



**Vilnius
universitetas**



Duomenų mokslo ir skaitmeninių Technologijų institutas

Informatikos inžinerijos krypties doktorantų konferencija

Veiklos ataskaita už 2021 m. gruodžio 1d. – 2022 m. kovo 24d.

Dalia BRESKUVIENĖ – Informatikos inžinerija T 007 doktorantė

Darbo vadovas – prof. habil. dr. Gintautas DZEMYDA

Doktorantūros pradžios ir pabaigos metai: 2021 - 2025

2021–2022m.

Vilniaus
universitetas

STUDIJŲ PLANAS IR JO VYKDYMO SUVESTINĖ

Studijų metai	Egzaminai ¹		Dalyvavimas konferencijose ²		Publikacijos ³		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė ⁴
I (2021/2022)	3	1	1	1			
II (2022/2023)	1		1		1		
III (2023/2024)			1		1		
IV (2024/2025)					1		
Iš viso:	4	1	3		3		

2021–2022m.

Vilniaus
universitetas

ATASKAITINIŲ METŲ DARBO PLANAS IR JO SUVESTINĖ

Egzaminai		
Planas	Įvykdyta	Būklė
Mašininis mokymasis	Mašininis mokymasis. 2022.02.17	Išlaikyta. Įvertinimas 10

Dalyvavimas konferencijose		
Planas	Įvykdyta	
Mokslinė konferencija „12th Conference: Data Analysis Methods for Software Systems“, 2021.12.02, Druskininkai	„ <i>Forbearance prediction using XGBoost and LightGBM Models</i> “, DAMSS21, 2021.12.02, Druskininkai	Nacionalinė konferencija

Mokslinių tyrimų ir disertacijos rengimo planas:

Vilniaus
universitetas

	Darbo pavadinimas	Atlikimo terminai
1.	<u>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</u>	
	1.1. Disertacijos tyrimo objekto detalizavimas.	2021 m. gruodžio mėn. – 2022 m. vasario mėn.
	1.2. Atlikti būdų klasifikatorių veikimo optimizavimui analitinę apžvalgą.	2021 m. gruodžio mėn. – 2022 m. spalio mėn.
	1.3. Nustatyti (identifikuoti) mokslines problemas, kylančias uždaviniuose, susijusiuose su klasifikavimo kokybės optimizavimu, o taip pat ir naudojant giliuosius neuroninius tinklus.	2022 m. kovo mėn. – 2022 m. spalio mėn.
	1.4. Tyrimo tikslo suformavimas.	2022 m. kovo mėn. – 2022 m. spalio mėn.
2.	<u>Mokslinio tyrimo vykdymas:</u>	
	2.1. Tyrimo metodikos sudarymas:	
	2.1.1. Tyrimo metodikos išskeltiems uždaviniams spręsti parinkimas;	2022 m. kovo mėn. – 2022 m. spalio mėn.
	2.1.2. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	2022 m. kovo mėn. – 2022 m. spalio mėn.
	2.2. Teorinis tyrimas:	
	2.2.1. Klasifikatorių efektyvumo galimybių tyrimas optimizuojant mokymo aibės taškų parinkimą.	2022 m. lapkričio mėn. – 2023 m. spalio mėn.
2.2.2. Giliųjų neuroninių tinklų panaudojimo galimybių optimaliai mokymo aibei rasti tyrimas.	2022 m. lapkričio mėn. – 2023 m. spalio mėn.	
2.3. Empirinis tyrimas:		
2.3.1. Sudarytų metodų pritaikymas praktinių uždavinių sprendimui.	2024 m. kovo mėn. – 2024 m. spalio mėn.	
2.3.2. Gautų duomenų analizė, rezultatų apibendrinimas, išvadų parengimas.	2024 m. spalio mėn. – 2025 m. vasario mėn.	
3.	<u>Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:</u>	
	3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas;	2024 m. spalio mėn. – 2025 m. vasario mėn.
	3.2. Analitinės disertacijos dalies parengimas;	2024 m. kovo mėn. – 2025 m. rugpjūčio mėn.
	3.3. Teorinės disertacijos dalies parengimas;	2024 m. kovo mėn. – 2025 m. rugpjūčio mėn.
	3.4. Eksperimentinės disertacijos dalies parengimas;	2024 m. kovo mėn. – 2025 m. rugpjūčio mėn.
3.5. Bendrųjų išvadų formulavimas.	2024 m. kovo mėn. – 2025 m. rugpjūčio mėn.	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2025 m. rugsėjo mėn.
5.	Daktaro disertacijos gynimas	2025 m. lapkričio mėn.

Disertacijos tema, tyrimo objektai ir tikslas

Preliminari disertacijos tema:

Klasifikatoriaus (nesubalansuotos) mokymo aibės optimizavimas, siekiant geresnės klasifikavimo kokybės.

Tyrimo objektai:

Finansinio sukčiavimo duomenų optimizavimas, siekiant tikslesnių mašininio mokymosi metodų rezultatų.

Tikslas:

Sukurti arba patobulinti nesubalansuotų duomenų optimizavimo metodą, skirtą anomalinių įvykių identifikavimui ir jų prevencijai.

Tyrimo uždaviniai

Vilniaus
universitetas

Identifikuoti tinkamus metodus nesubalansuotos mokymo aibės optimizavimui;

Identifikuoti aktualias mokslines problemas, kylančias uždaviniuose, susijusiuose su finansinio sukčiavimo aptikimu;

Sukurti arba patobulinti algoritmą nesubalansuotos mokymo aibės optimizavimui atsižvelgiant į naujo taško klasifikavimą;

Pritaikyti sukurtą arba patobulintą metodą realaus laiko duomenims ir atlikti gautų duomenų analizę, rezultatų apibendrinimą, išvadų parengimą.

Literatūros apžvalga

Nesubalansuotos mokymų aibės problemos mašininio mokymosi kontekste:



- Standartiniai mašininio mokymosi algoritmai yra dažniausiai nepritaikyti nesubalansuotoms duomenų aibėms [1];
- Nesubalansuotos duomenų aibės dydis dažniausiai kelia papildomų problemų [2];
- Triukšmas duomenyse sukvalifikuojamas kaip viena iš klasių [2]

Dažniausiai siūlomi sprendimo būdai:

- Didesnės klasės narių išretinimas mokymų aibėje (undersampling)[3]
- Mažesnės klasės narių generavimas/kopijavimas (oversampling) [4][5]

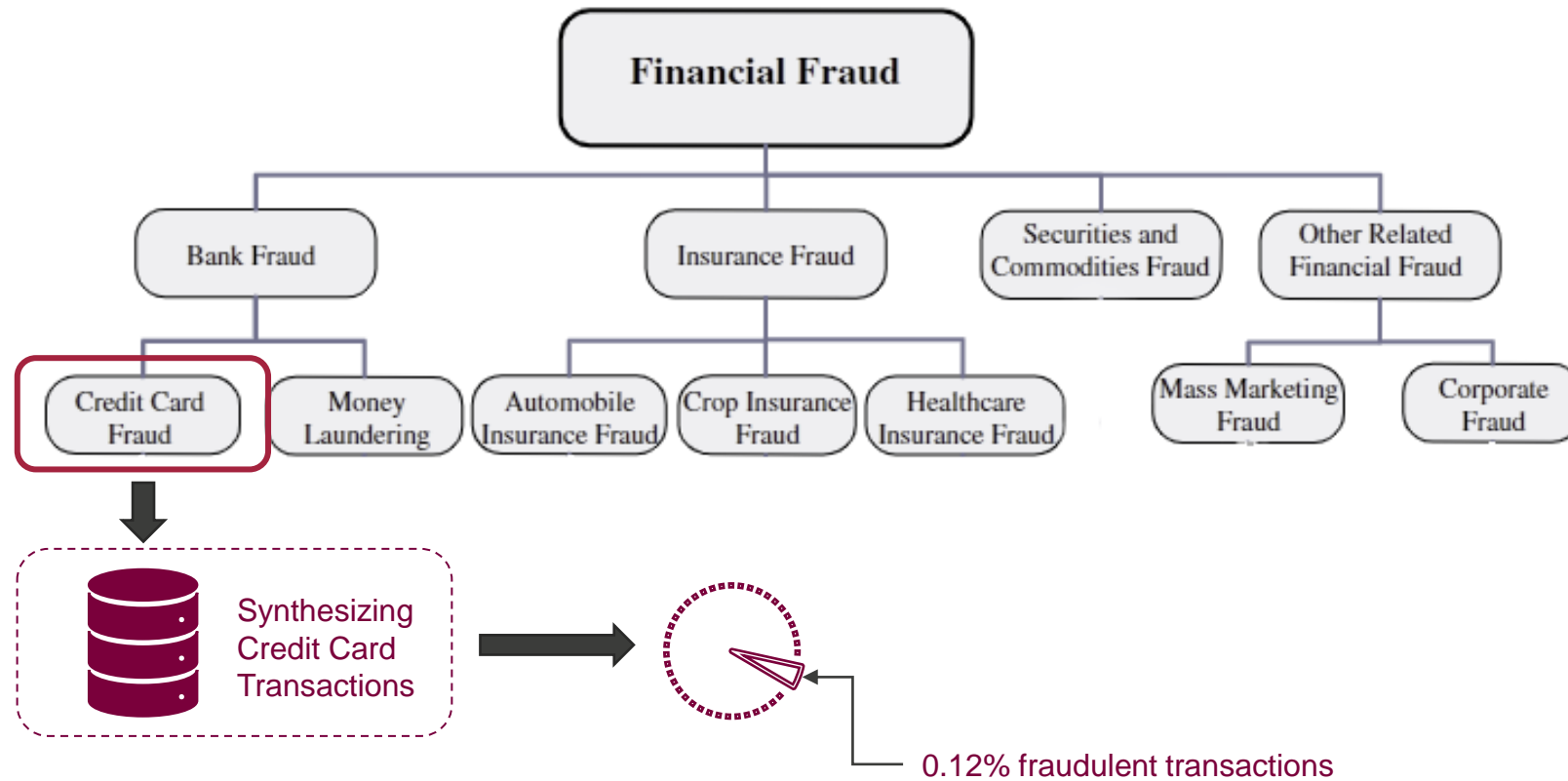


Literatūros šaltiniai

- 
- [1] F. Provost, "Machine learning from imbalanced data sets 101," *Proceedings of the AAAI'2000 Workshop on ...*, 2000.
- [2] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from Imbalanced Data Sets. 2018. doi: 10.1007/978-3-319-98074-4.
- [3] P. Kaur and A. Gosain, "Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise," in *Advances in Intelligent Systems and Computing*, 2018, vol. 653. doi: 10.1007/978-981-10-6602-3_3
- [4] S. H. Park and Y. G. Ha, "Large imbalance data classification based on MapReduce for traffic accident prediction," 2014. doi: 10.1109/IMIS.2014.6.
- [5]. J. Zhai, J. Qi, and C. Shen, "Binary imbalanced data classification based on diversity oversampling by generative models," *Information Sciences*, vol. 585, 2022, doi: 10.1016/j.ins.2021.11.058
- 

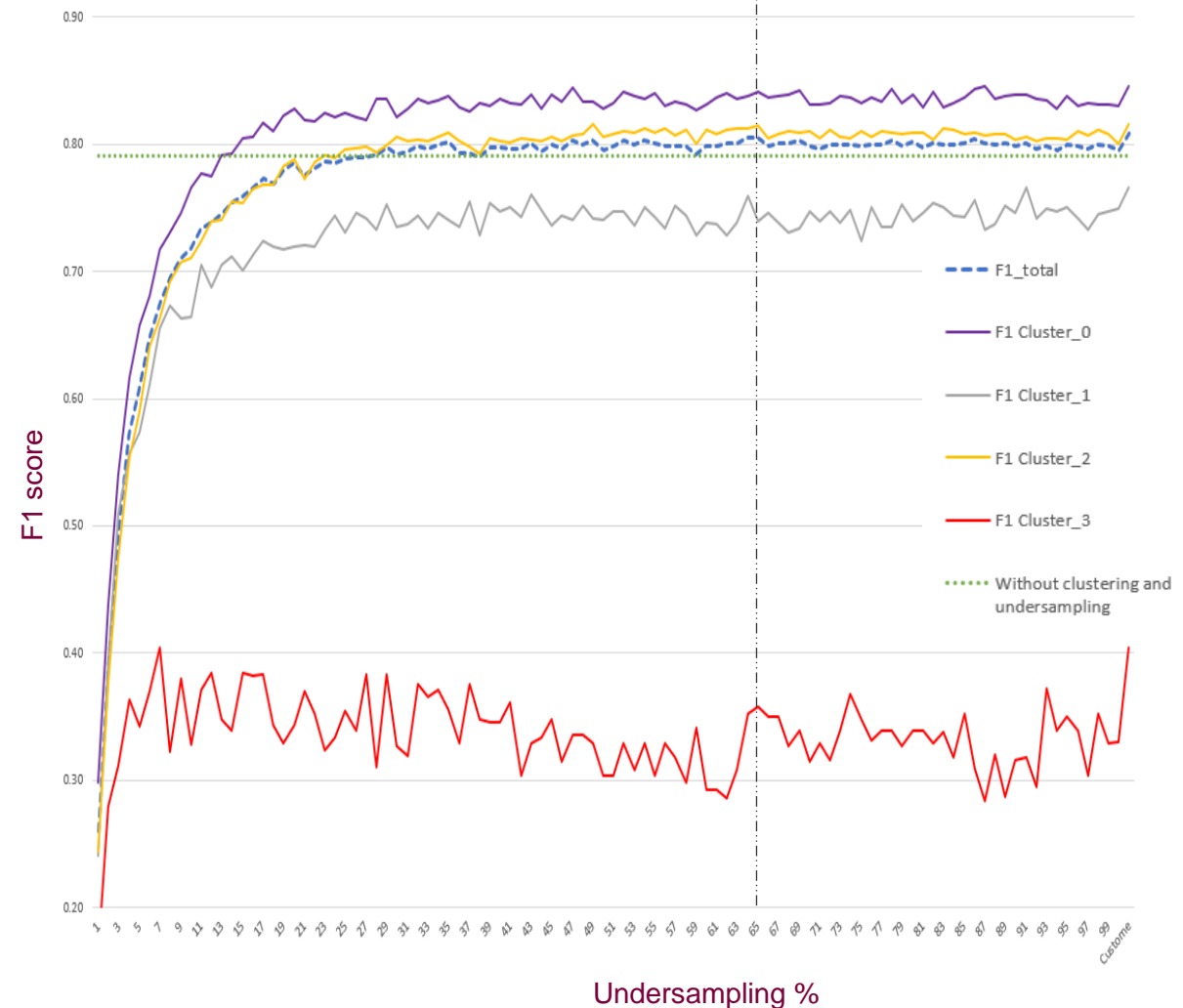
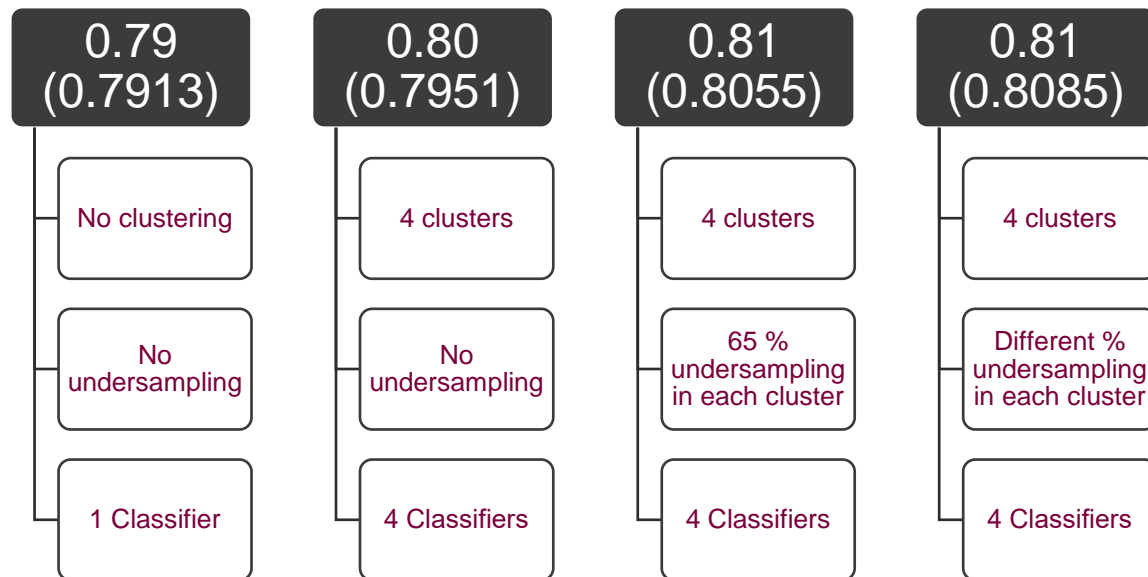
Eksperimentinių tyrimų rezultatai

Vilniaus
universitetas



Eksperimentinių tyrimų rezultatai

Vilniaus
universitetas



KITO PUSMEČIO DARBO PLANAS

Vilniaus
universitetas

Egzaminai:

- Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika
- Netiesiniai statistikos modeliai masinių duomenų analizėje

Publikacijos:

- Publikacija recenzuojamame mokslo žurnale arba konferencijų medžiagoje

Mokslinių tyrimų apžvalga ir analizė:

- Nustatyti mokslines problemas, kylančias uždaviniuose, susijusiuose su klasifikavimo kokybės optimizavimu
- Tyrimo tikslo suformavimas.

Mokslinio tyrimo vykdymas:

- Tyrimo metodikos sudarymas:
- Tyrimo metodikos iškeltiems uždaviniams spręsti parinkimas;
- Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.



**Vilnius
universitetas**

Klausimai?