MAGYAR AGRÁR- ÉS ÉLETTUDOMÁNYI EGYETEM

Mesterséges intelligencia módszerek alkalmazása az informatikai rendszerek biztonsági auditjában

DOI: 10.54598/001400

Barta Gergő

Gödöllő

2021

A doktori iskola

megnevezése: Gazdaság- és Regionális Tudományok Doktori Iskola

tudományága: gazdaság és regionális tudományok

vezetője: Prof. Dr. H. c. Popp József

MTA levelező tag

Magyar Agrár- és Élettudományi Egyetem

Gazdaságtudományi Intézet - Intézetigazgató

Gazdaság- és Regionális Tudományi Doktori Iskola vezető

Témavezető: Dr. Pitlik László

Egyetemi Docens, Tanszékvezető

Kodolányi János Egyetem, Fenntartható Gazdaság Intézet

Informatika tanszék

....

Az iskolavezető jóváhagyása A témavezető jóváhagyása

Tartalomjegyzék

1. A MUNKA ELŐZMÉNYEI, CÉLKITŰZÉSEK

Folyamatos és növekvő igény mutatkozik az üzleti folyamatok automatizálására (STEGMAN et al. 2019), mely ma már nem kizárólag a rutinmunkának tekinthető, vagyis pl. szabályalapú tevékenységek kiváltására korlátozódik, hanem a komplexebb, kreativitást és akár szubjektív értékítéletet is igénylő feladatok gépiesítésére is kiterjed. A jelen kor technológiái, például a Big Data, Mesterséges Intelligencia, Felhő-alapú Megoldások, valamint a hardveres erőforrások gyarapodása (pl. HPC – High Performance Computing) és teljesítményük fokozatos javulása lehetővé teszik a magas színvonalú automatizált döntéselőkészítést és munkavégzést, mely szakmai és kutatási terület évről évre egyre bíztatóbb eredményeket tudhat magáénak (pl. BODA 2019).

Mindazonáltal, az automatizálást támogató, integrált, üzleti-informatikai megoldások fejlesztése és az eddigi elavult rendszerek korszerűsítése időről időre kihívások elé állítják a technikai szakembereket, vezetőket és felhasználókat, mivel olyan kockázatokat (pl. a kibertérben tárolt üzleti adatvagyon bizalmasságának sérülése) hordoznak, melyek szükségessé teszik az információ kezelésével, biztonságtechnikájával és ellenőrzésével foglalkozó funkcionális egységek megjelenését - megteremtve az informatikai erőforrásokra irányuló kockázatmenedzsment folyamatot (BARTA – GÖRCSI 2021).

A kockázatelemzési eljárások végső kimenete egy stratégiai szintű döntés a szervezetben felmerülő informatikai kockázatok csökkentésére (VASVÁRI 2008). Az informatikai kockázatelemzés eredményére támaszkodó döntés lehet a felmerülő kockázatok enyhítésére irányuló intézkedések megtétele, azaz a megismert kockázatokat kontrollfolyamatokkal történő mérséklése, úgy, mint a megelőző óvintézkedések, korrektív- vagy felderítő tevékenységek bevezetése.

Az implementált informatikai kontrollok megléte még nem szavatolja a hibamentes működést, vagyis a vezetésnek rendszeresen meg kell győződnie arról, hogy a mitigáló intézkedések hatékonyan működnek, nem lehetséges azok megkerülése, illetve akaratlagos kijátszása (POMPON 2016).

A belső informatikai kontrollok hatékonyságának feltárására szolgáló funkcionális terület megnevezése az informatikai audit-csoport, melyet olyan üzleti és informatikai szakmai tudással rendelkező szakemberek alkotnak, akik elsődleges célja a belső informatikai kontrollkörnyezet ellenőrzése és folyamatos tesztelése, annak garantálására, hogy az informatikai folyamatok

teljes mértékben a szervezetek üzleti célkitűzéseit követik a lehető legnagyobb információbiztonsági szint fenntartása mellett. Az auditálás megjelenése és annak bevezetése egy szervezetben segíthet a kontrollhiányosságok feltárásában, azonban továbbra is fennáll az emberi tényező, azaz a véletlenül és/vagy szándékosan elkövetett hiba lehetősége, valamint az emberi erőforráskorlátok (BARTA 2018b). A legtöbb esetben szinte lehetetlen manuálisan végrehajtani az ellenőrzést annak dinamikus karakterisztikái végett.

Mindezért az auditorok egy előre meghatározott statisztikai mintaelemszámmal igyekeznek kiszűrni a nem-megfeleléseket, azaz az ellenőrzés csak részleges (BARTA 2018b). Továbbá, számos példa tetten érhető hazánkban és külföldön is, amikor az auditori munka minősége megkérdőjelezhető volt függetlenségi konfliktus, szervezeti érdekellentétek vagy az etikátlan üzleti magatartás következtében:

- Gondoljunk csak a 2001-es Enron botránya, melynek következtében a világ 5. legnagyobb auditor cége, az Arthur Andersen megszűnt, mivel az Egyesült Államok Legfelsőbb Bírósága felelősnek ítélte szakmailag és erkölcsileg is kifogásolható gyakorlatai miatt (GREENHOUSE 2005).
- Magyarországon is említhető negatív példa, ami kapcsolatba hozható az informatikai ellenőrzések hanyagságával:
    - 2015. február 24-én a Magyar Nemzeti Bank (MNB) felfüggesztette a Buda-Cash Bróker Zrt. működését (Magyar Nemzeti Bank 2015),
    - amely pár hétre rá magával sodorta a Hungária Értékpapír Zrt.-t és
    - Quaestor Értékpapír-kereskedelmi és Befektetési Zrt.–t, melyek együttvéve megközelítőleg 250 milliárd forintnyi fiktív kötvényt bocsájtottak piacra.

A felsorolt negatív példákból és az audithoz köthető kockázatokból egyértelművé válik, hogy a téma aktuális, és ez a közeljövőben még inkább hangsúlyos és jelentős lesz (vö. Ipar 4.0, MI-koalíció), mivel az informatikai iparosítás és üzleti automatizáció által nyerhető üzleti előnyök még inkább késztetni fogják a versenyben résztvevőket a technológiai megoldások implementálására (VINOGRADOV 2020).

Az auditori munka hatékonyságát növelendő, olyan automatizált megoldás szükséges, mely képes objektíven a hiányosságokra utaló jelek felkutatására (üzleti probléma), feldolgozására és értelmezésére, nem kizárólag szabályalapú értékelésére, hanem a komplex, háttérben meghúzódó

összefüggések feltárására is, melyet a klasszikus statisztikai módszerekkel már nem lehet hatékonyan kezelni. Ezen felsorolt elvárások gyakorlati kikényszerítése szakirodalmi kutatásaim alapján a mesterséges intelligencia fogalomkörébe illeszthetők, így a mesterséges intelligencia, a levont következtetések szerint, létjogosultsággal rendelkezhet a probléma megoldását illetően mindenkor a Knuth-i (1995) elvek követésével, ahol a Knuth-i elvárás látszólag egyszerű: „*Tudás/tudomány az, ami forráskódba átírható*”, mégis a XXI. század alapelvárásaként azonosítandó be (PITLIK et al. 2017).

Az intelligenciával ellátott szoftveres megoldások pl. tanulási mechanizmusaik révén képesek lehetnek a gyanús tevékenységek monitorozásában, ezzel hatékony módon valós értéket képezve az informatikai auditban. A felvetett probléma, tehát, az informatikai rendszerekben elkövetett kontrollhiányosságok, csalások, anomáliák, tehát gyanús események, mint informatikai kockázat fogalmának döntési helyzet-specifikus kialakítása mesterséges intelligenciák alkalmazásával, vagyis az élő emberi képességek számítógépes leképezésével, helyettesítésével, támogatásával.

A mesterséges intelligenciával ellátott szoftveres megoldások egyik legnagyobb próbatétele a rendszerek tanítására és tesztelésére felhasznált minták korlátozott elérhetősége és annak diverzitása, így szükségszerű olyan módszerek kutatása, melyek képesek a rendelkezésre álló adathalmazban rejlő maximális potenciál (információs többletérték) kiaknázására. A dolgozat egyik centrális eredménye az elérhető adatok felhasználási optimumának letapogatása, egyrészt a tanuló minta hatékonyabb feldolgozásán, másrészt a klasszikus tesztelési eljárások nélküli modell-preferencia levezetésén keresztül.

A kutatás célkitűzései és hipotézisei az alábbi pontok szerint strukturálhatók:

C1: A Knuth-i elvet követve az információbiztonsági auditok hatékonyságát növelendő, létrehozandó olyan mesterséges intelligenciával ellátott döntéstámogató rendszer (robot-auditor), mely automatizáltan a historikus információbiztonsági auditjelentésekből tanulva képes a kontrollhiányosságok és kontrollterületek közötti összefüggések matematikai feltárására és javaslattételeivel a potenciális emberi hibából fakadó észlelési kockázatok csökkentésére.

H1: Az információbiztonsági auditjelentések szöveges eredményeiből strukturált adatbázist alkotva és bemenetként a mesterséges intelligencia

fogalomkörébe illeszthető eszközökkel azt feldolgozva, az auditok során feltárni kívánt kontrollhiányosságok megléte a véletlen találgatásnál nagyobb valószínűséggel kimutathatók, azaz a kontrollhiányosságok konstellációi matematikailag értelmezhető összefüggéseket hordoznak magukban.

- **H1.1:** A gyanúgenerálás, mint megoldandó üzletileg értelmezett probléma sajátosságait értékelve, a kontrollhiányosságok detektálása megoldható felügyelt és felügyelet nélküli gépi tanuló eljárásokkal is.

- **H1.2:** A gyanúgenerálás teljesítménye fokozható hibrid megközelítésben, azaz a felügyelt és nem felügyelt módszerek együttes felhasználásának a kutatásban alkalmazott releváns performancia metrikái ideálisabb értékeket mutatnak, mint önálló alkalmazásban

- **H1.3:** A hibrid modell többlet-információs értéket teremtve képes az egyszerű modellek általánosító képességén javítani.

C2: A fejlesztendő mesterséges intelligenciával ellátott szoftveres robot-auditornak kényszerűen alkalmasnak kell lennie a rendelkezésre álló adathalmaz minél inkább az optimálishoz közeli felhasználására, mely által teljesítménye maximalizálható, azaz a cél a robot-auditor genetikai potenciáljának kiaknázása a tanulási adathalmaz irányított redukálása révén.

H2: A döntéstámogató rendszer genetikai potenciálja letapogatható hasonlóságelemzéssel ellátott kereső eljárással a tanításra alkalmazott adathalmaz irányított feldolgozásán keresztül, úgy, hogy a genetikai potenciálhoz vezető kereső eljárás a genetikus algoritmusok esetén alkalmazott véletlen mutáció és a populáció egyedeinek keresztezése nélkül is képes ideálisabb eredményt szolgáltatni.

C3: A tanulásra felhasznált adathalmaz információtartalmát növelendő, anti-diszkriminatív módon szükséges az egyes robot-auditor alternatívák teljesítményeinek összehasonlítása, a legjobb alternatíva kiválasztása, melyhez nem szükséges validációs és teszthalmaz elkülönítése a szokásos tesztelés általi adat/információ-vesztési gyakorlattal szemben.

H3: A mesterséges intelligenciával ellátott döntéstámogató rendszerek teljesítményalapon a gépi tanuló alkalmazások klasszikus tesztelési eljárásai nélkül is rangsorolhatók, a predikciók, mint generált gyanúforrások leíró tulajdonságainak érték-irány levezetésével és az ezen adatokat feldolgozó matematikai apparátussal, mely automatizáltan képes a preferált modellek objektív meghatározására.

2. ANYAG ÉS MÓDSZERTAN

A dolgozatban bemutatott kutatás alapos tervezési munkát igényelt, ezért annak folyamata előre meghatározott, részletes szakmai és tudományos módszertani irányelveknek megfelelően lett strukturálva a mindenkori célkitűzéseket fókuszban tartva. A tervezés stratégiai szintjén a legfontosabb az volt, amit minden mesterséges intelligencia-alapú projekt/kutatás esetén minimum-követelményként kell betartani: a „jó" fogalmát kell előre és minél pontosabban algoritmizáltan (Knuth-i alapon) definiálni.

2.1. Adatgyűjtés

A kísérletek elvégzéséhez az adatbázisok megtervezése és feltöltése volt a feladat, ahol kritikus elemként szolgált olyan adatbázisok reális, megbízható és objektív összeállítása, melyek már képesek a modellezési célkitűzések megvalósítására, azaz tiszta képet adnak arról, hogy melyek azok az információbiztonsági kontrollterületek, melyek a legnagyobb kockázatot jelentik a szervezetek számára a biztonságos üzemeltetés szempontjából. Mivel az információbiztonsági kitettségek, a kontrollok hiánya, és azok nem megfelelő üzemeltetése üzleti titok és a legtöbb szervezet számára sérülékenységet jelent azok nyilvánosságra hozatala, ezért a szervezetek a legtöbb esetben nem hajlandóak azok megfelelőségéről nyilatkozni, vagy eltakarva a teljes igazságot, képesek a hiányosságok ellenkezőjét is jelenteni akár szervezeten belül is félve a fegyelmi eljárások következményeitől.

A kutatáshoz felhasznált adatok forrása egy saját (megtervezett és feltöltött) adatbázis, melyben tartalmazott valós adatvagyont helyszíni vizsgálat (terepmunka) alatt gyűjtöttem két könyvvizsgáló, auditor és tanácsadó szervezettől – ezek hozzájárulása mellett, melyek mindegyike 500 főnél több munkavállalót foglalkoztat.

Az auditor szervezetek összesen 127 valós auditjelentést bocsátottak rendelkezésemre szöveges, anonimizált formában 2016-2020 között lezárult auditokról. Az auditjelentések 127 különböző egymástól független auditot reprezentálnak. A legrövidebb auditjelentés 5 oldalas, míg a leghosszabb 106 oldalas volt. Az adatok gyűjtésére 2020. április 1. és 2020. június 30. között került sor és összesen 3699 oldalnyi auditjelentés került feldolgozásra (átlagosan kerekítve 29 oldal auditjelentésenként).

Szükséges megjegyezni, hogy a disszertáció készítése során tudatosan kerültek kizárásra kérdőívre és/vagy interjúra alapozó adatgyűjtések, mivel

egyrészt, a kontrollhiányosságok meglétét egy vezető nem szándékozik külső féllel megosztani, másrészt, ebben a témakörben különösen csak a függetlenül gyűjtött adat az, ami reprodukálható, objektíven ellenőrizhető módon írja le a valóságot.

Az adatbázis nyers attribútumai az ISO/IEC 27001:2013 A melléklete alapján kerültek meghatározásra.

2.2. Alkalmazott algoritmusok és statisztikai eljárások

Az alábbi alfejezet szolgáltatja a hipotézisek bizonyításához felhasznált modellezési gyakorlatokat, melyek bemenetét az előző alfejezetben tárgyalt adatvagyon, valamint kapcsolódó számítási eredmények képezték.

Döntési fa

A döntési fa alapú eljárások az adathalmaz egy kitüntetett attribútumából (gyökér csúcs) kiindulva kérdések sorozatát fogalmazza meg, ahol mindegyik soron követő kérdés (belső csúcs) egy adott attribútum küszöbértékére (feltétel) vonatkozik. A döntési fa a döntéstámogató rendszer (robot-auditor) fejlesztésében került felhasználásra.

Neurális háló

A neurális hálók elméleti megközelítése az emberi agy biológiai neuronjainak gépi formába öltött megtestesítése, amiről a megnevezését is örökölte (KÁSA 2018). Hasonlóan a természetes neuronok felépítéséhez, a mesterséges neuronok is egymáshoz kapcsolódnak, ahol egy közvetítő közeg szállítja a feldolgozott információt neuronról-neuronra. A neuronok különböző rétegekben helyezkednek el: a bemeneti réteg fogadja a feldolgozáshoz szükséges adatokat, a kimeneti réteg szolgáltatja a predikciót, a közbenső rétegek, pedig transzformálják és segítségükkel jut el az információ a bemeneti neurontól a kimeneti neuroning. A kapcsolóelem a neuronok között a „dendrit”, mely létezhet bármely neuron pár között, de a kapcsolat mértéke a legtöbb esetben az, hogy minden neuron kapcsolódik minden soron követő neuronhoz. A transzferált információ az úton az egyik neurontól a másik felé „átalakul” a súlymátrixok és aktivációs függvények közbenjárásával, melynek összegzett és transzformált értéke a szinaptikus súly. A súlymátrixok a

neurális háló „tudásanyagát” tartalmazzák, és a neurális háló a megadott tanulási eljárás és az adatokban fellelhető minta szerint módosítja azt. A neurális háló a definiált veszteségfüggvény szerint kalkulálja a tényadat és becslés közötti különbséget, melyet visszafejt (pl. backpropagation) a háló súlymátrixainak módosítására (gradiens eljárás)(BARTA 2018a, BARTA – PITLIK 2018). A neurális háló a döntéstámogató rendszer (robot-auditor) fejlesztésében került felhasználásra.

Adaptív Boosting (ABM)

Az Adaptív Boosting (Adaptív Gyorsítás/Fokozás vagy AdaBoost vagy ABM – Adaptive Boosting Machine) egy együttes módszer, mely a hibásan osztályozott rekordok újrasúlyozásával, több alaposztályozó algoritmus felhasználásával egy erős osztályozót alkot. Az ABM egy szekvenciális együttes módszer, tehát az egyes alaptanulók egymásra épülnek (nem függetlenül operálnak), melyben a hangsúly a bemeneti adatokon összpontosul. Minden tanulásra felhasznált adathalmaz meghatározott mértékben az előző alaptanulótól függ, fokozatosan javítva a megelőző osztályozók által elkövetett hibákat. Az ABM az alaptanulók egyes predikcióit kombinálja legvégső osztályozásként, ahol veszi a részeredmények móduszát (többségi szavazás). Alapvetően döntési fáknál alkalmazott módszer, azonban karakterisztikái miatt bármilyen osztályozó szekvenciális összekapcsolására alkalmazható. Az ABM a döntéstámogató rendszer (robot-auditor) fejlesztésében került felhasználásra.

Gradiens Boosting (GBM)

A Gradiens Boosting (Gradiens Gyorsítás/Fokozás vagy GBM – Gradient Boosting Machine) az ABM algoritmushoz hasonlóan egy szekvenciális együttes módszer, tulajdonképp, az ABM egy finomhangolt variánsa, azonban a differenciálható veszteségfüggvény optimalizálásával (gradiens) törekszik az erős osztályozó algoritmus megalkotására (MASON et al. 1999, FRIEDMAN 2001). A GBM a döntéstámogató rendszer (robot-auditor) fejlesztésében került felhasználásra.

Kollaboratív szűrés

A módszer távolság/kapcsolat-metrikákkal képes a hasonló auditjelentések beazonosítására, evégett egy speciális klaszterező eljárásról van szó, ahol a leghasonlóbb objektumok egy listával térnek vissza, mely a feltételezett gyanús kontrollokat tartalmazza. A lista a két pl. leghasonlóbbnak ítélt objektum közötti különbségeket adja kimenetként, ahol súlyszámként az objektumok közötti hasonlóságok mértéke felhasználható. A döntéshozó preferenciája, hogy meghatározza az elfogadható hasonlóság küszöbértékét, vagy azon minimum és maximum auditjelentések számát, melyek a legközelebb találhatók a kiválasztott objektumhoz a többdimenziós térben. A kollaboratív szűrés a döntéstámogató rendszer (robot-auditor) fejlesztésében került felhasználásra.

Hasonlóságelemzés

A hasonlóságelemzés egyik eljárása a „mindenki másképp egyforma" elv matematikai megtestesítője, melyben a modellezés célja a legideálisabb objektum megtalálása, oly módon, hogy definiálunk egy fiktív célváltozót, mely értéke konstans, az objektumokat leíró attribútumok, pedig a megadott irány-preferenciák alapján rangsorolva vannak, azaz idealitás alapján attribútumonként sorba rendezhetők. Konstans célváltozó alkalmazásával az objektum-azonosságok kényszerként jelennek meg a modellezési folyamatban, mely az anti-diszkriminációs számítások, vagyis a hasonlósági skála központi eleme (norma-értéke). A hasonlóságelemzés optimalizáló eljárás, amely minden egyes attribútumhoz egy lépcsős függvényt approximál, mely a célváltozóhoz történő hozzájárulás mértékét határozza meg (BÁNKUTI 2010). A lépcsők (idealitás-súlyszámok) közötti különbségek minimuma mindenkor nagyobb kell, hogy legyen mint nulla, tehát különböző rangsorszámokhoz különböző lépcsős értékek kell, hogy tartozzanak. Hasonlóságelemzés által, így, a normától jelentősen eltérő objektumok úm. kiugró értékként jelentkeznek, az objektumok célváltozó értéke szerint rangsorolhatóvá válnak. A hasonlóságelemzés anti-diszkriminatív eljárása egy speciális neurális hálónak, módszertanilag felügyelet nélküli gépi tanuló algoritmusnak tekinthető.

2.3. Jóságmetrikák

A kutatási objektív eredményeinek megállapítására és a modellek jóságának mérésére az alábbi metrikákat alkalmazom:

- **Igaz pozitívak száma**: A modell igaz találatainak száma, amikor az helyesen eltalálta a fellépő gyanús kontrollt.
- **Igaz negatívak száma**: A modell igaz találatainak száma, amikor helyesen eltalálta, hogy adott kontroll nem gyanús eset.
- **Hamis pozitívak száma**: A modell hamis találatainak száma, amikor helytelenül arra a következtetésre jutott, hogy az adott kontroll gyanús eset, valójában pedig nem.
- **Hamis negatívak száma**: A modell hamis találatainak száma, amikor helytelenül arra a következtetésre jutott, hogy az adott kontroll nem gyanús eset, valójában pedig az volt.
- **Pontosság**: Az Igaz Pozitívak és Igaz negatívak összegének és az összes eset hányadosa.
- **Precizitás**: A Precizitás az Igaz pozitívak számának, és az Igaz pozitívak és Hamis pozitívak összegének aránya. Kiegyensúlyozatlan adathalmazok esetén preferált mutató, melynek célkitűzése a modell pontosság becslésének relevanciájára vonatkozik a „Mennyire érvényes az eredmény?” kérdés megválaszolásával.
- **Fedés**: A Fedés az Igaz pozitívak számának, és az Igaz pozitívak és Hamis negatívak összegének aránya. Kiegyensúlyozatlan adathalmazok esetén preferált mutató, melynek célkitűzése a modell pontosság becslésének relevanciájára vonatkozik a „Mennyire teljes az eredmény?” kérdés megválaszolásával.
- **F1-Pont**: Az F1-Pont a Precizitás és Fedés kombinációjából származtatott jóságmetrika, harmonikus középértéke.
- **Variancia**: A modell variancia mutatója fejezi ki a tanulás és tesztelési adathalmazon mért pontosságok közötti különbséget.
- **ROC-görbe és AUROC mutató:** A módszer alkalmazása lehetővé teszi az osztályozás igaz pozitív és hamis negatív találatok arányai közötti kompromisszum feltárását és értelmezését.
- **PR-görbe és AUPRC mutató**: A Precizitás és Fedés mutatók közötti kompromisszumot szemlélteti hasonlóan a ROC-görbékhez.

3. EREDMÉNYEK

A fejezet prezentálja a felállított hipotézisek bizonyítását a terepmunkán gyűjtött adatok elemzésére alapozva az előző fejezetben ismertetett matematikai apparátusok felhasználásával.

3.1. Gyanúgenerálás információbiztonsági kontrollhiányosságok detektálására

3.1.1. Felügyelt gépi tanuló modellek futtatása és kiértékelése

A felügyelt gépi tanuló algoritmusok modellezési szakaszában a transzformált adathalmazt három különböző eljárással dolgoztam fel:

- ABM (Adaptív Boosting) meta-tanulással ellátott döntési fa;
- GBM (Grádiens Boosting) meta-tanulással ellátott döntési fa;
- NN (Neurális Háló).

Az algoritmusok fejlesztését és „üzembehelyezését” követően az 1. táblázatban ismertetett, a tesztelési halmazon mért, eredményeket rögzítettem, melyet a keresztvalidációs kiértékelés szolgáltatott.

A legtöbb igaz pozitív találatot az ABM produkálta (80 db), majd az NN (74 db), legvégül a GBM, mely csak 46 esetben volt képes igaz pozitív találatot teljesíteni, ezzel nagyságrenddel a két másik alkalmazott módszer performanciája alatt maradt, mely a többi mutatóban is tetten érhető. Azonban, az igaz negatívak számát vizsgálva a GBM (954 db) felülkerekedett az ABM (919 db) és NN (916 db) algoritmusokon, mely vélelmezhetően azt támasztja alá, hogy a GBM alapértelmezetten a mintákat (kontrollokat) megfelelőnek értékelte, alacsony hamis pozitív találatokat produkált. Az F1-Pont fejezi ki a Fedés és Precizitás együttes harmóniáját egy 0-tól 1-ig terjedő skálán (melyet 100-al megszorozva százalékos teljesítményt kapunk, hasonlóan a Pontosság, Precizitás, Fedés és AUROC esetén), mely az ABM (0.57) alkalmazásnál a legkedvezőbb. Az ABM és GBM AUROC mutatói (0.85) a legideálisabbak, valamint a GBM algoritmus rendelkezik a legkisebb varianciával (0.04).

Összességében, az F1-Pont mutatót vizsgálva az ABM eljárás nevezhető ki a három algoritmus közül a legkielégítőbbnek, azonban ez vezetői döntéshez kötött. Amennyiben a cél a rendelkezésre álló erőforrások függvényében az igaz pozitív találatok maximalizálása annak árán is, hogy a rendszer számos hamis pozitívat generál, az ABM implementálása indokolt.

1. táblázat: Felügyelt gépi tanuló algoritmusok performancia metrikái

Performancia mutatók	Adaptív Boosting (ABM)	Gradiens Boosting (GBM)	Neurális Háló (NN)
Igaz pozitívak száma (db)	80	46	74
Igaz negatívak száma (db)	921	954	916
Hamis pozitívak száma (db)	42	9	47
Hamis negatívak száma (db)	79	113	85
Pontosság	0.89	0.89	0.88
Precizitás	0.66	0.83	0.61
Fedés	0.50	0.29	0.47
F1-Pont	0.57	0.43	0.53
Variancia	0.08	0.04	0.11
AUROC	0.85	0.85	0.82
AUPRC	0.58	0.61	0.56

Forrás: Saját szerkesztés

Általánosságban megállapítható, hogy mind a három módszer túlilleszkedett (Variancia > 0), azaz az általánosító képességük nem optimális, a tanulási halmazon mért pontosság meghaladja a tesztelési metrika eredményét, tehát az algoritmusok az adatban meghúzódó zajokat, valamint egyéb a kizárólag a tanulási halmazra jellemző karakterisztikákat is beépítettek az approximációba, mely várható volt. A variancia értéke a legkisebb a GBM esetében (0.04), míg a legnagyobb az NN modellben (0.11), ahol a tanulás 2,000 db minta környékén szinte maximális (az algoritmus konvergált), tehát a modell a tanulóhalmazon minden mintát helyesen értékelt, ami a magas túlilleszkedés és alacsony általánosító képesség jeleit vélelmezik, ezért éles környezetben megkérdőjelezhető az alkalmazás fenntarthatósága. A túlilleszkedés mértéke csökkenthető addicionális tanulási adatok beszerzésével, valamint a modellek regularizációjával.

Összefoglalva a kiértékelteket, az AUROC eredményeit vizsgálva kijelenthető, hogy mind a három alkalmazott módszer képes volt a kontrollhiányosságok között meghúzódó mintázat matematikai leképezésére, mivel a kívánt véletlen szint felett teljesítettek (az AUROC értéke meghaladta a 0.5-et). Továbbá, mivel az F1-Pont értéke értelmezhető és értéke nagyobb nullánál (az algoritmusok nem minden rekordot egy megadott osztályba soroltak), így kijelenthető, hogy a rendszerek teljesítménye jobb a véletlen találgatásnál, a kontrollhiányosságok együttes megléte között összefüggés állapítható meg.

3.1.2. Gyanúgenerálás felügyelet nélküli gépi tanuló módszerekkel

A felügyelet nélküli módszerek is alkalmazhatók gyanúgenerálása, azonban a teljesítmények mérése körülményes, mivel nincs visszaigazolt célváltozó az eljárások sajátosságainak köszönhetően. A kollaboratív szűrés és a gyanúgenerálás között párhuzam vonható, tehát a gyanúgenerálásra alkalmazott döntéselőkészítő rendszer, ajánlórendszerként is értelmezhető, ahol az auditjelentés számára kell mesterségesen a historikus adatok alapján gyanús kontrollt „ajánlani".

A kísérletben kollaboratív szűrést alkalmaztam, ahol az algoritmusok esetén három különböző távolság/kapcsolat-metrikát használtam, így három különböző modell állt elő: EUC, PEA és COS.

Tesztelésre 25 véletlen minta került kiválasztásra, mely a teljes populáció (127) 19.69%-a. Az algoritmusok futtatását követően a 2. táblázatban ismertetett eredményeket rögzítettem.

2. táblázat: Felügyelet nélküli gépi tanuló algoritmusok performancia metrikái

Performancia mutatók	EUC	PEA	COS
Igaz pozitívak száma (db)	16	18	20
Igaz negatívak száma (db)	2,739	2,721	2,710
Hamis pozitívak száma (db)	111	129	140
Hamis negatívak száma (db)	9	7	5
Pontosság	0.96	0.95	0.95
Precizitás	0.13	0.12	0.13
Fedés	0.64	0.72	0.80
F1-Pont	0.21	0.21	0.22

Forrás: Saját szerkesztés

A kiválasztott minta esetén az alkalmazott módszertan függvényében, az algoritmusok összesen potenciálisan 25 gyanús kontrollról voltak képesek a feltételezett kontrollhiányosságokról igaz pozitív véleményt alkotni. Ebből a COS (20 db) 80%-át, a PEA (18 db) 72%-át és az EUC (16 db) 64%-át tudta helyesen megítélni a könnyen előre megjósolható magas hamis pozitív találati arány mellett. A Fedés metrika, mely az igaz és hamis pozitívak arányát fejezi ki, kedvező értéket mutat mindhárom modell esetében, ami a mesterségesen kikényszerített látens célváltozó alkalmazásának köszönhető. Mivel a visszatérési listák az összes eltérést tartalmazzák mely a kiválasztott objektumok között felmerül, ezért a Precizitás alacsony szintje nem meglepő. Az F1-Pont közel azonos értéke nem enged szignifikáns különbségre

következtetni az eljárások között, minimális differenciával a COS tekinthető az F1-Pont aspektusából a legjobb módszernek.

3.1.3. Gyanúgenerálás hibrid megközelítésben

A hibrid modellezés alapmegközelítése a releváns szakirodalmat felhasználva és következtetéseket levonva, hogy az előző alfejezetben ismertetett felügyelet nélküli algoritmusok segítségével növeljük a becslések jóságmetrikáit, tehát a felügyelt módszerekbe beépítve az „ajánlórendszer ajánlásait", azok további hasznos bemenetre tegyenek szert, vélelmezhetően addicionális összefüggéseket tartalmazva.

A hibrid modellezésben a COS modell kimeneti értékeit építettem be a felügyelt módszerekbe (ABM, GBM, NN) ceteris paribus, a felügyelt eljárások minden konfigurációs beállításait megőrizve, kizárólag így a tanulási halmazon változtatást (bővítést) eszközölve. A COS modellt futtattam az összes auditjelentésen, bemenetet képezve a tanulási és teszthalmaz számára is. A modell a lehetséges 115 kontrollból, 98 esetben tért vissza gyanúmomentummal, így a tanulási halmaz összesen 98 bináris változóval egészült ki, mely a COS modell ajánlásait tartalmazza. Mivel a felügyelt eljárások beállításai identikusak, ezért objektíven meg lehet győződni a hibrid megoldás javító/rontó hatásairól.

A hibrid modellt éleskörnyezetben futtatva a felügyelt módszerek performancia metrikái az alábbiak szerint alakultak (3. 4. és 5. táblázat). A hibrid megközelítést nélkülöző módszereket „egyszerű" modelleknek neveztem el.

3. táblázat: Egyszerű és hibrid ABM performancia metrikái

Performancia mutatók	Egyszerű ABM	Hibrid ABM	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	80	89	+11.25%	Növelés	Javulás
Igaz negatívak száma (db)	921	944	+2.50%	Növelés	Javulás
Hamis pozitívak száma (db)	42	19	-54.76%	Csökkentés	Javulás
Hamis negatívak száma (db)	79	70	-11.39%	Csökkentés	Javulás
Pontosság	0.89	0.92	+3.37%	Növelés	Javulás
Precizitás	0.66	0.82	+24.24%	Növelés	Javulás
Fedés	0.50	0.56	+12.00%	Növelés	Javulás
F1-Pont	0.57	0.67	+17.54%	Növelés	Javulás
Variancia	0.08	0.06	-0.25%	Csökkentés	Javulás
AUROC	0.85	0.90	+5.88%	Növelés	Javulás
AUPRC	0.58	0.70	+20.69%	Növelés	Javulás

Forrás: Saját szerkesztés

4. táblázat: Egyszerű és hibrid GBM performancia metrikái

Performancia mutatók	Egyszerű GBM	Hibrid GBM	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	46	62	+34.78%	Növelés	Javulás
Igaz negatívak száma (db)	954	958	+0.42%	Növelés	Javulás
Hamis pozitívak száma (db)	9	5	-44.44%	Csökkentés	Javulás
Hamis negatívak száma (db)	113	97	-14.16%	Csökkentés	Javulás
Pontosság	0.89	0.91	+2.25%	Növelés	Javulás
Precizitás	0.83	0.93	+12.05%	Növelés	Javulás
Fedés	0.29	0.39	+34.48%	Növelés	Javulás
F1-Pont	0.43	0.55	+27.91%	Növelés	Javulás
Variancia	0.04	0.03	-0.25%	Csökkentés	Javulás
AUROC	0.85	0.85	0%	Növelés	Semleges
AUPRC	0.61	0.71	+16.39%	Növelés	Javulás

Forrás: Saját szerkesztés

5. táblázat: Egyszerű és hibrid NN performancia metrikái

Performancia mutatók	Egyszerű NN	Hibrid NN	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	74	76	+2.70%	Növelés	Javulás
Igaz negatívak száma (db)	916	926	+1.09%	Növelés	Javulás
Hamis pozitívak száma (db)	47	37	-21.28%	Csökkentés	Javulás
Hamis negatívak száma (db)	85	83	-2.35%	Csökkentés	Javulás
Pontosság	0.88	0.89	+1.14%	Növelés	Javulás
Precizitás	0.61	0.68	+11.48%	Növelés	Javulás
Fedés	0.47	0.47	0%	Növelés	Semleges
F1-Pont	0.53	0.56	+5.66%	Növelés	Javulás
Variancia	0.11	0.11	0%	Csökkentés	Semleges
AUROC	0.82	0.85	+3.66%	Növelés	Javulás
AUPRC	0.56	0.62	+10.71%	Növelés	Javulás

Forrás: Saját szerkesztés

A hibrid modellezés szemmel láthatóan beváltotta a hozzá fűzött reményeket, szinte kivétel nélkül javulás mutatkozik az összes metrikában (minimálisan a GBM AUROC mutatója esetén is, azonban a kerekítés ezt nem engedi láttatni).

Az ABM algoritmus legnagyobb teljesítménynövekedése a hamis pozitív találatok jelentős csökkentésében érhető tetten, mely hatással van a Precizitás mutatóra. A hamis pozitívak efféle hangsúlyos redukálása az audit részéről erőforrás (humán és pénzügyi) megtakarítást enged, azaz kevesebb hamisan feltételezett hiányosságokkal rendelkező kontroll újraértékeléséről és ismételt teszteléséről szükséges meggyőződni. Kiemelendő még az igaz pozitív találatok, valamint AUROC mutató növekedése, mely utóbbi a legkedvezőbb a három modell tekintetében.

A GBM eljárás igaz pozitív találatainak száma emelkedett drámaian, ezzel párhuzamosan a hamis pozitívak száma közel felére csökkent, mely az ABM-mel összehasonlítva is kicsivel több, mint negyede. A GBM erőssége így a minimális hamis pozitív találatok száma, illetve, a modell varianciája, mely jobb általánosító képességet feltételez, mely alapesetben kevésbé volt várható.

Az NN modell esetén is megfigyelhető a jóságmetrikák javulása, azonban kevésbe szembetűnő, mint az ABM és GBM modelleknél, mely valószínűsíthetően a modell eddig is feltárt (1. táblázat) erős túlilleszkedésének köszönhető, mely a hibridizációt követően még inkább hangsúlyos.

3.2. Genetikai potenciál keresése a gépi tanulás adatvagyonának redukált felhasználásával

A tanuló algoritmus tanulási halmazának optimális felhasználásával meghatározott genetikai potenciálja keresési algoritmussal közelítendő, melyben a keresés akkor járt sikerrel, ha a kiindulóponttól különböző, objektíven felismerhető, ideálisabb eredményt érünk el, egy vagy több lépésben úgy, hogy a genetikai potenciál, mint szélsőérték azonosítható be iteratív módon.

Az általam fejlesztett kereső eljárás az alábbi táblázatokban (6., 7. és 8. táblázat) közöltek szerint javította a hibrid ABM, GBM és NN algoritmusok teljesítményét, ahol az ABM esetében már az első iterációban megtörtént a maximum megtalálása, míg az NN módszernél ez a második iterációban következett be, minimális javulást hozva.

6. táblázat: A hibrid és a kereső eljárással javított ABM performancia metrikái

Performancia mutatók	Hibrid ABM	Javított Hibrid ABM	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	89	102	+14.60%	Növelés	Javulás
Igaz negatívak száma (db)	944	940	-0.42%	Növelés	Romlás
Hamis pozitívak száma (db)	19	23	+21.05%	Csökkentés	Romlás
Hamis negatívak száma (db)	70	57	-18.57%	Csökkentés	Javulás
Pontosság	0.92	0.93	+1.09%	Növelés	Javulás
Precizitás	0.82	0.82	0%	Növelés	Semleges
Fedés	0.56	0.64	+14.29%	Növelés	Javulás
F1-Pont	0.67	0.72	+7.46%	Növelés	Javulás
Variancia	0.06	0.06	0%	Csökkentés	Semleges

AUROC	0.90	0.86	-4.44%	Növelés	Romlás
AUPRC	0.70	0.71	+1.43%	Növelés	Javulás

Forrás: Saját szerkesztés

7. táblázat: A hibrid és a kereső eljárással javított GBM performancia metrikái

Performancia mutatók	Hibrid GBM	Javított Hibrid GBM	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	62	78	+25.81%	Növelés	Javulás
Igaz negatívak száma (db)	958	955	-0.31%	Növelés	Romlás
Hamis pozitívak száma (db)	5	8	+60.00%	Csökkentés	Romlás
Hamis negatívak száma (db)	97	82	-15.46%	Csökkentés	Javulás
Pontosság	0.91	0.92	+1.10%	Növelés	Javulás
Precizitás	0.93	0.91	-2.15%	Növelés	Romlás
Fedés	0.39	0.48	+23.08%	Növelés	Javulás
F1-Pont	0.55	0.64	+16.36%	Növelés	Javulás
Variancia	0.03	0.03	0%	Csökkentés	Semleges
AUROC	0.85	0.85	0%	Növelés	Semleges
AUPRC	0.71	0.68	-4.22%	Növelés	Romlás

Forrás: Saját szerkesztés

8. táblázat: A hibrid és a kereső eljárással javított NN performancia metrikái

Performancia mutatók	Hibrid NN	Javított Hibrid NN	Változás mértéke (%)	Cél	Változás hatása (javulás/romlás)
Igaz pozitívak száma (db)	76	80	+5.26%	Növelés	Javulás
Igaz negatívak száma (db)	926	926	0%	Növelés	Semleges
Hamis pozitívak száma (db)	37	37	0%	Csökkentés	Semleges
Hamis negatívak száma (db)	83	79	-4.82%	Csökkentés	Javulás
Pontosság	0.89	0.90	+1.12%	Növelés	Javulás
Precizitás	0.68	0.68	0%	Növelés	Semleges
Fedés	0.47	0.50	+6.38%	Növelés	Javulás
F1-Pont	0.56	0.58	+3.57%	Növelés	Javulás
Variancia	0.11	0.11	0%	Csökkentés	Semleges
AUROC	0.85	0.81	-4.71%	Növelés	Romlás
AUPRC	0.62	0.62	0%	Növelés	Semleges

Forrás: Saját szerkesztés

A táblázatok szemléltetik, hogy az algoritmusok egyes mutatói romlottak az F1-Pont javítása érdekében, mivel a kereső eljárás az F1-Pont ideálisabb

értékének letapogatására szolgált a többi metrika figyelmen kívül hagyásával. Kivétel nélkül ez az igaz pozitív (és így a hamis negatív) találatok arányának javításával volt elérhető. Az ABM és GBM alapú rendszerek igaz negatív és hamis pozitív találatainak száma csökkent, míg az NN esetében nem változott. Az AUROC mutatók a GBM kivételével romlottak, míg a varianciára az eljárás nem volt hatással, azaz ugyanolyan ütemben javult a tanulás pontossága, mint a tesztelésé. Összességében kijelenthető, hogy a kereső eljárás sikerrel járt, mind a három algoritmus esetén javult az F1-Pont a tanulóhalmaz optimumhoz közelibb felhasználásával.

3.3. Modell-preferencia levezetése klasszikus tesztelési eljárások nélkül

A tesztelés nélküli jóságmérés feltett szándéka, hogy a rendszer révén generált gyanús objektumok leíró tulajdonságaiból létrehozunk egy olyan adathalmazt, melyben több dimenzió mentén meghatározható, mely rendszerválaszokat preferáljuk (preferálja pl. a mindenkori döntéshozó, vagy a döntéstámogató rendszer maga) egy másiknál jobban. Tehát, melyek azok a jellemzők egy-egy rendszerválasz esetén, amelyek magasabb bizonyosságot (vö. konzisztenciát (PITLIK et al. 2017)) nyújthatnak a döntéstámogató rendszer felhasználójának. Egy olyan alkalmazás, mely nagy mennyiségű gyanút generál különböző kontrollokról, különböző kontrollterületekről, vélhetően elbizonytalanítja a döntéshozót és magasabb hamis pozitív eredményhez vezethet. Az alkalmazás, mely egységesen és homogén módon képes egy irányba mutatni (pl. egy kontrollt jelöl meg markánsan, mint gyanús eset, vagy hasonló struktúrájú/működésű kontrollok halmazát javasolja felülvizsgálatra) annak predikcióját vélhetően könnyebben lehet elfogadtatni, mivel a rendszer minden egyes részegysége más utak bejárásával is hasonló konklúzióra jut. A meghatározandó leíró tulajdonságok, ennél fogva, a sikeres és sikertelen modellezés minél szélesebb körű aspektusait kell, hogy definiálják modelljóságot leíró mérőszámok formájában. Mivel ezen aspektusok száma végtelen, így a klasszikus egytényezős optimalizálást kényszerűen fel kell, hogy váltsa egy többdimenziós érték-fogalom a Hartman-elvet finomhangolandó (PITLIK et al. 2020).

Saját modell-preferencia levezető algoritmus fejlesztésével felügyelet nélküli modellek esetén vizsgáltam 12 db modell viselkedését. A performancia metrikákat a 9. táblázat szemlélteti.

9. táblázat: Felügyelet nélküli modellértékelés összefoglaló táblázata

Modell azonosító	Naiv Átlagok	Naiv rangsor	Hasonlóság-elemzés idealitások	Hasonlóság-elemzés rangsor
R_NS_NRS_C	5.75	6	997	8
R_NS_NRS_E	**3.00**	**1**	**1012**	**1**
R_NS_NRS_P	5.75	6	1002	4
R_NS_RS_C	5.75	6	997	8
R_NS_RS_E	4.75	3	991	11
R_NS_RS_P	5.75	6	1002	4
R_S_NRS_C	6.00	10	996	10
R_S_NRS_E	3.75	2	1003	3
R_S_NRS_P	5.00	5	1008	2
R_S_RS_C	6.25	11	1002	4
R_S_RS_E	4.75	3	991	11
R_S_RS_P	6.25	11	1002	4

Forrás: saját szerkesztés

A táblázat alapján az olvasható le, hogy az R_NS_NRS_E a legideálisabb modell a felvetett gyanúgenerálási probléma megoldására az összes fejlesztett modell között, melyet mind a két rangsor (naiv átlagolás és hasonlóságelemzés is) alátámaszt.

Elvégezve a teszthalmaz kiértékelését a szimulált környezetben és összehasonlítva az eredményeket a leírtakkal, a következő táblázat szemlélteti a modellek találati arányaira vonatkozó összefoglaló eredményeket (10. táblázat).

10. táblázat: Felügyelet nélküli modellek értékelő táblázata

Modell azonosító	Pontosság	Precizitás	Fedés	F1-Pont
R_NS_NRS_C	0.78	0.12	0.42	0.14
R_NS_NRS_E	0.84	0.24	0.67	0.32
R_NS_NRS_P	0.83	0.13	0.42	0.16
R_NS_RS_C	0.78	0.12	0.42	0.14
R_NS_RS_E	0.61	0.22	0.92	0.30
R_NS_RS_P	0.83	0.13	0.42	0.16

R_S_NRS_C	0.75	0.11	0.42	0.13
R_S_NRS_E	0.81	0.22	0.67	0.30
R_S_NRS_P	0.80	0.12	0.42	0.14
R_S_RS_C	0.74	0.11	0.42	0.13
R_S_RS_E	0.56	0.18	0.92	0.25
R_S_RS_P	0.79	0.12	0.42	0.15

Forrás: Saját szerkesztés

Elvégezve az alkalmazott performancia metrikák (Pontosság, Precizitás, Fedés, F1-Pont) kiértékelését a hasonlóságelemzés által nyújtott anti-diszkriminatív matematikai apparátus által, ahol a norma értéke 1,000 (11. táblázat). Kijelenthető, hogy az R_NS_NRS_E modell mutatói összességében a legideálisabbak (1,016.60), tehát a visszacsatolás eredménye az, hogy jogosan lett a modell győztesként kihirdetve.

11. táblázat: Modelljóság becslés anti-diszkriminatív eljárással

ID	Modell azonosító	Pontosság	Precizitás	Fedés	F1-Pont	Y_0
1	R_NS_NRS_C	0.78	0.12	0.42	0.14	**995.60**
2	R_NS_NRS_E	0.84	0.24	0.67	0.32	**1016.60**
3	R_NS_NRS_P	0.83	0.13	0.42	0.16	**1005.60**
4	R_NS_RS_C	0.78	0.12	0.42	0.14	**995.60**
5	R_NS_RS_E	0.61	0.22	0.92	0.30	**1003.60**
6	R_NS_RS_P	0.83	0.13	0.42	0.16	**1005.60**
7	R_S_NRS_C	0.75	0.11	0.42	0.13	**986.60**
8	R_S_NRS_E	0.81	0.22	0.67	0.30	**1011.60**
9	R_S_NRS_P	0.80	0.12	0.42	0.14	**997.60**
10	R_S_RS_C	0.74	0.11	0.42	0.13	**985.60**
11	R_S_RS_E	0.56	0.18	0.92	0.25	**998.60**
12	R_S_RS_P	0.79	0.12	0.42	0.15	**997.60**

Forrás: Saját szerkesztés

A 9. táblázat mutatta be a modellek rangsorolását, mely az objektum-leíró tulajdonságok felhasználásával került levezetésre. Korrelációt számítva a 9. és 11. táblázatban ismertetett hasonlóságelemzés idealitásokkal (Y_0) a korrelációs együttható értéke: 0.42, mely közepes pozitív kapcsolatot feltételez. Kiemelendő, a modell-preferencia mind a két eljárás esetén az R_NS_NRS_E volt, független visszaigazolást nyert a levezetés eredményessége.

4. ÚJ ÉS ÚJSZERŰ TUDOMÁNYOS EREDMÉNYEK

Kutatásom eredményéül az alább felsorolt új és/vagy újszerű tudományos eredményeket rögzítettem a hipotézisekkel összhangban:

***H1**: Az információbiztonsági auditjelentések szöveges eredményeiből strukturált adatbázist alkotva és bemenetként a mesterséges intelligencia fogalomkörébe illeszthető eszközökkel azt feldolgozva, az auditok során feltárni kívánt kontrollhiányosságok megléte a véletlen találgatásnál nagyobb valószínűséggel kimutathatók, azaz a kontrollhiányosságok konstellációi matematikailag értelmezhető összefüggéseket hordoznak magukban.* **IGAZOLT**

- Az audit által közölt megállapítások, azaz kontrollhiányosságok, melyek szöveges riport formájában kerülnek közlésre a szervezetek vezetői és befektetői számára, logikus kontextusfüggő formában kategorizálhatók, a megállapítások tematikájára irányuló tudatos rendszerezést követően a gép számára strukturált formában átadható;
- Az auditok által dokumentált megállapítások együttes konstellációi között összefüggés tapasztalható mely objektíven, matematikai apparátusok felhasználásával kimutatható, így döntéstámogató rendszer építhető, mely képes az auditok hatékonyságát növelni azok objektív minőségbiztosítása révén:
    - Egyszerű és hibrid ABM F1-Pont értékei rendre: 0.57 és 0.67;
    - Egyszerű és hibrid GBM F1-Pont értékei rendre: 0.43 és 0.55;
    - Egyszerű és hibrid NN F1-Pont értékei rendre: 0.53 és 0.56;
- A kontrollhiányosságok együttes megléte alapján előre a döntéstámogató rendszer által nem feldolgozott adatokon a kontrollhiányosságok adott kontrollra becsülhetők, a véletlen találgatásnál (AUROC > 0.5) ideálisabb eredmény érhető el (4.2.6. alfejezet):
    - Egyszerű és hibrid ABM AUROC értékei rendre: 0.85 és 0.90, AUPRC értékei rendre: 0.58 és 0.70;
    - Egyszerű és hibrid GBM AUROC értékei rendre: 0.85 és 0.85, AUPRC értékei rendre: 0.61 és 0.71;
    - Egyszerű és hibrid NN AUROC értékei rendre: 0.82 és 0.85, AUPRC értékei rendre: 0.56 és 0.62;

***H1.1**: A gyanúgenerálás, mint megoldandó üzletileg értelmezett probléma sajátosságait értékelve, a kontrollhiányosságok detektálása megoldható felügyelt és felügyelet nélküli gépi tanuló eljárásokkal is.* **IGAZOLT**

- A kontrollhiányosságok detektálását osztályozási problémaként azonosítva, a gyanúgenerálás lehetséges felügyelt gépi tanuló modellezés keretében megvalósítani, ahol a dedikált célváltozó az adott kontroll megfelelőségére irányuló állapotot jelöli:
    - Egyszerű ABM F1-Pont értéke: 0.57;
    - Egyszerű GBM F1-Pont értéke: 0.43;
    - Egyszerű NN F1-Pont értéke: 0.53;
- A kontrollhiányosságok detektálását ajánlórendszerként azonosítva, a gyanúgenerálás lehetséges felügyelet nélküli gépi tanuló modellezés keretében megvalósítani, amely célváltozó hiányában, egy listával tér vissza az azonosított gyanúmomentumokról, melyet az auditjelentések hasonlósága alapján vél felfedezni (4.2.5. alfejezet):
    - EUC F1-Pont értéke: 0.21;
    - PEA F1-Pont értéke: 0.21;
    - COS F1-Pont értéke: 0.22;

***H1.2**: A gyanúgenerálás teljesítménye fokozható hibrid megközelítésben, azaz a felügyelt és nem felügyelt módszerek együttes felhasználásának a kutatásban alkalmazott releváns performancia metrikái ideálisabb értékeket mutatnak, mint önálló alkalmazásban.* **IGAZOLT**

- A kontrollhiányosságok detektálásának teljesítménye a felügyelt és felügyelet nélküli módszerek együttes alkalmazásával javítható, ahol a felügyelet nélküli ajánlórendszer javaslatot tesz a gyanúmomentumokról, melyet a felügyelt módszerek beépítenek a döntéselőkészítésbe, így addicionális információként csökken a bizonytalanság mértéke:
    - Egyszerű és hibrid ABM F1-Pont értékei rendre: 0.57 és 0.67;
    - Egyszerű és hibrid GBM F1-Pont értékei rendre: 0.43 és 0.55;
    - Egyszerű és hibrid NN F1-Pont értékei rendre: 0.53 és 0.56;

***H1.3**: A hibrid modell többlet-információs értéket teremtve képes az egyszerű modellek általánosító képességén javítani.* **RÉSZBEN IGAZOLT**

- A felügyelt gépi tanuló rendszerek varianciáit csökkenti a hibrid megközelítés, ezért általánosító képességük ideálisabb, mint önálló alkalmazásban:
    - Egyszerű és hibrid ABM Variancia értékei rendre: 0.08 és 0.06;
    - Egyszerű és hibrid GBM Variancia értékei rendre: 0.04 és 0.03;
    - Egyszerű és hibrid NN Variancia értékei rendre: 0.11 és 0.11 (a neurális háló esetén nem csökkent a variancia);

***H2**: A döntéstámogató rendszer genetikai potenciálja letapogatható hasonlóságelemzéssel ellátott kereső eljárással a tanításra alkalmazott adathalmaz irányított feldolgozásán keresztül, úgy, hogy a genetikai potenciálhoz vezető kereső eljárás a genetikus algoritmusok esetén alkalmazott véletlen mutáció és a populáció egyedeinek keresztezése nélkül is képes ideálisabb eredményt szolgáltatni.* **IGAZOLT**

- A hasonlóságelemzés által generált lépcsős függvények alkalmasak keresési eljárásokban történő felhasználásra;
- A döntéstámogató rendszer genetikai potenciálja kereshető a tanulóhalmazhoz tartozó racionális leíró attribútumok érték-irány levezetésével;
- A keresési eljárás meg tudja határozni a rendelkezésre álló populáció értékei alapján, hogy adott modell vélhetően elérte-e már a tanulási halmaz optimalizálási kísérlete révén genetikai potenciálját, valamint, hogy milyen attribútumok módosítása szükséges az ideális célváltozó irányába történő elmozduláshoz:
    - Az algoritmus a legelőkelőbb helyen álló súlyszámok összegeként megadja a modell genetikai potenciálját adott célváltozó tükrében (a GBM esetében a 11. iterációban az F1-Pont genetikai potenciálja: 0.71);
    - Az algoritmus meghatározta a súlyszámok különbségéből adódóan, mely attribútumok módosítása volt szükséges az ideális célváltozó irányába történő elmozduláshoz (pl. a GBM esetében az 1. iterációban az f_{10} attribútum meghatározott irány-preferencia szerinti növelésével lehetett elérni, ahol annak értéke 1.9 volt).
- A kereső eljárás véletlen mutáció nélkül is képes ideálisabb irányt megjelölni:
    - Az algoritmus véletlen mutáció nélkül már az 1. iterációban javította a GBM F1-Pont értékét 0.60833-ról 0.61925-ra, mely 1.80 %-os javulást jelentett;
    - Az algoritmus véletlen mutáció nélkül a 11. iterációban a GBM F1-Pont értékét a kezdeti 0.60833-ról 0.63673-ra javította, mely 4.57 %-os javulást jelentett.

- A kereső eljárás az egyedek keresztezése nélkül is képes ideálisabb irányt megjelölni:
    - Az algoritmus az egyedek keresztezése nélkül már az 1. iterációban javította a GBM F1-Pont értékét 0.60833-ról 0.61925-ra, mely 1.80 %-os javulást jelentett.

    - Az algoritmus az egyedek keresztezése nélkül a 11. iterációban a GBM F1-Pont értékét a kezdeti 0.60833-ról 0.63673-ra javította, mely 4.57 %-os javulást jelentett.
- A kereső eljárás nem igényel a nyitó populáción túl köztes/itérációnként új populációkat;
- Az irányok automatikusan minden iterációban újra paraméterezhetők (4.3.2. alfejezet).

***H3**: A mesterséges intelligenciával ellátott döntéstámogató rendszerek teljesítményalapon a gépi tanuló alkalmazások klasszikus tesztelési eljárásai nélkül is rangsorolhatók, a predikciók, mint generált gyanúforrások leíró tulajdonságainak érték-irány levezetésével és az ezen adatokat feldolgozó matematikai apparátussal, mely automatizáltan képes a preferált modellek objektív meghatározására.* **IGAZOLT**

- A modell-preferencia keresésére fejlesztett eljárás alkalmas az ideális modell megtalálására egy előre definiált modellhalmazból, mely független teszteléssel objektíven bizonyítható:
    - Az R_NS_NRS_E modell volt a legideálisabb, melynek hasonlóságelemzés idealitása: 1012 (a legmagasabb az összes modell között), melyet alátámasztott a függetlenül mért modelljóság becslés (a hasonlóságelemzés idealitása: 1017, a legmagasabb az összes modell között).
- A modellek önerősítő (konzisztencia) mechanizmusai (rendszerválaszok viselkedései) és modelljóság metrikák között összefüggés tapasztalható, így az ellentmondásosság felfedezése révén a teljesítmény becsülhetővé válik:
    - Korrelációt számítva a hasonlóságelemzés idealitásokkal (Y_0) a korrelációs együttható értéke: 0.42, mely közepes pozitív kapcsolatot feltételez.
- A döntéstámogató rendszer által közölt rendszerválaszokhoz tartozó attribútumok irány-preferenciának meghatározása objektivizálható, így lehetséges a döntéselőkészítési folyamatban minimalizálni az emberi tévesztéseket;

5. KÖVETKEZTETÉSEK ÉS JAVASLATOK

A dolgozatban leírtak alapján az alábbi pontok összegzik a kutatás és kísérletek által nyert gyakorlati alkalmazhatóság területeit:

- Audit, és auditot végző szakmai területek a dolgozatban közölt eljárásokat alkalmazni tudják, mely többek között hozzájárul a biztonságosabb informatikai üzemeltetéshez, jogszabályozói elvárások betartásához, valamint növelheti a szervezet befektetőinek bizalmát;
- A kontrollhiányosságok automatizált detektálása képes pl. az emberi hibából fakadó tévesztések felülvizsgálatára és felülbírálatára, objektívebb képet ad az ellenőrzés minőségéről befolyásmentesen (nem fél a következményektől);
- Az információbiztonsági környezet javítása révén a szervezetek kisebb kitettséget mutathatnak külső támadókkal, továbbá a munkatársak szándékos károkozásával szemben, mely növelheti a vevők a szervezet termékei/szolgáltatásai iránti bizalmát, tehát csökkenti a reputációs és kártérítési kockázatokat;
- Az audit megállapítások pontosításának révén leleplezhetővé válhatnak a csalások, szűkkeresztmetszetek, mérsékelhető a vállalati adatvagyon bizalmasságának, integritásának és rendelkezésre állásának kompromittálódásából fakadó kockázatok;
- Az ismertetett gépi tanuló rendszerek genetikai potenciálját kereső algoritmus általánosítható, nem kizárólag kontextusfüggő módon működik, hanem egyéb üzleti és szakmai problémák gépi tanulással feldolgozott megoldáskeresésére is felhasználható;
- A kereső eljárás a tanulóhalmaz méretének/összetételének optimalizálását célozta meg, azonban belátható módon, adott célváltozó esetén képes a célváltozóhoz köthető független változók elmozdulásai alapján, illetve azok érték-irány levezetésével ideálisabb irány kijelölésére, így jobb célváltozó-értékek lépésről lépésre történő megtalálására, valós/sztochasztikusan működő input-output-rendszerek esetében;
- A modell-preferencia matematikai levezetése és az ideális modell keresése hasonlóan általánosítható, így lehetőség nyúlik kontextus független módon az algoritmust felhasználni ott, ahol eddig a hermeneutikai alrendszer az emberi intuícióra hagyatkozott a több értékelési réteg aggregálása érdekében – így az automatizmus egyszerre garantálja az optimalizálást és a Knuth-i elvek betartását;
- Az irány-preferenciák automatizálása által csökkenthető az emberi tévesztés kockázata;

A kutatás során számos olyan feladat és potenciális kutatási irány fogalmazódott meg bennem, melyet szerettem volna elvégezni és dokumentálni, azonban idő és tartalmi korlátok révén nem volt rá lehetőség. Az alábbi pontok egyfajta jövőkép jelleggel foglalják össze azon javasolt kutatási irányokat, mellyel a dolgozatban közölt eredmények javíthatók, valamint a dolgozat alapot szolgált további kutatási célkitűzések meghatározásához:

- A dolgozat a kontrollhiányosságok detektálását osztályozási problémaként azonosította, azonban regressziós algoritmusokkal lehetőség adódhat a kontrollokhoz tartozó megállapítások számának approximációjára is;
- A kontrollhiányosságok becslésére az üzleti probléma bizalmasságának sajátosságai miatt, nem állt rendelkezésre számos olyan attribútum, mely képes lenne pontosítani az előrejelzéseket. Amennyiben van rá lehetőség, javasolt a kísérlet megismétlése, ahol elérhető az egyes szervezetek árbevétel, információbiztonsági költségvetés, szervezeti hierarchia, méret stb. adatai, vélelmezhetően javítana a rendszerek ítélőképességén;
- A kutatásban nem került elkülönítésre az egyes audittesztekhez köthető tervezet, implementáció és működési hatékonysági szinten mért kontrollhiányosságok, ezért a döntéstámogató rendszer képes lehet ezen információk finomhangolásával ítélkezni, vajon kontrolhiányosság tapasztalható már a kontroll kialakítási, implementációs vagy működtetési fázisaiban;
- A kutatásban strukturált, numerikus értékekkel feltöltött adatbázis került feldolgozásra, azonban a szövegesen megfogalmazott riportok további információt is tartalmazhatnak, mely segítheti az előrejelzések pontosságát, ezért javasolt egy olyan kísérletet is elvégezni, amely ötvözi a természetes nyelvi feldolgozás előnyeit a dolgozatban ismertetett módszerekkel;
- A kutatás nem kísérelte meg a tökéletes konfiguráció megtalálását, nem volt célja tartalmi korlátok miatt. Javasolt a konfigurációk optimális beállítását követően kiértékelni az algoritmusokat. Továbbá, a konfigurációk előkelőbb megtalálásához a dolgozatban ismertetett kereső eljárás alkalmasnak bizonyul;
- A hasonlóságelemzés teljesítményét növelendő, javasolt az algoritmus fejlesztése univerzális approximátorok pl. neurális háló által. Az anti-diszkriminatív modellek efféle javítása alkalmassá teheti őket a normától történő eltérés pontosításában.

6. MELLÉKLETEK

6.1. Irodalomjegyzék

1. BARTA, G. (2018a): Predicting Human Resource Attrition with Artificial Neural Networks. 55-66. p. In: ALMÁDI B. – GARAI-FODOR M. – SZEMERE P. T. (Szerk.): *Business as usual: Comparative socio-economic studies*. Budapest: Vízkapu Kiadó, 127 p.

2. BARTA, G. (2018b): The Increasing Role of IT Auditors in Financial Audit: Risks and Intelligent Answers. In: *Business Management and Education*, 16 (1) 81-93. p.

3. BARTA, G. – GÖRCSI, G. (2021): Risk Management Considerations for Artificial Intelligence Business Applications. In: *International Journal of Economics and Business research*, 21 (1) 87-106. p.

4. BARTA, G. – PITLIK, L. (2018): Startup felvásárlások multikulturális hátterének elemzése, avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez. 15-37. p. In: FARKAS A. (Szerk.): *A gazdaság kulturális szerkezete*. Gödöllő: Szent István Egyetemi Kiadó, 240 p.

5. BÁNKUTI, GY. (2010): About the method of component-based object comparison for objectivity. In: INTERNATION CONGRESS OF MATHEMATICIANS. (2010)(Hindustan). International Congress of Mathematicians: Abstracts, short communications, posters. Hindustan, Book Agency. p. 593-594.

6. BODA, M. A. (2019): Üzleti szimulációk és tanuló-rendszerek döntéshozatali mechanizmusai. Doktori disszertáció. Gödöllő: Szent István Egyetem, Gazdálkodás és Szervezéstudományok Doktori Iskola. 214 p.

7. FRIEDMAN, J. H. (2001): Greedy function approximation: A gradient boosting machine. In: *The Annals of Statistics*, 29 (5) 1189-1232. p.

8. GREENHOUSE L. (2005): Justices Unanimously Overturn Conviction of Arthur Andersen. https://www.nytimes.com/2005/05/31/business/justices-unanimously-overturn-conviction-of-arthur-andersen.html Keresőprogram: Google. Kulcsszavak: Arthur Andersen. Lekérdezés időpontja: 2020. 08. 08.

9. KÁSA, R. (2018): Neurális hálók alkalmazásának lehetőségei innovációs teljesítmény mérésére. In: *LOGISZTIKA - INFORMATIKA – MENEDZSMENT*, 3 (1) 60-73. p.

10. KNUTH, D. (1995): A=B. Előszó PETKOVSEK, M. – WILF, S. H. – ZEILBERGER, D. könyvében. Massachusetts: A K Peters/CRC Press. 224 p.

11. Magyar Nemzeti Bank (2015): A Jegybank azonnali hatállyal felfüggesztette a Buda-Cash Brókerház működési engedélyét és felügyeleti biztosokat rendelt ki. https://www.mnb.hu/sajtoszoba/sajtokozlemenyek/2015-evi-sajtokozlemenyek/a-jegybank-azonnali-hatallyal-felfuggesztette-a-buda-cash-broker
haz-mukodesi-engedelyet-es-felugyeleti-biztosokat-rendelt-ki
Keresőprogram: Google. Kulcsszavak: MNB sajtószoba, Buda-Cash Brókerház. Lekérdezés időpontja: 2020. 05. 02.

12. MASON, L. – BAXTER, J. – BARTLETT, P. – FREAN, M. (1999): Boosting algorithms as gradient descent. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (12.)(1999)(Cambridge). NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems. Cambridge, p. 512-518.

13. PITLIK, L. – RIKK, J. – GÁNGÓ, V. – TÓTH, CS. (2020): A távoktatás, mint kritikus oktatási üzem – IT-aspektusai, avagy felkészülés a duális képzésre. In: *Magyar Internetes Agrárinformatikai Újság*, 23 (266) 1-26. p.

14. PITLIK, L. – VARGA, Z. – BARTA, G. – LOSONCZI, GY. – PITLIK, L. (jun.) – PITLIK, M. – PITLIK, M. (2017): Magyar statisztikai régiók érintettségi sorrendje a szálláshelyek árbevételének havi adatai alapján eltérő módszertanokkal. In: VIDÉKFEJLESZTÉSI KONFERENCIA (1.)(2017)(Szarvas). Magyar vidék - perspektívák, megoldások a XXI. században. Szarvas, Szent István Egyetem Egyetemi Kiadó. p. 127-140.

15. POMPON, R. (2016): IT Security Risk Control Management. An Audit Preparation Plan. Washington: Apress. 311 p.

16. STEGMAN E. – GUEVARA, J. – MICHELOGIKANNAKIS, N. – FUTELA, S. – SHARMA, S. – KAUSHAL, S. (2019): IT Key Metrics Data 2020: Industry Measures — Executive Summary. https://www.gartner.com/document/3975995?ref=gfeed Keresőprogram: Google. Kulcsszavak: IT Key metrics. Lekérdezés időpontja: 2020. 08. 02.

17. VASVÁRI, GY. (2008): Vállalati (szervezeti) kockázatmenedzsment. Budapest: Információs Társadalomért Alapítvány. 183 p.

18. VINOGRADOV, SZ. (2020): A nemzeti versenyképesség puha tényezői, a társadalmi versenyképesség. 109-138. p. In: CSATH M. (Szerk.):

Versenyképesség: új elméleti és módszertani megközelítések. Budapest: Dialóg Campus Kiadó, 215 p.

Szabvány

19. ISO/IEC 27001 (2013): Information technology – Security techniques – Information security management systems – Requirements. International Standard. 23 p.

6.2. A doktori értekezés témakörében megjelent közlemények jegyzéke

Magyar nyelvű tudományos folyóiratcikkek

1. BARTA, G. (2020): Tanúsítványok értékelése ellátási láncok IT biztonsági megfelelésének vizsgálatára. In: *Logisztikai trendek és legjobb gyakorlatok*, 6 (1) 27-30. p.

2. GÖRCSI, G. – SZÉLES, ZS. – BARTA, G. (2019): Üzleti intelligencia megoldások alkalmazásának sikertényezői - A hazai szolgáltató szektor nagyvállalatainak körében végzett mélyinterjús kutatás. In: *Információs Társadalom: Társadalomtudományi Folyóirat*, 19 (2) 23-34. p.

Idegen nyelvű tudományos folyóiratcikkek

3. BARTA, G. (2018): Implementing and Evaluating Different Machine Learning Algorithms to Predict User Localization by the Strength of User Devices' Wi-Fi Signal. In: *SEFBIS Journal*, (12) 2-11. p.

4. BARTA, G. (2018): The Increasing Role of IT Auditors in Financial Audit: Risks and Intelligent Answers. In: *Business Management and Education*, 16 (1) 81-93. p.

5. BARTA, G. – GÖRCSI, G. (2021): Risk Management Considerations for Artificial Intelligence Business Applications. In: *International Journal of Economics and Business research*, 21 (1) 87-106. p.

Magyar nyelvű egyéb értékelhető cikkek

6. BARTA, G. (2017): A Mesterséges Intelligencia hatása az üzleti folyamatokra. In: *Magyar Internetes Agrárinformatikai Újság*, 20 (233) 1-15. p.

7. BARTA, G. – PITLIK, L. (2020): Hipotézis-tervezés PhD-disszertációkhoz - Konzisztens gépi tanuló modellezés beltéri felhasználói lokalizáció meghatározásának pontosítására. In: *Magyar Internetes Agrárinformatikai Újság*, 23 (263) 1-13. p.

8. BARTA, G. – PITLIK, L. (2018): A Titanic katasztrófa túlélőinek becslése döntési fa alapú gépi tanuló eljárással. In: *Magyar Internetes Agrárinformatikai Újság*, 21 (234) 1-24. p.

Magyar nyelvű, konferencia-kiadványokban megjelent publikációk

9. BARTA, G. – GÖRCSI, G. (2019): Csevegőrobotok a vállalati működésben. In: GAZDÁLKODÁS ÉS MENEDZSMENT TUDOMÁNYOS KONFERENCIA (3.)(2019)(Kecskemét). Versenyképesség és innováció. Kecskemét, Neumann János Egyetem Kertészeti és Vidékfejlesztési Kar. p. 912-917.

10. GÖRCSI, G. – BARTA, G. (2019): Az információs rendszer szerepe a döntési folyamatban. In: GAZDÁLKODÁS ÉS MENEDZSMENT TUDOMÁNYOS KONFERENCIA (3.)(2019)(Kecskemét). Versenyképesség és innováció. Kecskemét, Neumann János Egyetem Kertészeti és Vidékfejlesztési Kar. p. 252-256.

11. GÖRCSI, G. – BARTA, G. (2018): A CRM rendszerek szerepe a vevőkapcsolatok stratégiai kezelésében, vevőszegmentációs döntésekben. In: In: KÖZGAZDÁSZ DOKTORANDUSZOK ÉS KUTATÓK TÉLI KONFERENCIÁJA (4.)(2018)(Gödöllő). Közgazdász Doktoranduszok és Kutatók IV. Téli Konferenciája: Konferenciakötet. Budapest: Doktoranduszok Országos Szövetsége. p. 26-33.

12. PITLIK, L. – VARGA, Z. – BARTA, G. – LOSONCZI, GY. – PITLIK, L. (jun.) – PITLIK, M. – PITLIK, M. (2017): Magyar statisztikai régiók érintettségi sorrendje a szálláshelyek árbevételének havi adatai alapján eltérő módszertanokkal. In: VIDÉKFEJLESZTÉSI KONFERENCIA (1.)(2017)(Szarvas). Magyar vidék - perspektívák, megoldások a XXI. században. Szarvas, Szent István Egyetem Egyetemi Kiadó. p. 127-140.

Idegen nyelvű, konferencia-kiadványokban megjelent publikációk

13. BARTA, G. (2018): Artificial Intelligence: Blessing or Curse? In: BUSINESS AND MANAGEMENT SCIENCES: NEW CHALLANGES IN THEORY AND PRACTICE (2018)(Gödöllő). *Proceedings of the International Conference "Business and Management Sciences: New Challenges in Theory and Practice"*. Volume I. Gödöllő, p. 141-145.

14. BARTA, G. (2018): Challenges in the compliance with the General Data Protection Regulation: Anonymization of personal information and related information security concerns. In: INTERNATIONAL SCIENTIFIC CONFERENCES OF THE FACULTY OF MANAGEMENT (10.)(2018)(Krakkó). Knowledge – Economy – Society. Business, Finance and Technology as Protection and Support for Society: proceedings. Krakkó, Cracow University of Economics. p. 115-121.

15. BARTA, G. – GÖRCSI, G. (2020): Assessing and managing business risks for artificial intelligence based business process automation. In: SCIENTIFIC CONFERENCE ON CONTEMPORARY ISSUES IN BUSINESS, MANAGEMENT AND ECONOMIC ENGINEERING (6.)(2019)(Vilnius). 6th Internation Scientific Conference Contemporary Issues in Business, Management and Economics Engineering '2019: proceedings. Vilnius, Vilnius Gediminas Technical University Press. p. 823-832.

16. BARTA, G. – GÖRCSI, G. (2018): Artificial Intelligence and Audit: Why is it necessary to audit the intelligent decision support? In: KÖZGAZDÁSZ DOKTORANDUSZOK ÉS KUTATÓK TÉLI KONFERENCIÁJA (4.)(2018)(Gödöllő). Közgazdász Doktoranduszok és Kutatók IV. Téli Konferenciája: Konferenciakötet. Budapest: Doktoranduszok Országos Szövetsége. p. 225-234.

17. BARTA, G. – GÖRCSI, G. (2017): Intelligent Decision Making and Process Automation for Public Organizations. In: INTERNATIONAL SCIENTIFIC CORRESPONDENCE CONFERENCE (5.)(2017)(Nitra). Legal, economic, managerial and environmental aspects of performance competencies by local authorities. Nitra, Slovak University of Agriculture in Nitra. p. 30-37.

18. BARTA, G. – LUDVAI, N. – PUSKÁS, A. (2020): The analysis of data privacy incidents and sanctions in Europe after GDPR enforcement. In: INTERNATIONAL WINTER CONFERENCE OF ECONOMICS PHD STUDENTS AND RESEARCHERS (6.)(2020)(Gödöllő). VI.

International Winter Conference of Economics PhD Students and Researchers: Conference Proceedings. Budapest, Association of Hungarian PhD and DLA Students. p. 35-48.

Könyvrészletek

19. BARTA, G. (2018): Predicting Human Resource Attrition with Artificial Neural Networks. 55-66. p. In: ALMÁDI B. – GARAI-FODOR M. – SZEMERE P. T. (Szerk.): *Business as usual: Comparative socio-economic studies*. Budapest: Vízkapu Kiadó, 127 p.

20. BARTA, G. – PITLIK, L. (2018): Startup felvásárlások multikulturális hátterének elemzése, avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez. 15-37. p. In: FARKAS A. (Szerk.): *A gazdaság kulturális szerkezete*. Gödöllő: Szent István Egyetemi Kiadó, 240 p.

6.3. Egyéb témakörében megjelent közlemények jegyzéke

21. BARTA, G. (2015): Establishment of Enterprise Information Management Capability. In: INTERNATIONAL SCIENTIFIC CONFERENCES OF THE FACULTY OF MANAGEMENT (7.)(2015)(Krakkó). Knowledge-Economy-Society: Challenges for enterprises in knowledge-based economy. Krakkó, Cracow University of Economics. 29-36. p.

22. BARTA, G. – KARDOS, T. (2017): Integrált szervezeti rendszer bemutatása az Exxonmobil példáján keresztül. 49-56. p. In: SALAMONNÉ H. A. et al. (Szerk.): Ellátásilánc-menedzsment hallgatói esettanulmánykötet: Játszmák és stratégiai megoldások az ellátási lánc hálózatban. Budapest: Magyar Logisztikai Egyesület, 136 p.

23. BARTA, G. – ŁĘTEK, M. (2015): Equality in the Labor Market in Cracow. In: ACADEMIC INTERDISCIPLINARY CONFERENCE ON MODERN ECONOMICS AND SOCIAL SCIENCES (1.)(2015)(Tbilisi). International Academic Interdisciplinary Conference on Modern Economics and Social Sciences: proceedings. Tbilisi, Universal Publishing House. 281 p.

24. BARTA, G. – ŁĘTEK, M. (2015): Non-financial Performance Indicators as a New Approach to Measure the Value of Companies. In: INTERNATIONAL SCIENTIFIC CONFERENCES OF THE FACULTY OF MANAGEMENT (7.)(2015)(Krakkó). Knowledge-Economy-Society: Reorientation of paradigms and concepts of management in the contemporary economy. Krakkó, Cracow University of Economics. 247-254. p.

Megjelenés alatt álló közlemények jegyzéke

25. BARTA, G. (2021): Gépi tanuló rendszerek audit kihívásai. In: *Gazdaságinformatikai Kutatási és Oktatási Fórum*, (12).