COMPUTATIONAL SCIENCE AND TECHNIQUES
Volume 9, 2025, 679-697

Publisher: Klaipėda University
https://e-journals.ku.lt/journal/csat
Online ISSN: 2029-9966
DOI: 10.15181/csat.v9.2504

MAŠININIO MOKYMO METODŲ LYGINAMOJI NAŠUMO ANALIZĖ KIBERNETINIŲ INCIDENTŲ APTIKIMUI DAIKTŲ INTERNETO TINKLUOSE

Kostas Ereksonas

Klaipėdos Universitetas
Jūros technologijų ir gamtos mokslų fakultetas
kostas.ereksonas@edu.ku.lt

Santrauka. Augant naudojamų daiktų interneto įrenginių bei tinklų skaičiui, kyla ir naujos saugumo rizikos, susijusios su šiais daiktų interneto tinklais. Šiame darbe yra pristatoma lyginamoji įvairių mašininio ir gilaus mokymosi modelių klasifikavimo našumo analizė – tikslumo bei atpažinimo metrikos – naudojant CIC-IoT-IDS2023 duomenų rinkinį, skirtą apmokyti mašininio mokymo modelius aptikti prieš daiktų interneto įrenginius nukreiptas atakas. Taip pat pateiktiamos gaires tolimesniems darbams kuriant kibernetinių įsilaužimų aptikimo sistemą daiktų interneto tinkluose. Siūlomos gairės apima sprendimo medžių modelių hiperparametrų tobulinimą, realaus laiko įsilaužimų aptikimo komponento kūrimą bei testavimą naudojant kraštinius (*angl.* edge) įrenginius.

Raktiniai žodžiai: CIC-IoT-IDS2023, mašininis mokymas, gilusis mokymas, daiktų internetas, įsilaužimų aptikimo sistema

Įžanga

Interneto technologijoms tampant vis svarbesne visuomenės gyvenimo dalimi bei šių technologijų pagalba žmonėms keičiant gyvenimo, studijų bei darbo įpročius, įvairios kibernetinio saugumo grėsmės, su kuriomis tenka susidurti, tampa vis rimtesnės. Vienas iš svarbiausių bei neišvengiamų techninių iššūkių yra į kompiuterinį tinklą nukreiptų kibernetinių atakų aptikimas. Ypač svarbu yra aptikti naujas, dar anksčiau neatrastas kibernetines atakas. Ankstyvojo kibernetinių incidentų aptikimo sistema (*angl.* Intrusion Detection System – IDS), reikšmingas mokslinių tyrimų pasiekimas informacijos apsaugos srityje, geba aptikti tebevykstančia arba jau įvykusią kibernetinę ataką, nukreiptą prieš saugomą kompiuterinę sistemą [1].

Vystantis interneto bei kompiuterinių tinklų technologijoms bei joms tampant vis svarbesnėms modernioje visuomenėje, vis svarbesnis tampa kompiuterinio tinklo saugumo aspektas. Internetas bei komputeriniai tinklai tampa vis sudėtingesni bei labiau kompleksiški, todėl efektyvus įsibrovimų į kompiuterinius tinklus aptikimas bei kompiuterinio tinklo saugos užtikrinimas tapo labai aktualia šių laikų problema [2].

Pirminė kibernetinių incidentų aptikimo sistemos paskirtis yra aptikti kenkėjišką kompiuteriniais tinklais perduodamą duomenų srautą bei neleisti jam patekti į saugomą kompiuterinę sistemą. Vis dėlto, kompiuteriniais tinklais perduodamų duomenų skaičius vis auga, ko pasekoje kenkėjiškais tikslais siunčiamus duomenis yra lengviau paslėpti nuo

saugos sistemų. Kompiuteriniu tinklu perduodamas kenkėjiškų duomenų srautas gali būti panaudojamas kibernetinių atakų, nukreiptų prieš interneto puslapius, serverius bei pastaruosiuose saugomus vartotojų duomenis, įgyvendinimui [2].

1. Literatūros apžvalga

C. Yin, Y. Zhu, et al. kūrė kibernetinių atakų aptikimo modelį, paremtą giliojo mokymosi technologija. Giliojo mokymo modelis šiame darbe buvo realizuotas naudojant rekurentinius neuroninius tinklus (*angl.* Recurent Neural Network – RNN), buvo optimizuoti kurto modelio parametrai bei įvertintas modelio tikslumas dvejetainio (*angl.* binary) bei daugiaklasio (*angl.* multiclass) klasifikavimo uždaviniuose. Gautos tikslumas buvo palygintas su J48, dirbtinių neuroninių tinklų (*angl.* Artificial Neural Network – ANN), atsitiktinio medžio (*angl.* Random Forest), atraminių vektorių mašinos (*angl.* Support Vector Machine – SVM) bei kitų tradicinių mašininio mokymo algoritmų tikslumu bei buvo prieita išvados, jog pristatytas rekurentinių neuroninių tinklų modelis yra tikslesnis bei greitesnis už tradicinius mašininio mokymo modelius binariniame bei daugiaklasiame klasifikavime [1].

L. Chen, X. Kuang, et al. pristatė kompiuteriniame tinkle veikiančią kibernetinių įsilaužimų aptikimo sistemą (*angl.* Network Intrusion Detection System – NIDS), kuri buvo realizuota naudojant konvoliucinius neuroninius tinklus (*angl.* Convolutional Neural Network – CNN). Tyrimo metu buvo fiksuoti kompiuterinio tinklo srauto duomenys, o iš užfiksuotų duomenų buvo nustatytos aktualios apmokymui tinklo srauto savybės. Konvoliucinių neuroninių tinklų modelis buvo apmokytas tiek naudojant aktualias savybes, tiek ir neapdorotą tinklo srautą. Vertinant sukurto modelio, apmokyto CIC-IDS-2017 duomenų rinkiniu, tikslumą buvo nustatyta, jog konvoliucinių neuroninių tinklų modelis veikia tiksliau už SVM bei DBN modelius. Modelis, apmokytas neapdorotais duomenimis veikė tiksliau [2].

Pagrindinis tinklo saugumui užtikrinti skirtų sistemų tikslas yra užkirsti kelią internetų siunčiamų duomenų vagystei bei klastojimui – užtikrinti duomenų konfidencialumą bei integralumą. Internetas bei jį sudarantys kompiuteriniai tinklai tampa vis sudėtingesni, ko pasekoje atsiranda vis įvairesnių įsilaužimų bei ataku metodų. Dėl šios priežasties kibernetinių įsilaužimų sistemos gali būti ganėtinai netikslios bei neaptikti realių kibernetinių įsilaužimų ar fiksuoti normalų tinklo srautą kaip kenkėjišką [3].

Vienas iš H. Chen, Y. Liu ir kitų autorių šiame darbe minimų metodų, skirtų kibernetinių įsilaužimų aptikimo sistemai (*angl.* Intrusion Detection System – IDS) kurti yra Rekurentinis Neuroninis Tinklas (angl. Recurrent Neural Network – RNN) [3].

Dvejetainė kompiuterinio tinklo srauto klasifikacija (angl. Binary classification) – stebimo tinklo srauto elgsena, funkcionavimas vertinamas kaip normalus arba ne, t.y. turintis anomalijų [3].

Daugiaklasinė kompiuterinio tinklo srauto klasifikacija (angl. Multiclass classification) – išskiriamos penkios stebimo tinklo srauto elgsenos klasifikavimo kategorijos, remiantis NSL-KDD duomenų rinkiniu, pagal kurias stebimo tinklo srauto parametrus galima įvertinti kaip nekeliančius pavojaus saugomai sistemai, t.y. stebimas kompiuterinio tinklo srautas yra normalus, arba pagal aptiktus tinklo parametrus identifikuoti tam tikrą stebimo kompiuterinio tinklo srauto dalį kaip atitinkančią vieną iš keturių toliau pateiktų kibernetinių atakų tipų aprašymų [3]:

1. Paslaugų trikdymo atakos (Denial of Service – DoS) – kompiuterinė sistema yra užkraunama dideliu kiekiu tinklo srauto, ko pasekoje visi sistema nebegali atlikti savo pagrindinių funkcijų bei teikti paslaugų realiems vartotojams [3].
2. User to Root (U2R) – ataka, kurios metu, turint prieigą prie paprasto sistemos naudotojo paskyros, yra siekiama gauti prieigą prie sistemos administratoriaus

paskyros, turinčios praktiškai neribotas privilegijas prieiti prie sistemoje esančios informacijos bei atlikti sistemoje esančios informaicijos bei pačios sistemos pakeitimus [3].

3. Probe (Probing) – ataka, kurios tikslas yra perimti informaciją iš kompiuterinio tinklo [3].
4. Remote to Local (R2L) – ataka, kuria siekiama gauti vietinio (angl. local) vartotojo prieigą prie nutolusio (angl. remote) kompiuterinio įrenginio [3].

Pagrindiniai aprašomo tyrimo objektai: [3]

1. Remiantis rekurentinių neuroninių tinkle modeliu kuriamas bei implementuojamas kibernetinių incidentų aptikimo sistemos modelis [3].
2. Tiriama modelio greitaveika dvejetainėje bei daugiaklasinėje klasifikacijose, neuronų skaičiaus bei mokymosi greičio įtaka modelio tikslumui [3].
3. RNN metodo palyginimas su Naive Bayesian, Random Forest, Multi-Layer Perceptron, Support Vector Machine mašininio mokymo bei neuroninių tinkle metodais daugiaklasinėje kibernetinių incidentų kalsifikacijoje naudojant NSL-KDD duomenų rinkinį [3].
4. Greitaveikos bei tikslumo palyginimai [3].

H. Chen, Y. Liu, et al. pristato Bajeso tikimybių teorijos teorema paremtą neuroninių tinklų modelį (*angl.* Bayesian Progresive Neural Network – BPNN), kuriuo siekiama spręsti nepakankamo esamų modelių tikslumo bei didelio kiekio klaidingai teigiamų (*angl.* false positives) pranešimų problemas [3].

R. Singh, G. Srivstav sukūrė kibernetinių atakų aptikimo modelį, paremtą vienos klasės atraminių vektorių mašinos (*angl.* One Class SVM) algoritmu bei izoliaciniu aktyviuoju mokymus (*angl.* isolation based active learning), kur pagal tam tikrus parametrus duomenys yra sugrupuojami į atskirus klasterius. Šis modelis buvo apmokytas CIC-IDS2017 duomenų rinkinių, o šiuo duomenų rinkiniu apmokyto modelio tikslumas buvo geresnis už modelio, apmokyto UNSW-NB15 ir KDD99 duomenų rinkiniais [4].

T. Acharya, I. Khatri, et al. apžvelgė algoritmų, tokių kaip atsitiktinis miškas (*angl.* Random Forest – RF), J48, naivaus Bayeso (*angl.* Naive Bayes), Bayeso tinklo (*angl.* Bayesian Network) ar atraminių vektorių mašinos, efektyvumą aptinkant anomalijas bei dar nepažintas kibernetines atakas duotame kompiuterinio tinklo sraute. Taip pat šiuo tyrimu buvo nustatyta, jog tikslinių klasių (*angl.* target classes) sumažinimas duotame duomenų rinkinyje leido pasiekti didesnį kuriamo modelio tikslumą [5].

V. Bulavas apžvelgia duomenų vizualizavimo metodus ir galimą jų taikymą kibernetinių atakų aptikimui. Principinių komponentų analizė leidžia geriau pastebėti galimas anomalijas tiriamame kompiuteriniame tinkle, o sprendimų medžio (*angl.* Decision Tree) metodas įgalina vertinimo kriterijų apibrėžimą, kuriais naudojantis yra įvertinama, ar aptikta kompiuterinio tinklo anomalija bus atpažįstama kaip kibernetinis įsilaužimas į saugomą sistemą. Šiame darbe pristatyta įsilaužimų aptikimo sistema, apjungianti principinių komponentų analizę bei sprendimų medžio metodus leidžia aptikti paslaugos trikdymo (*angl.* Denial of Service) bei informacijos rinkimo (*angl.* probing) kibernetines atakas, tokias kaip Smurf, Satan, Neptune, Portsweep ar Ppsweep, didesniu nei ***95%*** tikslumu [6][7].

D. Nikolov, I. Kordev, et al. pristato kibernetinių incidentų aptikimo sistemą, paremtą rekurentinių neuroninių tinklų klasifikaciniu modeliu bei ilgalaikės-trumaplaikės atminties (*angl.* Long short term memory – LSTM) vienetais [8]. Q. Duan, X. Wei, et al. pristato konvoliucinių neuroninių tinklų modelį, skirtą apsaugoti Wi-Fi belaidžio interneto ryšį nuo kibernetinių atakų, o modelio tikslumas naudojant AWID duomenų rinkinį siekė 99% [9]. D. Hongwei ir W. Liang pristato branduolio principinių komponentų analizės (*angl.* Kernel

Principal Component Analysys – KPCA) metodą Bajeso tikimybių teorijos teorema paremtam neuroninių tinklų algoritmui. Siūlomo metodo tikslas – sumažinti duoto duomenų rinkinio matmenų (*angl.* dimensions) skaičių, pagreitinti apmokymo procesą bei pagerinti apmokyto modelio tikslumą [10]. R. Damasevicius, A. Venckauskas, et al. – KTU mokslininkai pristato naują sužymėtą duomenų rinkinį kibernetinių įsilaužimų aptikimo sistemai apmokyti – LITNET-2020. Šis duomenų rinkinys turi 85 tinklo srauto parametrus bei 12 skirtingų kibernetinių atakų tipų. Straipsnyje pateikiama statistinė duomenų rinkinio parametrų analizė bei įvairūs duomenų klasteriavimo metodai. Naudojantis šiuo duomenų rinkinių galima efektyviai aptikti visus 12 atakų tipų, o duomenys yra laisvai prieinami ir galimi naudoti moksliniais tikslais [11].

2. Duomenų rinkinys

Šiame darbe naudojamas Kanados kibernetinio saugumo institute surinktas kompiuterinio daiktų interneto tiklo duomenų rinkinys, skirtas apmokyti mašininio mokymo modelius aptikti įvairias bei plataus masto kibernetines atakas prieš daiktų interneto įrenginių tinklus. Šis duomenų rinkinys buvo surinktas puolant IoT tinklą su 105 įrenginiais 33 skirtingomis atakomis. Duomenų rinkinyje esančios atakos bei jų pasikartojimo dažniai vaizduojami 1 lentelėje [12]:

1 lentelė. Tinklo srauto tipai CIC-IoT-IDS2023 duomenų rinkinyje

Atakos Klasė	**Atakos Tipas**	**Atakos Dažnis**
Benign	Benign	1 398 195
Brute Force	Dictionary Brute Force	13 177
DDoS	ICMP Flood	7 247 733
	UDP Flood	5 448 030
	TCP Flood	4 527 598
	PSHACK Flood	4 121 654
	SYN Flood	4 085 880
	RSTFIN Flood	4 071 666
	SynonymousIP Flood	3 621 613
	UDP Fragmentation	288 780
	ACK Fragmentation	286 918
	HTTP Fragmentation	29 007
	Slowloris	23 552
DoS	UDP Flood	3 340 418
	TCP Flood	2 689 246
	SYN Flood	2 042 227
	HTTP Flood	72 334
Mirai	Greeth Flood	998 456
	UDP Plain	896 385
	Greip Flood	756 561
Recon	Host Discovery	135 322
	OS Scan	98 901
	Port Scan	82 850
	Vulnerability Scan	37 593
	Ping Sweep	2 282
Spoofing	ARP Spoofing	309 643
	DNS Spoofing	180 095
Web	Browser Hijacking	5 896
	Command Injection	5 448
	SQL Injection	5 287
	XSS	3 876
	Backdoor Malware	3 244
	Uploading Attack	1 258

3. Duomenų paruošimas

a. Duomenų normalizavimas

Apmokant mašininio mokymo modelį naudojant duomenų rinkinį, turintį daug skirtingo tipo duomenų, matuojamų skirtingo dydžio skalėmis, būtina normalizuoti visas duomenų vertes į vieną bendrą skalę. Vienas iš duomenų normalizavimo būdų – kiekvieno tipo duomenis normalizuoti individualiai bei nustatant, jog duomenų vidurkis $\mu = 0$, o standartinis nuokrypis $\sigma = 1$. ***Scikit-learn*** bibliotekoje šis metodas vadinamas ***StandardScaler()*** ir apskaičiuojamas pagal formulę [13]:

$$X_i' = \frac{X_i - \mu}{\sigma}$$

b. Tekstinių verčių kodavimas

Duomenų užkodavimas - tekstinių verčių kodavimas į skaitines vertes. Šiame darbe naudojamas ***one-hot*** duomenų kodavimo metodas, kurio esmė – duomenų stulpelyje esantį N skaičių tekstinių verčių (*angl.* categorical values) konvertuoti į N – dimensijų retą vektorių (*angl.* sparse vector), kurio visos vertės, išskyrus koduojamas vertes, yra prilyginamos nuliams. Koduojama vertė yra prilyginama vienetui [14].

c. Dimensijų mažinimo metodai

Dimensijų mažinimo metodai leidžia atrinkti duomenų rinkinio savybes, darančias didžiausią įtaką teisingai modelio prognozei. Tai išryškina duomenyse esančius įvairius dėsningumus bei sutrumpina laiką, reikalingą modeliui apmokyti.

i. Principinių komponentų analizė (Principal component analysis – PCA)

Principinių komponentų analizės statistinis metodas yra naudojamas duomenų sumažinimui iki pagrindinių savybių, kurios vadinamos principiniais komponentais. Principiniai komponentai yra apibrėžiami kaip kelios linijinės originalaus duomenų rinkinio kintamųjų kombinacijos, maksimaliai paaiškinančios varijaciją duomenyse. Šio proceso metu pateikiama duomenų rinkinio aproksimacija naudojant principinius komponentus [15][16].

ii. Linijinių diskriminantų analizė (Linear discriminant analysis – LDA)

Linijinis, prižiūrimas savybių ištraukimo algoritmas. Šio algoritmo veikimo principas paremtas naujos duomenų savybių erdvės nustatymu, į kurią būtų galima suprojektuoti duomenis su tikslu maksimizuoti klasių atskiriamumą. Iš duomenų rinkinyje esančių nepriklausomų savybių d yra ištraukiama k naujų nepriklausomų duomenų savybių, kurios labiausiai atskiria duomenų rinkinio klases. Iš žemiau pateiktų formulių galime matyti, jog, naudojant LDA algoritmą, yra sukuriamos dvi išmėtymo (scattter) matricos: (1) tarpklasinė matrica SM_b, nustatanti atstumą tarp kiekvienos klasės vidurkių ir (2) vienos klasės matrica SM_w, kuri gražina atstumą tarp klasės vidurkio ir joje esančių įrašų [16].

$$SM_b = \sum_{m=1}^{k} N_k(\mu_k - \mu)(\mu_k - \mu)^T$$

$$SM_w = \sum_{k=1}^{m} \sum_{x=1}^{n} N_k(X - \mu_k')(X - \mu_k')^T$$

4. Naudojami modeliai

a. Logistinė regresija

Linijinis prižiūrimo mašininio mokymo algoritmas, leidžiantis nustatyti tikimybę, jog turimi stebėjimų duomenys priklauso vienai ar kitai klasei. Logistinės regresijos modelį galima naudoti tiek binominio (dvejetainio), tiek ir multinominio (daugiaklasio) klasifikavimo uždaviniuose [17].

b. Atraminių vektorių mašina

Atraminių vektorių mašina (angl. Support Vector Machine – SVM) – prižiūrimo mašininio mokymo algoritmas, skirtas spręsti tiek klasifikavimo, tiek ir regresijos uždavinius. Pagrindinis tikslas naudojant šį algoritmą – rasti optimalią hiperplokštumą, maksimizuojančią atstumą tarp dviejų duomenų klasių.Taip pat, naudojant šį algoritmą išvengiama lokalaus minimumo atvejų [18].

c. K artimiausi kaimynai

K artimiausių kaimynų algoritmas (angl. K Nearest Neighbors – kNN) – prižiūrimo mašininio mokymo algoritmas, sukurtas 1951 metais mokslininkų Evelyn Fix ir Joseph Hodges. K artimiausių kaimynų algoritmas ekstensyviai naudojamas tokiose srityse, kaip duomenų gavyba (angl. data mining), rekomendacinės sistemos, ar daiktų interneto sistemos. Taip pat šis algoritmas yra dažnai naudojamas klasifikuojant tam tikrus žmogaus veiklos požymius, registruojant pasikartojančius artimiausius taškus ir atpažįstant tam tikrus veiklos šablonus. kNN algoritmą taip pat patikimai galima naudoti sistemose, skirtose aptikti įsilaužimus į sistemą ar gedimus gamybiniame procese [19].

d. Atsitiktinis miškas

Atsitiktinio miško klasifikatorius - jungtinis modelis, susidedantis iš tam tikro skaičiaus sprendimo medžių, kurių kiekvienas klasifikuoja tam tikrą poaibį iš duotos duomenų imties. Dažniausiai naudojami atsitiktinio miško variantai yra gradiento tobulinimo (*angl.* Gradient descent) metodu grįsti medžių algoritmai, tokie, kaip XGBoost ir LightGBM [20].

e. Tiesioginio sklidimo neuroniniai tinklai

Tiesioginio sklidimo neuroninis tinklas – tai neuroninis tinklas, kurio struktūra susideda iš kelių tarpusavyje sujungtų sluoksnių, kurių kiekvieną sudaro tam tikras skaičius perceptronų. Tiesioginio sklidimo neuroniniai tinklai gali būti vieno sluoksnio (*angl.* single layer feed-forward neural network) ir daugiasluoksniai (*angl.* muliti-layer feed-forward neural network). Vieno sluoksnio tiesioginio sklidimo neuroniniai tinklai turi tik 2 sluoksnius – įvesties bei išvesties. Tačiau įvesties sluoksnyje jokie skaičiavimai nėra atliekami, o įvesties signalas su tam tikrais jam priskirtais svoriais iš šio sluoksnio yra paduodamas į išvesties sluoksnį, kuriame yra apskaičuojmai išvesties signalai. Daugiasluoksniai neuroniniai tinklai turi bent vieną paslėptąjį sluoksnį (a*ngl.* hidden layer), kurio paskirtis – prasmingai įsiterpti tarp neuroninio tinklo įvesties ir išvesties bei prisidėti prie tikslesnio duoto uždavinio sprendimo Daugiasluoksnis tiesioginio sklidimo neuroninis tinklas gali būti naudojamas klasifikavimo uždaviniams spręsti, kai vienas ar keli paslėptieji sluoksniai naudojami klasifikavimui atlikti, o išvesties sluoksnyje esantys perceptronai reprezentuoja klasifikavimo uždavinyje apibrėžtas klases [21][22].

f. LSTM rekurentiniai neuroniniai tinklai

Ilgos trumpalaikės atminties modelis (*angl.* Long Short Term Memory – LSTM) – rekurentinio neuroninio tinklo atminties modelis skirtas išspręsti nykstančio bei sprogstančio gradiento problemas, pirmą kartą pristatytas mokslininkų S. Hochreiter ir J. Schmidhuber 1992 metais [23]. Ilgalaikės trumpalaikės atminties modelio pagrindas – atminties ląstelė (*angl.* memory cell), kuri pakeičia standartiniame rekurentiniame neuroniniame tinkle esančią rekurentinę paslėptąją būseną. Kiekviena atminties ląstelėje turi rekurentinį kraštą (a*ngl.* Recurrent Edge), kurio norimas svoris , taip siekiant išspręsti

nykstančio ar sprogstančio gradiento problemas. Visos vertės, susijusios su šiuo rekurentiniu kraštu, bendrai yra vadinamos ląstelės būsena (*angl.* Cell State) [24].

5. Našumo metrikos

Tikslumas (*angl.* Accuracy) – teisingai klasifikuotų įrašų skaičius nuo visų duomenų rinkinyje esančių įrašų [25]:

$$Tikslumas = \frac{TP + TN}{TP + TN + FP + FN}$$

Precizija (*angl.* Precision) – teisingai teigiamai klasifikuotų reikšmių skaičius nuo visų teigiamai klasifikuotų reikšmių [25]:

$$Precizija = \frac{TP}{TP + FP}$$

Atpažinimas (*angl.* Recall) – teisingai teigiamai klasifikuotų reikšmių skaičius nuo visų teigiamų reikšmių, esančių duomenų rinkinyje [25]:

$$Atpažinimas = \frac{TP}{TP + FN}$$

F1 – harmoninis precizijos ir atpažinimo vidurkis [24]:

$$F1 = 2 * \frac{Precizija * Atpažinimas}{Precizija + Atpažinimas}$$

Uždelsimas (*angl.* Latency) – laikas, reikalingas vienam įrašui klasifikuoti, matuojamas milisekundėmis.

Apmokymo laikas – laikas, reikalingas mašininio mokymo modeliui apmokyti.

6. Eksperimentas ir rezultatai

Eksperimento metu buvo siekiama nustatyti bei palyginti parinktų mašininio mokymo modelių našumą dvejetainio klasifikavimo (tarp įprastinio tinklo srauto ir kibernetinės atakos) bei daugiaklasio klasifikavimo (tarp įprastinio srauto ir įvairių kibernetinių atakų) uždaviniuose. Duomenų paruošimui, modelių apmokymui bei našumo įvertinimui buvo naudojama ***Python*** programavimo kalba bei ***Pandas***, ***Scikit-learn*** ir ***Tensorflow*** bibliotekos. Visuose apmokomuose modeliuose, išskyrus neuroninius tinklus, naudojami Scikit-learn bibliotekoje numatytieji (default) hiperparametrai. Tiesioginio sklidimo neuroninių tinklų architektūra vaizduojama 2 lentelėje:

2 lentelė. Tiesioginio sklidimo neuroninio tinklo architektūra

Parametras	Vertė
Sluoksnių skaičius	4
Įvesties sluoksnio neuronų skaičius	47
1 paslėpto sluoksnio neuronų skaičius	1024
2 paslėpto sluoksnio neuronų skaičius	512
Išvesties sluoksnio neuronų skaičius	8
Epochų skaičius	20

LSTM rekurentinio neuroninio tinklo architektūra vaizduojama 1 paveiksle (1 pav.):

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 20, 64)	28,672
dropout (Dropout)	(None, 20, 64)	0
lstm_1 (LSTM)	(None, 32)	12,416
dropout_1 (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 64)	2,112
dense_5 (Dense)	(None, 8)	520

1 pav. LSTM reukrentinio neuroninio tinklo architektūra

a. Dvejetainis klasifikavimas

Siekiant paspartinti mašininio mokymo modelių apmokymo procesą, dvejetainio klasifikavimo eksperimentui bei mašininio mokymo modelių apmokymui ir testavimui atlikti iš maždaug 40 milijonų įrašų, esančių CIC-IoT-IDS2023 duomenų rinkinyje, buvo paimtas duomenų poaibis su 60 tūkstančių įrašų, iš kurių 30 tūkstančių įrašų buvo įprastinio tinklo srauto, o likę 30 tūkstančių – įvairios kibernetinės atakos. Apmokymo duomenų rinkinys vaizduojamas 3 lentelėje:

3 lentelė. Apmokymo duomenų rinkinys dvejetainiam klasifikavimui

Tinklo srauto tipas	Įrašų skaičius
Įprastinis tinklo srautas	30,000
Paskirstyto paslaugos trikdymo atakos	22,304
Paslaugos trikdymo atakos	5,314
Mirai botnet atakos	1,769
Spoofing atakos	323
Informacijos rinkimo (recon) atakos	259
Tinklo atakos	18
Grubios jėgos atakos	13

Dvejetainio klasifikavimo našumo rezultatai vaizduojami 4 lentelėje:

4 lentelė. Dvejetainio klasifikavimo rezultatai

Modelis / Metrika	**Tiklsumas**	**Atpažinimas**	**F1**	**Uždelsimas (ms)**
Logistinė regresija	98.49%	98.36%	98.49%	0,0007
Atraminių vektorių mašina	99.10%	99.09%	99.09%	0,7434
K artimiausi kaimynai	99.22%	99.22%	99.22%	0,9381
Tiesioginio sklidimo neuroninis tinklas	99.31%	99.34%	99.34%	0,6930
Atsitiktinis miškas	99.60%	99.57%	99.57%	0,0348
LightGBM	99.63%	99.63%	99.63%	0,0132
XGBoost	99.64%	99.63%	99.63%	0,0048
LSTM	100.00%	100.00%	100.00%	0,6930

Tiksliausias modelis modelis buvo LSTM rekurentinių neuroninių tinklų modelis su 100% tikslumo ir atpažinimo rezultatais. Antrą geriausią rezultatą pademonstravo sprendimų medžiais grįsti modeliai – atsitiktiniai medžiai, XGBoost ir LightGBM, kurie viršijo 99.6% tikslumą. Praščiausias našumas – logistinės regresijos modelio, pasiekusio 98.49% tikslumą, tačiau turėjo trumpiausią uždelstumo laiką – 0,00007 milisekundės. Didžiausias uždelstumas buvo pademonstruotas k artimiausių kaimynų modelio.

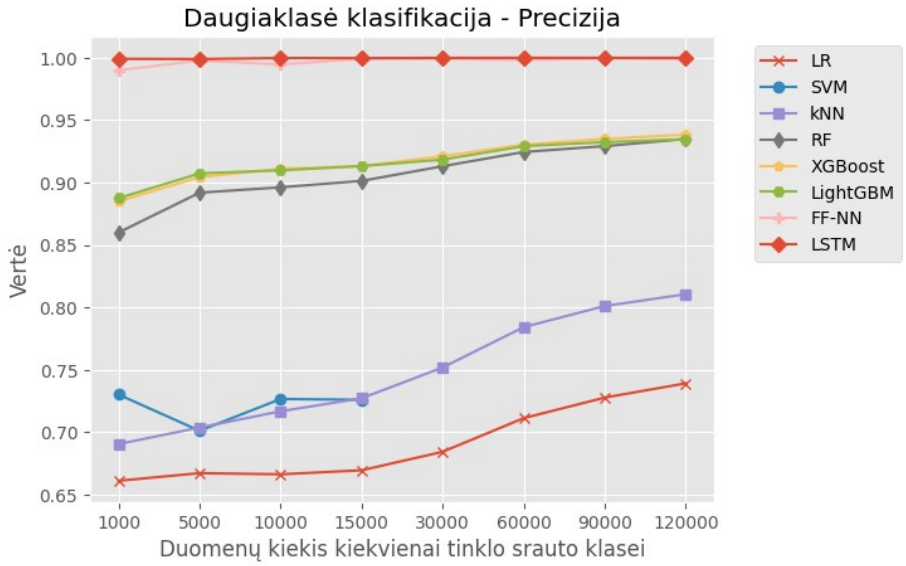
b. Daugiaklasis klasifikavimas

Daugiaklasio klasifikavimo našumui nustatyti buvo sukurti aštuoni CIC-IoT-IDS2023 duomenų rinkinio variantai su skirtingu skaičiumi įrašų kiekvienai tinklo srauto, esančio duomenų rinkinyje, klasei. Buvo sukurti šie duomenų rinkinio variantai:

1. 1,000 įrašų kiekvienai tinklo srauto klasei;
2. 5,000 įrašų kiekvienai tinklo srauto klasei;
3. 10,000 įrašų kiekvienai tinklo srauto klasei;
4. 15,000 įrašų kiekvienai tinklo srauto klasei;
5. 30,000 įrašų kiekvienai tinklo srauto klasei;
6. 60,000 įrašų kiekvienai tinklo srauto klasei;
7. 90,000 įrašų kiekvienai tinklo srauto klasei;
8. 120,000 įrašų kiekvienai tinklo srauto klasei.

Šių duomenų rinkinio variantų sukūrimo tikslas – įvertinti apmokomų modelių našumo pokytį kintant apmokymo duomenų skaičiui. Dėl ilgo apmokymo laiko, atraminių vektorių mašinos modelis nebuvo apmokomas su duomenų rinkinio variantais, kurių įrašų skaičius vienai tinklo srauto klasei viršijo 15 tūkstančių įrašų. Galiausiai, svarbu pabrėžti, jog CIC-IoT-IDS2023 duomenų rinkinys yra nesubalansuotas – grubios jėgos atakų yra 13,177 įrašų, o tinklo atakų yra 25,009 įrašai – ženkliai mažiau, nei kitų tinklo srauto klasių, esančių duomenų rinkinyje, įrašų. To pasekoje buvo vertinamas ne vien tik bendras modelių našumas daugiaklasio klasifikavimo uždavinyje, bet ir našumas klasifikuojant kiekvieną tinklo srauto klasę atskirai.

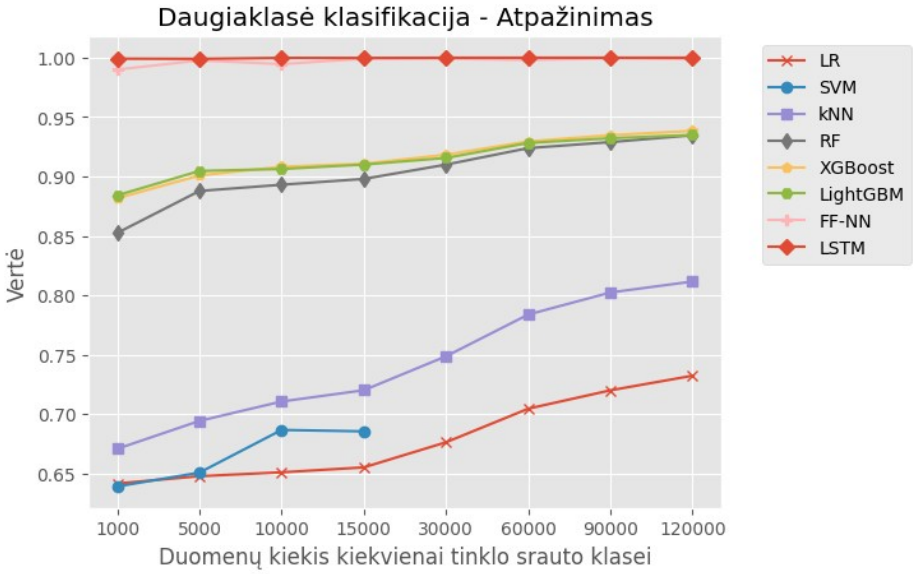
Bendro daugiaklasio klasifikavimo tikslumas vaizduojamas 2 paveiksle (2 pav.):



2 pav. Daugiaklasio klasifikavimo tikslumas

Visų modelių, išskyrus atraminių vektorių mašiną bei neuroninius tinklus, precizija kilo didinant apmokymo duomenų skaičių, o LSTM rekurentiniai neuroniniai tinklai pasiekė 99.91% tikslumą jau su tūkstančiu kiekvienos klasės įrašų. Iš visų sprendimų medžiais paremtų modelių, prasčiausią rezultatą pademonstravo atsitiktinio miško metodas, tačiau, didinant duomenų įrašų skaičių, sprendimo medžiais grįstų modelių rezultatai susivienodino. Prasčiausią precizijos rezultatą pademonstravo logistinės regresijos modelis.

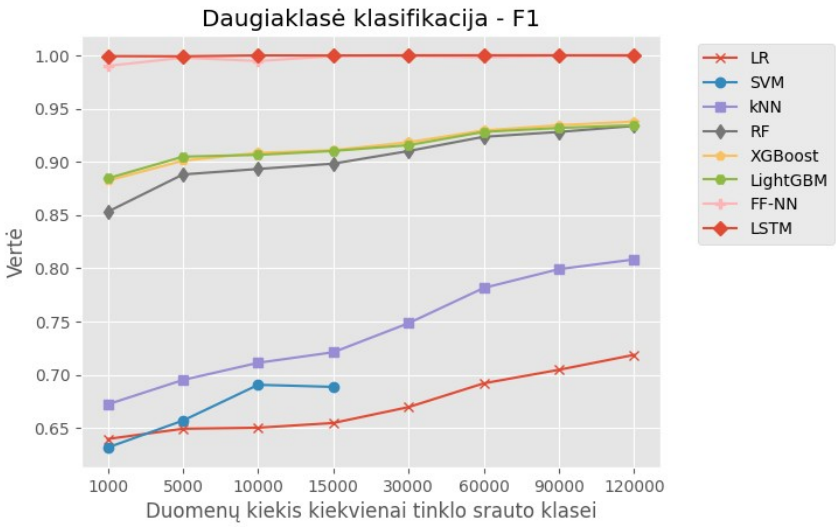
Bendro daugiaklasio klasifikavimo atpažinimas vaizduojamas 3 paveiksle (3 pav.):



3 pav. Daugiaklasio klasifikavimo atpažinimo rezultatas

Neuroninių tinklų modeliai pasiekė didesnę nei 99% atpažinimo vertė su sąlyginai mažais duomenų kiekiais, kurį išlaikė ir su didesniu duomenų kiekiu. Sprendimų medžiais paremtų modelių atpažinimo rezultatai nesiskiria nuo tikslumo rezultatų. Prasčiausią atpažinimo rezultatą parodė atraminių vektorių mašinos ir logistinės regresijos modeliai.

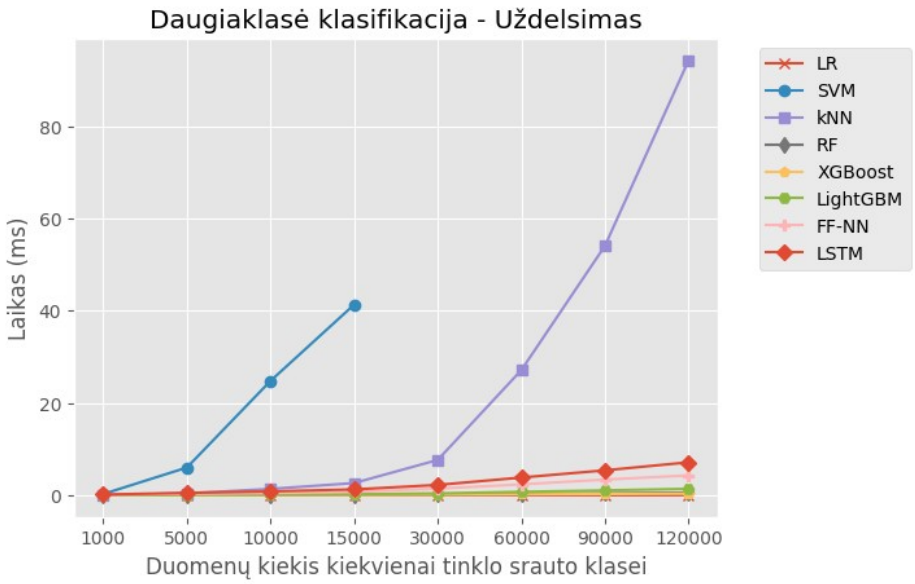
Bendro daugiaklasio klasifikavimo F1 vertė vaizduojama 4 paveiksle (4 pav.):



4 pav. Daugiaklasio klasifikavimo F1 įvertis

Naudojant harmoninį precizijos ir atpažinimo vidurkio įvertį, geriausius rezultatus pademonstravo neuroninių tinklų modeliai. Prasčiausius rezultatus parodė atraminių vektorių mašina bei logistinės regresijos modelis, o sprendimo medžiais grįstų modelių rezultatai supanašėja naudojant daugiau duomenų.

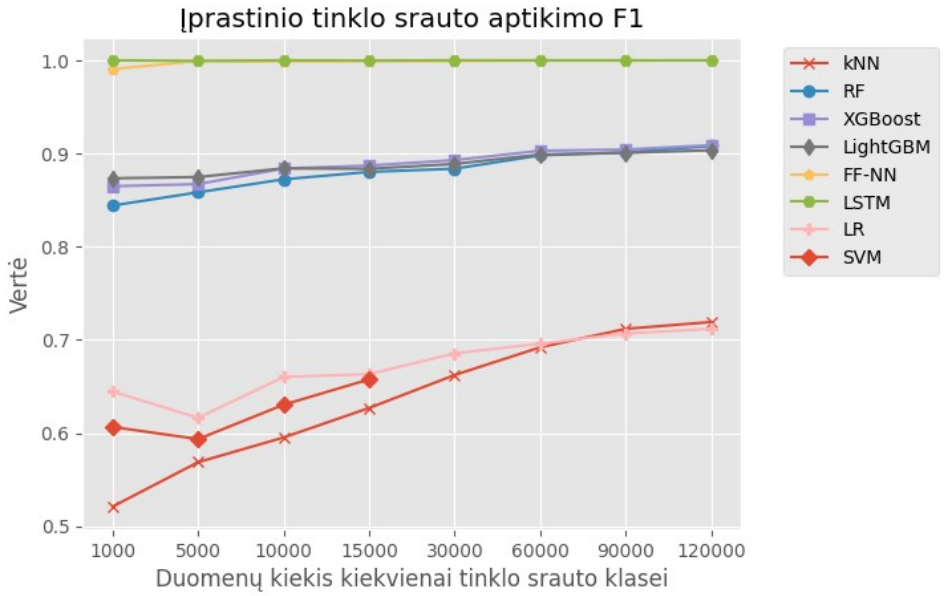
Bendro daugiaklasio klasifikavimo uždelstumas yra vaizduojamas 5 paveiksle (5 pav.):



5 pav. Daugiaklasio kasifikavimo uždelsimas

Modeliai, apmokyti su daugiau duomenų, užtrunka ilgesnį laiką, kad klasifikuoti naujus duomenis. Su 15000 duomenų kiekvienai klasei, ilgiausią uždelsimo laiką turėjo atraminių vektorių mašina – 41.29 milisekundės vienam pavyzdžiui klasifikuoti. Su 120000 duomenų kiekvienai klasei, ilgiausią uždelsimo laiką turėjo K artimiausių kaimynų modelis – 92.24 milisekundės vienam pavyzdžiui klasifikuoti. Mažiausią uždelsimą turėjo logistinės regresijos modelis – 0.02 milisekundės vienai klasifikacijai.

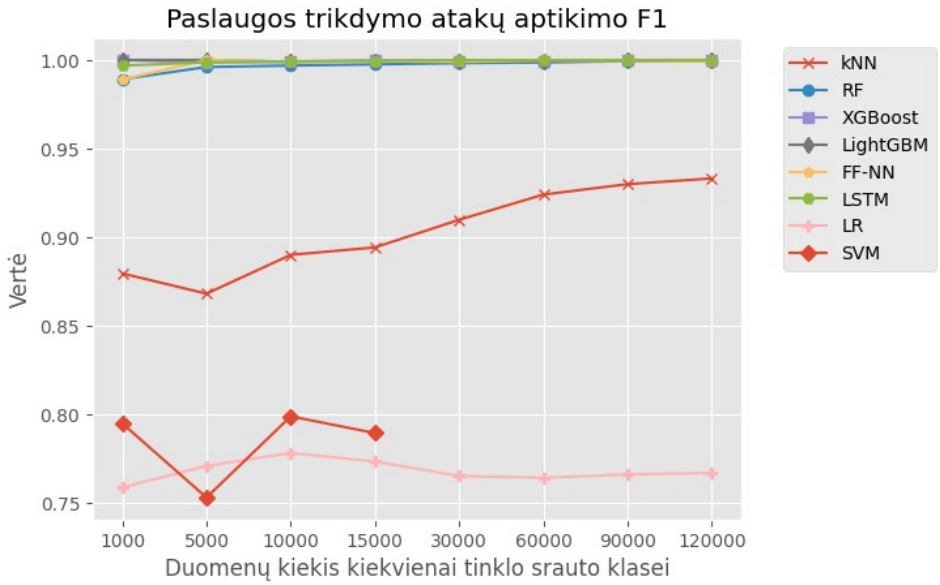
Įprastinio tinklo srauto aptikimo F1 vaizduojamas 6 paveiksle (6 pav.):



6 pav. Įprastinio tinkle srauto aptikimo F1 įvertis

Geriausią įprastinio tinklo srauto klasifikavimo rezultatą pasiekia neuroninių tinklų modeliai. Prasčiausias rezultatas – K artimiausių kaimynų modelio, tačiau, didejant duomenų kiekiui, šio modelio našumas susilygina su logistinės regresijos modelio našumu. Sprendimo medžiais grįstų modelių našumas klasifikuojant įprastinį tinklo srautą tarpusavyje yra panašus.

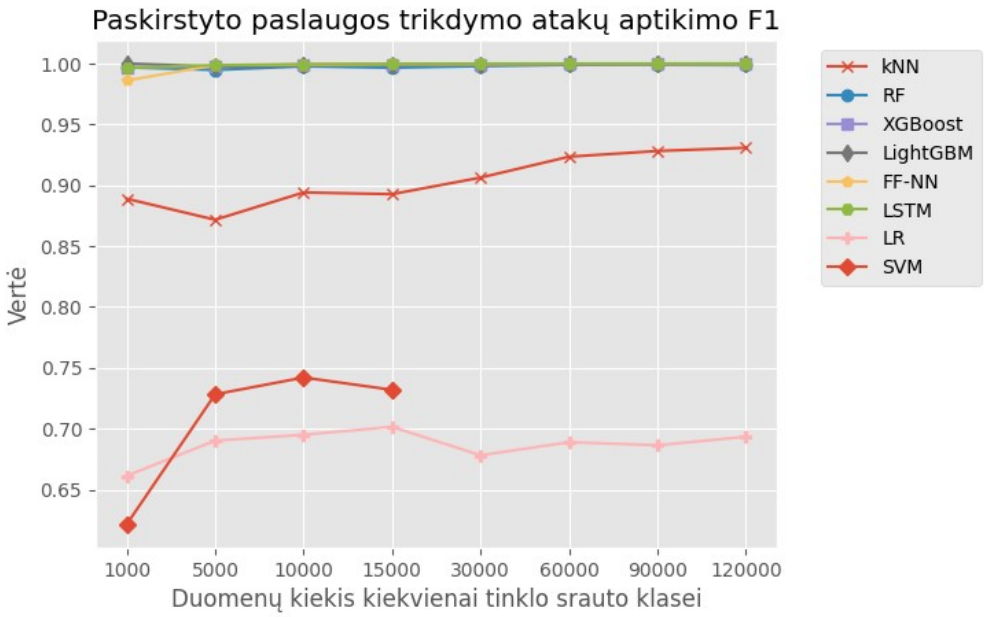
Paslaugos trikdymo atakų aptikimo F1 vaizduojamas 7 paveiksle (7 pav.):



7 pav. Paslaugos trikdymo atakų aptikimo F1 įvertis

Prasčiausią rezultatą pademonstravo logistinės regresijos bei atraminių vektorių mašinos modeliai. Naudojant daugiau duomenų apmokymui, sprendimų medžiais grįstų modelių našumas susilygina su neuroninių tinklų modelių našumu. Sprendimų medžių bei neuroninių tinklų modelių našumas aptinkant paslaugos trikdymo atakas, viršija 99.9%.

Paskirstyto paslaugos trikdymo atakų aptikimo F1 vaizduojamas 8 paveiksle (8 pav.):

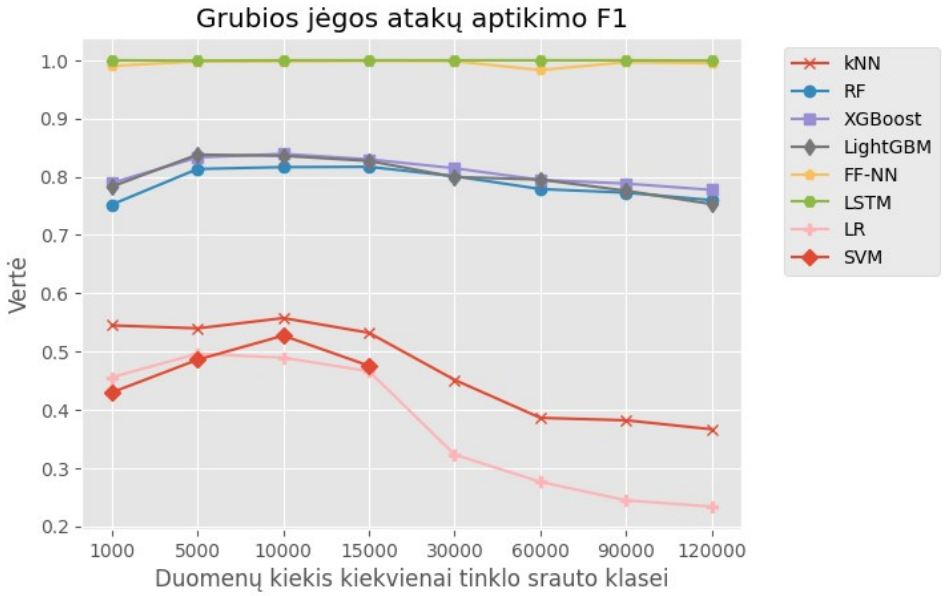


8 pav. Paskirstyto paslaugos trikdymo atakų aptikimo F1 įvertis

Prasčiausią rezultatą pademonstravo logistinės regresijos bei atraminių vektorių mašinos modeliai. Logistinės regresijos modelio bei atraminių vektorių mašinos rezultatai

klasifikuojant paskirstyto paslaugos trikdymo atakas yra prastensi nei įprastinių paslaugos trikdymo atakų klasifikavimo rezultatai. Naudojant daugiau apmokymo duomenų, sprendimų medžiais grįstų modelių našumas susilygina su neuroninių tinklų modelių našumu bei viršija 99.9%.

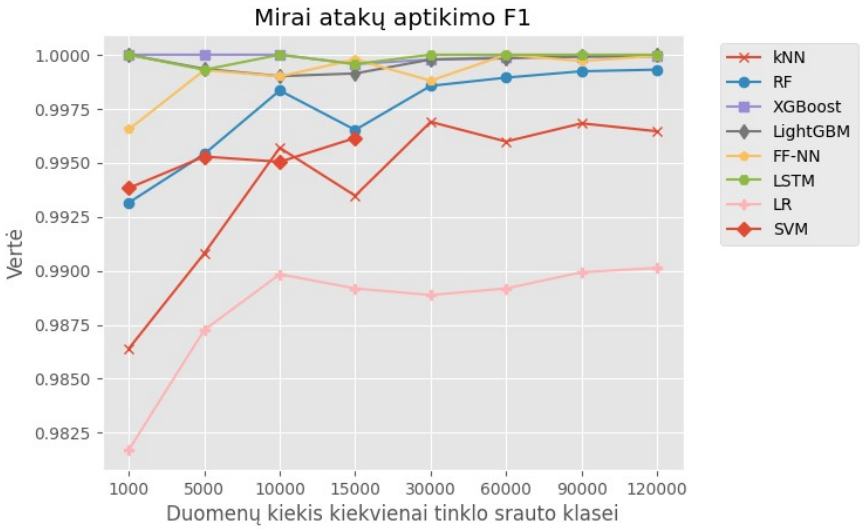
Grubios jėgos atakų aptikimo F1 vaizduojamas 9 paveiksle (9 pav.):



9 pav. Grubios jėgos atakų aptikimo F1 įvertis

Geriausias rezultatas pasiektas naudojant neuroninių tinklų modelius. Geriausi F1 rezultatai, naudojant sprendimo medžiais paremtus modelius, siekė 81% - 83%. Prasčiausias rezultatas pasiektas naudojant logistinės regresijos modelį. Naudojamame duomenų rinkinyje iš viso yra 13,177 grubios jėgos atakų įrašų. Didėjant kitų atakų įrašų skaičiui, grubios jėgos atakų aptikimo našumas mažėja naudojant visus modelius, išskyrus neuroninius tinklus.

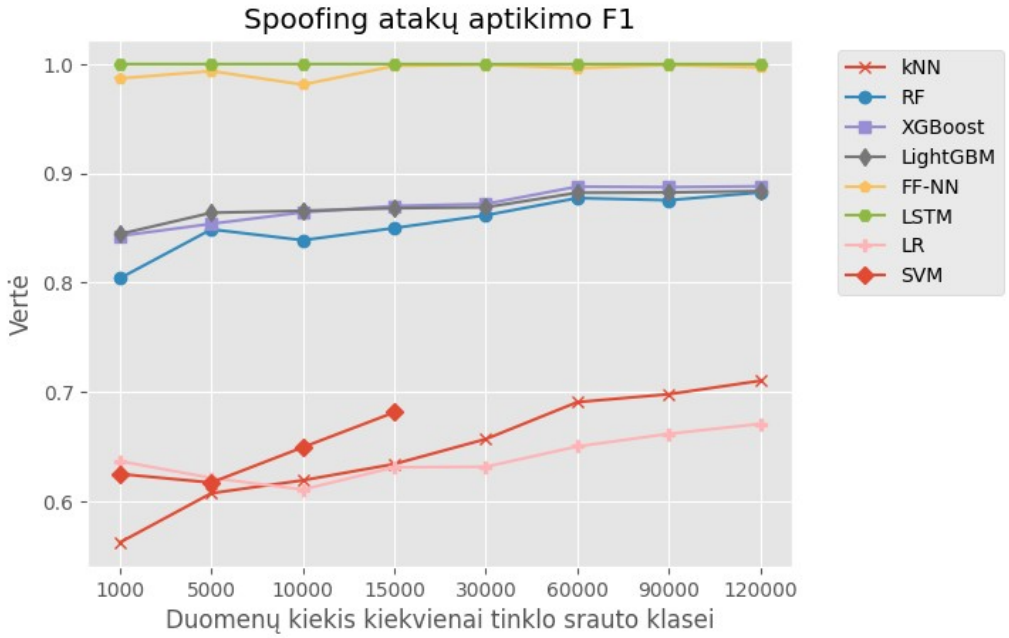
Mirai botnet atakų aptikimo *F1* vaizduojamas 10 paveiksle (10 pav.):



10 pav. Mirai botnet atakų aptikimo F1 įvertis

Visų apmokytų modelių našumas aptinkant Mirai botnet atakas viršijo 98%. Praščiausią rezultatą pademonstravo logistinės regresijos bei K artimiausių kaimynų modeliai. Likę modeliai viršijo 99.25% F1 rezultatą.

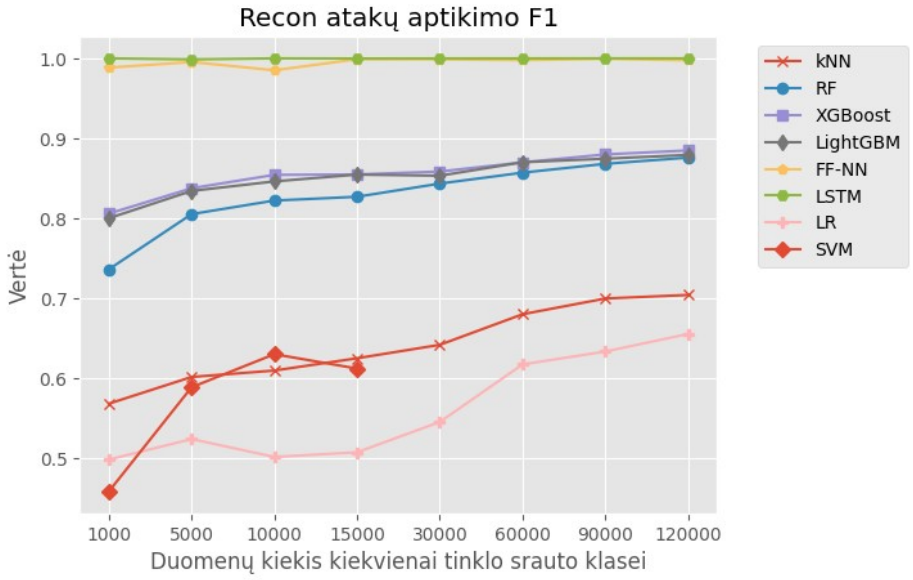
Spoofing atakų aptikimo F1 vaizduojamas 11 paveiksle (11 pav.):



11 pav. Spoofing atakų aptikimo F1 įvertis

LSTM neuroninių tinklų modelis viršijo 99.9% rezultatą nuo 1000 įrašų, o tiesioginio sklidimo neuroninis tinklas – nuo 15000 įrašų. Sprendimų medžių modelių F1 našumo rezultatai panašūs ir siekė 87%. Prasčiausias rezultatas buvo pademonstruotas logistinės regresijos, K artimiausių kaimynų ir atraminių vektorių mašinos modelių.

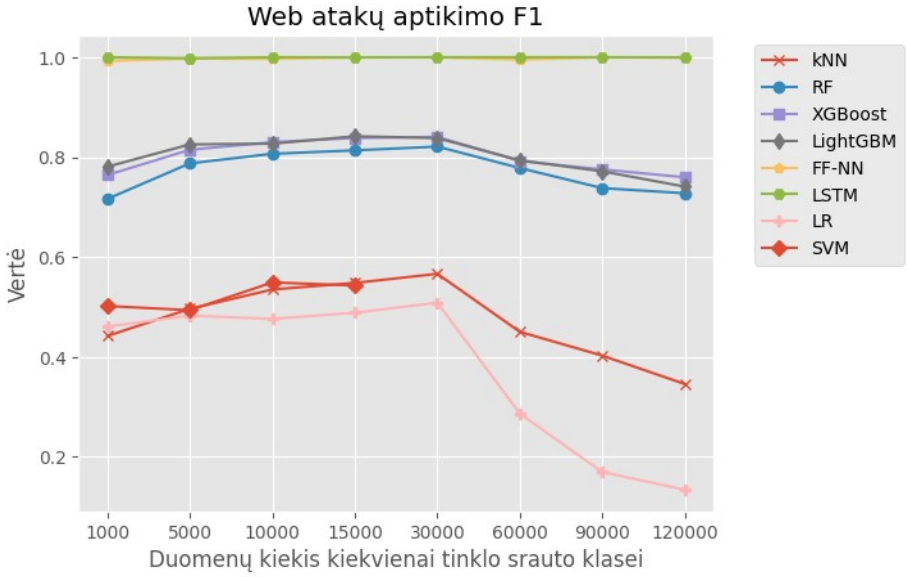
Informacijos rinkimo atakų aptikimo F1 vaizduojamas 12 paveiksle (12 pav.):



12 pav. Informacijos rinkimo atakų aptikimo F1 įvertis

Informacijos rinkimo (recon) atakos geriausiai klasifikuotos neuroninių tinklų bei sprendimų medžių modelių. Prasčiausią našumą pademonstravęs modelis – logistinė regresija.

Web atakų aptikimo F1 vaizduojamas 13 paveiksle (13 pav.):



13 pav. Web atakų aptikimo F1 įvertis

Web atakos, kaip ir grubios jėgos atakos, turi mažiau įrašų duomenų rinkinyje, nei kitų tipų atakos. Web atakų įrašų skaičius – 25009. Visi modeliai, išskyrus neuroninius tinklus, didinant kitų atakų skaičių apmokymo duomenų rinkinyje, prasčiau klasifikavo web atakas.

Išvados ir tolimesni darbai

Remiantis gautais rezultatais, teikiamos šios išvados:

1. Dvejetainiame klasifikavime praščiausią tikslumą pademonstravo logistinės regresijos modelis su 98.49% tikslumu, o LSTM rekurentinių tinklų modelis pasiekė 100% tikslumą.
2. Mažiausias uždelsimas (latency) dvejetainiame klasifikavime pasiektas naudojant logistinę regresiją, o didžiausias – K artimiausių kaimynų modelį.
3. Daugiaklasiame klasifikavime geriausias našumas buvo pasiektas naudojant tiesioginio sklidimo ir LSTM neuroninių tinklų modelius.
4. Naudojamas duomenų rinkinys CIC-IoT-IDS2023 yra nesubalansuotas – grubios jėgos bei web atakų įrašų skaičius yra mažesnis nei kitų atakų duomenų rinkinyje, ko pasekoje, naudojant tam tikrus modelius, daugiaklasiame klasifikavime yra pasiekiamas mažesnis minėtų atakų aptikimo našumas.
5. Neuroninių tinklų apmokymo laikas, naudojant mažesnį duomenų kiekį artimas sprendimo medžiais grįstų modelių apmokymo laikui, o tikslumas – geresnis. Vis dėlto, sprendimų medžiais grįstų modelių uždelsimas yra mažesnis.

Šį tyrimą pratęsti ir rezultatus papildyti būtų galima šiais eksperimentais:

1. Sprendimo medžiais grįstų modelių hiperparametrų optimizavimas (XGBoost, LightGBM) – tai leistų įvertinti, kokią įtaką modelio našumui turi jo parametrai. Šio eksperimento tikslas galėtų būti padidinti sprendimo medžių modelių tikslumą išlaikant klasifikavimo laiką trumpesnį nei neuroninių tinklų.

2. Tyrimą būtų galima išplėsti realaus laiko kibernetinių incidentų aptikimo komponentu – sukurti programą, galinčią ištraukti savybes iš neapdoroto kompiuterinio tinklo srauto bei paduoti šias savybes neprižiūrimo (unsupervised) mokymo modeliui. Neprižiūrimam mokymui galėtų būti naudojami autoenkoderiai, transformeriai ar K-vidurkių (K-means) algoritmas. Kartu su tikslumu į vertinimo metrikas gali būti įtrauktas laikas, reikalingas srauto apdorojimui bei klasifikuoti konkretų įrašą. Testavimui gali būti naudojami kraštiniai (*angl.* edge) įrenginiai.

Literatūros sąrašas

1. Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5:21954–21961, 2017.
2. Lin Chen, Xiaoyun Kuang, Aidong Xu, Siliang Suo, and Yiwei Yang. A novel network intrusion detection system based on cnn. In 2020 Eighth International Conference on Advanced Cloud and Big Data (CBD), pages 243–247, Dec 2020.
3. Haonan Chen, Yiyang Liu, Jianming Zhao, and Xianda Liu. Research on intrusion detection based on bp neural network. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pages 79–82, Jan 2021.
4. Richa Singh and Gaurav Srivastav. Novel framework for anomaly detection using machine learning technique on cic-ids2017 dataset. In 2021 International Conference on Technological Advancements and Innovations (ICTAI), pages 632–636, Nov 2021.
5. Toya Acharya, Ishan Khatri, Annamalai Annamalai, and Mohamed F Chouikha. Efficacy of machine learning-based classifiers for binary and multi-class network intrusion detection. In 2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS), pages 402–407, June 2021.
6. Viktoras Bulavas. Investigation of network intrusion detection using data visualization methods. In 2018 59th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), pages 1–6, Oct 2018.
7. Viktoras Bulavas. Applying machine learning methods for early detection of cyber security incidents. 2018.
8. Dimitar Nikolov, Iliyan Kordev, and Stela Stefanova. Concept for network intrusion detection system based on recurrent neural network classifier. In 2018 IEEE XXVII International Scientific Conference Electronics - ET, pages 1–4, Sep. 2018. 3
9. Qiang Duan, Xianglin Wei, Jianhua Fan, Long Yu, and Yongyang Hu. Cnn-based intrusion classification for ieee 802.11 wireless networks. In 2020 IEEE 6th International Conference on Computer and Communications (ICCC), pages 830–833, Dec 2020.
10. Ding Hongwei and Wan Liang. Research on intrusion detection based on kpca-bp neural network. In 2018 IEEE 18th International Conference on Communication Technology (ICCT), pages 911–915, Oct 2018.
11. Robertas Damasevicius, Algimantas Venckauskas, Sarunas Grigaliunas, Jevgenijus Toldinas, Nerijus Morkevicius, Tautvydas Aleliunas, and Paulius Smuikys. Litnet-2020: An annotated real-world network flow dataset for network intrusion detection. Electronics, 9(5), 2020.

12. Canadian Institute for Cybersecurity. CIC IoT dataset 2023, 2023. Accessed: March 26, 2025.
13. Dalwinder Singh and Birmohan Singh. Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97:105524, 2020.
14. Cedric Seger. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. PhD thesis, 2018.
15. Michael Greenacre, Patrick J. F. Groenen, Trevor Hastie, Alfonso Iodice D'Enza, Angelos Markos, and Elena Tuzhilina. Principal component analysis. Nature Reviews Methods Primers, 2(1):100, Dec 2022.
16. Farzana Anowar, Samira Sadaoui, and Bassant Selim. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). Computer Science Review, 40:100378, 2021.
17. Craig Starbuck. Logistic Regression, pages 223–238. Springer International Publishing, Cham, 2023.
18. Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine Learning, 20(3):273–297, Sep 1995.
19. Rajib Halder, Mohammed Uddin, Md Ashraf Uddin, Sunil Aryal, and Ansam Khraisat. Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. Journal of Big Data, 11, 08 2024.
20. Adele Cutler, David Cutler, and John Stevens. Random Forests, volume 45, pages 157–176. 01 2011.
21. Jo˜ao Alexandre Lˆobo Marques, Francisco Nauber Bernardo Gois, Jo˜ao Paulo do Vale Madeiro, Tengyue Li, and Simon James Fong. Chapter 4 - artificial neural network-based approaches for computer-aided disease diagnosis and treatment. In Akash Kumar Bhoi, Victor Hugo C. de Albuquerque, Parvathaneni Naga Srinivasu, and Gon¸calo Marques, editors, Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data, Intelligent Data-Centric Systems, pages 79–99. Academic Press, 2022.
22. Murat Sazli. A brief review of feed-forward neural networks. Communications Faculty Of Science University of Ankara, 50:11–17, 01 2006.
23. Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory. Neural Computation, 9:1735–1780, 11 1997.
24. Vahid Mirjalili Sebastian Raschka. Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn and TensorFlow 2. 3rd Edition. Packt Publishing, Birmingham, 2019.
25. Oona Rainio, Jarmo Teuho, and Riku Kl´en. Evaluation metrics and statistical tests for machine learning. Scientific Reports, 14(1):6086, Mar 2023

COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR INTRUSION DETECTION IN NETWORKS OF INTERNET OF THINGS DEVICES

Kostas Ereksonas

Klaipeda University

With the increasing number of Internet of Things devices and networks deployed, novel cyber security threats are emerging. This paper presents a comparative analysis of various machine learning and deep learning models, focusing on the precision and recall classification scores of the models trained to detect cyber attacks on an IoT network. The CIC_IoT_IDS2023 dataset was used for model training. Lastly, potential directions for future work on this topic are suggested, primarily concerning the optimisation of boosted trees, the addition of a real-time intrusion detection component, and the testing of the intrusion detection system's performance on edge devices.

Keywords: CIC_IoT_IDS2023, machine learning, deep learning, internet of things, intrusion detection system