

Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Regionális Politika és Gazdaságtan Doktori Iskola

**A gépi tanuláson alapuló fizetésképtelenséget előrejelző
modellek háttere és gyakorlati alkalmazása magyar kis- és
középvállalatok körében**

Doktori értekezés tézisei

Készítette: Ágoston Norbert

Témavezetők: Dr. Bedőházi Zita-Rozália

Dr. habil. Kruzslicz Ferenc
egyetemi docens

Pécs, 2024

Tartalomjegyzék

1. A témaválasztás indoklása és a kutatás jelentősége	1
2. A dolgozat felépítése céljai és hipotézisei	4
3. A kutatás módszerei és adatbázisa	9
4. Az értekezés tudományos eredményei és az eredmények hasznosíthatóságának lehetőségei	18
5. Jövőbeli kutatási irányok	22
A téziszűzetben felhasznált irodalom.....	24
Publikációk és konferencia előadások	30

Absztrakt

A doktori értekezés szerzője: Ágoston Norbert

Címe: A gépi tanuláson alapuló fizetésképtelenséget előrejelző modellek háttere és gyakorlati alkalmazása magyar kis- és középvállalatok körében

Témavezetők: Dr. Bedőházi Zita-Rozália

Dr. habil. Kruzslicz Ferenc, egyetemi docens

A vállalkozások kényszerű megszűnése, illetve csődje visszafordíthatatlan költségeket generál az érintettek körében. Ezért a gazdasági társaságok vezetőit, a pénzügyi intézményeket és biztosítókat folyamatosan foglalkoztatja a vállalati ügyfelek fizetőképessége. Emellett a csödkutatók is folyamatosan keresik a legjobban interpretálható módszereket és a legkiemelkedőbb teljesítményű eljárásokat és ezzel összhangban egy szüntelen vizsgálódás és tesztelés jellemzi a szakterületet. Mindez pedig módot kínál az újabb és újabb módszertani kísérletre és fejlesztésre, amely a jelen tanulmány egyik fő motivációjaként is szolgált. A disszertáció egyik célja olyan fizetésképtelenséget előrejelző modellek kifejlesztése, amely a magyar korlátolt felelősségű kis- és középvállalatokon, univerzálisan és hatékonyan alkalmazható. Ennek kivitelezése pedig olyan módszertani finomításokat tartalmaz, amely egy része a kutatások főáramában található és a disszertáció során hipotézisekkel teszteltem. Az első ilyen a mintavételi kérdések közé sorolható azonos vagy kiegyensúlyozatlan mintavétel problémája. A másik kérdésköre a disszertációnak a tanuló adathalmaz időbeli terjedelmére reflektál. Egy következő tétele a disszertációnak, a pénzügyi változóktól eltérő, alternatív magyarázó változók tárgyköre. Ez utóbbi egy kevésbé kutatott terület, a magyar szakirodalomban nem is igazán találunk említést róla. Meg kell említeni azt a tényt is, hogy a fokozott

elemzési igény és a megnövekedett számítástechnikai kapacitások lendületet adtak a hagyományostól eltérő módszerek elterjedésének. A doktori értekezés másik fő célja a vállalati fizetésképtelenség előrejelzésére alkalmas, a legkorszerűbb gépi tanuláson alapuló módszerek ismertetése és a magyar adatokra történő kalibrálása. A széles adatminta és a több párhuzamos modell illesztése robusztus eredményeket biztosított. Az első megállapításom szerint a magyar vállalati adatmintán az azonos osztálylétszámú modellfejlesztés javasolt. Továbbá a fizetésképtelenséget előre jelezni kíván időszakhoz legközelebb eső egy év adatai elegendőnek bizonyultak a hatékony teljesítmény eléréséhez. Figyelembe vettem a vállalati működés külső, mezoszintű változóit, amelyek közül lehetett találni olyant, amelyik magyarázta a fizetésképtelenséget, azonban az előrejelzési teljesítményt nem javította. Végezetül a legmagasabb teljesítményű gépi tanuláson alapuló modellt keresve az együttes módszercsalád eljárásai bizonyultak a leghatékonyabbnak.

Journal of Economic Literature (JEL) kódok: C45, C53, C58, F65, G33,

L25

Tárgyszavak: *felszámolás, fizetésképtelenség előrejelzés, pénzügyi teljesítmény, gépi tanulás, metamódszerek, boosting, bagging, random forest*

1. A témaválasztás indoklása és a kutatás jelentősége

A vállalati nyereségeket hajtó pezsgő lét mellett létezik a vállalatok sötét oldala, a kudarcokkal és sikertelenségekkel jellemezhető felszámolások területe. A felszámolások nem csak a vállalatok megszűnését, hanem sok esetben személyes célok és küzdelmek felbomlását, bizalmi viszonyok megszűnését, társadalmi konfliktusok kialakulását eredményezi. Minden sikertelen vállalatnak sajátos története van. Azonban ezeknek vannak olyan közös jellemzőik, amelyek által azonosítani tudunk bizonyos mintázatokat.

A felszámolás a vállalati életciklus utolsó szakasza, amelyet számos tényező előidézhethet. Az egyik legfőbb kiindulópontja a fizetéseképtelenség, amelyben a gazdasági tevékenység pénzügyi problémák miatt állt le, és gyakran a felszámolási eljárással végződik. Magát a fizetéseképtelenséget is számos mögöttes erő befolyásolhatja: környezet, iparág, méret, makrogazdasági körülmények, vállalatvezetés, verseny, életciklus, kultúra stb. A vállalatok kényszerű megszűnése, illetve csődje visszafordíthatatlan költségeket generál az érintettek körében. Veszteség éri a vállalat hitelezőit, tulajdonosait, vezetését, munkavállalóit, akik elveszíthetik munkahelyüket, továbbá a fizetéseképtelen vállalat vevőit és beszállítóit. Nem utolsó sorban az állami költségvetés is elesik a várható adóbevételektől, így egy sorozatos csődjelenség veszélybe tudja sodorni a teljes nemzetgazdaságot, emellett 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 – Mallikarjunappa, 2012).

A fizetéseképtelenségi kutatások hosszú múltra tekintenek vissza. Az alkalmazott módszerek kezdetben egyváltozós elemzések voltak, amit idővel a többváltozós módszerek helyettesítettek. Manapság a számítástechnikai kapacitások lehetővé teszik, hogy fejlett mesterséges intelligencia és gépi tanuláson alapuló módszerek elterjedjenek és innovatív módszertani fejlesztéseket eredményezzenek. Mivel az üzleti kultúra rendkívül sokszínű, a működési környezet állandóan változó, a probléma pedig nagymértékben összetett, ezért vég nélkül vannak feltáratlan területek és kihívások a csődelőrejelzés kutatásában (Fejér-Király, 2015).

A fizetéseképtelenség- és csődelőrejelzés szakirodalma egy nagyon gazdag kutatási terület, számtalan empirikus kutatással és megközelítéssel. A terület tanulmányozása lehetővé tette az interdiszciplinaritást, amelyben a közgazdaságtan, a pénzügy és számvitel, valamint informatika együttműködésében számos különböző perspektívából közelíthetjük meg a problémát. A kutatások szüntelen lendületét magyarázza az is, hogy a csőd és fizetéseképtelenség sokrétű problémája miatt, az elméleti keretet adó empirikus kutatások nem álltak össze egy egységes egésszé, hanem egy folyamatos keresés, kísérletezés jellemzi a területet. A kutatások nemzetközi szinten megfigyelhetőek, azonban a gyakorlatban széles körben alkalmazott modellek a híres amerikai kutatások eredményeinek köszönhetőek. Azonban a neves külföldi modellek változatlan formában történő átvételével körültekintően szükséges eljárni. Az amerikai vagy más külföldi vállalatokra készült csődelőrejelző modellek sajátos vállalati, működési és környezeti tulajdonságai okán nem feltétlen működnek jól hazai vállalatok fizetőképességének azonosítására. Ezzel összhangban Kristóf és Virág (2019) kihangsúlyozta, hogy magyar vállalatokra nem jók a külföldi csődmodellek, sokkal előnyösebb eredményeket érünk el, ha saját modellek

építünk. Mindez pedig megalapozta a kutatásom egyik alapvető célját, mely szerint a magyar vállalatok széles körére érvényes, fizetéképtelenséget előrejelző modelleket készítek, amelyek a módszerek részletezése mellett az érdekelt feleket is hasznos információval látja el a pénzügyi kockázatkezeléssel kapcsolatos döntéshozatali folyamatokhoz. Ez pedig lehetőséget teremt módszertani kérdések tesztelésére illetve vizsgálatára, amelyekből értékes következtetések vonhatók le, segítve a modellépítés döntéseinek meghozatalát.

A csőd- és fizetéképtelenséget előrejelző modellek egyik folyamatosan megújuló keresése a bemeneti magyarázó változók körül forog. A fizetéképtelenségi tanulmányok szinte kizárólag a vállalati pénzügyi adatokra támaszkodnak. Az éves beszámolókból kalkulált pénzügyi arányszámok adják a gerincét az előrejelző modelleknek. Az alkalmazandó pénzügyi mutatókra azonban nincs minden adatállományra egységesen érvényes szabály, továbbá az is zavart okoz, hogy a kutatók tengernyi mutató kalkulálását végezhetik el. A tanulmányban ötven előre definiált pénzügyi változó közül kiemelem azt a néhányat, amely a magyar KKV-k körében megbízhatóan és több perióduson keresztül, magyarázza a fizetéképtelenséget. Emellett felmerül a pénzügyi változóktól eltérő, alternatív magyarázó változók kérdésköre. A tágabb működési környezet változóinak integrálása a fizetéképtelenséget előrejelző modellekbe nem túl gyakori, a magyar csődkutatásban kifejezetten hiányzik. Így kísérletet teszek külső, mezoszintű magyarázó változók előrejelző modellekbe való integrálására.

A csődelőrejelző modellek építésére számos módszer és eljárás létezik a szakirodalomban. A gépi tanulás eszköztárából egy változatos palettát

mutatok be és alkalmazok a modellezés során. Az egyik a tartóvektor gépek (support vector machines, SVM), amelyek az osztályok elválasztására az optimális hipersíkot keresik és összetett problémák nagy pontosságú elválasztására képesek. A másik, nagy népszerűségnek örvendő módszer, a neurális hálózatok, amelyek összekapcsolt csomópontok komplex rétegeinek segítségével olyan kapcsolatokat és mintákat tudnak feltárni, amelyeket a hagyományos illetve lineáris módszerek kevésbé tudnak. Emellett az együttes módszerek (ensemble) vagy metamódszerként ismert eljárásokat alkalmaztam, nevezetesen a bagging, a boosting és a random forest algoritmusokat, amelyek az alaposztályozók kombinálásával kiemelkedő eredményeket érnek el. A felvonultatott módszerek teljesítményének egy megbízható összemérése a disszertáció végén kap helyet. A módszerek összehasonlító erősségeinek megvilágításával arra törekedtem, hogy hasznos információval lássam el az érdekelt feleket, amely végső soron hozzájárulhat ahhoz, hogy a vállalatok megerősítsék stratégiai döntéshozatali folyamataikat egy állandóan változó pénzügyi környezetben.

2. A dolgozat felépítése, céljai és hipotézisei

A csődelőrejelzés szakirodalmában nincs egy egységesen elfogadott elmélet, sem egyetlen univerzális fogalom a csőd meghatározására. A tudományterület képviselői azonosan értelmezik a csőd fogalmát a vállalati fizetéseképtelenség, a pénzügyi nehézség, a fizetés elmulasztása és a kudarc fogalmával. Ez a kutatások célját, rendeltetését és eredményét érdemben nem befolyásolja, azonban a bizonyos fogalmak közti finom eltéréseket érdemes szem előtt tartani. Ezért a tanulmány első részében a fizetéseképtelenség és csőd keretrendszerét ismerttettem, amelyben értelmezni tudjuk az előrejelzés szempontjait, szerepét és fontosságát.

Ezt követően a tanulmány terjedelmesebb része, a szakirodalmi áttekintés következett. Ez időrendi sorrendben a csődelőrejelzés és fizetésképtelenséghez kapcsolódó tanulmányok mérföldköveit ismerteti a kezdetektől a napjainkig. Mivel a csőd- és fizetésképtelenség előrejelzés egy nagyon gazdag és széles kutatási terület, számos mellékággal, ezért a témában írt összes tudományos írás ismertetése nem volt célja a dolgozatnak, hatékonysági és áttekinthetőségi szempontból sem. Mindössze az angol nyelven írt magasan hivatkozott és jelentős hatású művek kerültek bemutatásra, különösen kiemelve az újszerű módszertani megoldásokat és azok lényeges elemeit. A témában folytatott, magyar nyelven írt kutatások jelentős részét ismertettem, amelyek tapasztalataihoz kapcsolódva és eredményeiből merítve épült a jelen dolgozat is.

A disszertáció harmadik fejezete a fizetésképtelenség módszereinek áttekintését adta, részletesen kifejtve a mesterséges intelligencia és azon belül a gépi tanulás fejlett eszköztárát. A klasszikus módszerek mellett, az SVM, a neurális háló és a döntési fa módszercsaládba tartozó bagging, boosting és random forest módszerek működését és specifikációit ismerhettük meg, rávilágítva az előnyökre és az esetleges hátrányokra.

Tanulmányom céljaként tűztem ki egy kellően széles, magyar vállalati adatmintára épülő, általános érvényű fizetésképtelenséget előrejelző modellek felállítását és értékelését. A felépült modellek nem csak az előrejelző módszerek megértését gazdagítják, hanem akár a pénzügyi-akár a kereskedelmi hitelminősítések gyakorlati alkalmazása során segítséget nyújthatnak. A vállalatok tulajdonosai, vezetői és érdekeltjei számára szintén eszköztár lehet a saját- vagy partner vállalat fizetőképességének felmérésére. A további célok közt szerepelt a fizetésképtelenség előrejelzés módszertani

kérdéseinek tisztázása valamint újító jellegű előrelépések megtevése, főként olyan területen, ahol a szakirodalomban nincsenek egyértelmű utalások illetőleg ellentmondások mutatkoznak. Nem utolsó sorban pedig a gépi tanulás eszköztárának részletes bemutatása, gyakorlatba való ültetése és hasznosságának értelmezése volt a kutatás központi fókuszában.

A következőkben ismertetem a hipotéziseimet a magyar vállalatok fizetésektelenségének előrejelzésére, modellezésére és módszertanára vonatkozóan. A négy hipotézisem közül először a minta reprezentativitását veszem a fókuszba, és azt feltételezem, hogy a fizetésektelenség előrejelzésére kialakult azonos mintavétel nem önkényes és kényelmi szempontok miatt alakult ki, hanem a modellek végső teljesítményére van döntő hatással. Ez pedig nem mindig van explicit kimondva, a kutatók sokszor a szokásra hivatkozva végzik ezt a gyakorlatot.

H1: A reprezentatív mintával készült fizetésektelenséget előrejelző modell az azonos osztálylétszámú mintavétellel szemben kisebb hatékonyságú előrejelzésre képes a fizetésektelenség helyes besorolási arányát tekintve.

Az egyik módszertani vita, amely figyelmet érdemel, a mintavételhez kapcsolható. A csődelőjelzésben hagyományosan kialakult az azonos osztálylétszámú mintavétel, ahol a fizetésektelenség és egészségesek egyforma arányban szerepelnek a modellezési folyamatokban. Ezt egyrészt a kezdeti és széles körben elismert csőd tanulmányok alapozták meg (Altman, 1968; Edmister, 1972; Gombola et al., 1987), másrészt az azonos mintavételnek technikai és kényelmi okai is vannak. A kiegyensúlyozatlan mintavétellel csupán néhány módszertani tanulmány foglalkozott, és azok

sem mindig az valós eloszlások alapján határozták meg a mintát, hanem sok esetben egy köztes, önkényesen kiválasztott arány alapján (Berg, 2007; Oommen – Baise – Vogel, 2011; Alaka et al., 2018). Ezekhez kapcsolódik a tanulmány egyik kérdése, amelyben az azonos és valós eloszlásnak megfelelő mintavétel megbízhatóságát elemzem, amellyel a mintavételi döntés megalapozását kívánom megerősíteni.

H2: A több év adataiból épített, fizetéseképtelenséget előrejelző modellek átlagosan magasabb előrejelző teljesítményűek egy éves előrejelzési horizonton, mint az egyetlen üzleti év adatain épült modellek.

A második kérdésköre a disszertációnak a tanuló adathalmaz időbeli terjedelmére vonatkozik. Ha a modellek több évet átfogó adatok felhasználásával készülnek, akkor az előrejelzési teljesítmény javulását várhatjuk, mivel az információk szélesebb spektrumát ölelik fel, ezáltal árnyaltabb és pontosabb képet adhat a vállalatok pénzügyi állapotáról. Ezzel szemben az egy éves megközelítés egyszerű megoldást kínál és a legfrissebb, legrelevánsabb információkat foglalja magába. Arra találunk empirikus kutatásokat, hogy a megfigyelések több éves múltbeli adatainak a felhasználása javítja a modellek teljesítményét, szemben azzal, hogy csak egyetlen év adataira támaszkodunk, mivel figyelembe veszi a pénzügyi mutatók időbeli alakulását (Berg, 2007; Nyitrai, 2015; Fejér-Király – Ágoston – Varga, 2019). Azonban arra a kérdésre, hogy ez a teljesítmény javulás számottevő vagy csak egy-egy sajátos helyzetben érvényesül, nem kapunk rá egyértelmű utalást, ezért a probléma kibontása kitüntetett helyet kapott a disszertációban.

H3: A mezoszintű bemeneti adatok közül lesznek olyanok, amelyek szignifikáns hatással vannak a fizetésképtelenségre, azonban a hatásuk a pénzügyi mutatókkal szemben olyan kismértékű, hogy a mezoszintű változókat tartalmazó előrejelző modellek hasonló arányban képesek besorolni az egészséges és fizetésképtelen vállalatokat, mint a kizárólag vállalati pénzügyi adatokon alapuló fizetésképtelenséget előrejelző modellek.

A harmadik hipotézissel a vállalat belső pénzügyi adataitól eltérő, külső, mezoszintű inputok hatását vizsgálom a fizetésképtelenséget előrejelző modellek teljesítményére. A hagyományos megközelítés szerint az előrejelző modellek a vállalati pénzügyi adatokra támaszkodnak és csak néhány esetben találunk makrogazdasági, iparági és vállalatspecifikus változók felszámolásra gyakorolt hatásának elemzését a szakirodalomban. Ezekből az egyik felmerülő kérdés: vajon a külső inputok hatást gyakorolnak-e a fizetésképtelenségre, és ha igen az előrejelzési előny matematikailag kimutatható? Azt várnák, hogy a makrogazdasági, piaci és egyéb változókban olyan információk rejtőztek, amelyek nem tükröződnek a vállalati pénzügyi adatokban. A szakirodalmi megfigyelésekben azonban az alternatív típusú változók, a pénzügyi mutatókkal együtt gyakorolt hatása mérsékelt, ezért az előrejelző modellekben szinte észrevehetetlenek.

H4: A gépi tanuláson alapuló együttes (ensemble) módszerek előrejelzési pontossága szignifikánsan magasabb, mint az egyéb (logisztikus regresszió, SVM, neurális háló) előrejelzési módszereké egy- és több éves előrejelzési horizonton.

Végül a fizetésképtelenség előrejelzés örökös törekvésével összhangban, a leghatékonyabb módszer után kutatva, fogalmazom meg az utolsó

hipotézisem. Ebben a különböző módszerek teljesítményét mérem össze, és keresem a szakirodalomban gyakran felmerülő módszerek közül azt, amely a magyar vállalati adatokon a legjobban működik. A gépi tanulás tárházából az együttes módszerek ígéretesnek tűnnek, mivel egyesítik több tanuló eljárás képességét és ez által az összetett és nem lineáris problémákat könnyebben teljesítik (Dietterich, 1997). Feltételezem, hogy ez az előny állandó marad a logisztikus regresszió, az SVM, és még a neurális hálóval szemben is, több adatkészleten és különböző előrejelzési horizonton.

3. A kutatás módszerei és adatbázisa

A szakirodalomban általánosan alkalmazott módszereket és eljárásokat ismertettem, amelyeket relevánsnak ítéltam és így az empirikus vizsgálatokban is helyet kaptak. Ezek a módszerek a kutatás gerincét képezik, és egy keretet adnak a soron következő elemzésekhez. A csőd- és fizetéképtelenség statisztikai alapokon nyugvó klasszikus módszerei régóta fontos szerepet játszanak a vállalatok pénzügyi egészségének értékelésében. E módszerek közül megemlítem a hagyományos diszkriminanciaanalízist, amely Altman híres Z-score modelljével vált híressé. A módszer alkalmazását a számos korlátozó feltételek következtében elvettem. A másik klasszikus, de rendkívül széles körben alkalmazott eljárás a logisztikus regresszió, amely máig használatos eszköz a csődelőrejelzésre a rugalmasságának és magas hatékonyságának köszönhetően.

A fent említett diszkriminanciaanalízist és logisztikus regressziót gépi tanuláson alapul algoritmusok között említhetjük, azonban nem soroljuk az intelligens mesterséges intelligencia módszerek közé. Gépi tanulási eljárások között inkább a nemparametrikus módszereket értem és ilyenformán

ismertetem a döntés fán alapuló algoritmusokat, az SVM módszert és a neurális hálózatokat. Általános meghatározásban a gépi tanulás a mesterséges intelligencia azon részterülete, amely számítástechnikai alapokon nyugvó eljárásokkal lehetővé teszi olyan elemző adatmodellek létrejöttét, amelyek automatikusan, rejtett információk feltárására alkalmasak anélkül, hogy egyértelműen kódolnánk őket (Wang, 2017).

A Support Vector Machine általánosan kernelgépeknek nevezett tanulómódszert takar, amely a magyar terminológiában is főként SVM-ként használatos, azonban a „tartó vektor gépek” vagy „szupport vektorgépek” néven is ismert (Russell – Norvig, 2005). A módszer kifejlesztése Cortes és Vapnik (1995) szerzőpárhoz köthető. Az eljárás alkalmas lineáris és nem-lineáris elválasztásra.

A neurális hálózatok, a mesterséges intelligencia fő képviselői, amelyek alkalmazása a számítástechnika gyors fejlődésének köszönhetően jelentősen elterjedt. Felépítésük az emberi idegrendszer működésén alapszik, éppen ezért számos területen alkalmazható összetett problémák megoldására, például adatbányászat, jelfeldolgozás vagy karakter felismerés. A neurális háló matematikai modelljét McCulloch és Pitts (1943) dolgozta ki, alapul véve a biológiai neuron felépítését. Rosenblatt (1958) kifejlesztette az egyrétegű neuronhálót a perceptron algoritmussal, amely gyakorlatilag meghatározta a mesterséges neuronháló alkalmazásának kiindulópontját. A neurális hálók fő gyengesége, hogy „fekete dobozként” működnek. A gépi tanulási módszerek között a neuronháló az egyik, amelynél a legnehezebb megállapítani, hogy mely bemeneti tényezők mekkora súllyal befolyásolják a végső outputot, a fizetésképtelenségi modelleknél mely pénzügyi mutatók mekkora súllyal befolyásolják az osztályokba való besorolást (Balcean –

Ooghe, 2004). Másik gyenge pontja, hogy egy kiegyensúlyozatlan adatmintán mérsékelt teljesítmény várható, mivel hajlamos a besorolást a több adattal rendelkező osztályba sorolni, így a tesztminta előrejelzése torzított lesz. Továbbá a rejtett neuronok számának meghatározására nincs szabály, többnyire esetleges a megállapításuk. Az egyre több rejtett neuron számításigénye magas, azonban a kevés rejtett neuron nem képes megfelelő pontosságot elérni (Zhao et al., 2015).

A döntési fa módszer a rekurzív particionáló algoritmust alkalmazza, amely által klasszifikációs fákat állít elő egyszerű szabályok alapján. A fastruktúra létrehozása több lépéses folyamat, amelynek során a teljes mintát kétfelé osztja. Az elválasztást úgy valósítja meg, hogy a függő változó szempontjából az elválasztott csoporton belüli variancia a lehető legkisebb legyen, míg a csoportok közti variancia a legnagyobb legyen (Hátori, 2001). Addig állítja elő az elágazásokat, amíg van elválasztásra alkalmas változó. A fa tetején lévő első változó kiválasztása kulcsfontosságú. A módszer a magyarázó változók rangsorát elkészíti, szerre kipróbálja a változókat particionálásra és miután mindenik elválasztás megtörtént, az lesz a fa csúcán, amelyik a legkisebb hibát követi el. Majd a többi változó ugyanezzel a módszerrel kerül kiválasztásra, amíg fel nem épül a teljes fa (Virág et al., 2013). A döntési fa módszercsaládba sorolhatók azok az együttes (ensemble) módszerek, amelyek döntési fákon alapulnak: a bagging (zsákolás), a boosting (gyorsítás) és a random forest (véletlen erdő) metamódszerek. Ezeket különálló módszereknek tekintem. A bagging név a „bootstrap aggregálás” technika rövidítése, amely egy bizonyos független osztályozó eljárást foglal magába. A teljes adatmintából több különböző, véletlenszerűen kiválasztott almintát használ fel, visszatevéses eljárással, osztályozó eljárást alkalmaz rajta, majd végül a több különböző eredményt átlagolja. Ez pedig általában hatékonyabb

osztályozáshoz vezet, mint egy egyedülálló modell (Bodon – Buza, 2014). A létrehozott alminták bootstrap másolatok, amelyekre jellemző lehet, hogy az eredeti mintából egyes elemek többször előfordulhatnak, mások meg egyáltalán nem jelennek meg. Így átlagosan az eredeti tanulóadat 63,2%-a használandó a bootstrap alminták létrehozásánál (Breiman, 1996).

A boosting alapjait Schapire (1990) fektette le, amely mentén Freund – Shapire (1996) dolgozta ki az AdaBoost (Adaptive Boost) algoritmust, amelyet a következőkben ismertetek. Az alapgondolata hasonló a bagging eljáráshoz, hogy gyenge (weak) tanító eljárásokat használ erős tanító létrehozásához. Az adaptív elnevezés onnan ered, hogy a súlyozási paramétereket iteratív módon, az osztályozási teljesítmény alapján adaptív módon változtatja. Az osztályozóknak azonban bizonyos paramétereit rendszerint nem változtatjuk, például döntési fáknál azonos metszési módszert és vágási függvényt használunk (Bodon – Buza, 2014). A boostinggal a magyar terminológiában turbózás, fokozás vagy gyorsítás néven találkozhatunk (Russel – Norvig, 2005). A boosting a tanítóhalmazhoz hozzárendel egy $w_i \geq 0$ súlyt, amely az adott minta fontosságát reprezentálja. Az egyes osztályozások után a tanuló minta súlya változik az osztályozási teljesítménynek megfelelően. A következő osztályozó eljárás ezután az újrásúlyozott tanuló minta felhasználásával épül fel (Kim – Upneja, 2014). Hasonlóan mint a bagging, a boosting is számos statisztikai módszert és gépi tanuláson alapuló eljárást alkalmazhat osztályozásra, azonban gyakorta egy vágású döntési fa (döntési tönk, decision stump) alaptanulóval épül fel (James et al., 2013).

A random forest a magyar terminológiában véletlen erdő eljárásként is ismert, kifejezetten a döntési fa osztályozóhoz tervezett együttes módszer. A

módszer alapja a bináris fákat előállító klasszifikációs és regressziós fa (Classification and Regression Tree, CART). Több döntési fa előrejelzéseit átlagolja és az osztályba kerülést többségi szavazással dönti el (Breiman, 2001). A módszer a lehető legnagyobb metszés nélküli fákat hozza létre úgy, hogy mindenik elágazásnál a változók egy bizonyos halmazát használja fel, majd az elkészült nem korreláló fák eredményeit aggregálja (Figini – Savona – Vezzoli, 2016; Takács, 2021). A véletlen erdő a Bagging eljáráshoz hasonló, azonban annak továbbfejlesztett változatának tekinthető. Az alaptanuló algoritmus kizárólag döntési fa, amely több különböző bootstrap almintán fut le, ezért általában a teljes minta egyharmada nem kerül felhasználásra a fák kiépítésekor (Yeh – Chi – Lin, 2014). A döntési fával létrehozott bagging-nél a fák erősen korreláltak lehetnek, tehát ezek átlagolásával sem fog lényegesen csökkenni a variancia. A random forest megoldja ezt a problémát azzal, hogy minden vágásnál a tanulmányok egy részét használja, így korrelálatlan fákat hoz létre minden tanulási ciklusban (James et al., 2013). A kétszeres véletlenszerűsítés egyértelműen előnyösebb módszerré teszi a bagging módszernél. Először a megfigyelések véletlenszerű kiválasztása történik a döntési fa létrehozásakor, amely a bagging-nél is megjelenik. Majd a független változók köréből a döntési fa mindenik elágazásánál véletlenszerű mintavételre kerül sor (Nyitrai, 2021), és a legjobb elválasztást biztosítót meghagyjuk.

A rendelkezésre álló adatokból egy olyan minta összeállítása volt a cél, amely általánosan alkalmazható a magyar vállalatok fizetésektelenségének előrejelzésére, valamint felhasználható bizonyos módszertani kérdések megválaszolására. Az alább ismertetett keretek közt értelmezhetőek az eredmények.

A csőd kutatásoknál szokás szerint a fizetéképtelen vállalatokból indulunk ki mivel azok jelentik a szűk keresztmetszetet. Számosságban kevesebb van belőlük és sok esetben hiányosak a hozzájuk köthető információk. Ennek mentén az egyik megkötésem az volt, hogy a fizetéképtelen vállalatok esetén a fizetéképtelenségi eljárást megelőző közvetlen első évből álljanak a rendelkezésre pénzügyi adatok. Ez azért lényeges, mert a fizetéképtelenségre jellemző tulajdonságok rendszerint egyre inkább érvényüket veszítik, ahogy távolodunk a fizetéképtelenség időpontjától.

A vállalati méret alapján a mikro, kis és közepes méretű vállalatokra korlátozódtam, gazdasági forma szerint csak a korlátolt felelősségű társaságokat vizsgáltam, ezzel megtartva egy elég nagy merítési méretet. A foglalkoztatotti létszám tekintetében a mintába a minimum 2 és maximum 249 foglalkoztatott személlyel rendelkező vállalat kerülhetett be. További bekerülési kritérium volt, hogy a vállalat mérlegfőösszege és árbevétele legkevesebb 1-1 millió forint értékű legyen és legalább két lezárt üzleti évvel rendelkezzen. Csak az év végi, teljes üzleti évre vonatkozó adatokat vettem figyelembe, a részleges félévközi adatokat elvettem.

Az elemzett iparágak tekintetében nem volt kizárási feltétel, mindössze annyi, hogy kimaradtak a pénzügyi intézményekre vonatkozó, sajátos számviteli szabályokkal működő bankok és biztosítási társaságok mivel a vállalatoktól eltérő mérleg és eredménykimutatás szerkezettel rendelkeznek. Az iparági elhatárolás érdekében a nemzetgazdasági ágakat ismertetem, amelyeket az elsődleges tevékenység TEÁOR száma alapján határoztam meg. Hozzá kell tenni, azt hogy ilyen tekintetben a vállalatokat nagyon erőltetetten lehet besorolni egyetlen iparágba, mivel jellemzően egyszerre több tevékenységet folytathatnak. Gyakran előfordul, hogy bizonyos

másodlagos tevékenységek nagyobb árbevételt biztosítanak, mint a statisztikai főtevékenység, ezért az iparági besorolás félrevezető lehet (Deák, 2000). Mivel nem álltak rendelkezésre részletesebb adatok, ezért a statisztikai célra kijelölt főtevékenységet vettem alapul a vállalatok iparági besorolásánál.

A fizetéképtelenséget előrejelző modellek építőkövei a magyarázó változók, amelyek összetett kölcsönhatásainak feltárásával fontos betekintés nyerhetünk a fizetéképtelen és egészséges vállalatok rejtett pénzügyi mintázataiba. Mindez pedig lehetőséget teremt az általánosításra és az egyedi vállalatok problémás területein történő beavatkozásra.

A modellezés során a függő változó a fizetéképtelenség ténye, amely a felszámolási jogi eljárás alapján került megállapításra. Ez a kétértékű változó 1-es értéket kapott, amennyiben a vállalat fizetéképtelen és 0-át, ha fizetőképese.

A csődelőrejelzés szakirodalmában a magyarázó változók köre kimondottan a pénzügyi adatok felhasználásából adódik. A csőd kutatások több mint 90%-ban a vállalatokhoz kapcsolódó pénzügyi bemeneti változók figyelhetők meg du Jardin (2009) szerint. Ezzel egybevágó következtetést tudtam levonni a saját irodomelemző kutatásban is (Ágoston, 2022a). A makrogazdasági változók alkalmazásának az aránya 3-5% közötti, amelynél némileg magasabb arányú a piaci alapú bemeneti változók megjelenése a kutatásokban (du Jardin, 2009; Ágoston, 2022a).

A felhasznált független magyarázó változók körét a kutatási kérdéseknek megfelelően válogattam ki, így két fő csoportba sorolhatók. Az egyik csoportban vannak a vállalatok belső attribútumait leképző pénzügyi és

méretre vonatkozó jellemzők. A másik csoportban vannak a külső környezeti tényezőket megvilágító mutatók, amelyek a vállalatok székhelyén érvényesülő mezoszintű magyarázó változók. A kiválasztásban segítségemre volt a magyar csődelőrejelzésben megalapozott magyarázó változók Virág (2001), Virág és Kristóf (2005), Nyitrai (2014), Virág és Nyitrai (2014) munkáiból. Ezen kívül Bellovary et al. (2007) irodalomösszefoglaló munkája, Altman (1968), Carton és Hofer (2006), du Jardin (2015), Linag et al. (2016), Zięba et al. (2016) és a saját empirikus tapasztalatok alapján jelöltem ki a magyarázó változókat (Ágoston, 2022b; Ágoston, 2023; Ágoston, 2024). A mezoszintű változók kiválasztását Buchler, Kaiser és Jaeger (2012), és Arcuri, Brunetto és Levratto (2019) munkája inspirálta.

A kiválasztott mutatókat, a jövedelmezőség, likviditás, eladósodottság, hatékonyság, eszköz- és tőkeszerkezet, cash flow, méret és egyéb típusokba soroltam. Összesen ötven pénzügyi és egyéb, a vállalat belső információjához köthető változót vettem figyelembe (jelölésük X1-X50), emellett hét mezoszintű magyarázó változót választottam az elemzésekhez (jelölésük Y1-Y7). A pénzügyi magyarázó változók mindegyike az év végi záró mérleg- és eredménykimutatásból kalkulált hányados típusú arányszám. A vállalat belső jellemzőit leíró pénzügyi és méretre vonatkozó mutatókat nyolc típusba sorolhatjuk. A jövedelmezőséget tükröző mutatókból kilencet számítottam ki, köztük a ROA, a ROE, a ROS több változatát, a forgótőke arányos árbevételt és a bérarányos eredményt. A likviditást jellemző mutatók száma nyolc, csak egy néhány említve kalkuláltam a likviditása rátát, a gyorsrátát, a pénzhányad mutatót és a működő tőkéhez kapcsolódó mutatókat. Az eladósodottság mutatói közül a bonitás, a saját vagyon aránya, a kötelezettségek aránya mutatók emelhetők ki, összesen tizenkettő eladósodottsági mutatót választottam ki. A hatékonyság mutatói közül hatot

számítottam ki, köztük az eszközök-, készletek és vevők forgási sebesség mutatóit. Az eszköz és tőkeszerkezeti mutatók száma szintén hat, a cash flow alapú mutatók száma öt. A pénzügyi mutatókon túlmenően a vállalat mérlegfőösszegének és árbevételének a természetes alapú logaritmusát alkalmaztam, mint vállalatok méretére jellemző mutatót. Az egyéb típusú változók között említtem az alkalmazottak átlagos létszámát a vizsgált üzleti évben és a vállalat működésének korát években kifejezve.

A mezoszintű mutatók a vállalatok székhelye alapján kerültek kiválasztásra. Ezek a vállalat külső körülményeit leíró, adott településre jellemző változók. Az első közülük a település vállalatainak jövedelmezőségét jelző, egy vállalatra jutó árbevétel, amelynek feltételezett hatása a vállalati fizetéképtelenségre tekintve negatív. A helyi munkanélküliségi ráta az álláskeresők és a gazdaságilag aktív korú népesség hányadosa, amely egyben a helyi gazdaság egészségi állapotának a mutatója is. A munkanélküliségi ráta alacsony szintje kedvező gazdasági környezetet, valamint fokozott gazdasági aktivitást jelez, amelyben a lakossági fogyasztás által megnövekedett vállalati bevétel várható. Ezzel szemben az emelkedő munkanélküliségi ráta a fogyasztáson alapuló ágazatokban a kereslet csökkenéséhez és ezzel együtt a bevételek zsugorodásához vezet. A településen az ezer főre jutó bankfiókok és takarékszövetkezeti fiókok száma a pénzügyi infrastruktúra szintjét méri. A banki források elérhetőségei és költségei fontos szerepet játszanak abban, hogy a vállalatok finanszírozni tudják a működésüket. Azon kívül, hogy alapvető pénzügyi támogatást nyújtanak, lehetővé teszik a növekedést a befektetési lehetőségek finanszírozásával, valamint erősíthetik az ellenálló képességet gazdasági visszaesések idején. Távoli hozzáféréssel (mobilbank, internetbank) rendkívül sok banki tevékenység elvégezhető, azonban a bankfiókok jelentősége hosszú ideig még megmarad (Helmeczi, 2010).

A település önkormányzatához kapcsolódó mutatót is bevontam a modellezésbe, mivel a vállalatok telephelyválasztásának lényeges eleme. Az egy főre jutó önkormányzati adóbevételek, az önkormányzat pénzügyi erejét és a település fejlettségét tükrözi. A beszedett adók mértéke az ingatlanpiactól, a gépkocsiparktól illetve az ott élők gazdasági erejétől függ. Minél magasabb az érték, annál függetlenebb a helyi önkormányzat a központi költségvetéstől. A település megközelíthetőségét, a közúti közlekedés és szállítási lehetőségek helyzetét képviselő változók a főváros távolsága, és a legközelebbi gyorsforgalmi csomópont távolsága. A főváros mutató percben fejezi ki egy adott település távolságát a leggyorsabb közúton, a legközelebbi gyorsforgalmi csomópont mutató a legrövidebb úton fejezik ki a kilométerben mért távolságot. A város, kétértékű változó, a település jogállását tükrözi, amely a város és község elválasztásának mérésére szolgál azzal a céllal, hogy a városi/községi elhelyezkedés jelentőségét mérje a fizetéseképtelenség szempontjából.

4. Az értekezés tudományos eredményei és az eredmények hasznosíthatóságának lehetőségei

A hipotézisek tesztelésén keresztül a doktori értekezés számos ponton kapcsolódik a szakirodalmi vitákhoz és újszerű megállapításokat tesz, amelyek hozzájárulnak a tudományterület előremozdulásához.

A magyar vállalati adatmintán, a fizetéseképtelen és egészséges eloszlás alapján a reprezentatív mintavétel nem célravezető egy átlagosan jól teljesítő fizetéseképtelenséget előrejelző modell kifejlesztéséhez, sokkal inkább ajánlott az azonos osztálylétszám alkalmazása. A reprezentatív minta még megfelelő súlyozást követően is azt eredményezi, hogy a nagyobb számban

lévő egészséges vállalatokra idomulnak a tanuló algoritmusok, a statisztikai alapokon nyugvó módszerek és a mesterséges intelligencia alapú eljárások egyaránt. Ebből adódóan a modellek átlagos osztályozása kiváló lesz, azonban a fizetéképtelen megfigyeléseket képtelen helyesen osztályozni. Ezzel pedig az egyik leglényegesebb cél vesz el, a problémás vállalatok azonosítása. A szűkös adatállományoknál, vagy a kis mintából dolgozó elemzéseknél az azonos osztálylétszámon kívül nincs is egyéb opció. Bár a reprezentativitásra törekvő mintavétel megbízható képed ad a populációról, a fizetéképtelenség pontos előrejelzésében nem kívánatos. Az eredmény és a hozzá kapcsolódó módszertani ajánlás hozzájárul a szakirodalomban tapasztalható mintavételi vita előmozdításához.

2. A több pénzügyi év adatain készült modellek előrejelző teljesítménye átlagosan nem bizonyult magasabb teljesítményűnek a mindössze egy év adataival készült modellhez mérten. Ezek a megállapítások az előzetes feltevéssel ellentétesek, mely szerint a több év adatainak beépítése szükségszerűen jobb előrejelzési pontosságot eredményez. Ehelyett a modellek átlagos előrejelző ereje nem növekszik további múltbeli adatok bevonásával, hanem mintegy megegyezik az előrejelzést közvetlen megelőző egy év adatain mért eredményekkel. Ez pedig döntő fontosságú a csődelőjelző modellfejlesztési stratégiák kialakításának szempontjából és arra ösztönözi a téma művelőit, hogy a mennyiség helyett a minőségre fektessék a hangsúlyt. Érdeemes tehát a legfrissebb és legrelevánsabb adatokra támaszkodni és azokra építeni a modelleket, ami hatékonyabb modellépítést és időszerű eredményeket biztosít. Az is körvonalazódik, hogy a halmozódó, egyre bővülő adatállomány nem rendelkezik hozzáadott értékkel az előrejelzés tekintetében. Ennek a kérdésnek a bizonyítása további elemzést igényel és egy későbbi tanulmányban kerülhet kifejtésre.

3. A mezoszintű külső tényezők segítségével egy tágabb gazdasági összefüggésben értelmezhetjük a fizetéseképtelenség kialakulását. A település szintű munkanélküliségi ráta szignifikáns pozitív hatással van a fizetéseképtelenségre. Azonban átlagosan a pénzügyi mutatók és más mezoszintű változók mellett az előrejelzési teljesítményt nem fokozza. A munkanélküliségi ráta, mint a helyi üzleti környezet fokmérő mutatója minden adatállományon szignifikáns magyarázó változója a fizetéseképtelenségnek. Emellett a többi beválogatott mezo-változó, a helyi árbevétel, a fizikai bakfiókok száma és a település jogállása nem bizonyult szignifikánsnak és együttesen az előrejelző teljesítményt sem növelték. Sőt a modellek átlagos teljes besorolási aránya és az AUC-érték alapján a mezo-változókkal készült modellek alacsonyabb teljesítményűnek bizonyultak, mintha csak pénzügyi mutatókat önmagukban használtam volna. Az eredmények tükrében a vállalatok fizetéseképtelenségének előrejelzésére inkább a vállalat belső pénzügyi változói ajánlhatóak. Természetesen fontolóra vehetők az olyan nem pénzügyi mutatók is, mint az iparági tényezők, menedzsment minősége, piaci hangulat és trendek mutatói, amelyek hatással lehetnek a vállalatok fizetéseképtelenségére, de ugyanakkor a pénzügyi mutatókban nem tükröződnek, így általuk akár növelhető lehet az előrejelző pontosság.

4. Az alkalmazott módszerek közül átlagosan a legkiválóbb teljesítményt az együttes (ensemble) vagy metamódszereként ismert eljárások érték el, egyes és több éves előrejelzési horizonton egyaránt. Ezek közül nem lehet egyértelmű győztes módszert megállapítani azonban ki lehet emelni a random forest-et, amely több szempontot figyelembe véve átlagosan egy árnyalattal hatékonyabb teljesítményt ért el mint a módszercsaládba tartozó adaboost és bagging. A logisztikus regresszió, mint a fizetéseképtelenség

klasszikus módszere meglehetősen jól teljesített. A gyors tanulási képesség és könnyű interpretálhatósági előny mellett viszonyítási alapként kifejezetten ajánlott. A módszer eredmények mellé kiegészítendő, hogy a bemeneti változók skálázására érzékeny módszereknél, mint amilyen a lineáris SVM, javasolt a változók standardizálása, amellyel nem csak a tanulási ciklus gyorsítható fel, hanem a módszer teljesítménye is szignifikánsan javul.

5. A fizetésképtelenség azonosítására voltak olyan mutatók, amelyek valamennyi vagy csaknem mindegyik modellben konzisztensen szerepletek mint magyarázó változók. Az adott minta jellemzői mellett a következőket tekinthetjük a legjelentősebb változóknak a fizetésképtelenség magyarázatára, mivel a sokrétű változószelekciós eljárások gyakran fontos változónak jelölték meg. Ezek közül az első számú a vállalatok működési éveinek a száma (X50), amely az összes modellben szerepelt magyarázó változóként. A mutatónál a kisebb érték jelent fizetésképtelenségi kockázatot. A megfigyelések között a fizetésképtelen osztály középértékei 5-6 év volt, míg a fizetőképeseké 8-10 év. A számos adatbázison futtatott logisztikus regresszió esélyhányadosai alapján a működési évek egységnyi növekedése 4%-7%-al csökkenti a fizetésképtelenség esélyét ceteris paribus. A pénzügyi mutatók közül a leggyakrabban kiválasztott volt a készletekkel csökkentett forgóeszközök és összes eszköz aránya (X39), a pénzeszközök aránya a forgóeszközökön belül (X37) és a saját tőke összes kötelezettség aránya (X28). A készletekkel csökkentett forgóeszközök és összes eszköz aránya (X39) gyakorlatilag a pénzeszközök, követelések és értékpapírok együttes arányát mutatja az eszközök összértékén belül. A mutató egészséges középértéke 45%-50% között helyezkedik el, míg a fizetésképteleneké 56%-72% között, tehát ezek fokozott értéke növeli a fizetésképtelenség esélyét. Arra kérdésre, hogy a likvid eszközöknek a mutatónak a számlálójában

milyen arányok érvényesülnek az egészséges és fizetéképtelen osztályban, két megállapítást tehetünk. A rövid lejáratú likvid értékpapírok birtoklása a mintában szereplő vállalatok nagy kis hányadára jellemző (4%-5%), szinte valamennyi vállalatnál ez nulla értékű. A pénzeszközök forgóeszközökön belüli súlyára pedig az X37-el jelölt szintén fontosnak értékelt magyarázó változó ad iránymutatást. Mindez pedig rávilágít, hogy a pénzeszközök aránya a forgóeszközökön belül a fizetéképteleneknél számottevően alacsonyabb, mint az egészségeseknél. Ebből következően pedig visszautalhatunk az X39 változóra is, hogy a magasabb arányú likvid eszközök mögött a követelések nagy aránya szerepel, így követelémennyment problémákra utal a fenti két változó, amely miatt a fizetéképtelenség kialakul. Arra a kérdésre viszont, hogy ez a hitelpolitikai stratégiák elégtelenségéből, a követelés kezelés problémáiból, ügyfélkommunikációs hiányosságokból, körbetartozásokból vagy netán kevésbé diverzifikált ügyfélkörből és gyenge vagy teljesen hiányzó partnerkockázat értékelésből adódik, a rendelkezésre álló adatokból nem volt lehetőség választ adni. Erre egy következő, mélyebb fókuszú kutatás reflektálhat. További eredmények alapján a saját tőke/idegen tőke aránynál (X28) a 100% és afeletti érték tartozik az egészséges osztályhoz. A fizetéképtelenekre jellemző a mutató negatív értéke, amely egyben a sorozatos üzleti veszteségek indikátora is.

5. Jövőbeli kutatási irányok

A tanulmány több módszert ismertet és validál, amelyek jó előrejelzési teljesítményre képesek, azonban nem méríti ki az ezekben rejlő összes kapacitásokat. Nem beszélve az alternatív módszerek és hibrid megoldások teszteléséről. Az együttes módszereknél például a döntési fa alaptanuló helyett

meg lehetne fontolni a neurális háló alkalmazását. Meglepő eredménye a tanulmánynak, hogy a legnépszerűbb módszerek közé sorolható neurális háló csupán mérsékelt teljesítményt ért el. Ennek fokozására a legkorszerűbb mélytanuló neuronháló specifikálása nyújthat megoldást, amelynek végrehajtása szintén javasolható későbbi kutatások témájának.

A tanulmányok túlnyomó többségében a fizetésképtelenség vagy csőd állapotát a jogi eljárások alapján határozzák meg. A jövőbeni kutatásoknál javasolt a pénzügyi nehézség definícióját más megközelítésből is vizsgálni, amely a jövedelmezőség mutatóin vagy, a lehetőségekhez mérten, piaci mutatókon alapszik. Ez pedig hatékonyabb megoldást is biztosíthat olyan vállalatok esetén, amelyek a fizetésképtelenség szélén állnak, de ellenük még nem indult meg a jogi eljárás.

A modellek bemeneti adatát képző változók az éves jelentésből nyert pénzügy arányszámok. Az éves jelentésekben a mérleg csak egy adott pillanatban, az üzleti év végén érvényes értékeket mutatja, így az adott évben az állományok időbeli alakulása teljes mértékben ismeretlen. Továbbá az a kérdés is felmerülhet, hogy a kkv-k beszámolóit mennyire megbízhatóak, lehetnek-e benne hibák, torzítások, amelyek a végső következtetéseinket is elferdítik. Ugyanakkor egyelőre csak ezekre támaszkodhatunk, a pénzügyi mutatók kiváltására nincs más mód és úgy látszik, hogy nem is érdemes például külső változókkal helyettesíteni. A külső, mezoszintű változók alkalmazásával a pénzügyi változóktól egy fokkal távolabbi perspektívából vizsgáltam a vállalatok túlélését. Ettől még távolabbi fókuszpont is elképzelhető, amelyekre szintén nincs közvetlen hatásuk a vállalatoknak, ugyanakkor e külső jelenségek és tényezők erőteljesen befolyásolhatják a vállalatok teljesítményét és pénzügyi egészségét. Ilyenek például a

makrogazdasági események és sokkok, szabályzási változások vagy geopolitikai események.

Ajánlatos figyelembe venni a korlátozott számítási kapacitásokat is. A mai modern számítástechnikai eszközök kiváló teljesítményre képesek, azonban szem előtt kell tartani, hogy a gépi tanulás algoritmusai rendkívül időigényesek lehetnek és jelentős erőforrást emésztnek fel. A több tízezres rekordok összetett gépi tanulási folyamatokra az otthoni vagy irodai használatra kiadott számítógépek kevésbé alkalmasak.

A tanulmány térbeli szempontból statikus, figyelmen kívül hagyja a földrajzi elhelyezkedést a fizetéseképtelenség tekintetében. Jövőbeli kutatásoknál megfontolandó lehet a csőd és fizetéseképtelenség vizsgálatát térbeli alapokra helyezni, kísérletet tenni a földrajzi tér hatásának beágyazására az előrejelző modellekbe. Érdemes lehet megvizsgálni annak a lehetőségét, hogy a magyarázó változók hatásainak kiszűrése után, marad-e térbeli függőség a közeli és távoli elhelyezkedésű vállalatok között a gazdasági-társadalmi interakcióknak köszönhetően.

A tézisfüzetben felhasznált irodalom

- Alaka, H. A.** – Oyedele, L. O. – Owolabi, H. A. – Kumar, V. – Ajayi, S. O. – Akinade, O. O. – Bilal, M. (2018): Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, Vol. 94. pp. 164-184.
- Altman, E. I.** (1968): Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *The Journal of Finance*, Vol. 23. No. 4. pp. 589-609.

- Arcuri, G.** – Brunetto, M. – Levratto, N. (2019): Spatial patterns and determinants of firm exit: an empirical analysis on France. *The Annals of Regional Science*, Vol. 62. No.1. pp. 99-118.
- Balcaen, S.** – Ooghe, H. (2004): *Alternative methodologies in studies on business failure: do they produce better results than the classic statistical methods?* Vlerick Leuven Ghent Working Paper Series 2004/16. Vlerick Leuven Ghent Management School, University of Ghent, Ghent.
- Bellovary, J.** – Giacomino, D. – Akers, M. (2007): A review of bankruptcy prediction studies: 1930-present. *Journal of Finance Education*, Vol. 33. pp. 1-42.
- Berg, D.** (2007): Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, Vol. 23. No. 2. pp. 129-143.
- Bodon F.** – Buza K. (2014): *Adatbányászat* [on-line]. <https://dtk.tankonyvtar.hu/handle/123456789/12379> (2023.12.18.)
- Breiman, L.** (1996): Bagging predictors. *Machine Learning*, Vol. 24. No. 2. pp. 123-140.
- Breiman, L.** (2001): Random forests. *Machine Learning*, Vol. 45. No. 1. pp. 5-32.
- Buehler, S.** – Kaiser, C. – Jaeger, F. (2012): The geographic determinants of bankruptcy: evidence from Switzerland, *Small Business Economics*, Vol. 39. No. 1. pp. 231-251.
- Carton, R. B.** – Hofer, C. W. (2006): *Measuring organizational performance: Metrics for entrepreneurship and strategic management research*, Edward Elgar Publishing, Cheltenham, UK.
- Cortes, C.** – Vapnik, V. N. (1995): Support-Vector Networks. *Machine Learning*, Vol. 20. No. 3. pp. 273-297.

- Deák Sz.** (2000): A Porter-féle rombusz-modell főbb közgazdasági összefüggései. In: Lengyel I. (szerk): *Versenyképesség-regionális versenyképesség*. Szegedi Tudományegyetem Gazdaságtudományi Kar Közleményei. JATEPress. Szeged. pp. 67-87.
- Dietterich, T. G.** (1997): Machine-learning research. *AI magazine*, Vol. 18. No. 4. pp. 97-136.
- du Jardin, P.** (2009): Bankruptcy prediction models: How to choose the most relevant variables? *Bankers, Markets & Investors*, No. 98. pp. 39-46.
- du Jardin, P.** (2015): Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, Vol. 242. No. 1. pp. 286-303.
- Edmister, R. O.** (1972): An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, Vol. 7. No. 2. pp. 1477-1493.
- Fejér-Király G.** (2015): Bankruptcy prediction: A survey on evolution, critiques, and solutions. *Acta Universitatis Sapientiae, Economics and Business*, Vol. 3. No. 1. pp. 93-108.
- Figini, S.** – Savona, R. – Vezzoli, M. (2016): Corporate default prediction model averaging: A normative linear pooling approach. *Intelligent Systems in Accounting, Finance and Management*, Vol. 23. pp. 6-20.
- Freund, Y.** – Schapire, R. E. (1996): Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*, Bari, Italy.
- Gombola, M. J.** – Haskins, M. E. – Ketz, J. E. – Williams, D. D. (1987): Cash flow in bankruptcy prediction. *Financial Management*, Vol. 16. No. 4. pp. 55-65.
- Hámori G.** (2001): A CHAID alapú döntési fák jellemzői. *Statisztikai Szemle*, 79. évf. 8. sz. pp. 703-710.

- Helmecci I.** (2010): *A magyarországi pénzforgalom térképe*, MNB-tanulmányok 84. Magyar Nemzeti Bank, Budapest.
- James, G.** – Witten, D. – Hastie, T. – Tibshirani, R. (2013): *An introduction to Statistical Learning with Applications in R*. New York, Springer Science+Business Media, 426 p.
- Kim, S. Y.** – Upneja, A. (2014): Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling*, Vol. 36. pp. 354-362.
- Kristóf T.** – Virág M. (2019): A csődelőrejelzés fejlődéstörténete Magyarországon. *Vezetéstudomány*, 50. évf. 12. sz. pp. 62-73.
- 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*, Vol. 252. No. 2. pp. 561-572.
- McCulloch, W. S.** – Pitts, W. (1943): A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, Vol. 5. pp. 115-133
- Nyitrai T.** (2014a): Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében. *Statisztikai Szemle*, 92. évf. 4. sz. pp. 357-377.
- Nyitrai T.** (2015): Hazai vállalkozások csődjének előrejelzése a csődeseményt megelőző egy, két, illetve három évvel korábbi pénzügyi beszámolók adatai alapján. *Vezetéstudomány*, 46. évf. 5. sz. pp. 55-65.
- Nyitrai T.** (2021): A gépi tanulás módszereinek alkalmazása R-ben. *Statisztikai Szemle*, 99. évf. 2. sz. pp. 173-198.
- Oommen, T.** – Baise, L. G. – Vogel, R. M. (2011): Sampling bias and class imbalance in maximum-likelihood logistic regression. *Mathematical Geosciences*, Vol. 43. pp. 99-120.

- Rosenblatt, F.** (1958): The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, Vol. 65. No. 6. pp. 386-408.
- Russell, S.** – Norvig, P. (2005): *Mesterséges intelligencia modern megközelítésben*. Panem Könyvkiadó, Budapest.
- Schapire, R. E.** (1990): The strength of weak learnability. *Machine Learning*, Vol. 5. No. 2. pp. 197-227.
- Shetty, U.** – Pakkala, T. P. M. – Mallikarjunappa, T. (2012): A modified formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India. *Expert Systems with Applications*, Vol. 39. No.2. pp. 1988-1997.
- Takács O.** (2021): Nemek közötti bérkülönbségek Magyarországon: a véletlenerdő-és az OLS-becslésen alapuló Blinder–Oaxaca-dekompozíció eredményeinek összehasonlítása. *Statisztikai Szemle*, 99. évf. 1. sz. pp. 5-45.
- Virág M.** (2001): *Pénzügyi elemzés csődelőrejelzés*, Kossuth Kiadó, Budapest, 112 p.
- Virág M.** – Kristóf T. (2005): Az első hazai csődmodell újraszámítása neurális hálók segítségével. *Közgazdasági Szemle*. 52. évf. pp. 144-162.
- Virág M.** – Fiáth A. – Kristóf T. – Varsányi J. (2013): *Pénzügyi elemzés, csődelőrejelzés, válságkezelés*. Budapest, Kossuth Kiadó, 301 p.
- Virág M.** – Nyitrai T. (2014): Metamódszerek alkalmazása a csődelőrejelzésben. *Hitelintézeti Szemle*, 13. évf. 4. sz. pp. 180-195.
- Wang, N.** (2017): Bankruptcy prediction using machine learning. *Journal of Mathematical Finance*, Vol. 7. No. 4. pp. 908-918.

- Yeh, C.-C.** – Chi D.-J. – Lin, Y.-R. (2014): Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, Vol. 254. pp. 98-110.
- 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*, Vol. 58. pp. 93-101.
- Zhao, Z.** – Xu, S. – Kang, B. H. – Kabir, M. M. J. – Liu, Y. – Wasinger, R. (2015): Investigation and improvement of multi-layer perception neural networks for credit scoring. *Expert Systems with Applications*, Vol. 42. No. 7. pp. 3508-3516.

Publikációk és konferencia előadások

Megjelent publikációk

Ágoston N. (2024): A többskálás földrajzilag súlyozott regresszió alkalmazása a településszintű vállalati felszámolások hazai vizsgálatához, 2019. *Területi Statisztika*, 64. évf. 2. sz. pp. 125-149.

Ágoston N. (2023): Spatial analysis of firm exit in European regions. *Journal of East European Management Studies*, The CEE Region on a New Wave of Transition, Special Issue, pp. 165-188.

Ágoston N. – Fejér-Király G. (2023): A székelyföldi városi önkormányzatok költségvetésének elemzése. In: Nagy, Benedek (szerk.) *Csiki Közgaz: 20 éves a közgazdasági oktatás a csikszeredai Sapientián*. Mark House, Gyergyószentmiklós, pp. 65-89.

Madaras Sz. – Ágoston N. A rule-based expert system for financial investment decisions in Romania. In: Nagy, Benedek (szerk.): *16th Annual International Conference on Economics and Business, Challenges in the Carpathian Basin*, conference volume, Risoprint, Cluj-Napoca, pp. 328-338.

Ágoston N. (2022a): Külföldi csődelőrejelző módszerek szisztematikus irodalomelemzése. *Vezetéstudomány / Budapest Management Review*, 53. évf. 1. sz. pp. 69–89.

Ágoston N. (2022b): Mesterséges intelligencia és gépi tanulási módszerek a vállalati fizetésektelenség becslésére. *Statisztikai Szemle*, 100. évf. 6. sz. pp. 584-609.

Ágoston N. (2022): A kiegészítő mellékletek szerepe a hitelezési döntéstámogatásban. In: Benke, Mariann; Schmuck, Roland; Riedelmayer, Bernadett (szerk.) 3. *Farkas Ferenc Nemzetközi Tudományos Konferencia: „Menedzsment forradalmak”*, Pécsi Tudományegyetem Közgazdaságtudományi Kar Vezetés- és Szervezéstudományi Intézet, pp. 551-563.

Ágoston N. (2022): A vállalatok megszűnésének térbeli vizsgálata az európai régiókban. In: Veresné Prof. Dr. Somosi Mariann, Dr. Lipták Katalin, Harangozó Zsolt (szerk.) : *„Mérleg és Kihívások – Fenntarthatóság”*, XII. *Nemzetközi Tudományos Konferencia*, Miskolci Egyetem Gazdaságtudományi Kar, pp. 627-636.

Fejér-Király G. – Ágoston N. – Varga, J. (2019): Modelling the Financial Failure of Romanian Stock Companies. *Acta Universitatis Sapientiae, Economics and Business*, Vol. 7. No.1. pp. 65-82.

Konferencia előadások

Madaras Sz. – Ágoston N. (2023): *A rule-based expert system for financial investment decisions in Romania*. 16th Annual International Conference on Economics and Business: Challenges in the Carpathian Basin. Sapientia Hungarian University of Transylvania, Faculty of Economics, Socio-Human Sciences and Engineering, Csíkszereda, 2023. május 10-12.

Ágoston N. (2022): *A vállalatok megszűnésének térbeli vizsgálata az európai régiókban.* XII. Nemzetközi Tudományos Konferencia: „Mérleg és Kihívások – Fenntarthatóság”, Miskolci Egyetem Gazdaságtudományi Kar, Lillafüred, 2022. október 13-14.

Ágoston N. (2022): *A kiegészítő mellékletek szerepe a hitelezési döntéstámogatásban.* 3. Farkas Ferenc Nemzetközi Tudományos Konferencia: „Menedzsment forradalmak”, Pécsi Tudományegyetem Közgazdaságtudományi Kar, 2022 június 3-4.

Ágoston N. (2021): *Vállalati fizetéseképtelenség és csőd előrejelzés COVID-19 pandémia idején.* XIV. Pécsi Pénzügyi Napok, Post-Covid gazdaság kihívásai. Pécsi Tudományegyetem Közgazdaságtudományi Kar, 2022 április 28-29.

Ágoston N. (2021): *Vállalati fizetéseképtelenség modellezése többdimenziós skálázás segítségével.* 15th Annual International Conference on Economics and Business: Challenges in the Carpathian Basin, Global Challenges - Local Answers. Sapientia Hungarian University of Transylvania, Faculty of Economics, Socio-Human Sciences and Engineering, Csíkszereda, 2021. március 3-4.