Disszertációs Rész

A Szent István Egyetem Gazdálkodás és Szervezéstudományok Doktori Iskola 2018. évi komplex vizsgára

Barta Gergő

II. éves PhD hallgató

Neptun kód: N787OJ

Témavezető: Pitlik László

2018. június 11.

Tartalomjegyzék

1. Bevezetés

A 2016-os évben a doktori képzés központi reformon ment keresztül, melynek egyik eredményterméke a komplex vizsga bevezetése, melyet a hallgatóknak a második év végén kell teljesíteniük, bizonyítva megfelelő felkészültségüket és szakmai ismeretüket. A komplex vizsga szerves részét képezi a Disszertációs Rész megírása, mely kiemelt célja a saját kutatási témához kapcsolódó szakirodalmi-feldolgozás ismertetése, továbbá részletezi a kutatási módszereket, az alkalmazni kívánt modelleket és tervezett hipotéziseket. Jelen dokumentum célkitűzése a felvázolt pontok mélyreható bemutatása, vagyis annak demonstrálása, hogy a szakirodalom alapján a kutatási témám létező, releváns és innovatív eredmények produkálására képes. A Disszertációs Résszel végcélom, tehát az eddigi problémafeltáró kutatómunkám ismertetése, a tudományos megismerésre való törekvésem és kutatási szándékom, elhivatottságom igazolása.

A szakirodalom alapján olyan potenciális kutatási célok és hipotézisek realizálhatósága kerül bizonyításra, melyek kapcsán a tételes megoldásokról a Kutatás Módszertani Terv mutatja be az operatív részleteket, ami a Disszertációs Rész szerves részét képezi külön dokumentumban. Hipotézis-értékű egy-egy gondolat akkor, ha a szakirodalom kritikai értelmezése alapján logikai úton belátható, hogy az adott gondolat tudományos módszerekkel alátámasztott, objektív levezetése egy tervezett kutatási tevékenység keretei között új tudományos felismeréshez és innovációhoz vezet. Ennek feltétele a hipotézisek objektív jóságának (bizonyítottságának) mérni tudása. Egy-egy hipotézis kapcsán az alternatív megoldások objektív jósága is mérendő. Minden új megoldásnak az ismert régi megoldásokat mérhetően meg kell tudnia haladnia. A szakirodalomban leírtakat a szövegelemzés szigorú nyelvi, logikai szabályai szerint kell értelmezni, mely alapján a megfogalmazott hipotézisek levezethetők. Emellett egy-egy hipotézis kapcsán le kell tudni vezetni annak a mérésnek a mibenlétét, mely keretében megmérésre került a már létező szakirodalmi megoldások sorozata, mely meghaladásának esélyét a hipotézis bizonyíthatósága igazolja. A hipotézisek kapcsán új megoldásról kétféle esetben lehet beszélni: mennyiségi meghaladásról van szó, ha a szakirodalmi megoldások felülmúlása olyan eszközökkel történik, melyek ma is ismertek. Akkor lehet szó minőségi meghaladásról, amennyiben a mennyiségi meghaladás olyan módon történik, ami nem volt vélelmezhető a szakirodalomban tárgyalt kutatások trendjeként.

A kutatói munka eredményéül előálló értekezés jelen tervezett címe:

Mesterséges intelligenciák alkalmazása az informatikai rendszerek biztonsági auditjában

A téma sajátossága miatt a kutatómunka látszólag két önálló szakterületre összpontosít. Az első szakterület az információbiztonsági kérdésekre helyezi a hangsúlyt. A biztonsági incidensek anomáliaként értelmezendők, melyek elsődleges forrása a mindenkor rendelkezésre álló információrendszerek naplóállományai, melyet auditálva az anomáliák felderíthetők. Információrendszereket szűkítendő, a kutatás a vállalatirányítási rendszerekben fellelhető üzleti felhasználók által elkövetett, biztonsági incidensre utaló magatartások észlelésére összpontosít. A második szakterület a mesterséges intelligencia rendszerek fejlesztése, mely a szakirodalom szerint minden kétséget kizáróan általában véve megfelelő technikai eszköztárat biztosít anomáliák detektálására. Kutatói szándékom olyan modellek megalkotása, mely ezidáig publikált kutatási eredmények és az általam elvégzett kísérletek alapján alkalmasnak tekinthetők az anomáliák észlelésére előre definiált jóság kritériumok és metrikák szakirodalmi eredményeinek mennyiségi meghaladásával. A már ismert modellezési logikák egymással való összehasonlításának eredményeként az ismert hibridizáción túl olyan egyéb, új modellalkotási megoldások tervezett előállítása a cél, melyek minőségi meghaladását eredményezik a szakirodalmi „*benchmark*”-oknak. Hogy mely szerzőkhöz és milyen időszakokhoz kötődnek a meghaladni tervezett szintek, az a további fejezetekben részletesen kifejtésre kerül.

A dokumentum szerkezetéről: A dokumentum a kutatási témám rövid felvázolásával folytatódik, melyben levezetem az általam észlelt gyakorlati problémákat és bemutatom a kutatási célkitűzéseket, összefoglalót adok, mely alapján a dokumentum további részei teljes értelmet nyernek. A fejezet végén ismertetem a felállított hipotéziseket. A kutatási téma bemutatását követően a szakirodalom jelen állításait és kutatási eredményeit prezentálom három különböző fejezetben, annak érdekében, hogy ezek alátámasszák a kutatási téma létjogosultságát, relevanciáját, innovációs potenciálját indirekt és direkt módon. Az első fejezet, az alapvető fogalmak egységes rendszerű bevezetéseként, az információ értékét, az információbiztonság, és audit kérdéseit tárgyalja három külön alfejezetben, melynek célja az információbiztonság szükségességének levezetése, továbbá, azon szervezeti kihívásokat ismerteti a biztonságtechnika területén, melyek vállalati szinten várnak megoldásra az információs vagyon védelmének érdekében. A második fejezet a mesterséges intelligencia fogalomkörét definiálja, az intelligens rendszerek fejlődését tanulmányozza, az intelligens algoritmusok fejlesztéseire irányuló kutatói munkákat összegzi a fogalmi rend további stabilizálásának igényével. A harmadik szakirodalmi fejezet a felállított hipotézisek racionalitását hivatott alátámasztani, mely keretében elemzésre kerül a mesterséges

intelligencián és gépi tanuláson alapuló információbiztonsági anomáliák kutatási eredményei. A dokumentumot összefoglalással zárom.

2. A kutatási probléma összefoglalása

Növekvő igény mutatkozik az üzleti folyamatok automatizálására (Gartner.com, 2017/a), azonban az automatizálást támogató integrált üzleti informatikai megoldások fejlesztése és az eddigi elavult rendszerek korszerűsítése kihívások elé állítják a technikai szakembereket, vezetőket, felhasználókat és olyan kockázatokat hordoznak, mely szükségessé teszi az információ kezelésével, biztonságtechnikájával és ellenőrzésével foglalkozó funkcionális egységek megjelenését megteremtve az informatikai erőforrásokra irányuló kockázatmenedzsment folyamatot, melynek végső kimenete egy stratégiai szintű döntés a szervezetben felmerülő informatikai kockázatok csökkentésére (Vasvári, 2008). Az informatikai kockázatelemzés eredményére támaszkodó döntés lehet a fellépő kockázatok enyhítésére irányuló intézkedések megtétele (összhangban ISACA (2015/a) csoportosításával), azaz a megismert kockázatokat mitigáló kontrollokkal történő mérséklése, úgy mint a megelőző óvintézkedések (pl. alkalmazások jelszavas védelme) korrektív- (pl. adatvesztés esetén adatvisszatöltés) vagy felderítő (pl. naplóállományok ellenőrzése) tevékenységek bevezetése. Az implementált informatikai kontrollok megléte még nem szavatolja a hibamentes működést, a vezetésnek rendszeresen meg kell győződnie arról, hogy a mitigáló intézkedések hatékonyan működnek, nem lehetséges azok megkerülése, illetve kijátszása. Erre egy gyakorlati példa az alkalmazások jogosultságkezelését megkerülendő hibajavításra (éles programfejlesztésre) szolgáló jogokkal történő visszaélés, melyekkel az alkalmazás szintű biztonsági logikát felül lehet írni, mely, ha a naplófájlok írási jogkörrel való hozzáférésével párosul, akkor egy rosszindulatú felhasználó vagy támadó képes csalások kivitelezésére a rendszerben, úgy, hogy saját elektronikus lábnyomait törölni képes.

A belső informatikai kontrollok hatékonyságának feltárására szolgáló funkcionális terület megnevezése az informatikai audit csoport, melyet olyan üzleti és informatikai szakmai tudással rendelkező szakemberek alkotnak, melyek elsődleges célja a belső informatikai kontrollkörnyezet ellenőrzése és folyamatos tesztelése, annak meggyőződésére, hogy az informatikai folyamatok teljes mértékben a szervezetek üzleti célkitűzéseit kövessék a lehető legnagyobb információbiztonsági szint mellett. Az informatikai auditorok egyik legértékesebb forrása a gyanús tevékenységek felderítésére az informatikai rendszerek által generált naplóállományok. A naplóállományok ellenőrzése, mint felderítő informatikai kontroll implementálása segítségül szolgálhat, amennyiben teljessége, pontossága és megmásíthatatlansága biztosított, az informatikai rendszerekben történő jogosulatlan hozzáférések és csalások, azaz az anomáliák felderítéséhez, mivel a naplózó rendszerek

konfigurálhatók a felhasználók minden egyes hozzáféréseinek, adatmódosításainak, adattörléseinek, tevékenységeinek rögzítésére, értékes információs adatvagyont generálva az auditorok számára. A naplófájlok ellenőrzése, tehát, egy monitorozó tevékenység az anomáliák detektálására (ISACA, 2015/b).

Az anomália észlelés egyik legnagyobb kihívása, hogy a rendelkezésre álló naplóállományok folyamatos ellenőrzésének folyamata és monitorozása manuálisan nehezen kivitelezhető a naplóállományok óriási méretéből adódóan, amit az operációsrendszerek, adatbázisok és alkalmazások együttesen produkálnak, illetve ezek összehasonlítása és összefüggések keresése a jelentett informatikai incidensekkel. A másik probléma a manuális feldolgozással, hogy akadályokba ütközik az átlagostól eltérő tranzakciók kiszúrása pl. hétvégi könyvelések, mivel az elemzőnek ismernie kell az összes lehetséges potenciális incidensre okot adó kiváltó jeleket, mely eredhet üzleti és technikai oldalról egyaránt. A felhasználók kooperációiból adódó gyanús tevékenységek kiszűrése is akadály lehet, mivel a teljes naplóállományt ismerni kell, hogy az időben eltérő tevékenységek együttesen is kiértékelhető váljanak: pl. a szállítói törzsadatokért felelős személy megváltoztatja egy rekord bankszámlaszámát, egy másik felelős személy pedig arra utal. Harmadik kihívás az informatikai rendszerek és infrastruktúra folyamatos változásában testesül meg. Az új rendszerek és fejlesztések megjelenése vagy eddig nem ismert, vagy elvileg ismert, de nem várt sérülékenységet hordozhatnak magukban, így növelve a rendszerekben elkövethető csalások kockázatát: pl. a nem teljes körűen tesztelt új fejlesztések kártékony kódot tartalmazhatnak, vagy éppen kiskaput nyithatnak a rendszert támadók felé. Emelett, a külső behatolások fenyegetettsége is egy potenciális veszélyforrás, melyek detektálására, bár léteznek megoldások, (pl. behatolás-megelőző rendszerek) azonban a felhasználók bejelentkezési adatainak hanyag kezelése pl. jelszavak megosztása, jelszavak kiragasztása a monitorra, könnyű jelszavak megadása, általánosan ismert adminisztratív felhasználói nevek használata stb. újabb sebezhetőséget jelentenek az informatikai rendszerekben. Az ilyen adatok megszerzése harmadik fél által magában rejti a jogosulatlan hozzáférés kockázatát, amelyet a behatolás-védelmi eszközök nem feltétlen támadásként, hanem hiteles autentikációként értelmezhetnek.

Egyértelműen egy olyan automatizált tanuló megoldás szükséges, mely képes a csalásra utaló jelek felkutatására, feldolgozására és értelmezésére, mely tanulási mechanizmusai révén képes a gyanús bejelentkezések és tevékenységek monitorozásában, ezzel hatékony módon valós értéket képezve az informatikai auditban. Továbbá, maga a rendszer ellen irányuló belső és külső támadásokkal szemben is ellenállónak kell lennie, mivel a támadók első lépésben a

szervezet védelmi mechanizmusainak kiiktatását is megkísérelhetik, hogy az ne legyen képes riasztások formájában értesíteni az auditorokat az anomália bekövetkeztéről. 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. A felvetett probléma, összegezve, az informatikai rendszerekben elkövetett csalások és anomáliák, mint informatikai kockázat fogalmának döntési helyzet-specifikus kialakítása mesterséges intelligenciák alkalmazásával. Mivel a mesterséges intelligencia eszköztára számos modellt kínál különböző erősségekkel és gyengeségekkel, mely modellek a későbbiekben a dokumentumban kifejtésre kerülnek, ezért a különböző megközelítéseket alkalmazó módszerek hibridizált megoldására van szükség a gyengeségek kiküszöbölésére a magas teljesítmény elérése érdekében.

Összefoglalóan, a kutatási célkitűzésem olyan modellek megalkotása, melyek az informatikai rendszerekben felmerülő csalások felderítésére irányulnak a mesterséges intelligencia alapú fogalomalkotás lehetőségeit felhasználva a modellalkotásban és a modelljóság mérésben egyaránt, ahol a mindenkor rendelkezésre álló naplóállományokból kell a mesterséges kockázatfogalom megalkotása keretében ennek mértékét, normáját algoritmikusan levezetni. A cél az eddigi kutatási eredmények visszamérése és azok meghaladása. A fent említett gondolatok és a szakirodalom alapos áttanulmányozása alapján a következő hipotéziseket állítom fel:

H1: Különböző logikát/megközelítést alkalmazó mesterséges intelligencia módszertanok magasabb fokú hibridizációjával magasabb teljesítmény érhető el az incidensek/anomáliák detektálására vonatkozóan.

H2: A kutatásban elkészítendő egyes hibrid modellek képesek az előre definiált normális tevékenységek között meghúzódó összefüggések matematikai leképezésére, s ez által az anomáliák feltárására, kísérletekkel alátámasztott optimális parametrizálásával azok megtanulására a túlilleszkedést minimalizálva.

H3: A kutatásban elkészítendő hibrid modellek a többváltozós matematikai statisztikai módszerekkel szemben meghaladják az előzetesen definiált jóság metrikák aggregált szintjét.

H4: A kutatásban elkészítendő hibrid modellekben az input jelként felhasznált adatokat lehetséges olyan szinten transzformálni és a modellekbe lehetséges olyan védelmi mechanizmusokat beépíteni, hogy azok ellenállók legyenek egyes külső szándékos

manipulációkkal szemben, melyek célja az anomália-detektáló rendszer félrevezetése, vagyis az anomáliák normális magatartásként történő feltüntetése.

3. Szakirodalmi kutatási rész

3.1. Információbiztonság és audit

3.1.1. Az információ fogalmának megközelítései

Az információt egyes szakirodalmak az 5. termelési tényezőként tartják számon, többek között Farkasné és Molnár (2017), Kiss (2016), Margitay-Becht et al. (é. n.) és Boda et al. (2009), akik tanulmányukban az információt a tudással is azonosítják. Farkasné és Molnár (2007) egy sajátos termelési tényezőként kezeli az információt (a tudással együtt), melyet a tudományos kutatás termel, és hozzájárul a teljes társadalmi átalakuláshoz, mivel az áthatja a gazdasági folyamatok egészét. Hasonló véleményen van Mátyus (2015), aki hozzáteszi, hogy a technológiai fejlődések révén a gyors információáramlás és feldolgozás, és az információból kinyerhető tudás létrehozta az információs társadalmat.

Harkevics (1960) az információ értékéről beszél, csak akkor válik az információ értékké, amennyiben hozzájárul egy kitűzött cél megvalósításához. Az információ közvetítése, értelmezése és feldolgozása nagyban hozzájárul a vállalati döntéstámogatáshoz, annak eszköze, nélkülözhetetlen kiszolgáló eleme. Például, egy új termék bevezetéséhez piaci információra van szükség a versenytársakról, fogyasztókról, szabályozói környezetről, ajánlásokról, technológiáról stb. Ágoston és Szluka (1989) az információt a „*hatalom*” szóval kapcsolják össze, implikálva, hogy az információ előnyt jelent annak birtokosa számára. Capurro (1991), továbbá kiemeli, hogy az információ elveszíti a kapcsolatot az emberi világgal, arra utalva, hogy nem kizárólag ember és ember között létezik információáramlás, hanem ember és gép, illetve gép és gép között is zajlik információ csere, azaz „*informálódás*”. Fülöp (1996) az ipari termelés ki nem merülő tartalékaként kezeli. Munk (2007) az információt a valóság visszatükröződéseként értelmezi.

A szakirodalomban az előző bekezdésben megfogalmazottakat kiértékelve, véleményem szerint az információ a termelési tényezők egyik meghatározó eleme. A döntéstámogatáson keresztül beépül a szervezeti stratégia készítésébe, annak kivitelezésébe és a teljes stratégiai irányításba, ezért is kezelhetjük termelőeszközként, mert az információ feldolgozása hozzájárul a stratégiai döntéshozatalon keresztül a javak előállításához. Az információ fogalmát, azonban, a különböző tudományos szakterületek eltérően definiálják. Alapvetően az információelmélet, mint matematikai és hírközlési tudományterület, az információ feldolgozásával és annak értelmezésével foglalkozik, és mint annak egyik jeles képviselője, és a tudományterület atyja

Claude Shannon (1948) az információt az adó és vevő közötti valamilyen üzenet közléseként, eseményként nevezte meg. Forgó (2011) ezt azzal egészíti ki, hogy az információelmélet szerint, az információ a kommunikációs folyamat mennyiségi mértékegysége. Komenczi (2011) alapján, az információ egy az agyunkban meglévő elképzelés, ami magába foglalja a tudást, belátást, felismerést. Worth és Gross (1977) szociológiai szempontból közelítette meg az információ fogalmát, melyet társadalmi folyamatként írtak le, a jelek közleményként való észleléseként, melyből jelentésre lehet következtetni. A modernebb definíciók az információ és „*Big Data*[1]" fogalmát gyakran összekötik, melyben az információt felruházzák olyan leíró elemekkel, mint annak mennyisége, változatossága, redundanciája, mivel a jelen technológia és eszközök lehetőséget nyújtanak az adatok és információ folyamatos tárolására, gyűjtésére és másolására (Lin et al., 2016). A világháló mindennapi alkalmazásával, annak jelentős elterjedésével az információ megosztása másodpercek alatt zajlik. 2017 végével kb. 4.1 milliárd internet felhasználó volt jelen, mely 1052%-kal több, mint a 2000-ben mért adat (Internetworldstats.com, 2017). A hardveres tároló kapacitások árának folyamatos csökkenése miatt, egyre több információt tudunk tárolni, 1 gigabite adat 1980-ban átlagosan 200 ezer dollárba, 2000-ben 10 dollárba, 2009-ben csupán 10 centbe, míg 2017 végén fél centbe került (Mkomo.com, 2009)(Backblaze.com, 2017). A hardveres erőforrások efféle dramatikus csökkenése azt eredményezi, hogy a szervezetek képesek a keletkező információ permanens tárolására hosszú évekre visszamenőleg is.

3.1.2. Információrendszerek megjelenése

A szakirodalmi kutatásom és az ez idáig felsorakoztatott megfogalmazások alapján, álláspontom szerint az információt erőforrásként célszerű értelmezni a gazdálkodás és szervezéstudományok területén, mely gyökereiben hozzájárul hatékony feldolgozása és értelmezése által a szervezeti folyamatok eredményes működéséhez és a döntéshozatal elősegítéséhez. Termelési tényezőként való értelmezésében egyedülálló tulajdonságokkal rendelkezik. A többi termelési tényezővel ellentétben, gyakorlatilag korlátlanul sokszorosítható, az információ nem véges (nem fogy el), hasznosítása során annak állománya nem csökken, továbbá, lehetséges pillanatok alatt tovább terjeszteni, korunk technológiai szintje lehetővé teszi annak egyszerű tárolását és mindenkori rendelkezésre állását. Az

[1] A Big Data azt a jelen technológiai környezetet jelenti, melyben az adatállományok olyannyira komplexek, változatosak és nagymennyiségűek, hogy azt a hagyományos adatbázis-kezelő rendszerekkel nem lehetséges feldolgozni (Warden, 2011).

információ folyamatos raktározásából, azonban, csak akkor teremthető érték, ha abból a szervezet képes tudásanyagot létrehozni és azt hatékonyan felhasználni (Ward, 1998). A megnövekedett információmennyiség kezelésének szükségessége életre hozta az információmenedzsment tudományát, melynek központi eleme az információrendszer. Az információrendszerek kutatása az 50-esek évekre nyúlik vissza, melynek előfutárjai a termeléstámogató-rendszerek voltak, majd a 70-es években megjelentek a vezetőket támogató információrendszerek, a 90-es évektől, pedig már az alapvető üzleti folyamatokat is információrendszerek szolgálták ki (Ward, 1998). Jelenleg az információrendszerek képesek a teljes ellátási-lánc menedzsment funkcióit ellátni, integrált vállalatirányítási rendszerek által a globálisan működő szervezetek teljes üzleti folyamatát központilag lefedni, beleértve a különböző elektronikus kereskedelmi és elektronikus üzletviteli funkciókat, hatékony kommunikációs csatornát kiépítve a szállítók, vevők és az állami szervek felé is. B. Langefors (1973) szakirodalmi kutatásom alapján, a legelső volt, aki az információrendszer fogalmát definiálta, korai művében a döntéstámogatás szerepét emelte ki az információrendszereknek, mely publikálás óta az alapos fejlődésen ment keresztül. K. C. Laudon és J. P. Laudon (1991) már kiemelte az adatokra vonatkozó keresés, tárolás, továbbítás funkcióját, mely technikai oldalról hangsúlyozta az információrendszert. K. C. Laudon és J. P. Laudon (2015) két és fél évtized elteltével megfogalmaz egy üzleti definíciót is, melyben szervezeti és menedzselési megoldásnak nevezi az információrendszereket. Szepesné (2011, 3.old.) úgy fogalmaz, hogy az információrendszer „fő célja az információ-előállítása, vagyis olyan célorientált üzenetek létrehozása, amelyek a címzett számára újdonságot jelentenek, bizonytalanságot szüntetnek meg és feladataik, döntéseik teljesítésében segítséget nyújtanak.” Az információrendszer, tehát képes az üzleti adatok összegyűjtésére, továbbítására és azok teljes életciklus menedzselésére.

3.1.3. Az információ biztonsági kérdései

Mivel az információrendszerek gyakorta bizalmas, üzletileg kritikus információt kezelnek, ezért az üzleti rendszer sérthetetlenségének és bizalmasságának biztosítása kiemelt szerepet ölel fel a szervezetek életében. Erre világítanak rá jelen korunk szabványai is, mint az ISO 27001-

es szabvány család[2], CobIT 5[3], NIST 800-53[4], illetve a nemrég megjelent ISF kiadvány[5], a „*The Standard of Good Practice for Information Security 2018*". A szabályozói környezet Magyarországon és az Európai Unióban is magas hangsúlyt fektet az információbiztonsági előírások megfogalmazására és azok betartatására. Országunkban (és tulajdonképp hasonlóan az egész világon) kiemelendő, a pénzügyi szektor a legjobban szabályozott ágazat, melyet hazánkban a Magyar Nemzeti Bank kötelező érvénnyel rendszeresen ellenőriz. Információ-biztonságot érintő ajánlatok és rendeletek hazánkban többek között, az *MNB 7/2017. (VII.5.) számú ajánlása az informatikai rendszer védelméről, MNB 2/2017. (I.12.) számú ajánlása a közösségi és publikus felhőszolgáltatások igénybevételéről, MNB 19/2017. (VII.19) az elektronikus hírközlő eszközök auditjáról, 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról* stb. Illetve, az Európai Parlament és Tanács 2018. május 25-től hatályba léptette az *EU 2016/679 Általános Adatvédelmi Rendeletét,* melynek hatóköre kiterjed az Európai Unióban és az Európai Gazdasági Térségben működő minden egyes szervezetre megcélozva a személyes adatokat feldolgozó rendszerek biztonsági előírását és biztonságos üzemeltetését.

A biztonság fenntartásáért az információ-biztonsági szervezeti egység felelős a szervezeti felépítésben. Az információbiztonsági szervezet feladata az információrendszerek, a hálózat és egyéb fizikai és logikai eszközök védelmi intézkedéseinek meghatározása, betartatása és folyamatos nyomon követése. Az információbiztonsági osztálynak függetlenül külön kell működnie más szervezeti funkcionális egységektől, még az informatikai üzemeltetési osztálytól is, kikényszerítve a négyszem-elvet. Információbiztonságot érintő incidensek alapvetően három különböző forrásból érkezhetnek (Information Security Forum, 2014):

- külső, a szervezettől kívülálló csoportoktól, személyektől, melyek lehetnek kiberbűnözők, versenytárak ipari kémei, hobby hackerek, szélsőséges esetben terroristák;

[2] ISO/IEC 27001. (2013): Information technology – Security techniques – Information security management systems – Requirements. International Standard, 23 p.
[3] ISACA (2012): COBIT 5: A Business Framework for the Governance and Management of Enterprise IT. USA, ISACA, 94 p.
[4] NIST (2013): Secuirty and Privacy Controls for Federal Information Systems and Organizations. USA, NIST Special Publication 800-53, 462 p.
[5] Information Security Forum (2018): The Standard of Good Practice for Information Security 2018. Information Security Forum Limited, 338 p.

- belső, a szervezet belső dolgozói, akik szándékosan vagy véletlen bizalmas információt osztanak meg a külvilággal;

- természeti események, melyek az információ elérhetőségét és rendelkezésre állását veszélyeztetik pl. árvíz, mely megkárosítja a szerverparkot.

Az információbiztonsági szintet informatikai és szervezeti kontrollok implementálásával lehet növelni, melyek szükségességét, erősségét és rendszerességét a *szervezeti informatikai kockázatelemzés* által lehet érvényre juttatni. Az informatikai kockázatok becslésére a rendelkezésre álló naplóállományok megfelelő forrásként szolgálhatnak, mivel általuk elemezhetők a bekövetkezett anomáliák pl. biztonsági incidensek száma az adott pénzügyi évben. Tehát, az informatikai kockázatelemzés segítséget nyújt azon területek feltárásában, melyek további biztonsági óvintézkedéseket igényelnek. Az ISO (International Organization for Standardization – Nemzetközi Szabványügyi Szervezet) és az IEC (International Electrotechnial Commission – Nemzetközi Elektrotechnikai Bizottság) a 27000-es információs biztonságra vonatkozó szabványcsaládjában a következőként fogalmaz a kockázatról: „*A bizonytalanság hatása a célkitűzéseken*" (ISO/IEC 27000:2013(E), 2014, 8.old.). Ezt a hatást a szabvány pozitívként és negatívként is értelmezi, míg más szakirodalom (Vasvári, 2008, 15.old) a kockázatot kizárólag negatív hatású eseményként írja le, mely „*egy veszélyforrás képezte fenyegetés bekövetkezési lehetősége, amely kárkövetkezménnyel jár, és így kedvezőtlen hatást fejt ki az üzleti célokra.*" Az informatikai kockázat vagy kockázati tényező az ISACA Risk IT framework meghatározás alapján (ISACA, 2009, 11. old.) „*egy üzleti kockázat – azaz, az üzleti kockázat, mely szervezeti kereten belül kapcsolódik az információtechnológia felhasználásához, tulajdonlásához, működéséhez, bevonásához és befolyásolásához*." Értelmezésem szerint az informatikai kockázatkezelés legvégső célja, hogy felszínre kerüljenek a szervezetben azok a hiányosságok, melyek az információ, mint szervezeti vagyon és termelési tényező, egyes attribútumainak, mint pl. teljesség, pontosság, megbízhatóság, elérhetőség, bizalmasság, hitelesség stb. elvesztéséhez és sérüléséhez vezetnek, tehát azon pontok és üzleti folyamatok, ahol a legkevesebb erőfeszítés összpontosul az információ biztonságos tárolásán, feldolgozásán és továbbításán. Az 1985-ben alapított COSO (Committee of Sponsoring Organizations of the Treadway Comission – A „Treadway Comission" Támogató Szervezeteinek Bizottsága) az egyik legnagyobb szervezetnek számít ma, mely szakmai publikációkkal és egy széles körben elismert keretrendszerrel támogatja a szervezeti kockázatmenedzsmentet, mely átfogja az információ, mint vállalati érték és termelési tényezők

kockázatkezelését (Coso.org, é. n.). Az említett szervezet célja, hogy egy megalkotott modell biztosítása által (COSO modell) kockázatorientált megközelítésben bemutassa a szervezeti folyamatokat a vállalatok belső kontrollrendszere, irányítási rendszere, és a funkcionális területek tevékenysége alapján. Egy másik nagy amerikai szervezet, az ISACA (Information Systems Audit and Control Association – Információrendszer Audit és Kontroll Egyesület) hasonlóan több keretrendszert dolgozott ki. Fontos kiemelni a CobIT-ot (Control Objectives for Information and Related Technology – Információra és a Kapcsolatos Technológiára Vonatkozó Kontroll Célkitűzések), mely informatikai kontrollcélkitűzéseket határoz meg az üzleti kockázatok csökkentésére. 1996-ban publikálta első verzióját, melynek legfrissebb változata a CobIT 5 nevet viseli, és 2012-ben került kiadásra (ISACA, 2012). A másik ismert keretrendszer a Risk IT framework, mely a nagyvállalatoknak nyújt segítséget az informatikai kockázatok megfelelő kezeléséhez (ISACA, 2009). A COSO és Deloitte által kiadott *Risk Assessment in practice* publikációban (Curtis – Carey, 2012), a kockázatértékelés tevékenysége négy eljárásra bontható. Ehhez a folyamathoz inputként szolgál a kockázatok azonosítása, tehát a negatív események meghatározása. A kockázatértékelés első folyamata az értékelési kritériumok meghatározása, amely magában foglalhatja a kockázatok bekövetkezési valószínűségének becslését, a nem várt hatásokat, a kockázat előfordulásának gyakoriságát. A következő lépés a kockázatok kiértékelése, melyben az elvégzendő feladatok az értékek rendelése a kockázati tényezőkhöz a kritériumban definiáltaknak megfelelően, majd kvalitatív vagy kvantitatív módszerekkel a kockázatok rangsorának felállítása. A COSO külön tevékenységként kiemeli, hogy nem csupán az egyes kockázatok egyéni hatásait szükséges figyelembe venni, hanem a kockázatok közötti interakciót is, melyek hatásai együttesen még nagyobb kárt okozhatnak. Egy kisméretű vállalat kevesebb figyelmet fordíthat az akár pénzügyi kimutatásokat is érintő ütemezett és automatizált feladatok futtatásának szabályozására, ami ha párosul az összeegyezhetetlen szerepkörök nem megfelelő szétválasztásával, akkor bekövetkezhet azaz esemény, hogy egy egyszerű dialógus felhasználó változtatást indukál, és más időpontra helyezi a kritikusnak vélt feladatok elvégzését pl. a heti teljes mentéseket a következő hétre ütemezi, azaz a kutatás relevanciájával összefügg, normális tevékenységek sorozata által anomáliát, azaz információbiztonsági incidenst idéz elő. Üzemzavar esetén még nagyobb lehet a gond, mivel, ha nem sikerül visszaállítani az éppen aktuális adatokat, a szervezet kétheti adatállománya odaveszett, ezzel együtt az egész munkát ismételten el kell végeznie a személyzetnek. A kockázatértékelési folyamat utolsó pontja a kockázati rangsor felállítása, amely gyakran egy kockázati térkép elkészítésével ér véget. Az ISACA (2015/a) észlelési kockázatnak nevezi az informatikai auditban azt a kitettséget, amikor a

kontrolltesztelési módszertanok nem megfelelőek, azaz nem képesek egy adott informatikai kockázatot mitigáló kontrollok hatékonyságának felmérésére, ezáltal torz képet adva annak helyes működéséről.

Magyarországon és a történelemben is számos botrány kapcsolatba hozható az informatikai ellenőrzések hanyagságával. 2015. február 24-én a Magyar Nemzeti Bank felfüggesztette a Buda-Cash Bróker Zrt. működését, 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. Lukács János, a Magyar Könyvvizsgálói Kamara elnöke, a 2015. április 3-án az ATV-nek adott interjújában kiemelte a részletesebb informatikai ellenőrzések fontosságát és igényét a jövőbeli botrányok elkerülése végett, és könyvvizsgálók szerepét az ehhez hasonló események megelőzése érdekében. A Budacash és Questor cégek esetében az informatikai rendszerekben tárolt adatok kerültek meghamisításra, melynek révén történtek a csalások. Látszólag sem a belső, sem a külső fél nem használt alaposan megtervezett audit eljárásokat, mivel adott volt a lehetőség a fiktív kötvények kibocsájtására, amit a szervezetek információrendszerei valósnak kezeltek. Ez az eset, mind az alaposabb informatikai ellenőrzés bevonását, mind újabb jogszabályi megfelelőség szerepét újraértékelheti.

3.1.4. Az alfejezet összefoglalása

A saját kutatási téma vonatkozásában, az információról és információbiztonságról folytatott szakirodalomkutatás az alábbi pontokban erősítette meg kutatói munkám létjogosultságát:

- Az információ, az információrendszerekben tárolt adatok kritikus értékkel bírnak a szervezetek a számára, így azokat védeni kell a külső behatolókkal, belső munkatársakkal és a természeti katasztrófákkal szemben.
- A szervezeti információbiztonsági szint implementált kontrollokkal fokozható, azonban a kontrollok működési hatékonyságát folyamatosan nyomon kell követni, mivel lehetséges azok kikerülése. A fejlesztésre szoruló kontrollok és azok szintjét az információbiztonságra irányuló informatikai kockázatelemzés, mint üzleti támogató folyamat képes meghatározni.
- A belső informatikai audit a felelős szervezeti egység, mely feladata a kontrollok folyamatos ellenőrzése.

- Az audit egyik legértékesebb forrása a rendelkezésre álló naplóállományok, melyek, feltéve, hogy sértetlenségük biztosított, teljes képet tudnak adni az információrendszerekben lezajlott eseményekről, általuk detektálhatók az információbiztonságot érintő incidensek, mint pl. a jogosulatlan adathozzáférés, adatmódosítás, adatszivárogtatás, belső csalások és anomáliák észlelése. Az informatikai kockázatelemzés feladata, hogy a rendelkezésre álló naplóállományok felhasználásával meghatározza a szervezet adott időpontra vonatkozó kockázati kitettségét.
- Véleményem szerint, a rendszeres információrendszerek által produkált naplófájlok elemzésével az információbiztonsági incidensek jelentős része megelőzhető, a bekövetkezett negatív események (incidensek/anomáliák), pedig utólagosan a naplófájlok által kiértékelhetők, azok újbóli előfordulásai csökkenthetők, a naplófájlok alapján létjogosultságot nyert azonnal óvintézkedések megtételével.

3.2. A mesterséges intelligencia fogalomköre és fejlesztése

3.2.1. A mesterséges intelligencia definícióinak rendszerezése

A mesterséges intelligencia fogalma és a tudományterületet átfogó definíciók, algoritmusok és alkalmazási területek nem számítanak újkeletűnek a kutatók számára. Igaz, hogy a mesteréges intelligencia 2014 óta fokozott népszerűségnek örvend (mely a kutatási témában megnövekedett publikációk számából is igazolható), többek között köszönhető a számítógépek grafikus kártyáinak megnövekedett kapacitásának és feldolgozó képességének (Liu et al., 2017). A fogalom, mint „*mesterséges intelligencia*” megalkotása John McCarthynak köszönhető, aki 1956 nyarán két hónapos munkatalálkozót szervezett a számításelmélet és atomelmélet amerikai kutatóinak számára Dartmouthban. A munkatalálkozó előkészítő anyagában kiemelte a mesterséges intelligencia önállő tudományterületté válásának szükségességét (McCarthy et al., 1955). Ugyanakkor, a mesterséges intelligencia, első igazán jelentős elméletének megszületése Waren Sturgis McCulloch és Walter Pitts nevéhez fűződik, akik 1943-ban publikálták a perceptron modellt. Az agysejtek működését alapul véve, egy egyszerűsített bináris klasszifikációs eljárást ajánlottak, mely az agy tanulóképességét hivatott lemásolni (McCulloch és Pitts, 1943). Ez volt a legelső modell, mely a mai nap is ismert neurális háló alapjait képezi, és innen indult el annak széleskörű kutatása. Kiemelendő, hogy ezen az elven működik a jelenleg legmodernebb technológia a „*deep learning*”, melynek alkalmazási

területei kiterjednek az önvezető autók, virtuális asszisztensek, virtuális valóságok, automatikus játékgépek és ajánló rendszerek fejlesztésére is. Diamant (2016) álláspontja, hogy jelen korunk mesterséges intelligencia kutatásának fejlődése nagyrészben a deep learningnek köszönhető.

A mesterséges intelligenciát az évek során számos kutató definiálta. Russel és Norvig (2009) ezen definíciókat rendszerezte, és arra jutottak, hogy a mesterséges intelligencia definícióit két dimenzió mentén értelmezik a terület kutatói. A megfogalmazások csoportosítását az 1. táblázat szemlélteti.

1. táblázat: A mesterséges intelligencia definícióinak csoportosítása (Russel - Norvig, 2009)

Gondolkodó rendszerek	Emberi módon gondolkodó rendszerek	Racionálisan gondolkodó rendszerek
Cselekvő rendszerek	Emberi módon cselekvő rendszerek	Racionálisan cselekvő rendszerek
	Emberi teljesítmény	Racionalitás

A táblázat felső része a gondolkodásra, az észszerűségre helyezi a hangsúlyt, míg az alsó része a viselkedésre, a konkrét cselekvésre. A táblázat bal oszlopa az emberi képességet és teljesítményt helyezi előtérbe, a jobb oldala a racionalitást. Az emberi teljesítményt és racionalitást érdemes két részre bontani, mégpedig azért, mert az ember nem feltétlen mindig racionális döntést hoz, ami a számítógépek esetén sem minden esetben várható el, sőt nem is minden esetben fontos. Lehetséges, hogy egy adott problémára kizárólag közelítő megoldás elegendő, mivel a teljes optimum meghatározása vagy nagyon sok időbe telik, vagy nagyon költséges pl. magas hardver kapacitásra van szükség. A két dimenzió mentén, a mesterséges intelligencia fogalmait, így négy különböző részre lehet osztani:

- Emberi módon gondolkodó rendszerek
- Racionálisan gondolkodó rendszerek
- Emberi módon cselekvő rendszerek
- Racionálisan cselekvő rendszerek

Az emberi módon gondolkodó rendszerek fogalomkörhöz tartozik többek között, Bellman (1978), Haugeland (1985), Borgulya (1998), Shabbir és Anwer (2018) definíciói. A mesterséges intelligencia:

- „*az emberi gondolkodással asszociálható olyan aktivitások automatizálása, mint pl. döntéshozás, problémamegoldás és tanulás*” (Bellman, 1978, 3.old.)
- „*újszerű kísérlet, hogy a számítógépeket gondolkodásra késztessük*” (Haugeland, 1985, 2.old.).
- „*olyan módszerek, amelyek az emberi problémamegoldást, következtetési folyamatot, vagy heurisztikus megközelítéseket modelleznek*” (Borgulya, 1998, 11.old.)
- „*ma rendelkezik az emberi intelligencia képességeivel, utánozza azt, különböző feladatokat végez, amelyekhez gondolkodás és tanulás szükséges*” (Shabbir - Anwer, 2018, 1.old.)

A racionálisan gondolkodó rendszerek fogalomkörhöz tartozik többek között, Charniak és McDermott (1985), Winston (1992) és Araújo (2014) definíciói. A mesterséges intelligencia:

- „*a mentális képességek tanulmányozása számítási modellek segítségével*” (Charniak - McDermott, 1985, 6.old.)
- „*az észlelést, a következtetést és a cselekvést biztosító számítási mechanizmusok tanulmányozása*” (Winston, 1992, 6.old.)
- „*képes az alapvető logikai következtetések levonására*” (Araújo, 2014, 129.old.)

Az emberi módon cselekvő rendszerek fogalomkörhöz tartozik többek között, Kurzweil (1990), Rich és Knight (1991), Mata et al. (2018). A mesterséges intelligencia:

- „*olyan funkciókat teljesítő gépi rendszerek létrehozásának a művészete, amelyhez az intelligencia szükséges, ha azt emberek teszik*” (Kurzweil, 1990, 14.old.)
- „*annak tanulmányozása, hogy hogyan lehet a számítógéppel olyan dolgokat művelni, amiben pillanatnyilag az emberek a jobbak*” (Rich et al., 2009, 3.old.)
- „*egy olyan kiterjesztett tudományterület, mely lehetővé teszi a számítógépek részére, hogy problémákat oldjanak meg komplex biológiai folyamatok emulálásával, mint a tanulás, érvelés és önkorrekció*” (Mata et al., 2018, 43.old.)

A racionálisan cselekvő rendszerek fogalomkörhöz tartozik többek között, Poole et al. (1998), Nilsson (1998), Dilek et al. (2015), Yu és Kumbier (2017), Pfeffer et al. (2017) definíciói. A mesterséges intelligencia:

- „*az intelligens ágensek tervezésének a tanulmányozása*” (Poole et al., 1998, 32.old.)
- „*a műtárgyak intelligens viselkedésével foglalkozik*” (Nilsson, 1998, 1.old.)

- „*lehetővé teszi számunkra, hogy olyan autonóm számításitechnikai megoldásokat tervezzünk, melyek alkalmazkodni tudnak felhasználási környezetükhöz*” (Dilek et al., 2015, 21. old.)
- „*lényegében adatvezérelt. Az ember-gép együttműködésén keresztül statisztikai koncepciókat fogalmaz meg, és így adatokat generál, algoritmusokat fejleszt és értékeli az eredményeket*” (Yu - Kumbier, 2017, 6.old.)
- „*képes automatizálni feladatokat*” (Pfeffer et al., 2017, 1.old.)

A modernebb definíciók jelentős része összeköttetésbe hozza a mesterséges intelligenciát az automatizálással és az adatvezérelt döntéshozatallal, míg a korábbi, főleg 2014 előtti, megfogalmazások részemre inkább futurisztikusabb hangvételt teremtenek. Ennek oka, hogy a korábban is említett Big Data feldolgozása mesterséges intelligencia eszköztárával együtt magas szintű tudás kiépítésére képes a szervezeti életben, tehát a technológiák párhuzamos fejlődése egymás húzóerejével is rendelkeznek. Továbbá, a mesterséges intelligencia definícióiban ellentétek is meghúzódnak. Míg Pfeffer (2017) az automatizálásra helyezi a hangsúlyt, elvonatkoztatva az emberi intelligencia reperformálásának képességétől a mesterséges intelligenciát, addig egyes kutatók, többek között Mata et al. (2018), Shabbir és Anwer (2018) teljes egészében az emberi intelligenciával asszociálják azt. Pfeffer gondolatait megalapozó, Diamant (2016) érdekes módon azt részletezi, hogy a mesterséges intelligencia kutatásoknak nem feltétlenül kell az emberi biológia és az agy működését szimulálnia, nem az az elsődleges vizsgálandó terület, mivel az amőbák és baktériumok is képesek intelligens cselekvéseket produkálni, úgy, hogy biológiai aggyal nem rendelkeznek. Álláspontom szerint, mivel számos olyan algoritmus is létezik, melyek képesek kvázi intelligenciát szimulálni pl. logisztikus regresszió, szupport vektor gépek, döntési fák, és itt az intelligencia alatt megalapozott racionális döntést értek, az emberi intelligencia és a mesterséges intelligencia párhuzamosításának nincs feltétlen létjogosultsága. Az értelmezett és kielemzett szakirodalmi definíciók alapján, a következő saját megfogalmazást alkalmazom:

A mesterséges intelligencia a racionális döntések sorozatát hivatott támogatni, olyan matematikai eszköztárt kínálva, mellyel lehetséges a logikus következtetések levonása és számítási problémák optimalizálása.

A mesterséges intelligencia algoritmusokra általánosságban elmondható, hogy számításigényes folyamatok összessége, tehát, hiába létezett korábban elmélet, azt a gyakorlat kevésbé tudta követni és kevésbé tudta igazolni az elméletek létjogosultságát a hiányzó technológiai háttér

miatt. Neurális hálók, ahogy fentebb említésre került, már a 40-es években is léteztek koncepció szintjén, de jelen korunk fejlettsége adja a lehetőséget, hogy az azokban rejlő számítási erőt kihasználjuk és olyan alkalmazásokat hozzunk létre, melyek sokáig csak elméletben léteztek. Erre a legjobb példa az önvezető autót támogató szoftverkomponensnek, mely mélyített neurális háló alapon működnek (Maqueda et al., 2018).

Álláspontom szerint a mesterséges intelligencia, mint önálló tudományterület nem helyezhető el egyértelműen a Peter és Norvig (2009) által összeállított 4 kategória egyikében. A mesterséges intelligenciához jelenleg számos alterület tartozik, melyek lefedik az említett teljes két dimenziót, és az alterületekhez tartozó újabb és újabb algoritmusok, döntéstámogató rendszerek, intelligens alkalmazások száma folyamatosan növekszik, mely a mesterséges intelligencia sokszínűségét tovább diverzifikálják. Továbbá, kutatói munkám karakterisztikái miatt, számomra az információbiztonsági események és anomáliák detektálása a központi kérdés, így eddigi munkám főleg a mesterséges intelligencia azon ágának leszűkítésére összpontosított, mely a célkitűzéseim megoldására irányulnak. Eddigi szakirodalmi áttekintésem alapján, a teljesség igénye nélkül, az alábbi területeket lehet mesterséges intelligenciával hatékonyan kiaknázni, melyeket a 2. táblázat demonstrál.

2. táblázat: A mesterséges intelligencia kutatások összefoglaló táblázata különböző általam feldolgozott szakirodalom alapján

Kutatási terület	Alkalmazott eljárások	Témában megjelent tanulmányok
Természetes nyelvi feldolgozás	Hangfeldolgozás, Beszédfeldolgozás, Szövegfeldolgozás	(Suster et al., 2017), (Trivedi et al., 2017), (Vu et al., 2018), (Young et al., 2018)
Képfeldolgozás	Gépi tanulás	(Baygin et al., 2018), (Noyel - Jourlin, 2018), (Tutika et al., 2018)
Mintázatfelismerés	Gépi tanulás, Evolúciós optimalizálás, Adatbányászat	(Cho et al., 2018), (Kuznetsova et al., 2018), (Lin et al., 2017)
Irányítástechnika	Fuzzy rendszerek	(Damelin et al., 2015), (Syropoulos, 2014) (Vychodil, 2016)
Kereső eljárások	Genetikus programozás	(Behandish és Wu, 2014), (Martí et al., 2016)
Tudásreprezentáció	Szakértői rendszerek	(Ravuri et al., 2018),

		(Shah et al., 2017)
Döntéstámogatás	Gépi tanulás, Valószínűségi következtetési eljárások, Nem klasszikus logikák	(Veale et al., 2018)
Robotika	Helymeghatározás, Mozgástervezés, Hardveres vezérlés	(Chatzilygeroudis et al., 2016)
Digitális valóság	Virtuális valóság, Kiterjesztett valóság	(Stets et al., 2018)

Szakirodalmi kutatásom alapján arra a következtetésre jutottam, hogy mivel az anomália észlelés alapgondolata mögött a minták felismerése a végső cél, mint a normális viselkedés matematizálása és az eltérő magatartások felderítése, ezért az anomália észlelésére a mesterséges intelligencia egyik szűkebb köre, a gépi tanulás eszközei képesek megoldásokat nyújtani. Azonban, nem szabad elvonatkoztatni a mesterséges intelligencia más alterületeitől sem, mivel, az első hipotézisemmel összhangban, véleményem szerint a hibridizált rendszerek képesek lehetnek magasabb teljesítmény elérésére.

3.2.2. Gépi tanuló rendszerek fejlesztése

A gépi tanulás a mesterséges intelligencia egyik alterülete, melynek célja a rendelkezésre álló adathalmazban meghúzódó mintázatok felismerése, megtanulása, majd a mintázatok alapján új adatrekordok csoportosítása (Gartner.com, 2017/b). A gépi tanulás Raschka megfogalmazása alapján, „*a számítógépek felruházása a tanulás képességével*” (Raschka, 2015, 2.old.). Sokkal technikaibb definíciót ad meg Russel és Norvig (2005, 749.old.): „*A tanulás alapgondolata az, hogy a megfigyeléseket ne csak az ágens jelenlegi cselekvéseinek kialakítására használjuk, hanem arra is, hogy javítsuk a cselekvésre való jövőbeli képességeit*”. Ebben a definícióban megjelenik a jövő, azaz értelmezésem szerint a predikció megnevezése, miszerint a gépi tanulás egy olyan automatizált folyamat, mely hozzájárul a jövőre vonatkozó sejtések előrejelzéséhez. Dua és Du (2011, 24.old.) szerint a gépi tanulás „*egy tudományos modell építése, mely képes a jelenlegi adatokból tudást képezni*”. Hastie et al. (2009, 1.old.) egyszerűen „*az adatokból való tanulás*” miként jeként definiálja a fogalmat. A gépi tanulásra vonatkozó megfogalmazások összefoglalásaként én az alábbi definíciót szolgáltatom:

A gépi tanulás a számítógépek által történő adatfeldolgozás tudománya, melynek eredményterméke egy jövőbeli állapot minél pontosabb meghatározása.

A gépi tanulás alábontását a szakirodalom többé-kevésbé egységesen kezeli. Hastieet al. (2009) a gépi tanulást felügyelt és felügyelet nélküli tanulásként csoportosítja, míg pl. Raschka (2015) egy harmadik kategóriát is külön kiemel, ami a megerősítéses tanulás. A felügyelt tanulás egy meglévő adathalmaz, azaz historikusan összegyűjtött információ alapján mintázat keresési eljárások összessége, mely a mintázat feltárását követően hozzájárul egy adott célváltozó értékének előrejelzéséhez (Sagar, 2015). A felügyelt tanulásra példa a pénzügyi intézeteknél is alkalmazott profilozó eljárások, melyek alapján kategorizálásra kerül, hogy adott potenciális ügyfél milyen mértékben lesz képes a felvett hitelt törleszteni. A célváltozó ebben az esetben a kockázat egy 0-tól 1-ig terjedő skálán a pénzvisszafizetés lehetséges valószínűsége, mely, ha 1-hez közeli akkor a potenciális ügyfél kockázatosnak minősíthető, és a pénzintézet rendelkezhet úgy, hogy nem folyósít hitelt, míg 0-hoz közeli érték alacsony kockázatnak tudható be, tehát a pénzintézet folyósít hitelt az igénylőnek. A historikus adatok ebben az esetben lehetnek a igénylő neme, kora, lakcíme, havi nettó jövedelme stb. Egy másik korszerű alkalmazási területe a felügyelt tanulásnak az önvezető autók által alkalmazott automatizmusok. Historikus adatok alapján a gépi eljárás képes felismerni az önvezető autó előtt, mellett, hátul közlekedő más járműveket és eszerint pozícionálni önmagát, megelőzve a baleseteket. Értelmezi a közlekedési táblákat, észreveszi a gyalogosokat, tehát a cél, alkalmazkodni a külső környezethez (Maqueda et al., 2018). A gépi tanuló eljárások másik kategóriája a felügyelet nélküli tanulás. Az ebbe a csoportba tartozó eljárások a begyűjtött adatokban hasonlóan mintázatot keresnek, azonban a cél nem egy célváltozó becslése, hanem a mintázat alapján az adatok egyes kategóriákba való csoportosítása (Sajtos - Mitev, 2007). A klaszterező eljárások tipikusan felügyelet nélküli eljárások. A marketingkutatásban gyakorta használatosak a piaci szegmentációk feltárása, így azok elemzése és kiértékelése a felügyelet nélküli tanulási eljárások egyik gyakori alkalmazása. A megerősítéses tanulásban a rendszer feladata, hogy képes legyen a környezetéhez alkalmazkodni és optimális stratégiát válasszon egy adott költségfüggvény minimalizálása, vagy jutalomfüggvény maximalizálása érdekében (Peter - Norvig, 2009). Az automatizált sakkjáték a megerősítéses tanuláson alapuló gépi tanuló rendszer egy klasszikus példája.

A gépi tanuló eljárások kutatása széleskörű, és a szakirodalmi kutatásaim alapján arra a következtetésre jutottam, hogy a gépi tanuló rendszerek fejlesztése az információrendszerek fejlesztéséhez hasonlóan több fázisra bontható, mely fázisok tervezése azért indokolt, hogy a lehetséges döntési pontok (pl. adattranszformáció, modellválasztás, metrika választás) a modell objektíven mért jóságához a lehető legnagyobb mértékben hozzájáruljon. A gépi tanuló

eljárások fejlesztése esetén, összegezve, célszerű az általános alkalmazás- és információrendszer fejlesztési módszerekkel magas szinten párhuzamot vonni, azonban álláspontom szerint, ezt ki kell terjeszteni a tudományos modellek tervezésének és fejlesztésének módszertanával. Gépi tanuló rendszerek esetén is igaz az állítás, hogy az alkalmazást körültekintően, bevált iparági gyakorlatok és kutatási eredmények által igazolt módszertanok mentén érdemes fejleszteni. Szepesné (2010) az információrendszerek fejlesztését 3 részre osztja. Első szakaszban, melyet Szepesné (2010) előszakasznak nevez, történik a kiindulási helyzet elemzése, a feladat megfogalmazása, a költségelemzés, illetve az előnyök definiálása, tehát az információrendszer célkitűzésének meghatározása. A második szakasz a fejlesztés szakasza, melyben kivitelezésre kerül az adatbázis struktúra leírása, az alrendszerekre való bontás, a specifikációk véglegesítése és a rendszerkomponensek együttműködésének biztosítása. Az utolsó szakaszban, a felhasználói szakaszban, veszi kezdetét az alkalmazás használata, karbantartása, felülvizsgálata és optimalizálása. Kovács (2009) ajánlást fogalmaz meg a tudományos szimuláció és modellépítéssel kapcsolatban. Elsőként a modellalkotás célját kell meghatározni, majd a vizsgálandó rendszert kell definiálni, amit a modellkalibráció követ, azaz az inputként szolgáló adathalmazt kell a modellre szabni. Az utolsó lépés a validáció, az ellenőrzés szakasza, hogy a modell elérte-e az előzetesen definiált célját.

A felhasznált szakirodalmak áttekintésének segítségével az alábbiakban leírtak szerint érdemes megtervezni és fejleszteni a kutatási célok elérését szolgáló gépi tanuló alkalmazásokat, melyben figyelmet kell szentelni az iparági gyakorlatok és a tudományos céllal történő modellalkotás elvárásainak, tehát az általam alkalmazni kívánt modellalkotási folyamat a következő. A gépi tanuló alkalmazások fejlesztésének az első fázisa a gépi tanuló rendszer célkitűzésének definiálása. Ebben a szakaszban történik az adott üzleti probléma kiértékelése, és hogy az új alkalmazás milyen módon lesz képes az üzleti problémát megoldani, elhárítani, karbantartani. A célkitűzés meghatározásakor a felhasználni kívánt adatvagyont, annak beszerzési kritériumait, a függő és független változókat is meg kell határozni, majd még az adatgyűjtés költséges fázisa előtt kiértékelni tesztadatokkal, hogy valóban az alkalmazás tervezeti szinten úgy fog működni, ahogy azt annak felhasználói elvárják. Amennyiben a specifikáció megfelelő minőségű eredményeket produkál, úgy a következő fázis az adatok beszerzése. Adatokat lehetséges különböző forrásokból összegyűjteni, mely lehet az adatok vásárlása harmadik személytől, az adatok munkahelyen belüli kollekciója, internetes letöltése stb. Az adatok begyűjtése után az adatok transzformációjára van szükség.

Mivel az alkalmazni kívánt adathalmaz gyakorta nem áll megfelelő formában rendelkezésre, így azok tisztítása, transzformációja, tehát előfeldolgozása indokolt. Azon felül, hogy egy adott gépi modell teljesítményét is képes fokozni, a magas minőségű adattranszformáció ellenálóbb lehet külső támadások ellen, ahol az alkalmazás funkcionalitásának megkárosítása a támadó célja (Bhagoji et al., 2017). Az adatok előfeldolgozása esetén az alábbi kihívásokkal kell szembenézni:

- Hiányzó értékek az adathalmazban. Hastie et al. (2009) felveti, hogy hiányzó értékek esetén egyik lehetséges megoldás a hiányzó függő változó oszlopában szereplő adatok átlagát, móduszát vagy mediánját venni, majd a hiányzó értékeket azzal helyettesíteni. Raschka (2015) megjegyzi, hogy sok esetben a hiányzó rekord teljes törlése is indokolt lehet, mivel a hiányzó adatokkal való adatfeldolgozás a modell predikciós erejét csökkentheti, álösszefüggéseket indukálhat. Bealuc és Rosenthal (2018) egy klasszifikációs és regressziós döntési fából álló algoritmust javasol, mely kutatási eredményeik alapján magasabb teljesítménnyel szolgál, mint az adatok átlagolása, vagy egyéb statisztikai módszerekkel való kitöltése. Modelljükben a hiányzó adatok feltöltését egy másodrendű gépi tanuló eljárásnak kezelik, ahol a hiányzó értékeket kell előrejelezni. A hiányzó értékek esetén mérlegelendő, hogy az adatok eltávolítása mekkora kárt tehet a teljes adathalmaz értelmezhetőségében, amennyiben az csak egy kitüntetett függő változóra jellemző érdemes lehet kizárólag a változó elhanyagolása.

- Különböző mérési skálán mért adatok transzformálása: Az adathalmaz a céltól függően különböző mérési skálán szereplő adatokat tartalmazhat, melyek lehetnek nominális (névleges), ordinális (rendezéses), intervallum (különbség), vagy arányskálán mért adatok (Szűcs et al., 2008). A nominális és ordinális skálán mért adatok szövegszerű értékeket tartalmazhatnak, ezért ezen adatok kódolása szükséges, hogy a számítógép képes legyen az adatfeldolgozásra (Bodon, 2010). Au (2017) a „*one-hot-encoding*” módszert ajánlja, melynek lényege, hogy a szöveges értéket tartalmazó függő változókból új változók létrehozásával, a változó különböző értékeinek számával megegyezően, orvosolni lehet a problémát, ahol a változó csak ott vesz fel pl. 1 értéket, ahol az az eredeti változóban adott érték jellemző volt rá.

- Adattranszformáció a modell teljesítményének növelése érdekében: Hackeling (2014) felhívja a figyelmet, hogy különböző gépi tanuló módszerek érzékenyek, ha az adatok

nem ugyanazon a skálán szerepelnek. Ilyen pl. a neurális hálók, logisztikus regresszió, szupport vektor gépek, K-legközelebbi szomszéd algoritmusok, melyek performanciája nagyban függ az alkalmazott adattranszformációs eljárástól. Az adatok transzformációja lehetséges, többek között az adatok standardizálásával vagy normalizálásával.

- Függő változók statisztikai transzformálása: A függő változók nagyszáma performancia problémához vezethet, azon túl a változók értelmezése is megnehezedik, így Sajtos és Mitev (2007) a sok homogén jellemzővel rendelkező adathalmazban a változók közötti összefüggések feltárására a faktorelemzést javasolja, mely a változók redukcióját eredményezi.

- Függő változók kiválasztása: Az adatgyűjtés során olyan adatok kerülhetnek be az adathalmazba, melyek egyáltalán nem járulnak hozzá a célváltozó előrejelzéséhez, ezért azok kvázi használhatatlanok az üzleti cél elérése érdekében. Raschka (2015) különböző változó kiválasztási algoritmusokat tesztelt (Szekvenciális változó kiválasztási algoritmus és Véletlen erdő) és bizonyította, hogy a hardveres erőforrás performancia javulásán túl, melyet a kevesebb változó bevonása eredményezett, az alkalmazott tanuló eljárás predikciós ereje legalább 2%-kal javult.

- Adathalmaz felbontása tréning-, teszt-, és validációshalmazra: Ahhoz, hogy megfelelően validálni lehessen az elkészítendő modell performanciáját az adatokat tréning-, teszt-, és validációshalmazra szükséges bontani (Combs, 2016). A tréning adatok szolgálnak a modell tanítására, vagyis a mintázatok felfedésére és azok elraktározására. A teszthalmazon, mely független a tréninghalmaztól, elvégzett számítások adnak becslést a modell álalánosítóképességére, mivel olyan adatokat tartalmaz, melyet a modell korábban nem látott, így alkalmas a modell tesztelésére. A validációshalmaz alkalmazható a kiválasztott modellek további tuningolására, vagyis a belső paraméterek optimális kombinációjának megtalálására.

A gépi tanuló eljárások adatfeldolgozó szakaszát követően, a fejlesztési szakasz veszi kezdetét. A fejlesztési szakaszban kiválasztásra kerülnek az alkalmazni kívánt modellek, majd előre felállított jóságkritériumok mentén azok kiértékelése. Szakirodalmi kutatásom alapján összegyűjtésre kerültek a leggyakrabban alkalmazott gépi tanuló eljárások, melyeket kutatómunkámban is fel kívánok használni a hibrid modellezés keretében. Az alkalmazni szánt

modellek elemzésében és kiválasztásában az anomália detektálása, mint elsődleges cél van fókuszban. Az anomália észlelés többféle szempontból megközelíthető. Értelmezésem szerint, felfogható, mint egy klasszifikációs feladat, melyben az anomália egy külön osztályként jelenik meg, tehát a hibrid modellezésben olyan gépi tanuló eljárásokat szükséges alkalmazni az output becslésére, melyek bizonyítottan alkalmasak klasszifikációs feladat elvégzésére. Másodsorban, a nem ismert anomáliákat kiugró értékként is lehet kezelni, melynek következtében statisztikai módszerekkel is érdemes az anomáliaészlelést tanulmányozni. Harmadsorban, logikai következtetés eredményeképp, az anomáliák csoportosításának is van értelme (pl. azok elkülönítése a normális magatartásoktól), tehát empirikus kutatással vizsgálandó feladat a különböző csoportok, azaz klaszterek felderítése. Mindezek technikai megvalósítását szem előtt tartva, a modellépítés során alkalmazott output struktúra eltérő lehet, így olyan algoritmusokra van elsősorban szükség, melyek a kívánt struktúrák leképezésére alkalmasak. Ugyanakkor, a hibridizálás nem kizárólag közvetlenül a kimenet struktúrájára lehet hatással, lehetséges a különböző algoritmusokat a hibrid modell részfeladatainak optimalizálására is alkalmazni az egyes döntési pontok és feladatok megoldásában, melynek célja az egyes gépi tanuló módszerek előnyeinek és erősségeinek kiaknázása, melyek más algoritmusok esetén gyengeségnek bizonyulnak. Az alábbi kritériumok definiálásával és a szakirodalom tanulmányozásával összhangban, kutató munkámban az alábbi modelleket kívánom elsősorban alkalmazni:

1. **Logisztikus Regresszó (Logistic Regression).** A logisztikus regresszió egy olyan módszer, mely egy bináris output valószínűségét hivatott előrejelezni az inputváltozók alapján. Az algoritmus futása során az input változók egy súlyvektorral kerülnek összeszorzásra, melyet egy logisztikus függvény aktivál. A logisztikus regresszión alapuló gépi tanuló eljárás a súlyok folyamatos változtatásán keresztül tanul, mely az alkalmazás négyzetes hibáján keresztül realizálódik, vagyis a hiba nagysága korrekcióra kerül a logisztikus aktivációs függvény deriváltja szerint. Az alkalmazás előnye, hogy egyszerűen lehet a súlyókat változtatni alacsony komplexitása miatt, azonban alulteljesít nemlineáris problémák esetén (Elitdatascience.com, 2017).
2. **Szupport Vektor Gépek (Support Vector Machines, avagy SVM).** Az SVM alkalmazások fő célkitűzése az egyes osztályok közötti döntési határok maximalizálása. Ha az osztályok nem szeparálhatók lineárisan, akkor az SVM kernelizálható, ami egy magasabb dimenzionális térben képezi le a mintákat. Mivel az osztályokat elválasztó szeparátor maximalizálásra kerül, ezért az SVM alkalmazások kevésbé kitettek a túlilleszkedésnek, ugyanakkor a kernel kiválasztása korántsem triviális, ami megnöveli

az optimalizálásra szánt időt. Anomália észlelésre SVM alapú eljárást fejlesztettek Chen et al. (2005) és Zhang és Zulkernine (2006/b).

3. **Döntési Fa (Decision Tree).** A döntési fa alapú gépi tanuló eljárásokban a modell kérdések sorozatára válaszol, mely végső soron adott osztály meghatározásához vezet. A döntési fa az egyes attribútumok szerint kerül „szétdarabolásra" a célváltozó függvényében. A döntési fa magas varianciájú modelleket generál, így nagy az esélye a túlilleszkedésnek.
4. **Véletlen Erdő (Random Forest).** A véletlen erdők hasonlóak a döntési fák működéséhez, azonban több döntési fa általi predikciót kombinálnak, ahol a fák véletlen vektorok alapján kerülnek meghatározásra, így azok sokkal robusztusabbak, de ez sokkal lassúbbá teszi a tanulási folyamatot. Véletlen erdő alapú algoritmust fejlesztett anomália detektálására Zhang és Zulkernine (2006/a).
5. **Neurális Hálók (Neural Networks).** A neurális hálók alapkoncepciója, és amiről a nevét is kapta, az emberi agy biológiai neuronjait hivatott gépi formában megtestesíteni. Hasonlóan a természetes neuronok felépítéséhez, a mesterséges neuronok is kapcsolódnak egymáshoz, ahol egy közvetítő közeg szállítja a feldolgozott információt neuronról-neuronra. A neuronok különböző rétegekben helyezkednek el, az input réteg fogadja a feldolgozáshoz szükséges bemenő adatokat, az output réteg prezentálja a prediktált kimenetet, a közbelső rétegek, pedig transzformálják és segítségükkel jut el az információ az input neurontól az output neuroning. A kapcsolóelem a neuronok között a dendrit, mely létezhet bármely neuron pár között, de a kapcsolat mérete gyakorta az, hogy minden neuron kapcsolódik minden másik következő rétegbeli neuronhoz. A megküldött információ az úton az egyik neurontól a másik felé átalakul a dendritek súlyaitól függően, mely végső soron az adatok tudásanyagát tartalmazza, és melyet a neurális háló a megadott tanulási eljárás és az adatokban fellelhető minta szerint módosít. A fogadó neuron a hozzá beérkező adatokat összesíti, majd egy aktivációs függvény segítségével eldönti, hogy a számára fogadott input, milyen outputként távozik. Mivel a neurális hálók komplex számításokon alapulnak, ezért optimalizálása időigényes. Legnagyobb előnye, hogy univerzális függvény approximációra képes, ezáltal nemlineáris leképezésekre is alkalmazható (Chen et al., 2018). Neurális hálókkal vizsgált anomália felismerő rendszereket, többek között, Chen et al. (2005), Zhai et al. (2016) és Vu et al. (2018).
6. **K-legközelebbi Szomszéd (K-nearest Neighbors, avagy KNN).** A KNN algoritmus egy példája a kevés paraméterrel rendelkező gépi tanuló eljárásoknak, ezért azt „*lazy*",

azaz lusta algoritmusnak is nevezi a szakirodalom, tehát finomhangolása relatív gyors. A KNN a többségi szavazás elvén működik, azaz egy adott mintát a hozzá legközelebb álló ismert mintákhoz sorolja. KNN alapú gépi tanuló eljárást fejlesztett anomália észlelésre, többek között, Yang et al. (2018) és Abah et al. (2015).

7. **Bayesian Hálók (Bayesian Network).** A Bayesian hálók olyan tanuló eljárások, melyek gráfként szemléltetik a következtetés logikáját és csomópontjai valószínűségi változók. Az algoritmusok feltételes valószínűségtáblák alapján operál. Bayesian Hálót tervezett anomália észlelésre, többek között, Heard et al. (2010).
8. **Fuzzy mintázatfelismerés (Fuzzy Pattern Recognition).** Az algoritmus a fuzzy halmazelmélet és fuzzy logika alapján működik, mely a klasszikus logikával ellentétben engedi, hogy egy állítás 0 és 1 közötti bizonyossággal igaz legyen, ezért kifinomultabb logikai következtetéseket enged, ugyanakkor magas futásidőt igényel. Erre példa, Luo és Bridges (2000) fuzzy-logika alapú anomália detekciós alkalmazása.
9. **Evolúciós optimalizálás (Evolutionary Optimization).** Az evolúciós optimalizálás alapvetően kereső eljárás, melyet az evolúciós biológia inspirált. Egy adott problémára való megoldások halmazát egy kezdeti génpopulációként kezeli, melyek evolúciós folyamatok mentén válnak egyre jobb megoldássá. Ez a folyamat addig iterálódik, amíg egy előre definiált fitnessz-függvény nem lesz optimális, vagy a problémára egy elfogadható megoldás (szuboptimális válasz) születik. Anomália észlelés területén, Kaur et al. (2013) kiemelték az evolúciós algoritmusok jelentőségét a probléma megoldásában.
10. **Klaszterező eljárások.** A klaszterező eljárások az előre nem ismert minták felderítésében szolgálnak segítségül, mely algoritmusok futásának eredményeképp az egymásra legjobban hasonlító rekordok egy közös klaszterbe kerülnek. Hátrány lehet (pl. K-közép esetén), hogy a klaszterek számát előre kell definiálni, ezért a megfelelő paraméter kikísérletezése időigényes feladat lehet.

A modellalkotást követően azok kiértékelését kell elvégezni annak érdekében, hogy a teljesítményt mutatószámokkal ki lehessen fejezni. A jóság kritériumokra előzetesen metrikákat kell építeni, hogy objektivitása biztosítható legyen. Az anomália észlelés során az alábbi jóság metrikákat és kimutatásokat kívánom alkalmazni és elemezni:

1. **Relatív hiba.** A becslés abszolút hibáját adja meg a megfigyelt értékek tükrében.
2. **Helytelen becslések aránya.** A modell helyes becslésének valószínűsége.

3. **Álpozitív becslések aránya.** Az álpozitív találatok aránya. Az anomália észlelés során az egyik legfontosabb mutató, mely azt az információt szolgáltatja, hogy az igazolt anomáliákat milyen arányba fogadta el a rendszer normális magatartásként.
4. **Álnegatív becslések aránya.** Az álnegatív találatok aránya.
5. **Átlagos négyzetes hiba.** A kimenet négyzetes hibaösszege.
6. **Vevő Működési Karakterisztika (Received Operating Characteristic).** A klasszifikáció álpozitív és álnegatív aránya közötti kompromisszum grafikus megjelenítése.
7. **Validációs görbe.** Az egyes adathalmazok (tréning, teszt és validációs) tanulásának pontosságát hivatott bemutatni a minta időbeli feldolgozásának tükrében. Ha a tréninghalmaz és teszt-, és validációshalmaz görbéi közötti különbség jelentős, akkor a rendszer nagyvalószínűséggel túlilleszkedett, azaz megtanulta az adathalmazban jelentkező véletlen zajokat, vagy „memorizálta" a tréninghalmazt.
8. **Igazságmátrix.** Az igazságmátrix grafikusan szemlélteti az igaz pozitív és igaz negatív, valamint az álpozitív és álnegatív értékeket.
9. **F1-pontszám.** Az F1-pontszám egy 0 és 1 közötti érték, mely kétszer a modell pontossága és igaz pozitív mutatójának szorzata és a a modell pontossága és igaz pozitív mutatójának hányadosa.
10. **Tanulás gyorsasága.** A tanulás gyorsasága azt szemlélteti, hogy a modellt hányszor kellett lefuttatni a tréninghalmazon, amíg azt elfogadható szinten megtanulta.
11. **Keresztvalidáció.** A keresztvalidáció a modellek pontosságát mutatja különböző validációshalmazok véletlen kiosztásával.

A gépi tanuló módszerek jelentős része nagy mennyiségű paraméterrel rendelkezik, ezért a jóság metrikák legelső kiértékelése után szükségszerű kísérletet végezni, azaz „*tuningolni*" a modellek paramétereit az optimális teljesítmény elérése érdekében. Mivel a paraméterek száma és fajtája modellenként eltérő, ezért általánosságban az alábbi optimalizálási eljárásokat érdemes figyelembe venni Baesens et al. (2015) és Fawcett és Provost (2013) alapján:

1. **Túlilleszkedés csökkentése.** Túlilleszkedés akkor merül fel, ha a modell alaposan megjegyezte a tréninghalmazban meghúzódó mintákat vagy véletlen zajokat, azonban a teszthalmazon alulteljesít. A problémát a modell komplexitásának csökkentésével pl. függvény regularizációval lehet kivitelezni.
2. **Tanulási ráta növelése/csökkentése.** A tanulási ráta a modell konvergenciáját hivatott lassítani vagy gyorsítani. Magas tanulási ráta esetén a konvergencia gyorsabb, azonban

így könnyebben átugorhatók a hibafüggvény minimumpontjai, míg alacsony tanulási ráta esetén a konvergencia lassabb, azaz a tanulás költségesebb, ellenben, garantált a hibafüggvény egy lokális minimumának megtalálása. Fontos kiemelni, egyik esetben sem biztosított a globális minimumpont felfedezése, ezért a gépi tanuló eljárások súlyvektorainak véletlen változtatására empirikus kutatás javasolt.

3. **Tréningelés növelése.** Könnyen meglehet, hogy a tréninghalmaz egyszeri „*megmutatása*" a gépi tanuló eljárásnak nem jár sikerrel, vagyis nem képes a mintázatot teljes egészében leképezni, ezért a tréningelés számának növelése szükséges lehet a processzoridő kárára.
4. **Aktivációs függvény változtatása.** A klasszifikációs gépi tanuló algoritmusok egy része aktivációs függvényt alkalmaz a klasszifikáció valószínűségének meghatározására pl. szigmoid, tangens, RELU stb. Egy adott aktivációs függvény egy adott problémára nem biztos, hogy optimális eredményt tud adni, ezért empirikus úton javasolt kísérletezni különböző aktivációs függvényekkel.
5. **Architektúra változtatása.** Amennyiben az egyes modellparaméterek optimalizálása során is a modell alulteljesít, indokolt lehet a teljes architektúrát újragondolni.

A modell építése és optimalizálása, valamint a jóságmetrikák megfelelő teljesítményének elfogadása után a gépi tanuló rendszer éleskörnyezetbe való állítása következik, melyben a rendszer üzemszerű működése kezdetét veszi.

3.2.3. Az alfejezet összefoglalása

A saját kutatási téma vonatkozásában, a mesterséges intelligencia és gépi tanuló rendszerek fejlesztéséről folytatott szakirodalomkutatás az alábbi pontokban járult hozzá kutatómunkámhoz:

- A mesterséges intelligencia által kínált eszköztár alkalmasnak bizonyul az anomáliák észlelésében, mivel annak gépi tanuló rendszerekre irányuló kutatási eredményei alátámasztják annak létjogosultságát.
- A gépi tanuló rendszerek fejlesztése tervezési és optimalizálási feladatot igényel, kezdve az adatfeldolgozástól, a modellválasztáson keresztül a jóságmetrikák kiértékeléséig, ezért minden egyes részfeladat megalapozott kutatási eredményeken kell, hogy alapuljon.

- A különböző modellek eltérő logikát és megközelítéseket alkalmaznak egy adott probléma megoldására, így az egyes modellek erősségeit és gyengeségeit szükséges kielemezni, és olyan hibrid modelleket létrehozni, mely a felhasznált algoritmusok hátrányait képes egy másik algoritmus erősségeivel pótolni, így szándékszerint növelve a rendszer performanciáját és megbízhatóságát.

3.3. Hipotézisek racionalitását alátámasztó szakirodalom kutatás

3.3.1. Az anomáliaészlelés fogalmi kerete

Az információbiztonsági kihívások megoldása gépi tanuló módszerekkel széles szakirodalmat kínál, azonban a kutatások jelentős része a kiberbiztonsági védelmen belül a külső behatolások megfelelő időben való detektálására összpontosít. Denning (1987) bevezette az IDS (Intrusion Detection System azaz Behatolás-Érzékelő Rendszer) fogalmát, amely óta számos behatolás-érzékelő funkcióval ellátott rendszer fejlesztése valósult meg, melyekről szakirodalmi összefoglalásom is a következő fejezetekben értekezik. Ezen rendszereket a szakirodalom alapvetően két felé osztja (Dua és Du, 2011):

1. Az első az „Anomália Észlelés”, mely elsődleges célja a normálistól való eltérés detektálása, azaz egy új rekord a historikusan begyűjtött adathalmazhoz való összevetése és annak vizsgálata, hogy az mennyire különbözik vagy illeszkedik az eddig definiált mintázatba. Ez a technika lehetővé teszi a jelentősen eltérő viselkedéssel bíró objektumok megtalálását. Ez lehetséges, többek között, a hálózati forgalom monitorozásával, ahol a hálózati viselkedés naplófájljait alapul véve a cél a gyanús magatartás felderítése, azonban, az eltérő viselkedés még nem jelent azonnali anomáliát, ezért az anomália észlelés egyik legnagyobb hátránya, hogy az észlelt eltéréseket azonnal ki kell vizsgálni, így ezen rendszerek esetén általában magas az álpozitív találat. A másik jelen kutatói probléma az anomália észleléssel kapcsolatban, hogy a védekező rendszerek elterjedésével, azok működését a támadók is el kezdték tanulmányozni, így a kiberbűnözők igyekeznek a normális felhasználói magatartást lemásolni, amit a rendszer érvényesnek kategorizál, így az anomália tényét nem deklaráló naplóállományok kivizsgálása is indokolt lehet. Továbbá, álnegatív találat alatt, azt a találatot értjük, melyet a rendszer normális tevékenységként könyvelt el, azonban az anomália volt, ezért az álnegatív találatok is magas kockázatokat hordoznak magukban. Összefoglalva, az anomália érzékelő rendszerek esetén jelen

legnagyobb kihívás az álpozitív és álnegatív találatok minimalizálása, melyre Dua és Du (2011) is rávilágít.

2. A második csoport a „visszaélés felismerés” vagy „aláírás felismerés”, ahol a cél egy adott új viselkedésről eldönteni, hogy az kártékony-e vagy sem egy meglévő adatbázis alapján, amelyben a gyanús viselkedések definiálva vannak, és az algoritmusok kereséssel győződnek meg arról, hogy az adatbázisban tárolt mintázatba az új jel mennyire illeszkedik. Ezen az elven működnek a vírusírtók, melyek az ismert vírusok karakterisztikáit letárolják, majd a fájlok szkennelésekor ezt a definíciós adatbázishoz hasonlítják. A legnagyobb hátránya a kialakított módszereknek, hogy bár hatásosnak tűnnek az ismert tulajdonságokkal rendelkező támadásokkal szemben, egy új, eddig nem látott viselkedés esetén nem képesek annak kiszűrésére.

Kutatói munkám az anomália észlelés szakterületével foglalkozik, így a szakirodalmi kutatás további része főleg e téma köré épül, azonban az aláírás felismeréssel kapcsolatban is ismertetésre kerülnek a jelentős kutatási eredmények, mivel azok számottevő hatással bírtak az anomáliák észlelésére is. Az elemzett szakirodalmak a felállított hipotézisek mentén keresztül lettek strukturálva keretet adva azoknak az elméleteknek és kutatási eredményeknek, mely forrásként szolgálnak az egyes hipotézisem bizonyíthatóságához.

3.3.2. H1 hipotézis felismerését/racionalitását támogató szakirodalom

H1: Különböző logikát/megközelítést alkalmazó mesterséges intelligencia módszertanok magasabb fokú hibridizációjával magasabb teljesítmény érhető el az incidensek/anomáliák detektálására vonatkozóan.

Ahogy az korábban kifejtésre került, számos mesterséges intelligencia és gépi tanuló algoritmus alkalmazható különböző feladatok megoldására, melyek eltérő előnyökkel és hátrányokkal rendelkeznek. A cél olyan hibrid modellek megalkotása, melyek képesek az egyes gyengeségeket lefedni más tanuló algoritmusok erősségeivel hozzájárulva a modell teljesítőképességének növeléséhez. A felállított hipotézis azt tárgyalja, hogy a magasabb fokú hibridizáció magasabb performanciát eredményez, azonban a hibridizáció fokának meghatározása korántsem triviális feladat. Ahhoz, hogy a hibridizációt mérni lehessen, szükséges olyan objektíven kezelhető és mérhető modell-tulajdonságok kialakítása, melyekről

el tudjuk dönteni, hogy minél nagyobb/kisebb értékkel rendelkeznek, annál jobban/kevésbé járulnak hozzá a fejlesztett modell komplexitásához, hibridizálásához. Eddigi szakirodalmi kutatásom alapján nem találtam olyan objektív metrikát, mely egy gépi tanuló modell hibridizálásának fokát tudná matematizálni, ezért célom a hipotézissel a fogalom megalkotása annak objektív levezetése és visszamérhetőségének bizonyítása is egyaránt.

Hibrid rendszer alatt különböző, a következőkben is bemutatott szakirodalmi példák felsorakoztathatók, melynek nagy része az alapvető modellek kombinációját, vagy felügyelt és nem felügyelt modellek együttes alkalmazását értik, kutatások elérhetőek a mesterséges intelligencia más alterületeiből vett és a gépi tanuló algoritmusok együttes kombinációiból, illetve az anomália észlelés és aláírás felismerés közös működéséből eredő megoldásokból is, azonban, probléma hangsúlyozandó, a hibridizáció mércéje nem került még levezetésre.

Portnoy et al. (2001) klaszterező eljárással próbálták megállapítani audit fájlokból a betörés tényét és sajátosságát a hálózaton, viszonylag alacsony 50%-os performanciával, mely a modell teljesítőképességét tesztadaton mért pontos találatok számával határozták meg. A kutatás rávilágít egyrészt, hogy önmagukban a nem felügyelt tanulási módszerek nem tűnnek elégségesnek, továbbá, az egyszerű tradicionális modell önmagában nem volt elegendő a probléma megoldására. Yamanishi és Takeuchi (2001) hibrid módszert alkalmaztak felügyelt és nem felügyelt tanulási módszerek keresztezésével, mely publikációjukban az anomália, mint klasszifikációs probléma, pontozásos rendszerrel került kiértékelésre. A klasszifikációt, az adathalmazban meghúzódó mintázat alapján automatikusan feltárt logikai szabályok mentén értelmezték, azaz, logikai következtetések révén vezették le az anomália meglétét a rendelkezésre álló rendszernaplók attribútumai alapján. Átlagosan 71%-ban volt képes a modell az adathalmaz osztályainak pontos meghatározására. A nem felügyelt gépi tanuló eljárások további kutatásokban is előfordulnak, Eskin et al. (2002) és Dua és Du (2011) kifejezetten a felügyelt és nem felügyelt módszerek hibridizálását ajánlják. Álláspontjuk szerint a felhasználói szokások gyakran változnak, mely újabb mintákat generál a gyarapodó naplóállományba, így a tanuló eljárások paramétereit és összetételét is dinamikusan változtatni kell, különben nagyon magas lesz az álpozitív találatok aránya. Eskin et al. (2002) szerint nem felügyelt módszerekkel lehetséges a felhasználói viselkedések csoportosítása idősorosan is a változások összhangjában. Mahoney és Chan (2003) hálózati forgalmat elemeztek idősoros adatokon LERAD[6] szabály alapú tanuló algoritmussal, s alacsony, azaz 50%-os teljesítményt értek el, mely az elkészített

[6] A LERAD célja, hogy olyan feltételes szabályokat találjon, amely képes váratlan eseményeket idősorosan azonosítani (Mahoney és Chan, 2003).

modell klasszifikációs erejét mérte. Véleményem szerint, a kutatásból levont eredményeket mérlegelve, a szabály alapú algoritmus nem bizonyul önmagában hatékonynak, így annak más módszerekkel való hibridizálása magasabb teljesítményhez vezethet. Leung és Leckie (2005) klaszterező eljárással (fpMAFIA módszerrel[7]) elemeztek hálózati forgalmat, ami gyengébbnek bizonyult, mint a felügyelt tanulási eszközök, SVM-mel 94.9%-os, KNN algoritmussal 89.5%, azonban egy módosított (hibridizált) klaszterező eljárással 97.3%-os teljesítményt értek el az anomália előrejelzésének pontosságára. Álláspontom szerint a kutatási eredmény létjogosultságot teremt a hibrid rendszerek fejlesztésének részletesebb kutatását illetően. Zhang és Zulkernine (2006/a) az anomália észlelés és aláírás felismerés komponenseit hibridizálták, azaz az első olyan példa, mely nem algoritmusokat, hanem magát a modell logikákat ötvözte. A modell véletlen erdő alapú gépi tanuló eljárással kísérli meg az anomália észlelését, és magasabb performanciát tudtak a modellel elérni, amit elsősorban az aláírás felismerés kiszűr, majd az anomália észlelő komponens magasabb pontossággal elemez. Zhang és Zulkernine (2006/b) egy másik kísérletükben különböző algoritmusokat alkalmaztak behatolás érzékelésre, melyek kifejezetten gyengén szerepeltek. SVM-mel 67%-os teljesítményt, míg KNN-nel csupán 11%-os eredményt értek el, melyből triviálisan következik, hogy az algoritmusok önmagukban nem voltak képesek a feladatok megoldására. Bhuyan et al. (2011) összegyűjtötték koruk behatolás érzékelő rendszereinek megoldásait, és arra jutottak, hogy az alkalmazni kívánt modellek nem hatásosak teljes mértékben az anomáliák felderítésére, továbbá, szubjektív véleményük szerint, azon inkrementális tanuló megközelítések tűnnek ígéretesnek, melyek az adatbányászat, neurális hálók és küszöbérték alapú elemzés kombinált összetevőivel rendelkeznek. Yu és Parekh (2016) kiemelik a nem felügyelt gépi tanuló módszerek egyedüli alkalmazásának hiányosságait. Mivel a nem felügyelt algoritmusok címkézetlen adatokkal operálnak, ezért azok teljesítményének mérése megkérdőjelezhető, mert nincs előzetes felülvizsgált adat arra vonatkozóan, hogy egy kiugró érték valójában anomália vagy sem. Kutatásukban Bayesian hibrid modellt alkalmaznak, és arra a következtetésre jutnak, hogy a hibrid modell teljesít a legjobban hálózati adatokban kutatott anomáliák észlelése tekintetében. Alkasassbeh (2018) hibridizált neurális háló alapú eljárással vizsgál túlterheléses támadás áldozatául esett hálózatokat, és arra az eredményre jut, hogy a hibrid modell képes volt 98.4% performanciát produkálni, úgy, hogy kizárólag álnegatív hibákat vétett a rendszer. Shekhar és Akoglu (2018) „ensemble[8]” módszerrel anomáliát kerestek (kiugró értékként

[7] A saját fejlesztésű módszer lényege, hogy klaszterek halmazát határozzák meg az input adatokban frekvencia minták alapján (Leung és Leckie, 2003)

[8] Az ensemble módszerek, olyan tanuló algoritmusok, melyek alapvetően több algoritmus együttes predikciós

kezelve) bizalmas információ védelmére, mint pl. twitter bot detekció, e-mail spam felismerés stb. közel 90%-os teljesítményt elérve az osztályozás előrejelzésére. Sharma et al. (2018) közösségi hálók információ hozzáférési kontrolljait tanulmányozták neurofuzzy rendszerekkel, mely a neurális hálók és fuzzy logika hibrid modellje, 99% feletti anomália detekciós teljesítményt elérve, mely ráerősít a hibrid modellfejlesztés létjogosultságára. Vállalatirányítási rendszerből exportált naplófájlok alapján történő anomália észleléssel kapcsolatos publikációt nem találtam.

3.3.3. H2 hipotézis felismerését/racionalitását támogató szakirodalom

H2: A kutatásban elkészítendő egyes hibrid modellek képesek az előre definiált normális tevékenységek között meghúzódó összefüggések matematikai leképezésére, s ez által az anomáliák feltárására, kísérletekkel alátámasztott optimális parametrizálásával azok megtanulására a túlilleszkedést minimalizálva.

A felügyelt gépi tanuló rendszerek feladata az előre meghatározott értékű célváltozók alapján az egy osztályba sorolható rekordok közötti mintázatok felismerése. Azonban az, hogy a gépi tanuló alkalmazások bármely adathalmazra képesek ezt elvégezni közel sem triviális. Szakirodalmi kutatásomban bebizonyosodott, ahogy azt a következőkben leírt kutatási eredmények is igazolják, hogy egyes modellek képesek jobban egy adott adathalmazt megtanulni, mint a többiek, továbbá, arra is van példa, hogy egy tanuló algoritmus egyáltalán nem volt képes az adatok között meghúzódó összefüggések leképezésére pl. Zhang és Zulkernine (2006/b) esetében, ahogy az az előző alfejezetben kifejtésre is került. Továbbá, a tanulás foka és mértéke is sok esetben megkérdőjelezhető. A rendelkezésre álló adathalmazok tartalmazhatnak olyan változókat, melyek véletlen zajokkal vannak feltöltve, így fennáll a kockázat a zajok megtanulására is, így a rendszer álösszefüggéseket produkálhat. Az is egy lehetséges kockázat, hogy a kiválasztott tréning adathalmazt az algoritmus „*memorizálja*”, azaz annak predikciós ítélőképessége kizárólag a tréning halmaz elemeire lesz magas teljesítménnyel, míg az eddig nem látott objektumokon az alulperformál. Hipotézisemben állítottak szerint azt kívánom bizonyítani, hogy az általam elkészített hibrid modell elsősorban képes az összefüggések matematikai leképezésére, továbbá, paramétereit lehet optimalizálni,

erejét aggregálják, ezzel javítva az algoritmus általánosító képességét (Raschka, 2015).

ráadásul úgy, hogy az ellenálló legyen a külső zajoknak, és ne következzen be a túltanulás, azaz a tréningadatok kizárólagos memorizálása.

Ye et al. (2001) állítja, hogy egyetlen audit esemény nem elegendő az anomália észlelésére ezért szükséges az egyes események szekvenciális feldolgozása, azok egymásra épüléséből adódóan a mintázat csak idősoros modellek alkalmazásával nyerhető ki. Rejtett Markov Modellt és Döntési Fákat alkalmaznak Sun SPARC munkaállomásokról gyűjtött adatokon 250 audit eseményből, melyből 1613 normális tevékenység 526 pedig anomália. Az eredményük, hogy a tesztadatok növekedésével először exponenciálisan, majd lineárisan növekednek az álpozitív találatok. A kutatás arra enged következtetni, hogy anomália detektálása esetén a gépi tanuló rendszerek akkor képesek magas performancia elérésére, ha azok képesek az adatok közötti idősoros események kezelésére, ezáltal az egymásra épülő események között meghúzódó összefüggések leképezésére. Liao és Vemuri (2002) KNN alapú klasszifikációs rendszert fejlesztettek anomália észlelésre, melyben a rendszerhívások szövegbányászati megközelítését alkalmazták. A módszer előnye, hogy nincs szükség sokrétű kísérletezés elvégzésére, mivel a rendszert csak minimálisan kell paraméterezni az algoritmus sajátosságai miatt. 75%-os találati arányt értek el új támadások detektálása esetén, míg az ismert anomáliára utaló viselkedéseket 100%-ban meg tudták állapítani. A leszűrt konklúzió, hogy a KNN eljárás kevésbé mutat kitettséget a túlilleszkedésre, tehát érdemes, ezen előnnyel rendelkező komponenseit hibrid alkalmazásban felhasználni, azonban új anomáliák észlelése esetén alacsony teljesítményt mutatott, tehát predikciós ereje a hivatkozott kutatásban megkérdőjelezhető. Hu et al. (2003) SVM alapú megoldást javasoltak, összehasonlították azt a KNN algoritmussal és 18%-al magasabb teljesítményt értek el anomáliák osztályozásában. Minden esetben az álpozitív találatok száma 1% alatt volt, tehát a kísérlet megmutatta, hogy az SVM alkalmazás predikciós ereje anomália észlelés tekintetében jobban teljesít. Chen et al. (2005) szekvenciális rendszerhívásokat vizsgáltak SVM-mel és neurális hálókkal, a neurális hálók általánosító ereje gyengébbnek bizonyult tanulmányukban a tesztadatokon. Az SVM-ek kevésbé komplex algoritmusok, az elvégzett kísérletben a hibafüggvény globális minimumának megtalálása egyszerűbb volt a neurális hálóénál, mely arra enged következtetni, hogy a komplexebb algoritmusok alkalmazása nem feltétlenül a helyes irány a tanulási folyamatban. Soule et al. (2005) nagyméretű vállalati hálózati forgalmat elemeztek, melyben a normális forgalmat Kálmán-szűrővel[9] szűrték le. A kielemzésre ROC görbét használtak, mely az álpozitív és álnegatív eredmények közötti kapcsolatot is képes kimutatni. Kutatási

[9] A Kálmán-szűrő egy rekurzív becslő eljárás, mely változó rendszerek állapotáról ad predikciót (Kleemen, 1995).

eredményükre hivatkozva, az egyik konklúzió, hogy a ROC görbe elkészítése megfelelő jóság metrika, mert képes az álpozitív és álnegatív eredményeket közösen kezelni, ezáltal egy jól használható teljesítmény-mutatóként alkalmazható. A másik konklúzió, amelyet a szerzők ki is emelnek, hogy a tanuló algoritmusoknak egyik fontos feladata az anomália észlelésen túl, az anomália jelentésének gyorsasága, így olyan modelleket célszerű alkalmazni, melyek kiesési idő nélkül képesek az anomália azonosításának tényét közölni. Lakhina et al. (2005) Entrópia[10] metrikát alkalmazva felügyelet nélküli tanulással detektáltak anomáliát, mely alkalmasnak bizonyult a kisebb frekvenciával rendelkező támadások és az új támadások ellen egyaránt. Li et al. (2006) online hálózati monitorozás alapú anomália észlelést végeztek az attribútumokra vonatkozó dimenziócsökkentéssel, magas találati arányt értek el, 200-ból csupán 1 volt téves riasztás. A kutatás rámutatott, hogy a gépi tanulásban az adattranszformáció szerepe is jelentős a performancia növelésében, nem kizárólag a tanuló algoritmus tanulási folyamata. Abraham et al. (2007) az anomália észlelés szakirodalmában egy teljesen új megközelítéssel, genetikus programozással kísérleteztek ismert támadási mintákat felismerni, átlagosan 95%-os pontossággal sikerült a helyes klasszifikáció. Egyik konklúziója a kutatásnak, hogy a módszer nagy részben használható online megoldásként, mivel gépi kód szinten lehetséges az implementáció, másik, hogy a genetikus programozás alapvetően nem gépi tanuló eljárás, így a tanulási mechanizmusok teljesítményének fokozása érdekében, egyrészt érdemes megfontolni más mesterséges intelligencia eljárások alkalmazását, másrészt, véleményem szerint, a modellezést nem feltétlenül szükséges gépi tanuló módszerekre kizárólagosan szűkíteni, mivel más algoritmusok is rendelkezhetnek olyan komponensekkel, mely alkalmazható az anomáliaészlelésben és a tanulásban. Su et al. (2009) valós idejű fuzzy szabályozó alapú anomália észlelést javasol túlterheléses támadások ellen, ahol a döntések meghozatala 2 másodpercenként zajlik. Magas szintű performancia érhető el javaslatuk szerint fuzzy szabályozókkal. Heard et al. (2010) nagyméretű dinamikus hálózatokat vizsgált Bayesian gépi tanuló eljárással, kiemelendő, az algoritmus párhuzamosíthatósága miatt lehetséges a gyors futásidő (2 másodperc), mely lehetségessé teszi az óriáshálózatok gyors kivizsgálhatóságát anomália detektálás tekintetében.

Az előzőleg kiértékelt kutatási eredmények javarészt hálózati forgalom alapon és rendszerhívások által kísérelték meg az anomália észlelését az információrendszerekben, azonban az előző években kb. 2011-2012-től kezdődően az anomália detektálásának kutatási

[10] Az Entrópia metrika alkalmas egy adott attribútum információtartalmi hasznosságának becslésére (Raschka, 2015)

területei terjeszkedni látszódnak. A közösségi hálók, a mobileszközök, az apró internetre csatlakoztatható szerkezetek elterjedésével megnőtt az igény, és ezzel együtt a tudományos társadalom kutatási hajlandósága, hogy egyéb területeket is vizsgáljon, mely az anomália felderítésével és elemzésével foglalkozik. Takahashi et al. (2014) közösségi hálókon résztvevő felhasználók közötti említések anomáliáját kutatják és valószínűségi modellel igazolják, hogy egyes felkapott témák népszerűségét már annak korai szakaszában detektálni lehet, így az anomália észlelés megfogalmazását egy más közegbe emelik. Kaur et al. (2013) összegzik a tudományterületen megjelent főbb kutatási irányokat, és arra jutnak, hogy anomália észlelésben a legalkalmasabbnak bizonyuló algoritmusok közé sorolhatók a neurális hálók, fuzzy logikán alapuló klasszifikáló algoritmusok, evolúciós megoldások és a Bayesian hálózatok. Navaz et al. (2013/a) felhő alapú rendszerek túlterheléses támadásait vizsgálják, és többszálú behatolás-érzékelő rendszer alkalmazását javasolják, mely képes a megnövekedett adatok (Big Data) hatékony feldolgozására. Egy másik tanulmányukban információrendszerek szolgáltatás minőségét vizsgálják, és az anomáliát, mint szolgáltatásban bekövetkezett minőségi ingadozást kezelik (Navaz et al., 2013/b). Abdelrahman et al. (2013) mobilkörnyezetben vizsgálják az anomáliát, amely a mobileszközöket támadó kártékony kódok és támadások formájában testesül meg, egy új keretrendszert javasolnak implementálásra, melyet alkalmasnak találnak a mobileszközöket érintő újszerű kitettségek mitigálására. Siddiqui et al. (2015) hangsúlyozzák, hogy az anomália észlelés egyik legnagyobb problémája, hogy az anomáliát felderítő alkalmazások nem képesek magyarázatot adni, miért tekintettek egy tevékenységet anomáliának, csak az anomália tényét adja tudtára az elemzőnek. Ezen probléma feloldására egy általuk készített keretrendszert ajánlanak („Sequential Feature Explanation”), mely magyarázattal szolgál az anomáliát tartalmazó rekordok attribútumaira vonatkozóan, megkönnyítve az anomália értékelését. A kutatásban leszűrhető konklúzió, hogy üzletileg olyan alkalmazásokat érdemes fejleszteni, mely lehetővé teszi az anomáliák bekövetkezésének magyarázatát, nem kizárólag az anomália tényére hívják fel a figyelmet. Zhai et al. (2016) mélyített neurális hálókat alkalmaznak, mely a legelső publikációnak számít „deep learning” megközelítésben anomália észlelésre a szerzők által leírtak alapján. Három különböző típusú hálót fejlesztettek: Rekurrens, Konvolúciós és Teljes Kapcsolattal rendelkező neurális hálókat. Tanulmányukban arra jutottak, hogy a mély architektúrával rendelkező neurális hálóknak van létjogosultságok az anomália észlelés területén, mivel jobb általánosító képességgel rendelkeznek a keskenyebb hálóknál, azaz, jobban elkerülik a túltanulást. Richardson et al. (2018) nem felügyelt eszközökkel kíséreltek meg anomália észlelést, melyben úgy értelmezték a hálózati forgalmat, mint gépek közötti kommunikációt, ezért egy ez idáig egyedi módon,

természetes nyelvi feldolgozást alkalmaztak. Nyelvi feldolgozásra egy modell a rekurrens neurális hálók, mely alkalmas idősoros adatok elemzésére is. 3.1 millió támadást tartalmazó címkézett adathalmazt dolgoztak fel, a neurális háló modellt frekvencia alapú modellel összehasonlítva arra jutottak, hogy az utóbbi jobban alkalmazható volt anomália észlelésre, mert magasabb teljesítményt eredményezett, azonban a háló modellel kapcsolatban nem volt törekvés a hiperparaméterek optimális megtalálására. Tandiya et al. (2018) mélyített neurális hálókkal kerestek anomáliát wi-fi hálózatokban, az álpozitív találatok aránya 3% körüli, összesen 90%-os találati arányt produkált a modell, azonban nem alkalmaztak különböző gépi tanuló technikákat magasabb teljesítményt célozva pl. adattranszformáció, hiperparaméter hangolást sem. Golomb et al. (2018) teljesen egyedi módon IoT[11] eszközök sérülékenységét vizsgálták Blokklánc[12] technológiával, kb. 70%-os teljesítmény értek él.

3.3.4. H3 hipotézis felismerését/racionalitását támogató szakirodalom

H3: A kutatásban elkészítendő hibrid modellek a többváltozós matematikai statisztikai módszerekkel szemben meghaladják az előzetesen definiált jóság metrikák aggregált szintjét.

Az anomália észlelés információ-biztonsági alkalmazásai 1987-től kezdtek teret hódítani Denning (1987) behatolás-érzékelés definícióját követően. Mivel a gépi tanulás és mesterséges intelligencia módszerei nem terjedtek el széles körben akkoriban ebben a témakörben, többek között köszönhető annak, hogy a gépi tanulásban alkalmazott algoritmusok magas hardver performancia igénnyel rendelkeznek, a kezdeti kutatások főleg a matematikai statisztika módszereivel igyekeztek az anomáliák felderítésében. Kutatásomban, célom a hibrid modellek fejlesztése mellet, ugyanazzal az adathalmazzal a kísérleteket elvégezni többváltozós statisztikai módszerekkel is, feltételezésem szerint a gépi tanuló eljárások magasabb teljesítmény elérésére lesznek képesek a többváltozós módszerekkel szemben, mivel egyes algoritmusok pl. neurális hálók képesek nemlineáris leképzéseket is hatékonyan approximálni.

Smaha (1988) rendszerhívások[13] naplófájljait elemezte a Haystack alkalmazással, mely statisztikai módszereken alapuló megközelítést alkalmazott. Hasonlóan Smaha-hoz, sok más

[11] Az IoT (Internet of Things) eszközök, olyan technológiai újítások, melyek lehetőséget adnak a hétköznapi fizikai eszközök csatlakoztatására az internetre, ezáltal azok távolról is irányítható válnak (pl. intelligens hűtő, mosógép stb., melyeket számítógépről is el lehet érni, monitorozni, változtatni beállításukat) (Kurniawan, 2016).
[12] A Blokklánc (Blockchain) egy olyan elosztott adatbázison alapuló technológia, melynek célja bizonyos tranzakciókban (kriptofizetés, szerződéskötés stb.) a letagadhatatlanság biztosítása adablokkokból álló lista és kriptográfiai eszközök alkalmazása által (Bashir, 2017).
[13] A rendszerhívások az operációsrendszer és futtatott szoftverek közötti kommunikáció eszköze (Smaha, 1989).

kutató is a rendszerhívásokból indult ki, hogy a hálózat, illetve belső szoftverek elleni támadásokat észlelje. Ide tartozik, többek között, Ghosh et al. (1998), Ye et al. (2001) és Liao és Vemuri (2002). Warrender et al. (1999) különböző behatolás érzékelő modelleket épített, szabály-alapú algoritmusok és Rejtett Markov Láncok felhasználásával, mely szakirodalmi kutatásom alapján, az első olyan kísérlet volt, ahol gépi tanulás kategóriájába tartozó algoritmusok magasabb teljesítményhez vezettek a hagyományos statisztikai eljárásoknál. A 90-es évek végére a többváltozós statisztikai módszerek közé beékelődtek az intelligens megoldások, mely szakirodalmi kutatásom alapján, fokozatosan egyre nagyobb teret nyertek és kezdték leváltani az anomália kutatásában eddig alkalmazott paradigmákat. Első kiemelt szereplője ezen váltásnak Lee et al. (1999), historikusan gyűjtött hálózati forgalmat tanulmányoztak és egy adatbányászati keretrendszert javasoltak az anomália detektálásával foglalkozó kutatók részére. Ghosh et al. (1999) szoftver viselkedéseket vizsgáltak profilozással Elman hálót alkalmazva, kutatásuk eredménytermékeként előállított modell magas álpozitív arányt mutatott. Ye et al. (2002) behatolás érzékelő modellt terveztek, mely többváltozós statisztikai módszereken alapszik. 4 napig folyamatosan gyűjtött naplófájlokat elemeztek, amelyben csupán 0.42% volt az anomália szerepe. Ez magas teljesítményhez, de magas álpozitív találatokhoz vezetett. A publikáció általam leszűrt konklúziója, hogy a magas találati arány még nem jelenti a rendszer jóságát, ahhoz szofisztikáltabb metrikákat kell kialakítani. Feinstein et al. (2003) statisztikai módszereket alkalmaznak hálózati forgalom elemzésére, mely célja a túlterheléses támadások kiszűrése. Az erre épített modell magas megbízhatósággal működik és képes az anomáliát kiszűrni, ez azonban annak is köszönhető, hogy túlterheléses támadások esetén a cél az adott hálózat támadása, így folyamatos terhelések miatt a forrás IP címe ismertté válik, mely gyakoribb kéréseket intéz a céleszköz felé, így ez gyorsan észrevehető és blokkolható. Lakhina et al. (2004) hálózati csomagok elemzését végezték subspace módszerrel, mely többváltozós statisztikai elemzésen alapszik. Az adattranszformációt faktor-elemzéssel végezték és a cél a kártékony tevékenységek, hálózati hibák és a felhasználói viselkedések különválasztása volt, 8%-os álpozitív találattal, mely álláspontom szerint magasnak tekintendő. Ranjan és Sahoo (2014) módosított K-közép klaszterező eljárást ajánl különböző távoli külső támadások detektálása érdekében, melyben egyes támadás típusokat (szolgáltatásmegtagadás, távoli rendszertámadás, sérülékenység kihasználás, port szkennelés) különböző modellekkel vizsgál. A javasolt módosított klaszterező algoritmus teljesítménye rendre 96%, 90%, 71% és 70%, mely a hagyományos klaszterező eljárásokon felül teljesít. Fanaee-T et al. (2014) IP/TV hálózati anomália felderítésére többdimenziós megoldást javasol, mivel a hagyományos klaszterező és regressziós eljárások, továbbá a faktor-analízis kizárólag

kétdimenziós modellek esetén működtethető. Kiemeli, hogy a hálózati anomáliákat a felhasználó, attribútum, idő mentén érdemes megvizsgálni, mely 3 dimenziós térbe helyezi a detektálás szükségességét. Saját kutatásban (Barta és Pitlik, 2018) diszkriminancia-elemzéssel osztályoztunk startup szervezetek lehetséges felvásárlási karakterisztikáit, amelyet elvégeztünk neurális háló alapú modellel és hasonlóságelemzéssel, az osztályozás pontossága átlagosan 22%-kal volt több a gépi tanuló módszerek javára, mely létjogosultságot teremt a hipotézisben állítottak alapos kivizsgálásához.

3.3.5. H4 hipotézis felismerését/racionalitását támogató szakirodalom

H4: A kutatásban elkészítendő hibrid modellekben az input jelként felhasznált adatokat lehetséges olyan szinten transzformálni és a modellekbe lehetséges olyan védelmi mechanizmusokat beépíteni, hogy azok ellenállók legyenek egyes külső szándékos manipulációkkal szemben, melyek célja az anomália-detektáló rendszer félrevezetése, vagyis az anomáliák normális magatartásként történő feltüntetése.

A külső és belső támadók tudatában lehetnek, hogy az anomáliák észlelésére a támadni kívánt célpont különböző eszközöket alkalmaz, így feltételezhetjük, hogy a védelmi mechanizmusok kikerülése és azok támadása elsődleges szempontot élvezhet a csalás elkövetése előtt. Míg külső támadók a rendszer védelmét kevésbé ismerhetik és annak hatástalanítását tarthatják szem előtt, belső támadás esetén megvan a kockázatot, hogy egy adott munkavállaló, aki már ismeri az alkalmazást, esetleg hozzáférése is van, szándékosan félrevezeti a rendszer szűrő logikáját. Természetesen fennáll a veszély a nem szándékos manipulációnak is, amennyiben a rendszerhez hozzá kevésbé értő felhasználók is jogot kapnak, vagy menedzsment nyomásra a határidők betartása végett nem megfelelően kerül letesztelésre és hibák maradnak az alkalmazásban. Célom a hipotézissel annak bizonyítása, hogy lehetséges az elkészítendő hibrid modellekbe olyan védelmi mechanizmusokat építeni, mely az adatok véletlen, vagy szándékos manipulációja esetén is képes megfelelő performanciát biztosítani, felismerve és kiszűrve az adatokra irányuló megmásítás tényét.

Ghosh et al. (1998) egyszerű architektúrájú 3 rétegű neurális háló modellt prezentáltak anomália észlelésre, 30-szoros teljes adathalmaz feldolgozás és 30 különböző hálósúlyozással. Mivel csak normális viselkedést leíró adatokkal dolgoztak, így az anomália észlelésére véletlenszámokat generáltak, melyet az algoritmusok átlagosan 85%-os teljesítménnyel voltak képesek kiszűrni, azonban furcsa mód, kutatási eredményük nem zárult álpozitív találattal, csak

álnegatívval. Ez az eredmény arra enged következtetni, hogy ezen korai modell, szem előtt tartva a kor kutatási eredményeit, nem megfelelően diverzifikált tesztadathalmazt alkalmazott és a véletlenszámokkal feltöltött objektumok annyira eltérő eredményeket produkáltak, hogy az semmilyen formában sem volt hasonló egy normális viselkedéshez és valós személy által elkövetett tevékenységek sorozatához. Ahogy az fentebb is megemlítésre került, az a meglátásom, hogy külső támadások esetén gyakorta a támadó is igyekszik hasonló magatartást követni adott rendszer felhasználóihoz, hogy a beépített rendszer-automatizmusok ne kezeljék a tevékenységet kiugró eseményként, így elsődlegesen a normális tevékenységet leíró rekordokat szükséges olyan szinten transzformálni, hogy a szándékos manipuláció ténye azonnal azonosíthatóvá váljon. Mahoney and Chan (2002) újszerű, eddig nem ismert támadási formákat próbáltak felderíteni hálózati forgalom elemzése alapján. A feltételezésük, hogy a támadók a szoftverek és egyéb biztonsági konfigurációk sérülékenységeit célozzák, így a szokatlan körülményekre helyezik a hangsúlyt. Szabályalapú megoldást ajánlanak magas számú attribútumok felhasználásával, mellyel finomítani lehet a szabály alapú keresést. Véleményem szerint, egyrészt, hosszú távon alkalmazva, ez a módszer túltanuláshoz vezet, mely a gépi tanuló rendszer általánosító-képességének romlásában mutatkozik meg, másrészt, a gép tanuló rendszerek konfigurációjának elrejtésére, titkosítására van szükség, mivel a támadók nem csak az adatokat manipulálhatják, hanem a gépi tanuló rendszer paramétereit is. Kloft és Laskov (2010) rávilágítanak, hogy a gépi tanuló eljárások önmagukban is információ-biztonsági veszélynek vannak kitéve, amennyiben kártékony, adatot meghamisító kód kerül a rendszerbe. Szakirodalmi kutatásom alapján, időrendben, ez az első olyan tanulmány, mely a gépi tanuló rendszerek biztonsági kérdéseit vitatja. Kutatásukban lineáris és RBF kernel eljárás alapú eszközökkel vizsgálnak sérülékenységet tanuló rendszeren, mely átlagosan 95%-os teljesítményt produkált, vagyis ebben az arányban sikerült kiszűrniük automatizmusokkal, vajon egy adott komponenst támadták vagy nem.

3.3.6. Az alfejezet összefoglalása

A fejezetben tárgyalt szakirodalmi kutatás az alábbiakban támogatja kutatási céljaim eléréseit:

- Az anomália észlelés korai kutatásai legfőképp a külső támadásokra összpontosítanak hálózati forgalom és rendszerhívásokat elemezve, azonban kutatás, mely a szervezet munkatársai által generált tevékenységek alapján elemezné az anomália megjelenésének sajátosságait, nem elérhető.

- Az előző években, kb. 2011-től kezdve az anomália észlelés területe terjeszkedni látszik, melyben megjelenik, többek között, a közösségi háló és IoT eszközökkel kapcsolatos anomáliák felderítése is, azonban továbbra sincs előtérben a vállalaton belüli információbiztonságot érintő incidensek kiszűrésére modell és a jogosulatlan hozzáférésekkel kapcsolatban is csak minimális a kutatások száma.
- A **H1** hipotézisem racionalitását alátámasztják a hibrid modellezés témakörében megjelent növekvő tanulmányok és a tradícionális modellek hibridizációjával kapcsolatos modellkombinációk száma, azonban a hibrid modellalkotás objektív leképezését a kutatók nem kísérelték meg ez idáig.
- A **H2** hipotézisem racionalitását támogatják a kielemzett tanulmányok, így véleményem szerint az elkészítendő hibrid modellek, alapul véve az eddigi kutatási eredményeket és javaslatokat, alkalmasak lesznek a gépi tanulás matematizálására és optimalizálására a túlilleszkedést elkerülve a belső anomáliaészlelés területén is.
- A **H3** hipotézisem racionalitását alátámasztja az a trend, hogy az anomália észlelés területén a gépi tanuló módszerek egyre nagyobb népszerűséggel jelennek meg a kutatási témában a statisztikai módszereknél, illetve, hogy a gépi tanuló rendszerek egyes fajtái hatékonynak bizonyulnak nemleniáris leképezésekre és függvény-approximációra.
- A **H4** hipotézisem racionalitását szolgálja, hogy több kutató is felismerte, hogy a gépi tanuló rendszerek sérülékenyek, ezért szükséges azokat külső és belső támadóktól egyaránt védeni.

4. Összefoglalás

A Disszertációs Rész elsődleges céljául tűzte ki, hogy bemutassa a kutatói témához kapcsolódó szakirodalmakat, kutatási eredményeket, alkalmazni kívánt megközelítéseket és módszereket, melyek elméleti alapot, iránymutatást és bizonyosságot szolgáltattak, egyrészt, hogy az információbiztonságra irányuló anomáliák felfedezésének területével kiemelt módon szükséges foglalkozni, mivel az egy olyan modern szervezeti probléma, mely az üzleti automatizáció, digitalizáció és információs társadalom által generált negatív extern hatásban jelenik meg, másrészt, a feldolgozott szakirodalmak alátámasztást nyújtottak, hogy a felállított hipotézisek kellően racionálisak ahhoz, hogy azok objektív levezetése újszerű eredmények produkálására, innovációra, tehát tudományos felismerésre legyenek alkalmasak.

A dokumentumban felvázolásra került a kutatói probléma, mely keretei között ismertettem a téma aktualitását, jelenkori kihívásokat és az általam szándékozott megoldási javaslatokat. A továbbiakban a szakirodalmi kutatást három részre bontva, bemutattam az információbiztonsággal, mesterséges intelligenciával és anomália észleléssel kapcsolatos definíciókat, megállapításokat és kutatási eredményeket. Az első alfejezet alapján leszűrt konklúzió, hogy az információbiztonsági incidensek felderítése érdekében a mindenkor rendelkezésre álló információrendszerek által produkált naplóállományokból érdemes kiindulni, mivel azok képesek a rendszerekben eszközölt tevékenységek, a felhasználók elektronikus lábnyomainak leképezésére és elmentésére, így azok kielemzésével lehetőség adódik az információbiztonsági incidensek, azaz anomáliák detektálására. Mivel a naplóállományok manuális ellenőrzése erőforráskorlátokba ütközik, így automatizált megoldásokra van szükség, melyek képesek felismerni az anomália tényét. A második szakirodalmi alfejezetben kitisztázódott, hogy a mesterséges intelligencia képes olyan modelleket nyújtani, melyek tanulási mechanizmusaik révén historikus adatokat felhasználva különbséget tudnak tenni különböző tevékenységek között. A mesterséges intelligencia egyik alága a gépi tanulás alkalmazott eljárásai hatékonyan alkalmazhatóak minták felismerésére és osztályozására, azonban hibridizált eszközökkel, véleményem szerint, magasabb teljesítmény érhető el az anomália detektálására vonatkozóan, így a mesterséges intelligencia más területének bevonása is indokolt lehet pl. evolúciós optimalizálás, fuzzy logika stb. Harmadik alfejezetben összegyűjtöttem az anomália észlelésére irányuló kutatásokat és a hipotéziseim tükrében azok racionalitását vizsgáltam, mely részemre alátámasztó evidenciát szolgáltatott, hogy a felismert hipotézisek levezetése újszerű eredmények létrehozására alkalmasak.

5. Irodalomjegyzék

1. Abah J. – Waziri O. V. – Abdullahi M. B. – Arthur U. M. – Adewale O. S. (2015): A Machine Learning Approach to Anomaly-based detection on Android Platforms. International Journal of Network Security & Its Applications (IJNSA). VII. évf. 6. sz. 15-35. p.
2. Abdelrahman O. H. – Gelenbe E. – Gorbil G. - Oklander B. (2013): Mobile Network Anomaly Detection and Mitigation: The NEMESYS Approach. Information Sciences and Systems. 429-438 p.
3. Abraham A. – Grosan C. – Martin-Vide C. (2007): Evolutionary design of intrusion detection programs. International Journal of Network Security. 4(3): 328-339. p.
4. Alkasassbeh M. (2018): A Novel Hybrid Method for Network Anomaly Detection Based on Traffic Prediction and Change Point Detection. Journal of Computer Science. XIV. évf. 2. sz. 153-162 p.
5. Ágoston M. - Szluka, E. (1989): Tudni vagy nem tudni! Gondolatok egy nemzeti információpolitikához. Budapest, Műszaki könyvkiadó, 178 p.
6. Araújo B. A. (2016): Semantic Information and Artificial Intelligence. In: Müller V. C. (szerk.:) Fundamental Issues of Artificial Intelligence. Svájc, Springer International Publishing, 572 p., 129-140 p.
7. Barta G. – Pitlik L. (2018): Startup felvásárlások multikulturális hátterének elemzése, avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez. In: Farkas Attila (szerk.): A gazdaság kulturális szerkezete. Cultural Structure of Economic . 240 p. Konferencia helye, ideje: Gödöllő , Magyarország , 2017.05.12. Gödöllő: Szent István Egyetemi Kiadó, 15-37. p.
8. Baesens B. – Vlasselaer V. V. – Verbeke Q. (2015): Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection. USA, Wiley and SAS Business Series, 359 p.
9. Bashir I. (2017): Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks. Birmingham, Packt Publishing, 540. p.
10. Behandish M. – Wu Z. Y. (2014): Concurrent Pump Scheduling and Storage Level Optimization Using Meta-Models and Evolutionary Algorithms. Procedia Engineering, 70:103-112. p.
11. Bellman R. (1978): An introduction to artificial intelligence: can computers think? San Francisco, Boyd & Fraser Publishing Company, 146 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
12. Bhuyan M. H. – Bhattacharyya D. K. – Kalita J. K. (2011): Survey on Incremental Approaches for Network Anomaly Detection. International Journal of Communication Networks and Information Security (IJCNIS). III. évf. 3. sz. 226-239 p.
13. Boda Gy. – Juhász P. – Stocker M. (2009): A tudás mint termelési tényező. Köz-gazdaság. IV. évf. 3. sz. 117-132. p.
14. Bodon F. (2010): Adatbányászati algoritmusok. Budapest, Free Software Foundation, 278 p.
15. Borgulya I. (1998): Neurális hálók és fuzzy-rendszerek. Budapest – Pécs, Dialóg Campus Szakkönyvek, 226 p.

16. Capurro, R. (1991): What is Information Science for? Conception of Library and Information Science konferencián prezentált tanulmány. Finnország, Tampere, 1991.
17. Charniak E. – McDermott D. (1985): Introduction to Artificial Intelligence. Massachusetts, Addison-Wesley, 701 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
18. Chen W. H. – Hsu S. H. – Shen H. P. (2005): Application of SCM and ANN for intrusion detection. Computers and Operations Research. 32:2617-2634.
19. Combs A. (2016): Python Machine Learning Blueprints: Intuitive data projects you can relate to. 1st Edition, Kindle Edition. Birmingham, Pack Publishing, 332 p.
20. Curtis P. – Carey M. (2012): Risk assessment in practice. Deloitte &Touche LLP, 19 p.
21. Damelin S. B. – Gu Y. – Wunsch D. C. – Xu R. (2015): Fuzzy Adaptive Resonance Theory, Diffusion Maps and their applications to Clustering and Biclustering. Mathematical Modelling of Natural Phenomena. X. évf. 3. sz. 206-211. p.
22. Denning D. (1987): An intrusion-Detection Model. IEEE Transactions on Software Engineering. 13(2): 118-131. p.
23. Diamant E. (2017): Advances in Artificial Intelligence: Are you sure, wea re on the right track? Transactions on networks and communications. V. évf. 4. sz. 23-30. p.
24. Dilek S. – Cakir H. – Aydin M. (2015): Applications of artificial intelligence techniques to combating cyber crimes: a review. International Journal of Artificial Intelligence & Applications (IJAIA). VI. évf. 1. sz. 21-39 p.
25. Dua S. – Du X. (2011): Data Mining and Machine Learning in Cybersecurity. USA, Taylor and Francis Group, 223 p.
26. Eskin E. – Arnold A. – Prerau M. – Portnoy L. – Stolfo S. (2002): A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data. Applications of Data Mining in Computer Security. 1-20. p.
27. Farkasné F. M. – Molnár J. (2007): Közgazdaságtan I. Mikroökonómia. Debrecen, Debreceni Egyetem Agrár- és Műszaki Tudományok centruma. Agrárgazdasági és Vidékfejlesztési Kar, 274 p.
28. Fawcett T. – Provost F. (2013): Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. USA, O'Reilly Media, 414 p.
29. Feinstein L. – Schnackenberg D. – Balupari R. – kindred D. (2003): Statistical approaches to Ddos attack detection and response. Proceedings of DARPA Information Survivability Conference and Exposition. 303-314. p.
30. Forgó S. (2011): A kommunikációelmélet alapjai. Eger, Médiainformatikai Kiadványok, 172 p.
31. Fülöp G. (1996): Az információ. 2. bővített és átdolgozott kiadás. Budapest, Eötvös Loránd Tudományegyetem, Könyvtártudományi – Informatikai Tanszék kiadványa, 236 p.
32. Ghosh A. K. – Wanken J. – Charron F. (1998): Detecting anomalous and unknown intrusions against programs. Proceedings of the 1998 Annual Computer Security Applications Conference (ACSAC). 1-9. p.
33. Ghosh A. K. – Scwartbard A. – Schatz M. (1999): Learning program behavior profiles for intrusion detection USENIX Association. Proceedings of the 1st USENIX Workshop on Intrusion Detection and Network Monitoring. 1-13. p.
34. Harkevics A. A. (1960): O cennoszti informarcii. Problemu kibernetiki. Vüp. 4. Moszkva. Idézi: Fülöp, 1996.
35. Hastie T. – Tibshirani R. – Friedman J. (2009): The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second Edition. New York, Springer Science, 745 p.

36. Haugeland J. (1985): Artificial Intelligence: The Very Idea. Massachusetts, MIT Press, 299 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
37. Heard N. A. – Weston D. J. – Platanioti K. – Hand D. J. (2010): Bayesian Anomaly Detection Methods for Social Networks. The Annals of Applied Statistics. IV. évf. 2. sz. 645-662. p.
38. Huckeling G. (2014): Mastering Machine Learning With scikit-learn. Birmingham, Pack Publishing, 238 p.
39. Hu W. J. – Liao Y. H. – Vemuri V. R. (2003): Robust support vector machines fora nomaly detection in computer security. Proceedings of the International Conference on Machine Learning. 282-289. p.
40. Kaur H. – Singh G. – Minhas J. (2013): A Review of Machine Learning based Anomaly Detection Techniques. International Journal of Computer Applications Technology and Research. II. évf. 2. sz. 185-187. p.
41. Kiss T. J. (2016): A tudásgazdaság jellemzői Magyarország vonatkozásában. International Journal of Engineering and management Sciences (IJEMS). I. évf. 1. sz. 1-11 p.
42. Kloft M. – Laskov P. (2012): Security Analysis of Online Centroid Anomaly Detection. Journal of Machine Learning Research. 13:3681-3724. p.
43. Komenczi B. (2011). Információelmélet. Eger, Médiainformatikai Kiadványok, 131 p.
44. Kovács Z. (2009): Szimulációs eszközök és megoldások műszaki és gazdasági rendszerekben. II. „Innováció az egyetemi képzésben és kutatásban" Jubileumi Tudományos Konferencia. Balatonvilágos, előadásjegyzet.
45. Kurniawan A. (2016): Smart Internet of Things Projects. Birmingham, Pack Publishing, 258. p.
46. Kurzweil R. (1990): The age of intelligent machines. Massachusetss, MIT Press, 565 p.
47. Luo J. – Bridges S. (2000): Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection. International Journal of Intelligent Systems. 15(8):687-703. p.
48. Lakhina A. – Crovella M. – Diot C. (2004): Characterization of network-wide anomalies in traffic flows. Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement. 201-206. p.
49. Lakhina A. – Crovella M. – Diot C. (2005): Mining anomalies using traffic feature distributions. Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. 1-29. p.
50. Langefors B. (1973): Theoretical analysis of information systems. Auerbach, 502 p.
51. Laudon, K. C. - Laudon, J. P. (1991): Management Information Systems. A Contemporary Perspective. USA, Macmillan Pub., 336 p.
52. Laudon, K. C. - Laudon, J. P. (2015): Management Information Systems. Global Edition. USA, Pearson, 648 p.
53. Lee W. – Stolfo S. – Mok K. W. (1999): A framework for constructing features and models for intrusion detection systems. ACM Transactions on Information Systems Security. 227-260. p.
54. Leung K. – Leckie C. (2005): Unsupervised anomaly detection in network intrusion detection using clusters. ACSC '05 Proceedings of the Twenty-eighth Australasian conference on Computer Science. 38:333-342. p.
55. Liao Y.H. – Vemuri V. R. (2002): Use of k-nearest neighbor classifier for intrusion detection. Computer & Security. 21(5): 439-448. p.

56. Li X. – Bian F. – Crovella M. – Diot C. – Govindan R. – Iannaccone G. – Lakhina A. (2006): Detection and identification of networka nomalies using sketch subspaces. Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement. 147-152. p.
57. Lin Y. – Wang H. – Zhang S. – Li J. – Gao H. (2016): Efficient quality-driven source selection from massive data sources. The Journal of Systems and Software. 118:221-233.
58. Liu F. – Shi Y. – Li P. (2017): Analysis of the Relation between Artificial Intelligence and the Internet from the Perspective of Brain Science. Procedia Computer Science. 122:377-383.
59. Mahoney M. V. – Chan P. K. (2002): Learning nonstationary models of normal network traffic for detecting novel attacks. Proceedings of the Eight ACM SIGKDD International Joint Conference on Neural Networks. 1-48. p.
60. Mahoney M. V. – Chan P. K. (2003): Learning Rules for Anomaly Detection of Hostile Network Traffic. ICDM '03 Proceedings of the Third IEEE International Conference on Data Mining. 601-611. p.
61. Marti L. - Arsene F. – Laurent N. – Schoenauer M. (2016): Anomaly Detection with the Voronoi Diagram Evolutionary Algorithm. Parallel Problem Solving from Nature. 697-706. p.
62. Mata J. – Miguel I. – Durán R. J. – Merayo N. – Singh S. K. – Jukan A. – Chamania M. (2018): Artificial intelligence (AI) methods in optical networks: A comprehensive survey. Optical Switching and Networking. 28:43-57. p.
63. Mátyus I. (2015): Tudományos tudás az információs társadalomban. Oktatási segédanyag. TÁMOP-4.2.1.D-15/1/KONYV-2015-0002 azonosítószámú pályázat keretében készült. 17 p.
64. McCarthy J. – Minsky M. – Rochester N. – Shannon C. E. (1955): Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth, Tech. Rep., 3p.
65. McCullock W. – Pitts W. (1943): A Logical Calculus of the Ideas Immanent in Nervous Activity. The bulletin of mathematical biphysics, V. évf. 4. sz. 115-133. p.
66. Munk S. (2007): Katonai informatikai a XXI. század elején. Budapest, Zrínyi Kiadó, 264 p.
67. Navaz A. S. S. – Sangeetha V. – Prabhadevi C. (2013/a): Entropy based Anomaly Detection System to Prevent DDoS Attacks in Cloud. International Journal of Computer Applications. LXII. évf. 15. sz. 32-38. p.
68. Navaz A. S. S. – Ravi M. – Prabhu T. (2013/b): Preventing Disclosure of Sensitive Knowledge by Hiding Inference. International Journal of Computer Applications. LXIII. évf. 1. sz. 42-47. p.
69. Nillson N. J. (1998): Artificial Intelligence: A new Synthesis. San Mateo, Morgan Kaufmann, 513 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
70. Poole D. – Mackworth R. – Goebel R. (1998): Computational Intelligence. A logical approach. New York, Oxford University Press, 576 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
71. Portnoy L. – Eskin E. – Stolfo S. (2001): Intrusion detection with unlabeled data using clustering. Proceedings ofACM CSS Workshop on Data Mining Applied Security. 1-25. p.
72. Ranjan R. – Sahoo G. (2014): A new clustering approach for anomaly intrusion detection. International Journal of Data Mining & Knowledge Management Process. IV. évf. 2. sz. 29-38. p.

73. Raschka S. (2015): Python Machine Learning. Birmingham, Packt Publishing, 425 p.
74. Rich E. – Knight K. – Nair S. B. (2009): Artificial Intelligence. Third Edition. New York, McGraw-Hill, 568 p.
75. Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Második Kiadás. Budapest, Panem, 1206 p.
76. Russel S. – Norvig P. (2009): Artificial Intelligence: A Modern Approach. 3rd Edition. USA, Pearson, 1152 p.
77. Sagar G. V. R. (2015): Modeling of Artificial Neural Networks using Evolutionary Algorithms. Germany, Lambert Academic Publishing, 179 p.
78. Sajtos L. – Mitev A. (2009): SPSS kutatási és adatelemzési kézikönyv. Budapest, Alinea Kiadó, 402 p.
79. Shabbir J. – Anwer T. (2015): Artificial Intelligence and its Role in Near Future. Journal of Latex Class Files. IV. évf. 8. sz. 1-11. p.
80. Shannon C. E. (1948): A Mathematical Theory of Communication. The Bell System Technical Journal. 27:379-423.
81. Smaha S. E. H. (1988): AN intrusion detection system. IEEE Fourth Aerospace Computer Security Applications Conference. 37-44 p.
82. Soule A. – Salamatian K. – Taft N. (2005): Combining filtering and statistical methods for anomaly detection. Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement. 331-344. p.
83. Szepesné S. M. (2010): Rendszertervezés 1. Az információrendszer fogalma, feladata, fejlesztése. Székesfehérvár, TÁMOP – 4.1.2-08/I/A-2009-0027, Nyugat-magyarországi Egyetem, Geoinformatikai Kar, 15 p.
84. Su M. – Yu M. – Lin C. (2009): A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach. Computers and Security. 28(5): 301-309. p.
85. Szűcs I. – Lőkös K. – Obádovics Cs. – Szigetváriné J. É. Zs. – Bárdos I. K. – Széles Zs. – Késmárky-Gally Sz. – Mucsics L. – Mohamed Zs. (2008): A tudományos megismerés rendszertana. Budapest, Szent István Egyetem Kiadó, 272 p.
86. Vasvári Gy. (2008): Vállalati (szervezeti) kockázatmenedzsment. Budapest, Információs Társadalomért Alapítvány, 183 p.
87. Veale M. – Kleek M. V. – Binns R. (2018): Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 440-445. p.
88. Vu T. – Nguyen D. Q. – Nguyen D. Q. – Dras M. – Johnos M. (2018): VnCoreNLP: A Vietnamese Natural Language Processing Toolkit. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, NAACL 2018. 1-5 p.
89. Vychodil V. (2016): Parameterizing the Semantics of Fuzzy Attribute Implications by Systems of Isotone Galois Connections. IEEE Transactions on Fuzzy Systems. XXIV. évf. 3. sz. 645-660. p.
90. Ward, J. (1998): Az információ-rendszerek szervezési elvei. Budapest, Co-Nex Könyvkiadó, 243 p.
91. Warden P. (2011): Big Data Glossary: A Guide to the New Generation of Data Tools. USA, O'Reilly Media, 62 p.
92. Warrender C. – Forrest S. – Pearlmutter B. (1999): Detecting intrusions using system calls: Alternative data models. IEEE Symposium on Security and Privacy. 133-145. p.

93. Winston P. H. (1992): Artificial Intelligence. Third edition. Massachusetts, Addison-Wesley, 737 p. Idézve: Russel S. – Norvig P. (2005): Mesterséges Intelligencia modern megközelítésben. Budapest, Panem, 1206 p.
94. Worth S. - Gross L. (1977): Szimbolikus stratégiák. In: Busignies H. – Stent G. S. – Knapp M. L. (szerk.): Kommunikáció I. Válogatott tanulmányok. Budapest, Közgazdasági és Jogi Kiadó, 37-50 p.
95. Yamanishi K. – Takeuchi J. I. (2001): Discovering outlier filtering rules from unlabeled data: Combining a supervised learner with an unsupervised learnin. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 389-394. p.
96. Yang Z. – Chen N. – Chen Y. – Zhou N. (2018): A Novel PMU Fog Based Early Anomaly Detection for an Efficient Wide Area PMU Network. Elhangzott: Fog and Edge Computing (ICFEC), 2018 IEEE 2nd International Conference, Washington, 2018. május 1-3.
97. Ye N. – Li X. Y. – Chen Q. – Emran S. M. – Xu M. M. (2001): Probabilistic techniques for intrusion detection based on computer audit data. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans. 31(4): 266-274 p.
98. Ye N. – Emran S. M. – Chen Q. – Vilbert S. (2002): Multivariate statistical analysis of audit trails for host-based intrusion detection. IEEE Transaction on Computers 51:810-820. p.
99. Yu B. – Kumbier K. (2018): Artificial Intelligence and Statistics. Frontiers of Information Technology & Electronic Engineering, XIX. évf. 1. sz. 6-9. p.
100. Zhang J. – Zulkernie M. (2016/a): A hybrid network intrusion detection technique using random forests. Proceedings of the First International Conference on Availability, Reliability and Security. 262-269 p.
101. Zhang J. – Zulkernie M. (2006/b): Anomaly based network intrusion detection with unsupervised outlier detection. Elhangzott: IEEE International Conference on Communications, Törökország, 2006. június 11-15.

6. Egyéb feldolgozott irodalmak

1. 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról. 43 p.
2. 9/2017. (VII. 19.) MNB rendelet a pénzmosás és a terrorizmus finanszírozása megelőzéséről és megakadályozásáról szóló törvény végrehajtásának az MNB által felügyelt szolgáltatókra vonatkozó, valamint az Európai Unió és az ENSZ Biztonsági Tanácsa által elrendelt pénzügyi és vagyoni korlátozó intézkedések végrehajtásáról szóló törvény szerinti szűrőrendszer kidolgozásának és működtetése minimumkövetelményeinek részletes szabályairól. 12 p.
3. A Magyar Nemzeti Bank 2/2017. (I.12.) számú ajánlása a közösségi és publikus felhőszolgáltatások igénybevételéről. 18 p.
4. A Magyar Nemzeti Bank 7/2017. (VII.5.) számú ajánlása az informatikai rendszer védelméről. 40 p.
5. AZ EURÓPAI PARLAMENT ÉS A TANÁCS (EU) 2016/679 RENDELETE (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet). 88 p.
6. Au T. C. (2017): Random Forests, Decision Trees, and Categorical Predictors: The „Absent Levels" Problem. https://arxiv.org/pdf/1706.03492.pdf. Letöltés ideje: 2018. május 9.
7. Backblaze.com (2017): Hard Drive Cost Per Gigabyte. https://www.backblaze .com/blog/hard-drive-cost-per-gigabyte/ Letöltése ideje: 2018. április 12.
8. Bealuc C. – Rosenthal J. S. (2018): Handling Missing Values using Decision Trees with Branch-Exclusive Splits. https://arxiv.org/pdf/1804.10168.pdf. Letöltés ideje: 2018. május 1.
9. Baygin M. – Karakose M. – Sarimaden A. – Akin E. (2018): An Image Processing based Object Counting Approach for Machine Vision Application. https://arxiv.org/ftp/arxiv/ papers/1802/1802.05911.pdf. Letöltés ideje: 2018 április 12.
10. Bhagoji A. N. – Cullina D. – Sitawarin C. – Mittal P. (2017): Enhancing Robustness of Machine Learning Systems via Data Transformation. https://arxiv.org/pdf/1704.02654.pdf. Letöltés ideje: 2018. május 1.
11. Coso.org (é. n.): About us. https://www.coso.org/Pages/aboutus.aspx. Letöltés ideje: 2018. április 23.
12. Chatzilygeroudis K. – Cully A. - Mouret J. (2018): Towards semi-episodic learning for robot damage recovery. https://arxiv.org/pdf/1610.01407.pdf. Letöltés ideje: 2018. április 12.
13. Cho Y. – Bianchi-Berthouze N. – Marquardt N. – Julier S. J. (2018): Deep Thermal Imaging: Proximate Material Type Recognition in the Wild through Deep Learning of Spatial Surface Temperature Patterns. https://arxiv.org/ftp/arxiv/papers/ 1803/1803.02310.pdf. Letöltés ideje: 2018. április 12.
14. Elitdatascience.com (2017): Modern Machine Learning Algorithms: Strengths and Weaknesses. https://elitedatascience.com/machine-learning-algorithms. Letöltés ideje: 2018. május 17.
15. Fanaee-T H. – Oliveira M. – Gama J. – Malinowski S. – Morla R. (2018): Event and Anomaly Detection Using Tucker3 Decomposition. https://arxiv.org/pdf/1406.3266.pdf. Letöltés ideje: 2018. mákus 2.

16. Gartner.com (2017/a): Beyond Automation: Digitalization Changes Business Process Design and Execution. https://www.gartner.com/document/3112919?ref=solrAll&refval=203318124&qid=3ea460a1d0cd3fc2ecbebee9a460f36a. Letöltés ideje: 2018. május 09.
17. Gartner.com (2017b): Machine Learning: FAQ from Clients. https://www.gartner.com/document/3772097?ref=solrAll&refval=203325447&qid=e107afa5add2a6fdb686dd81 Letöltés ideje: 2018. május 17.
18. Golomb T. – Mirsky Y. – Elovici Y. (2018): CIoTA: Collaborative IoT Anomaly Detection via Blockchain. https://arxiv.org/pdf/1803.03807.pdf. Letöltés ideje: 2018. május 12.
19. Information Security Forum (2014): IRAM$_2$. The next generation of assessing information risk. Information Security Forum Limited, 103 p.
20. Information Security Forum (2018): The Standard of Good Practice for Information Security 2018. Information Security Forum Limited, 338 p.
21. Internetworldstats.com (2017): Internet Usage Statistics. https://www.internetworldstats.com/stats.htm. Letöltés ideje: 2018. április 10.
22. ISACA (2009): The Risk IT Framework. USA, ISACA, 106 p.
23. ISACA (2012): COBIT 5: A Business Framework for the Governance and Management of Enterprise IT. USA, ISACA, 94 p.
24. ISACA. (2015/a): CISA Review Manual 2015. USA, ISACA, 426 p.
25. ISACA. (2016/b): CISM Review Manual 14th edition. USA, ISACA, 283 p.
26. ISO/IEC 27000. (2014): Information technology – Security techniques – Information security management systems – Overview and vocabulary. International Standard, 31 p.
27. ISO/IEC 27001. (2013): Information technology – Security techniques – Information security management systems – Requirements. International Standard, 23 p.
28. Kuznetsova I. – Karpievitch Y. V. – Filipocska A. – Lugmayr A. – Holzinger A. (2018): Review of machine learning algorithms in differential expression analysis. https://arxiv.org/ftp/arxiv/papers/1707/1707.09837.pdf. Letöltés ideje: 2018. április 12
29. Lin Y. H. – Brady J. P. – Forman-Kay J. D. – Chan H. S. (2017): Charge Pattern Matching as a "Fuzzy" Mode of Molecular Recognition for the Functional Phase Separations of Intrinsically Disordered Proteins. https://arxiv.org/pdf/1707.08990.pdf. Letöltés ideje: 2018. április 12.
30. Lukács János (2015): Brókerbotrány: a könyvvizsgálók az Alkotmánybírósághoz fordulhatnak. Elhangzott: ATV, 2015. április 3.
31. Maqueda A. I. – Loquercio A. – Gallego G. – Garcia N. – Scaramuzza D. (2018): Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars. https://arxiv.org/pdf/1804.01310.pdf . Letöltés ideje: 2018. április 29.
32. Margitay-Becht A. – Daruka M. – Petró K. (é. n.): Mikroökonómia. http://kgt.bme.hu/files//BMEGT301004/Mikrookonomia_jegyzet.pdf. Letöltés ideje: 2018. április 10.
33. Mkomo.com (2009): A history of storage cost. http://www.mkomo.com/cost-per-gigabyte Letöltés ideje: 2018. április 12.
34. NIST (2013): Secuirty and Privacy Controls for Federal Information Systems and Organizations. USA, NIST Special Publication 800-53, 462 p.
35. Noyel G. – Jourlin M. (2018): Framework for colour and multivariate images. https://arxiv.org/pdf/1803.00764.pdf. Letöltés ideje: 2018. április 12.
36. Pfeffer A. – Ruttenberg B. – Kellogg L. – Howard M. – Call C. – O'Connor A. – Takata G. – Reilly S. N. – Patten T. – Taylor J. – Hall R. – Lakhotia A. – Miles C. – Scofield D. – Frank J. (2017): Artificial Intelligence Based Malware Analysis. https://arxiv.org/pdf/1704.08716.pdf. Letöltés ideje: 2018. április 23.

37. Ravuri M. – Kannan A. – Tso G. – Amatriain X. (2018): Learning from the experts: From expert systems to machine learned diagnosis models. https://arxiv.org/pdf/1804.08033.pdf. Letöltés ideje: 2018. április 12.
38. Richardson B. D. – Radford B. J. – David S. E. Sequence Aggregation Rules for Anomaly Detection in Computer Network Traffic. https://arxiv.org/pdf/1805.03735.pdf. Letöltés ideje: 2018. május 18.
39. Shah V. – Gulikers L. – Massoulié L. – Vojnovic M. (2017): Adaptive Matching for Expert Systems with Uncertain Task Types. https://arxiv.org/pdf/1703.00674.pdf. Letöltés ideje: 2018. április 12.
40. Sharma V. – Kumar R. – Cheng W. – Atiquzzaman M. – Srinivasan K. – Zomaya A. Y. (2018): NHAD: Neuro-Fuzzy Based Horizontal Anomaly Detection In Online Social Networks. https://arxiv.org/pdf/1804.06733.pdf. Letöltés ideje: 2018. május 30.
41. Shekhar S. – Akoglu L. (2018): Incorporating Privileged Information to Unsupervised Anomaly Detection. https://arxiv.org/pdf/1805.02269.pdf. Letöltés ideje: 2018. május 17.
42. Siddiqui M. A. – Fern A. – Dietterich T. G. – Wong W. (2015): Sequential Feature Explanations for Anomaly Detection. https://arxiv.org/pdf/1503.00038.pdf. Letöltés ideje: 2018. május 17.
43. Stets J. D. – Sun Y. – Corning W. – Greenwald S. (2018): Visualization and Labeling of Point Clouds in Virtual Reality. https://arxiv.org/pdf/1804.04111.pdf. Letöltés ideje: 2018. április 12.
44. Suster S. – Tulkens S. – Daelemans W. (2017): A Short Review of Ethical Challenges in Clinical Natural Language Processing. https://arxiv.org/pdf/1703.10090.pdf. Letöltés ideje: 2018. április 12.
45. Syropoulos A. (2014): Fuzzy Categories. https://arxiv.org/pdf/1410.1478.pdf. Letöltés ideje: 2018. április 12.
46. Takahashi T. – Tomioka R. – Yamanishi K. (2011): Discovering Emerging Topics in Social Streams via Link-Anomaly Detection. IEEE Transactions on Knowledge and Data Engineering. XXVI. évf. 1. sz. 1-18. p.
47. Tandiya N. – Jauhar A. – Marojevic V. – Reed J. H. (2018): Deep Predictive Coding Neural Network for RF Anomaly Detection in Wireless Networks. https://arxiv.org/pdf/1803.06054.pdf. Letöltés ideje: 2018. május 28.
48. Trivedi G. – Pham P. – Chapman W. – Hwa R. – Wiebe J. – Hochheiser. (2017): An Interactive Tool for Natural Language Processing on Clinical Text. https://arxiv.org/pdf/1707.01890.pdf. Letöltés ideje: 2018. április 12.
49. Tutika C. S. – Vallapaneni C. – Karthik R. – Bharath K. P. Muthi N. R. R. K. M. (2018): Image Segmentation and Processing for Efficient Parking Space Analysis. https://arxiv.org/ftp/arxiv/papers/1803/1803.04620.pdf. Letöltés ideje: 2018. április 12.
50. Vu H. – Nguyen T. D. – Phung D. (2018): Detection of Unknown Anomalies in Streaming Videos with Generative Energy-based Boltzmann Models. https://arxiv.org/pdf /1805.01090.pdf. Letöltés ideje: 2018. május 19.
51. Young T. – Hazarika D. – Poria S. – Cambria E. (2018): Recent Trends in Deep Learning Based Natural Language Processing. https://arxiv.org/pdf/1708.02709.pdf. Letöltés ideje: 2018. április 12.
52. Yu E. – Parekh P. (2016): A Bayesian Ensemble for Unsupervised Anomaly Detection. https://arxiv.org/pdf/1610.07677.pdf. Letöltés ideje: 2018. május 11.

53. Zhai S. – Cheng Y. – Lu W. – Zhang Z. M. (2016): Deep Structured Energy Based Models for Anomaly Detection. https://arxiv.org/pdf/1605.07717.pdf. Letöltés ideje: 2018. május 12.

7. Mellékletek

7.1. számú melléklet: MTMT-be rögzített publikációs jegyzék

MTMT adatbázis - Keresés - Közlemény t

2018

1. Barta Gergő
   A Mesterséges Intelligencia hatása az üzleti folyamatokra
   MAGYAR INTERNETES AGRÁRINFORMATIKAI ÚJSÁG (233) pp. 1-15. (2018)
   Folyóiratcikk /Szakcikk /Tudományos

2. Barta Gergő , Pitlik László
   Startup felvásárlások multikulturális hátterének elemzése, avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez
   In: Farkas Attila (szerk.)
   A gazdaság kulturális szerkezete. Cultural Structure of Economic . 240 p.
   Konferencia helye, ideje: Gödöllő , Magyarország , 2017.05.12 Gödöllő: Szent István Egyetemi Kiadó, 2018. pp. 15-37.
   (ISBN:978-963-269-705-5)
   Könyvrészlet /Konferenciaközlemény /Tudományos

3. Barta Gergő , Pitlik László
   A Titanic katasztrófa túlélőinek becslése döntési fa alapú gépi tanuló eljárással
   MAGYAR INTERNETES AGRÁRINFORMATIKAI ÚJSÁG (234) pp. 1-24. (2018)
   Folyóiratcikk /Szakcikk /Tudományos

4. Barta Gergő , Görcsi Gergely
   Artificial Intelligence and Audit: Why is it necessary to audit the intelligent decision support?
   In: Káposzta József (szerk.)
   Közgazdász Doktoranduszok és Kutatók IV. Téli Konferenciája: Absztraktkötet . 119 p.
   Konferencia helye, ideje: Gödöllő , Magyarország , 2018.02.02 -2018.02.03. Gödöllő: Szent István Egyetem, 2018. p. 118. 1 p.
   (ISBN:978-963-269-714-7)
   Befoglaló mű link(ek): Egyéb URL, Teljes dokumentum
   Könyvrészlet /Absztrakt / Kivonat /Tudományos

5. Barta Gergő , Görcsi Gergely
   A CRM rendszerek szerepe a vevőkapcsolatok stratégiai kezelésében, vevőszegmentációs döntésekben
   In: Káposzta József (szerk.)
   Közgazdász Doktoranduszok és Kutatók IV. Téli Konferenciája: Absztraktkötet . 119 p.
   Konferencia helye, ideje: Gödöllő , Magyarország , 2018.02.02 -2018.02.03. Gödöllő: Szent István Egyetem, 2018. p. 62. 1 p.
   (ISBN:978-963-269-714-7)
   Befoglaló mű link(ek): Egyéb URL, Teljes dokumentum
   Könyvrészlet /Absztrakt / Kivonat /Tudományos

6. Barta Gergő , Görcsi Gergely
   Intelligent Decision Making and Process Automation for Public Organizations
   In: Monika Gubanova (szerk.)
   Legal, economic, managerial and environmental aspects of performance competencies by local authorities, 2017 : 5th international scientific correspondence conference . Konferencia helye, ideje: , 2017.12.01 -2017.12.08.

Nyitra: Slovak University of Agriculture in Nitra, 2018. pp. 30-37.
(ISBN:978-80-552-1839 -7)
Könyvrészlet /Konferenciaközlemény /Tudományos

2017

7. Barta Gergő , Pitlik László
Startup felvásárlások multikulturális hátterének elemzése, avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez
MAGYAR INTERNETES AGRÁRINFORMATIKAI ÚJSÁG 20:(228) pp. 1-23. (2017)
Link(ek): Teljes dokumentum, Matarka
Folyóiratcikk /Szakcikk /Tudományos

8. Barta Gergő , Kardos Tamás
Integrált szervezeti rendszer bemutatása az Exxonmobil példáján keresztül
In: Salamonné Huszty Anna , Pónusz Mónika , Kozma Tímea (szerk.)
Ellátásilánc-menedzsment hallgatói esettanulmánykötet: Játszmák és stratégiai megoldások az ellátási lánc hálózatban . 136 p.
Budapest: Magyar Logisztikai Egyesület, 2017. pp. 49-56.
(ISBN:978-963-12-7836-1)
Befoglaló mű link(ek): Egyéb katalógus
Könyvrészlet /Szaktanulmány /Tudományos

9. Barta Gergő , Losonczi György , Pitlik László , Pitlik László , Pitlik Marcell , Pitlik Mátyás , Varga Zoltán
Magyar statisztikai régiók érintettségi sorrendje a szálláshelyek árbevételének havi adatai alapján eltérő módszertanokkal
In: Futó Zoltán (szerk.)
Magyar vidék - perspektívák, megoldások a XXI. században: I. vidékfejlesztési konferencia . 330 p.
Konferencia helye, ideje: Szarvas , Magyarország , 2017.10.05 Szarvas: Szent István Egyetem Egyetemi Kiadó, 2017. pp. 127-140.
(ISBN:978-963-269-686-7)
Befoglaló mű link(ek): Kiadónál
Könyvrészlet /Konferenciaközlemény /Tudományos

2016

10. Barta Gergő , Marta Łętek
Equality in the Labor Market in Cracow
In: Academy of Management and Administration in Opole , Georgian Polish Education Center International Academic Interdisciplinary Conference on Modern Economics and Social Sciences . Tbilisi: Publishing House UNIVERSAL, 2016. pp. 242-250.
(ISBN:978-9941-22-736-3)
Könyvrészlet /Konferenciaközlemény /Tudományos

2015

11. Barta Gergő
Estabilishment of Enterprise Information Management Capability
In: Renata Oczkowska , Grazyna Smigielska (szerk.)
Knowledge-Economy-Society: Challenges for enterprises in knowledge-based economy . 162 p.
Cracow: Cracow University of Economics, 2015. pp. 29-36.
(Knowledge-Economy-Society; Students Book.)
(ISBN:978-83-65173-34-8, 978-8365173-35-5)
Befoglaló mű link(ek): Teljes dokumentum
Könyvrészlet /Konferenciaközlemény /Tudományos

ISBN: 978-83-65173-34-8 (printed version) ISBN: 978-83-65173-35-5 (on-line pdf)

12. Barta Gergő , Marta Łętek
Non-financial Performance Indicators as a New Approach to Measure the Value of Companies

In: Bogusz Mikuła , Tomasz Rojek (szerk.)
Knowledge-Economy-Society: REORIENTATION OF PARADIGMS AND CONCEPTS OF MANAGEMENT IN THE CONTEMPORARY ECONOMY . 520 p.
Cracow: Cracow University of Economics, 2015. pp. 247-254.
(Knowledge-Economy-Society; 2.)
(ISBN:978-83-65173-30-0)
Befoglaló mű link(ek): Teljes dokumentum
Könyvrészlet /Konferenciaközlemény /Tudományos

ISBN: 978-83-65173-30-0 (printed version) ISBN: 978-83-65173-31-7 (on-line pdf)

Megjegyzés: Az MTMT-ben feltöltött cikkeken kívül jelenleg (2018. június 10.) 7 db tudományos publikációról rendelkezem befogadó nyilatkozatokkal, melyből:

- 2 db angol nyelvű ISSN-nel rendelkező nemzetközi folyóiratcikk;
- 1 db angol nyelvű ISBN-nel rendelkező könyvrészlet;
- 2 db angol nyelvű ISBN-nel rendelkező konferenciaközlemény;
- 2 db magyar nyelvű ISBN-nel rendelkező konferenciaközlemény.

A befogadó nyilatkozatok elektronikus és fizikai formátumban is leadásra kerültek a komplex vizsga jelentkezés napjáig.