

Közzététel: 2019. július 10.

A tanulmány címe:

CHAID-alapú felülvizsgált kategorizálás a csődelőrejelzésben

Szerző:

Nyitrai Tamás, a Budapesti Corvinus Egyetem egyetemi adjunktusa, E-mail: tamas.nyitrai@uni-corvinus.hu

DOI: <https://doi.org/10.20311/stat2019.7.hu0656>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c.) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:

„*Forrás: Statisztikai Szemle c. folyóirat 97. évfolyam 7. számában megjelent, Nyitrai Tamás által írt, 'CHAID-alapú felülvizsgált kategorizálás a csődelőrejelzésben' című tanulmány (link csatolása)*”

7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem esnek szükségképpen egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

CHAID-alapú felülvizsgált kategorizálás a csődelőrejelzésben

Nyitrai Tamás,

a Budapesti Corvinus Egyetem
egyetemi adjunktusa

E-mail: tamas.nyitrai@uni-corvinus.hu

A szerző CHAID- (chi-squared automatic interaction detector – khi-négyzet-alapú automatikus interakció-detektálás) módszeren alapuló, több lépésből álló adatbányászati megközelítést mutat be a csődelőrejelzés területén. Az eljárás során először a megfigyelt lehetséges magyarázó változók értékészletét kategorizálja, majd azt vizsgálja, hogy e kategóriákon belül hogyan alakultak a független változók értékei az idő múlásával. A tanulmány újdonságértéke abban rejlik, hogy az időben egymást követő adatok kategóriáinak sorszámaiból új változót képez, amely a fizetés-képtelenség bekövetkezésétől időben távolodva csökkenő súllyal veszi figyelembe az egyes magyarázó változók értékeit, majd az így kapott számsorokat használja fel az eredeti adatok helyett. Az empirikus vizsgálatok tapasztalatai arra utalnak, hogy ez az újszerű megközelítés magasabb teljesítményű csődelőrejelző modellek felállítását teszi lehetővé. A feltételezést a szerző robusztussági vizsgálatoknak is aláveti, amelyek megerősítik a rendelkezésre álló adatok tekintetében a tanulmány fő következtetését. A módszer bármely olyan klasszifikációs probléma kapcsán alkalmazható, ahol a magyarázó változók időbeli folyamatának magyarázó ereje lehet, de a korábbi értékeknek önmagukban az idő múlásával egyre kevésbé.

TÁRGYSZÓ:
Csődelőrejelzés.
CHAID.

DOI: 10.20311/stat2019.7.hu0656

A vállalatok jövőbeli fennmaradásának és fizetőképességének előrejelzése több tudományterületen is az aktuális kutatási kérdések körébe tartozik. Az érdeklődés forrása kezdetben alapvetően gazdasági eredetű volt: a hitelezők az esetleges jövőbeli nemfizetés kockázatát szeretnék volna minimalizálni. Tekintettel azonban arra, hogy a jelenbeli, illetve potenciális adósok esetén ezt egyidejűleg több magyarázó változó is indokolhatja, a többváltozós statisztikai módszerek alkalmazása megkerülhetetlen ezen a területen. Ezek a sajátosságok pedig maguk után vonták, hogy a témakör több olyan kapcsolódó diszciplína esetén is középpontba került, mint az információtechnológia, operációkutatás vagy az adatbányászat.

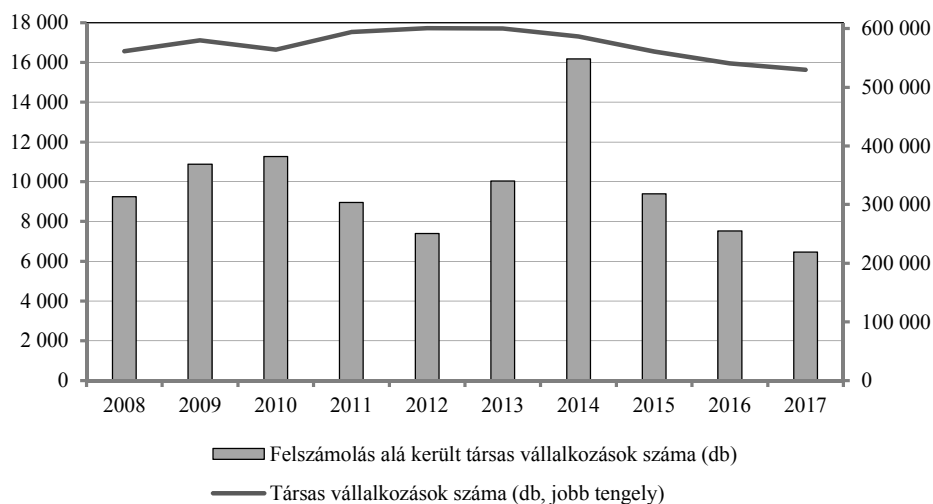
A csődelőrejelzés célja első ránézésre egyszerűnek tűnhet: olyan többváltozós függvény vagy szabályrendszer megalkotása, amely lehetővé teszi, hogy a korábbi tapasztalatok alapján egy jelen vagy jövőbeli adóst besoroljunk (klasszifikáljunk) a jó, azaz várhatóan fizetőképességű; illetve a rossz (várhatóan fizetéseképtelen) kategóriába. Mindez két szempontból vizsgálható: a megfigyelés tárgya lehet magánszemély vagy szervezet (gazdasági társaság, önkormányzat vagy akár ország). Mindkét esetben hatalmas mennyiségű nemzetközi szakirodalom áll rendelkezésre, melynek tételes áttekintésére méretéből adódóan nem vállalkozhat ez a tanulmány. A témakörben született tudományos kutatások számosságát jelzi, hogy az utóbbi időben évente jelennek meg a nemzetközi szakirodalom aktuális eredményeit összegző, áttekintő cikkek. A legfrissebbek közül ebbe a körbe tartozik például *Onay–Öztürk* [2018] írása. A bemutatott folyamatok nem hagyták érintetlenül a hazai szakirodalmat sem, mivel a *Statisztikai Szemlé*ben is növekvő számban jelennek meg olyan tanulmányok, amelyek a magyar vállalkozások jövőbeli nemfizetésének előre jelezhetőségét tárgyalják különböző szempontok szerint.

A témakör iránti növekvő tudományos érdeklődés egyik legfőbb tényezőjének – a hazai és a nemzetközi szakirodalom esetén is – a 2008-ban kezdődött világgazdasági válság következtében egyre gyakoribbá váló vállalati fizetéseképtelenség tekinthető. A KSH (Központi Statisztikai Hivatal) által gondozott „Statisztikai tükör” című kiadvány adatai szerint a működő, illetve a felszámolási eljárás alá kerülő társas vállalkozások száma Magyarországon a következők szerint alakult.

Az 1. ábrán látható, hogy 2014-ben kiugróan magas volt a felszámolás alá kerülő vállalkozások száma az azt követő és az azt megelőző évekhez viszonyítva egyaránt. Ugyan az ábrán bemutatott időszak utolsó éveit határozott csökkenés jellemzi, a vállalatok jövőbeli nemfizetését előre jelezni képes modellek fejlesztése vélhetően továbbra is fontos tudományos és gyakorlati feladat marad, mivel egyre több hírben merül fel egy esetleges újabb recesszió bekövetkezése a közeljövőben. Ebben az esetben a csődelőrejelző modellek kiemelt jelentőségűek, ugyanis *Ribeiro et al.* [2012] véleménye

szerint a legutóbbi válság kialakulásában jelentős szerepe volt annak, hogy a döntéshozók a partnerek jövőbeli nemfizetésének kockázatát nem mérték fel megfelelően.

1. ábra. A felszámolási eljárás alá került és a működő társas vállalkozások száma Magyarországon



Forrás: Itt és a 2. ábra esetén KSH [2009–2017]: *Statisztikai Tükör*. Budapest.

A magyar és a nemzetközi szakirodalomnak egyaránt sajátossága, hogy egy-egy tanulmány időnként több kérdést vet fel, mint amennyit megválaszolni képes. Ennek oka a témakör empirikus jellegében rejlik. Mivel az adósok jövőbeli nemfizetési kockázatának előrejelzésére vonatkozóan nem áll rendelkezésre objektív közgazdasági elmélet, így a témakör kutatói a múlt tapasztalataira építve keresik azokat a lehetséges változókat és többváltozós módszereket, amelyeket a lehető legnagyobb hatásfokkal lehet a jövőben alkalmazni.

Az utóbbi évek szakirodalmi tendenciái azonban egyre inkább arra utalnak, hogy módszertani oldalról már nem érhető el számottevő előrelépés a témakörben: az újabb klasszifikációs eljárások sok esetben már nem képesek jelentősen növelni a modellek teljesítményét a már jól ismert és széles körben alkalmazott módszerekhez viszonyítva; illetve többlet csak rendkívül nagy számításigénnyel vagy a modellek értelmezhetőségének kárára valósítható meg. Egyre több kritikai észrevétel¹ is meg-

¹ Például *Sariev–Germano* [2018] ennek kapcsán arra hívja fel a figyelmet, hogy az adatok modellezésre történő előkészítésének módja (például a kiugró értékek kezelése kapcsán) befolyásolhatja a különböző módszerekkel elérhető eredményt. *Jeong–Min–Kim* [2012] véleménye szerint a neurális hálók teljesítménye jellemzően alábecsült más módszerekhez képest, mivel az előbbieket teljesítményét döntően befolyásolja, hogy a kutató milyen és mennyi erőfeszítést tesz a legjobb teljesítményt lehetővé tevő paraméterkombináció azonosítása érdekében. Mivel az adatelőkészítési, illetve paraméterbeállítási feladatokra vonatkozóan nincs egységes keretrendszer, így azok önkényes kutatói megválasztása akár jelentősen is torzíthatja a módszertani összehasonlító elemzések következtetéseit.

jelenik a nemzetközi szakirodalomban arra vonatkozóan, hogy az egyes klasszifikációs eljárások teljesítményét összehasonlító tanulmányok nem nyújtanak objektív képet. Konszenzus kezd kialakulni abban a tekintetben, hogy általánosságban nem létezik olyan módszer, amelynek teljesítménye egyértelműen felülmúlná a más eljárásokkal elérhető eredményt. Ebből azonban még nem következik az, hogy a témában végzett kutatásoknak ne lehetne érdemi hozzáadott értéke. Egyrészt számottevő többlet nyerhető a már ismert módszerek innovatív felhasználásából, illetve kombinálásából; másrészt a rendelkezésre álló információk körének bővítéséből, valamint újszerű alkalmazásából.

Az utóbbi körbe tartoznak a modellek statikus jellegének feloldására tett kísérletek, amelyre az utóbbi években számos próbálkozás történt a témakör hazai és nemzetközi szakirodalmában egyaránt. Ezek közös vonása, hogy a módszertani összehasonlító elemzésekkel ellentétben a dinamizálási megközelítések eddig a legtöbb esetben eredményesnek bizonyultak.² Ez a tanulmány is ebbe a körbe sorolható. Célunk, hogy az utóbbi években megjelent hazai vonatkozású számítások eredményeit innovatív módon hasznosítsuk egy olyan több lépésből álló adatbányászati eljárás keretei között, amely lehetővé teszi a magyarázó változók időbeli folyamatának komplex figyelembevételét. Mindez azért jelent érdemi előrelépést a témában eddig megjelent kutatásokhoz képest, mert azok nem tekinthetők szó szerinti értelemben dinamikus modelleknek. Ez a tanulmány az eddigi kutatások eredményeinek integratív alkalmazásával ezt a hiányt szeretné pótolni.

Az első fejezetben az előbbi gondolatokat a hazai és nemzetközi szakirodalom aktuális tapasztalatai alapján kívánjuk alátámasztani, kitérve a jelen tanulmány alapját képező kutatási eredményekre. Az empirikus vizsgálathoz felhasznált adatbázis sajátosságait ismerteti a második fejezet. Ezt követően részletes leírást adunk a dinamikus modellezést lehetővé tevő módszertani keretről. A negyedik fejezetben a számítási eredményeket, az elvégzett robusztussági vizsgálatok tapasztalatait, illetve az azok alapján levont következtetéseket ismerheti meg az Olvasó. A tanulmány záró fejezete kitér a bemutatott empirikus vizsgálat korlátjaira, illetve a jövőbeli kutatási lehetőségek ismertetésére is.

1. Szakirodalmi háttér

A vállalatok jövőbeli fizetőképességének előrejelzése az 1960-as évek közepe óta áll a tudományos érdeklődés középpontjában. Az azóta eltelt több mint öt évtizedben

² Megjegyzendő ugyanakkor: a tudományos szakirodalom sajátossága, hogy túlnyomórészt olyan tanulmányok jelennek meg, amelyek valamely koncepció eredményességét kívánják igazolni. Az újszerű megközelítések esetleges „kudarcai” általában nem publikálnak a nemzetközi szakirodalomban.

olyan jelentős méretű szakirodalmat generált, melynek áttekintése túlmutat egy-egy önálló tanulmány keretein. Ezt igazolja, hogy az utóbbi években számos próbálkozás történt a nemzetközi publikációk³ különböző szempontok szerint történő összefoglalására, illetve azok alapján jövőbeli kutatási irányok kijelölésére. Ez a tanulmány az utóbbi tekintetben épít az idézett összefoglaló művekre, ugyanis azok között több olyan található, amely a csődelőrejelző modellek dinamizálását ígéretes jövőbeli kutatási irányként jelöli meg.

A folyamatszempléletű csődelőrejelzés fontossága nemcsak abból eredeztethető, hogy gyakorlati szempontból célszerűbb a vállalatok adatait korábbi időszakok megfelelő értékeinek tükrében megítélni, hanem abból is, hogy az utóbbi években több olyan tanulmány is napvilágot látott, amely arra utal: a pusztán módszertani oldalról végzett fejlesztési kísérletekben rejlő kutatási lehetőségek végesek. *De Andrés et al.* [2012] például a csupán módszertani oldalról tett javítási kísérletek kutatási irányát szuboptimálisnak tekintik, mert a többváltozós statisztikai eszközök közé tartozó logisztikus regresszió segítségével magasabb teljesítményű modellt tudtak felállítani, mint a mesterséges intelligencia körébe tartozó neurális hálókkal. *Sariev-Germano* [2018] arra hívják fel a figyelmet, hogy az egyes módszerek teljesítménye jelentősen függhet az adatok modellezésre történő előkészítésének módjától (kiugró értékek szűrése, hiányzó adatok kezelése, változószelekció), viszont ezt sok esetben figyelmen kívül hagyják a módszertani összehasonlító elemzésekben, ami felvetheti, hogy ezek a vizsgálatok esetleg nem adnak objektív és teljes képet a módszerek klasszifikációs teljesítményére vonatkozóan.

A vállalatok jövőbeli fizetéseképtelenségének előrejelzésével foglalkozó tanulmányok közös vonása a statikus jelleg, amely azt jelenti, hogy a modellek magyarázó változói között csak egyetlen – jellemzően a legutoljára megfigyelt – év adatait használják fel; figyelmen kívül hagyva az adatok időbeli tendenciájában rejlő esetleges további információkat. Ezt kezdetben az adatok egyik időszakról a másikra történő változásának modellbe építésével oldották meg; az utóbbi években azonban komplexebb módszertani megoldások kerültek a figyelem középpontjába. *Ciampi et al.* [2018] olasz vállalkozások pénzügyi folyamatait képezték le önszerveződő térképek segítségével, majd a jellegzetes csődfolyamatok alapján minősítve a megfigyeléseket arra a következtetésre jutottak, hogy a dinamikus megközelítéssel készített modellekkel jobb előrejelző teljesítmény adódik, mint a statikus magyarázó változók alapján. *Volkov et al.* [2017] Markov-láncok segítségével sűrítették egy mutatóba az adatok dinamikájából kinyerhető információkat, és azokat használták fel a statikus adatok mellett.

A most idézett két példa mellett számos további kutatás számolt be ígéretes eredményekről, különböző dinamizálási koncepciók alkalmazása kapcsán. Ezek

³ A teljesség igénye nélkül például: *Alaka et al.* [2018], *Habib et al.* [2018], *Chen–Ribeiro–Chen* [2016], *Kirkos* [2015], *Sun et al.* [2014].

közös sajátossága, hogy általában a mesterséges intelligencia rendkívül számításigényes módszereire épülnek; továbbá, hogy az így kapott modellek nem vagy csak nehezen értelmezhetők, ami a gyakorlati alkalmazás szempontjából képez jelentős akadályt. Jelen tanulmány koncepciója teljesen transzparens módon és viszonylag kis számítási kapacitás megkövetelése mellett teszi lehetővé a magyarázó változók időbeli folyamatának figyelembevételét.

A vonatkozó hazai szakirodalom mérete jóval kisebb és könnyebben áttekinthető. *Virág et al.* [2013] könyvükben külön alfejezetet szenteltek a magyar vonatkozások áttekintésének a kezdetektől a 2010-es évek kezdetéig. Összefoglalva megállapítható, hogy ezt az időszakot a hazai kutatás tekintetében is a módszertani összehasonlító elemzések dominálták. Ennek egyik jellegzetes példáját *Kristóf* [2005] tanulmánya jelenti, ahol a szerző azon tapasztalatát osztotta meg, hogy a mesterséges intelligencia eszközei közé sorolható neurális hálókkal magasabb előrejelző teljesítményű modellek állíthatók fel, mint a többváltozós statisztikai módszerek családjába tartozó diszkriminanciaanalízis vagy a logisztikus regresszió alkalmazásával az első hazai csődmodell alapját képező adathalmaz esetén.

A 2000-es évek második évtizedében a hazai szakirodalmat három tendencia jellemezte:

- a nemzetközi trendet követve magyar vállalati adatokon is további módszertani összehasonlító elemzések⁴ jelentek meg;
- kísérletek történtek a modellek teljesítményének javítására nem módszertani oldalról;⁵
- a magyar adatokon végzett számítások eredményei nemzetközi folyóiratokban⁶ láttak napvilágot; összekapcsolva ezzel a téma hazai és nemzetközi empirikus vizsgálatait.

A *Statisztikai Szemle* Olvasói legutóbb *Kristóf* [2018] munkájában olvashattak egy hazai vállalatok adatain végzett módszertani összehasonlító elemzésről, melynek célja a mesterséges intelligencia körébe tartozó esetalapú következtetés (case-based reasoning) módszerével felállított csődmodell klasszifikációs teljesítményének összevetése más eljárásokkal készített modellekével. Az idézett mű aktuális példát szolgáltat azon nemzetközi tendenciára vonatkozóan, amely szerint nem egyértelmű a mesterséges intelligencia eszköztárának fölénye a „hagyományosnak” tekinthető

⁴ Lásd például *Kristóf* [2018] munkáját az esetalapú következtetés módszerével kapcsolatban végzett összehasonlító elemzésről.

⁵ A döntési fákkal történő adatelőkészítés lehetőségére tudomásunk szerint először *Kristóf–Virág* [2012] hívták fel a figyelmet.

⁶ A kézirat benyújtásáig legutóbb *Nyitrai–Virág* [2018] publikált referált nemzetközi folyóiratban magyar vonatkozásban a csődelőrejelzés területén kutatási eredményeket.

többváltozós módszerekkel szemben, ugyanis *Kristóf* [2018] az esetalapú következtetés módszerével készített modelljének előrejelző képessége elmaradt a logisztikus regresszióhoz viszonyítva.

Ugyan az imént idézett szerző a legjobb eredményt a neurális hálók módszerével kapta, napjainkban számos kutató kérdőjelezi meg utóbbiak fölényét is. *Nyitrai–Virág* [2018] magyar és lengyel vállalkozások adatait is felhasználták összehasonlító elemzésükben annak igazolására, hogy megfelelő adatelőkészítés mellett a mesterséges intelligencia körébe tartozó neurális hálókhoz képest nagyobb előrejelző képességgel rendelkező modell állítható fel lineáris diszkriminanciaanalízis segítségével. A módszertani összevetéseken túlmenően *Kristóf–Virág* [2012] mutatott rá először magyar vonatkozású vállalatok jövőbeli fizetőképességének előrejelzése kapcsán a módszerek kombinációjában rejlő lehetőségekre. A lehetséges magyarázó változók mindegyikére egyváltozós döntési fákat állítottak fel, és az egyes fák által adott előrejelzéseket használták fel a későbbi modellekben magyarázó változóként. Az idézett szerzők a CHAID-módszerrel képzett egyváltozós döntési fák előrejelzéseire épített logisztikus regresszió esetén kapták a legjobb teljesítményt. Ez további példát szolgáltat arra, hogy megfelelő adatelőkészítés mellett a hagyományos módszertani keretek is lehetőséget adnak magas előrejelző teljesítményű modellek felállítására.

A csódmmodellek statikus jellegének feloldására a hazai szakirodalomban először *Nyitrai* [2014] tett kísérletet egy olyan mutató konstrukciója révén, amelynek célja, hogy az egyes magyarázó változók értékeit az adott megfigyelés időbeli szóródási terjedelméhez viszonyítva fejezze ki. Ugyan az idézett szerző a javasolt mutató alkalmazásával növelni tudta a modellek előrejelző képességét, valamint a későbbiekben további tanulmányok keretei között újabb robusztussági vizsgálatok eredményeivel kívánta demonstrálni a javasolt mutató gyakorlati hasznosságát, mégsem beszélhetünk ebben az esetben szó szerinti értelemben vett dinamizálásról. A *Nyitrai* [2014] által ajánlott mutatószám a dinamikának csak egyetlen pontját veszi figyelembe: a mutatószámok értékeit az egyes megfigyelések historikus minimumának és maximumának terjedelméhez viszonyítva fejezi ki; valódi folyamatelemzésről ebben az esetben nem beszélhetünk. E tanulmány ezt a hiányt kívánja pótolni azon koncepció alkalmazásával, melynek alapjait a hazai szakirodalomban az előzők mellett a következő kutatások képezik.

A metódus központi eszközét a CHAID-alapú döntési fa jelenti, amely *Hámori* [2001] munkáját követően vált népszerűvé a hazai szakirodalomban. Megjegyzendő, hogy a módszert viszonylag ritkán alkalmazzák a nemzetközi összehasonlító elemzésekben, melynek oka, hogy önálló klasszifikációs algoritmusként jellemzően gyengébb teljesítményt mutat, mint más módszerek. Ezt tapasztalták a nemzetközi szakirodalomban például *Serrano-Cinca–Gutiérrez-Nieto–Bernate-*

Valbuena [2018], de hasonló eredményeket közölt magyar vonatkozású elemzésében *Kristóf* [2018] is. *Hámori* [2016] hívta fel először a figyelmet arra, hogy a CHAID-módszerrel lehetőség van a folytonos magyarázó változók objektív módon történő kategorizálására. Az idézett szerző az egyes osztályközökbe kerülés tényét dummy változókkal kódolta, és a tőkeáttétel nevű pénzügyi mutató példáján demonstrálta: annak ellenére, hogy a mutató folytonos változóként nem volt szignifikáns a vállalatok fizetőképességének előrejelzése céljából felállított logisztikus regresszióra épülő modellben, addig a CHAID-módszer alapján generált kategóriákon alapuló dummy változók igen.

Jelen tanulmány *Hámori* [2016] megközelítésétől annyiban tér el, hogy a magyarázó változók értékkészletében képzett kategóriákat nem dummy változóként, hanem ordinális skálán mért adatként veszi figyelembe. Ennek oka, hogy ha a lehetséges magyarázó változók és az azok értékkészletén belül a CHAID-módszerrel képzett kategóriák száma nagy, akkor rendkívül sok dummy változó jönne létre, amely még nagyméretű adathalmaz esetén is számítási nehézséget okozhat. Ha az eredeti magyarázó változók helyett az azokból képzett kategóriák sorszámaikat ordinális változóként vesszük figyelembe, akkor a lehetséges magyarázó változók száma változatlan marad. A következő fejezetben látható lesz, hogy ebben az esetben problémát okozhat, ha a célváltozó (a fizetéseképtelenség bekövetkezése) és a magyarázó változók közötti kapcsolat nem monoton. Ezt a nehézséget a kategóriák átrendezésével kezeli az a koncepció, amelyet a 3. fejezetben mutatunk be.

2. Az empirikus vizsgálatához felhasznált adatok

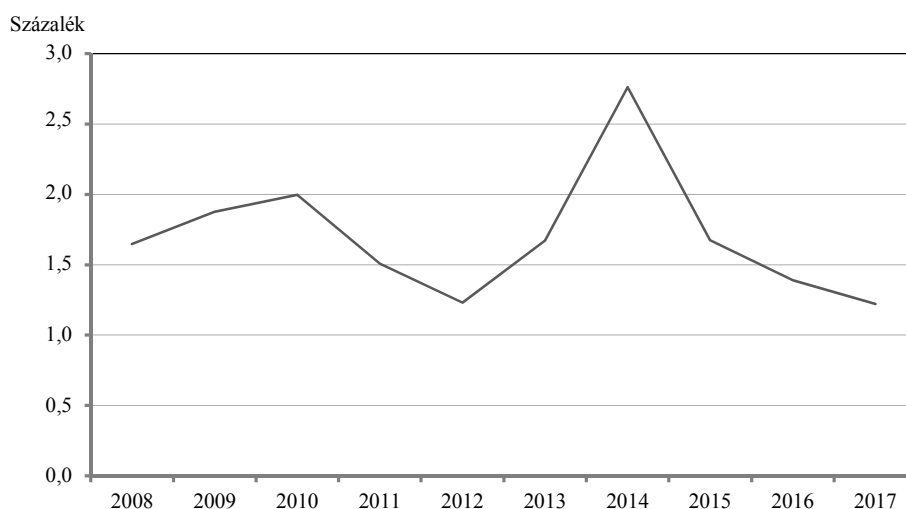
A modellezési koncepció empirikus vizsgálatához Magyarországon bejegyzett társas vállalkozások mintáját használtuk. Kiválasztási szempont volt, hogy a vállalkozásokra vonatkozóan rendelkezésre álljanak a számviteli beszámolók legalább nyolc egymást követő évre vonatkozóan a tanulmányban vizsgálni kívánt dinamizálási koncepció adatigénye miatt. További mintavételi feltétel volt a legalább nyolc év vonatkozásában a folyamatos működés, amit annak követelményével operacionálizáltunk, hogy a mintába csak olyan vállalkozás kerülhetett, amely a vizsgált évek mindegyikében realizált árbevételt.

A megfogalmazott követelményeknek eleget tevő vállalkozások pénzügyi adatait manuálisan gyűjtöttük össze a következő nyilvánosan hozzáférhető adatbázisokból. Azok a vállalkozások minősültek fizetéseképtelennek, amelyekkel szemben a Cégbegyűzők hatályos adatai alapján csőd- vagy felszámolási eljárás megindítására került

sor; azaz a fizetésképteliséget a jogi értelemben definiáltuk. Fizetőképesség⁷ azokat a vállalkozásokat tekintettük, amelyekkel szemben semmilyen⁸ eljárásra nem került sor a hivatalos nyilvántartások tanúsága szerint a mintavétel időpontjáig. A tanulmányban az eddigiek során és ezt követően is a „fizetésképtelenség” és „csőd”, illetve a „fizetőképesség” és „működő” kifejezéseket szinonimaként kezeljük.

A számviteli beszámolók adatainak forrása az Igazságügyi Minisztérium Elektronikus Beszámoló⁹ nevű oldala volt. Ha adott évben a felszámolási eljárás alá kerülő társas vállalkozások számát az összes társas vállalkozás számához viszonyítjuk, akkor képet kaphatunk a hazai „csődarány” alakulásáról. (Lásd a 2. ábrát.)

2. ábra. A felszámolás alá került társas vállalkozások aránya az összes társas vállalkozáshoz viszonyítva



A 2. ábra tanúsága szerint a társas vállalkozások „csődaránya” 1–2 százalékos között alakult az elmúlt évtizedben. Fontos hozzátenni, hogy az 1. és a 2. ábra a Magyarországon működő összes társas vállalkozás adatait tartalmazza, ebben a tanulmányban azonban csak olyan társas vállalkozásokat vizsgáltunk, amelyek legalább nyolc egymást követő éven át működtek. Erre az alcsoportra vonatkozóan nem állnak rendelkezésre adatok. A 2016-ra vonatkozó Statisztikai Tükör adatai szerint a társas

⁷ Megjegyzendő, hogy egy vállalkozás pénzügyi értelemben lehet fizetésképtelen, függetlenül attól, hogy vele szemben nem került sor sem csőd-, sem pedig felszámolási eljárás megindítására.

⁸ Csak olyan vállalkozások szerepelhettek a működő cégek között, amelyekkel szemben csőd-, felszámolási, végelszámolási, illetve kénysztörletési eljárásra sem került sor a mintavétel időpontjáig.

⁹ <https://e-beszamolo.im.gov.hu/oldal/kezdolap>

vállalkozások öt éves túlélési aránya 44 százalék, azaz a társas vállalkozások közel fele az első öt évben befejezi működését. Ebből feltételezhető, hogy az első öt év tekintetében nagyobb, azt követően pedig alacsonyabb a megszűnési, illetve „csődarány”, ami feltehetően még inkább érvényes tendencia a nyolc vagy több év vonatkozásában. Kiemelendő továbbá, hogy egy vállalkozás nem kizárólag fizetési képzetlenség miatt szűnhet meg, azaz az előbb említett 56 százalékos megszűnési rátának csak egy részét teszik ki a felszámolási eljárások. Jelen tanulmány kapcsán ezek azért lényeges információk, mert rámutatnak arra, hogy a vizsgálni kívánt sokaság mérete kicsi ahhoz, hogy az empirikus vizsgálatban alkalmazott nagy mintát igénylő CHAID-módszer adatigénye kielégíthető legyen oly módon, hogy a vizsgálódást egy-egy konkrét üzleti évre lehessen szűkíteni.¹⁰ Ebből adódóan az adatgyűjtés a 2007 és 2016 közötti periódust¹¹ ölelte fel. Az egy évtizedet átfogó időintervallum alkalmazását az erre szükségessé, hogy a mintában minden nemzetgazdasági ág, illetve minden méretkategória nemzetgazdasági súlyának megfelelően képviseltetve legyen, ugyanakkor kellő számban rendelkezésre álljanak felszámolási eljárás alá került vállalkozások az egyes alcsoportokban és a minta egészében egyaránt. A tízéves periódus alkalmazása azonban befolyásolhatja az eredményeket, ezért a teljes mintán végzett számításokat robusztussági vizsgálat keretében megismételtük a mintában legnagyobb arányt¹² kitevő 2010. év adataira külön is. Az eredményeket a 4. fejezet ismerteti.

A mintában szereplő társas vállalkozások száma 2098, amelyek körében a működők aránya 54 százalék volt. A megfigyelések nemzetgazdasági ágak szerinti eloszlását az 1. táblázat mutatja.

A társas vállalkozások méret szerinti megoszlását a 2. táblázat mutatja. Mivel a létszámra vonatkozóan sok esetben nem állt rendelkezésre információ,¹³ ezért a méretet a realizált nettó árbevétel összegével definiáltuk, és a vonatkozó törvényi értékhatárok¹⁴ alapján kategorizáltuk.

¹⁰ Ez csak az adott üzleti év szinte teljes körű megfigyelésével lenne megvalósítható.

¹¹ Az empirikus vizsgálatban a $t - 2$ -vel jelölt év adatai e periódus éveire vonatkoznak.

¹² Nagyságrendileg az adatállomány negyede.

¹³ A KSH adatai szerint 2018 novemberében a regisztrált társas vállalkozások száma 521 369, amelyek közül 257 735 esetben a vállalkozás létszáma nulla fő vagy a létszámadat ismeretlen (http://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_qvd022.html).

¹⁴ A kis- és középvállalkozásokról, fejlődésük támogatásáról szóló 2004. évi XXXIV. törvény 3. § (1)–(3) bekezdései.

1. táblázat

Az empirikus vizsgálathoz felhasznált minta megoszlása nemzetgazdasági ágak szerint, 2009–2018

Nemzetgazdasági ág	Fizetőképes társas vállalkozások száma	Fizetésképtelen társas vállalkozások száma	Összesen
Adminisztratív és szolgáltatást támogató tevékenység	61	27	88
Bányászat, kőfejtés	0	1	1
Egyéb szolgáltatás	13	7	20
Építőipar	87	133	220
Feldolgozóipar	102	180	282
Humánegészségügyi, szociális ellátás	30	3	33
Információ, kommunikáció	79	37	116
Ingatlanügyletek	125	64	189
Kereskedelem, gépjárműjavítás	290	276	566
Közigazgatás, védelem, kötelező társadalombiztosítás	1	0	1
Mezőgazdaság, erdőgazdálkodás, halászat	32	33	65
Művészet, szórakoztatás, szabadidő	17	23	40
Oktatás	15	2	17
Pénzügyi, biztosítási tevékenység	21	6	27
Szakmai, tudományos, műszaki tevékenység	183	45	228
Szálláshely-szolgáltatás, vendéglátás	38	48	86
Szállítás, raktározás	38	67	105
Villamosenergia-, gáz-, gőzellátás, légkondicionálás	2	1	3
Vízellátás, szennyvíz gyűjtése, kezelése, hulladék-gazdálkodás, szennyeződésmérséklés	6	5	11
<i>Összesen</i>	<i>1 140</i>	<i>958</i>	<i>2 098</i>

Megjegyzés. A vállalkozások pénzügyi adatai a 2007 és 2016 közötti időszakra vonatkoznak.

2. táblázat

Az empirikus vizsgálathoz felhasznált minta méret szerinti megoszlása

Vállalkozás mérete	Fizetőképes	Fizetésképtelen	Összesen
Mikro	975	823	1 798
Kis	120	89	209
Közép	35	40	75
Nagy	10	6	16
<i>Összesen</i>	<i>1 140</i>	<i>958</i>	<i>2 098</i>

A KSH adataival¹⁵ összevetve megállapítható, hogy a minta tevékenységi kör tekintetében jól reprezentálja a társas vállalkozások nemzetgazdaságon belüli megoszlását, mivel a legnagyobb arányban a kereskedelem, az ingatlanügyletek, illetve a szakmai, tudományos és műszaki tevékenységek képviseltetik magukat. A méret tekintetében az összevetés nehezebb, mivel a KSH a vállalkozások méret szerinti kategorizálását jellemzően a létszám alapján közli. Ez az információ azonban az empirikus vizsgálathoz felhasznált adatbázisban nem állt rendelkezésre. Ha feltételezzük, hogy a létszám, illetve a tanulmányban alkalmazott árbevétel szerinti megoszlás hasonlóan tekinthető, és eltekintünk azokról a társas vállalkozásokról, amelyek létszámadata nulla vagy nem ismert, akkor megállapítható, hogy az empirikus vizsgálathoz felhasznált adatok méret tekintetében is jól reprezentálják a nemzetgazdaság egészét.¹⁶ Az eltérés abban mutatkozik, hogy a vizsgált adathalmazban a közép-, illetve nagyvállalatok kismértékben felülreprezentáltak a mikro- és kisvállalkozások terhére. Az összevetések kapcsán azonban megjegyzendő, hogy az összehasonlítás alapját képező adatokat a KSH az összes társas vállalkozásra közli, jelen tanulmány azonban csak a legalább nyolc éve működő (árbevételt realizáló) vállalkozásokat vizsgálja. Erre az alcsoportra nyilvános statisztikák nem állnak rendelkezésre, így a teljes populáció értékeivel való összevetés csak akkor helytálló, ha feltételezhetjük, hogy a legalább nyolc éve működő társas vállalkozások méret és nemzetgazdasági ág szerinti megoszlása hasonló az összes társas vállalkozás arányaihoz.

A mintában szereplő megfigyelések pénzügyi helyzetének mérésekor a következő pénzügyi dimenziókat vettük figyelembe:

- Likviditás: likviditási ráta, likviditási gyorsráta, gyorsfizető képesség, pénzeszközök aránya a forgóeszközön belül, nettó forgótőke aránya.
- Adósságfinanszírozás:¹⁷ a működésből származó pénzeszközök rövid lejáratú és összes kötelezettséghez viszonyított aránya.
- Hatékonyság: eszközök, készletek, követelések, kötelezettségek forgási ideje.¹⁸
- Tőkeszerkezet: saját és idegen tőke aránya, bonitás.

¹⁵ Forrás: https://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_qvd020a.html

¹⁶ A KSH adataival összevetve (https://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_qvd022.html).

¹⁷ E mutatócsoport elemeinél a mérleg és az eredménykimutatás egyes sorait osztottuk. Mivel a mérleg adatai fordulónapra, az eredménykimutatás számai pedig egy beszámolási időszakra vonatkoznak, így az e körbe tartozó mutatóknál a mérlegből származó értékeket a nyitó- és zárómérleg átlagában vettük figyelembe. Ugyanígy jártunk el a hatékonysági mutatók esetén is. A többi változócsoporthoz a mutatók számítása csak egyetlen pénzügyi kimutatás (kizárólag mérleg vagy eredménykimutatás) adataiból történt, így azokban az esetekben a mérlegből származó számértékeket nem átlagoltuk.

¹⁸ A mutatót kétféle szemléletben is alkalmaztuk. Az egyik esetben a számításmód: Rövid lejáratú kötelezettségek átlagos állománya/Értékesítés nettó árbevétele. A másik esetben: Rövid lejáratú kötelezettségek átlagos állománya/(Anyagi és Személyi jellegű ráfordítások összege).

- Jövedelmezőség: árbevétel- és eszközarányos jövedelmezőség, eredménytartalék aránya.
- Méret: árbevétel és mérlegfőösszeg nagysága.

Mindezek összesen húsz lehetséges magyarázó változót jelentenek.

Nyitrai [2014] a statikus pénzügyi adatok dinamikájának megragadására a következő formulát javasolta:

$$\frac{X_{t-2} - X_{\min[t-3; t-m]}}{X_{\max[t-3; t-m]} - X_{\min[t-3; t-m]}} \quad /1/$$

Az /1/ képletben az X egy adott pénzügyi mutatószám értékét jelöli. A számítás a legutoljára megfigyelt év ($t-2$) adatát az azt megelőző időszak (m a rendelkezésre álló idősor tagszáma) szóródási terjedelmének szélsőértékeihez viszonyítja. Tekintsük példaként egy olyan vállalkozást az empirikus vizsgálathoz felhasznált mintából, amely 2011-ben került felszámolási eljárás alá. A 3. táblázat a korábban említett bonitás értékének időbeli alakulását mutatja.

3. táblázat

Egy 2011-ben felszámolási eljárás alá került vállalat bonitásának időszora

Év	Év jelölése	Bonitás
2010	$t-1$	-14,9
2009	$t-2$	12,7
2008	$t-3$	16,5
2007	$t-4$	2,8
2006	$t-5$	2,5

Alkalmazzuk az /1/-es képletet! A példában az X a bonitást jelenti. A felszámolási eljárás 2011-ben indult. Ezt az évet t jelöli. A tanulmányban kétéves előrejelzési időhorizontot alkalmazunk, ezért a $t-1$. év adatát figyelmen kívül hagyjuk, azaz a $t-2$. év adatát ($X_{t-2} = 12,7$) viszonyítjuk az azt megelőző – 2006 és 2008 közötti – időszak minimumához ($X_{\min[t-3; t-m]} = 2,5$) az /1/ formula számlálójában. A nevező pedig a 2006 és 2008 közötti időszak maximális ($X_{\max[t-3; t-m]} = 16,5$) és minimális értékének különbségként határozható meg.

Számszerűsítve:

$$\frac{X_{t-2} - X_{\min[t-3; t-m]}}{X_{\max[t-3; t-m]} - X_{\min[t-3; t-m]}} = \frac{12,7 - 2,5}{16,5 - 2,5} = 0,73.$$

Az eredményt úgy értelmezhetjük, hogy a vizsgált vállalkozás esetén a bonitás $t-2$. időpontbeli értéke az azt megelőző időszak szóródási terjedelmén belül¹⁹ helyezkedik el – közelebb a $t-2$. évet megelőző időszak maximumához, mint a minimumához.

A dinamizálásra javasolt /1/-es képlet bármely pénzügyi mutató esetén alkalmazható az időbeli tendencia számszerűsítésére. Az empirikus vizsgálatok során az /1/-es formulát a korábban bemutatott húsz pénzügyi mutató mindegyikére kiszámítottuk; így összességében negyven²⁰ potenciális magyarázó változó állt a modellezés során rendelkezésre.

3. A CHAID-alapú csődelőrejelzés lépései

A fejezet ismerteti a tanulmányban javasolt adatbányászati megközelítés lépéseit:

- a csődmodellépítés alapfogalmait;
- a CHAID-alapú kategorizálás részleteit;
- a javasolt koncepció folyamatábráját és gyakorlati megvalósítását.

3.1. A csődmodellekről

A csődelőrejelző modellek hagyományosan a vállalatok által készített számviteli beszámolók adataiból számítható pénzügyi mutatók értékein alapulnak. Ugyan a könyvelési tételek eredményeképp kialakuló adatok alapvetően a múltbeli vezetői döntések pénzügyi következményeit tükrözik (*Virág et al.* [2013]), azok bizonyos mértékben alkalmasak a vállalatok jövőbeli fizetőképességének előrejelzésére is. A modellépítés alapja annak megfigyelése, hogy a múltban mely vállalatok váltak fizetésképtelenné és melyek nem. A nemfizetés ténye objektív módon legtöbbször²¹ csak jogi oldalról ra-

¹⁹ Formálisan: a vizsgált vállalat esetén a bonitás értéke a $t-2$. évben az azt megelőző időszak szóródási terjedelmének 73. percentilisével egyezik meg.

²⁰ Húsz statikus mutató és azok /1/-es formula segítségével dinamizált változatai.

²¹ A vállalkozások hitelezői, illetve szállítói rendelkeznek múltra vonatkozó nemfizetési információkkal, azonban ezek az adatok nem nyilvánosak, így a modellezés jellemzően a jogi értelemben vett csődelőrejelzésére irányul a szakirodalomban.

gadható meg. Tanulmányunkban magunk is ezt a megközelítést választottuk. A mintavétel során alkalmazott pontos definíciót a 2. fejezetben mutattuk be.

Hazánkban a számviteli törvény²² az előző üzleti évre vonatkozó beszámoló közvetélteli kötelezettségének határidejét a fordulónaptól számított ötödik hónap utolsó napjaként határozza meg. Ebből következően a gyakorlatban legalább öthónapos késedelemmel ismerhetők meg az előző üzleti év adatai. Annak érdekében, hogy a felállított modellek valódi előrejelző képességgel rendelkezessenek, a fizetési képtelenség t . évbéli bekövetkezését valamely korábbi évből származó beszámoló adatainak alapján célszerű modellezni. A legtöbb esetben ez a $t - 1$. év adatait jelenti, de az előbb említett öthónapos késedelem miatt ebben a tanulmányban a $t - 2$. év adatait használtuk fel annak érdekében, hogy a felállított modellek a gyakorlati alkalmazás szempontjából hasznosabbak legyenek a hosszabb előrejelzési időhorizont révén. Mindez lehetőség nyújthat a vállalatok esetleges fizetési képtelenségében érdekeltek számára, hogy megelőző lépéseket tegyenek a potenciális veszteségek elkerülése, illetve minimalizálása érdekében.

A csődmodellezés statisztikai szempontból egy klasszifikációs feladat megoldását jelenti. Cél a fizetési képtelenség t . évbéli bekövetkezése (fizetőképesség vagy fizetési képtelenség) mint bináris célváltozó, illetve a valamely korábbi évből²³ vett magyarázó változók közötti sztochasztikus kapcsolat feltárása. Ezen összefüggés feltárására széles eszköztárat kínál a többváltozós statisztika, az adatbányászat és a mesterséges intelligencia is. Ezek közül a CHAID-módszert adatelőkészítési, a lineáris diszkriminancia-analízist pedig klasszifikációs céllal alkalmaztuk; utóbbi általános alakja:

$$z = a + b_1x_1 + b_2x_2 + \dots + b_kx_k.$$

A b együtthatók meghatározásának matematikai alapja, hogy az eljárás célja olyan csoportok képzése, amelyek esetén a belső szóródás minimális, a külső viszont maximális (Hajdu [2003]).

3.2. A CHAID-alapú kategorizálás

A csődmodellek magyarázó változói között használt pénzügyi mutatószámok matematikai szempontból folytonos változók, amelyek értékkészlete az esetek nagy részében nem korlátos. További lényeges sajátosság, hogy gyakran találkozhatunk kiugró értékekkel a csődelőrejelzési adatbázisokban. Az extrém értékek torzító hatását célszerű kiküszöbölni. E célra számos eljárás áll rendelkezésre. Ezek közül praktikus választás lehet a Hámori [2016] által javasolt CHAID-módszer szerinti kategorizálás. Ennek lényege, hogy a változók értékkészletét a decilisek alapján tíz osztályközre bontjuk, majd az így képzett kategóriák összes lehetséges párosítása közül iteratív módon egye-

²² Forrás: 2000. évi C. törvény a számvitelről 153. § (1) bekezdés.

²³ Esetünkben ez a két évvel korábbi évet jelenti.

sítjük azokat, amelyek esetében a legnagyobb szignifikanciaszinten elfogadható a statisztikai függetlenség feltételezése az adott kategóriába tartozás ténye és a fizetésképtelenség bekövetkezése szerint képzett két kategória²⁴ között. Ez az eljárás mindaddig ismétlődik, amíg létezik olyan kategóriapár, amely esetén a függetlenség hipotézise²⁵ elfogadható a felhasználó által meghatározott szignifikanciaszint mellett.

A könnyebb érthetőség kedvéért tekintünk példaként az empirikus vizsgálatban felhasznált adatbázis bonítás²⁶ nevű mutatójának CHAID-módszerrel történő kategorizálását! (Lásd a Függelék F1. ábráját.) Az F1. ábrán a működés tényét 0, a fizetésképtelenségét pedig 1 jelöli. Az eljárás során a mutató empirikus eloszlásának decilisei alapján képzett tíz kategóriából kiindulva három összevonást hajtottunk végre. Látható, hogy az egyes kategóriákban nem azonos a fizetésképtelen megfigyelések aránya. Az empirikus eredmények arra utalnak, hogy nemlineáris (U alakú) kapcsolat mutatkozik a bonítás értéke és a fizetésképtelenség bekövetkezésének valószínűsége között. *Hámori* [2016] ajánlása szerint az egyes kategóriákba kerülés tényét dummy változóként kódolva van lehetőség az adott változó modellbe építésére. Jelen tanulmány ettől annyiban tér el, hogy az egyes kategóriák sorszámait használja fel magyarázó változóként a csődelőrejelző modellekben. Ez azonban a korábban bemutatott nemlineáris kapcsolat miatt problémát jelenthet a klasszifikációs módszerként használt diszkriminancia-analízis későbbi alkalmazása során, mivel az utóbbi feltételezi a magyarázó tényezők többdimenziós normális eloszlását. A probléma kezelése szempontjából kézenfekvő megoldás lehet a kategóriák sorszámainak átrendezése annak függvényében, hogy az adott kategóriában milyen arányt képviselnek a fizetésképtelen vállalkozások. Csökkenő sorrend esetén ez a következő kategorizálást jelenti.

4. táblázat

A kategóriák újraszámozása a fizetésképtelen vállalatok arányának csökkenő sorrendje szerint

Kategória eredeti sorszáma	Osztályköz felső határa	Kategória új sorszáma
1	-2,25	1
2	0,01	3
3	0,44	7
4	0,81	6
5	1,32	5
6	2,08	4
7	végtelen	2

²⁴ Fizetőképesség vagy fizetésképtelenség.

²⁵ A függetlenségvizsgálat alapja a két minőségi ismérv közötti sztochasztikus kapcsolat elemzése során általánosan elterjedt χ^2 -próba, amelyet részletesen ismert *Hunyadi-Vita* [2004] munkája.

²⁶ Számítási mód: (Hosszú és Rövid lejáratú kötelezettségek)/Saját tőke.

Ehhez hasonló szemléletben kategorizálható és szükség esetén újraszámozható az összes potenciális magyarázó változó értékészlete a modellépítést megelőzően. A változók ily módon felállított kategóriarendszerei képezik a dinamizálás koncepciójának alapját, amelyet a következő alfejezetben mutatunk be. Az eljárás eredményeképp minden magyarázó változó értékészlete osztályközökre bomlik. Ezt felülvizsgált kategorizálásnak nevezzük, mivel az egyes osztályközök határait nem valamilyen egységes szabályt²⁷ követve határozzuk meg, hanem azon elvárás alapján, hogy a létrejövő csoportok a modellezési célváltozó (fizetőképes vállalat vagy sem) tekintetében a lehető leghomogénebbek legyenek.

3.3. Dinamizálás a CHAID-módszer szerint képzett kategóriák alapján

Az előző alfejezetben látott kategorizálás alkalmas az egyes vállalatok esetén a mutatók időbeli folyamatának számszerű leképezésére is. A 4. táblázatban látott kategorizálás az empirikus vizsgálathoz felhasznált adatbázis vállalatainak $t-2$. évbéli adatai alapján készült. Ez az osztályközös gyakorisági táblázat alkalmas arra is, hogy az egyes vállalatok adott mutatóit több év vonatkozásában is követni tudjuk. Példaként tekintsünk az empirikus vizsgálathoz használt adatbázisból egy felszámolás alá került vállalatot, amely esetén a bonitás értéke a fizetésektelenség bekövetkezésétől az időben távolodva rendre: 2,76; 1,94; 1,03. Ezen értékek kategória-sorszámai²⁸ a 4. táblázat szerint rendre: 2, 4, 5. Ha ezeket a számokat egymás mellé helyezzük, akkor 245-öt kapunk. Ez az eljárás alkalmas eszköz lehet az adott mutató esetén az időbeli tendencia számszerűsítésére, mivel Nyitrai [2015] empirikus vizsgálatai szerint a fizetésektelenség bekövetkezésétől időben távolodva a pénzügyi mutatók magyarázó ereje egyre kisebb mértékű.²⁹ Ebből adódóan célszerű azokat a felszámolás vagy csődeljárás megkezdésétől időben távolodva csökkenő súllyal figyelembe venni. Erre a célra alkalmas megközelítés lehet a kategória-sorszámok egymás mellé helyezése, mivel ekkor a legnagyobb helyi értéken a pénzügyi mutató legaktuálisabb értéke szerepel, a középső helyi értéken az egyvel korábbi év adata és végül a legkisebb helyi értéken a még egy évvel korábbi érték.

Ezt követően lehetőség nyílik az így képzett folyamatszemplétű mutatók magyarázó változóként történő felhasználására a csődelőrejelző modellek keretei között. Az Olvasóban azonban ezen a ponton joggal merülhet fel annak aggálya, hogy a kategória-sorszámok egymás mellé helyezése egyúttal azt a feltevést is maga után vonja,

²⁷ Például k darab egyforma hosszúságú osztályköz.

²⁸ Átrendezést követő új sorszámok.

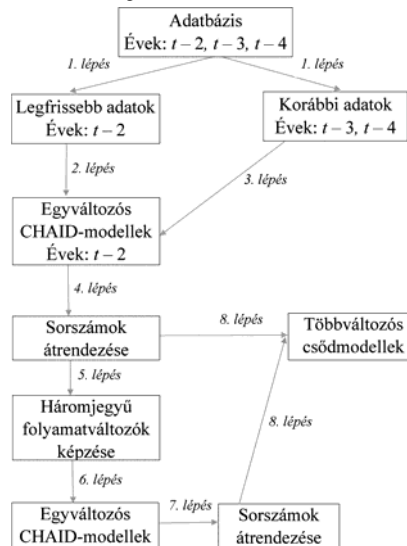
²⁹ Az empirikus vizsgálathoz felhasznált adatbázis esetén ezt magunk is megvizsgáltuk. Az eredmények megerősítették az idézett szerző tapasztalatait. A vonatkozó számítások eredményeit az 5. táblázat tartalmazza.

hogy a százas nagyságrendű változókból eredően a $t-2$. évből vett mutató „fontossága” többszázszorosa a $t-4$. évből vett adatnak. Ez nyilvánvalóan nem tartható feltevés, ezért itt ismét a CHAID-módszert hívjuk segítségül annak érdekében, hogy a kategória-sorszámok egymás mellé helyezéséből kapott mutatókat objektív módon kategorizáljuk annak vizsgálata céljából, hogy mely folyamatszoportok tekintetében képezhetők a homogén célváltozó csoportok. Ezt a bonítás esetén az F2. ábra szemlélteti.

A példában szereplő vállalkozás esetén a folyamatváltozó 245-ös értéke a második kategóriába esik. A rendelkezésre álló mintában az ilyen folyamatot követő vállalkozások 62,1 százaléka két éven belül fizetéképtelenné vált, amely jelzésértékű a vállalat jövőjére vonatkozóan. Az F2. ábrán látható, hogy az imént képzett folyamatváltozó tekintetében az egyes csoportokat monoton csökkenő csődarány jellemzi a nagyobb számmal jelölt kategóriák felé haladva. A bonítás statikus értékeinek kategorizáláshoz hasonlóan a most létrehozott kategóriák sorszámait is felhasználhatók a klasszifikációs módszerek magyarázó változói között. A folyamatváltozókat kategorizálásakor azonban éppúgy felmerülhet az a probléma, hogy az egymást követő kategóriákban a fizetéképtelen vállalatok aránya nem monoton növekvő vagy csökkenő. Ez problémát okozhat az olyan módszerek esetén, mint a lineáris diszkriminanciaanalízis. A bonítás statikus eseténél látott módszerrel (lásd a 4. táblázatot) azonban ezúttal is lehetőség van a kategóriák újraszámolására.

Az alfejezetben bemutatott eljárást sematikusan a 3. ábra foglalja össze.

3. ábra. A CHAID-alapú dinamikus csődelőrejelzési koncepció lépései



Megjegyzés. A 8. lépés esetén a vízszintes nyíl arra utal, hogy az empirikus vizsgálat során a $t-2$., $t-3$. és $t-4$. évek statikus értékeivel, valamint a folyamatszámított mutatók alkalmazásával is (felfelé mutató nyíl) készültek modellek. Az empirikus vizsgálat célja a statikus és a tanulmányban javasolt folyamatszámított mutatókra épített modellek előrejelző teljesítményének összevetése.

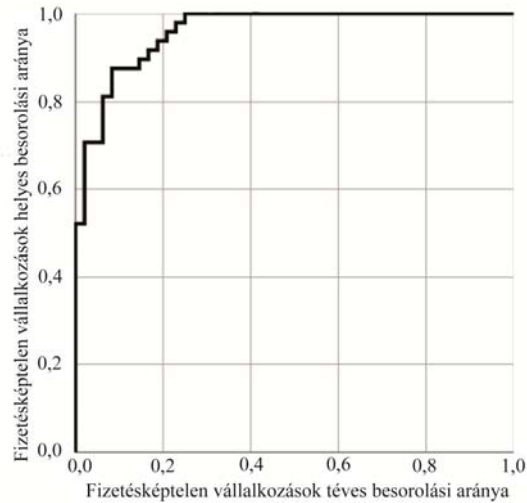
4. Eredmények

A kutatás célja két változócsoport összevetése volt: az egyik esetben a legutoljára megfigyelt év adatai szerepeltek a CHAID-módszer szerint kategorizálva; a másik esetben a lineáris diszkriminanciaanalízis magyarázó változóit a 3.3. alfejezetben bemutatott folyamatjellegű változók képezték. A teljes mintán végzett számítások eredményeit a 4.1., a következtetések megerősítése céljából végrehajtott robusztusági vizsgálatok tapasztalatait a 4.2. alfejezet foglalja össze.

4.1. A teljes mintán kapott eredmények

A modellezés célja a minél a pontosabb előrejelzés volt. A prediktív teljesítmény megítéléséhez a rendelkezésre álló adatkört két részre szükséges bontani: tanuló és tesztelő mintákra. Az objektivitás érdekében a felosztásnak véletlennek kell lennie, de egy egyszeri véletlenszerű felosztás is könnyen vezethet nem kellően megalapozott következtetések levonásához. Ennek érdekében a tanulmány eredményei és következtetései tízszeres arányos keresztvalidáción alapulnak, melynek lényege, hogy a teljes adathalmazt tíz egyenlő részre osztottuk fel oly módon, hogy az egyes részekben belül a fizetésképtelen vállalatok aránya megegyezett a teljes adathalmazzal jellemző aránnyal (Nyitrai [2014]). A keresztvalidáció során minden mintafelosztási egység egy alkalommal a tesztelő minta szerepét töltötte be, miközben a tanuló mintát a többi kilenc rész együttese alkotta. A tíz tesztelő mintán kapott eredmény átlagában jó eséllyel számíthatunk arra, hogy a vizsgált adathalmaz vonatkozásában objektíven ítéltük meg az adott klasszifikációs módszer vagy modellezési koncepció teljesítményét. A tanulmányban a modellek előrejelző képességét a csödelőrejelzés módszertanában általánosan elterjedt ROC- (receiver operating characteristic – kumulált besorolási pontosság) görbe alatti terület segítségével mérjük (Kristóf [2018]). Lineáris diszkriminanciaanalízis esetén az egyes vállalatok besorolása a fizetőképes és fizetésképtelen kategóriákba egy ún. vágópont segítségével történik. Ha a diszkriminanciafüggvény értéke a vágópont alá (fölé) esik, akkor a megfigyelést az egyik (másik) csoportba soroljuk. Ebből azonban következik, hogy a vágópont értékének megválasztásával befolyásolható a modell találati aránya. Annak érdekében, hogy a modellek teljesítményét ne egy önkényesen kiválasztott vágópont mellett ítéljük meg, lehetőség van a modell által adott helyes és téves besorolási arányok vizsgálatára az összes lehetséges vágópont mellett. Ennek grafikus megjelenítése a ROC-görbe, melynek vízszintes (függőleges) tengelyén a tévesen (helyesen) besorolt fizetésképtelen vállalatok aránya látható. Az SPSS szoftver által készített ROC-görbét szemlélteti a 4. ábra.

4. ábra. ROC-görbe az SPSS megjelenítésében



Megjegyzés. Az ábrán látható görbe a robusztussági vizsgálatok során készült.

A ROC-görbe alatti terület vonatkozásában a 0,5 érték a véletlen találgatást, az 1 (mint elméleti maximum) a tökéletes klasszifikációt szimbolizálja. A két szám között relatív módon mérhető egy modell besorolási képessége. A tíz felosztás tesztelő mintáin kapott eredményeket az 5. táblázat mutatja.

5. táblázat

A számítások eredményei (a ROC-görbe alatti terület aránya)

Mintafelosztás sorszáma	$t-2$. év	$t-3$. év	$t-4$. év	Folyamatváltozók három év adatai alapján	Folyamatváltozók négy év adatai alapján
1	0,867	0,806	0,755	0,880	0,898
2	0,893	0,847	0,781	0,905	0,895
3	0,877	0,808	0,732	0,877	0,868
4	0,916	0,822	0,766	0,930	0,922
5	0,887	0,816	0,797	0,890	0,918
6	0,847	0,805	0,751	0,851	0,862
7	0,891	0,830	0,819	0,901	0,900
8	0,872	0,799	0,781	0,879	0,889
9	0,884	0,852	0,788	0,890	0,885
10	0,834	0,777	0,731	0,871	0,883
Átlag	0,877	0,816	0,770	0,887	0,892
Szórás	0,022	0,021	0,027	0,020	0,018

Az 5. táblázatban $t - 2$. év jelöli a fizetéseképtelenség bekövetkezését megelőző második évet. Ennek az évnek az adatai képezték a statikus modellek változókörét. Hasonlóképpen, $t - 3$. és $t - 4$. év az előre jelezni kívánt eseményt megelőző három, illetve négy évvel korábbi éveket jelöli. Az empirikus vizsgálat részeként megvizsgáltuk, hogy miként alakul a modellek előrejelző teljesítménye, ha a három, illetve négy évvel korábbi pénzügyi mutatók statikus értékeire építünk modelleket. Az összes mintafelosztás esetén azt tapasztaltuk, hogy a fizetéseképtelenség bekövetkezésétől időben távolodva monoton csökken³⁰ a statikus pénzügyi mutatókra épített modellek prediktív képessége. Ez azonban nem zárja ki azt, hogy a vizsgált három év adatai együtt – a korábban bemutatott módon egy változóba sűrítve – releváns magyarázó erővel rendelkezzenek a jövőbeli fizetéseképtelenség bekövetkezése szempontjából.

Az 5. táblázat eredményei arra is rámutatnak, hogy a tanulmányban javasolt dinamizálási koncepció segítségével egyetlen mintafelosztás esetén sem adódott gyengébb eredmény, mint a legutoljára megfigyelt év ($t - 2$) statikus adatai alapján. A vizsgált adathalmaz egészét tekintve a tíz mintafelosztás átlagában a statikus mutatók alapján felállított modellekhez képest 1 százalékponttal nagyobb előrejelző képesség jellemezte a folyamatjellegű változókra épített modelleket. A páros t -próba nemparaméteres megfelelője³¹ esetén az átlagos ROC-görbe alatti területek között tapasztalt különbség szignifikánsan különbözik nullától (p -érték: 0,008), azaz a 4.3. fejezetben bemutatott folyamatjellegű változók használata szignifikánsan nagyobb előrejelző képességű modellek felállítását tette lehetővé a rendelkezésre álló adatbázis esetén, mint ami statikus pénzügyi mutatók alkalmazásával volt elérhető.

Az eddigiekben három egymást követő üzleti év adataiból képzett idősorok kapcsán vizsgáltuk a tanulmányban javasolt dinamizálási koncepció hatékonyságát. Felmerülhet azonban a kérdés: volna-e hozzáadott értéke annak, ha növelnénk az elemzett idősor hosszát. A vizsgált adathalmaz kapcsán empirikus választ úgy kerestünk, hogy a számításokat megismételtük az előbbiekkal teljesen azonos módon, viszont annyi változtatással, hogy a vizsgált idősort kiegészítettük a fizetéseképtelenség bekövetkezését megelőző ötödik év adatával. Az azokból képzett folyamatjellegű változókra épített modellek előrejelző képességét az 5. táblázat utolsó oszlopa mutatja. Az eredmények szerint ugyanazon tíz mintafelosztás átlagában további 0,5 százalékponttal magasabb ROC-görbe alatti terület mutatkozott a négy üzleti év trendjének információtartalmát magába sűrítő folyamatváltozók alkalmazása esetén, mint az egy évvel rövidebb idősorok használata mellett. Fontos azonban kiemelni, hogy nem minden mintafelosztásra volt érvényes ez: a tíz vizsgált esetből ötnél a rövidebb idősor alkalmazása bizonyult hatékonyabbnak. Mindez a vonatkozó statisztikai teszt eredményén is megmutatkozott, mivel a 0,5 százalékpontos eltérés nem minősült

³⁰ A tendencia a robusztussági vizsgálatok eredményeiben is megmutatkozott. Lásd a 6. táblázat megfelelő adatait.

³¹ Wilcoxon-próba.

szignifikánsnak,³² így a vizsgált adathalmaz esetén megállapítható, hogy nincs számottevő hozzáadott értéke három tagnál hosszabb folyamat megfigyelésének. Ezt a megállapítást erősíti az is, hogy az idősor hosszának növelése tovább növeli a bemutatott dinamizálási koncepció egyébként is jelentős adatigényét.

4.2. Robusztussági vizsgálatok

Az eddigi számítások eredményei a teljes adathalmazra vonatkoztak, ahol a különböző méretű és tevékenységi körű vállalatok együtt szerepeltek az elemzésben. Kérdéses lehet, hogy ez a mintabeli heterogenitás mennyiben befolyásolhatta az 5. táblázat adatait és az azok alapján levont következtetést. Ennek vizsgálata céljából az elemzést megismételtük a rendelkezésre álló adatbázis két legnagyobb rész-halmazán. Mivel a CHAID-módszer alapját képező függetlenségvizsgálat egy ún. „nagy mintás” próba (*Hunyadi-Vita* [2004]), így a tevékenységi kör és méret tekintetében csak azokban a kategóriákban van lehetőség az elemzés megismétlésére, ahol a megfigyelések száma elegendő³³ ahhoz, hogy tanuló és tesztelő minták használatára is lehetőség legyen. A tevékenységi kör tekintetében a „Kereskedelem, gépjárműjavítás” nemzetgazdasági ág vállalatai, a méret vonatkozásában pedig a mikro-vállalkozások tekinthetők ebből a szempontból a legalkalmasabbnak.

A robusztussági vizsgálatok számításai a teljes mintán végzett elemzésekkel azonos módon készültek. Az egyetlen különbség, hogy a „Kereskedelem, gépjárműjavítás” nemzetgazdasági ágba sorolt vállalatok esetén a CHAID-módszer alkalmazása során a kategóriák képzéséhez szükséges minimális elemszámot 50-ről 5-re³⁴ csökkentettük, mivel a minta ezen szegmensében jóval kevesebb adat állt rendelkezésre,

³² A Wilcoxon-próba p -értéke 20 százaléknál nagyobb.

³³ A tanuló és tesztelő minta nagyságára vonatkozóan nem áll rendelkezésre objektív iránymutatás. Általánosságban kijelenthető, hogy mindkét esetben a nagyobb elemszám preferált. A tanuló minta esetében azért, hogy minél több adat álljon rendelkezésre a modellezéshez, a tesztelő minta esetén pedig amiatt, hogy a véletlen minél kevésbé befolyásolja a tesztelési eredményeket, illetve hogy minél objektívebb képet kaphassunk a modell megbízhatóságára vonatkozóan olyan megfigyelések kapcsán, amelyek nem szerepeltek a modellépítéshez használt mintában. Saját megítélésünk szerint a tanulmányban javasolt modellezési koncepció adatigénye legalább 500 megfigyelés a tanulási mintára vonatkozóan. A rendelkezésre álló adatállomány kapcsán ennek a kritériumnak csak a „Kereskedelem és gépjárműjavítás”, illetve méret tekintetében a mikrovállalkozások kategóriája felelt meg.

³⁴ Az FI. ábrán bemutattuk, hogy a CHAID-módszerrel miként bontható kategóriákra a bonítás értékkészlete. Az ábrán látható, hogy a teljes adathalmaz több mint 2000 megfigyelésből állt, így a létrejött kategóriák is több száz megfigyelést tartalmaztak. Az SPSS nevű szoftver alapbeállítása, hogy a CHAID-módszer csak abban az esetben képez kategóriákat, ha azok elemszáma az 50 darabot meghaladja. Ez több ezer megfigyelés esetén megfelelő beállítás, viszont a „Kereskedelem és gépjárműjavítás” nemzetgazdasági ágat tekintve a rendelkezésre álló megfigyelések száma jóval alacsonyabb, mint a teljes adathalmaz esetén. A kategóriák létrehozásához szükséges elemszámot ezért célszerű volt csökkenteni annak érdekében, hogy az egyes változók értékkészletén belül két kategóriánál több jöjjön létre. Számításaink során a kategóriák képzéséhez szükséges minimális elemszámot 5 darabnál rögzítettük. Választásunk – objektív megkötések hiányában – saját tapasztalatainkon alapult.

mint a teljes adathalmaz esetén. A ROC-görbe alatti területek értékét a tíz tesztelő minta átlagában a 6. táblázat mutatja.

6. táblázat

A robusztussági vizsgálatok eredményei

Csoport	$t-2$. év adatai	$t-3$. év adatai	$t-4$. év adatai	Folyamatváltozók
Kereskedelem, gépjárműjavítás	0,863	0,804	0,764	0,881
Mikrovállalkozások	0,871	0,815	0,739	0,881

A számítások eredményei mindkét alminta esetén megerősítették a teljes adathalmaz alapján levont következtetést, miszerint a 3.3. alfejezetben bemutatott folyamat jellegű változók alkalmazásával javítható a csődelőrejelző modellek előrejelző képessége a csak statikus pénzügyi mutatókra épített modellekkel szemben.

Ugyan a megfelelő statisztikai teszt eredménye szerint a folyamatváltozók a vállalatok jövőbeli csődje, illetve felszámolása szempontjából szignifikáns információ-többletet hordoznak a statikus pénzügyi mutatókhoz képest, a bemutatott 1-2 százalékpontos növekmény a statisztikai szignifikancia ellenére is kevésnek tűnhet az Olvasó számára. Fontos azonban kiemelni, hogy a dinamizálási koncepciónak része a Nyitrai [2014] által javasolt mutatók alkalmazása is, amely a pénzügyi adatok tárgyévi értékeit az azt megelőző időszak historikus minimumához és maximumához viszonyítja. Kérdés, hogy ennek mekkora szerepe volt a vizsgált adatbázis esetén a 3.3. alfejezetben kifejtett dinamizálási koncepció eredményessége szempontjából. Ennek érdekében a teljes adathalmazon elvégzett számításokat megismételtük a Nyitrai [2014] által javasolt változók nélkül a statikus pénzügyi mutatókra, illetve a folyamatjellegű változók alkalmazásával is. A számítások minden egyéb tekintetben megegyeztek az előző fejezetekben elmondottakkal. Az eredményeket a 7. táblázat mutatja.

7. táblázat

A dinamizálási koncepció hozzáadott értéke

Változó	Átlag	Szórás
Statikus mutatók (D)	0,877	0,022
Statikus mutatók	0,856	0,034
Folyamatváltozók (D)	0,887	0,020
Folyamatváltozók	0,867	0,027

Megjegyzés. A ROC-görbe alatti terület aránya a tízszeres keresztvalidáció átlagában. (D) jelöli a Nyitrai [2014] által javasolt mutatók alkalmazását.

A számítások eredményei arra utalnak, hogy a pénzügyi mutatók értékeit célszerű azok historikus minimumának és maximumának tükrében megítélni, mivel mindkét vizsgált esetben 2 százalékponttal nőtt a modellek előrejelző képessége ahhoz képest, mint amikor nem szerepeltek a magyarázó változók között a Nyitrai [2014] által „dinamikus pénzügyi mutatóknak” nevezett arányszámok. A tanulmány 3.3. alfejezetében bemutatott CHAID-alapú adatbányászati megközelítés ezt további 1 százalékponttal növelte, azaz összességében 3 százalékponttal nagyobb előrejelző teljesítmény mutatkozott a vizsgált adatbázis esetén a bemutatott komplex dinamizálási koncepció alkalmazásával, mint amikor a modelleket csupán a pénzügyi mutatók statikus értékeire alapoztuk. Kiemelendő az is, hogy a tanulmányban bemutatott empirikus eredmények mindegyikére érvényes, hogy a dinamizálás a vizsgált adatbázis esetén nagyobb és stabilabb³⁵ teljesítményű csődmodellek felállítását tette lehetővé, mint az ahhoz viszonyított statikus megközelítések.

A 2. fejezetben említettük, hogy a vállalkozásokra vonatkozó pénzügyi adatok meglehetősen széles időperiódusból (2007–2016) származnak. Ebből adódóan joggal merülhet fel az Olvasóban, hogy a bemutatott dinamizálási koncepció hatékonysága mennyiben tulajdonítható az időbeli heterogenitásnak. Formálisan: azt kívántuk megvizsgálni, hogy miként módosulnak a korábban látott eredmények abban az esetben, ha a modellépítés alapját képező adatbázisban a legutoljára megfigyelt év ($t - 2$) egyúttal ugyanazt a naptári évet is jelenti; ugyanis a korábban látott elemzésekben az idősor záró tagja a 2007 és 2016 közötti periódus bármely évéből származhatott, amely befolyásolhatta az eredményeket.

A kérdés vizsgálata céljából a teljes adatkör azon részhalmazát használtuk fel, ahol a legnagyobb számban álltak rendelkezésre olyan megfigyelések, amelyek egyazon üzleti évben váltak fizetéképtelenné. Az adatbázisban szereplő csőd-, illetve felszámolási eljárás alá került vállalkozások közel egynegyede (486 darab) 2011-ben vált fizetéképtelenné. Ezt követően a rendelkezésre álló adathalmazból véletlenszerűen – a másik csoporttal azonos számosságban – kiválasztottuk a fizetőképes megfigyelések egy olyan részhalmazát, amelyek 2011-ben nem voltak fizetéképtelenek. Ebben az esetben mindkét csoportban az idősor utolsó – $t - 2$. – tagja a 2009. naptári év pénzügyi teljesítményét mutatta, melyhez az azt megelőző időszakot használtuk fel az időbeli tendenciát megragadó /1/-es formula számítása során. Az empirikus vizsgálat módja minden szempontból megegyezett a korábban bemutatott koncepcióval. Az eredményeket a 8. táblázat foglalja össze.

A 8. táblázatban „statikus modell” alatt azokat értjük, amelyek esetén csak az idősor utolsó ($t - 2$. évből származó statikus) adatai szerepeltek magyarázó változóként. „Folyamatmodellek” esetén azon modellek teljesítményét jelenítettük meg, amelyekben a magyarázó változók a tanulmányban javasolt dinamizálási koncepció segítségével készültek. Az elsődleges elemzésnek megfelelően ebben az esetben is tízszeres

³⁵ A teljesítmény tekintetében alacsonyabb szóródású.

arányos keresztvalidációt alkalmaztunk. Az eredmények megerősítik azt a korábban levont következtetést, miszerint a folyamatszempléletű mutatók segítségével növelhető a modellek előrejelző teljesítménye a statikus modellekhez képest. Az eredmények azonban arra is rámutatnak, hogy a vizsgált adatbázis esetén az időbeli heterogenitás a bemutatott dinamizálási koncepció hozzáadott értékét negatívan befolyásolta, mivel átlagosan 13 százalékponttal kaptunk jobb eredményt a folyamatszempléletű mutatókkal, mint a statikus pénzügyi adatokkal. Ez a különbség csak 1 százalékpont volt a tanulmány elsődleges elemzésében. Fontos azonban kiemelni, hogy ezt a robusztussági vizsgálatot csak jóval szűkebb adatkörön volt lehetőségünk elvégezni. A megfigyelések számának növelésével a későbbiekben további vizsgálatot igényel, hogy ez a tendencia más évek, illetve nagyobb elemszámú mintára is érvényes-e. Jelen tanulmány, valamint a rendelkezésre álló adatkör esetén azonban egyértelműen kimondható, hogy ez a robusztussági vizsgálat is megerősíti az elsődleges elemzésből levonható következtetést.

8. táblázat

*Az empirikus vizsgálatok eredményei
azonos záróév alkalmazásával*

Mintafelosztás	Statikus modell	Folyamatmodell
1	0,788	0,945
2	0,794	0,970
3	0,873	0,964
4	0,749	0,950
5	0,839	0,966
6	0,851	0,948
7	0,881	0,967
8	0,831	0,936
9	0,787	0,921
10	0,824	0,967
Átlag	0,822	0,953
Szórás	0,040	0,015

5. Összegzés

A tanulmány az utóbbi évek magyar és nemzetközi szakirodalmának széles körét alapul véve tett kísérletet a csödelőrejelző modellek teljesítményének javítására. A magyarázó változók körében felhasznált pénzügyi-számviteli mutatószámok ér-

tékkészletét a *Hámori* [2016] által javasolt CHAID-módszerrel kategorizáltuk. Az idézett szerző az egyes kategóriákat dummy változóként javasolta figyelembe venni a modellek független változói között. Azonban ez, a kategóriák, illetve a lehetséges változók nagy száma esetén, rendkívüli mértékben növeli a változószámot, ezért jelen tanulmány *Hámori* [2016] megközelítését annyiban módosította, hogy az egyes kategóriák sorszámait használta fel magyarázó változóként – szükség esetén átrendezve a kategóriákat annak érdekében, hogy azok sorszámai az egyes csoportokon belül a fizetéseképtelenség kockázatának monoton függvényei legyenek.

Az így képzett kategorizálás lehetővé teszi a pénzügyi mutatók időbeli változásának számszerű megjelenítését is. Az egyes évek kategória-sorszámából kísérleti jelleggel háromjegyű változókat képeztünk, majd azokat ismét a CHAID-módszer szerint kategorizálva hoztuk létre új lehetséges magyarázó változókat, amelyek már a magyarázó változók hároméves idősorának információtartalmát sűrítik egyetlen értékbe. Felhasználtuk a *Nyitrai* [2014] által javasolt „dinamikus pénzügyi mutatókat” is annak érdekében, hogy az egyes idősorok historikus minimuma és maximuma is megjelenjen viszonyítási alapként a modellekben.

A csődelőrejelzés területén elsőként megjelent többváltozós módszert, a lineáris diszkriminanciaanalízist alkalmaztuk több mint kétezer magyar vállalkozás pénzügyi adatait tartalmazó mintán. A vizsgálat célja az volt, hogy a 3.3. alfejezetben részletesen bemutatott dinamizálási koncepcióval létrehozott mutatók alkalmazásával kiderítsük, javítható-e a modellek előrejelző teljesítménye a pusztán statikus változókra épített modellekkel szemben. A ROC-görbe alatti terület tekintetében a számítások eredményei a rendelkezésre álló adathalmaz esetén statisztikailag szignifikáns teljesítménynövekedést mutattak, ami arra utal, hogy a statikus pénzügyi mutatók információtartalma a vállalati fizetéseképtelenség, illetve felszámolás előre jelezhetősége szempontjából növelhető a korábbi évek adataival a tanulmányban bemutatott dinamizálási koncepció keretei között. A következtetések megalapozottságát az adatbázis két legnagyobb részhalmozán (kereskedelmi tevékenység, illetve mikrovállalkozások) elvégzett számítások is megerősítették.

Az eredmények mellett fontos figyelmet fordítani a bemutatott empirikus vizsgálat korlátjaira is, amelyek további kutatási lehetőségeket vetnek fel:

– A számítások csak egyetlen, a magyar vállalkozások pénzügyi adatait tartalmazó minta esetén igazolták a bemutatott megközelítés hatékonyságát. A levont következtetések robusztusságának valódi próbája az lenne, ha más országok adatain, illetve magyar vállalkozások nagyobb elemszámú mintáin is hasonló eredmények adódnának. A kutatás ilyen irányú kiterjesztése azonban akadályba ütközik, mert jelenleg nem állnak rendelkezésre olyan adatbázisok, amelyek több időszakra vonatkozóan tartalmazzák fizetőképés és fizetéseképtelen társas vállalkozások pénzügyi adatait.

– A bemutatott empirikus vizsgálat megismételhető más klasszifikációs módszerek – például logisztikus regresszió, neurális hálók, metamódszerek (*Virág–Nyitrai* [2014]) – alkalmazásával is.

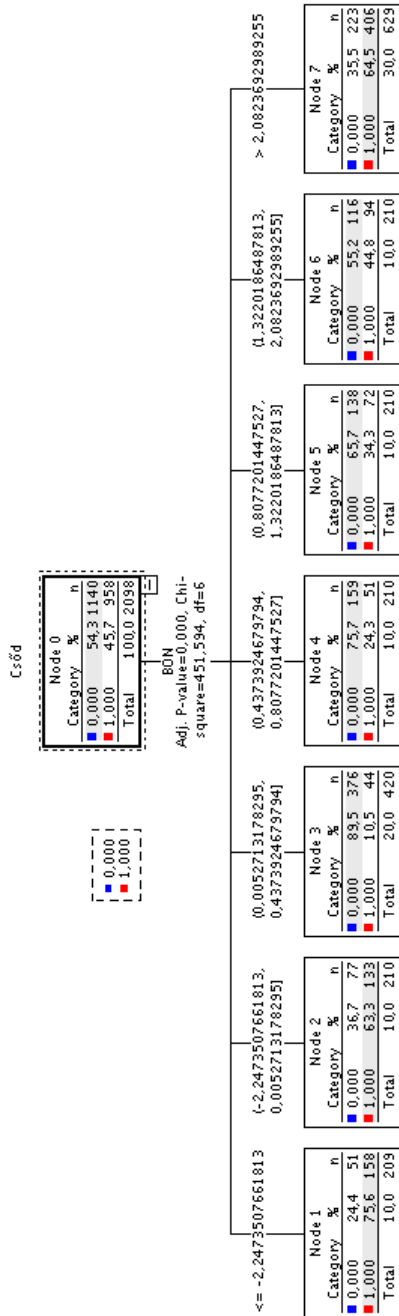
– Bármely dinamizálási koncepció jelentős adatigényt támaszt a modellezéssel szemben, mivel azonos vállalkozásokra vonatkozóan több egymást követő év adatára is szükség van. Induló, illetve fiatal vállalkozások esetén ez nem teljesíthető, így ezen csoportok vonatkozásában továbbra is fontos szerepe marad a modellek teljesítményének módszertani oldalról történő javításának, illetve a rendelkezésre álló információkör bővítésének.

– A bemutatott adatbányászati megközelítés nemcsak ezen a tudományterületen használható, hanem bárhol, ahol a magyarázó tényezők időbeli tendenciája is magyarázó erejű. A jövőben a tanulmányban javasolt koncepció hatékonyságát célszerű lenne vizsgálni más kutatási kérdések kapcsán is.

– Nemzetközi kutatások eredményei arra utalnak, hogy a csődelőrejelző modellek teljesítménye az idő múlásával gyengül (*Platt–Platt* [1990]). Az idézett szerzők azonban kutatásaikat a pénzügyi mutatók statikus értékeinek használatával végezték. Jövőbeli empirikus vizsgálatok tárgyát képezheti, hogy miként változik a tanulmányban javasolt dinamizálási koncepció segítségével létrehozott modellek előrejelző képessége az idő múlásával.

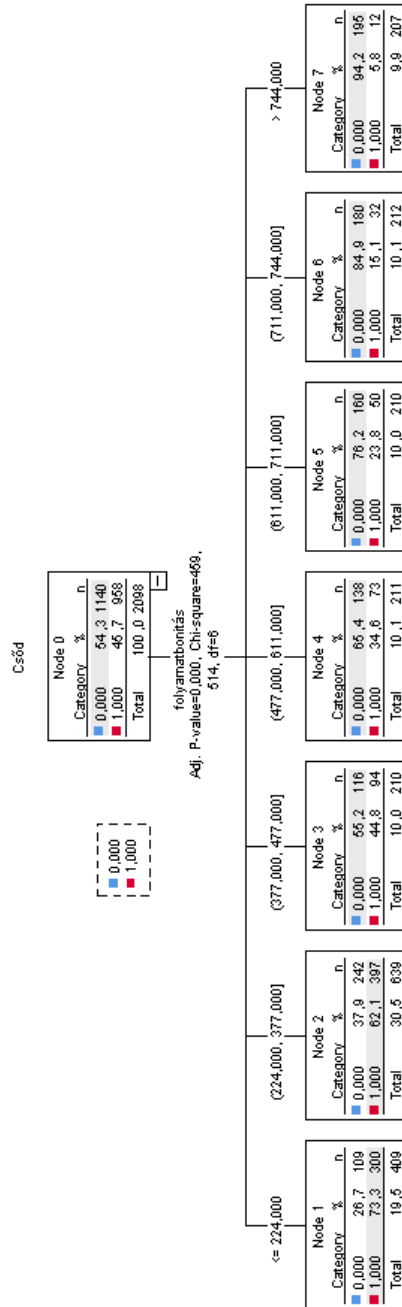
Függelék

F1. ábra. A bonítás értékesítlének kategorizálása CHAID-módszerrel



Megjegyzés: Itt és a következő ábránál a számítások az SPSS Statistics 25 szoftverrel készültek, az alapbeállítások alkalmazásával.

F2. ábra. A bonítás folyamathonításának kategorizálása



Irodalom

- ALAKA, H. A. – OYEDELE, L. O. – OWOLABI, H. A. – KUMAR, V. – AJAYI, S. O. – AKINADE, O. O. – BILAL, M. [2018]: Systematic review of bankruptcy prediction models: towards a framework for tool selection. *Expert Systems with Applications*. Vol. 94. March. pp. 164–184. <http://dx.doi.org/10.1016/j.eswa.2017.10.040>
- CHEN, N. – RIBEIRO, B. – CHEN, A. [2016]: Financial credit risk assessment: a recent review. *Artificial Intelligence Review*. Vol. 45. Issue 1. pp. 1–23. <http://dx.doi.org/10.1007/s10462-015-9434-x>
- CIAMPI, F. – CILLO, V. – FIANO, F. [2018]: Combining Kohonen maps and prior payment behavior for small enterprise default prediction. *Small Business Economics*. <https://doi.org/10.1007/s11187-018-0117-2>
- DE ANDRÉS, J. – LORCA, P. – DE COS JUEZ, F. J. – SÁNCHEZ-LASHERAS, F. [2011]: Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS). *Expert Systems with Applications*. Vol. 38. Issue 1. pp. 1866–1875. <https://doi.org/10.1016/j.eswa.2010.07.117>
- HABIB, A. – D’COSTA, M. – HUANG, H. J. – BHUIYAN, M. B. U. – SUN, L. [2018]: Determinants and consequences of financial distress: review of the empirical literature. *Accounting & Finance*. <http://dx.doi.org/10.1111/acfi.12400>
- HAJDU O. [2003]: Többváltozós statisztikai számítások. In: Hunyadi L. (szerk.): *Statisztikai módszerek a társadalmi és gazdasági elemzésekben*. Központi Statisztikai Hivatal. Budapest.
- HÁMORI G. [2001]: A CHAID alapú döntési fák jellemzői. *Statisztikai Szemle*. 79. évf. 8. sz. 703–710. old.
- HÁMORI G. [2016]: A magyarázóváltozók kezelésének egyes kérdései regressziós modellezés során. *Statisztikai Szemle*. 94. évf. 1. sz. 5–21. old.
- HUNYADI L. – VITA L. [2004]: Statisztika közgazdászoknak. In: Hunyadi L. (szerk.): *Statisztikai módszerek a társadalmi és gazdasági elemzésekben*. Harmadik kiadás. Központi Statisztikai Hivatal. Budapest.
- JEONG, C. – MIN, H. J. – KIM, M. S. [2012]: A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*. Vol. 39. Issue 3. pp. 3650–3658. <http://dx.doi.org/10.1016/j.eswa.2011.09.056>
- KIRKOS, E. [2015]: Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review*. Vol. 43. Issue 1. pp. 83–123. <http://dx.doi.org/10.1007/s10462-012-9367-6>
- KRISTÓF, T. – VIRÁG, M. [2012]: Data reduction and univariate splitting – Do they together provide better corporate bankruptcy prediction? *Acta Oeconomica*. Vol. 62. Issue 2. pp. 205–228. <http://dx.doi.org/10.1556/AOecon.62.2012.2.4>
- KRISTÓF T. [2005]: A csodelőrejelzés sokváltozós statisztikai módszerei és empirikus vizsgálata. *Statisztikai Szemle*. 83. évf. 9. sz. 841–863. old.
- KRISTÓF T. [2018]: A case-based reasoning alkalmazása a hazai mikrovállalkozások csodelőrejelzésére. *Statisztikai Szemle*. 96. évf. 11–12. sz. 1109–1128. old. <https://doi.org/10.20311/stat2018.11-12.hu1109>

- NYITRAI, T. – VIRÁG, M. [2018]: The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*. August. <http://dx.doi.org/10.1016/j.seps.2018.08.004>
- NYITRAI T. [2014]: Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében. *Statisztikai Szemle*. 92. évf. 4. sz. 357–377. old.
- NYITRAI T. [2015]: Hazai vállalkozások csődjének előrejelzése a csődeseményt megelőző egy, két, illetve három évvel korábbi pénzügyi beszámolók adatai alapján. *Vezetéstudomány*. 46. évf. 5. sz. 55–65. old.
- ONAY, C. – ÖZTÜRK, E. [2018]: A review of credit scoring research in the age of big data. *Journal of Financial Regulation and Compliance*. Vol. 26. Issue 3. pp. 382–405. <http://dx.doi.org/10.1108/JFRC-06-2017-0054>
- 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. Issue 1. pp. 31–51. <http://dx.doi.org/10.1111/j.1468-5957.1990.tb00548.x>
- RIBEIRO, B. – SILVA, C. – CHEN, N. – VIEIRA, N. – DAS NEVES, J. C. [2012]: Enhanced default risk models with SVM+. *Expert Systems with Applications*. Vol. 39. Issue 11. pp. 10140–10152. <http://dx.doi.org/10.1016/j.eswa.2012.02.142>
- SARIEV, E. – GERMANO, G. [2012]: An innovative feature selection method for support vector machines and its test on the estimation of the credit risk of default. *Review of Financial Economics*. November. <https://doi.org/10.1002/rfe.1049>
- SERRANO-CINCA, C. – GUTIÉRREZ-NIETO, B. – BERNATE-VALBUENA, M. [2018]: The use of accounting anomalies indicators to predict business failure. *European Management Journal*. Vol. 37. Issue 3. pp. 353–375. <http://dx.doi.org/10.1016/j.emj.2018.10.006>
- SUN, J. – LI, H. – HUANG, Q. H. – HE, K. Y. [2014]: Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling and featuring approaches. *Knowledge-Based Systems*. Vol. 57. February. pp. 41–56. <http://dx.doi.org/10.1016/j.knsys.2013.12.006>
- VIRÁG M. – KRISTÓF T. – FIÁTH A. – VARSÁNYI, J. [2013]: *Pénzügyi elemzés, csődelőrejelzés, válságkezelés*. Kossuth Kiadó. Budapest.
- VIRÁG M. – NYITRAI T. [2014]: Metamódszerek alkalmazása a csődelőrejelzésben. *Hitelintézeti Szemle*. 13. évf. 4. sz. 180–195. old.
- VOLKOV A. – BENOIT D. F. – VAN DEN POEL D. [2017]: Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems*. Vol. 98. June. pp. 59–68. <https://doi.org/10.1016/j.dss.2017.04.008>

Summary

The study presents a multi-step CHAID-based data mining approach in the field of bankruptcy prediction. The procedure first categorizes the range of possible explanatory variables and then it examines the evolution of independent variables within these categories with the lapse of time. The novelty of the study lies in creating a new measure from the number of categories of data coming

from subsequent years. The proposed new measure takes into account the values of each year with decreasing weight; that is, it gives higher (smaller) weight to newer (older) data. The study uses the so-obtained numbers as explanatory variables instead of the original data. The empirical results suggest that this novel approach makes the development of more accurate bankruptcy prediction models possible. This assumption is examined via robustness checks which support the main conclusion of the study with respect to the available data. The proposed methodology may be applied in solving any classification problem where the evolution of the explanatory variables may have predictive power but the previous years' values have smaller power as the time goes by.