számtalan példa található a szakirodalomban a 2000-es évektől kezdődően (lásd például Ahn et al. [2000], illetve Harris [2013] munkáit).

Fontos hangsúlyozni, hogy a keresztvalidációs technikával becsült modellteljesítmény is érzékeny lehet arra, hogy a rendelkezésre álló minta milyen

osztópontok segítségével került felosztásra n egyenlő részre. Annak érdekében, hogy egy véletlenszerűen kiválasztott osztópont-sor ne torzíthassa a modell becsült előrejelző képességét, alkalmazható a k -szor n -szeres keresztvalidáció. Ennek lényege, hogy az n egyenlő részmintára történő felosztást megismétlik k alkalommal, minden esetben más osztópont-sort alkalmazva. A modell teljesítményét a $k \times n$ darab részmintán elért találati arányok átlagában ítéldük meg ebben az esetben. A technika alkalmazására példát Alfaro et al. [2008] munkája nyújt.

Végezetül meg kell jegyezni, hogy a mintafelosztás nagyságára (a k illetve az n értékére) nincs objektív iránymutatás a szakirodalomban, azok értékeit a modellező határozza meg annak figyelembevételével, hogy a tesztelési céllal kihagyott minták mérete elég nagy legyen érdemi következtetések levonásához a modell teljesítménye tekintetében. A szakirodalmi áttekintés alapján az n értéke jellemzően 5 és 50 között mozog.

3. Többszörös véletlenszerű tanuló-tesztelő eljárás

Az előző módszerek mellett gyakran használatos még a minta többszörös véletlenszerű felosztása tanuló és tesztelő mintákra a modellező által rögzített arányban. Erre példát Min et al. [2011] tanulmánya szolgáltat. Az idézett szerzők 80:20 arányban osztották fel a rendelkezésükre álló adathalmazt tanuló és tesztelő mintákra 100 darab véletlenszerűen kiválasztott osztópont alkalmazásával. Ebben az esetben a modell előrejelző képességét a 100 darab tesztelő mintán elért találati arány átlagával adhatjuk meg. E módszer esetén szintén nincs objektív iránymutatás arra vonatkozóan, hogy milyen felosztást érdemes alkalmazni és azt hányszor célszerű ismételni. A 80:20, 70:30, 60:40 és 50:50 arányban történő felosztás közti

különbségeket több klasszifikációs módszer tekintetében Hu-Tseng [2007] és Hu [2009] vizsgálta. Eredményeik felhívják a figyelmet arra, hogy egyes módszerek esetén a modellek becsült előrejelző képességében tapasztalt különbség elérheti az 5 százalékpontot a felosztási aránytól függően. Arra a kérdésre azonban nem találtak egyértelmű választ, hogy melyik arányt célszerű választani, mivel az eredmények ellentmondásosak voltak. Ebből adódóan továbbra is egyetlen szempont marad a gyakorlati elemzés számára: a felosztási arány meghatározásánál figyelembe kell venni, hogy a tesztelő minta mérete elégséges legyen a modell előrejelző képességének megítéléséhez, a tanuló minta elemszáma pedig egy adekvát modell felállításához.

Az előző két validációs megközelítés ugyan rendkívül elterjedt, néha ezeket is bírálataz abból adódóan, hogy mindkét eljárás esetén egy tanuló mintán felállított modell teljesítményét egy tesztelő minta alapján ítélik meg, ami felveti annak lehetőségét, hogy a modellt – a tanuló minta által lehetővé tett mértékben – túlságosan a tesztelő mintán elért eredményre optimalizálják. E probléma elkerülése érdekében Chen et al. [2011] azt a megoldást választotta, hogy a keresztvalidációs eljáráson belül a mintát három részre osztotta: tanuló, tesztelő és validációs mintára 60:20:20 arányban. A modell felállítása a rendelkezésre álló megfigyelések 60 %-ának felhasználásával történt, oly módon, hogy a paramétereket úgy optimalizálták, hogy a tesztelő minta eredménye maximális legyen, majd a modell előrejelző képességét az addig nem használt 20 %-nyi validációs mintán elért eredmény alapján becsülték. A megközelítés előnye, hogy ekkor a modellteljesítmény felmérése olyan adatokon történik, amelyeket egyáltalán nem vettek figyelembe a modell felállítása során, melynek köszönhetően a mintaspecifikus eredmények esélye minimális lehet.

Zárásképp fontos megjegyezni, hogy a szakirodalomban nincs egységes álláspont a fenti validációs módszerek közti választást illetően sem. A leggyakrabban a keresztvalidációs technika kerül alkalmazásra, népszerűség tekintetében a második helyet a minta többszörös véletlenszerű felosztása foglalja el. Az egyet kihagyó megközelítés nagyon ritka, amely vélhetően a nagy számításigénynek tudható be. Az első két módszer esetén azonban érdekes, hogy a kutatók nem indokolják meg, hogy miért épp az egyik, vagy a másik módszert használják. Felmerülhet a kérdés, hogy tapasztalható-e közöttük érdemi eltérés, optimistább, vagy pesszimistább becslést ad-e a két módszer közül valamelyik a modellek előrejelző teljesítményére. A kérdésre magam is kerestem empirikus választ. A kutatás részleteit és eredményeit terjedelmi okok miatt nincs lehetőségem részletesen ismertetni, erre vonatkozóan egy másik munkámat (Nyitrai [2014]) ajánlom az Olvasó figyelmébe.

2.4. A csődelőrejelzés hazai szakirodalma

Magyarországi viszonyok között a csődelőrejelzés tudományos vizsgálatának lehetőségét a rendszerváltást követően 1991-ben a csődtörvény megjelenése teremtette meg (Virág et al. [2013]). Az első hazai csődmodellt Virág [1993] publikálta. Az azóta eltelt több mint két évtized során a csődelőrejelzés magyar vonatkozású szakirodalma is egyre jelentősebbé vált a gazdaságtudományi területen kompetens hazai folyóiratokban. A témakörben megjelent hazai publikációk eredményeit ismertetem ebben a fejezetben. Mivel a csődelőrejelzés magyar szakirodalma jóval szűkebb, mint a nemzetközi, továbbá amiatt, hogy az főleg a főáramú kutatási irány köré koncentrálódik, nincs lehetőség a dolgozatban eddig alkalmazott keresztmetszeti rendező elv alkalmazására. Helyette időrendi sorrendben

ismertettem a hazai tudományos folyóiratokban publikált csődelőrejelzéssel foglalkozó tanulmányok legfontosabb eredményeit.

2.4.1. A csődelőrejelzés kezdetei Magyarországon

A hazai csődelőrejelzés alapköveit Virág-Hajdu [1996] fektették le. Munkájukban 156 hazai iparvállalat adatait használták fel, melyekből 17 pénzügyi mutatószámot határoztak meg a mintába került megfigyelések 1990. és 1991. évi beszámolói alapján. A modellezéshez az idézett szerzők a lineáris diszkriminancia-analízist és a logisztikus regressziót alkalmazták, melyekkel rendre 77,9, illetve 81,8 %-os találati arányt értek el.

Ugyanez a szerzőpáros elkészített egy olyan csődmodell-családot is, amely a nemzetgazdaság egészére, tíz nemzetgazdasági ágra és 30 ágazatra vonatkozóan specifikált kiemelkedően magas találati arányokkal jellemezhető csődmodelleket több mint 10000 hazai vállalkozás adatai alapján. E szerzőpáros hívta fel a figyelmet először a hazai szakirodalomban arra, hogy a csődmodellek előrejelzéseit inkább tekinthetjük korai figyelmeztetésnek, mintsem szoros értelemben vett előrejelzésnek. Véleményük szerint a modellek eredményei azt tükrözik, hogy az egyes megfigyelések mennyire hasonlítanak egyik, vagy a másik csoport általános sajátosságaihoz.

2.4.2. Neurális hálók a hazai csődelőrejelzésben

A módszertani fejlődés nem hagyta érintetlenül a csődelőrejelzés hazai szakirodalmát sem. A 2000-es évek első évtizedének közepén Magyarországon is megjelentek a nemzetközi kutatásban főáramot képviselő módszertani összehasonlító elemzések,

melyek eredményei jellemzően összhangban vannak a globális szakirodalomban tapasztalt tendenciákkal.

A statisztikai modellek helyét az 1990-es évektől kezdték átvenni a mesterséges intelligenciára épülő eljárások. Hazánkban ez a folyamat a 2000-es évek első évtizedére tolódott. Az első hazai csődmodell adatbázisán Virág-Kristóf [2005] végzett összehasonlító elemzéseket, melyek célja annak vizsgálata volt, hogy hazai viszonyok közt is versenyképes alternatívát jelentenek-e a neurális hálók a hagyományos matematikai-statisztikai módszerekkel szemben.

A későbbiekben ugyanez a szerzőpáros szintén az első hazai csődmodell adatbázisát használta egy még kiterjedtebb komparatív elemzés lefolytatására, amelyben négy klasszifikációs módszer²⁰ előrejelző képességét vetették össze a Platt-Platt [1990] által javasolt iparági relatív mutatók alkalmazása mellett. Eredményeik ebben az esetben is azt mutatták, hogy a mesterséges intelligencia módszerei a magyarországi vállalatok körében is hatékonyabban képesek a csőd előrejelzésére, mint a hagyományos módszerek (Virág-Kristóf [2006]).

2.4.3. A legkorszerűbb módszerek megjelenése a hazai csődelőrejelzésben

A módszerkombinációk alkalmazása egy új és fejlődő trend a csődelőrejelzésben, ami a 21. század első évtizedének végén a hazai szakirodalomban is megjelent Virág-Kristóf [2009] munkája révén. Kismintás elemzésükben a fizetőképes és fizetéseképtelen megfigyeléseket, mint többdimenziós koordinátákat egy alacsonyabb dimenziószámú térbe képezték le többdimenziós skálázás segítségével, majd a

²⁰ diszkriminancia-analízis, logisztikus regresszió, döntési fák, neurális hálók

redukált dimenziószámú térben alkalmazták a logisztikus regresszió módszerét kiemelkedő találati aránnyal.

A csődelőrejelzés hazai szakirodalma napjainkra elérte a nemzetközi szakirodalom színvonalát a vizsgált kutatási kérdések és az alkalmazott módszerek tekintetében egyaránt. A főáram módszertani összehasonlító elemzései mellett ugyanis Kristóf-Virág [2012] révén megjelentek az adatelőkészítési feladatok fontosságát hangsúlyozó munkák is. Az idézett szerzők a CHAID módszer segítségével diszkrétizált, valamint a főkomponens-elemzés segítségével létrehozott mesterséges változók előrejelző képességét vizsgálták a csődelőrejelzésben leggyakrabban használt módszerek tekintetében. Eredményeik arra utalnak, hogy az adatok modellezésre történő előkészítése a vizsgált két megközelítés vonatkozásában módszerfüggő: különböző klasszifikációs eljárások esetén eltérő hatást gyakorolt a két vizsgált megközelítés alkalmazása a modellek találati arányára. Fontos azonban hangsúlyozni, hogy sem a CHAID-alapú diszkrétizálás, sem a főkomponens-elemzés, sőt a kettő együttes alkalmazása sem rontott a modellek előrejelző képességén – az esetek döntő többségében javított kisebb-nagyobb mértékben.

A legkorszerűbb módszerek közé tartozó SVM eljárás alkalmazására hazai vállalati adatokon tudományos publikáció keretében először Virág-Nyitrai [2013] tett kísérletet az első hazai csődmodell adatbázisán. A választást az indokolta, hogy ezen az adathalmazon korábban már több módszertani összehasonlító elemzés is készült, így az SVM hatékonyságát először e kismintás elemzés keretében vetették össze a hagyományos statisztikai eljárások és a neurális hálók teljesítményével. Az eredmények ebben az esetben is összhangot mutatnak a nemzetközi szakirodalom alapján kiolvasható tendenciákkal, ugyanis az SVM módszer a tanuló és a tesztelő mintán egyaránt magasabb találati arányra volt képes az első hazai csődmodell

adatbázisán, mint a neurális hálók és a hagyományos matematikai-statisztikai módszerek.

A korábban már hivatkozott halmazelméleten alapuló rough set theory (RST) klasszifikációs módszer teljesítményét szintén az első hazai csődmodell adatait felhasználva vizsgálta Virág-Nyitrai [2014]. A módszertani háttérrel bemutató alfejezetekben bemutattam, hogy a gépi tanulásra épülő eljárások jellemzően fekete dobozok, ami jelentősen behatárolja gyakorlati alkalmazási lehetőségüket annak ellenére, hogy relatíve magas klasszifikációs teljesítmény jellemzi őket. A modellek értelmezhetősége ugyanis fontos szempont a gyakorlati életben. Ebből a célból Martens et al. [2010] a döntési fákat generáló eljárások alkalmazását tartja célszerűnek. E körbe sorolható a Virág-Nyitrai [2014] munkájában alkalmazott RST módszer is, melynek vonatkozásában arra a kérdésre is választ kerestek a szerzők, hogy le kell-e mondani a modellek értelmezhetőségéről csődelőrejelzés területén a magasabb klasszifikációs teljesítmény érdekében. Az első hazai csődmodell adatai alapján az eredmény azt mutatta, hogy a könnyen értelmezhető „ha-akkor” típusú szabályokat generáló RST módszert alkalmazva az SVM algoritmussal azonos találati arány érhető el, így nem szükségszerű az átváltás a modellek interpretálhatósága és teljesítménye közt.

3. KUTATÁSI HIPOTÉZISEK

Az értékezés első fejezetében behatároltam a választott tudományterület legfontosabb fogalmait, a második fejezetben pedig részletes áttekintést adtam a csődelőrejelzés szakirodalmi fejlődéséről és aktuális helyzetéről egy újszerű keresztmetszeti megközelítésben. E fejezetben bemutatom azokat a vizsgálni kívánt kutatási hipotéziseket, melyeket a csődelőrejelzés nemzetközi szakirodalmából kirajzolódó tendenciák alapján relevánsak lehetnek a tudományterület jövőbeli fejlődése szempontjából.

3.1. Nyitott kutatási kérdések a szakirodalomban

A szakirodalmi áttekintés során bemutattam, hogy a tudományterület fejlődését alapvetően a módszertani fejlődés motiválja, így a témában megjelenő publikációk döntő többsége a különböző klasszifikációs eljárások teljesítményét veti össze a csődelőrejelzés területén – véleményem szerint ez tekinthető a csődelőrejelzés kutatásában a főáramnak. A nemzetközi publikációk eredményei alapján megállapítható, hogy ezen a területen a modellek pontosságának növelhetősége korlátozott, ugyanis az utóbbi években megjelenő tanulmányokban a legkorszerűbb módszerek használatából adódó növekedés az előrejelző teljesítményben általában csak kismértékű, továbbá sok esetben ez a többlet csak fekete doboz jellegű klasszifikációs eljárások alkalmazásával realizálható. Az interpretálhatóság azonban fontos szempont a gyakorlati alkalmazás szempontjából, amit véleményem szerint a tudományos kutatásban is figyelembe kell venni.

Fontos ugyanakkor azt is hangsúlyozni, hogy a csődelőrejelző modellek teljesítményében már egy kismértékű javulás is számottevő haszonnal (Wang-Ma [2012]) járhat az azokat alkalmazók számára. Ennek érdekében azonban célszerű az összes lehetséges módot megragadni a modellek előrejelző képességének javítására. Az értekezés legfontosabb hipotézise, hogy a csődelőrejelző modellek találati aránya nem csak módszertani oldalról javítható. Erre lehetőséget adhat az input változók körében felhasznált pénzügyi-számviteli mutatókban rejlő információtartalom minél nagyobb mértékű kihasználása, amelynek egyik lehetséges módja a statikus pénzügyi mutatók dinamikájának figyelembevétele.

3.1.1. Dinamikus pénzügyi mutatók a csődelőrejelzésben

Az értekezés szakirodalmi áttekintésének vezérfonalát a csődmodell-építés 1. ábrán bemutatott folyamatok képezte. A legnagyobb tudományos érdeklődés évtizedek óta a csődelőrejelzésre alkalmas klasszifikációs módszerek összehasonlító elemzését övezi, melyhez viszonyítva a folyamat többi lépésével relatíve keveset foglalkoznak a szakirodalomban.

Kimondottan érvényes ez az adatok modellezésre történő előkészítésére, amely csak az utóbbi években kezdett nagyobb hangsúlyt kapni a tudományos folyóiratokban. A szakirodalmi áttekintésben számos szerzőt idéztem, akik a csődelőrejelző modellek statikus jellegét súlyos problémának tartják, melynek megoldása egy ígéretes jövőbeli kutatási irány lehet. A probléma megoldására Berg [2007] javasolt egyszerű megoldást: a mutatószámok időbeli változását is felhasználták a modellek független változói között; a modellek statikus jellegét pedig úgy próbálta feloldani, hogy a

legutolsó megfigyelt év pénzügyi mutatói mellett az azt megelőző két év változóit is alkalmazta modelljében.

Módszertani oldalról tett kísérletet a csődelőrejelzés dinamizálására Du Jardin-Séverin [2012], akik a pénzügyi mutatók időbeli trendjét önszerveződő térképek segítségével szemléltették és klasszifikálták annak érdekében, hogy azonosítsák a vállalatok tipikus csődfolyamatait.

Véleményem szerint hasonló dinamizálás megvalósítható olyan változók kalkulációjával is, amelyek azt számszerűsítik, hogy hogyan viszonyul egy adott vállalkozás pénzügyi mutatójának legutoljára megfigyelt értéke az azt megelőző időszak megfelelő értékeihez. E célra egy lehetséges formula az alábbi:

$$\frac{X_{i,t-1} - X_{i,\min_{[t-2;t-n]}}}{X_{i,\max_{[t-2;t-n]}} - X_{i,\min_{[t-2;t-n]}}}$$

ahol

X a pénzügyi mutatószám

i a vállalat azonosítója

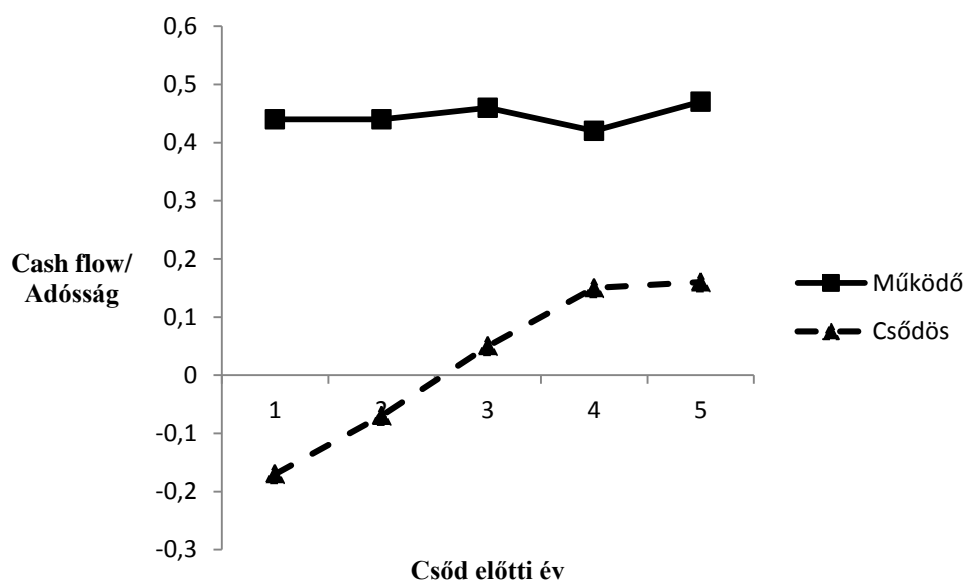
t az az év, amelyre vonatkozóan előrejelzést kívánunk adni

n a megfigyelt évek száma.

A fenti formulával definiált változókat a továbbiakban dinamikus pénzügyi mutatóknak (röviden dinamikus mutatóknak/változóknak) nevezem. Feltételezésem szerint ezek a változók releváns információt hordoznak a fizetőképesség és a fizetésképtelen megfigyelések megkülönböztetése céljából. A hipotézis alapját Beaver [1966] kutatási eredményei képezik. A csődelőrejelzés atyjának tekintett

szerző úttörő munkájában az alábbi ábrán szemléltette a Cash flow/Adósság mutató értékének időbeli alakulását a csőd előtti ötéves periódusban.

10. ábra: A Cash flow/Adósság mutató időbeli alakulása



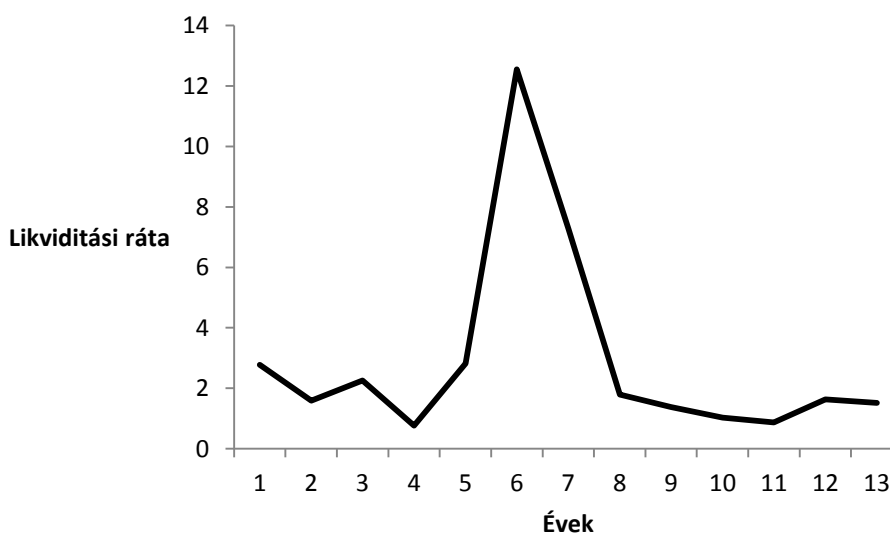
Forrás: Beaver [1966] alapján saját szerkesztés

Az ábrán folytonos vonallal egy működő, szaggatott vonallal pedig egy csődbe ment vállalat Cash flow/Adósság mutatójának időbeli alakulását szemléltette az idézett szerző. Az ábrát tekintve megállapítható, hogy a működő cég esetében a ráta értéke időben stabil és relatíve magas, szemben a másik vállalattal, amely esetében a mutató folyamatos csökkenést mutat a csődhöz közeledve. Az értekezésben javasolt dinamikus mutatók ezt az időbeli tendenciát képesek számszerűsíteni minden egyes pénzügyi mutató esetén.

3.1.2. A kiugró értékek kezelése

A dinamikus mutatók számításához olyan adatbázisra van szükség, amely valamennyi megfigyelési egységére vonatkozóan tartalmazza a vizsgált pénzügyi mutatók értékeit több évre visszamenőleg annak érdekében, hogy azok időbeli trendje értelmezhető legyen. Fontos kiemelni azonban, hogy a hazai csődös és működő vállalkozások körében feltehetően kevésbé tendenciózus a pénzügyi mutatók időbeli alakulása, mint ahogy azt a 10. ábra szemlélteti. Hosszabb idősorokat vizsgálva nagy valószínűséggel találhatunk olyan éveket, amelyek esetén a pénzügyi mutatók értéke „megtöri” a többi év vonatkozásában kirajzolódó trendet. Úgy is fogalmazhatunk, hogy az idősoron belül kiugró értékek is találhatók. Egy ilyen példát szemléltet a 11. ábra, amely egy, a kutatási adatbázisban szereplő vállalkozás likviditási rátájának időbeli alakulását mutatja.

11. ábra: Egy kiugró értéket tartalmazó mutatószám-idősor



Az ábrán látható, hogy a vizsgált vállalkozás likviditási rátája a vizsgált időszak tekintetében relatíve stabil. Ez alól kivételt képez a tárgyévet megelőző hatodik év, amikor a vizsgált mutató szélsőségesen magas (kiugró) értéket vett fel, melynek figyelembevétele jelentősen torzítaná a mutatószám-idősor elemzése alapján levonható következtetéseket.

Ennek elkerülése érdekében mutatószám-idősorok fentihez hasonló tagjait célszerűnek tartom helyettesíteni az idősor hozzá legközelebb eső, de már nem kiugró értékével. Természetesen ez a helyettesítés nem vonatkozik a csőd előtti utolsó évre, hiszen a csődös vállalatok gyakori ismertető jele a kiugróan alacsony, illetve magas érték bizonyos mutatók esetén a csőd, illetve felszámolási eljárás bekövetkezését megelőző évben. Ezek korrekciója ugyan statisztikai szempontból kezelhetőbbé tenné az adathalmazt, a csődös vállalatok felismerését azonban megnehezítené.

A helyettesítés megvalósítása érdekében azonban azt is definiálni szükséges, hogy mely értékek tekinthetők kiugrónak. Erre vonatkozóan nincs egyértelmű állásfoglalás a szakirodalomban, így statisztikai hüvelykujjszabályok alkalmazását tartom célszerűnek. Gyakran alkalmazott megközelítés, hogy akkor tekintenek egy megfigyelést kiugrónak, ha annak standardizált értéke a ± 3 szórás terjedelmen kívülre esik. Ezt a „szabályt” azonban véleményem szerint olyankor célszerű alkalmazni, amikor a megfigyelések száma relatíve magas (több száz, illetve ezer). Az általam vizsgálni kívánt mutatószám-idősorok azonban ennél jóval rövidebbek, így célszerűbbnek tartom egy „szigorúbb outlier definíció” alkalmazását a kiugró értékek hatékonyabb szűrése érdekében. Kísérleti jelleggel megvizsgáltam azt az esetet, amikor a mutatószám-idősorok értékei esetén minden olyan megfigyelés

kiugró értéknek minősül, amelynek az idősoron belül standardizált értéke a két szórás terjedelmen kívülre esik.

3.2. Az értekezésben vizsgált kutatási hipotézisek

A csődelőrejelzés témakörében a nyitott kutatási kérdések köre rendkívül széles, melyek közül számos példát mutattam be a szakirodalmi összefoglalásban. Empirikus tudományról lévén szó, egyértelmű válaszokat egyik kérdés tekintetében sem kaphatunk. Minél több elemzés jut azonban hasonló következtetésre, annál nagyobb megbízhatósággal tekinthetünk megválaszoltnak egy-egy kutatási kérdést. A szakirodalom bemutatása során terjedelmi okokból a legrelevánsabb témakörök bemutatására kellett szorítkoznom. E vizsgálati szempontoknak pedig csak egy még szűkebb részhalmazát volt lehetőségem empirikusan is elemezni.

A bemutatott szakirodalmi háttér ismeretében értekezésemben az alábbi hipotéziseket vizsgáltam.

1. A dinamikus pénzügyi mutatószámok között találhatók olyanok, amelyek a rendelkezésre álló mintában statisztikailag szignifikáns diszkrimináló erővel bírnak a csődös és a működő vállalkozások között.
2. A legutolsó megfigyelt év statikus pénzügyi mutatói, valamint az azokból képzett dinamikus ráták együttes alkalmazásával a rendelkezésre álló minta adatai alapján magasabb előrejelző képességű modell állítható fel, mint abban az esetben, amikor csak a statikus pénzügyi mutatókat szerepeltetjük a CHAID módszerrel képzett döntési fák input változói között.

3. A rendelkezésre álló minta adatai alapján a CHAID módszerrel felállított döntési fákra épülő csödelőrejelző modellek előrejelző képessége növelhető, ha a dinamikus mutatók számítása során az outlier értékeket helyettesítjük a hozzájuk legközelebb eső, de már nem kiugró értékekkel.

A harmadik hipotézis vizsgálata során az egyes mutatószám-idősorok *valamennyi* értékét a $[t-2; t-n]$ időintervallum átlaga és szórása alapján standardizáltam. Az így kapott standardizált adatsorokban azokat tekintettem kiugró értéknek, amelyek standardizált értéke a $(-2, 2)$ nyílt intervallumon kívülre esett.

A bemutatott hipotézisek vizsgálatát a döntési fát felállító CHAID eljárás alkalmazásával végeztem el. Ennek oka, hogy az empirikus vizsgálatokban gyakran alkalmazott SPSS szoftverben ez áll beépített funkcióként rendelkezésre. Másrészt a módszerválasztást az a korábban bemutatott szakirodalmi érv is befolyásolta, amely szerint a gyakorlati csödelőrejelzésben az olyan módszerek preferáltak, amelyek könnyen értelmezhető modelleket eredményeznek (Martens et al. [2010]). Az idézett szerzők erre a célra a döntési fákat felállító módszerek alkalmazását javasolják, mivel azok jellemzően nem támasztanak statisztikai előfeltevéseket a vizsgált adathalmazzal szemben, továbbá meglehetősen magas klasszifikációs teljesítménnyel is jellemezhetők, mindamellet könnyen értelmezhető „ha-akkor” típusú szabályok segítségével valósítják meg a klasszifikációt. E szempontok miatt esett a választásom a CHAID módszerre, amely szintén egy döntési fát eredményező eljárás.

4. AZ EMPIRIKUS VIZSGÁLATHOZ FELHASZNÁLT ADATOK

A bevezető fejezetben kiemelttem, hogy a csődelőrejelzés tudományos vizsgálatát Magyarországon az új évezred első évtizedének végéig a vállalati adatokhoz való szabad hozzáférés hátráltatta. Ez az akadály azonban 2009-től kezdve elhárult, ugyanis ettől az évtől kezdődően a hazánkban bejegyzett társas vállalkozások kötelesek pénzügyi beszámolóikat közzétenni az erre a célra jogszabályi szinten meghatározott honlapon, ahol a vállalatok neve és cégjegyzék száma alapján bárki keresheti a cégek legfontosabb számviteli dokumentumait.

E lehetőséget megragadva azt a célt tűztem magam elé, hogy az értekezésemben szereplő kutatási hipotéziseimet a lehető legfrissebb adatokon, továbbá a lehető legnagyobb elemszámú adatállomány alapján vizsgáljam meg. Az adatgyűjtés módját, szempontjait és az azok eredményeképp létrejött adatbázist mutatom be ebben a fejezetben.

4.1. Általános mintavételi kérdések

A felsorolt kutatási kérdések vizsgálatához saját adatgyűjtésből származó mintát állítottam össze. A mintavétel során az első kérdés az adatbázis mérete volt. A megfigyelések számát 1000 elemben maximalizáltam, ami kissé alacsonynak tűnet a Magyarországon működő vállalatok teljes számához viszonyítva. A relatíve alacsony esetszám választása azzal indokolható, hogy az adatok összegyűjtésére manuálisan került sor, mely a nyilvános adatforrások alkalmazása esetén meglehetősen időigényes feladat. Továbbá hangsúlyozni szükséges, hogy a vállalatok adatait nem

csak a legfrissebb üzleti év vonatkozásában, hanem 2001-ig visszamenőleg valamennyi év esetén összegyűjtöttem.

A minta méretét követően a második legfontosabb kérdés a minta összetétele a vizsgált csoportok (csődös, működő) tekintetében. Az adathalmazban azonos arányban találhatók működő és fizetéseképtelen megfigyelések. Az 50-50 %-os arányú mintafelosztás indoka, hogy a rendelkezésre álló esetszám vonatkozásában a reprezentativitás követelményének érvényesítése azt eredményezné, hogy a csődös vállalatok száma rendkívül alacsony lenne. Ez pedig felvetheti annak problémáját, hogy a modell számára nem áll elegendő mennyiségű információ rendelkezésre a csődös vállalatok felismerésére. Mivel a csődmodellezés elsődleges célját épp a csődös vállalatok minél hatékonyabb felismerése képezi (Du Jardin [2010]), a minta felosztása során az 50-50 %-os arányra esett a választásom az imént ismertetett probléma elkerülése érdekében.

A minta összeállítása során a harmadik kérdés a kutatási hipotézisekben szereplő fogalmak operacionalizálása. A csődelőrejelzésben azt szükséges definiálni, hogy milyen kritérium alapján tekintünk egy vállalatot csődösnek, illetve működőnek. Ennek azért van kiemelt jelentősége, mert – ahogy a fogalmi kereteket behatároló 1. fejezetben kifejtettem – a csőd fogalmát nem szó szerinti értelmezésben használják ezen a tudományterületen.

A minta összeállítása során azokat a vállalatokat tekintettem fizetéseképtelennek (csődösnek), amelyek az adatgyűjtés időpontjában felszámolási, vagy csődeljárás alatt álltak; és azokat működőnek, amelyek nem álltak csődeljárás, felszámolási eljárás, vagy végelszámolási eljárás alatt. Ezek tényét a Közigazgatási és Igazságügyi

Minisztérium által a www.e-cegjegyzek.hu oldalon nyilvánosan hozzáférhető hatályos cégjegyzék adatok alapján állapítottam meg.

4.2. Az adatgyűjtés folyamata

Az adatgyűjtés kiinduló információforrása a Céglözlöny volt, melynek számaiból véletlenszerűen választottam. Az azokban hirdetményt megjelentető vállalkozások köréből az alábbi szempontok figyelembevételével állítottam össze a mintát.

1. A mintavétel során alapvető szempont volt, hogy legalább három évre visszamenőleg rendelkezésre álljanak a vizsgált vállalkozások beszámolói (mérleg, eredménykimutatás). Ennek oka, hogy a kutatás fő célkitűzése annak vizsgálata, hogy bírnak-e diszkrimináló erővel azon változók, amelyek egy adott pénzügyi mutató értékét saját korábbi értékeinek tükrében mutatják. Ehhez azonban szükséges, hogy legalább három éves időszakra rendelkezésre álljanak a vállalkozások pénzügyi mutatói.
2. Nem kerülhettek mintába azon vállalkozások, amelyeknek volt olyan pénzügyi mutatószáma, amely nem mutatott időbeli szóródást, mivel ez lehetetlenné tette volna a dinamikus mutatók kiszámítását.
3. Szintén kimaradtak a mintából az olyan megfigyelések, amelyek legalább két egymást követő évben²¹ nem realizáltak árbevételt. Ennek oka, hogy az ilyen vállalkozások vélhetően nem folytatnak érdemi gazdálkodást, így mintába kerülésüknek torzító hatása lenne a modellek eredményeire.

²¹ Azok a vállalkozások, amelyek esetén csak egyetlen évben fordult elő, hogy nem realizáltak árbevételt, azért kerülhettek be a mintába, mert a csőd, illetve felszámolási eljárás alá került vállalkozásoknál gyakran tapasztaltam, hogy a felszámolási eljárás megkezdését megelőző évben nem realizáltak árbevételt. Ebből adódóan az ilyen esetek kizárása a mintából olyan válságjelek kiszűrését is jelentené, amely segítheti a csőd előrejelzését.

A mintába került vállalkozások pénzügyi adatait az általuk a számviteli törvény előírásai szerint közzétett beszámolók alapján számítottam ki, melyeket a korábban említett Minisztérium a <http://e-beszamolo.kim.gov.hu/Default.aspx> oldalon nyilvánosan hozzáférhetővé tesz.

Az adatgyűjtés során problémát jelentett, hogy a bemutatott nyilvános adatbázisok használata során nincs lehetőség arra, hogy a keresési feltételek között szerepeltessem a vállalatok tevékenységi körét, illetve méretét. Ebből adódóan az adatgyűjtés során előzetesen nem volt lehetőség e szempontok figyelembevételére; azonban ez nem tekinthető munkám egyedi sajátosságának, ugyanis a nemzetközi szakirodalomban a kis-, és középvállalkozások vonatkozásában csődmodellt felállító kutatók jellemzően szintén nem szűkítik le elemzésüket egy-egy iparágra.

Amint korábban bemutatam, az iparági sajátosságok sok esetben jelentős hatást gyakorolhatnak a mutatószámok értékeire, ami mérsékelheti azon modellek előrejelző képességét, amelyek modellépítési adatbázisában egyidejűleg szerepelnek megfigyelések különböző ágazatokból. Mivel az ágazati hovatartozás figyelembevételére az adatgyűjtés során előzetesen nekem sem volt lehetőségem, így az empirikus vizsgálatok során felállított modellek nem javasoltak gyakorlati előrejelzési célokra. Ez azonban véleményem szerint nem feltétlenül jelent problémát a kitűzött kutatási célok szempontjából, ugyanis munkámban a dinamikus pénzügyi mutatók alkalmazási lehetőségeit kívánom demonstrálni; tehát nem célom gyakorlati előrejelzésre is alkalmas csődmodell felállítása. A kutatás célja szempontjából egy heterogén minta még akár kedvező is lehet, ugyanis abban az esetben, ha egy ilyen adathalmazon a dinamikus mutatók alkalmazhatósága igazolható, jó eséllyel várható, hogy egy homogénebb minta esetén is javít a modellek előrejelző képességén a

dinamikus változók alkalmazása. Ennek vizsgálata egy jövőbeli kutatás tárgya lehet abban az esetben, ha az értekezés bemutatásra kerülő eredményei erre alapot adnak.

Mivel a felhasznált nyilvános adatbázisokban nincs lehetőség keresésre iparág és méret alapján, így a mintát nem lehetett előzetesen szűkíteni e változók tekintetében, azonban az adatok birtokában ezek a változók is szerepeltethetők a modellekben magyarázó tényezőként, mivel az adatgyűjtést követően a vállalatok mérete és a TEÁOR kód alapján definiált tevékenységi köre ismertté vált. Ebből az is következik, hogy utólagosan lenne lehetőség a minta szűkítésére iparág, illetve méret alapján, ez azonban oly mértékben csökkentené a rendelkezésre álló megfigyelések számát az iparág és méret alapján képzett csoportokban, hogy az nem lenne elégséges statisztikai modellek építésre és azok eredményei alapján érdemi következtetések levonására. Emiatt a kutatás során a teljes rendelkezésre álló adathalmazzal, mint egységes adatbázissal dolgoztam.

Hasonló a helyzet a vállalatok kora esetén. Ebben az esetben a felhasznált nyilvános adatbázisok ismét korlátot emeltek az elemzés elé, ugyanis azokból a megfigyelések kora nem határozható meg egyértelműen. Helyette azt vettem figyelembe, hogy a vizsgált vállalkozás hány évre visszamenőleg tette közzé beszámolóit az arra a célra törvényileg meghatározott weboldalon. Véleményem szerint ez a változó jó közelítést adhatja a vizsgált vállalkozás korának.

A csődmodellek magyarázó változóit – a szakirodalom évtizedes hagyományaival összhangban – a vállalatok mérlegeiben és eredménykimutatásaiban közzétett adatokból képzett hányados típusú mutatószámok köréből választottam ki. Ezek forrását részben a hazai empirikus kutatásokban gyakran alkalmazott mutatószámok (Virág et al. [2013]), részben saját megfontolásaim képezték. A vizsgálatban

alkalmazott pénzügyi mutatószámok nevét és számításmódját tartalmazza a 2. táblázat. A mutatók kalkulációjához szükséges adatokat a mintába került vállalatok mérlegeiben, illetve eredménykimutatásaiban közölt év végi záró értékek alapján vettem figyelembe.

2. Táblázat: Az empirikus vizsgálatban felhasznált mutatószámok neve és számításmódja

Mutatószám neve	Számítás módja
Likviditási ráta	Forgóeszközök/Rövid lejáratú kötelezettségek
Likviditási gyorsráta	(Forgóeszközök – Készletek)/Rövid lejáratú kötelezettségek
Pénzeszközök aránya	Pénzeszközök/Forgóeszközök
Cash flow/Kötelezettségek	(Adózás utáni eredmény + Értécsökkenési leírás)/Kötelezettségek
Cash flow/Rövid lejáratú kötelezettségek	(Adózás utáni eredmény + Értécsökkenési leírás)/Rövid lejáratú kötelezettségek
Tőkeellátottság	(Befektetett eszközök + Készletek)/Saját tőke
Eszközök forgási sebessége	Értékesítés nettó árbevétele/Mérlegfőösszeg
Készletek forgási sebessége	Értékesítés nettó árbevétele/Készletek
Követelések forgási ideje	Követelések/Értékesítés nettó árbevétele
Eladósodottság	Kötelezettségek/Mérlegfőösszeg
Saját tőke aránya	Saját tőke/Mérlegfőösszeg
Bonitás	Kötelezettségek/Saját tőke
Árbevétel-arányos nyereség	Adózás utáni eredmény/Értékesítés nettó árbevétele
Eszközarányos nyereség	Adózás utáni eredmény/Mérlegfőösszeg
Követelések/Rövid lejáratú kötelezettségek	Követelések/Rövid lejáratú kötelezettségek
Nettó forgótőke aránya	(Forgóeszközök – Rövid lejáratú kötelezettségek)/Mérlegfőösszeg
Vállalat mérete	Az eszközállomány természetes alapú logaritmus
Évek	A megfigyelt évek száma

A csődmodellekben gyakran használt mutatószámok közé tartozik a saját tőke arányos nyereség, mely sok esetben veti a fel a kettős negatív osztás²² problémáját (Kristóf [2008]), melynek kezelésére nincs egyértelműen preferált megoldás a szakirodalomban, ezért e mutatót nem vettem figyelembe a számítások során.

A hányados típusú mutatószámok másik jellemző problémája akkor merül fel, amikor a hányados nevezőjében nulla érték adódik. E problémát a gyakorlatban

²² Amikor egy hányados formában számítható pénzügyi mutató számlálója és nevezője is negatív.

gyakran kezelik úgy, hogy az ilyen adatokat hiányzó értéknek tekintik, melyeket a többi megfigyelés valamilyen középértékével, vagy azok valamelyik szélső percentilisével helyettesítik. Véleményem szerint azonban ezek a megközelítések nem feltétlenül visznek konzisztens értéket a csődelőrejelző modellekbe. Ezt a következő példa segítségével kívánom alátámasztani.

Tekintsünk egy olyan vállalkozást, amely rövid lejáratú kötelezettségeit mindig azonnal, vagy jellemzően minden évben a mérleg fordulónapját közvetlenül megelőzően teljesíti, ebből adódóan a mérleg fordulónapján nem rendelkezik rövid lejáratú kötelezettséggel, ami lehetetlenné teszi a likviditási ráta²³ kalkulációját. Tételezzük fel, hogy a példában szereplő vállalkozás jelentős forgóeszköz állománnyal is rendelkezik, ami lehetővé teszi számára, hogy egy később felmerülő esetleges „likviditási sokkot” képes legyen finanszírozni. Ha egy ilyen vállalkozás likviditási rátáját a mintában szereplő többi megfigyelés átlagával helyettesítenénk, akkor a megfigyelést a modell átlagos likviditású vállalkozásként tekintené, ami az adatai alapján nem helytálló. A másik lehetőség: valamely szélső percentilissel történő helyettesítés, amely már konzisztensebb információt visz a modellbe, de ekkor azonban mintaspecifikus, hogy egy konkrét mutatószám mivel kerül helyettesítésre.

A fentiekből kiindulva az adatok előkészítése során azt a megoldást alkalmaztam, hogy azokban az esetekben, ahol a nevező értéke nulla lenne, a nulla értéket eggyel²⁴ helyettesítettem. Így a példában szereplő vállalat likviditási rátája egy meglehetősen nagy értéket vesz fel, jelezve, hogy a vállalat likviditása rendkívül magas. Az ilyen megfigyeléseket a csődelőrejelzés gyakorlata általában mellőzni igyekszik, mivel a

²³ Likviditási ráta=Forgóeszközök/Rövid lejáratú kötelezettségek

²⁴ Mivel a beszámolók adatai ezer Ft-ban adottak, így az eggyel történő helyettesítés 1000 Ft-ot jelent.

szélsőségesen magas, illetve alacsony értékek – mint outlier megfigyelések – jellemzően torzítják a statisztikai modellek eredményeit. Egyetértve azonban McLeay és Omar [2000] azon megállapításával, mely szerint az outlier értékek és a normális eloszlástól való eltérés inkább alapvető sajátosságai a csődmodellek adatbázisának, mintsem ritka kivételek, olyan eljárásokat érdemes alkalmazni, amelyek segítségével lehetőség van az outlier értékek modellben tartására. Azok elhagyásával ugyanis a megfigyelések száma drasztikusan csökkenne, ami a modell számára felhasználható információ mennyiségét redukálná hasonló mértékben. Ez pedig azzal a következménnyel járhat, hogy a modell későbbi alkalmazása során kevésbé lenne alkalmas a szélsőséges értékkel bíró megfigyelések minősítésére.

Az outlier megfigyelések modellben tartásának szándéka is jelentősen befolyásolta a hipotézisek vizsgálata céljából használni kívánt módszer kiválasztását. A döntési fákat generáló technikák alkalmazásának előnye, hogy klasszifikációs teljesítményüket a kiugró értékek jelenléte nem befolyásolja (Twala [2010]).

A bemutatott mintavétel eredményeképp rendelkezésre álló adatbázisban tehát 1000 hazai vállalkozás pénzügyi mutatói szerepelnek összesen 7592 üzleti évre vonatkozóan a 2001-2012 közti időszakból. A minta egyenlő arányban tartalmaz fizetőképes és fizetéseképtelen vállalkozásokat, melyek pénzügyi helyzetét a 2. táblázatban látható 17 pénzügyi mutatószám segítségével, valamint azon évek számával mértem, amelyekre vonatkozóan az egyes megfigyelések pénzügyi adatai rendelkezésre álltak az adatok forrását képező nyilvános adatbázisban. A csődös vállalatok esetén a csőd, illetve felszámolási eljárás kezdetét megelőző év, a működő

vállalatok esetén pedig a legutolsó megfigyelt év²⁵ adatai alapján számított pénzügyi mutatószámokat nevezem statikus pénzügyi mutatóknak az értekezésben.

A 17 vizsgált pénzügyi mutató mindegyike esetén kiszámítottam a 3. fejezetben definiált dinamikus pénzügyi mutatókat, melyek a legutoljára megfigyelt év pénzügyi helyzetét hivatottak megítélni az azt megelőző időszak tükrében. Így az empirikus vizsgálatok számára 35 független változó áll rendelkezésre. Ezek alapstatisztikai mutatóit a Függelék F.1. táblázata tartalmazza.

²⁵ A mintavétel időpontjához képest legaktuálisabb (legfrissebb) év.

5. EMPIRIKUS VIZSGÁLAT

A feldolgozott és ismertetett szakirodalom alapján körvonalazható számos lehetséges kutatási irány közül a csődelőrejelző modellek dinamizálását választottam értekezésem tárgyául. Ezt a célt a 3. fejezetben bemutatott dinamikus pénzügyi mutatók használatával kívánom megvalósítani. Az ott bemutatott három kutatási hipotézis vizsgálatát a 4. fejezetben ismertetett minta adatai alapján végeztem el. Empirikus vizsgálataim eredményeit foglalom össze ebben a fejezetben a vizsgált hipotézisek szerinti bontásban.

5.1. Az első kutatási hipotézis vizsgálata

A 3. fejezetben bemutatott hipotézisek közül az első az alábbi:

„A dinamikus pénzügyi mutatószámok között találhatók olyanok, amelyek a rendelkezésre álló mintában statisztikailag szignifikáns diszkrimináló erővel bírnak a csődös és a működő vállalkozások között.”

E hipotézis vizsgálatát egy egyszerű statisztikai eszközzel, a két független mintából származó átlagok összehasonlítására szolgáló T-próba segítségével végeztem el. A módszer használatával kapcsolatban érdemes felhívni a figyelmet annak alkalmazási feltételeire:

- az összehasonlítani kívánt sokaságok egymástól függetlenek;
- az összehasonlítani kívánt sokaságokban az egyes pénzügyi mutatók normális eloszlást követnek;

- az összehasonlítani kívánt sokaságokban az egyes pénzügyi mutatók szórása közelítőleg azonos.

A három feltevés közül az első a mintavételi sajátosságokból adódóan teljesül ugyan, a másik kettő azonban bizonyítottan nem érvényes a pénzügyi mutatók esetén. Ennek ellenére a csődelőrejelzésben általános elfogadott e próba alkalmazása. Ezt erősíti Liang et al. [2015] kutatása is, akik a független mintás T-próba használatát javasolják az adatbányászati módszerek magyarázó változóinak szelekciójára. Empirikus vizsgálatukban az idézett szerzők nem tértek ki a teszt alkalmazási feltételeinek ellenőrzésére, amely a feldolgozott szakirodalom alapján általános gyakorlatnak mondható a csődelőrejelzés főáramában. Emiatt magam sem vizsgáltam az alkalmazási feltételek fennállását.

A vizsgált pénzügyi mutatók, valamint az azok idősoraiból képzett dinamikus változók alapstatisztikáit tartalmazza a függelék. F.1. Táblázata, ahol a „D_” előtag jelzi az egyes mutatószámok idősoraiból képzett dinamikus változók értékeit. A táblázatban *-gal jelöltem azokat a mutatókat, amelyek esetén a statisztikai elemzésekben általánosan elfogadott 5 %-os szignifikancia-szinten elvethető, hogy a csődös és a működő vállalatok megfelelő mutatóinak átlaga azonosnak tekinthető.

Az eredmények szerint a vizsgált mutatószámok közül 13 bizonyult szignifikánsnak a csődös és a működő vállalatok között. Ha a megfigyelt évek számát figyelmen kívül hagyjuk, akkor azt láthatjuk, hogy a statikus pénzügyi mutatók közül hét, míg a javasolt dinamikus mutatószámok köréből öt volt szignifikáns, azaz jelentős mértékben növekedett azon változók száma, amelyek ezen egyszerű statisztikai próba alapján jó előrejelző változónak minősülhetnek a fizetőképes és a fizetéseképtelen vállalkozások megkülönböztetésében.

A szignifikáns változók körén belül 3 csoportot lehet megkülönböztetni annak alapján, hogy minden pénzügyi mutatóhoz tartozik egy olyan változó, amely az adott mutató legutolsó megfigyelt értékét viszonyítja az azt megelőző időszak értékeihez, így vannak olyan pénzügyi mutatók

- amelyeknek csak a legutoljára megfigyelt év statikus értéke szignifikáns (Pénzeszközök aránya, Nettó forgótőke aránya);
- amelyek esetén a statikus érték és a javasolt dinamikus változó is szignifikáns (Követelések forgási ideje, Eladósodottság, Saját tőke aránya, Eszközarányos jövedelmezőség);
- amelyek esetén csak a dinamikus változó volt szignifikáns a két csoport megkülönböztetése szempontjából (CF/Adósság, Készletek forgási sebessége)

Az eredmények tehát arra utalnak, hogy bővült azon változók köre, amelyek felhasználhatók a hazai fizetőképes és fizetéseképtelen vállalkozások diszkriminálására. Az eredmények arra is rámutatnak, hogy a dinamikus változók nem helyettesítő, hanem sokkal inkább kiegészítő szerepet tölthetnek be a vállalkozások minősítésében, hiszen voltak olyan mutatók, amelyek esetén csak azok statikus értékei voltak szignifikánsak, viszont adódtak olyan változók is, amelyek statikus értékei nem, csak az azokból képzett dinamikus mutatók különböztek szignifikánsan.

Érdekes eredmény, hogy a csődmodellekben gyakran alkalmazott likviditási ráták, illetve a vállalat mérete nem minősült szignifikánsnak. Szembetűnő, hogy a rendelkezésre álló változók között relatíve alacsony a szignifikánsan különböző mutatók aránya; továbbá, hogy az egyes pénzügyi mutatók alapstatisztikai mutatói

közt találhatók olyanok, amelyek irreálisan magasnak tűnhetnek az Olvasó számára. Ennek oka egyrészt a minta iparág és méret vonatkozásában tapasztalható heterogenitása; másrészt az, hogy az adatgyűjtés során azokban az esetekben, ahol a nevező értéke 0 lenne, ott az adott pénzügyi adat eggyel került helyettesítésre, ami szélsőségesen magas, vagy éppen alacsony mutatószám értékeket generált. Különösen a forgási mutatóknál láthatunk extrémnek tűnő értékeket. Ezekben az esetekben ugyanis gyakran előfordult – főleg a csődös, illetve a gyenge pénzügyi helyzetben lévő működő vállalatoknál – hogy a csőd előtti évben nem, vagy csak nagyon alacsony összegű árbevételt realizáltak a megfigyelések. A helyettesítésből adódóan szélsőségesen magas, illetve alacsony értékű mutatók torzítónak tűnhetnek ugyan, azonban véleményem szerint ez a megközelítés hozzájárul a mutatókban rejlő információtartalom még magasabb szintű kihasználásához és az azokra épülő statisztikai modellek előrejelző képességének növeléséhez.

Ha a szignifikáns mutatószámok átlagait alaposabban szemügyre vesszük a Függelékben található F.1. táblázatban, akkor a mintába került hazai vállalkozások pénzügyi helyzetére vonatkozóan vonhatunk le következtetéseket. Általános megállapítás, hogy a szignifikáns mutatószámoknál a várakozásnak megfelelően alakul a változók átlagainak nagyságrendje, azaz például a pénzeszközök aránya átlagosan magasabb a működő vállalkozásoknál, mint a fizetésektelen megfigyeléseknél, de hasonló tendenciát tapasztalhatunk a többi mutató esetén is.

A csődelőrejelzés „atyjának” tekintett Beaver [1966] a leggyakrabban alkalmazott pénzügyi mutatószámokat hasonlította össze azok előrejelző képessége szempontjából. Eredményei azt mutatták, hogy a CF/Adósság mutatószám a „legjobb” ebben a tekintetben. E változó statikus értéke nem, de az azokból képzett dinamikus változó szignifikáns diszkrimináló erővel bír a hazai vállalkozások

körében is a fizetőképess és a fizetéseképtelen megfigyelések között. A táblázatban látható, hogy amíg a működő vállalkozások esetén e mutató értéke jellemzően a korábbi időszak minimuma és maximuma között található, addig a csődös vállalatoknál a legutolsó megfigyelt érték jóval az azt megelőző időszak minimuma alatt alakult, azaz a csődös vállalatok azonosításában nagy segítséget nyújt, ha megfigyeljük, hogy milyen mértékű a változás e mutató értékében az azt megelőző időszakhoz viszonyítva.

A forgási mutatók közül a követelések és a készletek forgási mutatóiról is hasonló megállapításokat tehetünk. A mintába került fizetéseképtelen vállalkozásokat az eredmények alapján rendkívül magas érték jellemzi a követelések forgási mutatóit tekintve. Kimondottan magasnak mondható a növekedés mértéke a csőd előtti utolsó évben az azt megelőző időszakhoz képest e változó esetén is. Ugyanilyen tendencia olvasható ki az adósság mutatókból is.

Érdekes szemügyre venni az eszközarányos-jövedelmezőség, valamint a nettó forgótőke mutatókat is, ugyanis e változók átlaga negatív lett a fizetéseképtelen és a fizetőképess vállalatok körében egyaránt. Ez vélhetően a mintavételi periódussal magyarázható, ugyanis a mintába került vállalkozások pénzügyi mutatószám-idősorainak utolsó tagjai 2009-2012 közti időszakból származnak, amely a hazánkat is sújtó világgazdasági recesszió időszaka. Ezt figyelembe véve már kevésbé mond ellent az előzetes várakozásoknak az, hogy a működő vállalkozásokat is átlagosan negatív jövedelmezőség jellemzi. A dinamikus ráták diszkrimináló ereje itt abban nyilvánul meg, hogy a fizetéseképtelen megfigyelések esetén e mutató még inkább negatív, illetve a „D_” előtaggal jelölt mutatók alapján az is megállapítható, hogy a csődöt megelőző időszakhoz képest drasztikus csökkenés jellemzi e mutatókat a fizetéseképtelen vállalatok esetén.

Végezetül említést érdemel a megfigyelt évek száma, amely ugyan nem pénzügyi mutató, mégis szignifikánsan különbözik a két csoportban. Az eredmények alapján megállapítható, hogy a mintába került fizetésektelen vállalatok jellemzően rövidebb „cégtörténettel” bírnak, azaz a fiatalabb vállalkozások nagyobb eséllyel mennek csődbe, mint az idősebb megfigyelések. E megállapítás összhangban van a tudományterület nemzetközi szakirodalmából kirajzolódó tendenciával is, de természetesen fel kell hívni a figyelmet, hogy e változó esetén a többi mutatóhoz képest nagyobb az esélye annak, hogy a mintavételi sajátosságok hatást gyakoroltak a változó átlagos értékeire a két csoportban.

A bemutatott eredmények alapján azt a következtetést lehet levonni, hogy az értekezésben javasolt dinamikus mutatószámok közt találhatók olyanok, amelyek szignifikánsan különböznek a hazai vállalkozások esetén a csődös és a működő vállalatok esetén, ezért az értekezés első hipotézisét elfogadtam.

5.2. A második kutatási hipotézis vizsgálata

Az értekezés második hipotézisét az alábbiak szerint fogalmaztam meg a 3. fejezetben:

„A legutolsó megfigyelt év statikus pénzügyi mutatói, valamint az azokból képzett dinamikus ráták együttes alkalmazásával a rendelkezésre álló minta adatai alapján magasabb előrejelző képességű modell állítható fel, mint abban az esetben, amikor csak a statikus pénzügyi mutatókat szerepeltetjük a CHAID módszerrel képzett döntési fák input változói között.”

Az eredmények ismertetése során fontosnak tartom hangsúlyozni, hogy értekezésem elsődleges célja a dinamikus pénzügyi mutatószámok bemutatása és alkalmazási lehetőségeinek empirikus bizonyítása. Tehát munkámban nem tűztem célul olyan modell felállítását, amely alkalmas lehet gyakorlati vállalatminősítési célokra. Emiatt a modellépítés során nem tértem ki olyan, a gyakorlati hitelkockázati modellezésben fontos vizsgálati szempontokra, mint az egyes megfigyelések vonatkozásában konkrét csődvalószínűségek becslése, illetve azok kalibrációja. Ilyen vizsgálatokra a rendelkezésre álló minta heterogenitása sem teremt lehetőséget. E heterogenitás azonban véleményem szerint eszközként szolgálhat a dinamikus mutatók létjogosultságának bizonyítására a csődelőrejelzés területén.

Az empirikus vizsgálatokat az SPSS Statistics 20 programmal végeztem. A CHAID eljárás alkalmazása során a felhasznált szoftver alapbeállításaitól csak annyiban tértem el, hogy a rendelkezésre álló adatbázis méretére tekintettel a szülő ágak létesítéséhez szükséges minimális elemszámot tíz, a gyerekágak létesítésének minimális elemszámát pedig öt darabban határoztam meg.

Annak érdekében, hogy elkerüljem a minta tanuló és tesztelő részre történő felosztásából eredő véletlenszerűségek torzító hatását, az elemzésben a tízszeres keresztvalidáció módszerét alkalmaztam. Ez azt jelenti, hogy a rendelkezésre álló 1000 elemű mintát véletlenszerűen felosztottam tíz egyenlő részre. A modelleket a teljes minta 90 %-ának megfelelő adatkörön állítom fel, majd azok előrejelző-képességét a maradék 10 %-nyi megfigyelés – mint tesztelő minta – felhasználásával ítélem meg oly módon, hogy a minta mindegyik tizede szerepeljen egyszer a tesztelő minta szerepében. Az eljárás eredményeképp tehát 10 döntési fa került felállításra,

melynek találati arányait²⁶ feljegyeztem a tanuló és tesztelő mintákon egyaránt, majd ezeket átlagoltam annak érdekében, hogy a modellezésből levont következtetések minél kevésbé legyenek tulajdoníthatók annak, hogy hogyan történt a minta felosztása tanuló és tesztelő mintákra. A tíz mintafelosztás átlagában kapott modellteljesítményt mutatja a 3. táblázat.

3. Táblázat: A CHAID módszerrel felállított döntési fák találati aránya tízszeres keresztvalidáció alkalmazásával

Minta	Csoport	Input változók		
		Pénzügyi mutatók	Dinamikus mutatószámok	A két változókör együtt
Tanuló	Fizetésképtelen	85,7%	83,2%	84,6%
	Fizetőképes	80,1%	77,8%	83,8%
	Összesen	82,9%	80,5%	84,2%
Tesztelő	Fizetésképtelen	76,6%	75,0%	76,8%
	Fizetőképes	71,0%	71,6%	74,6%
	Összesen	73,8%	73,3%	75,7%

A 10-szeresen keresztvalidált eredmények tükrében megállapíthatjuk, hogy a dinamikus mutatószámok önálló alkalmazásával nem javult a modellek előrejelző képessége. Sőt, a legtöbb esetben romlás volt tapasztalható a statikus pénzügyi mutatók önálló alkalmazásához képest. Ezzel szemben az esetek túlnyomó többségében növekedett a modellek találati aránya akkor, amikor a két változókört együtt alkalmaztam a modellekben független változóként. Ezt illusztrálандó a 3. táblázatban vastag betűvel emeltem ki azon besorolási pontosságokat, amelyek magasabbak a két változókör együttes alkalmazása esetén azok önálló felhasználáshoz képest. A javulás mértéke a tanuló minta esetén összességében 1,3,

²⁶ A modellek találati aránya alatt a helyesen besorolt megfigyeléseket értem az összes megfigyelés számához viszonyítva.

míg a tesztelő minta esetén 1,9 százalékpont. Az empirikus eredmények tehát megerősítik azt az előző szakaszban tett megállapítást, miszerint szinergikus hatás feltételezhető a statikus és a dinamikus mutatószámok között a csődelőrejelző modellekben.

A bemutatott empirikus eredmények alapján az értekezés második hipotézisét is elfogadtam.

5.3. A harmadik kutatási hipotézis vizsgálata

Az értekezésben vizsgált utolsó hipotézist korábban az alábbiak szerint fogalmaztam meg:

„A rendelkezésre álló minta adatai alapján a CHAID módszerrel felállított döntési fákra épülő csődelőrejelző modellek előrejelző képessége növelhető, ha a dinamikus mutatók számítása során az outlier értékeket helyettesítjük a hozzájuk legközelebb eső, de már nem kiugró értékekkel.”

E hipotézis kapcsán hangsúlyozni kell azt a 3. fejezetben már rögzített tényt, hogy a szakirodalomban nincs egységes definíció arra vonatkozóan, hogy mikor tekinthetünk egy értéket kiugróan magasnak vagy alacsonynak (outlier-nek). Ilyen helyzetben általában hüvelykujj-szabályokat alkalmaznak a statisztikai elemzésekben. Elterjedt gyakorlat az a megközelítés, mely szerint akkor tekinthető egy megfigyelés outlier-nek, ha annak standardizált értéke az öt, illetve három szórás terjedelmen kívül esik.

Ezek az értékek azonban csak akkor tekinthetők alkalmas „definíciónak” a kiugró értékek azonosítására, ha a megfigyelések száma kellően magas. Az általam végzett

vizsgálatban azonban a megfigyelések száma relatíve alacsony, ugyanis az egyes pénzügyi mutatók idősorának hossza három és tizenkét tag közt mozog. Ilyen alacsony tagszám mellett pedig az előbb említett határértékek túlságosan megengedők lehetnek a kiugró értékekkel szemben, amelyek jelentősen torzíthatják a dinamikus mutatók értékét. Emiatt véleményem szerint szigorúbb határt érdemes húzni a kiugró értékek azonosításához: választásom a két szórás szabály alkalmazására esett, melynek lényege, hogy azokat az értékeket tekintetem az egyes mutatószámok idősorán belül kiugrónak, melynek standardizált értéke a ± 2 szórás terjedelmen kívülre esik. A kiugró értékek dinamikus mutatókra gyakorolt torzító hatását, valamint a kiugró értékek korrekciójának módját szemléltetem a következő táblázatokban.

4. Táblázat: Egy kiugró értéket tartalmazó mutatószám idősor

Üzleti év	ROA	Standardizált ROA
2011	-4,9112	-2,3599
2010	-1,8360	-0,5296
2009	0,1153	0,6317
2008	0,1671	0,6626
2007	0,0403	0,5871
2006	-0,0924	0,5081
2005	-0,1939	0,4477
2004	0,0284	0,5800
2003	-0,4214	0,3123
2002	-5,5220	-2,7234
2001	-1,7470	-0,4766

A táblázat egy olyan vállalat ROA mutatójának időbeli alakulását mutatja, amely 2012-ben csődbe ment. Látható, hogy a csőd előtti évben lényegesen gyengébb volt a mutató értéke, mint az azt megelőző nyolc évben. Annak érdekében, hogy a csőd

előtti év szélsőségesen gyenge teljesítményének információtartalmát modellben lehessen tartani, a standardizálás során az átlag és a szórás kiszámításánál a csőd előtti év adatát nem vettem figyelembe, mivel az torzítóan hatna ezen alapstatisztikai mutatók értékeire.

Látható, hogy a ROA mutató értéke 2008-tól kezdődően romló tendenciát mutat, melynek mértéke a csődhöz közeledve rohamos. Láthatjuk ugyanakkor azt is, hogy 2002-ben a mutató még a csőd előtti év adatánál is alacsonyabb volt. Abban az esetben, ha erre az idősorra alkalmaznánk a dinamikus mutatók 3. fejezetben javasolt képletét, akkor a csőd előtti évre vonatkozóan az alábbi számítást kellene elvégezni:

$$\frac{-4,9112 - (-5,522)}{0,1671 - (-5,522)} = 0,107$$

Ez az érték úgy értelmezhető, hogy a cég legutóbbi üzleti évében realizált ROA mutatója az azt megelőző időszak megfelelő értékeihez viszonyítva alacsony értéket vett fel – a legutolsó megfigyelt évben csak 10 %-kal volt magasabb, mint az azt megelőző időszak minimuma, ha a legutolsó megfigyelt évet megelőző teljes időszak terjedelmét tekintjük 100 %-os viszonyítási alapnak. Mivel a vizsgált vállalkozás „túlélt” már ennél gyengébb mutatóval zárt évet (2002), így a dinamikus mutatók nem feltétlenül segítik, hogy reálisan lássuk a vállalat pénzügyi helyzetét, ugyanis ez a vállalkozás a következő évben felszámolási eljárás alá került.

Feltételezhető azonban, hogy a 2002-ben látott alacsony mutató még a vállalkozás kezdeti nehézségeinek, vagy valamilyen egyéb egyedi esetnek tudható be, ami nem eredményezte a vállalkozás felszámolását. Megfigyelhetjük azt is, hogy a mutató standardizált értéke ebben az évben a két szórás terjedelmen kívülre esik. Ebből

adódóan célszerűnek látszik ezt az értéket helyettesíteni az idősor egy olyan értékével, amely már nem minősül outliernek a két szórás „szabály” alkalmazásával.

A helyettesítés kapcsán fel kell hívni a figyelmet arra, hogy értekezésem alapkoncepciója, hogy a csőd előtti évben gyakran megfigyelhető kiugró értékek jelentősen segíthetik a csődös vállalatok azonosítását, ezért a legfrissebb pénzügyi adatok esetén kiemelt figyelmet fordítottam arra, hogy azok információtartalma változatlan formában felhasználható legyen a klasszifikációs módszer számára. Ennek érdekében az egyes vállalkozások pénzügyi mutatóinak idősoraiban az adatelőkészítési feladatok (átlag, szórás értékének meghatározása, kiugró értékek helyettesítése) során a csőd előtti év adatát nem vettem figyelembe.

Az előbbiekben tárgyalt példa esetén ez annyit jelent, hogy a 2002. év kiugró értékét nem a hozzá legközelebb eső 2011. év adatával helyettesítettem (mivel ebben az esetben ez a csődöt megelőző év), hanem azzal az értékkel, amely a helyettesítendő értékhez a legközelebb esik a legfrissebb adat figyelmen kívül hagyásával. Ebben az esetben ez a 2010. év mutatója. A helyettesítést követően a vizsgált idősor értékeit mutatja a következő táblázat, amelynek utolsó oszlopában már az idősor új elemeinek felhasználásával számított átlag és szórás alapján standardizált értékek szerepelnek.

5. Táblázat: A kiugró értéket tartalmazó idősor értékei a helyettesítést követően

Üzleti év	ROA	Standardizált ROA
2011	-4,9112	-5,2845
2010	-1,8360	-1,5345
2009	0,1153	0,8448
2008	0,1671	0,9080
2007	0,0403	0,7534
2006	-0,0924	0,5916
2005	-0,1939	0,4678
2004	0,0284	0,7389
2003	-0,4214	0,1905
2002	-1,8360	-1,5345
2001	-1,7470	-1,4260

A helyettesítést követően láthatjuk, hogy az idősor valamennyi értéke a két szórás terjedelmen belülre esik, így a kiugró értékek további kezelésére nincs szükség. Ha az így kapott idősorra alkalmazzuk a dinamikus mutatók ismertetésénél bemutatott formulát, akkor az alábbi számítást kell elvégeznünk:

$$\frac{-4,9112 - (-1,836)}{0,1671 - (1,836)} = -1,535$$

Az így kapott eredmény már azt tükrözi, hogy a vállalat legutolsó évben realizált ROA mutatója jóval alacsonyabb, mint az azt megelőző időszak megfelelő értékei. Tekintettel arra, hogy a vizsgált vállalkozás a következő évben ténylegesen felszámolási eljárás alá került, a dinamikus mutató imént kapott értéke már sokkal inkább szolgálja a következő évben bekövetkezett esemény előrejelzését, mint az, amely korábban a kiugró érték korrekciója nélkül adódott (0,107).

A bemutatott eset arra is empirikus példát mutat, hogy miért érdemes a statisztikai hüvelykujj-szabályoknál szigorúbb „definíciót” használni a kiugró értékek azonosítására. A vizsgált vállalat esetén a később helyettesített mutató standardizált

értéke -2,72 volt, azaz a hagyományos három szórás „szabály” keretei közt a hasonló nagyságú értékeket még nem helyettesítenénk, melynek következtében feltételezésem szerint a dinamikus mutatók előrejelző képessége kisebb lenne.

Az értékezés harmadik hipotézise azt vizsgálja, hogy mennyiben tekinthető az előbb bemutatott példa egyedi esetnek; más szóval növekszik-e a dinamikus mutatók előrejelző képessége a csődmodellekben, ha a bemutatott helyettesítést az összes megfigyelés valamennyi pénzügyi mutatója esetén elvégezzük.

A hipotézis vizsgálatát úgy hajtottam végre, hogy a dinamikus mutatók vonatkozásában elvégeztem a fent bemutatott outlier helyettesítést, majd az így kapott dinamikus változókat szerepeltettem a statikus változók mellett a CHAID módszerrel felállított csődmodellek független változói közt. A modellek tanuló és tesztelő mintái pontosan megegyeztek a második hipotézisben alkalmazott mintákkal, valamint ebben az esetben is ugyanazokat a beállításokat alkalmaztam a döntési fák felállításánál, mint a második hipotézis esetén annak érdekében, hogy a kapott eredmények összehasonlíthatók legyenek. A tízszeres keresztvalidációt követően adódott eredményeket a következő táblázat mutatja.

6. Táblázat: Az empirikus vizsgálat keretében felállított modellek találati aránya

Adathalmaz	Minta	Csoport	Input változók		
			Pénzügyi mutatók	Dinamikus mutatószámok	A két változókör együtt
Korrigálatlan	Tanuló	Fizetésképtelen	85,7%	83,2%	84,6%
		Fizetőképes	80,1%	77,8%	83,8%
		Összes	82,9%	80,5%	84,2%
	Tesztelő	Fizetésképtelen	76,6%	75,0%	76,8%
		Fizetőképes	71,0%	71,6%	74,6%
		Összes	73,8%	73,3%	75,7%
Korrigált	Tanuló	Fizetésképtelen	85,7%	80,0%	86,9%
		Fizetőképes	80,1%	81,8%	83,1%
		Összes	82,9%	80,9%	85,0%
	Tesztelő	Fizetésképtelen	76,6%	71,6%	80,2%
		Fizetőképes	71,0%	72,4%	75,8%
		Összes	73,8%	72,0%	78,0%

A táblázat első oszlopában a „Korrigált” kifejezés arra az esetre utal, amikor a kiugró értékeket az egyes mutatók idősoráiban helyettesítettem a hozzájuk legközelebb eső, de már nem kiugró értékkel a két szórás „szabály” alkalmazásával. A „Korrigálatlan” sorban lévő eredményeket azon modellek mutatták, amelyek dinamikus változói esetén korrekcióra nem került sor. Az ott látható eredmények azonosak a 3. táblázatban bemutatott eredményekkel.

Ebben az esetben is megállapítható, hogy a dinamikus mutatókat önállóan még a korrekciót követően sem célszerű a modellek magyarázó változóiként felhasználni, mivel az így kapott modellek előrejelző képessége jellemzően gyengébb, mint azon modelleké, amelyekben csak a statikus pénzügyi mutatószámok szerepelnek. Javulás ebben az esetben is csak akkor mutatkozik, ha a két változócsoporthoz együtt szerepeltetjük a modellekben. A korrekciót követően azonban a javulás mértéke nagyobb: a tanuló minta esetén 2,1, míg a tesztelő minta esetén 4,2 százalékpont. (A javulás mértéke a korrekció nélkül rendre 1,3 és 1,9 százalékpont volt.)

A bemutatott eredmények alapján az értekezés harmadik hipotézisét is elfogadtam, mivel a kiugró értékek korrekcióját követően kapott dinamikus mutatók alkalmazásával a vizsgált minta esetén növekedett a modellek előrejelző képessége azon esethez képest, amikor a dinamikus mutatók számításánál az outlier értékek kezelésére nem került sor.

5.4. Az empirikus vizsgálatok eredményeinek összegzése

A bemutatott eredmények alapján megállapítható, hogy függetlenül attól, hogy korrigáljuk-e a mutatók idősorokban található kiugró értékeit, a CHAID módszerrel felállított döntési fák előrejelző képessége érdemben növekszik abban az esetben, amikor a statikus pénzügyi mutatók mellett azt is szerepeltetjük a modellekben, hogy a legutoljára megfigyelt üzleti évben a mutatók értéke hogyan viszonyul az azt megelőző időszak megfelelő értékeihez. Az empirikus vizsgálatok arra is rámutattak, hogy célszerű a mutatószám idősorokban a kiugró értékeit helyettesíteni a hozzájuk legközelebb eső, de már nem kiugró értékekkel.

Érdekes eredmény, hogy a kiugró értékek kezelésétől függetlenül igaz, hogy összességében magasabb volt a fizetésképtelen vállalkozások helyes besorolási aránya, mint a fizetőképes vállalkozásoké. Ez az eredmény arra utal, hogy hazánkban a válság idején „könnyebb” a fizetésképtelen vállalatokat azonosítani, mint a fizetőképeseket. A meglepő eredményre magyarázatot adhat a függelékben található táblázat, ahol megfigyelhető, hogy a válság idején a hazai vállalkozások szolvenciájuktól függetlenül átlagosan negatív jövedelmezőséggel, illetve negatív nettó forgótőke aránnyal jellemezhetők. Továbbá az is látható, hogy a fizetőképes vállalatok is jellemzően gyengébb teljesítményt mutattak a legutoljára megfigyelt

évben az azt megelőző időszakhoz viszonyítva. Ezek ismeretében kevésbé meglepő eredmény, hogy a modellek nagyobb megbízhatósággal ismerték fel a fizetéseképtelen, mint a fizetőképes vállalatokat, mivel a recesszió idején a működő vállalkozásoknak is nagyobb eséllyel lehettek hosszabb-rövidebb ideig pénzügyi problémái.

6. ÖSSZEGZÉS

Értekezésemet a témaválasztás indoklásával, valamint a fogalmi keretek tisztázásával kezdtem. Ezt követően bemutattam a csődelőrejelzés hazai és nemzetközi szakirodalmát, az utóbbin belül körvonalazható fő kutatási irányokat és rámutattam néhány olyan nyitott kutatási kérdésre, amelyek közül választásom a csődelőrejelző modellek dinamizálására esett. Ezt a feladatot egy olyan formula segítségével kíséreltem meg megvalósítani, amely a vállalatok legutoljára megfigyelt pénzügyi mutatóit az azt megelőző időszak megfelelő értékeihez viszonyítja, lehetővé téve, hogy bonyolultabb módszertani megoldások alkalmazása nélkül figyelembe vegyük a pénzügyi mutatók időbeli trendjét a tudományterületen általánosan elterjedt statikus jellegű klasszifikációs módszerek keretei között.

A dinamikus mutatók előrejelző képességét egy saját adatgyűjtésből származó 1000 elemű minta adatain vizsgáltam; külön gondot fordítva a mutatószámok idősoraiban található kiugró értékek torzító hatásának kiküszöbölésére. Az empirikus vizsgálatok eredményei alapján megerősítést nyert azon hipotézisem, miszerint a csődelőrejelző modellek előrejelző képessége nem csak fejlettebb klasszifikációs módszerek alkalmazásával javítható. Erre lehetőséget teremt a pénzügyi mutatókban rejlő információtartalom eddigieknél magasabb szintű felhasználása. A CHAID módszerrel felállított döntési fák esetén azt tapasztaltam, hogy a modellek előrejelző képessége érdemben növekszik akkor, ha a csődmodellezésben általánosan elterjedt statikus pénzügyi mutatók mellett szerepeltetjük az értekezésben javasolt dinamikus változók értékeit is.

Munkám záró fejezetének céljai:

- elhelyezni az értekezést a csődelőrejelzés szakirodalmán belül, valamint bemutatni, hogy miben járult hozzá annak fejlődéséhez;
- ismertetni a dolgozatban kifejtett empirikus vizsgálat korlátait, hiányosságait;
- vázolni azokat a lehetséges kutatási irányokat, amelyeket a dolgozat eredményei alapján a jövőben vizsgálni érdemes.

6.1. Az értekezés hozzájárulása a tudományterület fejlődéséhez

A dolgozat legnagyobb terjedelmű fejezete a szakirodalmi áttekintést tárgyalta. A relatíve nagy terjedelem ellenére hangsúlyozni kell, hogy a szakirodalmi áttekintés közel sem tekinthető teljesnek. Ennek oka, hogy a témában megjelenő publikációk száma az utóbbi öt évben drasztikus ütemben növekszik. A hatalmas méretű szakirodalom feldolgozása során arra törekedtem, hogy értekezésemben minél átfogóbb képet tudjak adni a csődelőrejelzés fejlődéséről és jelenlegi helyzetéről, valamint a legfontosabb kutatási irányokról.

Más tudományokhoz hasonlóan ebben az esetben is adekvát szakirodalom fejlődését időrendi sorrendben ismertetni. A témában megjelenő publikációk szinte kivétel nélkül azt a rendezőelvet követik, hogy felidézik a csődelőrejelzés legfontosabb mérföldköveit, majd a vizsgálni kívánt részterülethez kötődő publikációkat ismertetik részletesebben. Az értekezésben hasonló elvet alkalmaztam, de azzal a különbséggel, hogy a témában eddig megjelent legfontosabb kutatásokat a csődmodell-építés egyes fázisaihoz rendeltem hozzá. Ennek alapját az képezi, hogy a csődelőrejelzésben napjainkra az a tendencia vált általánossá, hogy a kutatások csak a modellépítés egyes részterületeihez kötődően fogalmaznak és vizsgálnak kutatási kérdéseket.

Emiatt pedig már az egyes részterületek esetén is beszélhetünk külön szakirodalomról, amelyek persze nem tekinthetők egymástól függetlennek. Munkámban az egyes részterületek vonatkozásában kiemeltem a legfontosabb kutatási kérdéseket, valamint az azok kapcsán végzett főbb empirikus vizsgálatok eredményeit. Ezt a megközelítést a szakirodalom keresztmetszeti áttekintésének neveztam. Mivel hasonló koncepcióval eddigi kutatómunkám során nem találkoztam, ezt tekinthető dolgozatom egyik hozzáadott értékének a tudományterület fejlődéséhez.

Az eddigi megjelent kutatások alapján kiemeltem számos olyan vizsgálati területet, amelyet a szakirodalomban ígéretes jövőbeli kutatási iránynak tekintenek. Ezek közül a csődmodellek statikus jellege egy olyan kérdés, mely már régóta ismert, de érdemben csak az utóbbi években kezdtek vele foglalkozni. A probléma megoldására elsősorban – az értekezésben bemutatottakhoz hasonló – bonyolultabb módszertani megoldások születtek, melyek – vélhetően túlzott komplexitásuk miatt – az ígéretes eredmények ellenére sem terjedtek el általánosan a csődelőrejelzés szakirodalmában. Ebből kiindulva kutatásaim során egy egyszerű, de ugyanakkor hatékony módot kerestem arra vonatkozóan, hogyan lehetne a statikus pénzügyi mutatók időbeli tendenciáit is megjeleníteni a csődelőrejelző modellekben. A probléma megoldására a 3. fejezetben bemutatott dinamikus pénzügyi mutatószámok alkalmazását javasoltam.

Az empirikus vizsgálatok eredményei alapján arra a következtetésre jutottam, hogy a 4. fejezetben bemutatott mintában szereplő 1000 darab hazai vállalkozás vonatkozásában van létjogosultsága a dinamikus mutatók alkalmazásának, ugyanis a CHAID módszerrel felállított döntési fák előrejelző képessége érdemben javult a statikus mutatók időbeli tendenciáját tükröző változók alkalmazásával. Az empirikus

vizsgálatok keretein belül arra is rámutattam, hogy a dinamikus mutatók előrejelző képessége fokozható, ha az egyes vállalkozások pénzügyi mutatószámainak idősorán belül a kiugróan magas, illetve alacsony értékeket helyettesítjük a hozzájuk legközelebb eső, de már nem kiugró értékkel. Ezen eredmények megerősítik azon feltevésemet, mely szerint a csődelőrejelző modellek predikciós ereje nem csak újabb és korszerűbb klasszifikációs módszerek alkalmazásával javítható, mint ahogy ezt a szakirodalomban kirajzolódó főáramú kutatási irányvonal sugallja. Reális kutatási potenciál van a csődelőrejelzés alternatív fejlesztési lehetőségeinek terén is, melyre jelen értekezésben magam is példát találtam. Ebből adódóan munkámat a szakirodalom alternatív kutatási lehetőségei közt helyezem el. Tekintettel arra, hogy az áttekintett szakirodalomban nem talákoztam olyan tanulmánnyal, amely a csődmodellek dinamizálására hasonlóképp tett volna kísérletet, a dinamikus mutatók alkalmazását tekintem értekezésem azon aspektusának, amely a leginkább hozzájárulhat a tudományterület fejlődéséhez.

A modellek előrejelző képességének javítása mellett a dinamikus mutatók hiánypótló szerepet is betölthetnek a csődelőrejelzés elméleti háttérét illetően is. Korábban szó esett arról, hogy a csődelőrejelzés esetén nincs olyan egzakt elméleti háttér, mint a közgazdasági modellek mögött általában. A leggyakrabban hivatkozott elméleti keret Beaver [1966] nevéhez kötődik, aki a vállalkozásokat likvid eszközök víztározójaként értelmezve a csődöt a víztározó kiszáradásaként határozta meg. Fontos szempont lehet azonban ebből a szempontból, hogy a különböző vállalkozások nem tekinthetők teljesen azonos víztározóknak. Felfoghatók azonban likvid eszközök tavaiként, amelyek különböző méretűek és különböző földrajzi elhelyezkedésűek, így kiszáradásukat több tényező is befolyásolhatja. Még fontosabb, hogy e különbségek miatt a tavak nem hasonlíthatók össze közvetlenül. A

tó vízállása (a vállalkozás pénzügyi helyzete) ebben a keretben objektívebben megítélhető, ha azt nem csak abszolút értékben vizsgáljuk, hanem figyelembe vesszük, hogy a legutóbbi vízállás hogyan viszonyul a tó vízállásának korábbi értékeihez.

A dinamikus ráták alkalmazása tehát elméleti szempontból is előrelépést hozhat a csödelőrejelzés tudományterületén, bár hozzá kell tenni, hogy ez sem jelent általános elméleti keretet, mindössze egy kiegészítése lehet Beaver [1966] modelljének. Továbbra sem határozza meg azonban azt, hogy mely mutatószámok alapján lehet a vállalatok jövőbeli fizetéseképtelenségét előrejelezni. Ezen elméleti modell kidolgozása továbbra is a jövő nagy kutatási kihívásai közé tartozik.

6.2. Az értekezésben bemutatott elemzés korlátai

Ugyan az empirikus vizsgálatok eredményei ígéretesek, fel kell hívni a figyelmet a bemutatott elemzés korlátaira is.

Az empirikus vizsgálat alapját képező mintavétel során kérdés lehet, hogy mennyiben gyakorolt hatást a bemutatott eredményekre, hogy az adatgyűjtés során alkalmazott nyilvános adatbázisokban a keresést nem volt lehetőség szűkíteni a vállalatok mérete, kora vagy iparága alapján. Mivel a rendelkezésre álló adatbázis e szempontok szerint meglehetősen heterogén, ezért célszerű lenne a bemutatott elemzést a jövőben homogénebb adathalmazokon is elvégezni.

Bár a kutatási adatbázis nagynak tekinthető (több mint 7500 üzleti évre vonatkozóan tartalmaz adatokat), ennek ellenére mégis azt mondhatjuk, hogy relatíve kis mintán készült az elemzés, mivel az adathalmazban szereplő 1000 vállalkozás számosságát

illetően kevésnek tekinthető az országban működő valamennyi vállalkozás számához képest. Egy nagyobb mintán végzett hasonló vizsgálat fontos megerősítése lenne az értekezésben levont következtetéseknek. A mintaméret tekintetében még előnyösebb lenne egy olyan kutatás, melynek tárgyát a csődelőrejelzésben gyakran vizsgált amerikai, vagy távol-keleti tőzsdei cégek képezik. Kérdés lehet ugyanis, hogy a dinamikus mutatók előnyös tulajdonságai a vállalatok jövőbeli fizetőképességének előrejelzésében mennyire tekinthető egyedi, azaz kizárólag a magyar vállalkozásokra korlátozódó sajátosságnak.

Az empirikus vizsgálatok másik korlátja abból adódik, hogy a mintába csak olyan vállalkozások kerülhettek be, amelyeknek legalább három egymást követő évre vonatkozóan rendelkezésre álltak a pénzügyi mutatói. Ez pedig nyilvánvalóan azzal a következménnyel jár, hogy a három évnél fiatalabb vállalkozások esetén a dinamikus mutatók koncepciója nem alkalmazható. Emiatt az ilyen vállalkozások jövőbeli fizetőképessége csak a statikus pénzügyi mutatókra, illetve egyéb kvalitatív információkra építve jelezhető előre.

Szót kell ejteni a vizsgálatban alkalmazott módszertanról is. A dolgozatban kizárólag a CHAID alapú döntési fák kerültek alkalmazásra klasszifikációs módszerként. A választást az indokolta, hogy ezzel az eljárással relatíve könnyen és gyorsan meglehetősen magas találati aránnyal jellemezhető modell állítható fel. Továbbá jelentős szerepe volt a választásban annak, hogy véleményem szerint a kiugró értékeknek fontos szerepe van a csődelőrejelzésben, mivel azok segíthetik a közelgő fizetéseképtelenség előrejelzését a szélsőségesen gyenge pénzügyi teljesítményt tükröző pénzügyi adatok alapján. Az empirikus vizsgálatok során alkalmazott döntési fák előnye, hogy klasszifikációs teljesítményüket a kiugró értékek jelenléte nem befolyásolja, ami meglehetősen ritka sajátosság a klasszifikációs eljárások között.

Kérdéses lehet azonban, hogy a kapott eredmények mennyiben tudhatók be a választott klasszifikációs módszer sajátosságainak. Ebből a szempontból a jövőben célszerű lehet olyan vizsgálatokat is elvégezni, amelyekben a dinamikus mutatókat többféle klasszifikációs módszer input változói közt is szerepeltetjük.

Végezetül említést kell tenni arról, hogy munkám elsősorban elméleti vonatkozású abban az értelemben, hogy értekezésem fő célja a dinamikus mutatószámok koncepciójának ismertetése volt és nem olyan modellek felállítása, amelyek alkalmasak lehetnek gyakorlati vállalatminősítésre. Ebből adódóan több olyan kérdéskör kimaradt az értekezésből, amely a gyakorlati hitelkockázati modellezésben fontos szerepet játszik: ilyen például az egyes megfigyelésekre fennmaradási valószínűség kalkulációja, azok kalibrációja, vagy a felállított modellek stabilitásvizsgálata. Szintén jövőbeli vizsgálati kérdés, hogy e szempontokat mennyire képesek támogatni a dolgozatban bemutatott dinamikus mutatók.

6.3. Lehetséges jövőbeli kutatási irányok

A tudományos munkák esetén gyakori, hogy egy kutatás több kérdést vet fel, mint amennyit eredetileg meg kívánt válaszolni. Az értekezés végére érve magam is ezzel a ténnyel szembesültem.

Ahogy a dolgozat címe is jelzi, munkám elsődleges célja a dinamikus mutatószámok alkalmazási lehetőségeinek vizsgálata volt a csődelőrejelzés területén. Az értekezésben bemutatott empirikus vizsgálatok arra utalnak, hogy van létjogosultsága a dinamikus szemléletnek a vállalati jövőbeli fizetőképességének előrejelzésében. A dinamikus mutatók számításának koncepciója azonban sok helyen inkább intuitív „megérzéseken”, mint tudományos alapokon nyugszik. Ebből adódóan a bemutatott

elemzésnek számos olyan eleme van, mely a későbbiekben további vizsgálatok tárgyát képezheti. Ezek eredményeképp ugyanis a dinamikus mutatók előrejelző képessége tovább növelhető, lehetővé téve még pontosabb előrejelző modellek felállítását. Ezek közül ismertetek néhányat ebben a szakaszban.

1. A dinamikus mutatók számításánál a legutolsó megfigyelt év adatát viszonyítottam az azt megelőző teljes időszak megfelelő értékeihez. Kérdés azonban, hogy érdemes-e a mutatószámok teljes idősorát figyelembe venni. Nem lenne-e célszerűbb, ha az idősorok csak egy szakaszát – például az elmúlt 5 évet – tekintenénk viszonyítási alapnak? A rövidebb idősorok alkalmazását az is indokolhatja, hogy így csökkenne az idősorban előforduló kiugró értékek előfordulásának esélye, és így az azok kezelésével járó jelentős adatelőkészítési munka.

2. A dolgozatban egy konkrét példán keresztül szemléltettem az idősorokban lévő outlier értékek torzító hatását a dinamikus mutatókra. Az empirikus vizsgálatok eredményei arra is rámutattak, hogy célszerű az ilyen értékek helyettesítése a dinamikus mutatók jobb előrejelző képessége érdekében. Az outlierok azonosítása azonban ebben az esetben is egy általam „önkényesen” meghatározott elv alapján történt objektív definíció hiányában. Külön vizsgálat tárgyát képezheti, hogyan célszerű az outlier értékeket definiálni, hogy annak köszönhetően maximális legyen a dinamikus mutatókat is tartalmazó csődmodellek előrejelző képessége.

3. A dinamikus mutatók ötlete a viszonylagosság elvén alapul, melynek lényege, hogy a pénzügyi mutatók nem jelentenek abszolút kritériumot, azokat csak valamilyen alkalmas viszonyítási alap tükrében lehet objektíven megítélni (Virág et al. [2013]). E viszonyítási alapok közt gyakran utalnak az iparági középértékek használatára. A dinamikus mutatók koncepciója ezen a területen is új eredményeket

hozhat, ugyanis iparági átlagok ismeretében lehetőség van arra, hogy azt számszerűsítsük, hogyan alakulnak az időben a vállalkozások egyes pénzügyi mutatói az iparági tendenciák tükrében. A 2. fejezetben bemutatott Platt-Platt [1990] által javasolt iparági relatív mutatókra vonatkozóan például szintén alkalmazható az értekezésben bemutatott koncepció. Az így képzett dinamikus iparági relatív mutatók tovább növelhetik a csődmodellek előrejelző képességét.

FÜGGELÉK

F.1. Táblázat: Az empirikus vizsgálatban használt pénzügyi mutatók alapstatisztikai mutatói

Mutatószám neve	Csoport	Átlag	Szórás
Likviditási ráta	Fizetőképés	39,82	381,86
	Fizetéseképtelen	16,38	224,52
D_Likviditási ráta	Fizetőképés	1,38	9,11
	Fizetéseképtelen	22,70	348,44
Likviditási gyorsráta	Fizetőképés	38,96	381,77
	Fizetéseképtelen	16,09	224,47
D_Likviditási gyorsráta	Fizetőképés	1,40	9,17
	Fizetéseképtelen	19,21	312,11
Pénzeszközök aránya*	Fizetőképés	0,32	0,33
	Fizetéseképtelen	0,15	0,27
D_Pénzeszközök aránya	Fizetőképés	1,00	8,13
	Fizetéseképtelen	2,63	25,54
CF/Adósság	Fizetőképés	4,31	139,08
	Fizetéseképtelen	-0,92	11,74
D_CF/Adósság*	Fizetőképés	0,83	14,19
	Fizetéseképtelen	-2,28	25,69
CF/Rövid lejáratú kötelezettségek	Fizetőképés	4,46	139,14
	Fizetéseképtelen	-1,28	12,70
D_CF/Rövid lejáratú kötelezettségek	Fizetőképés	0,73	14,25
	Fizetéseképtelen	-0,69	8,54
Tőkeellátottság	Fizetőképés	2,65	24,46
	Fizetéseképtelen	105,98	2359,90
D_Tőkeellátottság	Fizetőképés	58,55	1259,06
	Fizetéseképtelen	-2,92	76,36
Eszközök forgási sebessége	Fizetőképés	4,76	28,33
	Fizetéseképtelen	14,49	171,90
D_Eszközök forgási sebessége	Fizetőképés	0,90	4,99
	Fizetéseképtelen	9,13	109,87
Készletek forgási sebessége	Fizetőképés	65334,50	688444,08
	Fizetéseképtelen	18316,01	72091,13
D_Készletek forgási sebessége*	Fizetőképés	54,23	685,75
	Fizetéseképtelen	693,64	5675,54
Követelések forgási ideje*	Fizetőképés	112,44	1184,25
	Fizetéseképtelen	2023,27	17177,09
D_Követelések forgási ideje*	Fizetőképés	94,90	1361,02
	Fizetéseképtelen	2827,33	20538,81

Eladósodottság*	Fizetőképes	17,64	244,48
	Fizetési képtelen	153,17	1206,91
D_Eladósodottság*	Fizetőképes	1,36	11,79
	Fizetési képtelen	263,86	2800,30
Saját tőke aránya*	Fizetőképes	-16,83	244,53
	Fizetési képtelen	-164,61	1230,10
D_Saját tőke aránya*	Fizetőképes	-0,35	11,77
	Fizetési képtelen	-298,95	3120,05
Bonitás	Fizetőképes	7,51	84,73
	Fizetési képtelen	212,76	4758,64
D_Bonitás	Fizetőképes	13,59	267,85
	Fizetési képtelen	0,02	14,98
Árbevétel-arányos nyereség	Fizetőképes	-323,96	7036,36
	Fizetési képtelen	-1549,67	12425,06
D_Árbevétel-arányos nyereség	Fizetőképes	-48,92	660,10
	Fizetési képtelen	-20988,26	358038,57
Eszközarányos nyereség*	Fizetőképes	-3,17	63,44
	Fizetési képtelen	-92,61	923,00
D_Eszközarányos nyereség*	Fizetőképes	-0,31	8,16
	Fizetési képtelen	-159,18	1353,89
Követelések/Rövid lejáratú kötelezettségek	Fizetőképes	21,35	249,91
	Fizetési képtelen	7,47	134,47
D_Követelések/Rövid lejáratú kötelezettségek	Fizetőképes	1,45	9,79
	Fizetési képtelen	1,54	28,50
Nettó forgótőke aránya*	Fizetőképes	-14,76	241,86
	Fizetési képtelen	-134,64	1130,71
D_Nettó forgótőke aránya	Fizetőképes	-0,10	10,92
	Fizetési képtelen	-208,37	2596,76
Méret	Fizetőképes	9,65	2,59
	Fizetési képtelen	9,61	2,86
D_Méret	Fizetőképes	0,47	1,03
	Fizetési képtelen	-5,23	91,05
Évek*	Fizetőképes	8,39	2,89
	Fizetési képtelen	6,80	2,82

HIVATKOZÁSOK

Abdou, H. – Pointon, J. [2011]: Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management*. Vol. 18., p. 59-88., DOI: <http://dx.doi.org/10.1002/isaf.325>.

Abellán, J. – Mantas, C. J. [2014]: Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*. Vol. 41., p. 3825-3830, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.12.003>.

Agarwal, V. – Taffler, R. [2008]: Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*. Vol. 32., p. 1541-1551, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2007.07.014>.

Agarwal, V. – Taffler, R. J. [2007]: Twenty-five years of the Taffler z-score model: does it really have predictive ability? *Accounting and Business Research*. Vol. 37., No. 4., p. 285-300, DOI: <http://dx.doi.org/10.1080/00014788.2007.9663313>.

Ahn, B. S. – Cho, S. S. – Kim, C. Y. [2000]: The integrated methodology of rough set theory and artificial neural networks for business failure prediction. *Expert Systems with Applications*. Vol. 18., No. 2., p. 65-74, DOI: [http://dx.doi.org/10.1016/S0957-4174\(99\)00053-6](http://dx.doi.org/10.1016/S0957-4174(99)00053-6).

Akkoc, S. [2012]: An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*. Vol. 222., p. 168-178, DOI: <http://dx.doi.org/10.1016/j.ejor.2012.04.009>.

Alfaro, E. – García, N. – Gámez, M. – Elizondo, D. [2008]: Bankruptcy forecasting: An empirical comparison of Adaboost and neural networks. *Decision Support Systems*. Vol. 45., p. 110-122, DOI: <http://dx.doi.org/10.1016/j.dss.2007.12.002>.

Alici, Y. [1995]: Neural networks in corporate failure prediction: The UK experience. In: Refenes, A. N. – Abu-Mostafa, Y. – Moody, J. – Weigend, A. (eds.): *Processing Third International Conference of Neural Networks in the Capital Markets*, London, October 1995, p. 393-406.

Altman, E. I. – Marco, G. – Varetto, F. [1994]: Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*. Vol. 18., p. 505-529, DOI: [http://dx.doi.org/10.1016/0378-4266\(94\)90007-8](http://dx.doi.org/10.1016/0378-4266(94)90007-8).

Altman, E. I. [1968]: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*. Vol. 23., No. 4., p. 589-609, DOI: <http://dx.doi.org/0.1111/j.1540-6261.1968.tb00843.x>.

Altman, E. I. [1977]: ZETA analysis. A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*. Vol. 1., p. 29-54, DOI: [http://dx.doi.org/10.1016/0378-4266\(77\)90017-6](http://dx.doi.org/10.1016/0378-4266(77)90017-6).

Altman, E. I. [2002]: *Corporate distress prediction models in a turbulent economic and BASEL II environment*. Salomon Center for the Study of Financial Institutions <New York, NY> / Credit & Debt Markets Research Program.

Anandarajan, M. – Lee, P. – Anandarajan, A. [2001]: Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. *Intelligent Systems in Accounting, Finance and Management*. Vol. 10., p. 69-81, DOI: <http://dx.doi.org/10.1002/isaf.199>.

Atiya, A. F. [2001]: Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results, *IEEE Transactions on Neural Networks*, Vol. 12, No. 4, p. 929-935.

Back, B. – Laitinen, T. – Sere, K. – van Wezel, M. [1996]: *Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms*. Technical Report No. 40. Turku Centre for Computer Science, Turku.

Bae, J. K. [2012]: Predicting financial distress of the South Korean manufacturing industries. *Expert Systems with Applications*. Vol. 39., p. 9159-9165, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.058>.

Baesens, B. – Setiono, R. – Mues, C. – Vanthienen, J. [2003]: Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science*. Vol. 49., No. 3., p. 312-329, DOI: <http://dx.doi.org/10.1287/mnsc.49.3.312.12739>.

Balcean, S. – Ooghe, H. [2006]: 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*. Vol. 38., p. 63-93., DOI: <http://dx.doi.org/10.1016/j.bar.2005.09.001>.

Barniv, R. – Agarwal, A. – Leach, R. [2002]: Predicting bankruptcy resolution. *Journal of Business Finance & Accounting*. Vol. 29., No. 3-4., p. 497-520, DOI: <http://dx.doi.org/10.1111/1468-5957.00440>.

Beaver, W. H. – McNichols, M. F. – Rhie, J. W. [2005]: Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*. Vol. 10., p. 93-122, DOI: <http://dx.doi.org/10.1007/s11142-004-6341-9>.

Beaver, W. H. [1966]: Financial ratios as predictors of failure. *Journal of Accounting Research*. Vol. 4., Supplement, p. 71-111, DOI: <http://dx.doi.org/10.2307/2490171>.

Becchetti, L. – Sierra, J. [2003]: Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*. Vol. 27., p. 2099-2120, DOI: [http://dx.doi.org/10.1016/S0378-4266\(02\)00319-9](http://dx.doi.org/10.1016/S0378-4266(02)00319-9).

Bellotti, T. – Matousek, R. – Stewart, C. [2011]: A note comparing support vector machines and ordered choice models' predictions of international banks' ratings. *Decision Support Systems*. Vol. 51., p. 682-687, DOI: <http://dx.doi.org/10.1016/j.dss.2011.03.008>.

Berg, D. [2007]: Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*. Vol. 23., p. 129-143, DOI: <http://dx.doi.org/10.1002/asmb.658>.

Bernhardsen, E. [2001]: *A model of bankruptcy prediction*. Working Paper, Financial Analysis and Structure Department, Research Department, Norges Bank, Oslo. <http://www.norges-bank.no/Upload/import/publikasjoner/arbeidsnotater/pdf/arb-2001-10.pdf>

Beynon, M. J. – Peel, M. J. [2001]: Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega*. Vol. 29., p. 561-576, DOI: [http://dx.doi.org/10.1016/S0305-0483\(01\)00045-7](http://dx.doi.org/10.1016/S0305-0483(01)00045-7).

Bharath, S. T. – Shumway, T. [2008]: Forecasting default with the Merton distance to default model. *The Review of Financial Studies*. Vol. 21., No. 3., p. 1339-1369, DOI: <http://dx.doi.org/10.1093/rfs/hhn044>.

Bioch, J. C. – Popova, V. [2001]: *Bankruptcy prediction with rough sets*. No. ERS-2001-11-LIS). Erasmus Research Institute of Management (ERIM).

Blanco, A. – Pino-Mejías, R. – Lara, J. – Rayo, S. [2013]: Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*. Vol. 40., p. 356-364, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.07.051>.

Bonfim, D. [2009]: Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance*. Vol. 33., p. 281-299, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2008.08.006>.

Bongini, P. – Ferri, G. – Hahm, H. [2000]: Corporate bankruptcy in Korea: Only the strong survive? *The Financial Review*. Vol. 35., No. 4., p. 31-50, DOI: <http://dx.doi.org/10.1111/j.1540-6288.2000.tb01428.x>.

Bose, I. [2006]: Deciding the financial health of dot-coms using rough sets. *Information & Management*. Vol. 43., p. 835-846, DOI: <http://dx.doi.org/10.1016/j.im.2006.08.001>.

Boyacioglu, M. A. – Kara, Y. – Baykan, Ö. K. [2009]: Predicting bank failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SFID) transferred banks in Turkey. *Expert Systems with Applications*. Vol. 36., p. 3355-3366, DOI: <http://dx.doi.org/10.1016/j.eswa.2008.01.003>.

Brealey, R. A. – Myers, S. C. [2005]: *Modern Vállalati Pénzügyek*. Panem, Budapest.

Brezigar-Masten, A. – Masten, I. [2012]: CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*. Vol. 39., p. 10153-10159, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.125>.

Brown, I. – Mues, C. [2012]: An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*. Vol. 39., p. 3446-3453, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.09.033>.

Cantner, U. – Dressler, K. – Krüger, J. J. [2006]: Firm survival in the German automobile industry. *Empirica*. Vol. 33., p. 49-60, DOI: <http://dx.doi.org/10.1007/s10663-006-9006-z>.

Cao, Y. [2012a]: MCELCCh-FDP: Financial distress prediction with classifier ensembles based on firm life cycle and Choquet integral. *Expert Systems with Applications*. Vol. 39., p. 7041-7049, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.01.043>.

Cao, Y. [2012b]: Aggregating multiple classification results using Choquet integral for financial distress early warning. *Expert Systems with Applications*. Vol. 39., p. 1830-1836, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.08.067>.

Carling, K. – Jacobson, T. – Lindé, J. – Roszbach, K. [2007]: Corporate credit risk modeling and the macroeconomy. *Journal of Banking & Finance*. Vol. 31., p. 845-868, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2006.06.012>.

Chava, S. – Jarrow, R. A. [2004]: Bankruptcy prediction with industry effects. *Review of Finance*. Vol. 8., No. 4., p. 537-569, DOI: <http://dx.doi.org/10.1093/rof/8.4.537>.

Chen, J. H. [2012]: Developing SFNN models to predict financial distress of construction companies. *Expert Systems with Applications*. Vol. 39., p. 823-827, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.07.080>.

Chen, M. Y. [2011]: Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*. Vol. 38., p. 11261-11272, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.02.173>.

Chen, N. – Riberio, B. – Vieira, A. – Chen, A. [2013]: Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications*. Vol. 40., p. 385-393, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.07.047>.

Chen, N. – Riberio, B. – Vieira, A. S. – Duarte, J. – Neves, J. C. [2011]: A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. *Expert Systems with Applications*. Vol. 38., p. 12939-12945, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.04.090>.

Chi, B. W. – Hsu, C. C. [2012]: A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. *Expert Systems with Applications*. Vol. 39., p. 2650-2661, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.08.120>.

Coats, P. – Fant, L. [1993]: Recognizing financial distress patterns using a neural network tool. *Financial Management*. Vol. 22., No. 3., p. 142-155.

Cubiles-De-La-Vega, M. D. – Blanco-Oliver, A. – Pino-Mejías, R. – Lara-Rubio, J. [2013]: Improving the management of microfinance institutions by using credit scoring models based on statistical learning techniques. *Expert Systems with Applications*. Vol. 40., No. 17., p. 6910-6917, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.06.031>.

Dambolena, I. G. – Khoury, S. J. [1980]: Ratio stability and corporate failure. *The Journal of Finance*. Vol. 35., No. 4., p. 1017-1026, DOI: <http://dx.doi.org/10.1111/j.1540-6261.1980.tb03517.x>.

Darayseh, M. – Waples, E. – Tsoukalas, D. [2003]: Corporate failure for manufacturing industries using firms specifics and economic environment with logit analysis. *Managerial Finance*. Vol. 28., No. 8., p. 23-36, DOI: <http://dx.doi.org/10.1108/03074350310768409>.

Das, S. R. – Hanouna, P. – Sarin, A. [2009]: Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance*. Vol. 33., p. 719-730, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2008.11.003>.

Deakin, E. B. [1972]: A discriminant analysis of predictors of failure. *Journal of Accounting Research*. Vol. 10., No. 1., p. 167-179, DOI: <http://dx.doi.org/10.2307/2490225>.

Demers, E. – Joos, P. [2007]: IPO failure risk. *Journal of Accounting Research*. Vol. 45., No. 2., p. 333-371, DOI: <http://dx.doi.org/10.1111/j.1475-679X.2007.00236.x>.

Dewaelheyns, N. – Van Hulle, C. [2006]: Corporate failure prediction modeling: Distorted by business groups' internal capital markets? *Journal of Business Finance & Accounting*. Vol. 33., Vol. 5-6., p. 909-931, DOI: <http://dx.doi.org/10.1111/j.1468-5957.2006.00009.x>.

Ding, Y. – Song, X. – Zen, Y. [2008]: Forecasting financial condition of Chinese listed companies based on support vector machines. *Expert Systems with Applications*. Vol. 34., p. 3081-3089, DOI: <http://dx.doi.org/10.1016/j.eswa.2007.06.037>.

Doumplos, M. – Gaganis, C. – Pasiouras, F. [2005]: Explaining qualifications in audit reports using a support vector machine methodology. *Intelligent Systems in Accounting, Finance and Management*. Vol. 13., p. 197-215, DOI: <http://dx.doi.org/10.1002/isaf.268>.

Du Jardin, P. – Séverin, E. [2012]: Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*. Vol. 221., p. 378-396, DOI: <http://dx.doi.org/10.1016/j.ejor.2012.04.006>.

Du Jardin, P. [2010]: Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*. Vol. 73., p. 2047-2060, DOI: <http://dx.doi.org/10.1016/j.neucom.2009.11.034>.

Duan, J. C. – Sun, J. – Wang, T. [2012]: Multiperiod corporate default prediction – A forward intensity approach. *Journal of Econometrics*. Vol. 170., p. 191-209, DOI: <http://dx.doi.org/10.1016/j.jeconom.2012.05.002>.

Duffie, D. – Saita, L. – Wang, K. [2007]: Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*. Vol. 83., p. 635-665, DOI: <http://dx.doi.org/10.1016/j.jfineco.2005.10.011>.

Elliott, R. J. – Siu, T. K. – Fung, E. S. [2014]: A double HMM approach to Altman Z-scores and credit ratings. *Expert Systems with Applications*. Vol. 41., p. 1553-1560, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.08.052>.

Erdal, H. I. – Ekinici, A. [2012]: A comparison of various artificial intelligence methods in the prediction of bank failures. *Computational Economics*. Vol. 42., No. 2., p. 199-215, DOI: <http://dx.doi.org/10.1007/s10614-012-9332-0>.

Fan, A. – Palaniswami, M. [2000]: A new approach to corporate loan default prediction from financial statements. In: *Proceedings of Computational Finance/Forecasting Financial Markets Conference CF/FFM-2000*, London.

Fedorova, E. – Gilenko, E. – Dovzhenko, S. [2013]: Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*. Vol. 40., p. 7285-7293, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.07.032>.

Feki, A. – Ishak, A. B. – Feki, S. [2012]: Feature selection using Bayesian and multiclass support vector machines approaches: Application to bank risk prediction. *Expert Systems with Applications*. Vol. 39., p. 3087-3099, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.08.172>.

Foreman, R. D. [2003]: A logistic analysis of bankruptcy within the US local telecommunications industry. *Journal of Economics and Business*. Vol. 55., p. 135-166, DOI: [http://dx.doi.org/10.1016/S0148-6195\(02\)00133-9](http://dx.doi.org/10.1016/S0148-6195(02)00133-9).

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., p. 269-291, DOI: <http://dx.doi.org/10.1111/j.1540-6261.1985.tb04949.x>.

Gaganis, C. – Pasiouras, F. – Spathis, C. – Zopounidis, C. [2007]: A comparison of nearest neighbours, discriminant analysis and logit models for auditing decisions. *Intelligent Systems in Accounting, Finance and Management*. Vol. 15., p. 23-40, DOI: <http://dx.doi.org/10.1002/isaf.283>.

Gaganis, C. [2009]: Classification techniques for the identification of falsified financial statements: A comparative analysis. *Intelligent Systems in Accounting, Finance and Management*. Vol. 16., p. 207-229, DOI: <http://dx.doi.org/10.1002/isaf.303>.

García, V. – Marqués, A. I. – Sánchez, J. S. [2012]: On the use of data filtering techniques for credit risk prediction with instance-based models. *Expert Systems with Applications*. Vol. 39., p. 13267-13276, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.05.075>.

Gersbach, H. – Lipponer, A. [2003]: Firm defaults and the correlation effect. *European Financial Management*. Vol. 9., No. 3., p. 361-377, DOI: <http://dx.doi.org/10.1111/1468-036X.00225>.

Gilbert, L. R. – Menon, K. – Schwarz, K. B. [1990]: Predicting bankruptcy for firms in financial distress. *Journal of Business Finance & Accounting*. Vol. 17., No. 1., p. 161-171, DOI: <http://dx.doi.org/10.1111/j.1468-5957.1990.tb00555.x>.

Glennon, D. – Nigro, P. [2005]: Measuring the default risk of small business loans: A survival analysis approach. *Journal of Money, Credit and Banking*. Vol. 37., No. 5., p. 923-947.

Grunert, J. – Norden, L. – Weber, M. [2005]: The role of none-financial factors in internal credit ratings. *Journal of Banking & Finance*. Vol. 29., p. 509-531, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2004.05.017>.

Gruszczynski, M. [2004]: Financial distress of companies in Poland. *International Advances in Economic Research*. Vol. 10., No. 4., p. 249-256, DOI: <http://dx.doi.org/10.1007/BF02295137>.

Hájek, P. [2011]: Municipal credit rating modeling by neural networks. *Decision Support Systems*. Vol. 51., p. 108-118, DOI: <http://dx.doi.org/10.1016/j.dss.2010.11.033>.

Harada, K. – Ito, T. – Takahashi, S. [2013]: Is the distance to default a good measure in predicting bank failures? A case study of Japanese major banks. *Japan and the World Economy*. Vol. 27., p. 70-82, DOI: <http://dx.doi.org/10.1016/j.japwor.2013.03.007>.

Harada, N. – Kageyama, N. [2011]: Bankruptcy dynamics in Japan. *Japan and the World Economy*. Vol. 23., p. 119-128, DOI: <http://dx.doi.org/10.1016/j.japwor.2011.01.002>.

Hardle, W. – Lee, Y. J. – Schafer, D. – Yeh, Y. R. [2009]: Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies. *Journal of Forecasting*. Vol. 28., p. 512-534, DOI: <http://dx.doi.org/10.1002/for.1109>.

Harris, T. [2013]: Quantitative credit risk assessment using support vector machines: Broad versus narrow default definitions. *Expert Systems with Applications*. Vol. 40., p. 4404-4413, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.01.044>.

Heiss, F. – Köke, J. [2004]: Dynamics in ownership and firm survival: Evidence from corporate Germany. *European Financial Management*. Vol. 10., No. 1., p. 167-195. DOI: <http://dx.doi.org/10.1111/j.1468-036X.2004.00244.x>.

Hillegeist, S. A. – Keating, E. K. – Cram, D. P. – Lundstedt, K. G. [2004]: Assessing the probability of bankruptcy. *Review of Accounting Studies*. Vol. 9., p. 5-34., DOI: <http://dx.doi.org/10.2139/ssrn.307479>.

Horta, L. M. – Camanho, A. S. [2013]: Company failure prediction in the construction industry. *Expert Systems with Applications*. Vol. 40., p. 6253-6257, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.05.045>.

Hsieh, T. J. – Hsiao, H. F. – Yeh, W. C. [2012]: Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm. *Neurocomputing*. Vol. 82., p. 196-206, DOI: <http://dx.doi.org/10.1016/j.neucom.2011.11.020>.

Hu, Y. C. [2009]: Bankruptcy prediction using ELECTRE-based single-layer perceptron. *Neurocomputing*. Vol. 72., p. 3150-3157, DOI: <http://dx.doi.org/10.1016/j.neucom.2009.03.002>.

Hu, Y-C. – Tseng, F-M. [2007]: Functional-link net with fuzzy integral for bankruptcy prediction. *Neurocomputing*, Vol. 70, p. 2959-2968, DOI: <http://dx.doi.org/10.1016/j.neucom.2006.10.111>.

Huang, S. C. – Tang, Y. C. – Lee, C. W. – Chang, M. J. [2012]: Kernel local Fisher discriminant analysis based manifold-regularized SVM model for financial distress prediction. *Expert Systems with Applications*. Vol. 39., p. 3855-3861, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.09.095>.

Huang, Z. – Chen, H. – Hsu, C. H. – Chen, W. H. – Wu, S. [2004]: Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*. Vol. 37., p. 543-558, DOI: [http://dx.doi.org/10.1016/S0167-9236\(03\)00086-1](http://dx.doi.org/10.1016/S0167-9236(03)00086-1).

Huynh, K. P. – Petrunia, R. J. – Voia, M. [2010]: The impact of initial financial state on firm duration across entry cohorts. *The Journal of Industrial Economics*. Vol. 58., No. 3., p. 661-689, DOI: <http://dx.doi.org/10.1111/j.1467-6451.2010.00429.x>.

Ioannidis, C. – Pasiouras, F. – Zopounidis, C. [2010]: Assessing bank soundness with classification techniques. *Omega*. Vol. 38., p. 345-357, DOI: <http://dx.doi.org/10.1016/j.omega.2009.10.009>.

Jagtiani, J. – Kolari, J. – Lemieux, C. – Shin, H. [2003]: Early warning models for bank supervision: Simpler could be better. *Federal Reserve Bank of Chicago Economic Perspectives*. Vol. 27., p. 49-60.

Jeong, C. – Min, J. H. – Kim, M. S. [2012]: A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systemes with Applications*. Vol. 39., p. 3650-3658, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.09.056>.

Jones, F. L. [1987]: Current techniques in bankruptcy prediction. *Journal of Accounting Literature*. Vol. 6., p. 131-164.

Jones, S. – Hensher, D. A. [2004]: Predicting firm financial distress: A mixed logit model. *The Accounting Review*. Vol. 79., No. 4., p. 1011-1038, DOI: <http://dx.doi.org/10.2308/accr.2004.79.4.1011>.

Jones, S. – Hensher, D. A. [2007]: Modelling corporate failure: A multinomial nested logit analysis for unordered outcomes. *The British Accounting Review*. Vol. 39., p. 89-107, DOI: <http://dx.doi.org/10.1016/j.bar.2006.12.003>.

Kennedy, K. – Mac Namee, B. – Delany, S. J. [2013]: Using semi-supervised classifiers for credit scoring. *Journal of the Operational Research Society*. Vol. 64., p. 513-529, DOI: <http://dx.doi.org/10.1057/jors.2011.30>.

Kiang, M. Y. [2003]: A comparative assessment of classification methods. *Decision Support Systems*. Vol. 35., p. 441-454, DOI: [http://dx.doi.org/10.1016/S0167-9236\(02\)00110-0](http://dx.doi.org/10.1016/S0167-9236(02)00110-0).

Kim, G. – Wu, C. H. – Lim, S. – Kim, J. [2012]: Modified matrix splitting method for the support vector machine and its application to the credit classification of companies in Korea. *Expert Systems with Applications*. Vol. 39., p. 8824-8834, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.007>.

Kim, H. S. – Sohn, S. Y. [2010]: Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*. Vol. 201., p. 838-846, DOI: <http://dx.doi.org/10.1016/j.ejor.2009.03.036>.

Kim, M. J. – Kang, D. K. [2012]: Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications*. Vol. 39., p. 9308-9314, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.072>.

Kolari, J. – Glennon, D. – Shin, H. – Caputo, M. [2002]: Predicting large US commercial bank failures. *Journal of Economics and Business*. Vol. 54., No. 4., p. 361-387, DOI: [http://dx.doi.org/10.1016/S0148-6195\(02\)00089-9](http://dx.doi.org/10.1016/S0148-6195(02)00089-9).

Koopman, S. J. – Lucas, A. [2005]: Business and default cycles for credit risk. *Journal of Applied Econometrics*. Vol. 20., p. 311-323, DOI: <http://dx.doi.org/10.1002/jae.833>.

Koyuncugil, A. S. – Ozgulbas, N. [2012]: Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications*. Vol. 39., p. 6238-6253, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.12.021>.

Kristóf Tamás [2008]: *Gazdasági szervezetek fennmaradásának és fizetőképességének előrejelzése*. Ph. D. értekezés, Budapesti Corvinus Egyetem, Budapest.

Kristóf, T. – Virág, M. [2012]: Data reduction and univariate splitting – Do they together provide better corporate bankruptcy prediction? *Acta Oeconomica*. Vol. 62., No. 2., p. 205-227, DOI: <http://dx.doi.org/10.1556/AOecon.62.2012.2.4>.

Laitinen, E. K. – Laitinen, T. [2000]: Bankruptcy prediction. Application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis*. Vol. 9., No. 4., p. 327-349, DOI: [http://dx.doi.org/10.1016/S1057-5219\(00\)00039-9](http://dx.doi.org/10.1016/S1057-5219(00)00039-9).

Lee, M. C. – To, C. [2010]: Comparison of support vector machine and back propagation neural network in evaluating the enterprise financial distress. *International Journal of Artificial Intelligence & Applications*. Vol. 1., No. 3., p. 31-43, DOI: <http://dx.doi.org/10.5121/ijaia.2010.1303>.

Lee, S. – Choi, W. S. [2013]: A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*. Vol. 40., p. 2941-2946, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.12.009>.

Lensberg, T. – Eilifsen, A. – McKee, T. E. [2006]: Bankruptcy theory development and classification via genetic programming. *European Journal of Operational Research*. Vol. 169., p. 677-697, DOI: <http://dx.doi.org/10.1016/j.ejor.2004.06.013>.

Leshno, M. – Spector, Y. [1996]: Neural network prediction analysis: The bankruptcy case, *Neurocomputing*, Vol. 10, No. 2., p. 125-147, DOI: [http://dx.doi.org/10.1016/0925-2312\(94\)00060-3](http://dx.doi.org/10.1016/0925-2312(94)00060-3).

Li, H. – Sun, J. [2011a]: Principal component case-based reasoning ensemble for business failure prediction. *Information & Management*. Vol. 48., p. 220-227, DOI: <http://dx.doi.org/10.1016/j.im.2011.05.001>.

Li, H. – Sun, J. [2011b]: Predicting business failure using support vector machines with straightforward wrapper: A re-sampling study. *Expert Systems with Applications*. Vol. 38., p. 12747-12756, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.04.064>.

Liang, D. – Tsai, C. F. – Wu, H. T. [2015]: The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*. Vol. 73., p. 289-297, DOI: <http://dx.doi.org/10.1016/j.knosys.2014.10.010>.

- Lin, F. – Liang, D. – Chen, E. [2011]: Financial ratio selection for business crisis prediction. *Expert Systems with Applications*. Vol. 38., p. 15094-15102, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.05.035>.
- Lin, F. – Liang, D. – Yeh, C. C. – Huang, J. C. [2014]: Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*. Vol. 41., p. 2472-2483, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.09.047>.
- Lin, R. H. – Wang, Y. T. – Wu, C. H. – Chuang, C. L. [2009]: Developing a business failure prediction model via RST, GRA and CBR. *Expert Systems with Applications*. Vol. 36., p. 1593-1600, DOI: <http://dx.doi.org/10.1016/j.eswa.2007.11.068>.
- Lin, T. H. [2009]: A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models. *Neurocomputing*. Vol. 72., p. 3507-3516, DOI: <http://dx.doi.org/10.1016/j.neucom.2009.02.018>.
- Löffler, G. – Maurer, A. [2011]: Incorporating the dynamics of leverage into default prediction. *Journal of Banking & Finance*. Vol. 35., p. 3351-3361, DOI: <http://dx.doi.org/10.1016/j.jbankfin.2011.05.015>.
- Lyandres, E. – Zhdanov, A. [2013]: Investment opportunities and bankruptcy prediction. *Journal of Financial Markets*. Vol. 16., p. 439-476, DOI: <http://dx.doi.org/10.1016/j.finmar.2012.10.003>.
- Manjón-Antolín, M. C. – Arauzo-Carod, J. M. [2008]: Firm survival: methods and evidence. *Empirica*. Vol. 35., p. 1-24, DOI: <http://dx.doi.org/10.1007/s10663-007-9048-x>.
- Marqués, A. I. – García, V. – Sánchez, J. S. [2012a]: Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications*. Vol. 39., p. 10916-10922, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.03.033>.
- Marqués, A. I. – García, V. – Sánchez, J. S. [2012b]: Exploring the behaviour of base classifiers in credit scoring ensembles. *Expert Systems with Applications*. Vol. 39., p. 10244-10250, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.092>.
- Martens, D. – Van Gestel, T. – De Backer, M. – Haesen, R. – Vanthienen, J. – Baesens, B. [2010]: Credit rating prediction using ant colony optimization. *Journal of Operational Research Society*. Vol. 61., No. 4., p. 561-573, DOI: <http://dx.doi.org/10.1057/jors.2008.164>.
- Martin, D. [1977]: Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*. Vol. 1., No. 3., p. 249-276, DOI: [http://dx.doi.org/10.1016/0378-4266\(77\)90022-X](http://dx.doi.org/10.1016/0378-4266(77)90022-X).
- McKee, T. E. – Greenstein, M. [2000]: Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*. Vol. 19., No. 3., p. 219-230, DOI: [http://dx.doi.org/10.1002/\(SICI\)1099-131X\(200004\)19:3](http://dx.doi.org/10.1002/(SICI)1099-131X(200004)19:3).

- McKee, T. E. – Lensberg, T. [2002]: Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research*. Vol. 138., No. 2., p. 436-451, DOI: [http://dx.doi.org/10.1016/S0377-2217\(01\)00130-8](http://dx.doi.org/10.1016/S0377-2217(01)00130-8).
- McKee, T. E. [2003]: Rough sets bankruptcy prediction models versus auditor signalling rates. *Journal of Forecasting*. Vol. 22., p. 569-586, DOI: <http://dx.doi.org/10.1002/for.875>.
- McLeay, S. – Omar, A. [2000]: The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios. *British Accounting Review*. Vol. 32., p. 213-230, DOI: <http://dx.doi.org/10.1006/bare.1999.0120>.
- Min, J. H. – Jeong, C. – Kim, M. S. [2011]: Tuning the architecture of support vector machine: The case of bankruptcy prediction. *International Journal of Management Science*. Vol. 17., No. 1., p. 19-43.
- Min, J. H. – Lee, Y. C. [2005]: Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*. Vol. 28., p. 603-614, DOI: <http://dx.doi.org/10.1016/j.eswa.2004.12.008>.
- Moro, R. – Hardle, W. – Aliakbari, S. – Hoffmann, L. [2011]: *Forecasting corporate distress in the Asian and Pacific region*. Working Paper, No. 11-08, Economics and Finance Working Paper Series, 2011 May, Internet: <http://www.brunel.ac.uk/about/acad/ssss/depts/economics>
- Mossman, C. E. – Bell, G. G. – Swartz, L. M. – Turtle, H. [1998]: An empirical comparison of bankruptcy models. *The Financial Review*. Vol. 33., No. 2., p. 35-54, DOI: <http://dx.doi.org/10.1111/j.1540-6288.1998.tb01367.x>.
- Nam, C. W. – Kim, T. S. – Park, N. J. – Lee, H. K. [2008]: Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*. Vol. 27., p. 493-506, DOI: <http://dx.doi.org/10.1002/for.985>.
- Nam, J. H. – Jinn, T. [2000]: Bankruptcy prediction: Evidence from Korean listed companies during the IMF crisis. *Journal of International Financial Management and Accounting*. Vol. 11., No. 3., p. 178-197, DOI: <http://dx.doi.org/10.1111/1467-646X.00061>.
- Nanda, S. – Pendharkar, P. [2001]: Linear models for minimizing misclassification costs in bankruptcy prediction. *International Journal of Intelligent Systems in Accounting, Finance and Management*. Vol. 10., p. 155-168, DOI: <http://dx.doi.org/10.1002/isaf.203>.
- Neophytou, E. – Charitou, A. – Charalambous, C. [2000]: *Predicting Corporate Failure: Empirical Evidence for the UK*. Department of Accounting and Management Science, University of Southampton, Southampton

Neophytou, E. – Mar Molinero, C. [2005]: Financial ratios, size, industry and interest rate issues in company failure: An extended multidimensional scaling analysis. *Kent Business School Working Papers 10*. University of Kent, Kent.

Neves, J. C. – Vieira, A. [2006]: Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review*. Vol. 15., No. 2., p. 253-271. DOI: <http://dx.doi.org/10.1080/09638180600555016>.

Nikolic, N. – Zarkic-Joksimovic, N. – Stojanovski, D. – Joksimovic, I. [2013]: The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications*. Vol. 40., p. 5932-5944, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.05.022>.

Nyitrai Tamás [2014]: Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében. *Statiztikai Szemle*, 92. évfolyam, 4. szám, 357-377. o.

Odom, M. D. – Sharda, R. [1990]: A neural network model for bankruptcy prediction. In: *Proceeding of the International Joint Conference on Neural Networks, San Diego, 17-21 June 1990, Vol. II.*, IEEE Neural Networks Council, Ann Arbor, p. 163-171.

Ohlson, J. A. [1980]: Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*. Vol. 18., No. 1., p. 109-131, DOI: <http://dx.doi.org/10.2307/2490395>.

Olmeda, I. – Fernández, E. [1997]: Hybrid classifiers for financial multicriteria decision making: The case of bankruptcy prediction. *Computational Economics*. Vol. 10., No. 4., p. 317-335, DOI: <http://dx.doi.org/10.1023/A:1008668718837>.

Ooghe, H. – Claus, H. – Sierens, N. – Camerlynck, J. [1999]: *International comparison of failure prediction models from different countries: an empirical analysis*. Department of Corporate Finance, University of Ghent, Ghent.

Orbe, J. – Ferreira, E. – Núñez-Antón, V. [2001]: Modelling the duration of firms in Chapter 11 bankruptcy using a flexible model. *Economic Letters*. Vol. 71., No. 1., p. 35-42, DOI: [http://dx.doi.org/10.1016/S0165-1765\(01\)00358-5](http://dx.doi.org/10.1016/S0165-1765(01)00358-5).

Oreski, S. – Oreski, D. – Oreski, G. [2012]: Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*. Vol. 39., p. 12605-12617, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.05.023>.

Oreski, S. – Oreski, G. [2014]: Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*. Vol. 41., p. 2052-2064, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.09.004>.

Paleologo, G. – Elisseeff, A. – Antonini, G. [2010]: Subagging for credit scoring models. *European Journal of Operational Research*. Vol. 201., p. 490-499, DOI: <http://dx.doi.org/10.1016/j.ejor.2009.03.008>.

Pantalone, C. C. – Platt, M. B. [1987a]: Predicting failure of savings & loan associations. *Real Estate Economics*. Vol. 15., No. 2., p. 46-64, DOI: <http://dx.doi.org/10.1111/1540-6229.00418>.

Pantalone, C. C. – Platt, M. B. [1987b]: Predicting commercial bank failure since deregulation. *New England Economic Review*. 1987 July/August, p. 37-47.

Parnes, D. [2007]: Applying credit score models to multiple states of nature. *The Journal of Fixed Income*. Vol. 17., No. 3., p. 55-71, DOI: <http://dx.doi.org/10.3905/jfi.2007.700304>.

Pasaribu, R. B. F. [2008]: Financial distress prediction in Indonesia Stock Exchange. *Journal of Economics, Business and Accounting*. Vol. 11., No. 2., p. 153-172.

Pendharkar, P. C. [2011]: Probabilistic approaches for credit screening and bankruptcy prediction. *Intelligent Systems in Accounting, Finance and Management*. Vol. 18., p. 177-193, DOI: <http://dx.doi.org/10.1002/isaf.331>.

Peng, Y. – Kou, G. – Shi, Y. – Chen, Z. [2008]: A multi-criteria convex quadratic programming model for credit data analysis. *Decision Support Systems*. Vol. 44., p. 1016-1030, DOI: <http://dx.doi.org/10.1016/j.dss.2007.12.001>.

Philosophov, L. V. – Philosophov, V. L. [2002]: Corporate bankruptcy diagnosis: An attempt at a combined prediction of the bankruptcy event and time interval of its occurrence. *International Review of Financial Analysis*. Vol. 11., p. 375-406, DOI: [http://dx.doi.org/10.1016/S1057-5219\(02\)00081-9](http://dx.doi.org/10.1016/S1057-5219(02)00081-9).

Pindado, J. – Rodrigues, L. F. [2004]: Parsimonious models of financial insolvency in small companies. *Small Business Economics*. Vol. 22., No. 1., p. 51-66, DOI: <http://dx.doi.org/10.1023/B:SBEJ.0000011572.14143.be>.

Ping, Y. – Yongheng, L. [2011]: Neighborhood rough set and SVM based hybrid credit scoring classifier. *Expert Systems with Applications*. Vol. 38., p. 11300-11304, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.02.179>.

Platt, H. D. – Platt, M. B. [1990]: Development of a class of stable predictive variables: The case of bankruptcy prediction. *Journal of Business Finance & Accounting*. Vol. 17., No. 1., p. 31-51, DOI: <http://dx.doi.org/10.1111/j.1468-5957.1990.tb00548.x>.

Pompe, P. P. M. – Bilderbeek, J. [2005]: Bankruptcy prediction: the influence of the year prior to failure selected for model building and the effects in a period of economic decline. *Intelligent Systems in Accounting, Finance and Management*. Vol. 13., p. 95-112, DOI: <http://dx.doi.org/10.1002/isaf.259>.

Quek, C. – Zhou, R. W. – Lee, C. H. [2009]: A novel fuzzy neural approach to data reconstruction and failure prediction. *Intelligent Systems in Accounting, Finance and Management*. Vol. 16., p. 165-187, DOI: <http://dx.doi.org/10.1002/isaf.299>.

Riberio, B. – Silva, C. – Chen, N. – Vieira, A. – Das Neves, J. C. [2012]: Enhanced default risk models with SVM+. *Expert Systems with Applications*. Vol. 39., p. 10140-10152, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.02.142>.

Rösch, D. [2003]: Correlations and business cycles of credit risk: Evidence from bankruptcies in Germany. *Financial Markets and Portfolio Management*. Vol. 17., No. 3., DOI: <http://dx.doi.org/10.1007/s11408-003-0303-2>.

Ruzgar, N. S. – Unsal, F. – Ruzgar, B. [2008]: *Predicting bankruptcies using rough set approach: The case of Turkish banks*. American Conference on Applied Mathematics, Harvard, Massachusetts, USA, March 24-26, 2008. Internet: <http://www.naun.org/main/NAUN/ijmmas/mmmas-60.pdf>

Saberi, M. – Mirtalaie, M. S. – Hussain, F. K. – Azadeh, A. – Hussain, O. K. – Ashjari, B. [2013]: A granular computing-based approach to credit scoring modeling. *Neurocomputing*. Vol. 122., p. 100-115, DOI: <http://dx.doi.org/10.1016/j.neucom.2013.05.020>.

Salcedo-Sanz, S. – Deprado-Cumplido, M. – Segovia-Vargas, M. J. – Pérez-Cruz, F. – Bousoño-Calzón, C. [2004]: Feature selection methods involving support vector machines for prediction of insolvency in non-life insurance companies. *Intelligent Systems in Accounting, Finance and Management*. Vol. 12., p. 261-281, DOI: <http://dx.doi.org/10.1002/isaf.255>.

Salchenberger, L. – Cinar, E. – Lash, N. [1992]: Neural networks: A new tool for predicting thrift failures. *Decision Sciences*. Vol. 23., No. 4., p. 899-916, DOI: <http://dx.doi.org/10.1111/j.1540-5915.1992.tb00425.x>.

Sánchez-Lasheras, F. – De Andrés, J. – Lorca, P. – De Cos Juez, F. J. [2012]: A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications*. Vol. 39., p. 7512-7523, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.01.135>.

Sarkar, S. – Sriram, R. S. [2001]: Bayesian models for early warning of bank failures. *Management Science*. Vol. 47., No. 11., p. 1457-1475, DOI: <http://dx.doi.org/10.1287/mnsc.47.11.1457.10253>.

Segovia-Vargas, M. J. – Gil-Fana, J. A. – Heras-Martínez, A. – Vilar-Zanón, J. L. – Sanchis-Arellano, A. [2003]: *Using rough sets to predict insolvency of Spanish non-life insurance companies*. Documentos de trabajo de la Facultad de Ciencias Económicas y Empresariales 03-02, Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales, No. 03-02.

Serrano-Cinca, C. – Guitérez-Nieto, B. [2013]: Partial least square discriminant analysis for bankruptcy prediction. *Decision Support Systems*. Vol. 54., p. 1245-1255, DOI: <http://dx.doi.org/10.1016/j.dss.2012.11.015>.

- Shetty, U. – Pakkala, T. P. M. – Mallikarjunappa, T. [2012]: A modified formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India. *Expert Systems with Applications*. Vol. 39., p. 1988-1997. DOI: <http://dx.doi.org/10.1016/j.eswa.2011.08.043>.
- Shin, K. S. – Lee, T. S. – Kim, H. J. [2005]: An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*. Vol. 28., p. 127-135, DOI: <http://dx.doi.org/10.1016/j.eswa.2004.08.009>.
- Sinha, A. P. – Zhao, H. [2008]: Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems*. Vol. 46., p. 287-299, DOI: <http://dx.doi.org/10.1016/j.dss.2008.06.013>.
- Sjovoll, E. [1999]: *Assessment of credit risk in the Norwegian business sector*. Norges Bank. Internet: <http://www.norges-bank.no/Upload/import/publikasjoner/arbeidsnotater/pdf/arb-1999-09.pdf>
- Slowinski, R. – Zopounidis, C. [1995]: Application of the rough set approach to evaluation of bankruptcy risk. *Intelligent Systems in Accounting, Finance and Management*. Vol. 4., No. 1., p. 27-41, DOI: <http://dx.doi.org/10.1002/j.1099-1174.1995.tb00078.x>.
- Sohn, S. Y. – Kim, H. S. [2007]: Random effects logistic regression model for default prediction of technology credit guarantee fund. *European Journal of Operational Research*. Vol. 183., p. 472-478, DOI: <http://dx.doi.org/10.1016/j.ejor.2006.10.006>.
- Sueyoshi, T. – Goto, M. [2009]: Methodological comparison between DEA (data envelopment analysis) and DEA-DA (discriminant analysis) from the perspective of bankruptcy assessment. *European Journal of Operational Research*. Vol. 199., p. 561-575, DOI: <http://dx.doi.org/10.1016/j.ejor.2008.11.030>.
- Sun, L. – Shenoy, P. P. [2007]: Using Bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research*. Vol. 180., p. 738-753, DOI: <http://dx.doi.org/10.1016/j.ejor.2006.04.019>.
- Sung, T. K. – Chang, N. – Lee, G. [1999]: Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*. Vol. 16., No. 1., p. 63-85.
- Swiderski, B. – Kurek, J. – Osowski, S. [2012]: Multistage classification by using logistic regression and neural networks for assessment of financial condition of company. *Decision Support Systems*. Vol. 52., p. 539-547, DOI: <http://dx.doi.org/10.1016/j.dss.2011.10.018>.
- Tam, K. Y. – Kiang, M. Y. [1992]: Managerial applications of neural networks: The Case of Bank Failure Predictions. *Management Science*. Vol. 38., No. 7., p. 926-947, DOI: <http://dx.doi.org/10.1287/mnsc.38.7.926>.

Tan, C. N. W. [1999]: *An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System*. School of Information Technology, Bond University, Australia

Telmoudi, F. – El Ghourabi, M. – Limam, M. [2011]: RST-GCBR-clustering-based RGA-SVM model for corporate failure prediction. *Intelligent Systems in Accounting, Finance & Management*. Vol. 18., p. 105-120, DOI: <http://dx.doi.org/10.1002/isaf.323>.

Trinkle, B. S. – Baldwin, A. A. [2007]: Interpretable credit model development via artificial neural networks. *Intelligent Systems in Accounting, Finance and Management*. Vol. 15., p. 123-147, DOI: <http://dx.doi.org/10.1002/isaf.289>.

Twala, B. [2010]: Multiple classifier application to credit risk assessment. *Expert Systems with Applications*. Vol. 37., p. 3326-3336, DOI: <http://dx.doi.org/10.1016/j.eswa.2009.10.018>.

Van Gestel, T. – Baesens, B. – Van Dijcke, P. – Garcia, J. – Suykens, J. A. K. – Vanthienen, J. [2006]: A process model to develop an internal rating system: Sovereign credit ratings. *Decision Support Systems*. Vol. 42., p. 1131-1151, DOI: <http://dx.doi.org/10.1016/j.dss.2005.10.001>.

Virág M. – Nyitrai T. [2014]: Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model, *Acta Oeconomica*, Vol. 64, No. 4., p. 419-440, DOI: <http://dx.doi.org/10.1556/AOecon.64.2014.4.2>.

Virág Miklós – Hajdu Ottó [1996]: Pénzügyi mutatószámokon alapuló csődmodell-számítások, *Bankszemle*, 15. évfolyam, 5. szám, 42-53. o.

Virág Miklós – Kristóf Tamás – Fiáth Attila – Varsányi Judit [2013]: *Pénzügyi elemzés, csődelőrejelzés, válságkezelés*. Kossuth Kiadó, Budapest.

Virág Miklós – Kristóf Tamás [2005]: Az első hazai csődmodell újraszámítása neurális hálók segítségével, *Közgazdasági Szemle*, 52. évfolyam, 2. szám, 144-162. o.

Virág Miklós – Kristóf Tamás [2006]: Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel, *Vezetéstudomány*, 37. évfolyam, 1. szám, 25-35. o.

Virág Miklós – Kristóf Tamás [2009]: Többdimenziós skálázás a csődelőrejelzésben. *Vezetéstudomány*. 40. évfolyam, 1. szám, 50-58. o.

Virág Miklós [1993]: *Pénzügyi viszonyszámokon alapuló teljesítmény-megítélés és csődelőrejelzés*. Kandidátusi értekezés, Budapest.

Virág Miklós [2004]: A csődmodellek jellegzetességei és története. *Vezetéstudomány*. 35. évfolyam, 10. szám, 24-32. o.

- Virág, M. – Nyitrai, T. [2013]: Application of support vector machines on the basis of the first Hungarian bankruptcy model. *Society and Economy*. Vol. 35., No. 2., p. 227-248, DOI: <http://dx.doi.org/10.1556/SocEc.35.2013.2.6>.
- Vukovic, S. – Delibasic, B. – Uzelac, A. – Suknovic, M. [2012]: A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*. Vol. 39., p. 8389-8395, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.01.181>.
- Wang, G. – Ma, J. – Yang, S. [2014]: An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*. Vol. 41., p. 2353-2361, DOI: <http://dx.doi.org/10.1016/j.eswa.2013.09.033>.
- Wang, G. – Ma, J. [2011]: Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Systems with Applications*. Vol. 38., p. 13871-13878, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.04.191>.
- Wang, G. – Ma, J. [2012]: A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. *Expert Systems with Applications*. Vol. 39., p. 5325-5331, DOI: <http://dx.doi.org/10.1016/j.eswa.2011.11.003>.
- Wang, Z. [2004]: Financial ratio selection for default-rating modeling: A model-free approach and its empirical performance. *Journal of Applied Finance*. Spring/Summer, p. 20-35.
- Wilson, R. L. – Sharda, R. [1994]: Bankruptcy prediction using neural networks. *Decision Support Systems*. Vol. 11., No. 5., p. 545-557, DOI: [http://dx.doi.org/10.1016/0167-9236\(94\)90024-8](http://dx.doi.org/10.1016/0167-9236(94)90024-8).
- Xiaosi, X. – Ying, C. – Haitao, Z. [2011]: The comparison of enterprise bankruptcy forecasting method. *Journal of Applied Statistics*. Vol. 38., No. 2., p. 301-308, DOI: <http://dx.doi.org/10.1080/02664760903406470>.
- Yang, R. Z. [2001]: *A new method for company failure prediction using probabilistic neural network*. Department of Computer Science, Exeter University.
- Yang, Y. [2007]: Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research*. Vol. 183., p.1521-1536, DOI: <http://dx.doi.org/10.1016/j.ejor.2006.10.066>.
- Yazici, M. [2011]: Combination of discriminant analysis and artificial neural networks in the analysis of credit card customers. *European Journal of Finance and Banking Research*. Vol. 4., No. 4., p. 1-10.
- Yeh, C. C. – Chi, D. J. – Hsu, M. F. [2010]: A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*. Vol. 37., p. 1535-1541, DOI: <http://dx.doi.org/10.1016/j.eswa.2009.06.088>.

Yim, J. – Mitchell, H. [2005]: A Comparison of Corporate Distress Prediction Models in Brazil: Hybrid Neural Networks, Logit Models and Discriminant Analysis, *Nova Economia Belo Horizonte*, Vol. 15, No. 1., p. 73-93.

Yoon, J. – Kwon, Y. S. – Lee, C. H. [2008]: Bankruptcy prediction for small businesses using credit card sales information: Comparison of classification performance. *Proceedings of the 9th Asia Pacific Industrial Engineering & Management Systems Conference*. Nusa Dua, Bali Indonesia, December 3-5, 2008

Yu, L. – Yao, X. – Wang, S. – Lai, K. K. [2011]: Credit risk evaluation using weighted least squares SVM classifier with design of experiment for parameter selection. *Expert Systems with Applications*. Vol. 38., p. 15392-15399, DOI: <http://dx.doi.org/10.1016/j.dss.2007.12.001>.

Yu, Q. – Miche, Y. – Séverin, E. – Lendasse, A. [2014]: Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing*. Vol. 128., p. 296-302, DOI: <http://dx.doi.org/10.1016/j.neucom.2013.01.063>.

Zavgren, C. V. [1985]: Assessing the vulnerability to failure of American industrial firms: A logistic analysis. *Journal of Business Finance & Accounting*. Vol. 12., No. 1., p. 19-45, DOI: <http://dx.doi.org/10.1111/j.1468-5957.1985.tb00077.x>.

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., No. 1., p. 16-32, DOI: [http://dx.doi.org/10.1016/S0377 2217\(98\)00051-4](http://dx.doi.org/10.1016/S0377 2217(98)00051-4).

Zhong, M. – Miao, C. – Shen, Z. – Feng, Y. [2014]: Comparing the learning effectiveness of BP, ELM, I-ELM and SVM for corporate credit ratings. *Neurocomputing*. Vol. 128., p. 285-295, DOI: <http://dx.doi.org/10.1016/j.dss.2005.10.001>.

Zmijewski, M. E. [1984]: Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*. Vol. 22. Supplement, p. 59-82. DOI: <http://dx.doi.org/10.2307/2490859>.

A TÉMÁBAN MEGJELENT SAJÁT PUBLIKÁCIÓK

Nyitrai Tamás [2014]: Növelhető-e a csőd-előrejelző modellek előrejelző képessége az új klasszifikációs módszerek nélkül? *Közgazdasági Szemle*, LXI. évfolyam, május, 566-585. o.

Nyitrai Tamás [2014]: Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében. *Statisztikai Szemle*, 92. évfolyam, 4. szám, 357-377. o.

Nyitrai Tamás [2015]: Hazai vállalkozások csődjének előrejelzése a csődeseményt megelőző egy, két, illetve három évvel korábbi pénzügyi beszámolók adatai alapján. *Vezetéstudomány*, 46. évfolyam, 5. szám, 55-65. o.

Virág M. – Nyitrai T. [2014]: Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model, *Acta Oeconomica*, Vol. 64, No. 4., p. 419-440, DOI: <http://dx.doi.org/10.1556/AOecon.64.2014.4.2>.

Virág Miklós – Nyitrai Tamás [2014]: Metamódszerek alkalmazása a csődelőrejelzésben. *Hitelintézeti Szemle*, XIII. évfolyam, 4. szám, 180-195. o.

Virág Miklós – Nyitrai Tamás [2015]: Csődelőrejelző modellek dinamizálása – a szakértői tudás megjelenítése a csődelőrejelzésben. *Vezetés és szervezet társadalmi kontextusban. Tanulmányok Dobák Miklós 60. születésnapja tiszteletére*. Szerkesztette: Bakacsi Gyula – Balaton Károly. Akadémiai Kiadó, Budapest, ISBN 978 963 05 9634 3, 284-304. o.

Virág, M. – Nyitrai, T. [2013]: Application of support vector machines on the basis of the first Hungarian bankruptcy model. *Society and Economy*. Vol. 35., No. 2., p. 227-248, DOI: <http://dx.doi.org/10.1556/SocEc.35.2013.2.6>.