

MESTERSÉGES INTELLIGENCIA A HITELKOCKÁZATI MODELLEKNÉL, AVAGY MIRE KÉPESEK A GÉPI TANULÁSI ALGORITMUSOK A HAGYOMÁNYOS MODELLEKHEZ KÉPEST

Rajka László – Pollák Zoltán¹

ABSZTRAKT

A mesterséges intelligencia által fémjelzett technológiai forradalommal a hitelkockázat kezelésének területén modellek egy új generációja jött létre, amelyeket összefoglalóan gépi tanuláson alapuló modelleknek nevezünk. A hitelkockázati modelleknek az elmúlt évtizedekben végbemenő fejlődésében a szakértői rendszerek jelentik a múltat, a hagyományos statisztikai modellek (például a logisztikus regresszió) képviselik a jelent, a gépi tanulási metódusok pedig valószínűsíthetően a jövőt. Jelen tanulmány célja ez utóbbi gépi tanulási modellek egyik legígéretesebb képviselőjének, az XGBoost klasszifikációs algoritmusnak az ismertetése és annak empirikus vizsgálata, hogy a jelenlegi iparági best practice-nek számító hagyományos modellezési módszerekhez képest milyen hatékonyságnövekedés érhető el a gépi tanulási algoritmusok segítségével. Vizsgálatunkban a mesterséges intelligencián alapuló módszertanok közül a Mesterséges Neurális Hálózat (ANN) és az XGBoost-modell egyaránt felülmúlta a logisztikus regressziós megközelítést a klasszifikációs hatékonyság tekintetében. Noha a gépi tanulási módszerek kiváló predikciós képességgel rendelkeznek, hátrányuk, hogy esetükben a döntési modellek értelmezése nehezebb hagyományos társaikhoz képest. A mesterséges intelligencián alapuló gépi tanulási módszerek „fekete doboz” jellege miatt a bankoknak jelenleg kevés lehetőségük van azok alkalmazására, ezért javasolnánk a jelenlegi – hagyományos modellekre szabott – szabályok és ajánlások újragondolását olyan módon, hogy az teret adjon a bankoknak a gépi tanulási modellek alkalmazására és ezáltal hitelkockázat-kezelésük hatékonyságának növelésére.

JEL-kódok: C13, C25, C51, C53, C58, G21

Kulcsszavak: mesterséges intelligencia, gépi tanulás, applikációs scoring, XGBoost, logisztikus regresszió, nemteljesítés valószínűsége

¹ *Rajka László* kockázatelemző, OTP Bank. E-mail: laszlo.rajka@gmail.com.

Pollák Zoltán levelező szerző, főiskolai docens, Budapesti Gazdasági Egyetem Pénzügyi és Számviteli Kar, Pénzügy Tanszék. E-mail: pollak.zoltan@uni-bge.hu.

1. BEVEZETÉS

A bankok legjelentősebb kockázataként a hitelkockázat kezelése és a később potenciálisan nem teljesítő hitelfelvevők szűrése kulcskérdés a profitábilis működés szempontjából, ezért a minél jobb predikációs képességgel rendelkező scoringmodellek alkalmazása kiemelt cél számukra.

Ez a fajta nyereségességi szempont volt az egyik hajtóereje a hitelkockázati rendszerek folyamatos fejlődésének az elmúlt évtizedekben. A 20. század második felét még eleinte a szubjektív elemekkel tarkított szakértői rendszerek uralták a hitelek kockázati értékelésében. Ezt a hegemoniát törték meg a számítástechnika és az egyre hatékonyabb algoritmusok fejlődésével a hagyományos statisztikai modellek, amelyek egyik zászlóshajójaként a logisztikus regresszió alapuló scoringrendszerek iparági best practice-szé váltak.

A mesterséges intelligencia előretörésével újabb forradalmi változások játszódnak le a szemünk előtt. Ez ráirányítja a figyelmet a hitelezési scoringrendszerek egy új generációjára, amelyet a gépi tanulási módszerek fémjeleznek. A technológiai fejlődés újabb mérföldkövének köszönhetően újabbnál újabb gépi tanuláson alapuló algoritmusok jelennek meg, amelyek közül az egyik legígéretesebbnek tűnő XGBoost-algoritmust helyeztük tanulmányunk középpontjába.

A profitabilitási szempont mellett a másik fontos hajtóereje a hitelkockázat-kezelés robbanásszerű fejlődésének a prudens bankműködés egyre szofisztikáltabb szabályozói elvárásrendszere volt. Az elmúlt évtizedek válságai között találunk olyanokat, amelyek a hitelkockázattal hozhatók összefüggésbe, így a nemzetközi szabályozó szervezetek is elkezdtek jobban koncentrálni az említett modellezési kérdéskörre (MNB, 2002).

A bázeli tőkeszámítási szabályrendszer és a banki kockázatkezelés részét képezi a belső tőkemegfelelés-értékelési folyamat (*ICAAP, Internal Capital Adequacy Assessment Process*), amelyben belső eljárás alapján méri fel a hitelintézetek kockázataikat, és számszerűsítik a fedezésükhöz szükséges tőkenagyságot. Ennek a folyamatnak az egyik alapját a nemteljesítési valószínűség (továbbiakban *PD, probability of default*) becslése jelenti. A hitelintézetekkel szemben elvárás, hogy ne eseti megoldásokat alkalmazzanak, hanem egységes modellezési eljárások alapján mérjék fel a hitelezési kockázatokat (MNB, 2018).

A szakirodalmat áttekintve az rajzolódik ki, hogy a bankok jelenleg még mindig jelentős részben a hagyományos technikákra támaszkodnak, mivel ezek értelmezése az üzletmenetben (és a szabályozó hatóság előtt) sokkal könnyebb. Mindközben a hitelkockázati modellezésben az elmúlt néhány évben megjelentek és fokozatosan egyre nagyobb szerepet kapnak a gépi tanulási módszerek, a döntési mechanizmusok átláthatósága esetükben viszont sokkal nehezebb.

Leegyszerűsítve úgy fogalmazhatnánk, hogy a hitelkockázati modellek fejlődéstörténetében a szakértői rendszerek jelentik a múltat, a hagyományos statisztikai modellek képviselik a jelent, a gépi tanulási metódusok pedig vélhetően a jövőt. Jelen tanulmány célja, hogy az utóbbi csoportba tartozó XGBoost klasszifikációs algoritmust bemutassuk, és egy valós hitelkártya-adatbázison megvizsgáljuk, hogy az iparági gyakorlatban jelenleg uralkodó, hagyományos modellezési módszerekhez képest milyen hatékonyságnövekedés érhető el a gépi tanulási algoritmusok felhasználásával.

Ennek megfelelően a vizsgált hipotézisünk a következő:

A gépi tanulási módszerek – és azon belül kiemelten az XGBoost-eljárás – klasszifikációs teljesítmény szempontjából képesek felülmúlni az iparági gyakorlatban jelenleg legerjedtebb, hagyományos statisztikai modelleket.

Fontos megjegyezni, hogy jelen tanulmánynak nem célja a lehető leghatékonyabb modell felállítása a vizsgált adatbázison. A modellezés jelen esetben csupán illusztrációs célokat szolgál, és a bemutatás mellett a kétféle megközelítés közötti hatékonyságbeli különbségre fókuszál, egyazon mintán, ugyanolyan körülmények között.

Kitérünk arra, hogy habár a gépi tanuláson alapuló módszerek hatékonyabbak lehetnek, de az alkalmazásuk további problémákat vethet fel. A modellek hatékonysága mellett így bemutatjuk a felhasználási és alkalmazhatósági szempontokat is.

A bevezetőt követően összefoglaljuk a szakirodalmi hátteret, a releváns kutatások eredményeit, továbbá a nemzetközi szervezetek által megfogalmazott elvárásokat a modellekkel kapcsolatosan. A tanulmány második felében részletesen ismertetjük az egyes modellezési módszertanokat, a modellezéshez és a modellek kiértékeléséhez szükséges metrikákat, végül pedig összehasonlítjuk az egyes módszerek alapján kapott eredményeket.

2. SZAKIRODALMI ÁTTEKINTÉS

Ebben a fejezetben szeretnénk bemutatni az aktuálisan fellelhető irányvonalakat és modellezési módszertanokat a PD-modellezés szakirodalmában. Kitérünk a modellek fejlődésére, az iparági best practice-ekre, illetve a leghatékonyabb becslési módszerekre. Szeretnénk továbbá ismertetni azt is, hogy az EBA, az MNB és a bázeli standardok jelenleg milyen irányvonalakat és elvárásokat fogalmaznak meg a modellezési eljárások kapcsán. Az irodalmi áttekintés célja tehát bemutatni párhuzamosan az elméleti, a gyakorlati, valamint a szabályozói oldal aktuális álláspontját a témában.

2.1. A modellezési módszertanok fejlődése

Az 1980-as években a pénzügyintézetek túlnyomórészt szubjektív elemzésre vagy másképpen szakértői rendszerekre támaszkodva végezték el a hitelek kockázati értékelését. A döntés során teljesen szubjektív módon, belső megítélésükre hagyatkoztak olyan szempontokat vizsgálva, mint például a hírnév, a tőke, az adóságállomány vagy a fedezet (Altman–Saunders, 1997).

Hasonló módon Sommerville–Taffler (1995) is arra a megállapításra jutott, hogy a kockázati értékelések során a bankárok szubjektív megítélése volt sokáig a jellemző, miközben a többváltozós hitelminősítő rendszereket alkalmazó intézmények hatékonyabban működtek. A pénzügyintézetek a század végére egyre inkább eltávolodtak a szubjektív szakértői rendszerektől, így sokkal inkább előtérbe kerültek a historikus adatokon alapuló becslési módszertanok.

Leginkább a 2000-es évek elején kezdtek megjelenni olyan tanulmányok, amelyek próbálták javítani a hagyományos modellek nemteljesítési valószínűségeinek becslési hatékonyságát. West (2000) cikkében több neurális háló-alapú modellt is összehasonlított a logisztikus regressziós eljárással. A lineáris elemzéshez képest ugyan a logisztikus regresszió alapuló modell jobb becslést adott, viszont a legjobb klasszifikációs képességgel a neurális háló-alapú modellezési eljárás rendelkezett vizsgálatában.

A kétezres évek elejét követően a számítástechnika fejlődésével egyre több olyan tanulmány készült, amely tovább próbálta javítani a vizsgált modellezési kérdéskör pontosságát. Huang et al. (2007) az adatbányász eszközök oldaláról közelítették meg a hitelkártya-kihelyezésekkel kapcsolatos PD-becslési módszereket. Az SVM (*support vector machine*) osztályozó algoritmus egy hibrid verzióját használták fel, amelynek az eredményei ígéretesnek bizonyultak az akkor használatos becslési módszerekhez képest.

Yu (2020) cikke is a hitelkártya-bedőlések hatékonyságát vizsgálta. Tanulmányában bevonta a modellezésbe az amerikai FICO Score adatait is. Arra jutott, hogy a gépi tanulási módszerek magasabb hatékonysággal rendelkeznek, illetve pontosabb becslést tudnak adni. Vizsgálatában a random forest algoritmus bizonyult a legpontosabbnak az adaboost, a döntési fák és a logisztikus regressziós eljárást megelőzve. A cikk arra is rávilágít, hogy az elérhető adatok mennyisége kulcsfontosságú pont a hitelkockázat modellezésének területén.

Angelini et al. (2008) bemutatják a neurális hálózatok alkalmazásának lehetőségeit a hitelkockázat értékelésének szempontjából. A tanulmány a mesterséges neurális hálózatok mellett az előrecsatolt tanulási mechanizmusokat is vizsgálja. Kutatásukban mindkét módszer esetében magas hatékonyságról számolnak be a becslési pontosság szempontjából. Azt is bemutatják, hogy a kockázatok értéke-

lésekor jelentős hibát okozhat a modellben, ha a zajos adatokat nem szűrik, így javasolják, hogy a modellezési eljárások használata előtt érdemes valamilyen normalizálást végrehajtani. A tanulmány azt is megállapítja, hogy van lehetőség a becslések további pontosítására, amely a paraméterek optimalizálásán keresztül érhető el. A cikk azt is javasolja, hogy érdemes lehet megtartani a hagyományos regressziós eljárásokat, és ezek kiegészítéseként és a becslések finomításaként alkalmazni a neurális hálózatokat.

Medema et al. (2009) a bázeli tőkeegyezményekhez kapcsolódó modellek validációs irányelveit egészítik ki egy gyakorlati kutatással. A tanulmány szerint egy jó PD-modellnek mind elméleti síkon, mind az adatok és a statisztika szintjén egyaránt meg kell állnia a helyét. A kutatásban javasolnak egy paramétervektort, amely a PD-k érvényesítésére alkalmazható. Cikkük felhívja arra is a figyelmet, hogy kiemelt módon kell kezelni a hiányzó adatokat, továbbá fontos figyelembe venni a modellek mintán kívüli hatékonyságát is, amelynek a méréséhez javasolják külső adatbázisok bevonását vagy bootstrap-eljárás alkalmazását.

Gurný–Gurný (2013) cikkükben a hagyományosnak mondható logit-, probit- és az LDA-modellek teljesítményét hasonlították össze. A tanulmányuk egyediségét az adja, hogy a modelleket 298 amerikai bank adatain tanították be, majd 100 más kereskedelmi bank adatain mérték a hatékonyságot. Így a tanulmány egy bankrendszer szintű modellt próbált meg felállítani, amely általánosságban jól leírja a banki bedőléseket. A modell rövid távú, 1-2 éves nemteljesítésekre becsült, és a logitmodell adta a leghatékonyabb előrejelzést, amelyhez igen impozáns ROC-görbe társult. A modellben felhasznált információk főleg pénzügyi adatok és ezek arányai vagy logaritmizált értékei voltak. A tanulmány azt is vizsgálta, hogy a kockázati modellekben a bemeneti adatok és a nemteljesítési események között milyen késleltetést érdemes alkalmazni.

A gépi tanulási algoritmusok felhasználását vizsgálták Butaru et al. (2016), ahol egy hat bank adatait tartalmazó, nagy hitelkártya-adatbázison mérték a hitelkockázati modellek teljesítményét. Kutatásukban 2009 és 2013 közötti adatokon hasonlították össze a döntésifa-alapú és a logisztikus regressziós eljárás eredményét. Megállapították, hogy a késedelmet a döntésifa-alapú modellek jobban magyarázzák, mint a regressziós eljárás. A tanulmány összehasonlította az egyes bankok kockázatkezelési gyakorlatait, és azt találták, hogy az üzletmenetből adódóan az egyes bankok között magas a heterogenitás a kockázati tényezők és a tényezők iránti érzékenység tekintetében. A tanulmány ennek alapján megállapítja, hogy nehézkes egy bankrendszer leíró modellt meghatározni. A szerzők kiemelik, hogy a hitelfortfólió jellemzői önmagukban nem mindig elegendők a késedelmek meghatározásához, mivel a bankok folyamatosan és aktívan kezelik a portfóliójukat.

Sirignano et al. (2016) kutatásukba 1995 és 2014 közötti jelzáloghitel-adatokat vontak be az Amerikai Egyesült Államokból. Az elemzéshez egy példa nélküli módon hatalmas, 120 millió megfigyelést tartalmazó adatsort használtak fel, amely egyaránt tartalmazott hitel- és hitelfelvevő-specifikus, illetve makrogazdasági változókat is. Megállapították, hogy a hitelfelvevő magatartása és a kockázati tényezők közötti kapcsolat nem tekinthető lineárisnak. Több esetben interakciókat figyeltek meg a vizsgált változók között. Egy mintán kívüli elemzés során kimutatták, hogy az adatokban talált nemlinearitások kezelése jelentősen javítja a hitel- és poolszintű kockázati előrejelzések pontosságát, a jelzálog-kereskedési stratégiák befektetési teljesítményét, valamint a jelzálog-fedezettű hitelek értékelését és fedezését. A modellezési eredmények szempontjából megállapították, hogy a neurális hálózatok felülmúlják a logisztikus regresszió adta eredményeket.

Venkatesh–Jacob (2016) a BayesNet, a Meta-Stacking, a Naive Bayes, a Random Forest, az SMO és a ZeroR algoritmusok előrejelzési erejét vizsgálták a hitelkockázat szempontjából. Tanulmányukhoz a University of California adatbázisát használták fel, amely egy tajvani bank hitelkártyaadatait tartalmazta. Az adatbázisban a hitelhez kapcsolódó pénzügyi információk mellett pár hitelfelvevőhöz kapcsolódó információt is felhasználtak a modellezés során. A tanulmány rögzíti, hogy az információs érték és a korrelációs elemzés hatékonyan felhasználható a magyarázó változók szűréséhez egy hitelkockázati modellezési probléma esetén. A kutatás eredménye, hogy a random forest, random tree és az IBK osztályozási algoritmusok adják a legjobb becslési eredményt, ám a többi módszer is közel 80 százalékos pontossággal dolgozik.

Addo et al. (2018) a mélytanulási algoritmusok teljesítményét vizsgálták a hitelkockázati modellezés szempontjából. Kutatásukban az iparági gyakorlatban uralkodó logisztikus regresszió, a random forest, a gradiens boosted modelleket és a neurális hálózatokat hasonlították össze. A tanulmány megállapítja, hogy a döntésifa-alapú modellek stabilabb becslést adnak a többretegű neurális hálózatoknál, és felülmúlják a hagyományos regressziós eljárásokat is. A szerzők bemutatják, hogy fontos az adatminőség ellenőrzése a tanító adaton, hiszen ez tartalmazhat torzításokat is a képzési osztályok kapcsán. Kiemelik, hogy érdemes több oldalról is vizsgálni a modellek összehasonlítását a validáció során, mivel a különböző modellek összevetése nehézkes. Például az AIC, BIC és R^2 mutatók nem minden modell esetén értelmezhetőek, így a tanulmányukban végül a ROC-görbét és az ebből képzett AUC- és GINI- értéket alkalmazzák a kiértékeléshez. A tanulmányban a szerzők kiemelik, hogy jelenleg a gépi tanulási algoritmusokhoz kapcsolódóan minél hamarabb szükség lenne a szabályozások kidolgozására, hiszen ezzel elkerülhetőek a fekete dobozos megoldások alkalmazásában rejlő hibák.

Moradi–Mokhatab (2019) tanulmánya az eddigi modellekhez képest egy fuzzy eljárást javasol a hitelkockázatok értékeléséhez. Véleményük szerint a modellek általában nem használják fel jól a havonta dinamikusán változó olyan külső adatokat, mint a politikai vagy gazdasági szankciók. Kutatásukban dinamikusán bővülő módon minden hónapban megvizsgálták a rossz ügyfeleket, és próbálták kiszűrni az olyanokat, akik nehézségekbe ütköztek a gazdasági ciklusok okozta negatív hatások miatt. A szerzők által javasolt kockázati modell felhasználható a hitelminősítésben úgy, hogy az információt tartalmaz a csőd vagy a hiperinfláció megjelenése kapcsán.

Wang et al. (2023) a hitelkockázati modellekben felhasznált jellemzők kiválasztására vonatkozó eljárásokat mélyítik el a gépi tanulási módszerek oldaláról, emellett pedig kockázatértékelő modelleket is bemutatnak a tanulmányukban. A leghatékonyabb modell kiválasztásának szempontjából kutatásuk az eddig bemutatott cikkekhez képest eltérő eredményt mutat, hiszen itt a hagyományosnak mondható logisztikusregresszió-eljárás felülmúlja az XGBoost és a döntési fák segítségével adott becsléseket.

Egy kifejezetten érdekes újdonság Rozo et al. (2023) cikke, amely a Covid19-világjárvány kapcsán teljesen online térbe szoruló világ hatásait vizsgálta a modellezésben. A modellben magyarázó változóként a böngészési adatokat, a webhelylátogatások számát és a bankfióki látogatásokat vették alapul a PD-becslésének módszertanában. A kutatásban arra jutottak, hogy a webes viselkedési adatokat felhasználva csökkenthetik a becslési hibákat a hagyományosnak nevezhető PD-becslési módszertannal készített modellekben. Eredményként tudták azt is felmutatni, hogy az aktív webes felhasználók között a mintában nem mutatkozott jelentős életkorbeli különbség, így az adat megfelelő módon használható a teljes mintán végzett becslésben.

A szakirodalmi áttekintés alapján megállapítható, hogy a modellezési eljárások jelentős fejlődésen mentek keresztül, párhuzamosan a technológia fejlődésével. Mind a hatékonyság, mind pedig a modellezési technikák terén nagy előrelépéseket és sokszínűséget tapasztalhattunk az elmúlt két évtizedben. Az elérhető adatok mennyisége és az adatok információs ereje is folyamatosan bővül, így a modellek is szélesebb körű adatpontokból tudnak kiindulni.

A gépi tanulási módszereken belül az XGBoost-algoritmus vizsgálata és alkalmazásának lehetőségei – amint azt a szakirodalmi áttekintő is mutatta – külföldön egyre jobban kutatott terület, a hazai szakirodalomból viszont teljesen hiányzik. Tanulmányunkkal szeretnénk ezt a rést betölteni, és új irányt nyitni a téma hazai kutatói számára.

2.2. Az iparági gyakorlat és a szabályozók iránymutatásai

A bankoknak a lakossági hitelezés volumenének növekedése és az élesedő verseny miatt pontosabb modelleket kell kifejleszteniük, így minimalizálva saját hitelezési veszteségeiket. A gazdag adatmennyiség lehetővé teszi számukra a hitelkockázati belső minősítésen alapuló módszerek hatékony alkalmazását. A PD-modellek a tőkekövetelmény meghatározása mellett a hitelkihelyezésről szóló döntés támogatásában is fontos szerepet játszanak.

A hitelkockázat szempontjából fontos megkülönböztetni a meglévő és az új ügyfeleket. A meglévő ügyfelekről sok folyamatosan frissülő adat áll rendelkezésre, viszont az új ügyfelekről kevés információval rendelkezik a hitelintézet. Új hitelígénylések esetén a bankok főként szociodemográfiai adatokra támaszkodhatnak, valamint amennyiben rendelkeznek bankközi historikus információkkal, úgy azokat is felhasználhatják. Az új ügyfelek hitelminősítő rendszereit jelentkezési (*application*) scoringnak nevezik, míg a meglévő ügyfelek esetén a hitelintézetek viselkedési (*behavioral*) scoringrendszereket alkalmaznak, amelyekben már a hitelhez köthető törlesztési adatokat is felhasználhatják.

A PD-becslésében a logitalapú modellek kifejezetten népszerűnek számítanak az iparági gyakorlatban, mivel a logisztikus együtthatók transzformálhatóak valószínűségi értékké. A logisztikus regresszióval becsült modelleket könnyen lehet akár scorecarddá is átalakítani, ahol a becsült valószínűségi értékeket pontszámokká konvertáljuk egy meghatározott skála mentén.

A modellezési eljárás során a bedőlés alapján képzett WoE (*weight of evidence*) értékek segítségével csoportok kerülnek kialakításra (Siddiqi, 2006). A WoE-értékek reprezentálják a részsokaságokon belüli bedölések arányát a teljes mintához képest.

Az így képzett scoringmodellek kiértékelése is kifejezetten fontos feladat. A gyakran használt eljárások közül kiemelendő a Kolmogorov–Szmirnov-statisztika (KS), a ROC-görbe és az ebből képzett GINI-érték (Kovács–Marsi, 2018). Ezeket a modellek hatékonyságának összehasonlításához fogjuk a későbbiekben felhasználni.

A banki minősítési rendszerek jellemzően hibrid módon működnek, tehát az applikációs és a viselkedési modelleken túl a PD jellege alapján is megkülönböztethetők az egyes hitelkockázati modellek. A bankok sokszor hibrid módon egyszerre alkalmaznak *Point in Time* (PiT) és *Through-the-Cycle* (TTC) modelleket is a kockázatok értékelésére, ezért szeretnénk bemutatni a két PD-típusban rejlő különbségeket.

Egy PiT-rendszerben a becslés függ a gazdasági ciklus pillanatnyi szakaszától, változói ciklikusak lehetnek. A PiT-rendszerben egy gazdasági visszaesés során a ciklusérzékeny mutatók használata (ezen mutatók értékeinek romlása) miatt az

adott ügyleteket tömegesen rosszabb minősítési kategóriákba migrálják. A PiT-rendszerekhez kapcsolódó tőkekövetelmény a „Point in Time” jelleg miatt időben ingadozó.

A TTC-rendszerekben sokkal inkább megjelenik a nevében is rejlő időszakokon átívelő becslési felfogás. Az intézmények a TTC-rendszerek felhasználása során igyekeznek kiszűrni a gazdasági ciklusok okozta volatilitást a nemteljesítési kockázatból, és mérni az ügyfél kockázatát a ciklus során. A TTC-minősítés nem reagál a gazdasági ciklus változásaira, csak a fundamentális jellemzők megváltozása okozhat ratingmigrációt. A TTC-rendszerek alkalmazása időben stabilabb tőkeszükséglethez vezet.

A jelenlegi, CRR-alapú tőkeszámításhoz használt modellekre vonatkozó piaci gyakorlat szerint érdemes a minősítési rendszereket a hitelintézeteknek úgy kialakítaniuk, hogy azok biztosítsák a minősítési kategóriák időbeni és a gazdasági ciklusokon keresztül átívelő becsléseinek a stabilitását. A belső minősítésen alapuló rendszerekben a TTC PD-becslését általában a lakossági szegmensben könnyű kivitelezni, hiszen a vállalati portfóliókat nehéz úgy figyelembe venni, hogy abból kiszűrésre kerüljenek a gazdasági ciklusok (Bíró–Nagy, 2018).

A European Banking Authority (2017) (továbbiakban: EBA) tanácsai szerint a PD-modellezés lépéseinek számos olyan eleme van, amelyeket szükséges szem előtt tartani a fejlesztés során. A statisztikai modellek használatakor szükséges figyelembe venni minden körülményt, amely az értékelés szempontjából releváns. Az EBA javaslata alapján a harmadik féltől származó minősítések felhasználása is hordozhat kockázatokat magában, így ezek felhasználásával óvatosan kell bánni. A modellépítés előtt szükséges tisztázni a kapcsolódó adatkövetelményeket, illetve az idősíkot, amelyen a modellezés történik. Fontos az adatok frissítésének gyakoriságát is figyelembe venni az elemzések során. Az irányelvek azt is kimondják, hogy a statisztikai megközelítések mellett érdemes az üzleti oldalon lévő szakértőket is bevonni, hogy a felhasznált információk megfelelő módon kerüljenek be az egyes becslő modellekbe, továbbá fontos kezelni az egy személynél lévő azonos hiteleket is.

Természetesen a Bank for International Settlements (2023) (továbbiakban: BIS) is előír minimális követelményeket, amennyiben egy hitelintézet belső minősítésen alapuló módszertant szeretne alkalmazni. A bázeli irányelvek szerint a hitelminősítési modellek sokszor csak mechanikus osztályozási eljárások mentén próbálnak egyfajta becslést adni a bedőlési valószínűségekre. Természetesen a modellek támogatják az üzletmenetet is, hiszen ezek felhasználásával akár automatizálható a hitelezési folyamat, viszont fontos megemlíteni, hogy a modellek sem tökéletesek, így az osztályozásban mindenképpen számolni kell becslési pontatlanságokkal. Így szükséges az adott modellek folyamatos monitoringja és a modell hatókörén kívüli információk figyelembevétele.

A belső minősítésen alapuló modellek módszertani elvárásai kapcsán a BIS (2023) nem ad részletes útmutatót, hanem általánosságban mutat irányt. A dokumentum szerint a bankoknak dokumentálniuk kell a nemteljesítés és veszteség sajátos, házon belül használt definícióit, és igazolniuk kell a konzisztenciát a bázeli standardokban meghatározott definíciókkal. Ha egy bank statisztikai modelleket alkalmaz a minősítési folyamatai során, akkor részletesen dokumentálnia kell annak módszertanát. A dokumentációnak pedig a következő standardoknak kell megfelelnie:

1. A modellnek pontosan kell megadnia a kategóriákhoz történő becslések, egyedi kötelezettségekhez történő becslések és kitétségekhez vagy poolokhoz való hozzárendelésének elméletének, feltevéseinek és/vagy matematikai és empirikus alapjait, valamint a modell becsléséhez használt adatforrás(oka)t.
2. A modell érvényesítésére szigorú statisztikai eljárásokat kell létrehozni (ideértve a más időszakra vonatkozó és a mintán kívüli, azaz független mintás teljesítményteszteket is).
3. A statisztikai modellekről szóló dokumentációnak meg kell jelölnie azokat a körülményeket, amelyek fennállása esetén a modell nem működik hatékonyan.

A fentebb pontokba szedett irányelveket tovább pontosítja a Magyar Nemzeti Bank (2018) (továbbiakban: MNB) által kiadott kézikönyv is. Összeségében elmondható, hogy nagyon hasonló elveket emel ki a hazai szabályozó is, összhangban a BIS (2023) dokumentumaival. Előírják a teljes körű, részletes dokumentációt, a fogalmak és eljárások pontos rögzítését, a megfelelő adatminőség biztosítását, az éves validációt, a modellek teljesítményének és stabilitásának rendszeres riportálását. A modellekkel kapcsolatosan fontos egy aktualizált leltár kialakítása, illetve folyamatos visszaméréssel kell igazolni a megfelelőséget. A modelleknek továbbá biztosítaniuk kell az érdemleges kockázati megkülönböztetést, illetve a portfólió minőségében bekövetkezett változásokat is jelezniük kell.

Az irányelvekben megfogalmazzák, hogy a ratingmodellek által meghatározott kategóriáknak tükröznie kell a monotonitást, tehát a rosszabb ratingkategóriákba a magasabb nemteljesítési rátával rendelkező ügyfeleket kell sorolni, míg az alacsonyabb kockázatok esetén jobb ratingkategóriákat kell meghatározni. Továbbá a hitelintézeteknek azt is mérniük kell, hogy időről időre milyen migráció figyelhető meg az egyes ratingkategóriák között, ahol azt is monitorozni szükséges, hogy a változások mögött valós portfólióromlás van-e, vagy a PiT-rendszer hibáiból adódó okokra vezethető vissza a jelenség. A tőke megfelelési kézikönyv a PiT minősítő rendszerek használata esetén TTC-rendszer alkalmazását is előírja. Az MNB kijelenti, hogy a megfogalmazott irányelvek mellett az adott modelleknek meg kell felelniük az EKB és a BIS belső minősítési rendszerekkel kapcsolatos irányelveinek is.

Fontos megemlíteni a nemzetközi felügyeleti hatóságok jövőbe mutató tevékenységét és ajánlásait is. Az EBA (2020) közzétett egy jelentést az Environmental, Social és Governance (környezeti, társadalmi és vállalati irányítási, továbbiakban: ESG) kockázatok intézményi alkalmazásának szükségességéről. Az EBA útmutatója kiemeli, hogy az ESG-kockázatok felhasználása a portfólió minőségének meghatározása során egy fontos szempont, viszont az olyan historikus adatokra támaszkodó modellek esetén, mint a PD- vagy LGD-becslés, a felhasználás igen nehézkes lehet. Az ESG alkalmazása a hitelkockázat-elemzésben még kezdetleges, viszont megjelentek már kutatások, amelyek bizonyítják, hogy létezik kapcsolat az ESG-események és a nemteljesítések között (Henisz–McGlinch, 2019).

3. A MODELLÉPÍTÉS FOLYAMATA ÉS AZ EREDMÉNYEK ÉRTÉKELÉSE

Ebben a fejezetben mutatjuk be a PD becslésre használt modellezési eljárásokat. Az első alfejezetben szeretnénk ismertetni a felhasznált adatbázist egy rövid leíró jellegű statisztikai leírással. A fejezet második részében bemutatjuk azokat a kiértékelési technikákat, amelyekkel a modellek eredményei összehasonlíthatóvá válnak. A harmadik részben az iparági best practice-nek megfelelően egy csoportosított változókat felhasználó logisztikus regressziós modell eredményeit ismertetjük. Végül az úgynevezett „fekete dobozos” megoldások hatékonyságát fogjuk megvizsgálni a becslés szempontjából. Az elemzéseket Python3 környezetben futtattuk (Rossum–Drake, 2009), ahol főként a Scikit-learn könyvtárra támaszkodtunk (Pedregosa et al., 2011). A kutatás során a hitelkártya-bedőléseket előrejelző modellek közül szeretnénk megkeresni a legkisebb hibával becsülő eljárást. A modellezés során egy applikációs PiT PD-modellt specifikálunk, majd ennek pontosságát a ROC-görbe, az AUC-érték és a GINI-index segítségével hasonlítjuk össze egy tesztmintán.

3.1. Az adatbázis bemutatása és leíró elemzése

A tanulmányban bemutatott modellezési eljárásokat és összehasonlításokat egy részletes és számos változóval rendelkező Kaggle (n. a.) hitelkártya-adatbázison készítettük el. Az adatbázis 122 változót tartalmazott, amelyek közül a célváltozó a nemteljesítés volt. A hitelbedőlés bináris változóként volt megadva, ahol a teljes mintán kerekítve 8 százalékos nemteljesítési (default) arány volt megfigyelhető.

Ahogy korábban ismertettük, a bankok használnak applikációs, illetve viselkedési modelleket is a bedőlési valószínűségeik becslésekor. Az adatbázis alapvetően applikációs információkból állt (azaz jelentkezési adatokat tartalmazott), viszont

megtalálhatók voltak benne korábbi jelentkezésekből adódó adatok is. Az applikációs adatok tekintetében főként a hiteligénylőt, illetve a tulajdonában lévő vagyoni elemeket leíró adatokat lehetett megtalálni, amelyek kiegészültek pár külső forrásból érkező információval. Mivel a korábbi jelentkezések a hitelügyletek kevesebb mint felénél voltak elérhetőek, így ezeket kizártuk az elemzésből. Továbbá minden olyan változót kizártunk, ahol a hiányzó adatok aránya magasabb volt 30 százaléknál. A korábbi jelentkezési adatok szűrésével 26 változó került ki, míg a 30 százalékos hiányzó adatarány segítségével végül 50 változóra szűkült az adatbázis. A modellépítés során ezt követően töröltük az olyan megfigyeléseket, amelyek a kiválasztott változók esetén hiányzó értéket tartalmaztak. A teljes minta eredetileg 307 511 darab rekordot tartalmazott, ez végül a szűréseket követően 263 947 megfigyelésre csökkent.

Kiemelt változó volt a hitel nagysága, amely 5603 darab egyedi megfigyelést tartalmazott. A legkisebb kiadott hitel 45 000, míg a legnagyobb 4 050 000 volt (az adatbázis nem adta meg az adatok pénznemét). A hitel nagyságának eloszlása balra ferde, jobbra elnyúló alakot vett fel, vagyis a kiadott hitelek jellemzően kis értékűek voltak, a hitelkihelyezések nagyságának növekedésével előfordulási gyakoriságuk fokozatosan csökkent. A kiadott hitelek volumenének átlaga közel 600 000 egység volt, melyhez 404 312 egységnyi szórás társult. Az eredményeket az 1. táblázatban mutatjuk be.

Az elemzés egyik következő kulcsváltozója az életkor volt a hitelkihelyezés időpontjában. A változó eloszlását tekintve egyenletesen található a mintában minden korosztály, a legidősebb hitelkártya-felhasználó 70 év körüli volt, míg a legfiatalabb 21 éves.

1. táblázat

A főbb változók leíró elemzése a szűrések után

	Az előző telefon cseréje óta eltelt napok száma	Az előző igazolvány cseréje óta eltelt napok száma	Hitel nagysága	Regisztrációs adatok változása óta eltelt napok száma	Jövedelem nagysága	Születés óta eltelt napok száma
Átlag	-988	-3 049	606 587	-4 990	171 124	-16 120
Szórás	833	1491	404 312	3 523	249 021	4 308
Minimum	-4 185	-7 197	45 000	-24 672	26 100	-25 201
25% percentilis	-1 603	-4 319	272 579	-7 479	112 500	-19 716
50% percentilis	-798	-3 335	521 280	-4 517	157 500	-15 816
75% percentilis	-286	-1 814	813 195	-2 007	202 500	-12 574
Maximum	0	0	4 050 000	0	117 000 000	-7 489

Forrás: saját szerkesztés

A teljes jövedelem tekintetében a minimum 26 100, míg a legmagasabb jövedelemérték 10^8 nagyságrendű volt. Természetesen nehéz meghatározni egy ilyen kiugró bevételérték valódiságát, de a modellezési eljárás során a WoE-csoportosítás segítségével az outlier-értékek problémáját kezeltük. A jövedelem tekintetében az átlagot 171 124 egység adta, melyhez 249 021 egységnyi szórás társult. A kihelyezett hitelösszegekhez hasonló módon tehát itt is igen heterogén volt a minta.

A mintában a kihelyezett hitelösszegek és a jövedelem között szoros pozitív irányú kapcsolatot lehetett megfigyelni, tehát a valóban magas jövedelemmel rendelkező ügyfelek használták ki nagyobb értékben a kiadott hitelkereteiket. Ez az összefüggés egyébként nem meglepő, mivel a nagyobb jövedelmű ügyfelek esetén jellemzően nagyobb hitelkereteket is tudnak biztosítani a bankok, így a lehívható és kihasználható összeg is nagyobb volt ezen ügyfelek esetén.

Érdekes információként meg lehet említeni a legutóbbi mobilcsere óta eltelt napok számát, amely a modellben végül egy szignifikáns változó lett. Itt a két végletet a 0, illetve a 4185 nap adta. Persze érdekes felvetés, de elképzelhető, hogy van olyan hitelkártyát használó ügyfél, aki 10 év alatt nem váltott telefont, azért megemlítjük, hogy ez is inkább kiugró értéknek számított. A mintában az ügyfeleknek átlagosan 988 napja vásárolták a telefonjukat, amely nagyjából 2,5 évet jelent. A legutóbbi telefonvásárlás óta eltelt napok számának szórása 833 nap, tehát a 2,5 éves átlaghoz 2,26 évnnyi szórás tartozik a telefoncsere tekintetében.

További lényeges változó volt a mintában a jelenlegi munkahelyen eltöltött napok száma. A munkanapok számában is nagyjából egyenletes megoszlást lehet megfigyelni, viszont itt volt jó pár adathiba, mivel ellentétes előjelű érték is megjelent az adatbázisban. Ezt a problémát is szűréssel kezeltük a modellezés során.

Az adatbázisban az is megtalálható volt, hogy az adott adós hány napja cserélte le a személyi igazolványát. A személyiigazolvány-csere tekintetében is egy nagyjából folytonos eloszlást lehet megfigyelni, viszont azért a régebbi személyik nagyobb arányban vannak jelen a mintában. Átlagosan a mintában lévő ügyfelek 8 éve cserélték le a személyi igazolványukat, és ehhez 4 évnnyi szórás társult.

További változó volt a hitelből megvásárolt áru értéke, amely a jövedelem és a hitelösszeghez hasonlóan egy balra ferde, jobbra elnyúló eloszlást vett fel. A változó átlaga 606 587 egység volt, amelyhez 404 312 egységnyi szórás társult. Szeretnénk kiemelni, hogy a hitelkártyával vásárolt áruk értékében a hitel összege egy részahalmazt képzett, és magas korreláció volt megfigyelhető a két változó között. Ezek alapján megállapítható, hogy a magasabb értékű árukhoz, nagyobb hitelösszegeket hívtak le a hitelkártyakeretükből az ügyfelek. A magas korreláció miatt szűrést kellett végeznünk a regressziós eljárásnál, amit a modellezési lépéseknél fejtünk ki részletesebben.

3.2. A modellek összehasonlításának módszertana

Mielőtt a modellezési eredményeket részleteznénk, szeretnénk bemutatni, hogy milyen metrikák mentén végeztük el a modellek teljesítményének összehasonlítását. Ahogy a bevezetőben is előrevetítettük, a tanulmány célja annak bemutatása, hogy az iparági gyakorlatban is megjelenő best practice-hez képest mennyiben tud javulást hozni a gépi tanulási módszerek alkalmazása a hatékonyság tekintetében. Ahhoz, hogy a modelleket összehasonlítsuk, fontos meghatározni egy jól megragadható, egységes mutatót, amely alkalmas a becslési teljesítmény összemérésére.

A Kolmogorov–Szmirnov-statisztika (KS) a hitelkockázati modellek validációjában olyan eszköz, amely segít az értékelőnek meghatározni, hogy a modell hogyan teljesít az eloszlásváltozók (például a hitelek vagy hitelkockázati mutatószámok) előrejelzésében. A KS-statisztika egy nem paraméteres teszt, amelynek a segítségével meghatározható, hogy két adatkészlet szignifikánsan különbözik-e egymástól. A KS-statisztika egy cutoff mentén vizsgálja a két külön részre vágott portfóliót (a modell által becsült bedőlt és a modell által becsült nem bedőlt ügyletek között). A két kumulatív eloszlásfüggvényt hasonlítja össze, és a két görbe közötti maximális érték adja a KS-statisztikát (Madar, 2015).

A hitelkockázat-elemzésben a kiértékeléshez gyakran használatos a *Receiver Operating Characteristic* (ROC) görbe és a görbe alatti terület (AUC, *Area Under Curve*), valamint az ebből képzett GINI-együttható. A ROC-görbe előállításához szükség van a PD-re vagy az ebből képzett pontszámra, illetve a valós bedőlési adatokra. Ezek után a modell által becsült PD-értékek mentén megvizsgáljuk, hogy az egyes cutoff-értékeknél mi a pontos találati arány a bedölések és nem bedölések mentén. Az úgynevezett *hit rate* és *false alarm rate* adja a két tengelyt, amely között felrajzolható a ROC-görbe. A ROC-görbe meredeksége egyben megmutatja azt is, hogy egy adott modell mennyire jó, hiszen ez választja el a valós bedöléseket és nem bedöléseket egymástól.

Fontos kiemelni, hogy a modellek jó összehasonlíthatóságához szükséges meghatározni egy jól megragadható együtthatót, amely megfelelően méri a diszkriminációs erőt. Erre tökéletes az AUC, amely a ROC-görbe alatti területet mutatja meg, így tehát minél nagyobb a görbe alatti terület, annál jobb az adott modell. Az összehasonlíthatóság kedvéért az AUC értékét még módosítani lehet a következő képlet szerint: $GINI = 2 \times AUC - 1$, amely metrika esetén ugyanúgy a legmagasabb érték kapcsolódik a legjobb szeparációs erővel rendelkező modellhez (Madar, 2015).

3.3. A hagyományos modellezési eljárás eredményei

Mielőtt a hagyományos modellezési eljárás eredményeit részleteznénk, szeretnénk bemutatni a kapcsolódó módszertant. Ezek használatához szükséges a változók felülvizsgálata, illetve regresszió esetén a változók kategorizálása. Ez a fajta kategorizálás lehetőséget nyújt a kockázati előrejelzések pontosabbá tételére, amelyhez a defaultrátát használjuk fel.

Az egyik legnépszerűbb eljárás a WoE (*weight of evidence*) módszer, amely a logisztikus térben keresi az optimumot, és a változókat egy kategorizált értékkel tölti fel. A WoE-csoportosítást a logisztikus regresszió előtt végeztük el annak érdekében, hogy a nemlineáris hatásokat megragadjuk. A binek meghatározásához döntésifa-alapú szegmentálást választottunk, amelyet szükség esetén a küszöbök módosításával tovább finomítottunk. A WoE-vágások lényege statisztikai megközelítéssel tehát az, hogy egy nagy sokaságból bizonyos információ mentén részsokeket képzünk, amihez felhasználjuk a nemteljesítési információt.

A WoE felhasználása azért előnyös, mert az így kapott helyettesített értékek megőrzik a változó nemlineáris összefüggéseit. A WoE-értékek alkalmazása továbbá azért is megfelelő lehet, mert a képzett csoportok mentén a következő aggregálással egy információs értéket (*information value*) is meghatározhatunk, amellyel az egyedi változók erejét lehet összehasonlítani.

Az *information value*-t mindig érdemes az adott modellezési probléma esetén külön értékelni, ahol a nagyobb érték erősebb magyarázóerővel rendelkező változót mutat. Az így kapott WoE-értékekkel létrehozható egy új adathalmaz, ahol már nem az eredeti értékeket tartalmazó értékkészlet szerepel, hanem a WoE-értékek. Ez a gyakorlatban azt jelenti, hogy amilyen sávba esnek a csoportosításnál az egyes értékek, a hozzájuk kapcsolódó WoE-értéket használjuk a továbbiakban a modellezésnél (Kovács–Marsi, 2018).

A hagyományosnak mondott iparági gyakorlatból a logisztikus regressziót választottuk, hogy felhasználásával egy becslést végezzünk a nemteljesítési valószínűségekre. A logisztikus regresszió egy statisztikai módszer, amelyet gyakran használnak kategorikus vagy bináris kimenetekkel rendelkező problémák megoldására. Ezt a módszert általában osztályozási feladatoknál alkalmazzák, ahol a cél meghatározni egy vagy több bemeneti változó alapján, hogy egy adott megfigyelés melyik kategóriába tartozik. A logisztikus regresszió a logisztikus függvény segítségével modellezi a kimeneti valószínűséget a bemeneti változók alapján. A bináris klasszifikációt használó „S” alakú logisztikus függvény 0 és 1 közötti valószínűségi értékeket ad vissza, ami felhasználható bináris problémák értelmezéséhez. A logisztikus függvény a következő módon írható fel:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_n x_n)}} \quad (1),$$

ahol a $p(x)$ jelöli egy esemény bekövetkezési valószínűségét, a β_0 pedig a bemeneti változók 0 értéke esetén a logit/logodds értéket adja. A β_n a bemeneti változókhoz tartozó súlyokat jelöli (koefficiens), amelyek azt mutatják, hogy mennyire befolyásolják a kimeneti valószínűséget a bemeneti változók értékei. Az x_n pedig a bemeneti változókat jelöli, amelyeket a logisztikus regressziós modell az esemény bekövetkezési valószínűségének becslésére használ.

A logisztikus regresszió alkalmazásának egyik lényeges lépése az, hogy egy valószínűségi értéket átalakítsunk bináris változóvá. Ezt úgy tehetjük meg, hogy megállapítunk egy küszöbértéket (*threshold*), amely felett az eseményt bekövetkezettnek tekintjük, és alatta pedig nem. A valószínűségi érték azért is jól felhasználható, mert egy gazdasági eseményhez kapcsolódó probléma esetén pénzügyi vonatkozása is van a *threshold* mértékének (Peng et al., 2002).

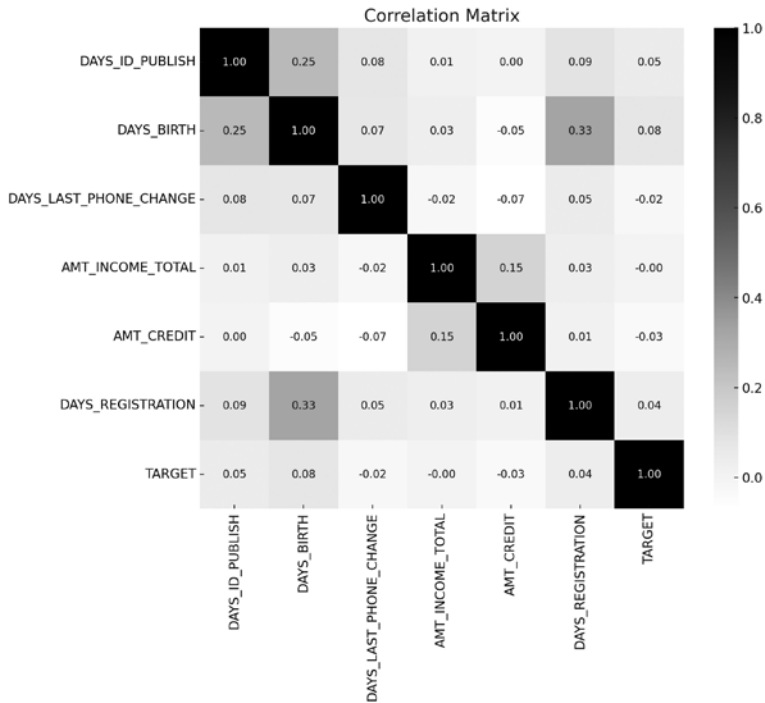
A modellépítés során első körben képezni kellett egy tanító és egy tesztmintát. A tanító mintát 70 százalékban, míg a tesztmintát 30 százalékban határoztuk meg, véletlenszerű kiválasztással. Ezek után a sok változóból kellett kiszűrni azokat a változókat, amelyek megfelelőek lesznek a becslés során. Ehhez a *scorecardpy* python könyvtárat használtuk fel (Shichen, 2023). A könyvtár automatikusan leképezte az összes változóra a csoportokat, és a vágások után az *information value* segítségével szűrést lehetett végezni. Szakértői alapon úgy ítéltük meg hogy a 0,08-as értéket meghaladó *information value*-val rendelkező változókat szeretnénk bent hagyni az elemzésben. Így végül 10 változó került kiválasztásra.

A törlesztőrészlet nagysága és a hitelkártyával megvásárolt termék ára magas korrelációt mutatott a kiadott hitelösszeggel, továbbá erősen korrelált a születés óta eltelt napok száma, illetve a munkahelyen eltöltött napok száma is. A regressziós eljárás során a magas korrelációval rendelkező változók modellben tartása felülreprezentáltságot eredményezhet, így az említett változókat kiszűrtük (Peng et al., 2002).

Az így megmaradt változók segítségével az 1. ábrán látható hőtérképet készítettük el a korreláció tekintetében. Jól látható, hogy a bent maradt változók között nem figyelhető meg sehol sem olyan mértékű korreláció, amely a becslés során torzításokat eredményezne. Fontos megemlíteni, hogy a bent maradt változók és a célváltozó kapcsolata a korrelációs elemzés alapján igen alacsonynak tekinthető, a legmagasabb érték is a 0,08-as korrelációs együtthatót veszi fel, így a modellezési eljárások során valószínűsíthetően a modell magyarázóereje alacsony lesz. A tanulmányban a célunk a modellek hatékonyságának összehasonlítása, így a gyengébb predikációs erő esetén is láthatóvá válik már a hatékonyabb módon becsülő modellek eredménye.

1. ábra

A logisztikus regresszióba kerülő változók korrelációja



Forrás: saját szerkesztés

Látható az is, hogy az idősebb emberek régebb óta rendelkeznek személyi igazolvánnyal, viszont ez az információ sem meglepő, mivel valószínűleg a fiataloknak a helyi szabályozások szerint gyakrabban kell megújítaniuk a személyi igazolványukat. A többi változó közötti kapcsolat szorossága elhanyagolhatónak tekinthető.

A modellezés előtt az adatokat igyekeztük újracsoportosítani, amennyiben az indokolt volt. Az information value tekintetében a születés óta eltelt napok száma adta a legnagyobb magyarázó erőt, ahol az IV 0,08-as értéket vett fel, de nem sokkal maradt el ettől a hitel nagyságáról és a legutóbbi telefonváltás óta eltelt napok számáról szóló információ.

A tanító mintán így a csoportosított változók segítségével építettünk egy logisztikus regressziós modellt, ahol ezek után ellenőriztük a p-értékek mentén, hogy az egyes változók szignifikáns módon magyarázzák-e a bedőlést. 5 százalékos szignifikancia mellett a kiválasztott változók mind szignifikánsnak bizonyultak. A kiértékeléshez – ahogy korábban is bemutattuk – több eljárást is lehet alkalmazni, amelyek közül

az AUC- és az ebből képzett GINI-értéket választottuk. A továbbiakban is ezt fogjuk felhasználni, hiszen ez alkalmas arra, hogy a modelleket a becslési hatékonyság szempontjából összehasonlítsuk, ami választ ad a kutatási kérdésre is.

2. táblázat

A logisztikus regresszióval készített becslés eredményei

Eljárás	GINI	AUC
Logisztikus regresszió	0,26	0,63

Forrás: saját szerkesztés

Amint az a 2. táblázatban látható, a logisztikus regressziós eljárással egy 0,26-os GINI-t tudunk elérni, ami 0,63-as AUC-értéket jelent. Ez nem nevezhető egyébként egy túl erősen becslő modellnek, viszont fontos kiemelni, hogy az adatbázisban főként kis értékű hitelkártya-kihelyezésekről beszélünk, ahol igen alacsony bedőlési arányt tudunk megfigyelni. A modellben továbbá kizárólag applikációs információkat találtunk, így a számlatörténetet és viselkedési adatokat nem tudtuk a mintában elemezni, ami szűkíti a becslési lehetőségeket.

3.4. Mesterségesintelligencia-alapú modellek

A következő részben a mesterséges intelligencia adatfeldolgozó képességének felhasználásával épített modellek eredményeit szeretnénk bemutatni. A probléma elemzésére több modell is készült, de csak azokat részletezzük, amelyek hatékonyabbnak bizonyultak a hagyományos megközelítésnél. Egy rövid módszertani bevezető után részletesen bemutatjuk a becslési hatékonyságra vonatkozó eredményeket.

3.4.1. Mesterséges neurális hálózat (ANN)

A mesterséges neurális háló biológiai alapokból táplálkozik: az idegrendszer is alapvetően egy bonyolult kapcsolt rendszer. Felépítése neuronokon alapszik, amelyek hálózatba vannak rendezve. A jelátvitel szinapszisokon keresztül történik, amikor a neuron dendritjében lévő sejt elektromos potenciálja elér egy küszöbértéket. Nagy jelentősége van annak is, hogy a szinapszisok közötti kapcsolatok erőssége milyen módon tér el (Pruves et al., 2019).

A mesterséges neurális háló működése is bizonyos aktivációs függvények szerinti súlyozáson alapszik. A szimuláció során pedig a hálózat egy súlymátrix kombinációi alapján próbál rátanulni az adatmintázatra. A folyamat során a fentebb említett összes változó jelenti a bemeneti adatokat, így az első, azaz a beme-

neti réteget 64 darab neuron alkotja. A modellezés során érzékelhető volt, hogy a predikciós pontosság vizsgálata esetén fontos befolyásoló tényező a második réteg (első rejtett réteg) neuronszámának bizonyos szintű növelése. Ide 10 darab neuron került végül, mivel ennél nagyobb számnál nem növekedett jelentősen a tanulás minősége. A második rejtett rétegbe csak 5 neuron került, végül pedig egy neuron jelentette a kimeneti réteget. A neuronrétegek teljesen kapcsolt lineáris kombinációkat alkotnak a mesterséges neurális hálózatok esetén.

Fontos azt is bemutatni, hogy a rejtett rétegekben található neuronok milyen aktivációs függvényeket használnak a tanulási mechanizmus során. Ehhez elsődlegesen a súlyozás mögött álló statisztikai folyamatot kell részletezni. A neuronok irányított kapcsolatokkal vannak összekötve. Ezek között a kapcsolatokat ellátjuk egy úgynevezett asszociált $w_{(j,i)}$ súllyal. Ezen súlyok alkotják végül magát a súlymátrixot. Minden egyes neuron (i egység) először a bemeneti kapcsolatok súlyozását végzi el a következő módon:

$$in_i = \sum_{j=0}^n W_{j,i} a_j \quad (2)$$

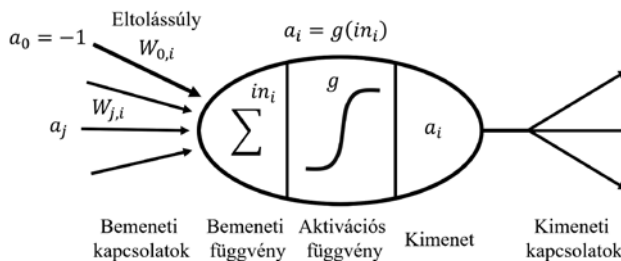
Az a_j érték az aktivációs érték kapcsolatát terjeszti a j -edik egységtől az i -edik felé. A neuronban az aktivációs függvény végzi el a súlyozást. Ennek során az a_0 mint zérus értéket -1 bemenetre állítja a modell, ehhez pedig egy eltolássúlyt ($w_{0,i}$) társít. Az új aktiváció így a következő képlet alapján történik:

$$a_i = g(in_i) = g\left(\sum_{j=0}^n W_{j,i} a_j\right) \quad (3),$$

ahol g jelöli az aktivációs függvényt. Az aktivációs függvény szempontjából fontos, hogy ne lineáris legyen, mivel akkor könnyen alakulhat a modell egy egyszerű lineáris függvénnyé. A neuron működését a 2. ábra foglalja össze.

2. ábra

Egy mesterséges neurális hálózatban részt vevő neuron működésének bemutatása



Forrás: Russal-Norvig (2005)

Az aktivációk szerint a rétegeket különböző módon határoztuk meg: az utolsó két kimeneti rétegben szigmoid aktivációs függvény szerint, míg az első kettőben az úgynevezett ratifikált lineáris egységfüggvény (*rectified linear unit*, továbbiakban: ReLU) szerint súlyozott a modell. A ReLU a negatív bemenetek során 0 értéket ad, a pozitívokat pedig vágatlanul hagyja, míg a szigmoid függvény értelmezhető egyfajta logisztikus függvényként is. A függvények segítségével határoztuk meg az egyes kimeneteket, amelyeket a küszöbpontok szerint állítottunk be. A függvény küszöbpontját az eltolási súlyok aktuális pontja állítja be. Eszerint akkor kerül aktivációra az egység, ha $\sum_{j=0}^n W_{j,i}a_j$ meghaladja a $w_{o,i}$ -t (Russel–Norvig, 2003).

3. táblázat

A mesterséges neurális hálózat (ANN) eljárás eredményei

Eljárás	GINI	AUC
Mesterséges neurális hálózat	0,33	0,67

Forrás: saját szerkesztés

A fentebb említett modell becslési eredményeit az előzőkhöz hasonlóan kiértékeltek a tesztmintán, és a GINI értéke ez esetben 0,33 lett, melyhez 0,67-es AUC-érték társult. Tehát a mesterséges neurálisháló-modell felülmúlta a logisztikus regressziós modell eredményeit. Az eredményeket a 3. táblázatban tüntettük fel. Fontos azt is megemlíteni, hogy az úgynevezett „fekete doboz” megoldások értelmezése, az eredmények magyarázata nagyon nehezen közelíthető meg, ez pedig olyan területen, mint az egészségügy vagy a pénzügyek, igen fontos kérdés (Maheshwari, 2018).

3.4.2. XGBoost

Az XGBoost az *Extreme Gradient Boosting* rövidítése, ahol a *Gradient Boosting* eljárást először Friedman (2001) írta le. Az eljárás alapját a felügyelt tanulás képezi. A felügyelt tanulás során egy klasszikus modellezési problémából indulunk ki, ahol rendelkezésre állnak a képzett magyarázó x_i változók, amelyek segítségével egy célt próbálunk meghatározni (y). Az eljárás során tehát meg kell határozni egy célfüggvényt, amelynek az optimalizálási feladatát látja el az XGBoost-metódus. Az ilyen célfüggvények két komponensből állnak: a tanító adatokon mért becslési veszteségből (*training loss*) és egy regularizációs tényezőből:

$$(\theta) = L(\theta) + \Omega(\theta) \quad (4),$$

ahol L jelöli a becslési veszteséget és Ω a regularizációs tényezőt. A tárgyalt esetben a becslési veszteség mutatja meg, mennyire pontos a modell predikciója, a

regularizációs tényező pedig segít elkerülni a túlillesztést, és szabályozza a modell összetettségét. Gyakran például az RMSE-értéket használják az L felírására, de az eljárás során a célfüggvény módosítható. A bemutatott módszer segítségével összehasonlíthatjuk a döntési fák, a random forest és a boosted forest eljárások teljesítményét, így formális módon tudjuk megközelíteni a paraméterek módosítását.

Az XGBoost-módszer a felügyelt tanulás mellett a döntési fák együttesén (CART) alapul, amelyekkel optimalizációs eljárás segítségével modellt építhetünk, legyen szó klasszifikációról vagy regresszióról. Például meghatározhatjuk, hogy valakinek tetszik-e egy számítógépes játék egy család tagjainak mintája alapján, figyelembe véve életkort, nemet, foglalkozást stb. A CART-modellekben az információk alapján a család tagjait különböző levelekbe sorolhatjuk, ahol a levelek valós pontszámokat tartalmaznak, ami erős eszköz az optimalizáláshoz. Mivel egyetlen fa nem ad elegendő információt, érdemes több fát együtt értékelni. Ahhoz, hogy jobban megértsük a fák alkalmazását, vegyünk egy példát egy két fából álló faegyüttesre. Az egyes fák előrejelzési pontszámait összeadjuk a végső pontszám eléréséhez, miközben a két fa kiegészíti egymást, ami így írható fel:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (5)$$

ahol K a fák száma, f_k egy függvény az F függvénytérben, és F tartalmazza az adott CART összes lehetséges kimenetét. Ahogy korábban részleteztük, meg kell találni a kombinációs halmazon a legpontosabb kimenetet. Ehhez az alábbi optimalizációs célfüggvényt használhatjuk:

$$(\theta) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \omega(f_k) \quad (6)$$

ahol $\omega(f_k)$ a fa összetettségét jelöli az f_k pedig az előzőekben definiált függvényt. Az említett faegyüttesek képzése hasonló a random forest módszerhez, mivel mindkettő faegyüttesek felhasználásán alapszik. A különbség a tanító adatokon történő képzés módjából ered. Ezért egy predikációs függvény mindkét eljárásban alkalmazható.

Természetesen felmerül a kérdés, hogy milyen paraméterek alapján különböztethetjük meg a faegyüttesek képzését, azaz milyen stratégiát alkalmazunk, amikor az f_i függvények tanulását és kiértékelését végezzük egy algoritmussal, amely tartalmazza a fa szerkezetét és a levél pontszámait. A fák struktúrájának tanulása sokkal összetettebb, mint egy egyszerű optimalizálási probléma, ahol csupán a gradienseket vesszük figyelembe, ezért additív stratégiát alkalmazunk. A gyakorlatban megnézzük, hogy eddig mit tanultunk, és hogyan javítható az, ami egy új fa létrehozását eredményezi. Ennek alapján lépésenként felírhatunk egy előrejel-

zési (predikciós) értéket, mint $y_i^{(t)}$. Így a következő módon írhatjuk le a módosított predikciós értéket:

$$\begin{aligned} y_i^{(0)} &= 0 \\ y_i^{(1)} &= f_1(x_i) = y_i^{(0)} + f_1(x_i) \\ y_i^{(2)} &= f_1(x_i) + f_2(x_i) = y_i^{(1)} + f_2(x_i) \\ &\dots \\ y_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = y_i^{(t-1)} + f_t(x_i) \end{aligned} \quad (7)$$

Az additív eljárásnál lényeges, hogy melyik fát válasszuk az egyes lépéseknél. Mindig azt a fát vesszük figyelembe, amelyik optimalizálja a célunkat (ezt adjuk hozzá):

$$\begin{aligned} obj^t &= \sum_{\{i=1\}}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{\{i=1\}}^t \omega(f_i) = \\ &\sum_{\{i=1\}}^n l(y_i, \hat{y}_i^{(t-1)} + f_{-t}(x_i)) + \omega(f_{-t}) + constant \end{aligned} \quad (8)$$

Ahogy korábban részleteztük, egy kitűzött célfüggvényt kell kijelölni, amelyet optimalizálni szeretnénk, és amellyel módosítani kell a fentebb részletezett függvényt. Az optimalizáló lépések mellett meg kell határozni egy regularizációs tényezőt, amely a fa összetettségét hivatott jelölni. Ezt a következő módon írhatjuk fel:

$$f_t(x) = w_{q(x)}, w \in R^T, q: R^d \rightarrow \{1, 2, \dots, T\}, \quad (9)$$

ahol w a levelek pontszámainak vektorát jelöli, míg a q egy olyan függvény, amely minden adatpontot hozzárendel egy megfelelő levélhez. A T a levelek számát adja. A komplexitást így a következő módon adhatjuk meg:

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (10)$$

A fastruktúra átszervezésével egy pontszámot kapunk, amit a szakirodalom strukturális pontszámként (*structure score*) említ. A strukturális értéket röviden így kaphatjuk meg:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (11)$$

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j}{H_j + \lambda} + \gamma T \quad (12)$$

Az egyenletben a w_j értékek függetlenek egymástól, és a legjobb értéke adja a legjobb $q(x)$ struktúrát, amely gyakorlatban azt adja meg, hogy egy adott fa meny-

nyire jól teljesít. Alapvetően egy adott fa struktúrájában a g_i és h_i statisztikákat a megfelelő levelekhez társítjuk, majd ezeket összeadjuk, és a képlet segítségével kiszámítjuk a fa teljesítményét. Ez a pontszám gyakorlatilag a döntési fa tisztasági mutatója, amely figyelembe veszi a modell komplexitását is.

Most, hogy képesek vagyunk mérni egy fa hatékonyságát, ideális esetben felsorolnánk az összes lehetséges fát, majd kiválasztanánk a legjobbat. A gyakorlatban azonban a fa egy-egy szintjét optimalizáljuk, tehát egy ágat két levélre osztva egy döntési fa szennyezettségi mérőszámát határozzuk meg:

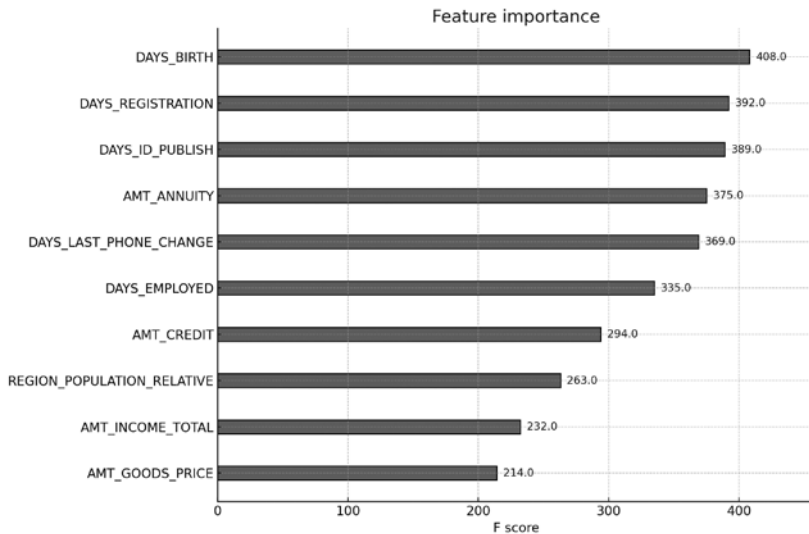
$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G_L^2 + G_R^2}{H_L + H_R + \lambda} \right] - \gamma \quad (13)$$

A képlet értelmezése során az első elem az új bal oldali levél pontszámát, a második elem az új jobb oldali levél pontszámát, a harmadik elem pedig az eredeti levél pontszámát jelenti, végül pedig a levélre vonatkozó regularizációt definiáljuk. Fontos megjegyezni, hogy ha a növekedési érték kisebb, mint γ , akkor nem érdemes a modellt bővíteni az adott ággal. Valós értékű adatok esetén ezzel közelítjük az optimális felosztást, ahol gyakorlatilag egy metszési technikával balról jobbra haladva elvégezzük az optimalizálást (xgboost developers, 2022a). A bemutatott eljárás az optimalizáció segítségével tudja a bemeneti információkat hatékonyan strukturálni, így nem véletlen, hogy az egyik legnépszerűbb módszer az online térben rendezett adattudományi versenyeken (Reinstein, 2017).

Az XGBoost-eljárás felhasználása során az alapvető paraméterbeállítások segítségével végzetük el a becslést, amely elérhető az xgboost developers (2022a) weboldalán is. A modellépítés során meghatározható a változók fontosságának (*feature importance*) értéke is, amely a becsléshez felhasznált információ fontosságának mértékéről ad visszajelzést egy F érték segítségével. A 3. ábrán látható, hogy a legfontosabb változók az életkor, a regisztrációs adatok változása óta eltelt napok száma, a banknál vezetett számla időtartama, a törlesztőrészlet nagysága volt az általunk épített XGBoost-modellben. Kevésbé fontos változóként jelent meg a teljes bevétel értéke, illetve a termékek értéke, amelyet megvásároltak a hitelkártyával, ám fontos kiemelni, hogy ezek is hasznos információt jelentettek az XGBoost-modellben, hiszen ezek is megfelelően magas értéknek számítanak az F érték tekintetében.

3. ábra

Az XGBoost-modellben felhasznált változók fontossági (F score) értékei



Forrás: saját szerkesztés

Az eljárás segítségével összesen 0,366-os GINI-értéket tudunk kapcsolni, amelyhez 0,68-as AUC társult. Az összehasonlítás szempontjából az XGBoost-modell adta a legjobb eredményt, ezzel felülmúlva az eddig épített összes modellt. A vizsgált modellek eredményét a 4. táblázatban tüntettük fel.

4. táblázat

A modellezési eljárások eredményeinek összehasonlítása

Eljárás	GINI	AUC
Logisztikus regresszió	0,26	0,63
Mesterséges neurális hálózat	0,33	0,67
XGBoost	0,37	0,68

Forrás: saját szerkesztés

Ahogy a mesterséges neurális hálózat (ANN) esetén, úgy az XGBoost-eljárás során is elmondható, hogy a modellépítés közben nem szükséges a változók szűrését külön elvégezni, mivel az XGBoost az információhordozó képességén alapuló jellemző fontosság becslést használja a modell tréningje során, ami lehetővé teszi a releváns változók azonosítását (Goodarzi et al., 2009). A gyakorlat azt mutatja,

hogy a változók kiszűrésének szakértői felülbírálatja nem javítja a modell hatékonyságát (xgboost developers, 2022b).

Összehasonlításképpen a logisztikus regressziós modellhez kiválasztott változók felhasználásával is építettünk egy XGBoost-modellt. A logisztikus megközelítéssel kapott 0,26-os GINI-hez képest 0,27-re javult az érték. Láthatóan ezek elmaradnak a teljes változókészlettel épített modell hatékonyságától.

Elmondható, hogy mind az ANN, mind az XGBoost modellezési szempontból jobban optimalizál, mint a WoE-beclsélen alapuló information value és a korrelációs szűrésen alapuló logisztikus regressziós eljárás. Az információkat jobban válogatják ki a gépi tanuláson alapuló modellek, viszont hátrányként megemlíthető, hogy ezek értelmezése sokkal nehezebb feladat, és túlmutat jelen tanulmány keretein.

4. ÖSSZEGZÉS

Tanulmányunkban bemutattuk, hogy a hitelkockázat modellezése a mai napig egy folyamatosan és intenzíven fejlődő tudományterület. Az elmúlt évtizedekben a számítástechnika fejlődésével a szakértői megközelítést egyre inkább felváltották az empirikus adatok tömegén alapuló, hagyományos statisztikai modellek.

A jelenleg is zajló, mesterséges intelligencia által fémjelzett technológiai forradalom a hitelkockázat területén a modellek egy új generációját indította útjára, amelyeket összefoglalóan gépi tanuláson alapuló modelleknek nevezünk.

A tanulmány egy hitelkártya-adatbázist felhasználva hasonlítja össze az applikációs (PiT) PD-modellezési eljárásokat a banki gyakorlatban szinte egyeduralgó, hagyományos módszertanok és a gépi tanulási modellek mentén. Empirikus vizsgálatunkban a logisztikus regressziós megközelítés esetében az információk értéke és a korreláció mentén végzett változósűrések után 7 változót választottunk ki. A változókat a WoE-módszertan szerint csoportosítottuk, és végül ezen értékekből képzett adatok segítségével futtattunk logisztikus regressziós eljárást a modellezés során.

A mesterséges intelligencián alapuló módszertanok közül a Mesterséges Neurális Hálózat (ANN) modell esetén a GINI együttható értéke szignifikánsan javult a hagyományos logisztikus regressziós beclséssel elért eredményekhez képest. Az általunk leghatékonyabbnak ítélt modellt végül az XGBoost-algoritmus adta.

Ahogy bemutattuk, az XGBoost és az ANN esetén nem szükséges feltétlenül a változók mentén előszűréseket végezni, hiszen a modell a tanulási mechanizmusa során automatikusan felülreprezentálja a fontos információkat. Ez is okozhatja,

hogy további hasznos információk segítségével a gépi tanulási módszerek hatékonysága erősebb a logisztikus regressziós megközelítésnél.

A gépi tanulási módszerek kiváló predikációs képessége mellett hátrányként meg kell említeni, hogy a döntési modellek értelmezése – annak fekete doboz jellegéből adódóan – nehezebb az iparági best practice-nek számító, hagyományos megközelítésekhez képest. Ennek kezelésére számos olyan programcsomag jelenik meg folyamatosan, amely egyszerűsíti a döntési mechanizmusok értelmezését a bonyolultabb modellek esetén. Ilyen csomag például a SHAP, amely az interakciós értékek segítségével tudja egyszerűsíteni az XGBoost-algoritmus döntési mechanizmusát (Lundberg, 2018). Véleményünk szerint az algoritmusok, valamint a modellértelmezési programcsomagok fejlődésével, illetve az elérhető információk bővülésével egyre inkább átveszik majd a vezető szerepet a gépi tanulási módszerek a hitelkockázati modellezés területén.

Végezetül szakpolitikai ajánlásunkat szeretnénk megfogalmazni a szabályozó és felügyeleti hatóságok számára. A mesterséges intelligencián alapuló gépi tanulási módszerek fekete doboz jellege miatt a bankoknak jelenleg nincs lehetőségük azok alkalmazására, a szabályozói és felügyeleti elvárásrendszer több pontján is problémát okoznak az eredmények és a döntési mechanizmus értékelési nehézségei. Javasolnánk a jelenlegi – hagyományos modellekre szabott – szabályok és ajánlások újragondolását olyan módon, hogy az teret adjon a bankoknak a gépi tanulási modellek alkalmazására. Az így elérhető – és a tanulmányunkban is bemutatott – hatékonyságnövelés hozzájárulhat a prudensebb működéshez, amely minden érintettnek egységesen elemi érdeke.

HIVATKOZÁSOK

- Addo, P. M. – Guegan, D. – Hassani, B. (2018): Credit risk analysis using machine and deep learning models. *Risks*, 6(2), 38, <https://doi.org/10.3390/risks6020038>.
- Angelini, E. – Di Tollo, G. – Roli, A. (2008): A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48(4), 733–755, <https://doi.org/10.1016/j.qref.2007.04.001>.
- Altman, E. I. – Saunders, A. (1997): Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21(11–12), 1721–1742, [https://doi.org/10.1016/S0378-4266\(97\)00036-8](https://doi.org/10.1016/S0378-4266(97)00036-8).
- Basel Committee on Banking Supervision (2023): Calculation of RWA for credit risk – CRE36: IRB approach: minimum requirements to use IRB approach.
- Butaru, F. – Chen, Q. – Clark, B. – Das, S. – Lo, A. W. – Siddique, A. (2016): Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72, 218–239, <https://doi.org/10.1016/j.jbankfin.2016.07.015>.
- European Banking Authority (2017): Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures.

- European Banking Authority (2020): On management and supervision of ESG risks for credit institutions and investment firms – EBA/REP/2021/18.
- Friedman, J. H. (2001): Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232, <https://doi.org/10.1214/aos/1013203451>.
- Goodarzi, M. – Deshpande, S. – Murugesan, V. – Katti, S. B. – Prabhakar, Y. S. (2009): Is feature selection essential for ANN modeling? *QSAR & Combinatorial Science*, 28(11–12), 1487–1499, <https://doi.org/10.1002/qsar.200960074>.
- Gurný, P. – Gurný, M. (2013): Comparison of credit scoring models on probability of default estimation for us banks. *Prague Economic Papers*, 22(2), 163–181, <https://doi.org/10.18267/j.pap.446>.
- Henisz, W. J. – McGlinch, J. (2019): ESG, Material Credit Events, and Credit Risk. *Journal of Applied Corporate Finance*, 31(2), 105–117, <https://doi.org/10.1111/jacf.12352>.
- Huang, C. L. – Chen, M. C. – Wang, C. J. (2007): Credit scoring with a data mining approach based on support vector machines. *Expert systems with applications*, 33(4), 847–856, <https://doi.org/10.1016/j.eswa.2006.07.007>.
- Kaggle (n. a.): Loan Defaulter Dataset. https://www.kaggle.com/datasets/gauravduttakiit/loan-defaulter?select=application_data.csv (letöltve: 2023.04.01.).
- Kovács, L. – Marsi, E. (szerk.) (2018): *Bankmenedzsment – banküzemtan*. Budapest: Magyar Bankszövetség. ISBN: 978-963-89653-2-5.
- Lundberg, S. (2018): Basic SHAP Interaction Value Example in XGBoost. https://shap.readthedocs.io/en/latest/example_notebooks/tabular_examples/tree_based_models/Basic%20SHAP%20Interaction%20Value%20Example%20in%20XGBoost.html (letöltve: 2023.06.11.).
- Madar, L. (2015): Scoring rendszerek hatásai a gazdasági tőkeszámítás során alkalmazott portfólió-modellek eredményeire (doktori disszertáció, Kaposvári Egyetem).
- Magyar Nemzeti Bank (2002): Tanulmányok a bankszektor középtávú fejlődési irányairól. *MNB Műhelytanulmányok*, 26.
- Magyar Nemzeti Bank (2018): ICAAP-ILAAP-BMA kézikönyv. Budapest: Magyar Nemzeti Bank. <https://www.mnb.hu/letoltes/icaap-ilaap-bma-kezikonyv-2018-januar.pdf>.
- Maheshwari, S. (2018): The Explainable Neural Network. Elérési link: <https://medium.com/@shagunm1210/the-explainable-neural-network-8f95256dcddb> (letöltve: 2023.05.25.).
- Medema, L. – Koning, R. H. – Lensink, R. (2009): A practical approach to validating a PD model. *Journal of Banking & Finance*, 33(4), 701–708, <https://doi.org/10.1016/j.jbankfin.2008.11.007>.
- Moradi, S. – Mokhtab Rafiei, F. (2019): A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, 5(1), 1–27, <https://doi.org/10.1186/s40854-019-0121-9>.
- Nagy G. – Biró G. (2018): Az IRB PD paraméterbecslés PIT- és TTC-problematikája a felügyeleti felülvizsgálatok tükrében. *Gazdaság és Pénzügy*, 5(3), 258–285.
- Pedregosa, F. – Varoquaux, G. – Gramfort, A. – Michel, V. – Thirion, B. – Grisel, O. (2011): Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12 (Oct), 2825–2830.
- Peng, C. Y. J. – Lee, K. L. – Ingersoll, G. M. (2002): An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14, <https://doi.org/10.1080/00220670209598786>.
- Purves, D. – Augustine, G. J. – Fitzpatrick, D. – Hall, W. – LaMantia, A. S. – White, L. (2019): *Neurosciences. De Boeck Supérieur*.
- Reinstein, I. (2017): XGBoost: A Top Machine Learning Method Explained in Kaggle Competitions. KDnuggets. <https://www.kdnuggets.com/2017/10/xgboost-top-machine-learning-method-kaggle-explained.html> (letöltve:2023.05.21.).

- Rossum, G. – Drake, F. L. (2009): *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Rozo, B. J. G. – Crook, J. – Andreeva, G. (2023): The role of web browsing in credit risk prediction. *Decision Support Systems*, 164, 113879, <https://doi.org/10.1016/j.dss.2022.113879>.
- Russell, S. – Norvig, P. (2005): Artificial intelligence. A modern approach. – MI. Mesterséges intelligencia – Tananyagbővítés – Hungarian Transalation. Panem Könykiadó, Budapest. <http://project.mit.bme.hu/> (letöltve: 2023.05.25.).
- Shichen, X. (2023): Credit Risk Scorecard- scorecardpy 0.1.9.6 (Python könyvtár). <https://github.com/ShichenXie/scorecardpy>.
- Siddiqi, N. (2006): Credit Risk Scorecards. NJ, Hoboken: John Wiley & Sons.
- Sirignano, J. – Sadhwani, A. – Giesecke, K. (2016): Deep Learning for Mortgage Risk. <https://arxiv.org/abs/1607.02470>.
- Somerville, R. A. – Taffler, R. J. (1995): Banker judgement versus formal forecasting models: The case of country risk assessment. *Journal of Banking & Finance*, 19(2), 281–297, [https://doi.org/10.1016/0378-4266\(94\)00051-4](https://doi.org/10.1016/0378-4266(94)00051-4).
- Venkatesh, A. – Jacob, S. G. (2016): Prediction of Credit-Card Defaulters: A Comparative Study on Performance of Classifiers. *International Journal of Computer Applications* (0975–8887), 145(7), <https://doi.org/10.5120/ijca2016910702>.
- xgboost developers (2022a): Introduction to boosted trees. <https://xgboost.readthedocs.io/en/stable/tutorials/model.html> (letöltve: 2023.05.20.).
- xgboost developers (2022b): Understand your dataset with XGBoost. <https://xgboost.readthedocs.io/en/stable/R-package/discoverYourData.html> (letöltve: 2023.06.10.).
- Wang, K. – Li, M. – Cheng, J. – Zhou, X. – Li, G. (2022): Research on personal credit risk evaluation based on XGBoost. *Procedia Computer Science*, 199, 1128–1135, <https://doi.org/10.1016/j.procs.2022.01.143>.
- West, D. (2000): Neural network credit scoring models. *Computers & Operations Research*, 27(11–12), 1131–1152, [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5).
- Yu, Y. (2020): The Application of Machine Learning Algorithms in Credit Card Default Prediction. In 2020 International Conference on Computing and Data Science (CDS), 212–218. *IEEE*, <https://doi.org/10.1109/cds49703.2020.00050>.