

A BOOSTING ÉS A VÉLETLEN ERDŐ MODELL BANKCSŐD-ELŐREJELZŐ KÉPESSÉGÉNEK ÖSSZEHASONLÍTÁSA

Bankfelügyeleti szempontú visszatekintés a 2008-as pénzügyi válságra

Sen, Safa¹

ABSZTRAKT

A tanulmány átfogóan elemzi a gazdasági stabilitás szempontjából is kiemelten fontos terület, a bankcsőd-előrejelzés terén használt modellek teljesítményét. Az Egyesült Államok szövetségi betétbiztosítási intézménye, a Federal Deposit Insurance Corporation (FDIC) adathalmazát felhasználva 950 bankot vizsgált, köztük 60 olyan intézményt, amely nem vészelte át a 2008-as pénzügyi válságot. Az elemzés 26 CAMEL-mutatósámmra alapuló bináris osztályozás keretében veti össze a boosting (javító) algoritmusok és a véletlen erdő (random forest) modellcsalád teljesítményét. A csődöt elkerülő bankok osztályozásában egyértelműen jobban teljesítenek a boosting-algoritmusoknál a véletlen erdő eljárások, amelyek az esetek 97%-ában helyesen osztályozták az adatokat. Ezek közül is kiemelkedő a regularizált véletlen erdő modell precizitása (0,988). A csődbe jutó bankok előrejelzése terén a véletlen erdő modellek, és különösen a regularizált modellváltozat nagy pontossággal azonosítják a valódi csődhelyzeteket. Mindezek alapján a véletlen erdő modellek képesek pontosan és megbízhatóan előrejelezni a bankcsődöt, ezért kulcsfontosságú szerepet kaphatnak a tévesen azonosított esetek minimalizálásában, ami elengedhetetlen a megalapozott előrejelzésekhez a bankszektorban.

JEL-kódok: C45, C53, G12, G17

Kulcsszavak: gépi tanulási modellek, bankcsőd, külső ellenőrzés, költségérzékeny erdő, XGBoost

¹ *Sen, Safa* PhD-hallgató, Miskolci Egyetem, Gazdaságtudományi Kar. E-mail: sen.safa@student.uni-miskolc.hu.

1. BEVEZETÉS

A gazdasági növekedés és stabilitás fenntartása szempontjából rendkívül fontos a bankcsődök tanulmányozása, amely összetett és sokrétű kutatási terület. A megtakarítókat a kölcsönzőkkel összekötő és a pénzügyi likviditást biztosító pénzügyi közvetítőként a bankok központi szerepet töltenek be a gazdaság szövevényében. Csődhelyzetbe kerülésük ugyanakkor mélyreható következményekkel jár nemcsak az érintett intézmény, de a tágabb gazdasági rendszer számára is. Tanulmányom a bankcsődök bonyolult sajátosságait, hatását és az azokat előrejelző modellek fejlődését járja körül.

A bankcsőd mindig is jelentős zavart keltett a gazdaságban, a pénzügyi szektoron túlmutató következményekkel. A bankok összeomlása akadályozza a tőke áramlását, ennek pedig olyan gazdasági következményei lehetnek, mint a fogyasztói és üzleti bizalomvesztés, a fogyasztás visszaesése, vagy rosszabb esetben akár gazdasági recesszió. A 2008-as pénzügyi válság szemléletesen példázza a bankcsődökben rejlő rendszerszintű kockázatokat és azok világméretű hatását.

A bankcsőd-előrejelzés jelentős fejlődésen ment keresztül az évek során. Meghatározóak voltak a terület úttörői, Meyer és Pifer (1970), Sinkey (1975) és Martin (1977) által alkalmazott korábbi módszerek: a lineáris és négyzetes diszkriminanciaelemzés, a faktoranalízis, valamint a logisztikus regresszió. E hagyományos módszerek elsősorban pénzügyi arányszámok és mutatók alapján mérték fel a potenciális kockázatokat. Azonban az egyre bonyolultabbá váló pénzügyi viszonyok megmutatták e modellek előrejelzési képességének határait. Egyértelművé vált, hogy kifinomultabb módszerekre van szükség.

A prediktív modellezés területén megjelenő, legújabb vívmányok között említhető a mesterséges neurális hálózatok, az operációkutatás, a hibrid intelligens modellek, a fuzzy logika és a szupportvektorgépek (SVM-ek) felhasználása. Quek, Zhou és Lee (2009) például újszerű fuzzy neurális hálózatot dolgozott ki a bankok csődjének előrejelzésére, bemutatva, hogy a számítógépes eljárásokat és a pénzügyi adatok rekonstruálását együttesen alkalmazva javítható az előrejelzések pontossága. Itt említhető még Jing és Fang (2018), akik a logitmodell és az adatbányászati módszerek eredményességét vetették össze az egyesült államokbeli bankok csődjének előrejelzésében, rámutatva, hogy bizonyos helyzetekben utóbbiak jóval hatékonyabbak.

A gépi tanulás és a mesterséges intelligencia megjelenésével jelentősen javult az előrejelző modellek teljesítménye. Ezekkel a korszerűbb módszerekkel terjedelmes adathalmazok is elemezhetők, felismerhetők összetett mintázatok és nagyobb pontossággal jelezhetők előre a potenciális bankcsődök, így árnyaltabb képet biztosítanak a bankok pénzügyi helyzetéről. Tung, Quek és Cheng (2004) neurális-fuzzy alapú korai bankcsőd-előrejelző rendszere, a GenSoFNN-CRI(S) hálózat például hatékonyan azonosította a pénzügyi nehézségek jeleit.

A 2008-as pénzügyi válság óta még nyilvánvalóbb a bankcsődök pontos előrejelzésének és megelőzésének fontossága. Az események jól szemléltették a globális pénzügyi rendszer összefonódottságát és egy-egy fontos pénzügyi intézmény összeomlásának kiterjedt világgazdasági hatását. A válság nyomán megszapordtak az előrejelző modellekkel és kockázatkezelési stratégiákkal kapcsolatos kutatások. Ez is mutatja, hogy a változó pénzügyi kockázatok lekötéséhez folyamatos módszertani megújulásra van szükség.

E módszertani fejleményekre építve tanulmányomban boosting és véletlen erdő algoritmusok felhasználásával újraelemzem és előre jelzem a 2008-as pénzügyi válságot a szövetségi bankfelügyelet alá tartozó amerikai bankok körében. Kutatásom tovább bővíti a bankcsőd-előrejelzés hatékony modelljeinek gazdag szakirodalmát.

Kezdeti feltételezéseink és céljaink is befolyásolhatják, hogy melyiket ítéljük a „legjobb” modellnek, ám az vitathatatlan, hogy a bankok csődbe jutásának pontos előrejelzése alapvető fontosságú a gazdasági stabilitás és növekedés fenntartásához.

A potenciális csődveszély korai felismerésével a bankok megelőző állami támogatáshoz juthatnak, így elkerülhető az összeomlásuk. Ennek a fordítottja is igaz: a közbelépés elmulasztásával az olyan bankok esetében, amelyeknél csődöt jeleznek előre a fent említett modellek, jelentősen megnő a pénzügyi intézmények stabilitását fenyegető kockázat, hiszen ha egy csődveszélyben lévő bank nem jut segítséghez, annak súlyos következményei lehetnek az egész pénzügyi rendszer stabilitására nézve.

2. A TANULMÁNY CÉLJA

A tanulmány súlypontja a vezető gépi tanulási modellek hatékonyságának értékelése a bankok csődjének előrejelzésében, nem a csőd mikroszintű okainak mélyelemzése. Ennek megfelelően elsősorban a fejlett elemzési eljárások és azok pénzügyi instabilitást előrejelző képessége felől közelítem meg a témát a csődöt kiváltó pontos tényezők taglalása helyett.

2.1. A 2008-as pénzügyi válság háttere

A világgazdaság történetének kiemelkedő jelentőségű eseménye, a 2008-as globális pénzügyi válság bonyolult pénzügyi mozgások és bankfelügyeleti mulasztások együtteséből alakulhatott ki. A válság az amerikai másodrendű jelzáloghitel-piaci helyzet 2007 nyarától kezdődő romlására vezethető vissza, amelyet a 2001–2002-es hitelválság nyomán fellépő trendek alapoztak meg. A hitelállomány soha nem látott mértékben bővült, a lakáscélú ingatlanok ára felszökött, és jelentősen meg-

nőtt a tőkeáttétel a pénzügyi rendszerben ezekben az években. Ezzel egy időben a gyors ütemben megjelenő pénzügyi és különösen az értékpapírosítási innovációknak köszönhetően jelentősen megerősödött a pénzügyi rendszer pénzteremtési képessége, ezzel azonban kockázatkezelési képessége nem tartott lépést (Mian-Sufi, 2009; Shiller, 2008).

Az USA-ban korábban bekövetkező hitelválságoktól eltérően az így kialakult globális pénzügyi válság világszinten érezhető következményekkel járt. Az amerikai másodrendű jelzáloghitel-piacon fellépő veszteségek gyorsan átgyűrűztek a nemzetközi pénzügyi ökoszisztémába. A bankokat sújtó, komoly veszteségek és likviditási problémák miatt a hiteleszköz-állomány értékelésével szemben széles körű bizonytalanság alakult ki, és drasztikusan visszaesett a bankközi hitelezés. A hitelezés újraindítása érdekében a központi bankok világszerte likviditástámogatási mechanizmusokkal és a csődveszélyben lévő bankok feltőkésítésével próbálták e hatásokat ellensúlyozni (Bernanke, 2010; Gorton, 2009).

A válság kiemelendő mozzanatai közé tartozik az egyesült királyságbeli Northern Rock állami tulajdonba kerülése 2008 februárjában a pénzintézetet érő betétkivonási rohamot követően – ami utoljára több mint egy évszázaddal azelőtt fordult elő az országban –, valamint az amerikai Bear Stearns befektetési bank felvásárlása a J.P. Morgan Chase által 2008 márciusában, pénzügyminisztériumi és központi banki irányítással (Sorkin, 2009). A válság megbénította az eszközfedezetű kereskedelmi értékpapírok és a repomegállapodások piacát is, ez pedig számos fedezeti alap csődjéhez vagy megszűnéséhez, illetve egy sor speciális célú befektetési alap (SIV) és eszközfedezetű kereskedelmi értékpapírokra szakosodott céltársaság (conduit) feloszlásához vezetett. Az így keletkező hitelezési veszteség világszinten több mint 1 ezermilliárd USD-ra rúgott (Financial Crisis Inquiry Commission, 2011).

A válság egyik fontos összetevője volt a hirtelen megugró lakáspiaci kereslet és jelzáloghitel-igénylés, ami részben a 2000-es évek elejét jellemző alacsony kamatkörnyezetnek volt köszönhető. A kereslet felhajtotta a lakáscélú ingatlanok árát, a másodrendű jelzáloghitelek magas hozama pedig vonzotta a befektetőket, köztük intézményi szereplőket is. E jelzáloghiteleket, amelyeket az első osztályú hitelfelvevőknek kínálnál jellemzően sokkal magasabb kamattal folyósítottak, egyre többen keresték értékpapírosítási céllal. Az értékpapírosítás során befektetésre nem ajánlott kategóriába tartozó eszközöket vontak össze, a cash flow-kat modellalapú becsléssel valószínűség alapján különítették el, majd a legbiztonságosabbnak ítélt cash flow-kból befektetésre ajánlott értékpapírokat alakítottak ki (Kothari, 2008).

A másodrendű jelzáloghitel-konstrukciókban sokszor a kezdeti alacsony „csalikamatokat” lényegesebben magasabb kamatok követték a későbbi években. Ezeket a hitelfelvevők – akik között lakossági ügyfeleket és spekulánsokat egyaránt

találunk – gyakran nem tudták fizetni, különösen, ha a csaliidőszak lejáratát után nem tudták újrafinanszírozni a hitelt. A keletkeztető és szétosztó (originate to distribute, OTD) bankmodell működéséből fakadóan a jelzáloghiteleken realizált veszteségeket a befektetők és nem a kibocsátó bankok viselték, ami a bankokat kevésbé motiválta arra, hogy kellő körültekintéssel járjanak el (Acharya–Richardson, 2009).

A késedelmes változó kamatozású másodrendű jelzáloghitelek megugrása tovább súlyosbította a válságot. 2007-re jelentősen megnőtt a 90 napon túli fizetési késedelmek aránya, ennek eredményeként számos másodrendűjelzálog-alapú értékpapírt leminősítettek. A kiváltó okok között említhető a kölcsönfelvevők rossz hitelképessége, a számos önerővel nem rendelkező elsőlakás-vásárló, valamint a csalikamatok széles körű alkalmazása (Foote–Gerardi–Willen, 2008).

Elterjedtek az olyan kockázatosabb jelzálogtermékek is, mint az anyagi fedezettel egyáltalán nem rendelkező hitelfelvevőknek nyújtott NINJA-hitelek és a jövedelem ellenőrzése nélkül folyósított ún. „liar” (hazug) hitelek, amelyek a laza hitelezési szabályok miatt teret engedtek a csalárd gyakorlatoknak. Tetézte a bajt a jelzáloghitel-brókerek díjazási rendszere, amely a hosszú távú törlesztés helyett a folyósított volumenre helyezte a hangsúlyt (Rajan, 2010).

A bankok a meglévő eszközeikből származó cash flow-kat változó hitelminősítésű ügyletrészsorozatokat formájában értékpapírosítva átadták a mérlegsemleges SIV-eknek. A módszer elméletben a kockázatok egyenletesebb elosztását szolgálta, de a gyakorlatban veszélyes sebezhető pontokat fedett fel a pénzügyi rendszer infrastruktúrájában (Ashcraft–Schuermann, 2008).

Erre válaszul a világ jegybankjai, és különösen az Egyesült Államokban a Federal Reserve innovatív likviditási intézkedések bevezetésével igyekeztek stabilizálni a pénzpiacokat. Az intézkedések keretében többek között hosszú távú hitelezési eszközöket hoztak létre, a kereskedelmi bankok számára is hozzáférhetővé tették a Fed likviditási hitelét (discount window), valamint támogatták a kereskedelmi értékpapírok vásárlását. A fontosabb amerikai kormányzati intervenciók között említhető a Term Auction Facility, a Primary Dealer Credit Facility, valamint a 2008. februári Economic Stimulus Act (Blinder–Zandi, 2010).

Összefoglalva, a modern kori gazdaságtörténet meghatározó eseményének számít a 2008-as globális pénzügyi válság kijózanítóan emlékeztetett a globális pénzügyi rendszeren belüli bonyolult kölcsönös függőségi viszonyokra és a rendszer sebezhetőségére. Az amerikai másodrendű jelzáloghitel-piac visszaeséséből kiinduló és a bonyolult pénzpiaci mozgások és szabályozói mulasztások együttállásának eredményeként elmélyülő válság széleskörű, nemzethatárokon átívelő gazdasági zűrzavart keltett. Nyilvánvalóvá tette a korszak pénzügyi innovációiban, köztük az értékpapírosításban, valamint a nagyarányú tőkeáttételben rejlő rend-

szerszintű kockázatokat, amelyek együttvéve meghaladták a pénzügyi intézmények és a szabályozók kockázatkezelési képességét.

A válság nemcsak a pénzügyi modellek és gyakorlatok hiányosságaira mutatott rá, hanem a hatékony kockázatkezelés és az éber szabályozói felügyelet alapvető fontosságára, valamint az etikus hitelnyújtási és -felvételi gyakorlatok szükségességére is. Jól szemléltette, hogy a hitelállomány gyors bővülése az alacsony kamatok fűtötte erőteljes lakáspiaci keresletfellendüléssel párosulva fenntarthatatlan eszközár-inflációhoz vezethet, amelyet piaci korrekciók követnek. A válság kiterjedt hatásaiból – a nagy bankok államosításától jelentős pénzügyi intézmények összeomlásáig – láthattuk, milyen súlyos következményekkel járnak e rendszer-szintű hibák.

Visszatekintve, a 2007–2009-es évek globális pénzügyi válsága tanulságos leckével szolgált a gazdasági és pénzügyi irányítás terén, felhívva figyelmünket a folyamatos felügyelet, az alkalmazkodóképesség, valamint a világ pénzügyi szervezetei közötti együttműködés szükségességére. Rámutat, hogy csak a múlt hibáiból tanulva tudjuk megvédeni a pénzügyi rendszert a jövőbeli válságoktól, hogy az eljövendő generációk ellenállóbb és stabilabb gazdasági környezetben élhessenek.

2.2. Szakirodalmi összefoglaló

Berger–Bouwman (2012): a tanulmány átfogóan elemzi, hogy az egyesült államokbeli bankok teljesítménye hogyan alakult a tőke függvényében a vizsgált 25 éves időtávlaton fellépő különböző gazdasági helyzetekben. Olyan korszerű statisztikai módszerekkel, mint a logisztikus regressziós túléléselemzés és a legkisebb négyzetek módszere, kimutatja, hogy a tőke minden gazdasági helyzetben jelentősen növeli a kis bankok túlélési esélyét és piaci részesedését. A tőke különösen a kis és közepes bankok számára létfontosságú a válságos időszakokban, azaz a banki stabilitás fenntartásában is kulcsszerepet játszik gazdasági visszaesés esetén.

Trussel–Johnson (2012): a tanulmány az amerikai bankok csődjéhez köthető pénzügyi mutatókat vizsgálja logisztikus regressziós módszerrel. A szerzők hat fő pénzügyi mutatószámot tárgyalnak, amelyekből összevont bankcsőd-előrejelző mérőszámot képeznek. Rendkívül hasznos megállapításokat tesznek a banki stabilitás előrejelzéséhez használható pénzügyi mérőszámokkal kapcsolatban, többek között, hogy az alapvető (Tier 1) tőke teljes eszközállományhoz viszonyított arányának növelése, valamint a nagyobb eszközarányos jövedelmezőség a két legfontosabb csőd-kockázat-csökkentő tényező.

Lu–Whidbee (2013): a tanulmány a 2000-es évek végén kialakult pénzügyi válságban közrejátszó tényezőket vizsgálja, köztük a bankok alapszabályát, holding-

társasági szerkezetét és a bankok sérülékenységet számszerűsítő különböző mérőszámokat. Az összes amerikai kereskedelmi bankra kiterjedő elemzés többek között a bankok megmentésére szánt pénzügyi forrásokat, valamint a banki tőkehányadokat és a likviditás mértékét azonosítja a bankok megmaradásában vagy csődjében döntő szerepet játszó tényezőként e viharos időszakban.

DeYoung–Torna (2013): a tanulmány a nem hagyományos banki tevékenységekből származó bevételek hatását elemzi az amerikai kereskedelmi bankok csődjére a pénzügyi válság idején. A szerzők több időszakos logitmodellel végzett vizsgálata alapján az értékpapír-forgalmazáshoz és a biztosítások értékesítéséhez hasonló, díj ellenében végzett nem hagyományos banki tevékenységek csökkentik a bankcsőd valószínűségét, míg az eszközalapú, így a kockázattitoke-szolgáltatások növelik azt. Eredményeikből többet tudhatunk meg arról, hogy a különböző üzleti modellek hogyan befolyásolják a bankok stabilitását.

Chiaromonte, Liu, Poli és Zhou (2016): a szerzők azt kutatják, hogy mennyire hatékonyan jelezhetők előre a Z-érték alapján a 2004 és 2012 között bekövetkezett bankcsődök. Eredményeik szerint a jövedelmezőséget, a tőkeáttételt és a jövedelem változékonyságát egyesítő Z-érték a bankcsődök 76%-át stabilan előrejelzi még hároméves időhorizonton is, azaz megbízható bankcsőd-előrejelző eszköznek tekinthető.

Cleary–Hebb (2016): a tanulmány diszkriminanciaelemzéssel vizsgálja 132 amerikai bank csődjét a 2002 és 2009 közötti időszakban. A szerzők magas, 92%-os előrejelzési pontossággal különítik el eredményesen a csődbe jutó és a csődöt elkerülő bankokat. Elemzésüket a 2010–2011-es időszakra kiterjesztve, hasonlóan nagy pontossággal jelzik előre a bankok csődjét.

Lu–Whidbee (2016): a válság időszakában 6236 amerikai kereskedelmi bankot vizsgáló tanulmány a központi beavatkozással megmenekülő vagy csődöt jelentő bankokra koncentrál. A szerzők számos hasonlóságot találnak a kimentett és a csődbe jutott bankok között, ami alapján azonos kockázati tényezőkön és sebezhető pontokon múlhatott, hogy átvészelték-e a válságot.

Serrano-Cinca, Fuertes-Callén, Gutiérrez-Nieto és Cuellar-Fernández (2014): az amerikai bankcsődöket 2009-től kezdve vizsgáló tanulmány számos elméletet mutat be a lehetséges okokról. A szerzők a strukturális egyenletek módszerét használva kimutatják, hogy a csődbe jutott bankokat nagyobb hitelállomány-növekedés, több ingatlanhitel, magasabb kockázati arány, nagyobb forgalom és alacsonyabb haszonkulcs jellemezte a csődöt elkerülő bankokhoz képest, ami alapján e tényezők szorosan összefüggnek a bankok csődjével.

Le–Viviani (2017): a szerzők tanulmányukban a hagyományos statisztikai és a modern gépi tanulási módszereket ötvöző átfogó bankcsőd-elemzést végeznek egy 3000 amerikai bankot tömörítő adathalmazon. Diszkriminanciaelemzést,

logisztikus regressziót, mesterséges neurális hálózatokat (ANN-ek), szupportvektorgépeket és a k-legközelebbi szomszéd (k-NN) módszert alkalmazva megállapítják, hogy a gépi tanulási módszerek, azon belül is az ANN-ek és a k-NN-módszer a hagyományos módszereknél hatékonyabban jelzik előre a bankcsődöt.

Gogas et al. (2018): a gépi tanulási modellekkel foglalkozó tanulmány alapja egy 1443 amerikai bankot számláló adathalmaz, amelyek közül 481 intézmény csődbe jutott 2007 és 2013 között. A szerzők a kétlépéses jellemzőkiválasztással és a szupportvektormodellel meggyőző, 99,22%-os előrejelzési pontosságot érnek el, ami ugyancsak azt bizonyítja, hogy a gépi tanulási eljárások lekörözhetik ezen a téren a hagyományos modelleket.

Carmona et al. (2019): a tanulmány az Extreme Gradient Boosting módszert alkalmazza a bankcsőd-előrejelzésre, 157 szövetségi bankfelügyelet alá tartozó amerikai kereskedelmi bankot vizsgálva 2001 és 2015 között. A szerzők 30 pénzügyi arányszámot értékelve megállapítják, hogy ezek némelyike, különösen az eredménytartalék és a kockázatalapú tőke, szoros kapcsolatban áll a fokozottabb bankcsőd-valószínűséggel.

Manthoulis et al. (2020): a statisztikai és gépi tanulási módszereket egyaránt felhasználó bankcsőd-előrejelző tanulmány 60 000 amerikai bankokra vonatkozó megfigyelésre alapul. Az eredmények alapján a diverzifikáló változók hatékonyan alkalmazhatók, valamint a sorrendiségen alapuló (ordinális) osztályozó modellek a bináris módszereknél jobban teljesítenek a bankcsőd-előrejelzésben.

Momparler et al. (2020): a szerzők 157 szövetségi bankfelügyelet alá tartozó amerikai kereskedelmi bankon végeznek fuzzy logikai és halmazelméleti alapú kvalitatív összehasonlító elemzést (fsQCA) a 2001 és 2015 közötti időszakban. Megállapítják, hogy nagyobb a csődveszély azoknál a bankoknál, amelyeket magas nemfizetési arány, illetve alacsonyabb kockázati fedezet és tőkésítettség jellemez, ami a jó eszközminőség és a megfelelő tőkeállomány jelentőségét igazolja.

2.3. A tanulmányban használt minta és arányszámok

A kutatáshoz az egyesült államokbeli Federal Deposit Insurance Corporation (FDIC) gondosan összeállított adathalmazát használtam, elsősorban a bankcsődökre vonatkozó bejegyzéseket tanulmányozva. Ez az adathalmaz volt bináris osztályozó elemzésem alapja, amellyel a bankokat két jól elkülöníthető csoportba kíséreltem meg besorolni: a csődöt jelentő és a tovább működő bankok csoportjába. A terjedelmes adathalmazban mintegy 950 különböző banki intézmény szerepel, anyabankok és leánybankok egyaránt. Ezek közül a 2008-as pénzügyi válság zűrzavaros időszakában mintegy 60 bank ment csődbe.

Az egészséges, azaz a csődöt elkerülő bankok azonosításához referenciaévként a pénzügytörténeti szempontból is meghatározónak tekinthető 2009-es évet választottam. Ebben az évben volt a legsúlyosabb a pénzügyi instabilitás, és a bankszektor is ebben a vízváltó évben szembesült a legnagyobb nehézségekkel. Választásomat arra az elgondolásra alapoztam, hogy ha az előrejelző modell képes a pénzügyi nyugtalanság tetőfokán pontosan elkülöníteni a csődbe ment és a csődöt elkerülő bankokat, akkor megbízhatónak és hatékonynak tekinthető.

Az elemzésbe 26 CAMEL (*capital adequacy* – tőke megfelelés, *asset quality* – eszközök minősége, *management quality* – vezetés képességei, *liquidity* – likviditás) arányszámot vontam be elsődleges változóként. Az előrejelző modellbe bevitt arányszámok összesen 24 700 adatpontot jelöltek ki. Mivel nyílt forrású, nyilvánosan hozzáférhető adatokat használtam, az adatgyűjtés nem ütközött anyagi korlátokba, így igen költséghatékony volt a kutatás. Emellett az adatok feldolgozásához és elemzéséhez is nyílt forráskódú szoftvert választottam: a széles körben alkalmazott és adatbányászati eredményeiről ismert WEKA programot. E döntések összhangban vannak azzal az alapelvvel, hogy a tudományos kutatást mindenki számára hozzáférhetővé és megismételhetővé kell tenni.

3. ALKALMAZOTT MÓDSZERTAN ÉS MODELLEK

3.1. XGBOOST

Az XGBoost vagy eXtreme Gradient Boosting elismert gépi tanulási eljárás, amely különösen alkalmas pénzügyi kockázatelemzési és bankcsőd-előrejelzési alkalmazásra. E továbbfejlesztett, lépcsőzetes építkezéssel „feljavított” döntési fák alapuló (gradient boosted decision tree) módszer előnye, hogy hatékonyan és rugalmasan kezelhetők vele nagy és összetett pénzügyi adathalmazok (Chen–Guestrin, 2016). A gyenge előrejelzési képességű modellek – lényegében döntési fák – kombinálásával megbízható előrejelző modellt létrehozó XGBoost-eljárás kimagaslóan teljesít a hitelkockázat-modellezésben és a bankcsőd-előrejelzésben. Ez a tulajdonsága kulcsfontosságú az egyéni hiteltörténettől az általánosabb gazdasági mutatókig számos különféle változót felvonultató pénzügyi elemzésekhez való alkalmazásához (Friedman, 2001). Az XGBoost egyik fő erőssége, hogy képes hiányzó értékekkel és különböző típusú adatokkal dolgozni, ami elengedhetetlen a gyakran hiányos pénzügyi adathalmazok esetében (Chen–He, 2015). Emellett az algoritmusba épített regularizációs eljárások biztosítják az összetett pénzügyi modellek esetében sokszor problémaként jelentkező túlillesztés elkerülését (Natekin–Knoll, 2013). Az XGBoost az egyes jellemzők fontosságához pontszámot rendel, ami különösen hasznos a hitelkockázat és a csődvalószínű-

ség legfontosabb tényezőinek azonosításában. A legjobb előrejelzési képességű pénzügyi arányszámok, irányítási mutatók és gazdasági tényezők megtalálásával a kockázatkezelési és szabályozási folyamatok kialakításában is segít (Huang–Chen–Wang, 2019). Empirikus kutatások támasztják alá, hogy pénzügyi területen az XGBoost jobb eredménnyel alkalmazható a hagyományos logisztikus regressziós és a véletlen erdő modelleknél. Pontosabban jelzi előre a hitelek nemteljesítését és a bankok nehéz helyzetbe kerülését, ezért a pénzügyi intézetek és a szabályozók által a leggyakrabban használt módszer (Malhotra–Malhotra, 2021).

3.2. Általánosított lineáris modell boosting

Az általánosított lineáris modellek (GLM-ek) elméletét a boosting algoritmusokkal egyesítő, általánosított lineáris modell boosting eljárás bizonyítottan hatékony a pénzügyi kockázatok értékelése és a bankcsőd-előrejelzés terén. Az eljárás jobb előrejelzési pontossága mellett megőrzi a hagyományos GLM-ek jó magyarázhatóságát, ami fontos szempont a pénzügyi modellezésben. A GLM-boosting kiválóan teljesít olyan kihívást jelentő, összetett pénzügyi feladatokban, mint a hitelkockázat-modellezés, a csőd-előrejelzés és a banki fizetéseképtelenség észlelése. Jól kezeli a különböző típusú adatokat, és bonyolult előrejelzőkimenet-kapcsolatokat, köztük nem lineáris viszonyok is modellezhetők vele, így pénzügyi felhasználásra is alkalmas (Hastie–Tibshirani–Friedman, 2009). Az eljárás iterációkkal illeszti a GLM-et az adatokhoz, amelyet a korábbi iterációk hibáinak korrigálásával tökéletesít. Az így létrejövő átfogó modellel észlelhetők a pénzügyi adatok olyan összetett mintázata is, amelyek felett a hagyományos lineáris modellek olykor elsiklanak (Bühlmann–Hothorn, 2007). A GLM-boosting iterációkat és az előrejelzők hatását szabályozó regularizációs funkciója felel a túlillesztés elkerüléséért, ami nélkül a modellek nem tudnak hatékonyan általánosítani az új adatokra (Schapire–Freund, 2012). Meg kell említeni továbbá, hogy a GLM-boostinggal változószelekció is végezhető. A nem fontos vagy felesleges változók kiiktatásával azonosítja a legnagyobb hatású előrejelzőket, így javítva a modell teljesítményét és magyarázhatóságát, ami alapvető fontosságú a pénzügyi döntéshozatalhoz (Tutz–Binder, 2006). Empirikus pénzügyi kutatások bizonyítják, hogy a GLM-boosting számos különböző célra, köztük a kockázati tényezők részletesebb megismerésére és a kedvezőtlen pénzügyi események pontos előrejelzésére is eredményesen felhasználható. Ezért kifejezetten jól alkalmazható a pontozásos hitelminősítésben és a bankcsőd-előrejelzésben, ahol a körültekintő döntéshozatalhoz elengedhetetlen az egyes kockázati tényezők szerepének pontos ismerete (Zou–Hastie, 2005).

3.3. LogitBoost

A logisztikus regressziót integráló algoritmusok, azon belül is ez a Friedman, Hastie és Tibshirani (2000) által kidolgozott megközelítés rendkívül eredményesnek bizonyult a gépi tanulás pénzügyi alkalmazásai, így például a hitelképesség-értékelés és az ügyfélszegmentálás terén. Kiválóan oldja meg a pénzügyi szektorban előforduló olyan gyakori bináris osztályozási feladatokat, mint a potenciális nemfizetők azonosítása vagy az ügyfelek különböző kockázati kategóriákba sorolása. A logisztikus regressziós modelleket iterációk útján tökéletesítő módszer átláthatóbbá teszi a pénzügyi adatok bonyolult összefüggéseit, ami nélkül lehetetlen a kockázatokat pontosan értékelni (Landwehr–Hall–Frank, 2005). E tulajdonság kiemelten fontos a pénzügyek világában, ahol a pontosság a költséges hibák elkerülésének kulcsa. Emellett a jellemzők előrejelzésekre gyakorolt hatásának jó átláthatósága biztosítja a módszer magas fokú magyarázhatóságát, ami pénzügyi környezetben szintén alapvető jelentőségű. A pénzügyi döntéseket, például a hitelkérelmek pozitív és negatív elbírálását megalapozó okok világos magyarázata a szabályozók és az érdekeltek számára egyaránt fontos (Hastie–Tibshirani–Friedman, 2009). A módszer a gyakorlatban igen eredményesnek bizonyult különböző pénzügyi felhasználási területeken, maga mögé utasítva más osztályozó eljárásokat például a csalások felderítésében és a hitelképesség értékelésében. Eredményességéhez nagyban hozzájárul, hogy összetett és nagy méretű adathalmazokkal is jól működik (Bühlmann–Hothorn, 2007).

3.4. Véletlen erdő

A pénzügyikockázat-modellezésben és a bankcsőd-előrejelzésben széles körben alkalmazzák Breiman (2001) együttes tanulási módszerek közé tartozó, véletlen erdő (random forest) algoritmusát. Az algoritmus több, az adatok és változók véletlen részhalmazára vonatkozó döntési fából képez egy erdőt. A módszerrel nagyobb előrejelzési pontosság érhető el, és kiküszöbölhető a túlillesztés veszélye, amely gyakori probléma a pénzügyi modellezésben. A véletlen erdő eljárás kiváló eredménnyel használható pénzügyikockázat-elemzésre, mivel a pénzügyi terület gyakorta nagy méretű, sokváltozós adathalmazait is képes kezelni. Hatékonyan dolgoz fel számos különféle adatfajtát, köztük a numerikus és kategorikus változókat, ezért optimális az összetett pénzügyi kockázatok elemzéséhez (Liaw–Wiener, 2002). A pénzügyi alkalmazások terén fontos előnyt jelent jellemzőkiválasztási képessége. A véletlen erdő számos fa átlagolásával képes megtalálni a nagyszámú pénzügyi mutató között a legnagyobb előrejelző értékkel rendelkezőket, ami pedig döntő fontosságú a pontozásos hitelminősítés és a csődelőrejelzés terén a jelentős kockázati tényezők beazonosításához (Díaz-Uriarte–De Andres, 2006).

A zsákolásos adatminta-vételezéssel (bagging) és véletlen jellemzőkiválasztással dolgozó véletlen erdő algoritmus csökkenti a varianciát és a torzítást, így pontos és az új adatokra jól általánosító modellnek számít. E tulajdonságok alapvetőek a bizonytalan környezetben kialakuló pénzügyi válságok és bankcsődök előrejelzéséhez (Cutler et al., 2007). A véletlen erdő módszer gyakorlati alkalmazása során eredményesebben jelezte előre a hitelek nemteljesítését és a bankok nehézségeit, mint a hagyományos módszerek és más gépi tanulási eljárások. A módszer megbízható teljesítményét és az összetett, nemlineáris pénzügyi viszonyok feltérképezésére való alkalmasságát már számos különböző gazdasági helyzetben igazolták (Hastie–Tibshirani–Friedman, 2009).

3.5. Regularizált véletlen erdő

A hagyományos véletlen erdő algoritmust regularizációs eljárásokkal kiegészítő, regularizált véletlen erdő (regularized random forest, RRF) módszer különösen hatékony a pénzügyikockázat-elemzést és a bankcsőd-előrejelzést is jellemző, többdimenziós adatkörnyezetben. A Deng (2013) által kidolgozott RRF-ben a véletlen erdő módszer megbízhatósága és jellemzőkiválasztási képessége a regularizálással egészül ki, ami a túlillesztés elkerülését és nagyobb pontosságot biztosít. A módszer nagyszámú változóval és akár összetett pénzügyi mutatókkal is jól működik, amelyek közül a leglényegesebbeket kiválasztva tökéletesíti a modellt (Meinshausen, 2007). A nagy jellemzőkészletekre jellemző problémát, a „dimenzionalitás átkát” a regularizálással kezeli, amely a faépítést vezérelve csökkenti a varianciát, és javítja a modell általánosítási képességét (Biau–Scornet, 2016). Az RRF eredményes pénzügyi alkalmazásának egyik fő oka jellemzőkiválasztási funkciója, amellyel képes elkülöníteni a legfontosabb változókat a kevésbé lényegesektől, ami pedig előfeltétele a pontos kockázatértékelésnek és döntéshozatalnak (Liaw–Wiener, 2002). A módszer megőrzi a véletlen erdő algoritmus minden erősségét, így képes különböző adatfajtákkal dolgozni, és számszerűsíti a változók fontosságát. Ezek igen hasznos tulajdonságok a gyakran hiányos pénzügyi adathalmazok kezeléséhez (Breiman, 2001). Gyakorlati alkalmazását tekintve az RRF igazoltan hatékony a pontozásos hitelminősítés terén, pontosabb előrejelzéseket biztosít a hitelek nemteljesítéséről és bankok nehézségeiről, ami hozzájárul a kockázatkezeléshez és a szabályozási követelmények betartatásához is (Strobl et al., 2009). Összességében az RRF, amely továbbfejlesztett jellemzőkiválasztási és regularizációs funkcióival hatékonyan dolgoz fel többdimenziós adatokat, jelentős fejlődést képvisel a pénzügyi kockázatok előrejelzése terén.

3.6. Költségérzékeny erdő

A véletlen erdő algoritmusához hasonlóan a költségérzékeny erdő is az együttes tanulási eljárások közé tartozik, annak egy kifinomult, kifejezetten a pénzügyi kockázatelemzés és bankcsőd-előrejelzés területére átdolgozott változata, ahol az előrejelzési hibáknak aszimmetrikus a költségvonzata. Az Elkan (2001) által kidolgozott algoritmus annyival bővül a hagyományos modellekhez képest, hogy tanítási folyamatában már figyelembe veszi az álpozitív és az álnegatív eredmények eltérő költségeit is, így jobban leképezi a valós pénzügyi döntéshozatali folyamatot (Ling–Sheng, 2008).

Pénzügyi területen különösen hasznos a módszer, hiszen egy bankcsőd azonosításának az elmulasztása (álnegatív eredmény) jelentősen nagyobb költséggel jár egy tévesen előrejelzett bankcsődnél (álpozitív eredmény). A költségesebb hibák kiküszöbölésére kihegyezett költségérzékeny erdő kifinomult szabályozói és gazdasági elemzőeszköz (Turney, 2000).

Egyik fő előnye, hogy a változó pénzpiacok folyamatosan átalakuló költségvonzataihoz alakítható, így kalibrálhatóságának köszönhetően eltérő gazdasági viszonyok között is jól alkalmazható (Zadrozny–Elkan, 2001). Empirikus kutatások igazolták, hogy a költségérzékeny erdő a hagyományos modelleknél pontosabb előrejelzéseket ad, így hozzájárul a pénzügyi intézmények és a szabályozók hatékony kockázatkezelési stratégiájának kialakításához (Khalilia–Chakraborty–Popescu, 2011).

3.7. Gépi tanulás

Gépi tanulás alatt egy sor olyan számítástechnikai módszert értünk, amelyekkel egy rendszer adatok alapján tanul, döntéseket hoz és előrejelzéseket készít. A gépi tanulás lényegében olyan algoritmusokat takar, amelyek konkrét feladatokra való célzott beprogramozásuk nélkül is képesek nagy méretű adathalmazokat feldolgozni, mintázatokat felismerni és a bevitt adatok alapján előrejelzéseket vagy döntéseket alkotni (Alpaydin, 2020). A gépi tanulási módszereket a felügyelt tanulás, a nem felügyelt tanulás és a megerősítéses tanulás fő kategóriáiba sorolhatjuk, amelyek mindegyike eltérő típusú problémák és adatszerkezetek kezelésére alkalmas (Goodfellow–Bengio–Courville, 2016).

3.8. Gépi tanulási teljesítménymutatók

Pontosság

A legegyszerűbb és legkézenfekvőbb teljesítménymutató a pontosság, a helyesen osztályba sorolt adatok és az adathalmaz összes adatának hányadosa. Bár könnyen értelmezhető, félrevezető is lehet, különösen olyan kiegyensúlyozatlan adathalmazok esetében, amelyek egyik osztálya aránytalanul nagyszámú adatot tartalmaz (Provost et al., 1998).

Precizitás (pozitív prediktív érték)

A precizitás azt adja meg, hogy a pozitívnak címkézett adatok közül hány volt valóban az. Ez különösen olyan helyzetekben fontos, ahol az álpozitív osztályozás súlyosabb következményekkel jár. A magas precíziós érték tehát az álpozitív eredmények alacsony arányát jelzi (Powers, 2011).

Felidézés (szenzitivitás, valódi pozitív predikciók aránya)

A felidézés mérőszáma azt mutatja, hogy a valóban pozitív adatok közül hányat talált meg a modell. Különösen akkor fontos, amikor a figyelmen kívül hagyott pozitív adatoknak (azaz az álnegatív eredményeknek) nagy a költségvonzata. A magas felidézési vagy szenzitivitási érték azt jelzi, hogy a pozitív adatok nagy részét megtalálta a modell (Powers, 2011).

Precizitás-szenzitivitás görbe

A precizitás-szenzitivitás görbe (PRC) a precizitást a szenzitivitás (felidézés) függvényében ábrázolja. A ROC-görbéhez hasonlóan a PRC-görbe a modell hatékonyságának teljes körű, átfogó értékelésére szolgál. A PRC-görbéhez tartozó AUC (görbe alatti terület) pontszám, amelyet AUC (PRC) rövidítéssel jelölünk, szintén hasznos mérőszám, amennyiben több osztályozót szeretnénk összehasonlítani. Saito és Rehmsmeier (2015) tanulmánya alapján a PRC-görbe kiegyensúlyozatlan adathalmazok esetében megbízhatóbban tájékoztat a ROC-görbénél a bináris osztályozók teljesítményéről. Ezért kutatásomban én is a PRC-görbét használtam.

Tévesztési mátrix

A tévesztési mátrix az osztályozási modell teljesítményét bemutató táblázat. A valódi pozitív, álpozitív, valódi negatív és álpozitív eredmények láthatók benne. Átlátható képet ad a modell teljesítményéről, amely különösen olyan osztályozási feladatok esetében hasznos, amikor több célosztályunk van (Stehman, 1997).

K-szoros keresztellenőrzés

Kutatásomban a Carmona et al. (2019) által felvázolt módszertant követtem, ezen belül az általuk is alkalmazott 10-szeres keresztellenőrzést végeztem el. Ez bevett statisztikai elemzési eljárás, különösen a gépi tanulási és adattudományos modellek validálása terén. A 10-szeres keresztellenőrzéshez először a teljes adathalmazt tíz egyforma nagyságú részre vagy „rétegre” osztjuk. A szegmentálás biztosítja, hogy az adathalmaz mindegyik része egyenlően érvényesüljön a validálási folyamatban, így csökken a torzítás, és megbízhatóbban értékelhető a modell. A folyamat tíz iterációja során az egyik réteg a teszhalmaz, a többi kilenc pedig a tanítóhalmaz szerepét tölti be. Ezt a forgatást módszeresen végezzük, azaz mindegyik réteg pontosan egyszer kerül sorra teszhalmazként, így a folyamat különböző szakaszaiban mindegyik adatpontot felhasználjuk tesztelésre és tanításra is. A módszer legfőbb előnye a módszeres forgatás, amellyel megbízhatóbban és alaposabban értékelhető a modell teljesítménye. Ezzel az eljárással a modellt számos különféle adatkörnyezetben tesztelhetjük. Mivel az iterációk során változik a tanító- és a teszhalmaz, a modell az adatok különböző részalmazain mutathatja meg tanulási és általánosítási képességét. Ez különösen hasznos, ha a modell teljesítményét valós élethelyzetekben értékeljük, amelyeket igen változó adatkészlet jellemezhet.

3.8.1. A boostingalgoritmusok eredményei

GLMBoost

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	915	96,32%	
Tévesen osztályozott adatok	35	3,68%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,97	0,991	0,989
Csődbe jutó	0,805	0,55	0,777
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	882	8	
Ténylegesen csődbe jutó	27	33	

XGBoost

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	918	96,63%	
Tévesen osztályozott adatok	32	3,37%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,981	0,983	0,990
Csődbe jutó	0,741	0,717	0,703
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	878	12	
Ténylegesen csődbe jutó	23	37	

LogitBoost

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	915	96,32%	
Tévesen osztályozott adatok	35	3,68%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,974	0,987	0,998
Csődbe jutó	0,755	0,617	0,782
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	875	15	
Ténylegesen csődbe jutó	17	43	

3.8.2. Véletlen erdő algoritmusok eredményei**Általános véletlen erdő**

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	925	97,37%	
Tévesen osztályozott adatok	25	2,63%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,982	0,99	0,989
Csődbe jutó	0,83	0,733	0,824
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	881	9	
Ténylegesen csődbe jutó	16	44	

Költségérzékeny véletlen erdő

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	925	97,37%	
Tévesen osztályozott adatok	25	2,63%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,984	0,988	0,999
Csődbe jutó	0,807	0,767	0,760
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	879	11	
Ténylegesen csődbe jutó	14	46	

Regularizált véletlen erdő

Mérszám	Érték	Százalék	
Helyesen osztályozott adatok	925	97,47%	
Tévesen osztályozott adatok	25	2,53%	
Osztály	Pontosság	Szenzitivitás	PRC-terület
Csődöt elkerülő	0,988	0,985	0,997
Csődbe jutó	0,79	0,817	0,755
Tévesztési mátrix	Csődbe jutónak ítélt	Csődöt elkerülőnek ítélt	
Csődöt ténylegesen elkerülő	877	13	
Ténylegesen csődbe jutó	11	49	

3.8.3. Az eredmények összehasonlítása

Gépi tanulási algoritmusok teljesítményének értékelése a csődöt elkerülő bankokra vonatkozó predikciók alapján

Algoritmus	Mérőszám	Érték	Százalék	Osztály	Pontosság	PRC-terület
Regularizált véletlen erdő	Helyesen osztályozott adatok	925	97%	Csődöt elkerülő	0,988	0,997
Költségérzékeny véletlen erdő	Helyesen osztályozott adatok	925	97%	Csődöt elkerülő	0,984	0,999
Általános véletlen erdő	Helyesen osztályozott adatok	925	97%	Csődöt elkerülő	0,982	0,999
XGBoost	Helyesen osztályozott adatok	918	97%	Csődöt elkerülő	0,981	0,999
LogitBoost	Helyesen osztályozott adatok	915	96%	Csődöt elkerülő	0,974	0,998
GLMBoost	Helyesen osztályozott adatok	915	96%	Csődöt elkerülő	0,97	0,989

Ha az algoritmusok teljesítményét a csődöt elkerülő bankokra vonatkozó predikcióik alapján részletesen elemezzük, az eredmények alapján a véletlen erdő változatai némiképp jobban szerepelnek a boostingeljárásoknál. Ezen belül a regularizált véletlen erdő, a költségérzékeny véletlen erdő és az általános véletlen erdő algoritmusok mindegyike mintegy 97%-os találati aránnyal osztályozta az adatokat, ami alapján megbízhatóan különböztetik meg a csődközeli és a stabil helyzetű bankokat. A boostingmódszerek, azaz a GLMBoost és a LogitBoost kissé elmaradnak mögöttük, de ezek az eljárások is elismerésre méltó, 96%-os pontossággal teljesítettek.

Ha a precizitást nézzük, amely biztosítja, hogy a stabilnak ítélt bankok valóban azok is legyenek, a legjobb eredményt a regularizált véletlen erdő nyújtotta, 0,988 precizitással osztályozva a csődöt elkerülő bankokat. Ez a magas precizitási érték az álpozitív eredmények kisebb előfordulási arányát jelzi, ami elengedhetetlen a csődveszélyben lévő bankok stabilként való téves azonosításának elkerüléséhez. A költségérzékeny és az általános véletlen erdő algoritmusok szintén jól teljesítenek, így elmondhatjuk, hogy a véletlen erdő módszertani keret egységesen

megbízhatóan azonosítja a csődöt elkerülő bankokat. A boostingeljárások és különösen a GLMBoost precizitása valamivel alacsonyabb, ezért a csődöt elkerülő bankok azonosításakor nagyobb arányban adhat álpozitív eredményt.

A precizitás-szenzitivitás (PRC) görbe alatti terület átfogó mérőszám, amely mind a modell precizitásáról, mind felidézési arányáról avagy szenzitivitásáról tájékoztat. A nagyobb PRC-terület azt jelzi, hogy a modell hatékonyan képes nagy precizitás mellett jó aránnyal azonosítani a valódi pozitív adatokat (azaz esetünkben a csődöt elkerülő bankokat). A véletlen erdő algoritmus változatai büszkélkedhetnek a legnagyobb PRC-területtel, ami alapján ezekkel a módszerekkel érhető el a legnagyobb egyensúly a csődöt elkerülő bankok helyes előrejelzése és a lehető legkevesebb álpozitív, illetve álnegatív predikció között. A boostingmódszerek esetében valamivel kisebb a PRC-terület, de ezek az eljárások is megbízható eredményt nyújtanak.

Összegezve: a véletlen erdő változatai felülmúlják a boostingmódszereket a csődöt elkerülő bankok osztályozásában, nagyobb precizitást, valamint jobb precizitás és szenzitivitás közötti egyensúlyt biztosítva. A regularizált véletlen erdő kimagasló pontossága jól ismert, míg az általános és a költségérzékeny változatok képviselik a legjobb kompromisszumot az álpozitív és az álnegatív előrejelzések kiküszöbölése terén. Bár a boostingmódszerek némiképp lemaradnak mögöttük, ezek is jó eredménnyel alkalmazhatók, különösen a paraméterek hangolásával, a GLMBoost pedig még jobb választásnak is bizonyulhat magyarázhatósága miatt. A modellválasztásnak mindig az adott alkalmazást jellemző igényekhez és kockázati profilhoz kell igazodnia.

Gépi tanulási algoritmusok teljesítményének értékelése a csődbe jutott bankokra vonatkozó predikciók alapján

Algoritmus	Mérőszám	Érték	Százalék	Osztály	Szenzitivitás	PRC-terület
Regularizált véletlen erdő	Tévesen osztályozott adatok	25	3%	Csődbe jutó	0,817	0,755
Költségérzékeny véletlen erdő	Tévesen osztályozott adatok	25	3%	Csődbe jutó	0,767	0,76
Általános véletlen erdő	Tévesen osztályozott adatok	25	3%	Csődbe jutó	0,733	0,824
XGBoost	Tévesen osztályozott adatok	32	3%	Csődbe jutó	0,717	0,703
LogitBoost	Tévesen osztályozott adatok	35	4%	Csődbe jutó	0,617	0,782
GLMBoost	Tévesen osztályozott adatok	35	4%	Csődbe jutó	0,55	0,777

Az algoritmusok bankcsőd-előrejelzési képességének részletes elemzése alapján mindegyik modell nagy pontossággal dolgozik, az adatok mindössze 3–4%-át osztályozzák tévesen. Az algoritmusok egyenletes teljesítménye igazolja, hogy képesek hatékonyan azonosítani a bankcsőd potenciális eseteit.

Ha szenzitivitásukat vizsgáljuk, amely a csődbe jutott bankok helyes azonosításában játszik központi szerepet, a regularizált véletlen erdő jár az élen, kiemelkedő, 0,817-es szenzitivitási értékkel, azaz ezzel a módszerrel azonosíthatók a leghatékonyabban a bankcsőd tényleges esetei. Ettől alig maradnak el a költségérzékeny véletlen erdő és az általános véletlen erdő 0,767-es, illetve 0,733-as szenzitivitási értékei. Az XGBoost, a LogitBoost és a GLMBoost esetében a felidézési arány alacsonyabb, tehát nagyobb valószínűséggel ítélik a csődveszélyben lévő bankokat tévesen stabil helyzetűnek.

További információkat kapunk a precizitás és a szenzitivitás egyensúlyát megadó fontos mérőszám, a PRC-görbe alatti terület alapján. A véletlen erdő algoritmus esetében mért legmagasabb, 0,824-es vonatkozó érték azt jelzi, hogy az optimális előrejelzési egyensúly ennél az algoritmusnál érvényesül. A regularizált és a költségérzékeny véletlen erdő modellek 0,755-ös, illetve 0,76-os értékükkel szintén jól teljesítenek a PRC-terület tekintetében, azaz az egyensúly ezekkel az eljárások-

kal is hatékonyan megőrizhető. A PRC-területre mért érték hasonlóan alakult az XGBoost és a GLMBoost esetében, a LogitBoost esetében pedig a kettő között helyezkedik el.

Az elemzésből összességében az derül ki, hogy a véletlen erdő módszer változatai különösen hatékonyak a bankcsődök pontos felismerésében, közülük pedig az általános véletlen erdőnél a legideálisabb a precizitás és a szenzitivitás egyensúlya. A boostingeljárások a felidézés tekintetében valamivel gyengébben, de így is megbízhatóan teljesítenek, azaz az adott alkalmazás követelményeitől és kontextusától függően ezek is jó választást képviselhetnek.

Összességében tehát a regularizált véletlen erdő kiemelkedően szenzitív a csődveszélyben lévő bankokra, azaz kisebb valószínűséggel siklik el felettük. Magas felidézési arányának köszönhetően különösen jól alkalmazható olyan helyzetekben, amikor a csődközeli bankok szem előtt tévesztésének jelentős a költségvonzata. Az általános véletlen erdő algoritmusnál figyelhető meg a precizitás és a szenzitivitás közötti optimális egyensúly, azaz a csődveszélyben lévő bankokat a lehető legkevesebb téves riasztás mellett képes azonosítani, amit az algoritmusnál mért legnagyobb PRC-terület is tükröz. Az alacsonyabb szenzitivitású és PRC-területű GLMBoost- és LogitBoost-módszerek szintén alkalmasak lehetnek bizonyos helyzetekben, különösen finomhangolással. A legmegfelelőbb modell kiválasztásához figyelembe kell venni az álnegatív és álpozitív eredmények következményeit, és az olyan nagy kockázatú környezetekben, mint a bankszektor, ahol a bankcsődök pontos észlelése különösen fontos, előnyben részesítendő a regularizált és az általános véletlen erdőhöz hasonló modellek.

4. ÖSSZEGZÉS

A gazdasági rendszerekben közvetítői szerepet betöltő bankok központi szerepet játszanak a pénzügyi stabilitás fenntartásában, ezért a bankcsőd kutatás kiemelt téma a pénzügyi szakirodalomban. A bankcsőd okainak megértése és előrejelzése alapvető nemcsak a szabályozási és felügyeleti lehetőségek javításához, de a csődök káros gazdasági hatásának megfékezéséhez is. Ezek az ismeretek hozzájárulhatnak a jövőbeli összeomlásokat hatékonyan megelőző stratégiák és szakpolitikák kidolgozásához, ezáltal pedig a bankrendszer és általában a gazdaság stabilitásának megóvásához.

A múltban jelentős fejlődésen ment keresztül a bankcsődök okainak tudományos kutatása, amelynek során a statisztikai módszerektől eljutottunk a kifinomultabb gépi tanulási és mesterséges intelligenciára alapuló megközelítésekig. A korábbi kutatások elsősorban diszkriminanciaelemzésre és logisztikus regressziós módszerekre alapultak, de a pénzpiacok komplexitásának fokozódásával ezek az el-

járások már nem tudtak kellően pontos előrejelzéseket adni. Alapvető változást hozott az olyan fejlettebb számítástechnikai eljárások előtérbe kerülése, mint a mesterséges neurális hálózatok és a szupportvektorgépek. E módszerek lehetővé tették az összetett pénzügyi adatok árnyaltabb elemzését, így javítva a potenciális bankcsődök felismerésére alkalmazott modellek előrejelzési képességét.

Átfogó elemzésem, amelyben a különféle gépi tanulási algoritmusok teljesítményét hasonlítottam össze a csődöt elkerülő és csődbe jutó bankokra vonatkozó előrejelzéseik alapján, különböző erősségeket tárt fel az egyes modellek esetében. A csődöt elkerülő bankokra vonatkozó predikciók szempontjából a véletlen erdő modell változatai, pontosabban a regularizált, a költségérzékeny és az általános véletlen erdő módszerek valamivel hatékonyabbnak bizonyultak a boostingeljárásoknál. Esetükben jobb volt a helyesen osztályozott adatok aránya, a precizitás, valamint a precizitás és szenzitivitás közötti egyensúly. Különösen jól szűrték az álpozitív eredményeket, azaz az említett algoritmusok által stabilnak ítélt bankok valóban azok is voltak.

Ugyanakkor a csődveszélyben lévő bankokat valamennyi előrejelzési modell nagy pontossággal azonosította, és csak az adatok alacsony százalékét osztályozta tévesen. A regularizált véletlen erdő algoritmus esetében azonban kiemelkedően magas szenzitivitási értéket mértünk, ami alapján ez a módszer különösen érzékeny a csődveszélyben lévő bankokra, és kisebb valószínűséggel téveszti szem elől azokat. Az általános véletlen erdő biztosítja a legjobb egyensúlyt a precizitás és a szenzitivitás között, amely az algoritmusnál mért legnagyobb PRC-területből is látható. Bár a boostingeljárások közé tartozó GLMBoost és LogitBoost esetében kapott szenzitivitási érték és PRC-terület valamivel kisebb, ettől függetlenül jó választást jelenthetnek, főleg olyan alkalmazásoknál, amelyek esetében különösen fontos szempont például a modell jó magyarázhatósága.

A hagyományos modellektől a fejlett gépi tanulási megközelítésekig bejárt út leköveti a pénzpiacok fokozódó bonyolultságát és a kifinomultabb elemzőeszközökre mutató igényt. A 2008-as pénzügyi válság, amely megmutatta a globális pénzügyi rendszer összefonódottságát és a pénzügyi intézmények csődjének messzemenő hatásait, szintén azt igazolta, hogy a bankcsődök pontos előrejelzése és megelőzése kulcsfontosságú.

Végső soron mindig az adott alkalmazás igényeitől és kockázati profiljától függ, hogy melyik a legmegfelelőbb modell a banki stabilitás előrejelzéséhez. A választásnál figyelembe veendő szempontok többek között az álnegatív és álpozitív ítéletek költségvonzata, a modell magyarázhatóságának elvárt foka és a felhasznált banki adatok jellegzetességei. A banki szférában elsődleges szempont a pontosság és a megbízhatóság, ezért itt alkalmasabbak lehetnek az olyan modellek, amelyeknek a csődhelyzetek iránti szenzitivitása és precizitása között optimális az egyensúly. Ilyen a regularizált és az általános véletlen erdő. Ugyanakkor előnyö-

sebb lehet valamelyik boostingmódszer, ha az ezeket jellemző erősségek jobban kiszolgálják az adott feladat operatív követelményeit. A módszertan folyamatos megújulása elengedhetetlen ahhoz, hogy lépést tudjunk tartani a változó pénzügyi kockázatokkal és napjaink összetett pénzügyi rendszereivel.

4.1. A kutatás korlátai

Eredendően egyfajta korlátot jelent, ha kizárólag CAMEL-arányszámokat használunk a bankcsőd-előrejelzésben és a banki kockázatcsökkentésben, különösen, ha figyelembe vesszük a pénzügyi intézmények bonyolultságát és dinamikus működési környezetüket. E korlát főként abban rejlik, hogy a CAMEL-arányszámok kvantitatív mérték révén nem feltétlenül tudják megragadni és leképezni a pénzügyi viselkedés és a vezetői döntéshozatal finomságait, pedig ezek jelentősen befolyásolják a bankok kockázati profilját és csődnek való kitettséget.

Alapvetően a CAMEL-arányszámok a szabályozó hatóságok és pénzügyi elemzők számára hasznos eszközök a bankok pénzügyi helyzetének és stabilitásának felmérésére, viszont a pénzügyi viselkedéssel és vezetői döntéshozattal összefüggésbe hozható kockázatokról nem tudnak teljes körű képet adni. Kutatásom korlátja ezért éppen az, hogy kizárólag ilyen arányszámokra alapul, és a banki működés és az emberi viselkedés kvalitatív vetületeit nem veszi figyelembe. Ennek eredményeként hiányos kockázatértékelést kaphatunk, és szem elől tévesztjük a nem számszerűsíthető tényezőkből fakadó kockázatokat.

E korlát kezelése érdekében jövőbeli kutatási irányként feltérképezhető, hogy miként lehetne a viselkedési pénzügyi mérőszámokat, valamint a bankirányítási gyakorlatok és a vezetés minőségének kvalitatív értékelését bevonni a kutatás jelenlegi kereteibe. Ez olyan új modellek kidolgozására is kiterjedhet, amelyekben a pénzügyi helyzetet értékelő kvantitatív mutatók mellett a vezetői döntéshozatal, az etikai szempontok és a vállalatirányítási struktúrák kvalitatív értékelése is helyet kapna. Ezeknek az összetett és sokrétű tényezőknek a bevonásával a szakterület kutatói és gyakorlati művelői átfogóbb szemléletű és hatékonyabb stratégiákat dolgozhatnak ki a bankcsődök előrejelzésére és a kockázatcsökkentő intézkedések végrehajtására.

A. FÜGGELÉK

A tanulmányban használt pénzügyi arányszámok (CAMEL-mutatók)

1. Jövedelemtermelő eszközök hozama
 2. Jövedelemtermelő eszközök kiadásai
 3. Nettó kamatmarzs
 4. Átlagos eszközállományhoz viszonyított nem kamatjellegű jövedelem
 5. Átlagos eszközállományhoz viszonyított nem kamatjellegű ráfordítások
 6. Hitelvesztésedet fedező eszközarányos céltartalék
 7. Eszközarányos nettó működési bevétel
 8. Eszközarányos nyereség
 9. Átlagos tőkéhez viszonyított eredménytartalék (csak az adott év eleje óta)
 10. Hitelállomány-arányos nettó leírások
 11. Jövedelemhez viszonyított nettó leírások (x)
 12. Költséghatékonysági mutató
 13. Egy alkalmazottra jutó összes eszköz (millió \$)
 20. Eszközarányos jövedelemtermelő eszközállomány
 15. Hitelek értékvesztése a hitelállományhoz képest
 16. Hitelek értékvesztése a hosszú lejáratú hitelállományhoz képest
 17. Hosszú lejáratú hitelek aránya
 18. Eszközarányos nettó hitelállomány
 19. Betétállományhoz viszonyított nettó hitelállomány
 20. Eszközarányos belföldi betétállomány
 21. Saját tőke aránya
 22. Kockázatalapú tőke aránya
 23. Átlagos eszközállomány
 24. Átlagos jövedelemtermelő eszközállomány
 25. Átlagos tőke
 26. Átlagos hitelállomány
-

HIVATKOZÁSOK

- Alpaydin, E. (2020): *Introduction to Machine Learning* (4th ed.). Cambridge, MA: MIT Press. ISBN: 9780262043793.
- Ashcraft, A. – Schuermann, T. (2008): Understanding the Securitization of Subprime Mortgage Credit. *Foundations and Trends in Finance*, 2(3), 191–309.
- Berger, A. N. – Bouwman, C. H. S. (2012): How Does Capital Affect Bank Performance During Financial Crises? *Journal of Financial Economics*, 109(1), 146–176.
- Bernanke, B. S. (2010): *The Courage to Act: A Memoir of a Crisis and Its Aftermath*. W. W. Norton & Company.
- Biau, G. – Scornet, E. (2016): A random forest guided tour. *TEST* 25, 197–227. <https://doi.org/10.1007/s11749-016-0481-7>.
- Blinder, A. S. – Zandi, M. (2010): The Financial Crisis: Lessons for the Next One. *Center on Budget and Policy Priorities*.
- Boycioglu, M. A. – Kara, Y. – Baykan, Ö. K. (2009): Predicting Bank Financial Failures Using Neural Networks, Support Vector Machines and Multivariate Statistical Methods: A Comparative Analysis in the Sample of Savings Deposit Insurance Fund (SDIF) Transferred Banks in Turkey. *Expert Systems with Applications*, 36(2), 3355–3366.
- Breiman, L. (2001): Random forests. *Machine Learning*, 45(1), 5–32.
- Bühlmann, P. – Hothorn, T. (2007): Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4), 477–505.
- Carmona, P. – Diranzo, F. J. C. – Momparler, A. (2019): Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61(C), 304–323.
- Chen, T. – Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Chiaromonte, L. – Liu, H. – Poli, F. – Zhou, M. (2016): How accurately can z-score predict bank failure? *Financial Markets, Institutions & Instruments*, 25(5), 333–360.
- Cleary, S. – Hebb, G. (2016): An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance*, 64, 101–111.
- Cutler, A. – Cutler, D. R. – Stevens, J. R. (2011): Random forests. In Zhang, Cha – Ma, Yunqian (eds.) (2011): *Ensemble Machine Learning*, 157–175, New York: Springer.
- Davis, Jesse – Goadrich, Mark (2006): The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning (ICML '06): Association for Computing Machinery, New York, NY, USA*, 233–240. <https://doi.org/10.1145/1143844.1143874>.
- Deng, H. (2013): Guided random forest in the RRF package. arXiv preprint arXiv:1306.0237.
- DeYoung, R. – Torna, G. (2013): Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, 22(3), 397–421.
- Díaz-Uriarte, R. – De Andres, S. A. (2006): Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7, 3.
- Elkan, C. (2001): The Foundations of Cost-Sensitive Learning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*.
- Erdal, H. I. – Ekinci, C. (2013): A New Approach to Turkish Banking Sector's Failure Prediction via Artificial Neural Network and Support Vector Machine. *Expert Systems with Applications*, 40(3), 887–897.

- Financial Crisis Inquiry Commission (2011): The Financial Crisis Inquiry Report. U.S. Government Printing Office.
- Foote, C. L. – Gerardi, K. S. – Willen, P. S. (2008): Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't. Federal Reserve Bank of Boston, Public Policy Discussion Papers.
- Freund, Y. – Schapire, R. E. (1996): Experiments with a New Boosting Algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning (ICML).
- Friedman, J. (2001): Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189–1232.
- Friedman, J. – Hastie, T. – Tibshirani, R. (2000): Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors): *Annals of Statistics*, 28(2), 337–407.
- Gogas, P. – Papadimitriou, T. – Agrapetidou, A. (2018): Forecasting Bank Failures: A Machine Learning Approach. *Annals of Economics and Finance*, 19(2), 421–455.
- Goodfellow, I. – Bengio, Y. – Courville, A. (2016): *Deep Learning*. Cambridge, MA: MIT Press.
- Gorton, G. (2009): *Slapped by the Invisible Hand: The Panic of 2007*. Prepared for the Federal Reserve Bank of Atlanta's 2009 Financial Markets Conference: Financial Innovation and Crisis, May 11–13.
- Hastie, T. – Tibshirani, R. – Friedman, J. (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer Series in Statistics.
- Huang, B. – Chen, J. – Wang, C. (2019): Financial crisis prediction model using ant colony optimization for XGBoost. *Expert Systems with Applications*, 122, 32–41.
- Khalilia, M. – Chakraborty, S. – Popescu, M. (2011): Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, 11, 51.
- Kothari, V. (2008): *Securitization: The Financial Instrument of the Future*. Hoboken, NJ: Wiley.
- Landwehr, N. – Hall, M. – Frank, E. (2005): Logistic Model Trees. *Machine Learning*, 59(1–2), 161–205.
- Le, L. – Viviani, J. L. (2017): An Empirical Analysis of Bank Failure Prediction in the United States. *Journal of Financial Research*, 40(2), 257–282.
- Liaw, A. – Wiener, M. (2002): Classification and regression by RandomForest. *R News*, 2(3), 18–22.
- Ling, C. X. – Sheng, V. S. (2008): Cost-Sensitive Learning and the Class Imbalance Problem. In Sammut, C. (eds.) (2010): *Encyclopedia of Machine Learning*. New York: Springer.
- Lu, W. – Whidbee, D. A. (2013): Bank structure and failure during the financial crisis. *Journal of Financial Economic Policy*, 5(3), 281–299.
- Lu, W. – Whidbee, D. A. (2016): US bank failure and bailout during the financial crisis Examining the determinants of regulatory intervention decisions. *Journal of Financial Economic Policy*, 8(3), 316–347.
- Malhotra, R. – Malhotra, D. K. (2021): Evaluating consumer loans using machine learning algorithms. *Journal of Risk and Financial Management*, 14(3), 105.
- Manthoulis, G. – Mylonakis, J. – Diacogiannis, G. (2020): Predicting Bank Failures: A Statistical and Machine Learning Approach. *Journal of Financial Risk Management*, 9(4), 305–323.
- Martin, D. (1977): Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking & Finance*, 1(3), 249–276.
- Meinshausen, Nicolai (2007): Relaxed Lasso. *Computational Statistics & Data Analysis*, 52(1), September, 374–393.
- Meyer, P. A. – Pifer, H. W. (1970): Prediction of Bank Failures. *The Journal of Finance*, 25(4), 853–868.
- Mian, A. – Sufi, A. (2009): The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *The Quarterly Journal of Economics*, 124(4), 1449–1496.

- Momparler, A. – Carmona, P. – Redondo, Y. (2020): Fuzzy-Set Qualitative Comparative Analysis (fsQCA) in Banking Failure Prediction. *Journal of Business Research*, 118, 253–263.
- Natekin, A. – Knoll, A. (2013): Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7, 21.
- Oppel, S. – Strobl, C. – Huettmann, F. (2009): Alternative Methods to Quantify Variable Importance in Ecology. University of Munich, Department of Statistics, Technical Report 65.
- Powers, D. M. (2011): Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*.
- Provost, F. – Fawcett, T. – Kohavi, R. (1998): The Case Against Accuracy Estimation for Comparing Induction Algorithms. In Proceedings of the 15th International Conference on Machine Learning (ICML, 1998).
- Rajan, R. G. (2010): *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton, New Jersey: Princeton University Press.
- Ravi, V. – Promodh, C. (2008): Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review. *European Journal of Operational Research*, 180(1), 1–28.
- Saito, T. – Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015 Mar 4;10(3): e0118432, DOI: 10.1371/journal.pone.0118432, PMID: 25738806; PMCID: PMC4349800.
- Serrano-Cinca, C. – Fuertes-Callén, Y. – Gutiérrez-Nieto, B. – Cuellar-Fernández, B. (2014): Path modelling to bankruptcy: Causes and symptoms of the banking crisis. *Applied Economics*, 46(31), 3798–3811.
- Shiller, R. J. (2008): *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It*. Princeton, NJ: Princeton University Press.
- Sinkev, J. F. (1975): A Multivariate Statistical Analysis of the Characteristics of Problem Banks. *Journal of Finance*, 30(1), 21–36.
- Sorkin, A. R. (2009): *Too Big to Fail: The Inside Story of How Wall Street and Washington Fought to Save the Financial System – and Themselves*. London: Allen Lane (Penguin Books).
- Stehman, S. V. (1997): Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1), 77–89.
- Thomson, James (1991): Predicting bank failures in the 1980's. *Economic Review*, 27, issue Q I, p. 9–20, <https://EconPapers.repec.org/RePEc:fip:fedcer:y:1991:i:qi:p:9-20:n:v.27no.1>.
- Trussel, J. J. – Johnson, L. (2012): A parsimonious and predictive model of the recent bank failures. *Academy of Banking Studies Journal*, 11(1), 15–30.
- Turney, P. D. (2000): Types of cost in inductive concept learning. Published in Proceedings of the Cost-Sensitive Learning Workshop at the 17th ICML-2000 Conference, Stanford, CA. July 2, 2000 NRC 43671.
- Tutz, G. – Binder, H. (2006): Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics*, 62(4), 961–971.
- Zadrozny, B. – Elkan, C. (2001): Learning and making decisions when costs and probabilities are both unknown. In KDD '01: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 204–213.
- Zou, H. – Hastie, T. (2005): Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320.