GAZDASÁGINFORMATIKA DOKTORI ISKOLA

TÉZISGYŰJTEMÉNY

Brunner Csaba

BEHATOLÁS DETEKTÁLÁS GÉPI TANULÁS ÁLTAL

című Ph.D. értekezéshez

TÉMAVEZETŐ:

Dr. Kő Andrea

egyetemi professzor

Dr. Fodor Szabina

egyetemi docens

BUDAPEST, 2020

INFORMATIKAI INTÉZET

TÉZISGYŰJTEMÉNY

Brunner Csaba

BEHATOLÁS DETEKTÁLÁS GÉPI TANULÁS ÁLTAL

című Ph.D. értekezéshez

TÉMAVEZETŐ:

Dr. Kő Andrea

egyetemi professzor

Dr. Fodor Szabina

egyetemi docens

TARTALOMJEGYZÉK

1. MOTIVÁCIÓ ÉS A KUTATÁS CÉLJA

Az igény az információrendszereket és informatikai erőforrásokat érintő visszaélésekkel szembeni védelemre már 1972-ben és 1980-ban is felmerült, amikor James P. Anderson körvonalazta, hogy az amerikai légierő egyre jobban tudatában van az információbiztonsági kihívásoknak (Anderson, 1972, 1980). Azóta a jelentett behatolások száma ijesztő mértékben növekedett, különösen a 2000-es évek korai szakaszától. Ezek súlyossága egyes jelentések, például (Beek *et al.*, 2019), szerint csak növekedett. A disszertáció készítésekor a legelterjedtebb kibertámadások a következők voltak:

- DDoS támadások a 2000-es évek korai szakaszában (Lau *et al.*, 2000; Smith, 2014), jelentős bevételkiesést okozva változatos szolgáltatások megszakításával.
- A DDoS támadásokkal kapcsolatban botnet fertőzések (Smith, 2014), legitim felhasználók elől számítási erőforrások eltulajdonítása, majd az erőforrások felhasználása illegális tevékenységekre.
- Specializált váltságdíj követelő szoftverek (Beek *et al.*, 2019), amik fontos információkat tesznek elérhetetlenné titkosítás segítségével és váltságdíjat követelnek annak feloldásáért.
- És újabban, deepfake támadások (Damiani, 2019; Statt, 2019), amikor mélytanuló modellek segítségével kiemelt pozícióban jelenlévő érdekelt feleket személyesítenek meg, hogy érzékeny információkhoz jussanak, vagy csalást kövessenek el.

A támadások mennyisége függ a vállalkozás által folytatott gazdasági tevékenység jellegétől is, a legveszélyeztetettebb ágazatok a pénzügyi szolgáltatási szektor, az egészségügy és a közoktatás. Számos módszer ismert ezen rosszindulatú tevékenységek megfékezésére, ezeket érdemes különböző rétegekbe szervezni. Az ilyen típusú szervezést a szakma a „defense in depth” kifejezéssel illette, ennek egyik eszköze lehet a gépi tanulás alkalmazása. Egy behatolás jól körül határolható viselkedésbeli mintákkal rendelkezik, ezek detektálása nem okozhat gondot egy specializált, gépi tanuló algoritmusokkal támogatott rendszer számára. Továbbá, bizonyos esetekben, mint a deepfake támadások, a gépi tanulás lehet az egyetlen hatásos módja a védekezésnek.

1.1. A KUTATÁS CÉLJAI

Mindezek ellenére a gépi tanuló technikák továbbra sem eléggé elterjedtek és használtak az IT biztonság területén belül. Ez adta a motivációt a hálózati behatolás detektálás tanulmányozására a gépi tanulás szemszögéből. A fő cél egy újszerű behatolás detektáló megoldás kifejlesztése volt gépi tanuló módszerek alkalmazásával. Ezt két további célra lehetett lebontani:

KC 1. Egy olyan behatolás detektáló modell létrehozása, ami a kiválasztott teljesítmény mutatók alapján képes felvenni a versenyt a kapcsolódó irodalomban bemutatott javaslatokkal. Teljesítmény alatt ebben a kontextusban a helyesen és helytelenül detektált támadások arányát kell érteni összehasonlítva a normál tevékenységgel, és fordítva. Ennek mérésére több alkalmas mutatószám is ismert, például az accuracy és a recall mutatók.

KC 2. Olyan, a gépi tanulás területén alkalmazható technikák feltárása, amelyek segítségével kiegyensúlyozatlan célváltozóval rendelkező komplex események detektálását lehet javítani. A behatolás detektálás megfelel a komplex események leírásának, mert a rendelkezésre álló adatok jelentős része a normál forgalom részét képezi, és csak egy apró hányada számít valóban rossz szándékú tevékenységnek.

1.2. KUTATÁSI KÉRDÉSEK

A kutatás céljai alapján az alábbi kérdések kerültek megfogalmazásra meg:

KK 1. A gépi tanulás egy jó megközelítés behatolások detektálására? Milyen modellek teljesítenek jól egy behatolás detektálási kontextusban?

KK 2. Melyik behatolás detektálási modell típus eredményesebb a vizsgált kontextusban?

KK 3. Milyen pontossági szint várható el egy behatolás detektálási feladat elvégzése során?

KK 1 a megfelelő gépi tanuló modell megtalálásában érdekelt, ami önmagában is egy kihívást jelentő feladat. Befolyásolja a választott behatolás detektálási módszer (szignatúra vagy anomália detektálás) éppúgy, mint a rendelkezésre álló adathalmaz vagy a választott mintavételezési eljárás.

A legelterjedtebb és legjobban teljesítő nem-gyűjteményes gépi tanuló algoritmusok a behatolás detektálás területén a döntési fa, a mesterséges neurális hálózat és a k-legközelebbi szomszéd algoritmusok. Ezek mindegyikét szignatúra detektálásra alkalmazzák, a fő hátrányok:

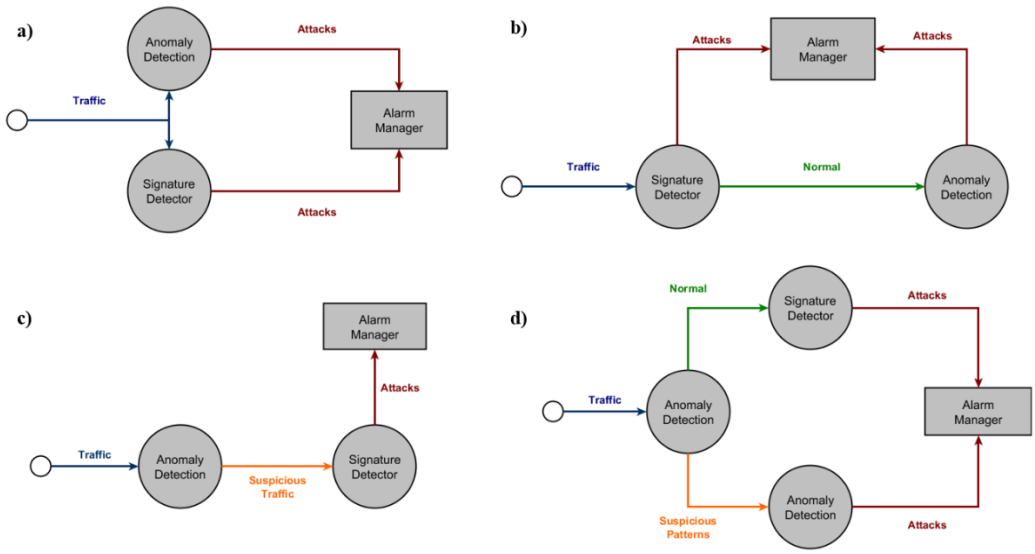
- A döntési fák hajlamosak a túltanulásra, instabil eredményt állítanak elő (egy apró változtatás a tanító adatokban teljesen eltérő döntési fákat eredményezhet) és gyengén teljesítenek kiegyensúlyozatlan tanító osztályok jelenlétében.
- A mesterséges neurális hálózatok, akárcsak a döntési fák, hajlamosak túltanulni és általában sok időt vesz igénybe a tanításuk.
- A k-legközelebbi szomszéd algoritmus gyorsan tanul, de a pontos előrejelzésekhez minden adatra szükségük van, ezért nehezen lehet őket több adatra alkalmazni.

Az előrejelzési pontosság, bár fontos, de nem az egyetlen jellemző, ami alapján a behatolás detektáló megoldásokat össze lehetne hasonlítani. A tanítási és az előrejelzési időtaram, illetve a modell hordozhatósága három további jellemző, amit figyelembe lehet venni az összehasonlítások során. Ezek kiértékelésére nem képezte a disszertáció részét az előrejelzések átfogóbb tanulmányozása érdekében.

KK 2 a kutatási területen belül egy aktuálisabb kérdés, amire (Dua and Du, 2016) is felhívta a figyelmet. Ezt a kérdést a két fő behatolás detektálási technika, a szignatúra és az anomália detektálás közötti különbségek jellemzik. Egyrészt, a szignatúra detektálás magas recall-t és alacsony téves riasztási rátát képes elérni, könnyű az ilyen algoritmusokat implementálni és gyorsan adnak előrejelzéseket. Ugyanakkor nem képesek új és ismeretlen támadásokat felismerni. Másrészt az anomália detektálás a normál forgalomból készít profilokat, majd a profilok alapján tesz különbséget a szokatlan vagy támadó és a normál forgalom között. Az anomália detektálás jobban fel tudja ismerni az új támadásokat, de nehezebb számára a megkülönböztetés a támadások és a szokatlan forgalom között, mivel az utóbbi tartalmazhat szokatlan de normál forgalmat is. Ezért az anomália detektálás magasabb téves riasztási rátát eredményez. Mivel a szignatúra és az anomália detektálás egymást kiegészítő módszerek, ezért érdemes lehet őket valamilyen módon kombinálni egy hibrid detektáló modellben.

Egy egyszerű kombinációja a kettő technikának nem elég, ennél célratörőbb megközelítést kell követni. A hibridizációra olyan modellek szolgálhatnak jó jelölteket,

amelyek nem hajtanak végre ellentétes műveleteket a szolgáltatott adatokon, mint például döntési fák és egyosztályos SVM modellek, vagy bármely tetszőleges autoencoder hálózat kombinációja egy tetszőleges mesterséges neurális hálózattal. Az integráció mikéntjét négy alternatíva közüli választásra egyszerűsítették (**Ábra 1**) (Dua and Du, 2016; Molina-Coronado *et al.*, 2020).



Ábra 1: A hibrid behatolás detektálás típusai. Forrás: (Molina-Coronado *et al.*, 2020)

A négy lehetségesen beazonosított hibrid behatolás detektálási technika:

- **Párhuzamos detektálás**: a szignatúra és anomália detektálási eredmények korrelálásával állít elő pontosabb detektálásokat (**Ábra 1.a**). A hálózati forgalmat támadásnak tekintik, ha akár az anomália, akár a szignatúra detektáló modell annak tekinti azt.
- **Szignatúra-anomália szekvenciális detektálás**: a detektálási képességet javítja azon ismeretlen támadásokon, amiket a szignatúra detektor nem ismert fel (**Ábra 1.b**).
- **Anomália-szignatúra szekvenciális detektálás**: a téves riasztási rátát csökkenti (**Ábra 1.c**). Az anomália detektor előbb megjelöli a gyanús forgalmat, majd azt a szignatúra detektor erősíti vagy cáfolja meg.
- **Összetett kevert detektálás**: bármely anomália és szignatúra detektálást alkalmazó megközelítés, amit a fenti kategóriák egyikéhez sem lehetett besorolni (például: **Ábra 1.d**).

Mivel a tanulmányozott adathalmazok mindegyike rendelkezett dedikált célváltozóval, ezért elsősorban szignatúra és hibrid detektálási eljárások jöhettek szóba az elemzések során.

A **KK 3** kérdéshez fűződően, a kapcsolódó irodalom elemzése alapján az alábbi kihívásokkal kell szembenézni behatolás detektálás tudományterületén belül:

- Az accuracy mutató túlnyomó használata kiegyensúlyozatlan osztályokkal rendelkező adatokon elért teljesítmény értékelésében
- A különböző cikkek saját mintákat vettek a kiválasztott adathalmazokból, ezzel megnehezítve, ha nem ellehetetlenítve az összehasonlítást az eredményeik között.
- A behatolás detektálásnak mindig fontos kérdése az alul reprezentált osztályok kezelése.

A modell teljesítmény mérését sokféleképpen el lehet végezni, ezért a javasolt modellek összehasonlítása érdekében minden cikknek két követelménynek kellet megfelelnie, a disszertáció feltételezéseinek összehasonlításához:

- **Nagyobb hangsúly a recall-on / detektálási rátán**: habár az accuracy az elterjedt mutató, de nem felel meg a detektálási teljesítmény mérésére kiegyensúlyozatlan adatokon. A recall e tekintetben egy jobb mutató. A disszertáció előnyben részesíti azokat a kutatásokat, ahol a recall mutatót alkalmazták a teljesítmény mérésére, szemben azokkal, amelyek accuracy-t használtak, noha az elterjedtsége miatt az accuracy mutatót nem lehet teljesen figyelmen kívül hagyni.
- Az adat mintavételezés a nehézségek és az előrejelzések szóródásának másik forrása a tanulmányozott kutatásokban. Más minták más modelleket eredményeztek eltérő mérésekkel. Ezért elsősorban olyan kutatások kerültek kiválasztásra, ahol a teljes dedikált teszt adathalmazokon tesztelték a modellek teljesítményét. Hasonlóképpen, a disszertációban bemutatott behatolás detektáló modellek tesztelésére a teljes teszt minta került felhasználásra, függetlenül attól, hogy a modellek milyen tanító adatokhoz fértek hozzá. Ez a mentalitás már az adat előfeldolgozás során is megjelent, ahol a teszt adatokon elvégzett transzformációk a tanító adatokból kapott számítások alapján történtek, elkerülve az információ szivárgást.

A kapcsolódó kutatásokkal való összehasonlítások mellett **KK 3** a behatolás detektálás lehetséges támogató módszereinek feltárásában is érdekelt. E módszerek:

- **Szintetikus mintavételezés**: mesterséges mintákat hoz létre az alul reprezentált osztályokból, növelve ezzel létszámukat az adathalmazon belül. Ilyen módszer például a SMOTE (synthetic minority oversampling technique - (Chawla *et al.*, 2002)) és változatai (SVM SMOTE - (Nguyen, Cooper and Kamei, 2009), SMOTE Tomek és SMOTE ENN - (Batista, Prati and Monard, 2004))
- **Haladó hiperparaméter optimalizálás**: a hiperparaméter optimalizálás egy automatizált döntés gépi tanuló algoritmusok olyan paraméterei között, amelyeket maga az algoritmus nem optimalizál, adottságként kezel. A hiperparaméter optimalizálásnak léteznek egyszerű módszerei is, de a disszertációban kifejezetten a fejlettebb, gaussian process-eken alapuló Bayes optimalizáció (Brochu, Cora and De Freitas, 2010; Snoek, Larochelle and Adams, 2012) és a TPE (tree-structured parzen estimators - (Bergstra *et al.*, 2011; Bergstra, Yamins and Cox, 2013)) módszer kerültek bemutatásra, a detektálási pontosság további javítása érdekében.
- **Gyűjteményes modellek**: módszerek egyedi gépi tanuló modellek kombinálására egy összesített osztályozóban az elfogultság (bias - boosting), a variancia (bagging) vagy mindkettő (stacking) csökkentésére (Smolyakov, 2017; Budzik, 2019).

2. HÁTTÉR

A behatolás és behatolás detektálás egy lehetséges definíciója: *„Bármely engedély nélküli kísérlet információ hozzáférésére, manipulálására, módosítására vagy megsemmisítésére vagy egy informatikai rendszer távoli kiaknázása spam küldésére vagy gépek feltörésére, módosítására. Egy IDS intelligens módon megfigyelés alatt tart minden, a számítási erőforráson zajló tevékenységeket, például a hálózati forgalmat vagy gépfelhasználást, hogy elemezze az eseményeket és válaszreakciókat állítson elő”* (Dua and Du, 2016, p. 10). Ez a definíció már figyelembe veszi a botnet hálózatokból fakadó tevékenységeket és lefedi a behatolás detektálás két fő típusát is. Egy másik definíció behatolás detektálási rendszerre: *„A behatolás detektálási rendszereket olyan kibertámadások feltárására vezették be, amelyek információrendszerekben okozhatnak károkat.”* (Molina-Coronado *et al.*, 2020, p. 2). Ebben a disszertációban az IDS fogalma alatt olyan rendszert kell érteni, amit egy tágabb külső hálózatból egy védett információrendszerhez beérkező nem engedélyezett hozzáférések detektálására terveztek. Egy ilyen rendszer működéséhez kulcsfontosságú, hogy a behatolásokat meg lehessen különböztetni a normál működéssel járó tevékenységektől.

A behatolás detektálási rendszerek alábbi két fő típusa között tesznek különbséget a védett célpont alapján (Scarfone and Mell, 2007; Dua and Du, 2016; Molina-Coronado *et al.*, 2020):

- **Hálózati** (network IDS - NIDS): hálózati szegmenseken vagy eszközökön áthaladó forgalmat figyel a hálózatra kötött eszközök felé intézett ártó szándékú forgalom felfedezése érdekében. A hálózati behatolás detektáló rendszereket általában DMZ-kben, intelligens tűzfalak részeként, VPN szervereken, távoli hozzáférés szervereken és vezeték nélküli hozzáférési pontokon alkalmazzák.

- **Hoszt alapú** (host IDS - HIDS): egy kiemelt rendszer erőforrásait figyeli meg gyanús viselkedés után kutatva. Ez a rendszer általában egy kritikus IT infrastruktúra eszköz, tipikusan egy alkalmazás vagy adatbázis szerver.

Adatbányászati szemszögből (Scarfone and Mell, 2007; Dua and Du, 2016; Molina-Coronado *et al.*, 2020) az alábbi IDS típusokat különböztette meg:

- **Szignatúra detektálás**: olyan IDS, ami egy ismert behatolás felfedezésekor riasztást ad. Az ismert támadásokat megbízható módon képes felismerni alacsony

téves riasztási rátával, de nem képes új támadásokat detektálni. A szignatúra detektáló rendszerek a már bekövetkezett támadásokból építenek mintákat, ezért szükségük van korábbi támadásokból származó adatokra.

- **Anomális detektálás**: akkor riaszt, ha a forgalom szignifikánsan eltér a normál forgalomból készített minták által meghatározott forgalomtól. Ennek következtében képesek detektálni ismeretlen támadásokat is cserébe magasabb téves riasztási rátával.

- **Hibrid detektálás**: az IDS teljesítmény javítása érdekében néhány kutató az anomália és szignatúra detektálás kombinálását javasolta egy hibrid megoldásban. A háttérben húzódó ötlet a két módszer előnyeinek kombinálása volt, mint az ismert támadások detektálása alacsony téves riasztási ráta mellet, miközben a rendszernek megtartja új támadások felismerésérnek a képességét is.

Az adatbányászatra is több definíciót használnak, az egyiket (Fayyad, Piatetsky-Shapiro and Smyth, 1996) adta meg, mint a tágabb adatbázisokban végzett tudásfeltárás (KDD – knowledge discovery in databases) folyamat egy lépése. A KDD-t mint a *„hasznos tudás adatokból való feltárásának átfogó folyamata”* lehet megfogalmazni (Fayyad, Piatetsky-Shapiro and Smyth, 1996, p. 40), az adatbányászatot pedig mint egy „statisztikai, matematikai és mesterséges intelligencia technikákat alkalmazó folyamat hasznos információk és későbbi tudás kinyerésére és azonosítására nagyméretű adathalmazokból”. Egy IDS szempontjából a rejtett tudás a forgalom küldőjének a mögöttes szándéka, az adat meg a beérkező hálózati forgalom maga. A cél az ártó szándékú forgalom elválasztása a normáltól.

Az adatbányászat és gépi tanulás fogalmait, értelmezéstől függően, gyakran használják egymás szinonimáiként. Ez a disszertáció a gépi tanulásra az alábbi definíciót használja: *„olyan kutatási terület, ami a számítógépeknek megadja a képességet a tanulásra anélkül, hogy kifejezetten arra programozták volna őket”* (Samuel, 1959, közvetett hivatkozás). Ennek alapján az adatbányászat az érdekes tudásminták feltárására szolgáló tevékenység a KDD folyamaton belül (Hiba! A hivatkozási forrás nem található.). A gépi tanuló algoritmusokat az adatbányászat gyakran, de nem kizárólagosan, a tudásminták előállítására használja fel.

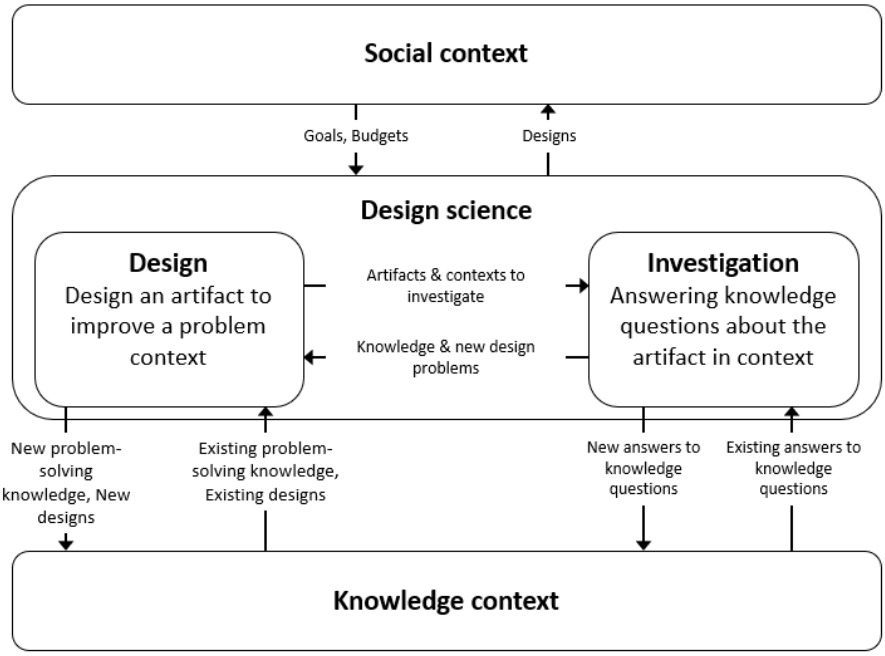
3. MÓDSZERTAN

A javasolt modellek megtervezésére, futtatására és kiértékelésére használt módszertan top-down megközelítést követ (**Ábra 2**). A disszertációban feltett kutatási kérdésekre a választ egy algoritmikus eredménytermék segítségével lehet megkapni. A design science célja, hogy kérdéseket válaszoljon meg egy eredménytermék tervének segítségével. A kutatási tudomány módszertanilag így illeszkedik a disszertációban szereplő kutatáshoz, létrehozva ezzel a tanulmány első módszertani alappillérét. Továbbá, mivel az eredménytermék egy gép tanuló modell, ezért a gépi tanuló modellek tervezésére, implementációjára és bevezetésére szolgáló CRISP-DM folyamatmodell fogalmait és megfontolásait is fel lehet használni, megalkotva a második módszertani alappillért. Végül a javasolt modellek tervei írják elő a modell elkészítés konkrét folyamatát a szükséges adat előkészítési, tanítási és kiértékelési lépésekkel, megalkotva ezzel a harmadik, legalsó szintjét a módszertani absztrakciónak.

Ábra 2: A disszertációban követett módszertan absztrakciós szintjei. Saját szerk.

3.1. A DESIGN SCIENCE ÉS A CRISP-DM

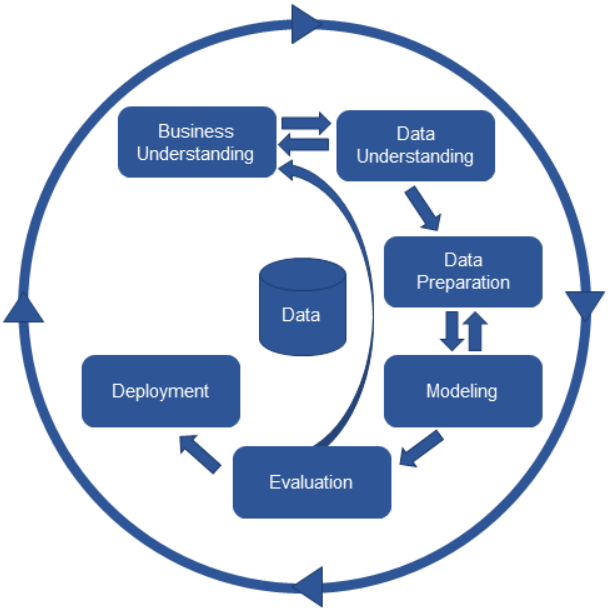
„A design science eredménytermékek tervezése és kivizsgálása éles környezetben. Az általunk tanulmányozott eredménytermékeket úgy tervezték meg, hogy interakcióba lépjenek környezetükben fellépő problémákkal annak érdekében, hogy javítsanak valamit a befogadó környezetben." (Wieringa, 2014, p. 3). A design science egyidejűleg jelenti kutatási kérdések megválaszolását és egy informatikai eredménytermék kifejlesztését (**Ábra 3**).



Ábra 3: A kutatási tudomány keretrendszere. Forrás: (Wieringa, 2014)

A kutatási tudományt annak szociális és tudás környezetei határozzák meg. A szociális környezet a célok és követelmények meghatározásáért és a kutatás erőforrásainak előteremtésért felelős érdekképviselők adják. Az ő elvárásuk az is, hogy a kutatási projekt valamilyen, a célokat teljesítő kézzel fogható eredménytermék kerüljön átadásra. A kutatás során a tudás környezetet használják vizsgálati és tervezési kihívások megoldására, illetve ezt a környezetet gazdagítják a projekt során elkészített tervek és a levont következtetések.

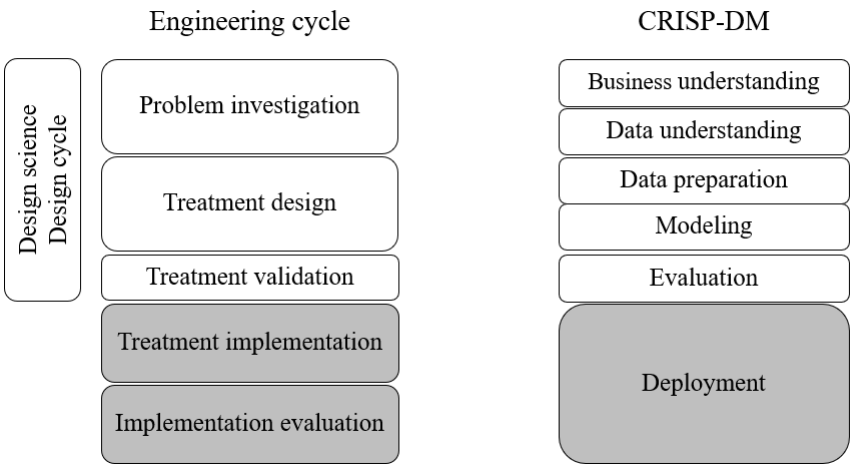
A kutatási tudomány két fő alappillérét a tervezési és mérnöki ciklusok alkotják, ezek felelősek az eredménytermékek létrehozásáért, kiértékeléséért, validációjáért és bevezetésért a célok elérése érdekében. A mérnöki ciklus egy racionális problémamegoldó folyamat egy működőképes termék leszállítására (**Ábra 5**), öt lépésből áll. A tervezési ciklus lefedi ennek az első három lépését, célja a termék tervének elkészítése.



Ábra 4: A CRISP-DM folyamatmodell. Forrás: (Chapman *et al.*, 2000)

Az adatbányászati projektet rendszeres kivitelezésére szükség van egy folyamatmodellre. Az egyik legnépszerűbb ilyen folyamatmodell a (Chapman *et al.*, 2000) tervezte CRISP-DM (cross industry standard process for data mining). A CRISP-DM folyamat (**Ábra 4**) az üzlet és az adatbányászat üzleti szükségességének alapos megértésével kezdődik és egy, az üzleti követelményeknek eleget tevő gépi tanuló modell bevezetésével ér véget.

A két fő módszertani alappillért az egymásnak logikailag megfelelő lépések kötik össze (**Ábra 5**). Például a probléma kivizsgálása a mérnöki ciklusban olyan tevékenységekkel jár együtt, amelyek hasonlítanak a CRISP-DM üzleti és adat megértési fázisaiban végrehajtott tevékenységeknek. Mivel a tervezési ciklus csak a mérnöki ciklus első három lépésében érdekelt, ezért a disszertáció a CRISP-DM folyamat lépéseit csak a kiértékelési fázisig hajtotta végre, a bevezetési feladatok hatáskörön kívül maradtak. Továbbá, a disszertáció körülményeiből fakadóan a szociális környezet részletes elemzésére nem kerülhetett sor, ezért az sem képezte a disszertáció részét.



Ábra 5: A mérnöki ciklus és CRISP-DM közötti kapcsolat. (Chapman et al., 2000; Wieringa, 2014) alapján.

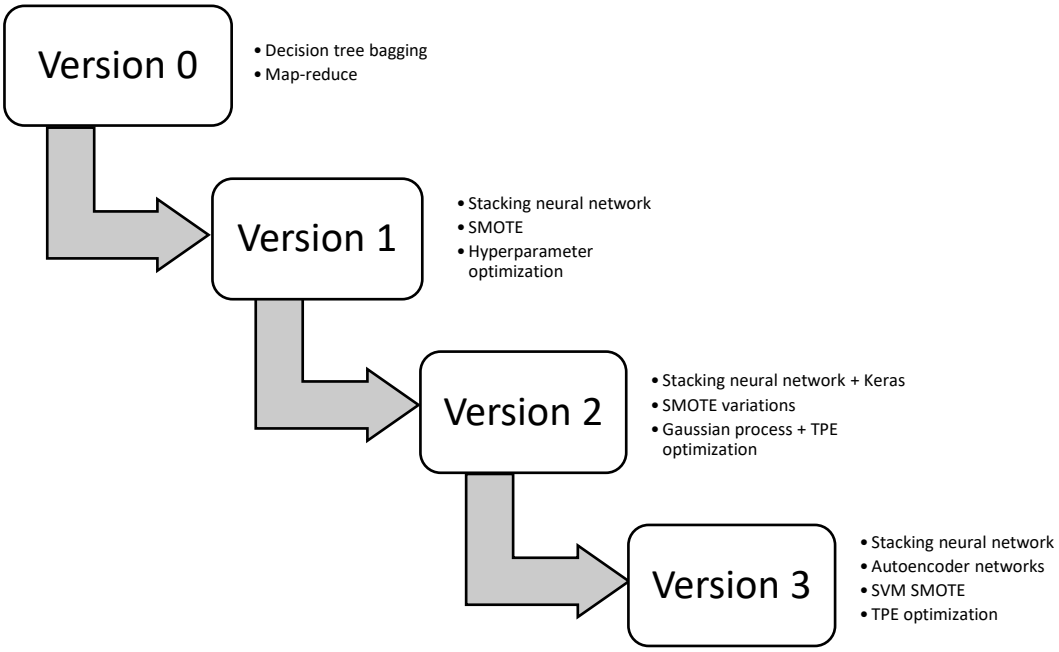
A két módszertan között különbségek is vannak, például kitűzött céljaikban. A design science célja nemcsak egy jól megtervezett működőképes eredménytermék leszállítása, hanem tudományos kérdések megválaszolása az eredménytermékkel, annak környezetével vagy a kettő közötti összefüggésekkel kapcsolatban is. Ezzel szemben a CRISP-DM célja sokkal gyakorlatiasabb. Egy olyan gépi tanuló modell leszállításában érdekelt, ami egy működő üzleti megoldás vagy szolgáltatás részét képezi, értéket teremtve ügyfélnek és a szervezetnek egyaránt. Tehát a CRISP-DM jobban összpontosít az üzleti környezetre és annak hatásaira, mintsem kutatási kérdések megválaszolására.

3.2. JAVASOLT MODELLEK

A harmadik és végső módszertani alappillér elemeit, a javasolt modellek terveit és értékelését iteratív lépésenként a CRISP-DM folyamat modellnek megfelelően, adatelőkészítésre, modellezésre, kiértékelésre lebontva kerültek bemutatásra. A tervek implementációja során néhány javaslat is felmerült a modellek továbbfejlesztését illetően, ezeket külön bekezdés dolgozta fel. A detektálási eredmények bemutatása és elemzése nem a tervezési fázisban, hanem külön fejezetben került feldolgozásra. Az egyes modell változatok tervei egy behatolás detektáló modell egyre kifinomultabb változatait adták meg (**Ábra 6**):

- **0. verzió** (V0 - prototípus): a modell első prototípusát (Brunner, 2017) körvonalazta, ahol egy map-reduce-szerű architektúrán tanított döntési fákból alkotott bagging osztályozó került kiértékelésre.

- **1. verzió** (V1 – neurális hálózatokból alkotott stacking gyűjteményes modell): ahol egy eltérő változókon tanított neurális hálózatokból alkotott stacking gyűjteményes modell épült fel. A modell teljesítményét egy robusztusabb mintavételezés és grid search hiperparaméter optimalizáció támogatta.
- **2. verzió** (V2 - áttérés Keras-ra és TensorFlow backend-re): a neurális hálózatokból alkotott gyűjteményes modell átköltöztetésre került Keras-ra TensorFlow backend mellett, javítva ezzel a tanítási folyamaton. A további javulás a detektálási pontosságban a TPE hiperparaméter optimalizáció használatától várható. Egy másodlagos cél e változat használatával a különböző SMOTE mintavételezési változatok, konkrétan a SMOTE ENN, a SMOTE Tomek és az SVM SMOTE kiértékelése volt.
- **3. verzió** (V3 – bővítés autoencoder hálózatokkal): a korábbi legjobban teljesítő elemek, mint az SVM SMOTE mintavételezés és a TPE optimalizáció, kibővítése mély autoencoder hálózatokkal, mely hálózatok csak a normál forgalmon tanultak, így hozva létre egy hibrid behatolás detektáló modellt.



Ábra 6: A disszertációban szereplő behatolás detektáló modell iterációi. Saját szerk.

A behatolás detektáló változatok két, egymással kapcsolatban álló, adatbázison tanultak és lettek kiértékelve, a KDD Cup 1999 (Stolfo *et al.*, 2000) és az NSL-KDD (Tavallaee *et al.*, 2009) adatbázisokot, amelyeket a mai napig behatolás detektálási modellek összehasonlításához viszonyítási alapként használnak. A két adatbázis között a

különbség, hogy az NSL-KDD adatbázis kijavította a KDD Cup 1999 redundanciáját, ami a korábban készített modellek elfogultságát okozta az ismétlődő megfigyelésekkel szemben. A javasolt modellek közül a V0 változat csak a KDD Cup 1999, a V1 mindkettőn és V2-től kezdve a többi változat már csak az NSL-KDD lett kiértékelve. Ugyan régi adatbázisok, de viszonyítási alapként még mindig értékesek, illetve azzal, ahogy az új támadások megjelenését modellezik olyan támadási kategóriákkal, amik csak a teszt adatokban jelennek meg.

4. A KUTATÁSI EREDMÉNYEK ÖSSZEFOGLALÓJA

Az eredmények (**Táblázat 1**, **Táblázat 2** és **Táblázat 3**) a V1-3 modell változatok mért teljesítmény mutatóit hasonlítják össze egymással és a kapcsolódó irodalomból válogatott különböző javaslatokkal. Az összehasonlítás kulcsmutatói a recall és az accuracy mérőszámok voltak. A **KK 1**-re válaszolva a gépi tanulás bebizonyítottan alkalmas módszer behatolások detektálására, de a valódi hatásosságban van különbség az egyes modellek között. Az elemzett irodalom alapján egyfajta hierarchia állítható fel az egyes modellek között kezdve az egyszerű, egy osztályozót tartalmazó modellektől, a gyűjteményes és hibrid megoldásokon át egészen az adatgeneráló módszerekkel (variációs autoencoder) kiegészített modellekig, ebben a sorrendben. Ezt bizonyítják a disszertáció V1-3-ig haladó modell változatai is.

	V1	V2			V3
		SMOT ENN	SMOTE Tomek	SVM SMOTE	
Normal	0,9452	0,9255	0,9198	0,9140	0,8367
DoS	0,8126	0,8259	0,8592	0,8438	0,7728
Probe	0,6898	0,5225	0,5580	0,5944	0,7732
R2L	0,2400	0,3258	0,3289	0,3109	0,3262
U2R	0,1045	0,3731	0,2985	0,2985	0,5821
Átlag	0,5584	0,5946	0,5929	0,5923	0,6582

Táblázat 1: Modell recall táblázat. Saját szerk.

KK 2-re térve, a recall mérések alapján (**Táblázat 1**) a V3 modell változat a többségi osztályokon elért teljesítményt feláldozva ért el jobb eredményeket az alul reprezentált osztályokon, ami kifejezetten igaz volt az U2R osztályra. Ez összességében emelte a modell makro-átlagolt recall értékét. Az accuracy számítások már nem hoztak hasonló eredményt, ott a V3 modell teljesített a legrosszabban (74,26%), miközben a V2 SMOTE Tomek hozta a legjobb eredményeket (78,34%-kal). A választott modell függ a behatolás detektálás céljától is, ahol egy támadás hibás osztályozásának normál forgalomként messzemenő következményei lehetnek. Ezeket a következményeket a recall mutató képes jobban megragadni, ezért a gyakorlatban a V3 modell alkalmazása indokolt.

Modell	Accuracy	Recall
KNN (Yang *et al.*, 2019)	76,51%	48,3%
Multinomial NB (Yang *et al.*, 2019)	78,73%	47,69%
RF (Yang *et al.*, 2019)	76,49%	48,84%
SVM (Yang *et al.*, 2019)	72,28%	45,88%
DNN (Yang *et al.*, 2019)	80,22%	52,77%
DBN (Yang *et al.*, 2019)	80,82%	53,61%

Modell	Accuracy	Recall
ROS-DNN (Yang *et al.*, 2019)	78,26%	49,59%
SMOTE-DNN (Yang *et al.*, 2019)	81,16%	51,49%
ADASYN-DNN (Yang *et al.*, 2019)	80,1%	51,47%
ICVAE-DNN (Yang *et al.*, 2019)	85,97%	62,66%
VGM + RF (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	73,61%	N/A
VGM + Logistic Regression (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	77,29%	N/A
VGM + Linear SVM (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	77,23%	N/A
VGM + MLP (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	79,26%	N/A
SVM SMOTE + RF (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	74,25%	N/A
SVM SMOTE + Logistic Regression (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	76,29%	N/A
SVM SMOTE + Linear SVM (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	77,99%	N/A
SVM SMOTE + MLP (Lopez-Martin, Carro and Sanchez-Esguevillas, 2019)	77,98%	N/A
Decision Tree (Yin *et al.*, 2017)	74,6%	N/A
NB (Yin *et al.*, 2017)	74,4%	N/A
RF (Yin *et al.*, 2017)	72,8%	N/A
NB Tree (Yin *et al.*, 2017)	75,4%	N/A
MLP (Yin *et al.*, 2017)	78,1%	N/A
RNN (Yin *et al.*, 2017)	81,29%	N/A
SAE + SMR (Javaid *et al.*, 2016)	79,1%	N/A
AE + SVM (Al-Qatf *et al.*, 2018)	80,48%	N/A
Javasolt V3 (AE + Stacking NN)	74,26%	65,82%
Javasolt V2 + SMOTE ENN	77,09%	59,46%
Javasolt V2 + SMOTE Tomek	78,34%	59,29%
Javasolt V2 + SVM SMOTE	77,75%	59,23%
Javasolt V1 (Stacking NN)	78,11%	55,84%

Táblázat 2: Recall és accuracy összehasonlítása külső forrásokkal. Saját szerk.

A kigyűjtött eredmények olyan cikkekből származnak, amelyek autoencoder hálózatokkal kiegészített hibrid behatolás detektáló eljárásokat elemeztek. Az elemzés során e kutatások egyszerűbb modell eredményeket is feltüntettek, ezek is bekerültek az összehasonlító táblázatokba (**Táblázat 2** és **Táblázat 3**). A **KK 3**-ra válaszolva, az összes modell változat a kapcsolódó irodaloban szereplő eredményekből számított átlag felett teljesített, ebből a V3 adta a legjobb recall-t. Továbbá (Yang et al., 2019) elérhetővé tett osztályszintű recall értékeket is, lehetővé téve a részletesebb elemzést is (**Táblázat 3**).

Modell	Normal	DoS	Probe	R2L	U2R
KNN (Yang *et al.*, 2019)	92,78%	82.25%	59.4%	3.56%	3.5%
Multinomial NB (Yang *et al.*, 2019)	96,03%	37.1%	82.61%	22.22%	0.5%
RF (Yang *et al.*, 2019)	97,37%	80.24%	58.53%	7.55%	0.5%

Modell	Normal	DoS	Probe	R2L	U2R
SVM (Yang *et al.*, 2019)	92,82%	74,85%	61,71%	0%	0%
DNN (Yang *et al.*, 2019)	96,1%	85,4%	65,3%	14,56%	2,5%
DBN (Yang *et al.*, 2019)	97,04%	83,11%	69,85%	12,56%	5,5%
ROS-DNN (Yang *et al.*, 2019)	92,61%	80,32%	56,26%	12,75%	6%
SMOTE-DNN (Yang *et al.*, 2019)	96,59%	82,19%	56,75%	10,93%	11%
ADASYN-DNN (Yang *et al.*, 2019)	96,43%	83,28%	59,81%	9,84%	8%
ICVAE-DNN (Yang *et al.*, 2019)	97,26%	85,65%	74,97%	44,41%	11%
Proposed V3 (AE + Stacking NN)	83,67%	77,28%	77,32%	32,62%	58,21%
Proposed V2 + SMOTE ENN	92,55%	82,59%	52,25%	32,58%	37,31%
Proposed V2 + SMOTE Tomek	91,98%	85,92%	55,80%	32,89%	29,85%
Proposed V2 + SVM SMOTE	91,40%	84,38%	59,44%	31,09%	29,85%
Proposed V1 (Stacking NN)	94,52%	81,26%	68,98%	24,00%	10,45%

Táblázat 3: Osztályszintű recall összehasonlítás. Forrás: (Yang et al., 2019) és saját szerk.

Az átlagos recall 95,5% volt a normál, 77,44% a DoS, 64,52% a probe, 13,84% az R2L és 4,85% az U2R osztályok esetében. A disszertációban javasolt modell változatok mind alulteljesítették az átlag értéket normál osztályon, a V3 modell kivételével túlteljesítették DoS forgalmon, a V2 modellek kivételével túlteljesítették probe támadásokon és mind túlteljesítették R2L és U2R osztályokon. A V3 a legrosszabb recallt adta normál és DoS támadásokon összehasonlítva a mérésekkel. Ugyanakkor jobban teljesített a probe, U2R és R2L támadások detektálásában. Elmondható, hogy a V3 modell a felülreprezentált osztályokon elért jó teljesítményt az alul reprezentált osztályokon elért jobb osztályozással kompenzálta.

A **KK 3**-mal kapcsolatban egy másodlagos cél olyan módszerek azonosítása volt, amik segíthetnek a modelleknek a pontosabb behatolás detektálás elérése érdekében. E célból a V2 változatot a SMOTE több változatával lettek tesztelve. Az elért eredmények alapján (**Táblázat 1**) egyetlen változat sem tudott szignifikánsan jobb eredményeket elérni a többinél, de a szintetikus mintavételezés alkalmazása határozottan javította a pontosságot. A másik alkalmazott technika TPE algoritmust alkalmazó fejlettebb hiperparaméter optimalizáció volt, amire példát a V2 és V3 modell változatok adtak.

5. PUBLIKÁCIÓK

Folyóiratcikkek

Brunner, C. (2019): A comparative study of Antminer+ and Decision Tree classification performances. *SEFBIS*, 13 szám, pp. 15-23.

Brunner, C. (2017): Processing Intrusion Data with Machine Learning and MapReduce. *ACADEMIC AND APPLIED RESEARCH IN MILITARY AND PUBLIC MANAGEMENT SCIENCE*, 16 szám, pp. 37-52.

Absztraktok konferencia kiadványokban

Brunner, C. (2019): Behatolás-detektálás Tensorflow Platformon. *16. Országos Gazdaságinformatikai Konferencia* (p. 32). Budapest: NJSZT Neumann János Számítógép-tudományi Társaság GIKOF Gazdaságinformatikai Kutatási és Oktatási Fórum.

Brunner, C. (2018): Hálózati behatolás detektálás Neurális hálózatok összevonásával. *OGIK'2018 Országos Gazdaságinformatikai Konferencia – Az előadások összefoglalói* (p. 75). Sopron, Alexander Alapítvány a Jövő Értelmiségéért.

Brunner, C. (2017): Ant Colony Algorithm in Data Mining. *XIV. Országos GazdaságInformatikai Konferencia* (p. 31). Győr: Alexander Alapítvány a Jövő Értelmiségéért.

Brunner, C. (2016): Behatolási adatok feldolgozása gépi tanulás és MapReduce segítségével. *XIII. Országos Gazdaságinformatikai Konferencia* (pp. 25-26). Dunaújváros: DUE Press.

Konferencia előadások

Brunner C. (2019): Behatolás-detektálás Tensorflow Platformon Előadva: *16. Országos Gazdaságinformatikai Konferencia;* Nov 8-9; Budapest, Magyarország

Brunner, C. (2018): Hálózati behatolás detektálás Neurális hálózatok összevonásával. Előadva: *OGIK'2018 Országos Gazdaságinformatikai Konferencia;* Nov 9-10; Sopron, Magyarország

Brunner, C. (2017): Ant Colony Algorithm in Data Mining. Poszter előadás: *XIV. Országos GazdaságInformatikai Konferencia*; Nov 10-11; Sopron, Magyarország

Brunner, C. (2016): Behatolási adatok feldolgozása gépi tanulás és MapReduce segítségével. Előadva: *XIII. Országos GazdaságInformatikai Konferncia*; Nov 11-12; Dunaújváros, Magyarország

Egyéb munkák

Brunner, C. (2019): Hálózati behatolás detektálás Neurális hálózatok összevonásával. In L. Bacsárdi, G. Bencsik, & Z. Pödör, *OGIK'2018 Országos Gazdaságinformatikai Konferencia - Válogatott közlemények*. Sopron, Magyarország: Alexander Alapítvány a Jövő Értelmiségéért.

Előkészületben:

Brunner, C., Kő, A., & Fodor, S. (2020): A novel ensemble method based on Neural Networks with hyperparameter optimization for Intrusion Detection. Kézirat

6. HIVATKOZÁSOK

Al-Qatf, M. *et al.* (2018) "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," *IEEE Access*. IEEE, 6, pp. 52843–52856.

Anderson, J. P. (1972) *Computer security technology planning study. volume 2*.

Anderson, J. P. (1980) "Computer security threat monitoring and surveillance," *Technical Report, James P. Anderson Company*.

Batista, G. E., Prati, R. C. and Monard, M. C. (2004) "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*. ACM, 6(1), pp. 20–29.

Beek, C. *et al.* (2019) *McAfee Labs Threats Report August 2019*. Available at: https://www.mcafee.com/enterprise/en-us/threat-center/mcafee-labs/reports.html.

Bergstra, J. S. *et al.* (2011) "Algorithms for hyper-parameter optimization," in *Advances in neural information processing systems*, pp. 2546–2554.

Bergstra, J., Yamins, D. and Cox, D. D. (2013) "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in science conference*, pp. 13–20.

Brochu, E., Cora, V. M. and De Freitas, N. (2010) "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*.

Brunner, C. (2017) "Processing Intrusion Data with Machine Learning and MapReduce," *Academic and Applied Research in Public Management Science*, 16(1), pp. 37–52.

Budzik, J. (2019) *Many Heads Are Better Than One: The Case For Ensemble Learning*. Available at: https://www.kdnuggets.com/2019/09/ensemble-learning.html (Accessed: September 29, 2019).

Chapman, P. *et al.* (2000) "CRISP-DM 1.0: Step-by-step data mining guide," *SPSS inc*, 16.

Chawla, N. V *et al.* (2002) "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Rearch*, 16, pp. 321–357.

Damiani, J. (2019) *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*, *Forbes*. Available at: https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/#3efedc652241 (Accessed: June 7, 2020).

Dua, S. and Du, X. (2016) *Data mining and machine learning in cybersecurity*. CRC press.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) "From data mining to knowledge discovery in databases," *AI magazine*, 17(3), p. 37.

Javaid, A. *et al.* (2016) "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pp. 21–26.

Lau, F. *et al.* (2000) "Distributed denial of service attacks," in *2000 IEEE international conference on systems, man and cybernetics.*, pp. 2275–2280.

Lopez-Martin, M., Carro, B. and Sanchez-Esguevillas, A. (2019) "Variational data generative model for intrusion detection," *Knowledge and Information Systems*. Springer, 60(1), pp. 569–590.

Molina-Coronado, B. *et al.* (2020) "Survey of Network Intrusion Detection Methods from the Perspective of the Knowledge Discovery in Databases Process," *arXiv preprint arXiv:2001.09697*.

Nguyen, H. M., Cooper, E. W. and Kamei, K. (2009) "Borderline over-sampling for imbalanced data classification," in *Proceedings: Fifth International Workshop on Computational Intelligence & Applications*, pp. 24–29.

Samuel, A. L. (1959) "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of*

Research and Development, 3(3), pp. 210–229.

Scarfone, K. and Mell, P. (2007) “Guide to Intrusion Detection and Prevention Systems (IDPS) Recommendations of the National Institute of Standards and Technology,” *Nist Special Publication*, 800–94, p. 127. doi: 10.6028/NIST.SP.800-94.

Smith, S. (2014) *5 Famous Botnets that held the internet hostage*. Available at: https://tqaweekly.com/episodes/season5/tqa-se5ep11.php (Accessed: June 7, 2020).

Smolyakov, V. (2017) *Ensemble Learning to Improve Machine Learning Results*. Aug. Available at: https://blog.statsbot.co/ensemble-learning-d1dcd548e936 (Accessed: March 20, 2019).

Snoek, J., Larochelle, H. and Adams, R. P. (2012) “Practical bayesian optimization of machine learning algorithms,” in *Advances in neural information processing systems*, pp. 2951–2959.

Statt, N. (2019) *Thieves are now using AI deepfakes to trick companies into sending them money*, *The Verge*. Available at: https://www.theverge.com/2019/9/5/20851248/deepfakes-ai-fake-audio-phone-calls-thieves-trick-companies-stealing-money (Accessed: June 7, 2020).

Stolfo, S. J. *et al.* (2000) “Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project,” in *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX’00*. IEEE, pp. 130–144.

Tavallaee, M. *et al.* (2009) “A Detailed Analysis of the KDD CUP 99 Data Set,” in *IEEE Symposium on Computational Intelligence for Security and Defense Applications - CISDA*. IEEE, pp. 1–6.

Wieringa, R. J. (2014) *Design science methodology for information systems and software engineering*. Springer.

Yang, Yanqing *et al.* (2019) “Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network,” *Sensors*. Multidisciplinary Digital Publishing Institute, 19(11), p. 2528.

Yin, C. *et al.* (2017) “A deep learning approach for intrusion detection using recurrent neural networks,” *Ieee Access*. IEEE, 5, pp. 21954–21961.