

KÜLFÖLDI CSŐDELŐREJELZŐ MÓDSZEREK SZISZTEMATIKUS IRODALOMELEMZÉSE

SYSTEMATIC LITERATURE REVIEW OF FOREIGN BANKRUPTCY PREDICTION METHODS

A vállalati fizetésképtelenség, csőd és pénzügyi nehézség vizsgálata egy intenzív kutatási terület, amelynek számos különböző gyakorlati eljárása megfigyelhető. A tanulmány a vállalati csődelőrejelzés külföldi szakirodalmát vizsgálja, a szisztematikus irodalomelemzés módszerével. A kutatás célkitűzése kettős, elsősorban megvizsgálja a vállalati csődelőrejelzés legjobban teljesítő módszereit, másodsorban felfedi az ehhez kapcsolódó leggyakoribb tényezőket, a magasan hivatkozott külföldi csőd kutatások alapján. Három tudományos adatbázist felhasználva, 105 szakirodalmi cikket dolgoztak fel, amelyek 1966 és 2017 közötti időszakban tettek közzé. A szakirodalmi áttekintés, hat módszercsalád összehasonlítását teszi lehetővé. Az eredmények azt mutatják, hogy a döntési fa módszercsalád fölülmúlja az SVM, a neuronháló és a hagyományos statisztikai módszereket. A közepes pontosságú módszerek közül a példányalapú módszercsalád és a logisztikus regresszió összemérésekor nem lehetett egyértelmű rangsort felállítani. A csőd tényezőinek vizsgálatánál körvonalazódott, hogy a hagyományos pénzügyi mutatók mellett alkalmazott piaci mutatók átlagosan nem vezetnek magasabb előrejelző pontossághoz, mint a csak kizárólag pénzügyi mutatókat tartalmazó modellek.

Kulcsszavak: csődelőrejelzés, szisztematikus irodalomelemzés, fizetésképtelenség, üzleti kudarc, pénzügyi nehézség

Research on corporate insolvency, bankruptcy and financial distress comprises an intensive research genre with numerous practical methods. The study investigates the foreign recognized literature on corporate bankruptcy prediction using the method of systematic literature review. The objective of this research is twofold, firstly to examine the best-performing methods of corporate bankruptcy prediction and secondly to reveal the most common factors based on the reviewed research. Using three scientific databases, 105 articles from 1966 to 2017 were reviewed. The literature review compares six families of methods. The results show that the decision trees exceed the SVM and neural network method and the traditional statistical methods. Between instance-based methods and logistic regression as methods with medium accuracy, no clear ranking could be established. Examining the factors of bankruptcy, we concluded that market indicators used next to financial indicators do not lead on average to higher forecasting accuracy than models including only financial indicators.

Keywords: bankruptcy prediction, systematic literature review, insolvency, business failure, financial distress

Finanszírozás/Funding:

A szerző a tanulmány elkészítésével összefüggésben nem részesült pályázati vagy intézményi támogatásban. The author did not receive any grant or institutional support in relation with the preparation of the study.

Szerző/Author:

Ágoston Norbert^a (agoston.norbert@ktk.pte.hu) PhD-hallgató

^aPécsi Tudományegyetem

A cikk beérkezett: 2021. 03. 22-én, javítva: 2021. 07. 22-én, elfogadva: 2021. 11. 09-én.

The article was received: 22. 03. 2021, revised: 22. 07. 2021, accepted: 09. 11. 2021.

A vállalatok kényszerű megszűnése, illetve csődje viszszafordíthatatlan költségeket generál az érintettek körében. Veszteség éri a vállalat hitelezőit, tulajdonosait, vezetőségét, munkavállalóit, akik elveszíthetik munkahelyüket, továbbá a fizetésképtelen vállalat vevőit és beszállítóit. Nem utolsó sorban az állami költségvetés is eszik a várható adóbevételektől, így egy sorozatos csődjelenség veszélybe tudja sodorni a teljes nemzetgazdaságot, emel-

lett jelentős társadalmi problémát is okoz. Ezért kiemelt hangsúlyt kapnak az olyan előrejelző módszerek, amelyek képesek nagy hatékonysággal, a potenciális csődhelyzet bekövetkezése előtt azonosítani a várható vállalati nehézségeket (Shetty, Pakkala & Mallikattjunappa, 2012).

A csőd és pénzügyi nehézség elemzése már régóta kiemelt kérdéskör a vállalatokat értékelő hitelintézetek, biztosítók és intézményi befektetők számára (Hu & Tseng,

2007). Kim & Kang (2010) véleménye szerint a csődelőrejelzés széles körben tanulmányozott téma a számviteli és pénzügyi kutatási területeken, amely kitüntetett jelentőséggel rendelkezik, és nagy hatással van a gazdaságra. A kutatók mellett a gyakorlati szakemberek és pénzügyi intézmények – a fentiekkel összhangban – folyamatosan keresik a legjobb módszereket, ügyfeleik fizetőképességének értékelésére (Barboza, Kimura & Altman, 2017). A csődelőrejelzés fontossága hatványozottan felértékelődik recessziós gazdasági környezetben, amikor a finanszírozók kockázatkerülése dominál (Nyitrai, 2014).

A tanulmány célkitűzése kettős, egyrészt megvizsgálja a vállalati csődelőrejelzés legjobb teljesítő módszereit, másfelől felfedi az ehhez kapcsolódó legfontosabb tényezőket a magasan hivatkozott csőd kutatások alapján. A módszerek teljesítményével kapcsolatban azt az előfeltevést fogalmaztam meg, hogy a korszerű, gépi tanulás módszercsaládba tartozó technikák előrejelző pontossága magasabb, mint a hagyományos statisztikai módszereké. A szakirodalom összehasonlító tanulmányai nem mindig hoznak egyértelmű győztes módszereket, némelyekben a gépi tanulási módszerek hatékonysági képességei tükröződnek (lásd Fan & Palaniswami, 2000; Barboza et al., 2017), ugyanakkor az ellenkezőjére is találunk példát (lásd Coats & Fant, 1993; Pompe & Feelders, 1997). A fenti feltevések tesztelésére és átfogó elemzésére a szisztematikus irodalomkutatás módszerét alkalmaztam.

A csődelőrejelzés tényezőivel kapcsolatban az a feltevés, hogy a hagyományos mérleg és eredménykimutatásból kalkulált vállalati pénzügyi mutatók mellett, makrogazdasági- és piaci változókat tartalmazó modellek pontosabban jelzik előre a vállalatok csődjét, mint a csak kizárólag pénzügyi mutatókat tartalmazó modellek. E feltevés arra alapozható, hogy a vállalati nehézségek és csőd nem mindig belső okokra, működési hibákra vezethető vissza, hanem külső, környezeti válságok is előidézhetnek csődhelyzeteket. Shumway (2001) sikerrel alkalmazott modelljeiben piaci alapú független változókat. Hernandez Tinoco & Wilson (2013) arra jutott, hogy a piaci mutatók hozzájárulása a csődmodellekhez jelentős, mivel a pénzügyi mutatókban nem tükröződő többletinformációt hordoznak, valamint a makrogazdasági változók bevonása a modellekbe teljesítménynövekedést okoz.

Szakirodalmi vonatkozások

A csődelőrejelzés hagyományos szakirodalmi összefoglaló tanulmányai kiválóan alkalmasak a terület körülírására, nemzetközi kitekintésben lásd Balcaen és Ooghe (2006), Bellovary, Giacomino & Akers (2007), Ravi Kumar & Ravi (2007) tanulmányait. Ez a fajta megközelítés több ponton szubjektív elemeket viselhet: a szerző személyes megítélésén alapulhat az irodalom összeválogatás és feldolgozás, amely bizonyos esetben nem megfelelő merítésű is lehet, így a következtetések értékelése torzításokat tartalmazhat.

A szisztematikus irodalomösszefoglaló tanulmányok dokumentált keretrendszerbe foglalják az összegyűjtött tanulmányokat, és ezért célzott kérdések megválaszolásá-

ra is alkalmasak. A csődelőrejelzés témakörében Appiah, Chizema & Arthur (2015) 83 cikket dolgoztak fel 11 országból és azt mutatták ki, hogy a statisztikai módszerek dominálják a szakirodalmat 68%-os átlagos besorolási pontossággal, valamint a fizetésektelenség előtti első év beszámolóit felhasználva készülnek a legjobb modellek. Ez utóbbit megerősíti Nyitrai (2015) megfigyelése, amely szerint a csőd előtti első év adatai magasabb információ-tartalommal rendelkeznek, mint a múltban távolabbi adatok. Alaka et al. (2018) 2010 és 2015 között publikált 49 folyóiratcikket elemeztek, és ezekből létrehoztak egy logikai rendszert az optimális módszer kiválasztásához. A szerzők arra jutottak, hogy nincs egyetlen uralkodó módszer, amely felülmúlná a többi, bizonyos helyzetben mindenik magában hordoz erősségeket. Szerintük a kifinomult hibrid módszerek lehetnek azok, amelyek abszolút előnyöket biztosítanak. Scherger, Terceño & Vigier (2019) 1968 és 2017 közötti csődelőrejelzéssel foglalkozó szakirodalmakról szisztematikus bibliográfiai összefoglalót készítettek, és kimutatták, hogy a témában megjelenő irodalmak 2005-től kezdődően erőteljes növekedésben vannak. Az elemzett 1104 dokumentumból mindössze hatnak volt 300 feletti hivatkozásszáma, amelyek jellemzően más tudományterületekre is hatással bíró munkák. Shi & Li (2019) 496 tanulmányt használtak fel egy szisztematikus irodalomelemzéshez és megvizsgálták a kutatók együttműködését a társszerzős cikkek mentén. Arra jutottak, hogy a csőd kutatók között kevés a kooperáció, továbbá azt mutatták ki, hogy a 2008-as globális gazdasági válság után a csődelőrejelző tudományos publikációk nagyfokú növekedést mutattak, a megvizsgált tanulmányokban a leggyakoribb módszerek a logisztikus regresszió és neurális hálózatok.

A jelen tanulmány a fentiek mellé ajánl egy magyar nyelvű alternatívát, 105 magas hivatkozásszámú, széles időperiódust és nemzetközi színteret lefedő angol nyelvű folyóiratcikk feldolgozásával, amelyben a módszerek teljesítményének több szempontú értékelése mellett a magyarázó változók szintetizált elemzése is megtalálható.

A tanulmány a magyar csőd kutatás tanulmányait nem tartalmazza, kizárólag a külföldi, angol nyelvű folyóiratcikkekre fókuszál. Az úttörő első magyar csődmodelleket és eredményeiket Virág (2004), valamint Virág és Dobé (2005) tanulmányok foglalják össze. A csődbejutás kialakulásának okait és folyamatait Kristóf (2005) munkájában találhatjuk, aki szervezetelméleti megközelítések részletezésével támogatja a csődelőrejelzés empirikus kutatásait. Néhány tanulmányt tetszőlegesen kiragadva, a hazai csődmodellek összehasonlító elemzéseit a következő munkákban találjuk: Virág és Kristóf (2005) tanulmányában a neurális háló, a logisztikus regresszió és a diszkriminanciaanalízis módszerét hasonlították össze. Virág és Nyitrai (2013) az SVM módszer és a neurális háló összevetését végezte el, Kristóf (2018) munkájában az esetalapú következtetés (case-based reasoning, CBR), a döntési fa, a logisztikus regresszió és a neurális háló módszerek összemérését ismerhetjük meg. Virág és Kristóf (2006) az iparági ráták alkalmazásán túlmenően négy különböző csődelőrejelző módszer, diszkriminan-

ciaanalízis, logisztikus regresszió, döntési fa és neurális háló, összehasonlítás is elvégezte. A csődelőrejelzés modellépítés nélküli módszertani összefoglalóját Szabadosné és Dávid (2005), Oravetz (2007) és Rattng (2015) tanulmányokban olvashatjuk. A magyarországi csődelőrejelzés 30 éves, teljes fejlődéstörténetéről átfogó képet Kristóf és Virág (2019), valamint Kristóf és Virág (2020) összefoglaló tanulmányokban találhat az olvasó. A szerzők megállapításaik alapján a magyar szakirodalom leggyakoribb eljárásai a logisztikus regresszió, a neurális háló és a döntési fák, ugyanakkor a módszertani előrelépéseket a mesterséges intelligencia, gépi tanulás, hibrid modellezés és adatbányászat jellemzi. A klasszikus diszkriminanciaanalízis módszer viszonylag egyszerű interpretálhatóságát beárnyékolja a mérsékelt teljesítmény, ennek ellenére létjogosultága megmarad, mivel jövőbeli kutatások során benchmark módszerként alkalmazható (Kristóf & Virág, 2019).

A szisztematikus irodalomkutatás módszere és alkalmazása

A szisztematikus irodalomkutatás (systematic literature review) egy alapvető eszköz az akadémiai tanulmányok elkészítésekor (Shi & Li, 2019). Ez egy olyan korszerű kutatási technika, amely kijelöli egy bizonyos tudományterület problémakörét, megjelöli az információs forrásokat, a keresési kulcsszavakat, a kizárési kritériumokat és az elvégzendő elemzést (Nicolas & Toval, 2009). A szisztematikus irodalomkutatás alapvetően különbözik a hagyományos, elbeszélő jellegű irodalomfeldolgozástól. A legfőbb előnye abban rejlik, hogy megbízható információkat kínál, módszertanilag átlátható és lehetővé teszi a tanulmány jövőbeli reprodukálását (Tranfield, Denyer & Smart, 2003).

A szisztematikus irodalomfeldolgozás során kis-, közép- és nagyvállalati csőd, fizetésképtelenség és pénzügyi nehézség előrejelzésére fókuszáltam, keresve a legmegbízhatóbb, legmagasabb előrejelző pontosságú modellek módszereit és azok jellemző változótípusait. Az irodalomkutatás és feldolgozás időszaka 2020. február 10-től 2020. július 31-ig tartott. A legrelevánsabb folyóiratcikkek azonosításához és a bibliográfiai kereséshez elektronikus adatbázisokra támaszkodtam, ezek a Web of Science (Clarivate Analytics), az EBSCO Business Source Premier és az EBSCO Academic Search Complete. A kulcsszavak kiválasztása előtt megfogalmaztam a keresési kritériumokat:

- kizárólag folyóiratcikk került be a vizsgálatba, kizártam a könyveket, a konferenciakiadványokat és egyéb dokumentumokat,
- kizárólag angol nyelven közzétett tanulmányok kerülhettek be, földrajzi határok és megjelenési időkorlát nélkül,
- a folyóiratcikkek közül kizártam a szakirodalom áttekintő tanulmányait (review paper),
- elegendő, ha egy tanulmány megjelenik az egyik adatbázisban, nem kell mindhárom adatbázisban szerepelnie.

A vállalati csődelőrejelzés témájának kulcsfontosságú tanulmányaihoz több fordulós folyamat során jutottam el, amelyet a következőkben bemutatok, ezzel lehetővé téve a tanulmány jövőbeli replikációját. A keresési kifejezések összefoglaló táblázata az 1. online mellékletben található. Első körben a lehetséges maximális kulcsszó meghatározása történt, a vállalathoz kapcsolódó kulcsszavak („corporate”, „firm”, „company”, „business”), és a csődelőrejelzéshez kapcsolódó kulcsszavak felhasználásával, amely nagyszámú találathoz vezetett. Ezután a kulcsszavak módosításával kizártam a hitel és bankcsődre utaló tanulmányokat, így a három adatbázis összes találatja több mint ezer tanulmány volt. A kevésbé lényeges kulcsszavak kiküszöbölésével, mint a „business”, a „forecast” és a „model”, a harmadik lépésben 676 találat volt a három adatbázisnak. A negyedik lépésben kiegészítettem a kulcsszavakat a „gépi tanulás” kifejezéssel, mivel érdeklődésem középpontjában az újszerű csődelőrejelző megoldások vannak. A négy keresési kifejezés után összesen 2998 tanulmányra leltem, a három elektronikus adatbázisból, amelynek összefoglaló táblázata az 1. online mellékletben található.

A 2. online melléklet a szisztematikus irodalomfeldolgozás folyamatábráját jeleníti meg, ismertetve a különböző kiválasztott és kizárt tudományos szakirodalmak számát és a műveletek magyarázatát. Alkalmazva a korábban bemutatott szűrőt, amely csak az angol nyelvű folyóiratcikkeket tartja meg, 984-re csökkent a kapcsolódó tanulmányok száma. A következő lépésben a különböző adatbázisok azonos találatának a kiszűrése következett, amelyben segítségemre volt a Zotero 5.0.84 hivatkozásmenedzsment-alkalmazás. A duplikátumok kiszűrése mellett kizártam az angol kivonattal, de nem angol nyelven közzétett tanulmányokat, így a tanulmányok száma 537-re csökkent.

A következő lényeges pont egy minőségi szűrő alkalmazása volt, amely során a száz legtöbbször hivatkozott cikkeket választottam ki, Google Scholar hivatkozásszám alapján. Az első legmagasabban hivatkozott tanulmány, több mint 3000 idézéssel, a századik, azaz legkevésbé hivatkozott tanulmány, 65 citációval rendelkezett, a Google Scholar szerint a 2020.02.25. – 2020.02.27. közötti időszakban. A száz legtöbbször hivatkozott szakirodalom mellé, szakértői javaslatra négy további szakirodalmat vontam be, amelyek egyenkénti hivatkozásszáma 560-130 közötti. Egy kiegészítő keresési eljárással, hólabdamódszerrel, a nagyon releváns szakirodalmak gyakran meghivatkozott tanulmányait kerestem meg, amelyek a szisztematikus kutatás során nem voltak megtalálhatók. Így öt további tanulmányt sikerült azonosítani és bevonni a feldolgozásba, amelyek egyenkénti hivatkozásszáma 1000 és 17000 közötti. Következésképpen az összes feldolgozásra kiválasztott szakirodalmak száma 109-re gyarapodott. A feldolgozás során négy tanulmány nem volt releváns, ugyanis bankcsődökkel kapcsolatos és szakirodalmi áttekintő tanulmányok voltak, így végül eltávolítottam őket, ezáltal a feldolgozott szakirodalmak száma 105-re módosult.

Az adatok összegyűjtése a következő fő pontok mentén történt, amely rendszerint lefedte az egyes szakirodalmak lényeges tartalmát:

- a tanulmány szerzői, kulcsszavai, megjelenési éve,
- a tanulmány célkitűzései, kutatási kérdései,
- a csőd fogalmának definiálása,
- a minta eredete, összetétele és mérete,
- a kutatási technika, amely segítségével kifejlesztették a csődelőrejelző modellt,
- az alkalmazott magyarázó változók száma és típusa, valamint a változó szelekció módszere,
- az eredmények, amelyek a besorolási, előrejelzési és modellilleszkedési pontosságokat tükrözik,
- az eredmények validációjának módja és eljárása.

A csődelőrejelzés gyakori módszerei

A következőkben röviden ismertetem azt az öt leggyakrabban használt módszert, melyek több mint 12 alkalommal voltak felfedezhetők az irodalomkutatás során, ezek a diszkriminanciaanalízis (DA), a logisztikus regresszió, a neurális háló, a Support Vector Machine (SVM) illetve a döntési fa (DT) módszercsalád.

A diszkriminanciaanalízis

A többváltozós diszkriminanciaanalízis módszere a megfigyeléseket előre meghatározott osztályokba sorolja be (Altman, 1968). A módszer megkeresi a jellemvonások olyan lineáris kombinációját, amely a legjobban elkülöníti a meghatározott csoportokat (Virág, 2001). Alkalmazása során azonosíthatjuk azokat a változókat, amelyek megkülönböztetik a megfigyelt osztályokat, azaz megvizsgálhatjuk, hogy a független változókkal becsülhető-e az osztályhoz való tartozás (Sajtos & Mitev, 2007). Az eljárás során új változók, úgynevezett diszkriminanciafüggvények jönnek létre úgy, hogy a lehető legnagyobb különbség legyen a csoportok között (Székelyi & Barna, 2008). A diszkriminanciafüggvény általános alakja a következő módon írható fel:

$$Z=c+w_1*x_1+w_2*x_2+...+w_n*x_n \quad (1)$$

„ahol „Z” a diszkriminanciaérték, amely által a besorolás megvalósítható, „c” a konstans, „w_i” a diszkriminancia-súlyok vagy együtthatók, „x_i” pedig a független változók. A diszkriminanciafüggvény (Z) egy kritikus érték, amely elválasztja a csődös és nem csődös vállalatokat egymástól. A diszkriminanciaanalízis számos feltételt támaszt az adatokkal kapcsolatban: a magyarázó változók normális eloszlása és függetlensége, valamint a független változók varianciájának a függő változók csoportjaiban megegyezőnek kell lennie (Virág & Kristóf, 2005).

A logisztikus regresszió

A kétértékű (dichotóm) logisztikus regresszió a csődtanulmányokban elterjedt eljárás, amely két, előre meghatározott csoport közti besorolást végez. A multinomiális logisztikus regressziónál a függő változónak kettőnél több értéke lehet. Logit transzformációt alkalmazva felírható a csőd bekövetkezésének valószínűsége a (2) képlet szerint:

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \sum(\beta_i * x_i))}} \quad (2)$$

ahol „P(y=1)” a csőd bekövetkezésének valószínűsége, „β_i” regressziós együtthatók, „x_i” független változók. A paraméterek becsüléséhez a logisztikus regresszió a maximum likelihood becslést használja, ami olyan együtthatókat keres, melyek a modell legjobb illeszkedését biztosítják. A módszer több lépéses eljárás során határozza meg a legjobb paramétereket (Székelyi & Barna, 2008). A diszkriminanciaanalízissel szemben előnyös, olyan szempontból, hogy nem feltételez többdimenziós normális eloszlást, és azonos variancia-kovariancia mátrixot (Virág & Kristóf, 2005).

A mesterséges neuronhálózatok

A neurális háló, a mesterséges intelligencia területéhez tartoznak, felépítésük az emberi idegrendszer mintáján alapszik (Herz, Krogh & Palmer, 1991). A neurális háló neuronok összekapcsolt láncolatából áll, amely neuron rétegeket hoz létre, ezáltal elkülöníthetünk bemeneti, kimeneti és köztes vagy más néven rejtett réteget. Mindenik szomszédos neuron összeköttetésben áll egymással, és mindenik összeköttetéshez egy súly tartozik, mely meghatározza a kapcsolat előjelét és erősségét. A mesterséges neuronháló tanulási folyamatában a kapcsolatok súlyai folyamatosan változnak addig, amíg egy leállítási feltétel nem teljesül, például ha a súlyok változása nagyon kicsi. A neuronok másik fő komponense az aktivációs függvény, ami a neuronok kimeneteleit szabályozza (Russel & Norvig, 2005).

Az utóbbi időben a neurális hálók alapuló mély tanulás (deep learning) a számítógépes számítási kapacitások növekedésének köszönhetően egyre népszerűbb gépi tanulási eljárás lett, és kiválóan alkalmas klasszifikációs feladatok megoldására (Zhao, Li & Yu, 2017). A mélytanulás olyan előrecsatolt hibavisszaterjesztő neurális hálózatokból épül fel amelyek, a ki- és bemeneti rétegek mellett egynél több köztes rejtett réteget tartalmaznak és ígéretes eredményekre képesek főként a kép-, hang- és szövegfelismerés terén (LeCun, Bengio, & Hinton, 2015).

A döntési fa módszercsalád

A döntési fa a gépi tanulási módszerek családjába tartozik, a rekurzív particionáló algoritmus eljárást alkalmazza, amely által klasszifikációs fákat állít elő egyszerű szabályok alapján. A fastruktúra létrehozása több lépéses folyamat, melynek során a teljes mintát kétfelé osztja úgy, hogy a lehető leghomogénebb osztályok jöjjenek létre (Virág, Fiáth, Kristóf & Varsányi, 2013). A fa létrehozásához számos algoritmust fejlesztettek ki, elsők közt az úgynevezett Iterative Dichotomiser 3 vagy ID3 (Quinlan, 1986), később újabb változatok jelentek meg Classifier 4.5 néven (Quinlan, 1993), majd megjelent a továbbfejlesztett kereskedelmi változat, a C5.0 (Quinlan, 1997). A CART (Classification And Regression Trees) algoritmus regresszió alapú döntési fa, bináris fákat állít elő (Breiman, Friedman & Olshen, 1984). A CHAID (Chi-squared Automatic Interaction Detector) döntési fa chi-négyzet statisztikával végzi az optimális eloszlást a csoportok között (Kass, 1980; Hámori, 2001), illetve az OCI (Oblique Classifier 1) algoritmussal készült döntési fákat Murty, Kasif, Salzberg & Beigel (1993) fejlesztette ki.

A döntési fa módszer családba sorolhatók azok az együttes (ensemble) módszerek, amelyek döntési fákon alapulnak: kiemelhetjük a Bagging (vagy zsákolás), a Boosting (vagy gyorsítás) és a Véletlen erdő (random forest) metamódszereket. A Bagging és Boosting együttes eljárások esetén a tanulólthalmaz szabályozásával építhetők ki az együttes módszerek. Előbbinél a teljes adatmintából visszatevésees eljárással több különböző, véletlenszerűen kiválasztott almintát felhasználva lefutatható egy alap-tanuló eljárás, majd végül a több különböző eredményt átlagolva megkapjuk a végeredményt (Breiman, 1996). Utóbbinál a véletlenszerűen kiválasztott almintákból kiépült modell az előzőleg létrejött modell teljesítményétől függ, olyan módon, hogy a tanuló esetek minden menet végén módosulnak: a helytelenül besorolt példányok súlyai növekednek, a helyesen besoroltak súlyai csökkennek. A legvégén az alaposztályozók súlyozott szavazata vagy átlaga alapján kapjuk a végső osztályozást (Freund & Schapire, 1996). A Véletlen erdő eljárás kimondottan a döntési fa osztályozóhoz tervezett együttes eljárás, amely több döntési fa előrejelzéseit átlagolja és az osztályba kerülést többségi szavazással dönti el (Breiman, 2001).

A Support Vector Machine

A Support Vector Machine (SVM) olyan mesterséges intelligencia alapú eljárás, amely alkalmas lineáris és nem-lineáris osztályozásra. Egy olyan matematikai függvény transzformáláson alapuló optimalizálási modell, amely az úgynevezett kernel függvény segítségével azonosítja a legnagyobb távolságot a legjobban hasonlító ellentétes besorolású megfigyelések között (Noble, 2006). Abban az esetben, ha lineárisan szeparálható a megfigyelés, akkor az osztályokat elválasztó felület vagy hipersík, egy lineáris függvény. Nem lineáris szeparációk esetében egy kernel függvény segítségével a megfigyeléseket az eredeti dimenzióból átranzformálja egy új, magasabb dimenziós számú térbe, ahol elvégezhető a szeparáció, azaz van egy lineáris hipersík (Russel & Norvig, 2005). A margó (margin) a hipersík alatt és felett helyezkedik el úgy, hogy meghatározza az osztályok határát. Az optimális hipersík az, amelyiknek a legnagyobb a margója. A margóhoz legközelebb lévő pontok az úgynevezett tartó- vagy szupport vektorok, melyek „tartják” az elválasztó síkot (Cortes & Vapnic, 1995).

Eredmények

A szisztematikus irodalomfeldolgozás során 105 magasan hivatkozott szakirodalmi cikket dolgoztam fel. A 3. online melléklet a kiválasztott szakirodalmakat és az általuk használt minták jellemzőit ismerteti. A beazonosított irodalmak 1966-2017 közötti időszakból származnak, az 1. ábra szerinti eloszlásban. A 60-as, 70-es és 80-as évekből szórványosan származik egy, legfeljebb három tanulmány, 1990-től kezdődően az 1997-es év kivételével, minden évből található legalább egy szakirodalom. A 2011-es évből a legtöbb, 9 tanulmányt vizsgáltam meg.

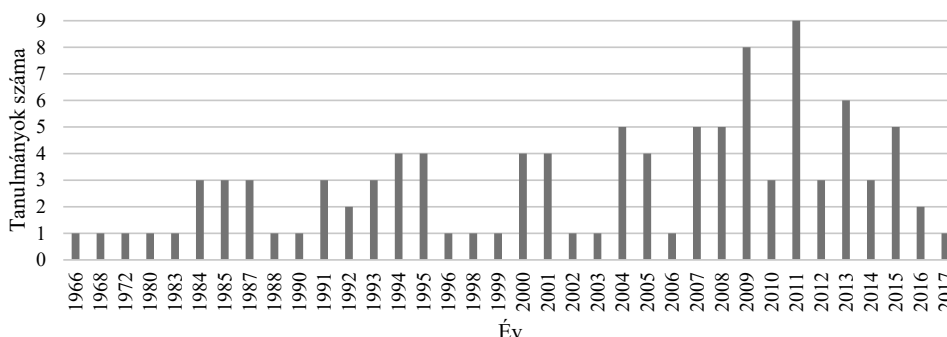
A 105 megvizsgált tanulmányból a legtöbb az Amerikai Egyesült Államok vállalatának mintáját dolgozta fel, összesen 39 ilyen tanulmány van. 12 tanulmány dél-koreai vállalati mintát, 11 tanulmány kínai vállalatok adatain dolgozott. Továbbá 7 Tajvan, 6 Egyesült Királyság és 3-3 Ausztráliához és Finnországhoz kapcsolódó tanulmány volt. Két tanulmány kötődött a következő országokhoz: Belgium, Spanyolország, Franciaország, Japán, Görögország, Irán, Lengyelország. Egy tanulmány kapcsolódott a következő országokhoz: Oroszország, Norvégia, Törökország, Olaszország, Szingapúr, Portugália, Argentína. További négy tanulmány több ország adatait egyidejűleg használta fel, jellemzően az Amerikai Egyesült Államok és Kanada, de volt Japán és Egyesült Államok, valamint Latin Amerika országainak adataira támaszkodó tanulmány is. Ezek összefoglaló tábláját lásd a 3. online mellékletben.

A tanulmányok igencsak változatos mintanagysággal dolgoznak. Előfordul 22 vállalat mintájából épülő modell (lásd Sandin & Porporato, 2007), viszont a több ezres mintanagyság sem ritka (lásd Berg, 2007; Zhou, 2013; De Andres, Lorca, de Cos Juez, & Sanchez-Lasheras, 2011). Mindössze 8 tanulmány jelölt meg a reprezentativitásra vonatkozó információt, ezek közül kettő jegyezte meg, hogy a minta reprezentatív a vizsgált sokaságra nézve: lásd Berg (2007), valamint Barboza, Kimura és Altman (2017) tanulmányokat.

A tanulmányok túlnyomó többsége ismertette a vizsgálati időhorizontot, habár öt tanulmány nem jelölte meg, hogy a minta melyik évből, avagy milyen időszakból származik. A legrövidebb vizsgált időperiódus egy év (lásd De Andres et al., 2011; Ciampi, 2015), a leghosszabb 38 év (lásd Chawa & Jarrow, 2004).

1. ábra

A feldolgozott szakirodalmak megjelenési év szerinti eloszlása



Forrás: saját számítás

Az adatbázisokat tekintve a leggyakoribb a Standard & Poors's Compustat vállalati statisztikai adatbázisa, amelyet 29 tanulmány jelölt meg információforrásként. A Bureau van Dijk és a Moody's adatszolgáltatót 6-6 tanulmányban említették meg. Meglepő módon 9 tanulmányban nem hozták nyilvánosságra az adatbázist, amely forrásul szolgált a minta összeállítására. A többi tanulmány egyéb adatforrást jelölt meg, a legjellemzőbb egy-egy tőzsde nyilvános adatainak igénybevétele vagy a cégjegyzék-hivatalok jelentései, de előfordulnak bankok, kutatóintézetek és egyéb sajátos adatszolgáltatók is. 11 tanulmány a mintában szereplő vállalatok adatait is közléteszi, a hivatalos elnevezésén kívül az utolsó pénzügyi jelentés dátumát, az összes eszközállományt, esetenként iparági besorolást nyilvánossá teszi, valamint előfordul, hogy bizonyos pénzügyi mutatószámokat is kapcsolnak a cégekhez, amely adatoldali szempontól megkönnyíti az adott tanulmány reprodukálhatóságát. A három legmagasabban hivatkozott Altman (1968), Beaver (1966) és Ohlson (1980) tanulmányok megingathatatlan sarokpontjai a témakörnek és referenciaként szolgálnak bármilyen jövőbeli kutatáshoz. Ezekon kívül a három legmagasabban hivatkozott tanulmány, amelyek közzétette a vállalatok adatait, a minta időperiódusát és a számszerű eredményeket az 1. táblázatban található.

megfogalmazás szerint az üzleti kudarc akkor következik be, amikor a vállalati tartalékok kimerülnek, és nem tudja teljesíteni a kötelezettségeit (Van Horne, 1977).

A szisztematikus irodalomfeldolgozás során azt találtam, hogy a tanulmányok a vállalati csődöt számos különböző meghatározással jelölik, ezért az átláthatóság érdekében szükség volt összefoglalni az eltérő definíciókat:

- *jogi* megközelítés szerinti csőd: hivatalos jogi eljárások útján lejátszódó csődeljárás, felszámolási eljárás, fizetéseképtelenségi eljárás, csődeljárás során történő felvásárlás vagy összeolvadás,
- *technikai* csőd: ebbe a kategóriába sorolódik a hitel nemteljesítés, a hitel újrastrukturálás, a kötelezettségek elmulasztása, a tőkeemelés az elegendő működő tőke biztosításához és az osztalék elmaradása,
- *gazdasági* megközelítés szerinti csőd: a tanulmányok bizonyos mutatók alapján határozták meg a csődöt, mint negatív eredmény, negatív működési pénzáramlás, elégtelen eszközarányos megtérülés (ROA) és elégtelen sajáttőke arányos megtérülés (ROE),
- *jelentés* alapú megközelítés: bizonyos jelentések alapján határozták meg a csőd létét, tőzsdei beszámoló, könyvvizsgálói jelentések, pénzügyi jelentések (éves pénzügyi jelentés elmaradása) mentén. Ebbe a kategóriába soroltam a kötvény- és részvényértékelő

1. táblázat

A legmagasabban hivatkozott tanulmányok, amelyek közzétették a minta összes jellemzőit

	Szerző (év)	Teljes minta	Csődös minta	Időhorizont	Minta országa	Google Scholar citáció (2020.02.25.)
1.	Platt & Platt (1990)	114	57	1972-1986	USA	537
2.	Gentry, Newbold & Whitford (1985)	66	33	1970-1980	USA	492
3.	Charitou, Neophytou & Charalambous (2004)	102	51	1988-1997	UK	447

Forrás: saját szerkesztés

A csőd fogalma

A feldolgozott szakirodalmakban a csőd, a fizetéseképtelenség, a gazdasági kudarc, a fizetés elmulasztása gyakran használt kifejezések a sikertelen vállalatok meghatározására, és jelentésük sok esetben összemosódik. A szakirodalom is megerősíti ezt a megállapítást. Constand & Yazdipour (2011) azt állítja, hogy a csőd fogalmát illetően nincs egy egyetemesen elfogadott meghatározás. Bruno & Leidecker (1988) kissé túlzó véleménye szerint nincs olyan két szakértő, aki az üzleti csőd fogalmában egyetértene. Sharma & Mahajan (1980) úgy véli a vállalati kudarc fogalmának megállapítása az egyik legnehezebb feladat a csőd-kutatók számára.

Az üzleti kudarc fogalma Greenwald (1973) gazdasági szótára szerint, az üzleti tevékenység abbahagyása önkéntesen vagy bírósági eljárások okán, amely veszteséget okoz a hitelezőknek. A tevékenység leállításának számos különböző oka lehet, például tőkevesztés, elégtelen nyereség, visszavonulás, azonban ha a hitelezői követelések maradéktalanul teljesítve lettek, akkor a megszűnő vállalat nem tekinthető kudarcosnak (Dun and Bradstreet, 1978). Más

(rating) szolgáltatók alapján meghatározott üzleti kudarcot és nehézséget.

A tanulmányok nagy része, (73%) a fenti négy fogalmi keret közül csak egyet használt, 14 tanulmány (13%) alkalmazott kettő vagy három fogalmi keretet a csőd definíciójára, 15 tanulmány (14%) pedig nem jelölte meg, hogy milyen fogalmi keret szerint határozták meg a vállalati kudarcot, csődöt. A csődös vállalatok leggyakoribb meghatározási kerete a jogi megközelítés, amelyet 67 tanulmány alkalmazott. A jelentésalapú megközelítést 20 tanulmányban fedezhetjük fel, amelyek főként a tőzsdei jelentésekre összpontosítottak, konkrétan olyan momentumokra, mint a tőzsdei kivezetés vagy kizárás. A technikai és gazdasági definíciót 10-10 tanulmány alkalmazta. A technikai meghatározást jellemzően más, elsősorban jogi megközelítés mellett alkalmazták, azzal a céllal, hogy a sikertelen vállalatok csoportján belül, a csődösök mellett más osztályokat hozzanak létre, például a fizetéseképtelen vállalatok csoportját.

A magyar jogi megközelítésben a csődeljárás és a felszámolási eljárás fogalmi merülnek fel a fizetéseképtelen

vállalatok azonosítására, amelyeket a többször módosított 1991. évi XLIX. törvény szabályoz. A csődeljárás olyan eljárás, amelynek során az adós fél fizetési haladékokat kezdeményez, és csődegyezséget kísérel meg. A felszámolási eljárás célja, a fizetéseképtelen adós jogutód nélküli megszüntetése, valamint a hitelezők a törvényben meghatározott módon történő kielégítése. Mindkét eljárás esetében a fizetéseképtelenség ténye fennáll (Kristóf, 2004).

A csődelőrejelzés módszereinek összehasonlítása

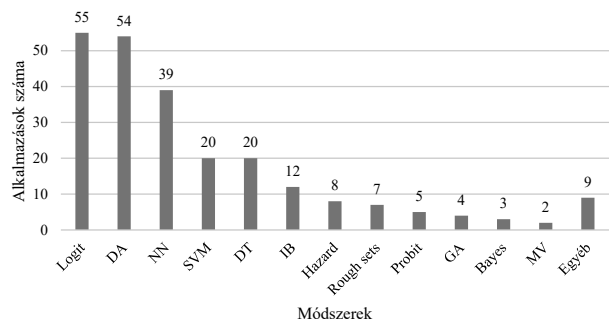
A 105 szakirodalmi cikk, 313 csődelőrejelző modellt mutat be, ebből adódóan az egy tanulmányra eső modellek mediánja 2, átlaga 2,98. A szakirodalmak 30%-a egyetlen módszert alkalmaz, 12 tanulmány logisztikus regressziót, 10 tanulmány diszkriminanciaanalízist és 9 tanulmány egyéb különböző módszereket használ.

Az összes 313 modelltől 238-nak van számszerűen ismertette az eredménye, amely az előrejelző pontosság, az illeszkedés, az AUC görbe alatti terület vagy a hibaarány bemutatásával valósul meg. E modelcsaládok gyakorisága a 2. ábrán látható. Amennyiben egy tanulmány egy adott eljárásalárhoz tartozó módszert egynél több alkalommal használt fel a modellezésre a finomhangolás vagy optimalizáció okán, a módszert csak egyszer vettem figyelembe. A diszkriminanciaanalízis és logisztikus regresszió eredményeit 54-55 tanulmányban fedezhetjük fel. A neurális háló módszerek közé kerültek besorolásra az egynél több rejtett réteget magába foglaló mély neuronhálók, mivel olyan kevés tanulmány alkalmazta, hogy nem volt rá mód önálló módszerként ismertetni. Ilyen tekintetben a neurális háló alapú módszer 39, a Support Vector Machine (SVM) 20 tanulmányban mutatkozott. A döntési fa alapú módszercsaládba a „statikus” rekurzív particionáló algoritmusok mellé bekerültek azok a „dinamikus” döntési fák is, amelyek az úgynevezett együttes (ensemble) metamódszereken alapulnak, ahol az alaptanuló algoritmusok homogén döntési fák, például a Véletlen erdő, a Bagging és a Boosting eljárások. A döntési fa eljárásalád besorolási pontossága 20 tanulmányban volt fellelhető. A példányalapú (Instance Based, IB) módszercsaládban olyan rokon módszerek kaptak helyet, mint az esetalapú következtetés (case based reasoning, CBR), a k-adik legközelebbi szomszéd (k-nearest neighbor, KNN), a tanuló vektor kvantálás (learning vector quantization, LVQ) és az önszerveződő térképek (self organizing maps, SOM), amelyek összesen 12 tanulmányt tettek összehasonlíthatóvá. A módszerek egy bizonyos részét kevesebb, mint tíz tanulmányban találhatjuk: a túlélés elemzést, a durva halmazok elméletén nyugvó eljárást, a probit elemzést, a genetikus algoritmusokat, a bayes-i eljárásaládot és a többségi szavazás módszerét. Az egyéb kategóriába tartozó 9 módszert csupán egy-egy tanulmányban fedezhetjük fel: Multivariate Adaptive Regression Splines (MARS), halmazott összeg módszer (Cumulative Sum, CUSUM), Együttes Robusz-tus Interpretatív Parszolás (Joint Robust Interpretive Parsing), Behavior-Knowledge Space, dichotóm osztályozás, Generalized Additive Model, Generalized Linear Model,

csoportos döntéshozói mechanizmus, Forward Intensity. A hibrid modelleknél a domináns módszerhez kapcsolom az adott tanulmány csődelőrejelző módszerét, így például Chaudhuri & De (2011) tanulmányban az SVM eljárásba integrált fuzzy logikán alapuló tagsági függvénnyel készült modell az SVM módszerek közé került besorolásra. Bizonyos esetekben a közzétett eredmények nem voltak összehasonlíthatók, például Neopythou és Mar Molinero (2004) tanulmányában a többdimenziós skálázás (MDS) segítségével létrejött logit modellnél nem állt rendelkezésre megfelelő teljesítménymutató.

2. ábra

Az alkalmazott módszerek gyakorisága, amelyek ismertették az eredményeket



Megjegyzés: Logit=logisztikus regresszió; DA=diszkriminanciaanalízis; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád; Hazard=túléléselemzés; Rough sets=durva halmazok; Probit=probit analízis; GA=genetikus algoritmus; Bayes=Bayes eljárásalád; MV=többségi szavazás.

Forrás: saját számítás

A módszerek pontossága

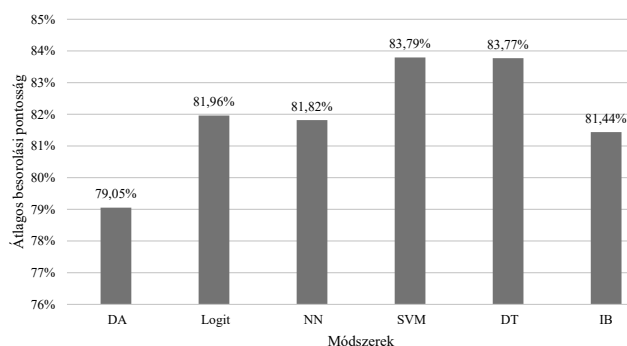
A szakirodalmi feldolgozás során az egyes módszerek átlagos besorolási pontosságait ismertetem, egy összehasonlító elemzés keretén belül. Az összehasonlíthatóságot nehezítette, hogy a tanulmányok néhol eltérő módszereket alkalmaznak a pontosságok bemutatására. A legtöbb tanulmány az átlagos besorolási pontosságot használta, amely a csődös vállalatok és a nem csődös vállalatok helyes besorolási arányát mutatja az összes, helyes és helytelen, besoroláshoz mérten. Néhány tanulmány csak a hibaarányokat ismertette, az elsőfajú hibát, amely a csődösök hibás besorolását jelzi és a másodfajú hibát, amely a nem csődösök hibás előrejelzését mutatja. Kevés tanulmány a pseudo R^2 -el és a ROC görbe alatti területtel (AUC) vizsgálta a modellek jóságát, hátrányosabb esetben a tanulmány egyáltalán nem ismertette a számszerű eredményeket. A tanulmányok 82,86%-a független tesztmintát használt, vagy egy a tanuló minta időperiódusától eltérő, „kívülről” időperiódust az előrejelző pontosság értékelésére.

Az összehasonlítás során az átlagos besorolási pontosságot használtam, mivel a legtöbb tanulmány ezzel értékeltette a módszereket. Számszerűen 238 modellnek van ismertette az eredménye, azonban az átlagos besorolási pontosság mutató 202 modell összehasonlítását tette lehetővé. Az összehasonlító vizsgálatból kimaradt

16 tanulmány 36 modellje, ugyanis egyéb teljesítmény-mutatót ismertettek a szerzők. Azon tanulmányoknál, ahol az első- és másodfajú hibaarányok voltak közölve, kiszámolható volt az átlagos besorolási pontosság mutató. Elsődlegesen a tesztmintán végzett modell eredményeket vettem figyelembe, és a csődhez legközelebbi időhorizont előrejelzését, amely túlnyomóan a csőd előtti első év volt. Amennyiben egy bizonyos módszercsaládot többször alkalmaztak egy tanulmányban a finomhangolás okán, vagy több mintán használták, akkor csak a legmagasabb pontosságú modellt tekintettem a tanulmány eredményének.

3. ábra

A leggyakoribb módszerek átlagos besorolási pontossága



Megjegyzés: DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád.

Forrás: saját számítás

Az irodalomkutatás 21 különböző csődelőrejelző módszercsaládot tárt fel, és összesen 202 modell eredményei hasonlíthatók össze az átlagos besorolási pontossággal, így egy módszer átlagosan 9,62 esetben merült fel, ezért a 9 és kevesebb alkalommal megjelenő módszereket nem vettem figyelembe. Ezen kívül azok az egyéb kategóriába tartozó módszerek, amelyek csak egyszer fordultak elő az irodalomkutatás során, kimaradtak az összehasonlításokból. A 3. ábra azon módszerek átlagos besorolási pontosságát mutatja, amelyek több mint 9 tanulmányban ismertették az adott módszer eredményeit.

A 3. ábrán az látszik, hogy az SVM módszere a legkiválóbb teljesítményű, közel 84%-os átlagos besorolási pontossággal, amelyet a döntési fa módszercsalád szorosan követ. A logisztikus regresszió harmadiknak tehető, átlagosan jobbnak bizonyult, mint a neuronháló teljesítménye. A példányalapú módszercsalád 81,44%-os átlagos pontosságot ért el, valamint a diszkriminanciaanalízisnek 79%-os az átlagos teljesítménye.

A 2. táblázat a leggyakoribb módszereknél alkalmazott teljes mintanagyság statisztikáit mutatja. Az alkalmazott minta elemszáma 22-től (lásd Sandin & Porporato, 2007) 100000-es vállalati mintanagyságig terjed (lásd Berg, 2007), a medián mintanagyság 166 és 254 között mozog az egyes módszereknél. A 22-23 elemű minta a statisztikai osztályozó módszereknél kevésnek tűnik, olyan tekintetben, hogy a gyakorlati ökölszabály a minimum 100 megfigyelés a tanuló mintában, azaz legyen legalább

2. táblázat

A teljes mintanagyság leíró statisztikái a megvizsgált módszereknél

	DA	Logit	NN	SVM	DT	IB
Minimum	22	23	23	36	23	23
Maximum	100 000	18 620	18 620	100 000	14 192	2 860
Medián	166	200	200	254	216	187
Átlag	3 389	1 559	1 559	4 703	1 894	641
Szórás	14 653	3 799	3 799	16 792	4 298	1 002

Megjegyzés: DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád.

Forrás: saját számítás

3. táblázat

A különböző mintanagyságú csoportok száma és az átlagos besorolási pontosságok

Mintanagyság csoport	Tanulmányok száma	Átlagos besorolási pontosság	Gyakori módszerek átlagos besorolási pontossága					
			DA	Logit	NN	SVM	DT	IB
1-99	20	81,96%	83,25%	79,84%	86,77%	94,57%	73,30%	92,94%
100-199	28	81,06%	78,88%	82,71%	83,67%	81,96%	88,00%	77,40%
200-499	20	80,92%	82,35%	79,85%	82,67%	81,69%	80,51%	81,97%
500-9.999	25	82,55%	74,46%	86,71%	80,07%	86,38%	87,37%	84,73%
10.000-100.000	12	76,45%	66,69%	79,06%	74,69%	81,49%	87,06%	N/A

Megjegyzés: DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád.

Forrás: saját számítás

50 fizetésképtelen megfigyelés egy azonos osztályarányú mintában (Engelmann, Hyden & Tache, 2003).

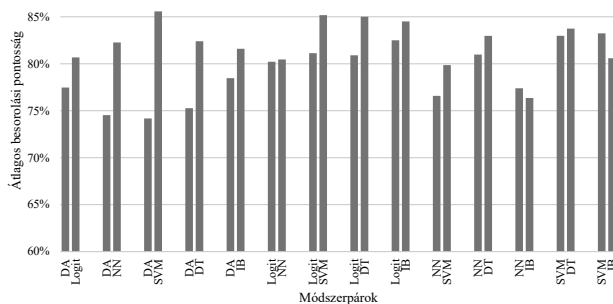
A mintanagyság szerint öt csoportba sorolva a tanulmányokat a 3. táblázatban tekinthetjük meg az átlagos besorolási pontosságokat. 100 alatti mintanagyság 20 tanulmányban figyelhető meg, az átlagos besorolási pontosság az SVM módszernél a legmagasabb. A 100-199 közötti mintanagyság a megvizsgált tanulmányok közt a leggyakoribb, itt a döntési fa eljáráscsalád teljesített átlagosan a legjobban. A 200-499 közötti mintánál a neuronháló bizonyult a legpontosabbnak. Az 500 feletti és a 10.000 feletti mintanagyságnál a döntési fa módszercsalád nyújtotta a legjobb teljesítményt. A példányalapú módszercsalád esetén nem volt mért eredmény a 10.000 mintaelemszám feletti csoportban, teljesítmény tekintetében az alsó-közép negyedbe sorolható a logisztikus regresszióval együtt. A diszkriminanciaanalízis majdnem mindenik mintacsoportban átlagosan alulmaradt.

Módszerpárok összehasonlítása

A közvetlenebb összehasonlításhoz, a módszerek páronkénti pontosságát tesztelhetjük, amelyek azonos tanulmányban, egyazon mintán lettek lefuttatva. A páronkénti összehasonlítások számát a 4. online mellékletben található táblázat foglalja össze, amely egyben azt is jelezi, hogy mely összehasonlítások ritkák, vagy hiányoznak a csodelőrejelzés szakirodalmának fő tanulmányaiból. Összesen 220 módszerpár összehasonlítás lehetőséget fel a szakirodalmakban, 47 féle módszerpár változathoz, így a páronkénti összehasonlítások átlaga 4,68. Ezért csak azokat az összehasonlítás párokat vettem figyelembe, ahol több mint négy eredménypár volt megtalálható, ez 14 módszerpár összemérését tette lehetővé, amelyek a 4. online melléklet táblázatában kiemelve szerepelnek.

4. ábra

A módszerek páronkénti átlagos besorolási pontosságai



Megjegyzés: Logit=logisztikus regresszió; DA=diszkriminanciaanalízis; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád.

Forrás: saját számítás

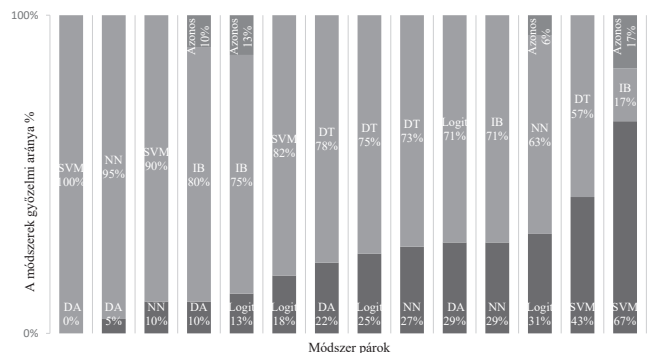
A 4. ábra a páronkénti összehasonlítások átlagos besorolási teljesítményét mutatja. A vizsgált módszerek közül a diszkriminanciaanalízis mindenik más módszerhez mérten alulmaradt. A másik gyengén teljesítő módszer a logisztikus regresszió, amely a páros összehasonlításban

átlagosan csak a diszkriminanciaanalízist múlta fölül, azonban kevéssel a neuronháló teljesítménye alatt maradt. A neurális háló az említett fenti két módszeren kívül a példányalapú eljáráscsaládot is fölülmúlta. Az SVM módszer az egyik legjobb teljesítményű, azonban a páros összemérésben átlagosan a döntési fa modellsalád látszik a legpontosabbnak, minden módszerhez viszonyítva jobbnak bizonyult.

A továbbiakban a módszerpárokat aszerint is elemezhetjük, hogy bizonyos tanulmányok, amelyek kettő vagy több módszert egyidejűleg teszteltek, két módszer tekintetében, melyiket jelölték meg pontosabbnak. Az átlagos besorolási pontosságokat használva a méréshez két tizedesjegy pontosságig, a páronkénti összehasonlítások számát a 4. online melléklet foglalja össze, kiemelve az elemzésben részt vett eseteket. A kivitelezés hasonló Alaka et al. (2018) tanulmányához.

5. ábra

A módszerek győzelmi aránya a közvetlen páronkénti összehasonlítások során



Megjegyzés: Logit=logisztikus regresszió; DA=diszkriminanciaanalízis; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú módszercsalád.

Forrás: saját számítás

Az 5. ábra a módszerek győzelmi arányát mutatja, amely az egyes módszereknél a páros összehasonlítás győztes esetszámának és az összes ugyanazon páros közvetlen összehasonlítás számának hányadosa. Az előző eredményekkel összhangban az látszik, hogy a döntési fa módszercsalád, bármely másik modellel összehasonlítva jobban teljesített, a tanulmányok nagyobb arányban jelölték meg pontosabb módszernek. A legszorosabb verseny az SVM-mel szemben volt 57%-os győzelmi aránnyal a döntési fa javára. Az SVM a döntési fa módszercsaládot kivéve, minden más módszerrel szemben jobban teljesített, ez megegyezik a 4. ábra jelzéseivel is. A példányalapú módszercsalád felülmúlta a neurális hálót és a logisztikus regressziót, az SVM módszerrel szemben pedig alulmaradt. A neuronháló a páros összehasonlításban 63% arányban pontosabbnak bizonyult, mint a logisztikus regresszió, és egy kis részben megegyező eredményre jutott a két módszer. A logisztikus regresszió és diszkriminanciaanalízis viszonyában 71%-os volt a logit győzelmi aránya. A diszkriminanciaanalízis a legkevésbé pontos technikának bizonyult, minden párosításnál nagy arányban alulmaradt.

A vizsgálatokat követően a különböző eredmények szintetizálásával a módszerek rangsora megállapítható. A 4. táblázat egy egyszerű pontszám módszer segítségével összefoglalja az eddigi eredményeket és rangsort állít a „legpontosabb” módszer tekintetében.

4. táblázat

Különböző vizsgálati szempontok pontszáma és a módszerek rangsora

	DA	Logit	NN	SVM	DT	IB
Egyéni	1	4	3	6	5	2
Mintanagyság csoport	1	3	4	5	6	2
Páros pontosság	1	2	4	5	6	3
Páros győzelmi arány	1	2	3	5	6	4
Összesen	4	11	14	21	23	11
Rangsor	5	4	3	2	1	4

Megjegyzés: DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú.

Forrás: saját számítás

A táblázat egyes sorai egytől hatig skálán kerültek pontozásra, az egyes a legrosszabb teljesítményt, a hatos a legjobb teljesítményt takarja. Négy különböző szempont szerint értékelhetjük a módszereket.

Az *egyéni* szempont a módszerek egyéni átlagos besorolási pontossága alapján történő értékelését jelenti, azaz a legnagyobb pontosságú modell kapott hat pontot, majd a következő ötöt és így tovább.

A *mintanagyság csoport* szerinti szempont az öt különböző minta csoport mérete alapján elért eredményeket tükrözi. A 3. táblázat alapján felírhatunk egy súlyozott pontszám táblázatot, amelyben a döntési fa módszere jön ki győztesnek, azt követi az SVM majd a neurális háló és így tovább.

A *páros pontosság*, a páronkénti átlagos besorolási pontosságok alapján való értékelést jelenti. A döntési fa módszercsalád mindenik módszert felülmúlja, tehát hat pontot kapott. Az SVM csak a döntési fa módszerrel szemben maradt alul tehát öt pontot kapott. A neurális háló felülmúlja a példányalapú módszercsaládot, a logisztikus regressziót és a diszkriminanciaanalízist ezért négy pontot kapott. A logisztikus regresszió és példányalapú következtetés viszonyában a logit maradt alul, a diszkriminanciaanalízis pedig a legalacsonyabb pontosságokat mutatta.

A *páros győzelmi arány* pedig a módszerek győzelmi aránya a közvetlen páronkénti összehasonlítások során, amely hasonló módon és hasonló eredményeket láthatunk, mint a páros besorolási pontosságoknál, annyi különbséggel, hogy a példányalapú módszercsalád nagyobb arányban volt sikeres módszer, mint a neurális háló ezért négy pontot kapott.

Az oszlopok összegzése után a legmagasabb pontszámot elérő módszer kapta a legmagasabb rangot. A feldolgozott szakirodalmak alapján a legmagasabb teljesítményű

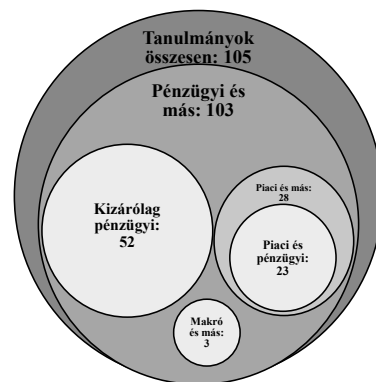
módszernak a döntési fa módszercsalád bizonyult, ezt követi az SVM módszere, majd a neurális háló. A példányalapú eljárásalád és a logisztikus regresszió a ranglétra azonos fokán állnak a negyedik helyen. A hagyományos statisztikai módszernek számító diszkriminanciaanalízis ez alapján a legkisebb teljesítményű módszer.

A magyarázó változók típusai

A szisztematikus irodomelemzés második célkitűzése a tanulmányokban felhasznált magyarázó változók hozzáadott értékének vizsgálata. A vállalatok teljesítményének mérésére leggyakrabban használt eszközök a pénzügyi kimutatások, és ez a csődelőrejelzésre is igaz, mivel az általánosan alkalmazott magyarázó változók kalkulált pénzügyi mutatószámok (Virág, 2001; Virág et al., 2013). Elemzőként fontos szem előtt tartani, hogy a mutatók a múlt adatait hordozzák, ugyanakkor a pénzügyi mutatók teljesítményértékelése relatív, hiszen ágazatonként, sőt vállalatonként változik a működési feltétel, nem mellesleg a különböző országok eltérő pénzügyi kimutatása nehezíti az összehasonlíthatóságot (Virág, 2001). A pénzügyi adatok bizonyos hiányosságuk ellenére domináns szerepet játszanak a csődmodellezésben (Nyitrai, 2017).

6. ábra

A magyarázó változók főbb típusai és a tanulmányok száma



Forrás: saját számítás

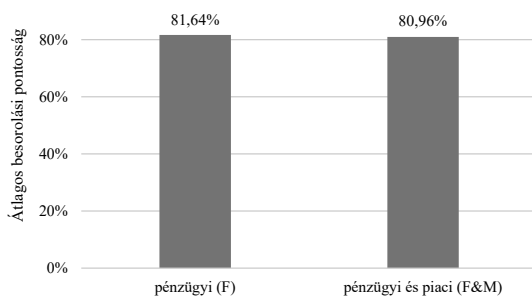
A tanulmányban megerősítést nyert, hogy a pénzügyi arányszámok a leggyakoribb magyarázó változók a csődelőrejelzésben. A 6. ábra szematikusan bemutatja a tanulmányok számát és a főbb változó típusok felmerülésének csoportosítását. A megvizsgált 105 tanulmányból 103-ban fordultak elő pénzügyi mutatók, mint magyarázó változók, ebből 52 tanulmányban csak kizárólag pénzügyi mutatókat találunk. 28 tanulmányban volt fellelhető piaci mutató, valamely más típusú változó mellett, ebből 23 tanulmányban piaci és pénzügyi mutatók együttes alkalmazása volt megfigyelhető. Önmagában csak piaci mutatókat alkalmazó kutatás nem volt fellelhető. Makrogazdasági változókat mindössze három tanulmány tartalmazott pénzügyi és egyéb mutatók mellett, ezért a makrogazdasági változók hozzáadott értékének mélyebb vizsgálata nem volt kivitelezhető. Kevés tanulmányban iparági, vállalatirányítási,

menedzsmentattitúd- és egyéb nem pénzügyi mutatókat is fellelhetünk, azonban ezek kis számosságukat tekintve nem voltak elegendőek egy átfogó elemzés lefolytatásához.

Az adatok átfedés nélkül a kizárólag pénzügyi, valamint a pénzügyi és piaci mutatókat alkalmazó tanulmányok összehasonlító vizsgálatát teszik lehetővé. Ha a második csoportban a pénzügyi és piaci változók mellett egyéb mutatók is szerepeltek a modellben, azok a tanulmányok nem kerültek be a vizsgálatba. A két változócsoporthoz azonos mintán való tesztelése egyetlen esetben sem fordult elő, azonban van rá példa, hogy egyetlen tanulmány több módszert alkalmazott ugyanazon mintán. A különböző módszereken alkalmazott pénzügyi, valamint pénzügyi és piaci mutatók alkalmazásának számát az 5. táblázat foglalja össze. A két mutatócsoport eltéréseit a módszerek átlagos besorolási pontosságával mérjük. Az 5. táblázatban összefoglalt 12 módszer átlagos besorolása a pénzügyi változók esetén közel 82%, a pénzügyi és piaci mutatók együttes alkalmazása esetén megközelítőleg 81%, amely a 7. ábrán szemléletesen látható.

7. ábra

A két vizsgált mutatócsoport átlagos besorolási pontossága



Forrás: saját számítás

A továbbiakban a mélyrehatóbb összehasonlításhoz a két változócsoporthoz, a kizárólag pénzügyi valamint a pénzügyi és piaci változók, azonos módszerrel elért pontosságait tesztelhetjük. Továbbra is az átlagos besorolási pontosságokat használjuk, és azokat a módszereket vesszük figyelembe, amelyeknél volt legalább 4-4 alkalmazás mindkét változócsoporthoz, így a diszkriminanciaanalízis, a logisztikus regresszió, a neurális háló, az SVM és a példányalapú módszereknél történt meg az összehasonlítás, amelyek az 5. táblázatban kiemelve szerepelnek.

A két vizsgált változócsoporthoz módszerekénti átlagos besorolási pontosságait a 8. ábra foglalja össze. A kizárólag pénzügyi mutatók átlagosan felülmúlták a piaci mutatókkal kombinált pénzügyi mutatókat a diszkriminanciaanalízis, a neurális háló és a példányalapú eljárás család esetében. A piaci és pénzügyi mutatócsoport csak a logisztikus regresszió módszerénél bizonyult átlagosan jobbnak a kizárólag pénzügyi mutatókat alkalmazó modellekkel szemben. Az SVM módszer esetén pedig nem lehet egyértelműen győztes változótípust meghatározni.

5. táblázat

A különböző módszerek pénzügyi, valamint pénzügyi és piaci magyarázó változók alkalmazásának a száma

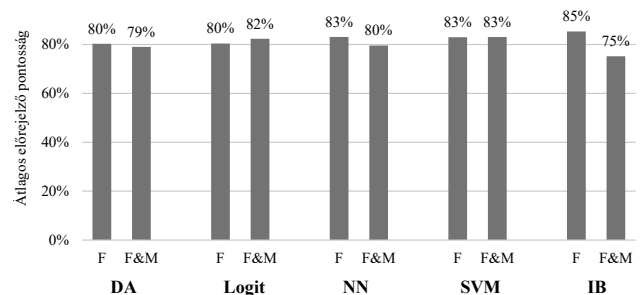
Mutató csoport \ Módszer	DA	Logit	NN	SVM	DT	IB	Hazard	Probit	Rough sets	GA	Bayes	MV	Összesen
pénzügyi (F)	27	21	21	9	9	6	1	3	3	3	1	1	105
pénzügyi és piaci (F&M)	11	9	9	4	2	4	3	2	1	1	1	1	48

Megjegyzés: DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú; Hazard=túléléselemzés; Rough sets=durva halmazok; Probit=probit analízis; GA=genetikus algoritmus; Bayes=Bayes eljárás család; MV=többségi szavazás.

Forrás: saját számítás

8. ábra

A két változócsoporthoz előrejelző pontossága különböző módszereknél



Megjegyzés: F=pénzügyi mutatók; F&M=pénzügyi és piaci mutatók. DA=diszkriminanciaanalízis; Logit=logisztikus regresszió; NN=neurális háló; SVM=Support Vector Machine; IB=példányalapú modellcsalád.

Forrás: saját számítás

Az eredmények szerint a megvizsgált tanulmányokban a pénzügyi mutatók mellett a piaci mutatók használatának átlagosan nem mutatkozik jelentős hozzáadott értéke az előrejelző pontosságok tekintetében. Ez ellentmond a tanulmány kezdeti elvárásának, azonban magyarázatot adhat arra, hogy miért kevésbé elterjedtek a piaci mutatószámok a csodelőrejelzés szakirodalmában.

A tanulmányokban alkalmazott változók száma 2 és 190 közötti, átlagosan tanulmányonként 25 változót jelöltek ki, a változószelekció után pedig átlagosan 16 változót használtak a létrejött modellek. A modellezésre használt változók kiválasztására a tanulmányok szerzői 54,29%-ban valamely előző tanulmányt említettek, jellemzően Altman (1968) ikonikus változóit. A tanulmányok 25,71%-a diszkriminanciaanalízis és logisztikus regressziót alkalmazott egy előzetes változószelekcióra, valamint 15,24% ad hoc változószelekciót alkalmazott, 4,76% pedig korrelációs statisztika alapján döntötte el, mely változókat tartsa meg a modellezésre. A 6. táblázat a leggyakrabban előforduló változókat mutatja, valamennyi pénzügyi adatból kalkulált arányszám.

6. táblázat

Az öt leggyakoribb változó a megvizsgált tanulmányokban

Változó	A tanulmányok száma, amely alkalmazta
Forgóeszközök/Rövid lejáratú kötelezettségek	55
Nettó árbevétel/Összes eszközállomány	52
Működő tőke/Összes eszközállomány	45
Adózott eredmény/Összes eszközállomány	45
Összes kötelezettség/Összes eszközállomány	38

Forrás: saját számítás

A legtöbbször fellelhető változó, az általános likviditási rátaként ismert forgóeszközök és rövid lejáratú kötelezettségek aránya, 55 tanulmányban volt megfigyelhető. Gyakoriság szerint második az eszközarányos árbevétel mutató. A harmadik helyen a működő tőke és összes eszköz aránya áll, ahol működő tőke a forgóeszközök és rövid lejáratú kötelezettségek különbsége. Azonos előfordulással láthatjuk az eszközarányos nyereség (ROA) jövedelmezőségi mutatót. A kötelezettségek aránya eladósodottsági mutató 38 tanulmányban volt megfigyelhető. A leggyakoribb piaci alapú mutatók az egy részvényre jutó nyereség (EPS), az egy részvényre jutó pénzáramlás, a részvényárfolyam volatilitása és a piaci kapitalizáció volt, amelyeket egyenként kevesebb, mint tíz tanulmányban lehetett felfedezni.

A változók széles palettáját látva érdekes kérdés, hogy az egyre több független változó nagyobb mértékben hozzájárul a modell pontosságához? Kétváltozós lineáris regresszióanalízis segítségével tesztelhető a változók számának hatása a besorolási pontosságokra, ahol a függő változó az átlagos besorolási pontosság, a független változó a magyarázó változók száma változószelekció után. Ha egy tanulmány több különböző modell eredményét ismertette, azok átlagát alkalmaztam függő változónak. 16 tanulmánynál nem volt fellelhető az átlagos besorolási pontosság, azok kimaradtak az elemzésből. Az eredmények alapján az R² determinációs együttható 0,26%, és a regressziós modell nem szignifikáns (p-érték=0,633), tehát nem figyelhető meg valós kapcsolat a magyarázó változók száma és az átlagos besorolási pontosság között a vizsgált tanulmányokban.

Összefoglalás

Szisztematikus irodalomkutatás segítségével a csődelőrejelzés magasan hivatkozott tanulmányait vizsgáltam, három tudományos adatbázist felhasználva. A vizsgálat egy nagyon széles időszakot és földrajzilag egy nemzetközi színteret fed le, azzal a céllal, hogy feltárjam a csődelőrejelzés szakirodalmában fellelhető legrangosabb módszereket.

Az irodalomfeldolgozás során körvonalazódott, hogy a csődelőrejelzés szakirodalmában nincs egy egységesen elfogadott fogalom a csőd meghatározására, a tanulmányok azonosan értelmezik a csőd fogalmát a vállalati fizetési képtelenség, a pénzügyi nehézség, a fizetés elmulasztása és a kudarc fogalmával.

A tanulmány hat elterjedt módszer részletes összehasonlítását tartalmazza, és rávilágít, hogy a legáltalánosabb módszer a csődelőrejelzésben a többváltozós diszkriminanciaanalízis és a logisztikus regresszió. Ennek egyszerű magyarázata, hogy a két legrégebbi módszerről van szó. A kapott eredmények alapján kimutatható, hogy átlagosan mind a hat vizsgált módszer alkalmas a csődelőrejelzésre, a véletlen találgatásnál lényegesen jobb teljesítményre képesek, 80% vagy afeletti átlagos besorolási pontosságokkal. Az összehasonlító eredmények azt mutatják, hogy a döntési fa eljáráscsaládba tartozó módszerek kiemelkedő teljesítményre képesek. A gépi tanulási módszerek családjába tartozó SVM eljárás a második „legjobb” módszernek mondható, míg a neuronháló a harmadik. A példányalapú eljáráscsalád és logisztikus regresszió tekintetében nincs egyértelmű győztes, mindkettő teljesítménye mérsékeltnak tekinthető. A diszkriminanciaanalízis a legkevésbé pontos módszer a vizsgáltak közül. Ez az eredmény egybecseng a magyar csődelőrejelzés átfogó elemzésével, amelyben Kristóf és Virág (2019) a hagyományos diszkriminanciaanalízis modelteljesítmény problémáira hívja fel a figyelmet.

Egy másik kérdésköre a tanulmánynak, a csődelőrejelzés szakirodalmában alkalmazott magyarázó változók és a hozzájuk köthető előrejelző pontosság kapcsolatának vizsgálata. A feldolgozott szakirodalmak 98%-a pénzügyi mutatókat alkalmaz a vállalati csőd becslésére, valamint a pénzügyi mutatók mellett a piaci mutatók alkalmazásával, átlagosan közel azonos előrejelző pontosságok érhetőek el, mint a csak kizárólag pénzügyi mutatókat tartalmazó modelleknél. Az eredmények alapján a piaci mutatók átlagosan nem rendelkeznek többletinformációval és jelentős hozzáadott értékkel a pénzügyi mutatókhoz viszonyítva a csőd előrejelzésére. Az elemzés alapján az sem bizonyítható, hogy egyre több magyarázó változó, vagy egyre nagyobb mintanagyság felhasználása a modellek magasabb pontosságát eredményezné.

A tanulmány az objektív értékelésre összpontosít, amely mellet lényeges a korlátait is szem előtt tartani. Az eredmények egy általános tájékoztatásként kezelendők a csődelőrejelzés témájában a legelterjedtebb módszerekről és a kapcsolódó szempontokról. A vállalati csődelőrejelzés egy nagyon népszerű kutatási terület, nagyszámú tudományos munkával rendelkezik, amelyen belül a magyar nyelvű és az angol nyelvtől eltérő más nyelven közzétett színvonalas szakirodalmak felülvizsgálatára nem volt mód.

Felhasznált irodalom

Abidali, A. F. & Harris, F. (1995). A methodology for predicting company failure in the construction industry. *Construction Management and Economics*, 13(3), 189-196.
<https://doi.org/10.1080/01446199500000023>

- Ahn, H. & Kim, K. J. (2009). Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Applied Soft Computing*, 9(2), 599-607.
<https://doi.org/10.1016/j.asoc.2008.08.002>
- 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*, 94, 164-184.
<https://doi.org/10.1016/j.eswa.2017.10.040>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
<https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I., Eom, Y. H. & Kim, D. W. (1995). Failure prediction: evidence from Korea. *Journal of International Financial Management & Accounting*, 6(3), 230-249.
<https://doi.org/10.1111/j.1467-646x.1995.tb00058.x>
- Alfaro, E., García, N., Gámez, M. & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, 45(1), 110-122.
<https://doi.org/10.1016/j.dss.2007.12.002>
- Anandarajan, M., Lee, P. & Anandarajan, A. (2001). Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. *Intelligent Systems in Accounting, Finance & Management*, 10(2), 69-81.
<https://doi.org/10.1002/isaf.199>
- Appiah, K. O., Chizema, A., & Arthur, J. (2015). Predicting corporate failure: a systematic literature review of methodological issues. *International Journal of Law and Management*, 57(5), 461-485.
<http://dx.doi.org/10.1108/IJLMA-04-2014-0032>
- Aziz, A., Emanuel, D. C. & Lawson, G. H. (1988). Bankruptcy prediction-an investigation of cash flow based models (1). *Journal of Management Studies*, 25(5), 419-437.
<https://doi.org/10.1111/j.1467-6486.1988.tb00708.x>
- Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63-93.
<https://doi.org/10.1016/j.bar.2005.09.001>
- Baldwin, J. & Glezen, G. W. (1992). Bankruptcy prediction using quarterly financial statement data. *Journal of Accounting, Auditing & Finance*, 7(3), 269-285.
<https://doi.org/10.1177/0148558x9200700301>
- Barboza, F., Kimura, H. & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417.
<https://doi.org/10.1016/j.eswa.2017.04.006>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
<https://doi.org/10.2307/2490171>
- Bellovary, J., Giacomino, D. & Akers, M. (2007). A review of bankruptcy prediction studies: 1930-present. *Journal of Finance Education*, 33, 1-42. https://publications.marquette.edu/cgi/viewcontent.cgi?article=1025&context=account_fac
- Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2), 129-143.
<https://doi.org/10.1002/asmb.658>
- Beynon, M. J. & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega*, 29(6), 561-576.
[https://doi.org/10.1016/s0305-0483\(01\)00045-7](https://doi.org/10.1016/s0305-0483(01)00045-7)
- Boritz, J. E., Kennedy, D. B. & Albuquerque, A. D. M. E. (1995). Predicting corporate failure using a neural network approach. *Intelligent Systems in Accounting, Finance and Management*, 4(2), 95-111.
<https://doi.org/10.1002/j.1099-1174.1995.tb00083.x>
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification and regression trees*. California: Wadsworth.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
<https://doi.org/10.1007/bf00058655>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/a:1010933404324>
- Bruno, A. V. & Leidecker, J. K. (1988). Causes of new venture failure: 1960s vs. 1980s. *Business Horizons*, 31(6), 51-56.
[https://doi.org/10.1016/0007-6813\(88\)90024-9](https://doi.org/10.1016/0007-6813(88)90024-9)
- Casey, C. & Bartczak, N. (1985). Using operating cash flow data to predict financial distress: Some extensions. *Journal of Accounting Research*, 23(1), 384-401.
<https://doi.org/10.2307/2490926>
- Charitou, A., Neophytou, E. & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), 465-497.
<https://doi.org/10.1080/0963818042000216811>
- Chaudhuri, A. & De, K. (2011). Fuzzy support vector machine for bankruptcy prediction. *Applied Soft Computing*, 11(2), 2472-2486.
<https://doi.org/10.1080/0963818042000216811>
- Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537-569.
<https://doi.org/10.1007/s10679-004-6279-6>
- Chen, W. S. & Du, Y. K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2), 4075-4086.
<https://doi.org/10.1016/j.eswa.2008.03.020>
- Chen, M. Y. (2011a). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications*, 62(12), 4514-4524.
<https://doi.org/10.1016/j.camwa.2011.10.030>
- Chen, M. Y. (2011b). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9), 11261-11272.
<https://doi.org/10.1016/j.eswa.2011.02.173>

- Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal of Business Research*, 68(5), 1012-1025. <https://doi.org/10.1016/j.jbusres.2014.10.003>
- Coats, P. & Fant, L. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, 22(3), 142-155. <https://doi.org/10.2307/3665934>
- Constand, L. R. & Yazdipour, R. (2011). Firm failure prediction models: a critique and a review of recent developments. In Yazdipour, R. (Ed.), *Advances in Entrepreneurial Finance: With Applications from behavioral Finance and Economics* (pp. 185-204). New York: Springer Science and Business Media. https://doi.org/10.1007/978-1-4419-7527-0_10
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/bf00994018>
- Dewaelheyns, N. & Van Hulle, C. (2006). Corporate failure prediction modeling: Distorted by business groups' internal capital markets? *Journal of Business Finance & Accounting*, 33(5-6), 909-931. <https://doi.org/10.1111/j.1468-5957.2006.00009.x>
- 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*, 38(3), 1866-1875. <https://doi.org/10.1016/j.eswa.2010.07.117>
- Delen, D., Kuzey, C. & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, 40(10), 3970-3983. <https://doi.org/10.1016/j.eswa.2013.01.012>
- Dimitras, A. I., Slowinski, R., Susmaga, R. & Zopounidis, C. (1999). Business failure prediction using rough sets. *European Journal of Operational Research*, 114(2), 263-280. [https://doi.org/10.1016/s0377-2217\(98\)00255-0](https://doi.org/10.1016/s0377-2217(98)00255-0)
- Du Jardin, P. & Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems*, 51(3), 701-711. <https://doi.org/10.1016/j.dss.2011.04.001>
- Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), 286-303. <https://doi.org/10.1016/j.ejor.2014.09.059>
- Duan, J. C., Sun, J. & Wang, T. (2012). Multiperiod corporate default prediction-A forward intensity approach. *Journal of Econometrics*, 170(1), 191-209. <https://doi.org/10.1016/j.jeconom.2012.05.002>
- Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635-665. <https://doi.org/10.1016/j.jfineco.2005.10.011>
- Dun and Bradstreet (1978). *The Business Failure Record*. New York: Dun and Bradstreet Inc.
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, 7(2), 1477-1493. <https://doi.org/10.2307/2329929>
- Engelmann, B., Hayden, E., & Tasche, D. (2003). *Measuring the discriminative power of rating systems*. Discussion Paper, Series 2. Banking and financing Supervision. Deutsche Bundesbank, Frankfurt. <https://www.bundesbank.de/resource/blob/704150/b9fa10a16dfff-3c98842581253f6d141/mL/2003-10-01-dkp-01-data.pdf>
- Etemadi, H., Rostamy, A. A. A. & Dehkordi, H. F. (2009). A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications*, 36(2), 3199-3207. <https://doi.org/10.1016/j.eswa.2008.01.012>
- Fan, A. & Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Vol. 6*, (pp. 354-359). Piscataway: IEEE. <https://doi.org/10.1109/ijcnn.2000.859421>
- Fedorova, E., Gilenko, E. & Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40(18), 7285-7293. <https://doi.org/10.1016/j.eswa.2013.07.032>
- Fletcher, D. & Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information & Management*, 24(3), 159-167. [https://doi.org/10.1016/0378-7206\(93\)90064-z](https://doi.org/10.1016/0378-7206(93)90064-z)
- Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *International Conference on Machine Learning*, 96, 148-156. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.6252&rep=rep1&type=pdf>
- Frydman, H., Altman, E. I. & Kao, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance*, 40(1), 269-291. <https://doi.org/10.1111/j.1540-6261.1985.tb04949.x>
- Geng, R., Bose, I. & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, 241(1), 236-247. <https://doi.org/10.1016/j.ejor.2014.08.016>
- Gentry, J. A., Newbold, P. & Whitford, D. T. (1985). Classifying bankrupt firms with funds flow components. *Journal of Accounting Research*, 23(1), 146-160. <https://doi.org/10.2307/2490911>
- Gombola, M. J., Haskins, M. E., Ketz, J. E. & Williams, D. D. (1987). Cash flow in bankruptcy prediction. *Financial Management*, 16(4), 55-65. <https://doi.org/10.2307/3666109>

- Greenwald, D. (1973). *The McGraw-Hill Dictionary of Modern Economics: A Handbook of Terms and Organizations*. New York: McGraw-Hill Book Company.
- Grice, J. S. & Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, 54(1), 53-61. [https://doi.org/10.1016/s0148-2963\(00\)00126-0](https://doi.org/10.1016/s0148-2963(00)00126-0)
- Hámori, G. (2001). A CHAID alapú döntési fák jellemzői. *Statisztikai Szemle*, 79(8), 703-710. https://www.ksh.hu/statszemle_archive/2001/2001_08/2001_08_703.pdf
- Hernandez Tinoco, M. & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30, 394-419. <https://doi.org/10.1016/j.irfa.2013.02.013>
- Hertz, J. – Krogh, A. & Palmer, R. (1991). *Introduction to the theory of neural computing*. New York: Addison Wesley.
- Hillegeist, S. A., Keating, E. K., Cram, D. P. & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), 5-34. <https://doi.org/10.1023/b:rast.0000013627.90884.b7>
- Horta, I. M. & Camanho, A. S. (2013). Company failure prediction in the construction industry. *Expert Systems with Applications*, 40(16), 6253-6257. <https://doi.org/10.1016/j.eswa.2013.05.045>
- Hu, Y-C. & Tseng, F-M. (2007). Functional-link net with fuzzy integral for bankruptcy prediction. *Neurocomputing*, 70, 2959-2968. <https://doi.org/10.1016/j.neucom.2006.10.111>
- Hua, Z., Wang, Y., Xu, X., Zhang, B. & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 33(2), 434-440. <https://doi.org/10.1016/j.eswa.2006.05.006>
- Huang, S. M., Tsai, C. F., Yen, D. C. & Cheng, Y. L. (2008). A hybrid financial analysis model for business failure prediction. *Expert Systems with Applications*, 35(3), 1034-1040. <https://doi.org/10.1016/j.eswa.2007.08.040>
- Johnsen, T. & Melicher, R. W. (1994). Predicting corporate bankruptcy and financial distress: Information value added by multinomial logit models. *Journal of Economics and Business*, 46(4), 269-286. [https://doi.org/10.1016/0148-6195\(94\)90038-8](https://doi.org/10.1016/0148-6195(94)90038-8)
- Jones, S. & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review*, 79(4), 1011-1038. <https://doi.org/10.2308/accr.2004.79.4.1011>
- Jones, S. & Hensher, D. A. (2007). Modelling corporate failure: A multinomial nested logit analysis for unordered outcomes. *The British Accounting Review*, 39(1), 89-107. <https://doi.org/10.1016/j.bar.2006.12.003>
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127. <https://doi.org/10.2307/2986296>
- Kim, M. J. & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications*, 37(4), 3373-3379. <https://doi.org/10.1016/j.eswa.2009.10.012>
- Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal*, 31(3), 441-468. <https://doi.org/10.1080/0264206802712848>
- Korol, T. (2013). Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, 31, 22-30. <https://doi.org/10.1016/j.econmod.2012.11.017>
- Kristóf, T. (2004). Mesterséges intelligencia a csődelőrejelzésben. *Jövőtanulmányok*, 21.
- Kristóf, T. (2005). Szervezetek jövőbeni fennmaradása különböző megközelítésekben. *Vezetéstudomány*, 36(9), 15-23. <https://doi.org/10.14267/VEZTUD.2005.09.02>
- Kristóf, T. (2018). A case-based reasoning alkalmazása a hazai mikroállalkozások csődelőrejelzésére. *Statisztikai Szemle*, 96(11-12), 1109-1128. <https://doi.org/10.20311/stat2018.11-12.hu1109>
- Kristóf, T. & Virág, M. (2019). A csődelőrejelzés fejlődéstörténete Magyarországon. *Vezetéstudomány*, 50(12), 62-73. <https://doi.org/10.14267/veztud.2019.12.06>
- Kristóf, T. & Virág, M. (2020). A comprehensive review of corporate bankruptcy prediction in Hungary. *Journal of Risk and Financial Management*, 13(2), 35. <https://doi.org/10.3390/jrfm13020035>
- Laitinen, E. K. (1992). Prediction of failure of a newly founded firm. *Journal of Business Venturing*, 7(4), 323-340. [https://doi.org/10.1016/0883-9026\(92\)90005-c](https://doi.org/10.1016/0883-9026(92)90005-c)
- Laitinen, E. K. (1993). Financial predictors for different phases of the failure process. *Omega*, 21(2), 215-228. [https://doi.org/10.1016/0305-0483\(93\)90054-o](https://doi.org/10.1016/0305-0483(93)90054-o)
- Lau, A. H. L. (1987). A five-state financial distress prediction model. *Journal of Accounting Research*, 25(1), 127-138. <https://doi.org/10.2307/2491262>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lee, K., Booth, D. & Alam, P. (2005). A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems with Applications*, 29(1), 1-16. <https://doi.org/10.1016/j.eswa.2005.01.004>
- Lee, S. & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, 40(8), 2941-2946. <https://doi.org/10.1016/j.eswa.2012.12.009>
- Leshno, M. & Spector, Y. (1996). Neural network prediction analysis: The bankruptcy case. *Neurocomputing*, 10(2), 125-147. [https://doi.org/10.1016/0925-2312\(94\)00060-3](https://doi.org/10.1016/0925-2312(94)00060-3)

- Li, H. & Sun, J. (2008). Ranking-order case-based reasoning for financial distress prediction. *Knowledge-based Systems*, 21(8), 868-878. <https://doi.org/10.1016/j.knosys.2008.03.047>
- Li, H. & Sun, J. (2009). Majority voting combination of multiple case-based reasoning for financial distress prediction. *Expert Systems with Applications*, 36(3), 4363-4373. <https://doi.org/10.1016/j.eswa.2008.05.019>
- Li, H. & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples—Evidence from the Chinese hotel industry. *Tourism Management*, 33(3), 622-634. <https://doi.org/10.1016/j.tourman.2011.07.004>
- Liang, D., Lu, C. C., Tsai, C. F. & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561-572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- Lin, T. H. (2009). A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models. *Neurocomputing*, 72(16-18), 3507-3516. <https://doi.org/10.1016/j.neucom.2009.02.018>
- Lin, F., Liang, D., Yeh, C. C. & Huang, J. C. (2014). Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, 41(5), 2472-2483. <https://doi.org/10.1016/j.eswa.2013.09.047>
- Liu, D., Li, T. & Liang, D. (2014). Incorporating logistic regression to decision-theoretic rough sets for classifications. *International Journal of Approximate Reasoning*, 55(1), 197-210. <https://doi.org/10.1016/j.ijar.2013.02.013>
- Luoma, M. & Laitinen, E. K. (1991). Survival analysis as a tool for company failure prediction. *Omega*, 19(6), 673-678. [https://doi.org/10.1016/0305-0483\(91\)90015-1](https://doi.org/10.1016/0305-0483(91)90015-1)
- Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management*, 33(1), 8-20.
- Mayew, W. J., Sethuraman, M. & Venkatachalam, M. (2015). MD&A disclosure and the firm's ability to continue as a going concern. *The Accounting Review*, 90(4), 1621-1651. <https://doi.org/10.2308/accr-50983>
- McKee, T. E. (2000). Developing a bankruptcy prediction model via rough sets theory. *Intelligent Systems in Accounting, Finance & Management* 9(3), 159-173. [https://doi.org/10.1002/1099-1174\(200009\)9:3%3C159::aid-isaf184%3E3.0.co;2-c](https://doi.org/10.1002/1099-1174(200009)9:3%3C159::aid-isaf184%3E3.0.co;2-c)
- McKee, T. E. & Greenstein, M. (2000). Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 19(3), 219-230. [https://doi.org/10.1002/\(sici\)1099-131x\(200004\)19:3%3C219::aid-for752%3E3.0.co;2-j](https://doi.org/10.1002/(sici)1099-131x(200004)19:3%3C219::aid-for752%3E3.0.co;2-j)
- McKee, T. E. (2003). Rough sets bankruptcy prediction models versus auditor signalling rates. *Journal of Forecasting*, 22(8), 569-586. <https://doi.org/10.1002/for.875>
- Mensah, Y. M. (1983). The differential bankruptcy predictive ability of specific price level adjustments: some empirical evidence. *Accounting Review*, 58(2), 228-246. <https://www.jstor.org/stable/246832>
- Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research*, 22(1), 380-395. <https://doi.org/10.2307/2490719>
- Min, J. H. & Jeong, C. (2009). A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36(3), 5256-5263. <https://doi.org/10.1016/j.eswa.2008.06.073>
- Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. (2009). The PRISMA Group Preferred Reporting Items for Systematic Reviews and Meta Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7). <https://doi.org/10.1371/journal.pmed.1000097>
- Mossman, C. E., Bell, G. G., Swartz, L. M. & Turtle, H. (1998). An empirical comparison of bankruptcy models. *Financial Review*, 33(2), 35-54. <https://doi.org/10.1111/j.1540-6288.1998.tb01367.x>
- Murthy, S. K., Kasif, S., Salzberg, S. & Beigel, R. (1993). OCl: A randomized algorithm for building oblique decision trees. *Proceedings of AAAI*, 93, 322-327. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.1037&rep=rep1&type=pdf>
- Nam, J. H. & Jinn, T. (2000). Bankruptcy prediction: Evidence from Korean listed companies during the IMF crisis. *Journal of International Financial Management & Accounting*, 11(3), 178-197. <https://doi.org/10.1111/1467-646x.00061>
- Nam, C. W., Kim, T. S., Park, N. J. & Lee, H. K. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, 27(6), 493-506. <https://doi.org/10.1002/for.985>
- Neophytou, E. & Mar Molinero, C. (2004). Predicting corporate failure in the UK: a multidimensional scaling approach. *Journal of Business Finance & Accounting*, 31(5-6), 677-710. <https://doi.org/10.1111/j.0306-686x.2004.00553.x>
- Nicolas, J. & Toval, A. (2009). On the generation of requirements specifications from software engineering models: a systematic literature review. *Information and Software Technology*, 51, 1291-1307. <https://doi.org/10.1016/j.infsof.2009.04.001>
- Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, 24(12), 1565-1567. <https://doi.org/10.1038/nbt1206-1565>
- Nyitrai, T. (2014). Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében. *Statisztikai Szemle*, 92(4), 357-377.
- Nyitrai, T. (2015). Hazai vállalkozások csődjének előrejelzése a csődeseményt megelőző egy, két, illetve három évvel korábbi pénzügyi beszámolók adatai alapján. *Vezetéstudomány*, 46(5), 55-65. <https://doi.org/10.14267/veztud.2015.05.06>

- Nyitrai, T. (2017). Stock és flow típusú számviteli adatok alkalmazása a csődelőrejelző modellekben. *Vezetéstudomány*, 48(8-9), 68-77.
<https://doi.org/10.14267/veztud.2017.09.07>
- Nyitrai, T. (2019). Dynamization of bankruptcy models via indicator variables. *Benchmarking: An International Journal*, 26(1), 317-332.
<https://doi.org/10.1108/bij-03-2017-0052>
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
<https://doi.org/10.2307/2490395>
- Oravecz, B. (2007). Credit scoring modellek és teljesítményük értékelése. *Hitelintézeti Szemle*, 6(6), 607-627.
- Peel, M. J. & Peel, D. A. (1987). Some further empirical evidence on predicting private company failure. *Accounting and Business Research*, 18(69), 57-66.
<https://doi.org/10.1080/00014788.1987.9729348>
- 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*, 17(1), 31-51.
<https://doi.org/10.1111/j.1468-5957.1990.tb00548.x>
- Platt, H. D. & Platt, M. B. (1991). A note on the use of industry-relative ratios in bankruptcy prediction. *Journal of Banking & Finance*, 15(6), 1183-1194.
[https://doi.org/10.1016/0378-4266\(91\)90057-s](https://doi.org/10.1016/0378-4266(91)90057-s)
- Platt, H. D., Platt, M. B. & Pedersen, J. G. (1994). Bankruptcy discrimination with real variables. *Journal of Business Finance & Accounting*, 21(4), 491-510.
<https://doi.org/10.1111/j.1468-5957.1994.tb00332.x>
- Pompe, P. P. M. & Feelders, A. J. (1997). Using machine learning, neural networks, and statistics to predict corporate bankruptcy. *Computer-Aided Civil and Infrastructure Engineering*, 12(4), 267-276.
<https://doi.org/10.1111/0885-9507.00062>
- Pompe, P. P. & Bilderbeek, J. (2005). The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business Venturing*, 20(6), 847-868.
<https://doi.org/10.1016/j.jbusvent.2004.07.003>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
<https://doi.org/10.1007/bf00116251>
- Quinlan, J. R. (1993). *C4.5: Programming for machine learning*. Morgan Kaufmann Publishers, San Mateo, California.
- Quinlan, J. R. (1997). *C5.0 and see 5: Illustrative examples*. RuleQuest Research. <http://www.rulequest.com>
- Rafiei, F. M., Manzari, S. M. & Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications*, 38(8), 10210-10217.
<https://doi.org/10.1016/j.eswa.2011.02.082>
- Ratting, A. (2015). Fizetésképtelenség-előrejelzési megközelítések. *Társadalom és Gazdaság*, 7(3), 53-73.
<https://doi.org/10.21637/gt.2015.3.04>
- Ravi Kumar, P. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180(1), 1-28.
<https://doi.org/10.1016/j.ejor.2006.08.043>
- Russell, S. & Norvig, P. (2005). *Mesterséges intelligencia modern megközelítésben*. Budapest: Panem Könyvkiadó.
- Ryu, Y. U. & Yue, W. T. (2005). Firm bankruptcy prediction: experimental comparison of isotonic separation and other classification approaches. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(5), 727-737.
<https://doi.org/10.1109/tsmca.2005.843393>
- Sajtos, L. & Mitev, A. (2007). *SPSS kutatási és adatelemzési kézikönyv*. Budapest: Alinea Kiadó.
- Sandin, A. R. & Porporato, M. (2007). Corporate bankruptcy prediction models applied to emerging economies. *International Journal of Commerce and Management*, 17(4), 295 – 311.
<https://doi.org/10.1108/10569210710844372>
- Scherger, V., Terceño, A. & Vigier, H. (2019). A systematic overview of the prediction of business failure. *International Journal of Technology, Policy and Management*, 19(2), 196-211.
<https://doi.org/10.1504/ijtpm.2019.100601>
- Sharma S. & Mahajan, V. (1980). Early Warning Indicators of Business Failure. *Journal of Marketing*, 44(4), 80-89.
<https://doi.org/10.1177/002224298004400412>
- Shetty, U., Pakkala, T. P. M. & Mallikarjunappa, T. (2012). A modified formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India. *Expert Systems with Applications*, 39(2), 1988-1997.
<https://doi.org/10.1016/j.eswa.2011.08.043>
- Shi, Y. & Li, X. (2019). An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. *Intangible Capital*, 15(2), 114-127.
<https://doi.org/10.3926/ic.1354>
- Shin, K. S. & Lee, Y. J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, 23(3), 321-328.
[https://doi.org/10.1016/s0957-4174\(02\)00051-9](https://doi.org/10.1016/s0957-4174(02)00051-9)
- Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135.
<https://doi.org/10.1016/j.eswa.2004.08.009>
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101-124.
<https://doi.org/10.1086/209665>
- Sun, J. & Li, H. (2008). Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers. *Expert Systems with Applications*, 35(3), 818-827.
<https://doi.org/10.1016/j.eswa.2007.07.045>
- Sun, J. & Li, H. (2009). Financial distress early warning based on group decision making. *Computers & Operations Research*, 36(3), 885-906.
<https://doi.org/10.1016/j.cor.2007.11.005>

- Sun, J., Jia, M. Y. & Li, H. (2011). AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. *Expert Systems with Applications*, 38(8), 9305-9312. <https://doi.org/10.1016/j.eswa.2011.01.042>
- Szabadosné, Németh Zs. & Dávid, L. (2005). A kis- és középvállalati szegmens mulasztási valószínűségének előrejelzése magyarországi környezetben. *Hitelintézeti Szemle*, 4(3), 39-58. https://www.bankszovetseg.hu/Content/Hitelintezeti/53Szabadosne_David.pdf
- Székelyi, M. & Barna, I. (2008). *Túlélő készlet az SPSS-hez*. Budapest: Typotex.
- Takahashi, K., Kurokawa, Y. & Watase, K. (1984). Corporate bankruptcy prediction in Japan. *Journal of Banking & Finance*, 8(2), 229-247. [https://doi.org/10.1016/0378-4266\(84\)90005-0](https://doi.org/10.1016/0378-4266(84)90005-0)
- Teng, H. S. S., Bhatia, G. S. & Anwar, S. (2011). A success versus failure prediction model for small businesses in Singapore. *American Journal of Business*, 26(1), 50-64. <https://doi.org/10.1108/19355181111124106>
- Theodossiou, P. (1991). Alternative models for assessing the financial condition of business in Greece. *Journal of Business Finance & Accounting*, 18(5), 697-720. <https://doi.org/10.1111/j.1468-5957.1991.tb00233.x>
- Theodossiou, P. T. (1993). Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures. *Journal of the American Statistical Association*, 88(422), 441-449. <https://doi.org/10.1080/01621459.1993.10476294>
- Tian, S., Yu, Y. & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89-100. <https://doi.org/10.1016/j.jbankfin.2014.12.003>
- Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207-222. <https://doi.org/10.1111/1467-8551.00375>
- Tsukuda, J. & Baba, S. I. (1994). Predicting Japanese corporate bankruptcy in terms of financial data using neural network. *Computers & Industrial Engineering*, 27(1-4), 445-448. [https://doi.org/10.1016/0360-8352\(94\)90330-1](https://doi.org/10.1016/0360-8352(94)90330-1)
- Virág, M. (2001). *Pénzügyi elemzés csődelőrejelzés*. Budapest: Kossuth Kiadó.
- Virág, M. (2004). A csődmodellek jellegzetességei és története. *Vezetéstudomány*, 35(10), 24-32. http://unipub.lib.uni-corvinus.hu/4551/1/VT_2004n10p24.pdf
- Virág, M. & Dóbé, S. (2005). A hazai csődmodell-család alkalmazása ágazati centroidokra. *Vezetéstudomány*, 36(4), 45-54. <https://doi.org/10.14267/VEZTUD.2005.04.05>
- Virág, M. & Kristóf, T. (2005). Az első hazai csődmodell újraszámítása neurális hálók segítségével. *Közgazdasági Szemle*, 52(2), 144-162. http://unipub.lib.uni-corvinus.hu/3465/1/Kszemle_CIKK_744.pdf
- Virág, M. & Kristóf, T. (2006). Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel. *Vezetéstudomány*, 37(1), 25-35. http://unipub.lib.uni-corvinus.hu/1048/1/vt_2006n1p25.pdf
- Virág, M., Fiáth, A., Kristóf, T. & Varsányi, J. (2013). *Pénzügyi elemzés, csődelőrejelzés, válságkezelés*. Budapest: Kossuth Kiadó.
- Virág, M. & Nyitrai, T. (2013). Application of support vector machines on the basis of the first Hungarian bankruptcy model. *Society and Economy*, 35(2), 227-248. <https://doi.org/10.1556/socec.35.2013.2.6>
- Wang, Y. & Campbell, M. (2010). Business Failure Prediction for Publicly Listed Companies in China. *Journal of Business & Management*, 16(1), 75-88. <http://jbm.johogo.com/pdf/volume/1601/JBM-1601-04-full.pdf>
- Wang, G., Ma, J. & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, 41(5), 2353-2361. <https://doi.org/10.1016/j.eswa.2013.09.033>
- Wilson, R. L. & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11(5), 545-557. [https://doi.org/10.1016/0167-9236\(94\)90024-8](https://doi.org/10.1016/0167-9236(94)90024-8)
- Xiao, Z., Yang, X., Pang, Y. & Dang, X. (2012). The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory. *Knowledge-Based Systems*, 26, 196-206. <https://doi.org/10.1016/j.knsys.2011.08.001>
- Xu, X. & Wang, Y. (2009). Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications*, 36(1), 366-373. <https://doi.org/10.1016/j.eswa.2007.09.040>
- Youn, H. & Gu, Z. (2010). Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model. *International Journal of Hospitality Management*, 29(1), 120-127. <https://doi.org/10.1016/j.ijhm.2009.06.007>
- Zhao, Y., Li, J. & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, 66, 9-16. <https://doi.org/10.1016/j.eneco.2017.05.023>
- Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, 41, 16-25. <https://doi.org/10.1016/j.knsys.2012.12.007>
- Zięba, M., Tomczak, S. K. & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58, 93-101. <https://doi.org/10.1016/j.eswa.2016.04.001>
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82. <https://doi.org/10.2307/2490859>

Melléklet

1. melléklet

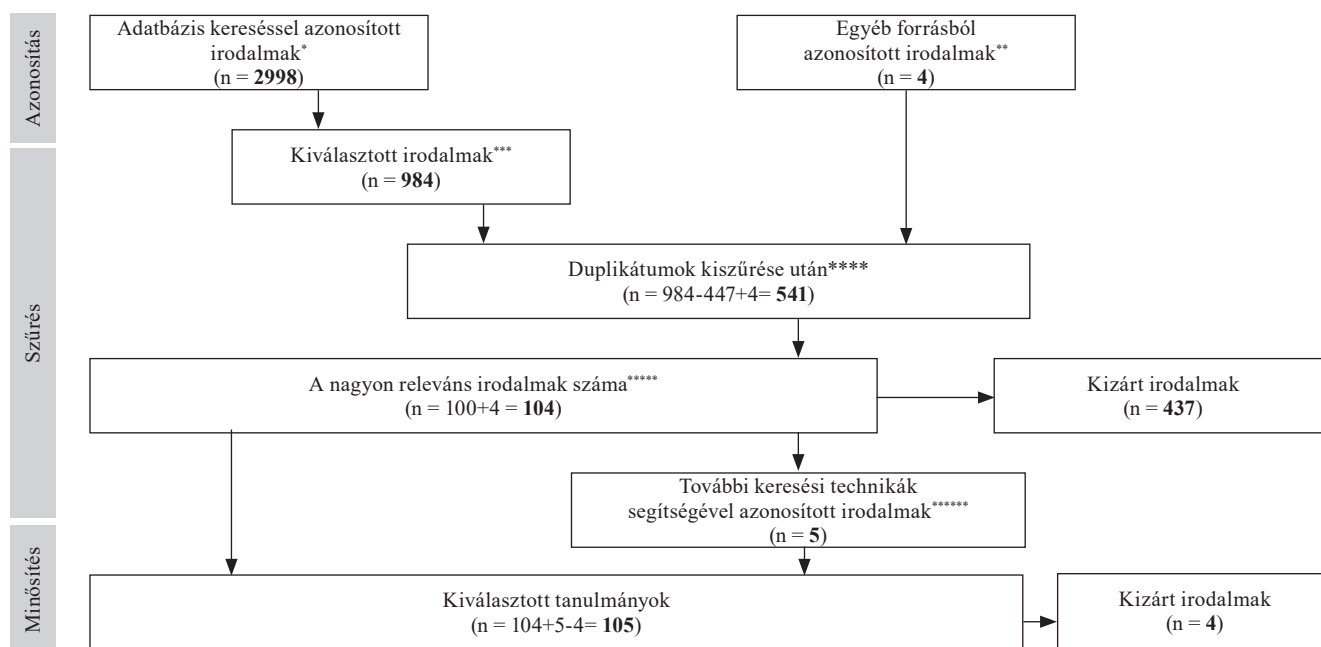
Keresési kifejezések és a lekérdezések találati eredményei

Keresési kifejezés [a]:	((„Corporate” OR „Firm” OR „Company” OR „Business”) AND („Bankruptcy predict*” OR „Bankruptcy forecast*” OR „Bankruptcy model” OR „Insolvency predict*” OR „Insolvency forecast*” OR „Insolvency model” OR „Distress predict*” OR „Distress forecast*” OR „Default predict*” OR „Default forecast*” OR „Failure predict*” OR „Failure forecast*”))											
Keresési kifejezés [b]:	((„Corporate” OR „Firm” OR „Company” OR „Business”) AND („Bankruptcy predict*” OR „Bankruptcy forecast*” OR „Bankruptcy model” OR „Insolvency predict*” OR „Insolvency forecast*” OR „Insolvency model” OR „Distress predict*” OR „Distress forecast*” OR „Default predict*” OR „Default forecast*” OR „Failure predict*” OR „Failure forecast*”) NOT(Credit OR Bank))											
Keresési kifejezés [c]:	((„Corporate” OR „Firm” OR „Company”) AND („Bankruptcy predict*” OR „Insolvency predict*” OR „Distress predict*” OR „Default predict*” OR „Failure predict*”) NOT(Credit OR Bank))											
Keresési kifejezés [d]:	((„Corporate” OR „Firm” OR „Company”) AND („Bankruptcy predict*” OR „Insolvency predict*” OR „Distress predict*” OR „Default predict*” OR „Failure predict*”) NOT(Credit OR Bank) AND („Machine learn*”))											
Cikkek száma	Keresési kifejezés [a]			Keresési kifejezés [b]			Keresési kifejezés [c]			Keresési kifejezés [d]		
	WoS	BSP	ASC	WoS	BSP	ASC	WoS	BSP	ASC	WoS	BSP	ASC
Adatbázisonként	502	498	223	452	394	204	330	248	98	22	17	10
Kifejezésenként	1223			1050			676			49		
Mindösszesen	2998											
WoS=Web of Science; BSP= Business Source Premier; ASC= Academic Search Complete												

Forrás: saját szerkesztés

2. melléklet

A szisztematikus irodalomfeldolgozás folyamatábrája



Megjegyzések:

* A szám az összes irodalmat jelöli, amelyek a különféle adatbázisokból származnak (ASC, BSP, WoS).

** A szakértők által ajánlott irodalmak száma.

*** Szűrő alkalmazása: csak angol nyelvű folyóiratcikk.

**** Zotero hivatkozásmenedzsmnt szoftver használatával. Ezen a szinten kizártam a cím nélküli, valamint angol nyelvű absztrakttal, de nem angol nyelven közzétett szakirodalmakat.

***** Az egyes irodalmak értékelése cím, absztrakt alapján, majd az első 100 szakirodalom kiválasztása Google Scholar hivatkozásszám alapján.

***** Kiemelkedő szakirodalmak hivatkozásai alapján és hólabda módszerrel.

Forrás: saját szerkesztés, a PRISMA módszertan alapján (Moher, Liberati, Tetzlaff & Altman, 2009)

A szisztematikus irodalomlemezés tanulmányai és a minták jellemzői

Nr.	Szerző és év	Mintavétel időszaka	Minta országa
1.	Abidali - Harris [1995]	1978-1986	Egyesült Királyság
2.	Ahn - Kim [2009]	1996-2000	Dél-Korea
3.	Alfaro et al. [2008]	2000-2003	Spanyolország
4.	Altman [1968]	1946-1965	USA
5.	Altman et al. [1995]	1990-1993	Dél-Korea
6.	Anandarajan et al. [2001]	1989-1996	USA
7.	Aziz et al. [1988]	1971-1982	USA
8.	Baldwin - Glezen [1992]	1977-1983	USA
9.	Barboza et al. [2017]	1985-2013	USA és Kanada
10.	Beaver [1966]	1954-1964	USA
11.	Berg [2007]	1996-2000	Norvégia
12.	Beynon - Peel [2001]	1997-1998	Egyesült Királyság
13.	Boritz et al. [1995]	1971-1984	USA
14.	Casey - Bartzak [1985]	1971-1982	USA
15.	Charitou et al. [2004]	1988-1997	Egyesült Királyság
16.	Chaudhuri - De [2011]	2001-2002	USA
17.	Chava - Jarrow [2004]	1962-1999	USA
18.	Chen [2011a]	2000-2010	Tajvan
19.	Chen [2011b]	2000-2007	Tajvan
20.	Chen - Du [2009]	1999-2006	Tajvan
21.	Ciampi [2015]	2008	Olaszország
22.	De Andres et al. [2011]	2007	Spanyolország
23.	Delen et al. [2013]	2005-2011	Törökország
24.	Dewaechevyns - Van Hulle [2006]	1996-2001	Belgium
25.	Dimitras et al. [1999]	1986-1993	Görögország
26.	du Jardin [2015]	2003-2012	Franciaország
27.	du Jardin - Severin [2011]	1996-2003	Franciaország
28.	Duan et al. [2012]	1991-2011	USA
29.	Duffie et al. [2007]	1980-2004	USA
30.	Edmister [1972]	1954-1969	USA
31.	Etemadi et al. [2009]	1998-2005	Irán
32.	Fan - Palaniswami [2000]	nincs jelölve	Ausztrália
33.	Fedorova et al. [2013]	2007-2011	Oroszország
34.	Fletcher - Goss [1993]	nincs jelölve	USA
35.	Frydman et al. [1985]	1971-1981	USA
36.	Gieng et al. [2015]	2008-2011	Kína
37.	Gentry et al. [1985]	1970-1980	USA
38.	Gombola et al. [1987]	1967-1981	USA
39.	Grice - Ingram [2001]	1988-1991	USA
40.	Hillegeist et al. [2004]	1980-2000	USA
41.	Horta - Camanho [2013]	2008-2010	Portugália
42.	Hua et al. [2007]	1999-2004	Kína
43.	Huang et al. [2008]	2001-2004	Tajvan
44.	Johnsen - Melicher [1994]	1970-1983	USA
45.	Jones - Hensher [2007]	1996-2003	Ausztrália
46.	Jones - Hensher [2004]	1996-2003	Ausztrália
47.	Kim - Kang [2010]	2002-2005	Dél-Korea
48.	Kim [2011]	1995-2002	Dél-Korea
49.	Korol [2013]	2000-2007 és 1996-2009	Lengyelország és Latin Amerika (Mexikó, Argentína, Brazília, Chile, Peru)
50.	Laitinen [1993]	1986-1988	Finnország
51.	Laitinen [1992]	1980-1985	Finnország
52.	Lau [1987]	1971-1977	USA
53.	Lee et al. [2005]	1995-1998	Dél-Korea
54.	Lee - Choi [2013]	2000-2009	Dél-Korea
55.	Leshno - Spector [1996]	1984-1988	USA
56.	Li - Sun [2008]	2000-2005	Kína
57.	Li - Sun [2010]	2000-2005	Kína
58.	Li - Sun [2012]	1998-2010	Kína
59.	Liang et al. [2016]	1999-2009	Tajvan
60.	Lin et al. [2014]	2000-2008	Tajvan
61.	Lin [2009]	1998-2005	Tajvan
62.	Liu et al. [2014]	nincs jelölve	Egyesült Királyság
63.	Luoma - Laitinen [1991]	nincs jelölve	Finnország
64.	Lussier [1995]	nincs jelölve	USA
65.	Mayew et al. [2015]	1995-2012	USA
66.	McKee [2000]	1986-1988	USA
67.	McKee [2003]	1991-1997	USA
68.	McKee - Greenstein [2000]	1981-1990	USA
69.	Mensah [1983]	1975-1980	USA
70.	Mensah [1984]	1972-1980	USA
71.	Min - Jeong [2009]	2001-2004	Dél-Korea
72.	Mossman et al. [1988]	1980-1991	USA
73.	Nam et al. [2008]	1991-2000	Dél-Korea
74.	Nam - Jinn [2000]	1997-1998	Dél-Korea
75.	Neophytou - Mar Molinero [2004]	1988-1999	Egyesült Királyság
76.	Ohlson [1980]	1970-1976	USA
77.	Peel - Peel [1987]	1982-1985	Egyesült Királyság
78.	Platt - Platt [1990]	1972-1986	USA
79.	Platt - Platt [1991]	1972-1987	USA
80.	Platt et al. [1994]	1982-1988	USA
81.	Pompe - Biderbeek [2005]	1986-1994	Belgium
82.	Rafiei et al. [2011]	2007-2008	Irán
83.	Ryu - Yue [2005]	1996-2001	USA
84.	Sandin - Porporato [2007]	1990-1998	Argentína
85.	Shin et al. [2005]	1996-1999	Dél-Korea
86.	Shin - Lee [2002]	1995-1997	Dél-Korea
87.	Shumway [2001]	1962-1992	USA és Kanada
88.	Sun et al. [2011]	2000-2008	Kína
89.	Sun - Li [2008]	2000-2005	Kína
90.	Sun - Li [2009]	2002-2004	Kína
91.	Takahashi et al. [1984]	1961-1977	Japán
92.	Teng et al. [2011]	2009-2010	Szingapúr
93.	Theodossiou, P.	1980-1984	Görögország
94.	Theodossiou [1993]	1967-1986	USA
95.	Fian et al. [2015]	1980-2009	USA
96.	Tsukuda - Baba [1994]	1970-1990 és 1986-1991	Japán
97.	Wang et al. [2014]	1970-1982 és 1997-2001	USA
98.	Wang - Campbell [2010]	2000-2008	Kína
99.	Wilson - Sharda [1994]	1975-1982	USA
100.	Xiao et al. [2012]	2004-2009	Kína
101.	Xu - Wang [2009]	1999-2005	Kína
102.	Youn - Gu [2010]	2000-2006	Dél-Korea
103.	Zhou [2013]	1981-2009	Észak-Amerika és Japán
104.	Zieba et al. [2016]	2007-2013	Lengyelország
105.	Zmijewski [1984]	1972-1978	USA

Forrás: saját szerkesztés

A módszerek páronkénti közvetlen összehasonlításának a száma

	Logit	DA	NN	SVM	DT	IB	Hazard	Rough Sets	Probit	GA	Bayes	MV	Összesen
Logit	-	21	16	11	8	8	4	3	3	1	2	1	78
DA	21	-	22	8	9	10	3	3	3	3	-	-	61
NN	16	22	-	10	11	7	2	2	3	2	3	2	42
SVM	11	8	10	-	7	6	-	1	-	-	3	2	19
DT	8	9	11	7	-	4	-	3	1	1	2	1	12
IB	8	10	7	6	4	-	1	1	1	-	1	-	4
Hazard	4	3	2	-	-	1	-	-	-	-	-	-	-
Rough Sets	3	3	2	1	3	1	-	-	1	-	1	1	3
Probit	3	3	3	-	1	1	-	1	-	-	-	-	-
GA	1	3	2	-	1	-	-	-	-	-	-	-	-
Bayes	2	-	3	3	2	1	-	1	-	-	-	1	1
MV	1	-	2	2	1	-	-	1	-	-	1	-	-
Összesen:	78	61	42	19	12	4	-	3	-	-	1	-	220
Átlagos módszer összehasonlítások száma: $220 / 47 = 4,68$													

Megjegyzés: A kiemeléssel jelölt módszerpárokat vettem figyelembe, amelyeknél több mint 4 páros összehasonlítás volt. Logit=logisztikus regresszió; DA=díszkriminanciaanalízis; NN=neurális háló; SVM=Support Vector Machine; DT=döntési fa; IB=példányalapú; Hazard=túléléselemzés; Rough sets=durva halmazok; Probit=probit analízis; GA=genetikus algoritmus; MV=többségi szavazás.

Forrás: saját számítás