

Minden Önről vagy a vezetésről szól?

IoT-alapú kockázatértékelések tervezése

Yi-Jen (Ian) Ho
[†ian.ho@psu.edu](mailto:ian.ho@psu.edu)
[u](#)

Siyuan Liu
[†sx168@psu.edu](mailto:sx168@psu.edu)
[u](#)

Jingchuan Pu
[†pucj@psu.edu](mailto:pucj@psu.edu)
[u](#)

Dian Zhang
[††zhangdian@ln.edu](mailto:zhangdian@ln.edu)
[hk](#)

[†] Smeal College of Business, Pennsylvania State University, University Park, PA

^{††} Gazdaságtudományi Kar, Lingnan Egyetem, New Territories, Hong Kong

Absztrakt

A kockázatértékelések kritikus küldetések mindenfajta műveletirányítás számára. A fejlett technológiák bevezetésével a kockázatok pontosabb értékelésére zavaró alkalmazásokat fejlesztenek ki. Ilyen paradigmaváltás zajlik különböző iparágakban, különösen az autóbiztosítási üzletágban. A statikus demográfiai adatokon alapuló általános gyakorlattal szemben a mobilitás-alapú kockázatértékelés a kockázatok az IoT-alapú telematikán keresztül nyert dinamikus vezetési adatok alapján jelzi előre. Ez a tanulmány újszerű megoldást javasol a járművezetők kockázatainak megismerésére. Kifejezetten a mobilitás két aspektusát vesszük figyelembe, nevezetesen a vonás és a pálya vonalait, amelyeket a GPS, a fedélzeti diagnosztika és a járműbe épített kamerák valós idejű módon figyelnek. A vonások a járművezetők jellegzetes vezetési viselkedésére (stílusára) utalnak, míg a pályák a jármű mozgássorozataiból és az utazások kontextuális tényezőiből állnak. A kettőből kivont jellemzőket választjuk ki egy olyan előrejelző modell optimalizálásához, amely mind az út, mind a járművezető szintjén értékeli a kockázatokot. A finomszemcsés vezetési adatok és a baleseti jelentések felhasználásával azt találjuk, hogy a viselkedési jegyek jelentős szerepet játszanak a balesetek előrejelzésében. Azt is észrevesszük, hogy a járművezetők kockázatai heterogének és idővel változhatnak. Egy sor empirikus validáció során a javasolt megoldás felülmúlja a jelenlegi gyakorlatot és a korábbi szakirodalomban figyelembe vett alternatív módszereket. Megmutatjuk, hogy a mobilitás-alapú modellek következetesen jobbak a gyakorlati szakemberek által általában elfogadott demográfiai alapú modelleknél. Mennyiségi szempontból megközelítésünk 43,45%-kal (41,82%), 29,46%-kal (27,87%), illetve 26,24%-kal (25,27%) javítja a logisztikus regresszió, a támogató vektorgép és a véletlen erdők felidézését (pontosságát). Végül, a javasolt megoldás robusztus és számítási szempontból hatékony a szokatlan felhasználási esetekre (pl. ferde és kis képzési minták). Az eredmények számos vezetői implikációt és tervezetet nyújtanak az autóbiztosítási ágazat számára az IoT-alapú kockázatértékelések operacionalizálásához az 5G kommunikáció elkövetkező korszakában.

Kulcsszavak: kockázatértékelés, használat-alapú biztosítás, a dolgok internete, fedélzeti diagnosztika, fedélzeti kamera

1. Bevezetés

A kockázatértékeléseket, a bizonytalanság hatásának azonosítását, értékelését és rangsorolását a különböző operatív kockázatkezelések első lépésének tekintik, például a beszerzés (pl. Nagali et al. 2008), a termelés (pl. Cole et al. 2017), a készletezés (pl. Michalski 2009) és a szállítás (pl. Blom et al. 2001) esetében. A kockázatelemzés kritikus, de a pontos információk hiánya miatt kihívást jelent (Ale 2016). A dolgok internete (IoT) és a mesterséges intelligencia (AI) a működési kockázatok minimalizálása érdekében finom nyomokat és rejtett bizonytalanságokat rögzít és modellez (Choi & Lambert 2017; Kumar et al. 2018). A Virgin Atlantic például Boeing 787-es flottával és IoT-n keresztül összekapcsolt teherszállító berendezésekkel rendelkezik (Accelya 2016). Az ilyen IoT-alapú kapcsolatok a késési kockázatok 20%-os csökkenéséhez vezetnek. A HSBC a mesterséges intelligenciával támogatott kockázatfelismerést pénzmossás elleni megoldásával példázza (Mejia 2020). A mesterséges intelligencia több pénzmossási mintát azonosít, mint a hagyományos, szabályalapú módszer, és 20%-kal csökkenti a téves pozitív téves besorolás kockázatát. A technológia-alapú kockázatértékelés nemcsak az általános működési hatékonyságot javítja, hanem bizonyos üzleti modelleket is felborít, különösen az autóbiztosítási ágazatban. Úttörők kísérleteznek a használat alapú biztosítással (UBI), amely a tényleges vezetés alapján számítja fel a díjakat. A 23%-os CAGR-rel növekvő UBI piaci részesedést rabol a hagyományos gyakorlattól, amely a biztosítási díjakat a tipikus demográfiai adatok alapján árazza (ReportLinker 2020). Az autóbiztosítási üzletág ilyen paradigmaváltása kezdi felhívni magára a tudományos élet figyelmét. A korábbi tanulmányok vagy a balesetek és a kontextuális tényezők összekapcsolására, vagy a baleseti kockázatok előrejelzésére összpontosítanak a járművezetők személyes adatainak felhasználásával. Jun et al. (2011) például pozitív összefüggést talált a sebesség és a balesetek között. Guo és Fang (2013) k-means-t alkalmaznak a járművezetők személyiség típusok szerinti klaszterezésére, majd a kockázatok előrejelzésére. Míg a korábbi tanulmányok egy kezdeti lépést tesznek a kockázatok előrejelzésére korlátozott információk és hagyományos modellezési megközelítések felhasználásával, az IT-alapú kockázatértékelések fejlesztését a legmodernebb technológiákkal együtt fel kell gyorsítani. Így sürgős szükség van egy olyan újszerű megoldásra, amely kihasználja az IoT és az AI előnyeit a kockázatok valós idejű értékeléséhez, ahogyan Chen és Jiang (2019) véli hogy a vezetésfelügyeleti technológiák kihasználása hatékonyságnövekedést eredményez az iparágban.

Természetesen felmerül egy alapvető kérdés: *vajon egy biztosítási szolgáltató kihasználhatja-e az IoT-telematikát a hagyományos üzleti modell felforgatására?* Miközben megfigyeljük az UBI megjelenését, továbbra sem világos, hogy

Az UBI hatékonyabb a kockázatértékelésben. E tekintetben olyan általánosítható keretrendszert javasolunk, amely egy újszerű *mobilitás-alapú kockázatértékelést* (MRA) valósít meg. Egy ilyen IoT-alapú megoldás várhatóan jelentősen javítja a prediktív teljesítményt, és felgyorsítja az iparág átalakulását. A kívánt kialakítás előtt a következő kérdésekre kell választ adnunk: (i) *milyen információkat kell gyűjteni az IoT segítségével?* (ii) *mely tényezők a legjelentősebbek a kockázati előrejelzésekben?* (iii) *milyen egyedi követelményeknek kell megfelelnie egy robusztus modellnek az autóbiztosítási üzletágban?*

Először is, a vállalkozásoknak mindig előnyös, ha a hatékony döntéshozatal érdekében finom szemcseméretű adatokat szereznek be. A megfelelő technológiák hiánya miatt az autóbiztosítók csak korlátozott személyes adatokhoz (pl. életkor és az önbevallás szerinti vezetési szokások) férnek hozzá, ami nemkívánatos eredményekhez, például jelentős bevételkieséshez és diszkriminációhoz vezet.¹ A korábbi szakirodalom új adatbevittelt vizsgált a baleseti kockázatok értékelésére, a naturalisztikus vezetéstől (pl. Guo és Fang 2013) a járműbe épített felvételekig (pl. Boyce & Geller 2002; Wahlberg 2004). Közösen megerősítik, hogy a nagy felbontású adatok javítják a modell előrejelzéseit. Egy korszerű rendszernek tehát mindenféle releváns információt figyelembe kell vennie, tekintettel a manapság gyorsabb adatátvitelre. A demográfiai alapú elemzéseken alapuló általános gyakorlattal szemben mi egy mobilitás-alapú megoldást javasolunk, amely a globális helymeghatározó rendszeren (GPS), a fedélzeti diagnosztikán (OBD) és a járműbe épített kamerákon (IVC) keresztül gyűjtött átfogó viselkedési adatokat is magában foglalja.

Másodszor, kritikus, de kihívást jelent a legjelentősebb balesethez hozzájáruló tényezők azonosítása a hatalmas mennyiségű kivont jellemzők közül, bár a több információ általában növeli a prediktív teljesítményt. A meglévő szakirodalom hajlamos az összes rendelkezésre álló jellemzőt felhasználni a modellek kalibrálásához (pl. Guo és Fang 2013; Wu et al. 2004). Egy ilyen konyhanyelven működő regressziós megközelítés csak akkor megvalósítható, ha a jellemzők száma kezelhető. Azonban gyakran túlilleszkedést toboroz az előrejelzésekben, ahogy egyre több prediktort veszünk figyelembe. Az 5G adatkommunikáció nagy mennyiségű adatot biztosít gyors és olcsó hozzáféréssel, ezért szükség van egy robusztus megoldásra, amely beágyazza a funkcióválasztást a tanulásba. Ebben a tekintetben kifejezetten modellezzük a funkcióválasztási folyamatot, mint a prediktív motor optimalizálásának egyik

kritériumát. Ezenkívül érdemes megjegyezni, hogy

¹ A ház 2019-ben elfogadja az autóbiztosítási diszkrimináció tilalmáról szóló törvényt, hogy bemutassa az esetleges visszaélészerű, demográfiai alapú árazás miatti esetleges diszkriminációt.

úgy tűnik, hogy a korábbi tanulmányok figyelmen kívül hagyják a téves osztályozások költségeit, különösen a hamis negatív eseteket (azaz a balesettel járó, biztonságosnak ítélt pályát). A javasolt modell alkalmazhatóságának általánossá tétele érdekében egy költségérzékeny célfüggvényt is beépítünk a modell kalibrálásába.

Harmadszor, a kockázatok előrejelzése a modellezés szempontjából természetesen osztályozási problémaként van megfogalmazva. A kívánatos tervezésnek meg kell felelnie az autóbiztosítási kontextusban felmerülő egyedi igényeknek, mivel az előrejelzések gyakorlásához különböző lehetőségek állnak rendelkezésre. Különösen fontos az előrejelzési teljesítmény (azaz a pontosság és a visszahívás) és a számítási hatékonyság (azaz az idő) közötti egyensúly megteremtése. A baleseti kockázatok valós idejű előrejelzése jelentősen eltér a szokásos előrejelzési gyakorlatoktól, mivel a vezetési viselkedés és az útviszonyok idővel változhatnak. A megoldásnak rugalmasnak kell lennie az adatfeldolgozás és a modellképzés terén, hogy kezelni tudja az ilyen vezetési dinamikát. A rendszernek emellett többszintű előrejelzésre is képesnek kell lennie. A jelenlegi gyakorlatok csak az út vagy a járművezető szintjén gyűjtik össze az adatokat. Ezzel szemben a javasolt megoldásunk a pályaszintű kockázatok meghatározásával kezdődik, és azokat vezetői szintű kockázatokként szintetizálja. A rugalmas modellezési megközelítés növeli a modell teljesítményét és új előrejelzési képességeket fejleszt ki (pl. pályaközpontú előrejelzések). Lehetővé teszi a biztosítótársaságok számára, hogy útvonal-szintű termékeket tervezzenek, amelyek tökéletesek lehetnek a taxiszolgáltatás, az autókölcsönzés és az autómegosztás (pl. Zipcar) számára. Ebben a kutatásban ezt a két igényt tekintjük egy olyan hatékony modell tervezési alapelvének, amely valós időben jósolja meg a többszintű kockázatokat.²

A fenti technikai megfontolások mellett az autóbiztosítás kontextusa ikonikus bemutatóhelyként szolgál a kutatók és a szakemberek számára annak megértéséhez, hogy a technológiák hogyan határozzák meg újra a legjobb gyakorlatot. Az ágazat alapvetően szolgáltatásorientált üzletág, és az autóbiztosítási szolgáltatók alapvető kompetenciája a pontos kockázatértékelésből ered. A legtöbb biztosítótársaság azonban komoly bevételkieséssel szembesül a járművezetők nem megfelelő profilalkotása miatt. A State Farm 7 milliárd dollárt veszített az autóbiztosítási kockázatvállalásán (2016Simpson 2017). Az információintenzív döntéshozatal jellegéből adódóan az autóbiztosítási ágazat jobban profitál a nagy adatokból és a fejlett technológiákból, mint más szolgáltatási ágazatok. IT-alapú

² A valós idejű rendszert Chen és Gurestrin (2016) szerint valós idejű feldolgozás elérése érdekében elosztott környezetben valósítjuk meg. Minden egyes klsztergépben 8 virtuális magot, 30 GB RAM-ot és két 80 GB-os SSD helyi lemezt használunk. A javasolt megoldást az Apache Spark MLLIB keretrendszerével integráljuk az XGBoost4J-Spark segítségével.

a termékinnovációk továbbra is megzavarják a jelenlegi üzleti modelleket, mivel az UBI kannibalizálja a piaci részesedéseket (Global Market Insights 2019). Ezért motiváltak vagyunk arra, hogy olyan újszerű informatikai megoldást javasoljunk, amely növeli az alapvető működési hatékonyságot az autóbiztosítási üzletágban.

A mobilitás-alapú kockázatértékelés fent említett lehetőségeinek feltárása érdekében a GPS, a fedélzeti diagnosztika és a járműbe épített kamerák által megfigyelt finom szemcsés vezetési viselkedést kezdjük vizsgálni. A GPS nyomon követi a járművezető minden egyes útvonalát, és összefoglalja a használati szokásait. A fedélzeti diagnosztikai rendszer a jármű érzékelőihez kapcsolódik, és rögzíti a biztonsági figyelmeztetéseket és a motor állapotát. Az arcfelismerő IVC-k figyelik a járműben végzett tevékenységeket és a fáradtság szintjét. A korábbi szakirodalom (pl. Xie et al. 2017) és a területi szakértők alapján a háromból kinyert szemantikai jellemzők átfogó készletét operacionalizáljuk. Kifejezetten a vezetési mobilitás két aspektusát vesszük figyelembe, nevezetesen a *vonás* és a *trajektóriát*. A vonások a járművezetők megkülönböztető vezetési viselkedésére (stílusaira) utalnak, míg a trajektória a jármű mozgássorozataiból áll, az utazás során a kontextuális tényezőkkel együtt.

Számos tervezési kihívással kell szembenéznünk, bár a fejlesztés egyszerűnek tűnik. A megoldásnak magas dimenziós adatokkal kell foglalkoznia, és többszintű elemzéseket kell végeznie. Előrejelző motorként egy XGBoost-alapú modellt fejlesztünk ki, amely az utazás és a járművezető szintjén is figyelembe veszi a kockázatokat. A gradiens boosting család választását a jellemzőválasztással való zökkenőmentes integrációnak, a nagy adatmennyiségek kezelésének (azaz a párhuzamos feldolgozásnak) az időhatékonyságának, valamint a ferde és ritka adatokkal szembeni robusztusságának tulajdonítjuk. A modell proaktívan beépíti az optimalizálásba a téves besorolás költségérzékenységét is.

Az MRA-megoldást valós vezetési adatok felhasználásával képezzük ki és értékeljük. A Fortune 500 biztosítótársaságok egyikének partnereként működünk együtt. A biztosító 2015 óta Kína egyik nagyvárosában IoT-alapú felügyeleti rendszer telepítését hajtja végre ügyfelei járművein. Hozzáférünk a teljes járművezetői állományból véletlenszerűen kiválasztott járművezetők 398 30 hetes vezetési pályafelvételeihez. A járművezetői pálya szintű adatok átfogó vezetési adatokból állnak, az utazási kilométertől kezdve a fáradtsági szintig. A rendőrségi és vállalati jelentésekben szereplő kapcsolódó baleseteket alapigazságként operacionalizáljuk. A járművezetők demográfiai adatait is összegyűjtjük. A

modell paramétereinek megtanulása érdekében 10-szeres keresztvalidálást végzünk, és számos robusztus eredményt kapunk magából a kalibrált modelltől és az empirikus elemzésekből.

Az értelmezhető modell azt a fontos üzenetet hordozza, hogy a viselkedési jellemzők relevánsabbak a baleseteknél

mint a pálya összefüggései. A funkcióválasztási folyamat felsorolja a balesetekhez leginkább hozzájáruló tényezőket, köztük a gyorsulást, a telefonhasználatot, az éles kanyarokat, a sebességet és az ásitást. Emellett megállapítjuk a járművezetők kockázatainak kiemelkedő heterogenitását, tekintettel az egyedi vezetési viselkedésre. Ezeket a heterogén kockázatokat nem lehet jól megragadni egy demográfiai alapú modellel. Ha jobban ráközelítünk, azt is észrevesszük, hogy a járművezetők kockázatai idővel változnak. Az ilyen váratlan minták a járművezetők különböző adaptív viselkedési jellemzőit mutatják (azaz a biztonságosabb vagy veszélyesebb vezetésre való önérvényesítés), amelyeket a modellezés szempontjából gondosan kezelni kell.

Egy sor empirikus validáció során először egy 2x2 értékelési keretet vizsgálunk a bemeneti jellemzők és a modellezési választások tekintetében. A javasolt megoldás (azaz a mobilitásból tanult XGBoost-alapú modell) a jobb választás. Egyrészt a mobilitásalapú megoldások következetesen jobbak a demográfiai alapúaknál (azaz a hagyományos gyakorlatnál), ami rávilágít a mobilitási adatok értékére. Másfelől a mi MRA-modellünk javítja az összes előrejelzési teljesítménymutatót, összehasonlítva más tanulási benchmarkokkal. Mennyiségileg a javasolt modell 43,45%-kal (41,82%), 29,46%-kal (27,87%) és 26,24%-kal (25,27%) magasabb felidézési (pontossági) értéket hoz, mint a logisztikus regresszió, a támogató vektorgép és a véletlen erdők.³ Másodszor, hogy teszteljük a modell robusztusságát a szokatlan felhasználási esetekre (azaz a ferde és kis képzési adatokra), a balesetek eloszlásának és a képzési adatok méretének manipulálásával rekonstruáljuk a képzési halmazt. Az eredmények azt mutatják, hogy az XGBoost-alapú kialakítás továbbra is a legrobusztusabb megoldás. Harmadszor, az alternatív modell-összehasonlítások rávilágítanak a többszintű modellezés értékére, amikor a javasolt modellt alternatív specifikációkkal (azaz út- és járművezető-központú) állítjuk szembe. Végül bemutatjuk a javasolt keretrendszer számítási hatékonyságát, amely kisebb képzési halmazt és rövidebb számítási időt igényel, miközben tisztességes teljesítményt ér el.

Ez a tanulmány több szempontból is hozzájárul a szakirodalomhoz. Racionalizáljuk a tárgyak internetének üzleti értékét működési szempontból. Korábbi kutatások dokumentálták, hogy az IoT-szerű technológiák, mint például a rádiófrekvenciás azonosítás (RFID), összekapcsolják a termékeket és az embereket, hogy működési szinergiájuk legyen. Továbbá azzal érvelünk, hogy a valós idejű döntéshozatalhoz szükséges intelligencia kiegészíti az IoT-t, hogy bomlasztó katalizátor legyen, mivel

³ A visszahívást választottuk elsődleges mérőszámnak, mivel a téves negatív esetek magasabb költséggel járnak, mint a téves pozitív esetek.

Olsen és Tomlin (2019) elképzelése. Mi tehát egy újszerű IoT-alapú megoldás kidolgozásával járulunk hozzá a szakirodalomhoz, amely jelentős technikai kihívások mellett a pontos kockázatértékelésekből operatív hatékonyságot nyer. A javasolt megoldás az adatok egyedisége, a modellezés újszerűsége és az üzleti alkalmazások tekintetében különbözik a kapcsolódó szakirodalomtól. Az elsők között vagyunk, akik átfogóan kivonjuk a járművön belüli tevékenységeket a balesetek előrejelzéséhez. Másodszor, a balesetek előrejelzését többszintű osztályozási problémaként keretezzük, míg a korábbi tanulmányok főként egyszintű elemzéseket végeznek (pl. Xie et al. 2017; Paefgen et al. 2014). Modellünk a kockázatok alakulását is figyelembe veszi, amelyet a szakirodalomban általában figyelmen kívül hagyni látszik. Végül bevezetjük a tanulási analitikát, hogy kezeljük az elfoglaltság-variáció kompromisszumot mind a társulások, mind az előrejelzések szükségletei esetében.

A megállapítások a gyakorlati szakemberek számára is megvalósítható stratégiákhoz vezetnek. Először is, megoldásunk bemutatja, hogy egy gépjármű-biztosítási szolgáltató hogyan használja ki az IoT-t az alapvető kompetenciái fejlesztésére. Az átfogó mobilitási adatok a legmagasabb adatfelbontást biztosítják, ami jelentősen javítja az előrejelzéseket. A valós idejű kockázatértékelések egyre kívánatosabbak, mivel a gyorsabb adatátvitel és a magasabb elemzési képességek költséghatékonyak az 5G hálózat korában. Másodszor, a gyakorlati szakembereknek fel kell ismerniük a kockázatok figyelemre méltó fejlődését. Az ilyen dinamikus változások figyelmen kívül hagyása a műveletek hatékonyságának hiányát eredményezi. A biztosítók azonban hozzászórtak ahhoz, hogy egy járművezető pontszámát féléves alapon állandó értéként modellezzék. Például, bár a Progressive magáévá teszi az UBI koncepcióját, Snapshot programja még mindig féléves díjakat számít ki egy fix, 30 napos időszak vezetési adatai alapján. Harmadszor, megoldásunk új biztosítási terméktervezéshez vezet. A GPS-követés segítségével a szolgáltatók a kockázatokat az út szintjén profilozhatják. Ez segít újra feltalálni az útszintű termékeket, amelyek tökéletesek a taxiszolgáltatások, az autóbérlés és az autómegosztás számára. Végül pedig az autógyártók tanulhatnak a tanulmány tanulságaiból. Az eredmények azt mutatják, hogy a figyelemelterelés és a fáradtság (a rendellenes gyorsítások, az éles kanyarok és a nagy sebesség mellett) jelentős szerepet játszik a balesetekben. A járművezetők és az utasok további védelme érdekében az autógyártók az arcfelismeréssel rendelkező IoT-képes eszközöket, például az itt vizsgált IVC-t, beépíthetik a fedélzeti online felügyeletbe. A rendszer proaktívan figyelmezteti a járművezetőket, amint

kockázatos vezetési magatartást észlel. Egy ilyen intelligens rendszer a gyártók, a járművezetők és a biztosítók számára egyaránt előnyös helyzetet eredményez.

A tanulmány további része a következőképpen folytatódik. Először az elméleti alapokat fejlesztjük ki a következők áttekintésével

a szakirodalom a 2. szakaszban. Az elméleti útmutatást követően a szakaszban meghatározzuk a kockázatértékelés problémáját, és egy új, IoT-alapú keretrendszert javasolunk a kockázatok előrejelzésére a szakaszban. A szakasz az adatok, eredmények és értékelések megvitatásával 4.részletezi az empirikus validálást. Ezt a kutatást mind az elméleti hozzájárulásokkal, mind a vezetői következményekkel zárjuk a szakaszban. 6.

2. Kapcsolódó irodalom

Először is áttekintjük a diszruptív technológiákkal kapcsolatos korábbi munkákat a műveletirányítás és az információs rendszerek szakirodalmában. Ezek a tanulmányok segítenek megérteni, hogy a fejlett technológiák hogyan teszik lehetővé az értékteremtést és a költségsökkentést általában. Ezután a szállítási műveletek kapcsolódó kutatásaira összpontosítunk, amelyek a járművezetők kockázatait értékelik különböző archív adattípusok felhasználásával.

2.1. IoT-alapú diszruptív technológia

Az akadémiai szféra az informatikai beruházások néhány évtizedes exponenciális növekedése miatt folyamatosan jelentős erőfeszítéseket tesz az informatika üzleti működésben betöltött értékének igazolása érdekében. Mithas et al. (2012) úgy véli, hogy az IT jelentős üzleti értéket teremt a bevételek növelése tekintetében. A bevételtermelő képességet a *termelékenység növekedése* (pl. Hitt & Brynjolfsson 1996) és az ellátási lánc koordinációjában a *költségsökkentés* (pl. Rai et al. 2006; Dong et al. 2009; Rai & Tang 2014) révén növelik. Ez a felfogás a technológiai fejlődéssel együtt folytatódik. A tárgyak internete és az 5G adatkommunikáció így várhatóan trilliárd dolláros hatást gyakorol majd az üzleti innovációra és működésre (McKinsey 2015). Olsen és Tomlin (2019) még azt is elképzelik, hogy az IoT, az embereken, gépeken és termékeken lévő, egymással összekapcsolt, *intelligens* érzékelők hatalmas tömege valós időben hatékonyabb döntéseket hoz majd.

Maga az IoT divatos kifejezés, miközben a hasonló koncepción alapuló alkalmazásokat is elfogadták. Az IoT-szerű technológiák hozzájárulása az üzleti értékekhez mindenütt jelen van. A rádiófrekvenciás azonosítás jól példázza az ilyen alkalmazásokat. Az RFID a gyakorlatban újradefiniálja a készletgazdálkodás folyamatát, ami arra ösztönzi a kutatókat, hogy megvizsgálják, mi lehet az optimális felhasználási eset (pl. Lee & Özer 2007; Heese 2007). Az RFID és a kontextuális tényezők közötti

kölcsönhatás figyelembevételére stilizált elemzési modelleket dolgoztak ki. Delen és munkatársai (2007) bemutatják, hogy az RFID nyomon követi a kulcsfontosságú teljesítménymutatókat (pl. átfutási idő egy kiskereskedelmi üzletben), és lehetővé teszi, hogy

az elosztóközpont és a kiskereskedelmi üzlet közötti koordináció. Ugyanebben az irányban a mobiltechnológiák és a helyalapú szolgáltatások a viselkedési kapcsolatokat a gépközpontú interakciókon túlra is kiterjesztik, nagy mennyiségű adatot generálva. Ezek a viselkedési adatok képessé teszik a vállalkozásokat a kereslet előrejelzésének gyakorlására (pl. Feng & Shanthikumar 2017; Ban & Rudin 2019) és a szükséges intézkedések előzetes megtételére. Emellett az empirikusok a különböző marketingstratégiák bevételnövelését vizsgálják a mobilitási adatok felhasználásával. Brynjolfsson et al. (2013) és Caro és Sadr (2018) például azt vizsgálja, hogy a mobil és a kiterjesztett valóság online jellegű vásárlási élményként fokozza a fogyasztók offline vásárlását. Ho és társai (2020) azt mutatják, hogy a geofencing hirdetések hatékonyan csábíthatják a fogyasztókat a helyi vállalkozások meglátogatására és az azt követő vásárlásra. Ghose et al. (2019) trajektória-alapú mobilkuponokat használnak az értékesítés előmozdítására egy bevásárlóközpontban. Röviden, az emberek, gépek és termékek közötti, technológiával lehetővé tett összekapcsolódásokból mindig új típusú hatékonyságnövekedés létezik.

A tárgyak internete nemcsak a mindenütt jelenlévő összekapcsolhatósága, hanem intelligens potenciálja miatt is egyedülálló. Más szóval, az IoT üzleti értékeket hoz létre (i) a csatlakoztatott érzékelők és eszközök segítségével, amelyek átfogóan gyűjtik és továbbítják a nagy mennyiségű adatot, valamint (ii) a mesterséges intelligencia és a gépi tanulás segítségével, amelyek zökkenőmentesen feldolgozzák az összegyűjtött adatokat és időben reagálnak. Egyértelmű, hogy egy IoT-megoldástól elvárják, hogy valós idejű, nagy dimenziójú adatokat gyűjtsön és dolgozzon fel előzetesen, és azonnali válaszokat adjon (Kumar et al. 2019; Olsen & Tomlin 2019). Amint azt Shacklett (2017) bemutatja, egy IoT-alapú élelmiszer-ellátási lánc olyan környezeti tényezőket, mint a hőmérséklet és a páratartalom, felügyel. Ennek megfelelően hozza meg az élelmiszer-utánpótlással, tárolással és szállítással kapcsolatos döntéseket az intelligens érzékelők és egy RFID-képes vállalati rendszer integrálásával. Bár egyszerűnek tűnik, egy IoT-alapú megoldás kifejlesztése sosem könnyű feladat. A kiskereskedelem témakörében például Ghose et al. (2019) feltárja, hogy technikailag kihívást jelent egy helymeghatározó rendszer megvalósítása a fogyasztók útvonalának valós idejű előrejelzésére.

Ennek megfelelően egy újszerű megoldás megtervezésével járulunk hozzá az IoT üzleti értékének feltörekvő témájához a szállítási műveletekben. Konkrétan a probléma relevanciája miatt a baleseti kockázatértékelésre összpontosítunk, mivel Chen és Jiang (2019) analitikusan kimutatták, hogy a vezetési

viselkedés részleteinek kihasználása jelentősen javítja a kockázatértékelést és a működési hatékonyságot egy autóbiztosítási ökoszisztémában. Adott

a jelenlegi gyakorlatot megzavaró újszerű megoldás hiánya miatt egy általánosítható, modellfüggetlen keretrendszert dolgozunk ki két rendszerösszetevő, nevezetesen az IoT-alapú adatelőfeldolgozó és a mesterséges intelligencia alapú előrejelző motor integrálására. Ezzel megtesszük az első lépést, hogy inspiráljuk a tudományos életet az IoT-ben rejlő lehetőségek teljes körű kiaknázására az üzemeltetésirányításban.

2.2. IoT-alapú kockázatértékelés és előrejelzés

A kockázatértékelés és -kezelés fontosságát a műveletirányítással foglalkozó szakirodalom is elismeri. A kutatók különböző területeken vizsgálják a kockázatokat (Sodhi et al. 2012), többek között a szolgáltatások (pl. Cohen 2018), a készletek (pl. Kouvelis & Li 2019), az ellátási lánc koordinációja (pl. Gao et al. 2017), a gyártás (pl. Ding et al. 2016) és a szállítás (pl. Chung et al. 2017) területén. Függetlenül attól, hogy milyen reaktív stratégiákat hajtanak végre a kockázatok mérséklésére, a kockázatkezelés mindig a bizonytalanság felmérésével kezdődik, ami előfeltételezi az átfogó információkhoz való hozzáférést. Amint azt Choi et al. (2016) tárgyalja, a tárgyak internete a legnagyobb felbontású nagy adatmennyiséget biztosítja, amely a kockázatértékelés első lépésének megtételéhez általában szükséges. Ez a tanulmány a közlekedésre összpontosít. A többi operatív tevékenységgel összehasonlítva a közlekedés, különösen a forgalomirányítás a viselkedési dinamika miatt bonyolult döntéshozatalt foglal magában. A kapcsolódó viselkedési bizonytalanságot jobban azonosítani és értékelni lehetne, miközben több adatot gyűjtenek. Ezért úgy véljük, hogy a kapcsolódó vállalkozások különösen profitálnak a fejlett technológiákból a közlekedéssel kapcsolatos kockázatok (pl. baleseti kockázatok) értékelésében. Ebben a tanulmányban azt kívánjuk bemutatni, hogy az IoT és a mesterséges intelligencia hogyan segít a járművezetők kockázatainak hatékony osztályozásában. A javasolt megoldás hozzájárul a szakirodalomhoz az adatok egyedisége, a modellezés újszerűsége és az üzleti innovációk terén. A táblázat kiválasztja azokat a kapcsolódó munkákat, amelyek a baleseti kockázatok azonosítására összpontosítanak, és értékelés a három szempontot illetően.

Először is, egy exkluzív jellemzőkkel rendelkező, valós világbeli adathalmazt kapunk, és figyelembe vesszük a sokrétű vezetési viselkedést. A korábbi szakirodalom a baleseteket egy sor általános tényezőnek tulajdonítja, amelyeket offline GPS-nyomkövetőkkel és fedélzeti diagnosztikával gyűjtöttek össze (pl. Xie et al. 2017; Jun et al. 2007), a vezetési sebességtől a napszakig. A járművezetők alapos

jellemzéséhez szemantikus vezetési jellemzőket és különböző járművön belüli tevékenységeket veszünk figyelembe, amelyeket az IVC-k arcfelismeréssel felismernek. Egy szerencsében lévő biztosítási szolgáltatóval együttműködve valós időben 500, hozzáférünk ezekhez a jellemzőkhöz a járművezető-út szintjén. A finomszemcsés adathalmaz

mélyrehatóbb elemzéseket tesz lehetővé, mint a meglévő tanulmányok, amelyek vagy a járművezetői szintű aggregált adatokra (pl. Guo & Fang 2013; Wu et al. 2014), vagy az utazási szintű keresztmetszeti adatokra támaszkodnak.

Másodszor, ez a tanulmány a modellezés szempontjából különbözik a korábbi szakirodalomtól. A meglévő kutatások általában egyetlen szinten (azaz utazás vagy járművezető) végeznek elemzéseket. A többszintű adatok természetének figyelmen kívül hagyása negatív statisztikai következményekhez és a heterogenitás gyenge megértéséhez vezethet, ha van ilyen (Osborne 2000). Az egyszintű modell az egyéni profilalkotást is akadályozza a kontextuális hatásoktól. Az elméleti útmutatásokat követve (pl. Machin & Sankey 2008; Toledo et al. 2008; Musicant et al. 2010) a mobilitást két szemantikai összetevőre (azaz viselkedési vonásokra és pálya jellemzőire) osztottuk fel egy többszintű hierarchiában. Modellünk nem csak az egyes járművezetők kockázatát profilozza azáltal, hogy viselkedési kockázatait az utazás szintjén aggregálja, hanem a környezeti tényezőket is beépíti az utazások között. Ez a modellezési megközelítés a korábbi munkákhoz képest (pl. Paefgen et al. 2014; Wu et al. 2014) szigorúbban magyarázza a balesetet okozó tényezőket. Modellünk a kockázat előrejelzésében is előnyt jelent. A statikus előrejelzési feladatoktól eltérően (pl. Guo & Fang 2013; Xie et al. 2017), a baleseti kockázatok előrejelzése természetesen viselkedésbeli változásokat és kontextuális dinamikát foglal magában, amelyek az idő múlásával fejlődnek. Többszintű modellünk mind az út-, mind a járművezetői szintű kockázatokat képes gyakorolni, miközben pontosan kontrollálja a fejleményeket. Az előrejelzések segíthetnek a meglévő termékek (pl. rendszeres féléves díj) rövidebb idő alatt történő átárazásában és új termékek (pl. utazásalapú biztosítás) kialakításában.

Végül egy új modellt javasolunk, amely kiválóan alkalmas asszociációk és előrejelzések készítésére. Az asszociáció elemzése viszonylag nagy előfeszítésű modellt igényel az adatgeneráló folyamat hátrafelé történő kikövetkeztetéséhez; ezzel szemben az előrejelzések gyakorlása viszonylag nagy varianciájú modellt részesít előnyben, hogy a helyes azonosítót válassza ki az eredmények halmazából előre (Dietterich et al. 1995). A kettő közötti konfliktus a gépi tanulásban az előítélet-variancia dilemmára utal. Míg a korábbi tanulmányok olyan magas előfeszítésű modelleket fogadnak el, mint a regresszió (pl. Paefgen et al. 2014) és a klaszterezés (pl. Guo & Fang 2013), mi egy kiegyensúlyozott tanulási modellt dolgozunk ki. Modellünk a tárgyalt többszintű tanulás optimalizálása

érdekében a funkcióválasztás segítségével kezeli a kompromisszumot. Röviden, egy új tanulási technika bevezetésével, valamint egyedi adatokkal és újszerű modellezési megközelítéssel járulunk hozzá a szakirodalomhoz, hogy megvilágítsuk a kockázatértékelést és -kezelést.

Táblázat A baleseti kockázatelemzés kiválasztott1. tanulmányai

	Referenciaadatok		IVC	Utazás	Modellezés Vezető	Idő*	Egyesület	Alkalmazás	Előrejelzés	Kutatás fókusz
Ez a tanulmány	✓	✓	✓	✓	✓	✓	✓	✓		Újszerű IoT-alapú megoldás kifejlesztése a baleseti kockázatok előrejelzése
Xie et al. (2019)				✓	✓	✓				Újszerű IoT-alapú megoldás kifejlesztése a baleseti kockázatok előrejelzése pályák segítségével
Xie et al. (2017)	✓			✓				✓ A		baleseti kockázatok földrajzi-rácsos szint előrejelzése a
Paefgen et al. (2014)	✓	✓			✓		✓			A balesetek és a röppályaadatok társítása
Wu et al. (2014)		✓	✓		✓		✓			A forgalmi események és a demográfiai adatok összekapcsolása
Guo és Fang (2013)		✓	✓		✓		✓	✓ A		balesetek és a demográfiai adatok összekapcsolása és a baleseti kockázatok előrejelzése k-közép segítségével
Jun et al. (2010)	✓				✓		✓			A balesetek és a vezetési sebesség összefüggése
Musicant et al. (2007)	✓	✓		✓	✓		✓			A balesetek és a vezetési idő összefüggése
Jun et al. (2011)							✓			A balesetek és a vezetési szokások összekapcsolása

Megjegyzés. * Az idő a kockázat időbeli alakulásának figyelembevételére utal.

** Xie et al. (2019) közlekedési videókat használnak a járműpályák konstruálásához.

<https://ssrn.com/abstract=3661062>

3. Mobilitás-alapú kockázati profilalkotás problémája

Ebben a szakaszban a *mobilitás-alapú kockázatelemzést* osztályozási problémaként fogalmazzuk meg, és a problémát három, egymással összefüggő modulra bontjuk. Ezután röviden áttekintjük a javasolt keretrendszert, amely ennek megfelelően magában foglalja a három komponensét.

3.1. A probléma meghatározása

Legyen $U = \{u\}_i, i = 1, 2, \dots, n$, a járművezetők halmaza, és $S = \{s\}_{ij}, j = 1, 2, \dots, m$, az n járművezető által megtett m utazás halmaza. A járművezető ui kockázatot rij vesz részt az sij utazáson, ahol a balesetet a vezetési magatartásával (pl. Toledo et al. 2008) és az utazással kapcsolatos összefüggésekkel (pl. Musicant et al. 2010; Jun et al. 2011) összefüggőnek tekintjük. A rij tehát két szemantikai komponensből, a viselkedési és a kontextuális kockázatokból áll.

A korábbi munkák többsége általában a kettőt egyesíti, és a kockázatelemzéseket vagy a vezetői vagy utazási szinten (pl. Guo & Fang 2013). Ez az egyszintű modellezési megközelítés azonban nemkívánatos statisztikai következményekhez vezethet (Osborne 2000). Többszintű elemzésekre van szükség, mivel a viselkedési és kontextuális dinamika különböző ütemben tarthat fenn. A baleseti kockázatokat többszintű módon elemezzük, tekintettel a hierarchikus jellegre mind viselkedési, mind adatmodellezési szempontból. Így a járművezető *viselkedési jellemzői* a járművezetés sajátos stílusára utalnak, amely nem változó az egyes utak során, de megkülönböztethető a többi járművezető stílusától. Ezzel szemben a *pályakörnyezeti összefüggések* olyan környezeti feltételeket rögzítenek, amelyek az egyes utak során változhatnak. Ennek megfelelően az i járművezető kockázatát rij a j utazáson a következőképpen határozzuk meg

$$r_{ij} = r^B + r_{ij}^T + e_{ij}, \quad (1)$$

ahol r^B és r a viselkedési jellemzőkből és a pályakövetési összefüggésekből származó kockázatokot T jelentik. IoT- az olyan eszközök, mint a fedélzeti diagnosztika, gazdag mobilitási információkat gyűjtenek az út szintjén. A mobilitási adatokat, M , operacionalizálják és két megfelelő jellemzőcsoportra osztják. A viselkedési jellemzők, $B = \{\mathbf{b}\}_{ij}$, a kapcsolódó viselkedési állapotok halmaza az utazás szintjén, míg a trajektória jellemzők, $T = \{\mathbf{t}\}_{ij}$, az utazási szintű jellemzők vektora.

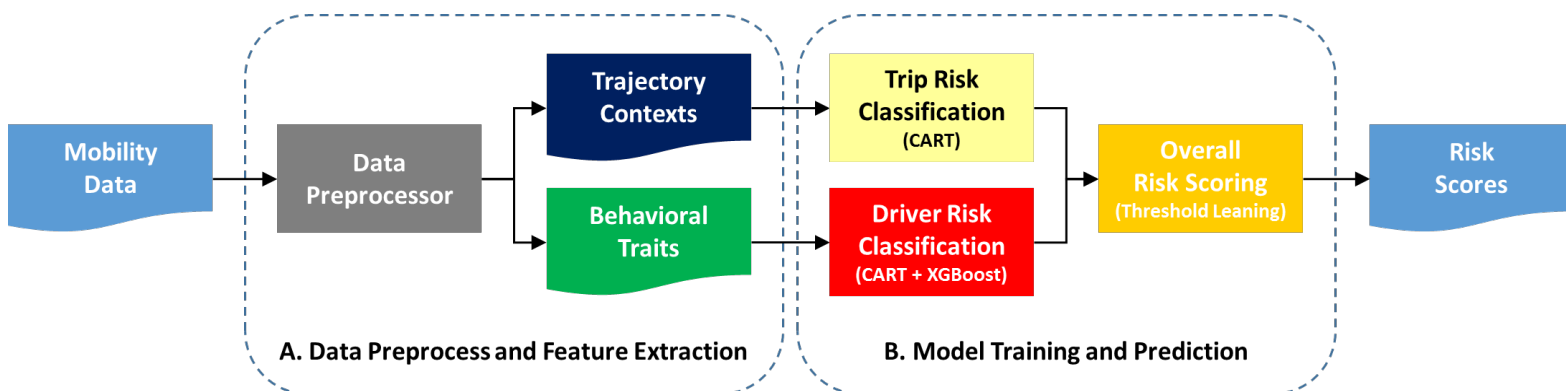
MRA probléma. Adott U járművezetők halmaza, meglévő útjaik S jellemzőivel B és T , az i járművezető r_{ij} baleseti kockázatának előrejelzése a j útra a hibák minimalizálásával, ahol a kockázatot az (1) egyenlet határozza meg.

Három konkrét feladatot kell elvégeznünk, mivel a mobilitás-alapú kockázatértékelést előrejelző feladatnak tekintjük, beleértve (i) az utazás kockázati osztályozását (TRC), (ii) a járművezető kockázati osztályozását (DRC) és (iii) az általános kockázati pontozást (ORS). Az utazási szintű kockázatok értékelésével kezdjük. Az S baleseti címkék ismeretében a TRC modul egy T trajektórijellemzőkből tanult osztályozó. Végül a modul a kimeneteként az S minden egyes útját valószínűséggel prediktív módon címkézi. Másodsor, a járművezetői kockázatosztályozó az egyes járművezetők kockázatát a B viselkedési jellemzőkből tanult utazások kockázatainak aggregálásával profillozza. A DRC modul a járművezető szintjén az előre jelzett valószínűségi címkéket adja ki. Végül, a címkék ismeretében az ORS modul kalibrálja a paramétereket, és a TRC és DRC kimeneteinek súlyozásával megjósolja a kockázati pontszámokat, $R-t$.

3.2. Keretrendszer áttekintése

A javasolt keretrendszert az 1. ábra szemlélteti. Az A. és B. rész az adatok előfeldolgozására és a modellképzésre összpontosít. Az A. részben az adatelőfeldolgozó az IoT által gyűjtött mobilitási adatokat pályaösszefüggések és viselkedési jellemzők szerint kategorizálja. Érdemes megjegyezni, hogy a viselkedési jellemzők lehetnek torzak, ritkák és a járművezetők között eltérőek. A B. részre áttérve, az utazás kockázati osztályozóját a trajektórijellemzők segítségével képezzük ki ennek megfelelően, míg a járművezető kockázati osztályozója a viselkedési jellemzőkből tanul. Az ORC végül megjósolja az általános kockázati osztályzatokat.

1. ábra. A javasolt keretrendszer áttekintése



A javasolt keretrendszer adat- és modellfüggetlen, hogy további adatokat (pl. érdekeltségi pontok jellemzőit) is befogadhatunk, és a modulokat más tanulási módszerekkel is megvalósíthatjuk. Ez a rugalmas keretrendszer nem csak konkrét közlekedési beállításokra, például a járatkésések értékelésére,

hanem általános

operatív kontextusban, amíg a pályaminták kritikus szerepet játszanak a hasznosságban, mint például a fogyasztók offline preferenciaprofilja és a helyi bűncselekmények felderítése.

4. MRA algoritmus

A valós idejű előrejelzéshez mobilitás-alapú megoldást tervezünk a baleseti kockázatok értékelésére a hatékony előrejelző teljesítmény és a hatékony számítási idő érdekében. Egy adatelőfeldolgozót fejlesztünk ki az előre meghatározott jellemzőknek az IoT által gyűjtött adatokból történő kinyerésére. Ezután az MRA algoritmust osztályozóként modellezzük az osztályozási és regressziós fa (CART), az eXtreme Gradient Boosting (XGBoost) és a küszöbérték-tanulás segítségével.

4.1. Adat előfeldolgozás

Tegyük fel, hogy az IoT-képes eszközök folyamatosan töltik fel az adatokat az MRA-rendszerbe. Az előfeldolgozó az időbélyegzett mobilitási adatokat (pl. földrajzi koordináták és arcképek) útvonal-szintű jellemzők halmazává alakítja. Egy utazás (azaz egy útvonal) a következőképpen definiálható.

Meghatározás. Az i járművezető s_{ij} utazása s tételeinek $s_{ij} = \{(t^1, t^1s), \dots, (t^p, t^ps)\}$ időben rendezett halmaza, jjjjj

ahol $t^p = (p^lat, p^long)$ a p hely földrajzi koordinátái, és t^ps a megfelelő időbélyeg. jjjjj

A szakterületi ismeretek alapján az S -t három további feltétel teljesítésével operacionalizáljuk, mint például (i) az utazás időtartama legalább négy percre tart ($t^ps - t^1s \geq \text{perc}4$), és (ii) a t^1s és t^ps sebességek t^1s és t^ps sebességek 0, jjjjj és (iii) az utazáson belüli bármely ideiglenes megállásnak négy percnél rövidebbnek kell lennie. Ezt követően kiszámíthatjuk az egyes utak kilométereit és időtartamát.⁴

Az előfeldolgozó a fedélzeti diagnosztikából és a járműfedélzeti kamerákból származó szemantikus jellemzőket az S megadott időbélyegegek segítségével képezi le. Míg az OBD-adatok útvonal szinten vannak összesítve, az előfeldolgozó egy előre betanított arcfelismerő modellt alkalmaz az IVC által gyűjtött videók elemzésére.⁵ Az adatfeldolgozás végén a Trip s_{ij} mobilitási jellemzők halmazával rendelkezik, $m_{ij} = \{b_{ij}, t_{ij}\}$, ahol a b_{ij} és a t_{ij} a viselkedésre és a trajektóriára utal.

⁴ A Haversine-képlet segítségével kiszámítjuk az utazáson belüli összes szegmens távolságát, és összegezzük azokat.

⁵ Az arcfelismerő algoritmust mély rekurrens neurális hálózatra alapozva valósítjuk meg. A modellt előzetesen betanítjuk 200 járművezető 10 szakértő által megjelölt 1 hetes vezetési videóinak felhasználásával. A figyelembe vett címkék között szerepel a telefonhasználat, dohányzás, ásitás és szemlehungyás. Az előre betanított modell 96%-os pontosságot és 97%-os visszahívást eredményez.

jellemzői, illetve. Ezután megkezdjük a baleseti kockázatok értékelését az utazás szintjén, miközben a rendőrségi és vállalati baleseti jelentéseket, $Y = \{y\}_{ij}$, tekintjük alapigazságnak, ahol $y_{ij} = 1$, ha az s_{ij} utazás balesetben érintett, és $y_{ij} = 0$, egyébként.

4.2. Utazási kockázat besorolása

A korábbi szakirodalom arra ösztönöz bennünket, hogy az utazási kockázatok osztályozásához figyelembe vegyük a pálya jellemzőit (T) és az alapigazságot (Y). Döntésfa-alapú modellt alkalmazunk az S címke előrejelzésére. Konkrétan a CART-ot (Breiman et al. 1984) alkalmazzuk Gini tisztátalanság-számítással, hogy olyan kockázati valószínűségeket kapjunk, amelyek könnyen integrálhatók a teljes kockázati pontozási modellbe. Az osztályozót úgy képezzük ki, hogy kiszámítjuk a T egyes jellemzőinek Gini-indexét, $Gini(S, T^k)$, ahol T^k a T k^{th} eleme. A faépítési folyamat iteratív módon keresi azt a jellemzőt, amelynek az index értéke a legalacsonyabb.

$$Gini(S, T^k) = \frac{|S|_1}{|S|} Gini(S)_1 + \frac{|S|_2}{|S|} Gini(S)_2 \quad .(2)$$

ahol S_1 és S_2 az S részhalmazai, és $|S|_1$, $|S|_2$ és $|S|$ az S_1 , S_2 és S halmazban lévő utazások számát jelölik.

Intuitív módon T^k az S -t S_1 -re, S_2 -re osztja, és az egyes részhalmazok Gini-tisztasága a következőképpen fogalmazódik meg

$$Gini(S^*) = 1 - \Pr(S^* = 1)^2 - \Pr(S^* = 0)^2 \quad .(3)$$

Ha T^k bináris változó, akkor a fa az értéke alapján osztja fel az S -t; ha nem, akkor a részhalmazok felosztása a $Gini(S, T^k)$ értékét minimalizáló küszöbérték keresésével történik. A CART osztályozó a kockázatokat, r^T , az út szintjén a trajektórijellemzők, T , segítségével adja ki.

4.3. A járművezető kockázati besorolása

Két feladatot kell elvégeznünk a járművezetők viselkedési jellemzőinek profilozásához. Konkrétan: (i) osztályozzuk az utazási kockázatokat a viselkedési jellemzők (B) segítségével, és (ii) aggregáljuk ezeket az osztályozott kockázatokat az utazás szintjén, hogy megkapjuk a viselkedési jellemzőket a járművezető szintjén. Először is, megismételjük a TRC gyakorlatát a B figyelembe vételével, hogy a B

r^B . Egyszerű, de naiv dolog a járművezető összes utazási kockázatát átlagolni, hogy hozzáférjünk a kockázatához. Ez a megközelítés a járművezető heterogenitásának figyelmen kívül hagyása miatt soha

nem előnyös, ami jelentősen csökkenti az előrejelző erőt.

Más szóval, egy szemantikai viselkedési jellemző hatása a járművezetők között eltérő lehet.⁶ Ebben a tekintetben Chen és Guestrin (2016) nyomán egy stilizált eXtreme Gradient Boosting algoritmust fejlesztünk ki a variancia-előny kompromisszum kezelésére. Az XGBoost-alapú modellünk egy költségfüggvényt tartalmaz, amely a járművezetők közötti eltérés enyhítésével javítja az előrejelzéseket. Emellett érdemes megjegyezni, hogy a DRC modell hatékonyan kezeli a ferde és ritka adatokat, mivel az XGBoosttól örökölte a párhuzamos fa-tanulást.

Tegyük fel, hogy az XGBoost-alapú modell az s_{ij} utazást $x_{s_{ij}}$ kockázattal ($y = 1$) balesetet okozó utazásként jósolja. Az előre jelzett kockázatot a SoftMax függvény segítségével átalakítjuk a megfelelő valószínűséggé, $\text{Pr}S_{ij}(y = 1)$:

$$\text{Pr}S_{ij}(y = 1) = \frac{e^{x_{s_{ij}}(y=1)}}{e^{x_{s_{ij}}(y=1)} + e^{x_{s_{ij}}(y=0)}} \quad (4)$$

A költség-érzékeny cél elérése érdekében (He and Garcia 2009) a következő kereszt-entrópia függvényt (Shore and Johnson 1980) tervezzük meg és minimalizáljuk a modell paramétereinek megtanulása érdekében

$$\Psi = \lambda - \sum_{i=1}^n \sum_{j=1}^{m_i} I_{s_{ij}}(y = 1) \ln \left(\frac{I_{s_{ij}}(y = 1)}{\text{Pr}S_{ij}(y = 1)} \right) + \sum_{i=1}^n \sum_{j=1}^{m_i} I_{s_{ij}}(y = 0) \ln \left(\frac{I_{s_{ij}}(y = 0)}{\text{Pr}S_{ij}(y = 0)} \right) \quad (5)$$

ahol λ az összes faosztályozó regularizált paramétere. $I_{s_{ij}}(y_{ij} = 1)$ egy jelzőfüggvény, amelyben

$$I_{s_{ij}}(y_{ij} = 1) = \begin{cases} 1, & \text{ha } y_{ij} = 1 \\ 0, & \text{egyébként} \end{cases} \quad (6)$$

Végül az iteratív gradiensfa-növelő feldolgozást végezzük el Ψ optimalizálására az első- és másodrendű közelítések (azaz a gradiens **gredS** és a Hess-mátrix **HessS**) segítségével, amelyek a következőképpen vannak megadva.

$$\mathbf{gred} = \frac{\partial \Psi}{\partial x_{s_{ij}}(y=1)} + \frac{\partial \Psi}{\partial x_{s_{ij}}(y=0)} = [\text{Pr}S_{ij}(y = 1) - I_{s_{ij}}(y = 1)] + [\text{Pr}S_{ij}(y = 0) - I_{s_{ij}}(y = 0)] \quad (7)$$

és

$$\mathbf{Hess} = \frac{\partial^2 \Psi}{\partial x_{s_{ij}}^2(y=1)} + \frac{\partial^2 \Psi}{\partial x_{s_{ij}}^2(y=0)} = 2\text{Pr}S_{ij}(y = 1) \cdot \text{Pr}S_{ij}(y = 0) \quad (8)$$

illetve. Az XGBoost-alapú osztályozóval rendelkezünk, amely a megfelelő kockázatokat, r_i^B , a vezetónél adja ki.

szint a viselkedési jellemzők segítségével, *B*.

⁶ Például: A járművezető tapasztalt vezető, aki gyorsan vezet, míg *B* járművezető amatőr vezető, de hajlamos gyorsan vezetni. Bár a gyorsajtás általában véve veszélyes, ennek a viselkedési állapotnak a hatása a kettőnél eltérő lenne.

4.4. Általános kockázati pontszámítás

A TRC és DRC modulok kimeneteit (azaz r^T és r^B) vesszük alapul a kockázati pontszámok előrejelzéséhez.

j

figyelembe kell venni, hogy a baleseti kockázatok megfigyeletlen tendenciák miatt változhatnak, ami a kockázatok heti rendszerességgel történő értékelését indokolja.

alapon. Ebben a tekintetben meg kell kapnunk az i meghajtó r^B viselkedési kockázatát és az r^T útvonalkockázatot a

$iwjw$

j a w . héten. Összesen W hétből kiindulva lineáris súlyozási vektort alkalmazunk a következőképpen

$$\omega(|\omega| = W, \omega_q = q) \quad (9)$$

az i járművezető időben változó kockázatának időbeli alakulásának kiegyenlítésére. A közelmúltbeli minták fontosságának kiemelése érdekében időbeli súlyozásra kerül sor, majd ezt követi a reprezentatívabb hetekre összpontosító jelentős súlyozás. Ezáltal azt várjuk, hogy a súlyozott viselkedési jegyek és a pályaösszefüggések robusztusabbak lesznek.

Végül r^B és r^T szintetizáljuk az i járművezető j útra vonatkozó kockázati pontszámának megjósolására egy lineáris

ij

specifikáció,
mint

$$r_{ij} = \alpha^B r + \beta r^T \quad (10)$$

ahol r és r^T normalizált, és α és β a két típus szignifikanciáját jelző paraméterek.

ij

kockázatok. Bayes-féle optimalizálási módszert (Snoek et al. 2012) használunk az α és β becslésére, θ , a balesetek címkézésének küszöbértékének lehorgonyzásával.

5. Empirikus validálás

Elemzéseket végzünk módszerünk értékeléséhez IoT-képes eszközökön keresztül gyűjtött valós adatok felhasználásával. Ezt a szakaszt az adatok részletezésével kezdjük. A kalibrált modell értelmezése után tárgyaljuk az értékelési stratégiánkat, a teljesítménymérőket, a benchmark módszereket és az empirikus eredményeket.

5.1. Adatok

A Fortune 500-as listáján szereplő autóbiztosítási szolgáltatóval együttműködve jutunk hozzá a finom szemcsés adatokhoz. Ügyfelei viselkedésének jobb megértése érdekében a szolgáltató IoT-alapú technológiákat alkalmaz, és megkezdi a megfigyelőrendszerek telepítési tervének megvalósítását Kína

egyik nagyvárosában a 2015. Gyűjtjük a mobilitási adatokat a következő személyekről

véletlenszerűen kiválasztott 398 járművezető⁷ IoT-képes GPS-nyomkövetők, fedélzeti diagnosztikai eszközök és a járműbe épített kamerák segítségével augusztus és 2017március között Ahogy az2018. ábra is mutatja2, a telepített eszközök akkor lépnek online állapotba és kezdik el figyelni a járműve mobilitását, amikor a járművezető beindítja a motort. A GPS-nyomkövető másodpercenként15 feltölti a földrajzi koordinátákat, és segít a sávváltások, éles kanyarok, valamint az út kilométereinek és időtartamának szemantikai jellemzőire következtetni. Az OBD valós időben rögzíti a viselkedési állapotokat, beleértve a gyorsításokat, a kemény fékezéseket és a frontális ütközésre vonatkozó figyelmeztetéseket. Az arcfelismerő IVC felismeri a járműben végzett tevékenységeket (pl. telefonálás, dohányzás, ásítás és szemlehungyás). A fenti GPS-kódolt útvonalakat a vállalat földrajzi információs rendszeréhez képezzük le, hogy további kontextuális tényezőket, például sebességkorlátozást nyerjünk ki. Az időbélyegzők az utazás napszakát jelzik (pl. Paefgen et al. 2014).

Ábra IoT-képes2. eszközök

(a) Fedélzeti diagnosztika GPS-követéssel



(b) Járműbe épített kamera



A 2. táblázat ezeket a szemantikus vezetési állapotokat 11 viselkedési és 5 pálya jellemzőbe csoportosítja. Ezután a rendőrség és a vállalatok baleseti jelentéseit használjuk fel a megalapozott igazság megalkotásához. A 3. ábra egy járművezető egy adott nap több útvonalát mutatja be, ahol a piros szegmens egy balesetet okozó útra utal. Ezenfelül a demográfiai adatokat a jelenlegi gyakorlat (azaz a demográfiai alapú profilalkotás) megismétlése érdekében szerezzük be, beleértve az életkort, a nemet, a jövedelmet, az iskolai végzettséget, a vezetési időt, a baleseti előzményeket stb.

⁷ A számítási korlátozások miatti esetleges szelekciós torzítás elkerülése érdekében véletlenszerűen választottuk ki a járművezetők 0,1%-os mintáját a járművezetői állományból. A minták a teljes állományt reprezentálják, tekintettel a két demográfiai eloszlás statisztikailag azonos eloszlására.

A mobilitási jellemzők táblázatos leírása2.

Változó	Leírás
Az érdeklődés eredménye	
<i>Crash</i>	A balesetben való részvétel bináris mutatója
Viselkedés	
<i>Sebesség</i>	A teljes út átlagsebessége (<i>kilométer/óra</i>)
<i>Gyorsítások</i>	# gyorsulások $\geq 1,8 \text{ méter/sec}^2$
<i>HardBreaks</i>	# Lassulások száma $\geq 1,8 \text{ méter/sec}^2$
<i>SharpTurns</i>	# A fordulatok száma, ahol a szögsebesség $\geq 30 \text{ radián/sec}$
<i>LaneChanges</i>	# A fordulatok száma, ahol $10 \text{ radián/sec} < \text{szögsebesség} < 30 \text{ radián/sec}$
<i>FrontAlerts</i>	# frontális ütközésre figyelmeztető jelzések
<i>Dohányzás</i>	# dohányzási magatartás
<i>PhoneUsing</i>	# a telefonnal kapcsolatos viselkedések
<i>Yawning</i>	# of yawns
<i>EycClosing</i>	# of eye-closing $\geq \text{sec}1$
<i>HighEngineRPM</i>	# motorfordulatszám $>$ a jármű alapértelmezett felső motorfordulatszám
Pálya	
<i>Kilométerek száma</i>	Az utazás teljes távolsága (<i>kilométer</i>)
<i>Időtartam</i>	Az utazás teljes menetideje (<i>perc</i>)
<i>SpeedLimit</i>	Egy út átlagos közúti sebességhatára (<i>kilométer</i>)
<i>Nappali</i>	Nappali utazás bináris mutatója
<i>Hétköznap</i>	Hétköznap utazás bináris mutatója

A mozdonyvezető röppályájának ábrázolása3.



A táblázatban összefoglaljuk a mobilitási jellemzők statisztikáit az út szintjén A *baleset3*., az előrejelzések eredménye, az *i* járművezető *j* útján a *t* héten bekövetkezett balesetre utal. A *baleset* átlaga 0,0013, ami azt jelzi, hogy a teljes adathalmazban a baleseti arány 0,13%. A járművezetők átlagosan a következő sebességgel közlekednek

20.29 (km/óra) 0 és 132 közötti tartományban. Egy út során a rendellenes gyorsulások, az éles gyorsulások átlagos számai

fordulatok és kemény törések 1,48, 2,87 és 2,35. A megfelelő tartományok és a standard eltérések jelentősen eltérnek egymástól. A járművezető figyelmeztetést 1.89 kap a szenorból, hogy tartsa a biztonságos távolságot. Megfigyelhető, hogy a járművezető figyelmét 1,63, illetve 0,09 alkalommal vonja el a telefon és a dohányzás. Az *ásítás* és az *EycClosing* a fáradtság szintjét jelzi; a járművezetők átlagosan *ásítást* 1.78 és *hosszú* 0.25 szemlezárást végeznek. Ami a pálya jellemzőit illeti, egy tipikus pálya hossza 9,98 kilométer és 27,32 perc. Az átlagos sebességhatár az összes út során 34,22 (km/óra). Az utazások 78%-a nappal történik, 87%-uk pedig hétköznap. A demográfiai adatok összefoglalása azt mutatja, hogy a járművezetők 68%-a férfi. A járművezetők átlagéletkora 41 év, míg az átlagos vezetési idő körülbelül 16 év.

Táblázat összefoglaló 3.statisztikák

Változó	Átlag	Std. Dev.	Min.	Max.
Az érdeklődés eredménye				
<i>Crash</i>	0.0013	0.0035	0	1
Mobilitás				
<i>Sebesség</i>	20.29	45.22	0	133
<i>Gyorsítások</i>	1.48	3.44	1	26
<i>HardBreaks</i>	2.87	7.33	1	23
<i>SharpTurns</i>	2.35	4.19	1	21
<i>LaneChanges</i>	5.34	6.28	4	49
<i>FrontAlerts</i>	1.89	2.73	0	18
<i>Dohányzás</i>	0.09	1.78	0	20
<i>PhoneUsing</i>	1.63	6.01	0	34
<i>Yawning</i>	1.78	3.57	0	19
<i>EycClosing</i>	0.25	2.73	0	9
<i>Kilométerek száma</i>	9.98	7.99	1.50	67.50
<i>Időtartam</i>	27.32	25.29	5	126
<i>SpeedLimit</i>	34.22	6.38	15	120
<i>Nappali</i>	0.78	0.23	0	1
<i>Hétköznap</i>	0.87	0.56	0	1
Demográfiai adatok				
<i>Gender</i>	0.68	0.10	0	1
<i>Kor</i>	41.10	7.39	23	62
<i>Jövedelem</i>	5,421.21	1,200.23	4,589	9,890
<i>Oktatás</i>	12.77	2.23	9	18
<i>Tenure</i>	15.61	6.95	1	37
<i>CrashHistory</i>	2.11	1.02	0	10
<i>Népesség</i>	5.31	4.20	0.57	33.92

5.2. A modell kalibrálása és értelmezése

Az MRA algoritmus paramétereit a következő folyamat során kalibráljuk. A hibákat minimalizáló modellparaméterek megtanulása érdekében 10-szeres keresztvalidálást végzünk. A kockázati pontszámok előrejelzése előtt a paraméter

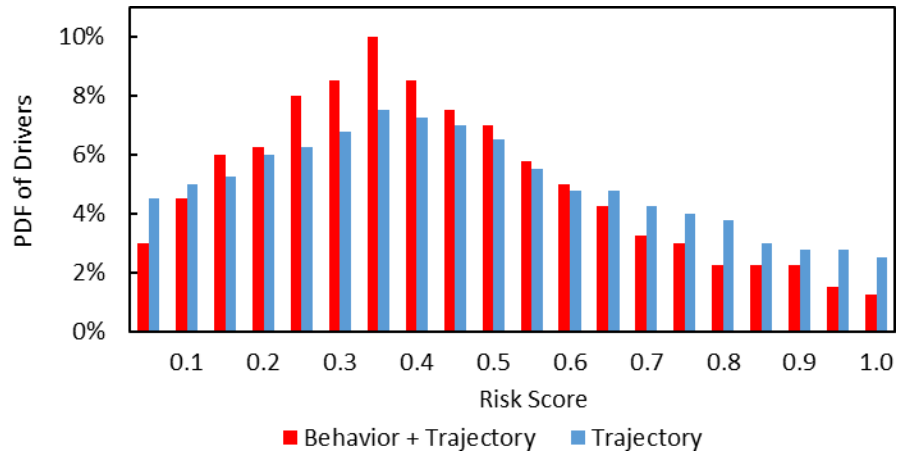
A α , β és θ értékeket α és β kezdeti értékkel becsüljük meg A0.5 α és β értékek 0.5.megválasztása úgy tekinthető, mint a viselkedési, illetve a kontextuális kockázatok által biztosított információk egyenlő súlyozása. A θ szintén a

0.5 a nemkívánatos torzítások elkerülése érdekében. Snoek et al. (2012) nyomán a hármat a Bayes-féle optimalizációs módszerrel kalibráljuk. Intuitív módon a paraméterek finomhangolása a modelteljesítményre adott visszajelzés alapján történik. A modellkalibrálás során számos értelmezhető betekintést nyerünk a vezetési viselkedésbe.

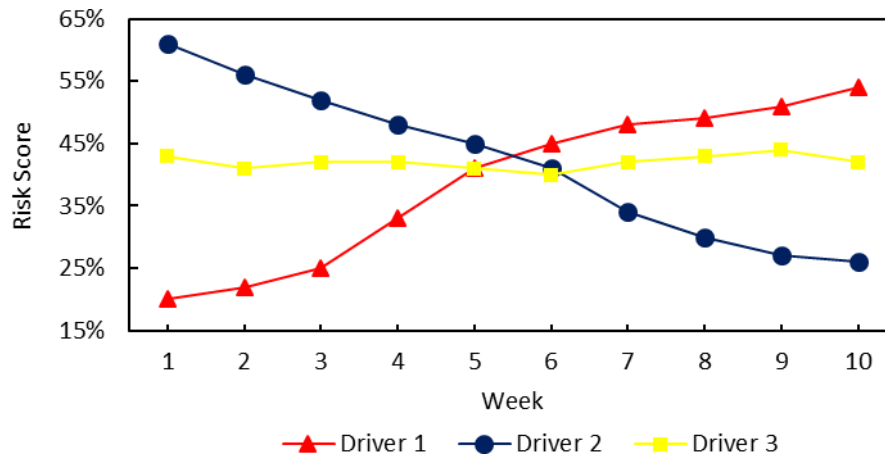
Először is, a járművezetők kockázati pontszámai heterogének, ha a vezetési állapotukat vesszük figyelembe. A 4. ábra a) panelje összehasonlítja a viselkedési jellemzők figyelembe vételével és anélkül, ami arra utal, hogy ezek kritikus szerepet játszanak a járművezetők rejtett heterogenitásának jellemzésében. Megmutatjuk, hogy a viselkedési jellemzők jelentősen javítják az előrejelzési teljesítményt. Másodszor, megerősítjük, hogy a balesetekben való részvételre való hajlam idővel változik. A 4. ábra b) panelében három járművezető jellegzetes fejlődési mintázatát ábrázoljuk. Az 1. járművezető kezdetben alacsony kockázati pontszámmal rendelkezik, de egyre kockázatosabbá válik, míg a 2. járművezető ellentétes viselkedést mutat. Az ilyen váratlan mintázatok a járművezetők eltérő adaptív (azaz önerősítő) vezetési tulajdonságait mutatják. Ezeket a dinamikusan agresszív (defenzív) kockázatok a (9) egyenletben simítjuk. Végül, a jellemzők kiválasztási folyamatából kapott jellemzők fontosságának vizsgálatával következtethetünk a legjelentősebb balesetet okozó tényezőkre.

Ábra A járművezetők kockázatai4.

(a) A kockázati pontszámok heterogenitása

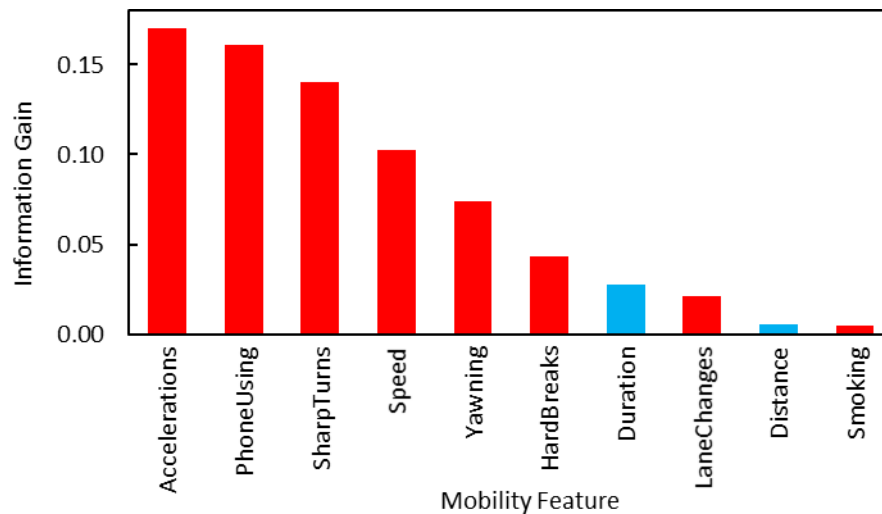


(b) A kockázati pontszámok alakulása



Amint az 5. ábra mutatja, a piros színű viselkedési jellemzők jelentősebbek, mint a kék színű pálya jellemzői. Ennél is fontosabb, hogy a figyelemeltereléssel (pl. telefonálás) és fáradtsággal (pl. ásítás) történő vezetés is végzetes lehet, míg a rendellenes gyorsítások, az éles kanyarok és a sebesség várhatóan veszélyes lehet.

Ábra Jellemző 5.fontossága



5.3. Validálás és modell-összehasonlítás

A javasolt megoldást a következő empirikus kísérletekben alternatív modellspecifikációkkal hasonlítjuk össze. A kutatás elsődleges célja annak vizsgálata, hogy az IoT-alapú kockázatértékelés hatékonyabb-e a jelenlegi legjobb gyakorlat (azaz a demográfiai alapú profilalkotás) megszakításával. Mivel a kockázatjóságot osztályozási problémaként fogalmazzuk meg, Kent et al. (1955) nyomán a pontosságot és a visszahívást választjuk teljesítménymérőnek. A következő értekezésben nagyobb figyelmet fordítanánk a visszahívásra, mint a pontosságra, mivel

az autóbiztosítási ágazatban a költséges téves negatív esetekhez. A gépi tanulás szakirodalmát követve három klasszikus, reprezentatív osztályozót választunk referencia-módszerként. Konkrétan a logisztikus regressziót (pl. Tu 1996), a támogató vektor gépet (Cortes & Vapnik 1995) és a véletlen erdőket (Ho 1995) vesszük figyelembe.⁸ A szakirodalmat követve finomhangoljuk az egyes benchmarkok modellparamétereit. Ezáltal tisztességesen össze tudjuk hasonlítani a modelleket.

Először egy 2x2 értékelési keretet vizsgálunk meg a bemeneti adatok és a modellezési választások tekintetében. Amint az ábra szemlélteti⁶, a függőleges dimenzió a mobilitás és a demográfia alapján képzett módok összehasonlítását jelenti. Eközben a javasolt modellt más osztályozókkal hasonlítjuk össze a vízszintes dimenzióban, azonos képzési jellemzők mellett.

6. ábra. 2x2 értékelési keretrendszer

	XGBoost-based model	Alternative models	
Mobility	Proposed solution	Logistic Regression Support Vector Machine Random Forests	Mobility
Demographics	XGBoost-based model	Logistic Regression Support Vector Machine Random Forests	Demographics

Az eredményeket a 4. táblázatban foglaljuk össze. Az eredmények egyrészt rávilágítanak a mobilitásalapú profilalkotás előnyeire. A mobilitásból tanult modellek jelentősen javítják a metrikákat (azaz a pontosságot és a visszahívást) a demográfiai adatokon alapuló modellekhez képest. Konkrétan, a javasolt konstrukció 45,86%-kal magasabb visszahívást generál, mint egy hasonló, de demográfiai alapon képzett XGBoost modell. A másik oldalon, miközben mobilitási adatokat veszünk figyelembe, a többszintű modellezési megközelítésünk következetesen felülmúlja a logisztikus regressziót (LG), a támogató vektor gépet (SVM) és a véletlen erdőket (RF), 43,45%-kal, 29,46%-kal és

⁸ A klasszikus három modellnek megfelelő három fejlett modellt is figyelembe veszünk, beleértve az osztályokkal súlyozott LR-t, az egymásba ágyazott SVM-et és a költségérzékeny súlyozott RF-t. Az MRA és a három modell közötti összehasonlításokat a Függelékben foglaljuk össze. Az eredmények megerősítik, hogy az MRA felülmúlja a fejlett modelleket. Üdvözljük a felülvizsgálati csoport által javasolt lehetséges referenciaértékeket.

26,24%, illetve. Az a) panelben a pontosságra vonatkozóan is hasonló megállapítások találhatók. A fenti empirikus adatok arra utalnak, hogy a legjobb teljesítmény csak akkor érhető el, ha az IoT által gyűjtött mobilitási adatokat egy mesterséges intelligencia alapú előrejelző motorra alkalmazzuk, ahogyan azt a javasolt MRA-megoldás teszi.

4. táblázat. Modell-összehasonlítás - jellemzők és modellek

(a) Precíziós

	Fő modell	Logisztikus regresszió	Támogató vektor gép	Véletlen erdők
Mobilitás	88.10%	62.12%	68.90%	70.33%
Demográfiai adatok	62.01%	50.19%	54.01%	54.42%

(b) Visszahívás

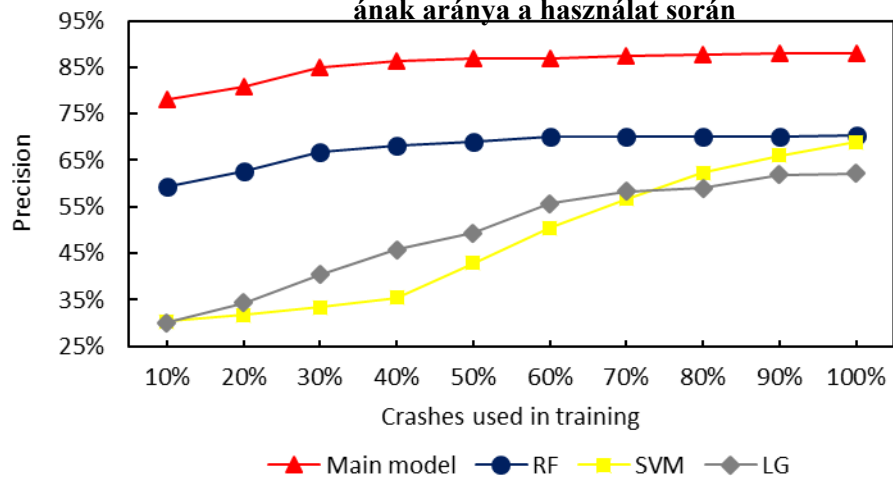
	Fő modell	Logisztikus regresszió	Támogató vektor gép	Véletlen erdők
Mobilitás	87.98%	61.33%	67.96%	69.69%
Demográfiai adatok	60.32%	49.12%	53.11%	53.79%

Másodszor, szeretnénk tesztelni, hogy az MRA-megoldás robusztus-e a szokatlan felhasználási esetekkel szemben. A gépek számára nagy kihívást jelent nem informatív (azaz ferde vagy kis méretű) képzési adatokból tanulni. Ezért a gyakorlóhalmazt a pozitív címkék (azaz a balesetet okozó utazás) eloszlásának manipulálásával rekonstruáljuk a gyakorlóhalmazban. Ebben a tekintetben szélsőséges eset lenne, ha a pozitív eseteknek csak 10%-át használnánk fel a képzésben. A 7. ábrán a négy modell teljesítményét ábrázoljuk a manipulált eloszlás tekintetében. A javasolt megoldás a legrobusztusabb a ferde adatokkal szemben. Az a) és b) panel azt mutatja, hogy az XGBoost-alapú modell hatékonyan tanul. A pozitív címkézésű minták kevesebb mint 30%-ának felhasználásával tisztességes pontossággal és visszahívással rendelkezik (azaz 80% feletti). Az RF, bár gyengébb, mint a mi modellünk, mégis jobb tanulási hatékonyságot mutat, mint az SVM és az LR. Emellett kíváncsiak vagyunk arra, hogy az MRA algoritmus hogyan kezeli a kis képzési adatok problémáját. A modelleket kisebb, kevesebb vezetőből álló adathalmazokkal képezzük, a teljes képzési minták 10%-ától kezdve. A javasolt modell a pontosság és a visszahívás tekintetében dominál az egyes referenciaalgoritmusok felett, amint azt a (c) és (d) panel mutatja. A modell teljesítménye kezdetben 60%-os, a pontosság és a visszahívás gyorsan 80%-ra javul, és

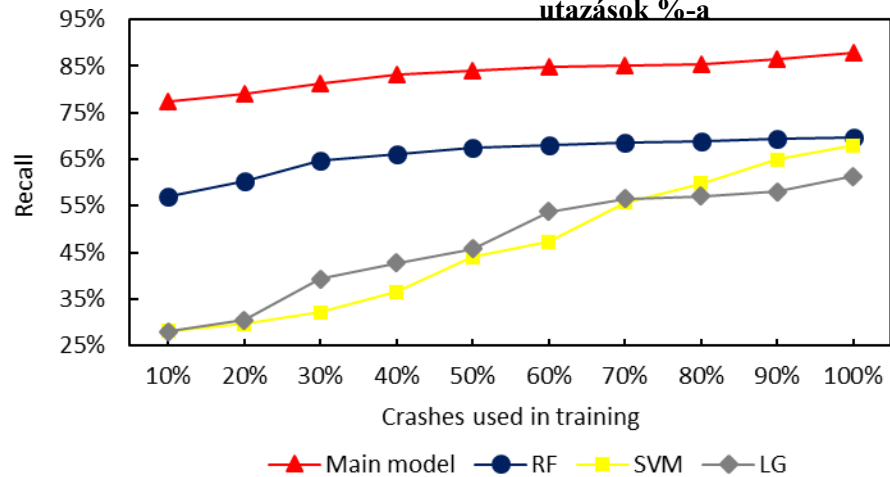
a teljes képzési halmaz 10%-ának, 50%-ának, illetve 100%-ának felhasználásával eléri a 88%-os telítési pontot. E két eredmény sorozat együttesen azt a következtetést vonja le, hogy a javasolt modell robusztus mind a ferde, mind a kis méretű képzési halmazokkal szemben.

ábra Modell 7.összehasonlítás változó képzési adatokhoz

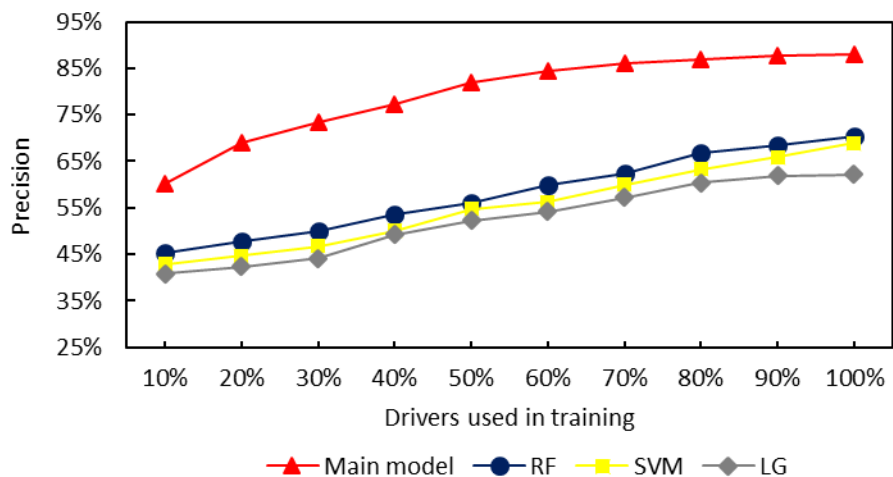
(a) Precizitás vs. a kockázatos utazások %-ának aránya a használat során



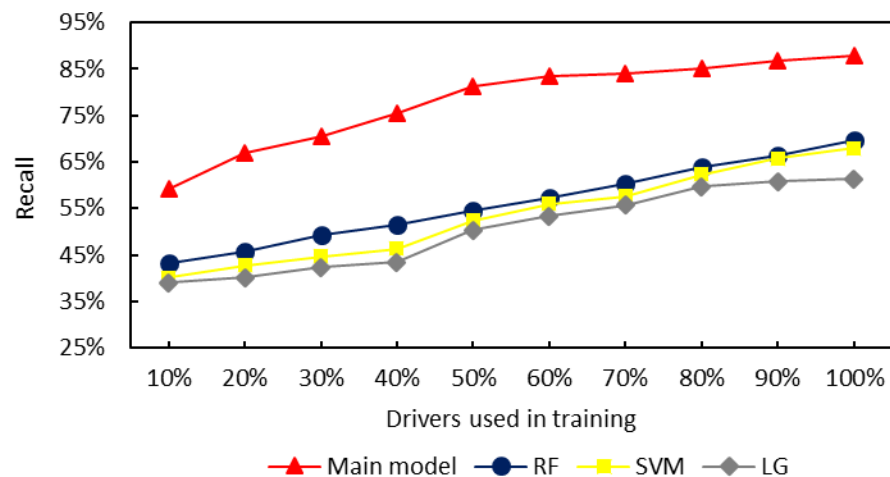
(b) Visszahívás vs. a használatban lévő kockázati utazások %-a



(c) Precizitás vs. a használatban lévő járművezetők %-a



(d) Visszahívás vs. a használatban lévő járművezetők %-a



Elektronikusan elérhető a következő címen:
<https://ssrn.com/abstract=3661062>

Eddig már megmutattuk az MRA modell előnyét az alternatív osztályozókkal szemben. Szükséges azonban, hogy az alternatív modellspecifikációk összehasonlítását ugyanazon a megoldási koncepción belül nagyító alá vegyük. Amint azt a 4.3. szakaszban tárgyaltuk, az elméleti útmutatást követve a mobilitási adatokat (i) a vezetési jellemzők vezetői szintű profilozására használt viselkedési tényezőkre és (ii) az útvonalkockázatok útvonalszintű profilozására használt kontextuális tényezőkre osztjuk. Az (i) és (ii) elemeket szintetizáljuk, hogy a (10. egyenlet segítségével megkapjuk a fő modellt.) Két változatos modellt határozunk meg, amelyekben a viselkedési és kontextuális kockázatokhoz vagy a járművezető szintjén, vagy az utazás szintjén teljes egészében hozzáférünk. A (10. egyenletet a járművezető- és az utazásközpontú modellek esetében a következőképpen írjuk át

$$r_{ij} = \alpha^{Br} + T \quad \beta r, (1)$$

1) ii

és

$$r_{ij} = \alpha^{Br} + T \quad \beta r, (1)$$

2) jj

illetve. Mindhármát értékeljük, és az eredményeket az 5. táblázatban foglaljuk össze. A fő modell kiemelkedik, és a legjobb teljesítményt eredményezi, amelyet a járművezető-központú modell követ. Mennyiségileg az előbbi 9,44%-kal (7,66%-kal) növeli az utóbbi felidézési (pontossági) értékét. Végül is egy ilyen robusztus modellezési megközelítés a teljes megoldás szíve.

Táblázat 5. Modell-összehasonlítás - Alternatív modellezési lehetőségek

Modellezési specifikációk	Fő modell <i>Viselkedés+Trajektóri</i> <i>aj</i>	Vezetőközpontú <i>Viselkedés+Trajektóri</i> <i>ai</i>	Utazás-központú <i>Viselkedésj+Trajektór</i> <i>iaj</i>
Precíziós	88.10%	81.83%	74.89%
Visszahívás	87.98%	80.39%	73.29%

Végül az egyes modellek feldolgozási idejéről a táblázatban számolunk be. Bár a 6. logisztikai regresszió a leggyorsabb a négy modell közül egyszerűsége miatt, a gyorsabb feldolgozási sebességért cserébe a prediktív teljesítményt adja. Ez emlékeztet bennünket az előrejelző teljesítmény és a számítási idő közötti kompromisszumra. Egyfelől a javasolt modell előnyben részesül, mert hasonló számítási idő mellett sokkal jobban teljesít a másik háromnál. Másrészt a javasolt modell továbbra is a legjobb választás marad, mivel a benchmarkoknál kisebb képzési halmaz használatával ugyanolyan teljesítmény elérése

érdekében meg tudja spórolni a feldolgozási időt. Amint az ábra szemlélteti⁸, csak másodpercekre²⁷ van szüksége ahhoz, hogy a pontosság és a visszahívás mennyiségi szempontból 60%-os legyen, míg a másik háromnak több időre van szüksége.

mint másodpercek140, hogy ugyanazt a teljesítményt elérjük.

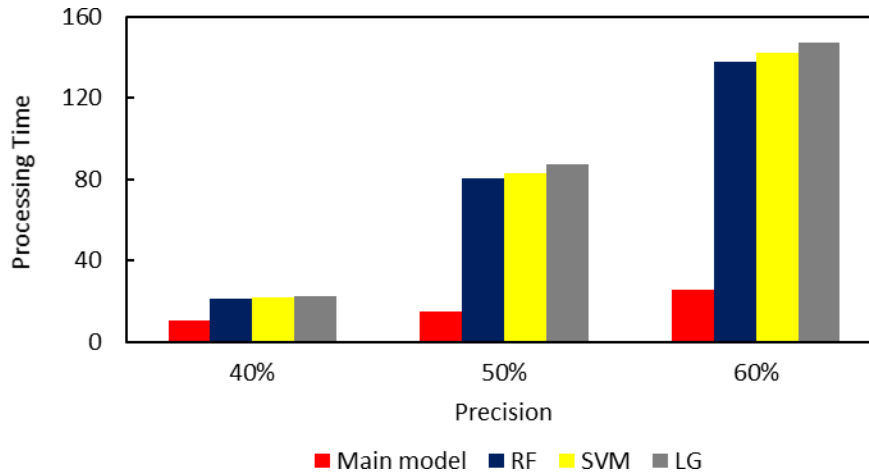
6. táblázat. Feldolgozási idő

Hetek adatai	1	5	10	15	20	25	30
Fő modell	6.48	35.40	67.67	101.93	130.33	180.36	201.74
Véletlen erdők	6.68	41.12	87.34	135.57	180.73	221.91	270.85
Támogató vektor gép	5.36	26.40	58.40	88.22	119.11	149.96	177.40
Logisztikus regresszió	5.24	27.48	60.28	90.17	120.60	151.73	180.09

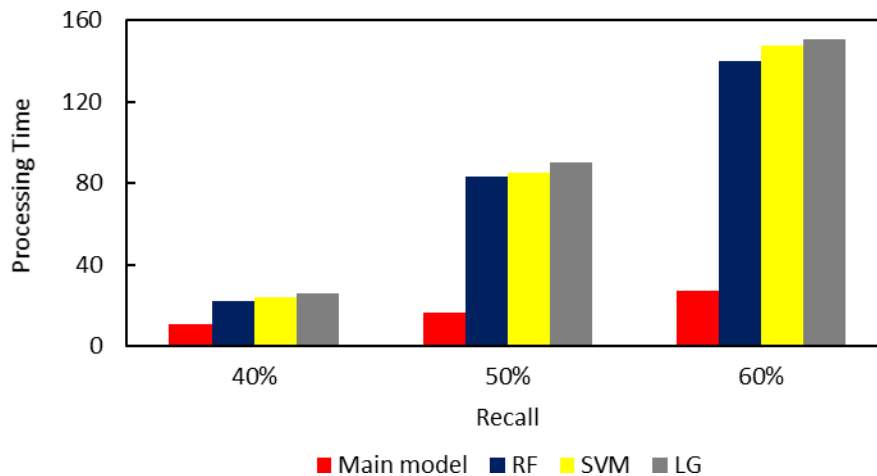
Megjegyzés. A feldolgozási időt a számítási környezetben mérik: Intel i7-6850K @ 3,60GHZ, 64 GB DDR4 memória, valamint 1 TB SSD és 4 TB HDD merevlemezek.

Ábra Modell8. összehasonlítás változó teljesítmény esetén

(a) Feldolgozási idő vs. pontosság



(b) Feldolgozási idő vs. Visszahívás



6. Következtetés és jövőbeli kutatás

A legújabb technológiák felhasználásával újszerű tervet javasoltunk a baleseti kockázatok előrejelzésére. A hagyományos demográfiai alapú gyakorlattól eltérően a javasolt keret a mobilitás alapú modellezést használja ki. Konkrétan, a finom szemcsés adatokat különböző IoT-képes eszközökről gyűjtjük, beleértve a GPS-nyomkövetőket, a fedélzeti diagnosztikát és a járműbe épített kamerákat. A járművezetők kockázatainak osztályozásához egy robusztus előrejelző motort fejlesztünk ki, amely a járművezető-központú viselkedésekkel kapcsolatos baleseti kockázatokat a pályaközpontú összefüggésekkel együtt veszi figyelembe. Egy XGBoost-alapú modellt valósítunk meg, amely rugalmasan képes a balesetek előrejelzésére mind a trajektória, mind a járművezető szintjén. A javasolt megoldást valós adatokra alkalmazva számos robusztus eredményt kapunk magából a kalibrált modelltől és a modellértékelésekből.

A modell értelmezhető eredményei egyrészt arra utalnak, hogy a vezetési tulajdonságok jelentős szerepet játszanak a balesetekben. A legnagyobb hatást gyakorló jellemzők közé tartoznak a rendellenes gyorsítások, az éles kanyarok és a figyelemelterelés. Észrevesszük a vezetési viselkedés markáns heterogenitását, amelyet korábban nem lehetett statikus demográfiai adatokkal megragadni. A nagyított heti elemzések továbbá azt mutatják, hogy a járművezetők kockázatai dinamikusan fejlődnek. Az ilyen váratlan minták a járművezetők különböző adaptív (azaz önmagukat erősítő) vezetési tulajdonságait mutatják. Más szóval, a járművezetők idővel agresszívvá (defenzívebbé) válhatnak.

A javasolt megoldást egy sor empirikus validáció során összehasonlítjuk az alternatívákkal. Először egy 2x2 értékelési keretet veszünk figyelembe a bemeneti jellemzők és a modellezési választások tekintetében. Az eredmények együttesen bizonyítják, hogy a javasolt megoldás felülmúlja a többi megközelítést. A mobilitásból tanult modellek jelentősen javítják a pontosságot és a visszahívást a demográfiai adatokból tanult modellekhez képest. Konkrétan, az MRA megoldás 45,86%-kal (42,07%) magasabb visszahívást (pontosságot) eredményez, mint a demográfiai adatokon alapuló XGBoost-alapú modell. A mobilitási jellemzők figyelembevételével a megközelítésünk az LR, SVM és RF visszahívását is javítja 43,45%-kal, 29,46%-kal, illetve 26,24%-kal. Másodsor, érdekel minket a modell robusztusságának tesztelése szokatlan felhasználási esetekre (azaz ferde és kis képzési adatokra). A képzési halmazt úgy rekonstruáljuk, hogy kísérletezünk a balesetek különböző eloszlásával és a benne szereplő járművezetők különböző számával. Az empirikus értékelések azt mutatják, hogy a javasolt

megoldás a legrobosztusabb. Harmadszor, a megoldáson belüli összehasonlítás rávilágít a többszintű modellezés értékére, amikor a teljes modellt összehasonlítjuk a

egyszerűbb alternatívák (azaz a pályaközpontú és a járművezető-központú modellek). Végül, a javasolt modell számítási szempontból hatékonyabb a kisebb gyakorlókészletre vonatkozó követelményeket illetően, kiváló teljesítménnyel.

Ez a tanulmány egy újszerű megoldás kidolgozásával járul hozzá a szakirodalomhoz, amelyben a tárgyak internete és a mesterséges intelligencia javítja a működési hatékonyságot a baleseti kockázatok felmérésében. Bár technikailag kihívást jelent, a javasolt megoldás három szempontból is egyedinek bizonyul a korábbi kutatásokhoz képest: adatok, modellezés és alkalmazások. Először is IoT-képes érzékelőket használunk (pl. járműbe épített kamera) a saját adatok gyűjtésére, amelyek valós időben figyelik a viselkedési részleteket, míg a korábbi tanulmányok az OBD-n keresztül archivált adatokra támaszkodnak (pl. Paefgen et al. 2004). Ezek a nagy felbontású adatok lehetővé teszik továbbá a kockázatok többszintű profilozását. Az egyszintű elemzésekkel (pl. Xie et al. 2017) ellentétben a mi járművezetői útprofilozásunk elkülöníti a viselkedési kockázatokat a kontextuális veszélyektől (pl. sebességkorlátozások), hogy pontosabban kalibráljuk a járművezetők heterogenitását. Tényezzük a kockázatosztályozás meglévő fejlődését, amelyet általában figyelmen kívül hagynak (pl. Wu et al. 2014). Végül, figyelembe vesszük a torzítás-variáció dilemmát, hogy megfeleljen a társítási és előrejelzési következtetések igényeinek. Bevezetünk egy hatékony tanulási technikát a naiv modellek teljesítményének jelentős javítására (pl. Guo & Fang 2013).

Ezek az eredmények relevánsak az autóbiztosítással kapcsolatos üzletágak szempontjából, és számos új irányt mutatnak a szakemberek számára. Először is, elengedhetetlen a mobilitási adatok beépítése a kockázattértékelésekbe, miközben az Egyesült Államokban a főbb szereplők még mindig haboznak. A mobilitás alapú modellezés jelentősen javítja a kockázati előrejelzések teljesítményét, amit az empirikus validációk is megerősítenek. A biztosítókat arra ösztönzik, hogy proaktívan fogadják el ezt az új legjobb gyakorlatot az alapvető kompetenciák táplálása érdekében. A mobilitás-alapú megoldásba történő beruházás a hatékonyságnövekedés (azaz a kockázatvállalás költségcsökkentése) révén indokolható, mivel az IT-katalizátorok (pl. IoT és AI) költségei idővel csökkennek. A várakozásoknak megfelelően ez a paradigma az 5G hálózat gyorsabb adatkommunikációjával még erőteljesebbé válhat a valós idejű előrejelzésben. Másodszor, a gyakorlati szakembereknek figyelniük kell a kockázatok alakulására. A dinamikus változások figyelmen kívül hagyása alul- vagy túlbecsüli az összeomlási

kockázatokat, ami a műveletek hatékonyságának csökkenéséhez vezet. A biztosítók azonban hozzászoktak ahhoz, hogy egy járművezető pontszámát féléves szinten állandónak modellezzék, függetlenül attól, hogy statikus demográfiai adatokat vagy korlátozott mobilitási adatokat használnak. A *Progressive Snapshot* programja például egy rögzített, 30 napos vezetési időszak összevont adatait értékeli a féléves díjakhoz. Ez a féléves profilalkotás olyan származékos problémákat okoz, mint az erkölcsi kockázat. Ezért azt javasoljuk, hogy a

a gyakorlat folyamatosan figyelemmel kíséri a járművezetők viselkedését, és dinamikusan modellezi a megfelelő kockázatokat. Harmadszor, megoldásunk új terméktervezéshez vezethet az autóbiztosítási ágazat számára. A GPS-követés segítségével a biztosítók a kockázatokat az út szintjén profilozhatják, ahogyan azt az MRA-megoldás pályaközpontú összetevője is teszi. Ez az új képesség segít újra feltalálni a taxiszolgáltatások, az autókölcsönzés és az autómegosztás (pl. Zipcar) számára tökéletes útvonal-szintű termékeket. Meggyőző felhasználási eset lenne, ha az autómegosztó cégek nem előre meghatározott átalánydíjat számítanának fel, hanem utólag, személyre szabott útszintű felárat.

Emellett az autógyártók tanulhatnak az eredményekből a járművezetők és az utasok további védelme érdekében. A figyelemelterelés és a fáradtság egyaránt veszélyes, bár a baleseteket általában a helytelen cselekvéseknek tulajdonítják, például a rendellenes gyorsításnak és az éles kanyaroknak. A figyelemelterelés és a fáradtság figyelmen kívül hagyása akkor válik végzetessé, ha nem történik további intézkedés. Ez a tanulmány egy olyan megoldást mutat be, amellyel megelőzhető a járművezetők ilyen veszélyeztetése. A járműbe épített kamerák arcfelismerés segítségével figyelik a járművezetők koncentrációs szintjét azáltal, hogy észlelik a figyelemelterelésre utaló viselkedésformákat (pl. telefonálás és dohányzás) és a fáradtság jeleit (pl. a szemek lehunyása és ásítás). Az autógyártók számára ideális lesz az IoT-képes IVC-ket valós időben integrálni a felhőalapú felügyeletbe. A rendszer proaktív módon képes a szükséges beavatkozásokat, például a kormány melegítését és rezgőre állítását megtenni, amint a nemkívánatos viselkedést észleli. Úgy véljük, hogy egy ilyen intelligens rendszer az autógyártók, a járművezetők és a biztosítók számára egyaránt előnyös helyzetet eredményez.

Ami a korlátokat illeti, a nagy felbontású mobilitási adatokhoz való hozzáférés során az átfogó pályaösszefüggések hiánya miatt szenvedhetünk. Elismerjük, hogy a további trajektórijellemzők felhasználása az egyik módja a modellteljesítmények javításának, ahogyan Jia et al. (2018) és Paefgen et al. (2014) vizsgálja az érdeklődési pontok, illetve az útviszonyok hatásait. Alternatív megoldásként a vezetési viselkedés kihasználása a másik út a gyakorlással kapcsolatos kockázati előrejelzésekhez. A kettő kiegészíti egymást, az extrinsic (azaz környezeti) jellemzőkre és az intrinsic (azaz viselkedési) tényezőkre összpontosítva. Ebben a tekintetben a megoldásunkat úgy terveztük, hogy rugalmasan beépítsen minden új környezeti jellemzőt, amikor csak rendelkezésre áll, hogy felerősítse az előrejelző erőt.

Emellett a technológiai fejlődéssel párhuzamosan mindig vannak erősebb algoritmusok is.

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=3661062>

Elvárható, hogy a prediktív teljesítmény tovább javuljon, miközben az XGBoost jelenleg az egyik legkedveltebb osztályozónak számít. Javasolunk tehát egy általánosítható, modellalapú

független keretrendszer. Más szóval, a szakemberek kevésbé lesznek korlátozva abban, hogy a kontextus-specifikus feladatok elvégzéséhez háttér tanulási módszereket válasszanak. Mint elismertük, a mi megoldásunk nem lehet minden idők legjobbja egy adott kontextusban, de általánosan alkalmazható a különböző iparágakban az előrejelzésekre. Az utolsó korlátozás az árazással kapcsolatos. Nagyon hatékony működési szempontból a termékdíjak finomhangolása az MRA-megoldás által profilozott járművezetői kockázatok felhasználásával. A kutatási együttműködés terjedelme miatt azonban az árképzés vizsgálatában korlátozottak vagyunk. Rendszerünk az első ugródeszkeként szolgál a későbbi műveletek hatékonyságának növeléséhez; végül is a biztosítási üzletág lényege a fedezeti ügylet, amelyhez a kockázatok azonosítására van szükség első, kritikus lépésként. A jövő irányaként a biztosítási szolgáltató nyereségének maximalizálására törekszünk a meglévő termékek árazásával és a már pontos előrejelzéseken alapuló új, utazási szintű termékek kiváló tervezésével.

E korlátozások ellenére ez a tanulmány az egyik első, amely a robusztus kockázatértékeléshez IoT-alapú megoldást fejlesztett ki. Az általánosítható keret hasznos vezetői implikációkkal szolgál arra vonatkozóan, hogy az autóbiztosítási ágazat hogyan használja az IoT-t és az AI-t a legjobb gyakorlat újradefiniálására, a biztosítási termékek újbóli feltalálására és a meglévő üzleti modell megzavarására az 5G hálózat közelebbi korszakában.

Hivatkozás

- Ale B (2016) Kockázatelemzés és nagy adatok. *Safety and Reliability* **36** (3):153-165.
- Accelya (2016) 5 nagyszerű mód, ahogy a légitársaságok használják a dolgok internetét (2020. május 20-i hozzáférés): <https://w3.accelya.com/blog/5-great-ways-airlines-are-using-the-internet-of-things>.
- Ban GY, Rudin C (2019) The big data newsvendor: Gyakorlati meglátások a gépi tanulásból. *Operations Research* **67**(1):90-108.
- Blom HAP, Bakker GJ, Blanker PJG, Daams J, Everdij MHC, Klompstra MB (2001) Accident risk assessment for advanced air traffic management.
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.
- Brynjolfsson E, Hu Y, Rahman MS (2013) Versenyzés az omnichannel kiskereskedelem korában. *MIT Sloan Management Review* **54** (4):23-29.
- Caro F, Sadr R (2018) Internet of things (IoT) in retail: A kereslet és a kínálat összekapcsolása. *Business Horizons* **62**(1):47-54
- Chen T, Guestrin C (2016) XGboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 785-794.
- Choi TM, Chan HK, Yue X (2016) Az üzleti műveletek és a kockázatkezelés nagy adatelemzésének legújabb fejleményei. *IEEE Transactions on Cybernetics* **47** (1):81-92.
- Choi TM, Lambert JH (2017) Advances in risk analysis with big data. *Risk Analysis* **37**(8):1435-1442.
- Chung SH, Ma HL, Chan HK (2017) A légitársasági munkaerő-felhasználás késéskockázatának kaszkádosítása személyzetpárosítással és menetrend-optimalizálással. *Risk Analysis* **37**(8):1443-1458.
- Cohen MC (2018) Big data and service operations. *Production and Operations Management* **27**(9):1709-1723.
- Cole S, Giné X, Vickery J (2017) Hogyan befolyásolja a kockázatkezelés a termelési döntéseket? Evidencia egy terep kísérletről. *The Review of Financial Studies* **30**(6):1935-1970.
- Cortes C, Vapnik VN (1995) Support-vector networks. *Machine Learning* **20**(3):273-297.
- Delen D, Hardgrave BC, Sharda R (2007) RFID for better supply-chain management through enhanced information visibility. *Production and Operations Management* **16**(5):613-624.
- Dietterich TG, Kong EB (1995) Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Technikai jelentés, Oregon State University, Corvallis, OR.
- Ding X, Tian Y, Yu Y (2015) A beltéri vezeték nélküli érzékelőhálózatokon alapuló valós idejű nagy adatgyűjtési algoritmus az ipari műveletek kockázatelemzésére. *IEEE Transactions on Industrial Informatics* **12** (3):1232-1242.
- Dong S, Xu SX, Zhu KX (2009) Kutatási jegyzet - Információs technológia az ellátási láncokban: Az it-alapú erőforrások értéke a versenyben. *Information Systems Research* **20**(1):18-32.

- Feng Q, Shanthikumar JG (2018) Hogyan fejlődhet a termelés- és műveletirányítási kutatás a nagy adatok korában. *Production and Operations Management* **27**(9):1670-1684.
- Gao L, Yang N, Zhang R, Luo T (2017) Dynamic supply risk management with signal- based forecast, multi- sourcing, and discretionary selling. *Production and Operations Management* **26**(7):1399-1415.
- Ghose A, Li B, Liu S (2019) Mobile targeting using customer trajectory patterns. *Management Science* **65**(11):5027-5049.
- Global Market Insights (2019) Usage-based insurance market revenue to hit usd 115 bn by 2026 (accessed on July 27, 2020): <https://www.prnewswire.com/news-releases/usage-based-insurance-market-revenue-to-hit-usd-115-bn-by-2026-global-market-insights-inc-300970226.html>.
- Guo F, Fang Y (2013) Egyéni járművezetői kockázatértékelés naturalisztikus vezetési adatok felhasználásával. *Accident Analysis & Prevention* **61**:3-9.
- He H, Garcia EA (2019) Tanulás kiegyensúlyozatlan adatokból. *IEEE Transactions on Knowledge and Data Engineering* **21**(9):1263-1284.
- Heese HS (2007) A leltári nyilvántartás pontatlansága, a kettős marginalizáció és az RFID bevezetése. *Production Operations Management* **16**(5):542-553.
- Hitt LM, Brynjolfsson E (1996) Productivity, business profitability, and consumer surplus: three different measures of information technology value. *MIS Quarterly* **20** (2):121-142.
- Ho TK (1995) Random Decision Forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition*.
- Ho Y-J, Dewan S, Ho Y-C (2020) Distance and Local Competition in Mobile Geofencing. *Information Systems Research* (megjelenés alatt).
- Jia, R, Khadka A, & Kim I (2018) Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis & Prevention* **121**:223-230.
- Jun J, Ogle J, Guensler R (2007) A balesetekben való részvétel és az időbeli-térbeli vezetési viselkedés aktivitási mintái közötti kapcsolatok: a globális helymeghatározó rendszerrel rendelkező járművek adatainak felhasználása. *Transportation Research Record* **2019** (1):246-255.
- Jun J, Guensler R, Ogle J (2011) A balesetben részt vevő és a balesetben részt nem vevő járművezetők között megfigyelt sebességminták közötti különbségek: A járműbe épített megfigyelési technológia alkalmazása. *Transportation Research Part C: Emerging Technologies* **19**(4):569-578.
- Kouvelis P, Li R (2019) Integrated risk management for newsvendors with value-at-risk constraints. *Manufacturing & Service Operations Management* **21** (4):816-832.
- Kumar S, Mookerjee V, Shubham A (2018) Research in operations management and information systems interface. *Production and Operations Management* **27**(11):1893-1905.
- Lee H, Özer Ö, (2007) Unlocking the value of RFID. *Production Operation Management* **16**(1):40-64.
- Machin MA, Sankey KS (2008) A fiatal járművezetők személyiségjegyei, kockázatérzékelése és vezetési magatartása közötti kapcsolatok. *Accident Analysis & Prevention* **40**(2):541-547.

- Mejia N (2020) Artificial Intelligence at HSBC - 2 Use-Cases (hozzáférés: 2020. május 20.): <https://emerj.com/ai-sector-overviews/artificial-intelligence-at-hsbc>.
- Michalski G (2009). A készletgazdálkodás optimalizálása a működési kockázatkezelés részeként. *Economic Computation and Economic Cybernetics Studies and Research* 213-222.
- Mithas S, Tafti A, Bardhan I, Goh JM (2012) Information technology and firm profitability: mechanisms and empirical evidence. *MIS Quarterly* **36** (1):205-224.
- Nagali V, Hwang J, Sanghera D, Gaskins M, Pridgen M, Thurston T, Mackenroth P, Branvold D, Scholler P, Shoemaker G (2008) Procurement risk management (PRM) at Hewlett-Packard company. *Interfaces* **38** (1):51-60.
- Olsen TL, Tomlin B (2020) Ipar 4.0: lehetőségek és kihívások az üzemeltetésirányítás számára. *Manufacturing & Service Operations Management* **22** (1):113-122.
- Toledo T, Musicant O, Lotan T (2008). Járműbe épített adatrögzítők a járművezetők viselkedésének nyomon követésére és visszajelzésére. *Transportation Research Part C: Emerging Technologies* **16** (3):320-331.
- Osborne JW (2000) A hierarchikus lineáris modellezés előnyei. *Gyakorlati értékelés, kutatás és értékelés* **7**(1):1.
- Paefgen J, Staake T, Fleisch E (2014) Multivariate exposure modeling of accident risk: A Pay-as-you-drive biztosítási adatokból nyert meglátások. *Transportation Research Part A: Policy and Practice* **61**:27-40.
- Rai A, Patnayakuni R, Seth N (2006) A digitálisan támogatott ellátási lánc integrációs képességek hatása a cégek teljesítményére. *MIS Quarterly* **30** (2):225-246.
- Rai A, & Tang X (2014) Research commentary-information technology-enabled business models: A conceptual framework and a coevolution perspective for future research. *Information Systems Research* **25**(1):1-14.
- ReportLinker (2020) A használat alapú biztosítási piac az előrejelzések szerint a 2019-es becsült 24,0 milliárd USD-ról 2027-re eléri a 125,7 milliárd USD-t, 23,0%-os CAGR mellett (2020. május 20-i hozzáférés): <https://www.globenewswire.com/news-release/2020/05/13/2032379/0/en/The-usage-based-insurance-market-is-projected-to-reach-USD-125-7-billion-by-2027-from-an-estimated-USD-24-0-billion-in-2019-at-a-CAGR-of-23-0.html>.
- Shacklett M (2017) Hogyan fogja a dolgok internete átalakítani az élelmiszer-ellátási láncokat. *Food Logistics* (hozzáférés: 2020. július 27.): <https://www.foodlogistics.com/technology/article/12382105/how-the-internet-of-things-will-transform-food-supply-chains>.
- Shore J, Johnson R (1980) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory* **26**(1)26-37.
- Simpson AG (2017) A State Farm 2016-ban 7 milliárd dollárt veszített az autón (hozzáférés: 2020. június 19.): <https://www.insurancejournal.com/news/national/2017/03/01/443224.htm>.
- Sodhi MS, Son BG, Tang CS (2012) A kutatók perspektívái az ellátási láncok kockázatkezeléséről. *Termelés- és műveletirányítás* **21** (1):1-13.
- Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 2951-2959.

- Tu JV (1996) A mesterséges neurális hálózatok és a logisztikus regresszió használatának előnyei és hátrányai az orvosi kimenetel előrejelzésében. *Journal of clinical epidemiology***49** (11):1225-1231.
- Xie K, Ozbay K, Kurkcu A, Yang H (2017) A gyalogosokat érintő közlekedési balesetek elemzése nagy adatok felhasználásával: A hozzájáruló tényezők vizsgálata és a gócpontok azonosítása. *Risk analysis***37** (8):1459-1476.
- Xie K, Ozbay K, Yang H, & Li C (2019) Automatikusan kinyert járműpálya-adatok bányászata proaktív biztonsági elemzésekhez. *Transportation Research Part C: Emerging Technologies* **106**:61-72.
- Wu KF, Agüero-Valverde J, Jovanis PP (2014) Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis & Prevention* **72**:210-218.
- Zhang DJ, Dai H, Dong L, Wu Q, Guo L, Liu X (2019) A pop-up üzletek értéke a kiskereskedelmi platformokon: Evidence from a field experiment with Alibaba. *Management Science* **65**(11):5142-5151.

Függelék

Három fejlett modellt valósítunk meg az LR, SVM és FR paradigmákban, mint alternatív viszonyítási pontok, beleértve az osztályokkal súlyozott LR-t (King and Zeng 2001)⁹, a beágyazott SVM-et (Lee and Scott 2010)¹⁰ és a költségérzékeny súlyozott RF-t (Devi et al. 2019)¹¹. Hasonlóképpen összehasonlítjuk az MRA modellt a hárommal, és az A1. táblázatban összefoglaljuk a pontosság és a visszahívás összehasonlításait. Az eredmények megerősítik a javasolt modell fölényét.

A1. táblázat. Modell-összehasonlítás - alternatív referenciaértékek

(a) Precíziós

	Fő modell	Osztályonkénti súlyozású LR	Beágyazott SVM	Költségérzékeny súlyozott RF
Mobilitás	88.10%	63.09%	69.27%	71.53%
Demográfiai adatok	62.01%	51.28%	55.24%	55.44%

(b) Visszahívás

	Fő modell	Osztályonkénti súlyozású LR	Beágyazott SVM	Költségérzékeny súlyozott RF
Mobilitás	87.98%	61.30%	68.72%	70.86%
Demográfiai adatok	60.32%	49.34%	54.27%	54.39%

⁹ King G, Zeng L (2001) Logisztikus regresszió ritka események adataiban. *Political Analysis* 9(2):137-163.

¹⁰ Lee G, Scott C (2010) Nested support vector machines. *IEEE Transactions on Signal Processing* 58(3):1648-1660. Devi¹¹ D, Biswas SK, Purkayastha B (2019) A cost-sensitive weighted random forest technique for credit card

fraud detection. *Proceedings of the 10th International Conference on Computing, Communication and Networking
Technológiák.* (10. nemzetközi konferencia a számítástechnikáról, kommunikációról és hálózatokról).