Hagyományos versus MI-alapú csalásfelderítés: Költséghatékonyság a gépjármű-biztosítások területén*

Benedek Botond – Nagy Bálint Zsolt

A vállalati gyakorlat és a különböző iparági jelentések mind azt mutatják, hogy a gépjárműbiztosítási csalások igen gyakoriak, éppen ezért a hatékony csalásfelderítés igencsak fontos. Tanulmányunkban azt vizsgáljuk, hogy a napjainkban elterjedt MI-alapú csalásfelderítő módszerek pénzügyi (költséghatékonysági) szempontból hatékonyabbak-e, mint a hagyományos statisztikai-ökonometriai eszközökön alapuló módszerek. Eredményeink alapján arra a nem várt következtetésre jutottunk, hogy a jelenleg a szakirodalomban megtalálható MI-alapú és valós adatbázison tesztelt gépjármű-biztosítási csalásfelderítési módszerek kevésbé költséghatékonyak, mint a hagyományos statisztikai-ökonometriai módszerek.

Journal of Economic Literature (JEL) kódok: G22, C14, C45

Kulcsszavak: gépjármű-biztosítás, biztosítási csalás, csalások felderítése, költségérzékeny döntéshozatal, adatbányászat

1. Bevezetés

A biztosítási csalás következményei komoly hatással vannak a biztosítási ágazatra. A csalás bizalmatlanságot kelt az iparággal szemben, gazdasági károkat okoz, és befolyásolja az általános megélhetési költségeket. Az egyesült államokbeli Biztosítási Információs Intézet[1] (*III 2021*) jelentése szerint a biztosítási csalások összköltsége az Egyesült Államokban 2015 és 2019 között évi 38 és 83 milliárd dollár között volt. Ez azt jelenti, hogy egy átlagos amerikai családnak a biztosítási csalások miatt évente 800–1 400 dollár többletkiadása keletkezik. A Brit Biztosítók Szövetsége[2] kiemeli, hogy 2020-ban az Egyesült Királyságban feltárt jogtalan kártérítési igények értéke 1,1

* A jelen kiadványban megjelenő írások a szerzők nézeteit tartalmazzák, ami nem feltétlenül egyezik a Magyar Nemzeti Bank hivatalos álláspontjával.

Benedek Botond: Babeş-Bolyai Tudományegyetem, adjunktus. E-mail: botond.benedek@econ.ubbcluj.ro
Nagy Bálint Zsolt: Babeş-Bolyai Tudományegyetem, egyetemi docens. E-mail: balint.nagy@econ.ubbcluj.ro

A kutatás a Magyar Tudományos Akadémia 86/12/2022/HTMT számú tudományos kutatási ösztöndíjának támogatásával valósult meg.

A magyar nyelvű kézirat első változata 2022. november 2-án érkezett szerkesztőségünkbe.

DOI: https://doi.org/10.25201/HSZ.22.2.77

[1] Insurance Information Institute (III)
[2] Association of British Insurers (ABI)

milliárd font volt (*ABI 2021*). Ha konkrétan a gépjármű-biztosítási csalásokat nézzük, az USA-ban és Nyugat-Európában 7–10 százalék, a kelet-közép-európai régiókban 10–20 százalék, míg Kínában a biztosítási kötvények 18–20 százaléka érintett (*ABI 2021; III 2019*).

Kétségtelen tehát, hogy a biztosításicsalás-azonosítás egy gazdaságilag igen fontos vizsgálódási terület. Tanulmányunkban 24 olyan tudományos folyóiratcikket és 3 konferenciaanyagot elemeztünk a gépjármű-biztosítási csalások felderítésével kapcsolatban, amelyeket a Web of Science adatbázisa indexelt 1990 és 2022 között. Ezek alapján megállapíthatjuk, hogy ez a kutatási terület még igencsak kidolgozatlan. Alapos szakirodalma van a klasszikus statisztikai-ökonometriai csalásazonosító módszereknek, valamint a mesterséges intelligencián (MI) és gépi tanuláson alapuló modelleknek, ám hiányzik egyrészt ezek szisztematikus összehasonlítása, másrészt pedig a csalásazonosítás költséghatékonyságának vizsgálata. A magyar nyelvű szakirodalom szintén hiányos e területen, nincs általánosan elfogadott definíciója a biztosítási csalásnak, ahogy a jogtalan gépjármű-biztosítási követelésnek sem.

Éppen ezért jelen tanulmány három területen kíván hozzájárulni a gépjárműbiztosítási csalásazonosítással kapcsolatos tudásunk fejlesztéséhez:

- A nemzetközi és magyar szakirodalom átfogó elemzése után amellett érvelünk, hogy bármely csalásazonosító rendszer teljesítményét költséghatékonyságának tükrében kell megítélni. Más tudományterületeken, ahol elterjedtté vált a mesterséges intelligencia (MI) alkalmazása, pl. az egészségügyben, ezek a költséghatékony megközelítések már dominánssá váltak (*Lee et al. 2017*; *Hill et al. 2020*).
- Tekintettel a szakirodalomban elérhető számos MI-módszer elterjedésére, úgy gondoljuk, égető szükség van egy szisztematikus metatanulmányra, amely képes bemutatni e modellek rangsorát, és összehasonlítani azokat pénzügyi teljesítményük szempontjából. Nem utolsósorban, nem találtunk olyan tanulmányt, amely azt vizsgálta volna, hogy a napjainkban elterjedt MI-alapú módszerek pénzügyi (költséghatékonysági) szempontból hatékonyabbak-e, mint a hagyományos statisztikai-ökonometriai eszközökön alapuló módszerek.
- Végül pedig a témához kapcsolódó (kvázi nem létező) magyar nyelvű szakirodalomhoz szeretnénk hozzájárulni, minimum olyan általánosan elfogadható definíciókkal, melyekből az olvasónak egyértelművé válik, hogy mi a biztosítási csalás vagy épp a jogtalan gépjármű-biztosítási követelés.

Miután áttekintjük a releváns szakirodalmat (*második rész*), részletesen bemutatjuk elméleti keretünket (*harmadik rész*) a költségmegtakarításra vonatkozó, *Benedek és szerzőtársai* (*megjelenés alatt*) által javasolt számítási módszerrel együtt. A *negyedik részben* a kiválasztott csalásazonosító módszerekre és költséghatékonyságukra koncentrálunk, és összehasonlítjuk a hagyományos statisztikai és a gépi tanuláson

alapuló csalásfelderítési módszereket egymással, majd bemutatjuk a hőtérképekkel végzett, részletes érzékenységelemzésünk eredményeit. Az utolsó részben levonjuk a következtetéseket.

2. Szakirodalmi áttekintés

Szakirodalmi áttekintésünket azzal kezdjük, hogy meghatározzuk, hogy a továbbiakban mit értünk biztosítási csalás és jogtalan gépjármű-biztosítási követelés alatt. A Jogi Információs Intézet[3] (*LII 2023*) meghatározása és Massachusetts állami jogszabályok értelmében[4] (melyek az angol nyelvű szakirodalomban a legszélesebb körben elfogadottak) biztosítási csalásnak tekinthető minden olyan cselekedet, amelyet azzal a szándékkal hajtanak végre, hogy a biztosítótól jogtalan kifizetésre tegyenek szert. A rendőrség és az ügyészség általában a biztosítási csalások két formáját különbözteti meg: (1) a biztosított jószágban történő szándékos károkozást (hard fraud) és (2) az okirat-hamisítást (soft fraud). A biztosított jószágban történő szándékos károkozás a két forma közül a ritkább, akkor beszélhetünk róla, amikor az elkövető szándékosan vagyonmegsemmisítést idéz elő azzal a céllal, hogy később szert tegyen a kártérítési összegre. Okirat-hamisításról akkor beszélhetünk, amikor a szerződő eltúlozza az egyébként jogos követelést, vagy amikor valótlant állít és/vagy elhallgat bizonyos feltételeket és körülményeket. Ha kimondottan a gépjármű-biztosítási piacot vizsgáljuk, jogtalan követelésről akkor beszélünk, ha (1) a biztosított olyan baleset miatt nyújt be kárigényt, amely meg sem történt, (2) több kárigényt nyújt be egyetlen baleset kapcsán, (3) kárigényt nyújt be nem az autóbaleset során keletkezett károk fedezésére, (4) hamisan jelenti be a sérülések miatti bérkieséseket/egészségügyi kezelések költségét vagy (5) magasabb autójavítási költségeket jelent be, mint amennyibe a javítás ténylegesen került (*LII 2023; MR 1993*).

2.1. Nemzetközi szakirodalom

Az egyik legkorábbi tanulmányban *Weisberg és Derrig* (*1991*) a lehetséges csalási mutatókat (red flags) relatív gyakoriságuk alapján szedte listába. Ebben a testi sérülésekkel kapcsolatos biztosításokhoz tartozó kártérítési igények (65 lehetséges jellemzőjéből) 18 objektív jellemzőt használtak fel a jogtalan követelések azonosítására. Ennek ellenére az alkalmazott módszer egyszerűsége csupán korlátozott sikert eredményezett. A csalási mutatók és a jogtalan követelések besorolási problémájának *Derrig és Ostaszewski* (*1995*) vizsgálatából kiderül az is, hogy az egyes szakértők részéről nincs egyetértés a jogtalan követelések tekintetében. Ezért a biztosítók számára fuzzy alapú besorolási technikát javasolnak. *Weisberg és Derrig* (*1998*) tesztelte a lehetséges csalási mutatók hasznosságát, számszerűsítette a standard

[3] Legal Information Institute (LII)
[4] Massachusetts Regulation (MR)

vizsgálati technikák hatékonyságát, és feltérképezte a vállalatok képességét a csalások további kinyomozására.

Belhadji et al. (*2000*) olyan „szakértői rendszert" mutatott be, amely segíti a biztosítótársaságok alkalmazottait a döntéshozatalban. Az eszköz közvetlenül nem alkalmazható egy adott biztosítóra, mert a felhasznált paraméterek iparági adatokból készült számításokból származnak, ennek ellenére fontos lépést jelentett a napjainkban elterjedt adatbányászati és mesterséges intelligencia alapú csalásfelderítési modellek irányába.

Az *Artís et al.* (*1999*; *2002*) által bemutatott újszerű megközelítés (diszkrét választású modell) már azt tesztelte, hogy a biztosított tulajdonságai és a baleset körülményeinek jellemzői milyen hatást gyakorolnak a csalás elkövetésének valószínűségére. Emellett ezek a kutatások már nagy hangsúlyt fektettek a téves besorolási problémára is. Az alkalmazott modell jellege és a valós gépjármű-biztosítási adatsorok jellemzői miatt a jogtalan kárigényeket felül kellett súlyozzák a becslésben. Ezzel megnyitották az utat az aszimmetrikus adatsorok (mint például a gépjármű-biztosítási csalások) különböző túl- vagy alulsúlyozási technikákkal történő vizsgálatához. Ezzel párhuzamosan *Viaene et al.* (*2002*) összehasonlította a különböző csalásfelderítési módszerek teljesítményét. A tanulmány készítői csak az anyagi károkra vonatkozó indikátorokat használták, mert a vizsgálati folyamat korai szakaszában csak ezek állnak rendelkezésre.

Miután *Artís et al.* (*1999*; *2002*) megnyitotta a kaput a túl- vagy alul-mintavételezési technikák előtt, és *Viaene et al.* (*2002*) bevezette a korai stádiumú indikátorok használatát, számos szerző bemutatott valamilyen túl- vagy alul-mintavételezésen alapuló osztályozási módszert (főleg anyagi károk esetén). Például *Pérez et al.* (*2005*) összehasonlította a konszolidált fa algoritmusának teljesítményét a jól ismert C4.5 algoritmusok teljesítményével egy túl-mintavételezett valós autóbiztosítási adatbázison. *Bermúdez et al.* (*2008*) egy aszimmetrikus logit modellt javasolt, amely képes volt kezelni a kiegyensúlyozatlan adatkészleteket. Néhány évvel később a kutatók két új megközelítést javasoltak a többségi osztály alul-mintavételezésére a kiegyensúlyozatlan adatkészletekben az osztályozók teljesítményének javítására. Az első megközelítésben *Sundarkumar et al.* (*2015*) az egyosztályú szupport vektorgép (one-class support vector machine, OCSVM) alapú alul-mintavételezést, míg a második megközelítésben *Sundarkumar – Ravi* (*2015*) a k-fordított legközelebbi szomszédság (k-nearest neighbour, KNN) és az OCSVM együttes alkalmazását javasolta.

Šubelj et al. (*2011*) újszerű szakértői rendszert mutatott be a közösségi hálózatok elemzésével, melynek célja az volt, hogy csalók csoportjait azonosítsa, és ne csupán néhány elszigetelt gépjármű-biztosítási csalást. *Farquad et al.* (*2012*) módosított aktív tanuláson alapuló megközelítést alkalmazott annak érdekében, hogy „ha…, akkor" típusú szabályokat állítsanak fel egy szupport vektorgép „fekete dobozából"

az ügyfélkapcsolat-kezeléshez. *Gepp et al.* (*2012*) a döntési fát, a túlélési elemzést és a diszkriminanciaanalízis módszertanát hasonlította össze a *Wilson* (*2009*) által alkalmazott logisztikus regresszióval. A *Tao et al.* (*2012*) által javasolt megközelítés újdonsága abban rejlett, hogy minden biztosítási igényt egyszerre lehetett két kategóriába (jogos és jogtalan) sorolni két különböző valószínűséggel.

Yan – Li (*2015*) a gépjármű-biztosítási csalások felderítését a kiugró értékek és jelenségek (outliers) észlelési problémájaként közelítette meg. Ezért egy továbbfejlesztett kiugróérték-azonosítási módszert javasoltak, amely a legközelebbi szomszéd algoritmus visszametszési (pruning) szabályokkal kiegészített változatán alapul. *Nian és szertőtársai* (*2016*) egy nem felügyelt spektrális rangsorolási (Spectral ranking algorithm, SRA) módszert javasoltak az anomáliák kimutatására. *Shaeiri és Kazemitabar* (*2020*) továbbfejlesztette az SRA-megközelítést, és bemutatott egy olyan implementációs módszertant, amely lehetővé teszi az SRA valós idejű alkalmazását nagy adatkészleteken. *Li et al.* (*2018*) az egyes osztályozókat több osztályozó rendszerbe egyesítette a klasszifikációs pontosság növelése érdekében. *Wang és Xu* (*2018*) mély neurális hálózaton és látens Dirichlet-allokáción (LDA) alapuló szövegelemzést javasolt.

Végül néhány szerző szigorúan pénzügyi szempontból közelítette meg a gépjármű-biztosítási csalások felderítésének problémáját, és nagy hangsúlyt fektetett a költségérzékeny kárbesorolásra. Például *Phua és szerzőtársai* (*2004*) összehasonlították az általuk javasolt megközelítés teljesítményét különböző, széles körben használt technikákkal, és bizonyították a javasolt módszer jobb teljesítményét a költségmegtakarítás szempontjából. *Viaene et al.* (*2007*) a hibaarány (téves besorolás) minimalizálása helyett a vizsgálati folyamat költségeire összpontosított, és kimutatta, hogy a költségérzékeny csalásszűrés nyereséges megközelítés lehet a vagyon- és balesetbiztosító társaságok számára. Végül *Zelenkov* (*2019*) is egy költségérzékenység alapú megközelítést javasolt, de egy példafüggő, költségérzékeny metaalgoritmust, az AdaBoost-ot (adaptive boosting), amely nemcsak a különböző besorolási hibákhoz (mint a korábbi tanulmányokban), hanem az egyes kártérítési esetekhez is eltérő költségeket rendelt.

A kapcsolódó nemzetközi szakirodalom átfogóbb ismertetése, beleértve a legfontosabb a csalás azonosítására használt mutatókat, a leggyakrabban használt adatbázisokat, illetve a csalásazonosítással kapcsolatos legaktuálisabb kihívások megtalálhatók *Benedek et al.* (*2022*) munkájában.

2.2. Magyar nyelvű szakirodalom

A magyar nyelvű szakirodalomból teljes mértékben hiányzik a csalásfelderítő módszerek alkalmazása, sőt egyáltalán a biztosítási csalás mint tudományos kutatási téma. Ebben a vonatkozásban mindenképpen premier értékű a jelen tanulmány.

Mivel teljesen hiányzik a biztosítási csalás tudományos kutatása magyar nyelven, röviden áttekinthetünk néhány olyan magyar nyelvű irodalmat, ahol egyáltalán a mesterséges intelligencia és gépi tanulás módszereit alkalmazzák gazdasági-pénzügyi problémákra.

Az első gazdasági-pénzügyi MI-alkalmazások a vállalati csődelőrejelzési modellek terén láttak napvilágot: logisztikus regressziók és faktoranalízis kombinációját alkalmazta *Hámori* (*2001*), akinek modelljében az osztályozási pontosság 95,3 százalék volt. *Virág – Kristóf* (*2005*) neurális háló alapú modellt alkalmazott csődelőrejelzésre, felhasználva a több neurális réteg (4) és a backpropagation algoritmus nyújtotta előnyöket. A neurális hálóval elért eredmények pontossága néhány százalékponttal meghaladta a lineáris diszkriminanciaanalízis és a logisztikus regresszió segítségével elért eredményeket. *Virág – Nyitrai* (*2013*) alkalmazta először a szupport vektorgép (SVM) módszert magyar vállalatok adatain. Különböző kernel függvényeket alkalmazva az SVM-el 5 százalékkal jobb teljesítményt értek el, mint neurális hálókkal. *Virág – Nyitrai* (*2014*) metamódszerek (ensemble methods), az AdaBoost és a bagging (bootstrap aggregating) teljesítményét hasonlította össze C4.5 döntési fák alkalmazásával közel ezer magyar vállalat 2001 és 2012 közötti adatait felhasználva. Eredményeik szerint a bagging nyújtotta a jobb teljesítményt, de nagyon kevéssel előzve meg az Adaboostot. A frissebb alkalmazások közül megemlítjük *Ágoston* (*2022*) tanulmányát, amely SVM, bagging és véletlen erdő (random forrest) algoritmusokat alkalmaz csődelőrejelzésre budapesti és pécsi városrégiókhoz tartozó cégek mintáját felhasználva. A mintán kívüli osztályozási mutatók pontossága alapján a véletlen erdő bizonyul győztesnek.

A csődelőrejelzésen kívüli, de gazdaságon belüli MI-tanulmányok közül érdemes megemlíteni még a következőket: *Muraközy* (*2018*) amellett érvel, hogy a predikciót előtérbe helyező gépi tanulás és az oksági viszonyrendszereket tanulmányozó ökonometria nem egymást helyettesítő, hanem inkább kiegészítő empirikus tudományágak. *Farkas et al.* (*2020*) a gépi tanulás felhasználási lehetőségeiről értekezik a mezőgazdaságban. A mesterséges intelligencia alkalmazását a vállalati gazdaságtan területein (menedzsment, marketing) is tetten érhetjük: *Benedek* (*1999*) a marketingakciók hatékonyságát elemzi statisztikai és adatbányászati módszerekkel, míg *Danyi* (*2019*) a mesterséges intelligencia alkalmazásának várható hatásait szemlézi az árazási politikákban és stratégiákban. *Bánkúty-Balog* (*2020*) a nemzetközi versenyképesség kontextusában méri fel a MI elterjedésének geoökonómiai hatásait Magyarországra nézve. Végezetül *Csillag et al.* (*2022*) a gépi tanuláson alapuló strukturális témamodellezés (STM) módszerével értékelte a környezeti témák médiában való előfordulását.

3. Fogalmi és elméleti háttér

A gépjármű-biztosítási csalások azonosítása egy bináris osztályozási probléma, ezért bármely osztályozó algoritmus teljesítménye leírható az *1. táblázatba* foglalt találati mátrixszal[5].

1. táblázat
Bináris osztályozó találati mátrixa és az értékelésben használt teljesítménymutatók

		Várható érték		Teljesítmény-mutatók
		Jogtalan követelés	Jogos követelés	
Tényleges érték	Jogtalan követelés	Valódi pozitív (*TP*)	Álnegatív (*FN*)	**Szenzitivitás (TPR):** $\frac{TP}{TP+FN}$
	Jogos követelés	Álpozitív (*FP*)	Valódi negatív (*TN*)	**Specificitás (TNR):** $\frac{TN}{FP+TN}$
Teljesítmény-mutatók		**Precizitás (PPV):** $\frac{TP}{TP+FP}$	**Negatív prediktív érték (NPV):** $\frac{TN}{TN+FN}$	**Becslési pontosság (ACC):** $\frac{TP+TN}{TP+FP+TN+FN}$
		F-mérték: $\frac{(1+\beta^2)*TPR*PPV}{\beta^2*TPR+PPV}$		

Megjegyzés: F-mérték esetében β a precizitás és a szenzitivitás relatív fontosságának beállítására szolgáló együttható.

A találati mátrixból különféle teljesítménymutatók vezethetők le. Az osztályozó teljesítményének legszélesebb körben használt mérőszámai a becslési pontosság (accuracy, ACC), a szenzitivitás (sensitivity, TPR) a specificitás (specificity, TNR) és az F-mérték (F-score). Azonban ezeknek a mérőszámoknak is megvannak a maguk korlátai, főleg olyan aszimmetrikus adathalmazokon, mint a gépjárműbiztosítási csalások. Az egyes teljesítménymutatók részletes ismertetése, valamint az esetleges korlátok bemutatása megtalálható *Benedek és szerzőtársai* (*megjelenés alatt*) munkájában.

Mindazonáltal vállalati szempontból a teljesítménymutatókkal kapcsolatos összes probléma leküzdésének egyik lehetséges módja az egyes osztályozók működési költségeinek számszerűsítése a különböző osztályozók teljesítményének vizsgálata helyett. Ez a megközelítés egyszerű összehasonlíthatóságot tesz lehetővé, és figyelembe tudja venni a különféle téves jelzésekből származó költségeket.

[5] A találati mátrixok módszertana és elmélete egészen *Green – Swets* (*1966*) munkájáig vezethető vissza.

Ráadásul a legtöbb biztosító sokkal fontosabbnak tartja a felderítési folyamatból eredő költségek minimalizálását, mint az osztályozó hibaarányának minimalizálását.

Egy (fél)automatizált csalásfelderítési rendszer költségmegtakarításának számszerűsítéséhez két meghatározó tényezőt kell figyelembe venni: (1) a rendszerek folyamatos használatából adódó költséget és (2) az alternatív rendszer működési költségét. A rendszerek folyamatos használatából adódó költségek egy része a csaláselemző részleg új feladatainak ellátásához szükséges munkaerő költsége. A legfontosabb tétel azonban ezen a téren a rendszer téves jelzéséből származó költség. Ha egy jogos követelést jogtalannak minősít a rendszer, a biztosító fizet a szükségtelen nyomozásért (hiszen a rendszer csak megjelöli a potenciális jogtalan követelést, de ezt egy szakértőnek ellenőrizni és bizonyítani kell). Hasonlóképpen, ha egy jogtalan követelést jogosnak minősít a rendszer, a biztosító fizet a csalónak. Figyelembe véve a biztosítók által feldolgozott nagyszámú kárigényt, a rendszer téves jelzéséből származó költségek igen jelentősek lehetnek. Az alternatív rendszer működési költségeinek meghatározásakor *Phua et al.* (*2004*) azt javasolja, hogy számoljunk azzal az alternatívával, amikor a biztosító nem tesz lépéseket a követelések jogszerűségének ellenőrzése érdekében, és egyszerűen kifizeti az összes igényelt kártérítést. Tehát a *Phua et al.* (*2004*) által javasolt (1) egyenlettel adott megközelítés bármely rendszer költségmegtakarításának (CSDM – cost saving of the decision method) számszerűsítésére a következő:

$$CSDM = NA - (MC + FAC + NC + HC) \tag{1}$$

ahol *NA* a „no action cost", azaz annak az alternatívának a költsége, amikor a biztosító nem tesz lépéseket a követelések jogszerűségének ellenőrzésére. Továbbá az álnegatívak költsége (misses cost, MC), az álpozitívak költsége (false alarms cost, FAC), a valódi negatívak költsége (normals cost, NC) és a valódi pozitívak költsége (hits cost, HC) a következő:

$$MC = NFN * ACA;$$

$$FAC = NFP * (ACI + ACA);$$

$$NC = NTN * ACA;$$

$$HC = NTP * ACI,$$

ahol NFN az álnegatív esetek, NFP az álpozitív esetek, NTN a valódi negatív esetek, míg NTP a valódi pozitív esetek száma, továbbá az átlagos kártérítést ACA (average claim amount) és az átlagos nyomozási költséget ACI (average cost per investigation) jelöli.

Viaene és szerzőtársai (*2007*) nem egy rendszer költségmegtakarítását, hanem annak (2) egyenlettel adott működési költségeit (OC) határozták meg, mindazonáltal

a ráfordítások meghatározásának módja azonos a *Phua et al.* (*2004*) által bemutatottakkal.

$$OC = MC + FAC + NC + HC \quad (2)$$

Ami vállalati szempontból fontos, hogy a szerzők mindkét esetben azzal a feltételezéssel dolgoznak, hogy a valódi negatív (TN) esetek nem jelentenek többletkiadást (azaz egy valódi negatív eset többletköltsége 0) a biztosítók számára, hiszen ezekben az esetekben a normál kártérítési folyamatról van szó. Azonban interjúink[6] során az iparági szakértők rávilágítottak, hogy a gyakorlatban ezeknek a valódi negatív eseteknek is többletköltsége van. Hasonló eltérés tapasztalható a vállalati gyakorlat és a szakirodalom között a valódi pozitív esetek költségszámítása során is. A szakirodalom szerint valódi pozitív esetekben a biztosító nem fizet a biztosítottnak, azaz az egyetlen költség, amely felmerül, a nyomozással kapcsolatos. A vállalati gyakorlatban azonban más a helyzet. Amint azt több korábbi tanulmány kimutatta (például *Derrig – Ostaszewski 1995; Weisberg – Derrig 1998*), a gépjármű-biztosítási csalások túlnyomó többsége az úgynevezett túlárazott[7] kártérítési igényekből áll. Interjúalanyaink rávilágítottak arra, hogy a szakirodalommal ellentétben, a vállalati gyakorlatban ritkán fordul elő, hogy a biztosító teljes mértékben megtagadja a kifizetést. Általában a kért összegnél kisebb összeget ajánlanak fel az azonosított túlárazott kártérítési igények esetén. Ennek számos oka van, például a hosszadalmas és költséges bírósági folyamat, vagy épp a negatív marketing.

Tekintettel a szakirodalom és a vállalati gyakorlat közötti, fentebb ismertetett különbségekre, a gépjármű-biztosítási csalások felderítésével kapcsolatos valós költségek meghatározására a *Benedek és szerzőtársai* (*megjelenés alatt*) által javasolt, a (3) egyenlettel adott számítási módszert javasoljuk:

$$CSDM = NA - (MC + FAC + NC + HC) \quad (3)$$

ahol *NA* a „no action cost”[8]. Továbbá:

$$MC = NFN * (ACA + AAC);$$

$$FAC = NFP * (ACI + ACA);$$

$$NC = NTN * (ACA + AAC);$$

$$HC = NTP * (ACA - ASCIFC + ACI);$$

[6] Három mélyinterjút készítettünk romániai biztosítótársaságok vezetőivel és multinacionális biztosítótársaságok szakértőivel a gépjármű-biztosítási csalások kapcsán. Ezt követően elkészítettünk egy 22 kérdést tartalmazó kérdőívet, amelyet az UNSAR (The National Association of Insurance and Reinsurance Companies in Romania) minden partnerintézetéhez eljuttattunk.

[7] Olyan esetek, amikor a biztosított vagy épp a szakszervíz a valós javítási költségnél nagyobb összeget tüntet fel a kártérítési igény során.

[8] A tanulmányban a *Puha et al.* (*2004*) által bemutatott megközelítést alkalmazzuk, de az általunk javasolt költségszámítási módszer akkor is tökéletesen működne, ha a „no action cost” helyett egy alternatív rendszer működési költségeivel dolgoznánk.

ahol az átlagos feldolgozási költséget AAC (average administrative cost) és az átlagos megtakarítást az azonosított jogtalan követelések esetén ASCIFC (average savings in case of identified fraudulent claims) jelöli.

Végezetül meg kell említenünk a csalásmegelőzési programok prevenciós hatását, hatékony prevenció nélkül ugyanis az idő múlásával a díjakat emelni kell olyan mértékben, hogy a jogtalan kifizetéseket is „elbírják", így előbb-utóbb a biztosítási díj elér egy olyan szintet, amely már nem lesz versenyképes a piacon. Ebben a tanulmányban a sokkal nehezebben kvantifikálható prevenciós hatás nem szerepel explicite, de nem is befolyásolná érdemben az eredményt, hiszen a prevenció költségei mind a klasszikus statisztikai módszerek, mind a mesterséges intelligencián alapuló módszerek jövedelmezőségét csökkentik rövid távon.

4. Eredmények

4.1. A kiválasztott módszerek költséghatékonysági metaanalízise

A szakirodalom tanulmányozása és a kutatási rések feltárása után elvégeztük a kiválasztott módszerek metaanalízisét, amelynek segítségével képessé válunk a gépjármű-biztosítási csalásazonosító módszerek rangsorolására, összehasonlítására. Elsőként e módszerek költségmegtakarítási képességét számoltuk ki a javasolt költségmegtakarítás-számítási módszerrel.

A kezdetben azonosított 24 folyóiratcikk és 3 konferenciaanyag kiválasztása mögötti logika kettős volt. Egyrészt csak azokat a tanulmányokat vettük figyelembe, melyeket a Web of Science indexelt, másrészt szem előtt tartottuk azt is, hogy az 1999–2012 közötti hagyományos statisztikai-ökonometriai megközelítést használó modellek teljesítményét kívánjuk összehasonlítani azon 2012–2022 közötti MI alapú modellek teljesítményével, melyeket valós adathalmazon teszteltek. Az így azonosított 27 tanulmány közül néhány azonban pusztán elméleti jellegű volt, és nem kínált konkrét csalásazonosítási módszert. Más tanulmányok szerzői (például *Pathak et al. 2005; Padmaja et al. 2007; Bhowmik 2011; Xu et al. 2011; Karamizadeh – Zolfagharifar 2016; Badriyah et al. 2018*) valós vállalati adatkészletek használata nélkül végezték kutatásaikat. Végül több olyan tanulmány is volt, amelyben a szerzők nem mutatják be a találati mátrixot, így ezen kutatások esetén nem tudtuk meghatározni a költségszámítási módszerünkhöz szükséges inputokat.

Figyelembe véve a fenti korlátokat, mindössze 12 kutatás maradt a mintánkban, amelyben minden szükséges adat megtalálható az egyes modellek költségmegtakarítási képességének meghatározásához. A 12 cikkben a szerzők összesen 35 különböző módszert javasolnak és hasonlítanak össze, a teljes lista megtalálható a *Melléklet 4. táblázatában*.

Mivel az elemzett tanulmányokban a jogtalan követelések aránya különböző, az adatbázisok mérete nagyon eltérő, mi több, a felhasznált 7 adatbázis közül 2 az Egyesült Államokból, 1 Kanadából, 2 Spanyolországból, 1 Oroszországból és 1 Szlovéniából származik, első lépésként egy általános keretrendszert építettünk fel, ahol feltételezzük, hogy egy biztosító 10 000 kártérítési igényt dolgoz fel, melyből 10 százalék jogtalan. A metaváltozókat, mint például az átlagos nyomozás költségét vagy az átlagos kártérítési összeget a korábban már említett kérdőíves felmérés alapján határoztuk meg. A kérdőívet öt romániai biztosítótársaság töltötte ki maradéktalanul, melyek együttes piaci részesedése közel 70 százalék. Jelen tanulmányban az öt biztosító által megadott értékek piaci részesedéssel súlyozott átlagával dolgoztunk. Ezek alapján az átlagos nyomozási költség 145 USD, az átlagos kártérítési összeg 2 420 USD, az azonosított jogtalan követelések esetén az átlagos megtakarítás 485 USD, míg az átlagos feldolgozási költség 12 USD.

A *2. táblázatban* három különböző forgatókönyv esetén foglaltuk össze a 35 módszer költségmegtakarítási képességét. A táblázat 2–7. sora az adott forgatókönyv bemeneti paramétereit mutatja. Ezek azok a bemeneti metaparaméterek, amelyek értékei biztosítási szakemberektől származnak, és amelyek minden egyes klasszikus statisztikai, illetve MI-n alapuló módszer esetén mindvégig konstansok. A 8. sor jelenti a legfontosabb sort, az outputot, hiszen ezt már a metaparaméterek konkrét algoritmus-paraméterekkel való kölcsönhatásaival és feldolgozásával nyerjük. Azaz egy algoritmus végső működési költsége egyenlő a találati mátrix alapján meghatározott különböző kategóriákba (álpozitív, álnegatív) tartozó kártérítési igények száma szorozva az adott kategóriához tartozó metaparaméter (átlagos nyomozási költség, átlagos kártérítési összeg) konstans értékével. Gazdasági nyelvezetben a 8. sor azt mutatja, hogy a 35 módszer közül hánynak volt nagyobb a működési költsége, mint az alternatíva működési költsége, azaz ha a biztosító nem vizsgálta a kártérítési igények jogosságát és egyszerűen csak kifizette a beérkező kártérítési igényeket. Intuícióval ellentétesen itt a legjobb forgatókönyvnek az minősül, ahol a legnagyobb a jogtalan követelések aránya, hiszen ebben az esetben egy kevésbé hatékony módszer is nagyobb költségmegtakarítást tud elérni.

2. táblázat
A jogtalan kártérítési igények azonosítására használt módszerek költséghatékonysága

	Legvalószínűbb forgatókönyv	Legrosszabb forgatókönyv	Legjobb forgatókönyv
	35 modell	**35 modell**	**35 modell**
Jogtalan követelések aránya (%)	10	5	20
Átlagos kártérítés nagysága (USD)	2 420	2 420	2 420
Átlagos nyomozási költség (USD)	145	193	97
Átlagos feldolgozási költség (USD)	12	12	12
Átlagos megtakarítás az azonosított jogtalan követelések esetén (USD)	485	315	1 213
Azon módszerek száma, melyek működési költsége magasabb, mint a „no action cost”	27	31	0

Megjegyzés: A legrosszabb és legjobb forgatókönyvek esetén a biztosító társaságok által megadott szélsőértékekkel dolgoztunk.

Hangsúlyozzuk, hogy a *2. táblázatban* összefoglalt adatok jól mutatják a javasolt költségmegtakarítás-számítási módszer fontosságát vállalati szempontból. Míg a *Phua et al.* (*2004*) által javasolt költségmegtakarítás-számítási módszer szinte az összes modellt költséghatékonynak minősíti, az általunk javasolt módszer (amely a valós csalásfelderítési folyamat során felmerülő költségeket veszi figyelembe), a legvalószínűbb forgatókönyv esetén is csupán a modellek 22,85 százalékát minősíti költséghatékonynak, míg a legrosszabb forgatókönyv esetén csak a módszerek 11,42 százaléka minősíthető költséghatékonynak szemben a *Phua et al.* (*2004*) által javasolt megközelítés 94,28 (illetve 68,57) százalékával.

4.2. A csalást azonosító módszerek költségmegtakarítási képességeinek hőtérképei

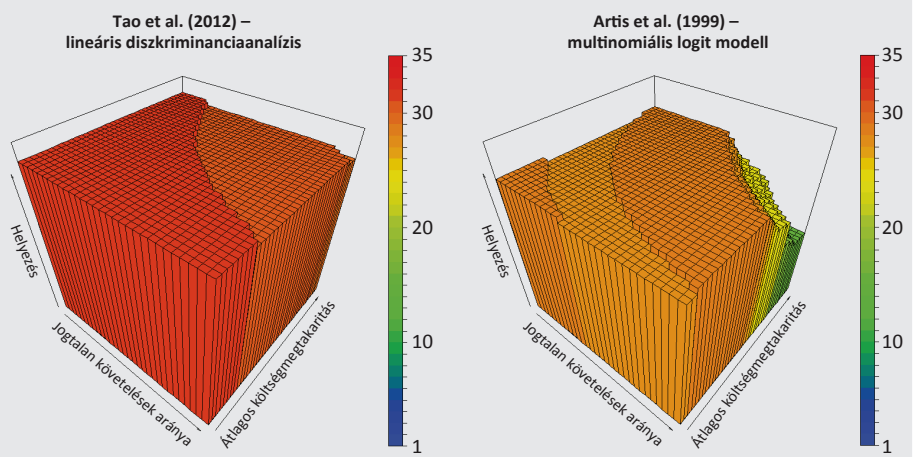
A metaanalízis során feltárt, meglehetősen meglepő eredmények láttán fontos lépésnek tartottuk az egyes csalásfelderítési módszerek további alapos elemzését és annak vizsgálatát, hogy az egyes módszerek milyen körülmények között válhatnak előnyösebbé, mint társaik. Ennek a megközelítésnek az egyik oka, hogy a metaanalízis során használt bemeneti paraméterek (például jogtalan követelések aránya, átlagos nyomozási költség) függvényében a csalásfelderítési módszerek költséghatékonysága jelentős eltéréseket mutat. A másik ok az, hogy bizonyos felderítési módszerek egyes biztosítók számára használhatatlanok, mivel ezek a csalásfelderítési módszerek olyan inputokat (a baleset jellemzői, rendőrségi/orvosi jelentések, balesetről készült fényképek) használnak, amelyek az adott biztosító számára nem (vagy még nem) állnak rendelkezésre.

Annak érdekében, hogy a csalásfelderítési módszerek egyedi jellemzőit minél inkább figyelembe tudjuk venni, valamint a metaanalízist a bemeneti paraméterek minél szélesebb választéka esetén elvégezhessük, így vizsgálva a módszerek

teljesítményét, 3 különböző szimulációt futtattunk, és hőtérképeket készítettünk az eredmények megjelenítéséhez.

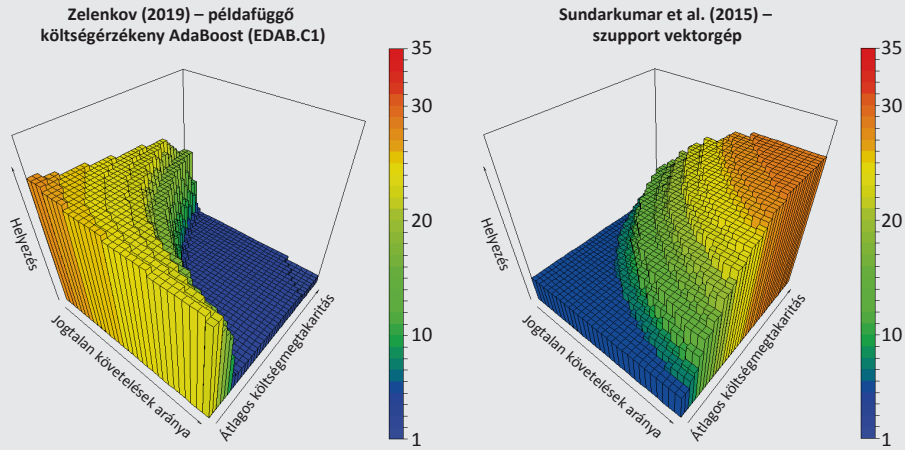
Az első szimulációban 145 dolláros állandó nyomozási költséget feltételeztünk, míg a jogtalan követelések arányát és az átlagos megtakarítást az azonosított jogtalan követelések esetén változtattuk. Ez a megközelítés nagyon hasznos lehet azon biztosító társaságoknak, amelyek fix nyomozási költséggel dolgoznak (például amiatt, hogy egy erre szakosodott külső vállalatot bíznak meg a nyomozás lefolytatásával, és egy előre megbeszélt árat fizetnek minden egyes kártérítési igény jogosságának vizsgálatáért), hiszen nagyon könnyen el tudják dönteni, hogy az adott piaci körülmények mellett mely módszer lehet a leghatékonyabb számukra. Például, ha egy biztosítótársaság nem tudja használni a *Tao et al.* (*2012*) vagy *Bermúdez et al.* (*2008*) által javasolt csalásfelderítési módszereket – mert nem állnak rendelkezésére a modell alkalmazásához szükséges input paraméterek –, de olyan piacon tevékenykedik, ahol magas a jogtalan követelések aránya és alacsony az átlagos megtakarítás az azonosított jogtalan követelések esetén, az *Artís et al.* (*1999*) által javasolt multinomiális logit modell optimális választás lehet (*1. ábra*), hiszen szinte ugyanolyan jó teljesítményt nyújt, mint a *Tao et al.* (*2012*) által javasolt módszer. Hasonlóképpen, bármely biztosító társaság könnyen kiválaszthatja a legmegfelelőbb módszert a jogtalan kártérítési igények aránya és az azonosított jogtalan követelések átlagos megtakarítása alapján. Alacsony jogtalan követelési rátával és alacsony átlagos megtakarítással rendelkező piacon működő társaságok esetén a *Zelenkov* (*2019*) által javasolt módszer jobbnak tűnik, mint a *Sundarkumar et al.* (*2015*) által javasolt, lásd a *2. ábrát*.

1. ábra
A Tao et al. (2012) és Artís et al. (1999) által javasolt modellek költségmegtakarítási képessége hőtérképen



Megjegyzés: A Tao et al. (2012) által javasolt lineáris diszkriminanciaanalízis modell, valamint az Artís et al. (1999) által javasolt multinomiális logit modell költségmegtakarítási képessége összehasonlítva az elemzett 35 modell költségmegtakarítási képességével különböző forgatókönyvek esetén.

2. ábra
A Zelenkov (2019) és Sundarkumar et al. (2015) által javasolt modellek költségmegtakarítási képessége hőtérképen



Megjegyzés: A Zelenkov (2019) által javasolt példafüggő költségérzékeny AdaBoost (EDAB.C1) modell, valamint a Sundarkumar et al. (2015) által javasolt szupport vektorgép modell költségmegtakarítási képessége összehasonlítva az elemzett 35 modell költségmegtakarítási képességével különböző forgatókönyvek esetén.

A második szimulációnál az azonosított jogtalan követelések megtakarítását állandónak tekintettük (485 dollár), és változott a nyomozás költsége, valamint a jogtalan követelések aránya. A harmadik szimulációban a jogtalan követelések aránya állandó (10%), és változtattuk a nyomozás költségét, valamint az azonosított jogtalan követelések átlagos megtakarítását.

4.3. A hagyományos statisztikai és gépi tanuláson alapuló módszerek összehasonlítása az átlagos költségmegtakarítás szempontjából

A metaanalízis és a hőtérképek után részletes, nem parametrikus rangkorrelációs elemzést végeztünk a különböző csalásfelderítési módszerekről. A Spearman-féle rangkorrelációk részletes ismertetésétől most eltekintünk, megtalálhatóak *Benedek et al.* (*megjelenés alatt*) tanulmányban. Az összefüggések nagyságrendje és szignifikanciája egyértelműen azt mutatja, hogy a jelen tanulmányban alkalmazott teljesítménymérők az elemzett csalásfelderítési módszerek következetes rangsorolását eredményezik (részletek a *Melléklet 3. táblázatában*).

A tanulmány talán legérdekesebb kérdése, hogy a mesterséges intelligencia alapú kimutatási módszerek lényegesen költséghatékonyabbak-e, mint a hagyományos statisztikai-ökonometriai eszközök.

Nyilvánvalóan a mesterséges intelligencia és a hagyományos statisztika-ökonometria módszerei mind ugyanannak a generikusan adattannak (data science) nevezett

tudományterületnek a fejezetei, és mint ilyen, a közöttük húzódó határ meglehetősen szubjektív és képlékeny, különös tekintettel az MI dinamikus fejlődésére, amely a szemünk előtt zajlik. Például a legtöbb gépi tanulás kurzus a lineáris és logisztikus regresszió módszertanával kezdődik, ami ugyanúgy része bármely standard ökonometria tantervnek. Mindazonáltal mi a következő megkülönböztetést alkalmaztuk a tanulmányban: MI vagy gépi tanulásos módszernek tekintettünk minden olyan módszert, amelyet a szakirodalomban az MI terminológia megjelenése után fejlesztettek ki. Ezért pl. a lineáris és logisztikus regressziót, lineáris diszkriminancia-analízist a hagyományos kategóriába soroltuk (hiszen ezekhez nincs szükség big datara, neurális hálókra) míg a genetikus algoritmusokat, neurális hálókat stb. az MI kategóriába soroltuk.

Első lépésként kiszámítottuk az említett két módszercsoport átlagos költségmegtakarításának különbségeit, és a különbségek statisztikai szignifikanciáját teszteltük a Mann–Whitney-féle nem-paraméteres teszt segítségével. Ezeket az összehasonlításokat a bemeneti paraméterek kombinációinak széles skáláján (összesen 10 780) hajtottuk végre, így egy szintetikus kereszttáblázatot kaptunk az átlagos vizsgálatonkénti költség és az azonosított csalárd követelések átlagos megtakarításai között.

A *Melléklet 5. táblázatában* jól látható, hogy az átlagos költségmegtakarítás a kombinációk túlnyomó többségénél magasabb a hagyományos statisztikai módszereknél[9] (az eltérések pozitívak és szignifikánsak), mint az MI-alapú módszereknél, ezért arra a következtetésre jutottunk, hogy meglepő módon egyelőre nem indokolt a biztosítótársaságok számára a jelentős többletbefektetés a mesterséges intelligencia alapú csalásfelderítési algoritmusokba. Ez persze nem jelenti azt, hogy ezeknek a cégeknek ne lenne szükségük szoftveres támogatásra a működésük során, csupán annyit, hogy a legtöbb esetben elegendő a hagyományos statisztikai szoftver.

5. Következtetések

Kutatásunkban rámutattunk arra, hogy a szakirodalom hiányos a gépjármű-biztosítási csalások felderítési módszereinek költséghatékonysági vizsgálata terén. Mi több, a feltörekvő piacok esetén, teljesen hiányzik a gépjármű-biztosítási csalások felderítésével kapcsolatos szakirodalom. Éppen ezért, jelen tanulmányban a *Benedek és szerzőtársai* (*megjelenés alatt*) által javasolt módszert alkalmaztuk a gépjármű-biztosítási csalásazonosítás költségmegtakarítási potenciáljának korrekt kiszámításához. A javasolt módszer minden olyan költséget figyelembe vesz, amely egy valós csalásfelderítési folyamat során felmerül (különös hangsúlyt fektetve arra,

[9] Bár ebben a tanulmányban nem célunk a hagyományos statisztikai és az MI-módszerek implementálási költségeinek vizsgálata, nagy valószínűséggel a hagyományos módszerek költségvonzata ilyen téren is alacsonyabb, ami tovább erősíti a *Melléklet 5. táblázatában* megfigyelhető következtetéseket.

hogy jogtalan vagy részben jogtalan követelés esetén a biztosító általában nem tagadja meg teljes mértékben a kifizetést, hanem részleges kártérítést ajánl fel).

A költséghatékonysági vizsgálat során 12 különböző forrásból származó 35 csalásfelderítési módszer metaanalízisét végeztük el, és arra a következtetésre jutottunk, hogy a legtöbb, jelenleg a szakirodalomban megtalálható gépjármű-biztosítási csalásfelderítési módszer nem jövedelmező. Emellett arra is rámutattunk, hogy a hagyományos statisztikai módszereken alapuló megközelítések egyelőre jobban teljesítenek az MI alapú módszereknél. Azaz, a biztosítótársaságok számára egyelőre nem indokolt a jelentős többletbefektetés a mesterséges intelligencia alapú csalásfelderítési algoritmusokba, a legtöbb esetben elegendő a hagyományos statisztikai szoftverek nyújtotta lehetőségek használata. Ez az eredmény összhangban van a *Benedek és szerzőtársai* (*megjelenés alatt)* által bemutatottakkal. Azaz, a hagyományos statisztikai módszerek alkalmazása gazdaságosabb a jelen tanulmányban vizsgált mintán is (2012 előtti hagyományos statisztikai módszerek versus 2012 utáni MI alapú megközelítések). Ezzel az eredménnyel jelen tanulmány egyfajta robusztusság vizsgálataként működik és megerősíti a korábbi kutatási eredményeket.

A kutatás legfontosabb korlátja, ami egyben továbbfejlesztési lehetőség is, az, hogy a metaanalízisben szereplő bemeneti paraméterek mögött korábbi, más-más kártérítési mintákon betanított algoritmusok állnak. Az igazán meggyőző bizonyítékot akkor kaphatnánk, ha ugyanezen algoritmusokat egytől egyig lefuttathatnánk ugyanazon a mintán.

Felhasznált irodalom

Ágoston Norbert (2022): *Mesterséges intelligencia és gépi tanulási módszerek a vállalati fizetésképtelenség becslésére*. Statisztikai Szemle, 100(6): 584–609. https://doi.org/10.20311/stat2022.6.hu0584

Artís, M. – Ayuso, M. – Guillén, M. (1999): *Modelling different types of automobile insurance fraud behaviour in the Spanish market*. Insurance: Mathematics and Economics, 24(1–2): 67–81. https://doi.org/10.1016/S0167-6687(98)00038-9

Artís, M. – Ayuso, M. – Guillén, M. (2002): *Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims*. Journal of Risk and Insurance, 69(3): 325–340. https://doi.org/10.1111/1539-6975.00022

Association of British Insurers (2021): *No Time to Lie*. https://www.abi.org.uk/news/news-articles/2021/10/detected-fraud-2020/. Letöltés ideje: 2021 november 15.

Badriyah, T. – Rahmaniah, L. – Syarif, I. (2018): *Nearest neighbour and statistics method based for detecting fraud in auto insurance*. International Conference on Applied Engineering (ICAE), Batam, Indonesia, pp. 1–5. https://doi.org/10.1109/INCAE.2018.8579155

Bánkúty-Balog Lilla (2022): *A mesterséges intelligencia elterjedésének geoökonómiai hatásai és Magyarország*. Külgazdaság, 66 (7–8): 102–130. https://doi.org/10.47630/KULG.2022.66.7-8.102

Belhadji, E.B. – Dionne, G. – Tarkhani, F. (2000): *A Model for the Detection of Insurance Fraud*. The Geneva Papers on Risk and Insurance – Issues and Practice*,* 25(4): 517–538. https://doi.org/10.1111/1468-0440.00080

Benedek Gábor (1999): *Mesterséges intelligencia az üzleti világban: Marketingakciók hatékonyságának elemzése statisztikai és Data Mining módszerekkel*. Vezetéstudomány-Management and Business Journal, 30(11): 33–36.

Benedek, B. – Ciumas, C. – Nagy, B.Z. (2022): *Automobile insurance fraud detection in the age of big data – a systematic and comprehensive literature review*. Journal of Financial Regulation and Compliance, 30(4): 503–523. https://doi.org/10.1108/JFRC-11-2021-0102

Benedek, B. – Ciumas, C. – Nagy, B.Z. (megjelenés alatt): *On the cost-efficiency of automobile insurance fraud detection methods – A meta-analysis*. Global Business Review, közlésre elfogadva, megjelenés alatt.

Bermúdez, L. – Pérez, J.M. – Ayuso, M. – Gómez, E. – Vázquez, F.J. (2008): *A Bayesian dichotomous model with asymmetric link for fraud in insurance*. Insurance: Mathematics and Economics, 42(2): 779–786. https://doi.org/10.1016/j.insmatheco.2007.08.002

Bhowmik, R. (2011): *Detecting auto insurance fraud by data mining techniques*. Journal of Emerging Trends in Computing and Information Sciences, 2(4): 156–162.

Csillag J. Balázs – Granát P. Marcell – Neszveda Gábor (2022): *A környezeti kérdésekre irányuló médiafigyelem és az ESG-befektetések*. Hitelintézeti Szemle, 21(4): 129–151. https://doi.org/10.25201/HSZ.21.4.129

Danyi Pál (2018): *A mesterséges intelligencia alkalmazása az árazásban*. Marketing & Menedzsment, 52(3–4): 5–18.

Derrig, R. A. – Ostaszewski, K. M. (1995): *Fuzzy techniques of pattern recognition in risk and claim classification*. Journal of Risk and Insurance, 62(3): 447–482. https://doi.org/10.2307/253819

Farkas Gábor – Magyar Péter – Molnár András – Zubor-Nemes Anna (2020): *Adatbányászati módszerek alkalmazása a mezőgazdaságban – a gépi tanulás felhasználási lehetőségei.* Gazdálkodás: Scientific Journal on Agricultural Economics, 64(1): 15–24.

Farquad, M.A. – Ravi, V. – Raju, S.B. (2012): *Analytical CRM in banking and finance using SVM: a modified active learning-based rule extraction approach.* International Journal of Electronic Customer Relationship Management, 6(1): 48–73. https://doi.org/10.1504/IJECRM.2012.046470

Gepp, A. – Wilson, H.J. – Kumar, K. – Bhattacharya, S. (2012): *A Comparative Analysis of Decision Trees Vis-a-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection*. Journal of Data Science, 10(3): 537–561. https://doi.org/10.6339/JDS.201207_10(3).0010

Green, D.M. – Swets, J. A. (1966): *Signal detection theory and psychophysics* (1 ed., Vol. 1). New York: Wiley.

Hámori Gábor (2001): *A fizetésképtelenség előrejelzése logit-modellel*. Bankszemle, 45(1–2): 65–87.

He, H. – Bai, Y. – Garcia, E. – Li, S. (2008): *ADASYN: Adaptive synthetic sampling approach for imbalanced learning*. IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, pp. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969

Hill, H. R. – Sandler, B. – Mokgokong, R. – Lister, S. – Ward, T. – Boyce, R. – Farooqui, U. – Gordon, J. (2020): *Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm*. Journal Of Medical Economics, 23(4): 386–393. https://doi.org/10.1080/13696998.2019.1706543

III (2019): Insurance Information Institute: *Insurance Fact Book.* Insurance Information Institute. https://www.iii.org/sites/default/files/docs/pdf/insurance_factbook_2019.pdf

III (2021): Insurance Information Institute: *Background on: Insurance fraud.* https://www.iii.org/article/background-on-insurance-fraud. Letöltés ideje: 2021 november 20.

Karamizadeh, F. – Zolfagharifar, S. A. (2016): *Using the Clustering Algorithms and Rule-based of Data Mining to Identify Affecting Factors in the Profit and Loss of Third Party Insurance, Insurance Company Auto.* Indian Journal of Science and Technology, 9(7): 1–9. https://doi.org/10.17485/ijst/2016/v9i7/87846

Lee, A. – Taylor, P. – Kalpathy-Cramer, J. – Tufail, A. (2017): *Machine Learning Has Arrived!* Ophthalmology, 124(12): 1726–1728. https://doi.org/10.1016/j.ophtha.2017.08.046

LII (2023): Legal Information Institute: *Insurance Fraud.* Legal Information Institute, Cornell Law School. https://www.law.cornell.edu/wex/insurance_fraud. Letöltés ideje: 2023. április 26.

Li, Y. – Yan, C. – Liu, W. – Li, M. (2018): *A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification*. Applied Soft Computing, 70(September): 1000–1009. https://doi.org/10.1016/j.asoc.2017.07.027

Massachusetts Regulation (1993): *Division of Insurance Regulations*. Massachusetts government. https://www.mass.gov/service-details/division-of-insurance-regulations. Letöltés ideje: 2023. április 26.

Muraközy Balázs (2018): *Gépi tanulás, predikció és okság a közgazdaság-tudományban.* Magyar Tudomány, 179(7): 1027–1037. https://doi.org/10.1556/2065.179.2018.7.10

Nian, K. – Zhang, H. – Tayal, A. – Coleman, T. – Li, Y. (2016): *Auto insurance fraud detection using unsupervised spectral ranking for anomaly.* The Journal of Finance and Data Science, 2(1): 58–75. https://doi.org/10.1016/j.jfds.2016.03.001

Padmaja, T. M. – Dhulipalla, N. – Bapi, R.S. – Krishna, P.R. (2007): *Unbalanced data classification using extreme outlier elimination and sampling techniques for fraud detection.* 15th International Conference on Advanced Computing and Communications (ADCOM 2007), Guwahati, India, pp. 511–516. https://doi.org/10.1109/ADCOM.2007.74

Pathak, J. – Vidyarthi, N. – Summers, S.L. (2005): *A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims.* Managerial Auditing Journal, 20(6): 632–644. https://doi.org/10.1108/02686900510606119

Pérez, J.M. – Muguerza, J. – Arbelaitz, O. – Gurrutxaga, I. – Martín, J.I. (2005): *Consolidated Tree Classifier Learning in a Car Insurance Fraud Detection Domain with Class Imbalance.* In: Singh, S. – Singh, M. – Apte, C. – Perner, P. (szerk.): *Pattern Recognition and Data Mining*. ICAPR 2005. Lecture Notes in Computer Science, vol 3686. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11551188_41

Phua, C. – Alahakoon, D. – Lee, V. (2004): *Minority report in fraud detection: classification of skewed data.* ACM Sigkdd Explorations Newsletter, 6(1): 50–59. https://doi.org/10.1145/1007730.1007738

Shaeiri, Z. – Kazemitabar, S. J. (2020): *Fast unsupervised automobile insurance fraud detection based on spectral ranking of anomalies.* International Journal of Engineering, 33(7): 1240–1248. https://doi.org/10.5829/ije.2020.33.07a.10

Šubelj, L. – Furlan, Š. – Bajec, M. (2011): *An expert system for detecting automobile insurance fraud using social network analysis*. Expert Systems with Applications, 38(1): 1039–1052. https://doi.org/10.1016/j.eswa.2010.07.143

Subudhi, S. – Panigrahi, S. (2017): *Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection.* Journal of King Saud University – Computer and Information Sciences, 32(5): 568–575. https://doi.org/10.1016/j.jksuci.2017.09.010

Sundarkumar, G.G. – Ravi, V. (2015): *A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance.* Engineering Applications of Artificial Intelligence, 37(January): 368–377. https://doi.org/10.1016/j.engappai.2014.09.019

Sundarkumar, G.G. – Ravi, V. – Siddeshwar, V. (2015): *One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection*. IEEE International Conference on Computational Intelligence and Research (ICCIC), Madurai, India, pp. 1–7. https://doi.org/10.1109/ICCIC.2015.7435726

Tao, H. – Zhixin, L. – Xiaodong, S. (2012): *Insurance fraud identification research based on fuzzy support vector machine with dual membership.* International Conference on Information Management, Innovation Management and Industrial Engineering, Sanya, pp. 457–460. https://doi.org/10.1109/ICIII.2012.6340016

Viaene, S. – Ayuso, M. – Guillen, M. – Van Gheel, D. – Dedene, G. (2007): *Strategies for detecting fraudulent claims in the automobile insurance industry.* European Journal of Operational Research*,* 176(1): 565–583. https://doi.org/10.1016/j.ejor.2005.08.005

Viaene, S. – Derrig, R. A. – Baesens, B. – Dedene, G. (2002): *A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection.* Journal of Risk and Insurance*,* 69(3): 373–421. https://doi.org/10.1111/1539-6975.00023

Virág Miklós – Kristóf Tamás (2005): *Az első hazai csődmodell újraszámítása neurális hálók segítségével.* Közgazdasági Szemle, 52(2) 144–162.

Virág Miklós – Nyitrai Tamás (2013): *Application of support vector machines on the basis of the first Hungarian bankruptcy model.* Society and Economy, 35(2): 227–248. https://doi.org/10.1556/SocEc.35.2013.2.6

Virág Miklós – Nyitrai Tamás (2014): *Metamódszerek alkalmazása a csődelőrejelzésben*. Hitelintézeti Szemle, 13(4): 180–195. https://hitelintezetiszemle.mnb.hu/letoltes/8-virag-nyitrai-2.pdf

Wang, Y. – Xu, W. (2018): *Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud.* Decision Support Systems*,* 105(January): 87–95. https://doi.org/10.1016/j.dss.2017.11.001

Weisberg, H. – Derrig, R. (1991): *Fraud and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts*. Journal of Insurance Regulation*,* 9(4): 497–541.

Weisberg, H. – Derrig, R. (1998): *Quantitative methods for detecting fraudulent automobile bodily injury claims*. Risques, 35: 75–99.

Wilson, J.H. (2009): *An analytical approach to detecting insurance fraud using logistic regression*. Journal of Finance and Accountancy, 85(150): 1–15.

Xu, W. – Wang, S. – Zhang, D. – Yang, B. (2011): *Random rough subspace based neural network ensemble for insurance fraud detection.* Fourth International Joint Conference on Computational Sciences and Optimization, Kunming and Lijiang City, China, pp. 1276–1280. https://doi.org/10.1109/CSO.2011.213

Yan, C. – Li, Y. (2015): *The Identification Algorithm and Model Construction of Automobile Insurance Fraud Based on Data Mining.* Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), Qinhuangdao, China, pp. 1922–1928. https://doi.org/10.1109/IMCCC.2015.408

Zelenkov, Y. (2019): *Example-dependent cost-sensitive adaptive boosting*. Expert Systems with Applications, 135: 71–82. https://doi.org/10.1016/j.eswa.2019.06.009

Melléklet

3. táblázat
A rangsorok közötti Spearman-rangkorrelációs együtthatók különböző paraméterek alapján

	Összmeg-takarítás	Szenzitivitás	Specificitás	Precizitás	Negatív prediktív érték	Becslési pontosság	F-mérték
Összmeg-takarítás	1,000						
Szenzitivitás	0,069 (0,731)	1,000					
Specificitás	0,831 (49,41)***	–0,346 (–2,57)**	1,000				
Precizitás	0,924 (24,56)***	0,047 (0,48)	0,871 (19,15)***	1,000			
Negatív prediktív érték	0,254 (2,47)	0,951 (33,51)***	–0,028 (–0,41)	0,252 (2,78)**	1,000		
Becslési pontosság	0,947 (98,34)***	–0,081 (–0,62)	0,957 (38,93)***	0,942 (25,87)***	0,135 (1,57)	1,000	
F-mérték	0,828 (19,11)***	0,278 (4,01)***	0,599 (6,85)***	0,792 (15,68)***	0,616 (6,29)***	0,732 (11,03)***	1,000

*Megjegyzés: A negatív prediktív érték meghatározására használt formula a következő: TN/(FN+TN). Student t-statisztikák zárójelben. *10%-os szinten szignifikáns; ** 5%-os szinten szignifikáns; *** 1%-os szinten szignifikáns.*

4. táblázat
A vizsgált 35 csalásfelderítési módszer, valamint ezek szenzitivitása és specificitása

Szerző	Módszer neve	Szenzitivitás	Specificitás
***Artís et al.* (*1999*)**	multinomiális logit modell	0,6614	0,9065
	beágyazott multinomiális logit modell	0,3209	0,8132
***Belhadji et al.* (*2000*)**	probit regresszió – 10%-os küszöb	0,6940	0,9145
	probit regresszió – 20%-os küszöb	0,5373	0,9596
***Artís et al.* (*2002*)**	logit regresszió nem teljes körűen megfigyelt függő változóval (logit regression with omission error)	0,7793	0,6994
	logit regresszió teljes körűen megfigyelt függő változóval (logit regression without omission error)	0,7703	0,7094
***Bermúdez et al.* (*2008*)**	Bayes-i aszimmetrikus logit modell (Bayesian skewed logit model)	0,8515	0,9968
	standard logit és Bayes-i logit modellek	0,8515	0,6043
***Wilson* (*2009*)**	logit regresszió	0,5918	0,8163
***Šubelj et al.* (*2011*)**	közösségi hálózat elemzése	0,8913	0,8667
***Tao et al.* (*2012*)**	lineáris diszkriminanciaanalízis	0,7392	0,9738
	másodfokú diszkriminanciaanalízis	0,7933	0,9767
	naiv Bayes-i (naive Bayesian)	0,8351	0,9815
Farquad et al. (*2012*)	MALBA (logisztikus) - 1000 extra példány	0,8838	0,5534
	MALBA (normal) - 1000 extra példány	0,8811	0,5588
	ALBA - 1000 extra példány	0,8784	0,5656
	MALBA - 1000 extra példány	0,8848	0,5560
Sundarkumar et al. (*2015*)	döntési fa	0,9552	0,5658
	többrétegű perceptron	0,4859	0,7889
	szupport vektorgép	0,9400	0,5639
	valószínűségi neurális háló	0,9173	0,5533
	csoportos adatkezelés módszere (group method of data handling)	0,7362	0,7148
Sundarkumar – Ravi (*2015*)	valószínűségi neurális háló	0,8750	0,5894
	többrétegű perceptron	0,6458	0,7189
	döntési fa	0,9074	0,5869
	csoportos adatkezelés módszere	0,5686	0,8020
	szupport vektorgép	0,9189	0,5839
Subudhi – Panigrahi (*2017*)	GAFCM – DT	0,6625	0,8765
	GAFCM – SVM	0,6970	0,8471
	GAFCM – MLP	0,6107	0,8400
	GAFCM – GMDH	0,5727	0,7976
Zelenkov (*2019*)	példafüggő költségérzékeny AdaBoost (EDAB.C1)	0,2510	0,9301
	példafüggő költségérzékeny AdaBoost (EDAB.C2)	0,5900	0,7327
	példafüggő költségérzékeny AdaBoost (EDAB.C2-ROC)	0,4477	0,8050
	példafüggő költségérzékeny AdaBoost (EDAB.C3)	0,2510	0,9301

Megjegyzés: félkövérrel jelölve a hagyományos statisztikai-ökonometria modellek

5. táblázat

Átlagos költségmegtakarítási különbségek a hagyományos statisztikai és a mesterséges intelligencia alapú azonosítási módszerek között

ASCIFC	ACI										
	100	110	120	130	140	150	160	170	180	190	200
160	73 100 (46)***	87 426 (45)***	101 753 (47)***	116 079 (48)***	130 405 (50)***	144 732 (51)***	159 058 (51)***	173 384 (51)***	187 711 (53)***	202 037 (53)***	216 363 (53)***
180	73 283 (45)***	87 610 (45)***	101 936 (46)***	116 262 (47)***	130 589 (48)***	144 915 (50)***	159 241 (51)***	173 568 (51)***	187 894 (51)***	202 221 (53)***	216 547 (53)***
200	73 467 (46)***	87 793 (46)***	102 120 (45)***	116 446 (46)***	130 772 (47)***	145 099 (48)***	159 425 (50)***	173 751 (51)***	188 078 (49)***	202 404 (51)***	216 730 (51)***
220	73 650 (44)***	87 977 (46)***	102 303 (46)***	116 629 (45)***	130 956 (46)***	145 282 (47)***	159 608 (48)***	173 935 (50)***	188 261 (51)***	202 588 (50)***	216 914 (51)***
240	73 834 (49)***	88 160 (46)***	102 487 (46)***	116 813 (45)***	131 139 (45)***	145 466 (47)***	159 792 (47)***	174 118 (48)***	188 445 (50)***	202 771 (50)***	217 097 (51)***
260	74 017 (49)***	88 344 (43)***	102 670 (47)***	116 996 (46)***	131 323 (45)***	145 649 (45)***	159 975 (47)***	174 302 (47,5)***	188 628 (48)***	202 955 (50)***	217 281 (50)***
280	74 201 (44)***	88 527 (49)***	102 853 (46)***	117 180 (46)***	131 506 (45)***	145 833 (45)***	160 159 (46)***	174 485 (47)***	188 812 (48)***	203 138 (48)***	217 464 (50)***
300	74 384 (42)***	88 711 (46,5)***	103 037 (43)***	117 363 (47)***	131 690 (46)***	146 016 (45)***	160 342 (45)***	174 669 (46)***	188 995 (47)***	203 322 (48)***	217 648 (48)***
320	74 568 (41)***	88 894 (48)***	103 220 (47)***	117 547 (46)***	131 873 (45)***	146 200 (46)***	160 526 (45)***	174 852 (45)***	189 179 (46)***	203 505 (47)***	217 831 (48)***
340	74 751 (42)***	89 078 (44)***	103 404 (47)***	117 730 (43)***	132 057 (46)***	146 383 (46)***	160 709 (45)***	175 036 (45)***	189 362 (45)***	203 689 (47)***	218 015 (47)***
360	74 935 (43)***	89 261 (42)***	103 587 (49)***	117 914 (46)***	132 240 (46)***	146 567 (45)***	160 893 (46)***	175 219 (45)***	189 546 (45)***	203 872 (46)***	218 198 (47)***
380	75 118 (47)***	89 445 (41)***	103 771 (46)***	118 097 (49)***	132 424 (44)***	146 750 (46)***	161 076 (46)***	175 403 (46)***	189 729 (45)***	204 056 (45)***	218 382 (46)***
400	75 302 (50)***	89 628 (42)***	103 954 (44)***	118 281 (46,5)***	132 607 (44)***	146 934 (46)***	161 260 (46)***	175 586 (46)***	189 913 (45)***	204 239 (45)***	218 565 (45)***
420	75 485 (51)***	89 812 (43)***	104 138 (42)***	118 464 (49)***	132 791 (49)***	147 117 (44)***	161 443 (47)***	175 770 (46)***	190 096 (46)***	204 423 (45)***	218 749 (45)***
440	75 669 (54)***	89 995 (43)***	104 321 (41)***	118 648 (45)***	132 974 (47)***	147 301 (44)***	161 627 (46)***	175 953 (46)***	190 280 (46)***	204 606 (46)***	218 932 (45)***
460	75 852 (60)***	90 179 (47)***	104 505 (41)***	118 831 (44)***	133 158 (49)***	147 484 (47)***	161 810 (44)***	176 137 (47)***	190 463 (45)***	204 790 (46)***	219 116 (45)***
480	76 036 (61)***	90 362 (50)***	104 688 (42)***	119 015 (42)***	133 341 (48)***	147 668 (49)***	161 994 (44)***	176 320 (46)***	190 647 (45)***	204 973 (46)***	219 299 (46)***
500	76 219 (61)***	90 546 (52)***	104 872 (43)***	119 198 (41)***	133 525 (44)***	147 851 (46,5)***	162 177 (47)***	176 504 (44)***	190 830 (47)***	205 157 (45)***	219 483 (46)***
520	76 403 (62)***	90 729 (53)***	105 055 (45)***	119 382 (41)***	133 708 (44)***	148 035 (49)***	162 361 (49)***	176 687 (43)***	191 014 (46)***	205 340 (47)***	219 666 (46)***
540	76 586 (65)***	90 913 (57)***	105 239 (47)***	119 565 (42)***	133 892 (42)***	148 218 (47,5)***	162 544 (47)***	176 871 (46)***	191 197 (44)***	205 523 (47)***	219 850 (45)***
560	76 770 (66)***	91 096 (60)***	105 422 (50)***	119 749 (43)***	134 075 (41)***	148 402 (44)***	162 728 (48)***	177 054 (49)***	191 381 (43)***	205 707 (46)***	220 033 (47)***
580	76 953 (73)***	91 280 (60)***	105 606 (51)***	119 932 (44)***	134 259 (41)***	148 585 (44)***	162 911 (49)***	177 238 (47)***	1915 64 (45)***	205 890 (44)***	220 217 (47)***
600	77 137 (73)***	91 463 (61)***	105 789 (51,5)***	120 116 (45)***	134 442 (42)***	148 769 (42)***	163 095 (46)***	177 421 (46,5)***	191 748 (47)***	206 074 (43)***	220 400 (46)***
620	77 320 (76)**	91 647 (62)***	105 973 (54)***	120 299 (47)***	134 626 (42)***	148 952 (41)***	163 278 (44)***	177 605 (49)***	191 931 (49)***	206 257 (44)***	220 584 (44)***
640	77 504 (77)**	91 830 (65)***	106 156 (60)***	120 483 (50)***	134 809 (43)***	149 136 (41)***	163 462 (44)***	177 788 (48)***	192 115 (47)***	206 441 (47)***	220 767 (43)***
660	77 687 (82)**	92 014 (66)***	106 340 (60)***	120 666 (51)***	134 993 (43)***	149 319 (41)***	163 645 (42)***	177 972 (45)***	192 298 (46)***	206 624 (49)***	220 951 (44)***
680	77 871 (87)**	92 197 (68)***	106 523 (60)***	120 850 (52)***	135 176 (46,5)***	149 503 (42)***	163 829 (41)***	178 155 (44)***	192 482 (49)***	206 808 (47)***	221 134 (47)***
700	78 054 (90)**	92 381 (73)***	106 707 (61)***	121 033 (54)***	135 360 (47)***	149 686 (43)***	164 012 (41)***	178 339 (44)***	192 665 (48)***	206 991 (46,5)***	221 318 (49)***

*Megjegyzés: ASCIFC: átlagos megtakarítás az azonosított jogtalan követelések esetén, ACI: átlagos nyomozási költség. Mann-Whitney U-statisztika zárójelben. *10%-os szinten szignifikáns; **5%-os szinten szignifikáns; ***1%-os szinten szignifikáns.*