



ATASKAITINĖ INFORMATIKOS KRYPTIES DOKTORANTŲ  
KONFERENCIJA 2021 M. KOVO 26 D.

# ATASKAITA

DOKTORANTAS VIKTORAS BULAVAS  
INFORMATIKA (N009)

VADOVAS: PROF. HABIL. DR. GINTAUTAS DZEMYDA

KONSULTANTAS: DR. VIRGINIJUS MARCINKEVIČIUS

DOKTORANTŪROS LAIKOTARPIS 2017 M. - 2021 M.

# Disertacijos tyrimo objektas, tikslai ir planuojami gauti rezultatai

- ▶ Preliminari disertacijos tema ir tyrimo objektas:
  - ▶ **Mašininio mokymo metodų taikymas ankstyvajam kibernetinių incidentų aptikimui**
- ▶ Tyrimo tikslai:
  - ▶ Sukurti arba patobulinti mašininio mokymosi grįstą metodą, skirtą ankstyvajam kibernetinių incidentų aptikimui
- ▶ Planuojami gauti rezultatai:
  - ▶ Panaudoti parinktus metodus, siekiant prognozuoti bei valdyti ankstyvąjį kibernetinių incidentų etapą

# Tarptautiniai renginiai: pranešimas

Pranešimas seminare  
AI Bay / DIH4.AI  
Seminar, Gdańsk, 2021  
sausio 28 d.

"Imbalance curse: case  
study of machine  
Learning on  
cybersecurity data sets"



# Kiti pristatymai

- ▶ Pristatymas Gdanskio technologijų universiteto Elektronikos ir informatikos fakulteto darbuotojams 2020 spalio 15 d.
- ▶ Metinė ataskaitinė informatikos krypties doktorantų konferencija 2020 m. spalio 21 d.

# Publikacijos

- ▶ Informatica „Study of supervised machine learning algorithms performance on network intrusion detection data with rare classes“ (Įteikta)

# Ketvirtųjų mokslo metų darbo planas

5. Atskirų daktaro disertacijos dalių (analizės rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas (2020 m. spalio– 2021 m. gegužė):

5.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas.

5.2. Analitinės disertacijos dalies parengimas.

5.3. Teorinės disertacijos dalies parengimas.

5.4. Eksperimentinės disertacijos dalies parengimas.

5.5. Bendrųjų išvadų formulavimas.

6. Daktaro disertacijos parengimas ir svarstymas padalinyje (2021 m. birželis).

7. Daktaro disertacijos pateikimas (2021 m. rugsėjis).

- ▶ Planuojama parengti dvi mokslines tyrimų publikacijas (recenzuojamame leidinyje, WoS su Impact Factor).



**Vilnius  
University**



# RETŲ KLASIŲ APTIKIMO KIBERNETINIO SAUGUMO RINKINIUOSE GERINIMAS DIRBTINIŲ DUOMENŲ ĮTERPIMO IR KLASIFIKAVIMO KLAIDOMS JAUTRIAIS METODAIS

# Pristatymo turinys

- ▶ Trumpai apie nesubalansuotų klasių problematiką ir tyrimui pasirinktus duomenų rinkinius
- ▶ Nesubalansuotų duomenų rinkinių paruošimo mašininiam mokymui galimybės
- ▶ Klasifikavimas klaidoms jautriais metodais
- ▶ Duomenų klasių disbalansui atskleisti tinkami rodikliai
- ▶ Tyrimo rezultatų aptarimas





© <https://www.krastozinios.lt/2019/11/11>

# Tyrimo tikslas

- ▶ Parinkti priemones, naudotinas duomenų paruošimui ir mašinos mokymui kuo tiksliau atpažinti retus kibernetinio saugumo įvykius, nes siekiama išvengti praleidimų (FN) ir apsipažinimų (FP).

# Eksperimento eiga

- ▶ Lyginamajam tyrimui paruošti ir dirbtiniais duomenimis algoritmų mokymo tikslu papildyti duomenys iš kibernetinio saugumo tyrimams skirtų tinklo lygio duomenų šaltinių CIC-IDS2017, CSE-CIC-IDS2018 ir LITNET-2020;
- ▶ Python aplinkoje, įgyvendinant klasifikavimo klaidoms jautrius mokymo metodus, parengti mašinų mokymo algoritmai;
- ▶ Parinkti mokymo hiperparametrai ir atliktas validavimas modelių nematytais anksčiau duomenimis;
- ▶ Suskaičiuoti algoritmų subalansuoto tikslumo (balanced accuracy score) ir geometrinio subalansuoto tikslumo (G-mean) rodikliai, palygintas su ir be dirbtinių duomenų panaudojimo gautas efektyvumas.

# Nesubalansuotų klasių problematika

- ▶ Mašininio mokymo algoritmai, jei jų teisingai nenukreipti, mokosi ir išmoksta atpažinti didžiausias klases.
- ▶ Jei matuosime modelių efektyvumą tradiciniais tikslumo matais, tai turėdami kelis šimtus teigiamų pavyzdžių tarp keliasdešimt milijonų, net visiškai ignoruodami retas klases, gausime artimą vienetui pagrindinės klasės aptikimo tikslumą, proporcingą piktybinių įrašų daliai nesibaigiančiame gerų įrašų sraute.
- ▶ Toks rezultatas netinka kibernetinių incidentų tyrimo atveju, nes tyrimui beveik nesvarbūs gerybiniai tinklo duomenys. Kibernetinių incidentų prevencijos uždaviniuose siekiame kuo kokybiškiau aptikti ir sustabdyti būtent retuosius įvykius, todėl siekiame įvertinti būtent retųjų klasių atpažinimo tikslumą.
- ▶ Gerybiniai duomenys reikalingi tiek, kiek jie padeda apmokyti mašinas atpažinti piktybinius įvykius.

# CIC-IDS duomenų rinkiniai

- ▶ Kanados kibernetinio saugumo tyrimų centro duomenų rinkiniai CIC-IDS2017 ir CSE-CIC-IDS2018 (<https://www.unb.ca/cic/datasets/index.html>)
- ▶ Duomenų rinkinius sudaro nuasmeninti tinklo srautų duomenys PCAP ir .csv formate. Lentelių formatu pateikiami ir CIC-Flowmeter priemone parengti įvairūs išvestiniai agregatai ir atakos tipas.
- ▶ Daugiau informacijos: Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

# LITNET 2020 duomenų rinkinys

- ▶ Lietuvos akademinio tinklo LITNET duomenų rinkinys LITNET-2020 (<https://dataset.litnet.lt/data.php>)
- ▶ Duomenų rinkinius sudaro nuasmeninti tinklo srautų duomenys NetFlow v9 formatu (rfc3954) ir duomenys lentelės .csv formatu, papildyti išvestiniais požymiais ir atakos tipu.
- ▶ Daugiau informacijos: Damasevicius, Robertas, 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 (Switzerland)* 9, no. 5 (2020). <https://doi.org/10.3390/electronics9050800>.

# LITNET-2020 rinkinio požymiai

- ▶ ts,te,td laiko įrašai, tokie kaip: srauto pradžia ir pabaiga, trukmė
- ▶ sp,dp srauto šaltinio ar gavėjo prievadai
- ▶ pr srauto protokolas (tcp, udp, rdp, fcp ir kiti)
- ▶ flg TCP ženklai (ACK, SYN, FIN, Reset, Push, Urgent)
- ▶ ipkt,ibyt gaunamų duomenų kiekis (paketais, baitais)
- ▶ opkt,obyt siunčiamų duomenų kiekis (paketais, baitais)
- ▶ in,out įvesties ar išvesties sąsajos SNMP numeris
- ▶ sas,das šaltinio ar gavėjo autonominės sistemos (AS) numeris
- ▶ dtos,stos šaltinio ar gavėjo paslaugos tipas

# CIC-IDS2017 turinys

Klasė	Įrašų skaičius	Dalis		Klasė	Įrašų skaičius	Dalis
Benign	2'273'097	80,3%		DoS Slowhttptest	5'499	0,2%
DoS Hulk	231'073	8,1%		Bot	1'966	0,07%
PortScan	158'930	5,6%		Web Attack - Brute Force	1'507	0,05%
DDoS	128'027	4,5%		Web Attack - XSS	652	0,02%
DoS GoldenEye	10'293	0,2%		Infiltration	36	0,001%
FTP-Patator	7'938	0,2%		Web Attack - Sql Injection	21	0,0007%
SSH-Patator	5'897	0,2%		Heartbleed	11	0,0004%
DoS Slowloris	5'796	0,2%				
					Viso: ~3 mln.	



# CSE-CIC-IDS2018 turinys

Klasė	Įrašų skaičius	Dalis	Klasė	Įrašų skaičius	Dalis
Benign	13'484'708	83.1%	DoS attacks-SlowHTTPTest	139'890	0.86%
DDOS attack-HOIC	686'012	4.2%	DoS attacks-GoldenEye	41'508	0.26%
DDoS attacks-LOIC-HTTP	576'191	3.5%	DoS attacks-Slowloris	10'990	0.068%
DoS attacks-Hulk	461'912	2.8%	DDOS attack-LOIC-UDP	1'730	0.011%
Bot	286'191	1.7%	Brute Force -Web	611	0.0038%
FTP-BruteForce	193'360	1.2%	Brute Force -XSS	230	0.0014%
SSH-Bruteforce	187'589	1.1%	SQL Injection	87	0.0005%
Infiltration	161'934	1.0%			
			Viso:	~16 mln.	

# LITNET-2020 turinys

Klasė	Įrašų skaičius	Dalis	Klasė	Įrašų skaičius	Dalis
Benign	36'423'860	91.971%	ICMP Flood	23'256	0.059%
SYN Flood	1'580'016	3.990%	HTTP Flood	22'959	0.058%
Code Red	1'255'702	3.171%	Scan	6'232	0.016%
Smurf	118'958	0.300%	Reaper Worm	1'176	0.003%
UDP Flood	93'583	0.236%	Spam	747	0.002%
LAND DoS	52'417	0.132%	Fragmentation	477	0.001%
W32.Blaster	24'291	0.061%			
			Viso:	~39 mln.	

# Pristatymo turinys

- ▶ Trumpai apie nesubalansuotų klasių problematiką ir tyrimui pasirinktus duomenų rinkinius
- ▶ Nesubalansuotų duomenų rinkinių paruošimo mašininiam mokymui galimybės
- ▶ Klasifikavimas klaidoms jautriais metodais
- ▶ Duomenų klasių disbalansui atskleisti tinkami rodikliai
- ▶ Tyrimo rezultatų aptarimas

# Nesubalansuotų klasių problema

- Vadinkime duomenų rinkinius ženkliai nesubalansuotais, kai rinkinyje yra klasių, kurių įrašų dažnis yra daugiau nei tūkstantį kartų mažesnis, nei neigiamos klasės įrašų dažnis.

Imbalance category <sup>1</sup>	CIC-IDS2017	CSE-CIC-IDS2018	LITNET-2020
Modest <(10:1)	8.16%	0.00%	0.00%
High <(1000:1)	11.39%	16.85%	7.83%
Extreme >(1000:1)	0.15%	0.08%	0.20%
Total Malignant	19.7%	16.9%	8.0%

<sup>1</sup> Share of records in imbalance category

# Retų klasių pavyzdžiai

CIC-IDS2017		CIC-IDS2018		LITNET-2020	
Class <sup>1</sup>	Share	Class	Share	Class	Share
Bot	0.0695%	DoS-Slowloris	0.0677%	W32.Blaster	0.0660%
Brute Force-Web	0.0532%	LOIC-UDP <sup>2</sup>	0.0107%	ICMP Flood	0.0638%
Brute Force-XSS	0.0230%	Brute Force-Web	0.0038%	HTTP Flood	0.0630%
Infiltration	0.0013%	Brute Force-XSS	0.0014%	Scan	0.0170%
SQL Injection	0.0007%	SQL Injection	0.0005%	Reaper Worm	0.0032%
Heartbleed	0.0004%			Spam	0.0021%
				Fragmentation	0.0013%
Total Extreme >(1000 : 1)	0.15%		0.08%		0.20%

<sup>1</sup> See Table 3 for explanation. <sup>2</sup> DDOS attack.

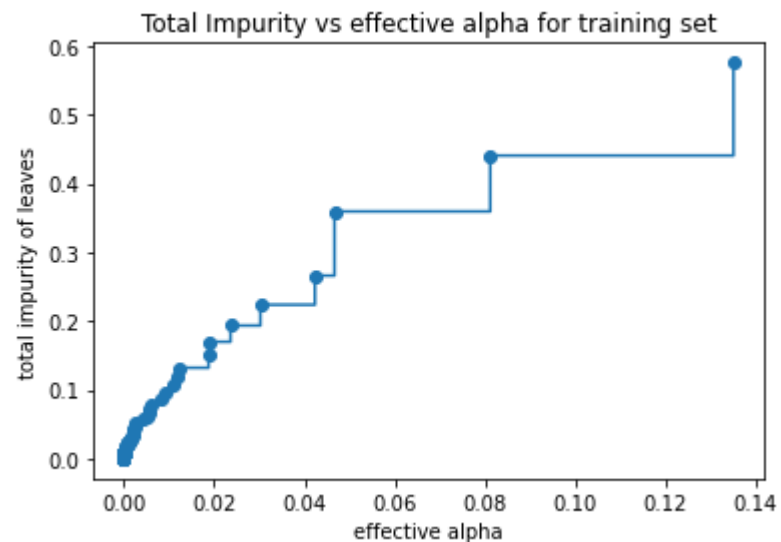
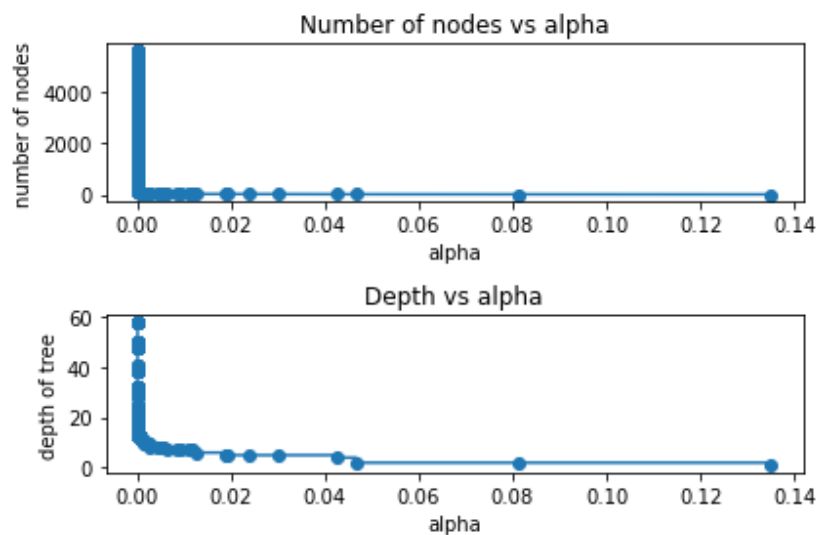
# Duomenų paruošimas

- ▶ Atrenkant įrašus algoritmų mokymui:
  - ▶ Paskaičiuojant mokymosi kreives įvertintas duomenų kiekis, pakankamas visiems panaudotiems algoritmams apmokyti
  - ▶ Atsitiktiniu būdu mažinant iki parinkto kiekio sumažintas dominuojančių klasių įrašų kiekis
  - ▶ Piktybiniai įrašai padalinti į dvi dalis
  - ▶ Sudaryti du ~0.6 mil. įrašų dydžio rinkiniai, mokymui ir testavimui
  - ▶ Kategoriniai duomenys perkoduoti
- ▶ SMOTE (dirbtinė pavyzdžių sukūrimo technika) pagalba retesnėms nei 1:1000 piktybinėms klasėms įterpiant dirbtinius įrašus sudarytas antras testavimo rinkinys

# Medžių ansamblių mokymas

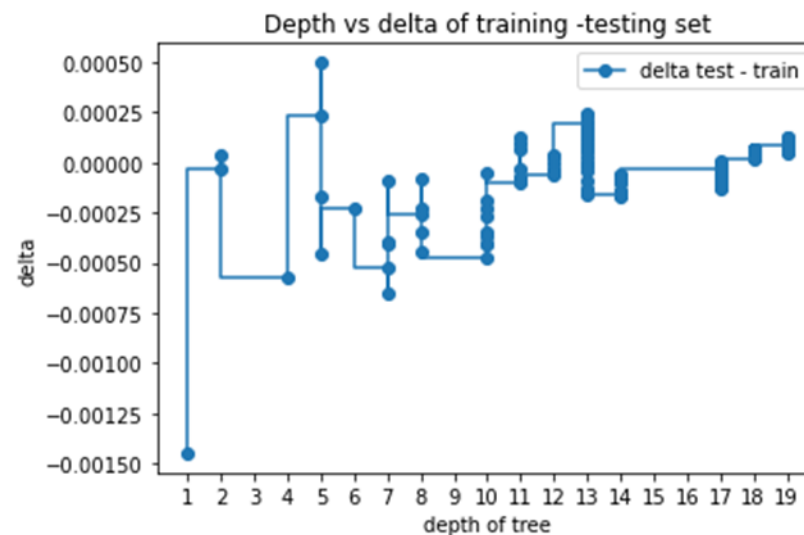
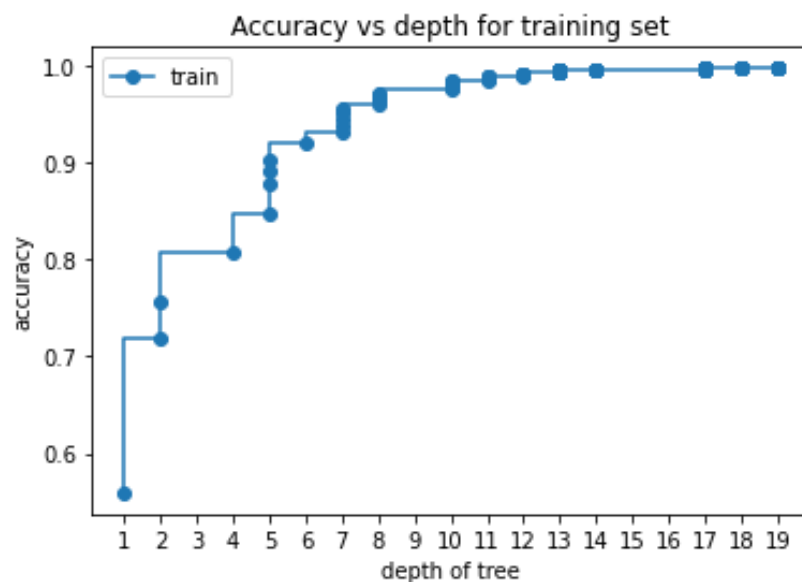
- ▶ Medžių mokymui buvo atliktas kelių lygių parametrų parinkimas
- ▶ Pirmas parinkimo žingsnis – medžio gylio ir lapų skaičiaus, tyrumo ir šakotumo parinkimas didžiausių kaštų kelio analizės metodu (Breimanas, J. Friedmanas, R. Olshenas ir C. Stone 1984), įgyvendintu scikit-learn `cost-complexity-pruning-path` funkcija
- ▶ Didėjant alfa genima daugiau. Taip sukuriamas geriau apibendrinantis sprendimų medis
- ▶ Vykdytame tyrime medžiams parinkti šie parametrai:  
`min_samples_leaf=8, max_features= 0.5, n_estimators=156, max_depth = 15, ccp_alpha = 0.00001, class_weight='balanced'`

# Medžio genėjimo parametrų parinkimas

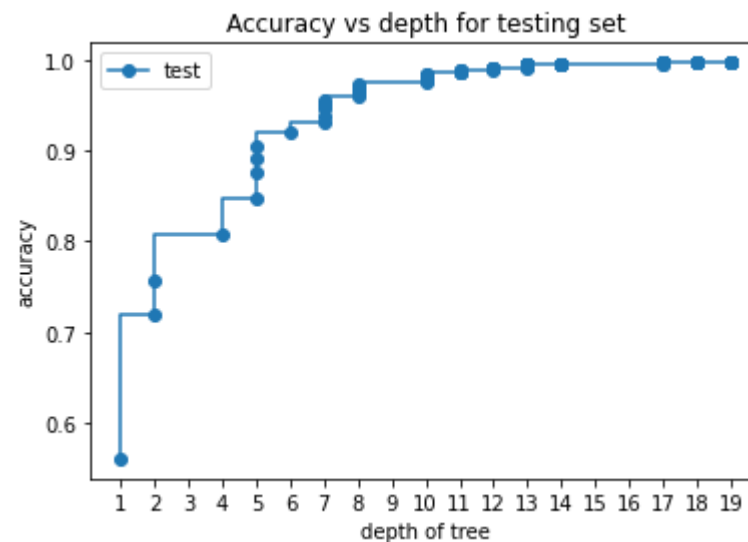
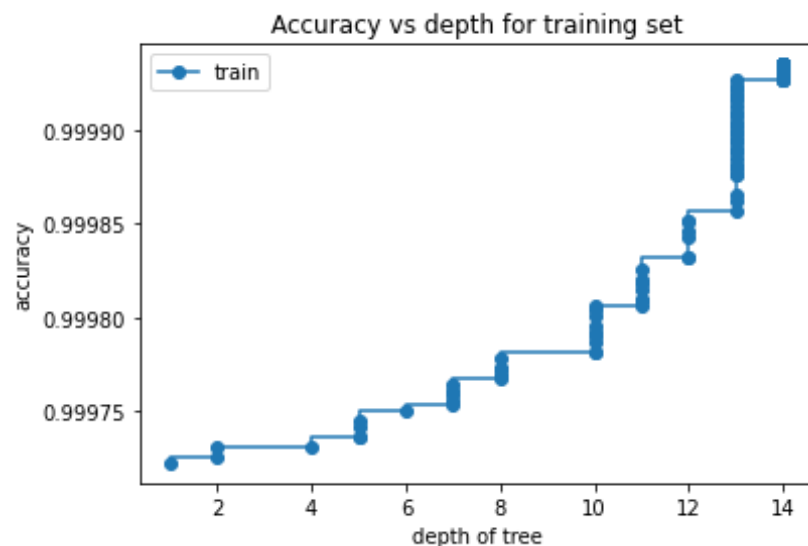




# Medžio genėjimo parametru parinkimas



# Medžio genėjimo parametų parinkimas



# SMOTE procesas

- ▶ SMOTE - Synthetic Minority Over-sampling Technique, Dirbtinė pavyzdžių sukūrimo technika, Chawla et al (2002).
  - ▶ Parenkame klases, kurių įrašų norėtume daugiau;
  - ▶ Parenkame svarbiausias skaitmenines savybes (naudojame koreliacijos analizę, Kbest, VIF ar kitą metodą);
  - ▶ Parenkame  $k$  artimiausių taškų;
  - ▶ Tarp kiekvienos taškų poros esantį vektorių dauginame iš atsitiktinio skaičiaus, esančio intervale  $]0;1[$ , kartojame, kol visos poros perrinktos;
  - ▶ Kartojame, kol pasiektas norimas taškų skaičius.

# Pristatymo turinys

- ▶ Trumpai apie nesubalansuotų klasių problematiką ir tyrimui pasirinktus duomenų rinkinius
- ▶ Nesubalansuotų duomenų rinkinių paruošimo mašininiam mokymui galimybės
- ▶ **Klasifikavimas klaidoms jautriais metodais**
- ▶ Duomenų klasių disbalansui atskleisti tinkami rodikliai
- ▶ Tyrimo rezultatų aptarimas

# KLASIFIKAVIMO KLAIDOMS JAUTRŪS METODAI

- ▶ Klasifikavimo klaidoms jautrių metodų (Cost-sensitive learning) esmė ta, kad esant klasių disbalansui, didžiųjų klasių atpažinimo klaidos nureikšminamos, o mažųjų klasių atpažinimo klaidos sureikšminamos.
- ▶ Populiariausias būdas tai padaryti yra keičiant mokymosi svorius.
- ▶ Euristinė taisyklė, patvirtinta sprendimų medžių logika, yra apversti svorius atvirkščiai proporcingai klasių proporcijoms rinkinyje. Tarkime, jei didžioji klasė atstovaujama santykiu 1000:1, tai reikia apversti proporciją ir vertinti tą vieną mažosios klasės įrašą kaip likusį tūkstantį.
- ▶ Sprendimų medžiuose algoritmas modifikuotas taip, kad gali atsižvelgti į klasės svorį, pasirenkant šakų dalinimą, papildomai sureikšminant mažąsias klases.
- ▶ scikit-learn bibliotekoje, jei tarp modelio parametrų matote parametą `class_weight`, nustatymas 'balanced' padės pasiekti norimo rezultato.

# Adaboost rezultatai LITNET-2020 (1 iš 2)

	pre	rec	f1	geo	iba	sup
http_f	0.905	1	0.95	0.999	0.998	11289
icmp_f	0.975	1	0.987	1	1	710
icmp_smf	0.999	1	1	1	1	7321
none	1	0.989	0.994	0.994	0.988	308395
smtp_b	0.989	0.997	0.993	0.999	0.997	373
tcp_land	0.991	0.999	0.995	1	0.999	6463
tcp_red_w	0.999	1	1	1	0.999	166154
tcp_syn_f	0.993	1	0.996	1	0.999	14437

# Adaboost rezultatai LITNET-2020 (2 iš 2)

	pre	rec	f1	geo	iba	sup
tcp_udp_win_p	1	0.999	1	1	0.999	3116
tcp_w32_w	1	0.99	0.995	0.995	0.989	100
udp_0	0.927	1	0.962	1	1	153
udp_f	0.962	0.999	0.98	0.998	0.996	45859
udp_reaper_w	0.957	0.985	0.971	0.992	0.983	587
avg / total	0.994	0.994	0.994	0.997	0.993	564957

# Pristatymo turinys

- ▶ Trumpai apie nesubalansuotų klasių problematiką ir tyrimui pasirinktus duomenų rinkinius
- ▶ Nesubalansuotų duomenų rinkinių paruošimo mašininiam mokymui galimybės
- ▶ Klasifikavimas klaidoms jautriais metodais
- ▶ Duomenų klasių disbalansui atskleisti tinkami rodikliai
- ▶ Tyrimo rezultatų aptarimas



# Pasirinkti efektyvumo rodikliai

- ▶ Šiame tyrime algoritmų efektyvumo palyginimui buvo pasirinkti subalansuoto tikslumo matas (balanced accuracy score) iš scikit-learn bibliotekos (Pedregosa ir kt. 2011) ir jautrumo geometrinis vidurkis (G-mean) iš nesubalansuoto mokymosi bibliotekos imbalanced-learn (G. Lemaitre, F. Nogueira, D. Oliveira, C. Aridas 2017).
- ▶ Rodikliai pasirinkti remiantis literatūros apžvalga ir dėl jų paprastumo ir tinkamumo klasių disbalansui kibernetinio saugumo srityje palyginimui.
- ▶ Klasikinis tikslumas priklauso nuo algoritmo pasiektų rezultatų didžiausiose klasėse ir atspindi jų rezultatus.

# Subalansuotas tikslumo rodiklis

- ▶ Subalansuotas tikslumo rodiklis vertinant kiekvienai klasei  $i \in k$  kiekvieną prognozę ar ji tikrai priklauso prognozuojamai klasei (TP), ar priklauso kitai klasei (TN), neteisingai priskirta klasei (FP), arba neteisingai nepriskirta klasei (FN):

$$\text{BAS} = \frac{\sum_{i=1}^k \frac{\text{jautrumas}_i + \text{selektyvumas}_i}{2}}{k},$$

kur

$$\text{jautrumas}_i = \frac{TP_i}{TP_i + \sum_k FN_{k,i}} \quad \text{selektyvumas}_i = \frac{TN_i}{TN_i + \sum_k FP_{k,i}}$$

- ▶ Atkūrimas, jautrumas = Recall, Sensitivity; Selektyvumas, specifiškumas – specificity, selectivity

# Metrikų paaiškinimas

- ▶ TP – True Positive – teisingai numatomų įrašų skaičius,
- ▶ FP – False Positive – neteisingai numatomų teisingų įrašų skaičius,
- ▶ (Precision, Pr) tikslumas, teisingai numatomoji vertė: Tai yra teisingai klasifikuotų atakų srautų (TP) santykis su visais klasifikuotais srautais (TP + FP).
- ▶ (Recall, Rc) arba jautrumas: Tai yra teisingai klasifikuotų išpuolių srautų (TP) santykis su visais sugeneruotais srautais (TP + FN).
- ▶  $F_1$  matas ( $F_1$ ): tai harmoninis tikslumo ir jautrumo vidurkis.

# Metrikų paaiškinimas

$$F_{1i} = \frac{2}{\frac{1}{Pr_i} + \frac{1}{Rc_i}}$$

$$Pr_i = \frac{TP_i}{TP_i + \sum_{i=1}^k FP_{i,k}} \quad Rc_i = \frac{TP_i}{TP_i + \sum_{i=1}^k FN_{i,k}}$$

# Geometrinis jautrumo vidurkis

- ▶ Geometrinis jautrumo vidurkis (G-mean) yra k-ojo laipsnio šaknis iš atskirų klasių jautrumų sandaugos:

$$\bar{G} = \sqrt[k]{\prod_{i=1}^k \text{jautrumas}_i}$$

- ▶ Vertinant šiuo rodikliu, blogas modelio jautrumas bet kurios klasės atžvilgiu lemia blogą bendrą rezultatą.

# MLP retųjų klasių rezultatai prieš ir po SMOTE

Klasė	n_pre	s_pre	n_geo	s_geo	n_iba	s_iba	sup
udp_reaper_w	0	0.778	0	0.972	0	0.94	587
smtp_b	0.631	0.912	0.766	0.988	0.563	0.973	373
udp_0	0.412	0.341	0.667	0.718	0.42	0.491	153
tcp_w32_w	0	0.285	0	0.969	0	0.934	100
avg / total	0.955	0.98	0.964	0.986	0.932	0.971	564957

Pastaba: čia n – netaikytas SMOTE, s – su SMOTE duomenimis

# BAS rodiklis CIC-IDS2017 rinkmenai

---

Modelis	Be SMOTE	Su SMOTE
Adaboost (ADA)	89.70%	96.88%
Decision Tree Classifier (CART)	88.67%	96.65%
Gradient Boosting Classifier (GBC)	81.10%	95.90%
K-Nearest Neighbors (KNN)	91.70%	96.20%
Multi Layer Perceptron (MLP)	86.00%	90.57%
Random Forest Classifier (RFC)	89.80%	96.98%

---

# Rodikliai LITNET-2020 rinkmenai

---

G-mean	Be SMOTE	Su SMOTE
K-Nearest Neighbors (KNN)	65.14%	83.77%
Multi Layer Perceptron (MLP)	44.59%	97.54%

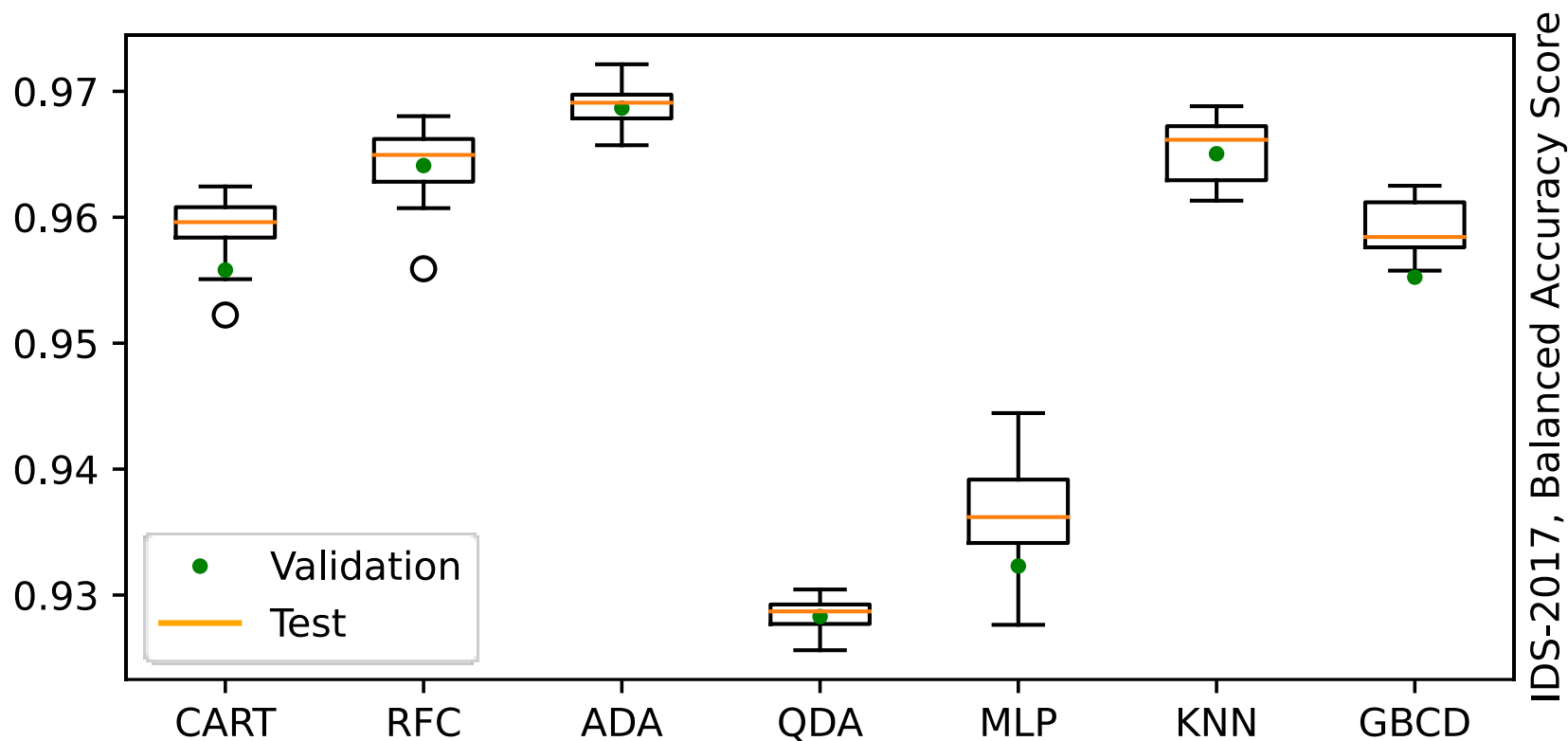
---

BAS	Be SMOTE	Su SMOTE
K-Nearest Neighbors (KNN)	80.64%	87.99%
Multi Layer Perceptron (MLP)	77.17%	97.63%

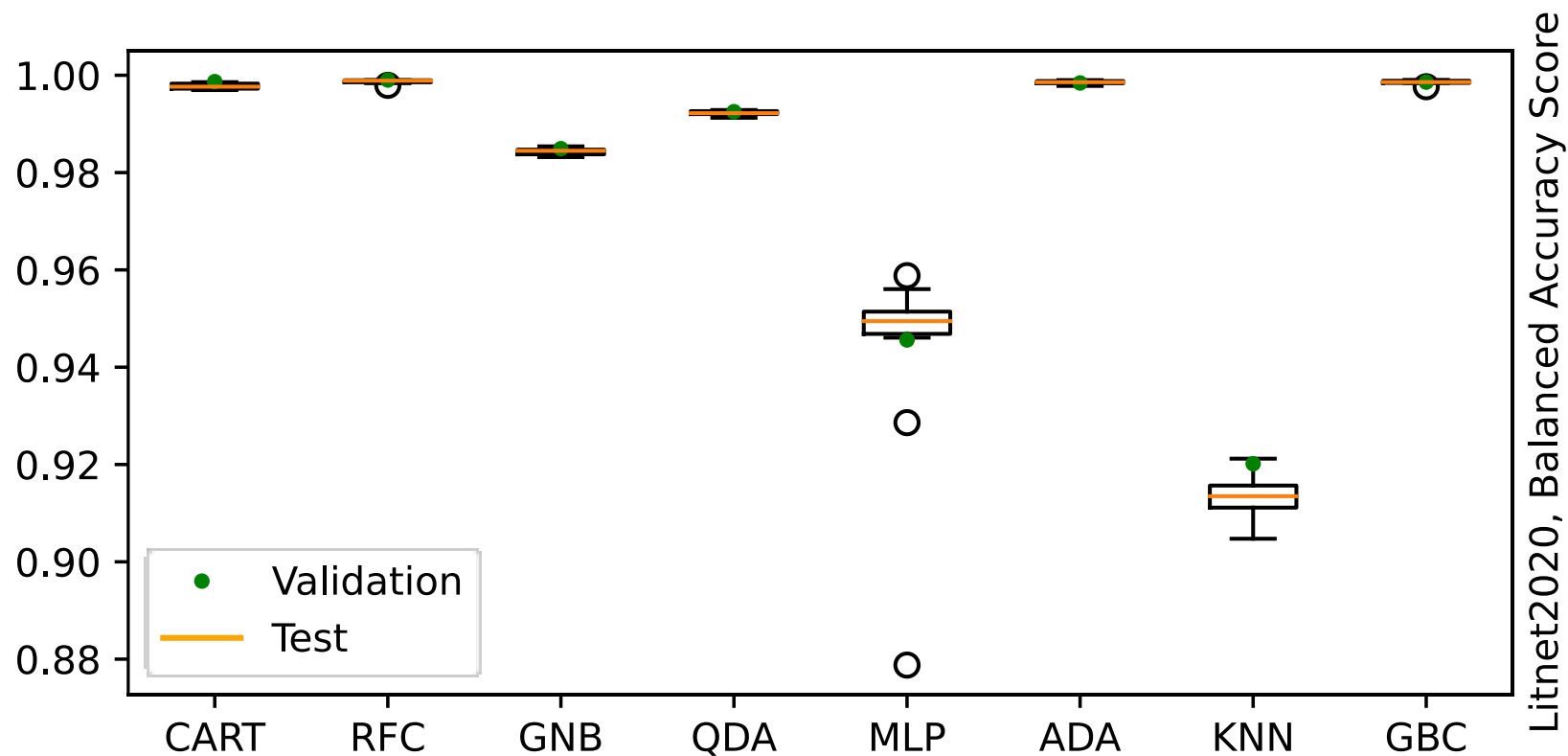
---



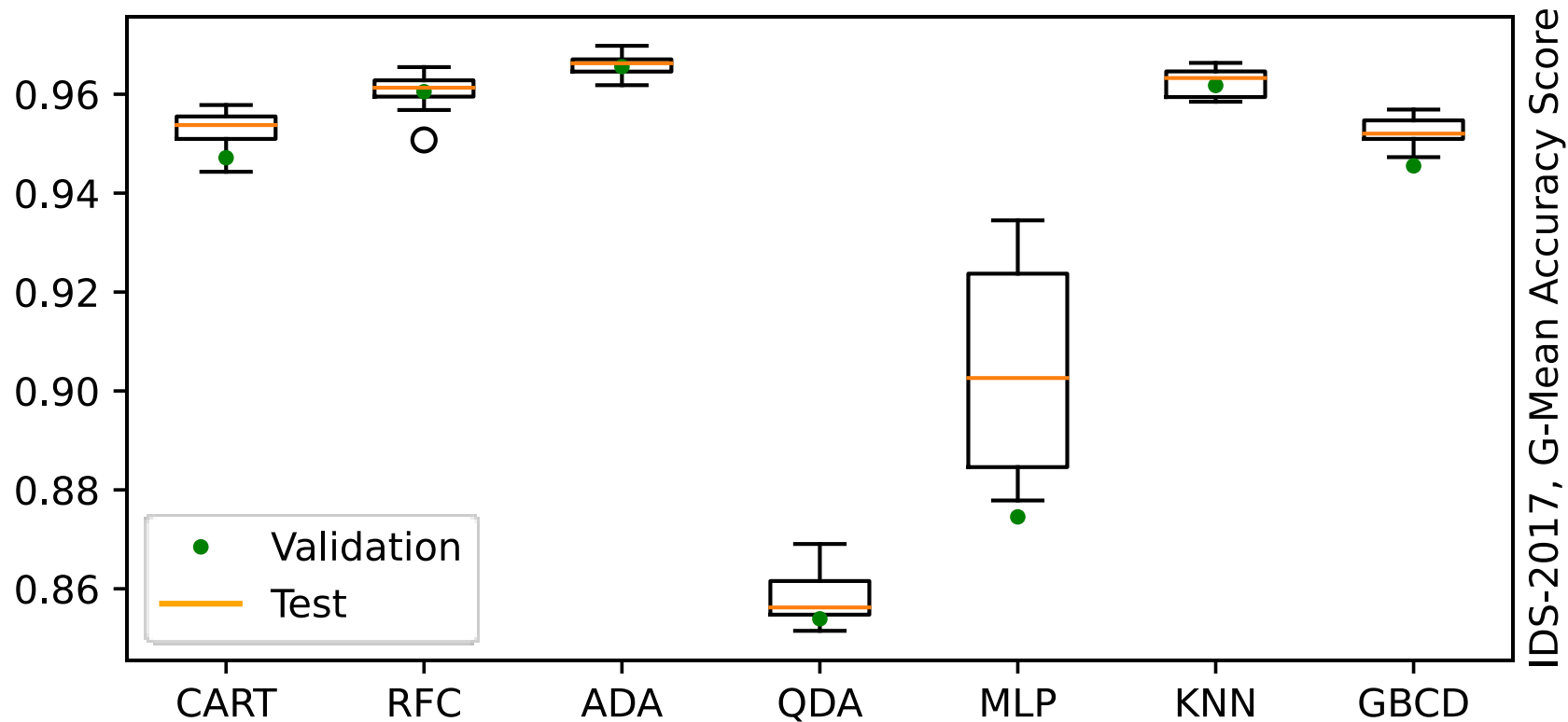
# Balanced Accuracy Score for IDS2017



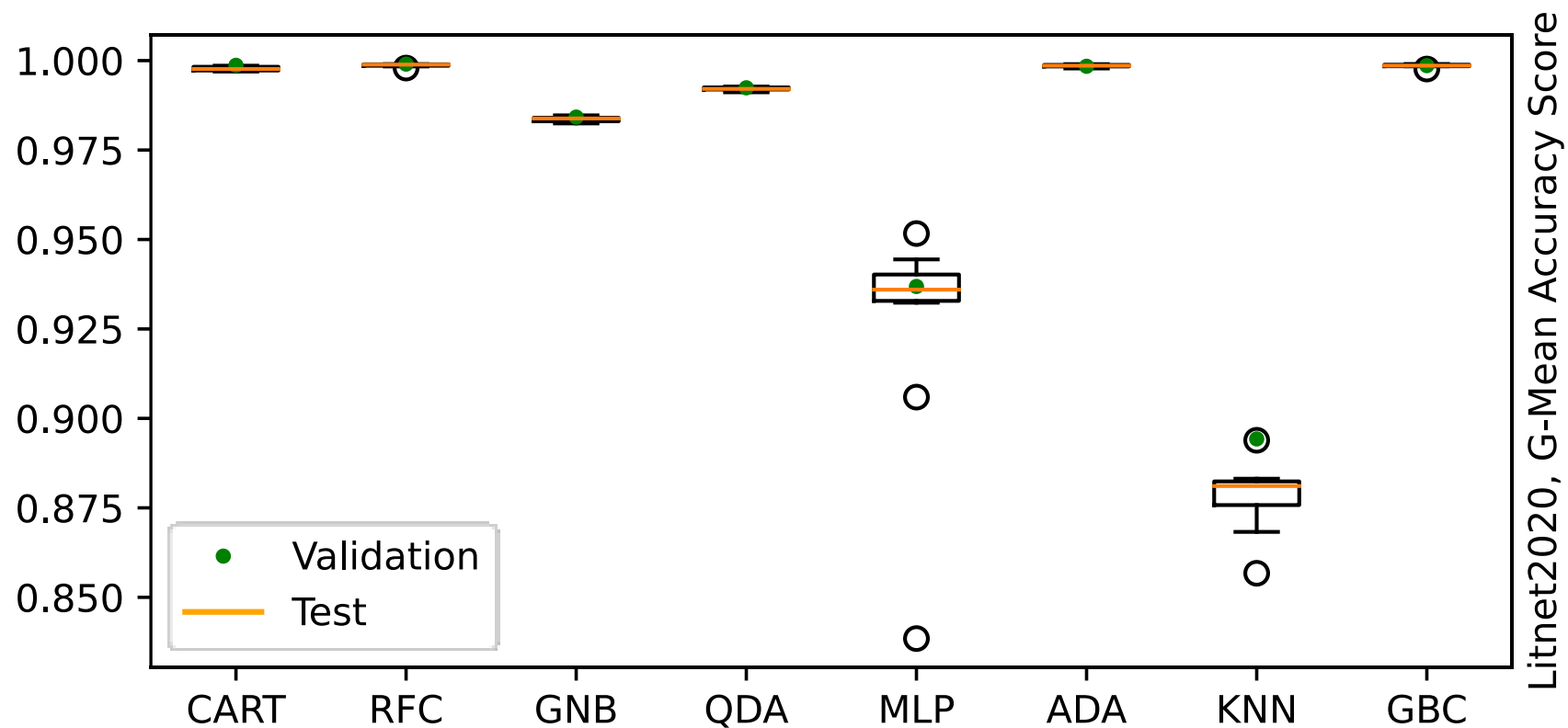
# Balanced Accuracy Score for LITNET-2020



# G-Mean Accuracy Score for IDS-2017



# G-mean LITNET-2020



# Vidutinės kvadratinės prognozės paklaidos išskaidymas į tendencingumą ir dispersiją

BAS	CIC-IDS2017			CIC-IDS2018			LITNET-2020			Rank	
Model	Mean	Std	Rank	Mean	Std	Rank	Mean	Std	Rank	Total	Best
ADA <sup>1</sup>	0.999	0.0002	1	0.940	0.0007	2	0.997	0.0002	1	4	1
CART	0.996	0.0008	4	0.936	0.0009	5	0.995	0.0010	2	11	3
GBC	0.997	0.0009	3	0.937	0.0005	4	0.989	0.0029	4	11	3
GNB	0.744	0.0670	8	0.845	0.0035	8	0.958	0.0007	6	22	8
KNN	0.994	0.0001	5	0.940	0.0006	3	0.956	0.0005	7	15	5
MLP	0.980	0.0009	6	0.928	0.0012	6	0.930	0.0217	8	20	7
QDA	0.932	0.0003	7	0.910	0.0059	7	0.978	0.0005	5	19	6
RFC	0.998	0.0001	2	0.941	0.0007	1	0.995	0.0002	3	6	2

# Rezultatai

- ▶ Dirbtinių duomenų įterpimas ir klasifikavimo klaidoms jautraus metodo taikymas padėjo parinktų modelių mokymosi procesui aptikti retas klases, kas patvirtinta testavimu su visiškai nematytais duomenimis.
- ▶ Sprendimų medžiais ir jų ansambliais (CART, ADABOOST, RandomForest, Gradient Boosting Classifier) pavyko pasiekti geresnių rezultatų.
- ▶ MLP reikės įdiegti Keras priemonėmis, nes toje bibliotekoje įgyvendintas klaidoms jautrus metodas.
- ▶ Norint rasti tinkamą šiuolaikinį neuroninių tinklų modelį, reikia atlikti tolesnius tyrimus, tačiau sunku tikėtis, kad jis bus greitesnis nei sprendimų medžiai.

# Laikas jūsų klausimams

Kontaktams:

[viktoras.bulavas@mii.stud.vu.lt](mailto:viktoras.bulavas@mii.stud.vu.lt)



# Ketvirtųjų mokslo metų darbo planas

5. Atskirų daktaro disertacijos dalių (analizės rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas (2020 m. spalį– 2021 m. gegužė):

5.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas.

5.2. Analitinės disertacijos dalies parengimas.

5.3. Teorinės disertacijos dalies parengimas.

5.4. Eksperimentinės disertacijos dalies parengimas.

5.5. Bendrųjų išvadų formulavimas.

6. Daktaro disertacijos parengimas ir svarstymas padalinyje (2021 m. birželis).

7. Daktaro disertacijos pateikimas gynimui (2021 m. rugsėjis).

- ▶ Planuojama parengti dvi mokslines tyrimų publikacijas (recenzuojamuose leidiniuose, WoS su Impact Factor).
- ▶ Erasmus+ (savanoriška) praktika Gdanskio technologijų universiteto Elektronikos, telekomunikacijų ir informatikos fakultete, 2019 rugsėjo 25 – 2020 sausio 31 d.



AČIŪ UŽ DĒMESĪ!

Viktoras Bulavas

E-mail: viktoras.Bulavas [eta] stud.mii.vu.lt

► Laikas klausimams!