MISKOLCI EGYETEM

GÉPÉSZMÉRNÖKI ÉS INFORMATIKAI KAR

HATVANY JÓZSEF INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA

RENDELLENESSÉG ALAPÚ BEHATOLÁS ÉRZÉKELŐ RENDSZEREK GÉPI TANULÁSI MÓDSZERREL TÖRTÉNŐ TANÍTÁSA

című PhD értekezés

TÉZIS FÜZETE

KÉSZÍTETTE:
Göcs László
okleveles informatika szakos tanár

DOKTORI ISKOLA VEZETŐ:
Prof. Dr. Szigeti Jenő
egyetemi tanár

TÉMAVEZETŐ:
Dr. habil. Johanyák Zsolt Csaba
főiskolai tanár

Miskolc, 2023.

Védési bizottság

Elnök:	Prof. Dr. Szigeti Jenő,	ME, egyetemi tanár
Tartalék elnök:	Prof. Dr. Kovács László	ME, egyetemi tanár
Tagok:	Prof. Dr. Kovács Szilveszter	ME, egyetemi tanár
	Dr. Pásztor Attila	NJE, főiskolai tanár
	Dr. Király Zoltán	DE, egyetemi docens
Póttag:	Dr. habil. Vásárhelyi József	ME, egyetemi adjunktus
Titkár és tag:	Dr. Hornyák Olivér	ME, egyetemi docens
Póttitkár és tag:	Dr. Veres Péter	ME, egyetemi adjunktus
Bírálók:	Dr. habil. Laufer Edit	ÓE, egyetemi docens
	Dr. Vincze Dávid	ME, egyetemi docens
Pótbíráló:	Prof. Dr. Kővári Attila	EKKE, egyetemi tanár

Tartalomjegyzék

1. Kutatási célok és motiváció

Több mint 10 éve foglalkozom informatikai biztonság szakterülettel. Az egyetemi oktatás mellett vállalatoknak szaktanácsadóként és informatikai igazságügyi szakértőként számtalan esetben találkozom az IT rendszerek sérülékenységeivel és azok problémáival. A jelenleg működő informatikai rendszerek legnagyobb kihívása a biztonság megléte és annak felügyelete. A számítógépes és hálózati biztonsági rendszerek folyamatosan támadásnak vannak kitéve, melyek már szervezett támadások.

Az emberi erővel való monitorozás szinte lehetetlen a mai rendszereknél, ezért fontos szerepet játszik az automatizálása ennek a területnek. A bekövetkezett események, azon belül is a támadások észlelésére a Behatolás Érzékelő Rendszerek (Intrusion Detection System - IDS) vannak rendszeresítve.

Kutatási motivációm az irodalomkutatásban bemutatott módszerekre, eredményekre és technológiákra építve olyan módszer megtalálása, mely az IDS rendszerek tanításában segít. A vizsgálat az egyik leggyakoribb támadási módra, a Brute-Force támadás felismerésére irányul. Az IDS-ek konfigurálására és a legmegfelelőbb algoritmusok meghatározására léteznek tanító adathalmazok, melyek tartalmaznak különféle hálózati kommunikációkat, beleértve a támadási kommunikációkat is.

A célom az, hogy a kiválasztott adathalmaz segítségével olyan osztályozó algoritmust találjak, amely hatékonyan és pontosan tud meghatározni egy esetleges hálózati támadást. Ennek érdekében egyik kitűzött célom az adathalmaz előfeldolgozása, majd a jellemzők fontossági sorrendjének meghatározása különböző módszerek normalizált értékszámainak átlagai alapján, valamint súlyozott átlag módszer alapján végzett rangsorolás segítségével. Ennek köszönhetően több paraméterrel lehet az osztályozó algoritmusokat tanítani és tesztelni, így még pontosabb értékelési szempont alakul ki ahhoz, hogy a legmegfelelőbb algoritmus kerüljön meghatározásra. A kutatásom elméleti eredménye egy olyan módszer kidolgozása, mely segítségével az IDS rendszerek megfelelő paramétereinek és algoritmusainak beállításaival a gyakorlatba való implementálással hatékonyan detektálhatóak a támadások.

2. Alkalmazott módszerek

2.1. Adathalmaz feldolgozás

Az adathalmaz feldolgozása során, különösen nagydimenziós esetekben, kiemelkedően fontos a megfelelő előfeldolgozás és dimenziócsökkentés alkalmazása. A nagydimenziós adatok gyakran tartalmaznak sok felesleges vagy zajos információt, ami negatív hatással lehet az analízis pontosságára és hatékonyságára. Az előfeldolgozási lépések, mint például a hiányzó adatok kezelése, az outlier értékek azonosítása és kezelése, valamint a normalizálás vagy skálázás segítenek tisztább és megbízhatóbb adatok létrehozásában. Emellett a dimenziócsökkentés technikái lehetővé teszik, hogy a nagy mennyiségű változót kevesebb, de lényeges dimenzióban jelenítsük meg az adatokat. Ez javítja az értelmezhetőséget, csökkenti a zajt, és segít az analízis és modellezés hatékonyságának növelésében.

Az adatcsökkentési szakasz a jellemzők kiválasztására és a dimenziócsökkentésre összpontosít, ami számos előnnyel járhat. Az egyik legfontosabb előny az, hogy számos adatbányászati algoritmus jobban működik, ha a dimenziók száma - az adatok attribútumainak (oszlopainak) száma - kisebb. Ez részben azért van így, mert a dimenziócsökkentés kiküszöböli az irreleváns attribútumokat és csökkenti a zajt. Egy másik előnye, hogy érthetőbb modellhez vezethet, mivel kevesebb jellemző lesz benne. Ezenkívül a csökkentett adatmennyiség kevesebb tárhelyet és kevesebb időt igényel a feldolgozásához

A tanulmányban ismertetett kutatás során használt adathalmaz a CSE-CIC-IDS2018 on AWS [1], amelyet a Canadian Institute for Cybersecurity laboratórium hozott létre. Ez az adathalmaz azért lett kiválasztva, mert a kutatásom kezdeti szakaszában ez volt a legfrissebb adathalmaz, a kutatásban szereplő támadásokat tartalmazza, és megfelel a kutatáshoz szükséges összes kritériumnak (pl. teljes forgalom, címkézés stb.). Az adathalmaz feldolgozás az alábbi lépésekkel történt meg:

1. Adattisztítás magában foglalja az érvénytelen vagy hiányzó adatokat tartalmazó sorok (rekordok) törlését, az azonos értékű oszlopok törlését (pl. olyan oszlopok, ahol minden érték nulla), az osztályozás szempontjából irrelevánsnak ítélt jellemzők (oszlopok) törlését.
2. Adattranszformáció jelenti a kategorikus adatok numerikus adatokká történő átalakítását, normalizálást és az adathalmaz felosztást.
3. Normalizálás a numerikus oszlopok közös skálára való átalakításából áll.
4. Adatok felosztása tanító-és tesztminták létrehozására.

Eredmény:

Az adathalmaz előfeldolgozása során jelentős dimenziócsökkentést értem el, hiszen a 3 kiindulási adathalmaz oszlopszáma 80 volt, és ez redukálódott 69-re. Továbbá az adattranszformáció segítségével támadástípus alapján már könnyedén szét lehetett osztani az adatállományokat különböző fájlokba, így a további vizsgálatokat már támadási típusonként folytattam.

1. táblázat Tanítóhalmazok a dimenziócsökkentés után

Fájlnév	Sorok száma	Oszlopok száma
dataset-ftp-tr.csv	171 433	69
dataset-ssh-tr.csv	170 280	69
dataset-web-tr.csv	417 592	69
dataset-xss-tr.csv	417 211	69
dataset-sql-tr.csv	417 068	69

2. táblázat Teszthalmazok a dimenziócsökkentés után

Fájlnév	Sorok száma	Oszlopok száma
dataset-ftp-ts.csv	85 716	69
dataset-ssh-ts.csv	85 140	69
dataset-web-ts.csv	209 101	69
dataset-xss-ts.csv	208 720	69
dataset-sql-ts.csv	208 577	69

2.2. Jellemzők kiválasztása

A jellemzők kiválasztása a legrelevánsabb attribútumok megtalálására összpontosít, amelyek segítségével hatékony osztályozás vagy előrejelzés végezhető [2] [3] [4].

Hozzájárul a probléma dimenzionalitásának csökkentéséhez és így az erőforrásigény (tárolás, számítás) csökkenéséhez, valamint javíthatja a gépi tanuló algoritmusok teljesítményét [5], azaz gyorsabb képzés, csökkentett túlillesztés, és esetenként jobb előrejelző képesség érhető el.

A többtényezős kiválasztás (EFS - Ensemble Feature Selection) olyan technika, amely több jellemző választó algoritmus erősségeit használja ki, hogy javítsa a jelentős jellemzők azonosítását egy adathalmazban. Az együttes jellemző választás előnyei közé tartozik a fokozott osztályozási pontosság, a csökkent túlillesztés és a kiválasztott jellemzők nagyobb stabilitása. Ez a megközelítés különösen előnyös lehet a gépi tanulás által vezérelt alkalmazásokban, például a behatolás érzékelő rendszerekben, ahol a jellemzők sokfélesége hatással lehet a modell pontosságára és tanítási

időtartamára. A különböző jellemző választó algoritmusok előnyeinek egyesítésével az együttes jellemzőválasztás megkönnyítheti az adott feladat szempontjából legfontosabb jellemzők azonosítását, ami hatékonyabb és eredményesebb adatelemzést eredményez. Összességében az EFS hatékony és népszerű technika az adatok kiválasztására, amely javíthatja a modell pontosságát és csökkentheti a redundanciát [6].

A kutatási munka során használt jellemzőkiválasztási módszerek:

- Információnyereség (Information Gain) [7] [8] [9]
- Nyereségarány (Gain Ratio) [10] [11]
- Relief [12]
- Szimmetrikus bizonytalanság (Symmetric Uncertainty) [13] [14]
- Khí-négyzet próba [15]
- Varinavaanalízis (ANOVA) [16]

A hat jellemzőválasztási módszert mind az öt adatkészletre alkalmaztam 30 egyetemi laboratóriumi számítógépen, valamint az ELKH felhőszolgáltatások [17] segítségével. Bár több feladatot párhuzamosan végeztem, a teljes folyamat hónapokat vett igénybe.

Minden egyes adatkészlet és minden egyes módszer esetében normalizáltam a jellemzőkiválasztási folyamat végén kapott jellemzőpontszám-értékeket. Ezután a végső jellemzőpontszámot minden egyes adatkészlet esetében külön-külön a normalizált pontszámok átlagaként számoltam ki. Ezt követően az értékekhez rangsorolási küszöbértéket határoztam meg 0,05-től kezdve, növelve 0,05 lépéssel 0,55-ig. Minden egyes küszöbértékhez azokat a jellemzőket választottam ki, amelyek pontszáma magasabb az adott küszöbértéknél, így csökkentett számú különböző jellemzőcsoportokat határoztam meg (lásd 1. táblázat).

3. táblázat Jellemzőszámok csökkentésének eredményei a rangsorolási küszöbértékekkel

Küszöbérték	FTP	SSH	WEB	XSS	SQL
0,05	56	59	65	65	66
0,10	43	53	60	57	64
0,15	32	48	60	57	60
0,20	23	29	58	51	57
0,25	21	22	56	46	48
0,30	13	17	50	36	37
0,35	8	7	44	31	31

Küszöbérték	FTP	SSH	WEB	XSS	SQL
0,40	3	2	34	27	26
0,45	2	2	23	10	12
0,50	2	1	9	6	4
0,55	2	1	1	1	2

A küszöbértékekhez meghatározott jellemzők segítségével gépi tanulás alapú osztályozó algoritmusok vizsgálatát végeztem el annak érdekében, hogy alacsony jellemezőszám mellett elfogadható vagy jó osztályozási eredményt érjek el. Minden adathalmaz esetében 5 osztályozó algoritmust vizsgáltam különböző osztályozási teljesítménymérőkkel az adott küszöbértékek alkalmazása esetén kiválaszott jellemzőkkel. Minden osztályozónál a tanító és teszthalmazok vizsgálatával létrejött Accuracy, Precision, Recall, F1 teljesítményértékek (0-1 közé eső szám, ahol az 1 a legjobb teljesítményt mutatja) számtani átlagát figyelembe véve kiválasztottam a legnagyobb értéket. Így meghatároztam, minden adathalmaz esetén azt a küszöbértéket, ahol a legkisebb jellemzőszámmal jó osztályozási eredményt érek el.

Eredmény:

Az eredeti adatokban található 69 jellemzőből a meghatározott küszöbértékeknek megfelelően mindegyik támadástípusra meghatározásra került a csökkentett jellemzők darabszáma, melynek eredménye a 2. táblázatban látható.

4. Táblázat Jellemzőszámok csökkentésének eredményei a rangsorolési küszöbértékekkel

Adathalmaz	Küszöbérték	Jellemzők darabszáma
FTP	0,35	8
SSH	0,35	7
WEB	0,35	44
SQL	0,40	26
XSS	0,45	10

2.3. Gépi tanulás alapú osztályozási algoritmusok

Az osztályozási módszereket arra használják, hogy megjósolják egy objektumpéldány osztályát egy jellemzővektor alapján. A gépi tanuláson alapuló osztályozási algoritmusok olyan modelleket építenek fel, amelyek képesek címkézett adathalmazokból tanulni, és ezeket felhasználni az új, nem látott adatpontok osztályának előrejelzésére. Ebben a vizsgálatban öt különböző osztályozó algoritmust használtam:

- Logisztikus regresszió (Logistic Regression) [18]
- Naive Bayes [19] [20]
- Döntési Fa (Decision Tree) [21] [22]
- Véletlen erdő (Random Forest) [23]
- Tartóvektor-gép (SVM) [24]

Az öt osztályozó képzése és tesztelése az Orange 3.34 program segítségével történt, amely egy nyílt forráskódú adatvizualizációs, gépi tanulási és adatbányászati eszközkészlet A bináris osztályozás célja, hogy a tanító adathalmaz alapján tanuljon egy osztályozó algoritmust, amely képes új, ismeretlen példákra is osztályozni. A tanító adathalmazban minden példa rendelkezik egy címkével (osztályozási címkével), ami meghatározza a helyes osztályt. Az osztályozó algoritmusok teljesítményének értékelésére a 5. táblázatban szereplő előrejelzési szempontok láthatóak.

5. táblázat osztályozók értékelési szempontjai

Címke tulajdonság értéke	Besorolási érték	Előrejelzés
0	0	TN
0	1	FP
1	0	FN
1	1	TP

Az összes osztályozót a gyakorló és tesztminták alapján értékeltem négy, azaz az osztályozási pontosság (Accuracy), a pontosság (Precision), a fedés (Recall) és az F-mérték (F1) mérőszámmal, melyek az alábbi képlettekkel számíthatóak ki:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}, \tag{1}$$

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP}, \tag{2}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN}, \tag{3}$$

$$F1 = \frac{2(\mathrm{Pr}ecision \cdot \mathrm{Re}call)}{\mathrm{Pr}ecision + \mathrm{Re}call} \quad . \tag{4}$$

Az osztályozók teljesítményei alapján mindegyik adathalmazra meghatároztam egy olyan küszöbértéket, ahol a legkisebb jellemzőszámokkal a legjobb osztályozási teljesítményt értem el. Mindegyik adathalmazt figyelembe véve a legkisebb küszöbérték a 0,35. Ez alapján az öt adathalmazra megállapított jellemzőcsoportokkal vizsgáltam a 0,35-0,55 közötti küszöbértékeknél az osztályozók teljesítményét. Ahhoz, hogy megtaláljam a legjobban teljesítő osztályozót a legkevesebb jellemzőszámok mellett, minden osztályozónál a tanító es teszt halmazok vizsgálatával létrejött teljesítmény értékek (Accuracy, Precision, Recall, F1) számtani átlaguk közül a legjobb értéket vettem.

Eredmény:

Meghatároztam, minden adathalmaz esetén azt a küszöbértéket, ahol a legkisebb jellemzőszámmal jó osztályozási eredményt érek el. Minden egyes támadástípus esetében külön listát határoztam meg (lásd 6. táblázat) a küszöbértékekhez tartozó releváns jellemzőkkel. Minden egyes jellemzőt a sorszámával ábrázoltam. A táblázat minden sora azokat a jellemzőket tartalmazza, amelyek pontszáma nagyobb vagy egyenlő volt második cellában megadott küszöbértéknél. Ezáltal a jellemzők fontossági sorrendjével a támadások beazonosítását lehet hatékonyabbá tenni.

6. táblázat A legkevesebb jellemzőszámokkal elérhető legjobb osztályozók támadástípusonként

Adathalmaz	Küszöbérték	A legjobb osztályozó	Jellemzők darabszáma	Jelemzők sorszáma
FTP	0,35	Véletlen erdő	8	*02,17,19,35,00,44,56,59*
SSH	0,35	Véletlen erdő	7	*00,02,17,19,57,56,59*
WEB	0,35	Döntési Fa	44	*16,20,10,49,66,67,35,38,56,64,34,27,07,09,11, 14,15,50,25,60,62,02,17,19,37,63,06,33,55,18, 58,04, 05,53,54,03,21,22,23,24,52,32,65,57*
SQL	0,40	Véletlen erdő	26	*05,26,53,56,25,02,17,19,35,16,18,27,28,34, 06,23,30,55,29,21,22,24,57,37,11,14*
XSS	0,45	Döntési Fa	10	*37,56,33,32,03,11,52,04,54,58*

2.4. Súlyozott átlaggal végzett többtényezős módszer

Az előző vizsgálat során azok az esetek, ahol a betanított osztályozók gyenge teljesítményt mutattak, arra ösztönöztek, hogy tovább vizsgáljam a súlyozott átlagolású megközelítést. A súlyozott együttes rangsorolás a minták értékelésének széles körben használt megközelítése, amely lehetővé teszi az egyes komponensek differenciált értékelését azok jelentősége, fontossága, erőssége vagy bármely más, súlyként említett kritérium alapján. Több jellemző rangsorolási módszer hozzájárulásának figyelembevételével a jellemzők pontszámainak súlyozott átlagát az 5. egyenlet segítségével számoltam ki. Ez az egyenlet egy átfogó értékelési pontszámot ad, amely az együttes értékelését tükrözi.

$$R_{WA} = \frac{R_{IG} \cdot w_{IG} + R_{GR} \cdot w_{GR} + R_{SU} \cdot w_{SU} + R_{\chi^2} \cdot w_{\chi^2} + R_{\mathrm{Re}} \cdot w_{\mathrm{Re}} + R_{AN} \cdot w_{AN}}{w_{IG} + w_{GR} + w_{SU} + w_{\chi^2} + w_{\mathrm{Re}} + w_{AN}}, \qquad (5)$$

ahol az R_{WA} a jellemzőnek az együttes módszerrel számított pontszáma, R_{IG}, R_{GR}, R_{SU}, $R_{\chi 2}$, R_{Re}, R_{AN} az együttesbe bevont egyedi jellemző rangsorolási módszerek által kapott normalizált jellemző pontszámok, míg a w_{IG}, w_{GR}, w_{SU}, $w_{\chi 2}$, w_{Re}, w_{AN} az ezekhez a módszerekhez tartozó súlyokat jelentik.

A súlyok optimális kombinációjának meghatározása kihívást jelentő feladat, mivel a pontszámok kiszámításából eredő különböző jellemzőgyűjtemények kiértékeléséhez jelentős időre van szükség. Ezért szükségessé válik a súlyok optimalizálása minimális számú próbálkozással.

Ez a felismerés vezetett a Taguchi-módszer néven ismert kísérlettervezési (DoE) technika felhasználásához. Ezt a Genichi Taguchi által az 1950-es években kifejlesztett megközelítést eredetileg a gyártóiparban alkalmazott minőségirányítás és tervezés [25] célozta. Az optimális paraméterbeállítás meghatározásához a Taguchi-módszer a "paramétertervezés" fogalmát alkalmazza. Ennek során a folyamatváltozókat előre meghatározott értéktartományokhoz rendelik, teszteket végeznek, és optimalizálják azokat. A kutatás során hat független változót kell kipróbálni, mindegyiket két szinten. Ezért a $L_8 2^7$ ortogonális tervezési tervet alkalmaztam. Minden egyes faktor esetében két szintet használtam. A súlykeresési tér jobb, minimális kísérletekkel történő feltárásának megkönnyítése érdekében a súlyváltozók (a DoE-ban faktoroknak nevezett) két szintjéhez a kiválasztott DoE-designban szereplő súlyváltozókhoz 0,0233 és 0,2336 súlyértéket rendeltem (lásd 7. ábra). E választás hátterében az állt, hogy egymástól jelentősen távol eső értékeket használtam.

7. táblázat Meghatározott súlyértékek

	WIG	WGR	WSU	WKhi	WRe	WAN
1	0,023256	0,023256	0,023256	0,023256	0,023256	0,023256
2	0,023256	0,023256	0,023256	0,232558	0,232558	0,232558
3	0,023256	0,232558	0,232558	0,023256	0,023256	0,232558
4	0,023256	0,232558	0,232558	0,232558	0,232558	0,023256
5	0,232558	0,023256	0,232558	0,023256	0,232558	0,023256
6	0,232558	0,023256	0,232558	0,232558	0,023256	0,232558
7	0,232558	0,232558	0,023256	0,023256	0,232558	0,232558
8	0,232558	0,232558	0,023256	0,232558	0,023256	0,023256

Elsősorban azokra az esetekre irányítottam a figyelmet, ahol a korábbi, számtani átlagot használó vizsgálat nem hozott kielégítő eredményt. Itt két célom volt:

1. vagy kevesebb jellemzőt tartalmazó jellemzőkészletek azonosítása az eredeti osztályozási teljesítmény megtartása mellett, vagy
2. olyan jellemzőkészletek keresése, amelyek osztályozási pontosság (Accuracy), a pontosság (Precision), a fedés (Recall) és az F-mérték (F1) mint teljesítménymérők segítségével javíthatják az osztályozási teljesítményt.

Eredmény:

Minden egyes adathalmazhoz a meghatározott csökkenett számú jellemzők a 8. táblázatban láthatóak.

8. táblázat Súlyozott átlaggal kapott jellemzők csoportja

Adathalmaz	Súlyozott átlaggal kapott jellemzők csoportja
FTP	*19, 02, 17, 56, 59*
SSH	*33, 32, 00, 56, 57, 59*
WEB	*32, 56, 07, 50, 09, 11, 65, 14, 37, 53, 05, 58, 57*
XSS	*57, 56*
SQL	*56, 43, 47, 57, 37, 11, 14*

2.5. Osztályozás CatBoost algoritmus segítségével

A korábban bemutatott és használt Naive Bayes, Decision Tree, Random Forest, Logistic Regression és SVM olyan osztályozó és modellező módszerek, amelyek hosszú ideje részei a gépi tanulás és adatbányászat eszköztárának. Kutatásom folytatásaként meg kívántam vizsgálni, hogy az utóbbi időben sok területen sikerrel alkalmazott Gradient Boost megközelítés egyik megvalósítása, a CatBoost algoritmus képes-e a hagyományos osztályozókkal közel azonos vagy azoknál jobb osztályozási eredményeket nyújtani.

Minden támadás típus esetében a CatBoost osztályozóval ugyanazokat a tanító- és tesztadathalmazokat használtam, valamint olyan jellemzőket, amelyeket korábban az egyes jellemzők pontszámainak súlyozott átlaggal történő összesítésével azonosítottam.

Eredmény:

Mindegyik adathalmaz esetében a tanító és teszt halmazok vizsgálatával létrejött Accuracy, Precision, Recall, F1 teljesítmény értékek számtani átlagát figyelembe véve határoztam meg egy átlagos osztályozási teljesítmény számot, amely 0-1 közé eső szám, ahol az 1 a legjobb teljesítményt mutatja. Ez alapján létrejött egy összehasonlítás (lásd 9. táblázat) a Naive Bayes, Logisztikus regresszió, Tartóvektor-gép, Döntési fa, Véletlen erdő osztályozó algoritmusok és a CatBoost algoritmus között.

9. Táblázat A Catboost összehasonlítása

Adathalmaz	Jellemzők száma	Osztályozó	Osztályozó teljesítmények átlaga
FTP	5	Naive Bayes	0,9942
		Logisztikus regresszió	0,9984
		Döntési fa	1,0000
		Tartóvektor-gép	0,9998
		Véletlen erdő	1,0000
		CatBoost	**1,0000**
SSH	6	Naive Bayes	0,9999
		Logisztikus regresszió	0,9940
		Döntési fa	1,0000
		Tartóvektor-gép	0,9999
		Véletlen erdő	1,0000
		CatBoost	**1,0000**
SQL	7	Naive Bayes	0,2499
		Logisztikus regresszió	0,5252
		Döntési fa	0,9826
		Tartóvektor-gép	0,7323
		Véletlen erdő	0,9913
		CatBoost	**0,9694**
XSS	2	Naive Bayes	0,2498
		Logisztikus regresszió	0,2498
		Döntési fa	0,9657

Adathalmaz	Jellemzők száma	Osztályozó	Osztályozó teljesítmények átlaga
		Tartóvektor-gép	0,3183
		Véletlen erdő	0,9697
		CatBoost	**0,9197**
WEB	13	Naive Bayes	0,5017
		Logisztikus regresszió	0,2494
		Döntési fa	0,9363
		Tartóvektor-gép	0,2071
		Véletlen erdő	0,8994
		CatBoost	**0,8994**

3. Új tudományos eredmények összefoglalása

Kutatásom kezdeti szakaszában a különböző IDS-rendszerek alapvető jellemzőit és funkcióit vizsgáltam [S15] [S11]. Ezt követően figyelmemet az anomália alapú IDS-rendszerekre irányítottam, különös tekintettel az osztályozó moduljuk tanítási folyamatára. Ez a folyamat jellemzően nagy adatminták felhasználását jelenti, amelyek mind jóindulatú, mind rosszindulatú forgalmi adatokat tartalmaznak. Kutatásom során több adathalmazt is megvizsgáltam, végül találtam egy megfelelőt (CSE-CIC-IDS2018 az AWS-en), amely nemcsak megfelelt a kritériumoknak, hanem friss is volt, így ideális a tanításhoz.

Az adatkészlet kiválasztása és számos előfeldolgozási lépés végrehajtása után a vizsgálatom a jellemzőválasztásimódszerek köré összpontosult. Itt elsődleges eredményem a hat különböző módszerrel kapott normalizált jellemzőpontszámok számtani átlaga alapján a jellemzők rangsorolása volt (ld. 1. tézis). A továbbiakban a kutatásom a jellemzőkiválasztást célozta meg, azzal a céllal, hogy küszöbértékeket határozzak meg a számtani átlagon alapuló többtényzős (ensemble) módszerrel kapott pontszámok számára. A cél az volt, hogy olyan releváns jellemzőkészletet határozzak meg, amely elegendő információt szolgáltat az osztályozó modul számára (ld. 2. tézis).

Ezt követően azzal a feltételezéssel folytattam a kutatást, hogy az egyes jellemzők pontszámainak súlyozott átlagon alapuló többtényezős módszer alkalmazása potenciálisan javíthatja az osztályozási teljesítményt, vagy legalábbis csökkentheti a szükséges jellemzők számát. Ezen hipotézis igazolása érdekében egy Taguchi típusú kísérlettervet alkalmaztam a szükséges kísérletek számának alacsony szinten tartása érdekében. A kísérleti eredmények megerősítették a hipotézist (ld. a 3. tézis).

Kutatásomat azzal a hipotézissel folytattam, hogy a korábban használt öt közismert osztályozó algoritmus által elért osztályozási teljesítményt a viszonylag új CatBoost algoritmus alkalmazásával felül lehet múlni, vagy legalábbis meg lehet közelíteni. A kísérleti eredmények ezt a hipotézist is megerősítették (ld. 4. tézis).

1. TÉZIS

Egy IDS rendszer tanítására használt adatbázis segítségével olyan módszert dolgoztam ki, ami alkalmas a beazonosításhoz szükséges jellemzők fontossági sorrendjének meghatározására az Információnyereség, Nyereségarány, Szimmetrikus bizonytalanság, Relief, Khi-négyzet próba és a Varianciaanalízis módszerek normalizált értékszámainak átlagai alapján végzett rangsorolás segítségével.

A tézisemhez tartozó publikációm a következő: [S3]

2.TÉZIS

Átlagolással kapott jellemző-értékszámokhoz kapcsolódóan meghatároztam azokat a küszöbértékeket, amelyek segítségével beazonosítható a jellemzők azon minimális darabszámú csoportja, amely mellett jó teljesítményű osztályozók taníthatók be.

A tézisemhez tartozó publikációm a következő: [S3]

3. TÉZIS

Megmutattam, hogy a különböző módszerekkel előállított jellemzőértékszámok súlyozásával képzett általános mérőszám alkalmazásával ugyanolyan vagy jobb osztályozási eredmények érhetőek el csökkentett számú figyelembe vett jellemző mellett.

A tézisemhez tartozó publikációm a következő: [S4]

4. TÉZIS

Megmutattam, hogy a CatBoost algoritmus alkalmazásával legalább olyan jó osztályozási eredmények érhetőek el bizonyos hálózati támadások esetében, mint a Logisztikus regresszió, Naive Bayes, Tartóvektor-gép, Döntési fa, Véletlen erdő osztályozótípusok használata mellett.

A tézisemhez tartozó publikációm a következő: [S5]

4. Az elért eredmények hasznosíthatósága

Egy hálózati kommunikációt (beleértve a normál és támadási eseteket) vizsgálva a súlyozott együttes jellemzőkiválasztási módszer segítségével a meghatározott jellemzőcsoportokban szereplő jellemzők segítségével bizonyos támadási típusokra már tényleges adatok, információk konfigurálhatóak egy IDS rendszer szenzorai számára. Az eredetileg 80 jellemzővel rendelkező hálózati kommunikációs adatokból csak a 10. táblázatban szereplő jellemzők csoportjait kell figyelembe venni ahhoz, hogy jó osztályozási eredményt érjen el egy IDS a megfelelő támadások detektálására.

10. táblázat A meghatározott releváns jellemzők tulajdonságai

Sorszám	Megnevezés	A jellemzők működésbeni jelentése
00	*Dst Port*	A célállomás portja, ahová az adatcsomagokat küldik.
02	*Flow Duration*	Az adatfolyam időtartama az első és az utolsó adatcsomag között.
05	*TotLen Fwd Pkts*	Az összes előre irányuló (forrás felé) adatcsomag mérete összesen.
07	*Fwd Pkt Len Max*	Az előre irányuló adatcsomagok közül a leghosszabb csomag mérete.
09	*Fwd Pkt Len Mean*	Az előre irányuló adatcsomagok átlagos mérete.
11	*Bwd Pkt Len Max*	A visszafelé irányuló adatcsomagok közül a leghosszabb csomag mérete.
14	*Bwd Pkt Len Std*	A visszafelé irányuló adatcsomagok méretének szórása.
17	*Flow IAT Mean*	Az adatfolyam közötti időközök átlagos hossza.
19	*Flow IAT Max*	Az adatfolyam közötti időközök maximális hossza.
32	*Fwd Header Len*	Az előre irányuló adatcsomagok fejlécének mérete.
33	*Bwd Header Len*	A visszafelé irányuló adatcsomagok fejlécének mérete.
37	*Pkt Len Max*	A legnagyobb adatcsomag mérete az összes csomag közül.
43	*RST Flag Cnt*	A RST zászlóval ellátott adatcsomagok száma.
47	*ECE Flag Cnt*	Az ECE zászlóval ellátott adatcsomagok száma.
50	*Fwd Seg Size Avg*	Az előre irányuló adatcsomagok átlagos szegmensmérete.
53	*Subflow Fwd Byts*	Az előre irányuló adatok összmérete az alközlési áramlatokban.
56	*Init Fwd Win Byts*	Az előre irányuló kezdeti ablakméret a TCP kapcsolatban.
57	*Init Bwd Win Byts*	A visszafelé irányuló kezdeti ablakméret a TCP kapcsolatban.
58	*Fwd Act Data Pkts*	Az előre irányuló effektív adatcsomagok száma.
59	*Fwd Seg Size Min*	Az előre irányuló adatcsomagok minimális szegmensmérete.
65	*Idle Std*	Az időtartamok közötti inaktivitás időtartamok szórása az adatfolyamban.

5. További kutatási irányok

Az IDS-rendszerekkel kapcsolatos kutatásomat az adatfúzión alapuló megoldások kialakítása irányában kívánom folytatni. Itt az adatfúzió olyan folyamatot jelent, amely során több forrásból vagy szenzorból származó adatokat kombinálnak és elemeznek annak érdekében, hogy növeljék a hálózati behatolások észlelésének pontosságát. Ide tartozik a tűzfalak, behatolásérzékelő-rendszerek (IDS), naplófájlok, hálózati forgalomadatok és más releváns források információinak integrálása. Az adatfúzió célja az, hogy kihasználja a különböző adatforrások előnyeit, és javítsa a behatolásérzékelő-rendszerek észlelési képességeit. Az egyes szenzorok vagy észlelési módszerek által esetleg fel nem ismert mintázatok és anomáliák valószínűleg könnyebben azonosíthatók több adatfolyam kombinálásával, ami megbízhatóbb behatolásérzékelést eredményezhet.

A kutatás előkészítéseként munkahelyemen, a kecskeméti Neumann János Egyetemen előkészítés alatt áll egy Informatikai Adatlabor kialakítása, melynek része lesz egy Szakértői és Kiberbiztonsági labor, ahol kialakításra kerül egy komplett informatikai laborkörnyezet, a kanadai laborhoz hasonlóan.

Egy teljes informatikai infrastruktúra lenne modellezve, normál és támadási kommunikációkkal. Egy ilyen rendszer kialakításával és az adatfúzió megvalósításával egy új, saját mintaadathalmazt tudnék létrehozni, amit az IDS-rendszereket kutatók szabadon használhatnának. Ezen adathalmaz felhasználásával a disszertációban bemutatott osztályozókon kívül más megoldások vizsgálatát is tervezem.

6. Summary

During the initial phase of my investigation, I delved into the fundamental characteristics and functionalities of various IDS systems [S15] [S11]. Subsequently, I directed my attention towards Anomaly-based IDS systems, specifically focusing on the training process of their classification module. Typically, this process involves utilizing big data samples that describe both benign and malicious traffic scenarios. Throughout my research, I explored multiple datasets, eventually discovering a suitable one (CSE-CIC-IDS2018 on AWS) that not only met the criteria but was also recent, making it ideal for training purposes.

Once the dataset was chosen and several preprocessing steps were executed, my investigation centered around feature selection methods. The primary outcome here was the ranking of features based on the arithmetic mean of normalized feature scores obtained from six distinct methods (refer to Thesis statement 1). Moving forward, my research targeted feature selection, aiming to establish threshold values for the scores obtained through the arithmetic mean based ensemble method. The objective was to identify a relevant set of features that would furnish sufficient information for the classifier module (refer to Thesis statement 2).

Subsequently, I pursued my investigation with the assumption that utilizing an ensemble method based on weighted aggregation of individual feature scores could potentially enhance classification performance or, at the very least, reduce the number of required features. To validate this hypothesis, I adopted a Taguchi-type DoE design, conducting a low number of trials. The experimental results confirmed the hypothesis (refer to Thesis statement 3).

Continuing my research, I operated under the hypothesis that the classification performance achieved by the five well-known classification algorithms used previously could be surpassed, or at least matched, by employing the relatively new CatBoost algorithm. The experimental results also confirmed this hypothesis (refer to Thesis statement 4).

Irodalomhivatkozás

[1] R. B. Basnet, R. Shash, C. Johnson, L. Walgren, and T. Doleck, ‘Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks’, *J. Internet Serv. Inf. Secur*, no. 9, pp. 1–17, 2019.

[2] Z. J. Viharos, K. B. Kis, Á. Fodor, and M. I. Büki, ‘Adaptive, hybrid feature selection (AHFS)’, *Pattern Recognition*, vol. 116, p. 107932, 2021, doi: 10.1016/j.patcog.2021.107932.

[3] K. Muhi and Z. C. Johanyák, ‘Dimensionality reduction methods used in Machine Learning’, *M\Huszaki Tudományos Közlemények*, vol. 13, no. 1, pp. 148–151, 2020, doi: 10.33894/mtk-2020.13.27.

[4] T. Dobján and E. D. Antal, ‘Modern feature extraction methods and learning algorithms in the field of industrial acoustic signal processing’, in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, IEEE, 2017, pp. 000065–000070. doi: 10.1109/sisy.2017.8080589.

[5] N. S. Chauhan, ‘Decision Tree Algorithm—Explained’. 2020. [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html/https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html/

[6] V. Bolón-Canedo and A. Alonso-Betanzos, ‘Ensembles for feature selection: A review and future trends’, *Information Fusion*, vol. 52, pp. 1–12, Dec. 2019, doi: 10.1016/j.inffus.2018.11.008.

[7] G. Ayyappan, D. C. Nalini, and D. A. Kumaravel, ‘Efficient mining for social networks using Information Gain Ratio based on Academic dataset’, *International Journal of Civil Engineering and Technology*, vol. 8, no. 1, 2017.

[8] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, *Feature selection for high-dimensional data*. Springer, 2015. doi: 10.1007/978-3-319-21858-8.

[9] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, ‘COMPARATIVE STUDY OF ATTRIBUTE SELECTION USING GAIN RATIO AND CORRELATION BASED FEATURE SELECTION’.

[10] R. P. Priyadarsini, M. Valarmathi, and S. Sivakumari, ‘Gain ratio based feature selection method for privacy preservation’, *ICTACT Journal on soft computing*, vol. 1, no. 4, pp. 201–205, 2011, doi: 10.21917/ijsc.2011.0031.

[11] S. J. Pasha and E. S. Mohamed, ‘Ensemble Gain Ratio Feature Selection (EGFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction’, in *2020 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India: IEEE, Feb. 2020, pp. 590–596. doi: 10.1109/ICICT48043.2020.9112406.

[12] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, ‘Relief-based feature selection: Introduction and review’, *Journal of biomedical informatics*, vol. 85, pp. 189–203, 2018, doi: 10.1016/j.jbi.2018.07.014.

[13] B. Singh, N. Kushwaha, O. P. Vyas, and others, ‘A feature subset selection technique for high dimensional data using symmetric uncertainty’, *Journal of Data Analysis and Information Processing*, vol. 2, no. 04, p. 95, 2014, doi: 10.4236/jdaip.2014.24012.

[14] S. Bakhshandeh, R. Azmi, and M. Teshnehlab, ‘Symmetric uncertainty class-feature association map for feature selection in microarray dataset’, *Int. J. Mach. Learn. & Cyber.*, vol. 11, no. 1, pp. 15–32, Jan. 2020, doi: 10.1007/s13042-019-00932-7.

[15] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, *Feature selection for high-dimensional data*. Springer, 2015.

[16] M. Kumar, N. K. Rath, A. Swain, and S. K. Rath, ‘Feature selection and classification of microarray data using MapReduce based ANOVA and K-nearest neighbor’, *Procedia Computer Science*, vol. 54, pp. 301–310, 2015, doi: 10.1016/j.procs.2015.06.035.

[17] M. Héder *et al.*, ‘The Past, Present and Future of the ELKH Cloud’, *InfTars*, vol. 22, no. 2, p. 128, Aug. 2022, doi: 10.22503/inftars.XXII.2022.2.8.
[18] M. Maalouf, ‘Logistic regression in data analysis: an overview’, *IJDATS*, vol. 3, no. 3, p. 281, 2011, doi: 10.1504/IJDATS.2011.041335.
[19] H. Zhang and J. Su, ‘Naive Bayesian Classifiers for Ranking’, in *Machine Learning: ECML 2004*, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., in Lecture Notes in Computer Science, vol. 3201. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 501–512. doi: 10.1007/978-3-540-30115-8_46.
[20] F.-J. Yang, ‘An Implementation of Naive Bayes Classifier’, in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA: IEEE, Dec. 2018, pp. 301–306. doi: 10.1109/CSCI46756.2018.00065.
[21] B. Charbuty and A. Abdulazeez, ‘Classification based on decision tree algorithm for machine learning’, *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[22] R. Davidson, ‘Reliable inference for the Gini index’, *Journal of Econometrics*, vol. 150, no. 1, pp. 30–40, May 2009, doi: 10.1016/j.jeconom.2008.11.004.
[23] L. Breiman, ‘Random Forests’, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[24] I. Steinwart and A. Christmann, *Support vector machines*, 1st ed. in Information science and statistics. New York: Springer, 2008.
[25] A. Freddi and M. Salmon, *Design Principles and Methodologies: From Conceptualization to First Prototyping with Examples and Case Studies*, 1st ed. 2019. in Springer Tracts in Mechanical Engineering. Cham: Springer International Publishing: Imprint: Springer, 2019. doi: 10.1007/978-3-319-95342-7.

Saját publikációk

Folyóiratcikkek

S1. **László, Göcs**; Attila, Pásztor; Zsolt, Csaba Johanyák: Computer network solutions in modern industrial environment ANNALS OF FACULTY OF ENGINEERING HUNEDOARA - INTERNATIONAL JOURNAL OF ENGINEERING 10: 1 pp. 75-80., 6 p. (2022)
S2. **László, Göcs**; Zsolt, Csaba Johanyák; Péter, András Agg: Protection of Computer Laboratories in Educational Institutions ACTA TECHNICA CORVINIENSIS – BULLETIN OF ENGINEERING 9: 2 pp. 93-98., 6 p. (2016)
S3. **László, Göcs**; Zsolt, Csaba Johanyák: Identifying Relevant Features of CSE-CIC-IDS2018 Dataset for the Development of an Intrusion Detection System, 2023 http://gocslaszlo.hu/phd/tezis_1_2.pdf (publikálás alatt)
S4. **L. Göcs** and Z. C. Johanyák, ‘Feature Selection with Weighted Ensemble Ranking for Improved Classification Performance on the CSE-CIC-IDS2018 Dataset’, Computers, vol. 12, no. 8, p. 147, Jul. 2023, doi: 10.3390/computers12080147.
S5. **László, Göcs**; Zsolt, Csaba Johanyák: Catboost algorithm based IDS cassification module for brute force attacks", ANNALS OF FACULTY OF ENGINEERING HUNEDOARA: INTERNATIONAL JOURNAL OF ENGINEERING 21: 3 pp. 13-18.,6 p. (2023)

Konferenciaközlemények

S6. **Göcs, László**; Johanyák, Zsolt Csaba: Adatbázis feldolgozása IDS rendszerek tanításához Kutatás és innováció 2021: GAMF Közlemények tanulmánykötete Kecskemét, Magyarország: Neumann János Egyetem GAMF Műszaki és Informatikai Kar (2021) pp. 401-406., 6 p.

S7. **Göcs, László**; Pásztor, Attila; Johanyák, Zsolt Csaba: Modern ipari környezet informatikai hálózati lehetőségei a rendelkezésre állás biztosítása érdekében GRADUS 8: 3 pp. 147-156., 10 p. (2021)

S8. **Göcs, László**; Johanyák, Csaba; Kovács, Szilveszter: IDS rendszerek fuzzy logikával In: Keresztes, Gábor; Kohus, Zsolt; Szabó P., Katalin; Tokody, Dániel (szerk.) Tavaszi Szél 2017 Konferencia. Nemzetközi Multidiszciplináris Konferencia: Absztraktkötet Budapest, Magyarország: Doktoranduszok Országos Szövetsége (DOSZ) (2017) 477 p. p. 311

S9. Agg, Péter András; Johanyák, Zsolt Csaba; **Göcs, László**: Szoftver által definiált hálózatok áttekintése In: Bitay, Enikő (szerk.) A XXI. Fiatal Műszakiak Tudományos Ülésszaka előadásai Kolozsvár, Románia: Erdélyi Múzeum Egyesület (EME) (2016) 452 p. pp. 57-60., 4 p.

S10. **Göcs, László**; Johanyák, Zsolt Csaba; Kovács, Szilveszter: Csapda a hálózaton GRADUS 3: 2 pp. 55-60., 6 p. (2016)

S11. **László, Göcs**; Zsolt, Csaba Johanyák; Szilveszter, Kovács: Review of Anomaly-Based IDS algorithms In: AlumniPress - AlumniPress (szerk.) TEAM 2016: Proceedings of the 8th International Scientific and Expert Conference Trnava, Szlovákia: Alumni Press (2016) 360 p. pp. 58-63., 6 p.

S12. **László, Göcs**; Zsolt, Csaba Johanyák: Virtualization in Network Administration Education In: Kucsinka, Katalin; Kiss, Alexandra; Veres, Erika (szerk.) Matematikát oktatók és kutatók nemzetközi tudományos konferenciája Beregszász, Ukrajna: II. Rákóczi Ferenc Kárpátaljai Magyar Főiskola (2016) 78 p. p. 52

S13. Agg, P; **Göcs, L**; Johanyák, Zs Cs; Borza, Z: Csomagszűrés CISCO routereken ACL-ek segítségével GRADUS 2: 2 pp. 104-111., 8 p. (2015)

S14. **Göcs, László**; Johanyák, Zsolt Csaba: Vállalati informatikai biztonság szerepe napjainkban In: Bitay, Enikő (szerk.) A XX. Fiatal Műszakiak Tudományos Ülésszaka előadásai: Proceedings of the XX-th International Sciencific Conference of Young Engineers Kolozsvár, Románia: Erdélyi Múzeum Egyesület (EME) (2015) 356 p. pp. 155-158., 4 p.

S15. **László, Göcs**; Zsolt, Csaba Johanyák: Survey on intrusion detection systems In: Prof, Aleksandar Sedmak; Zoran, Radakovic; Simon, Sedmak; Snezana, Kirin (szerk.) Proceedings of TEAM 2015: 7th International Scientific and Expert Conference of the International TEAM Society Beograd, Szerbia: University of Belgrade, Faculty of Mechanical Engineering (2015) 650 p. pp. 167-170., 4 p.

S16. Zsolt, Csaba Johanyák; Piroska, Gyöngyi Ailer; **László, Göcs**: A simple fuzzy control design for series hybrid electric vehicle In: Andrea, Ádámné Major; Lóránt, Kovács; Zsolt, Csaba Johanyák; Róbert, Pap-Szigeti (szerk.) Proceedings of TEAM 2014: 6th International Scientific and Expert Conference of the International TEAM Society Kecskemét, Magyarország: Kecskeméti Főiskola Gépipari és Automatizálási Műszaki Főiskolai Kar (2014) 499 p. pp. 159-164., 6 p.

S17. **Göcs, László**: Informatikai biztonság In: Ferencz, Árpád; Borsné, Pető Judit; Lipócziné, Csabai Sarolta; Kovács, Lóránt (szerk.) AGTEDU 2011: a Magyar Tudomány Ünnepe alkalmából rendezett 12. tudományos konferencia Kecskemét, Magyarország: Kecskeméti Főiskola (2011) 406 p. pp. 143-148., 6 p.

Egyéb publikációk

S18. **Göcs, László**: Covid19 hatása az informatikai rendszerekre (2020) AGTEDU 2020, 2020. november 12., Előadás,
S19. **Göcs, László**: Adataink és az okos eszközök - kémek a lakásban? (2019) AGTEDU 2019, 2019. november 13., Előadás,
S20. **Göcs, László**; Johanyák, Zsolt Csaba: Címkézett adatbázis IDS rendszerekhez (2018) AGTEDU 2018, Kecskemét: 2018. november 15., Előadás,
S21. **Göcs, László**: A digitális világ biztonságos használata: Internet és informatikai biztonság (2018) Hírös Szabadegyetem, Kecskemét, 2018. április 11., Előadás,
S22. **László, Göcs**: Importance of passwords in IT security (2018) WCNCI 2018 (Workshop on Computer Networks and Computational Intelligence), 2018. október 16., Előadás,
S23. **Göcs, László**; Johanyák, Zsolt Csaba; Bors, Ádám: A blokklánc technológia (2017) AGTEDU 2017: Magyar Tudomány Ünnepe: 2017. november 16., Neumann János Egyetem Gazdálkodási Kar, Szolnok, Előadás, Megjelenés: Magyarország,

Oktatási anyagok

S24. **Göcs, László**: Szerveroldali megoldások Linux környezetben (Ubuntu 20.04 LTS) Kecskemét, Magyarország: Neumann János Egyetem (2021) ISBN: 9786155817915
S25. Johanyák, Zsolt Csaba; **Göcs, László**: Windows hálózati adminisztráció a gyakorlatban, 153 p. (2014)
S26. Johanyák, Zsolt Csaba; Kovács, Péter; **Göcs, László**: Linux hálózati adminisztráció a gyakorlatban, 113 p. (2013)