

Bankrobotika: mesterséges intelligencia és gépi tanulás alapú banki kockázatkezelés

Pénzmosás és terrorizmusfinanszírozás megakadályozása

Prisznyák, Alexandra

Pécsi Tudományegyetem, Nemzetközi Bankárképző Központ Zrt.

alexandra.prisznyak@gmail.com

ÖSSZEFOGLALÓ

A Pénzügyi Akciócsoport pénzmosással és terrorizmusfinanszírozással összefüggő országvizsgálat alapján leminősítette Magyarország R15-ajánlásnak (új technológiák alkalmazása) való megfelelését. Ezzel párhuzamosan a Magyar Nemzeti Bank 2020–2021 között számos, Magyarországon működő kereskedelmi bank esetében állapított meg bírságot a pénzmosás és terrorizmusfinanszírozás megelőzésével kapcsolatos hiányosságok miatt. A tanulmány hiánypótló elemzéseként a banki kockázatkezelés, pénzmosás és terrorizmusfinanszírozás megelőzésért felelős területek nagymértékben kiegyensúlyozatlan adatállományán operáló felügyelt (osztályzás, regresszió), felügyelet nélküli (klaszteranalízis, anomália-észlelés) és hibrid gépi tanulási modelleket, alkalmazott algoritmusokat vizsgálja. A szerző hangsúlyozza, hogy nincs egyetlen ideális algoritmus. A választást a mögöttes működési logika mellett számos összehasonlítható tényező támogatja. A modellépítés az üzleti IT- és vizionárius menedzsment hibrid szempontjainak kialakítását igényli.

KULCSSZAVAK: mesterséges intelligencia, gépi tanulási algoritmusok, banki kockázatkezelés, pénzmosás és terrorizmusfinanszírozás megakadályozása, felügyelt/felügyelet nélküli tanulás

JEL-KÓDOK: C45, C80, G21, G32, O33

DOI: https://doi.org/10.35551/PSZ_2022_2_8

Turing Computing Machinery and Intelligence publikációjának híres tételmondatával „*Can machines think?*” (Tudnak-e a gépek gondolkodni?) 1950-ben útnak indította a mesterséges intelligencia és a kapcsolódó technológiák fejlődését. Az Artificial Intelligence (AI, továbbiakban: mesterséges intelligencia) az Európai Bizottság (2018) definícióját tekintve „*azokra a rendszerekre utal, amelyek intelligens viselkedést mutatnak ki környezetük elemzésével és bizonyos célok elérése érdekében bizonyos fokú autonómiával.*” Ernyőfogalomként az AI kiterjed a Machine Learning (ML, továbbiakban: gépi tanulás) folyamataira is (ECB, 2020). Az ISO/IEC 38505-1:2017 a gépi tanulást olyan folyamatként határozza meg, amely meglévő adatok alapján, algoritmusok használatán keresztül lehetővé teszi a jövőre vonatkozó előrejelzések készítését. A ML tehát az AI olyan részhalmozaként definiálható, amely a nagy mennyiségű adatállományokon (továbbiakban: Big Data) alapulva – szoftverek, algoritmusok tanulási folyamata révén – alkalmas előre beprogramozott menettű feladatok végrehajtására (EBA, 2020; European Parliament, 2020).

Az FSB (2017) a mesterséges intelligencia banki területeken megjelenő alkalmazásával összefüggésben a következőket emeli ki:

- *front office*: támogató megoldások (erős ügyfélazonosítás, chatbotok, szerződéskötések, piaci kereskedés, modellezés, hatás-elemzés, egyebek),
- *middle office*: hitelemzés, scoring tevékenység, ügyfélminősítés, ügyfélprofil kialakítása, egyebek,
- *back office*: banki kockázatkezelési tevékenység (stresszteszt, pénzmosás és terrorizmusfinanszírozás megakadályozása, megfelelőségbiztosítás, anomáliák azonosítása, modellvalidáció, egyebek).

Szikora és Nagy (2020) megállapítja, hogy az AI banki folyamatokban történő alkalmazása leggyakrabban az ügyfélminősítés, a hitel-

bírálat, a személyre szabott pénzügyi szolgáltatások, a csalások és visszaélések felderítése/megelőzése, a szerződések átvilágítása, valamint a jogi átvilágítás területekre terjed ki.

A szerző elsődleges célja a banki pénzmosás és terrorizmusfinanszírozás megakadályozása területen alkalmazott gépi tanulási módok, technikák, algoritmusok áttekintése és alkalmazásukkal, összehasonlításukkal kapcsolatos következtetések levonása.

MAGYARORSZÁGI HELYZETKÉP

A Covid-19-pandémia indukálta digitális technológiák térnyerése, továbbá a digitális éra növekvő számú pénzügyi bűncselekményeinek hatására a mesterséges intelligencia, a gépi tanulás és kapcsolódó technológiai megoldások – mint az Advanced Analytics (AA, továbbiakban: fejlett elemzési eszközök) – térnyerése kihívás elé állította a bankszektor. Ennek következtében a mesterséges intelligencia és a gépi tanulás az Anti-Money Laundering (AML, továbbiakban: pénzmosás megelőzése), Counter Financing Terrorism (CFT, továbbiakban: terrorizmusfinanszírozás megakadályozása), pénzügyi csalásmegelőzés, valamint a compliance (továbbiakban: jogszabályi megfelelésbiztosítás) területeken is megjelent és gyorsuló ütemben fejlődik (Van Wegberg, Oerlemans, Van Deventer, 2018; Johari és társai, 2020).

A pénzügyi rendszerek pénzmosás és terrorizmusfinanszírozás céljára való felhasználásának megelőzését számos nemzetközi szervezet, így az ENSZ, Financial Action Task Force (Pénzügyi Akciócsoport, továbbiakban: FATF), az Európai Unió, az Európa Tanács, illetve szakértői bizottsága a MONEYVAL, valamint számos nemzetközi szervezet – így az IMF, a Világbank, illetve a Bázeli Bankfelügyeleti Bizottság – egyaránt kiemelt ügyként kezeli. A pénzmosás és terrorizmusfinan-

szírozás megakadályozása tekintetében az EU 2015/849 irányelve mérvadónak tekinthető, s kiterjed a Customer Due Dilligence vizsgálatokra (CDD, továbbiakban: ügyfélátvilágítás), így a KYI (Know Your Intermediary, továbbiakban: Ismerd meg a közvetítőt!), illetve KYC (Know Your Customer, továbbiakban: Ismerd meg az ügyfeled!) politikákra is (BIS, 2001). A jogszabályi megfelelést biztosítandó, az (EU) 2017/565 irányelv (2014/65/EU irányelv kiegészítése) 22. cikke a szervezeti struktúrának és nyújtott szolgáltatásnak megfelelő, testreszabott és független compliance funkció működtetését és kockázatalapú monitoring rendszer üzemeltetését írja elő. Az említett EU-irányelvek, számos egyéb jogszabállyal, rendelettel és ajánlással egészülnek ki nemzetközi és nemzeti szinten. Az EU-s szintű, illetve nemzetközi jogszabályok (irányelvek, rendeletek, ajánlások) hazai jogba történő átültetése a Pénzmosás és a terrorizmus finanszírozása megelőzéséről és megakadályozásáról szóló 2007. évi CXXXVI. törvény, illetve a kapcsolódó jogszabályok módosításával valósul meg. A hazai implementációt a hatályban lévő, egyéb jogszabályok és a Magyar Nemzeti Bank (MNB) ajánlásai támogatják. A hatályos hazai pénzmosás elleni szabályozás a Büntető Törvénykönyvben és a 2017. évi LIII., a pénzmosás és terrorizmus finanszírozása megelőzéséről és megakadályozásáról szóló törvényben található.

A párizsi székhelyű – 1989-ben alapított – Financial Action Task Force (Pénzügyi Akciócsoport, továbbiakban: FATF), kormányközi szervezetként globális szinten 39 tagország pénzmosás és terrorizmusfinanszírozás-ellenes tevékenységét segíti elő, amelyhez Magyarország az Európai Bizottság tagsági viszonyán keresztül lazán kapcsolódik. Működése a pénzmosás és terrorizmusfinanszírozás megakadályozása érdekében nemzetközi együttműködések megteremtését, valamint kockázatalapú megközelítésének megfelelő ajánlások

kidolgozását és globális szintű végrehajtását célozza. Az FATF-ajánlásoknak (R40, illetve +9 speciális ajánlás) történő megfelelés valamennyi FATF-tag (így Magyarország) esetében rendszeres időközönként kiértékelésre kerülnek (FATF, 2012, 2021, 2021b).

Az FATF-ajánlásokkal összhangban, Magyarország az elmúlt évtizedben sikeresen fejlesztette az összefüggő szabályozási környezetet (FATF, 2021a; Tóth, 2018). Az FATF által is elfogadott országjelentés alapján a vizsgált ajánlások (R40) tekintetében Magyarország 2019–2021 időszakban javulást ért el, amelyet az 1. táblázat tartalmaz (FATF, 2021a).

Az FATF-ajánlások értékelése: (C) megfelel, (LC) többnyire megfelel, (PC) részben megfelel, (NC) nem felel meg minősítéssel történik. Az ajánlások tekintetében végső cél a PC-minősítésű ajánlások javítása legalább LC/C-szintre, ellenkező esetben a plenáris ülés Megfelelésnövelő Eljárásra (Compliance Enhancing Procedure, CEP) tesz javaslatot.

A 2019-es országminősítést alapul véve 2021-ben a következő változások jelentkeztek:

- teljesítettnek minősülnek a következő ajánlások: R4, R9, R20, R29, R30;
- többnyire megfelelnek a következő ajánlások: R1, R2, R3, R5, R6, R7, R10, R11, R12, R14, R16, R17, R19, R21, R22, R23, R25, R26, R27, R28, R31, R33, R34, R35, R36, R37, R38, R39, R40;
- javítandó terület: R13, R15 (új technológiák alkalmazása), R18, R24, R32.

A szerző felhívja a figyelmet arra, hogy – a mesterséges intelligencia és kapcsolódó technológiák alkalmazásával összefüggésben fontos változás 2019-hez képest a pénzmosás és terrorizmusfinanszírozás megakadályozását támogató – 'Új technológiákra vonatkozó ajánlás' (R15) leminősítése (C-minősítésről PC-minősítésre). Az R15-ös ajánlás célja azon új technológiák banki implementációjának ösztönzése, amelyek felkészítik az intézményeket a digitális korszak újfajta pénzmosás és terroriz-

MAGYARORSZÁG FATF AJÁNLÁSOKNAK VALÓ MEGFELELŐSÉGE 2019–2021 KÖZÖTT

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
2019	LC	LC	LC	C	LC	LC	LC	PC	C	LC
2021	LC	LC	LC	C	LC	LC	LC	PC	C	LC
	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
2019	LC	PC	PC	LC	C	LC	LC	PC	LC	C
2021	LC	LC	PC	LC	PC	LC	LC	PC	LC	C
	R21	R22	R23	R24	R25	R26	R27	R28	R29	R30
2019	LC	LC	LC	PC	LC	LC	LC	LC	C	C
2021	LC	LC	LC	PC	LC	LC	LC	LC	C	C
	R31	R32	R33	R34	R35	R36	R37	R38	R39	R40
2019	LC	PC	LC	LC	LC	LC	LC	LC	LC	LC
2021	LC	PC	LC	LC	LC	LC	LC	LC	LC	LC

Forrás: saját szerkesztés FATF-alapján (2021a)

musfinanszírozási módszereire, növelve ezáltal a pénzügyintézetek kockázatkezelési tevékenységének hatékonyságát. Amint az látható, az új technológiák alkalmazására irányuló ajánlások tekintetében Magyarország teljesítménye elmarad a várt szinttől.

A Magyar Nemzeti Bank 2020, 2021 során a jelentős készpénzforgalommal járó tevékenységeknél szükséges pénzmosás-megelőzési kontrollok (kockázatazonosítás, kezelés, folyamatszabályozás, belső ellenőrzés) hatékonyságának ellenőrzése során jelentős bírsággal sújtotta azon Magyarországon bejegyzett kereskedelmi bankokat, amelyeknél hiányosságokat tárt fel (1. ábra).

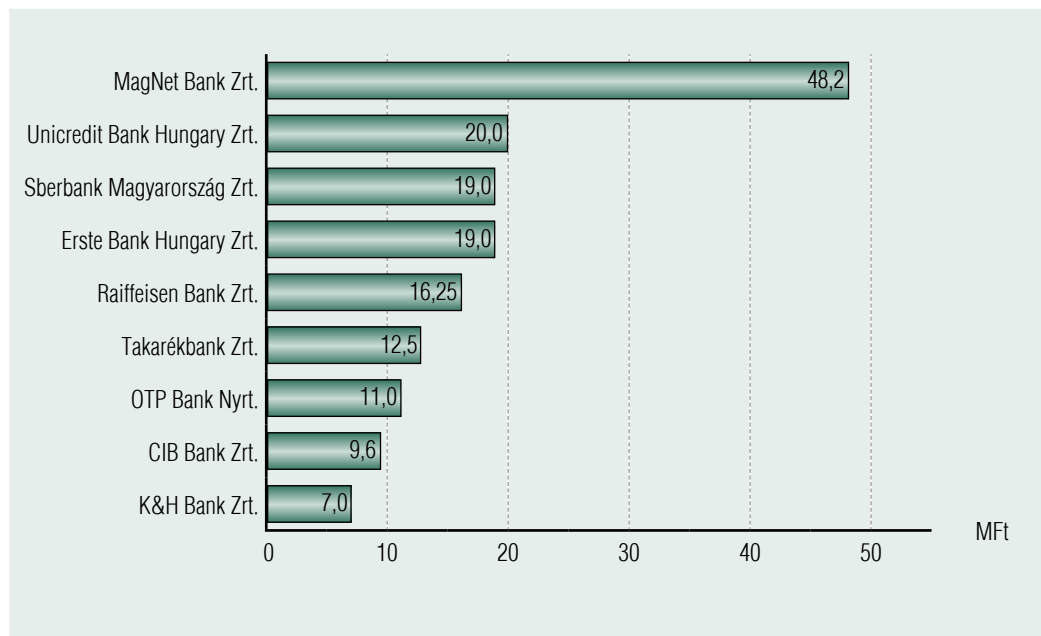
A mögöttes okok közül számos olyan folyamatszervezési és információelemzési, jelentési kötelezettségben tapasztalható hibákra vezethető vissza, amelyek kiküszöbölése humán

mesterséges intelligencia interakció révén jelentősen csökkenthető volna.

AML-, CFT-VIZSGÁLAT ÁLTALÁNOS FOLYAMATA ÉS A GÉPI TANULÁS LEHETŐSÉGEI

A névtelen és fiktív nevekre nyitott számlavezetés elkerülése érdekében a pénzügyintézetek az ügyfelek azonosságának ellenőrzését biztosító szisztémát, rendszert működtetnek az üzleti kapcsolat létrehozásakor és az egyedi tranzakciók indítása során. A kockázati kategóriánként eltérő mértékű és rendszerességű CDD-azonosítást végző jelzőrendszer az ügyfél azonosságának ellenőrzését külső és belső adatbázisok, egyéb benyújtott dokumentumok, továbbá független forrásdokumentumok révén

MBN FELÜGYELTI BÍRSÁG MÉRTÉKE A PÉNZMOSÁS-MEGELŐZÉSI SZABÁLYOK MEGSÉRTÉSÉÉRT 2020–2021 KÖZÖTT A MAGYARORSZÁGON MŰKÖDŐ KERESKEDELMI BANKOKNÁL



Forrás: saját összeállítás

támogatja. Az ügyfeladatok nagy mennyiségű állománya az FATF-ajánlásoknak megfelelően jellemzően olyan információkat tárol, amelyek lehetővé teszik az ügyfelek megbízható folyamatos azonosítását (CDD), továbbá a végrehajtott tranzakciók vizsgálatát.

A pénzmosási tevékenységre utaló mintázatot felismerő riasztási modellek a gépi tanulás révén képesek az ügyfeladatok, tranzakciós és egyéb háttérinformáció felhasználásával valószínűséget társítani a pénzmosás bekövetkezéséhez. Amennyiben az adott eseményt a rendszer „gyanús” címkével jelöli, úgy további vizsgálat céljából egy magasabb szakértői szintre történő delegáció valósul meg a folyamatban (2. ábra). Amennyiben a további szakértői elemzés az illetékes hatóság felé továbbítandó gyanús esetként jelöli meg az adott ügyletet, úgy az eset a bejelentési és egyéb compliance

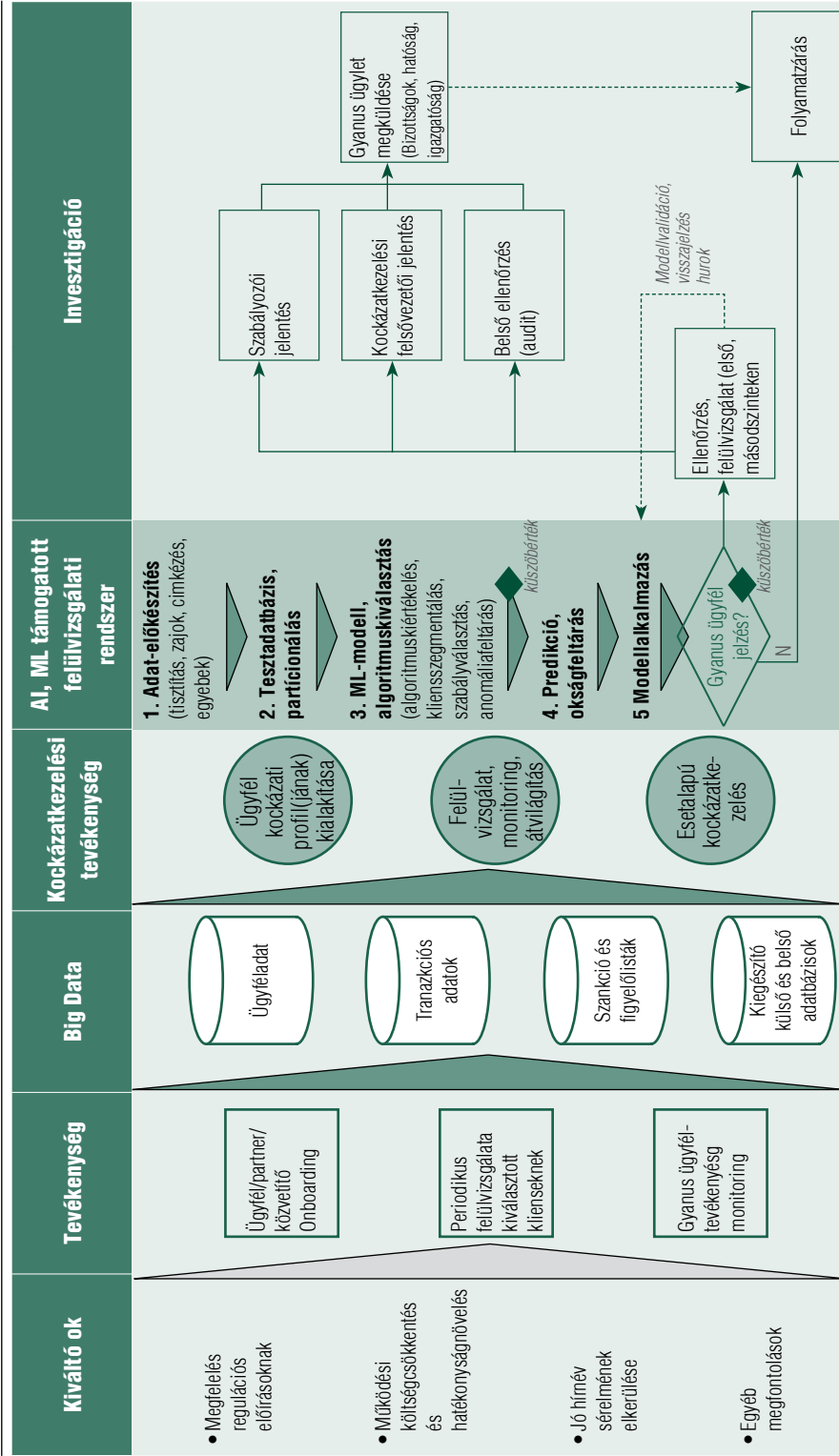
megfelelőség biztosítása érdekében jelentésre kerül (Jullum, Løland, Huseby, 2020).

Az alkalmazott attribútumok köre (szabályozói oldali információk, nyílt külső források, szakértői ismeretek, pénzügyi bűncselekmények múltbeli adatai) a modellépítő szakmai álláspontja alapján egyéb kiegészítő információval bővíthet a modell javítása céljából (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano, 2021; Rouhollahi, 2021; Chen et al., 2018). A pénzmosás megelőzése területen felhasznált jellemző attribútumokat a 2. táblázat szemlélteti.

Az attribútumok körének megfelelő megválasztása, valamint a redundáns, irreleváns attribútumok eltávolítása a ML-modellépítés fontos szakasza. Az irreleváns adatok eltávolítása javíthatja az algoritmus tanulási pontosságát, továbbá csökkenheti a számítási időigényt, va-

2. ábra

GÉPTANULÁS-ALAPÚ PÉNZMOSÁS ÉS TERRORIZMUSFINANSZÍROZÁS ELLENŐRZÉSI FOLYAMATA



Forrás: szerző saját ábrája

AML-, CFT-JELZÉSRE ÉPÍTETT ML-MODELL TIPIKUS ATTRIBÚTUMAI

Attribútum kategória	Változó
Ügyféllel kapcsolatos attribútumok	Ügyfél típusa (jogi személy, magánszemély, egyéb) Ügyfélszegmens Politikai közszereplői státusz (PEP) Életkor Állampolgárság Beáramló pénzeszköz (jövedelem) forrása Igénybe vett terméktípus Gazdasági tevékenység Ügyfél belépésétől eltelt idő Üzleti jellegű adatok (tulajdonosi szerkezet, részesedések)
Tranzakcióval kapcsolatos attribútumok	Tranzakció típusa Tranzakcióba bevont ügyfelek nevei (küldő, fogadó fél) Tranzakció gyakorisága Tranzakció dátuma, ideje Tranzakció összege Pénznem Átlagos összege Célbank Tranzakciós kód Fiók (ügyfél) típusa, Tranzakciónyilatkozat
Egyéb attribútumok	Termék/szolgáltatás típusa Földrajzi terület (kitettség) Jogi személy képviselője
Kapcsolati háló	Kliensek összefüggő üzleti kapcsolathálózata

Forrás: a szerző saját összeállítása a feldolgozott szakirodalmak alapján

lamint hozzájárulhat az összefüggések pontosabb megértéséhez.

OPTIMALIZÁCIÓS PROBLÉMA: ALGORITMUSVÁLASZTÁS ÉS ÖSSZEHASONLÍTÁS

A pénzügyintézetek a pénzmosás és terrorizmusfinanszírozás megelőzése, felderítése érdekében alkalmazott hagyományos elemző technikák

esetében számos problémával szembesülnek: magas IT-költségek, erőforrás-igényes elemzési technikák, magas hamis pozitív találati arány, a bűnözők viselkedési mintáját azonosító szabályok rugalmatlansága, dinamizmusának hiánya (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano, 2021). A hagyományos módszerekben bekövetkezett fejlődés mögöttes okaként, a számítógépes teljesítmény javulása, a mesterséges intelligencia és adatbányászat alkalmazásának elterjedése, a mesterséges

intelligencia és kapcsolódó technológiák fejlődése, továbbá az alkalmazott jogszabály szigorodása húzódik meg (Watkins et al., 2003).

A gépi tanulás a felügyelt, a felügyelet nélküli megerősítő tanulási módszerek révén olyan hatékony elemzési megoldást szolgáltat, amely a hagyományos relációs adatbázisok által feldolgozott strukturált adatokon túlmenően (dokumentumorientált/NoSQL-adatbázisok révén) tovább növeli az elemzéshez felhasználható adatok körét. A modellépítés során alkalmazott attribútum kiválasztásának kulcskérdése, hogy mely(ek) képes(ek) az egyes változók értékeiből más változók értékeit korrekten megbecsülni.

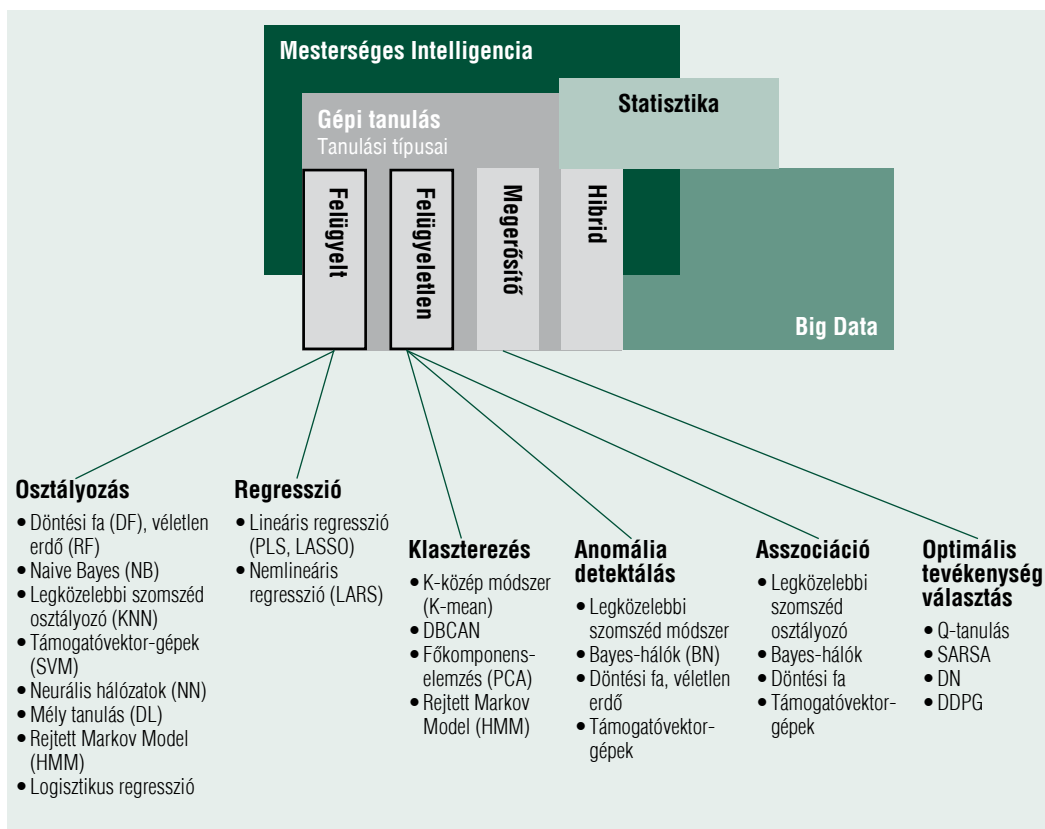
A mesterséges intelligencia ernyőfogalma alá tartozó gépi tanulási algoritmusok csoportosítását a 3. ábra szemlélteti.

Modell tanulása esetén fontos a rendelkezésre álló adatállomány nagysága, amelynek függvényében célszerű megválasztani a tanulási módját, így kijelölni a lehetséges algoritmusok körét (Savage et al., 2016; Zhang, Trubey, 2019; Chen et al., 2018).

Felügyelt gépi tanulás esetén az algoritmus olyan tanító adathalmaz révén tanul, amelyek címkékkel rendelkeznek (például a múltban igazoltan pénzmosás és terrorizmusfinanszírozás esetek alapján). Az (1) klasszifikáció és (2) regresszió a felügyelt tanulási módszerek közé

3. ábra

AI-ERNYŐFOGALOM, ML TANÍTÁSI MÓDOK ÉS ALGORITMUSOK



Forrás: saját ábra, EBA (2020) riport alapján

sorolandó. A felhasznált adatbázisból mintázatok alapján tanított modell (algoritmus) képes megfelelő pontossággal előre jelezni egy új, még nem ismert objektum (input, X) osztályba tartozását (output, Y) vagy azt, hogy milyen ismérvekkel, tulajdonságokkal fog rendelkezni: $(Y) = f(X)$.

A gyanús tranzakciók címkézése felügyelt tanulás esetén jellemzően szakértők segítségével, vagy címkéző módszerek (például Snorkel-modell) révén valósul meg.

Címkézési példa bizonyos tranzakciótípus, tranzakcióösszeg célszág alapján:

ase when Transaction_Type = 'Payment transfer' and Transaction_Amount > x then 'True'

when Country_Code in ('Ethiopia', 'Kenya') and Transaction_Amount > x then 'True' else 'False' and as Suspicious

ahol a felsorolt országok az FATF-feketelistán lévő országai közül kerültek kiválasztásra, illetve x az adott pénznyben denominált fizetőeszköz kockázati küszöb szempontból meghatározott limit értéke.

Felügyelet nélküli tanulás esetén az algoritmus automatikusan alakít ki osztályokat az alapján, hogy előre nem ismert összefüggéseket, asszociációs kapcsolatokat és döntési stratégiákat képez az adathalmazban felismert összefüggések, mintázatok alapján. Ellentétben a felügyelt tanulással, ez esetben nincs előre kijelölve, hogy a minta adatbázis adott objektuma mely kritériumok alapján kerül csoportosításra. Vagyis, az objektumok nincsenek előre felcímkézve, azok címkézését az algoritmus az adatbázisban feltárt (rejtett) mintázatok alapján végzi. Jellemző felügyelet nélküli módszerek: (1) klaszteranalízis, (2) anomáliaészlelés. Az adott üzleti probléma specifikálását követően az algoritmusválasztáshoz a felhasználó a rendelkezésre álló adatok körét (adatmeny-

nyiség), struktúráját, a kiugró adatokat (anomáliák) sokaságon belüli arányát (esetszám) tekintetbe véve dönthet. Döntését továbbá a modellépítést követően kiértékelési eszközök támogatják. A különböző gépi tanulási algoritmusok eltérő bemeneti adatformátumot és attribútumválasztást igényelnek. Az eltérő ML-tanulási folyamatok indukálják, hogy az adott algoritmus tanulási folyamata, adatfelhasználása eltérő az adott probléma megoldása érdekében (lásd 3., 4., 5. táblázat).

A 3., 4., 5. táblázat alapján megállapítható, hogy az osztályozási módszer során alkalmazott tipikus algoritmusok a következők: logisztikus regresszió, legközelebbi szomszéd módszer, (mesterséges) neurális hálózatok, Naive Bayes, döntési fa, (illetve változatai: XGBoost, pGBRT, FP-growth, véletlenerdő-modell), SVM, Bayes logisztikus regresszió, Bayesian network. Míg a regressziós eljárások tekintetében: a Maximum Likelihood logisztikus regresszió, került a vizsgált szakirodalmi mintán alkalmazásra. A klaszteranalízis során a következő algoritmusok alkalmazása figyelhető meg: K-közép módszer, neurális hálózatok, Neural Gas, SOM-algoritmus. Anomália detektáció során pedig az iForest algoritmus került alkalmazásra a vizsgált forrásokban.

A 3., 4., 5. táblázatból kitűnik, hogy a ML-modellek teljesítményértékelése jellemzően valamely teljesítményindikátor alapján, illetve szakértő validációján keresztül valósul meg (teljesítményindikátor-oszlop).

Tekintve, hogy a 3., 4., 5. táblázatban szereplő algoritmusok és eredmények az alkalmazott adatbázisok, modellépítési eljárás (adattisztítás, parametrizáció, egyebek) eltérést mutat, így a táblázatokban szereplő algoritmusok eredményének összevetése nem lehetséges, az ugyanis feltételezné az azonos adatbázison való futtatás lehetőségét. Ugyanakkor képet szolgáltatnak a pénzmosás és terrorizmusfinanszírozás területen alkalmazott megoldások széles spektrumáról.

FELÜGYELT GÉPI TANULÁSI MÓDSZEREK ÉS ALGORITMUSOK

Szerző(k), év	ML-algoritmusok	Alkalmazott adatházis	Teljesítményindikátor	Eredmény
Zhang és Trubey (2019)	SVM, ANN, Maximum Likelihood/ Bayes logisztikus regresszió, döntési fa, véletlenerdő-modell	Amerikai pénzügyi adatok	AUC, ROC-görbe	Legjobb teljesítmény: ANN-algoritmus. Az SVM és RF a logisztikus regressziónál jobb teljesítményt nyújt. Az SVM jól teljesít a lineárisan nem elválasztható csoportok osztályozásában.
Wang, és Yang (2007)	Döntési fa	Ügyfél tranzakciós adatok	Hamis pozitív arány	A döntési fa korlátozottan hatékony (nem azonosít megfelelően minden gyanús tranzakciót).
Chen és Guestrin (2016)	XGBoost, pGBRT	4 publikus adatkészletet alkalmaz	Futási (végrehajtási) idő, pontosság	Az XGBoost futási ideje, erőforrásigénye alacsonyabb, mint a pGBRT-é, és kielégítő eredményt nyújt.
Wei és társai. 2012	ContrastMiner	Ausztráliai bank adatbázisa	Riasztási : észlelési arány	A ContrastMiner algoritmus nagy, kiegyensúlyozatlan adatalomány esetén jelentősen javítja a pontosságot.
Khan és társai (2013)	Bayesi-hálózatok	Valós pénzügyi adatbázis	Hamis pozitív arány	Ügyfélviselkedési minizátozhoz kötött „gyanús” jelölés magasabb szintű szakmai vizsgálat céljából.
Kannan és Srinath (2017)	TBOD és AROMLD algoritmus	Valós banki adatbázis	Érzékenység, pontosság, specifikás, futási (végrehajtási) idő	A TBOD pontosabb, ugyanakkor számítási komplexitása és végrehajtási ideje az összteljesítményt csökkenti. Alternatívaként az AROMLD futási ideje alacsonyabb.
Jullium, Løland, Huseby (2020)	XGBoost	Anonim banki adatok	Brier-pontszám, AUC-görbe, PPP	Az XGBoost alkalmazása teljesítménynövekedést eredményez (AUC, PPP, TPR) tekintetében.
Patiil, Dharwadkar (2017)	Mesterséges neurális hálózatok	Nyilvánosan elérhető német hitelfelvevői adatbázis	Átlagos négyzetes gyökeltérés, pontosság	A modell jó eredményeket ér el a besorolási pontosság tekintetében.
Álvarez és társai (2017)	Logisztikus regresszió, döntési fák, neurális háló, véletlenerdő-modell	Valós tranzakciós adatokat tartalmazó adatbázis	Pontossága	Az adatalományban lévő zaj eltávolításával javítható az osztályozó algoritmus teljesítménye. Az RF a többenél magasabb osztályozói teljesítményt biztosít.
Savage és szerzői társai (2016)	K-legközelebbi szomszéd algoritmus, SVM, RF	Ausztrál Tranzakciójelentés, és az Elemzési Központnak jelentett tranzakcióadatok	ROC-görbe, pontosság, FTP-ára	Az SVM-nél az RF valamivel jobb teljesítményt ad. A kapcsolatalemzés növeli a jelenlegi rendszer hatékonyságát.
Deng és társai (2012)	Naive-Bayes, SVM	Pénzügyi tranzakciós adatok	Pontosság, szakértői validáció	A szkevencialis aktív tanulási módszer alkalmazása magasabb teljesítményt eredményezett a Naive Bayes és SVM-modellek teljesítményénél.
Luo (2014)	FP-growth algoritmus	Nagyméretű generált tranzakciós adatalomány	Pontosság	A tranzakciós szám növekedésével a modell teljesítménye javul.

Forrás: saját szerkesztés

4. táblázat

FELÜGYELET NÉLKÜLI GÉPI TANULÁSI MÓDSZEREK ÉS ALGORITMUSOK

Szerző(k), év	Alkalmazott algoritmusok	Alkalmazott adatbázis	Teljesítményindikátor	Eredmény
Alexandre Balsa (2018)	K-közép algoritmus	Pénzüntézet valós adatbázisa alapján	ROC-görbe, illetve banki szakemberek (validáció)	Szabálygenerálásnál a J48 „JPART” algoritmus biztosította a legjobb eredményt. Ugyanakkor a precizitás mértéke elmaradt a várttól.
Khac et al. (2010)	Neurális hálózatok	BEP Bank befektetési alapokhoz kapcsolódó tranzakciós adataival	banki szakértői validáció	A modell esetében a paramterek megválasztása fontos a modell teljesítménye és futási (végrehajtási) idő szempontjából.
Drezewski, Sepielak, Filipkowski (2012)	FP-Growth, FPClose, FPMMax, Sequence Miner, BIDE, BIDEMax	Bankszámlakivonatokat tartalmazó adatbázis	végrehajtási idő	A klaszterelemzés és létrehozott fűrtek sikeresen használható a pénzüsmosás felderítésére.
Rocha-Salazar, Segovia-Vargas, Camacho-Miliano (2021)	K-közép, Neural Gas, Strict, SOM-algoritmus	Mexikói pénzüntézetek adatbázisa	Calinski-Harabasz-index; abnormalitás indikátor, pontosság, ERR, ACC	Addicionális, nem tranzakciós változók modellbe integrálása javítja az előrejelzés pontosságát, és csökkenti az emberi erőforrás szükségletet, költségeket.

Forrás: saját szerkesztés

5. táblázat

HIBRID MODELLEK

Szerző(k), év	ML-algoritmusok	Alkalmazott adatbázis	Teljesítményindikátor	Eredmény
Rouhollahi (2021)	Klasszifikációs algoritmusok: logisztikus regresszió, legközelebbi szomszéd módszer, véletlen erdő, neurális hálózatok, Naive Bayes, Anomáliaszűrés: iForest	Strukturált banki tranzakciós adatai	Pontosság, precizitás, fedés, F1-mérték	Legmagasabb pontosság: neurális hálózat (futási ideje hosszabb, mint a többi algoritmusé). Fedés szempontjából legmagasabb érték: RF. Anomáliaszűrésnél legjobb eredmény: iForest. Az osztylvázás jobb teljesítményt produkált mint az anomáliaszűrés, ugyanakkor a kettő kombinációja magasabb precizitást és alacsonyabb humánerőforrás-szükségletet eredményezett.

Forrás: saját szerkesztés

KÖVETKEZTETÉSEK

A szerző felhívja a figyelmet arra, hogy nincs egyetlen ideális algoritmus. Lényegében minden algoritmussal készített előrejelzés felfogható egy optimalizálási problémaként, mivel egy adott célfüggvény optimalizálása a cél. Az algoritmusválasztás a megoldandó probléma vizsgálata alapján történik. Míg a lineáris regressziós modell célja az előrejelzések és a tényleges érték közötti (négyzetes) eltérés minimalizálása, addig az SVM-algoritmus lineáris kategorizálást hajt végre a hipersíkon egy elválasztó sík segítségével, amelynek eredményeképpen kapott margó (vagyis az elválasztó síkkal párhuzamos hipersíkkal meghatározott tér, amely nem tartalmaz tanító adatpontokat) a lehető legnagyobb. A véletlenerdő-modellben lévő független döntési fák egy véletlenszerű mintát alapul véve hozzák meg egyéni döntésüket, amelyek végül többségi szavazás révén megadják a klaszifikációs probléma megoldását. Míg a Naiv Bayes osztályozó a bemeneti vektorértékek alapján kalkulálja az adat osztályba tartozásának valószínűségét.

Több lehetséges megoldás esetén az egyes algoritmusok hatékonyságának mérése és összehasonlítása a következő kritériumok alapján lehetséges:

- (1) időbeli komplexitás (betanításhoz szükségbe vett idő,
- (2) végrehajtási időszükséglet,
- (3) memória igény/tárigény szükségesség (futás során szükséges memória),
- (4) párhuzamos műveletvégzés (több művelet egyidejű végrehajtása, több gépen való futtatás) lehetősége,
- (5) parametrikusság,
- (6) linearitás.

A 6. táblázat a véletlenerdő-modell, a legközelebbi szomszéd módszer, az SVM, K-központ és a lineáris regresszió algoritmus említett tényezők mentén történő összehasonlítását tartalmazza.

Megjegyzendő, hogy a táblázatban párhuzamos műveletvégzés szempontjából „nem” értékkel jelölt algoritmusok esetében számos módszer létezik a párhuzamos műveletvégzési képesség kialakításához.

A vizsgált algoritmusok összehasonlítását alapul véve elmondható, hogy az üzleti cél, továbbá az algoritmus mögöttes elméleti vonatkozásainak összhangja, a rendelkezésre álló adatállomány minősége egymástól elidegeníthetetlen kérdéskörként jelentkeznek. Az algoritmusok előrejelzésének kombinációja (hibridizáció, ensemble modellek) a varianciahiba trade off csökkentésén és az előrejelzési pontosság javításán keresztül megbízható megoldást kínálhat az egyoldalú elemzés elkerülése érdekében. Ugyanakkor figyelembe kell venni, hogy nem eredményez automatikusan magasabb teljesítményű, pontosságú modellt.

KONKLÚZIÓ

A pénzmosás és terrorizmusfinanszírozás megelőzése területeken alkalmazott gépi tanulási modellek validációja során a humán erőforrás nélkülözhetetlen. Ugyanakkor a gépi tanulási modellek jelentősen hozzájárulnak a munkaidő felszabadításához, a munkaező munkavégzésének támogatásához. Nincs egyetlen ideális algoritmus. Az üzleti szempont, az IT-terület, továbbá a vizionárius menedzsment együttműködéséből származó hibrid nézőpont figyelembevétele elengedhetetlen előfeltétele a ML-modellek folyamatba történő integrációjának és sikeres hasznosításának. Az algoritmusok közötti választást az egyes algoritmusok mögöttes elméleti vonatkozásai, továbbá egyéb összehasonlításul szolgáló faktorok támogatják. Hangsúlyozandó ugyanakkor, hogy a modellépítés jelentős részét (mintegy 80 százalékát) kitevő adatelőkészítési munkálatok is jelentősen befolyásolhatják a kapott előrejelzési eredményeket. Szakterület függően eltérő

6. táblázat

KIVÁLASZTOTT ALGORITMUSOK ÖSSZEHASONLÍTÁSA SZÁMÍTÁSI KOMPLEXITÁS TÉNYEZŐI ALAPJÁN

Összehasonlítás szempontjai	RF	K-NW	SVM	Lineáris regresszió	Logisztikus regresszió	Naive Bayes	K-közép
Tanítási idő	$O(n \times \log(n) \times m \times T)$	$O(k \times n \times m)$	$O(n^2)$ alacsony P esetén, míg $O(n^3)$ magas P esetén	$O(m^2(n+m))$	$O(n \times m)$	$O(n \times m)$	$O(l \times CL \times n \times m)$
Sok megfigyelés esetén lassú	Sok megfigyelés esetén lassú	Sok megfigyelés esetén lassú	magas P esetén tanítási idő hosszú	Tanítási ideje hosszú, végrehajtási ideje rövid	Rövid, hatékony végrehajtási idő (főleg kis adatbázison)	Alacsony tanítási idő	Nagyobb adatkészletnél lassabb
Végrehajtási idő	$O(D \times T \times T)$	$O(n \times m)$	$O(S \times m)$	$O(m)$	$O(m)$	$O(o \times m)$	$O(n \times m + CL \times m)$
Tárigény	$O(D \times T \times T)$	$O(n \times m)$	$O(n^2)$ alacsony	$O(m)$ alacsony	$O(m)$ alacsony	$O(o \times m)$ alacsony	$O((CL \times n)) \times m$
Párhuzamos műveletvégzés	Igen	Igen	Nem	Nem	Nem	Nem	Nem
Parametrikusság	Nem parametrikus	Nem parametrikus	Nem parametrikus	Parametrikus	Parametrikus	Parametrikus	Parametrikus
Linearitás	Nemlineáris	Nemlineáris	Lineáris/Nemlineáris (kernel)	Lineáris	Lineáris	Lineáris	Lineáris

Megjegyzés: A táblázatban felhasznált tényezők rövidítései a következők:
n: tanító adathalmaz adatmennyiség
P: büntetési paraméter (léves besorolás esetén járó büntetés paramétere)
k: szomszédok száma
*D*o*T:* döntési fák mélysége
l: iteráció száma

Forrás: saját szerkesztés

mértékű az adott optimalizációs problémához kapcsolódó események részaránya a sokaságon belül, amelynek függvényében a modellező eltérő gépi tanulási módszer (felügyelt, felügyelet nélküli, megerősítő tanulás), illetve algoritmus alkalmazása mellett dönthet.

LIMITÁCIÓ

A vizsgált szakterület esetében alkalmazott tipikus algoritmusok összehasonlítására valós banki adatbázis felhasználásán keresztül a tanulmány folytatásaként kerülhet sor. ■

IRODALOM

- ALEXANDRE, C., BALSÁ, J. (2018). A Multi-Agent System Based Approach to Fight Financial Fraud: An Application to Money Laundering. *Preprints*, 2018010193, <http://doi.org/10.20944/preprints201801.0193.v1>
- ÁLVAREZ-JAREÑO, J. A., BADAL-VALERO, E., PAVÍA, J. M. (2017). Using machine learning for financial fraud detection in the accounts of companies investigated for money laundering. Working Papers 2017/07, Economics Department, Universitat Jaume I, Castellón (Spain), <https://ideas.repec.org/p/jau/wpaper/2017-07.html>
- CHEN, T., GUESTRIN, C. (2016). XgBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 785–794, <http://dx.doi.org/10.1145/2939672.2939785>
- CHEN, Z. et al. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, Vol. 57, pp. 245–285, <http://doi.org/10.1007/s10115-017-1144-z>
- DENG, X., V. ROSHAN, J., A. SUDJIANTO, JEFI WU, C. F. (2012). Active Learning Through Sequential Design, With Applications to Detection of Money Laundering. *Journal of the American Statistical Association*, Vol. 104, 2009, Issue 487, <https://doi.org/10.1198/jasa.2009.ap07625>
- DREZEWSKI, R., SEPIELAK, J., FILIPKOWSKI, W. (2012). System supporting money laundering detection. *Digital Investigation*, Vol. 9, Issue 1, pp. 8–21, <https://doi.org/10.1016/j.diin.2012.04.003>
- JOHARI, R. J., ZUL, N. B., TALIB, N., HUSSIN, S.A.H.S. (2020). Money Laundering: Customer Due Diligence in the Era of Cryptocurrencies. Proceedings of the 1st International Conference on Accounting. *Management and Entrepreneurship*, (ICAMER, 2019) <https://doi.org/10.2991/aebmr.k.200305.033>
- JULLUM, M., LØLAND, A., HUSEBY, R. B., ÅNONSEN, G., LORENTZEN, J. (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control*, Vol. 23, No. 1, pp. 173–186, <https://doi.org/doi/10.1108/JMLC-07-2019-0055/full/html>
- KANNAN, S. R., SOMASUNDARAM. K. K. (2017). Autoregressive-based outlier algorithm to detect money laundering activities. *Journal of Money Laundering Control*, Vol. 20, No. 2, pp. 190–202, <https://doi.org/10.1108/JMLC-07-2016-0031>
- KHAN, N. et al (2013). A Bayesian approach for suspicious financial activity reporting. *International Journal of Computers and Applications*, 35(4), <https://doi.org/10.2316/Journal.202.2013.4.202-3864>

- LE-KHAC, N. A. et al. (2010). A Data Mining-Based Solution for Detecting Suspicious Money Laundering Cases in an Investment Bank. 2nd International Conference on Advances in Databases. *Knowledge, and Data Applications*, DBKDA, pp. 235–240, <https://doi.org/10.1109/DBKDA.2010.27>.
- LUO, X. (2014). Suspicious Transaction Detection for Anti-Money Laundering. *International Journal of Security and Its Applications*, Vol. 8, No. 2, pp. 157–166, <https://doi.org/10.14257/ijisia.2014.8.2.16>
- PATIL, P. S., DHARWADKAR, N. V. (2017). Analysis of banking data using machine learning. International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). pp. 876–881, <https://doi.org/10.1109/I-SMAC.2017.8058305>
- ROCHA-SALAZAR, J-D-J., SEGOVIA-VARGAS, M-J., CAMACHO- MIÑANOM M-D-M. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications*, Vol. 169, 114470, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114470>
- ROUHOLLAHI, Z., BEHESHTI, A., MOUSAEIRAD, S., GOLUGURI, S. R. (2021). Towards Artificial Intelligence Enabled Financial Crime Detection. ArXiv abs/2105.10866, pp. 538–546, <https://doi.org/10.1145/3487664.3487740>
- SAVAGE, D., WANG, Q., CHOU, P., ZHANG, X., YU, X. (2016). Detection of money laundering groups using supervised learning in networks. AAAI-17 Workshop on AI and Operations Research for Social Good, Australia, <https://doi.org/10.1108/JMLC-07-2019-0055>
- TÓTH Z. B. (2018). Magyarország válaszlépései a pénzmosással és terrorizmusfinanszírozással kapcsolatos kihívásokra. *Polgári Szemle*, 14. évf., 1–3. szám, 418–427. oldal, <https://doi.org/10.24307/psz.2018.0832>
- WANG, S. N., YANG, J. G. (2007). A Money Laundering Risk Evaluation Method Based on Decision Tree. IEEE. The 6th International Conference on Machine Learning and Cybernetics, 1, pp. 283–286, ISBN:978-1-4244-0972-3, <https://doi.org/10.1109/ICMLC.2007.4370155>
- WATKINS, R. C., REYNOLDS, K. M., DEMARA, R. E., GEORGIPOULOS, M., GONZALEZ, A. J., EAGLIN, R. (2003). Tracking dirty proceeds: Exploring Data Mining Technologies as Tools to Investigate Money Laundering. *Police Practice and Research: An International Journal*, Vol. 4, No. 2, pp. 163–178, ISSN 1561-4263, <https://doi.org/10.1080/15614260308020>
- WEGBERG, VAN R., OERLEMANS, J.-J., DEVENTER, VAN O. (2018). Bitcoin money laundering: mixed results? An explorative study on money laundering of cybercrime proceeds using bitcoin. *Journal of Financial Crime*, Vol. 25, No. 2, pp. 419–435, <https://doi.org/10.1108/JFC-11-2016-0067>
- WEI, W., LI, J., CAO, L., OU, Y., CHEN, J. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web* 16, pp. 449–475 (2013), <https://doi.org/10.1007/s11280-012-0178-0>
- ZHANG, Y., TRUBEY, P. (2019). Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection. *Computational Economics*, 54, pp. 1043–1063, <https://doi.org/10.1007/s10614-018-9864-z>
- BIS (2001). Consultative Document. Customer Due Diligence for Banks. Technical report, <https://www.bis.org/publ/bcbs85.htm>

FATF (2012). 40 Recommendations. Online: <https://www.fatf-gafic.org/documents/fatf-40r>

FATF (2019). Hungary's progress in strengthening measures to tackle money laundering and terrorist financing. Moneyval 3rd Follow Up Report Hungary, <https://www.fatf-gafi.org/publications/mutualevaluations/documents/fur3-hungary-2019.html>

FATF (2021). Opportunities and Challenges of New Technologies for AML/CFT, FATE, Paris, France, <https://www.fatf-gafi.org/publications/fatfrecommendations/documents/opportunities-challenges-newtechnologies-aml-cft.html>

FATF (2021a). Anti-money laundering and counter-terrorist financing measures Hungary 4th Enhanced Follow-up Report April 2021, <http://www.fatf-gafi.org/media/fatf/documents/reports/fur/Moneyval-FUR-Hungary-2021.pdf>

FATF (2021b). International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation, The FATF recommendations. FATE, Paris, France, <http://www.fatf-gafi.org/recommendations.html>

FSB (2017). Artificial intelligence and machine learning in financial services Market developments and financial stability implications. <https://www.fsb.org/wp-content/uploads/P011117.pdf>

Jogszabályok

2017. évi LIII. törvény, a pénzmosás és terrorizmus finanszírozása megelőzéséről és megakadályozásáról. <https://njt.hu/jogszabaly/2017-53-00-00>

A bizottság (EU) 2017/565 felhatalmazáson alapuló rendelete. <https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32017R0565&from=de>