

Kiugró értékek kezelése logisztikus regresszió alapuló csődelőrejelzési modellek esetén

Szántó Tünde Katalin¹

Összefoglalás

A banki nemfizetési kockázatok kezelésének elsődleges eszköze a potenciális ügyfelek hitelminősítése. Jelen tanulmány fókuszában a hitelezői scorecardok 95%-ának elkészítésekor is alkalmazott logisztikus regresszió módszere áll. A kutatás célja annak meghatározása, hogy a kiugró adatokra rendkívül érzékeny módszer alkalmazása során a kiugró értékek kezelése mennyire javítja a modellek besorolási pontosságát, valamint, hogy a kiugró értékek kezelésének mely módszere eredményezi a legmagasabb besorolási pontosságot. Továbbá, hogy egy olyan minta esetében, amely nem egyenlő arányban tartalmaz fizetőképes és fizetésképtelen vállalkozásokat, milyen szempontok alapján kell meghatározni a modellek cut-off értékét. Az elemzés egy 1677 építőipari vállalkozásból álló mintán került lefolytatásra. A kapott eredmények alapján a kiugró értékek kezelése jelentősen javítja a modellek előrejelző-képességét, a kiugró értékek kezelésére a kiugró értékek helyettesítése a hozzájuk legközelebb eső, nem kiugró értékkel bizonyult a leghatékonyabbnak. A cut-off meghatározása esetén a legmagasabb besorolási pontosságot eredményező érték használata nem megfelelő, hiszen ez az elsőfajú hibák arányának megnövekedéséhez vezethet. Ennek az értéknek az optimalizációja attól függhet, hogy az adott pénzügyi intézet milyen mértékű hitelezési kockázatot vállal a kihelyezett hiteleinek portfóliójában.

JEL-KÓDOK: G33, C6, G17

KULCSSZAVAK: csődelőrejelzés, nemfizetési kockázat, logisztikus regresszió, kiugró értékek kezelése

DOI: https://doi.org/10.35551/PFQ_2023_3_5

¹ PhD-hallgató, Szegedi Tudományegyetem, Gazdaságtudományi Kar,
szanto.tunde.katalin@o365.u-szeged.hu

Bevezetés

Az utóbbi évtizedekben egyre inkább megnőtt a vállalati csődelőrejelzés fontossága, a japán és skandináv bankrendszerekben a 2000-es évek elején bekövetkező banki krízisek legfőbb oka ugyanis a hitelezett vállalkozások csődje volt, amely rávilágított arra, hogy mennyire fontos az ügyfelek túlélőképességének vizsgálata hitelezés esetén. A csődelőrejelzési modellek legfontosabb alkalmazói tehát bankok, de alkalmazásuk hasznos lehet könyvvizsgáló cégek, illetve kötvényértékesítő cégek számára is (Virág 2004). Jelenleg a koronavírus-járvány utóhatására, valamint a háborús helyzet következményeként az egész világgazdaságot nagy bizonytalanság jellemzi, ami még fontosabbá teszi a vállalkozások hatékony csődelőrejelzését. Banki szempontból különösen fontos a hatékony hitelminősítés, hiszen a koronavírusjárvány kapcsán a bankok működési költségei jelentős mértékben megnöttek (Doma–Kozma 2022).

Magyarországon 2022 harmadik negyedében rekordszámú felszámolási eljárás indult, a szeptemberben indított felszámolások száma négyszerese az egy évvel korábbi értéknek (Opten 2022).² Tanulmányunk témája magyarországi építőipari vállalkozások logisztikus regresszió alapuló csődelőrejelzése. Választásunk oka, hogy az építőipar a magyar nemzetgazdaság egy meghatározó iparága, mind a bruttó hozzáadott érték tekintetében, mind a vállalkozások számában növekedés tapasztalható az iparágban (KSH 2021/a). A logisztikus regresszió csődelőrejelzés területén való alkalmazása már az 1970-es években elterjedt, a banki hitelezői scorecardok 95%-a azonban még mindig ezen a módszeren alapul, a bankok tehát előszeretettel használják hitelezési döntéseik meghozatala során a módszert (Nyitrai–Virág 2017). A logisztikus regresszió egy rendkívül népszerű technika a csődelőrejelzés területén, ugyanis nem követeli meg a változók normális eloszlását, ugyanakkor nagyon érzékeny a kiugró adatokra, amely főként a csődbe kerülő vállalkozásokra jellemző, amiatt ezen értékek kezelése mindenképpen szükséges modellépítés során (Nyitrai 2017). A kiugró értékek kezelésének két legelterjedtebb gyakorlata a kiugró értékek helyettesítése a hozzájuk legközelebb eső, de már nem kiugró értékkel, valamint a kiugró adatokat mutató vállalkozások mintából való kizárása (Nyitrai–Virág 2017, Nyitrai 2017).

A vizsgálat alapjául szolgáló minta egy nagy elemszámú, 1677 építőipari vállalkozásból álló adatsor, a mintában szereplő vállalkozások közül 1293 cég még működő, 384 pedig olyan vállalkozás, amely ellen jelenleg is felszámolási eljárás zajlik. A beszámoló forrása a Crefoport adatbázis. A mintában nem egyenlő arányban szerepelnek a működő és fizetéképtelen vállalkozások, amely összhangban áll azzal, hogy a valóságban is jóval nagyobb arányt képviselnek a jól működő, fizetőképes vállalkozások a gazdaságban.

2 Forrás: <https://www.opten.hu/kozlemenyek/meglodultak-a-felszamolasi-eljarasok-fizeteskptelenseget-jelez-vagy-covid-utohatas> Letöltés dátuma: 2023.03.05.

A CSÖDELŐREJELZÉS TÖRTÉNETE

Csőd alatt jogi értelemben fizetési képtelenséget értünk, a csőd azt az eseményt jelöli, amikor a vállalkozás nem képes határidőre eleget tenni fizetési kötelezettségének. A csőd azonban nem egy hirtelen fellépő helyzet, hanem egy hosszabb folyamat, a pénzügyi nehézségek időszakának egy lehetséges kimenetele. Pénzügyi nehézségek akkor lépnek fel a vállalkozás életében, amikor értékrombolás alakul ki az eszközök nem megfelelő hatékonyságából, vagy a rosszul kialakított eszközportfólióból adódóan. Az értékrombolás miatt csökken az eszközportfólió piaci értéke, ez pedig megnöveli a vállalkozás finanszírozási áttételét. Ezek a hatások likviditási problémák kialakulásához vezetnek, majd a folyamat eredményeként fizetési képtelen állapot alakulhat ki, ami jogi értelemben véve csődöt jelent (Pálinkó–Svoób 2016).

A pénzügyi nehézségek és a csőd azonban nem feltétlenül jelentkeznek együtt. Elképzelhető, hogy a vezetőség még időben felismeri az értékrombolást és a cég működésének átszervezésével sikeresen elejét veszi a fizetési képtelen állapot kialakulásának. Az is előfordulhat azonban, hogy valamilyen külső sokk hatására minden előjel nélkül alakul ki a csőd. Ez leggyakrabban olyan vállalkozások esetében fordul elő, amelyek alacsony tőke- és eszközállománnyal rendelkeznek, valamint működésüket rövid lejáratú forrásokból finanszírozzák. Recessziós időszakban a rövid lejáratú források megújítása nehezebbé válik, emiatt hirtelen likviditási hiány alakulhat ki és beállhat a fizetési képtelen állapot (Pálinkó–Svoób 2016).

Magyarországon kétféle eljárás létezik fizetési képtelenség esetén. A csődeljárás egy reorganizációs típusú eljárás, célja, hogy az adós a hitelezőivel való megegyezés után újrászervezze működését és ezt követően is gazdálkodást folytasson. A felszámolási eljárás ezzel szemben egy likvidációs típusú eljárás, amelynek célja nem az adós fél működésének hatékony újrászervezése, hanem az adós vállalkozás jogutód nélküli megszüntetése, amelynek során a hitelezők követeléseinek minél teljesebb körben való kielégítésére törekszenek (Piller 2013).

A legtöbb bank a hitelezési gyakorlatában alkalmaz valamilyen statisztikai alapú minősítési rendszert (scoring rendszert), amely a csődbe kerülés valószínűségét határozza meg. Ezek a minősítési rendszerek sokféle inputtal dolgozhatnak. Az általános tapasztalat azt mutatja, hogy azok a változók működnek a legjobban a csődbe kerülés valószínűségének meghatározásakor, amelyek a hitelt felvevők viselkedését írják le, ezen változók magasabb pontosságot eredményeznek, mint például a pénzügyi mutatószámok alkalmazása. Továbbá a számviteli adatok gyengeségeként jegyezhető meg, hogy a mérleg- és eredménykimutatás adatok bizonyos körülmények között manipulálhatók (Cziglerné 2020). A probléma azonban az, hogy nem léteznek olyan adatbázisok, amelyek segítségével a hitelfelvevők viselkedése megfigyelhető lenne (Mikolasek 2018). A csődelőrejelzés gyakorlatában emiatt általános a pénzügyi mutatószámok használata, hiszen a beszámoló adatai bárki számára nyilvánosan elérhetők. A scoring rendszerek objektív tényezőkön alapulnak és vizsgálatuk kiterjed a vállalkozások működésének egész területére, ezáltal átfogó képet ad a gazdálkodásukról (Zéman–Hegedűs–Molnár 2018).

A XX. század utolsó harmada előtt még nem állt rendelkezésre megfelelő informatikai háttér, valamint olyan fejlett statisztikai módszerek, amelyek pontos csődelőrejelzési modellek felállítását tették volna lehetővé, ennek ellenére már ekkor is próbáltak olyan módszereket találni, amelyek segítségével egy vállalat jövőjét előre lehetett vetíteni. Ekkoriban már csődbe került vállalkozások különböző pénzügyi mutatószámait próbálták összehasonlítani és ezek alapján következtetéseket levonni a vállalatok jövőjével kapcsolatban (Virág–Kristóf 2006). 1930-ban például a Bureau of Business Research 29 cég 24-féle pénzügyi mutatószámát hasonlította össze annak érdekében, hogy meghatározza, milyen hasonlóságok figyelhetők meg a csődbe került vállalkozások esetében (Bellovary–Giacomino–Akers 2007).

Az első modern csődelőrejelzési modellt Beaver alkotta meg 1960-ban. Munkája egyváltozós diszkriminancia-analízisre épül. A módszer lényege, hogy egy pénzügyi mutatószám megvizsgálásával dönti el az adott vállalatról, hogy a fizetésképtelen vagy a túlélő kategóriába sorolható-e be. Mintájába 79 fizetéképes és 79 fizetésképtelen vállalatot emelt be. Összesen 30 pénzügyi mutatószámot vizsgált meg és arra a következtetésre jutott, hogy az egészséges és fizetésképtelen vállalatok a Cash flow / Eszközök, Cash flow / Adósság és Nettó árbevétel / Adósság mutatószámok tekintetében különböznek a leginkább egymástól (Virág 2004). A legmegbízhatóbb eredményt a Cash flow / Összes eszköz arány mutatójával tudta elérni, amely segítségével 90%-os pontossággal tudta a csődöt előre jelezni egy évvel a fizetésképtelenség bekövetkezése előtt (Virág–Kristóf 2006). Az egyváltozós diszkriminancia-analízis hátránya azonban, hogy gyakran ellentmondásos eredményre vezet, ilyenkor az egyik pénzügyi mutatószám alapján túlélőként ítélt meg a vállalat, egy másik alapján azonban fizetésképtelenként (Virág 2004).

Edward I. Altman 1968-ban publikálta modelljét, amely már többváltozós diszkriminancia-analízisre épül. Az általa használt minta 33 fizetésképtelen és 33 fizetéképes vállalatból állt, ezek a vállalatok kis és közepes méretűek voltak, ebben az időben ugyanis nagy vállalatok csak ritkán kerültek csődbe. Összesen 22 pénzügyi mutatószámot vizsgált meg, amelyekből végül ötöt emelt modelljébe, ami egy lineáris függvényanalízis, az öt változót objektív arányszámokkal súlyozva azok összege egy „Z” értéket ad meg. „Z” értékét egy meghatározott cut-off ponthoz viszonyítva meghatározható, hogy a vállalat a túlélő vagy csődös kategóriába sorolandó (Virág 2004).

Bár a többváltozós diszkriminancia-analízis használata úttörőnek bizonyult, alkalmazása során problémát jelent, hogy használatának feltétele, hogy a változók statisztikailag függetlenek legyenek, pénzügyi mutatószámok között azonban gyakran multikollinearitás áll fenn, ami ezt a kitéltet megszegi. Ezen kívül fontos feltétel még, hogy a mutatószámok normális eloszlást kövessenek. Ezt a kitéltet oldja fel a logisztikus regresszió alapú csődelőrejelzés, amely nem igényli a változók normális eloszlását. A módszer maximum likelihood módszerével egy logisztikus regresszió-függvényt illeszt a megfigyelésekre (Virág–Kristóf 2006). Logisztikus regresszió alapuló csődmodellt először Ohlson használt csődelőrejelzésre.

Az 1980-as években kezdődött meg a probit regresszió alkalmazása a csődelőrejelzés területén, elsőként Zmijewski épített probit regresszióra alapuló modellt (Kristóf–Virág 2019). Mintájába 800 fizetőképes és 40 fizetésképtelen vállalatot vont be.

A modell három változót használ, az eszközarányos jövedelmezőséget, a kötelezettségek eszközökhöz viszonyított arányát, valamint a likviditási rátát. Az eredeti mintára vonatkoztatva kimagasló, 98%-os besorolási pontosságot ért el (Zmijewski 1984).

A döntési fákat csődelőrejelzés területén először Frydman, Altman és Kao alkalmazta 1985-ben. A döntési fák alkalmazása rendkívül népszerű, hiszen alkalmazásához nem szükséges, hogy a korábban taglalt statisztikai feltételek teljesüljenek (Kristóf–Virág 2019). Népszerű döntési fán alapuló módszer a rekurzív parcionáló algoritmus. A módszer egyváltozós elválasztással dolgozik, lépésről lépésre kétféle bontja az adatokat, így képezve a faágakat. A kiinduló adatsor egy olyan minta, amelyben előre ismert, hogy mely vállalatok tartoznak a fizetőképességű, és melyek a fizetéképtelen kategóriába. Az eljárás egyesével vizsgálja meg a változókat, a leginkább elválasztó értékkel rendelkező változók mentén építi fel a fát annak érdekében, hogy a lehető leghomogénebb osztályokat hozza létre. Az adatokat a függő változó szempontjából csoportosítva a csoportokon belül a variancia minimalizálására, míg az egyes csoportok között a variancia maximalizálására törekszik a módszer (Virág–Kristóf 2006).

Egy másik népszerű, döntési fán alapuló módszer a khí-négyzet-alapú automatikus interakció-detektálás (CHAID). Ez az eljárás intervallumokra bontja a magyarázó változó értékkészletét, majd páronként vizsgálva az osztályközöket eldönti, hogy az osztályközök és a bennük szereplő vállalatok beosztása (csődös vagy fizetőképességű) függetlenek-e egymástól, ha függetlenek, a két osztályközt egységesíti. A folyamat addig folytatódik, ameddig csak olyan osztályközök maradnak, amelyek nem függetlenek statisztikailag. A folyamat eredményeként a magyarázó változó értékkészlete osztályközökre bomlik fel (Nyitrai 2017).

Az 1990-es években került előtérbe mesterséges intelligencia, ezen belül a neurális hálók használata a csődelőrejelzésben. A neurális hálók párhuzamos, osztott működésre képes információ-feldolgozó eszközök, amelyek hardver vagy szoftver alapon működnek. A hálók több, egymáshoz kapcsolódó neuronból állnak és szemben a korábban tárgyalt módszerekkel, tanulással szerzik meg a képességüket. A neuronok összekapcsolódásának módja minden háló esetében különböző. A neuronok három fő rétegből állnak, a bemeneti rétegből, a köztes rétegből és a kimeneti rétegből. A neurális hálók példákön keresztül tanulnak, a kellőképpen megedzett hálók pedig már használhatók előrejelzésre más adatokon is (Kristóf 2005). Ezt a módszert először Odom és Sharda használták vállalati fizetéképtelenség előrejelzésére. Háromrétegű neurális hálóra épülő modelljük felülmúlta a korábbi módszerekkel felépített modellek besorolási pontosságát. A neurális hálókra épülő modellek azóta is folyamatosan fejlődtek és napjainkban is népszerű módszert jelentenek (Kristóf–Virág 2019). Alkalmazásukkor azonban probléma lehet a túltanulás jelensége. Ez azt jelenti, hogy a hálózat a tanulás folyamán nem az általános problémát tanulja meg, hanem annak a mintának a sajátosságait, amelyen készült. Ilyen esetekben más adatbázison már nem használható hatékonyan a modell. A probléma elkerülése érdekében az adatbázist tanulási és tesztelő adatbázisokra szokás felosztani. A háló edzése, tanítása a tanulási mintán történik, majd megnézik, hogy milyen eredményeket ér el

a számára addig még ismeretlen tesztelő mintán. Ha a tesztelő mintán is hasonló a besorolási pontossága a tanulási mintán kapottéhoz, akkor a tanulás eredményesnek tekinthető (Virág–Kristóf 2006).

A 2000-es évek elejétől kezdtek el használni a neuro-fuzzy rendszereket a vállalati fizetésképtelenség előrejelzésére, amelyek besorolási pontossága felülmúlta a hagyományos neurális hálókra épülő modellekét (Kristóf–Virág 2019). Az adatbányászati módszerek folyamatos fejlődésével egyre pontosabb eredmények érhetőek el, azonban több kiritka is érte a gépi tanulásra épülő eljárásokat. Az egyik megfogalmazott kritika a feketedoboz-probléma, azaz, hogy a modellezéskor csak a modell inputjai, valamint a számítás eredménye, az outputok ismertek, az azonban nem, hogy az egyes változók milyen arányban szerepelnek a modellben. Másik probléma, hogy a változók statisztikai szignifikanciáját nem lehet tesztelni (Nyitrai 2014).

Látható tehát, hogy a csődelőrejelzés kezdete óta sokféle módszer került alkalmazásra ezen a területen, du Jardin (2010) becslései szerint összesen több, mint ötvenféle módszer használatával, valamint több, mint ötszázféle pénzügyi változó alkalmazásával publikáltak csődelőrejelzési modelleket (Nyitrai 2017). Az öt legnépszerűbb módszernek a többváltozós diszkriminanciaanalízis, a logisztikus regresszió, a neurális háló, a contingent claims analízis, valamint az egyváltozós elemzés voltak (Kiss–Kosztópulosz–Szládek 2021).

Az alkalmazható technikák magas száma miatt manapság a meglévő módszerek tökéletesítésére törekszenek a kutatások (Nyitrai 2015). Ezzel összhangban kutatásunk egyik célja annak megvizsgálása, hogy a kiugró értékek többféle kezelésének lehetősége közül melyik eredményez magasabb besorolási pontosságot.

A vizsgált minta és módszertan

A tanulmányunk alapjául szolgáló minta egy nagy elemszámú, összesen 1677 építőipari vállalkozásból álló adatsor. A mintában szereplő vállalkozások közül 1293 cég még működő, 384 pedig olyan vállalkozás, amely ellen jelenleg felszámolási eljárás zajlik, ezzel is érzékeltetve, hogy a valóságban a fizetőképes vállalkozások nagyobb arányt képviselnek a gazdaságban, mint a csődbe kerülők. Az adatok forrása a Crefoport adatbázis.

Az építőipar helyzete Magyarországon

Az építőipar három ágazatot foglal magába: a 41-es TEÁOR-kódú épületek építése, 42-es TEÁOR-kódú egyéb építmények építése, valamint 43-as TEÁOR-kódú speciális szaképítés tevékenységi körrel rendelkező vállalkozások nevezhető építőipari vállalkozásnak (KSH 2021/a).

Az építőipar a nemzetgazdaság egy meghatározó iparága. 2021-ben a három ágazat részaránya a bruttó hozzáadott értékből 6,3% volt. Az iparág jelentőségét jelzi, hogy a regisztrált vállalkozások 7%-a építőipari vállalkozás, valamint a foglalkoztatottak 8%-a dolgozik az iparágban. A koronavírus-járvány miatti gazdasági visszaesés

az építőipar területén is érzékeltette hatását, 2020-ban ugyanis a termelés volumene jelentősen csökkent a 2019-es értékhez képest, 2021-ben azonban 13%-kal növekedett a termelési volumen a korábbi évhez képest (KSH 2021/a).

Az építőipar abból a szempontból tekinthető különösen kockázatosnak, mert az iparágban nagymértékű lánctartozások figyelhetők meg (Limpek–Kosztopoulos–Balogh 2016). Emiatt egy-egy vállalkozás csődje láncreakciót indíthat el az üzleti partnerek között.

A mintába olyan jelenleg is működő vállalkozásokat emeltünk be, amelyek a vizsgált időszakban legalább 5 főt foglalkoztattak, valamint legalább 2013 óta, azaz minimum 9 éve működnek. Ennek oka, hogy a jól működő, fizetőképes vállalkozások működésük első néhány évében gyakran hasonlítanak pénzügyi szerkezetükben a csődös vállalkozásokhoz, ami a kutatás során torzító lehet (du Jardin 2010). Összesen 1293 működőképes vállalkozást emeltünk a mintába.

A mintába 384 olyan vállalkozás került be, amely ellen jelenleg felszámolási eljárás folyik. Ezek olyan vállalkozások, amelyek beszámolóit a csődöt megelőzően legalább 3 évre visszamenőleg rendelkezésre álltak.

A vizsgált mintát véletlenszerű választással tanuló és tesztelő mintára osztottuk. A tanuló minta használatával készítettük el a modelleket. Ez az adatbázis összesen 1188 működő és 339 csődbe került vállalkozást tartalmazott. A tesztelő minta segítségével a modellek működését ellenőriztük független vállalkozásokra vonatkoztatva. Ez a minta 105 túlélő és 45 csődös vállalkozást foglalt magába.

Módszertan

Tanulmányunkban a logisztikus regresszió módszerét alkalmaztuk. Ennek oka, hogy egyes becslések szerint a hitelezői scorecardok 95%-a ezen a módszeren alapszik (Nyitrai–Virág 2017). A logisztikus regresszió jól használható magyarázó változók és a bináris válaszadás valószínűsége között. Az eredményváltozó egy dummy változó, ami vállalati csődjelölés esetén a fizetőképes vagy fizetéképtelen kategóriákat jelenti. A folyamat során a maximum likelihood módszerével egy logisztikus regressziófüggvény kerül illesztésre a megfigyelésekre. A független változók súlyozásával egy 'Z' értéket kapunk, amely a vállalatok csődbe jutásának valószínűségét fejezi ki (Virág–Kristóf 2006).

A logisztikus regressziós formula a következőképp írható fel:

$$Pr(\text{fizetőképes}) = \frac{e^z}{1+e^z} = \frac{e^{\beta_0 + \sum(\beta_j Z_j)}}{1+e^{\beta_0 + \sum(\beta_j Z_j)}} \quad (\text{Virág–Kristóf 2006})$$

A módszer nagy előnye, hogy nem igényli a változók normális eloszlását, valamint az egyező kovariancia mátrixokat a két osztályban. További előny, hogy a változók közötti heteroszkedaszticitás csak kis mértékben torzítja az eredményeket, így a vizsgált változókat nem szükséges bonyolult matematikai transzformációk alá vonni (Rácz–Tóth 2021). A módszer alkalmazásakor fontos a változók számának ésszerű csökkentése. Ezt a legtöbbször a backward elimination eljárással hajtják végre. A módszer egyesével hagyja el a modellt kevésbé szignifikáns változóit. Egy-egy

változó elhagyása után mindig újraszámolja a regressziós együtthatókat és a p-értékeket, amíg csak kellően szignifikáns változók maradnak. A végső modell a kollinearitást, a szignifikanciát és a besorolási pontosságot együttesen figyelembe véve készül el. A regressziós paraméterek meghatározása után fontos lépés a cut-off érték meghatározása. Ez a függvény függő változójának az az értéke, amelyhez viszonyítva a vállalatokat eldönthető, hogy a csődös vagy a fizetőképes kategóriába sorolandók be (Virág–Kristóf 2006).

Az eljárás hátránya, hogy érzékeny a kiugró értékekre, amelyek azonban a pénzügyi mutatók körében gyakran előfordulnak, különösen csődbe került vállalkozások esetén jellemző. Emiatt fontos az adatbázis kiugró adatainak kezelése a kutatást megelőzően (Nyitrai 2017).

Egy magyarországi építőipari vállalkozások csődelőrejelzésére alkalmazható modell megalkotása

Tanulmányunk egyik célja, hogy összehasonlítsuk a kiugró adatok többféle lehetséges kezelésének besorolási pontosságra gyakorolt hatását.

A modellépítés során vizsgált változók

Modellünkbe a szakirodalomban leggyakrabban előforduló pénzügyi mutatószámokat vontuk be. Az 1. táblázat a megvizsgált mutatószámokat és kiszámításuk módját mutatja be.

A logisztikus regresszió alkalmazásakor problémát jelenthet a változók között fennálló multikollinearitás, emiatt ennek vizsgálata szükséges a modell felállítása előtt (Kristóf 2005). A változók közötti multikollinearitás kiszűrésére a varianciainflációs tényezőt (VIF) használtuk. Egy változó VIF-értéke a korrelációs mátrix inverzének megfelelő diagonálisbeli értékéből kapható meg, a mutatószám azt becsüli meg, hogy mekkora mértékben növekszik a regressziós együtthatók varianciája a multikollinearitás miatt (Vörösmarty–Dobos 2020). A szakirodalomban nincs egyetértés annak tekintetében, hogy a VIF mekkora értékétől beszélhetünk multikollinearitásról. A leggyakrabban alkalmazott határérték az 5, így a végső modellbe olyan változót nem emeltünk be, amely VIF-értéke meghaladta volna az 5-öt.

1. táblázat: A vizsgált mutatószámok és számításuk módja

	Mutató megnevezése	Számítás módja
X_1	Likviditási ráta	Forgóeszközök / Rövid lejáratú kötelezettségek
X_2	Likviditási gyorsráta	(Forgóeszközök - Készletek) / Rövid lejáratú kötelezettségek
X_3	Cash flow / Kötelezettségek	(Adózott eredmény + Értékcsökkenési leírás) / Kötelezettségek
X_4	Cash flow / Rövid lejáratú kötelezettségek	(Adózott eredmény + Értékcsökkenési leírás) / Rövid lejáratú kötelezettségek
X_5	Tőkeellátottság	(Befektetett eszközök + Készletek) / Saját tőke
X_6	Forgóeszközök aránya	Forgóeszközök / Mérlegfőösszeg
X_7	Eszközök forgási sebessége	Értékesítés nettó árbevétele / Mérlegfőösszeg
X_8	Készletek forgási sebessége	Értékesítés nettó árbevétele / Készletek
X_9	Követelések forgási ideje	Követelések / Értékesítés nettó árbevétele
X_{10}	Eladósodottság	Kötelezettségek / Mérlegfőösszeg
X_{11}	Saját tőke aránya	Saját tőke / Mérlegfőösszeg
X_{12}	Vagyonarányos nyereség	Adózott eredmény / Saját tőke
X_{13}	Bonitás	Kötelezettségek / Saját tőke
X_{14}	Árbevétel-arányos nyereség	Adózott eredmény / Értékesítés nettó árbevétele
X_{15}	Eszközarányos nyereség	Adózott eredmény / Mérlegfőösszeg
X_{16}	Követelések / Rövid lejáratú kötelezettségek	Követelések / Rövid lejáratú kötelezettségek
X_{17}	Nettó forgótőke aránya	(Forgóeszközök - Rövid lejáratú kötelezettségek) / Mérlegfőösszeg
X_{18}	Vállalat mérete	Az eszközállomány természetes alapú logaritmus
X_{19}	Befektetett eszközök hosszú lejáratú kötelezettségekkel fedezett aránya	Hosszú lejáratú kötelezettségek / Befektetett eszközök

Forrás: saját szerkesztés

Az elkészült modellek

A kutatás során összesen három modellt készítettünk el. Egy modellt a kiugró értékek kezelése nélkül készítettünk el, majd ezt követte a kiugró értékek kezelésével készült modellek összeállítása. A szakirodalomban nincs egyetértés annak tekintetében, hogy mi tekinthető kiugró (outlier) adatnak. Gyakran statisztikai hüvelykujj-szabályok alkalmazásával történik a kiugró értékek definiálása, ilyenkor azokat az értékeket tekintik kiugrónak, amelyek valamely szórássterjedelmen kívül esnek. Ezzel a megközelítéssel azonban az a probléma, hogy a kiugró adatok kezelése után a változók szórásának értéke megváltozik, így az új szórással számolva olyan mutatószámok is kiugrónak minősülnek, amelyek előtte nem. Az újonnan meghatározott

szórásokkal való ellenőrzést addig kell folytatni, ameddig már nem adódik új kiugró érték a szórás megváltozása után (Nyitrai–Virág 2017). Emiatt tanulmányunkban a kiugró értékek meghatározására az SPSS program beépített funkcióját használtuk, amely meghatározza a kiugró adatokat a mintára vonatkoztatva.

A kiugró adatok kezelésére szintén nincs egy egyöntetűen elfogadott módszer a kutatók körében. Két olyan eljárás van, amelyet a leggyakrabban alkalmaznak. Az outlier adatok kezelése történhet oly módon, hogy a kiugrónak minősülő értéket helyettesítik a hozzá legközelebbi eső, de már nem kiugró értékkel (Nyitrai–Virág 2017). Egy másik módszer a kiugró adatokat tartalmazó megfigyelések elhagyása a mintából (Nyitrai 2017). A tanulmányban mindkét módszert alkalmazzuk, majd összevetjük hatékonyságukat.

A 2. táblázat adataiból látható, hogy 1 évvel a csőd előtt a tanuló mintán a kiugró értékek törlésével kapott modell osztályozta legmagasabb besorolási pontossággal a minta elemeit. A túlélő modellek 85,73%-át, míg a csődös cégek 82,32%-át osztályozta pontosan, ezáltal a teljes mintára vonatkoztatva 85,21%-os besorolási pontosságot ért el a modell. Látható, hogy az időhorizont tágulásával a csődös céges vonatkozásában mindhárom modell egyre gyengébb előrejelző erővel bír. 3 évvel a csőd előtt a kiugró értékek kezelése nélkül kapott modell 67,42%-os, a kiugró értékek helyettesítésével kapott modell az 58,7%-os, míg a kiugró értékek törlésével kapott modell 49,75%-os pontossággal dolgozik a fizetésektelen cégek körében.

2. táblázat: Az elkészült modellek besorolási pontossága

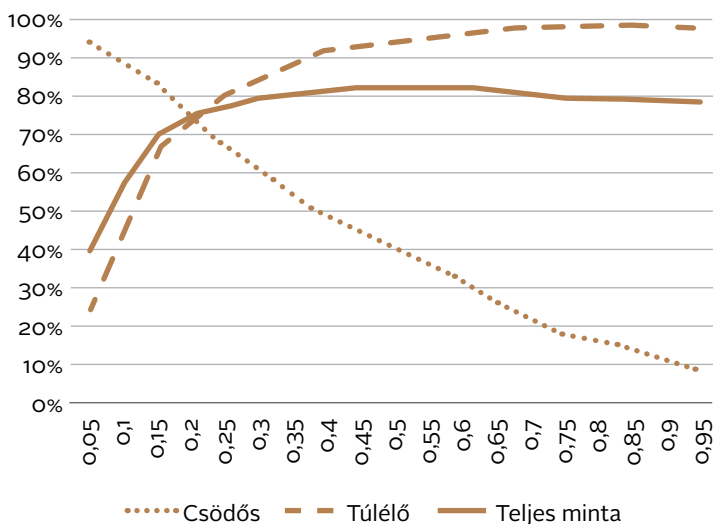
		1 évvel a csőd előtt			2 évvel a csőd előtt			3 évvel a csőd előtt		
		Helyesen besorolt, db	Tévesen besorolt, db	Pontosság, %	Helyesen besorolt, db	Tévesen besorolt, db	Pontosság, %	Helyesen besorolt, db	Tévesen besorolt, db	Pontosság, %
Kiugró értékek kezelése nélkül	túlélő	907	281	76,34	874	313	73,63	801	387	67,42
	csődös	261	78	76,99	245	95	72,06	260	79	76,70
	összesen	1168	359	76,48	1119	408	73,29	1026	461	69,48
Kiugró értékek helyettesítése	túlélő	967	221	81,40	942	245	79,36	911	277	76,68
	csődös	265	74	78,17	226	114	66,47	199	140	58,70
	összesen	1232	295	80,68	1168	359	76,49	1110	417	72,69
Kiugró értékek törlésével	túlélő	937	156	85,73	901	192	82,43	800	293	73,19
	csődös	163	35	82,32	140	58	70,71	99	100	49,75
	összesen	1100	191	85,21	1041	250	80,64	899	392	69,64

Forrás: saját szerkesztés

A cut-off értékek meghatározása a modellek esetében

A vizsgált minta nem egyenlő arányban tartalmazott fizetőképes és fizetéképtelen vállalkozásokat: a vizsgált vállalatok közül 1188 cég még működő, 339 pedig olyan vállalkozás, amely ellen jelenleg felszámolási eljárás zajlik. Ez jól tükrözi, hogy a gazdaságban is nagyobb arány képviselnek a túlélő vállalkozások, mint a csődbe jutók. Csődelőrejelzési modellek megalkotásakor általánosnak tekinthető az a gyakorlat, hogy azt a cut-off értéket rendelik a modellhez, amely a maximális besorolási pontosságot eredményezi. Ez az eljárás azonban csak abban az esetben alkalmazható, ha a vizsgált mintában 50-50%-ban szerepelnek csődbe kerülő és túlélő vállalkozások.

1. ábra: A kiugró adatok kezelése nélkül kapott modell besorolási pontossága különböző cut-off értékek mellett



Forrás: saját szerkesztés

Erre jó példa lehet a kiugró értékek kezelése nélkül kapott modell. A teljes mintára vonatkozó maximális besorolási pontosság 0,55-ös cut-off érték mellett adódna, ekkor a minta cégeinek 86%-a kerül helyes besorolásra. A modell a túlélő vállalkozások 97%-át sorolja be helyesen e mellett a cut off-érték mellett. A csődös vállalkozások körében azonban a helyesen besorolt vállalkozások aránya mindössze 37,09% (1. ábra). Az optimális cut-off értékek kiválasztásakor nem biztos, hogy az az érték lesz célravezető, amelyik a teljes besorolási pontosságot maximalizálja. Ha a teljes hibaarány minimalizálása az elsődleges cél, olyan cut-off érték kerül meghatározásra, amely a működő vállalkozásokat a legmagasabb pontossággal osztályozza, ami azonban nem veszi figyelembe a mintában alacsonyabb arányban szereplő fizetéképtelen vállalkozások besorolási pontosságát. Ez az elsőfajú hibák arányának megnövekedéséhez vezethet, tehát valójában csődbe kerülő vállalatokat túlélőként osztályoz

a modell. Mivel a csődelőrejelzési modellek elsődleges felhasználói a bankok, ez a jelenség jelentős károkat okozhat (Virág 2004). Ha a modell alapján túlélőnek adódó, valójában azonban hamarosan csődbe kerülő vállalkozásnak a bank hitelt folyósít, elveszítheti kihelyezett tőkéket és esetleges kamatbevételeit. A bank természetesen másodfajú hibák jelenlétekor (tehát amikor egy valójában túlélő vállalkozást tévesen csődösként osztályoz) is veszteségeket szenved el, hiszen profittól esik el, azonban az elsőfajú hibák magasabb aránya nagyobb zavarokat okoz a bankok működésében (Zavgren 1985).

Arra sem lehet hagyatkozni, hogy egy olyan cut-off értéket válasszunk, amely az elsőfajú hibákat minimalizálja, hiszen ebben az esetben a másodfajú hibák arány növekedhet meg. A kiugró értékek helyettesítése nélkül kapott modell esetében a csődös cégek körében 0,05-ös cut-off érték mellett érhető el a legmagasabb besorolási pontosság, 95,55%. Ebben az esetben azonban a túlélő cégek körében mindössze 25,34% a találati arány.

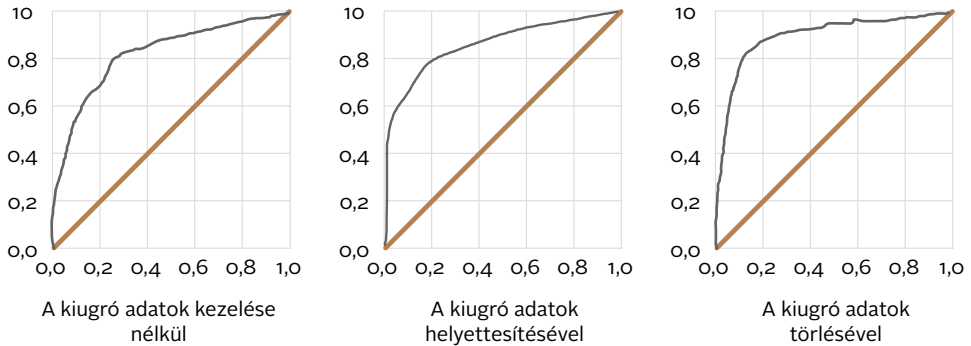
Az optimális cut-off meghatározásakor tehát egy olyan értéket kell választani, amely kellően magas besorolási pontosságot eredményez, az elsőfajú hibák alacsonyan tartása mellett. Bankok esetében ez a választás függhet attól, hogy az adott pénzügyi intézmény mekkora nemfizetési kockázatot hajlandó vállalni kihelyezett hiteleinek portfóliójában.

Az elkészült modellek összehasonlítása

Az elkészült modelleket elsőként a kumulált besorolási pontosságok görbéje (ROC) alapján hasonlítottuk össze. A ROC-görbe azt mutatja meg, hogy a modell által meghatározott besorolási értékek mennyire felelnek meg az eredeti besorolásnak különböző cut-off értékek mellett. A görbe értékét egy 45°-os egyeneshez kell hasonlítani, minél inkább elválik a görbe az egyenestől, azaz minél nagyobb a görbe alatti terület, annál pontosabban működik a modell (Virág-Kristóf 2009). A kiugró adatok kezelése nélkül kapott modell esetében a ROC-görbe alatti terület 83,5%, a kiugró adatok helyettesítésével kapott modell esetében ez az érték már magasabb, 85,2%-os. A kiugró adatok törlésével adódó modell esetében 90,8% a ROC alatti terület, tehát a modell 90,8%-os pontossággal osztályoz egy, a mintából véletlenszerűen kiválasztott vállalkozást (2. ábra). A ROC-görbék alapján tehát a kiugró értékek törlésével kapott modell alkalmazása tűnik a legcélravezetőbbnek.

A csődmodellek esetében fontos, hogy ne csak a tanuló mintán működjenek magas besorolási pontossággal, hanem független vállalkozások csődelőrejelzésére is alkalmasak legyenek, ezért a kapott modelleket ez független, tesztelő mintán is ellenőriztük. Ez a minta 150 db építőipari vállalkozást tartalmaz, amelyből 45 vállalkozás felszámolás alatt áll, 105 cég pedig működő, fizetőképességű vállalat.

2. ábra: Az elkészült modellek ROC-görbéi



Forrás: saját szerkesztés

A 3. táblázat az elkészült csődelőrejelzési modellek besorolási pontosságát mutatja be a tanuló és a tesztelő minta vonatkozásában. A táblázatban látható, hogy a tesztelő minta vonatkozásában a kiugró értékek kezelése nélkül adódó modell működött a legalacsonyabb besorolási pontossággal.

3. táblázat: Az elkészült csődelőrejelzési modellek pontossága a tesztelő és tanuló mintán

		Túlélő cégek besorolása, %	Csődös cégek besorolása, %	Teljes minta pontossága, %
Kiugró adatok kezelése nélkül	Tanuló minta	76,34	76,99	76,44
	Tesztelő minta	81,95	71,11	80,00
Kiugró adatok helyettesítésével	Tanuló minta	81,40	78,17	80,68
	Tesztelő minta	88,78	73,33	86,00
Kiugró adatok törlésével	Tanuló minta	85,73	82,32	85,21
	Tesztelő minta	88,78	68,89	85,20

Forrás: saját szerkesztés

Az outlier adatokat tartalmazó vállalkozások mintából való kizárásával adódó modell besorolási pontossága felülmúlja a másik két modell előrejelző képességét a tanuló mintán. A tesztelő mintán azonban a csődös cégekre vonatkoztatva alacsonyabb pontossággal működött, mint a másik két modell az elsőfajú hibák aránya tehát magasabb. Ennek oka, hogy a kiugró adatok leginkább a csődbe kerülő vállalkozásokra jellemzők, az eredeti tanuló minta fizetésképtelen vállalkozásainak 42%-a rendelkezett valamilyen kiugró értékkel és került kizárára a mintából. Ezeknek a vállalkozásoknak a mintából való kizárása erős információ- és adatvesztést okozott,

amely információ a modell elkészülésekor nem került felhasználásra, emiatt tapasztalható, hogy bár a modell a tanuló mintában szereplő fizetésképtelen vállalkozásokat kiemelkedően, 82,32%-os pontossággal osztályozta, a tesztelő mintán jóval alacsonyabb hatékonysággal működött: a csődös cégek vonatkozásában 68,89%-os találati aránnyal osztályozta a tesztelő minta fizetésképtelen részét.

Mindezek alapján igazoltnak tekinthető az az állítás, miszerint logisztikus regresszió alapuló csődelőrejelzési modell elkészítésekor fontos az adatok alapos előkészítése és a kiugró értékek kezelése. Magyarországi építőipari vállalkozások logisztikus regresszió alapuló csődelőrejelzésére egy olyan modell a legalkalmasabb, amely a kiugró adatok helyettesítéssel való kezelésével készült el.

Összegzés

A pénzügyintézetek a hitelezési scorecardok elkészítéséhez leggyakrabban a logisztikus regresszió módszerét alkalmazzák. Tanulmányunkban logisztikus regresszió segítségével készítettünk el három csődelőrejelzési modellt annak érdekében, hogy megvizsgálhassuk, hogy a kiugró értékekre érzékeny módszer pontosságát mennyire javítja, ha a modellezés előtt a minta kiugró adatait kezeljük, valamint, hogy a kiugró adatok törlése, valamint a kiugró értékek helyettesítése a hozzájuk legközelebb eső, de már nem kiugró értékkel módszerek közül melyik alkalmazásával érhető el magasabb besorolási pontosság. Tanulmányunkban azt is megvizsgáltuk, hogy egy olyan minta esetében, amely nem egyenlő arányban tartalmaz fizetőképessé és fizetésképtelen vállalkozásokat, mi alapján érdemes cut-off pontot választani.

Az elkészült modellek rávilágítottak, hogy a kiugró értékek kezelése jelentősen javítja a modellek előrejelző-képességét, a kiugró értékek kezelésére a kiugró értékek helyettesítése a hozzájuk legközelebb eső, nem kiugró értékkel bizonyult hatékonyabbnak.

A cut-off értékek meghatározásakor nem az az érték a célravezető, amely a legmagasabb besorolási pontosságot adja, hiszen ekkor a mintában nagyobb arányban szereplő túlélő vállalkozások találati arányát veszi figyelembe, ez pedig magas elsőfokú hibaarányhoz vezethet. Ez a csődelőrejelzési modellek elsődleges felhasználói, vagyis a bankok számára káros jelenség, ezért egy olyan cut-off érték meghatározása a célravezető, amely kellően magas besorolási pontosságot eredményez az elsőfajú hibák alacsonyan tartása mellett. Ennek az értéknek az optimalizációja attól függhet, hogy az adott pénzügyintézet milyen mértékű hitelezési kockázatot vállal a kihelyezett hiteleinek portfóliójában. Érdekes jövőbeli kutatási lehetőség lehet annak megvizsgálása, hogy a bankok portfólióikban milyen tényezők alapján határozzák meg a cut-off értékét. ■

Irodalom

1. Bellovary, J. L., Giacomino, D. E., Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present, *Journal of Financial Education*, Vol. 33, pp. 1-42.
2. Cziglerné Erb E. (2020). A reziduálisjövedelem-modell újbóli megjelenése a vállalatok és a beruházási projektek értékelésében, *Pénzügyi Szemle*, 3, pp. 430-442.
3. Doma, T., Kozma, N. (2022). A Covid 19-járvány hatása a magyarországi bankok működési kockázati veszteségeire, *Gazdaság és Pénzügy*, 9, pp 356-375. <https://doi.org/10.33926/GP.2022.4.3>
4. du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy, *Neurocomputing*, 73, pp. 2047-2060. <https://doi.org/10.1016/j.neucom.2009.11.034>
5. Kiss G., Kosztopulosz A., Szládek D. (2021). A magánkiadások hatása a hazai egészségügyi diagnosztikai szolgáltatók pénzügyi helyzetére, *Köz-Gazdaság*, 16, pp. 115-132. <https://doi.org/10.14267/retp2021.04.08>
6. Kristóf, T. (2005). A csődelőrejelzés sokváltozós statisztikai módszerei és empirikus vizsgálata, *Statisztikai Szemle*, 9, pp. 841-863.
7. Kristóf, T., Virág, M. (2019). A csődelőrejelzés fejlődéstörténete Magyarországon, *Vezetéstudomány*, 12, pp. 62-73.
8. KSH (2021). Helyzetkép, 2021 – Építőipar, Központi Statisztikai Hivatal, Budapest
9. Limpek, Á., Kosztopulosz, A., Balogh P. (2016). Késedelmes fizetés, tartozási láncok-A Dél-Alföld régió kis-és középvállalkozásainak pénzügyi kultúrája, *Statisztikai Szemle*, 94, pp. 365-387. <https://doi.org/10.20311/stat2016.04.hu0365>
10. Mikolasek, A. (2018). A hitelkockázati modellek alkalmazásának néhány problémája, *Gazdaság és Pénzügy*, 3, pp. 248-257.
11. Nyitrai, T. (2015). Hazai vállalkozások csődjének előrejelzése egy, két, illetve három évvel korábbi pénzügyi beszámoló adatai alapján, *Vezetéstudomány*, 5, pp. 55-65. <https://doi.org/10.14267/veztud.2015.05.06>
12. Nyitrai, T. (2017). Stock és flow típusú számviteli adatok alkalmazása a csődelőrejelző modellekben, *Vezetéstudomány*, 48, pp. 68-77. <https://doi.org/10.14267/veztud.2017.09.07>
13. Nyitrai, T., Virág M. (2017). A pénzügyi mutatók időbeli tendenciájának figyelembevétele logisztikus regresszióra épülő csődelőrejelző modellekben, *Statisztikai Szemle*, 1, pp. 5-28. <https://doi.org/10.20311/stat2017.01.hu0005>
14. Pálinkó, É., Svoób Á. (2016). A vállalati csőd bekövetkezésének fő okai és a csődhöz vezető folyamat, *Pénzügyi Szemle*, 4, pp. 528-543.
15. Piller, Zs., (2013). A fizetéseképtelenségi eljárások illeszkedési módjai nemzetközi összehasonlításban, *Pénzügyi Szemle*, 2, pp. 151-164.
16. Rácz, T. A., Tóth, B. (2021). A hazai önkormányzatok pénzügyi zavarai az adósságkonszolidáció és az önkormányzati rendszer reorganizációjának tükrében, *Pénzügyi Szemle*, 1, pp. 88-108. https://doi.org/10.35551/psz_2021_1_5

17. Virág M, (2004). A csődmodellek jellegzetességei és története, *Vezetéstudomány*, 10, pp. 24-32.
18. Virág, M., Kristóf T. (2006). Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel, *Vezetéstudomány*, 37, pp. 25-35. <https://doi.org/10.14267/veztud.2006.01.04>
19. Virág, M., Kristóf, T. (2009). Többdimenziós skálázás a csődmodellezésben, *Vezetéstudomány*, pp. 50-29. <https://doi.org/10.14267/veztud.2009.01.05>
20. Vörösmarty, Gy., Dobos, I. (2020). A vállalatméret hatása a zöldbeszerzési gyakorlatra, *Statisztikai Szemle*, 4, pp. 301-323. <https://doi.org/10.20311/stat2020.4.hu0301>
21. Zavgren, C. V. (1985). Assessing the vulnerability to failure of American industrial firms: a logistic analysis, *Journal of Business Finance & Accounting*, 12(1), pp. 19-45.
22. Zéman, Z., Hegedűs, Sz., Molnár, P. (2018). Az önkormányzati vállalkozások hitelképességének vizsgálata credit scoring módszerrel, *Pénzügyi Szemle*, 2, pp. 182-200.
23. Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research*, Vol. 22, pp. 59-82.